ASME
SETTING THE STANDARD

Journal of Engineering for Gas Turbines and Power

Published Bimonthly by ASME

VOLUME 130 • NUMBER 2 • MARCH 2008

Transactions of the ASME, Journal of Engineering for Gas Turbines and Power (ISSN 0742-4795) is published bimonthly (Jan., Mar., May, July, Sep, Nov.) by The American Society of Mechanical Engineers, Three Park Avenue, New York, NY 10016. Periodicals postage paid at New York, NY and additional mailing offices.
POSTMASTER: Send address changes to Transactions of the ASME, Journal of Engineering for Gas Turbines and Power, c/o THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 22 Law Drive, Box 2300, Fairfield, NJ 07007-2300.
CHANGES OF ADDRESS must be received at Society headquarters seven weeks before they are to be effective. Please send old label and new address.
STATEMENT from By-Laws. The Society shall not be responsible for statements or opinions advanced in papers or ... printed in its publications (B7.1, par. 3).

RESEARCH PAPERS

(Contents continued on inside back cover)

This journal is printed on acid-free paper, which exceeds the ANSI Z39.48-1992 specification for permanence of paper and library materials. ♾™
♻ 85% recycled content, including 10% post-consumer fibers.

(Contents continued)

Journal of Engineering for Gas Turbines and Power MARCH 2008

Volume 130, Number 2

(Contents continued on facing page)

Journal of Engineering for Gas Turbines and Power MARCH 2008

Volume 130, Number 2

TECHNICAL BRIEFS

DESIGN INNOVATION

The ASME Journal of Engineering for Gas Turbines and Power is abstracted and indexed in the following:

AESIS (Australia's Geoscience, Minerals, & Petroleum Database), Applied Science & Technology Index, Aquatic Sciences and Fisheries Abstracts, Civil Engineering Abstracts, Compendex (The electronic equivalent of Engineering Index), Computer & Information Systems Abstracts, Corrosion Abstracts, Current Contents, Engineered Materials Abstracts, Engineering Index, Enviroline (The electronic equivalent of Environment Abstracts), Environment Abstracts, Environmental Science and Pollution Management, Fluidex, INSPEC, Mechanical & Transportation Engineering Abstracts, Mechanical Engineering Abstracts, METADEX (The electronic equivalent of Metals Abstracts and Alloys Index), Pollution Abstracts, Referativnyi Zhurnal, Science Citation Index, SciSearch (The electronic equivalent of Science Citation Index), Shock and Vibration Digest

A Novel Multifunctional Energy System (MES) for CO_2 Removal With Zero Energy Penalty

Hongguang Jin
Institute of Engineering Thermophysics,
Chinese Academy of Sciences,
Beijing 100080, China
e-mail: hgjin@mail.etp.ac.cn

Wei Han
Institute of Engineering Thermophysics,
Chinese Academy of Sciences,
Beijing 100080, China;
Graduate School, Chinese Academy of Sciences,
Beijing 100049, China

Lin Gao
Institute of Engineering Thermophysics,
Chinese Academy of Sciences,
Beijing 100080, China

This paper proposes a novel, multifunctional energy system (MES), in which hydrogen and electricity are cogenerated and about 90% of CO_2 is removed. By integrating the methane/steam reforming reaction and combustion of coal, the natural gas and coal are utilized synthetically, and coal is burned to provide high-temperature thermal energy to the methane/steam reforming reaction. Afterwards, the resulting syngas enters a pressure swing adsorption (PSA) unit to separate about 70% of hydrogen, thereby significantly increasing the concentration of carbon dioxide from nearly 20% to 43% in the PSA tail gas. As a result, the overall efficiency of the new system becomes 63.2%. Compared to a conventional natural gas-based hydrogen plant and a coal-firing steam power plant without CO_2 removal (the overall efficiency of the two systems is 63.0%), the energy penalty for CO_2 removal in the new system is almost totally avoided. Based on the graphical exergy analysis, we propose that the integration of synthetic utilization of fossil fuel (natural gas and coal) and the CO_2 removal process plays a significant role in zero energy penalty for CO_2 removal and its liquefaction in the MES. The result obtained here provides a new approach for CO_2 removal with zero or low thermal efficiency reduction (energy penalty) within an energy system. [DOI: 10.1115/1.2799532]

Introduction

Carbon dioxide (CO_2) is the greenhouse gas that has the largest contribution from human activities and is released into the atmosphere mainly by combustion of fossil fuels, such as coal, oil, and natural gas (NG). Since fossil fuels are likely to remain as the main source of primary energy for many decades, we should pay special attention to finding ways to capture CO_2 [1]. The current viewpoint of system synthesis proposes several approaches to CO_2 sequestration: (1) The removal of CO_2 from the exhausts of conventional power stations; (2) the use of enriched CO_2 mixtures as working fluid with pure oxygen combustion; and (3) precombustion decarburization with coal gasification and NG/steam reforming. One of the challenges for these approaches is controlling the large thermal efficiency reduction (energy penalty) for the CO_2 separation and liquefaction.

The postcombustion approach has been considered a hot topic by many researchers [2–5]. Since the flue gas is at atmospheric pressure, a chemical absorption technique must first be considered. Thermal energy (steam extracted from steam turbine) is required for the regeneration of absorbent, leading to a decrease of turbine work output. The reduction of thermal efficiency is about eight percentage points for NG-based power plants [2,3], which mainly comes from the rather low mole concentration (about 3%) of CO_2 in the flue gas, although the CO_2 concentration can be increased to 6% with the exhaust recirculation. For coal-based power plants, the thermal efficiency reduction was 12–13 percentage points [4,5], which is mainly caused by the high carbon content of coal. The main efforts in this approach are mainly dependent on the improvement of techniques, i.e., chemical absorption technology.

In the oxy-fuel combustion approach, the enriched CO_2 mixtures substitute the air as a working fluid, in which the thermal efficiency reduction mainly relies on the efficiency of the air separation unit (ASU) instead of the CO_2 separation process. The efficiency reduction is about 7–11 percentage points for coal-based power plants [6,7] and about ten percentage points for NG-based power plants [5]. The key potential of this approach depends on both process technologies and system syntheses, including ways to decrease the energy consumption for generating pure oxygen and ways to effectively combine it with other systems. The precombustion approach is considered as a promising technique to recover CO_2 from coal-based power plants. Before combustion, the fossil fuel is converted to syngas (a mixture of CO and H_2) before CO is transformed into CO_2 through a water-gas shift reaction. Due to the high concentration of CO_2 and low quantity of gas to be treated, the energy consumption for CO_2 separation might be decreased. The thermal efficiency reduction is about seven percentage points for coal-based power plants [8] and 7–10 percentage points for NG-based power plants [9,10]. Extra energy consumption for the reforming or partial oxidation processes is the main reason for the high energy penalty [11]. A key point that should be emphasized is that precombustion provides a potential for the system synthesis to decrease the energy penalty of CO_2 removal. Therefore, the energy consumption is so large in the CO_2 removal process, such that it is the main energy penalty of the overall system. Nevertheless, although the improvement of the separation technologies may decrease the energy penalty, it is not a task that can be easily performed, even with the latest technologies.

Based on the exergy analysis, the exergy destruction of the conversion processes from fossil fuel to thermal energy (i.e., from chemical energy to thermal energy), such as combustion and partial oxidation fuel, is still the main contributor for the exergy destruction in the energy systems [9,10,12]. Moreover, the carbon composition of the fuel is also converted into carbon dioxide in these energy conversion processes. It is worthy to note that integrating the use of fossil fuel and CO_2 removal process may provide a feasible way to decrease the energy penalty of CO_2 removal.

This paper thus aims to (1) propose a new MES with a lower

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 8, 2007; final manuscript received June 13, 2007; published online March 3, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea, and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007.

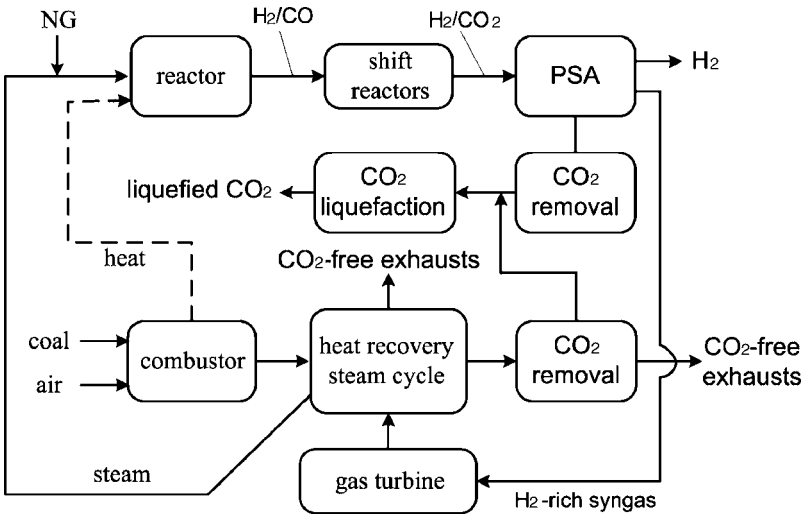


Fig. 1 Conceptual scheme of the multifunctional energy system with CO_2 removal

energy penalty; (2) identify the features of the new system; and (3) clarify the internal phenomena of the new system by graphical exergy analysis.

Novel MES With CO_2 Removal

Conceptual Design of the Novel MES. The novel MES with CO_2 removal is conceptually represented in Fig. 1. The new system integrates an NG-based hydrogen plant and a coal-fired steam cycle, and separates about 90% of carbon dioxide generated from fuel at the same time.

One of the features of the MES is the synthetic utilization of coal and natural gas by the dual fuel reforming method, which was introduced in detail in a previous paper [12]. Coal and natural gas can be used in quite different ways, relying on synthetic utilization. For example, the thermal energy released from the burning of coal (the dashed line shown in Fig. 1) was provided to a reforming reactor, which drove the methane reforming with steam. Thus, the entire natural gas input system can act as a reactant for hydrogen production. Unlike in the MES, about 25% of input natural gas was burned directly in the reformer to provide thermal energy for the reforming reaction in a single NG-based hydrogen plant (NHP). The synthetic utilization of coal and natural gas may decrease the exergy destruction of the conversion process from fossil fuel to thermal energy and may even lower the impact of CO_2 removal on the system performance.

The other feature of the new system is the integration of the hydrogen plant and the thermal cycle. In MES, the PSA tail gas was introduced to a gas turbine as fuel, instead of burning in the reformer of the single hydrogen plant. Due to the high turbine inlet temperature, the exergy destruction of combustion of tail gas might decrease. Moreover, the surplus heat of the hydrogen plant was also recovered and utilized efficiently by the thermal cycle, and the thermal energy (steam) required by the hydrogen plant was provided from the thermal cycle at the proper temperature and pressure. The point emphasized here is that the integration provides a potential to remove CO_2 with lower energy consumption. In the MES, the carbon dioxide was removed just before the gas turbine.

Configuration of the New MES With CO_2 Removal. Figure 2 is a flowchart of the new MES with CO_2 removal, which inte-

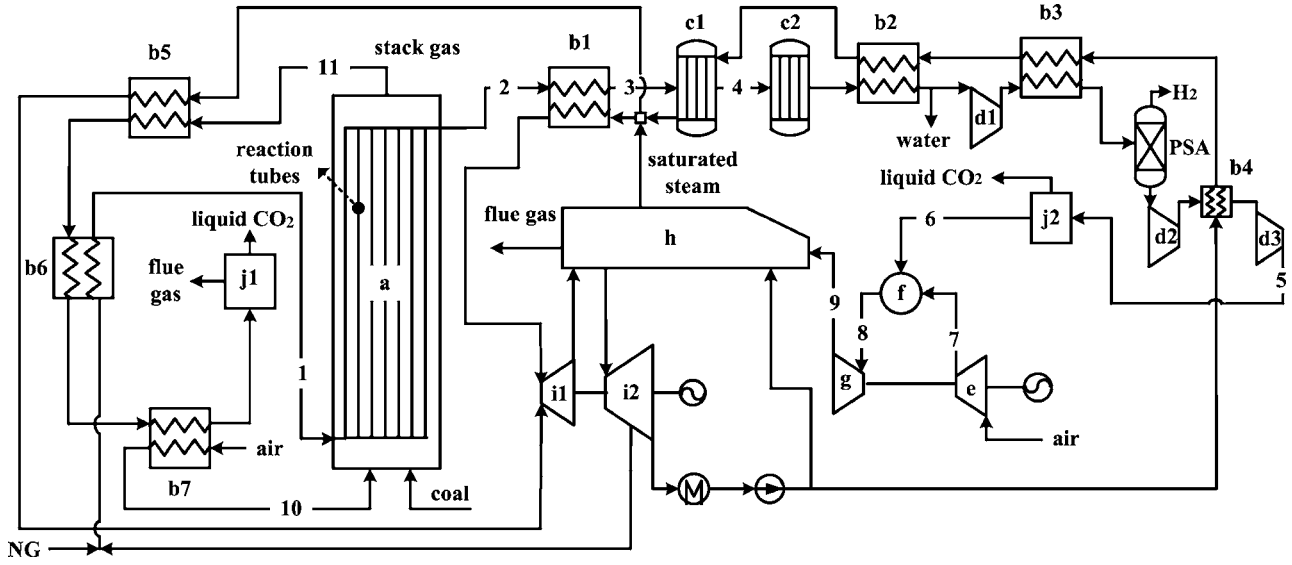


Fig. 2 System configuration of the new MES system with CO_2 removal

grates a coal-fired power plant and a hydrogen plant based on natural gas. This new system is mainly composed of five subsystems, including a dual fuel reformer (a), heat exchangers (b1–b7), a H_2 production subsystem (c1, c2, and PSA unit), a power generation subsystem (d–i), and CO_2 separation, and liquefaction units (j1 and j2). The dual fuel reformer (a) is one of the key components of the MES, and is comprised of reaction tubes and a hearth. The catalyst for the methane/steam reforming is inside the tubes, where the reforming reaction takes place. The coal is burned with preheated air (10) in the hearth (outside of the tubes) of the dual fuel reformer, which provides high-temperature thermal energy for the reforming reaction. The preheated mixture of the natural gas and the steam extracted from the steam turbine (i2) then enters the reaction tubes of the dual fuel reformer, where the mixture is converted into syngas. The syngas is then cooled to 355°C in a heat exchanger (b1) before it enters the shift reactor (c1), and high-temperature superheated steam is simultaneously generated in it (b1). Most of the CO in syngas is converted into CO_2 through two stage shift reactors (c1 and c2), while simultaneously producing high-pressure saturated steam in c1. Finally, the syngas is compressed to about 0.55 MPa through a syngas compressor (d1), before being piped to the PSA unit after cooling in a heat exchanger (b3) to produce pure hydrogen, where about 68% of hydrogen is separated. The tail gas discharged from the PSA unit is compressed by an intercooled two-stage compressor (d2 and d3). The high-pressure syngas is then converted into H_2-rich syngas in the physical absorption [13] and liquefaction unit, in which over 90% of CO_2 in the syngas is separated and liquefied. The surplus heat of stack gases from the burning of coal in the dual fuel reformer is recovered by several heat exchangers (b5–b7). In the heat exchanger of b5, the high-temperature thermal energy of the stack gas is used to superheat high-pressure saturated steam generated by the heat recovery steam generator (HRSG) The middle-temperature thermal energy is used to preheat the reactants of the reforming reaction in a heat exchanger (b6), and the low-temperature thermal energy is used to preheat combustion air in the heat exchanger of b7. Finally, the stack gas enters the chemical absorption (MEA solution) and liquefaction unit (j1) to remove about 90% of CO_2 [14,15].

The H_2-rich syngas generated in the physical absorption unit (j2) is sent to a combined cycle as fuel. The exhaust gas of the gas turbine then enters the HRSG (h) to generate high-pressure saturated steam and reheat the outlet steam of the high-pressure steam turbine (i1). The CO_2 chemical absorption process requires thermal energy for the regeneration of the solvent, which is provided by the steam extracted from the steam turbine. The CO_2 physical absorption process needs power for CO_2 compression and the solution pump, delivered by the MES itself. The CO_2, separated from flue gas of coal and natural gas, is then compressed to a pressure of 15 MPa by a three-stage intercooling compressor before being liquefied in a condenser, thereby making CO_2 ready for storage or disposal.

Evaluation of MES

Description of Reference Systems. With the multi-input of natural gas and coal, the evaluation of the new MES will be based on its comparison to the single systems. Usually, coal is utilized as fuel for direct combustion in steam power plants and NG is used in hydrogen plants. In this study, the coal-fired steam cycle (CSC) and NG-based hydrogen plant (NHP) are selected as reference systems without CO_2 removal. Two systems are considered as reference systems with CO_2 removal: A coal-fired steam plant based on circulating fluidized bed (CFB) boiler with CO_2 removal from exhaust gas (CSCR), and an NG-based hydrogen plant with CO_2 removal from exhaust gas of reformer (NHPR).

Figures 3(*a*) and 3(*b*) show the layouts of the reference plants. Figure 3(*a*), moreover, is a flowchart of CSCR. The coal-fired steam plant is a subcritical steam unit with a steam superheating

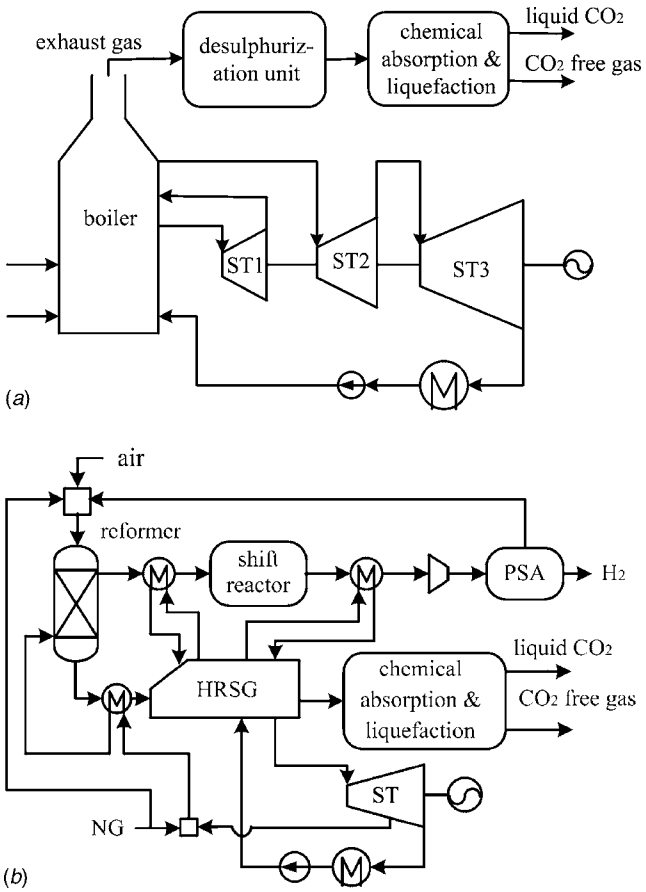


Fig. 3 (*a*) Coal-fired steam plant with CO_2 removal by exhaust gas treatment; and (*b*) hydrogen plant with CO_2 removal by exhaust gas treatment

and the reheating temperature of 540°C for a medium-sized plant, a maximum pressure of 14 MPa, and a condensing pressure of 0.007 MPa. The flue gas from the boiler is cooled and scrubbed in a gas desulphurization unit and then sent to a chemical absorption and liquefaction unit, in which CO_2 is recovered and liquefied. Figure 3(*b*) is a flowchart of the NHPR. Natural gas is first reformed by the steam (extracted from a steam turbine) at 0.2 MPa and absorbs high-temperature thermal energy from its combustion with tail gas from the PSA unit. Next, after being cooled in a syngas cooler, the component of CO in the product of the reforming reaction is converted into CO_2 through a shift reactor. The shifted products are cooled in a heat exchanger and compressed to 0.55 MPa, and the compressed gas enters the PSA unit to separate 80% of the hydrogen. The tail gas of the PSA unit is fed to the reformer as fuel, while the flue gas of the reformer is cooled in the HRSG and is used to generate the superheated steam for the steam turbine. Finally, the CO_2 in the flue gas is removed and liquefied in a chemical absorption and liquefaction unit.

Evaluation of the New MES. In the MES, the power subsystem is based on a heavy-duty gas turbine of current technology with a turbine inlet temperature of 1290°C and a pressure ratio of 15.5. The steam, which is saturated at a pressure of 12 MPa and reheated at a pressure of 3.9 MPa, is generated in the HRSG, and the temperatures of both the superheated and reheated steam are kept at 535°C. To obtain high conversion of methane, which is necessary for CO_2 removal, a near-atmospheric pressure is adopted and the temperature of the reaction is kept around 750°C. The steam/methane ratio is assigned as 3.0, since the unconverted steam will be used in the following shift reactor. The other assumptions in the reforming process are the same as those of

Table 1 Main assumptions throughout all calculations

Chemical process	
Reforming temperature, °C	750
Reforming pressure, MPa	0.15
Ratio of steam to methane	3
Surplus air ratio of coal combustion	1.22
Pressure loss of reformer of hot side, %	12
Pressure loss of reformer of cold side, %	10
Isentropic syngas compressor	0.85
High shift temperature, °C	335
Low shift temperature, °C	200
PSA inlet pressure, MPa	0.55
Hydrogen separation ratio, %	68
Pressure loss of heat exchangers, %	3
Steam to chemical absorption, MPa	0.3
Number of intercoolers for CO_2 comp.	2
Power plant	
Pressure loss of heat exchangers, %	3
Isentropic efficiency of air compressor	0.88
Isentropic efficiency of gas turbine	0.9
Isentropic efficiency of pump	0.8
Pinch point in HRSG, °C	10
Pressure loss of HRSG (gas side), %	3
Pressure loss of HRSG (steam, water)	10
Condensation pressure, MPa	0.007

Kesser [16]. The main assumptions here are reported in Table 1.

MES was studied using the commercial software Aspen Plus. During the simulation, the composition of natural gas was assumed to be methane, and the composition of coal was assumed to be $CH_{0.3}O_{0.17}N_{0.03}$. The pressure, temperature, mass flow, and composition of the main points in Fig. 2 are listed in Table 2.

The performance comparisons of the MES to that of the reference systems without CO_2 removal (CSC+NHP) and reference systems with CO_2 removal (CSCR+NHPR) are listed in Table 3. With the same quantity of natural gas and coal inputs, the MES produced hydrogen output of 344.0 MW, which was 40 MW higher than that of the reference systems without CO_2 removal. However, the electricity production of the MES and the reference system with CO_2 removal were decreased by 38.3 MW and 54.7 MW, respectively. When the CO_2 removal ratio was 90.1%, the overall efficiency of the new system reached to about 63.2%. Compared with the reference system without CO_2 removal, the overall efficiency of the MES was slightly increased by 0.2 percentage points. On the contrary, the efficiency of the reference systems with the same CO_2 removal was decreased by 8.5 percentage points. Unlike the large efficiency decrease of the conventional energy system for CO_2 removal, the MES separated CO_2 without an energy penalty.

To reveal the internal phenomena of the system, an exergy analysis was performed for the MES and the reference systems with CO_2 removal, and the results are listed in Table 4. The exergy analysis was also based on the assumption that the same quantity of natural gas and coal were consumed in the MES and the reference systems with CO_2 removal.

Table 3 Overall performances of the MES and reference systems

	System		
Items	CSC+ NHP	MES	CSCR+ NHPR
Natural gas input	427.5	427.5	427.5
Coal input	214.6	214.6	214.6
Hydrogen output	304.0	344.0	304.0
Power output from NG	17.6	—	0.9
Power output from coal	82.8	—	44.8
Net power output	100.4	63.7	45.7
Power to physical abscissa	—	3.8	—
Heat to chemical abscissa	—	78.4	163.1
Power for CO_2 liq.	—	13.4	16.7
CO_2 removal rate, %	—	90.1	90.1
Overall efficiency, %	63.0	63.2	54.5

The exergy efficiency of the new system was 60.42%, which was 8.38 percentage points higher than that of the reference systems (CSCR+NHPR). Comparing the exergy distributions of the MES to the reference systems in Table 4, the exergy of the hydrogen production and electricity production were increased by 37.96 MW and 17.97 MW, respectively. The exergy destruction of 60.35 MW was brought by the gas turbine in the MES. However, 86.74 MW of the exergy was destroyed in the boiler in the reference systems. The exergy destruction in the CO_2 separation process of the MES (physical and chemical separation) was decreased by 17.59 MW compared to the reference systems with CO_2 removal.

The exergy destruction of the conversion process of fossil fuel to thermal energy was also obviously decreased. In addition, through system synthesis, the exergy destruction of the CO_2 separation process was also decreased by means of the current separation technologies. When the energy utilization and CO_2 separation are integrated, the key problem of high energy penalty for CO_2 removal may be perfectly resolved.

Discussion

Efficient Utilization of the Chemical Energy of Coal and NG. In order to reveal the internal phenomena of the key processes within the MES, the graphical exergy analysis (EUD methodology) was adopted [17]. This section will analyze the utilization of coal and natural gas in detail.

First, we focused on the utilization process of coal in the reference systems with CO_2 removal and the MES. In the reference systems, coal was burned in the hearth of the CFB boiler to produce steam, and the exergy destruction of 86.74 MW was signifi-

Table 2 Parameters of the main points of the MES

Point	G	p	T	CH_4	CO	CO_2	H_2	H_2O	O_2	N_2
Fig. 2	Kg/s	MPa	°C	Percent molar composition (%)						
1	37.33	0.15	549.7	25	—	—	—	75.0	—	—
2	37.33	0.135	750	0.2	10.1	6.4	56.0	27.3	—	—
3	37.33	0.131	355	0.2	10.1	6.4	56.0	27.3	—	—
4	37.33	0.125	335	0.2	1.9	14.6	64.1	19.2	—	—
5	24.5	1.85	211.9	0.6	0.7	42.9	55.6	0.2	—	—
6	3.34	1.80	40.0	0.4	0.4	1.7	97.3	0.3	—	—
7	147.8	16	429.2	—	—	—	—	1.0	20.8	78.2
8	151.2	15.52	1290	—	—	0.9	—	12.5	13.1	73.5
9	170.6	0.103	606.4	—	—	0.8	—	11.1	14.0	74.1
10	85.7	0.12	200.0	—	—	—	—	1.0	20.8	78.2
11	92.5	0.105	900	—	—	16.3	—	3.6	3.7	76.4

Table 4 Exergy analysis of the MES and reference systems

Items	MES		CSCR+NHPR	
	MW	%	MW	%
Exergy input of NG	451.78	67.67	451.78	67.67
Exergy input of coal	215.89	32.33	215.89	32.33
Exergy output				
Hydrogen	339.75	50.89	301.79	45.20
Net electricity	63.65	9.53	45.68	6.84
Exergy destruction				
Reformer	77.07	11.54	72.50	10.86
Boiler	—	—	86.74	12.99
Syngas shift reactors	4.92	0.74	3.57	0.54
Syngas compressor	2.43	0.36	1.26	0.19
GT compressor	4.36	0.65	—	—
GT combustor	45.54	6.82	—	—
GT expansion	10.45	1.57	—	—
Steam turbine	3.89	0.58	7.18	1.07
Heat exchangers	33.70	5.05	51.07	7.65
Exhaust discharge	4.66	0.70	2.5	0.37
PSA unit	4.47	0.67	2.83	0.36
Physical CO_2 separation	4.19	0.63	—	—
Chemical CO_2 separation	20.10	3.01	41.88	6.27
CO_2 liquefaction	6.07	0.91	7.30	1.09
Auxiliaries	5.17	0.77	6.55	0.98
Separated CO_2	37.26	5.58	37.26	5.58
Exergy efficiency, %	60.42		52.04	

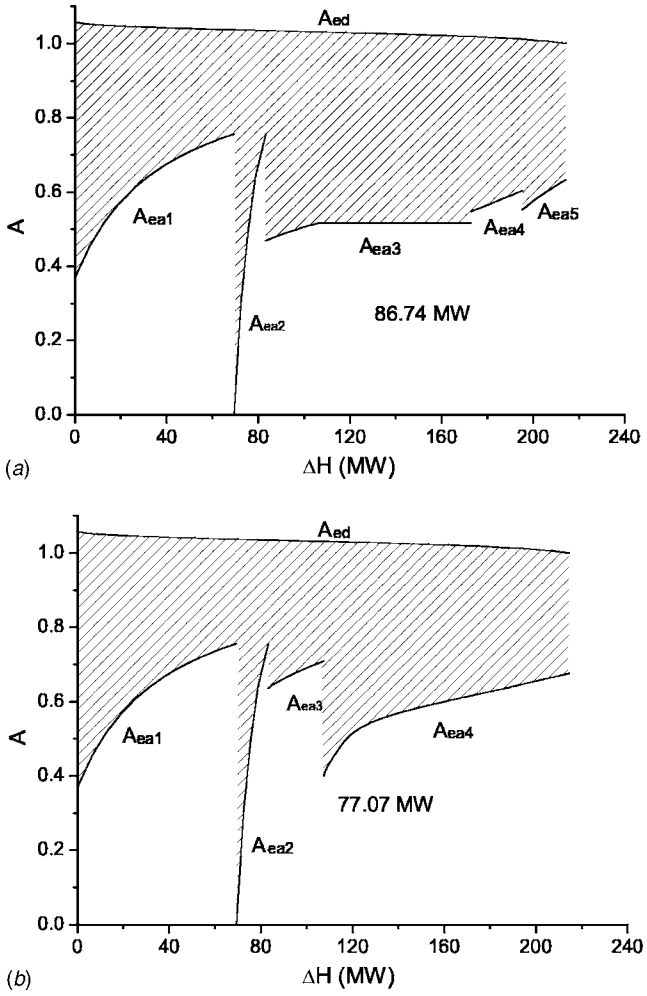


Fig. 4 (*a*) EUD for combustion of coal in the hearth of CFB boiler of reference systems with CO_2 removal; and (*b*) EUD for combustion of coal in hearth of dual fuel reformer of MES

cant in this process. Figure 4(*a*) illustrates the energy utilization diagram (EUD methodology) for the utilization of coal in the hearth of the CFB boiler in reference systems. As shown by the curve of A_{ed}, the oxidation of coal in the boiler acted as the energy donor. The energy acceptors include the preheating of air (curve A_{ea1}) and coal (curve A_{ea2}), the evaporation of water (curve A_{ea3}), the superheating of steam from 387.6°C to 479.5°C (curve A_{ea4}) and the reheating of steam from 393.6°C to 540°C (curve A_{ea5}). The shaded area between the energy donor and acceptors represents the exergy destruction in the hearth of the CFB boiler. The average energy level of coal oxidation was about 1.03, while the average energy level of evaporation of water was about 0.51, corresponding to the evaporation temperature of 342.5°C and the average energy level of the steam superheating process and reheating process were 0.57 and 0.59, respectively. From Fig. 4(*a*) we can disclose that the big difference between the energy levels of energy donor and acceptors causes the large exergy destruction in the conventional boiler.

Meanwhile, Fig. 4(*b*) shows the exergy-destruction distribution in the dual fuel reformer. The oxidation of coal also acted as an energy donor (curve A_{ed}), and the methane/steam reforming reaction (curve A_{ea4}) acted as one of the energy accepters. In the dual fuel reformer, the average energy level of the heating of a mixture of methane and steam was 0.68, and the average energy level of the reforming reaction was about 0.60. Compared to Figs. 4(*a*) and 4(*b*), the average energy level of the energy acceptors in the dual fuel reformer was higher than that in the boiler, which resulted in the decrease of exergy destruction in the burning of coal by 11.2% (9.67 MW).

Similar to coal, natural gas was also used by the MES in a quite different way compared to the reference systems. In the NG-based hydrogen plant (NHPR), about 25% of the natural gas was combusted to drive the reforming reaction, while in the MES, this part of natural gas was used as a reactant of the reforming reaction to produce hydrogen, with the PSA tail gas sent to the combustor of the gas turbine after CO_2 removal. Figures 5(*a*) and 5(*b*) illustrate the difference between combustion of the decarbonized tail gas in the combustor of the gas turbine and that of natural gas in the reformer.

In Fig. 5(*a*), the curve of A_{ed1} represents the oxidation (combustion) of natural gas in the reformer acting as donator and four streams (curves A_{ea1} through A_{ea4}) acting as the energy acceptors: The preheating of air (curve A_{ea1}) and natural gas (curve A_{ea2}), the preheating of the reactant of reforming (curve A_{ea3}), and the methane/steam reforming reaction (curve A_{ea4}). The area below the curve of A_{ed2} represents the exergy destruction in the mixing process of air and natural gas. Figure 5(*b*), meanwhile, shows the exergy-destruction distribution of the combustion process of the decarbonized tail gas in the combustor of the gas turbine. The oxidation of the decarbonized tail gas acted as an energy donor (curve A_{ed1}), and two feed streams, air (curve A_{ea1}) and tail gas (curve A_{ea2}), acted as the energy acceptors. The area below the curve of A_{ed2} represents the exergy destruction in the mixing process of air and tail gas.

Comparing Figs. 5(*a*) and 5(*b*), the distinctive difference is located in the energy acceptors. In Fig. 5(*a*) the average energy level of the heating processes was 0.62 and the average energy level of the reforming reaction was about 0.60. In Fig. 5(*b*) the average energy level of the preheating of fuel and air were 0.64 and 0.72, respectively. Compared to the combustion of natural gas in the reformer, the average energy level of the energy acceptors in the combustion of the decarbonized tail gas was increased, which, in turn, led to a great decrement in exergy destruction. The difference of the exergy destruction between Figs. 5(*a*) and 5(*b*) is 36.96%, which indicates that the natural gas is used more efficiently in the MES.

In general, by combining the utilization of coal and natural gas, the exergy destruction in the conversion processes of fossil fuel

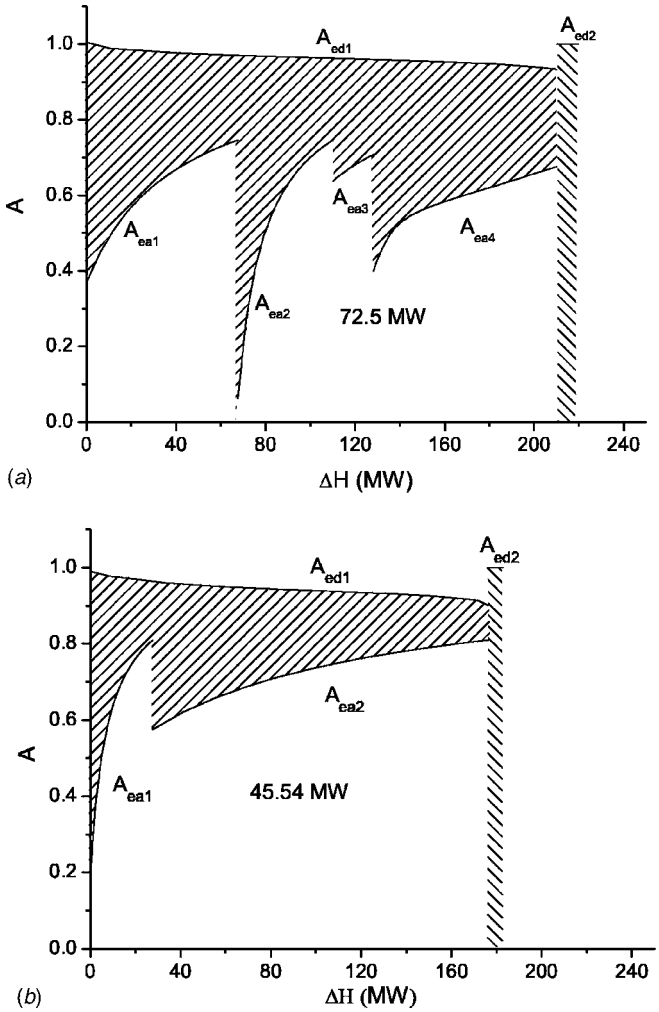


Fig. 5 (*a*) **EUD for combustion of NG in reformer of reference systems with CO_2 removal; and** (*b*) **EUD for combustion of tail gas in gas turbine combustor of MES**

into thermal energy was decreased by 36.63 MW. The decrease of exergy destruction led to the excellent performance of the new system. Thus, the exergy efficiency of the MES was increased by 5.5 percentage points.

Integration of Energy Utilization and CO_2 Separation. The total energy penalty for CO_2 removal was 8.4 percentage points in the reference systems, but the decrease of the energy penalty in the MES was mainly based on the two steps of integration: (1) The integration of the hydrogen plant and the thermal cycle, and (2) the synthetic utilization of coal and natural gas in the dual fuel reformer (discussed above).

In the reference system (NHPR), the tail gas of the PSA unit was directly burned in the reformer, and then the CO_2 was removed from the flue gas by MEA absorption process, as shown in Fig. 3(*b*). The steam of 84.7 MW with the exergy of 22.3 MW was extracted from the steam turbine for the regeneration of MEA solution. It resulted in a decrease of 3.3 percentage points of the exergy efficiency of the reference systems.

Unlike the reference systems with CO_2 removal, the hydrogen plant and the thermal cycle were integrated in the proposed MES, and the PSA tail gas was fed into the gas turbine, instead of the reformer. Before entering the gas turbine, the PSA tail gas needed to be compressed to a pressure of 1.8 MPa. It is to be noted that the partial pressure of the carbon dioxide increased from 0.043 MPa to 0.77 MPa after compression, which provides the possibility of decreasing the energy consumption for CO_2 removal by means of system synthesis. In the MES, this portion of CO_2 was removed just before the gas turbine with exergy consumption of 4.2 MW, and the exergy efficiency of the MES was only decreased by about 0.6 percentage points. In other words, relying on the integration of the hydrogen plant and the thermal cycle, the MES has less energy penalty of 2.7 percentage points for CO_2 removal than that in the reference system (NHPR).

With the exergy efficiency increment of 5.5 percentage points contributed by the synthetic utilization of coal and natural gas, and the exergy-destruction reduction achieved in the CO_2 separation process of 2.7 percentage points, the MES can obtain a benefit of a total of 8.2 percentage points, which just makes up the energy penalty for CO_2 removal in the reference systems.

Further Considerations. The system configuration proposed here was conceived for CO_2 removal from the MES. For the power section, the gas turbine needed just a little modification to adapt to the decarbonized fuel and was not a technological obstacle [10,18]. The conventional steam turbine can be adopted without modification. For the hydrogen generation unit, the small- and middle-scale PSA unit is commonly used throughout the world, but a large-scale PSA unit needs to be further researched and developed.

For the fuel treatment section, the coal-fired methane/steam reformer was the key equipment, and the structure of the dual fuel reformer was similar to a boiler, except that the stream flowing inside the reaction tubes is changed into the reactant of the reforming reaction. Since the condition of the reformer (the surface temperature of the reaction tubes) is quite different from the boiler, the high temperature corrosion and fouling on the reaction tubes should be taken into account. Techniques, such as the treatment of coal to remove most of the components of sulfur, ash, and alkali metals, and the proper combustion system, to minimize the corrosion and fouling, should be adopted. The shift reactors, chemical absorption unit, and physical absorption unit can adopt state of the art technologies.

The large energy consumption for the CO_2 removal will increase the consumption of fossil fuel with the same output of products. For the single NG-based hydrogen plant, CO_2 removal would add approximately 18–33% to the unit cost of hydrogen, in which the cost of feedstock accounts for 35–60% [1]. Since the MES can remove about 90% of CO_2 with zero energy penalty, the unit cost of hydrogen will only add about 11–20% in the MES, compared to the single hydrogen plant with CO_2 removal.

To remove the CO_2 from conventional energy systems, new equipment will be needed and extra energy will be consumed. Accordingly, the cost of products would be increased, which is unacceptable for the current energy system. However, it should still be noted that the MES may implement zero energy penalty for CO_2 removal based on the system synthesis, and that the system proposed here may lead to a new generation energy system toward more efficient greenhouse gas control.

Conclusions

This paper proposed a novel multifunctional energy system that can produce electricity and hydrogen together, with the recovery of CO_2 through the synthetic utilization of natural gas and coal by dual fuel reforming. This new system is expected to achieve a net overall efficiency of 63.2% and realize near-zero energy penalty for CO_2 recovery, compared with the conventional NG-based hydrogen plant and coal-fired plant with CO_2 recovery. The graphical exergy analysis (EUD methodology) revealed that the synthetic utilization of natural gas and coal brings about the reduction of exergy destruction in the conversion of fossil fuel into thermal energy, which accounts for 5.5 percentage points corresponding to the input exergy of the system. Moreover, this significant improvement in the energy utilization can make up the energy penalty for CO_2 removal compared to the conventional NG-based

hydrogen plant and coal-fired plant with CO_2 recovery. In addition, the combination of the hydrogen production plant and the thermal cycle in the new system enables a reduction in the exergy destruction of CO_2 separation, accounting for 2.7 percentage points corresponding to the input exergy of the system. Finally, the proposed MES provides a potential to recover CO_2 with near-zero energy penalty, and offers a new approach to simultaneously resolve the efficient utilization of fossil fuel and the control of greenhouse gas emissions based on system synthesis.

Acknowledgment

This work has been sponsored by the Natural Science Fundamental Research Program of China (Grant Nos. 50520140517 and 50706051).

Nomenclature

MES = multifunctional energy system
CSC = coal fired steam cycle
NHP = natural gas-based hydrogen plant
CSCR = coal fired steam power plant with CO_2 removal by exhaust gas treatment
NHPR = natural gas-based hydrogen plant with CO_2 removal by exhaust gas treatment
EUD = energy-utilization diagram

References

[1] Bert, M., Ogunlade, D., de Heleen, C., Manue, L., and Leo, M., 2005, "Carbon Dioxide Capture and Storage," IPCC Special Report, ISBN 92-9169-119-4.
[2] Gambini, M., and Vellini, M., 2003, "CO_2 Emission Abatement From Fossil Fuel Power Plants by Exhaust Gas Treatment," ASME J. Eng. Gas Turbines Power, **125**, pp. 365–373.
[3] Smelser, S. C., Stock, R. M., Mc Cleary, G. J., Booras, G. S., and Stuart, R. J., 1991, "Engineering and Economic Evaluation of CO_2 Removal From Fossil-Fuel-Fired Power Plants—Volume 1: Pulverized Coal-Fired Power Plants," EPRI Report No. IE-7365.
[4] Giittlicher, G., and Pruschek, L., 1997, "Comparison of CO_2 Removal Systems for Fossil-Fuelled Power Plant Processes," Energy Convers. Manage., **38** pp. s173–s178.
[5] Bolland, O., and Mathieu, P., 1998, "Comparison of Two CO_2 Removal Options in Combined Cycle Power Plants," Energy Convers. Manage., **39**, pp. 1653–1663.
[6] Chiesa, P., and Lozza, G., 1999, "CO_2 Emission Abatement in IGCC Power Plants by Semi Closed Cycles: Part A—With Oxygen-Blown Combustion," ASME J. Eng. Gas Turbines Power, **121**, pp. 635–641.
[7] Okawa, M., Kimura, N., Kiga, T., Takano, S., Arai, K., and Kato, M., 1997, "Trial Design for a CO_2 Recovery Power Plant by Burning Pulverized Coal in O_2/CO_2," Energy Convers. Manage., **38**, pp. S123–S127.
[8] Chiesa, P., and Consonni, S., 1999, "Shift Reactors and Physical Absorption for Low-CO_2 Emission IGCCs," ASME J. Eng. Gas Turbines Power, **121**, pp. 295–305.
[9] Lozza, G., and Chiesa, P., 2002, "Natural Gas Decarbonization to Reduce CO_2 Emission From Combined Cycles—Part I: Partial Oxidation," ASME J. Eng. Gas Turbines Power, **124**, pp. 82–88.
[10] Lozza, G., and Chiesa, P., 2002, "Natural Gas Decarbonization to Reduce CO_2 Emission From Combined Cycles—Part II: Steam-Methane Reforming," ASME J. Eng. Gas Turbines Power, **124**, pp. 89–95.
[11] Bolland, O., and Undrum, H., 2003, "A Novel Methodology for Comparing CO_2 Capture Options for Natural Gas-Fired Combined Cycle Plants," Adv. Environ. Res., **7**, pp. 901–911.
[12] Jin, H., Han, W., and Gao, L., 2006, "Multi-Functional Energy System (MES) With Multi Fossil Fuels and Multi Products," ASME J. Eng. Gas Turbines Power, in press.
[13] Doctor, R. D., Molburg, J. C., and Thimmapuram, P. R., 1996, "KRW Oxygen-Blow Gasification Combined Cycle: Carbon Dioxide Recovery, Transport, and Disposal," ANL/ESD-34, Argonne National Laboratory, Argonne, IL 60439.
[14] Alie, C., Backham, L., Croiset, E., and Douglas, P. L., 2005, "Simulation of CO_2 Capture Using MEA Scrubbing: A Flowsheet Decomposition Method," Energy Convers. Manage., **46**, pp. 475–487.
[15] Geuzebroek, F. H., Schneiders, L. H., Kraaijveld, G. J. C., and Feron, P. H. M., 2004, "Exergy Analysis of Alkanolamine-Based CO_2 Removal Unit With Aspenplus," Energy, **29**, pp. 1241–1248.
[16] Kesser, K. F., Hoffman, M. A., and Baughn, J. W., 1994, "Analysis of a Basic Chemically Recuperated Gas Turbine Power Plant," ASME J. Eng. Gas Turbines Power, **116**(4), pp. 277–284.
[17] Ishida, M., and Kawamura, K., 1982, "Energy and Exergy Analysis of a Chemical Process System With Distributed Parameters Based on the Energy-Direction Factor Diagram," Ind. Eng. Chem. Process Des. Dev., **21**, pp. 690–695.
[18] Chiesa, P., Lozza, G., and Mazzocchi, L., 2005, "Using Hydrogen as Gas Turbine Fuel," ASME J. Eng. Gas Turbines Power, **127**, pp. 73–80.

Laboratory Studies of the Flow Field Characteristics of Low-Swirl Injectors for Adaptation to Fuel-Flexible Turbines

R. K. Cheng

D. Littlejohn

Environmental Energy Technologies Division,
Lawrence Berkeley National Laboratory,
Berkeley, CA 94720

W. A. Nazeer

K. O. Smith

Solar Turbines Incorporated,
2200 Pacific Highway,
San Diego, CA 92101

The low-swirl injector (LSI) is a simple and cost-effective lean premixed combustion method for natural-gas turbines to achieve ultralow emissions (<5 ppm NO_x and CO) without invoking tight control of mixture stoichiometry, elaborate active tip cooling, or costly materials and catalysts. To gain an understanding of how this flame stabilization mechanism remains robust throughout a large range of Reynolds numbers, laboratory experiments were performed to characterize the flowfield of natural-gas flames at simulated partial load conditions. Also studied was a flame using simulated landfill gas of 50% natural gas and 50% CO_2. Using particle image velocimetry, the nonreacting and reacting flowfields were measured at five bulk flow velocities. The results show that the LSI flowfield exhibits similarity features. From the velocity data, an analytical expression for the flame position as function of the flowfield characteristics and turbulent flame speed has been deduced. It shows that the similarity feature coupled with a linear dependency of the turbulent flame speed with bulk flow velocity enables the flame to remain relatively stationary throughout the load range. This expression can be the basis for an analytical model for designing LSIs that operate on alternate gaseous fuels such as slower burning biomass gases or faster burning coal-based syngases.
[DOI: 10.1115/1.2795786]

Introduction

Lean-premixed (LP) combustion technologies have been adopted by virtually every industrial gas turbine manufacturer as a dry low NO_x (DLN) method to meet emission regulations being implemented in the US and in many regions worldwide. However, to meet more stringent ultralow emissions standards being proposed, the DLN combustors have to operate at conditions near the lean limit of their stability domains where noise, instability, and flame blowoff can seriously affect engine performance and reliability. To mitigate these potential problems, much effort has been devoted to explore remedies based on passive control (e.g., fuel and/or air staging) [1], and active control (e.g., autofeedback loop) [2,3]. Other alternatives invoke more costly exhaust gas clean up or catalytically assisted combustion. Undoubtedly, these new schemes would lead to elaborate combustion devices requiring tightly controlled sensors and actuators as well as many auxiliary components. For engines that operate on gaseous fuels other than natural gas, combustion instability problems are further exacerbated due to the variability of the fuel contents and fuel air ratios. Therefore, the injectors as well as the combustors have to be optimized or reengineered to accommodate the changes in the combustion properties.

One promising solution for fuel-flexible gas turbines is a nascent premixed combustion technology that operates on a unique low-swirl combustion concept. Originally developed at the Lawrence Berkeley National Laboratory as a small laboratory research burner for fundamental studies (15 kW), a good understanding of its operating principle has been obtained [4–9]. This patented combustion method exploits the aerodynamic properties of the propagating turbulent premixed flames. It is a simple, robust, and readily adaptable technology to meet stringent emission targets without significantly altering the system configurations, or impact on efficiency, turndown, and costs. Low-swirl combustion has been commercialized for industrial process heaters. Low-swirl burners (LSBs) of 150 kW–7.5 MW (0.5–25 MMBtu/h) with ultralow emissions of 4–7 ppm NO_x and CO (both at 3% O_2) within 10:1 turndown have been available since late 2003. Central to the commercialization pathway was the basic knowledge gained from laboratory studies that has provided critical information for scaling as well as resolving system integration issues [10,11]. Low-swirl combustion is also being developed for natural-gas turbines. Rig tests of a prototype low-swirl injector (LSI) for 5–7 MW engines show it to be a cost-effective solution that enables DLN turbines to meet the emission targets of <5 ppm (at 15% O_2) for both NO_x and CO. [12,13].

The low-swirl combustion concept is readily adaptable for burning other hydrocarbons and hydrogen enriched fuels. The approach is to adjust the flowfield to accommodate the differences in the turbulent flame speeds of these premixed flames. The process requires basic knowledge of the combustion properties for each fuel, e.g., flame speeds and flame temperatures. Equally significant is a good understanding of how these flames couple with and respond to the turbulent flowfield. Therefore, laboratory investigations to characterize these flame and their flowfield properties are essential. The validity of this approach has been demonstrated by development of fuel-flexible industrial LSBs. Prototypes have been tested with propane, ethylene, natural-gas diluted with flue gases (up to 30%) [11], and refinery gases having large hydrogen constituencies (up to 50% H_2) [14].

The main issue for adapting the LSI to fuel-flexible gas turbines stems from the significant diversity in the constituents of the alternative fuels and the variations in combustion properties and heat contents. To learn how to adjust the LSB flowfield to burn these slower burning and faster burning fuels (compared to natural gas), the first step is to obtain the baseline knowledge on how the

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 25, 2007; final manuscript received September 6, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), Barcelona, Spain, May 8–11, 2006. Paper No. GT2006-90878.

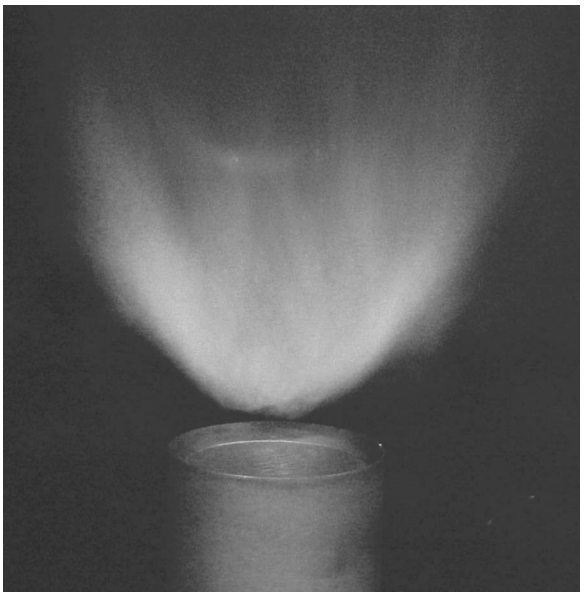


Fig. 1 LSI produces a freely propgating lited flame

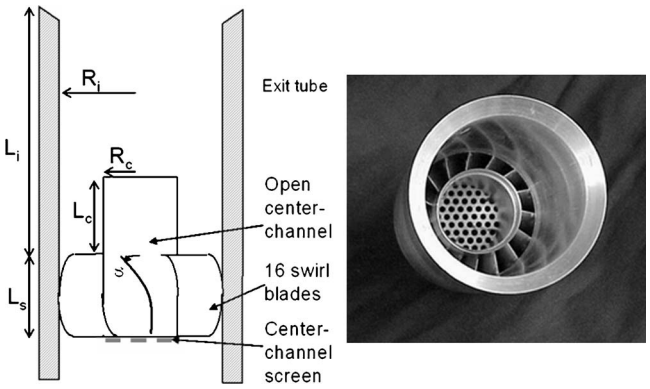


Fig. 2 Schematics and top view of a laboratory LSI

flowfield of natural-gas flame evolves with velocity. Toward this goal, laboratory experiments have been performed to characterize the changes in the LSI flowfield with flow Reynolds numbers and combustion heat release. Using particle image velocimetry (PIV), the distributions of the mean velocities, the rms fluctuations, and shear stresses have been obtained at five bulk flow velocities under atmospheric conditions. The corresponding nonreacting flows were also investigated. In addition, the feasibility of the current LSI prototype to accept alternate fuels was demonstrated by firing with a simulated landfill gas of 50% natural gas and 50% CO_2.

Low-Swirl Injector Configuration

Premixed flames consume the reactants in the form of self-sustaining reacting waves that propagate at flame speeds controlled by the mixture compositions, the thermodynamic conditions, and the turbulence intensities. In contrast, nonpremixed diffusion flames do not propagate (i.e., move through the reactants medium) because burning occurs only at the mixing zone formed between the fuel and the oxidizer streams. The low-swirl combustion method exploits the wavelike behavior of premixed turbulent flames and is valid only for premixed combustion. The premise is to capture a fast moving turbulent premixed flame as a "standing wave" that remains stationary without the need for a flame anchor (Fig. 1). This can be accomplished by letting the premixed flame burns freely in a divergent flow region produced in a weakly swirling flow.

The critical requirement is to set the swirl rate at a sufficiently low level to inhibit vortex breakdown—a precursor to the formation of flow reversal and strong recirculation [15]. Under this situation, the primary role of the swirling motions is to induce mean radial aerodynamic strain to expand the flow radially when it exits the burner. Radial flow expansion is accompanied by a linear decrease in the mean axial velocity. This velocity "downramp" is the key feature that provides a robust configuration for a premixed turbulent flame to freely propagate and settle at a position where the local flow velocity is equal and opposite to the flame speed. The flame does not flashback because it cannot propagate faster than the velocity at the exit. Blowoff is also mitigated because the divergent flow provides a broad region where the flame naturally settles. More importantly, mixture inhomogeneity or slight flow transients cause only a shift in the flame position and the likelihood of flameout is minimized. This is a self-adjusting mechanism for the flame to withstand transient and some abrupt changes.

For gas turbines, the key LSI component is a patented swirler (Fig. 2). It has an outer annular swirled section and an open center channel to allow a portion of the reactants to remain unswirled. The center-channel flow is a critical component whose presence retards recirculation. To achieve the appropriate flow divergence rate involves balancing the swirled and the unswirled flows. This is done by fitting a perforated screen to the center channel. Another key parameter is the swirler recess distance, L_i. It controls the residence time of the interaction between the swirled and unswirled flows and has a secondary effect on the flow divergence and thus the overall LSI performance.

Because an annular vane swirler is a standard component of most high-swirl injectors (HSIs), the LSI prototype utilizes an existing swirler designed for HSI and the details are given in Ref. [12]. The significant alteration from HSI to LSI is the removal of the solid centerbody to form the center channel. To configure the LSI for optimum flame stability at ultralean conditions, we invoked the engineering rules developed for atmospheric flow LSBs. These rules are specified in terms of the swirl number S and swirler recess distance L_i. The LSI swirl number is defined as [10]

$$S = \frac{2}{3}\tan\alpha\frac{1-R^3}{1-R^2+[m^2(1/R^2-1)^2]R^2} \quad (1)$$

where α is the vane angle, R is the ratio of the center channel R_c and the injector radii R_i and m is the mass flux ratio of the unswirled and swirled flows. Here, α and R in Eq. (1) are fixed by the swirler design. Using the same procedure for LSB development, the swirler was fitted with exit tubes of various lengths to vary L_i and different centerbody screens to vary S. These prototypes were evaluated at open atmospheric conditions. Using the lean blowoff and flame position as the criteria, a LSI design with swirl number $S=0.5$ and $L_i=9.5$ cm was configured for a 6.3 cm diameter injector. This is consistent with the $0.4<S<0.55$ guideline developed for designing LSBs.

Next, the LSI was evaluated with a combustor liner at simulated full load (430°C, 15 atm) and partial loads (230°C, 6 atm) [12]. Successful firing of the LSI at gas turbine conditions showed the direct relevance of the open flame experiments to gas turbine development. More significantly, the LSI NO_x emissions are 60% lower than those of a HSI (Fig. 3). These results demonstrate that the LSI can attain low emissions far away from the lean blowoff limits and at conditions less conducive to combustion oscillations. The lowest levels of NO_x and CO (<5 ppm) are comparable with those from much more costly and less durable catalytic options. Therefore, the implementation of the LSI is expected to be low cost and free of concerns about complex controls, durability, and maintenance [13].

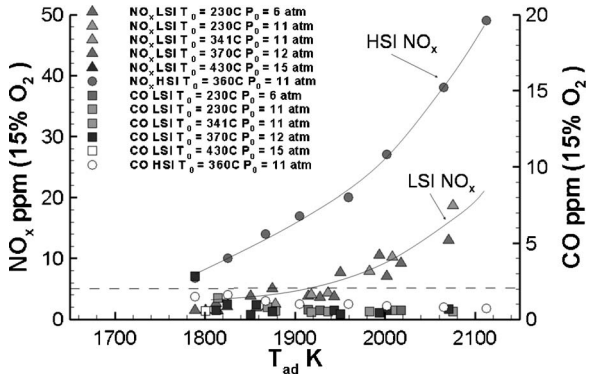


Fig. 3 Emissions of LSI and HIS from rig tests at partial and full load gas turbine conditions

Johnson et al. [12] also compared the flowfields produced by the LSI and the HSI. They showed that the lifted and detached LSI flame originates in the central region of the flowfield that is free of swirl and flow recirculation. In contrast, HSI imparts swirl to the entire flow of the reactants. Its centerbody promotes the formation of a tight recirculation zone in its wake. The HSI flame attaches to the rim of the centerbody and the hot combustion products trapped in the wake are essential for holding and continuously igniting the fresh reactants. Therefore, swirling and recirculating fluid motion dominate the HSI flame. Although the swirl numbers of the HSI ($S=0.73$) and the LSI differ by a seemingly trivial amount, due to the fundamental differences in the flowfields and the flame stabilization mechanism, their overall combustion behaviors as well as the emissions are quite different.

Experimental Setup and Diagnostics

Figure 4 shows a schematic of the experimental setup. For the current study, the LSI was mounted vertically on top of a 30 cm long pipe fitted centrally in a 12.7 cm diameter, 33 cm tall cylindrical flow settling chamber. This chamber is from a commercial LSB and the mounting of the LSI replicates the commercial configuration. The fuel and air mixtures were delivered through the side of the chamber and entered from the bottom of the central pipe. A honeycomb section at the pipe entrance acts as flow straightener. Air supplied by a fan blower is mixed with pipeline natural gas (NG) in a commercial venturi premixer manufactured by the Maxon Corp. The seeding flow for the PIV measurements was added to the premixture near the chamber entrance. This setup utilized a commercial burner control system manufactured by the Honeywell International Corp. It consists of a set of flow and flame sensors, control dampers, and a spark igniter. Startup and shutdown followed standard prescribed procedures. The experimental conditions were set and monitored by a personal computer (PC) control algorithm interfaced with the Honeywell system and an electronic flow controller for the PIV seeds. This experimental setup affords a maximum flow rate of 4500 LPM with heat output of up to 300 kW.

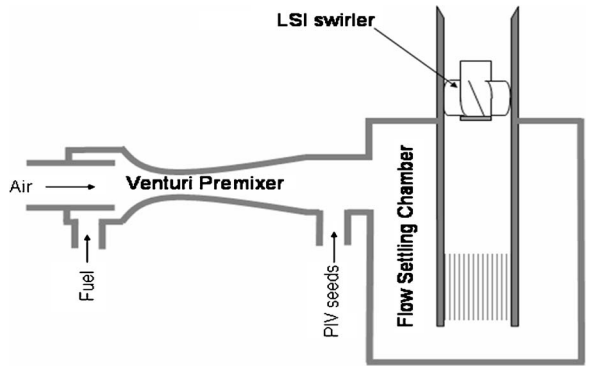


Fig. 4 Schematics of the LSI setup

Flowfield information was obtained using a PIV system consisting of a New Wave Solo PIV laser with double 120 mJ, 6 ns pulses at 532 nm and a Kodak/Red Lake ES 4.0 digital camera with 2048×2048 pixel resolution. The optics captured a field of view of approximately 13×13 cm^2 covering the near field as well as the far field of the flames with 0.065 mm/pixel resolution. A cyclone type particle seeder seeded the air flow with 0.5 μm Al_2O_3 particles, which can track velocity fluctuations up to 10 kHz [16].

Data acquisition and analysis were performed using software developed by Wernet [17]. Because of the complex and 3D nature of the swirling flowfield, care had to be taken to optimize interframe timing, camera aperture setting, light-sheet thickness, and seed density to ensure high data fidelity. Using a portion of the light sheet with approximately 1.1 mm thickness (away from the 0.3 mm waist produced by the 450 mm spherical lens) and a short interframe time (30–35 μs) helped to freeze the out of plane motion of seed particles. Sets of 224 image pairs were recorded for each experiment. The PIV data were processed using 64×64 pixel cross-correlation interrogation regions with 50% overlap. This rendered a spatial resolution of ≈2 mm.

As in Ref. [12], the PIV experiments were conducted in the open and at atmospheric pressure. Although enclosed flames give closer facsimiles of gas turbine combustion, a previous study [18] shown that the open and enclosed LSB flowfields are not significantly different when the enclosure diameter is three times larger than that of the LSB. The 3:1 expansion ratio guideline has been adopted in commercial LSBs and in the combustor liner for LSI. Due to the significant decreases in signal to noise ratios when capturing the PIV images through an optically accessible window, the need to perform enclosed flame experiments does not seem to be fully justified.

Results

Table 1 lists the experimental conditions. The nonreacting flows and NG/air flames were chosen to study the evolution of the LSI flowfields within a 3:1 turndown range. Though these velocities are relatively low compared to those in a typical gas turbine, they are relevant due to the fact that the characteristics of the LSI flames such as flame positions, flame discharge angles, and lean blowoff limits have been shown to be consistent with those at elevated T_0 and P_0 [12]. To investigate the fuel-flexible capability of the present LSI design, Flame 6 utilizes a fuel of 50% NG and 50% CO_2 to simulate a typical landfill gas with medium Wobbe index of 448 Btu/ft^3 (19 mJ/M^3). This is about half the Wobbe index of NG. Set at $\phi=0.8$, Flame 6 has the same adiabatic flame temperature as the $\phi=0.7$ NG flames ($T_{ad}=1845$ K) but with a lower S_L of 0.13 m/s instead of 0.196 m/s.

Mean Flowfields. Figure 5 compares the normalized mean velocity vectors and the normalized shear stresses $\overline{uv}/u'v'$ measured in Flows 1 and 5, and Flames 1 and 5. These conditions are at the minimum and maximum Re that we have investigated. The results obtained for Flows 1 and 5 (top) show the typical features of the flowfield generated by the LSI. In the near field ($x<40$ mm), the nonswirling flow through the center channel generates a relative flat velocity distribution around the axis ($-20<r<20$ mm). It is surrounded by a faster outward spreading swirling flow supplied by the swirl annulus. The expanding swirling motion generates an adverse axial pressure gradient and induces flow divergence within the nonswirling center core flow. As a result, flow velocity decreases with increasing x as shown by the shortening of the velocity vectors downstream. Within the nonswirling center core, the shear stress levels are very low because the turbulence is con-

Table 1 Experimental conditions

Case	U_0 (m/s)	Re	Fuel	ϕ	Output (kW)
Flow 1	6.83	27802	None	0.0	0
Flow 2	10.12	41211	None	0.0	0
Flow 3	15.11	61509	None	0.0	0
Flow 4	19.12	77809	None	0.0	0
Flow 5	22.74	92571	None	0.0	0
Flame 1	7.70	31339	NG	0.7	55
Flame 2	10.09	41088	NG	0.7	73
Flame 3	15.88	64652	NG	0.7	113
Flame 4	18.99	77317	NG	0.7	139
Flame 5	21.97	89434	NG	0.7	158
Flame 6	7.85	31973	NG/CO2	0.8	58

trolled by the perforated plate placed over the center channel. High shear stresses occur in the outer regions where the swirling flow entrains and mixes with ambient air. The outlines of the high stress regions also illustrate that the discharge angles of the nonreacting flows are not sensitive to Re. In the far field of Flow 5, a near zero region is found at about $x=70$ mm. This feature together with slight increases in shear stresses in the central region at $x>100$ mm indicates the presence of a very weak flow recirculation zone [7,12]. In contrast, a weak recirculation zone is not found in Flow 1. The results indicate that the appearance of a weak recirculation bubble in the far field is one of the main Re dependent characteristics of the LSI flowfield.

The development of flow recirculation with increasing Re is shown again by the velocity vectors of Flames 1 and 5 (Fig. 5, bottom row). For Flame 1 at a low Re, combustion heat release accelerates the flow velocity in the far field such that the small velocity deficit found on the corresponding Flow 1 at $x=120$ mm does not exist. For Flame 5, combustion has the opposite effect in that it enhances the flow recirculation. A stagnation point is seen at $x=82$ mm and negative flow velocities are observed at $x=120$ mm. As discussed in previous publications [7,12], the weak recirculation zones in LSI and LSB are far away from the flame. For Flames 1 and 5, the leading edge of the flame brushes are at $x=30$ and 20 mm, respectively. Therefore, flow recirculation has little relevance to the flame stabilization mechanism. The flowfields of Flames 1 and 5 also show that the discharge angles of the flow are similar and independent of the flow velocity. Visual observations of the LSI flames during the enclosed test rigs confirmed the same behavior.

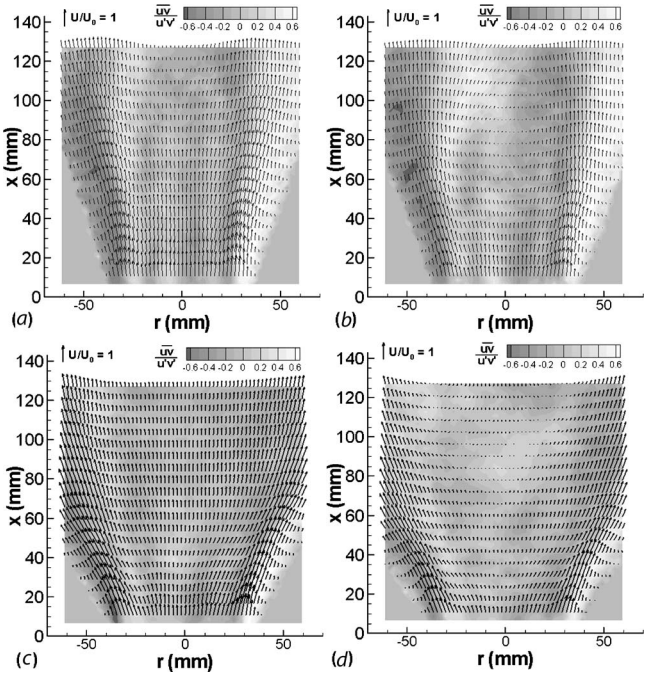


Fig. 5 Normalized mean velocity vectors and normalized shear stress for (*a*) Flow 1, (*b*) Flow 5, (*c*) Flame 1 and (*d*) Flame 6

Axial Profiles. The freely propagating flame generated by the LSI settles naturally at the position with the central region where the local flow velocity matches that of the turbulent flame speed S_T. Therefore, the centerline velocity profiles are useful for understanding how the LSC mechanism remains robust under a wide range of flow and mixture conditions. Quantitative features of the flowfields and the flame, e.g., mean strain rate S_T and flame brush thickness can be deduced from these profiles.

Shown in Figs. 6 and 7 are the centerline axial profiles of U, u', and v' for Flows 1–5 and Flames 1–5, respectively. In Fig. 6(*a*), the nonreacting U profiles all exhibit the characteristic linear velocity decay beginning at the burner exit. The development of the weak recirculation zone is illustrated by the appearance of velocity troughs on the profiles of Flows 3–5. The velocity troughs move closer to the LSI exit with increasing Re. Although the minimum U on these profiles did not drop below zero, the corresponding large values of u' shown on Fig. 6(*b*) indicate the occurrence of intermittent flow reversal that would be consistent with the initiation of a weak recirculation zone. From Figs. 6(*b*) and 6(*c*), it can be seen that the turbulence at the LSI exit is anisotropic with u' higher than v'. For Flows 1 and 2, the anisotropy is slight and the u' and v' profiles do not show significant turbulence decay with increasing x. A lack of turbulence decay indicates the influence of mean strain imposed by the divergent flow. This is the same behavior found in LSBs using tangential entry air-jet swirlers [5]. Turbulence anisotropy at the exit becomes higher at larger Reynolds numbers (Flows 3–5) where u' is 50% higher than v' for Flow 5. An additional feature is that the u' and v' profiles for the large Reynolds number cases increases with x. The u' profiles of Flows 3–5 attain their maximum values at the point where the U profile drops to its minimum. These increases are consistent with the formation of the weak recirculation zones. Correspondingly, the v'/U_0 of Flows 3–5 continue to increase and surpass u'/U_0 in the far field.

In Fig. 7(*a*), the effects of combustion heat release on the U profiles for Flames 1–5 are readily discernable. In the near field region adjacent to the LSI exit all U profiles retain the linear decay feature but with steeper rates of the velocity decline. Further downstream, the positions where these U profiles deviate from the linear decline correspond to the leading edges of the turbulent flame brushes. Therefore, the local velocity at these points are defined as the turbulent flame speed S_T because the flame brush is locally normal (in the mean) to the approach flow. Within the flame brushes, the mean velocities increase slightly before resuming the declining trends in the postflame regions. The

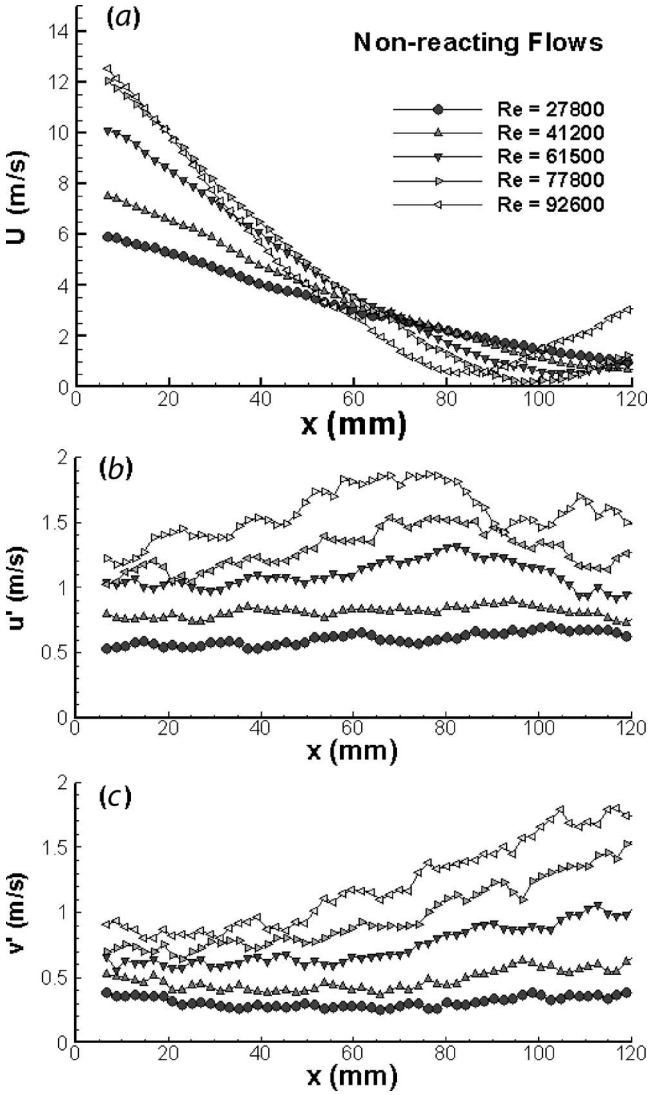


Fig. 6 Axial profiles of the nonreaction flows

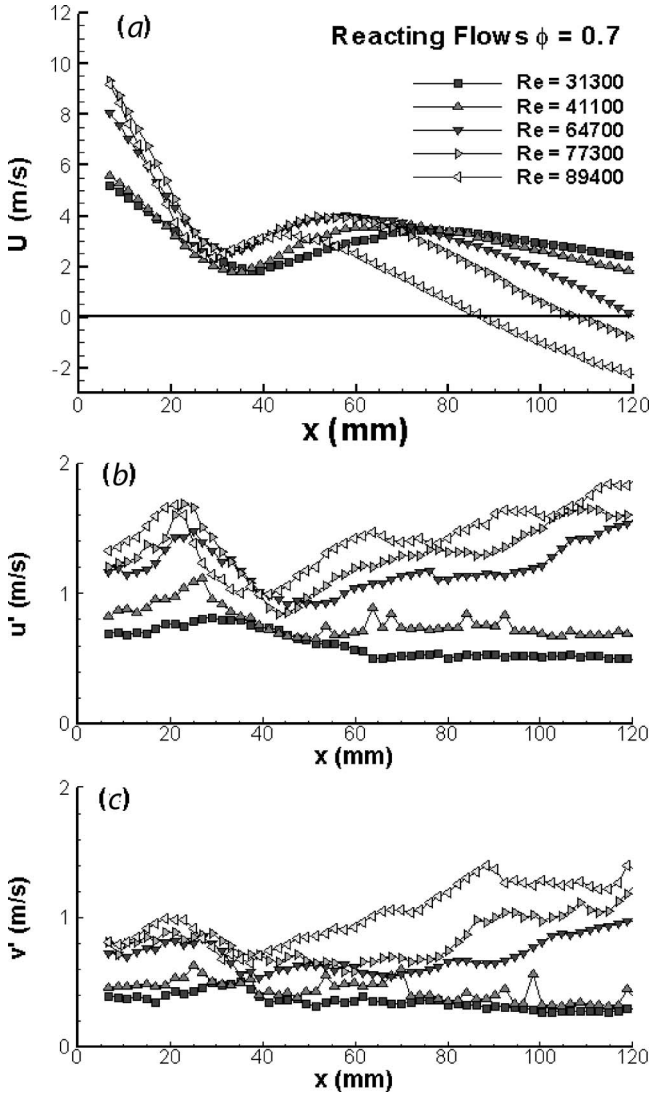


Fig. 7 Axial profiles of the reacting flows

formation of the weak recirculation zones in flames at higher Re is now clear in the U profiles of Flames 4 and 5 where flow reversals occur in the far field ($x>80$ mm). A shift in the position of the forward stagnation point (i.e., where U crosses zero) toward smaller x with increasing Re is consistent with the trend shown by the nonreacting profiles in Fig. 6(a).

The features of the u' and v' profiles in Figs. 7(b) and 7(c) show that the flames affect turbulence fluctuations in the postflame regions as well as in the reactants upstream of the flame brushes. In the near field below the flame brushes ($x<20$ mm), the turbulence fluctuations are anisotropic as in the nonreacting cases. Except for Flame 1 at the lowest Re we have investigated, all the u' and v' profiles rise slightly toward the flame with the increases more pronounced on the u' profiles. This type of flame generated turbulence has been observed in other turbulent premixed flames and is associated with the pressure fluctuations due to the fluctuating wrinkled flame fronts. Within the flame brushes ($20<x<40$ mm), both u' and v' profiles drop to lower levels followed by increases in the postflame region ($x>40$ mm). These are considered small changes as the overall turbulence levels measured in the LSI flames are not significantly different than in the corresponding nonreacting flow.

Turbulent Flame Speed and Similarity Features. The turbulent flame speed S_T is the most significant parameter for the low-swirl combustion method because its flame stabilization mechanism relies on this fundamental flame property. Although the turbulent flame speed has been an important research topic, the results reported in the combustion literature are very scattered and not consistent because of the differences in experimental configurations as well as the methods to extract the turbulent flame speeds from the measurements. However, the definition of the turbulent flame speeds from LSB and LSI is less ambiguous due to its unattached flame brush that is locally normal to the approach flow. The crucial turbulent flame speed for engineering development of the LSB and LSI is the one measured at the centerline because it is where the flame originates.

Our previous laboratory studies of the turbulent flame speed using jet LSBs have resulted in a set of turbulent flame speed data that show a linear correlation between the normalized parameter u'/S_L and S_T/S_L [9]. This linear behavior is quite unlike those found in other systems such as expanding flames inside a combustion chamber or in a Bunsen type conical flame. The flame speed correlations derived for these systems deviate from linearity and level off at high turbulence intensities. Although, this so-called "bending effect" has been a subject of many studies, it does not have a significant implication on their basic operation. However, if this bending effect occurs in low-swirl combustion sys-

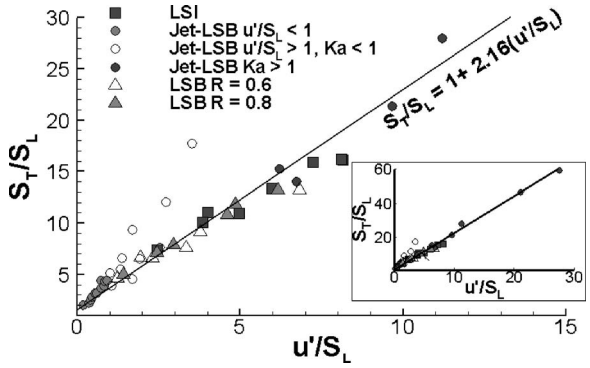


Fig. 8 Turbulent flame speed correlation

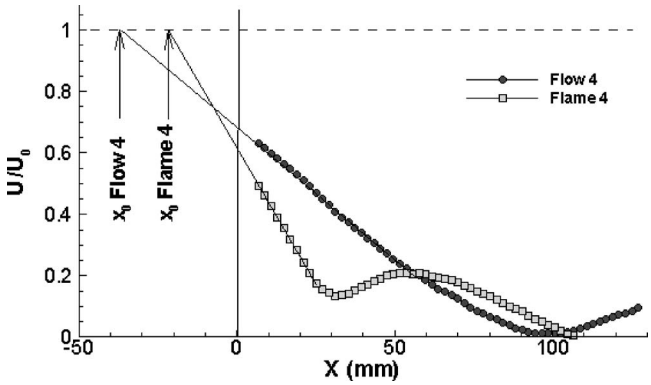


Fig. 9 Definition of virtual origin x_0

tems, it would mean a downstream shift of the LSB flame with increasing Reynolds number leading to eventual blowoff even for flames near stoichiometry.

In Fig. 8, S_T deduced from the PIV data for Flames 1–6 and from the two flames reported by Johnson et al. [12] are compared with the results reported in Ref. [9]. Also shown are data obtained in two 5.08 cm diameter vane LSBs with center channels $R=0.8$ and $R=0.6$. The $R=0.8$ vane LSB has been used in two of our previous studies [10,11]. For the LSB and LSI, S_T can be defined as the velocity normal to the leading edge of the flame brush. From previous investigations using velocity (LDV) and scalar (Mie scattering from oil aerosol) measurements [6], it has been shown that the flame's leading edge corresponds to the position on the centerline U profile where the velocity distribution deviates from its initial linear decay. Therefore, the first inflection point on the axial velocity profiles is a convenient marker for the leading edge position of the flame brush and the velocity at that point is S_T.

The results in Fig. 8 show that despite the variations in burner configuration, range of flow conditions, stoichiometry, and fuels (ethylene, methane, and NG/CO_2 blend), S_T in LSI and LSB has an unequivocal linear correlation with u'. The only exceptions are three outlying points from the jet-LSB. Otherwise, the inclusion of the 20 new data points from the two LSBs and the LSI causes only a small change in the empirical constant for the turbulent flame speed correlation from 2.12 to 2.16. These results clearly show that the two implementations of low-swirl combustion, i.e., tangential injection swirler and vane swirler, are compatible. More importantly, they indicate that the linear behavior of the turbulent flame speed is a characteristic combustion property central to the operation of low-swirl combustion. Therefore, the measurement and correlation of the turbulent flame speed is significant for building a scientific foundation for further engineering development of this combustion technology.

Additionally, the ability to sustain a stable flame strongly depends on the characteristics of the divergent flow and how it evolves with Reynolds number and responds to heat release. To quantify these changes, two parameters have been extracted from the centerline U profile. Figure 9 compares the normalized U/U_0 profiles of Flow 4 and Flame 4 at $\mathrm{Re}\approx 77{,}000$. It can be seen that the flame not only increases the mean axial aerodynamic stretch rate a_x representing the steeper slope of the linear velocity decline region downstream of the LSI exit but also changes the position where the flow divergence originates. The virtual origin x_0 obtained by linear extrapolation of the near field velocity distribution provides a reference point for the divergent flow. Its physical representation is similar to that of the forward stagnation point for a fully developed recirculation bubble in a high-swirl flow.

The values of a and x_0 obtained for Flows 1–5 and Flames 1–5 are shown in Fig. 10. In Fig. 10(a), both sets of data show x_0 decreasing with increasing Reynolds number to indicate a shift of the divergent flow structure into the injector. This trend is analogous to the recirculation bubble of a high-swirl flow moving slightly upstream with increasing Reynolds number [15]. From data obtained in a vane-LSB with $R=0.8$, we found that x_0 eventually become insensitive to Reynolds number and this behavior is also similar to that of a high-swirl recirculation bubble. The leveling trends of x_0 at higher values of Reynolds number in Fig. 10(a) strongly suggest that the structural shift of the divergent flow will eventually cease.

The increases in a_x due to combustion are shown in Fig. 10(b) where the values for the flows are about −0.008 (1/mm) and those for the flames are about −0.017 (1/mm). Both sets of data show that a_x is not sensitive to Reynolds number and gives further evidence that divergent flow structures remain the same despite the changes in their positions. Moreover, the constant a_x values indicate that the mean velocity profiles have a high degree of similarity in the near field. This can be seen in Figs. 11 and 12 when the U/U_0 and q'/U_0 profiles are plotted using x_0 as the origin. In Fig. 11(a) all the nonreacting flow data collapse onto a

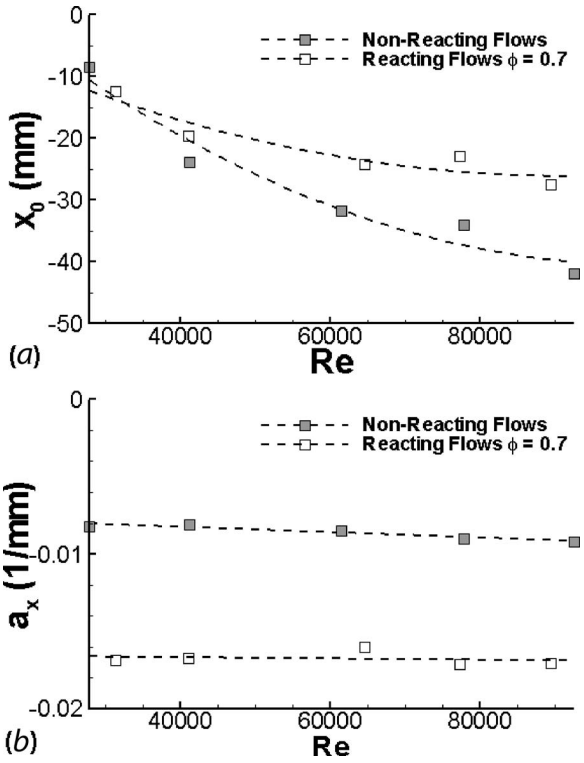


Fig. 10 Virtual origin x_0 and axial stretch rate a_x deduced from the axial velocity profiles

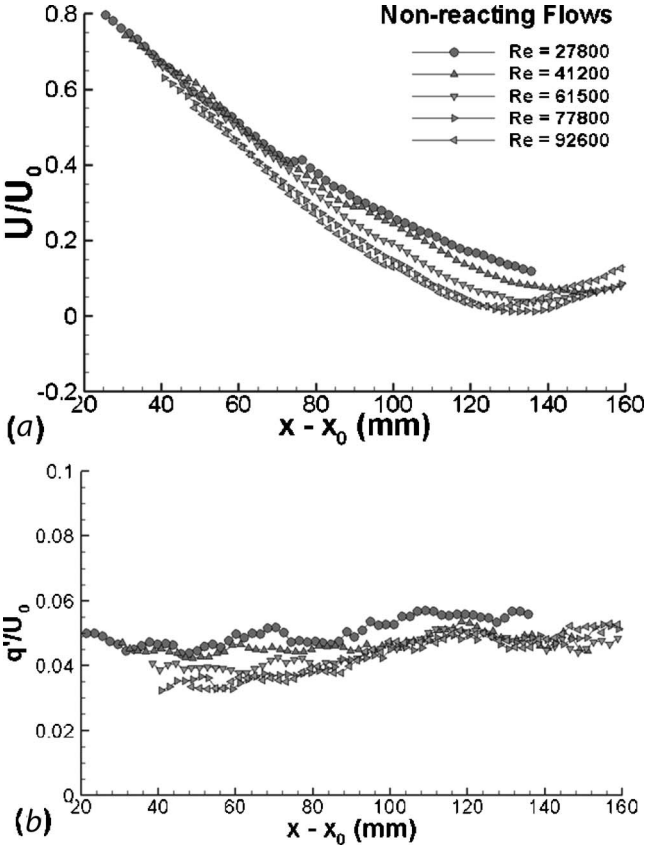


Fig. 11 Normalized nonreacting flow profiles shifted by x_0

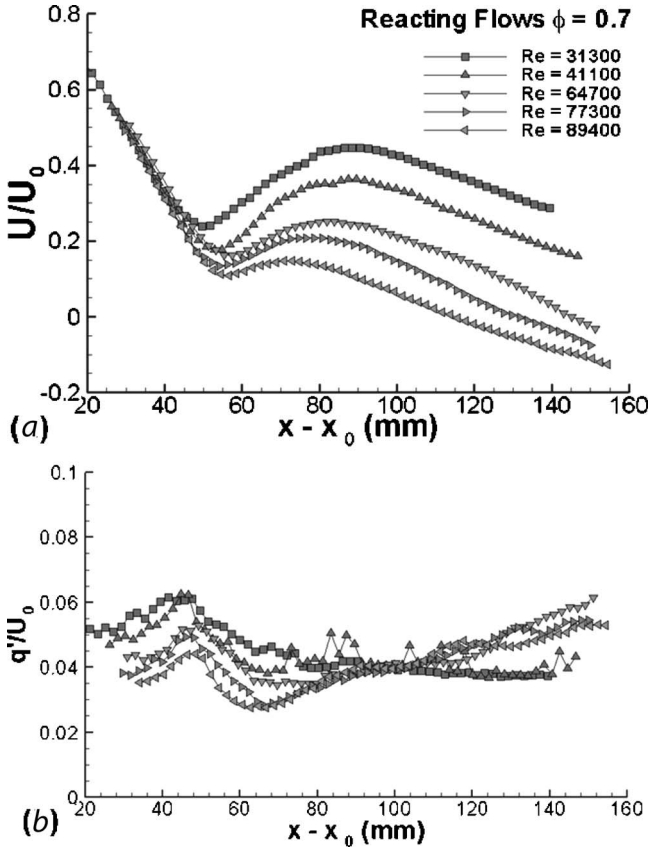


Fig. 12 Axial reacting flow profiles shifted by x_0

consistent trend. The fact that the minimum points on the profiles for the three flows with higher Re coincide at $x-x_0=130$ mm confirms that the overall structure of the divergent flow remains unaltered. In Fig. 11(*b*), the q'/U_0 profiles collapse to a flat distribution to show that the turbulence is nondecaying and scales directly with U_0. The reacting profiles shown in Fig. 12(*a*) illustrate that the similarity behavior prevails in the near field of $x-x_0<40$ mm. Except for Flame 1 at a low Reynolds number, all other flame positions, i.e., the locations where the profiles deviate from linearity, cluster around $x-x_0=50$ mm. This shows again that the divergent flow structures in the reactants upstream of the flame are consistent. In Fig. 12(*b*), q'/U_0 data are more scattered than in the nonreacting flow but their overall trends are not significantly different.

Radial Profiles. The normalized radial profiles obtained at $x=20$ mm for Flows 1–5 and Flames 1–5 are shown, respectively, in Figs. 13 and 14. These positions are within the reactants and the profiles represent the flow features entering the flame brushes. In Fig. 13(*a*), all U/U_0 profiles are characterized by relatively flat velocity distributions generated by the center channel flow at $-20<r<20$ mm. The two higher peaks flanking the flat regions are produced by the swirl annulus. Although the locations of the two peaks and their U/U_0 values at the peaks are the same for all cases, the mean values within the central region decrease with increasing Re. This is associated with the shifting of the overall structure of the divergent flow deeper into the injector. In contrast, the V/U_0 profiles in Fig. 13(*b*) collapse onto a consistent shape to show that the radial velocity component V also exhibits similarity. These V/U_0 profiles have a linear region centered at $r=0$ characteristic of flow divergence. The slope of the linear region indicates a mean radial aerodynamic stretch rate a_r of −0.0047 (1/mm). This is about half of the axial stretch rate a_x of −0.008 (1/mm) shown in Fig. 10(*b*). This ratio is in accord with those of other divergent flows such as that produced by flow impingement on a stagnation plate. The q'/U_0 profiles in Fig. 13(*c*) are similar and show very high turbulent kinetic energies in the outer regions when the swirling flow interacts with the ambient air.

For Flames 1–5, the features of the U/U_0 profiles of Fig. 14(*a*) are the same as those of the nonreacting flows of Fig. 13(*a*). The quantitative differences are a slight enlargement of the central flat region accompanied by an overall lowering of the velocity levels, and the two peaks shifting slightly outward. Similarity features are also found on the V/U_0 profiles in Fig. 14(*b*). However, they have a larger slope within the center region compared to the nonreacting cases. The mean radial aerodynamic stretch rate a_r deduced from the linear region at the center is −0.0088 (1/mm). This value is again about half of the axial aerodynamic stretch rate of −0.017 (1/mm) showing that the overall features of the near field divergent flows are not altered by combustion heat release. In Fig. 14(*c*), the q'/U_0 profiles show center regions where the turbulence intensity remains uniform. It is due to this uniformity and to the absence of steep mean velocity gradients with high shear stresses that the LSI can sustain the weaker flames at very lean conditions.

Flame With Low Wobbe Index Fuel. Flame 6 utilized a fuel mixture of 50% NG and 50% CO_2 to simulate the operation of the LSI with medium heating value landfill gas. As discussed above, at $\phi=0.8$ the laminar flame speed of Flame 6 is lower than that of with the NG Flames 1–5 at $\phi=0.7$. However, the adiabatic flame temperatures of the two mixtures are the same. The flow conditions of Flame 6 are similar to Flame 1 with Re≈31,000 and the centerline profiles from the two flames are compared in Fig. 15. Both U/U_0 profiles in Fig. 15(*a*) have the same characteristic feature. However, the locations of the points where they deviate from the initial linear decline show that the flame brush of Flame 6 shifted slightly downstream compared to Flame 1. The turbulent

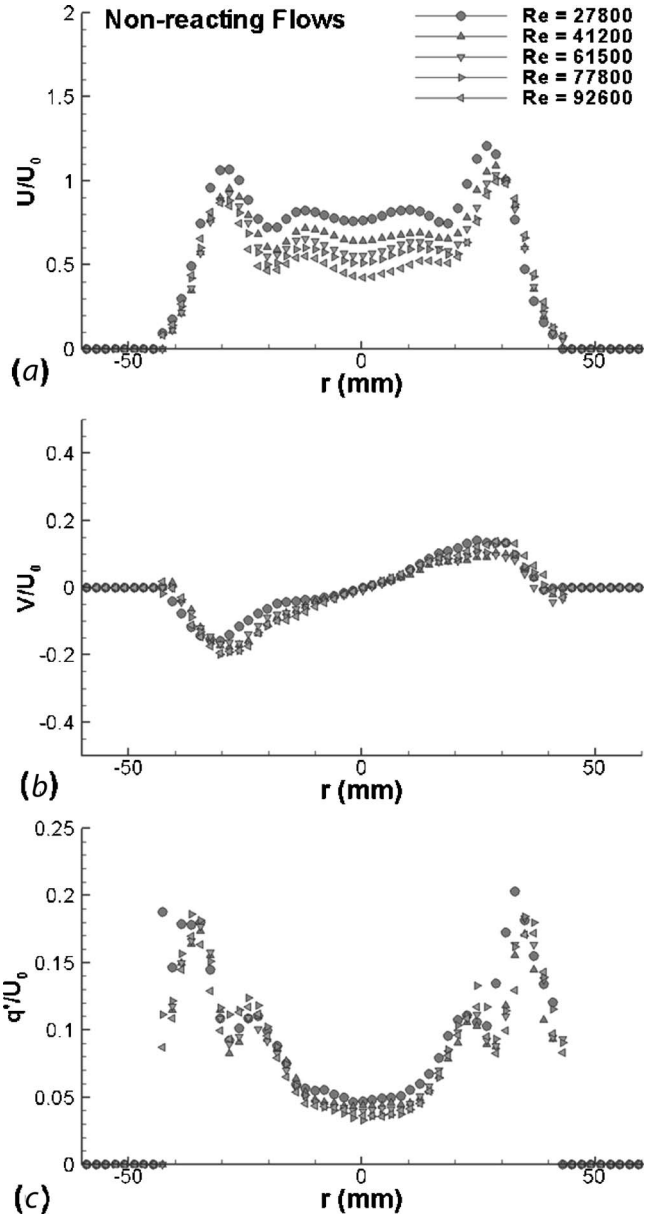


Fig. 13 Nonreacting radial profiles at $x=20$ mm

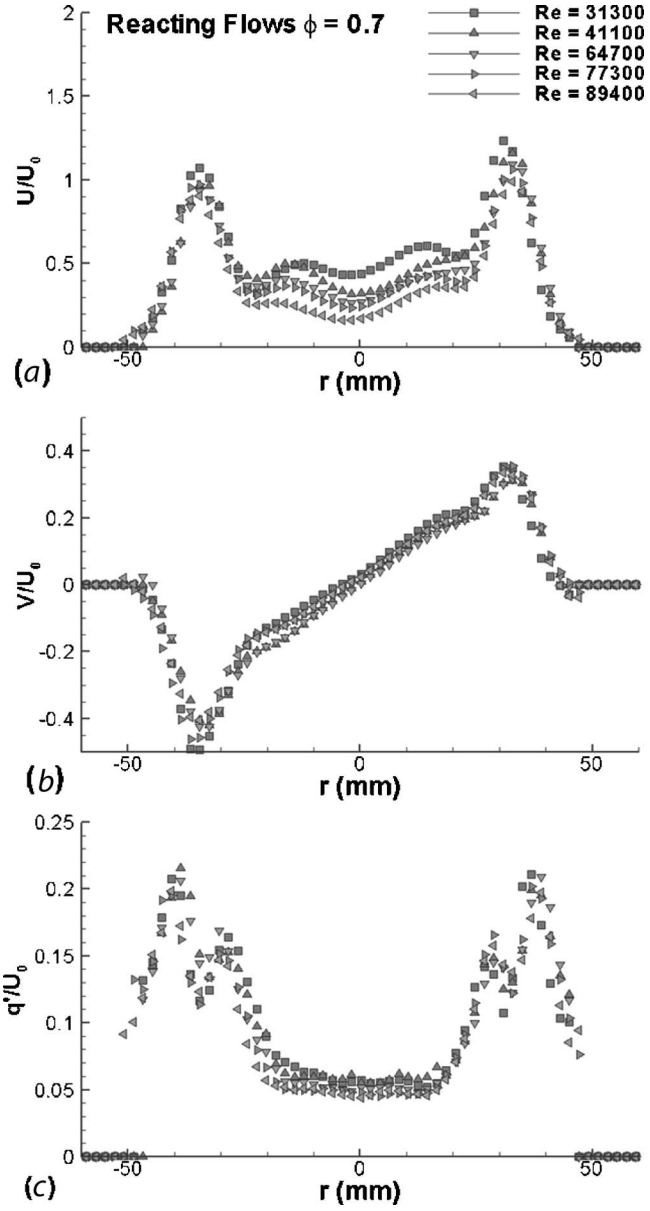


Fig. 14 Reacting radial profiles at $x=20$ mm

flame speed as indicated by the local minima is also lower.

The difference in the centerline U/U_0 profiles of Flames 1 and 6 gives a clear demonstration of the self-adjusting flame stabilization mechanism provided by the LSI. The lower laminar speed S_L for the Flame 6 mixture means that the flame has a lower turbulent flame speed when subjected to the same turbulence intensity as for Flame 1. The divergent flowfield allows the flame to shift to a different position downstream where the local flow velocity matches the lower turbulent flame speed. This demonstrates that low-swirl combustion provides a robust mechanism that is dependent primarily on the flame speed of the mixture and the characteristics of the divergent flowfield. The mechanism is not highly sensitive to the fuel constituents and compositions provided that the premixed turbulent flame behaves as a propagating wave without significant local quenching. Therefore, knowledge of the laminar flame speed and better understanding of how the divergence flow characteristics vary with swirler configurations will provide the basic knowledge for engineering the LSI to accept fuels with both high and low heating values.

The q'/U_0 profiles of Fig. 15(*b*) are shown for completeness as there are no significant differences between the two except for the fact that Flame 6 has a slightly lower values of q'/U_0 at the leading edge of the flame brush. However, from Fig. 8 the S_T and u' deduced for Flame 6 are in excellent agreement with those of the ethylene, methane and NG flames. Therefore, the turbulent flame speed correlation equation should also be a useful empirical input for fuel-flexible LSI designs.

Discussion

The features of the LSI near field shown by the PIV data exhibit a high degree of similarity that can explain why the flame remains stationary regardless of the flow velocity U_0. We start by invoking an equality for the axial velocity at the leading edge position of the flame brush x_f,

$$U_0 - \frac{dU}{dx}(x_f - x_0) = S_T \tag{2}$$

where x_0 is the virtual origin of the flow divergence as defined in Fig. 9. Equation (2) simply states that the local velocity at x_f within a divergent flow with U_0 and an axial stretch rate of dU/dx

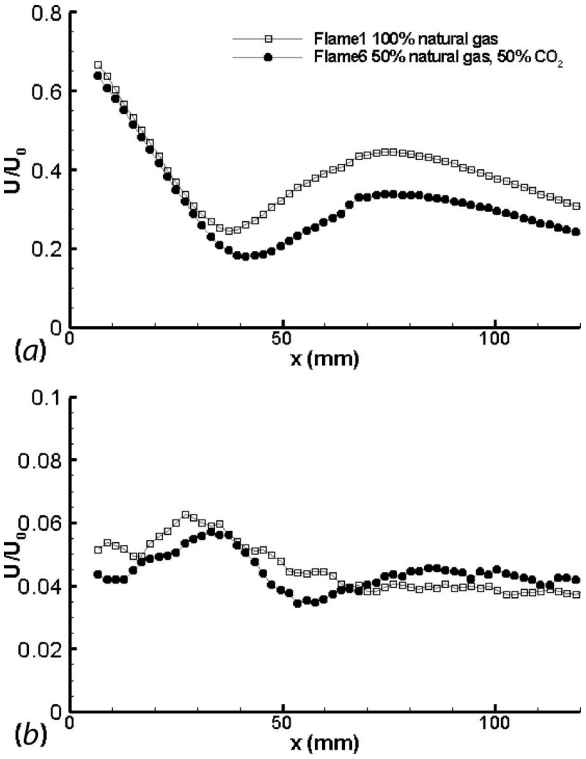


Fig. 15 Axial profiles of Flames 1 and 6

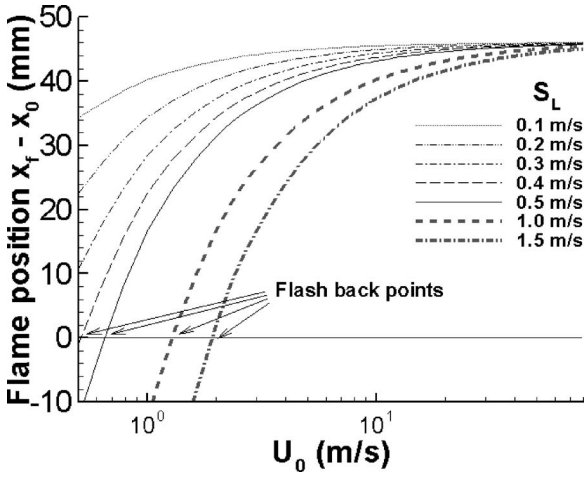


Fig. 16 Flame positions estimated for a range of S_L using the empirical parameters from experimental measurements

is equal to the turbulent flame speed. From Fig. 8, we showed that S_T of the LSB and LSI increases linearly with the turbulence fluctuation u' such that $S_T = S_L + 2.16u'$. Substituting this correlation into Eq. (2) and dividing both sides by U_0 results in

$$1 - \frac{dU}{dx}\frac{(x_f - x_0)}{U_0} = \frac{S_T}{U_0} = \frac{S_L}{U_0} + \frac{2.16u'}{U_0} \tag{3}$$

On the LHS of Eq. (3), $dU/dx/U_0$ is the normalized axial divergence rate a_x. As shown in Fig. 10, it has a constant value of −0.017 (1/mm) due to similarity. On the far RHS of Eq. (3), $S_L/U_0 + 2.16u'/U_0$—the contributions from the first term becomes small for large U_0 because values of S_L for typical gaseous fuels are from 0.1 m/s to 1.5 m/s. The second term on the RHS is the dominating term that is a constant because u' is controlled by the perforated plate and u' increases linearly with U_0 according to classic theory on turbulence production. For the current data set Flames 1–5, S_L is constant. Equation (3) indicates that changes in their flame positions $x_f - x_0$ become progressively smaller when U_0 becomes large. This is clearly shown on Fig. 12(*a*) where except for Flame 1 at low Re, all other values of $x_f - x_0$ are at 50 mm.

In Fig. 16, Eq. (3) is plotted using empirical values of $dU/dx/U_0 = a_x = -0.017$ (1/mm), $u'/U_0 = 0.1$ for a range of S_L representing CH_4/air flames of $0.6 < \phi < 1.0$ ($0.1 < S_L < 0.5$ m/s) and H_2/air flames near stoichiometry ($1.0 < S_L < 1.5$ m/s). It can be seen that significant changes in flame position occur at $U_0 < 10$ m/s. For $U_0 > 10$ m/s, all flame positions converge. This expression explains why the flame remains stationary at typical gas turbine flow velocities of up to $U_0 = 80$ m/s. Equation (3) also shows that the flashback occurs at velocities where the flame positions decrease rapidly to zero. This information will provide the combustion engineers with a useful means to estimate flashback and determine the minimum operating condition.

This analysis shows that laboratory experiment can provide valuable insights to understand and explain the observations of the LSI flame behavior at gas turbine conditions. Therefore, the development of fuel-flexible LSIs will greatly benefit by further laboratory investigations so that the LSI can be modified to accommodate the variations in flame speeds and combustion properties. This should be a very cost-effective process because it does not need extensive and elaborate developmental steps involving computational fluid dynamics. Our next step will be to continue with studying the current LSI using high and low heating value fuels. The results will provide the information to further refine the turbulent flame speed correlation to include the data from alternate fuels and to quantify the relationships between the changes in divergent flowfield characteristics, flame position, lean blowoff, and emissions with the various combustion properties. From these preliminary results, it will be possible to gain the insights needed to vary the LSI design for burning a range of a variety of fuels with a range of heating values.

Conclusions

A laboratory investigation has been conducted to characterize the nonreacting and reacting flowfields produced by a preproduction ultralow emission LSI. The objective is to quantify and analyze the flow and flame structures to obtain the scientific underpinning for guiding the engineering development of the LSI for fuel-flexible turbines. Using PIV, the distributions of the mean and turbulent fluctuation velocities have been measured in NG flames ($\phi = 0.7$) at five bulk flow velocities from 6.8 m/s to 23 m/s. These conditions simulate the partial load conditions of a typical gas turbine within a 3:1 turndown range. Also investigated are the corresponding nonreacting flowfields. The feasibility of operating this LSI with low heating value gases was also demonstrated by firing with a simulated land fill gas consisting of 50% NG and 50% CO_2.

The mean velocity vectors show that the overall structures of the nonreacting and reacting flowfields are not highly sensitive to changes in the flow Reynolds number Re. In the central near field region, all flowfields exhibit the characteristic features of divergent flows with the flow expanding radially and the axial velocity declining linearly with increasing x. With combustion heat release, the discharge angle of the expanding flow increases slightly but remains relatively invariant with Re. In the far field, the development of a weak recirculation zone is shown by the appearance of velocity troughs in the nonreacting flows and slight flow reversals in the reacting flows.

The axial profiles were analyzed to determine the flame and flow parameters significant to the flame stabilization mechanism. The turbulent flame speeds S_T obtained at the centerline of the NG and simulated landfill gas flames are in excellent agreement with the ethylene/air and methane/air flames from a LSB with a tan-

gential air swirler and from two other vane LSBs. Collectively, all the data show an unequivocal linear $S_T = S_L(1+2.16u')$ correlation despite the variations in burner configuration, range of flow conditions, stoichiometry, and fuels. This linear behavior is a unique combustion characteristic of the LSI.

The axial mean aerodynamic stretch rates for the reacting and nonreacting flows are not sensitive to Reynolds number. However, the presence of the flame increases the axial stretch rate. The virtual origins of the flow divergence were also extrapolated from the axial profiles and show a consistent shift of the divergent flow structure deeper within the LSI throat as Reynolds number increases. However, the leveling trends of the virtual origins at higher values of Reynolds number strongly suggest that the flow structure shift will eventually cease.

The similarity features of the reacting flows and the nonreacting flows were shown clearly by the normalized axial and radial profiles. The fact that both the radial mean aerodynamic stretch rates are about half those of the corresponding axial mean aerodynamic stretch rates indicates that the basic divergent flow structures are not altered by combustion heat release.

The features of the LSI flowfield obtained by the PIV measurements exhibit a high degree of similarity that explains why the flame remains stationary despite changes in flow velocity. By invoking an equality for the axial velocity at the leading edge position of the flame brush, an expression is obtained in terms of the virtual origin, the normalized mean axial aerodynamic stretch rate, and the turbulent flame speed correlation. Due to the similarity behavior of the flowfields and the linear dependence of S_T on turbulence, the expression shows that under typical gas turbine conditions of $U_0 > 10$ m/s, the flame position is not sensitive to changes in the laminar flame speed associated with changing stoichiometry or fuel type.

This analysis shows that laboratory experiments can provide valuable insights to understand and explain the operation of the LSI at gas turbine conditions. Therefore, the development of fuel-flexible LSIs will greatly benefit by further laboratory investigations so that the LSI can be modified to accommodate the variations in flame speeds and combustion properties.

Acknowledgment

This work was supported by the US Department of Energy, Energy Efficiency and Renewable Energy, Distributed Energy Resources under Contract No. DE-AC02-05CH11231.

Nomenclature

a = normalized flow divergence (1/mm)
L_i = swirler recess distance
$m = m_c/m_s$ = mass flux ratio
m_c = mass flux through center channel
m_s = mass flux through swirl annulus
q' = 2D turbulent kinetic energy $= 1/2(u'^2+v'^2)^{1/2}$
Re = Reynolds number $2R_iU_0/\nu$
$R = R_c/R_i$ = ratio of the center channel radius R_c to injector radius R_i
r = radial distance
S = swirl number
S_L = laminar flame speed
S_T = turbulent flame speed
T_{ad} = adiabatic flame temperature
U_0 = bulk flow velocity
U = axial velocity
u' = axial rms velocity
v = radial rms velocity
uv = shear stress
x = axial distance from injector exit
x_f = leading edge position of the flame brush
x_0 = virtual origin of divergent flow

References

[1] Richards, G. A., Straub, D. L., and Robey, E. H., 2003, "Passive Control of Combustion Dynamics in Stationary Gas Turbines," J. Propul. Power, **19**(5), pp. 795–810.
[2] Mongia, H. C., Held, T. J., Hsiao, G. C., and Pandalai, R. P., 2003, "Challenges and Progress in Controlling Dynamics in Gas Turbine Combustors," J. Propul. Power, **19**(5), pp. 822–829.
[3] Muruganandam, T. M., Nair, S., Scarborough, D., Neumeier, Y., Jagoda, J., Lieuwen, T., Seitzman, J., and Zinn, B., 2005, "Active Control of Lean Blowout for Turbine Engine Combustors," J. Propul. Power, **21**(5), pp. 807–814.
[4] Chan, C. K., Lau, K. S., Chin, W. K., and Cheng, R. K., 1992, "Freely Propagating Open Premixed Turbulent Flames Stabilized by Swirl," Sym. (Int.) Combust., [Proc.], **24**, pp. 511–518.
[5] Bedat, B., and Cheng, R. K., 1995, "Experimental Study of Premixed Flames in Intense Isotropic Turbulence," Combust. Flame, **100**(3), pp. 485–494.
[6] Cheng, R. K., 1995, "Velocity and Scalar Characteristics of Premixed Turbulent Flames Stabilized By Weak Swirl," Combust. Flame, **101**(1–2), pp. 1–14.
[7] Plessing, T., Kortschik, C., Mansour, M. S., Peters, N., and Cheng, R. K., 2000, "Measurement of the Turbulent Burning Velocity and the Structure of Premixed Flames on a Low Swirl Burner," Proc. Combust. Inst., **28**, pp. 359–366.
[8] Shepherd, I. G., and Cheng, R. K., 2001, "The Burning Rate of Premixed Flames in Moderate and Intense Turbulence," Combust. Flame, **127**(3), pp. 2066–2075.
[9] Cheng, R. K., Shepherd, I. G., Bedat, B., and Talbot, L., 2002, "Premixed Turbulent Flame Structures in Moderate and Intense Isotropic Turbulence," Combust. Sci. Technol., **174**(1), pp. 29–59.
[10] Cheng, R. K., Yegian, D. T., Miyasato, M. M., Samuelsen, G. S., Pellizzari, R., Loftus, P., and Benson, C., 2000, "Scaling and Development of Low-Swirl Burners for Low-Emission Furnaces and Boilers," Proc. Combust. Inst., **28**, pp. 1305–1313.
[11] Littlejohn, D., Majeski, M. J., Tonse, S., Castaldini, C., and Cheng, R. K., 2002, "Laboratory Investigation of an Untralow NOx Premixed Combustion Concept for Industrial Boilers," Proc. Combust. Inst., **29**, pp. 1115–1121.
[12] Johnson, M. R., Littlejohn, D., Nazeer, W. A., Smith, K. O., and Cheng, R. K., 2005, "A Comparison of the Flowfields and Emissions of High-swirl Injectors and Low-swirl Injectors for Lean Premixed Gas Turbines," Proc. Combust. Inst., **30**, pp. 2867–2874.
[13] Nazeer, W. A., Smith, K. O., Sheppard, P., Cheng, R. K., and Littlejohn, D., 2006, "Full Scale Testing of a Low Swirl Fuel Injector Concept for Ultra-Low NOx Gas Turbine Combustion Systems," Paper No. GT2006-90150.
[14] Littlejohn, D., and Cheng, R. K., 2006, "Fuel Effects on a Low-swirl Injector for Lean Premixed Gas Turbines," Proc. Combust. Inst., **31**, pp. 3155–3162.
[15] Beer, J. M., and Chigier, N. A., 1972, *Combustion Aerodynamics*, Applied Science, London.
[16] Mellings, A., 1997, "Tracer Particles and Seeding for Particle Image Velocimetry," Meas. Sci. Technol., **8**, pp. 1406–1416.
[17] Wernet, M. P., 1999, "Fuzzy Logic Enhanced Digital PIV Processing Software," *18th International Congress on Instrumentation for Aerospace Simulation Facilities*, Toulouse, France.
[18] Yegian, D. T., and Cheng, R. K., 1998, "Development of a Lean Premixed Low-Swirl Burner for Low NOx Practical Applications," Combust. Sci. Technol., **139**(1–6), pp. 207–227.

Development and Application of an Eight-Step Global Mechanism for CFD and CRN Simulations of Lean-Premixed Combustors

Igor V. Novosselov

Philip C. Malte[1]

Energy and Environmental Combustion Laboratory,
Department of Mechanical Engineering,
University of Washington,
P.O. Box 352600,
Seattle, WA 98195-2600
e-mail: malte@u.washington.edu

In this paper, the development of an eight-step global chemical kinetic mechanism for methane oxidation with nitric oxide formation in lean-premixed combustion at elevated pressures is described and applied. In particular, the mechanism has been developed for use in computational fluid dynamics and chemical reactor network simulations of combustion in lean-premixed gas turbine engines. Special attention is focused on the ability of the mechanism to predict NO_x and CO exhaust emissions. Applications of the eight-step mechanism are reported in the paper, all for high-pressure, lean-premixed, methane-air (or natural gas-air) combustion. The eight steps of the mechanism are as follows: (1) oxidation of the methane fuel to CO and H_2O, (2) oxidation of the CO to CO_2, (3) dissociation of the CO_2 to CO, (4) flame-NO formation by the Zeldovich and nitrous oxide mechanisms, (5) flame-NO formation by the prompt and NNH mechanisms, (6) postflame-NO formation by equilibrium H-atom attack on equilibrium N_2O, (7) postflame-NO formation by equilibrium O-atom attack on equilibrium N_2O, and (8) post-flame Zeldovich NO formation by equilibrium O-atom attack on N_2.
[DOI: 10.1115/1.2795787]

Introduction

Computational fluid dynamics (CFD) modeling of lean-premixed gas turbine combustors has relied on the use of global chemical kinetic mechanisms for the prediction of heat release. This includes the use of mechanisms of one to four steps drawn from the combustion literature, e.g., Refs. [1–4]. Most of the mechanisms do not include a reverse step for carbon monoxide; thus, it is difficult to obtain convergence of CO in the burnout zone of the combustor, and the ability to predict CO exhaust emission is lost. Furthermore, the global mechanisms do not include nitric oxide (NO) kinetics. Although CFD packages include post-processing for NO, especially for thermal Zeldovich NO, the details of flame-formed NO are lacking. Flame-formed NO lies at the heart of NO_x and its control for advanced gas turbine engines with single-digit emissions.

Our goal has been to develop a global mechanism that permits the prediction of ppm emissions of CO and NO_x as well as provides reasonable results for heat release and patterns for temperature and major species in the lean-premixed combustor. Focus is on the industrial combustor. Thus, the pressure range of our mechanism is 5–20 atm, and the inlet air temperature is set accordingly. The mechanism is valid for fuel-air equivalence ratios (phi) from about 0.45 to 0.75 and for mean residence times from slightly greater than blowout to full combustor time. The fuel is methane or natural gas sufficiently high in methane so that it can be reasonably simulated as methane.

The eight-step mechanism builds on a five-step mechanism we previously developed [5]. The five-step mechanism used two NO steps, one for the flame zone and one for the postflame zone, and was restricted to single pressures, whereas the eight-step mechanism has a variable pressure and five NO steps. Also, the database used for the development of the five-step mechanism was smaller than that used for the eight-step mechanism.

The balance of the paper is divided into the following sections:

- discussion of the key points about NO formation in lean-premixed combustion, which uses results from our CFD modeling of a generic can-type gas turbine combustor;
- development of the eight-step mechanism, including a discussion of the database used for obtaining the mechanism, and a listing and explanation of the mechanism;
- application of the eight-step mechanism in CFD modeling of an experimental bluff body combustor; and
- application of the eight-step mechanism in chemical reactor network (CRN) modeling of a test rig gas turbine combustor.

NO Formation in Lean-Premixed Combustion

Prior to discussing the development of the eight-step mechanism, it is helpful to briefly review the formation of NO in lean-premixed combustion. (Please note that the terms "NO formed" and "NO_x emitted" can be used interchangeably since a fraction of the NO formed is oxidized to NO_2 within the combustor and engine. Thus, the computation of NO formed is equivalent to NO_x emitted.)

For this discussion, a generic, can-type, swirl-stabilized, air-back-sided cooled, lean-premixed combustor for the industrial gas turbine engine is assumed. A commercial CFD package (FLUENT 6.2) is used to solve a 2D, axisymmetric rendition of the combustor. Details are given in the Ph.D. thesis of Novosselov [6]. The upper, forward part of the combustor is shown in the CFD results pictured below in Figs. 1(*A*)–1(*H*). The eight-step mechanism provides the global reactions and rates. For each reaction step, the slower of two rate choices is selected by the CFD: either the chemical kinetic rate (as provided by the eight-step mechanism) or the mixing controlled rate as calculated by the eddy breakup model [7]. In this case, since the combustion is premixed, the reactant-product eddy breakup rate, rather than the fuel-air rate, is

[1]Corresponding author.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 26, 2007; final manuscript received June 21, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007, Paper No. GT2007-27990.

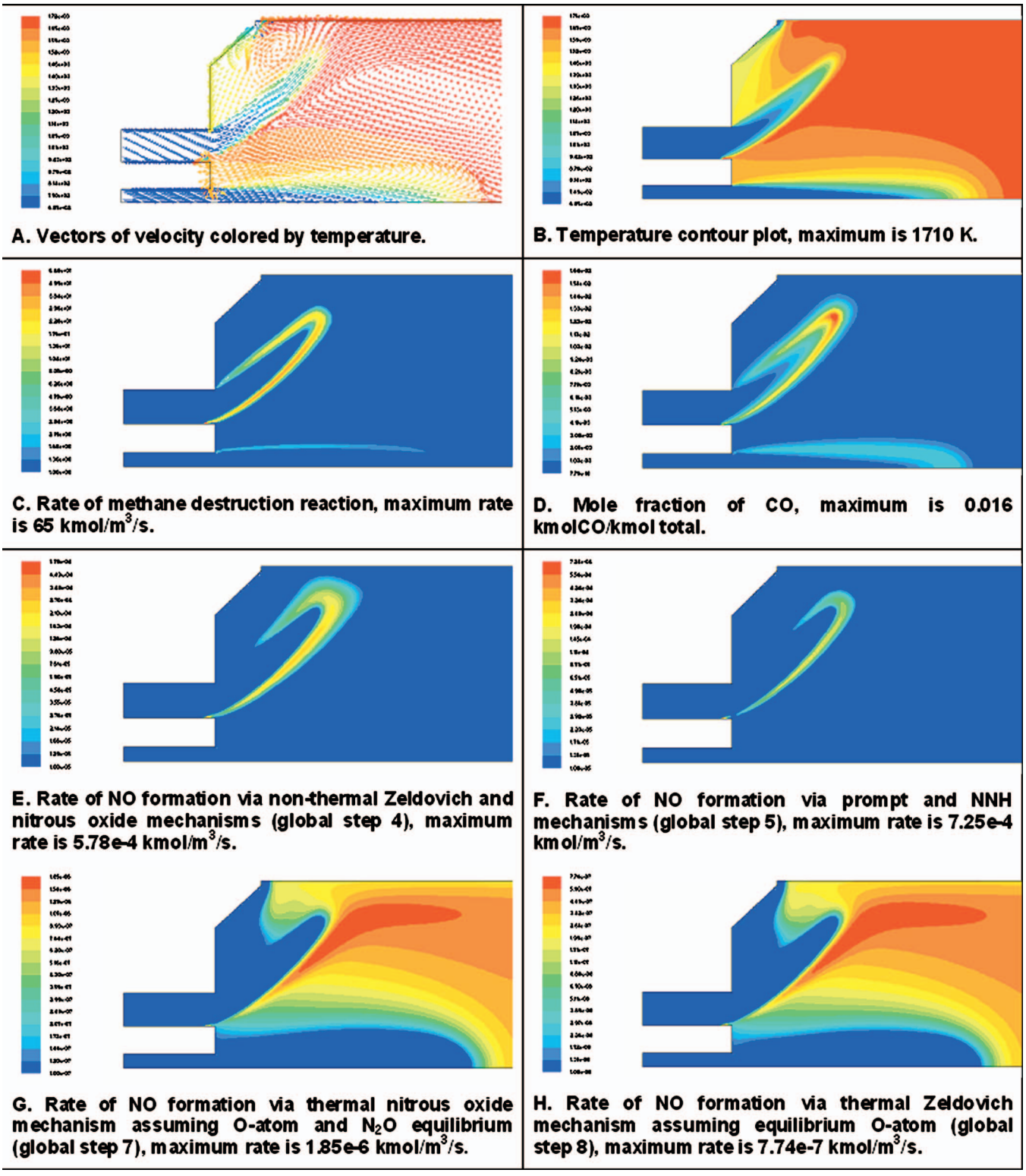


Fig. 1 CFD solutions for the generic, can-type, lean-premixed, gas turbine combustor. Upper, forward section of combustor shown. Flow is from left to right. Injectors are at the left side. Pilot injector is shown on the centerline. Weak pilot case: 50% of neutral pilot. The main premixer injector is located above the centerline.

selected. The first reaction step is significantly affected by this choice; that is, the initial oxidation of the methane to water vapor and CO tends to be mixing controlled in these combustors, except in regions of intense shear. The oxidation of CO can experience either kinetic or mixing control, though the tendency is kinetic control. NO formation is kinetically controlled.

Figure 1(*A*) shows the velocity vectors, and Fig. 1(*B*) shows the temperature contours of the combustor. Also seen is the flame structure (Figs. 1(*C*) and 1(*D*)). Swirled air-fuel mixture exits the main premixer injector shown above the centerline of the combustor. The premixer injector is also solved in CFD. For the results plotted in Figs. 1(*A*)–1(*H*), a uniform fuel-air ratio is assumed

Table 1 NO mechanisms and formation rates for lean-premixed combustion

Zeldovich mechanism	
Rxn 1	$N_2+O \Rightarrow NO+N$
Rxn 2	$N+O_2 \Rightarrow NO+O$
Rxn 3	$N+OH \Rightarrow NO+H$
Rate	$d[NO]/dt=2k_1[N_2][O]$
Nitrous oxide mechanism	
Rxn 4	$N_2+O+M \Rightarrow N_2O+M$
Rxn 5	$N_2O+O \Rightarrow NO+NO$
Rxn 6	$N_2O+H \Rightarrow NO+NH$
Note	Under a lean-premixed combustion, quantitative oxidation of NH to NO is assumed.
Note	Several reactions [6], not shown here, convert N_2O back to N_2.
Rate	$d[NO]/dt=2k_5[N_2O][O]+2k_6[N_2O][H]$
Prompt NO	
Rxn 7	$N_2+CH \Rightarrow HCN+N$
Note	Under a lean-premixed combustion, quantitative oxidation of HCN and N to NO is assumed.
Note	N reacts to NO by Rxn's 2 and 3.
Rate	$d[NO]/dt=2k_7[N_2][CH]$
NNH mechanism	
Rxn 8	$N_2+H \Rightarrow NNH+H$
Rxn 9	$N_2+H+M \Rightarrow NNH+M$
Rxn 10	$NNH+O \Rightarrow NO+NH$
Rxn 11	$NNH+O \Rightarrow N_2O+OH$
Note	As noted above, the oxidation of NH to NO is assumed quantitative
Note	Several reactions [6], not shown here, convert NNH back to N_2.
Note	The NO rate for the NNH chemistry does not include Rxn 11.
Rate	$d[NO]/dt=2k_{10}[NNH][O]$

across the premixer. Centrifugal force pulls the premixed stream radially outward after it leaves the premixer, and the mixture is ignited by mixing with the recirculation zones: mainly with the large on-axis recirculation zone but also by the smaller dome recirculation zone. The methane destruction rate (Fig. 1(*C*)) and the CO concentration (Fig. 1(*D*)) indicate the location of the flame zone; this is the main region of nonequilibrium chemical reaction. This combustor operates very lean: the phi of the main premixer injector is 0.475, and that of the centerline pilot (which is premixed) is taken as $\frac{1}{2}$ of this (i.e., 50% pilot is assumed in this case). This results in a relatively low peak temperature: 1710 K. The pressure is 16 atm.

Figures 1(*E*)–1(*H*) show how and where the NO forms. Under the conditions represented, about 50% of the NO forms via the Zeldovich and nitrous oxide mechanisms within the flame zone (Fig. 1(*E*)) and about 25% forms in the flame zone by the prompt and NNH mechanisms (Fig. 1(*F*)). The mechanisms are listed in Table 1. The lumping of mechanisms together is done on the basis of where they are active in the flame zone: the prompt-NNH NO forms earlier in the flame zone than the Zeldovich-nitrous oxide NO. Furthermore, the prompt-NNH NO has the greatest maximum rate. However, since its region of impact is not as large as the Zeldovich-nitrous oxide NO, its contribution is about one-half that of the Zeldovich-nitrous oxide NO. The NO formed within the flame zone is termed nonthermal in order to distinguish it from thermal NO formed in the postflame zone, which assumes that the free radical species such as O atom, H atom, and OH radical are at a local thermochemical equilibrium condition. In the flame zone, however, the free radicals are at significantly greater concentrations, termed nonequilibrium or superequilibrium. This drives the rates of NO formation to high levels within the flame zone. Thus, it is important to maintain the flame zone as thin as practically feasible in order to curtail NO formation in a single-digit emission lean-premixed combustor. Globalizing the Zeldovich and nitrous oxide reactions into one step for the flame zone and the prompt and NNH reactions into another step is shown in the next section.

The NO formation is brought to 100% by noting that 25% is formed by the thermal steps in the postflame zone of the combustor (Figs. 1(*G*) and 1(*H*)). Although the maximum rates of thermal-NO formation are two to three orders of magnitude less than the maximum rates of NO formation within the flame zone, since the postflame expanse of the thermal-NO formation is large relative to the flame zone, its contribution sums up to about 25%. Figure 1(*H*) shows the thermal Zeldovich NO, which is the textbook method of calculating NO formation, e.g., see Ref. [8]. However, under the present conditions, there is a second essentially equal contributor to postflame thermal NO: thermal nitrous oxide NO (Fig. 1(*G*)) which assumes that N_2O and O, both at local equilibrium concentration, react to form NO.

Development and Explanation of the Eight-Step Mechanism

The database used to generate the eight-step mechanism began with two high-pressure jet stirred reactor (JSR) experiments, one at the University of Washington by Rutar-Shuman [9] for pressures up to 6.5 atm and the other at the Swiss Federal Institute of Technology by Bengtsson [10] for pressures up to 20 atm. The fuel in each study was methane. Rutar and Malte [11] compared both sets of experiments to simple CRN modeling, assuming two perfectly stirred reactors (PSRs) in series for the JSR of Rutar and a PSR followed by a plug flow reactor (PFR) for the JSR of Bengtsson. The GRI 3.0 mechanism [12] was used for both JSRs. The volume of PSR1 relative to that of PSR 2 of the Rutar JSR depended on the experimental Damkölher number. For the Bengtsson JSR, the PSR and PFR volumes were fixed as 88% and 12%, respectively, of the total reactor volume. Experimental temperatures were used. The modeling results were found to compare very favorably with the experimental results for CO, NO_x, and N_2O. The modeling also provided estimates of the free radical concentrations within the reactors and permitted the measured NO_x to be interpreted in terms of the four contributing mechanisms (Table 1).

Novosselov [13] extended the database by running a wide range of lean-premixed combustion conditions, all for methane fuel, using CRNs (i.e., CRN modeling). The pressure was selected from 5 atm to 20 atm. The inlet air temperature was set assuming a compression from 1 atm and 15°C to the pressure of interest through a compressor of 85% efficiency. The chemical reactors and residence times used are catalogued in Table 2. This process yielded a large computer-generated database.

From these CRN computer runs, the results of value for developing the eight-step global mechanism are the following:

- temperature: T (K),
- concentrations (kmol/m^3) of major species and OH: CH_4, CO, CO_2, O_2, H_2O, OH, and N_2,
- concentrations (kmol/m^3) of species involved in NO formation (see Table 1): CH, N_2O, NNH, O, and H, and
- rates (kmol/m^3 s) of CH_4 and CO oxidation, and dissociation of CO_2.

The elemental rate data for Rxn's 1, 5, 6, 7, and 10 in Table 1 are also required. The data are taken from GRI 3.0. Although Rxn 11 has recently received attention in the literature, e.g., Ref. [14], it is not included in this development since it is not part of GRI 3.0.

Step 1 in the eight-step mechanism is the oxidation of methane.

Table 2 CRN schemes and residence times used in developing the eight-step mechanism

Reactor schemes (adiabatic)	CH_4 oxidation	CO oxidation	CO_2 dissociation	Flame NO by Zeldovich & N_2O	Flame NO by prompt & NNH
PSR: blowout to 3 m	X	X	X	X	X
PSR at blowout+PSR to 3 ms	X	X		X	X
PSR from blowout to 3 ms+PFR to 3 ms				X	
PSR from blowout to 3 ms+PFR approaching $CO \Leftrightarrow CO_2$ equilibrium		X	X		
PSR from blowout to 3 ms+PFR with air addition approaching $CO \Leftrightarrow CO_2$ equilibrium		X	X		

The stoichiometry is expressed by the following chemical equation, and the CRN solutions provide the database for the CH_4 loss rate (oxidation rate) R_1:

$$CH_4 + 3/2O_2 \Rightarrow CO + 2H_2O$$

When evaluating the loss of hydrocarbon material in the CRN, our methodology is to include methyl radical (CH_3) with the CH_4. For most cases, other hydrocarbon containing species in the combustion field are negligibly small compared to CH_4 and CH_3. (The global mechanism has not yet been designed to consider sub-ppm levels of formaldehyde.)

Step 2 is the oxidation of CO. The stoichiometry is expressed by the following chemical equation, and the CRN solutions provide the database for the CO loss rate (oxidation rate) R_2:

$$CO + 1/2O_2 \Rightarrow CO_2$$

At the elemental level, this step occurs principally by the forward rate of the reaction

$$CO + OH \Leftrightarrow CO_2 + H$$

Step 3 is the dissociation of CO_2. The stoichiometry is expressed by the following chemical equation, and the CRN solutions provide the database for the CO_2 loss rate (dissociation rate) R_3:

$$CO_2 \Rightarrow CO + 1/2O_2$$

At the elemental level, this step occurs principally by the reverse of the reaction: $CO+OH \Leftrightarrow CO_2+H$.

Step 4 is the major lumped route to flame-NO formation: the Zeldovich and nitrous oxide mechanisms operating within the flame zone. The stoichiometry is expressed by the following chemical equation:

$$N_2 + O_2 \Rightarrow 2NO$$

The rate of NO formation is found from the CRN database per rates of Table 1. This gives the step-4 rate as:

$$R_4 = 2k_1[N_2][O] + 2k_5[N_2O][O] + 2k_6[N_2O][H]$$

Step 5 is the second lumped route to flame-NO formation: the prompt and NNH mechanisms operating within the flame zone. The stoichiometry is again expressed by the chemical equation: $N_2+O_2 \Rightarrow 2NO$.

The rate of NO formation is found from the CRN database per rates of Table 1. This gives the step-5 rate as

$$R_5 = 2k_7[N_2][CH] + 2k_{10}[NNH][O]$$

The species concentrations and temperatures required for R_4 and R_5 are provided by the CRN solution.

So far, a large database of global rate data for R_1-R_5 has been generated. This could be used via a look-up-table methodology in CFD computations. However, the goal here is to obtain global rate expressions for each of the five steps. This is explained after the three remaining steps of the mechanism are developed. The remaining steps (6–8) cover thermal-NO formation; thus, they depend on equilibrium thermochemistry. These steps only involve the Zeldovich and nitrous oxide mechanisms, since only these survive into the postflame zone. Furthermore, the thermal effect is subtracted from step 4 so that double counting does not occur. In all of these steps, the stoichiometry is expressed by the chemical equation from above: $N_2+O_2 \Rightarrow 2NO$.

Step 6 accounts for H-atom attack on N_2O, where both the H atom and N_2O are assumed to be at local equilibrium concentration. Thus, the rate is

$$R_6 = 2k_6[N_2O]_e[H]_e$$

$[N_2O]_e$ is expressed in terms of $[N_2]$ and $[O_2]$ and the equilibrium constant between N_2O, N_2, O_2, and $[H]_e$ is expressed in terms of $[H_2O]$ and $[O_2]$ and the equilibrium constant between H, H_2O, and O_2. The result of these manipulations, leading to the rate expression used, is given in Table 3. (Strictly, the N_2, O_2, and H_2O should also be denoted as equilibrium concentrations, but under lean-premixed postflame conditions, the difference between equilibrium and kinetic values for these major species is small.)

Step 7 accounts for O-atom attack on N_2O, where both the O atom and N_2O are assumed to be at local equilibrium concentration. Thus, the rate is

$$R_7 = 2k_5[N_2O]_e[O]_e$$

$[O]_e$ is expressed in terms of $[O_2]$ and the equilibrium constant between O and O_2. The result of the manipulations, leading to the rate expression used, is given in Table 3. Step 7 augmented by step 6 represents thermal nitrous oxide NO.

Step 8 accounts for O-atom attack on N_2, where the O atom is assumed to be at local equilibrium concentration. Additionally, N-atom steady-state is assumed valid, and the reverse rates of the Zeldovich mechanism are assumed negligible (which is clearly valid for lean-premixed combustion). This is textbook thermal Zeldovich NO (e.g., see Ref. [8]). The rate is

$$R_8 = 2k_1[N_2][O]_e$$

The result of the manipulations, leading to the rate expression used, is given in Table 3.

The general equation used for the global reaction rates, R_1-R_5, is

$$R = 10^{n+mP}[A]^{a+xP}[B]^{b+yP}[C]^{c+zP} \exp\{-(T_a + T_{a1}P)/T\}$$

where $T_a+T_{a1}P$ is the activation temperature (K), T is the combustion temperature (K), [ ] is the species concentration ($kmol/m^3$), P is the pressure (atm), and R is the reaction rate

Table 3 Eight-step global mechanism for a lean-premixed combustion in industrial gas turbine engines (note: Rxn 5: log=log base 10)

$CH_4+3/2O_2 \Rightarrow CO+2H_2O$ $R_1=10^{13.354-0.004628P}[CH_4]^{1.3-0.01148P}[O_2]^{0.01426}[CO]^{0.1987}\exp\{-(21{,}932+269.4P)/T\}$
$CO+1/2O_2 \Rightarrow CO_2$ $R_2=10^{14.338+0.1091P}[CO]^{1.359-0.0109P}[H_2O]^{0.0912+0.0909P}[O_2]^{0.891+0.0127P}\exp\{-(22{,}398+75.1P)/T\}$
$CO_2 \Rightarrow CO+1/2O_2$ $R_3=10^{15.8144-0.07163P}[CO_2]\exp\{-(64{,}925.8-334.31P)/T\}$
$N_2+O_2 \Rightarrow 2NO$ (flame NO by Zeldovich and nitrous oxide mechanisms) $R_4=10^{14.122+0.0376P}[CO]^{0.8888-0.0006P}[O_2]^{1.1805+0.0344P}\exp\{-(46{,}748+126.6P)/T\}$
$N_2+O_2 \Rightarrow 2NO$ (flame NO by prompt and NNH mechanisms) $R_5=10^{29.8327-4.7822\log(P)}[CO]^{2.7911-0.04880P}[O_2]^{2.4613}\exp\{-(61265+704.7P)/T\}$
$N_2+O_2 \Rightarrow 2NO$ (thermal NO by H-atom attack on N_2O) $R_6=10^{14.592}[N_2][H_2O]^{0.5}[O_2]^{0.25}T^{-0.7}\exp(-69158/T)$
$N_2+O_2 \Rightarrow 2NO$ (thermal NO by O-atom attack on N_2O) $R_7=10^{10.317}[N_2][O_2]\exp(-52861/T)$
$N_2+O_2 \Rightarrow 2NO$ (thermal NO by O-atom attack on N_2) $R_8=10^{14.967}[N_2][O_2]^{0.5}T^{-0.5}\exp(-68899/T)$

(kmol/m^3 s). n, m, a, b, c, x, y, z, T_a, and T_{a1} are coefficients and parameters to be determined.

Regression analysis on the CRN database is performed to obtain the global rate expressions for R_1-R_5. Following the selection of the species dependences for each global rate, the natural logarithm of the global rate expression is written, and then least squares analysis is conducted to obtain the coefficients and parameters. The global rate expressions that give the best agreement to the CRN database are listed in Table 3 below. As a surrogate for the flame free radicals, CO is used (see Polifke et al. [15]). R_1, R_2, R_4, and R_5 depend on [CO]. As expected, R_1 (methane oxidation) also depends on $[CH_4]$ and $[O_2]$. R_2 (CO oxidation) also depends on $[H_2O]$ and $[O_2]$ because of the importance of the hydroxyl radical in the elementary chemistry. R_3 (CO_2 dissociation) is found to depend only on $[CO_2]$. Fine tuning of R_2 and R_3 has been conducted so that the $CO \Leftrightarrow CO_2$ equilibrium can be obtained in the burnout zone of the combustor. The best fits for both flame-NO steps (R_4 and R_5) depend on [CO] as a surrogate for the free radical chemistry and on $[O_2]$.

Figure 2 provides an example of an agreement between a global rate expression and the CRN database. In this case, R_5 is plotted (note: CH mechanism means prompt mechanism). As shown in Table 2, this case used the single-PSR and two-PSRs-in-series CRN schemes. Postflame CRN schemes were not used because CH and NNH do not survive the flame. Typically, expect for a few outlying points at low rates, the agreement is very good. Although developed for methane combustion, steps 2–8 should be valid when other gaseous fuels are burned in lean-premixed combustion turbines. However, the mechanism is not valid for 1 atm combustion, for which free radical concentrations increase. See Nicol et al. [5] for a 1 atm global mechanism for lean-premixed methane oxidation with NO formation.

Application to Bluff Body Combustor

The high-pressure, lean-premixed, methane-fired, bluff body combustor of Bucher et al. [16] is modeled with CFD and the eight-step global mechanism. Figure 3 shows the cross-sectional area of the actual combustion chamber viewed from the exit plane, where the gas sampling probe is located. Two-dimensional

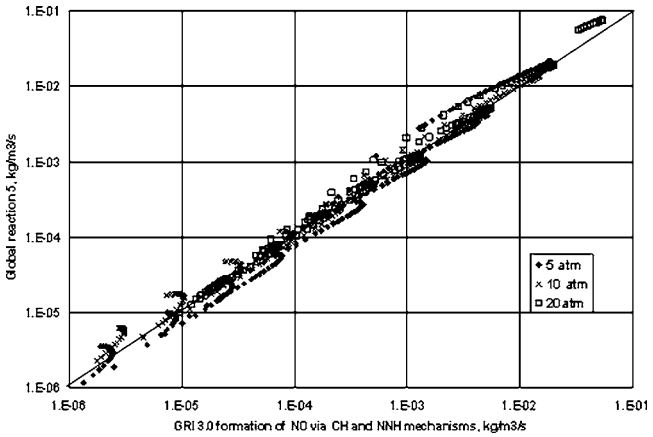


Fig. 2 Comparison of R_5 by a global rate expression of Table 3 (*y* axis) versus rates computed by GRI 3.0 (*x* axis)

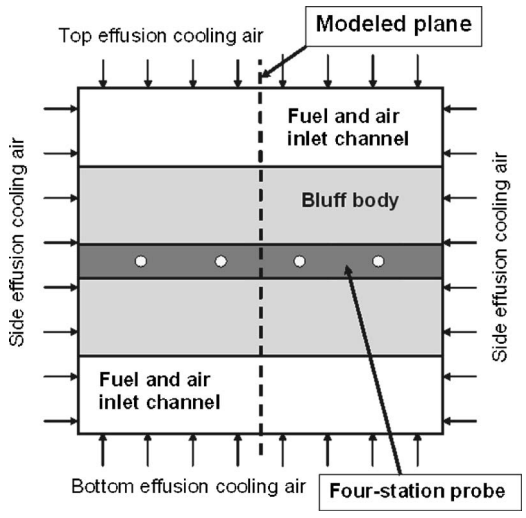


Fig. 3 View of the bluff body combustor from the exit plane. Shown is the gas sampling probe.

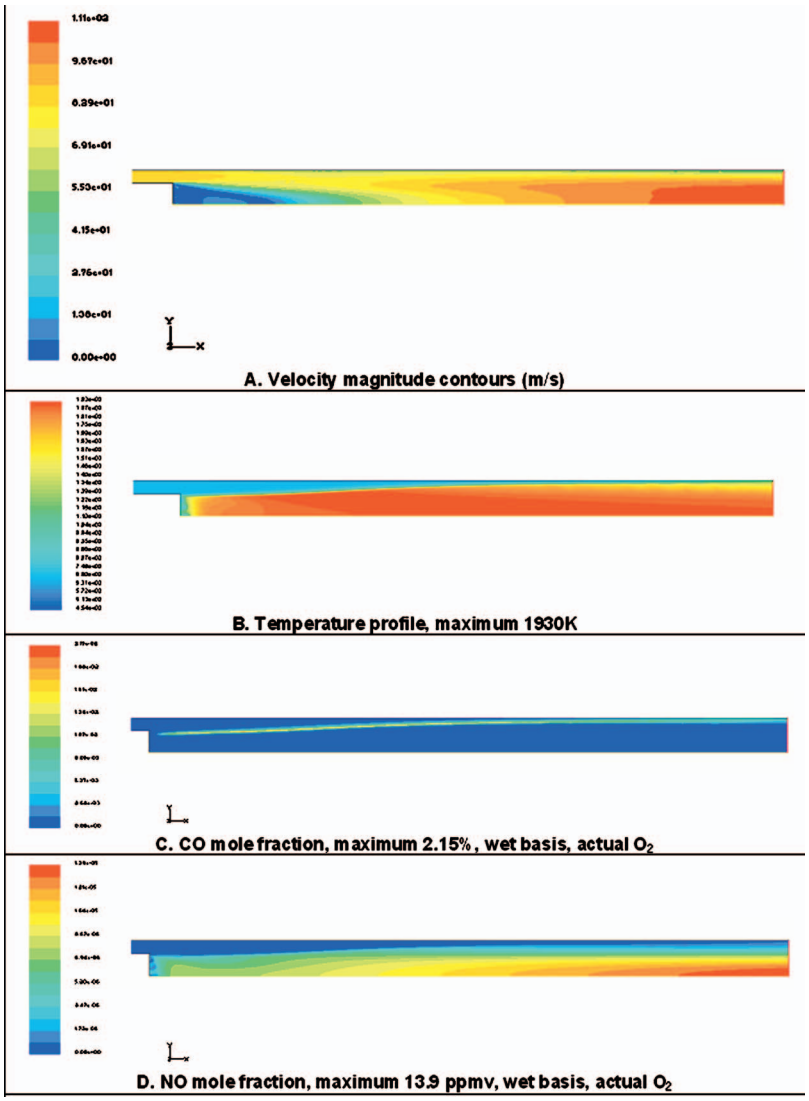


Fig. 4 CFD solutions for the bluff body combustor: methane fuel, 0.59 premixer phi, 14.3 atm pressure, and 678 K inlet air temperature.

CFD simulations of the combustor are shown in Figs. 4(*a*)–4(*d*). Premixed methane and air enter at the left of Figs. 4(*a*)–4(*d*), flow through the channel along the top of the bluff body (the bottom channel is not modeled), enter the combustion chamber, and create a large recirculation zone at the base of the bluff body (as indicated by the region of low velocity magnitude in Fig. 4(*a*)). The blockage ratio created by the bluff body is 0.63. The inlet air temperature and pressure are 678 K and 14.3 atm, respectively, and the nominal air mass flow rate is 1.08 kg/s.

The air is split between the main air (that is, the premixer air) and the combustor wall cooling air. Impingement/effusion cooling is used. Bluff body cooling air is drawn from the main air. Most of the methane is premixed into the main air well upstream of the channels running along the bluff body (see Bucher et al. [16]). Additionally, a few percent of the methane is injected via small jets located at the corners of the bluff body. Although the eight-step mechanism has not been designed to model stoichiometric (and near-stoichiometric) combustion of fuel jets, the impact of using the eight-step mechanism outside of its range is very small in this case because of the small percentage of pilot fuel burned.

The two-dimensional CFD simulations (Figs. 4(*a*)–4(*d*)) of the upper half of the combustor extend from combustor inlet to outlet. The position of the flame zone is best indicated in Fig. 4(*b*), as the temperature rapidly increases from that of the fuel-air mixture up to combustion temperature, and in Fig. 4(*c*), as the zone of peak CO concentration. This is one of the leaner cases treated: premixer phi is 0.59. The CFD indicates a flame zone starting near the bluff body corner and extending to the upper wall. Significant CO (1–2% by mole) is computed in the flame zone. At the exit plane, the CFD indicates a large CO concentration near the wall (because of the leanness of this case, the wall CO at the exit appears to be caused by effusion air quenching of flame CO). This CO falls rapidly as the wall region is departed. Nitric oxide is found spread across most of the combustor vertical dimension, though the concentration is greatest in the center and falls to zero at the air cooled wall (see Fig. 4(*d*)).

Maximum NO formation rate occurs in the flame zone. Peak flame-NO rates are 2.2×10^{-3} kmol/m^3 s by step 4 and 1.8×10^{-4} kmol/m^3 s by step 5. Peak thermal-NO rates in the post-flame zone are 2.6×10^{-5} kmol/m^3 s by step 7 and 3.5×10^{-5} kmol/m^3 s by step 8.

The CFD modeling of the bluff body combustor is performed with the assumptions and conditions listed in Table 4.

When modeling the combustor in two dimensions rather than in three dimensions, it is necessary to make adjustments in the wall

Table 4 Modeling assumptions and boundary conditions for the bluff body combustor

Computational domain	2D unstructured grid with 110,000 cells
Solver	Segregated RANS with species transport and volumetric reactions
Turbulence closure	Reynolds stress model
Convergence scheme	Second order (QUICK)
Pressure velocity coupling	Pressure implicit splitting of operators (PISO)
Wall treatment	Standard wall function
Heat loss	Convection and radiation heat transfer for top wall
Radiation heat transfer	Discrete ordinate (DO) model
Chemical kinetic rates	Eight-step mechanism
Chemical mixing rates	Eddy breakup rates

cooling. Effusion slots are modeled rather than effusion holes. Since there are no sidewalls in the 2D modeling, only two surfaces receive cooling air: most of the cooling air acts on the top surface of the combustor (the bottom surface is not modeled), while a smaller amount is used to cool the base of the bluff body. Unless the ratio of cooling air to main air is reduced, too much air will cool the top surface and possibly quench the flame. This reduction in cooling air acts to maintain the phi of the flame zone to that of the actual combustor, but increases the phi of the postflame zone and the overall phi of the modeled combustor.

Modeling and measurement of CO are compared in Fig. 5. When the premixer phi is less than 0.63, the CFD-area-average CO is considerably greater than the CFD-centerline CO. This is the situation depicted in Fig. 4(*c*). For greater premixer phi's, however, and thus for greater combustion temperatures, the two CFD CO results converge, and then for the highest phi run, the centerline CO exceeds the area-average CO. Now, CO is caused by the dissociation of CO_2, leading to enhanced CO in the hot gases found in the center of the combustor. For the highest phi's, the CO emission is thermodynamically controlled, whereas at the lower phi's, it is kinetically controlled.

For the leanest phi's, the experimental CO is seen to lie at about 1/3 "height" between the centerline CFD result (which is very

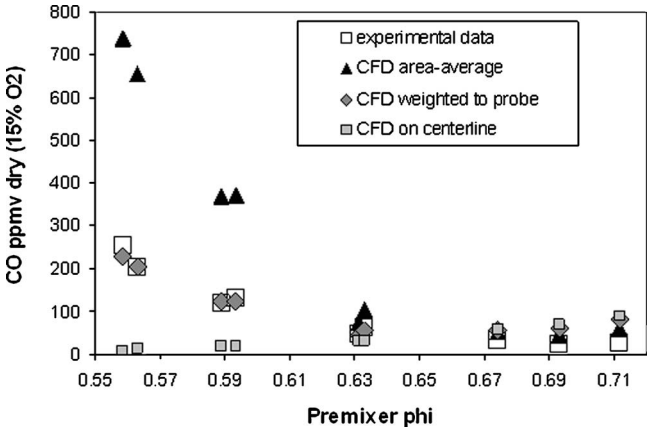


Fig. 5 CFD results for CO in comparison to measurements for the bluff body combustor

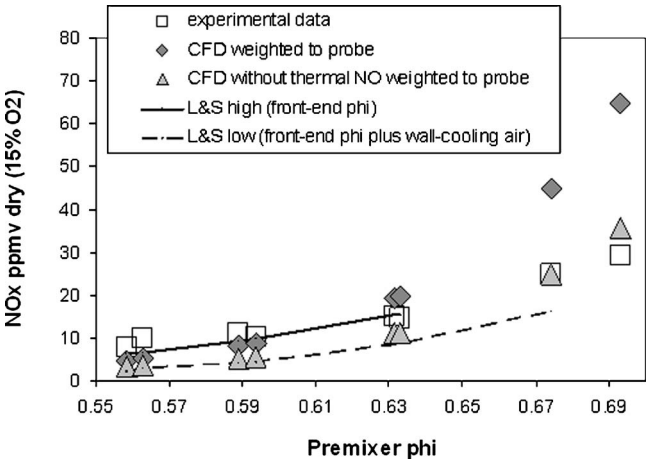


Fig. 6 CFD results for NO in comparison to measurements for the bluff body combustor Calculations of Leonard and Stegmaier NO_x also plotted.

small) and the area-average CFD result (which is significantly influenced by the large amount of CO near the upper wall). This suggests the following relation for weighting the CFD results to the probe:

$$CO_{probe} \cong 0.7CO_{center} + 0.3CO_{average}$$

Novosselov [6] finds similar coefficients by considering diffusion of gas from the combustor wall toward the probe.

For the highest phi's examined, experimental CO is about 25 ppmv (dry 15% O_2), whereas the CFD results are in the 50–80 ppmv range (dry 15% O_2). Very likely, the difference is caused by enhanced dissociation of CO_2 in the postflame zone of the 2D CFD combustor, which as argued above may run hotter than the experimental combustor. Additionally, the opportunity for oxidation of the experimental CO cannot be ruled out as the hot combustion gases are drawn into the sampling probe and dwell there.

The comparison of modeling and measurement of NO_x is shown in Fig. 6. The measured NO_x of Bucher et al. [16] is seen to increase smoothly from 8–10 ppmv (dry 15% O_2) at the leanest case (phi of 0.56) up to 30 ppmv (dry 15% O_2) at the highest phi plotted (0.69). As shown first by Bucher et al. [16], the experimental NO_x can be modeled closely using the correlation of Leonard and Stegmaier [17]. Leonard and Stegmaier operated a porous-plate burner over a wide range of lean-premixed conditions (including elevated pressures) up to about 10 ppmv NO_x (dry 15% O_2) and 1950 K, and showed NO_x emission correlating well with a single variable: adiabatic equilibrium flame temperature. This approach is used in deriving the L&S curves in Fig. 6. For "L&S high," the front-end phi (which is based on the premixer phi plus the small corner jets' fuel), inlet air and fuel temperatures, and pressure are used to calculate the adiabatic equilibrium flame temperature. Then, the NO_x is determined from the Leonard and Stegmaier straight-line fit, of the following form:

$$\ln(NO_x) = a + bT$$

For "L&S low" in Fig. 6, the front-end phi is adjusted by adding cooling air from the top wall, up to the point where the flame approaches the wall. The effect of this is to decrease the flame temperature and the L&S NO_x. The L&S NO_x curves are truncated once they exceed the range of the Leonard and Stegmaier NO_x database.

Figure 6 also shows the CFD generated NO results. The procedure is the same as that used for CO: that is, the centerline and area-average 2D CFD results for NO are generated. These values are then multiplied, respectively, by 0.7 and 0.3 and added to give

the probe NO. The CFD-probe NO is significantly less sensitive to the choice of the coefficients (so long as they add to 1) than the case of CO for the lean mixtures.

The CFD results in Fig. 6 also include the effect of turbulent fluctuations on NO formation. This follows from a separate CFD study of lean-premixed combustion using PDF modeling (Novosselov [6]), from which it is deduced that for the leanest phi's an additional 20–25% NO could form, and for the highest phi's the increase could be 10–15%. This implies a greater impact of turbulent fluctuations and, thus, a larger correction to the steady-state NO for flame NO than for thermal NO in lean-premixed combustors.

Two sets of CFD NO results are plotted in Fig. 6: the upper one for all NO steps included and the lower one with the thermal-NO steps removed.

The following observations are drawn.

- The 2D CFD-simulated NO, using all steps, shows a reasonably close agreement to the experimental NO_x when the premixer phi is less than about 0.63.
- For the highest values of phi examined, the CFD, using all steps, overpredicts the measured NO_x. The overprediction is in the thermal NO, presumably because the postflame zone of the 2D combustor may run hotter than the actual combustor.
- The CFD solution with the thermal-NO steps removed, forming only flame NO by steps 4 and 5, shows a good agreement with L&S low over the range of validity of the Leonard and Stegmaier NO_x.

Application to Gas Turbine Combustor

Above, we showed CFD results for a generic, lean-premixed, gas turbine combustor. In that modeling, a 2D structured grid of 31,000 cells was used. The Reynolds stress turbulence closure model with a quadratic pressure strain was used; other conditions were similar to those listed in Table 4 for the bluff body combustor.

Additionally, lean-premixed, engine test rig combustors have been modeled by multiple-element CRN models. The development of a 31-element CRN for an annular test rig combustor with air-back-side cooling is reported in the 2006 paper by Novosselov et al. [18]. The development of the 31-element CRN is based on a 3D CFD solution of a sector of the annular combustor. The CFD solution provides insight and information useful for selecting the size and nature of the CRN elements and mass exchange between the elements. The CRN can treat either uniform or nonuniform fuel-air ratio profiles at the premixer-injector outlet and takes into account velocity nonuniformity at the premixer outlet. It can also treat premixed pilot flames. The 3D CFD uses the eight-step mechanism. On the other hand, the CRN is capable of quickly running either a full chemical kinetic package (GRI 3.0 in this study) or a global mechanism (the eight-step global mechanism in this study).

The 3D CFD solutions have been used mainly to guide the CRN development, whereas the CRN output contains the information useful for a comparison of the eight-step predictions of CO and NO_x to those of full GRI 3.0 solutions and the engine test rig data.

Table 5 shows the CO comparison and Fig. 7 shows the NO_x comparison. The eight-step mechanism shows a very good agreement to the full GRI 3.0 mechanism for CO and NO_x emission calculations when applied to the industrial gas turbine combustor through the 31-element CRN. The slightly greater amount of NO_x (actually NO) obtained with the eight-step mechanism is the result of a slightly higher combustion temperature obtained with the global chemistry compared to GRI 3.0. That is, the full mechanism has somewhat more endothermicity that is lacking in the global mechanism. This deficiency can be overcome easily by adding a small amount of an inert species to the methane used with the global mechanism. Both mechanisms show a very good agreement to the NO_x measured for the engine test rig combustor. This provides confidence in the 31-element CRN, which was the objective of the paper by Novosselov et al. [18], and confidence in the eight-step mechanism, which is the purpose of the current paper.

Table 5 Comparison of CO emissions for test rig combustor running lean premixed [18]

Pilot level	35–185% of neutral
GRI 3.0 in CRN	1.82 ± 0.03 ppmv (dry 15% O_2)
8-step in CRN	1.96 ± 0.03 ppmv (dry 15% O_2)

The CO emissions measured for the engine test rig are within a few ppmv of the CRN calculations.

Conclusions

In this paper, the development of the eight-step mechanism for methane oxidation with flame and postflame-NO formation for lean-premixed, high-pressure combustion is explained. The eight steps are given. In addition to the dominance by flame NO, a thermal nitrous oxide step is shown to form essentially equal amounts of NO to thermal Zeldovich NO in the postflame zone. The range of validity of the eight-step mechanism is 5–20 atm pressure, air inlet temperature corresponding to compression to this pressure range, and 0.45–0.75 phi. Combustors with pilot flames of phi somewhat greater than 0.75 have been modeled with the mechanism, but since this is outside of the development range, the percentage of pilot fuel burned should be small; otherwise, inaccuracies may occur in the computed NO_x emission. Two gas turbine lean-premixed combustors, one a virtual generic combustor and the other an engine test rig combustor, have been modeled with the eight-step mechanism using both CFD and CRN techniques. CFD modeling of the generic combustor is used to show how and where NO forms in a single-digit NO_x combustor. Modeling of the test rig combustor shows a very good agreement to the measured CO and NO_x, especially the NO_x. This provides confidence in the eight-step mechanism. A bluff body combustor is also modeled in CFD using the eight-step mechanism, with modeling results comparing favorably to the measurements of CO and NO_x. Nonetheless, this combustor has presented a challenge when comparing modeling with measurement because of the nonuniformity of the exit plane gas composition.

Topics for future work include the application of the eight-step mechanism to other databases for lean-premixed combustion and

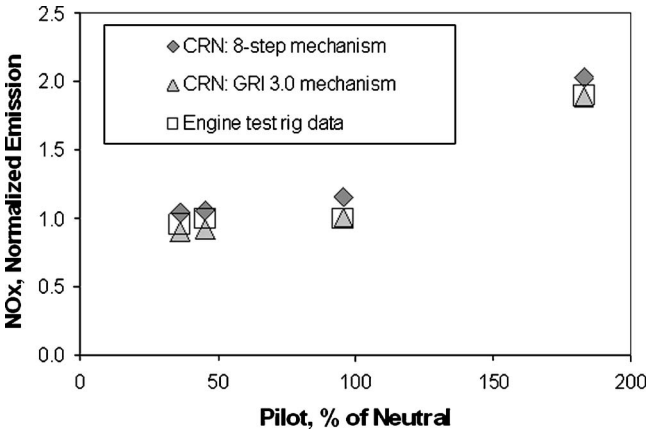


Fig. 7 Comparison of modeled and measured NO_x for the engine test rig combustor. NO_x emission is normalized by test rig emission for the neutral pilot. At neutral, the pilot has the same phi as the main premixer.

the development of the mechanism for alternative gaseous fuels of interest for gas turbine engines, including natural gas high in NMHC, LNG, syngases, and syngases and hydrogen blended with natural gas.

Acknowledgment

This paper is based on the Ph.D. thesis of Dr. Igor Novosselov [6].

References

[1] Westbrook, C. K., and Dryer, F. L., 1981, Combust. Sci. Technol., **27**, pp. 31–43.
[2] Westbrook, C. K., and Dryer, F. L., 1984, Prog. Energy Combust. Sci., **10**, pp. 1–57.
[3] DuPont, V., Pourkashanian, M., and Williams, A., 1993, J. Inst. Energy, **66**, p. 20.
[4] Hautman, D. J., Dryer, F. L., Schug, K. P., and Glassman, I., 1981, Combust. Sci. Technol., **25**, pp. 219–235.
[5] Nicol, D. G., Malte, P. C., Hamer, A. J., Roby, R. J., and Steele, R. C., 1999, Trans. ASME: J. Eng. Gas Turbines Power, **121**, pp. 272–280.
[6] Novosselov, I. V., 2006, "Chemical Reactor Network Modeling of Combustion Systems," Ph.D. thesis, University of Washington, Seattle, WA. (This may be accessed online at the address www.energy.washington.edu, denoting "laboratory," followed by "publications.")
[7] Magnussen, B. F., and Hjetager, B. H., 1977, "On Mathematical Models of Turbulent Combustion With Special Emphasis on Soot Formation and Combustion," *Proceedings of the 16th Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, pp. 719–729.
[8] Heywood, J. B., 1988, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.
[9] Rutar-Shuman, T., 2000, "NOx and CO Formation for Lean-Premixed Methane-Air Combustion in a Jet-Stirred Reactor Operated at Elevated Pressure," Ph.D. thesis, University of Washington, Seattle, WA.
[10] Bengtsson, K. U. M., 1998, "Experimental and Numerical Study of the NOx Formation in High-Pressure Lean Premixed Combustion of Methane," Ph.D. thesis, Swiss Federal Institute of Technology, Zürich, Switzerland.
[11] Rutar, T., and Malte, P. C., 2002, ASME J. Eng. Gas Turbines Power, **124**, pp. 776–783.
[12] GRI MECH 3.0 (http://www.me.berkley.edu/gri_mech/).
[13] Novosselov, I. V., 2002, "Eight-Step Global Kinetic Mechanism of Methane Oxidation With Nitric Oxide Formation for Lean Premixed Combustion Turbines," MSME thesis, University of Washington, Seattle, WA. (This may be accessed online at the address www.energy.washington.edu, denoting "laboratory," followed by "publications.")
[14] Konnov, A. A., and De Ruyck, J., 2001, Combust. Flame, **125**, p. 1258.
[15] Polifke, W., Dobbeling, K., Sattelmayer, T., Nicol, D. G., and Malte, P. C., 1996, Trans. ASME: J. Eng. Gas Turbines Power, **118**, pp. 765–772.
[16] Bucher, J., Edmonds, R. G., Steele, R. C., Kendrick, D. W., Chenevert, B. C., and Malte, P. C., 2003, "The Development of a Lean-Premixed Trapped Vortex Combustor," ASME Paper No. GT-2003-38236.
[17] Leonard, G., and Stegmaier, J., 1994, ASME J. Eng. Gas Turbines Power, **116**, pp. 542–546.
[18] Novosselov, I. V., Malte, P. C., Yuan, S., Srinivasan, R., and Lee, J. C. Y., 2006, "Chemical Reactor Network Application to Emissions Prediction for Industrial DLE Gas Turbine," ASME Paper No. GT2006-90282.

Vortex Breakdown in Swirling Fuel Injector Flows

Adrian Spencer

James J. McGuirk

Department of Aeronautical and Automotive Engineering,
Loughborough University,
Loughborough, Leicestershire LE11 3TU, U.K.

Kris Midgley
Renault F1 Team Ltd,
Enstone, Oxfordshire, OX74EE U.K.
e-mail: j.j.mcguirk@lboro.ac.uk

It is well known that the process of vortex breakdown plays an important role in establishing the near-field aerodynamic characteristics of fuel injectors, influencing fuel/air mixing and flame stability. The precise nature of the vortex breakdown can take on several forms, which have been shown in previous papers to include both a precessing vortex core (PVC) and the appearance of multiple helical vortices formed in the swirl stream shear layer. The unsteady dynamics of these particular features can play an important role in combustion induced oscillations. The present paper reports an experimental investigation, using particle image velocimetry (PIV) and hot-wire anemometry, to document variations in the relative strength of PVC and helical vortex patterns as the configuration of a generic fuel injector is altered. Examples of geometric changes that have been investigated include: the combination of an annular swirl stream with and without a central jet; variation in geometric details of the swirler passage, e.g., alteration in the swirler entry slots to change swirl number, and variations in the area ratio of the swirler passage. The results show that these geometric variations can influence: the axial location of the origin of the helical vortices (from inside to outside the fuel injector), and the strength of the PVC. For example, in a configuration with no central jet (swirl number $S=0.72$), the helical vortex pattern was much less coherent, but the PVC was much stronger than when a central jet was present. These changes modify the magnitude of the turbulence energy in the fuel injector near field dramatically, and hence have an important influence on fuel air mixing patterns. [DOI: 10.1115/1.2799530]

Introduction

Although vortex breakdown was first observed in a study of delta wing leading edge aerodynamics (Peckham and Atkinson [1], the relevance to the highly swirling flows produced by gas turbine combustion system fuel injectors soon became apparent (e.g., Gupta et al. [2]). The term vortex breakdown refers to: "*an abrupt and drastic change in flow structure which is characterised by a sudden axial deceleration that occurs above a certain level of swirl, leading to the formation of a free stagnation point followed by a separation region with turbulence behind it*," Wang and Rusak [3]. Because of the complex nature of the phenomenon and its significance for several important engineering applications, it has received considerable attention in the literature for the last 50 years, covering experimental, theoretical, and computational investigations. Two comprehensive reviews have been produced by Escudier [4] and Lucca-Negro and O'Doherty [5]. Nevertheless, there is still no general theory of vortex breakdown or agreement on its fundamental nature.

Experimentally, Sarpkaya [6] and Leibovich [7] discovered seven distinct types of vortex breakdown; however, in turbulent flows only two, the spiral and bubble modes, were observed. Sarpkaya [6] also showed that the effect of an adverse axial pressure gradient was to move the vortex breakdown further upstream, or if extreme enough, even to initiate it. In addition to the appearance of a recirculation region in the bubble mode, evidence has been gathered that vortex breakdown can display strongly time-dependent characteristics, whereby the recirculation zone itself becomes unstable and starts to precess about its own axis (a precessing vortex core (PVC), Syred and Beer [8]). Finally, an important concept that appears in most explanations of vortex breakdown is the idea of "criticality" (Benjamin [9]), i.e., whether the flow axial velocity is, throughout the flow, greater than the phase velocity of longitudinal inertial instability waves (supercritical), or less than this (subcritical). In the latter case, a mechanism exists whereby the downstream conditions may affect the upstream flow. Indeed, perhaps a useful analogy for the occurrence of vortex breakdown is that it allows a supercritical flow to transition to a subcritical one, so that it may then gradually adjust to downstream conditions, in the same way that a supersonic flow requires a shock wave to be formed, so that a subsonic flow is created that can then sense the downstream pressure field. It is for this reason that care has to be taken in both experimental and numerical studies of vortex breakdown to pay attention to the region of influence of downstream conditions. Escudier et al. [10] have recently shown the importance of this for highly swirling flows downstream of a nozzle or orifice.

The significance of vortex breakdown for the design of gas turbine fuel injectors that employ high swirl for flame stabilization reasons is well known. Li and Gutmark [11], in measurements on a triple annular swirler typical of future fuel lean module injectors, have shown that the dynamics of vortex breakdown cause interactions between vortical structures in the flow associated with breakdown and acoustical instability modes. The measurements show that the vortical structures are highly dependent on swirler geometry. The fundamentally unsteady nature of vortex breakdown and its association with instability modes of the flow has attracted the attention of computational fluid dynamics (CFD) model developers, particularly using large eddy simulation (LES). Studies of highly swirling flow in confined (Wegner et al. [12]) and unconfined (Garcia-Villalba and Froehlich [13]) geometries have recently been reported. The former noted the appearance of what was described as "*a double-cell precessing vortex core*," but it is not clear whether this phrase is mixing up the appearance of rotating spiral vortex structures and a genuine PVC. The latter study observed that the introduction of a central pilot jet inside an annular swirl stream had little influence on the average flow, but spectra extracted from the LES predictions indicated dramatic effects on the instantaneous spiral vortex structures. This result is very significant, but it needs to be confirmed that it is correct for

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 1, 2007; final manuscript received July 4, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea, and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007.

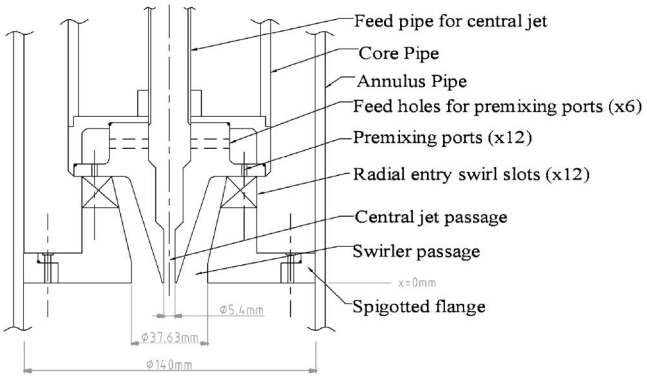


Fig. 1 Datum fuel injector geometry

more practically relevant confined swirl flow. Finally, in a detailed experimental study that included both a central jet and an annular swirl stream, Midgley et al. [14,15] deduced from PIV measurements that multiple spiral vortex structures were observed wrapped around a central recirculation zone (CRZ). This should perhaps be interpreted as a vortex breakdown in which both spiral and bubble modes occurred simultaneously. The origin of the spiraling vortex structures was traced to a separation event inside the fuel injector, and the "two strong/two-weak" double helix form of the spiral vortices was explained by the need for the secondary flowfield of the vortex structures to satisfy continuity within the annulus. As noted in other works, the near fuel injector turbulence field was strongly influenced by the appearance of the spiraling vortex structures. Based on the above, the measurements of Midgley et al. [15] have been extended to explore the influence of a central jet in a confined flow, and the effect of the swirl and swirl passage geometry on vortex breakdown.

Test Facility

The test rig used for the present measurements is the same as in the study reported by Midgley et al. [15]; details are given in Refs. [14,15]. Tests were conducted in both water and air flow, but superior PIV data were obtained in water flow and these are reported here, with hot wire data from airflow tests. The radially fed swirler (12×30 deg angled slots) produced an annular swirlstream (outer diameter $D_s=37.63$ mm) at fuel injector exit (Fig. 1). The inner diameter of the annulus was formed by the outer wall of a central jet nozzle (inner diameter $D_j=5.4$ mm). The standard test condition corresponded to swirler and jet Reynolds numbers of 8×10^4 and 2.63×10^4 defined using D_s and D_j and the bulk mean velocities based on geometrical exit areas: $V_{x,j}=4.88$ m/s, $V_{x,s}=2.13$ m/s. Note, however, that the discharge coefficient of the swirl passage was measured at 0.5, so the peak swirlstream velocity was about 1.5 $V_{x,s}$. In the data reported in [15], the central jet had always been present; the current experiments concentrated initially on the case with no central jet, but an unchanged swirlstream condition. To allow further exploration of changes in the swirlstream conditions, a modular injector was designed (Fig. 2). This allowed changes to the radial swirler slot angle (α_1), and the angles of the central and outer body (α_2,α_3) to assess the effects of swirl passage shape. In all tests with the modular injector, no central jet was present. Note also that the annular swirl passage in the modular geometry did not contain an internal corner, as present in the datum geometry. This was intentional to allow assessment of whether the presence of this internal corner in the datum geometry was a "trigger" to the internal vortex breakdown, as suggested in [15]. By selecting $\alpha_2=17$ deg and $\alpha_3=9.5$ deg, (Case 1) the injector exit plane bulk axial and swirl velocities of the swirlstream were the same in datum and modular fuel injectors. By varying (α_2,α_3), the bulk average axial velocity ratio of the swirl passage flow between outlet and inlet

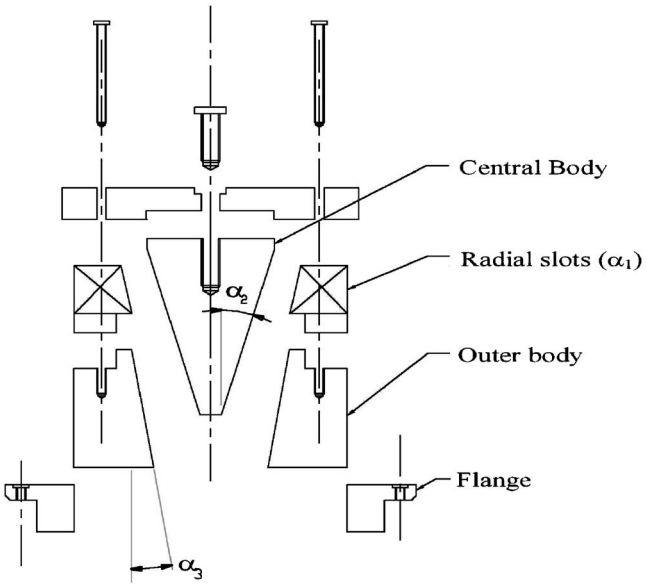


Fig. 2 Modular fuel injector geometry

($V_{x,\mathrm{so}}/V_{x,\mathrm{si}}$), and similarly the swirl velocity ratio ($V_{t,\mathrm{so}}/V_{t,\mathrm{si}}$), could be altered to investigate the effect of different levels of acceleration and "spin up" within the swirlstream passage on the flow produced. Table 1 shows the various cases considered. Case 6 used central body and outer body modules of constant radius (14/25 mm) to provide a parallel swirl passage configuration. Finally, to assess the impact of the swirl number on the flow, different radial slot modules were constructed with various values of α_1. Table 2 shows the resulting swirl numbers as deduced from exit plane PIV measurements.

Instrumentation

The PIV instrument comprised a LaVision system with a twin Nd: YAG laser generating a 1 mm thick light sheet via spherical and cylindrical lenses, and a Kodak Megaplus ES 1.0 camera with 1008×976 pixel resolution, running at 15 Hz (a few measurements at 1 kHz were also made with an upgraded camera/PIV system). Laser/camera synchronization, image acquisition, and data processing using the cross-correlation technique were controlled by Davis software running on a dual processor Pentium III PC. 20 μm polyamide particles (almost neutrally buoyant in water) were used as light scattering particles. The PIV instrumenta-

Table 1 Modular injector geometry parameters

	α_2 (deg)	α_3 (deg)	$V_{x,\mathrm{so}}/V_{x,\mathrm{si}}$	$V_{t,\mathrm{so}}/V_{t,\mathrm{si}}$
Case 1	17	9.5	1.25	1.56
Case 2	20.5	8.5	1.3	1.56
Case 3	13.5	11.5	1.43	1.56
Case 4	13.5	8.5	1.29	1.65
Case 5	20.5	11.5	1.21	1.43
Case 6	0	0	1.0	1.0

Table 2 Radial swirl module slot angles and swirl numbers

α_1 (deg)	S
30	0.72
20	0.57
10	0.47
0	0

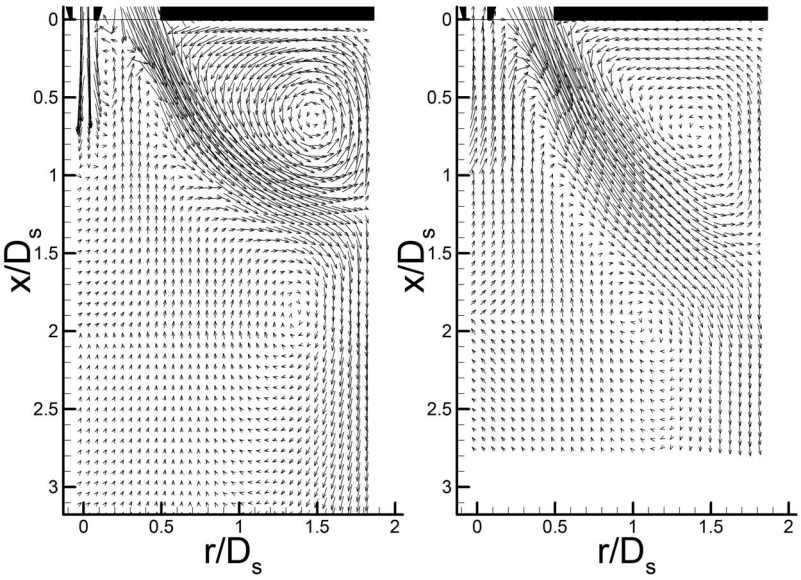


Fig. 3 Overall flow structure; datum injector (a) with fuel jet; (b) without fuel jet

tion parameter setup was selected following the recommendations of Keane and Adrian [16]. Laser power and optical system/data collection parameters were optimized by minimizing the interframe time Δt to reduce truncation errors in velocity evaluation, but not setting Δt so small as to incur precision errors in particle location due to subpixel interpolation. The interframe time (particle shift) was set so that the inplane displacement was less than $1/4$ of the PIV interrogation cell used for cross correlation and, where possible, led to out-of-plane motion less than $\delta z/4$, where δz is the camera depth of field. The optimum particle image diameter is around 2 pixels. For 20 μm particles, this was achieved in the present experiment when the field of view (FoV) was smaller than 80×80 mm^2. In most data collected, there were at least ten particle pairs per interrogation window. The interrogation cell used a deformed multipass grid (LaVision [17]) of initial size 64×64 pixels, reducing to a final size of 32×32 pixels (with 50% overlap). The interframe delay time was around 500 μs for the larger FoV images used for initial identification of flow structures (160×160 mm^2) and 20–50 μs for the smallest FoV (20×20 mm^2). This minimized out-of-plane data loss and maximized the inplane displacement, which was typically from 3–8 pixels. The same approach to data validation was applied as described in [15]; a range of FoVs was used that allowed optimization of PIV system parameters to the length scale of local dynamically important flow structures. The local turbulence length scales were estimated using the technique developed by Spencer and Hollis [18], and the measured turbulence levels corrected (increased) using the formula proposed in [18] if the size of the PIV interrogation cell was not small enough, relative to the local integral length scale. In all the data presented here for statistical quantities, time averaging over 650 PIV frames has been carried out.

Results and Discussion

Comparison of Flowfield With and Without Central Jet. Figure 3 presents the mean velocity vectors in the x-r plane measured in the datum injector geometry. In comparison with the overall flow structure for the case with a central jet shown in Fig. 3(a) [15], the no jet case in Fig. 3(b) shows that the central recirculation zone (CRZ) extends right up to the injector exit plane, the corner recirculation is weaker and less well defined, with the swirl cone attachment point on the outer wall further downstream ($x/D_s=1.5$ compared to the value $x/D_s=1.2$, reported in [15]). The absence of the central jet changes the turbulence structure in the centerline region in the first 2–3 swirler diameters dramatically. Figure 4 compares measured radial RMS turbulence inten-

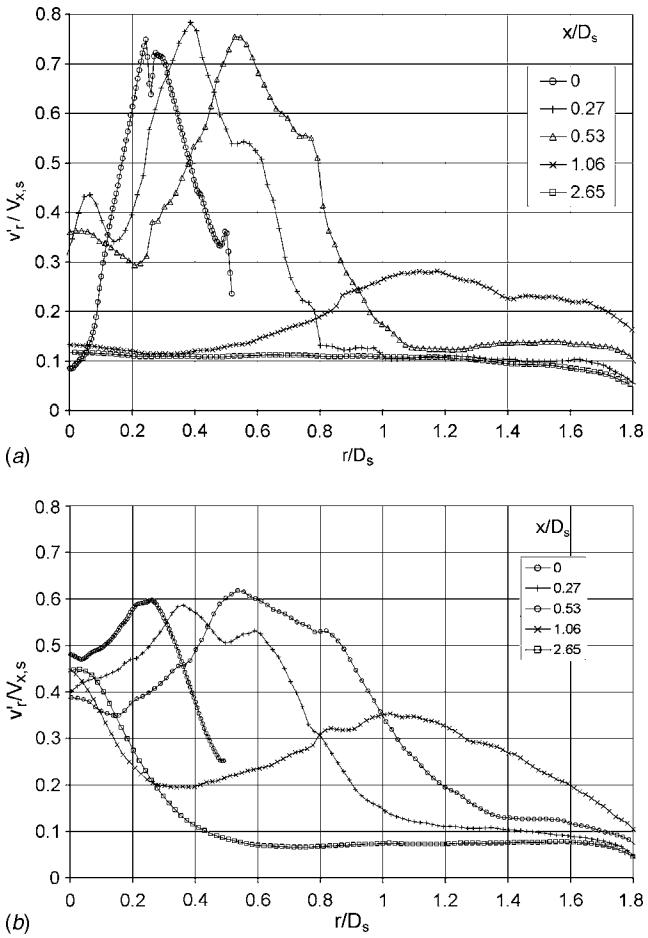


Fig. 4 Radial RMS profiles; (top) with central jet; (bottom) no central jet

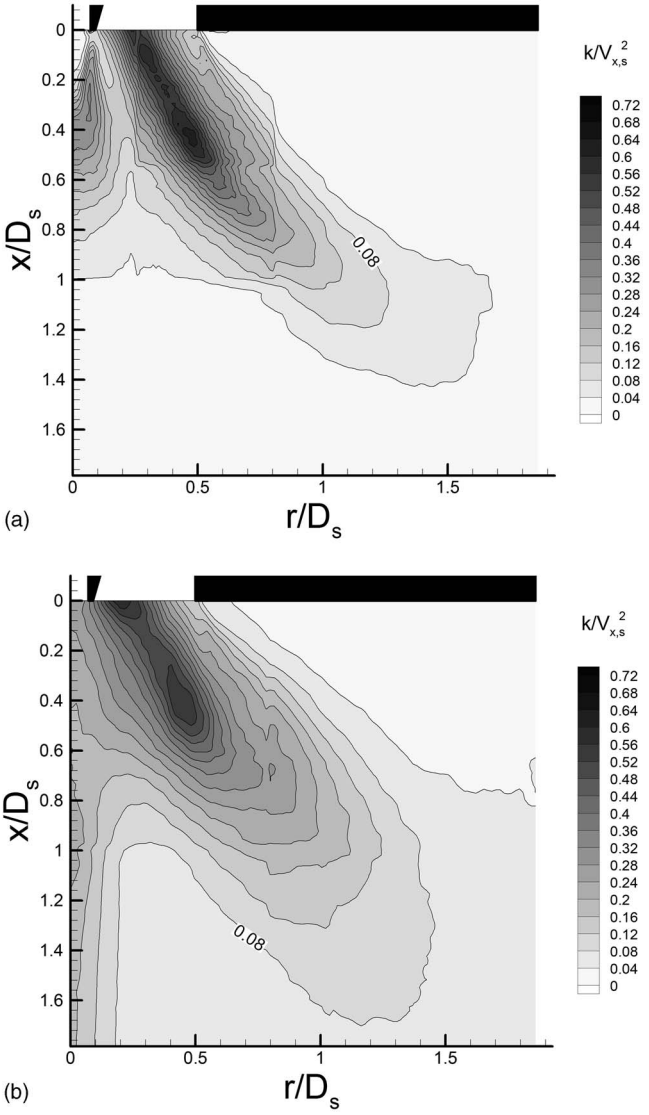


Fig. 5 Turbulence kinetic energy contours; (top) with central jet; (bottom) no central jet

sity profiles in this region for flows with and without a central jet (similar changes are observed in axial and circumferential RMS data). Outboard of $r/D_s \sim 0.3$, the results are quite similar in the two flows. A striking difference is observed, however, closer to the centerline. With a central jet, the turbulence levels start low, rise to a peak at $x/D_s \sim 0.5$ and then decay rapidly beyond this. This structure is connected to the high axial velocity gradients in the region of $x/D_s = 0.5-1.0$ associated with the free stagnation point between forward flowing central jet and backflow in the CRZ. Of course with no central jet, as noted above, the CRZ reaches back to the injector exit plane. The turbulence levels in the no-jet case are quite different. Figure 4 shows that these start high and remain very high ($\sim 45\%$) right down to $x/D_s \sim 3$. Comparison with measured axial and circumferential turbulence intensities (Midgley [19]), indicates that the radial and circumferential turbulence levels are significantly higher (a factor of ~ 2) compared to the axial data. This is recognizable as an effect of the unsteady fluctuations associated with a PVC (that would influence radial and tangential motions more than axial), which are picked up by the time-averaging process as "turbulence." The implication is that the presence of a central jet stabilizes the CRZ and damps its tendency to precess; more measurements to confirm that the no central jet case does have a strong PVC will be presented below. Further illustration of the changed fluctuating velocity field brought about in the near-injector region by the omission of the central jet is provided in Fig. 5, which compares the measured turbulence kinetic energy fields. In both cases, the peak turbulence levels are associated initially with the inner edge of the swirl-stream shear layer (and identified in [15] as caused by the rotating multiple vortex system). For no central jet, however, the maximum value is lower (0.6 compared to 0.75). The swirl shear layer is clearly more diffused in the no jet case, and the streak of high fluctuations measured on the centerline and stretching downstream is again the contribution from the PVC, which is clearly stronger in the no jet case. It is important to note again here that, for the high values measured in the shear layer edge, and on the centerline in the no jet case, these should not be considered as genuine turbulence, since a large contribution is due to high energy periodic events. A further piece of evidence that shows the no-jet case is markedly more unstable than the central jet case is contained in Fig. 6; this shows the PDF of the outer wall attachment point of the swirl cone. The larger variance in this ($x/D_s = 0.5-2.5$) than when a central jet is present ($x/D_s = 0.75-1.75$)

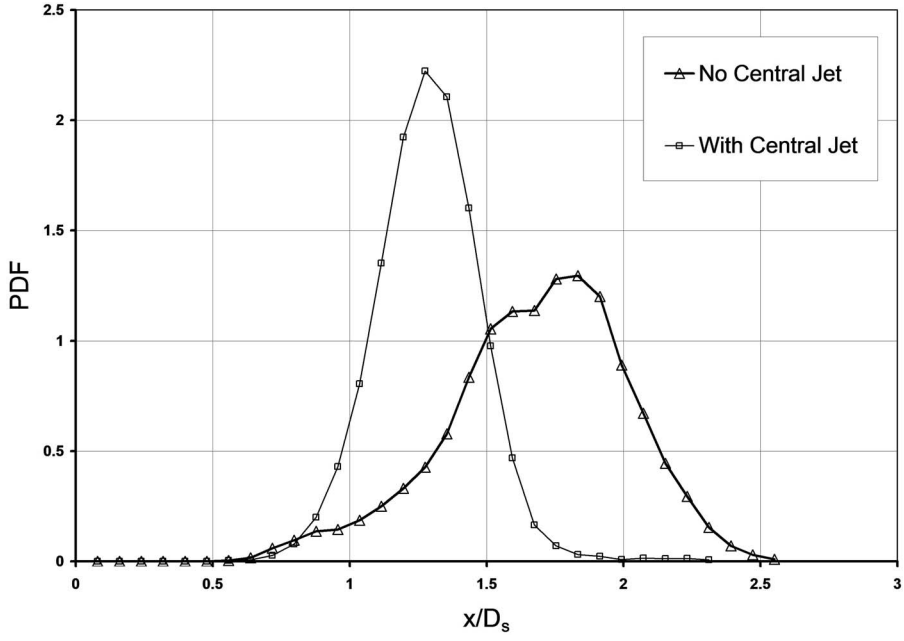


Fig. 6 PDF of spatial attachment point of swirl cone

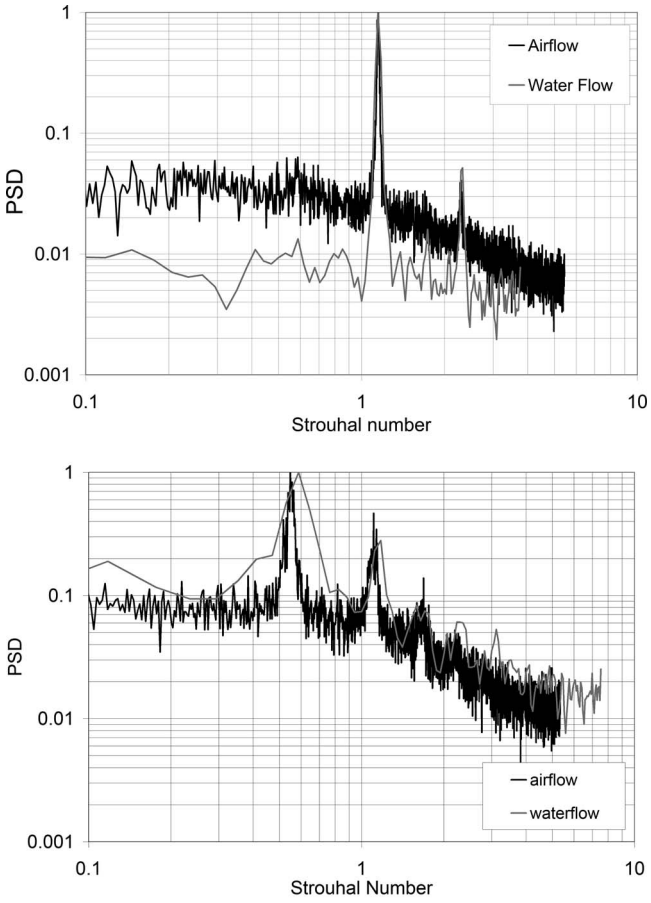


Fig. 7 PSD of axial velocity in swirlstream shear layer (top) with central jet; (bottom) no central jet

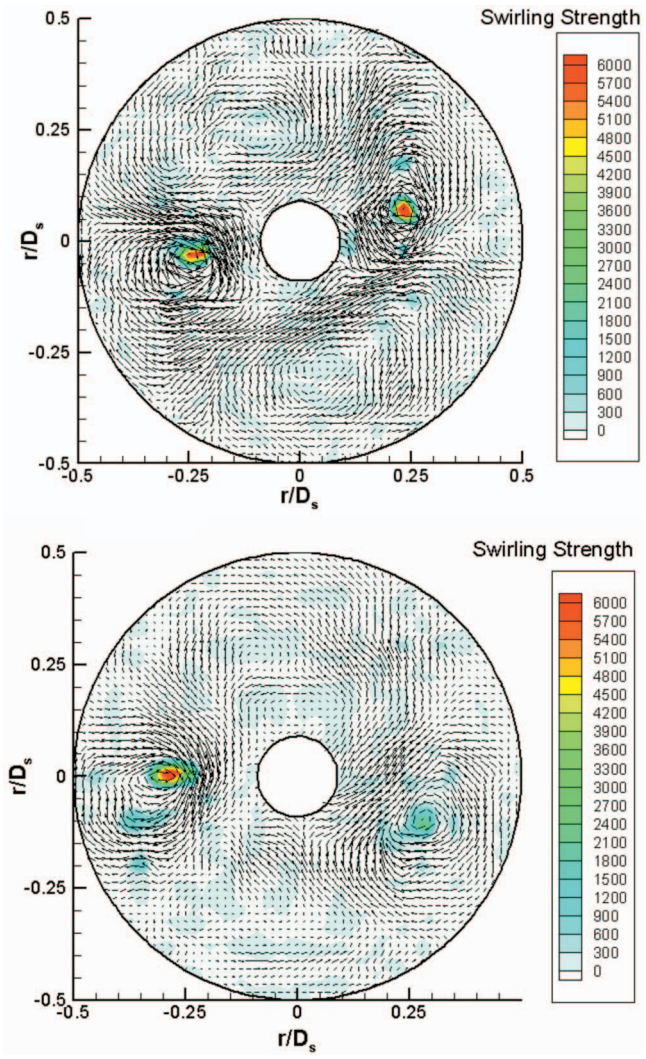


Fig. 8 Conditionally averaged velocity vectors in $r-\theta$ plane at $x/D_s=0.0$; (top) with central jet; (bottom) no central jet

indicates a more vigorous "flapping" of the swirl stream cone, once again probably induced by the strong PVC. Finally, Fig. 7 compares power spectral density (PSD) measurements of the axial velocity taken at a point within the inner swirlstream shear layer edge at injector exit ($x/D_s=0.27$, $r/D_s=0.27$, $\theta=0$ deg). These data were taken in an airflow experiment with a hot wire at 10 kHz resolution, and also in the water flow experiment with high speed PIV (1 kHz); good agreement between the two experimental techniques is seen.

Strongly periodic events are identifiable in both spectra. In the flow with a central jet, the Strouhal numbers at which these occur (St=1.32 and 2.71) were identified in [15], with the rotation of a two strong/two weak double vortex pair system originating with a breakdown event inside the fuel injector. With no central jet, the largest peaks appear at Strouhal numbers of 0.6 and 1.2. The energy is also distributed over a broader range of frequencies near these peaks. The appearance of multiple peaks confirms that a vortex pair system must still be present in the no jet case, although probably weaker as was seen in the reduced turbulence energy maximum value in Fig. 5. Following the technique described in [15], conditionally averaged vectors were derived for the two cases at injector exit and are shown in Fig. 8, superimposed on a swirl strength parameter. The two strong/two weak vortex pattern is very obvious in the central jet data; with no central jet, only two vortices can be clearly identified, and these are also much less coherent with quite unequal swirl strength parameters. As a consequence of this, the main periodic feature identified as these structures rotate is the passage of the strongest vortex past the measurement point, with a weaker signal from the passage of the vortex pair (since these are less coherent), and a much weaker signal from the hardly visible second vortex pair. This explains why the Strouhal peaks in the no jet case are at half the value of the with jet data.

Further, an additional periodic event is now present in the no jet case that was not present with a jet, namely the PVC. In order to extract quantitative details on the PVC, the decision was taken to look further downstream (at $x/D_s=2.65$), where evidence for a PVC had first been seen in the turbulence statistics. Instantaneous PIV data of velocity vectors in the $r-\theta$ plane are shown in Fig. 9 at two instants of time. The aerodynamic center of the swirling flow can be easily identified, and is clearly precessing in time about the geometric center (identified by the black dot). The variation in time of the angular location of the aerodynamic center was taken from the instantaneous PIV data and is shown in Fig. 10. The sawtooth pattern indicates a coherent rotational motion at the very low frequency of around 0.75 Hz, a Strouhal number of only ~0.01, as defined in the above data using swirl passage exit parameters. If the Strouhal number is redefined (St_{duct}) using the downstream duct diameter (140 mm) as a length scale and the associated bulk mean velocity, then St_{duct} takes on a value of ~0.7, which is in good agreement with the data of Syred [20] on PVC characteristics at a bulk swirl number of around 0.75, as in the present case.

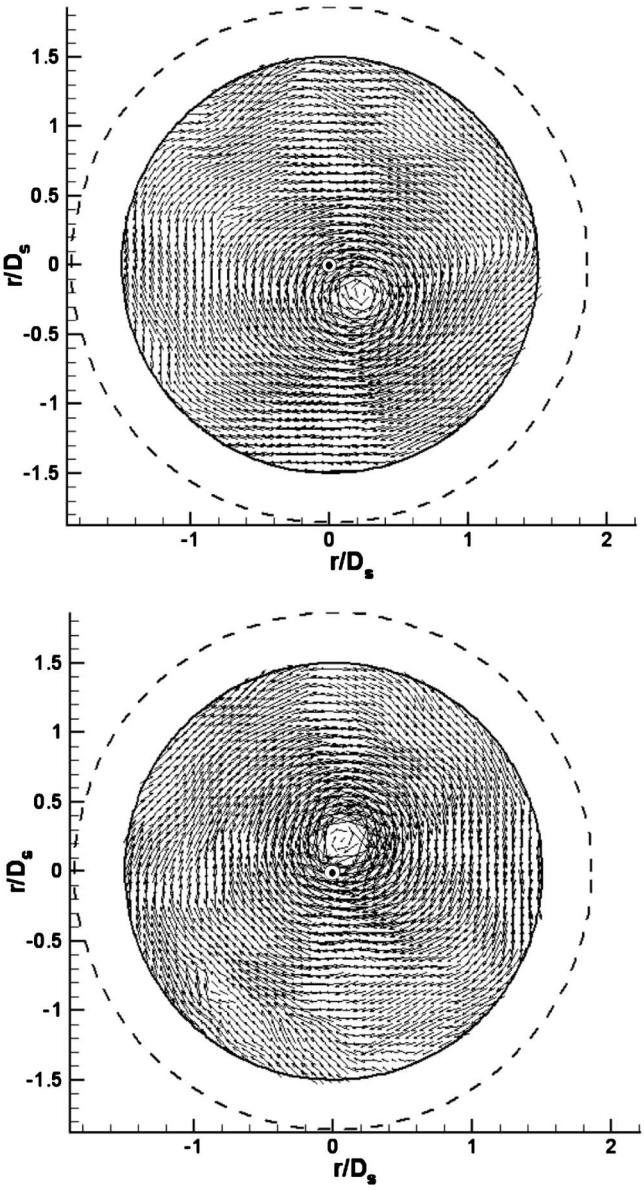


Fig. 9 Instantaneous velocity vectors in $r-\theta$ plane at x/D_s =2.65, no central jet

Modular Injector Geometry Data. The first data taken with the modular injector were for Case 1, which was set up to be the same as the no jet flow case discussed above, except for the removal of the internal corner. Figure 11 presents a series of spectra measured at x/D_s=0.27 with a hot wire in airflow for Case 1 at 25 kHz resolution with swirl slots at α_1=30 deg. Two spikes in the spectra are observed at the same Strouhal numbers as in the datum injector geometry at all radial locations measured. This result shows conclusively that the separation event inside the injector identified in [15] as the origin of the double vortex pair was not associated with the internal corner. By reducing the swirl slot angle (α_1), and hence swirl number, the change in spectral characteristics may be observed in Fig. 12. The intensity of the dominant spikes reduces with swirl level until at α_1=10 deg, where no spikes are observed. Although the data are shown here for the Case 1 swirl passage geometry, the same result was obtained for all Cases studied. The swirl number clearly affects the appearance of the internal vortex breakdown, creating the multiple vortex structures, but the passage geometry has no discernible influence.

Measurements reported in [15] with the datum injector had suggested that the origin of the vortex structures was inside the injector. Using the modular injector geometry, the improved optical access allowed this to be confirmed, see, Fig. 13, where the mean separation point is measured as x/D_s=−0.25. This separation point is also mainly determined by the swirl number, moving downstream to x/D_s=−0.1 for α_1=20 deg, and by α_1=10 deg there is no internal separation (Fig. 14). This unequivocally links

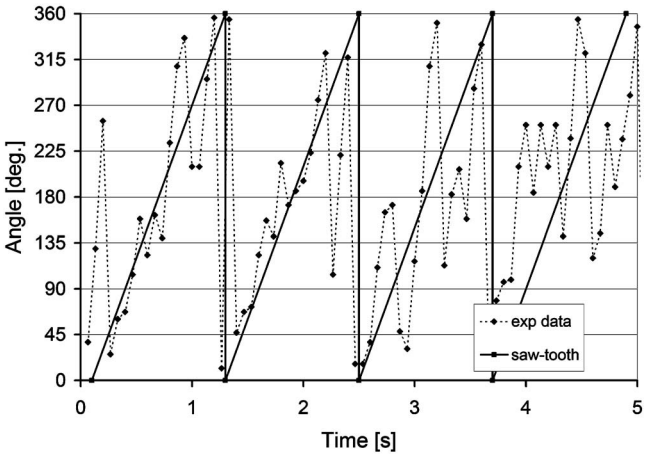


Fig. 10 Angular location of swirl pattern aerodynamic center at x/D_s=2.65, no central jet

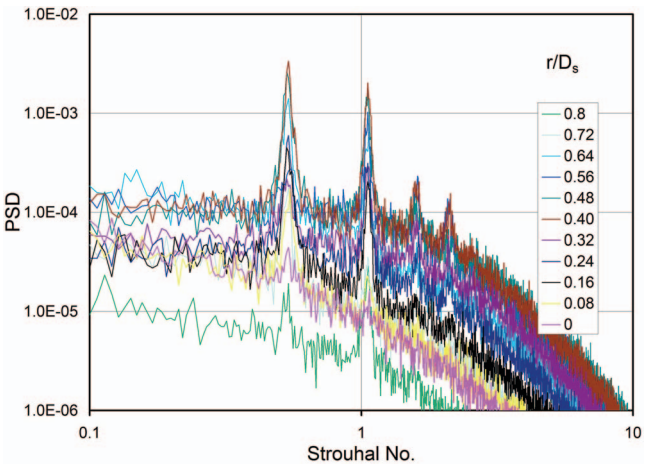


Fig. 11 PSD in modular injector, Case 1, x/D_s=0.27; α_1 =30 deg

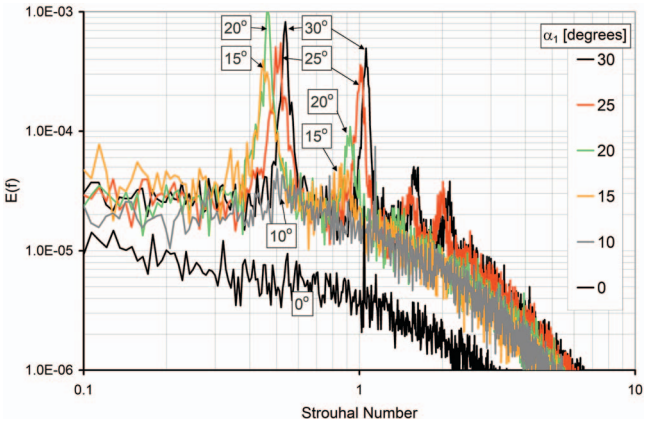


Fig. 12 PSD in modular injector, Case 1; various slot angles

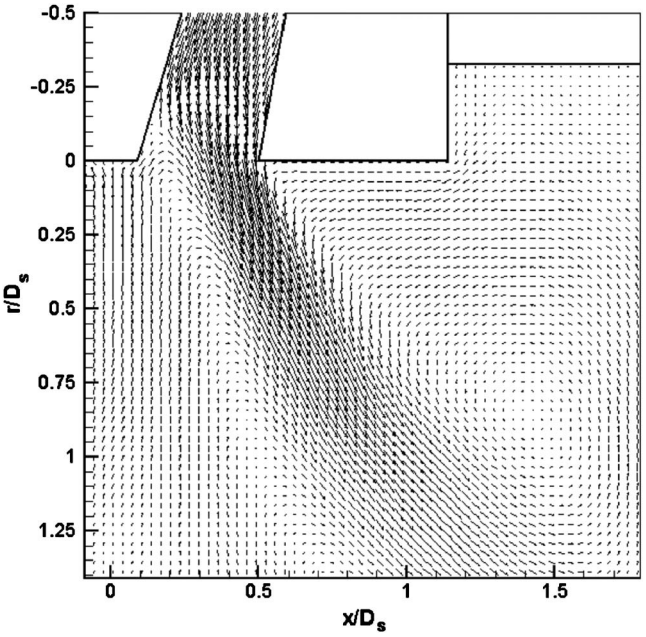


Fig. 13 Velocity vectors, modular injector, Case 1; $\alpha_1=30$ deg

the internal flow separation and the multiple spiraling vortex structures. As a final piece of evidence, Fig. 15 shows the measured velocity vectors for Case 6 with a parallel walled swirl passage, superimposed on the turbulence energy contours. For $\alpha_1=30$ deg and a parallel passage, the resulting swirl number was higher than used before at 0.95. However, there is now no separation inside the injector and the peak turbulence level at 0.35 at the injector exit plane is half that observed previously. From hot wire data taken at $x/D_s=0.27$, providing the PSD data shown in Fig. 16, evidence was found of spikes in the spectra but for this geometry only for $\alpha_1>20$ deg. The implication is that the spiraling vortices have still appeared at high enough swirl numbers, but their origin is now outside the injector. The most likely explanation is that the formation point of the vortices is determined by an interaction between the forward flow in the swirlstream passage and the backflow occurring in the CRZ, with the swirl number being the dominant parameter controlling this, and the swirl passage geometry playing an indirect role, since changes here will influence the precise swirl number at injector exit.

Conclusions

A detailed experimental study has been reported of the near-field injector flowfields created as various flow and geometric parameters of the injector were varied. The main emphasis was on the effect of the presence or absence of a central nonswirling jet in addition to the annular swirlstream. The absence of the jet was shown to lead to large destabilization of the flow. The CRZ began to precess creating a PVC, and large levels of velocity fluctuations were measured as a consequence on the centerline for a significant region downstream of the injector exit plane. The PVC unsteadiness influenced the precise nature of the vortex breakdown, which was confirmed to occur inside the injector for the baseline flow conditions studies. The swirler cone itself also became more unstable without a central jet, producing a large increase in the variation in time of its attachment location on the confining geometry outer wall. Geometry variations in the swirl passage were examined but had only minor influence on the flow characteristics, except for their influence on the injector exit swirl number, which was the dominating parameter. Variations in the swirl number were shown to lead always to a spiral form of vortex breakdown with multiple vortex structures wrapped around the CRZ. Depend-

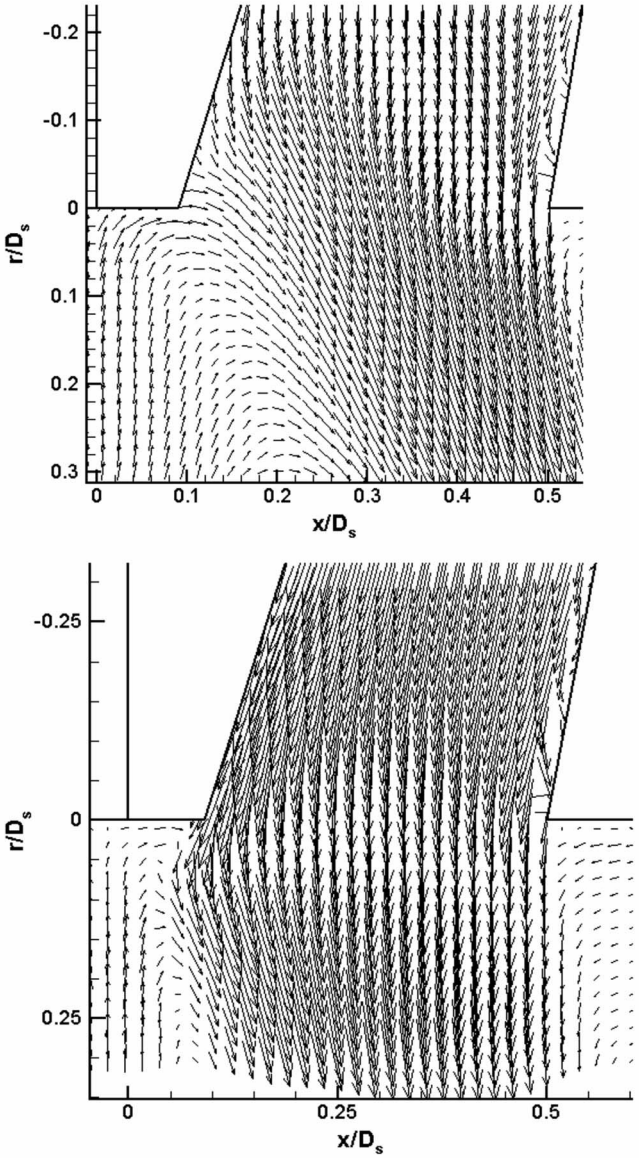


Fig. 14 Velocity vectors, modular injector, Case 1; (top) $\alpha_1=20$ deg; (bottom) $\alpha_1=10$ deg

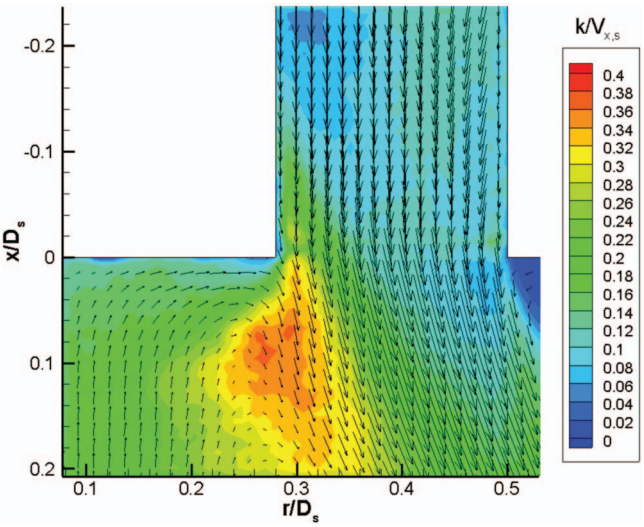


Fig. 15 Velocity vectors and *k* contours, modular injector, Case 6; $\alpha_1=30$ deg

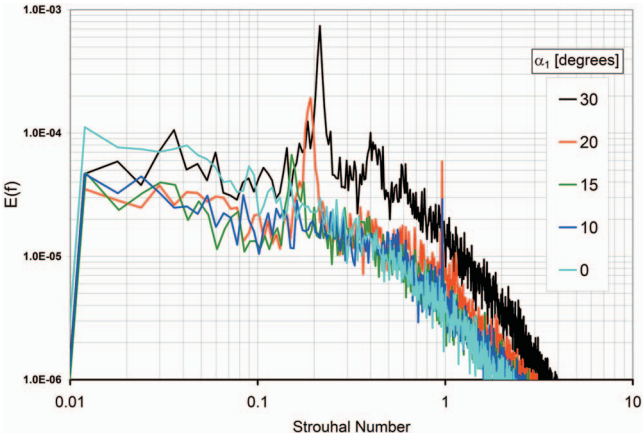


Fig. 16 PSD for modular injector, Case 6; various slot angles

ing on the geometry, the precise value of swirl number at which these spiraling structures appeared varied, as did the location of their first appearance, which could be both inside and outside the injector.

Acknowledgment

The work reported here was carried out in the Loughborough UTC in Combustion Aerodynamics with funding provided by EU project "MOLECULES" (No. G4RD-CT-2000-00402).

Nomenclature

D = cylindrical combustor duct diameter
D_j = central jet diameter
D_s = annular swirler passage diameter
f = frequency of periodic peaks in PSD data
k = turbulence kinetic energy, or direction k
L_{ij}^k = length scale of i/j velocity fluctuation correlation in k direction
r = separation vector
R_{ij} = spatial correlation function
S = swirl number
St = dimensionless Strouhal number ($fD_s/V_{x,s}$)
u_i, u_j = velocity fluctuations in i, j directions
V_x, V_r, V_t = time-mean velocities in x, r, θ directions
$V_{x,j}$ = bulk axial velocity of central jet stream
$V_{x,s}$ = bulk axial velocity of annular swirl stream
v_x', v_r', v_t' = fluctuating RMS in x, r, θ directions
x, r, θ = cylindrical polar coordinates
Δt = interframe time
δz = camera depth of field

References

[1] Peckham, D. H., and Atkinson, S. A., 1957, "Preliminary Results of Low Speed Wind Tunnel Tests on a Gothic Wing of Aspect Ratio 1.0," ARC Technical Report CP No. 508, TN No. Aero 2504.
[2] Gupta, A. K., Lilley, D. G., and Syred, N., 1984, *Swirl Flows*, Abacus Press, Cambridge, MA.
[3] Wang, S., and Rusak, Z., 1997, "The Dynamics of a Swirling Flow in a Pipe and Transition to Axisymmetric Vortex Breakdown," J. Fluid Mech., **3**, pp. 177–223.
[4] Escudier, M., 1988, "Vortex Breakdown: Observations and Explanations," Prog. Aerosp. Sci., **25**, pp. 189–229.
[5] Lucca-Negro, O., and O'Doherty, T., 2001, "Vortex Breakdown: A Review," Prog. Energy Combust. Sci., **27**, pp. 431–481.
[6] Sarpkaya, T., 1971, "On Stationary and Travelling Vortex Breakdowns," J. Fluid Mech., **45**, pp. 545–559.
[7] Leibovich, S., 1978, "The Structure of Vortex Breakdown," Annu. Rev. Fluid Mech., **10**, pp. 221–246.
[8] Syred, N., and Beer, J. M., 1972, "The Damping of Precessing Vortex Cores by Combustion in Swirl Generation," Astronaut. Acta, **17**, pp. 783–801.
[9] Benjamin, T. B., 1962, "Theory of the Vortex Breakdown Phenomenon," J. Fluid Mech., **14**, pp. 593–629.
[10] Escudier, M., Nickson, K., and Poole, R., 2006, "Intensely Swirling Turbulent Pipe Flow Downstream of an Orifice: The Influence of an Outlet Contraction," *13th Int. Symp. on Appl. of Laser Techniques in Fluid Mechanics*, Lisbon, Portugal.
[11] Li, G., and Gutmark, E. J., 2003, "Geometry Effects on the Flow Field and the Spectral Characteristics of a Triple Annular Swirler," ASME Paper No. GT-2003-38799.
[12] Wegner, B., Janis, B., Sadiki, A., Dreizler, A., and Janicka, J., 2005, "Study of Flow and Mixing in a Generic GT Combustor Using LES," *Eng. Turb. Modelling and Measurements 6*, W. Rodi, ed., Elsevier Ltd., Oxford, U.K., pp. 731–739.
[13] Garcia-Villalba, M., and Froehlich, J., 2005, "On the Sensitivity of a Free Annular Swirling Jet to the Level of Swirl and a Pilot Jet," *Eng. Turb. Modelling and Measurements 6*, W. Rodi, ed., Elsevier Ltd., Oxford, U.K., pp. 845–854.
[14] Midgley, K., Spencer, A., and McGuirk, J. J., 2003, "PIV Measurements of Confined Swirling Flow," *Proc. of 3rd TSFP Symposium*, Sendai, Japan.
[15] Midgley, K., Spencer, A., and McGuirk, J. J., 2005, "Unsteady Flow Structures in Radial Swirler Fed Fuel Injectors," ASME J. Eng. Gas Turbines Power, **127**, pp. 755–764.
[16] Keane, R. D., and Adrian, R. J., 1992, "Theory of Cross-Correlation Analysis of PIV Images," Appl. Sci. Res., **49**, pp. 191–215.
[17] LaVision GmbH, 2000, "PIV Flowmaster Manual."
[18] Spencer, A., and Hollis, D., 2005, "Correcting for Sub-Grid Filtering Effects in Particle Image Velocimetry Data," Meas. Sci. Technol., **16**, pp. 2323–2335.
[19] Midgley, K., 2005, "An Isothermal Experimental Study of the Unsteady Fluid Mechanics of Gas Turbine Fuel Injector Flowfields," Ph.D. thesis, Loughborough University.
[20] Syred, N., O'Doherty, T., and Froud, D., 1994, "The Interaction of the Precessing Vortex Core and Reverse Flow Zone in the Exhaust of a Swirl Burner," Bussei Kenkyu, **208**, pp. 27–35.

Reduced Reaction Mechanisms for Methane and Syngas Combustion in Gas Turbines

N. Slavinskaya

M. Braun-Unkhoff

P. Frank

DLR,
Institute of Combustion Technology,
Pfaffenwaldring 38-40,
Stuttgart, Germany

Two reduced reaction mechanisms were established that predict reliably for pressures up to about 20 bar *the heat release for different syngas mixtures including initial concentrations of methane. The mechanisms were validated on the base of laminar flame speed data covering a wide range of preheat temperature, pressure, and fuel-air mixtures. Additionally, a global reduced mechanism for syngas, which comprises only two steps, was developed and validated, too. This global reduced and validated mechanism can be incorporated into CFD codes for modeling turbulent combustion in stationary gas turbines.* [DOI: 10.1115/1.2719258]

Keywords: mechanism reduction, heat release, syngas, flame speed

Introduction

The introduction of syngas combustion technology into the new class of advanced high firing temperature and high efficiency gas turbines leads to a distinct economical advantage in comparison with the existing technology. Improvements on the fuel flexibility of the syngas combustion technology with optimization of the design will provide for a maximum of acceptable range in the variation of fuel composition and conditions. Using advanced syngas combustion technology in high firing temperature engines gives the opportunity to use solid fuels very efficiently at low emission levels compared with other technologies and therefore enables the use of low quality feedstock for power production. At existing syngas combustion systems the back-up fuel, in particular natural gas, is burned in diffusion mode. Especially for the high firing temperature combustion systems, special measures have to be made in order to keep the NOx emission for back-up fuel operation below the currently acceptable limit of 25 ppm and to generate a potential for further reduction.

To overcome these problems there is an urgent need for a reliable simulation of combined cycles of co-firing natural gas and syngas. The prediction of the combustion behavior of fuels with increased hydrogen content from different feedstock like coal, refinery residues, and biomass is a necessary precondition for establishing numerical tools in order to verify a proper design already at early development stages.

Therefore, reduced and validated reaction mechanisms were established that predict reliably the heat release for different syngas flames, partially including initial concentrations of methane, under lean to stoichiometric conditions and pressures up to about 20 bar. The ultimate aim was to generate reaction models suitable for implementation into CFD codes for modeling turbulent combustion in high efficiency gas turbines.

Modeling Simulations for Methane Combustion

Reduced Reaction Mechanisms From Literature. For modeling turbulent combustion, CFD codes need reduced kinetical mechanisms. In the present work different reduced reaction mechanisms of methane combustion existing in the literature for the prediction of experimental data for flame speed in CH_4/air systems were checked. Six of these were taken into consideration [1–6]. The number of species and reactions of all reduced mechanisms are given in Table 1.

The results of modeling compared with experimental data [7–12] are presented in Figs. 1–3. Besides the full mechanism [13], the reduced mechanisms [1,6] lead to similar good agreement with experimental [7–10] flame speed data from atmospheric laminar methane-air flames (Fig. 1). But none of the reduced mechanisms was able to predict correctly measured laminar flame speed data for methane at elevated pressures (Figs. 2 and 3) for different stoichiometric ratios, especially for lean mixtures, which are most interesting for our investigation.

Selection of the Detailed Reaction Mechanism. As a first base reaction model, the well proved GRI 3.0 [13] mechanism, which comprises 53 species and 325 elementary reaction steps, was chosen. Figures 1–3 show the excellent agreement between measured flame speed data of laminar premixed methane-air flame [7–12] over a large variety of ϕ=fuel/air ratios, preheat temperatures, and pressures. Consequently, the detailed reaction mechanism GRI 3.0 [13] that gave the best agreement was used as the base model for constructing a reduced mechanism. As combustion in gas turbines is mainly conducted under fuel lean conditions at $\varphi \approx 0.5$, special attention should be paid to the calculation of lean mixtures in order to validate a reaction mechanism appropriate for these conditions. Unfortunately, no experimental data are available for this range of the air-fuel ratio.

Development of a Reduced Mechanism for Methane Combustion

Optimized Reduced Mechanism for Methane Combustion. Based on the foregoing results it was decided to reduce the selected full mechanism (GRI 3.0) with respect to the desired conditions (high pressures, lean condition). Special attention was paid to the number of species and reactions incorporated in the reduced mechanism as this will limit its use in commercial CFD codes. Of course, in order to keep time for simulation low and to enable the use of commercial CFD codes like CFX [14] and Fluent [15], a reduction of species and reactions to less than 20 species and about 100 reactions is needed, especially under turbulent conditions.

Therefore, a reduced "skeletal" mechanism was constructed for describing the heat exchange of methane/air flames within a wide range of parameters. Skeletal mechanisms contain only the most

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received August 29, 2005; final manuscript received January 15, 2007; published online January 30, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2005: Land, Sea and Air (GT2005), June 6–9, 2005, Reno, NV. Paper No. GT2005-68287.

Table 1 Reduced reaction mechanisms

Reference	Species	Reactions	Remarks
Earn et al. [1]	19	46	
Boni and Penner [2]	19	23	
Jazbec et al. [3]	17	28	
Sher and Refael [4]	19	39	Irreversible
Bilger and Starner [5]	18	58	
Kazakov and Frenklach [6]	21	84	GRI 1.2 red

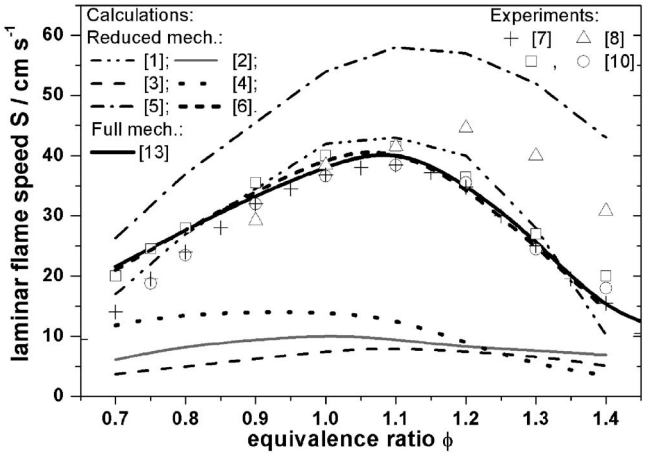


Fig. 1 Laminar flame speed of methane/air mixtures at atmospheric pressure and a preheat temperature T_0=298 K. Comparison between *measurement* (symbols) [7,8,10] and *calculation* (curves) with different *reduced* mechanisms [1–6] and *full* mechanism (GRI_3.0 [13]).

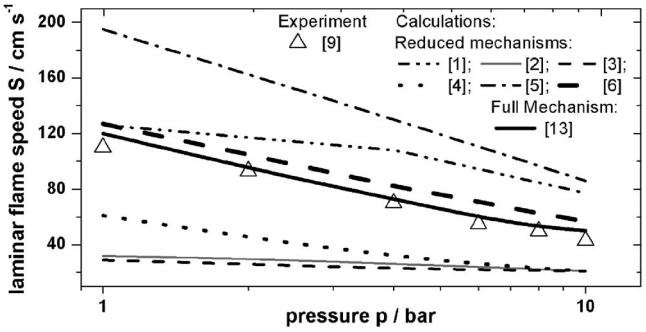


Fig. 2 Laminar flame speed of stoichiometric methane/air mixtures at T_0=573 K at p=1–10 bar. Comparison between experiment (symbols [9]) and calculation (curves).

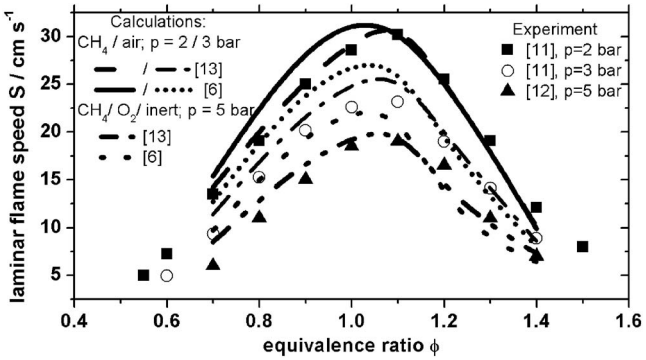


Fig. 3 Laminar flame speed of methane/air and methane/oxygen/inert flames at elevated pressures for a preheat temperature T_0=300 K. Comparison between measurement (symbols) [11,12] and calculation (curves).

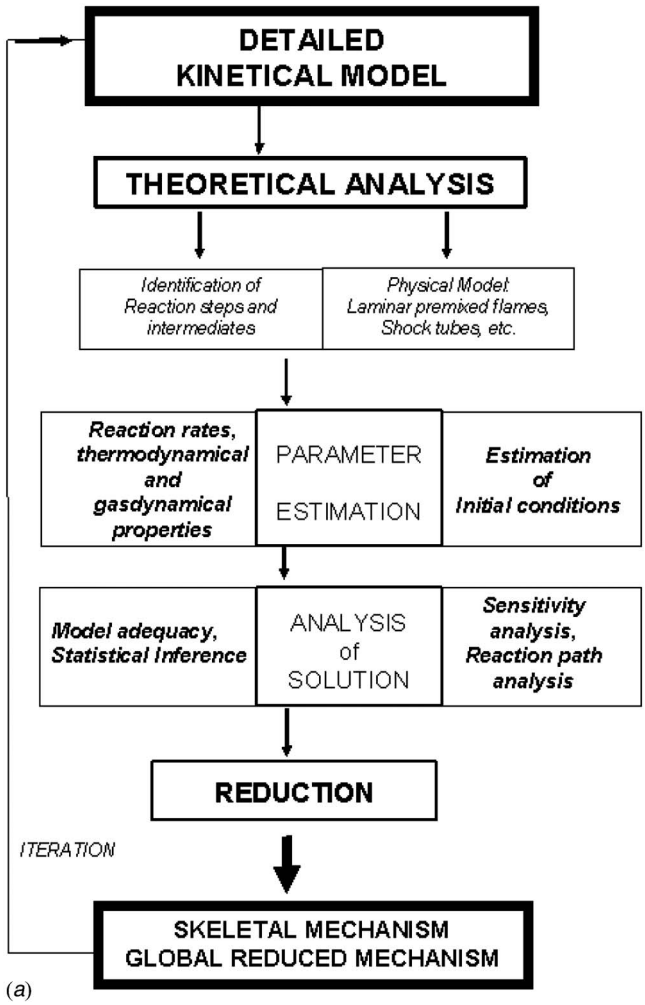


Fig. 4 (*a*) Strategy for reduction of a reaction mechanism. (*b*) Elimination of redundant species: sensitivity plot for an atmospheric methane/air flame; T_0=298 K, φ=1.

essential species and reaction steps for syngas and methane, respectively. The identification of redundant species and reactions was made using tools of the KINALC package [16] according to the strategy presented on Fig. 4(*a*). The detailed reaction mechanism was studied mainly by sensitivity and rate of production analysis for different conditions in order to reduce the numbers of

Table 2 Elaborated reduced reaction mechanisms

Reference	Species	Reactions	Remarks
This work; Reduced mechanism I	19	86	Based on GRI 3.0
This work; Reduced mechanism II	12	20	Irreversible; without C_xH_y
This work; Global reduced Mechanism	5	2	Irreversible; syngas flame, high p; lean

species and reactions. As an example, Fig. 4(*b*) presents the selection of the redundant components on the base of overall sensitivity coefficients [16] for a group of species and the temperature, at a certain height above burner surface.

Thus, a reduced mechanism (reduced mechanism I, Table 2) for methane/air combustion involving 19 species (H_2, H, O, O_2, OH, H_2O, HO_2, CH_2, CH_3, CH_4, CO, CO_2, HCO, CH_2O, CH_2OH, CH_3O, CH_3OH, N_2, Ar) and 86 reactions was obtained. On the base of sensitivity and reaction rate analysis to flame speed, it was found that by modifying the rate coefficients of two reactions $OH+CO \rightarrow CO_2+H$ *and* $HCO+M \rightarrow CO+H+M$, a much better agreement was achieved (Figs. 5–8). These new data were taken from the CEC data evaluation by Baulch et al. [17]. All the other kinetic parameters are identical to the base model [13]. This improved reduced mechanism I leads to only slightly reduced flame speed data for stoichiometric mixtures (also shown in Figs. 5–7).

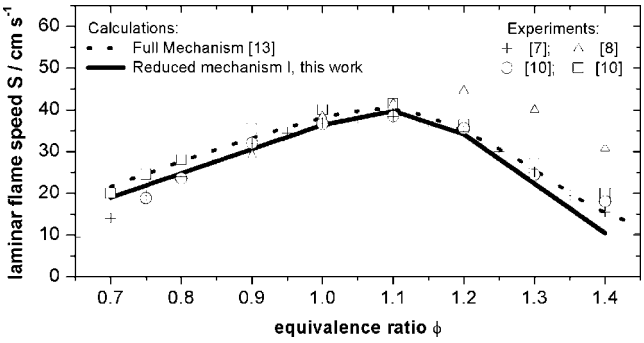


Fig. 5 Laminar flame speed of atmospheric methane/air flames for T_0=298 K. Comparison between measured (symbols) [7–10] and calculated (curves) values.

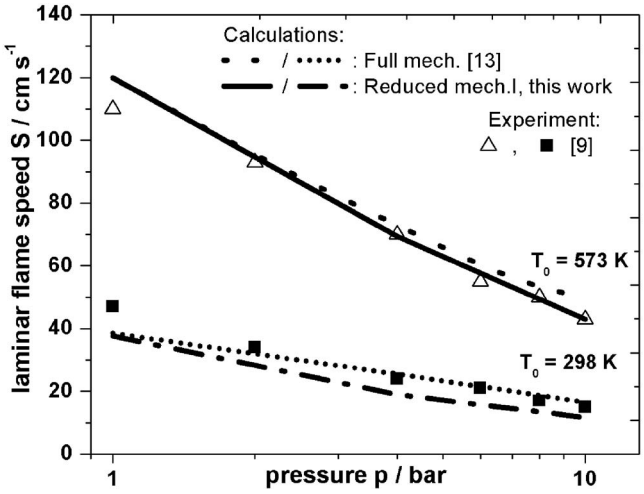


Fig. 6 Stoichiometric methane air flames at different pressures and preheat temperatures. Comparison between measurement (symbols) [9] and calculation (curves).

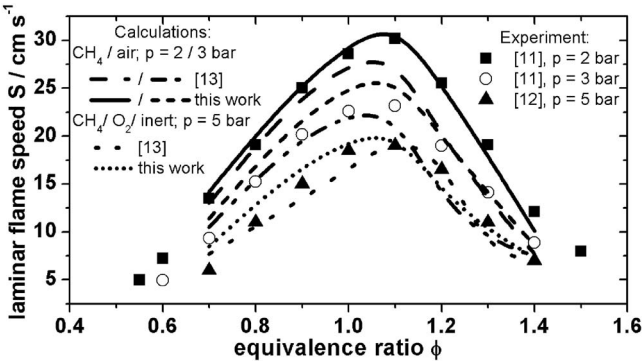


Fig. 7 Laminar flame speed of methane/air flames at elevated pressures at a preheat temperature T_0=300 K. Comparison between *measured* (symbols) [11,12] and *calculated* (curves) values.

Development of a Reduced Mechanism for Syngas Combustion

CO/Air-Flames With Addition of H_2 and/or CH_4. The calculation of the laminar flame speed of CO-air mixtures with the addition of methane by using the reduced mechanism I (Table 2) discussed above gives significantly better agreement with the measured data than the full [13] and reduced [6] mechanisms from literature over the whole range of mixture composition (Fig. 8).

For the development of a skeletal mechanism for modeling the flame speed of different syngas air mixtures containing CO, H_2 and H_2O experiments from different groups were used [18–21]. As the simulation of the flame speed data in Figs. 9–13 revealed that the reduced mechanism I (Table 2) was not able to reproduce under all conditions the flame speed profiles for mixtures with higher hydrogen content, the reduced mechanism GRI 1.2 [6] was used for comparison (Figs. 9–11 and 13).

Using these mixtures for validating the reaction mechanisms, it was found that the reduced mechanism II elaborated for syngas/air flames could be further reduced to 12 species (inclusively Ar and N_2) and 20 irreversible reactions (Table 3). In order to achieve best agreement with data from flames with high hydrogen fuel content, the rate coefficients of reactions 8, 16, and 17 were taken from Refs. [22,23]. This reaction mechanism now leads to a better agreement between experiment and prediction than the published GRI 1.2 mechanism (Figs. 9–11 and 13) in spite of the fact that the elaborated syngas skeletal mechanism contains significantly fewer species (12 instead of 21 species).

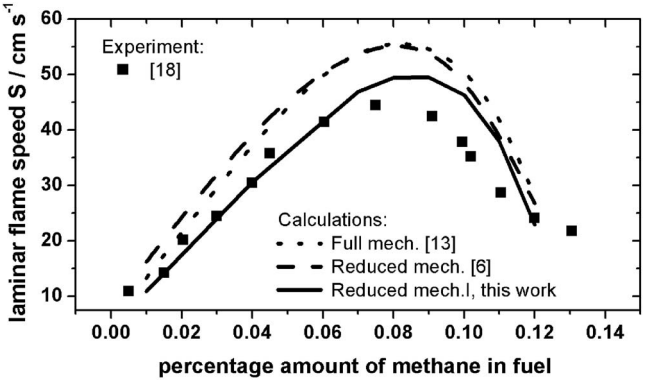


Fig. 8 Laminar flame speed of atmospheric premixed carbon monoxide/methane air flames with different amounts of methane ($X_{CO}+X_{CH4}$=0.15) at T_0=300 K. Symbols: experiment [18], curves: calculations.

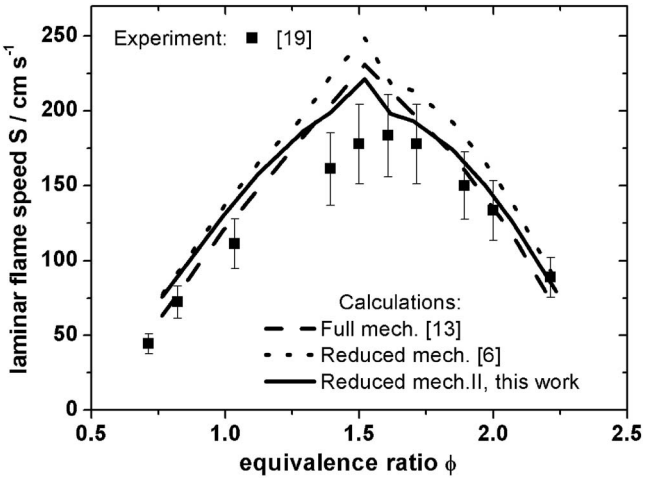


Fig. 9 Laminar flame speed of atmospheric premixed 50% CO/50% H_2 air flames at T_0=298 K. Symbols: experiment [19]; curves: calculations.

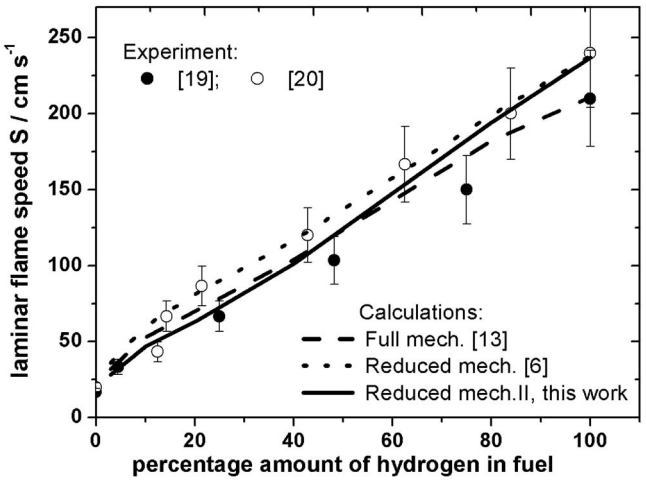


Fig. 10 Laminar flame speed of atmospheric premixed CO/air flames with different amounts of hydrogen at a preheat temperature T_0=298 K. Symbols: experiment [19,20]; curves: calculations.

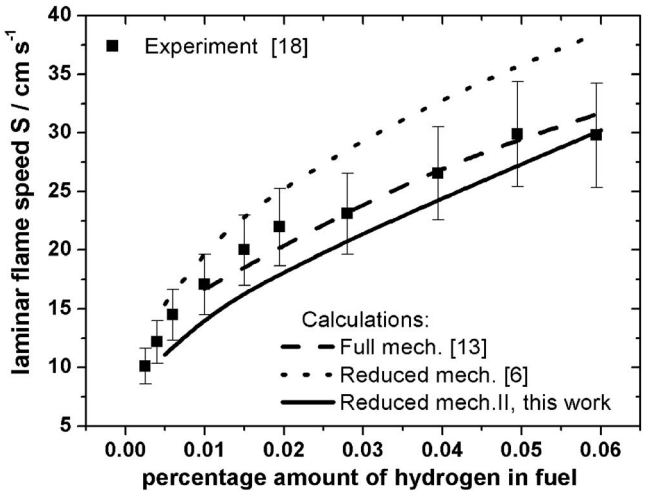


Fig. 11 Laminar flame speed of atmospheric premixed carbon monoxide/hydrogen air flames with different amounts of hydrogen ($X_{CO}+X_{H2}$=0.20) at T_0=300 K. Symbols: experiment [18]; curves: calculations.

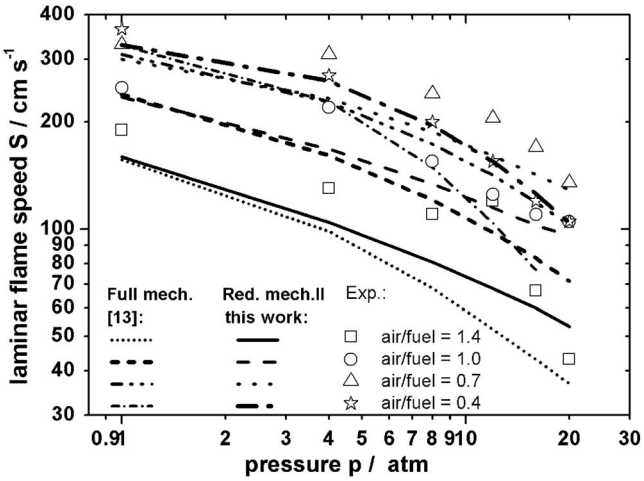


Fig. 12 Laminar flame speed of premixed syngas flames (0.37 CO, 0.35 H_2, 0.25 CO_2, 0.03 N_2) at different pressures and air/fuel ratios for T_0=573 K. Symbols: experiment [21]; curves: calculations with full [13] and reduced mechanism II (this work, Table 2).

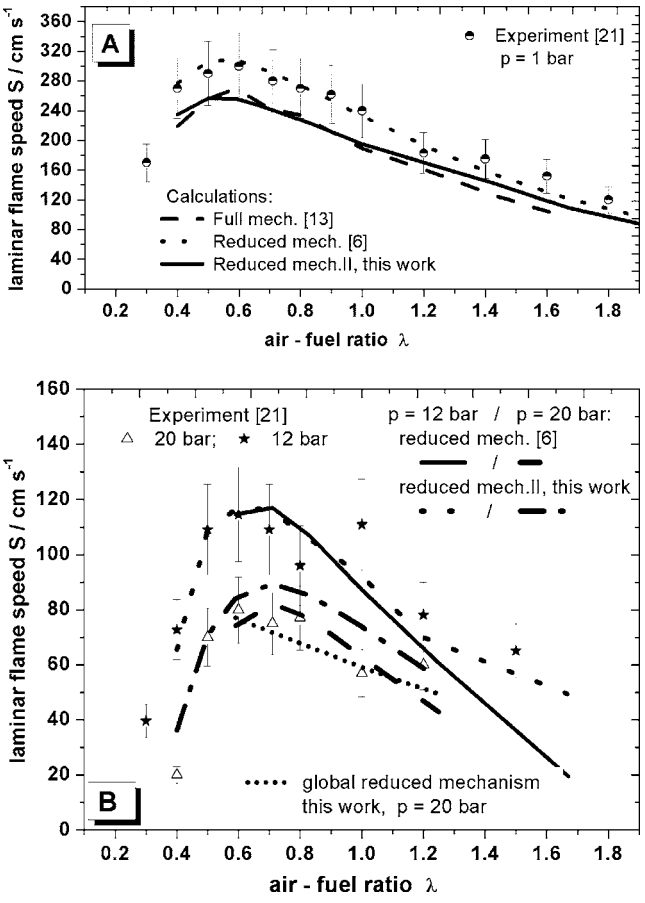


Fig. 13 Laminar flame speed of premixed laminar syngas flames (0.28 H_2, 0.296 CO, 0.20 CO_2, 0.20 H_2O, 0.024 N_2) at atmospheric (A) and at elevated (B) pressures at T_0=573 K. Symbols: experiment [21]; Curves: calculations with full [13], reduced mechanism II (this work, Table 2), and global reduced mechanism (this work, reactions I+II).

Table 3 Developed skeletal mechanism for syngas combustion. $k = A \cdot T^n \exp(-E_a/T)$ in cm³, s, K.

Reactions	A	n	E_a	Reference
1. $O+H_2 \rightarrow H+OH$	3.87E+04	2.7	3151.0	[13]
2. $H+OH \rightarrow O+H_2$	2.91E+04	2.64	2236.0	[13]
3. $O+HO_2 \rightarrow OH+O_2$	2.00E+13	0.0	0.0	[13]
4. $H+O_2+M \rightarrow HO_2+M$	3.50E+16	−0.41	−558.0	[13]
5. $H+O_2 \rightarrow O+OH$	2.65E+16	−0.67	8576.0	[13]
6. $O+OH \rightarrow H+O_2$	6.96E+13	−0.27	−108.0	[13]
7. $2H+H_2O \rightarrow H_2+H_2O$	6.00E+19	−1.25	0.0	[13]
8. $H+OH+M \rightarrow H_2O+M$ $H_2O/3.5/CO/0.75/$ $CO_2/2.5/O_2/0.5/$	2.20E+22	−1.95	0.0	[22]
9. $H+HO_2 \rightarrow O+H_2O$	3.97E+12	0.0	338.0	[13]
10. $H+HO_2 \rightarrow O_2+H_2$	4.48E+13	0.0	538.0	[13]
11. $H+HO_2 \rightarrow 2OH$	8.40E+13	0.0	320.0	[13]
12. $H+HCO \rightarrow H_2+CO$	7.34E+13	0.0	0.0	[13]
13. $OH+H_2 \rightarrow H+H_2O$	2.16E+08	1.51	1726.0	[13]
14. $2OH \rightarrow O+H_2O$	3.57E+04	2.4	−1062.0	[13]
15. $O+H_2O \rightarrow 2OH$	4.49E+06	2.1	7977.0	[13]
16. $OH+CO \rightarrow H+CO_2$	1.30E+07	1.3	−380.0	[23]
17. $H+CO_2 \rightarrow OH+CO$	8.38E+13	0.02	12781.0	[23]
18. $HCO+M \rightarrow H+CO+M$ $H_2O/3.5/CO/0.75/$ $CO_2/2.5/O_2/0.5/$	1.87E+17	−1.0	8556.0	[13]
19. $H+CO+M \rightarrow HCO+M$ $H_2O/3.5/CO/0.75/$ $CO_2/2.5/O_2/0.5/$	1.18E+15	−0.35	127.0	[13]
20. $HCO+O_2 \rightarrow HO_2+CO$	1.35E+13	0.0	201.0	[13]

Global Mechanism Generated for Syngas Flames. As some CFD codes need a further reduced mechanism for describing the combustion processes in gas turbines, a global reduced mechanism for syngas was also elaborated. With the KINALC code [16] the analysis of lifetimes of species and the error of application of the quasi-steady-state approximation (QSSA) were performed and QSS species were selected. Using the nonlinear set of algebraic equations obtained after that selection, concentrations of QSS species were produced as functions of the concentrations of non-QSS species and rates of elementary reactions. With linear algebra [24], a global mechanism that comprises five species including educt and product species and two global reactions was elaborated. The reactions rates for the global mechanism were approximated for a pressure of 20 bar to CHEMKIN format and can be used with the CHEMKIN III package.

$$\text{I } 2H_2 + O_2 \rightarrow 2H_2O$$

$$R_I = 10.8 \times 10^{-7} T^{6.1} \exp(-4876.7/T) \cdot [H_2]^{2.0}$$

$$\text{II } CO + O_2 + H_2 \rightarrow CO_2 + H_2O$$

$$R_{II} = 20.15 \times 10^{-8} T^{5.9} \exp(-3070.5/T) \cdot [O_2][CO]^{1.4}$$

For the prediction of the laminar flame speed of a syngas composition at elevated pressures under air rich conditions, good agreement between global and reduced mechanisms is achieved (Fig. 13). It is also to be seen that the global mechanism is only valid for a relatively narrow mixture range $\phi = 0.8-1.6$ at 20 bar and fails to predict the decrease in flame speed in the low air-fuel ratios.

Conclusion and Outlook

The detailed mechanism GRI 3.0 was reduced for modeling flame speed data in CH_4, CH_4+CO/air and syngas/air mixtures.

A reduced reaction mechanism for CH_4, CH_4+CO/air mixtures with 19 species and 86 reactions was constructed that predicts reliably the heat release for pressures up to about 10 bar over a wide range of air/fuel ratios.

For different syngas/air mixtures containing CO, H_2, and H_2O, a skeletal mechanism with 12 species and 20 irreversible reactions was developed and validated for pressures up to 20 bar and different mixture contents and fuel/air ratios.

These two generated reduced mechanisms lead to a better agreement with experimental data than reduced mechanisms taken from literature.

Starting from a skeletal mechanism constructed for syngas mixtures, a global two-step mechanism with five species was elaborated for a restricted range of parameters ($p = 20$ bar, air/fuel ratio 0.6–1.2).

These validated mechanisms can be used in CFD calculations to model turbulent combustion under practical conditions in gas turbines using methane or syngases as fuels.

Acknowledgment

This work was carried out within the EU project HEGSA FP5, Project No. NNE5-2001-00644; financial support is gratefully acknowledged. The authors wish to thank K. Huang for her help in performing calculations.

References

[1] Ern, A., Douglas, C. C., and Smooke, M. D., 1991, "Reduced Kinetic Mechanisms and Asymptotic Approximations for Methane-Air Flames," *Lecture Notes in Physics*, Springer Verlag, New York, p. 384.
[2] Boni, A. A., and Penner, R. C., 1977, "Sensitivity Analysis of a Mechanism for Methane Oxidation Kinetics," Combust. Sci. Technol., **15**, pp. 99–106.
[3] Jazbec, M., Fletcher, D. F., and Haynes, B. S., 2000, "Simulation of the Ignition of Lean Methane Mixtures Using CFD Modelling and a Reduced Chemistry Mechanism," Appl. Math. Model., **24**, pp. 689–696.
[4] Sher, E., and Refael, S., 1988, "A Simplified Reaction Scheme for the Combustion of Hydrogen Enriched Methane/air Flame," Combust. Sci. Technol., **59**, pp. 371–389.
[5] Bilger, R. W., and Starner, S. H., 1990, "On Reduced Mechanisms for Methane-Air Combustion in Nonpremixed Flames," Combust. Flame, **80**, pp. 135–149.
[6] Kazakov, A., and Frenklach, M., 1994, GRI 1.2_red, http://www.me.berkeley.edu/drm/.
[7] Van Maaren, A., Thung, D. S., and De Goey, L. P., 1994, "Measurement of Flame Temperature and Adiabatic Burning Velocity of Methane/Air Mixtures," Combust. Sci. Technol., **96**, pp. 327–344.
[8] Gibbs, G. J., and Calcote, H. F., 1959, "Effect of molecular Structure on Burning Velocity," J. Chem. Eng. Data, **4**(3), pp. 226–237.
[9] Eberius, H., and Kick, T., 1992, "Stabilization of Premixed, Conical Methane Flames at High Pressure," Ber. Bunsenges. Phys. Chem., **96**, pp. 1416–1419.
[10] Zhu, D. L., Egolfopoulos, F. N., and Law, C. K., 1988, "Experimental and Numerical Determination of Laminar Flame Speeds of Methane/(Ar, N_2, CO_2)-Air Mixtures as Function of Stoichiometry, Pressure, and Flame Temperature," Proc. Combust. Inst., **22**, pp. 1537–1545.
[11] Egolfopoulos, F. N., Cho, P., and Law, C. K., 1989, "Laminar Flame Speeds of Methane-Air Under Reduced and Elevated Pressures," Combust. Flame, **76**, pp. 375—391.
[12] Rozenchan, G., Tse, S. D., Zhu, D. L., and Law, C. K., 2001, "Laminar Burning Rates and Markstein Lengths of CH4/O2/Inert Mixtures at High Pressure," Paper No. AIAA -2001–1080.
[13] Smith, G. P., Golden, D. M., Frenklach, M., Moriarty, N. W., Eiteneer, B., Goldenberg, M., Bowman, C. T., Hanson, R. K., Song, S., Gardiner Jr., W. C., Lissianski, J., and Qin, Z., 1999, GRI 3.0 mechanism, Version 3.0 7/30/99. See http://www.me.berkeley.edu/gri_mech
[14] ANSYS CFX-5, http://www-waterloo.ansys.com/cfx
[15] Fluent Inc., Fluent 6.1; see www.fluent.com
[16] Turanyi, T., "KINALC: a CHEMKIN base program for kinetic analyses," available at http://www.chem.leeds.ac.uk/Combustion/kinalc.htm. Tomlin, A. S., Turanyi, T., and Pilling, M. J., 1997, "Mathematical Tools for the Construction, Investigation and Reduction of Combustion Mechanism," *Low Temperature Combustion and Autoignition*, M. J. Pilling and G. Hancock, eds., Elsevier, Amsterdam, pp. 293–437.
[17] Baulch, D. L., Cobos, C. J., Cox, R. A., Frank, P., Hayman, G., Just, Th., Kerr,

J. A., Murrells, T., Pilling, M. J., Troe, J., Walker, R. W., and Warnatz, J., 1994, "Evaluated Kinetic Data for Combustion Modelling Supplement I," J. Phys. Chem. Ref. Data, **23**, pp. 847–1031.
[18] Vagelopoulos, C. M., and Egolfopoulos, F. N., 1994, "Laminar Flame Speeds and Extinction Strain Rates of Mixture of Carbon Monoxide With Hydrogen, Methane and Air," Proc. Combust. Inst., **25**, pp. 1317–1323.
[19] McLean, I. C., Smith, D. B., and Taylor, S. C., 1994, "The Use of Carbon Monoxide/Hydrogen Burning Velocities to Examine the Rate of the CO+OH reaction," Proc. Combust. Inst., **25**, pp. 749–757.
[20] Günther, R., and Janisch, G., 1971, "Messwerte der Flammengeschwindigkeit von Gasen und Gasgemischen," Chem.-Ing.-Tech., **43**, pp. 975–978.
[21] Lauer, G., and Leuckel, W., 1994, AG Turbo, 4. Status Seminar, University Karlsruhe.
[22] Mueller, M. A., Yetter, R. A., and Dryer, F. L., 1998, "Measurement of the Rate Constant for $H+O_2=HO_2+M$," Proc. Combust. Inst., **27**, pp. 177–184.
[23] Yetter, R. A., and Dryer, F. L., 1991, "A Comprehensive Mechanism for Carbon Monoxide/Hydrogen/Oxygen Kinetics," Combust. Sci. Technol., **79**, pp. 97–128.
[24] Gorsky, W. G., 1984, in *Planning of Kinetical Experiments*, Nauka, Moscow, p. 241 (in Russian).

Engine Design and Operational Impacts on Particulate Matter Precursor Emissions

Stephen P. Lukachko
e-mail: sluka@mit.edu

Ian A. Waitz
e-mail: iaw@mit.edu

Gas Turbine Laboratory, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139

Richard C. Miake-Lye
e-mail: rick@aerodyne.com

Robert C. Brown

Center for Aerothermodynamics, Aerodyne Research, Inc., Billerica, MA 01821

Aircraft emissions of trace sulfur and nitrogen oxides contribute to the generation of fine volatile particulate matter (PM). Resultant changes to ambient PM concentrations and radiative properties of the atmosphere may be important sources of aviation-related environmental impacts. This paper addresses engine design and operational impacts on aerosol precursor emissions of SO_x *and* NO_y *species. Volatile PM formed from these species in the environment surrounding an aircraft is dependent on intraengine oxidation processes occurring both within and downstream of the combustor. This study examines the complex response of trace chemistry to the temporal and spatial evolution of temperature and pressure along this entire intraengine path after combustion through the aft combustor, turbine, and exhaust nozzle. Low-order and higher-fidelity tools are applied to model the interaction of chemical and fluid mechanical processes, identify important parameters, and assess uncertainties. The analysis suggests that intraengine processing is inefficient. For in-service engine types in the large commercial aviation fleet, mean conversion efficiency* (ε) *is estimated to be 2.8–6.5% for sulfate precursors and 0.3–5.7% for nitrate precursors at the engine exit plane. These ranges reflect technological differences within the fleet, a variation in oxidative activity with operating mode, and modeling uncertainty stemming from variance in rate parameters and initial conditions. Assuming that sulfur-derived volatile PM is most likely, these results suggest emission indices of* 0.06–0.13 g/kg *fuel, assuming particles nucleated as* $2H_2SO_4 \cdot H_2O$ *for a fuel sulfur content of* 500 ppm. [DOI: 10.1115/1.2795758]

Keywords: particulate matter, sulfate, nitrate, emissions, inventory, environmental impact

1 Introduction

This paper addresses aircraft engine design and operational impacts on aerosol precursor emissions of oxidized sulfur and nitrogen species (i.e., $SO_x = SO_2 + SO_3 + H_2SO_4$ and $NO_y = NO + NO_2 + HONO + HNO_3$, respectively), focusing on sulfur trioxide (SO_3) due to its significant role in particulate matter (PM) production relative to other emissions. Influential parameters and phenomena that control intraengine oxidation are identified. Modeling techniques are outlined and uncertainties in predicted trace emission evolution are assessed. Estimates for volatile particulate mass generated in the near-field exhaust of current in-service commercial aircraft are presented as a function of operating mode. Our goal is to synthesize these investigations and present a first identification of technology trends in volatile PM emissions. This will enable an inventory development for aviation PM and the assessment of associated air quality and climate effects. These results update and expand upon an earlier work [1] with improved, validated modeling techniques and revised chemistry.

1.1 Aircraft Particle Emissions. Fine particles are emitted and generated in aircraft engine exhaust in size ranges constituting the nucleation, Aitken, and accumulation modes of a typical PM size distribution. Direct and indirect atmospheric effects from aviation-sourced PM constitute environmental risks of an uncertain magnitude [2,3].

Aviation-sourced PM emissions evolve over a large range of spatial and temporal scales. Aircrafts directly emit nonvolatile carbonaceous particles (soot) that are generated over millisecond time scales within the combustor. In situ sampling of older in-service airframe-engine combinations suggest that soot emissions at cruise are characterized by a log normal distribution with a median diameter in the range of 20–60 nm, a geometric standard deviation on the order of 1.5–1.75, and a number emission index (EI) of $0.1 \times 10^{15} - 6 \times 10^{15}$ particles/kg fuel [4–7]. Under current U.S. air quality rules, these particles would be defined as PM2.5 (particles smaller than a nominal 2.5 μm aerodynamic diameter). Assuming a soot density of 1500–1800 kg/m^3, these data imprecisely place nonvolatile EIPM ($EIPM_{nv}$) in the range of 0.002–2 g/kg fuel, with a mean value of ~0.2 g/kg fuel. The scaling of soot concentration with engine performance suggested in Döpelheuer [8] indicates that this EI range should also be representative of operational modes other than cruise.

Sulfate and nitrate production is initiated by gas phase oxidation of SO_2 and NO that begins in the postcombustion intraengine flow. This gives a unique role to trace species chemical processing through the combustor dilution zone, turbine, and exhaust nozzle, which is as important to the formation of volatile PM emissions as is the influence of combustor fluid mechanics on total NO_x ($NO + NO_2$), CO, and HC emissions. Our analysis addresses precursor emissions of trace nitrogen and sulfur oxides, which are formed within the engine over time scales on the order of 10 ms. These emissions contribute to the generation of volatile fine PM formed in the engine plume at longer time scales (10 ms to < 1 s) compared to soot formation in the combustor.

As the exhaust plume mixes with the atmosphere, additional (secondary) PM may continue to form over days to weeks in air masses moving regionally hundreds of kilometers from the source. This is the result of further oxidation of emitted SO_x,

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received November 21, 2005; final manuscript received February 14, 2007; published online February 29, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2005: Land, Sea and Air (GT2005), June 6–9, Reno, NV, 2005, Paper No. GT-2005-59112.

NO_y, and HC species, typically via condensation or absorption onto existing particles. In this respect, soot continues to play a part in microphysical and chemical processes after emission. Over longer time scales, deposition occurs and additional aqueous and gaseous conversion processes can become important (e.g., reaction with ambient ammonia). This secondary processing has a different consequence than near-field plume and intraengine oxidation, which are, to first order, of greater importance to local perturbations in ambient PM concentration.

Volatile particles are smaller than nonvolatiles but similarly described by a log normal distribution, with median diameters in the range of 1–15 nm, standard deviation on the order of 1.5, and number EIs 10–100 times greater than for soot. Using the modeling assessment described in this paper, we estimate the production of the aerosol precursors SO_3 and nitrous acid (HONO) within the engine for in-service technology. Results are also presented for NO_2, which is the source of another aerosol precursor, nitric acid (HNO_3), that can be produced in the exhaust plume. Results are presented as conversion efficiencies (ε), molar ratios defined as in

$$\varepsilon_{SO_3} = SO_3/SO_x$$

$$\varepsilon_{HONO} = HONO/NO_y$$

$$\varepsilon_{NO_2} = NO_2/NO_y \quad (1)$$

Based on SO_x precursors available at the engine exit, our analysis suggests that volatile EIPM ($EIPM_v$) can be in the range of 0.06–0.13 g/kg fuel, assuming particles nucleated as $2H_2SO_4 \cdot H_2O$ for a fuel sulfur content FSC of 500 ppm. These emission indices are comparable to those for soot.

If all NO_y precursors available at the engine exit were also converted to PM, the $EIPM_v$ may increase by a factor of 10–20. This is an unlikely event. One analysis of plume processing estimates emitted HONO uptake peaks at perhaps 10% for stratospheric temperatures, but the absorption is short lived in the presence of SO_3 [9] and would undoubtedly peak at values lower than 10% in the troposphere where temperatures are higher. Our estimates of $EIPM_v$ do not include PM contributions sourced to HC species. Aircrafts emit a range of gaseous hydrocarbons, which may also nucleate in the near-field region and add to the volatile PM mass [10,11]. Aircrafts also emit engine oil, which contributes directly to volatile particulate mass by a currently unknown amount.

1.2 Outline. The discussion proceeds as follows. Section 2 describes the role of SO_x and NO_y emissions in aerosol formation, their consequent impacts on the atmosphere, and what is currently known about aircraft aerosol precursor emissions. We then model the interaction of chemical and fluid mechanical processes using both low-order and higher-fidelity tools. Our approach is to build an understanding of the key parameters and uncertainties in trace emissions prediction in order to establish a firmer connection between technology and environmental impact.

Section 3 discusses the methods used to determine physical drivers and to develop estimates for trace species emissions, and details the parameter space investigated. Sections 4 and 5 focus on cycle-level effects, employing a low-order, comparative timescale methodology and 1D models over averaged flow properties through the combustor, turbine, and exhaust nozzle to investigate fundamental trends in SO_x and NO_y oxidations as a function of engine technology and flight condition.

Section 6 addresses two areas of uncertainty in these results. Section 6.1 looks at the error in the specification of initial conditions and reaction parameters. Further investigation with higher-fidelity, unsteady 2D computational tools presented in Sec. 6.2 focuses on the role of temperature and velocity nonuniformities associated with turbine fluid mechanical phenomena in determining trace species chemistry. This provides a quantification of uncertainties related to a methodological choice in modeling intraengine processing. Section 7 synthesizes the analyses of Secs. 3–6 and presents estimates for $EIPM_v$.

2 Aircraft Volatile Particulate Matter Emissions

Various mechanisms lead to the production of particulate precursors within the engine. Oxidized nitrogen species originate primarily from the high temperature oxidation of atmospheric nitrogen in the combustor. Total sulfur emissions are predictable functions of fuel composition and emerge from the primary zone as SO_2 in lean conditions. Sulfur emissions are thus controlled by fuel consumption to a greater extent than NO_x. The formation of precursors to volatile PM, including SO_3 and HONO, initiates within the combustor and continues downstream through the turbine and exhaust nozzle [1,12–14]. The response of trace chemistry to the temporal and spatial evolution of temperature and pressure through the turbine is complex and presents both computational and experimental challenges. Total emissions are related to the technological characteristics of the aircraft (weight, aerodynamic efficiency, and engine overall efficiency), its operational use, and details of the combustor, turbine, and nozzle design.

2.1 Precursors, Microphysics, and Effects. Experimental and modeling studies have highlighted the role of trace emissions of SO_3 in the formation of high number densities of fine aerosol particles observed in the exhaust streams of several aircrafts [5–7,11,15–17]. Particle concentrations are correlated with the level of oxidized fuel sulfur in the exhaust [6,18]. At exhaust temperatures and lower temperatures, SO_3 converts to H_2SO_4 in the presence of exhaust water vapor [19]. In the plume, new volatile sulfate particles can be formed by binary homogeneous nucleation of H_2SO_4 with emitted water vapor [20], accentuated by concomitant chemiion emissions [21]. Nucleated sulfate particles grow via coagulation and uptake of water vapor [22,23]. Because nitrous and nitric acids have high saturation vapor pressures relative to sulfuric acid, they have a lesser tendency to nucleate new particles and thus contribute to PM primarily via uptake on existing particles. Hydrocarbons can also contribute via uptake and may additionally nucleate as an independent PM source [24], but thermodynamic conditions are not favorable for this process at the HC levels typically emitted [25]. Both HNO_3 and H_2SO_4 are emitted in concentrations orders of magnitude smaller than SO_3 and HONO; thus, the evaluation of volatile aviation PM should focus on the latter species. The results of our analysis point out that the production of H_2SO_4 from emitted SO_3 is greater in magnitude than the new sulfuric acid production in the plume. Although the production of HNO_3 occurs at rates about an order of magnitude lower than that for H_2SO_4 at exhaust and ambient conditions [26], HNO_3 can play a role in plume PM processing [27]. We thus also examine intraengine production of NO_2, the chemical precursor to nitric acid. Note that condensed matter can further increase through heterogeneous nucleation of sulfuric acid on soot and metal surfaces, activated by adsorption of oxidized sulfur, in the presence of water vapor [16]. This analysis does not address such interactions between nonvolatile PM and aerosol precursors.

Particles formed can perturb the mass and size distribution of the background atmospheric aerosol if scavenged, alter the chemical makeup of the upper atmosphere through heterogeneous chemistry, and perhaps freeze, persisting as contrail particles. Microphysical processes involving emitted and generated particles can lead to changes in contrail optical properties, but have minimal apparent impact on formation tendency [11]. However, the incidence of persistent contrails is expected to increase as aircraft engines become more efficient [28]. Measurements suggest the presence of radiative impacts from aviation contrails and related cirrus [29], but the effect of aerosols on clouds is not well understood [30]. Particles may play a part in cirrus formation near flight tracks, via contrail evolution into cirrus, for example, and as a

result of the greater spatial coverage that results, aviation-induced cirrus impacts are expected to be larger than contrail impacts [31–33]. Comparatively, direct sulfate radiative forcing has been estimated as a relatively small negative impact [3]. Although there is high uncertainty, the collective instantaneous radiative impact of contrails and clouds is estimated to be greater than that of aviation CO_2 and NO_x emissions combined. Accounting for all aviation perturbations, models suggest that the present day aviation particulate-related contribution to total anthropogenic forcing from aircrafts is greater than 2% [3,34]. This contribution is expected to increase with air transportation growth.

The role of aviation PM in determining local air quality is poorly understood, partially due to an incomplete characterization of particulate emissions from in-service aircrafts. Fine PM is associated with the incidence of mortality and illness, including both acute and chronic respiratory and cardiac health end points, and can lead to visibility impairment. Upon deposition, PM contributes to ecosystem damage resulting from acidification as well as soiling and other material damages [2].

2.2 Measurement and Modeling. Experimental data characterizing sulfur species in engine exhaust consist primarily of concentrations inferred from in-flight plume measurements [6,7,11] and a few direct measurements at and downstream of the engine exit plane [35–37], mostly at higher power conditions for older in-service commercial and military aircrafts. Inferred levels indicate an apparently broad range of SO_2 to H_2SO_4 oxidation. However, a detailed analysis of the instrument responses and age of the sampled air has refined the estimates of oxidative conversion to 0.5–5% of the fuel sulfur [38]. In-flight trends and ground measurements suggest that oxidation efficiency is dependent on engine technology and operating point [11,37]. For comparison, measurements of land-based, marine, and laboratory test gas turbines burning diesel fuel demonstrate oxidation to SO_3 of between 2% and 22% [39].

Measurements of NO and NO_2 are routine in engine development and certification, but there are few measurements of their oxidation products, HONO and HNO_3. These species have been predicted to be present in aircraft exhaust [1,13], and measurements of HNO_3 and HONO [40,41] have indicated that conversion of NO to HONO and NO_2 to HNO_3 at the engine exit plane amounts to a few percent or less at higher power conditions. Intraengine conversion of NO to NO_2 has been estimated from measurements at much higher levels, up to ~25% [42].

Because nucleation rates are high for H_2SO_4, given typical fuel sulfur levels, modeling investigations of the microphysical processes that lead to the formation of volatile aerosols emphasize the development of oxidized sulfur through the aircraft plume and wake. These investigations find that known gaseous pathways yield only 1–2% oxidation within the near-field plume (<1 s after emission) for a range of aircraft engine configurations, physical approximations, and chemical assumptions [22,43]. Modeling investigations have also shown that compared to the plume, sulfur oxidation can be more vigorous within the aircraft engine as a result of gaseous chemical processes through the combustor dilution zone, turbine, and exhaust nozzle. Upper bound chemical kinetic analyses indicate that SO_3 formation via atomic oxygen is less than 6% of SO_x within the combustors used in aircraft [12] and industrial applications [44]. Previous studies have also suggested that SO_3 formation via OH and O may result in an upper limit 10% oxidation through the turbine and nozzle [1,12–14,45].

Accounting for key factors influencing postcombustion intraengine processing, revised trace chemistry, and the thermodynamic behavior of in-service aircraft types under typical operational conditions, this study finds that intraengine conversion efficiencies of SO_2 to SO_3 for commercial aircrafts are within this upper bound limit. Our modeling analyses presented here are also consistent with reported experimental results.

3 Methods

Our analysis uses a simulation methodology for aerosol precursor behavior that has been previously introduced. Reacting flow simulations are exercised with fidelity appropriate to the magnitude of flow nonuniformity, particularly in temperature, along the gas path. This study finds that low-order models of the active chemistry, where kinetics are driven by averaged flow parameters specified as a function of time, capture the impact of fluid properties on trace NO_y chemistry within the engine. Nonuniformities can, however, have a significant influence on trace SO_x chemistry. In order to estimate the magnitude of the effect on sulfate precursor emissions, a more complex fluid dynamical representation is constructed. These higher-fidelity computations need only be applied to regions of the engine flow path that, because of thermodynamic conditions and favorable residence time, exhibit trace species activity which could be influenced by nonuniformity of the intraengine flow. We use comparative time-scale analyses such as those discussed in Sec. 4 to identify regions along the gas path that fit this description, and observe that it is usually not necessary to increase fidelity for engine sections downstream of the high-pressure turbine (HPT). For this analysis, a 2D unsteady computation for the first stage of the HPT is developed to evaluate multidimensional effects. Lower fidelity, 1D simulations suffice as the alternative for the remaining sections of the gas path.

Additional methodological details are described in the following sections. In Whitefield et al. [41], estimates for emission levels using this approach correctly captured trends and matched experimental data within measurement uncertainty for the conversion of NO to HONO at both the combustor and nozzle exits of a test engine. This provided support that OH-driven oxidation of NO_x can be modeled with acceptable accuracy. Similar OH-driven kinetics also control the oxidation of SO_2 to SO_3, for which direct measurements are not available. However, favorable results for estimating HONO and the availability of appropriate SO_x kinetic parameters suggest that the modeling of SO_x oxidation should also provide good estimates of the exit sulfur emission speciation. Section 4 further explores the foundations and applications for the methodology described, using a comparison of physical and chemical time scales to establish the relative importance of various controls on oxidation and relating such influences to aircraft engine design parameters.

3.1 Parameter Specification. Initial species and flow conditions for the time-scale, flow-averaged 1D, and higher resolution 2D unsteady computation conditions were specified to highlight technological and operational trends in particulate emissions from the current commercial aviation fleet. A number of engine thermodynamic cycles were developed to represent the range of in-service engine types on large commercial aircrafts. Temperature and pressure profiles as a function of time were specified for each of five operating modes: the four certification settings—idle/taxi, approach, climb, and take-off—and a fifth setting to represent the altitude cruise condition. These were used directly as the flow parameter specification for the time-scale and 1D simulations discussed in Secs. 4 and 5. Table 1 summarizes these cycles by thrust class.

The 2D analyses reviewed in Sec. 6.2 for a representative single turbine stage were specified at flow conditions also spanning the parameter range of large in-service commercial engines. Simulation parameters were varied to highlight fluid mechanical impacts on oxidation through the postcombustor flow path for a range of fuel sulfur levels, power conditions, and pattern factors. Table 2 summarizes the conditions examined.

3.2 Chemistry and Initial Conditions. A prerequisite for the simulation capability is the availability of finite rate kinetic mechanisms that apply to conditions intermediate between combustion and atmospheric chemistry. The full mechanism used for the computations described here was truncated from the larger set

Table 1 Cycle parameter specification for time-scale and flow-averaged 1D gas path simulations. $EINO_x$ as NO_2 and EIS=0.5 g/kg fuel. EICO and OFAR different for each cycle and condition. Common initial species conditions are (a) $NO/NO_x=1$, (b) $SO/SO_x=0$, and (c) no HC or H_2. Note that combustion efficiency is consistent with EICO

Class (kN)	<100		100–200		200–400		>400	
Thrust (kN)	60	65	100	170	205	230	250	420
Cycle No.	1	2	3	4	5	6	7	8
Take-off								
P_3 (atm)	16	15	23	25	20	27	27	37
T_4 (K)	1350	1200	1520	1510	1400	1560	1540	1940
$EINO_x$ (g/kg fuel)	35	19	55	17	36	30	27	20
Climb								
P_3 (atm)	14	12	19	22	17	24	23	31
T_4 (K)	1300	1140	1440	1450	1340	1510	1480	1840
$EINO_x$ (g/kg-fuel)	29	16	42	14	27	24	22	16
Cruise								
P_3 (atm)	6.0	5.7	8.0	9.0	7.1	7.3	9.7	13
T_4 (K)	1170	1030	1340	1340	1270	1210	1370	1710
$EINO_x$ (g/kg fuel)	5.1	7.1	7.1	6.3	8.2	7.6	9.4	5.9
Approach								
P_3 (atm)	6.0	5.7	7.1	9.2	7.6	10	9.8	13
T_4 (K)	960	870	1010	1080	1010	1140	1110	1440
$EINO_x$ (g/kg-fuel)	9.3	8.4	10	6.0	7.3	10	11	5.5
Idle								
P_3 (atm)	2.6	2.6	3.3	3.7	3.3	4.4	4.0	5.3
T_4 (K)	750	710	840	870	830	940	900	1220
$EINO_x$ (g/kg) fuel	3.7	8.4	6.0	4.0	2.9	4.2	4.7	2.4

developed in Mueller et al. [46] based on previous studies and new laboratory experiments (see also Ref. [47]). A reaction list can be found in Table 7.

For computations addressing chemistry aft of the combustor, initial trace species concentrations are determined at the turbine inlet using the pressure, temperature, and overall fuel-air ratio (OFAR) relevant to the conditions specified in Tables 1 and 2; combustor SO_x or NO_y oxidation is initialized using a similar methodology, but at a stoichiometric FAR. To best identify technology trends given the complexity of the combustor flow-chemistry environment, we chose to parametrize trace chemical activity in the combustor dilution zone in Sec. 5.1 using a separate set of computations to provide a full account of aerosol precursor

Table 2 Parameter specification for 2D unsteady turbine stage simulations. Common initial species conditions are (a) $NO/NO_x=1$, (b) $SO/SO_x=0$, and (c) no HC or H_2. $EINO_x=15$ g/kg fuel and EIS=0.5 g/kg fuel for all cases. Note that combustion efficiency is consistent with EICO

Parameter specifications	FSC (ppm)	Pattern factor	Range T_4 (K)
Higher power cases			
$T_4=1900$ K; $P_3=22$ atm; $T_{surf}=850$ K (NGV), 775 K (rotor)			
Case 1	500	30%	1550–2250
Case 2	500	10%	1800–2000
Case 3	10	30%	1550–2250
Case 4	10	10%	1800–2000
Lower power cases			
$T_4=1300$ K; $P_3=7$ atm; $T_{surf}=640$ K (NGV), 590 K (rotor)			
Case 5	500	30%	1100–1500
Case 6	10	30%	1100–1500

levels at the engine exit. It is possible to simply add these results to the oxidation estimated to occur from the turbine inlet to nozzle exit [12,13]; sensitivity parameters discussed in Sec. 6.1 suggest that the low levels of SO_3 and HONO formed in the combustor have a minor effect on subsequent oxidation, translating instead into equivalent conversion at the engine exit.

Estimation of species initial conditions at the turbine inlet starts with equilibrium concentrations at the given flow conditions and FAR. Concentrations for species known not to be in equilibrium (e.g., NO_x, SO_x, and CO) are then changed to levels measured during certification tests of the engines on which the representative cycles are based while maintaining elemental balance. This calibration of emission indices is followed by a short constant temperature and pressure kinetic calculation adjustment to allow the species to achieve a quasisteady state. This final step removes any nonphysical, rapid readjustments of concentrations in the mixture that result from the imposed changes to the mixture. These rapid readjustments occur for reactive species that are sensitive to the distribution of NO_x, SO_x, and CO and have relatively short characteristic time scales. Although oxidation within the SO_x and NO_y families may be significant, species family total EIs are basically invariant through the engine. Fuel sulfur levels are set at a constant 0.5 g/kg fuel (500 ppm) for all 1D analyses, but vary as specified in Table 2 for the 2D simulations (10–500 ppm). All other species initial conditions, including OH, O, HO_2, and H, are set to their equilibrium values at the specified temperature and pressure. The initial ratios of NO/NO_x and SO_2/SO_x are set to 1.0.

4 Determinants of Intraengine Oxidation

The evolution of aerosol precursors is controlled by the thermodynamic potential for formation (referenced to equilibrium,

Table 3 Reduced reaction sets describing SO_3, HONO, and NO_2 kinetics. Notes: *=Irreversibility (→) in a reaction was considered only when differences between forward and backward rates differed by more than three orders of magnitude. For reactions 4 and 6, Troe form used here is different from the original references. †=Efficiencies for R1: N_2=1.3, H_2O=10; for R4: H_2O=10

Reaction* troe=Troe form parameters low=Low-pressure parameters			A (cm mol s K) E (cal/mol) A	b	E_a
SO_3					
1	$SO_2+O(+M)\rightarrow SO_3(+M)$	(†)	9.20E+10	0.00	2.384E+03
	troe: n/a	|low→	4.00E+28	−4.00	5.250E+03
2	$SO_3+O\rightarrow SO_2+O_2$		0.44E+12	0.00	0.610E+04
3	$SO_2+OH=SO_3+H$		0.49E+02	2.69	0.238E+05
4	$SO_2+OH(+M)=HOSO_2(+M)$	(†)	1.21E+12	0.00	0.000E+00
	troe: 0.35/1E−30/1E30	|low→	1.87E+31	−4.61	2.050E+03
5	$HOSO_2+O_2=SO_3+HO_2$		7.80E+11	0.00	0.656E+03
NO_2					
6	$NO+O(+M)\rightarrow NO_2(+M)$		1.30E+15	−0.75	0.000E+00
	troe: 0.957/1E−90/1551E03	|low→	4.72E+24	−2.87	1.551E+03
7	$NO_2+O\rightarrow NO+O_2$		1.05E+14	−0.52	0.000E+00
8	$NO_2+H\rightarrow NO+OH$		1.32E+14	0.00	3.620E+02
9	$NO+HO_2=NO_2+OH$		2.11E+12	0.00	−4.790E+02
HONO					
10	$HONO+OH\rightarrow NO_2+H2O$		1.70E+12	0.00	−5.200E+02
11	$NO+OH(+M)=HONO(+M)$		1.99E+12	−0.05	−7.210E+02
	troe: 0.62/1E−90/1E+90	|low→	5.08E+23	−2.51	−6.760E+01
12	$HONO+O=OH+NO_2$		1.20E+13	0.00	5.961E+03

and changing with T and P), chemical kinetic rates, and the flow residence time available to complete reactions. Engine cycle, internal turbomachinery fluid mechanics, and engine size determine temperature, pressure, and residence time. For trace species interactions through the combustor dilution zone, turbine, and exhaust nozzle, heat release is negligible except in conditions where the combustor is operated at FARs ratios higher than those employed in current commercial aircraft engines. Since oxidation activity generally decreases with lower temperatures and pressures, power extraction in the turbine acts as a limit on the extent of intraengine oxidation. Such constraints are not fully countered by the increase in thermodynamic potential, as equilibrium levels of oxidation products increase with lower temperatures. There are also kinetic limits on oxidation, such as the availability of O, OH, and HO_2, which evolve primarily through self-reactions [13,46,47]. These species, along with NO_x, SO_x, CO, and HC, are found at the turbine inlet in concentrations established by combustion and dilution processes in the combustor. With temperature and pressure, they determine reactivity. Within the engine, heterogeneous oxidation pathways are of minimal consequence to aerosol precursor production [48]. Thus, gaseous chemistry is our focus.

4.1 Kinetic Influences. In Table 3, reduced sets of reactions specific to SO_3, HONO, and NO_2 formation are listed, which highlight the primary kinetic influences on postcombustion oxidation. Reactions were selected from the reaction mechanism listed in Table 7 based on ranked contribution to production rates for the species of interest over the postcombustion flow path for the range of engine cycle parameters listed in Table 1. While the reduced reaction sets listed in Table 3 preserve the relevant formation dynamics, their use in a predictive manner can result in deviations of up to 25%. In addition to the nitrous acid production outlined, the production of sulfuric and nitric acids depends on the formation of SO_3 ($SO_3+H_2O\rightarrow H_2SO_4$) and NO_2 ($NO_2+OH\rightarrow HNO_3$) described by these kinetics.

There are two pathways leading to SO_3 formation represented in Table 3, via OH and less importantly via O, mediated to a minor extent by consumption via H. For SO_3, reactions 1 and 2 (R1 and R2) constitute a partial recombination of atomic oxygen to form molecular oxygen, $O+O=O_2$. However, the rate of destruction of SO_3 by R2 is typically much less than the rate of production by R1. R3 is a reductive reaction, but the limited availability of atomic hydrogen limits the strength of this destruction pathway to approximately ten times less than R2. R4 and R5 represent the formation of SO_3 via an OH pathway, which dominates SO_3 dynamics in the regime intermediate between atmospheric conditions and combustor conditions.

Similar to SO_x chemistry, O and OH are central to NO_y chemistry. Although NO_2 and HONO are necessarily coupled in the full mechanism shown in Table 7, they are isolated here to highlight the relevant kinetics. Like SO_3, R6 and R7, which contribute to NO_2 kinetics, represent $O+O=O_2$, but compared to the corresponding dynamic in the sulfur chemistry, the resultant impact on NO_2 can be either positive or negative. However the formation of NO_2 is typically dominated by R9 via HO_2 as opposed to the O pathway. Destruction by H has a larger role than for SO_3, but is still relatively minor. R11 involving OH primarily determines the formation of HONO.

Characteristic time scales for SO_3, HONO, and NO_2 formation can be derived from the sets of reactions listed in Table 3, as shown in Eq. (2). Terms in Eq. (2) depend on the specification of H, O, OH, HO_2, and O_2 concentrations, as well as reservoir species such as N_2 and H_2O to account for efficiencies. The coefficients C_x depend primarily on temperature. For Eq. (2), $HOSO_2$ in R5 is determined by a pseudo steady-state equilibrium using R4 and R5 in a relative abundance of O_2. Equation (2) also assumes that NO_x and SO_x concentrations remain constant.

$$\tau_X[\mathrm{OH}] = \left\{ C_{\mathrm{OH}} + C_{\mathrm{const}}\frac{1}{[\mathrm{OH}]} + C_{\mathrm{O}}\frac{[\mathrm{O}]}{[\mathrm{OH}]} + C_{\mathrm{HO_2}}\frac{[\mathrm{HO_2}]}{[\mathrm{OH}]} + C_{\mathrm{H}}\frac{[\mathrm{H}]}{[\mathrm{OH}]} \right\}^{-1} \tag{2}$$

	SO_3	HONO	NO_2
C_{OH}	$k_4\left(\frac{k_5[\mathrm{O_2}]}{k_{4b}+k_5[\mathrm{O_2}]}\right)$	k_1+k_{3b}	k_{4b}
C_{count}	0	k_{2b}	0
C_{O}	k_1+k_2	k_3	k_1+k_2
$C_{\mathrm{HO_2}}$	$k_{5b}\left(\frac{k_{4b}}{k_{4b}+k_5[\mathrm{O_2}]}\right)$	0	k_4
C_{H}	k_{3b}	0	k_3

At equilibrium for typical turbine conditions, the molar ratios of O/OH and HO_2/OH are of order 0.01, and H/OH is of order 0.00001. To illustrate the relative contributions of OH, HO_2, O, and H to oxidation reactivity in the turbine, relative trends for the terms in Eq. (2) using these approximate equilibrium molar ratios are shown in Fig. 1 for a range of temperatures. In Fig. 1, the larger the relative value, the faster the chemistry and the greater the contribution. At any given temperature, pathways involving HO_x are generally most consequential to trace chemical reactivity; atomic oxygen has a role, but for the equilibrium ratios relevant to these cycles, it is a minor contributor, requiring an increase in concentration by a factor of 100 to be of comparable importance along the gas path.

O_x and HO_x equilibrate within the engine on time scales faster than consumption by trace families. As a result, the relative consumption of reactive oxidation species associated with the SO_x and NO_y chemistries, for a given rate of temperature change, is of primary consequence to conversion efficiencies (ε). The blue shaded areas in Fig. 1 show the typical range of temperatures and characteristic chemical time scales found in the thermodynamic environments of the high-pressure turbine (HPT) and low-pressure turbine (LPT). Specific examples of time-scale trends as

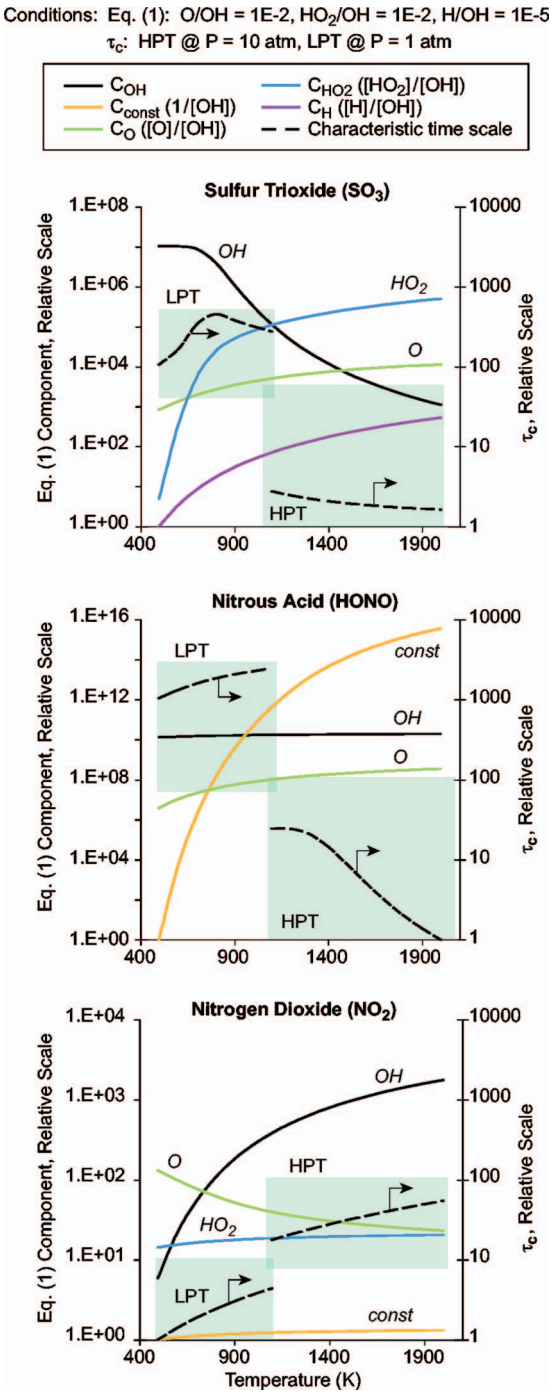


Fig. 1 Chemical kinetic drivers of trace SO_3, HONO, and NO_2 chemistry

a function of temperature for constant pressures characteristic of the LPT and HPT (1 atm and 10 atm, respectively) are shown with dashed lines. In general, despite large equilibrium values for SO_3, HONO, and NO_2 at the flow conditions of the engine exit, oxidation time scales are too long to allow significant conversion along the engine gas path. For common in-service engines, the characteristic time scale for SO_3 is approximately 100 ms, increasing to values greater than 1 s at the nozzle exit. Comparatively, the flowthrough times for the postcombustor flow path are much shorter, typically in the range of 1–10 ms.

4.2 Conversion Potential. To exemplify the trade between thermodynamic potential and kinetic time scale as a function of

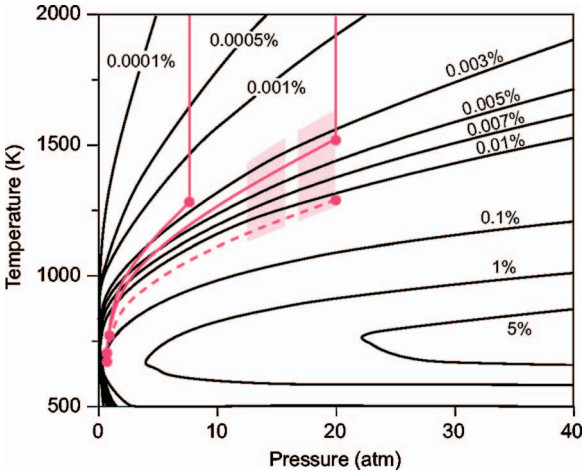


Fig. 2 Conversion potential ($\Delta\varepsilon_{SO_3}$) as a function of temperature and pressure. To calculate ε', the fuel hydrogen-carbon ratio (H/C) is set at 2, and the FSC is specified at 500 ppm, both representative of Jet A fuel.

temperature and pressure, a comparison of fluid and chemical time scales for SO_3 production is constructed in Eq. (3), defining a conversion potential $\Delta\varepsilon = \Delta SO_3 / SO_x$. The definition assumes that $[SO_3] \gg [H_2SO_4]$, which is true for all but a few parameter ranges, as discussed in Sec. 5. Fluid mechanical and chemical kinetic influences are represented through a Damköhler (Da) number, and thermodynamic potential is defined as the equilibrium conversion efficiency (ε').

$$\Delta\varepsilon = \begin{cases} \mathrm{Da}[\varepsilon' - \varepsilon] & \text{for } \mathrm{Da}_{SO_3} \leq 1 \\ \varepsilon' & \text{for } \mathrm{Da}_{SO_3} > 1 \end{cases}$$

$$\mathrm{Da} = \tau_{\mathrm{flow}} / \tau_{SO_3}$$

$$\varepsilon' = \varepsilon \text{ at equilibrium} \tag{3}$$

Equilibrium SO_x depends on temperature and pressure, shifting toward SO_3 at low temperatures, and is only weakly dependent on pressure, increasing from <10% at the combustor exit to ~90% at the nozzle exit. Increasing the fuel sulfur level generally decreases the resulting ratios at equilibrium, but as will be evident in later sections, fuel sulfur level has a minimal consequence for oxidation levels. As defined, Da>1 indicates that 95% of ε' is reached within the reference flow residence time, and Da<1 indicates that only a fractional conversion is attained.

Figure 2 maps $\Delta\varepsilon$ for a small increment in a flow time of ~0.1 ms (a flowthrough time for a single blade row) and an initial converted fraction of 0%. Iso-$\Delta\varepsilon$ contours indicate oxidation potential >1% over this time scale only for a range of temperatures between 600 K and 900 K and pressures over ~5 atm. Conditions of this description are experienced by limited portions of the core flow, specifically in the combustor dilution zone flow at low power, and areas interacting with cooling flows, at locations typically within the combustor or high-pressure turbine. The bulk flow evolution through the turbine and beyond is generally outside this region for all engine operating modes.

The strength and location of regions of high oxidation potential would necessarily evolve as S(VI) is formed (ε increases as the gas path is traversed). Moving from the upper right of Fig. 2 to its lower left traces a generalized path of oxidation potential that follows the change of temperature and pressure from the turbine inlet to the nozzle exit. Two sample thermodynamic paths are shown, one representative of the take-off condition and one for cruise. Also shown on Fig. 2 is the path of constant pressure processing through the combustor dilution zone. We will consider

this further in Sec. 5.1. The shaded areas represent the influence of spanwise temperature nonuniformities in the turbine, as discussed in Sec. 6.2.

The form of the contours in Fig. 2 is representative of results for NO_2 and HONO as well, with the exception that peak conversion potential occurs at lower temperatures for NO_2, and even lower for HONO. In addition, the contours tend to be more spread for NO_2 and less so for HONO relative to SO_3, a consequence of the relative range of chemical time scales over the gas path (see Fig. 1). As a result, HONO oxidation tends to exhibit a greater dependence on power setting than with SO_x. HONO production is typically limited to a smaller axial section of the flow path at lower temperatures than SO_3 or NO_2. On this account, HONO production should also exhibit a greater sensitivity to temperature than SO_3. However, cross-stream temperature nonuniformities, which play a relatively important role in SO_3 production, are less significant in the lower pressure flow path where HONO is formed.

5 Technological and Operational Trends

In this section, estimates of postcombustion SO_3, HONO, and NO_2 production through the combustor dilution zone, turbine, and exhaust nozzle are obtained using 1D chemical kinetic calculations through averaged gas path temperature and pressure profiles derived from the eight-cycles described in Table 1. Section 5.1 develops a parametrization of trace chemical activity in the combustor dilution zone. The objective is to place bounds on the likely conversion efficiency at the turbine inlet for a range of combustor designs operating in the current fleet. Section 5.2 then addresses trace chemistry through the expansion gas path. Together, these analyses enable the identification of technological and operational trends. Uncertainties in these results are discussed in Sec. 6, which evaluates the robustness of the result to changes in rate parameters and initial conditions, and the impact of introducing more realistic fluid mechanics.

5.1 Combustor Oxidation. Two parameters were independently varied to examine the range of potential oxidation levels at the combustor exit—residence time (τ_{res}) and rate of dilution air addition. In one case, referred to here as the *quick quench* case, all dilution air is added over a period of 1 ms to reach the turbine inlet temperature, and then trace chemistry is allowed to continue for additional residence times of 1–15 ms. In the other case, dilution air is added at a constant rate for a total residence time τ_{res} of 2–16 ms (the *constant quench* case). This parameter space represents upper and lower bounds on the technologies associated with the engine types simulated, but the quick quench case is more likely representative of an in-service combustor design. Note that as cycle T_4 increases, the amount of combustor dilution air added necessarily decreases and the response to dilution schedule becomes less important.

For the analysis, the upstream boundary is defined by the location of peak combustion temperature at a stoichiometric FAR. The physical location of this point may vary among combustor designs, but it uniformly defines when an oxidative environment would be achieved, and this is a necessary precondition to precursor production. Other than the specification of the initial stoichiometric FAR, the methodology described previously in Sec. 4 to establish species conditions is employed using the same values for $EINO_x$ and EIS listed in Table 1. For each case, combustor conversion efficiency was explored using a 1D model (averaged, fully mixed conditions) varying temperature profiles at constant pressure. The temperature profiles are specified by the different dilution schedules, enforcing different rates of temperature change from a flame temperature estimated using T_3 calculated for the cycle, to T_4 as listed in Table 1.

The results of this simplified representation of the combustor suggest that attention must be paid to dilution zone processing, which is likely the primary location for sulfate aerosol precursor

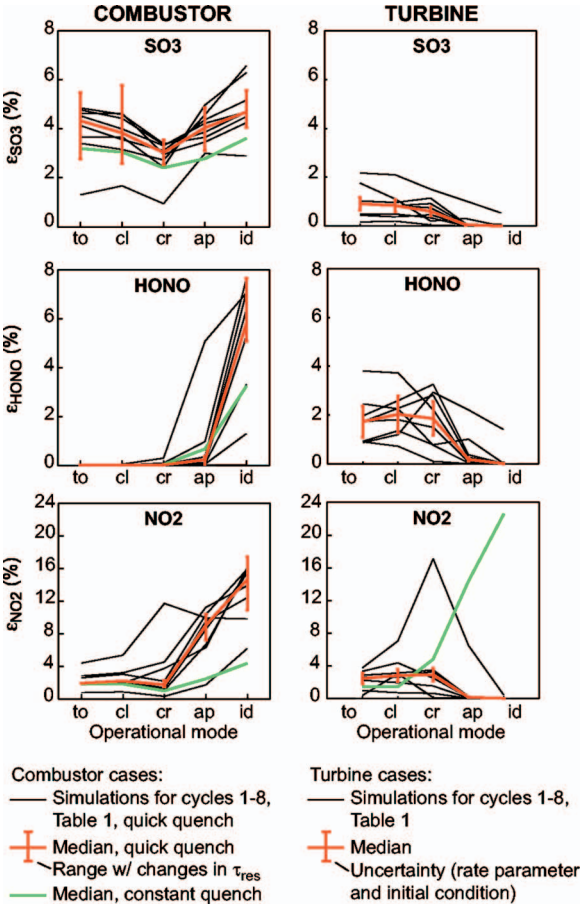


Fig. 3 Intraengine conversion efficiencies (ε_{SO_3}, ε_{HONO}, and ε_{NO_2}) estimated using 1D flow-averaged simulations

production over all modes, and to low power conditions for nitrate precursors. Figure 3 shows results for conversion efficiency for the quick quench case where τ_{res} is 6 ms. The median combustor ε_{SO_3} is highlighted in red. For comparison, median ε_{SO_3} is shown in green for the constant quench case. These medians span a range of 2.5–4.5% over the five operating modes.

The dependence on cycle temperatures and pressures is not monotonic. A clearer picture of trends in SO_3 production in the combustor can be described with reference to the vertical profile section of thermodynamic paths drawn in Fig. 2 representing the dilution zone flow. These constant pressure profiles fall within higher contours of conversion potential as T_4 is reduced; however, lower operating temperatures are correlated with lower operating pressures, which follows the contours of $\Delta\varepsilon$. The influence of pressure and, by implication, of altitude is evident in Fig. 3 comparing ε_{SO_3} for the cruise condition to the four LTO modes. The range shown by the red whiskers in Fig. 3 shows the variation of median ε_{SO_3} with τ_{res}. Higher ε_{SO_3} is correlated with higher τ_{res}, but the range is narrower than the variance sourced to the rate parameter and initial condition uncertainty discussed in Sec. 6.1. The higher the τ_{res} in the combustor, the more likely it is that equilibrium ε_{SO_3} will be found at the turbine inlet.

Combustor ε_{SO_3} is higher than the estimates for median ε_{SO_3} through the rest of the gas path. The combustor plays a more limited role in the production of HONO and NO_2 where dilution zone processing is dominant only for lower power conditions. For NO_y, the impact of residence time is not monotonic, suggesting the more complex kinetic response to the thermodynamic environment seen for oxidation through the turbine and exhaust nozzle. A significant combustor role in determining ε_{HONO} in comparison to

downstream engine sections is observed only at the idle condition (6% median at idle versus <0.1% at other conditions). Median ε_{NO_2} at take-off, climb, and cruise is comparable to subsequent oxidation, but similar to HONO, there is a much stronger combustor oxidation contribution at lower power conditions (median 8% at approach and 14% at idle).

5.2 Turbine and Exhaust Nozzle Oxidation. The impact of cycle design on ε through the turbine and exhaust nozzle is reflected primarily through the influence of peak temperature and, to a lesser extent, of peak pressure. As cycles become hotter and reach higher overall pressure ratios, postcombustor trace chemical activity will tend to increase. The use of cooling air has also increased as a necessary enabler for hotter cycles, with more blade rows in the HPT and LPT subject to some protection to achieve life requirements. This further increases the opportunity for oxidation as described later in Sec. 6.2. However, importantly, turbine design choices set the time rate of the change of temperatures and pressure along the gas path. This rate is a function of blade loading, which has increased since the introduction of the jet engine, tending to reduce the opportunity for large conversion; the number of stages is reduced and residence time at any one temperature and pressure decreases. For similar size engines shown in Table 1, intraengine residence time has generally decreased historically (but increases with engine size/thrust class).

Figure 3 summarizes technology and operational trends in aerosol precursor production along the turbine and exhaust nozzle gas path. The figure shows the range of ε over all cycles as a function of operating mode. The red line indicates the median trend with propagated uncertainties, discussed further in Sec. 6.1, denoted by the whiskers. While thermodynamic conditions would support high conversion efficiencies, the results show the strong kinetic limitations of the turbine environment on aerosol precursor production. In Fig. 3, conversion efficiencies generally increase with power level or, more generally, with higher peak temperature and pressure, whether this is due to cycle design or operating mode. This is most universally applicable to ε_{SO_3}, but there is significant variability in the results.

Part of the variability can be explained by differences in Da, as given by Eq. (3). Oxidation ratios are plotted against Da in Fig. 4. Approximate geometries and gas path velocities define the flow residence time, and the chemical time scale is based on Eq. (2) specified for conditions at the turbine inlet. For SO_2 oxidation to SO_3, increasing Da nonlinearly increases conversion efficiency. It should not be expected that a single scale can explain all variabilities, but the goodness of fit displayed ($R^2 > 0.98$) suggests that for SO_3, the overall cycle design and operating point are factors most important in distinguishing technological differences in ε_{SO_3}. The variability among engine types in intraengine temperature and pressure as a function of time as well as the influence of other trace chemical activities (with the exception of HO_x) are less influential.

In contrast, the correlation with Da is weaker for NO_y species, suggesting that such details have a more significant impact in determining technological trends for HONO and NO_2. Outliers are highlighted in Fig. 4. For NO_2, in particular, the kinetic rates can get faster or slower with progress through the gas path, and are sensitive to the temperature and pressure variation as a function of time, which leads to the more varied behavior for the different modes. In reference to Fig. 2, this means that the gradient of formation potential contours can either be in the direction of changes in temperatures and pressures along the gas path or against them, depending on the mode and cycle. This is a significant source of variance in the trends shown in Fig. 3.

Sensitivity analyses presented in the next section further suggest that $EINO_x$ influences ε_{NO_2} and ε_{HONO} to a significant degree, and this is an additional source of variability in the results. In contrast, ε_{SO_3}, ε_{NO_2}, and ε_{HONO} are all minimally impacted by EIS over the parameter space examined, which indicates independence from FSC. However, absolute emissions of SO_3 are not independent and vary proportionally with FSC. For SO_3 and HONO, the greatest conversion in the bulk flow is localized to the latter part of the HPT and to early part of the LPT, earlier for SO_3 and later for HONO. Formation rates for NO_2 are almost constant with location through the postcombustion flow path. Because the formation of NO_y species occurs in the LPT, scaling based on turbine inlet parameters as in Fig. 4 is less useful.

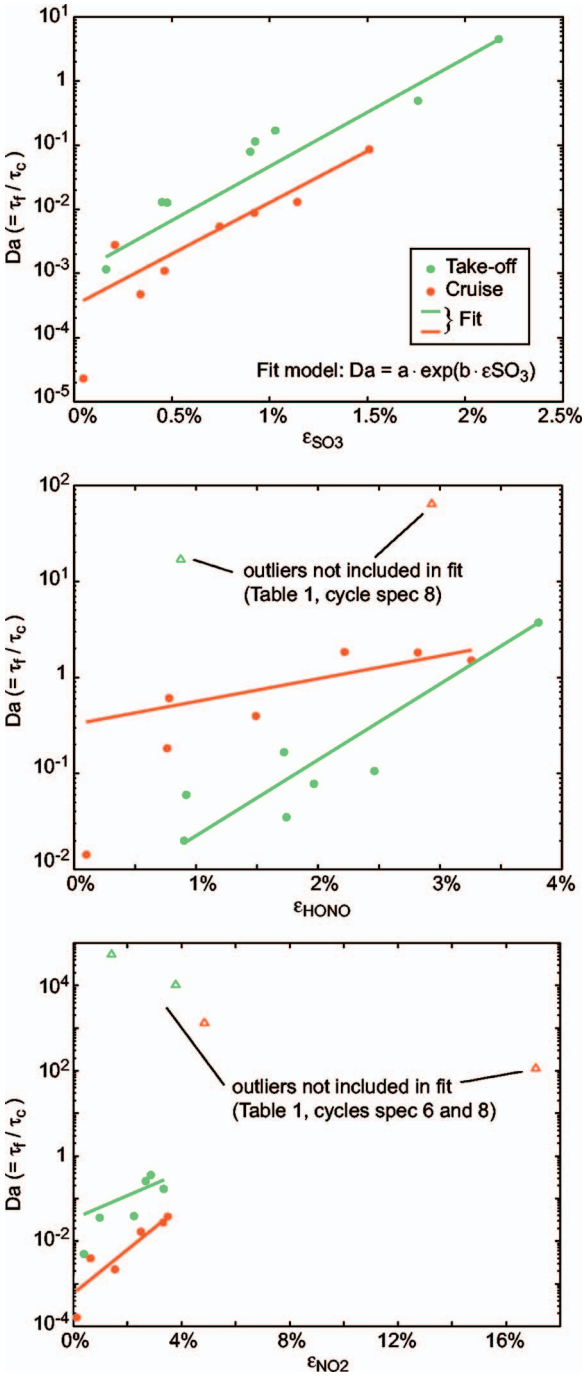


Fig. 4 Explanation of variance in conversion efficiency trends as a function of technology and flight condition using Da

Overall, ε_{SO_3} sourced to trace chemistry through the turbine and exhaust nozzle is typically 0.5–1% at higher power operating conditions, falling to values <0.1% at low power modes. Only at the lowest temperature, the lowest pressure over all cycles examined (typical of idle power for an older in-service engine) is the con-

version of SO_3 to H_2SO_4, a significant fraction of SO_x oxidation. Trends for conversion to NO_y are more dependent on cycle design, as results exhibit both increasing and decreasing ε_{NO_2} and ε_{HONO} as a function of operating mode. This is a result of the sensitivity of the NO_y kinetics to changes in temperature and pressure discussed previously in this section and is most evident for large engines. A good example in Fig. 3 is where the largest thrust class simulated (highlighted in green) exhibits a monotonic increase in ε_{NO_2} with decreasing operating mode. Apart from these exceptions, oxidation to HONO is typically 2% of NO_y at high power, falling to values similar to SO_3 at idle and approach, ~0.1%. Similarly, conversion to NO_2 in the turbine is typically 2.5% for higher power modes and <0.1% at low power. Note that NO_y trends with power condition are generally opposite in the combustor. Calculated HNO_3 is produced at less than 0.3% NO_y for all cases.

6 Uncertainty and Bias

In this section, we estimate the magnitude of three important sources of uncertainty in modeling trace species emissions from gas turbine engines: uncertainty in chemical kinetic rate parameters, uncertainty in initial conditions, and modeling uncertainty associated with the fidelity with which turbomachinery fluid mechanics are represented. Uncertainties in rate parameters and initial conditions, fundamental to all attempts to estimate intraengine conversion efficiency, are addressed first in Sec. 6.1 using the calculations summarized in Sec. 5 as a basis. Sensitivities to initial conditions outlined in Sec. 6.1 will clarify the importance of combustor processing to subsequent oxidation through the turbine and exhaust nozzle. Section 6.2 then addresses the influence of turbomachinery fluid mechanics on conversion efficiency. From these results, we can infer the potential role of flow temperature and velocity nonuniformities in the combustor dilution zone. Figure 2 indicates the impact of these nonuniformities by the hatched area. In Fig. 2, rate parameter and initial condition uncertainties would change the locations of iso-$\Delta\varepsilon$ contours.

6.1 Rate Parameters and Initial Conditions. Our analysis of sensitivity to rate parameter and initial condition uncertainties focuses on the 1D treatment of the turbine flow presented in Sec. 5. Simulations were performed in which rate parameters and inlet species levels were individually perturbed for the 1D analyses summarized in Fig. 3 to determine changes in exit plane emission levels. First, as shown in Eq. (4), exit plane relative sensitivities S_β of a given species to a given model parameter, β, were computed by running the 1D model with perturbation $\delta\beta$. Then, relative sensitivities were multiplied by the estimated uncertainty $\Delta\beta$ in the given parameter and summed in quadrature to yield the overall relative uncertainty ($\Delta X/X$) in the predicted species emission level, as shown in

$$S_\beta = \frac{1}{X_\beta}\left(\frac{\Delta X_\beta}{\delta\beta}\right)$$

$$\frac{\Delta X}{X} = \sqrt{\sum_\beta (S_\beta \Delta\beta)^2} \tag{4}$$

Uncertainties in rate parameters were obtained from literature reviews. Uncertainties in species initial conditions were specified as variances in EI (g/kg fuel) by a relative standard deviation. These values represent the uncertainty in the original data used to specify emissions for the cycles simulated and are derived from variance in certification measurements and expected fuel composition. Specifications are summarized in Table 4.

The sensitivity to initial conditions is summarized in Table 5. In the table, S is the average sensitivity over all engine cycles and operating conditions, presented as the ratio of a percent change in an independent species at the turbine inlet (columns) referenced to the percent change in a dependent species at the engine exit (rows) on a molar basis. Results for S are highlighted in the blue columns. $\Delta S(T)$ and $\Delta S(\tau)$ are, respectively, the change in S from the lowest to the highest temperature cycle measured by T_4 (which correlates with the oldest to the newest technology, and low power to high power) and from the shortest residence time to the longest. In Table 6, a negative trend is highlighted in red and a positive trend highlighted in green. Entries with insignificant values, here distinguished by an absolute value smaller than 0.045, are left blank.

Table 4 Specifications for rate parameter and initial condition uncertainty

Rate parameters	
Reaction	$\Delta \log_{10}(k)$
$H+O_2=O+OH$	0.3
$HO_2+OH=H_2O+O_2$	0.3
$H_2O+O=2OH$	0.2
$SO_2+O(+M)=SO_3(+M)$	0.4
$SO_3+H_2O=H_2SO_4$	0.5
$SO_3+O=SO_2+O_2$	0.4
$SO_2+OH=SO_3+H$	0.4
$SO_2+OH(+M)=HOSO_2(+M)$	0.4
$HOSO_2+O_2=SO_3+HO_2$	0.5
$NO+O(+M)=NO_2(+M)$	0.5
$NO_2+O=NO+O_2$	0.1
$NO_2+H=NO+OH$	0.1
$NO+HO_2=NO_2+OH$	0.4
$HONO+OH=NO_2+H_2O$	0.3
$NO+OH(+M)=HONO(+M)$	0.3
$HONO+O=OH+NO_2$	1.0
Initial conditions	
Species	1σ EI (unless noted otherwise)
CO_2	0.4%
H_2O	2.5%
O_2	1.0%
NO_x	15%
CO	21%
SO_x	35%
OH	OH/3 to 3 OH

Sensitivities are generally smaller than 1, and most of the sensitivity parameters that have significant values are correlated with turbine inlet temperature or residence time or both. Self-sensitivities are in the order of 1 and are dominant compared with other values, signifying the low level of interaction between SO_x and NO_y chemistries. Sensitivities of all species to O, H, and HO_2 are ~0. Given the discussion in Sec. 4, the sensitivity to OH is expectedly larger, but still much less than 1, reinforcing the idea that SO_x and NO_y chemistries are a minor draw on the radical pool in comparison to reactions within the HO_x family or among reactive oxidants generally. The correlation of sensitivity parameters with T_4 and residence time for all species and cycles can be summarized as shown in Table 6. This is an alternate way of visualizing results for $\Delta S(T)$ and $\Delta S(\tau)$.

The analysis for rate parameters employed the full reaction mechanism listed in Table 7, but we focus on major species inputs and dominant reactions identified using molar production rates and parametric sensitivity gradients. Figure 5 plots an estimate for the relative standard deviation for key species, which sums contributions from variance in initial conditions and uncertainty in rate parameters. Similar estimates are also provided for ε_{SO_3}, ε_{NO_2}, and ε_{HONO}. The filled bars show the median values. The whiskers mark the minimum and maximum values to depict the potential range in relative uncertainty with changes in engine technology or operating condition.

The relative uncertainty in NO_y and SO_x species is seen to

Table 5 Sensitivity coefficients. Values of $|S|$ smaller than 0.045 are considered insensitive to the perturbations and are not shown. For correlation coefficients (R^2) smaller than 0.2, $\Delta S(T)$ and $\Delta S(\tau)$ are randomly scattered and not shown

	perturb O_2			perturb CO		
Species	S	$\Delta S(T)$	$\Delta S(\tau)$	S	$\Delta S(T)$	$\Delta S(\tau)$
NO_2				0.214		
HONO	−0.057	0.137	0.060	−0.065		0.264
HNO_3	−0.039	0.273	0.157	0.095		
SO_3	0.125	0.205	0.127			
H_2SO_4	−0.005	0.273	0.189			

	perturb CO_2			perturb H_2O		
Species	S	S	S	S	$\Delta S(T)$	$\Delta S(\tau)$
NO_2						
HONO				0.060	0.205	
HNO_3				0.088	0.205	0.157
SO_3				−0.094		−0.106
H_2SO_4	−0.045	−0.055	−0.050	0.922		−0.133

	perturb NO_y			perturb SO_x		
Species	S	$\Delta S(T)$	$\Delta S(\tau)$	S	$\Delta S(T)$	$\Delta S(\tau)$
NO_2	0.805	−0.273	−0.294			
HONO	0.942					
HNO_3	0.710	−0.752				
SO_3	−0.036			0.999		
H_2SO_4	−0.028			0.998		0.013

	perturb OH		
Species	S	$\Delta S(T)$	$\Delta S(\tau)$
NO_2	0.043	−0.137	
HONO	0.231		−0.504
HNO_3	0.373	−0.956	−0.830
SO_3	0.063		−0.080
H_2SO_4	0.061		

increase with oxidation state and is largest for the inorganic acids HNO_3 and H_2SO_4. As a fraction of the total uncertainty, rate parameter variance is a small contribution relative to the impact of initial conditions for the major combustion products as well as NO and SO_2. Rate parameter variance is roughly equal in importance to initial condition uncertainty for NO_2 and HONO. For SO_3, H_2SO_4, and HNO_3, rate parameter uncertainties are dominant. To understand how these uncertainties impact ε for the turbine and exhaust nozzle, Fig. 3 shows error bars around the median result in Fig. 5 that represent uncertainty in ε. The magnitude of the variance due to rate parameter and initial condition uncertainty is less than the differences across engine types.

6.2 Temperature and Velocity Nonuniformities. Fluid mechanical phenomena have an important influence on oxidation in the SO_x family in the early stages of the turbine. In general, flow nonuniformities in the turbine flow can result in an increased opportunity for oxidation. To estimate the magnitude of impact, and thus the uncertainty in using a 1D approximation, we demonstrate these influences using higher-fidelity computational tools to estimate the impact of fluid-chemical interactions through the first stage of a HPT typical of current in-use engines. Turbomachinery reacting flows are computationally demanding. The scenarios presented were developed to capture the primary phenomena that affect trace chemistry for the high power operating conditions where ε_{SO_3}, ε_{NO_2}, and ε_{HONO} through the turbine and exhaust nozzle are highest. By inference, it should also be expected that temperature and velocity nonuniformities in the combustor should be similarly important for SO_x chemistry, but models addressing the relevant dilution jet and liner flows have not yet been developed.

Table 6 Marginal trends

Trend	$\uparrow T$		$\uparrow \tau$	
Species	SO_x	NO_y	SO_x	NO_y
O_2	$S\uparrow$	$S\uparrow$	$S\leftrightarrow$	$S\uparrow$
CO	$S\leftrightarrow$	$S\leftrightarrow$	$S\uparrow$	$S\downarrow$
CO_2	$S\downarrow$	$S\downarrow$	$S\leftrightarrow$	$S\uparrow$
H_2O	$S\leftrightarrow$	$S\downarrow$	$S\uparrow$	$S\uparrow$
OH	$S\leftrightarrow$	$S\downarrow$	$S\leftrightarrow$	$S\downarrow$

The reacting flow solution for the multidimensional case was produced using the CNEWT code [1]. CNEWT combines well-established computational fluid dynamics (CFD) and chemical kinetics solution mechanisms within a structure capable of calculating reacting internal flows. CNEWT is built on the NEWT turbomachinery CFD code [49], which employs a vertex-centered, finite-volume solution method incorporating a Runge-Kutta time discretization scheme to solve the full Reynolds-averaged Navier–Stokes, conservation of mass, and conservation of energy equations for three-dimensional geometries. Chemical mechanisms are represented using a system of ordinary differential equations (ODEs) handled through CHEMKIN routines and solved using VODE [50,51]. To complete the fluid-chemical integrations under the

Table 7 Reaction mechanisms. Notes: * Efficiencies for reactions 1,3,5,6,9, and 10: H_2=2.5, H_2O=12, CO=1.9, CO_2=3.8, AR =0.75. † Efficiencies for reactions 23, 24, 25, 28, and 30: AR=0.75. ‡ Efficiencies for reactions 40 and 44: H_2=2.5, H_2O=12, CO =1.9, CO_2=3.8. ** Efficiences for reactions 54: N_2=1.3, H_2O=12. †† Efficiencies for reactions 55: H_2O=10. ‡‡ Efficiencies for reactions 57: N_2=1.5, H_2O=10.

	Reaction troe=Troe form parameters low=low-pressure form		A (cm mol sec K) E (cal/mol) A	b	E_a
1*	$H_2+M=H+H+M$		4.57E+19	−1.4	104,440
2	$O+H_2=H+OH$		5.08E+04	2.7	6290
3*	$O+O+M=O_2+M$		6.16E+15	−0.5	0
4	$H+O_2=O+OH$		1.91E+14	0.0	16,440
5*	$H+O_2(+M)=HO_2(+M)$		1.48E+12	0.6	0
	troe 0.5/1E−30/1E30	⟨low→	3.482E+16	−0.411	−1115
6*	$H+O+M=OH+M$		4.71E+18	−1.0	0
7	$OH+H_2=H_2O+H$		2.16E+08	1.5	3430
8	$H_2O+O=OH+OH$		2.97E+06	2.0	13,400
9*	$H_2O_2(+M)=OH+OH(+M)$		2.95E+14	0.0	48,380
	troe 0.5/1E−90/1E90	⟨low→	0.12E+18	0.0	45,500
10*	$OH+H+M=H_2O+M$		2.21E+22	−2.0	0
11	$HO_2+O=O_2+OH$		3.25E+13	0.0	0
12	$HO_2+H=H_2O_2$		1.66E+13	0.0	823
13	$HO_2+H=OH+OH$		7.08E+13	0.0	295
14	$HO_2+OH=H_2O+O_2$		2.89E+13	0.0	−497
15	$HO_2+HO_2=H_2O_2+O_2$	(duplicate)	4.20E+14	0.0	11,982
16	$HO_2+HO_2=H_2O_2+O_2$	(duplicate)	1.30E+11	0.0	−1629
17	$H_2O_2+O=OH+HO_2$		9.55E+06	2.0	3970
18	$H_2O_2+H=H_2O+OH$		2.41E+13	0.0	3970
19	$H_2O_2+H=HO_2+H_2$		4.82E+13	0.0	7950
20	$H_2O_2+OH=H_2O+HO_2$	(duplicate)	1.00E+12	0.0	0
21	$H_2O_2+OH=H_2O+HO_2$	(duplicate)	5.80E+14	0.0	9557
40	$HCO+O_2=CO+HO_2$		7.58E+12	0.0	410
41	$HCO+O=CO+OH$		3.02E+13	0.0	0
42	$HCO+H=CO+H_2$		7.23E+13	0.0	0
43	$HCO+OH=CO+H_2O$		3.02E+13	0.0	0
44	$CO+O(+M)=CO_2(+M)$		1.80E+10	0.0	2384
	troe: n/a	⟨low→	1.35E+24	−2.788	4191
45	$CO+O_2=CO_2+O$		2.53E+12	0.0	47,700
46*	$CO+OH=CO_2+H$		1.40E+05	1.9	−1347
47	$CO+HO_2=CO_2+OH$		3.01E+13	0.0	23,000
48	$NO+HCO=HNO+CO$		7.33E+12	0.0	0
49	$NO_2+HCO=CO+HONO$		1.26E+23	−3.3	2354
50	$NO_2+HCO=H+CO_2+NO$		8.43E+15	−0.8	1927
51*	$NO_2+CO=CO_2+NO$		9.03E+13	0.0	33,780
52	$SO_3+O=SO_2+SO_2$		4.40E+11	0.0	6100
53	$SO_3+SO=SO_2+SO_2$		1.00E+12	0.0	4000
54**	$SO_2+O(+M)=SO_3(+M)$		9.20E+10	0.0	2384
	treo: n/a	⟨low→	4.00E+28	−4.0	5250
55††	$SO_2+OH(+M)=HOSO_2(+M)$		1.21E+12	0.0	0
	troe 0.35/1E−30/1E+30	⟨low→	1.87E+31	−4.61	2050
22	$HNO+H=NO+H_2$		4.40E+11	0.7	650
23†	$NO+O(M)+NO_2(+M)$		1.30E+15	−0.8	0
	troe 0.957/1E−90/8.332E+03	⟨low→	4.72E+24	−2.87	1551
24	$NO+O(+M)=NO_2(+M)$		1.52E+15	−0.4	0
	troe 0.82/1E−90/1E+90	⟨low→	3.1E+19	−1.32	735.2
25†	$NO+OH(+M)=HONO(+M)$		1.99E+12	−0.1	−721
	troe 0.62/1E−90/1E+90	⟨low→	5.08E+24	−2.51	−67.6
26	$NO_2+H_2=HONO+H$		7.33E+11	0.0	28,810
27	$NO_2+O=O_2+NO$		1.05E+14	0.0	0.0
28†	$NO_2+O(+M)=NO_3(+M)$		1.33E+13	0.0	000
	troe 0.826/1E−90/3.191E+0.3	⟨low→	1.49E+29	−4.08	2467
29	$NO_2+H=NO+OH$		1.32E+14	0.0	362
30†	$NO_2+OH(+M)=HNO_3(+M)$		2.41E+13	0.0	0
	treo 0.837/1E−90/1657E+03	⟨low→	6.42E+32	−5.49	2350
32	$HO_2+NO=NO_2+OH$		2.11E+12	0.0	−479
32	$NO_2+NO_2=NO_3+NO$		9.64E+09	0.7	20,920
33	$NO_2+NO_2=2NO+O_2$		1.63E+12	0.0	26,120
34	$HNO+O=OH+NO$		1.81E+13	0.0	0
35	$HNO+OH=H_2O+NO$		1.30E+07	1.9	956
36	$HNO+NO_2=HONO+NO$		6.02E+11	0.0	1987
37	$HONO+O=OH+NO_2$		1.20E+13	0.0	5961
38	$HONO+OH=H_2O+NO_2$		1.70E+12	0.0	520
39	$HCO+M=H+CO+M$		1.86E+17	−1.0	17,000
56	$SO_2+OH=SO_3+H$		4.90E+01	2.7	23,800
57‡	$SO+O(+M)=SO_2(+M)$		3.20E+13	0.0	0
	troe 0.55/1E−30/1E+30	⟨low→	0.29E+25	−2.9	0
58	$SO+OH=SO_2+H$		5.20E+13	0.0	0
59	$SO+OH+M=HOSO+M$		8.00E+21	−2.2	830
60	$SO+O_2=SO_2+O$		6.20E+03	2.4	3050
61	$HOSO+M=SO_2+H+M$		5.90E+34	−5.7	50,900
62	$HOSO+OH=SO_2+H_2O$		1.00E+12	0.0	0
63	$HOSO+O_2=SO_2+HO_2$		1.00E+12	0.0	1000
64	$HSO_2+M=SO_2+H+M$		1.20E+28	−4.1	18,900
65	$HSO_2+M=HOSO+M$		1.10E+21	−2.0	29,900
66	$HOSO_2=HOSO+O$		5.40E+18	−2.3	10,630
67	$HOSO_2+M=SO_3+H+M$		3.20E+16	−0.8	53,700
68	$HOSO_2+H=SO_2+H_2O$		1.00E+12	0.0	0
69	$HOSO_2+O=SO_3+OH$		5.00E+12	0.0	0
70	$HOSO_2+OH=SO_3+H_2O$		1.00E+12	0.0	0
71	$HOSO_2+O2=SO_3+HO_2$		7.80E+11	0.0	656
72	$SO_2+NO_2=SO_3+NO$		6.30E+12	0.0	27,000
73	$SO_3+H_2O=H_2SO_4$		7.23E+08	0.0	0

passive chemistry approximation in CNEWT, the flow and chemistry algorithms are decoupled, with the latter parallelized to better handle large systems of equations.

Computational grids were constructed so as to emphasize only circumferential and axial flows, thus establishing what is essentially a 2D computation for the nozzle guide vane (NGV) and first HPT rotor. The radial dimension was ignored because the key fluid mechanics that impact chemistry are captured in midspan. This includes the impact of a cool boundary layer on oxidation, the impact of a nonuniform temperature profile at the combustor exit, the bulk evolution of species through the freestream, and the unsteady mixing of NGV blade wakes through the rotor. Three pitches for the NGV and five for the rotor were calculated simultaneously in separate calculations to match the periodicity of a typical engine turbine stage. These calculations were connected using an inlet mask to propagate the NGV flow into the rotor

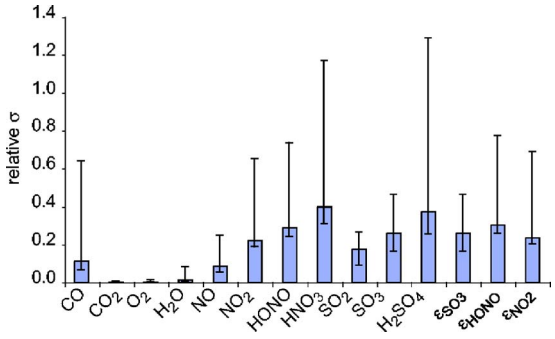


Fig. 5 Relative uncertainties in species concentrations and conversion efficiencies

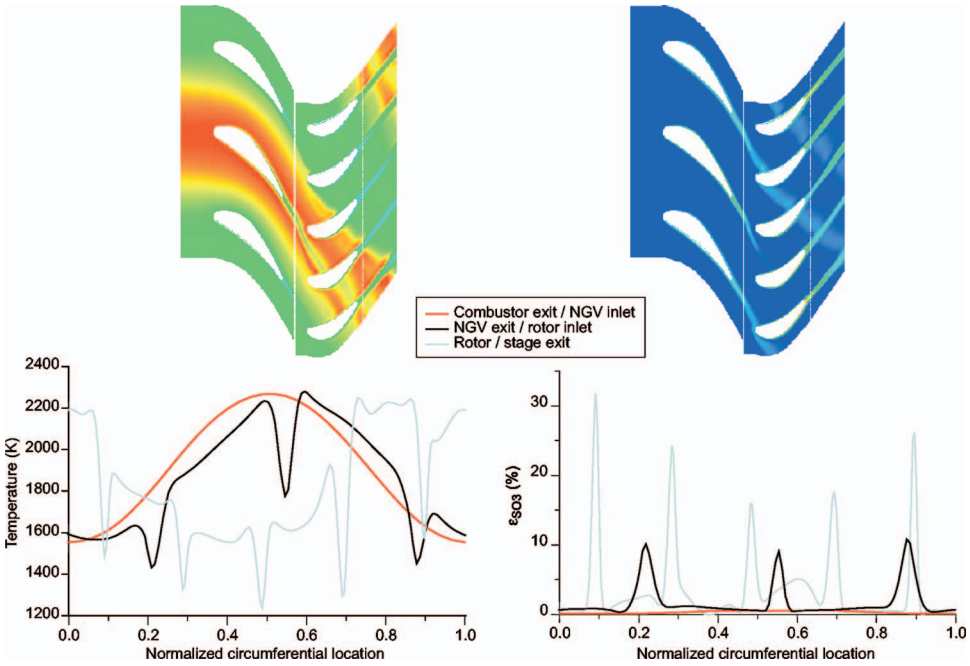


Fig. 6 Distribution of SO_3 and temperature across a turbine stage

domain. Blade surfaces are set to a lower temperature boundary condition relevant to cooling air, but no film cooling mass is added in the solutions.

The variation in combustor exit properties represents the impact of typical exit nonuniformity, specified as a sinusoidal circumferential temperature profile with magnitude dependent on the given pattern factor. For a high efficiency, nearly adiabatic combustor, changes in temperature at the combustor exit are related to the variation in the initial FAR of the exhaust arriving at the exit. To calculate the FAR profile, an adiabatic flame temperature calculation was performed to find the FAR associated with the temperature at each point on the combustor exit profile. The local FAR, temperature, and pressure thus constitute the necessary components for the initial condition specification conducted for each point specified for the temperature profile.

Figure 6 shows the distribution of SO_3 across the NGV and a snapshot across the rotor for the high power, high pattern factor and high sulfur case outlined in Table 2. Radial and circumferential temperature nonuniformities at the combustor exit plane act to segregate the chemically active regions of the turbine flow. This influence is evident with a tenfold increase in oxidation to SO_3 as the highest temperature combustor flow moves over the cooled center blade, but there are smaller changes for the top and bottom blades (which are also cooled) where temperatures are lower. This difference occurs because of the increased residence time near the blade surface at temperatures favorable to SO_3 formation. For the equivalent low power case, the difference is more pronounced (although the enhancement is quantitatively smaller) and the increase happens only for sections exposed to the hotter flow.

Between blade rows, wake mixing occurs at constant pressure, and dilution of the oxidized sulfur with higher temperature air moves kinetics in the wake in the opposite direction from the near-blade effect. However, the mixing time scale of wakes, which is generally on the order of the flow time through a full component or the full engine, is restricted by the row spacing. Thus, the boundary layer enhancement persists as a permanent augmentation. This can be seen in Fig. 6. Similar features to the NGV exit are found at the stage exit, and it can be seen that additional flow has been exposed to the cooled blade as upstream wakes are chopped and migrate through the rotor passage. Unsteady chopping of wakes into downstream blade rows enhances mixing by increasing the interfacial surface area, but this is a minor effect. The net impact is an increase in oxidation through the stage, resulting from temperature gradients near the blade.

Comparatively, the bulk flow exhibits very little change through the stage, which is the result a 1D analysis would suggest. This difference is illustrated notionally in Fig. 2. The solid line represents the 1D analysis, the evolution of mean temperature and pressure through the engine. The shaded area depicts the range of temperatures across the blade row. Only a small fraction of the flow moves through these regions, but it can be seen that the cooler regions enable higher conversion efficiencies. The dashed line is included to suggest the persistence of these temperature changes through the engine. Since trace chemical activity is slow compared to turbulent time scales, the dominant role of turbulence should be in determining mixing rates, and we expect the k-ε model employed to approximately capture this effect. Increased turbulence will increase mixing, affecting temperature gradients in the flow, and decreasing the overall conversion efficiency.

Similar features are evident for both HONO and NO_2, but since HONO production potential is significant only later in the gas path at the conditions simulated, temperature and residence time nonuniformities in this set of calculations have less impact on HONO as compared to NO_2 and SO_3. This can be seen in Fig. 7, which summarizes ensemble-averaged results from all multidimensional simulations outlined in Table 2. Sensitivities and uncertainties in rate parameters and initial conditions discussed in the previous section apply equally to these results. In particular, we would expect to find a similar influence from changing the $EINO_x$, as presented previously.

The lack of HONO response is generally applicable only to the extent that nonuniformities are less significant at locations downstream in the gas path where HONO is more active. For SO_3, the impact of power setting is the most important factor, suggesting that nonuniformity effects predominate at higher power modes and for more recent engines. The influence of pattern factor depends on the response of oxidation rates to temperature variations, with a 1% increase in ε_{SO_3} and a 0.2% decrease in ε_{NO_2} resulting from a factor of 3 decrease in the pattern factor. Except for the low power condition, variation in the sulfur level has an insignificant impact based on our estimation of the absolute accuracy lim-

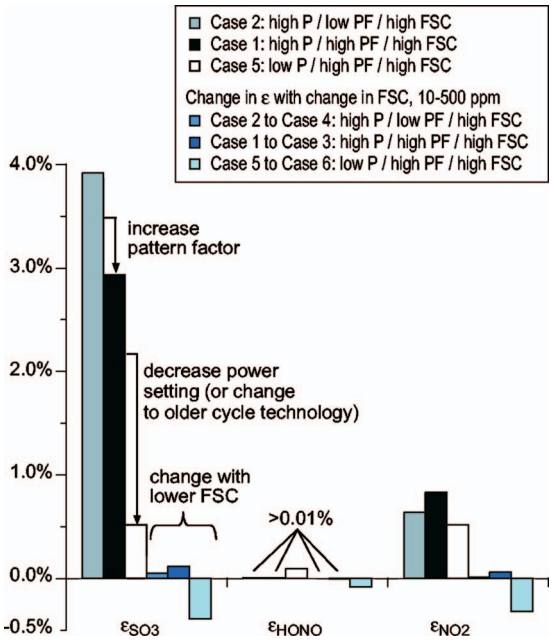


Fig. 7 Effect of nonuniformity as a function of power setting (P), pattern factor (PF), and FSC

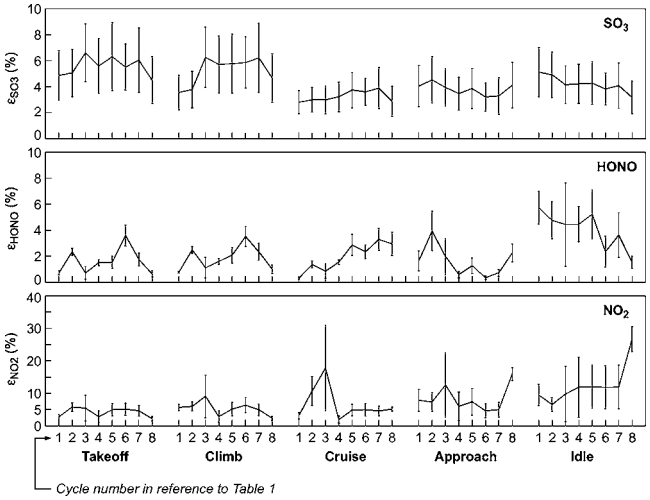


Fig. 8 Best estimate conversion efficiencies (ε_{SO_3}, ε_{HONO}, and ε_{NO_2}) as a function of technology and operating mode

its of the numerical calculation. At low power, altering the fuel sulfur level from 500 ppm to 10 ppm results in a 0.4% reduction in ε_{SO_3}, which is comparable to the total oxidation through the stage at high FSC. However, in context, note that most conversion for the lower power conditions occurs within the combustor.

Secondary flows, mass injection through the blade, and end wall boundary layers were not modeled, but we can provide some evaluation of the direction of influence. End wall cooling can enhance stage oxidation since the cooled surface area at the blade row hub and casing is comparable to the cooled blade surface area. For mass injection, sensitivity results suggest that added O_2 would tend to increase the production of SO_3 but decrease that of HONO and NO_2. However, cooling schemes are generally of high film effectiveness across the blade surface, and the resulting unmixedness may keep the freestream flow at a higher temperature than depicted in the solutions shown in Fig. 6, reducing reactivity and thus oxidation [52]. Comparatively, this change in the temperature to which the SO_x chemistry would be exposed is stronger than the impact of compositional changes (O_2), suggesting that the augmentations in conversion shown would be lower with mass injection represented.

Based on a comparison between 1D model approximations and the 2D solutions, it is estimated that flow nonuniformities account for an additional 0-2% SO_x conversion to SO_3 for each cooled turbine stage. The higher end of this range is relevant to high power conditions or, equivalently, higher temperature cycles. This can be a significant fraction of the total conversion efficiency realized at the engine exit and thus represents a bias error in using 1D analyses to predict exit emissions of sulfate aerosol precursors. In-service engines typically carry one to two cooled stages.

While the simulations show spatial impacts on NO_y chemistry through the turbine stage, a summary conclusion regarding HONO and NO_2 is not clear. There is generally no oxidation to HONO at high power conditions through the turbine stage, and a small negative influence from temperature and velocity nonuniformities at low power of 0% to −0.5% for the stage. This is despite the similarly enhanced near-blade oxidative activity as in the case of SO_3. Downstream temperature forcing in regions of the flow path where HONO is more active would be lower in the absence of cooling. In reference to the technology trends outlined above, this suggests that a 1D analysis captures the bulk of the HONO activity through the turbine and exhaust nozzle and further emphasizes that for low power conditions or older engines, likely conversion to HONO after the combustor would be minimal. The magnitude of impact from temperature and velocity nonuniformities on NO_2 is comparable to the stage conversion, but the sign of impact for NO_2 was ambiguous, being positive for high power and negative for low power. Given the trends exhibited in the 1D analysis, it should be recognized that the design of the turbine is a significant influence on the sign of the result, and we can at most estimate here that nonuniformities represent added variability in the 1D simulations of ±1%.

7 Summary

Aerosol precursors form through the combustor dilution zone, turbine, and exhaust nozzle of gas turbine engines. The intraengine environment is more important to the production of condensable volatile PM in the area near an aircraft than emissions processing in the engine plume. However, due to an inefficient combination of thermodynamic and kinetic factors, there is, overall, little opportunity for the production of SO_3, the most likely of the precursors to result in volatile PM.

Comparing combustor ε_{SO_3} to conversion magnitudes through the postcombustor gas path suggests that SO_3 production in older technology engines would tend to be located in the combustor, whereas for more recent cycle designs, the turbine and exhaust nozzle have a more prominent role. Since the combustor is likely the dominant source of precursors for most power conditions, further research should focus on a more detailed investigation of the combustor.

Best estimates for aerosol precursor production from in-service commercial engines are summarized in Fig. 8. The figure shows the conversion efficiencies for each of the eight cycles examined at each of the five operating modes simulated (Table 1). The values shown result from a Monte Carlo simulation sampling from distributions based on the results of this analysis. The distribution of combustor oxidation is specified uniform, bounded by the minimum and maximum estimated ε over the cooling schedules and residence times summarized in Fig. 3. A random value chosen from this distribution is added to the ε estimated for the turbine and exhaust nozzle, as shown in Fig. 3. Finally, a uniformly random value is selected in the interval 0–2% for ε_{SO_3} and added for cases similar to the high power conditions simulated to account for the impact of temperature and velocity nonuniformities. Similarly, a random value from the interval −0.5% to 0% for ε_{HONO} is

added for low power cases, and a random value between −1% and 1% is added to all cases for ε_{NO2}. The conversion efficiency estimated in this fashion is then perturbed to account for uncertainties related to uncertainty in rate parameters and initial conditions using the relevant value from Fig. 5.

The mean results are consistent with measurements of sulfur and nitrogen precursors. Although HONO and NO_2 oxidation can be on the order of 10%, particularly at low power conditions, we would not expect nitrate contributions to particulate mass until well after the plume mixes with the atmosphere. Thus, the focus rests upon the conversion efficiency ε_{SO_3} to evaluate volatile PM emissions that may impact the surroundings near an aircraft. In contrast to NO_y species, SO_x chemistry is active over the entire operational range of aircraft currently in the fleet. The trends examined here suggest that mean ε_{SO_3} is limited to the range of 2.8%–6.5%. This reflects technological differences within the fleet, the variation in oxidative activity with operating mode, and modeling uncertainty. Note an additional 1–2% conversion to SO_3, and up to 1% for HONO (and NO_y) may be realized in the plume.

Since fuel flow increases with power setting, the SO_3, NO_2, and HONO emission rates (e.g., Kg/s) will be higher at take-off and climb than that suggested by the conversion efficiencies in Fig. 8. Subsequently, for the landing take-off cycle, higher levels of sulfate in the near-field plume can be expected along the departure portion of a flight profile as opposed to landing. Assuming that sulfur-derived volatile PM is most likely, these results suggest emission indices of 0.06–0.13 g/Kg fuel sourced to intraengine conversion, assuming particles nucleated as $2H_2SO_4 \cdot H_2O$ for a FSC of 500 ppm (EIS of 0.5 g/Kg fuel). These values are similar to EIs for a nonvolatile particulate, as interpreted from measurements. Thus, on a mass basis, the impact of aviation volatile PM on local air quality is likely to be roughly comparable to the impact of aviation soot.

Acknowledgment

This work is the product of several years of NASA support and a more recent funding through the University of Missouri-Rolla Center of Excellence for Aerospace Propulsion Particulate Emissions Reduction. Many thanks go to Dr. Chowen Wey, Dr. Nan-Suey Liu, and Dr. Thomas Wey of the NASA Glenn Research Center for their thoughtful interactions with our group. The foundations of this work were established through the tireless efforts of current and former graduate students at MIT. Many thanks go to Tony Chobot, Pierre Dakhel, Sookhyun Han, and Chi Kin (Benny) Yam for their contributions to the understanding of aviation PM emissions. Thank you also to Dave Dassault of Aerodyne and Jun Luo of Delta Search Laboratories for their assistance in assembling the literature review and uncertainty analysis presented in this paper.

References

[1] Lukachko, S. P., Waitz, I. A., Miake-Lye, R. C., Brown, R. C., and Anderson, M. R., 1998, "Production of Sulfate Aerosol Precursors in the Turbine and Exhaust Nozzle of an Aircraft Engine," J. Geophys. Res., [Atmos.], **103**(D13), pp. 16159–16174.

[2] U.S. Environmental Protection Agency, EPA, 2004, "Air Quality Criteria for Particulate Matter, Volumes 1 and 2," National Center for Environmental Assessment-RTP Office, Office of Research and Development, Report No. EPA/600/P-99/002bF.

[3] IPCC 1999, Intergovernmental Panel on Climate Change, 1999, *Aviation and the Global Atmosphere: A Special Report of the Intergovernmental Panel on Climate Change*, J. E. Penner, D. H. Lister, D. J. Griggs, D. J. Dokken, and M. McFarland, eds., Cambridge University Press, Cambridge, UK.

[4] Schumann, U., Schlager, H., Arnold, F., Ovarlez, J., Kelder, H., Hov, Ø., Hayman, G., Isaksen, I. S. A., Staehelin, J., and Whitefield, P. D., 2000, "Pollution From Aircraft Emissions in the North Atlantic Flight Corridor: Overview on the POLINAT Projects," J. Geophys. Res., [Atmos.], **105**(D3), pp. 3605–3631.

[5] Anderson, B. E., Cofer, W. R., Bagwell, D. R., Barrick, J. W., Hudgins, C. H., and Brunke, K. E., 1998, "Airborne Observations of Aircraft Aerosol Emissions. I: Total Nonvolatile Particle Emission Indices," Geophys. Res. Lett., **25**(10), pp. 1689–1692.

[6] Miake-Lye, R. C., Anderson, B. E., Cofer, W. R., Wallio, H. A., Nowicki, G. D., Ballenthin, J. O., Hunton, D. E., Knighton, W. B., Miller, T. M., Seeley, J. V., and Viggiano, A. A., 1998, "SO_x Oxidation and Volatile Aerosol in Aircraft Exhaust Plumes Depend on Fuel Sulfur Content," Geophys. Res. Lett., **25**(10), pp. 1677–1680.

[7] Anderson, B. E., Cofer, W. R., Barrick, J. D., Bagwell, D. R., and Hudgins, C. H., 1998, "Airborne Observations of Aircraft Aerosol Emissions. II: Factors Controlling Volatile Particle Production," Geophys. Res. Lett., **25**(10), pp. 1693–1696.

[8] Döpelheuer, A., 2001, "Quantities, Characteristics and Reduction Potentials of Aircraft Engine Emissions," *World Aviation Congress and Exposition*, Society of Automotive Engineers, Seattle, WA, Paper No. SAE 2001-01-3008.

[9] Karcher, B., 1997, "Heterogeneous Chemistry in Aircraft Wakes: Constraints for Uptake Coefficients," J. Geophys. Res., [Atmos.], **102**(D15), pp. 19119–19135.

[10] Yu, F. Q., Turco, R. P., and Karcher, B., 1999, "The Possible Role of Organics in the Formation and Evolution of Ultrafine Aircraft Particles," J. Geophys. Res., [Atmos.], **104**(D4), pp. 4079–4087.

[11] Schumann, U., Arnold, F., Busen, R., Curtius, J., Karcher, B., Kiendler, A., Petzold, A., Schlager, H., Schroder, F., and Wohlfrom, K. H., 2002, "Influence of Fuel Sulfur on the Composition of Aircraft Exhaust Plumes: The Experiments SULFUR 1–7," J. Geophys. Res., [Atmos.], **107**(D15), 4247.

[12] Brown, R. C., Anderson, M. R., MiakeLye, R. C., Kolb, C. E., Sorokin, A. A., and Buriko, Y. Y., 1996, "Aircraft Exhaust Sulfur Emissions," Geophys. Res. Lett., **23**(24), pp. 3603–3606.

[13] Tremmel, H. G., and Schumann, U., 1999, "Model Simulations of Fuel Sulfur Conversion Efficiencies in an Aircraft Engine: Dependence on Reaction Rate Constants and Initial Species Mixing Ratios," Aerosp. Sci. Technol., **3**(7), pp. 417–430.

[14] Starik, A. M., Savel'ev, A. M., Titova, N. S., and Schumann, U., 2002, "Modeling of Sulfur Gases and Chemiions in Aircraft Engines," Aerosp. Sci. Technol., **6**(1), pp. 63–81.

[15] Reiner, T., and Arnold, F., 1993, "Laboratory Flow Reactor Measurements of the Reaction $SO_3+H_2O+M-H_2SO_4+M$-Implications for Gaseous H_2SO_4 and Aerosol Formation in the Plumes of Jet Aircraft," Geophys. Res. Lett., **20**(23), pp. 2659–2662.

[16] Karcher, B., Busen, R., Petzold, A., Schroder, F. P., Schumann, U., and Jensen, E. J., 1998, "Physicochemistry of Aircraft-Generated Liquid Aerosols, Soot, and Ice Particles. 2. Comparison With Observations and Sensitivity Studies," J. Geophys. Res., [Atmos.], **103**(D14), pp. 17129–17147.

[17] Schröder, F. P., Karcher, B., Petzold, A., Baumann, R., Busen, R., Hoell, C., and Schumann, U., 1998, "Ultrafine Aerosol Particles in Aircraft Plumes: In Situ Observations," Geophys. Res. Lett., **25**(15), pp. 2789–2792.

[18] Schumann, U., Strom, J., Busen, R., Baumann, R., Gierens, K., Krautstrunk, M., Schroder, F. P., and Stingl, J., 1996, "In Situ Observations of Particles in Jet Aircraft Exhausts and Contrails for Different Sulfur-Containing Fuels," J. Geophys. Res., [Atmos.], **101**(D3), pp. 6853–6869.

[19] Kolb, C. E., Jayne, J. T., Worsnop, D. R., Molina, M. J., Meads, R. F., and Viggiano, A. A., 1994, "Gas-Phase Reaction of Sulfur-Trioxide With Water-Vapor," J. Am. Chem. Soc., **116**(22), pp. 10314–10315.

[20] Karcher, B., Peter, T., and Ottmann, R., 1995, "Contrail Formation—Homogeneous Nucleation of H_2SO_4/H_2O Droplets," Geophys. Res. Lett., **22**(12), pp. 1501–1504.

[21] Yu, F. Q., and Turco, R. P., 1997, "The Role of Ions in the Formation and Evolution of Particles in Aircraft Plumes," Geophys. Res. Lett., **24**(15), pp. 1927–1930.

[22] Brown, R. C., MiakeLye, R. C., Anderson, M. R., Kolb, C. E., and Resch, T. J., 1996, "Aerosol Dynamics in Near-Field Aircraft Plumes," J. Geophys. Res., [Atmos.], **101**(D17), pp. 22939–22953.

[23] Karcher, B., 1998, "Physicochemistry of Aircraft-Generated Liquid Aerosols, Soot, and Ice Particles. 1. Model Description," Geophys. Res. Lett., **103**(D14), pp. 17111–17128.

[24] Karcher, B., 1999, "Aviation-Produced Aerosols and Contrails," Surv. Geophys., **20**(2), pp. 113–167.

[25] Dakhel, P. M., Lukachko, S. P., Waitz, I. A., Brown, R. C., and Miake-Lye, R. C., 2007, "Post-Combustion Evolution of Soot Properties in an Aircraft Engine," J. Propul. Power, **23**(5), pp. 942–948.

[26] Miake-Lye, R. C., Brown, R. C., Anderson, M. R., and Kolb, C. E., 1994, "Calculations of Condensation and Chemistry in an Aircraft Contrial," *Impact of Emissions From Aircraft and Spacecraft Upon the Atmosphere*, U. Schumann and D. Wurzel, eds., *Proceedings of an International Scientific Colloquium*, Cologne, Germany, April 18–20, DLR-Mitteilung 94-06, Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen and Cologne, Germany.

[27] Gleitsmann, G., and Zellner, R., 1999, "The Aerosol Dynamics of $H_2O-H_2SO_4-HNO_3$ Mixtures in Aircraft Wakes. a Modeling Study," Phys. Chem. Chem. Phys., **1**(24), pp. 5503–5509.

[28] Detwiler, A. G., and Jackson, A., 2002, "Contrail Formation and Propulsion Efficiency," J. Aircr., **39**(4), pp. 638–644.

[29] Travis, D. J., Carleton, A. M., and Lauritsen, R. G., 2002, "Climatology: Contrails Reduce Daily Temperature Range—A Brief Interval When the Skies Were Clear of Jets Unmasked an Effect on Climate," Nature (London), **418**(6898), pp. 601–601.

[30] Boucher, O., 1999, "Air Traffic May Increase Cirrus Cloudiness," Nature (London), **397**(6714), pp. 30–31.

[31] Meyer, R., Mannstein, H., Meerkotter, R., Schumann, U., and Wendling, P., 2002, "Regional Radiative Forcing by Line-Shaped Contrails Derived From

Satellite Data," J. Geophys. Res., [Atmos.], **107**(D10), 4113.
[32] Marquart, S., Ponater, M., Mager, F., and Sausen, R., 2003, "Future Development of Contrail Cover, Optical Depth, and Radiative Forcing: Impacts of Increasing Air Traffic and Climate Change," J. Clim., **16**(17), pp. 2890–2904.
[33] Minnis, P., Ayers, J. K., Palikonda, R., and Phan, D., 2004, "Contrails, Cirrus Trends, and Climate," J. Clim., **17**(8), pp. 1671–1685.
[34] Schumann, U., 2003, "Aviation, Atmosphere, and Climate—What Has Been Learned," *European Conference on Aviation, Atmosphere and Climate (AAC), Proceedings of an International Conference*, European Commission, Friedrichshafen, Germany, pp. 349–351.
[35] Arnold, F., Stilp, T., Busen, R., and Schumann, U., 1998, "Jet Engine Exhaust Chemiion Measurements: Implications for Gaseous SO_3 and H_2SO_4," Atmos. Environ., **32**(18), pp. 3073–3077.
[36] Curtius, J., Arnold, F., and Schulte, P., 2002, "Sulfuric Acid Measurements in the Exhaust Plume of a Jet Aircraft in Flight: Implications for the Sulfuric Acid Formation Efficiency," Geophys. Res. Lett., **29**(7), 1113.
[37] Katragkou, E., Wilhelm, S., Arnold, F., and Wilson, C., 2004, "First Gaseous Sulfur (VI) Measurements in the Simulated Internal Flow of an Aircraft Gas Turbine Engine During Project Partemis," Geophys. Res. Lett., **31**(2), 2117.
[38] Karcher, B., Turco, R. P., Yu, F., Danilin, M. Y., Weisenstein, D. K., Miake-Lye, R. C., and Busen, R., 2000, "A Unified Model for Ultrafine Aircraft Particle Emissions," J. Geophys. Res., [Atmos.], **105**(D24), pp. 29379–29386.
[39] Harris, B. W., 1990, "Conversion of Sulfur-Dioxide to Sulfur-Trioxide in Gas-Turbine Exhaust," Trans. ASME: J. Eng. Gas Turbines Power, **112**(4), pp. 585–589.
[40] Arnold, F., Scheid, J., Stilp, T., Schlager, H., and Reinhardt, M. E., 1992, "Measurements of Jet Aircraft Emissions at Cruise Altitude. I. The Odd-Nitrogen Gases NO, NO_2, HNO_2 and HNO_3," Geophys. Res. Lett., **19**(24), pp. 2421–2424.
[41] Brundish, K., Clague, A. R., Wilson, C. W., Miake-Lye, R. C., Brown, R. C., Wormhoudt, J., Lukachko, S. P., Chobot, A. T., Yam, C. K., Waitz, I. A., Hagen, D. E., Schmid, O., and Whitefield, P., 2007, "Evolution of Carbonaceous Aerosol and Aerosol Precursor Emissions Through a Jet Engine," J. Propul. Power, **23**(5), pp. 959–970.
[42] Haschberger, P., and Lindermeir, E., 1997, "Observation of NO and NO_2 in the Young Plume of an Aircraft Jet Engine," Geophys. Res. Lett., **24**(9), pp. 1083–1086.
[43] Miake-Lye, R. C., Martinez-Sanchez, M., Brown, R. C., and Kolb, C. E., 1993, "Plume and Wake Dynamics, Mixing, and Chemistry Behind a High-Speed Civil Transport Aircraft," J. Aircr., **30**(4), pp. 467–479.
[44] Hunter, S. C., 1982, "Formation of SO_3 in Gas-Turbines," J. Eng. Power, **104**(1), pp. 44–51.
[45] Wilson, C. W., Clague, A. R., Pourkashanian, M., and Ma, L., 2004, "Chemical Kinetic Modeling of the Evolution of Gaseous Aerosol Precursors Within a Gas Turbine Engine," *ASME Turbo Expo 2004*, American Society of Mechanical Engineers, Vienna, Austria, Paper No. GT2004-53704.
[46] Mueller, M. A., Yetter, R. A., and Dryer, F. L., 2000, "Kinetic Modeling of the $CO/H_2O/O_2/NO/SO_2$ System: Implications for High Pressure Fall-Off in the $SO_2+O(+M)=SO_3(+M)$ Reaction," Int. J. Chem. Kinet., **32**(6), pp. 317–339.
[47] Glarborg, P., Kubel, D., DamJohansen, K., Chiang, H. M., and Bozzelli, J. W., 1996, "Impact of SO_2 and NO on CO Oxidation Under Post-Flame Conditions," Int. J. Chem. Kinet., **28**(10), pp. 773–790.
[48] Brown, R. C., Miake-Lye, R. C., Lukachko, S. P., and Waitz, I. A., 2002, "Heterogeneous Reactions in Aircraft Gas Turbine Engines," Geophys. Res. Lett., **29**(10), 1424.
[49] Dawes, W. N., 1992, "The Simulation of 3-Dimensional Viscous-Flow in Turbomachinery Geometries Using a Solution-Adaptive Unstructured Mesh Methodology," ASME J. Turbomach., **114**(3), pp. 528–537.
[50] Brown, P. N., Byrne, G. D., and Hindmarsh, A. C., 1989, "Vode—A Variable-Coefficient Ode Solver," SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput., **10**(5), pp. 1038–1051.
[51] Kee, R. J., Rupley, F. M., and Miller, J. A., 1991, "CHEMKIN-II: A FORTAN Chemical Kinetics Package for the Analysis of Gas-Phase Chemical Kinetics, SAND89–8009," Sandia National Laboratories, Report No. SAND89-8009.
[52] Lukachko, S. P., Kirk, D. R., and Waitz, I. A., 2003, "Gas Turbine Engine Durability Impacts of High Fuel-Air Ratio Combustors, Part I: Potential for Secondary Combustion of Partially Reacted Fuel," ASME J. Eng. Gas Turbines Power, **125**(3), pp. 742–750.

The Effect of Liquid-Fuel Preparation on Gas Turbine Emissions

Sosuke Nakamura[1]
Vince McDonell
Scott Samuelsen

UCI Combustion Laboratory,
University of California,
Irvine, CA 92697-3550

*The emissions of liquid-fuel fired gas turbine engines are strongly affected by the fuel preparation process that includes atomization, evaporation, and mixing. In the present paper, the effects of fuel atomization and evaporation on emissions from an industrial gas turbine engine were investigated. In the engine studied, the fuel injector consists of a coaxial plain jet airblast atomizer and a premixer which consists of a cylindrical tube with four mixing holes and swirler slits. The goal of this device is to establish a fully vaporized, homogeneous fuel/air mixture for introduction into the combustion chamber and the reaction zone. In the present study, experiments were conducted at atmospheric pressure and room temperature as well as at actual engine conditions (*0.34 MPa, 740 K*) both with and without the premixer. Measurements included visualization, droplet size, and velocity. By conducting tests with and without the premixing section, the effect of the mixing holes and swirler slit design on atomization and evaporation was isolated. The results were also compared with engine data and the relationship between premixer performance and emissions was evaluated. By comparing the results of tests over a range of pressures, the viability of two scaling methods was evaluated with the conclusion that spray angle correlates with fuel to atomizing air momentum ratio. For the injector studied, however, the conditions resulting in superior atomization and vaporization did not translate into superior emissions performance. This suggests that, while atomization and the evaporation of the fuel are important in the fuel preparation process, they are of secondary importance to the fuel/air mixing prior to, and in the early stages of the reaction in, governing emissions.* [DOI: 10.1115/1.2771564]

Introduction

To meet increasingly stringent emissions regulations, combustors for the next generation of advanced gas turbine engines are being designed to reduce pollutant formation while maintaining efficient performance. In order to achieve low emissions combustion, many strategies are being considered. One strategy that is now common is the use of lean premixed combustion for gaseous fuels [1,2]. By operating under well mixed and at lean conditions, reaction temperatures can be reduced both locally and on average. For liquid fired systems, achieving low emissions requires not only sufficient mixing of fuel vapor and air, but also sufficient time for atomization and vaporization. If liquid droplets enter the reaction zone, combustion of the vapor produced by these droplets can take place near stoichiometric conditions depending upon the local conditions [3]. As a result, the preparation of the fuel/air mixture for liquid fired systems is inherently more complicated than it is for gaseous fuels [4].

Operating fuel lean, with prevaporized fuel and premixing with air (LPP), has demonstrated low emission levels [1,5,6]. LPP involves the introduction of a uniformly lean mixture of fuel vapor and air into the combustor. There are several investigations about evaporation and mixing of liquid fuel in the LPP system [5,7,8]. The process in the present system is typical of LPP systems and involves several relatively discrete steps—a twin-fluid atomization approach, followed by vaporization and mixing. Twin-fluid atomization is commonly used in gas turbine applications to enhance mixing and the production of fine droplets with relatively low liquid pressure drops [9,10]. The rate of vaporization/mixing is determined largely by the enthalpy and fluid mechanics of the combustion and swirl air.

The study of spray phenomena for gas turbines is challenging due to the difficulties of acquiring information at engine conditions. Droplet size distribution, velocity, and evaporation are critical for fuel-air mixing inside the combustor, but it is difficult to investigate these characteristics at actual engine conditions. As a result, it is common to conduct measurements in lower pressure and nonreacting test rigs. However, many aspects of the behavior of spray at lower pressure conditions may not be representative of the behavior in the engine. Literature regarding scaling methods for two-phase phenomenon in gas turbine combustion is sparse, but work describing strategies for identifying operating conditions that enable measurements at low pressure and temperature while conserving certain flow quantities such as momentum ratio, pressure loss, Re, M, or St can be found [11–15]. To date, no studies that systematically explore the relative behavior of atomization, vaporization, and mixing at engine conditions and relating them to emissions performance have been conducted. As a result, the current study has been undertaken.

The objective of the present study is to relate the atomization, vaporization, and mixing phenomena occurring within a fuel injector/premixer assembly to emissions performance of a small gas turbine operated on Diesel Fuel No. 2 (DF-2).

Approach

The overall approach taken in the present study to accomplish the objectives is as follows:

- Select an injector that is suitable to meet the objective of the present study;
- Characterize the emissions produced by the gas turbine engine operating at baseline conditions and at conditions where the atomization is systematically varied;

[1]Visiting Scientist, Mitsubishi Heavy Industries, Ltd., Hyogo, Japan.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 31, 2006; final manuscript received October 31, 2006; published online February 29, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2006: Land, Sea and Air (GT2006), Barcelona, Spain, May 8–11, 2006, Paper No. GT2006-90730.

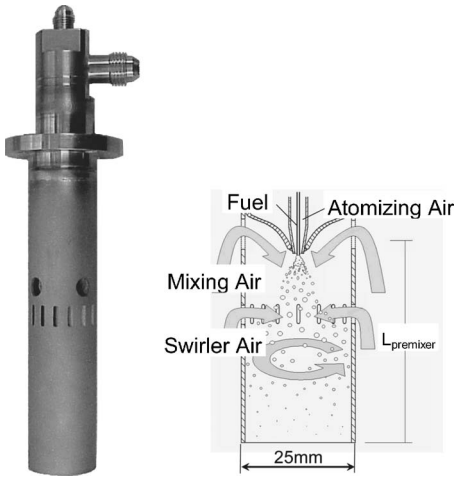


Fig. 1 Liquid atomizer/premixer assembly

- Characterize the spray produced by the engine fuel injector in the absence of the premixer hardware under engine and scaled conditions;
- Characterize the spray produced by the engine injector/premixer assembly under engine and scaled conditions;
- Analyze the results and assess the relative roles of atomization, vaporization, and mixing in the emissions performance; and
- Analyze the results to evaluate the relative merit of difference scaling strategies for the injector operation in producing "engine like" behavior.

Experiment

Injector. The injector selected for this study (Fig. 1) is the the fuel atomizer/premixer assembly used in the Capstone Turbine Corporation C30 microturbine generator (MTG). The injector was selected because the configuration of the atomizer is very simple. Figure 1 also illustrates a mechanistic representation of the spray phenomena that occur within the assembly and are the basis for the objectives of the study.

The premixer has a length $L_{premixer}$, which is the length the atomized fuel droplets have to vaporize and mix with air before they are reacted. The residence time $t_{residence}$ and evaporation time t_{vap} are key parameters that determine the extent to which the fuel has vaporized prior to entry into the combustion chamber. The spray is produced by an air-blasted plain jet atomizer that features centerline injection of a liquid column surrounded by a coflowing high velocity annulus of air. The premixer incorporates a radial swirler to help aid in mixing and distributing the spray. Air is added through four round orifices around the atomizer to provide for combustion downstream of the injector as well as to improve mixing. Each air stream can potentially play an important role in the preparation of the fuel/air mixture and therefore influence the formation of pollutants in the gas turbine combustor. The swirl and mixing air are referred to as "combustion air."

Injector experiments were carried out on two test rigs, each of which is described briefly in this section.

Rig 1: MTG (Microturbine Generator). The first test rig is the recuperated Model C30 liquid fired microturbine generator. The gas turbine is shown in Fig. 2. The Capstone C30 Microturbine Generator shows a photograph of the integrated gas turbine/recuperator along with a cross section through the combustor in a plane perpendicular to the engine shaft. The plane shown cuts through the three fuel injectors which introduce fuel and air tangentially into the combustor.

Mtg Instrumentation. The MTG was retrofitted with additional instrumentation to monitor conditions inside the fuel injector. One of the three injector assemblies was fabricated to allow the connection of pressure (pressure transducers) and temperature (thermocouple) measuring devices near the airblast nozzle exit of the injector during operation. The instrumented assembly is shown in the lower left corner of Fig. 2. Data obtained with this instrumentation were used to establish the conditions to simulate in performing atomization studies of a single injector. An additional air line was added in order to make it possible to increase atomizing air flow rate.

Emissions Monitoring. Exhaust emissions were measured with a Horiba PG 250 emissions analyzer via an extractive sample probe centered at the exit plane of the exhaust stack. Due to the recuperator, the products of combustion emitted were completely mixed and were highly uniform across the exhaust stack. Emissions were characterized for 50–100% load operating conditions. The accuracy of the measurements is ± 0.25 ppm NO_x and ± 2 ppm CO based on instrument specification. The analyzers were zeroed and spanned before and after each measurement campaign and revealed negligible drift and bias.

Rig 2: High-Pressure Atomization Test Rig. Experiments were conducted in a facility designed to produce conditions found in conventional and advanced aero-engine combustors. A picture of the facility is shown in Fig. 3 and is described in more detail elsewhere [16–18]. The pressure vessel and window arrangement used is designed to withstand actual engine conditions. The test section (described below) is mounted centered within the main pressure vessel as shown and injects downward at a plane where various optical ports are available. The entire vessel is suspended from a 2D horizontal traverse system that allows diagnostics to be fixed to minimize alignment issues. A seal block located at the center of the top flange allowed vertical traversing of the test article within the vessel. All motion is monitored by a magnetic pickup and a precision readout. dc motors are remotely controlled from within the facility control room. The system allows positioning to within 0.2 mm.

In the present study, experiments were conducted on two dif-

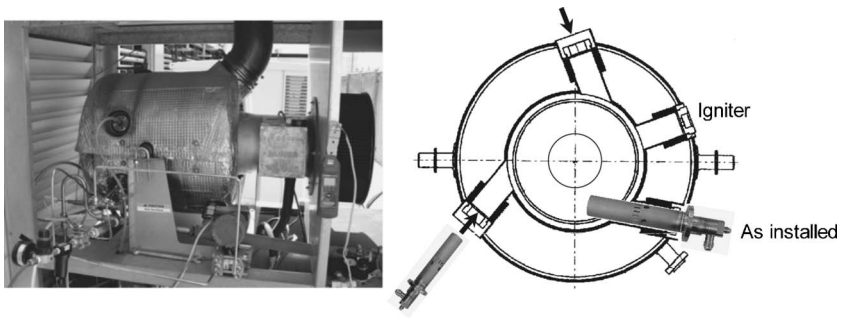


Fig. 2 Capstone C30 microturbine generator

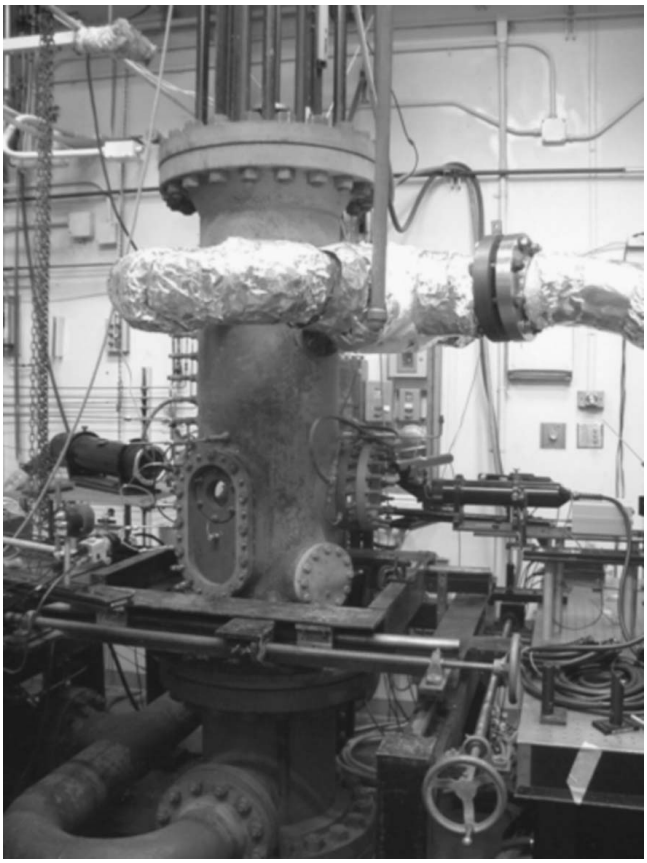


Fig. 3 High-pressure facility

ferent injector configurations. The first configuration utilized the fuel injector/premixer assembly as shown in Figs. 1 and 4. The air box is designed to allow the swirl and mixing air to be admitted into the premixer in a manner similar to that which occurs in the engine. The second configuration consists of the fuel injector atomization elements in the absence of the premixer as shown in Fig. 5. This "airblast nozzle only" configuration was implemented to isolate the role of the atomizing air on the atomizer performance.

All atomization tests were conducted by using the same fuel as the engine and the same injector hardware as the engine. Comparing the results from the two configurations just described allows the effect of atomizing air and combustion air on atomization and evaporation to be isolated.

Figure 6 presents details regarding the internal flow distribution of the air for the atomizer/premixer assembly and for the atomizer alone. Since independent control over two preheated air streams was not available, a flow distribution plate was inserted into the chamber ("screen plate"). The plate: (1) provides screen air to suppress recirculation of mist back up into the measurement plate; and (2) mimics the boundary conditions utilized in the previous studies. Since temperatures at the engine conditions achieved within the high-pressure facility were well above the autoignition

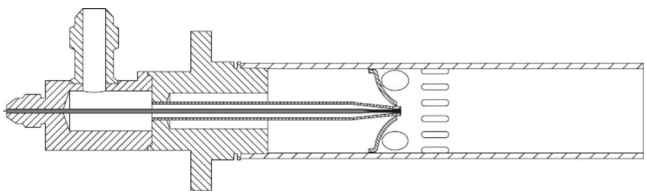


Fig. 4 Atomizer/premixer assembly

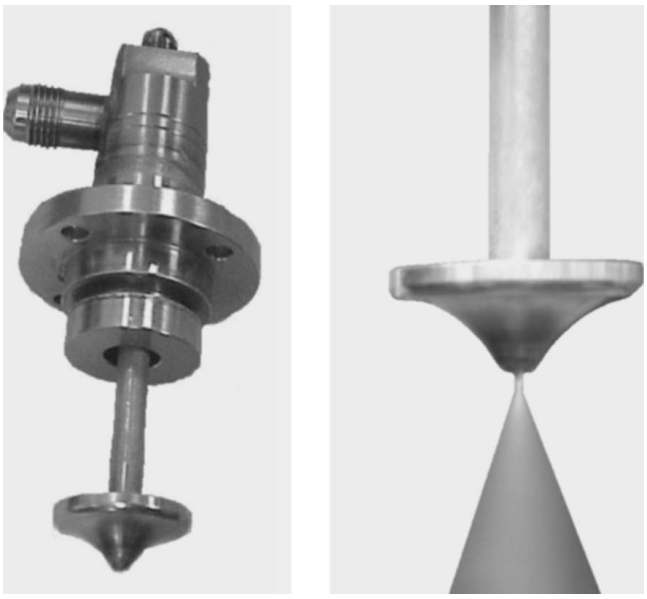


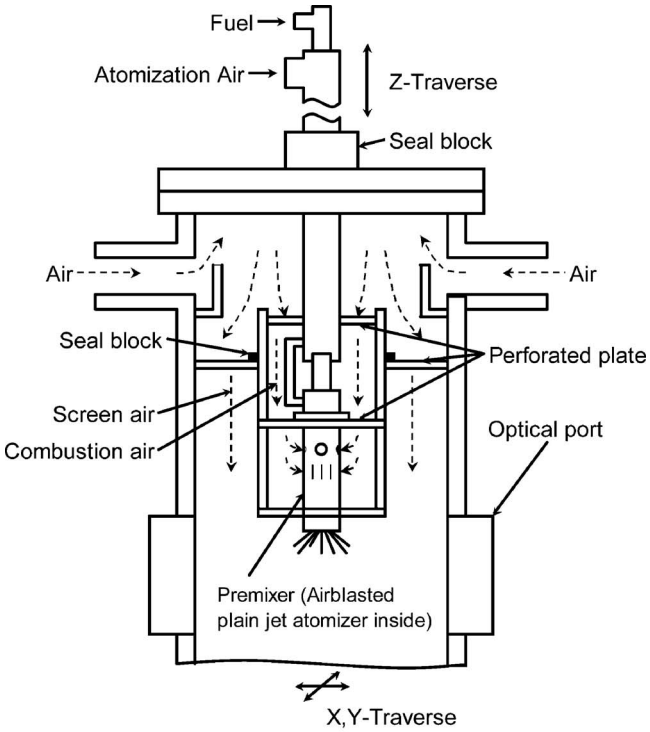
Fig. 5 Injector without premixer tube

temperature for DF-2 (~500 K) [19], nitrogen was mixed into air to weaken the mixture. Nitrogen was added such that the oxygen concentration in screen air, combustion air, and atomizing air was reduced from 21% to 9–12% for all tests conditions. Autoignition was not observed during the test campaign.

Figure 7 presents a cross section of the vessel at the optical access height. Thirty eight-mm-thick fused silica windows are used to facilitate optical access at the pressures and temperatures of interest for the diagnostics utilized which are described briefly in this section. Internal window cooling/purge was used to maintain the clarity of the windows at the conditions studied.

Phase Doppler Interferometry. Droplet size, radial and axial velocity, and volume flux distributions were measured with a two-component phase Doppler interferometer (PDI). A frequency domain processor was used (Aerometrics Model RSA 1000). A fiber optic coupled transmitter was used with 40 mm beam spacing for both sets of beams. A 30-deg forward scatter receiver position was utilized to collect the light refracted by the droplets. While this angle maximizes signal levels, it does lead to some uncertainty associated with the change in droplet refractive index as the drop temperature increases [18]. The uncertainty associated with this effect, however, does not impact the conclusions drawn from the PDI measurements. Measurements were taken for the tests with "atomizer only" at three different locations downstream of the atomizer exit ($Z=25$ mm, 50 mm, and 75 mm).

Visualization. A 1.5-mm-thick laser sheet was formed by directing the multiline beam from a 5 W Ar+ laser into a −6.3 mm focal length cylindrical lens. As shown in Fig. 7, the sheet is oriented such that it can intersect the plane containing the injector centerline (the sheet is fixed in space relative to the chamber such that when the chamber is traversed 75 mm laterally the sheet will intersect the injector centerline). By recording the scattering of light sheet by the spray, an estimate of the spray evaporation can be obtained. Digital images were obtained at 30 frames/s with 1/100–1/2000 s exposure times using a Hitachi IEEE-1394 camera (Model KP-D20BU). One hundred individual frames from the *.avi file were extracted to *.tif format and then averaged (Media Cybernetics ImagePro Plus Ver. 5.1). In addition, single frame digital images were obtained with a consumer grade 3.2-megapixel camera.



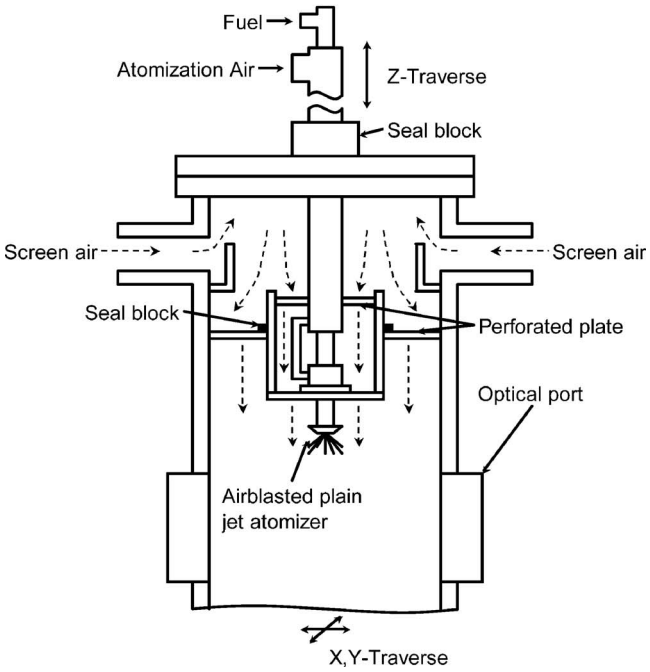


Fig. 6 Pressure vessel internal airflow schematic

Test Conditions

MTG & Atomizer/Premixer Assembly (Phase 1). All tests were conducted at a condition equivalent to 100% engine load. In addition, the atomizing air to liquid mass flow rates (ALR) was varied from 0.3 to 1.2. Typical engine conditions at 100% are shown in Table 1. All tests carried out in the high-pressure vessel were set to the same values found in the engine except for the combustion air temperature. The highest premixer inlet temperature that could be attained was 740 K.

Atomizer Only Tests (Phase 2). All tests were carried out at the following conditions shown in Table 2. Tests with high temperature were conducted to isolate the spray characteristics without mixing air and swirler air. Tests at room temperature were done to verify the scaling method and to investigate the pressure effect on atomization.

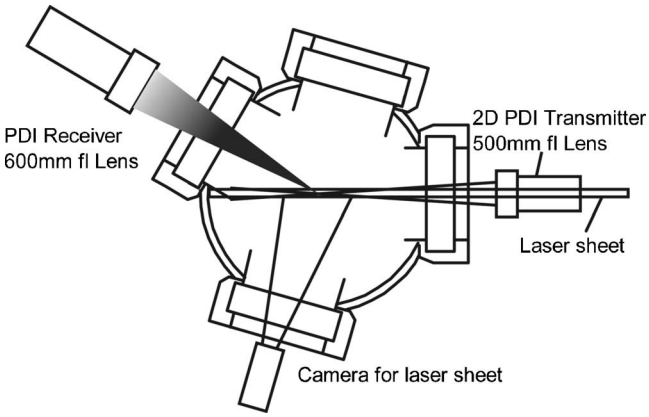


Fig. 7 Cross section at optical access height

Results

Baseline Performance of MTG. Emission measurement was obtained for the C30 MTG operated on DF-2 with a set of commercial injectors. Results versus load setting are presented in Fig. 8 for NO and CO emissions. CO emissions decrease with load and are below 10 ppmvd at 15% O_2 for 50–100% load operation. NO emissions increase with load and are approximately 15 ppmvd at 15% O_2 at maximum power output. ALR at maximum power output (100% load) is set to 0.3.

To help establish the degree to which the premixer is performing optimally in terms of prevaporizing and premixing the fuel and air prior to entry into the combustor, the baseline emissions measurements were compared to results determined for a "nearly perfect" prevaporized and premixed combustion system [20]. Figure 9 shows NO_x as a function of average reaction temperature for the baseline results along with the generalized data obtained for previous mixers (shown by small symbols and a black line). The upper slanted line shows the estimated data for diesel fuel using data obtained by other workers [21]. According to the results shown in Figs. 8 and 9, it appears that vaporization and mixing in the C30 liquid fuel injector can be improved. Note that results for

Table 1 Engine conditions and Phase 1 test conditions

Parameter	Setting
Power output (kW)	25
Combustion air press. (MPa)	0.34
Premixer inlet temp. in engine/high-pressure test rig (K)	810/740
Atomizing air temp. in engine/high-pressure test rig (K)	305/420
$m_{\text{combustion air}}$ (kg/min)	1.82
Primary equivalence ratio	0.52
m_{fuel} (kg/min)	0.053
$m_{\text{atomizing air}}$ (kg/min)	0.016–0.064
ALR	0.3–1.2

Table 2 Test conditions for Phase2

Parameter	Setting
Screen air press. (MPa)	0.1–0.9
Screen air temp. (K)	300, 740
m_{fuel} (kg/min)	0.027–0.184
$m_{\text{atomizing air}}$ (kg/min)	0.008–0.064
ALR	0.3–1.2

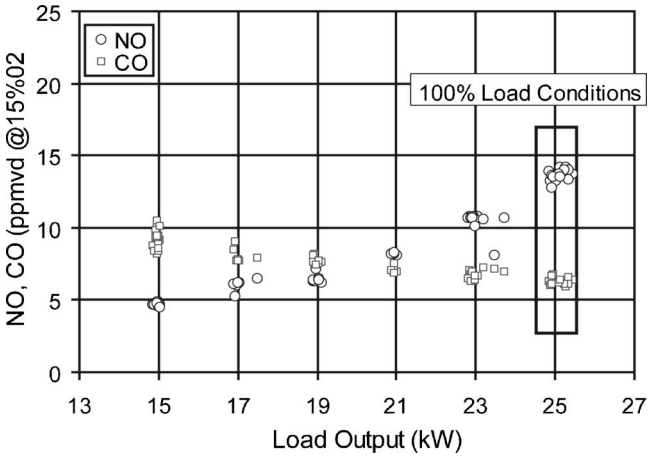


Fig. 8 Emission versus load output

a natural gas fired 60 kW MTG operating at similar conditions and with a similar fuel injection approach but with excellent premixing [22] affirms this opportunity. As a result, an opportunity to improve the emissions performance through an improvement in the vaporization (e.g., reducing droplet size) and premixing of the injector is apparent.

Spray Behavior

Uncertainty. An evaluation of spray symmetry and repeatability was carried out to help establish uncertainty in the PDI measurements. Measurements were carried out at 300 K and 0.1 MPa in the pressure vessel for two orthogonal traverses at a distance of 25 mm downstream of the injector with fuel flow rate of 0.027 kg/min and ALR of 0.3. The results are shown in Fig. 10 which presents error bars based on the differences observed between the two radial profiles obtained. The ratio of volumetric flow rate that is determined by integrating the volume flux profile to metered flowrate of liquid fuel is 0.53–0.86–0.78 for $Z=25-50-75$ mm. The ratio is within experimental uncertainty at 50 mm and 75 mm. At 25 mm, the spray density is high enough to cause some drops to be missed. However, it has been shown that a lack of mass conservation based on flux measurements does not mean that measurements of drop size and velocity are incorrect [23]. It was observed at higher temperatures and pressures that the symmetry of the spray structure changed somewhat. It is hypothesized that thermal expansion of fuel and atomizing the air line at the tip of the atomizer alters the spray structure.

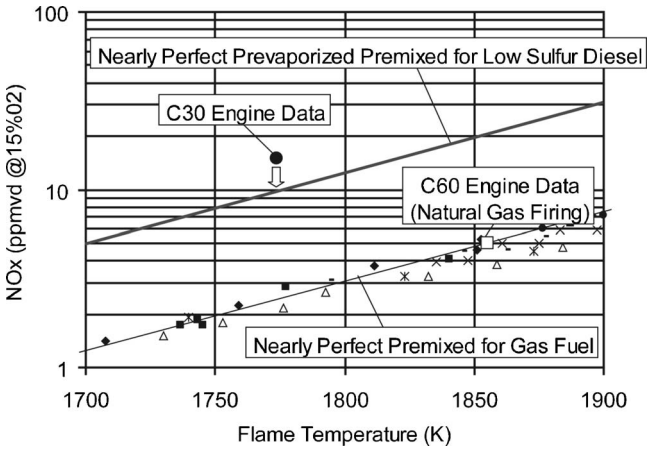


Fig. 9 Effects of nonuniform fuel and air mixing on NO_x formation

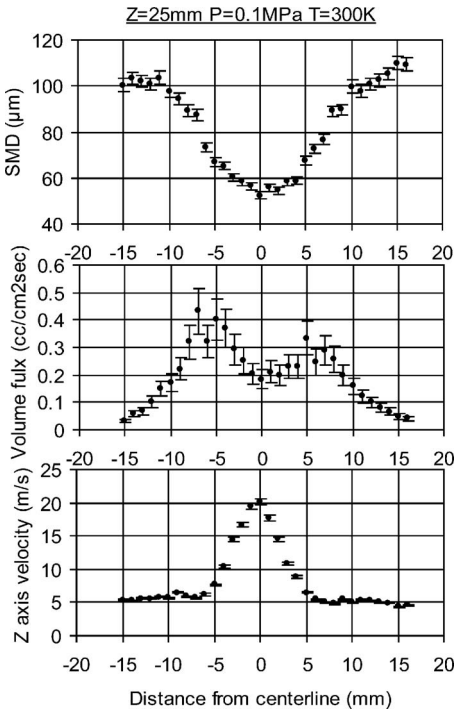


Fig. 10 Symmetry assessment and uncertainty of spray measurement

Premixer Spray Visualization. Figure 11 presents images of the spray structure obtained from scattering of a laser sheet light in the plane of the centerline of the system at the engine condition (0.34 MPa, 740 K, constant fuel flow rate 0.053 kg/min). The 100 frame time averaged images shown in Fig. 11 reveal little or no droplet scattering. Careful interrogation of the video images does reveal the presence of some drops at ALR=0.3 from time to time (e.g., Fig. 11(*a*)). At higher ALRs, no drops are observed exiting the premixer assembly. This result is interesting in that it suggest that, at the engine condition, some droplets exit the injector prior to complete vaporization.

Spray Behavior in Absence of Premixer. Spray images obtained from scattering of a laser sheet and digital camera downstream of the atomizer are shown in Fig. 12. Images were taken in same manner as described above. As shown in the time averaged image of laser sheet scattering and the digital still image (Fig. 12(*a*)), the droplets are obviously vaporizing as they move downstream. However, for the case corresponding to the engine condition (Fig. 12(*a*)), it is apparent that droplets are still present at $Z=75$ mm, which corresponds to the premixer exit plane. Hence, the results observed for the atomizer only correspond well to the results for the atomizer/premixer assembly. In both cases, the results suggest that liquid droplets exit the premixer assembly at engine conditions. This could be a source of the NO_x observed in excess of the perfect mixing case shown in Fig. 9.

PDI measurements of droplet size (presented as Sauter mean diameter, D_{32}), volume flux (volume of liquid passing through the interferometric probe volume cross section per unit time), and velocity vectors are presented in Figs. 13–15, respectively.

Figure 13 presents radial profiles of the droplet size distribution D_{32}. As mentioned above, due to thermal expansion of the fuel and an atomizing air line, the radial profiles are not symmetric, but still provide some important insight into the characteristics of the spray. It is worth noting that the change in symmetry at engine conditions was observed visually as well as in the PDI profiles. Hence, the lack of symmetry in the profiles is not a result of measurement anomalies or errors. Compared to $Z=25$ mm, with

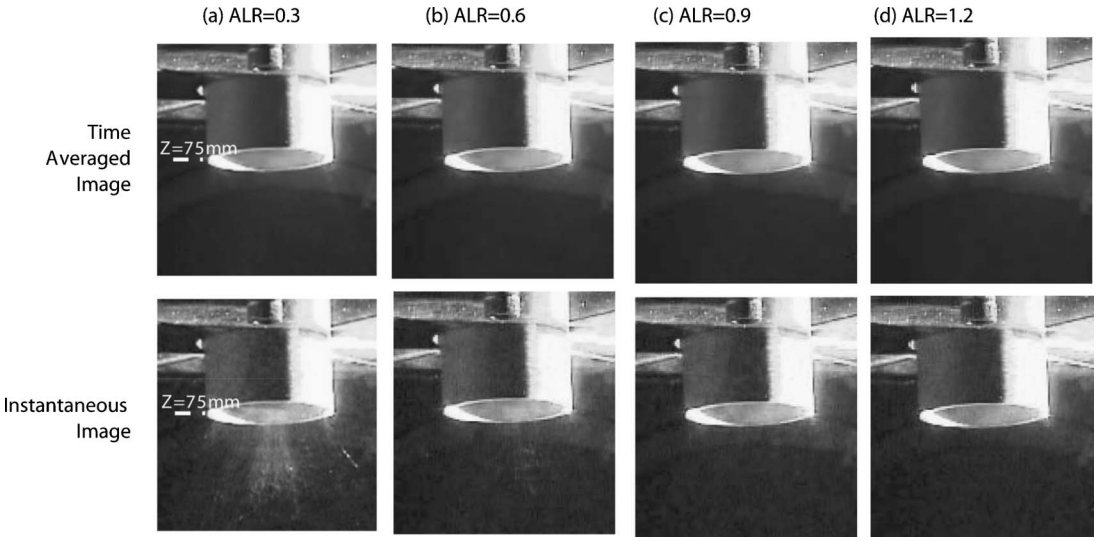


Fig. 11 Time averaged and short exposure laser sheet scattering images at the exit of premixer for actual engine condition cases (fuel flow rate=0.053 kg/min, *P*=0.34 MPa, *T*=740 K)

the profiles at $Z=50$ mm and 75 mm an increase in the distribution D_{32} is observed. At the conditions studied (740 K), small droplets rapidly vaporize, leaving a larger proportion of larger drops which therefore increases the D_{32}.

Profiles of volume flux are shown in Fig. 14 and again suggest the presence of an asymmetry. The results also reveal that droplets do exist at an axial location corresponding to the exit of premixer ($Z=75$ mm). Furthermore, the results indicate that, as ALR increases, the droplet volume flux decreases.

Hence, at ALR of 1.2, the droplets are smaller and the volume flux of liquid fuel is lower in number when compared with the engine condition (ALR=0.3).

Although the results obtained for the spray in the absence of the premixer suggest the presence of drops at 75 mm, the addition of the 740 K swirl and mixing air plays a significant role in the further evaporation of the drops. This is concluded by comparing the results from Fig. 11 with Fig. 12. Hence, the vaporization characteristics of the spray produced by the atomizer cannot be taken to represent the behavior within the premixer. For example, previous studies suggest secondary atomization due to combustion air may be occurring based on Weber number analysis [24].

Influence of Atomizing Air (ALR=0.3–1.2). Figures 13–15 also show the influence of ALR. By changing the ALR, the atomization behavior of the spray can be altered while maintaining the rest of the parameters constant. It is important to note that the atomizing air flow rate is only 1–2% of the total premixer air flow depending upon the ALR used. As ALR increases a number of consistent trends are observed. First, Fig. 13 shows that increasing the ALR substantially reduces the droplet sizes. Second, Fig. 14

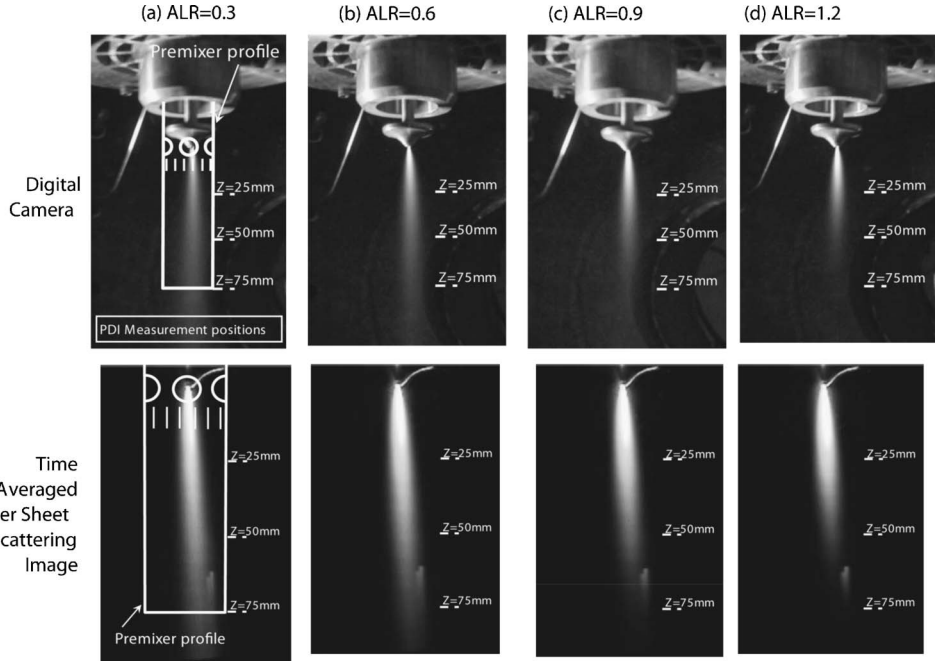


Fig. 12 Digital camera and time averaged laser sheet scattering images of the spray plume for actual engine condition cases (fuel flow rate=0.053 kg/min, *P*=0.34 MPa, *T*=740 K)

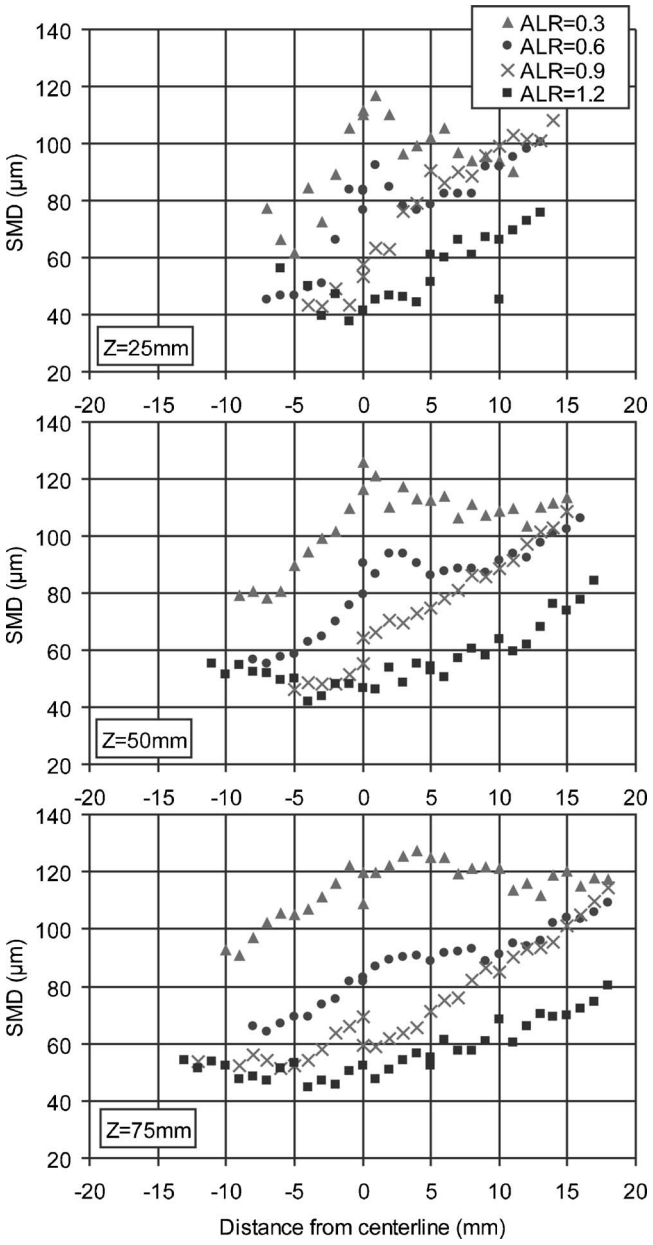


Fig. 13 Radial profiles of D_{32} for actual engine condition cases

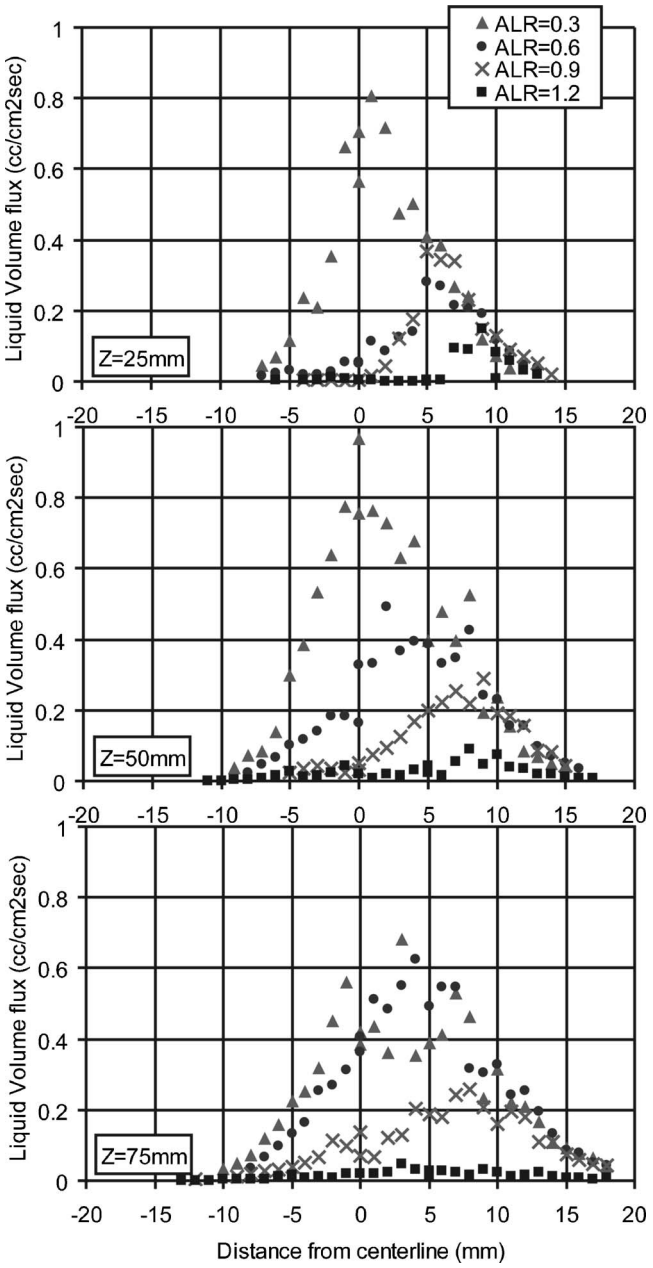


Fig. 14 Radial profiles of volume flux for actual engine condition cases

shows that the volume flux is significantly reduced, which is attributed to the smaller droplet sizes. Finally, the axial velocity of the droplets is substantially higher as ALR increases as shown in Fig. 15. Since the D_{32} is lower and axial velocities are higher, vaporization will occur more quickly due to the higher surface area and increased convective mass transfer. As a result, it is not surprising that the volume flux at each axial location is reduced with ALR. This is further corroborated with the visualization results shown in Fig. 11.

According to these results, it can be concluded that evaporation and mixing of fuel vapor and air (combustion air=mixing air +swirler air) will be promoted at higher ALR. Given the trends observed with ALR in terms of vaporization and mixing, it is reasonable to expect that the emissions performance (at least NO_x) for higher atomizing air ALRs should be superior when compared to the baseline conditions.

In order to verify the effect of ALR on emissions, ALR was varied at 100% load condition using the MTG with an additional atomizing air circuit added to independently vary atomizing air flow. In the tests, ALR was changed from 0.3 to 1.2 at constant 100% power output, constant fuel flow rate, and constant combustion air flow rate. The results are shown in Fig. 16. Interestingly, the results are completely the opposite of the expected tendencies. According to the results from atomization tests, it is considered that higher ALR can promote both evaporation and eventually mixing. In contrast, NO increased with higher ALR. NO emissions are approximately 35 ppmvd at 15% O_2 at ALR=1.2. CO remains almost constant over the entire range of ALR.

Possible causes for this nonintuitive behavior are hypothesized as follows. Spray images obtained from the scattering of a laser sheet reveal the presence of a strong recirculation zone one diameter downstream of the premixer (Fig. 17), produced by the radial swirler. Therefore, it can be assumed that the flame will be anchored within this region over the entire range of ALR.

The lower ALR cases have greater residence time (=mixing time) than higher ones due to the difference in droplet velocity as

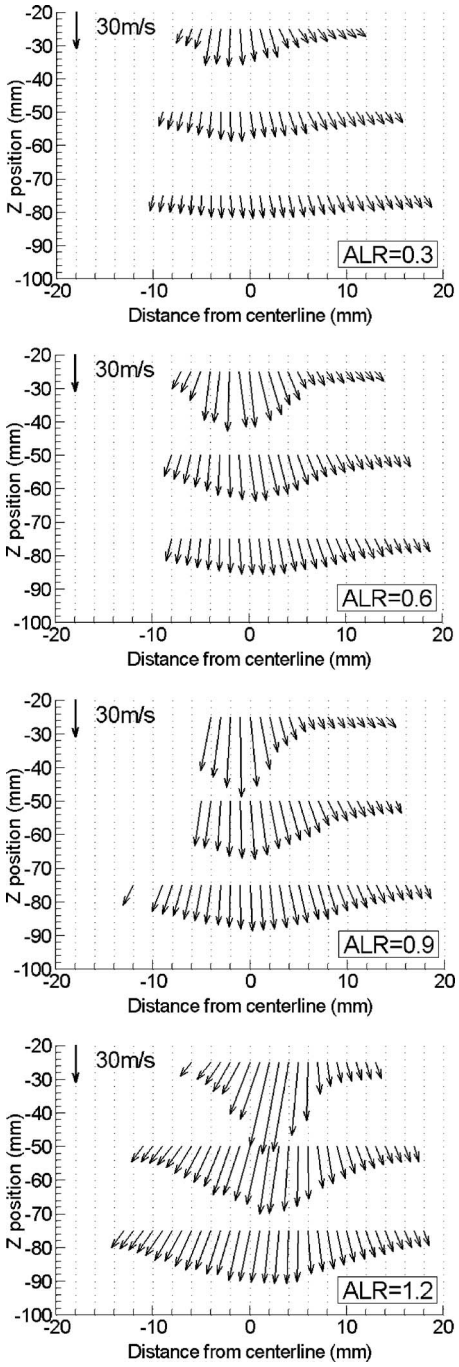


Fig. 15 Velocity vectors in *R-Z* plane for actual engine condition cases

shown in Fig. 15. Even if droplets are present at the exit of the premixer, sufficient time exists for vaporization and mixing before reaching the flame. On the other hand, the higher ALR cases have shorter residence time than lower ones due to the same reason above. Even if spray evaporates before reaching the flame there is not enough time for mixing. Therefore, one possible cause for the NO_x increase with higher ALR is that the mixing process is dominated by residence time, and NO_x increases with higher ALR due to higher droplet velocities. Another possible cause is that the momentum interaction between fuel spray and combustion leads to locally rich areas with increasing ALR.

Evaluation of Scaling Methods. Even though experiments were carried out at actual engine conditions, additional studies

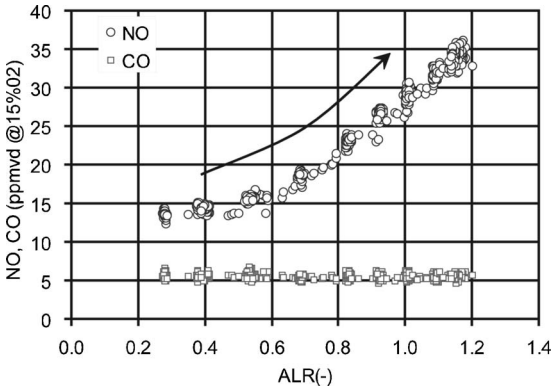


Fig. 16 Emission versus ALR at 100% load

were carried out to investigate the viability of two atomization scaling approaches. As a first step, pressure scaling was evaluated. As a result, experiments were carried out at various pressure conditions in order to simulate droplet size distribution and spray angle. Tests were carried out for the "atomizer only" configuration at room temperature. First, the actual engine condition was set as the standard condition (P=0.34 MPa) and atomizing air and fuel flow rate were changed at each pressure according to scaling strategy. Then the profiles of droplet size were compared with the standard condition.

Strategy 1. Since the configuration of the atomizer is similar to a plain jet airblast atomizer used to develop the empirical correlation shown in Eq. (1) [25], it was considered as a first step to estimate dominant factors controlling spray characteristics, especially for D_{32}.

$$\frac{D_{32}}{d_o} = 0.48\left(\frac{\sigma}{\rho_A U_R^2 d_o}\right)^{0.4}\left(1+\frac{1}{\mathrm{ALR}}\right)^{0.4} + 0.15\left(\frac{\mu_L^2}{\sigma\rho_L d_o}\right)^{0.5}\left(1+\frac{1}{\mathrm{ALR}}\right) \quad (1)$$

where D_{32}=the Sauter mean diameter (m); d_o=liquid discharge orifice diameter (m); σ=surface tension (kg/s^2); U_R=relative velocity (coflowing); ρ_A=density of air (kg/m^3); ALR=air to liquid mass flow ratio; μ_L=liquid viscosity (kg/m s); and ρ_L=liquid density (kg/m^3).

Tests were carried out in room temperature, therefore thermal property of liquid fuel can be assumed to be identical at each

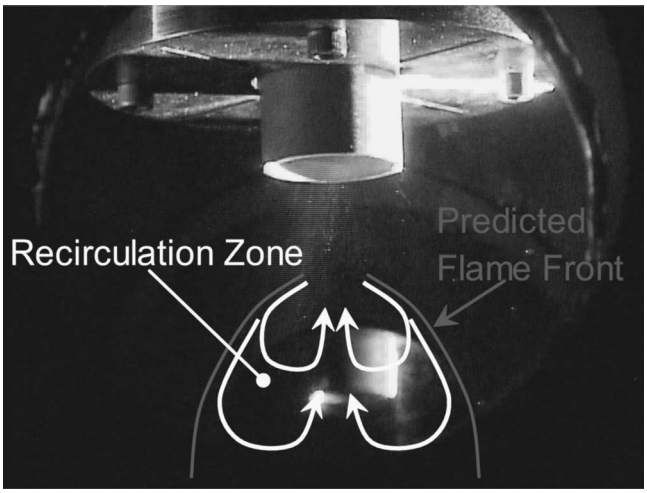


Fig. 17 Recirculation zone downstream of the premixer

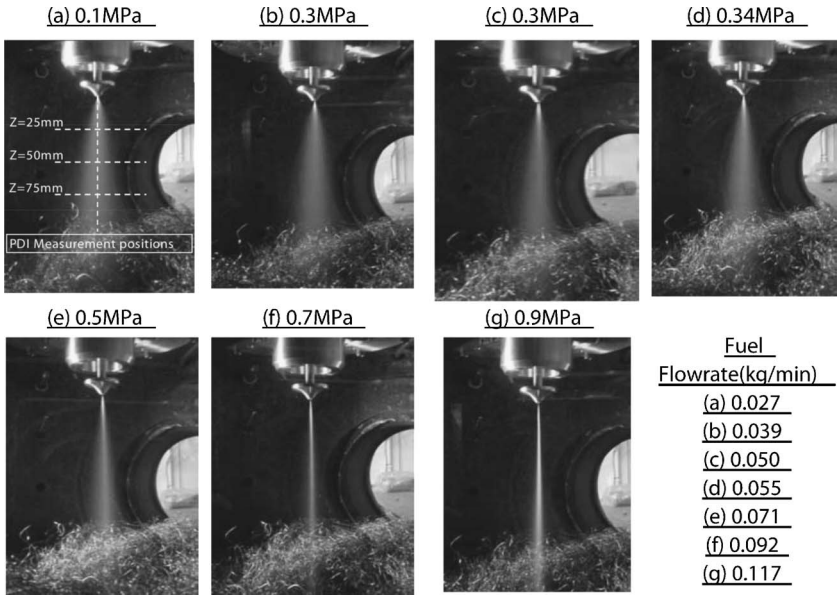


Fig. 18 Digital camera images at the exit of atomizer for Strategy 1 at different pressure (0.1–0.9 MPa) 300 K cases

pressure with the assumption of its incompressibility.

According to Eq. (1), conserving $\rho_A U_R^2$ and ALR (which is equivalent to conserving the initial kinetic energy of the atomizing air) should result in the same D_{32}.

Strategy 2. The second strategy involved increasing atomizing air flow rate to maintain the ratio of set pressure to standard pressure. This method was established in order to keep the momentum ratio of atomizing air and screen air constant so that the spray angle was considered to be the same as the standard condition. ALR was also conserved.

Figures 18 and 19 show still images of the spray obtained by a digital camera. The results shown in Figs. 18 and 19 can provide a quantified comparison of spray angle. Figure 20 shows a typical image of spray at 0.1 MPa along with a representation of the spray angle measurement procedure. A line profile was drawn across the image at an axial distance corresponding to 25 mm downstream of the atomizer. Then a line was drawn intersecting the center at the exit of the atomizer and the location on the line profile at which the intensity value was found to be saturated. The same process was repeated on the other side of the spray, and the included angle between the two lines is considered to be the spray angle.

It is observed that the spray cone exhibits a systematic collapse with higher pressure. This tendency can be observed in previous work using an air blast simplex nozzle [26]. According to their work, the spray angle correlates with fuel to atomize air momentum ratio. Figure 21 shows plots of spray angle at 25 mm downstream plotted against fuel to air momentum ratio. In the present study, when matching momentum and ALR (Strategy 1), at higher pressure the atomizing air flow rate is increased to keep $\rho_A U_R^2$ constant. However, in order to keep ALR constant, the fuel flow rate was also increased. Therefore, the fuel to atomizing air momentum ratio was increased. Because of the relative effect of $\rho_A U_R^2$ and ALR on fuel to atomizing air momentum ratio, this strategy doesn't predict spray angle behavior correctly. The same basic phenomenon occurs when considering the second scaling strategy as well.

Figures 22 and 23 present the radial profiles of D_{32} for each strategy. The radial profiles included data from a few points beyond the centerline as a check on symmetry. These tests were carried out at room temperature, hence asymmetry due to thermal expansion is not observed.

The droplet size distribution and spray angle did not match the results obtained at engine pressures (P=0.34 MPa) using either strategy. For ambient pressures above 0.2 MPa, a local peak in droplet size is evident at the centerline at Z=25 mm, 50 mm, 75 mm corresponding to poor atomization of the central liquid core. At 0.1 MPa the profile reveals a local minimum at the centerline and local maximums at the extreme edge of the spray.

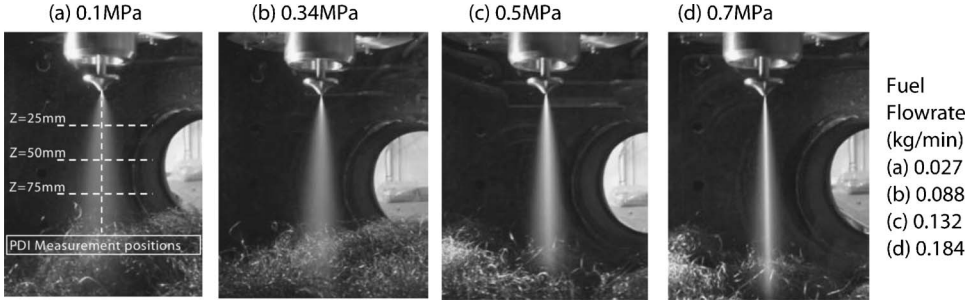


Fig. 19 Digital camera images at the exit of atomizer for Strategy 2 at different pressure (0.1–0.7 MPa) 300 K cases

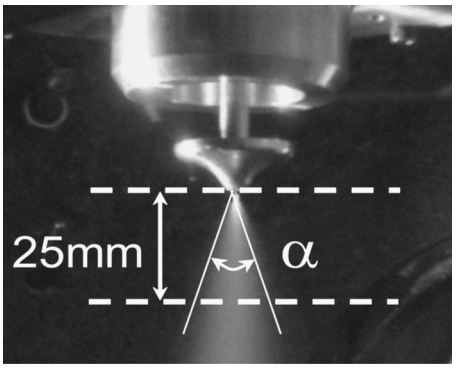


Fig. 20 Typical image of spray angle measurement technique

Conclusions

This study represents the first documented investigation between fuel preparation and emissions at actual engine conditions by using a practical injector in both a controlled, well instrumented, laboratory test rig and in a practical gas turbine engine. The performance of a fuel injector in preparing the fuel/air mixture for combustion was investigated in the laboratory at actual engine conditions using PDI and flow visualization. The emissions produced by a commercial microturbine generator operated with this same injector at the same condition studied in the laboratory. In the laboratory, two strategies for scaling the atomization from 0.1 MPa to 0.9 MPa were evaluated. One strategy was to match ρ AUR2 and ALR, and the second strategy was to keep the momentum ratio of atomizing air and screen air constant. Conclusions drawn from the study are as follows:

For the range of pressures and temperatures studied, the spray angle for the air blasted plain jet injector used correlates well with fuel to atomizing air momentum ratio, a result that will prove useful for combustor and injector designers that are challenged with careful consideration of fuel preparation.

At engine combustor inlet conditions, the combustion air, which consists of mixing air and swirler air, significantly promotes the evaporation process compared to the atomizing air alone. As a result, evaporation rates based primarily upon the atomization process will not provide correct results.

Observations in the laboratory that would normally be expected to lead to a reduction in NO emission, produced, in fact, a higher emission of NO in the practical engine. For example, higher ALR was found in the laboratory injector studies to markedly reduce droplet size and increase vaporization rates with the expectation of improved homogeneity of the resultant fuel/air mixture, but to produce higher NO emission from the practical engine. This implies that a design strategy focused on improving atomization of

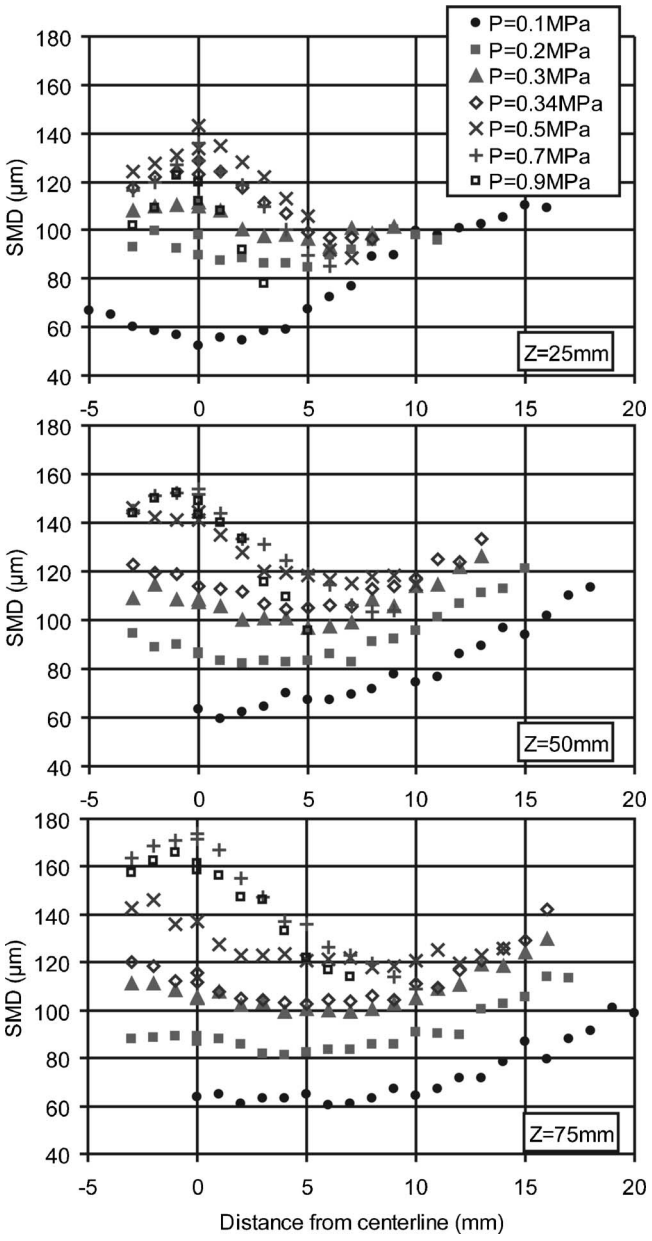


Fig. 22 Radial profiles of D_{32} for Strategy 1

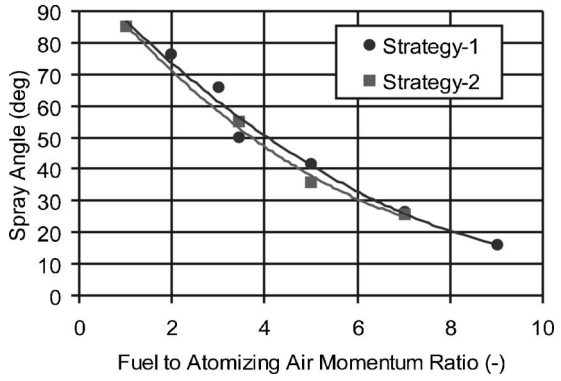


Fig. 21 Spray angle measured downstream of the atomizer as a function of fuel to air momentum ratio

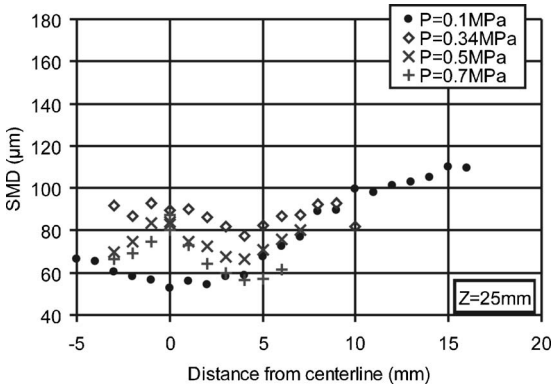


Fig. 23 Radial profile of D_{32} for Strategy 2

the spray such that no droplets exit the premixer is not sufficient to infer emissions performance. Indeed, in the present case, emissions were minimized at a condition that was observed to result in droplets exiting the premixer. It is inferred that, in advanced injector designs, the emissions are likely dictated by the fuel/air mixture properties (e.g., temporal and spatial distribution of homogeneity) prior to and within the primary zone of stabilization.

Acknowledgment

The authors acknowledge the support of the California Energy Commission (Contract No. 500-00-020), Mitsubishi Heavy Industries, and Capstone Turbine Corporation. The authors would also like to acknowledge the great contributions of the University of California, Irvine Combustion Laboratory (UCICL), staff and students including Mr. Richard Hack, Mr. Josh Mauzey, Mr. Christopher D. Bolszo, Mr. Brandon J. Masuda, Mr. Steven R. Hernandez, Mr. Nikhil K. Kar, Mr. Patrick M. Couch, and Mr. Peter L. Therkelsen for assistance with the preparation and operation of the test facility.

Nomenclature

D_{32} = Sauter mean diameter, m
d_o = liquid discharge orifice diameter, m
σ = surface tension, kg/s^2
U_R = relative velocity (coflowing)
ρ_A = density of air, kg/m^3
ALR = atomizing air to liquid mass flow ratio
μ_L = liquid viscosity, kg/m s
ρ_L = liquid density, kg/m^3
L_{premixer} = premixer length, m
Re = Reynolds number
M = Mach number
St = Stokes number
LPP = lean premixed prevaporized
ppmvd = parts per million by volume on a dry basis

References

[1] Lefebvre, A. H., 1999, *Gas Turbine Combustion*, 2nd ed., Taylor & Francis, Philadelphia, PA.
[2] Richards, G. A., McMillian, M. M., Gemmen, R. S., Rogers, W. A., and Cully, S. R., 2001, "Issues for Low-Emissions, Fuel Flexible Power Systems," Prog. Energy Combust. Sci., **27**, pp. 141–169.
[3] Chiu, H. H., and Liu, T. M., 1977, "Group Combustion of Liquid Droplets," Combust. Sci. Technol., **17**, pp. 127–142.
[4] Lefebvre, A. H., 1995, "The Role of Fuel Preparation in Low-Emission Combustion," ASME J. Eng. Gas Turbines Power, **117**, pp. 617–654.
[5] Behrendt, T., Heinze, J., and Hassa, C., 2003, "Experimental Investigation of a New LPP Injector Concept for Aero Engines at Elevated Pressures," *Proceedings Turbo Expo 2003*, Atlanta, GA.
[6] Zaralis, N., Ripplingera, T., Hohmann, S., Hettelb, M., Merkleb, K., Leuckelb, W., Klosec, G., Meierc, R., Kochc, R., Wittigc, S., Carld, M., Behrendtd, T., Hassad, C., Meiere, U., Lückerathe, R., and Strickere, W., 2002, "Low NO_x Combustor Development Pursued Within the Scope of Engine 3E German National Research Program in a Cooperative Effort Among Engine Manufacturer MTU University of Karlsruhe and DLR German Aerospace Research Center," Aerosp. Sci. Technol., **6**(7), pp. 531–544.
[7] Becker, J., and Hassa, C., 2003, "Liquid Fuel Placement and Mixing of Generic Aeroengine Premix Module at Different Operating Conditions," ASME J. Eng. Gas Turbines Power, **125**, pp. 901–908.
[8] Rachner, M., Brandt, H. E., and Hassa, C., 1996, "A Numerical and Experimental Study of Fuel Evaporation and Mixing for Lean Premixed Combustion at High Pressure," *Proceedings 25th Symposium (International) on Combustion*, Vol. 2, pp. 2741–2748.
[9] Lefebvre, A. H., 1998, *Atom and Sprays*, Hemisphere, New York, NY.
[10] Georjon, T. L., and Reitz, R. D., 1999, Atomization Sprays, **9**, pp. 231–254.
[11] Jermy, M. C., Hussain, M., and Greenhaigh, D. A., 2003, "Operating Liquid-Fuel Airblast Injectors in Low-Pressure Test Rigs: Strategies for Scaling Down the Flow Conditions," Meas. Sci. Technol., **14**, pp. 1151–1158.
[12] Sayre, A. N. Dugue, J., Weber, R., Domnick, J., and Lindenthal, A., 1994, "Characterization of Semi-Industrial-Scale Fuel-oil Spray Issued from a Y-Jet Atomizer," J. Inst. Energy, **67**, pp. 70–77.
[13] Mcvey, J. B., Kennedy, J. B., and Russell, S., 1989, "Application of Advanced Diagnostics to Airblast Injector Flows, J. Eng. Gas Turbines Power, **111**, pp. 53–62.
[14] Lefebvre, A. H., 1999, *Gas Turbine Combustion*, Taylor and Francis, Philadelphia, PA.
[15] Wang, H. Y., McDonell, V. G., Sowa, W. A., and Samuelsen, G. S., 1993, "Scaling of the Two-Phase Flow Downstream of a Gas Turbine Combustor Swirl Cup: Mean Quantities," ASME J. Eng. Gas Turbines Power, **115**, pp. 453–460.
[16] Leong, M. Y., McDonell, V. G., and Samuelsen, G. S., 2001, "Effect of Ambient Pressure on an Airblast Spray Injected Into a Crossflow," J. Propul. Power, **17**(5), pp. 1076–1084.
[17] Leong, M. Y., Smugeresky, C. S., McDonell, V. G., and Samuelsen, G. S., 2001, "Rapid Liquid Fuel Mixing for Lean Burning Combustors: Low Power Performance," ASME J. Eng. Gas Turbines Power, **123**, pp. 574–579.
[18] McDonell, V. G., Seay, J. E., and Samuelsen, G. S., 1994, "Characterization of the Non-Reacting Two-Phase Flow Downstream of an Aero-Engine Combustor Dome Operating At Realistic Conditions," ASME IGTI Meeting, The Hague, Netherlands, June, Paper No. 94-GT-263.
[19] Glassman, I., 1996, *Combustion*, 3rd. ed., Academic, New York.
[20] Leonard, G., and Stegmaier, J., 1994, "Development of an Aeroderivative Gas Turbine Dry Low Emission Combustion System," ASME J. Eng. Gas Turbines Power, **116**, pp. 542–546.
[21] Lee, J. C.Y., Malte, P. C., and Benjamin, M. A., 2001, "Low NO_x Combustion for Liquid Fuels: Atmospheric Pressure Experiments Using a Staged Prevoporizer-Premixer," ASME Paper No. 01-GT-0081.
[22] Phi, V. M., Mauzey, J. L., McDonell, V. G., and Samuelsen, G. S., 2004, "Fuel Injection and Emissions Characteristics of a Commercial Microturbine Generator," ASME Paper No. GT-2004-54039.
[23] Edwards, C. F., and Marx, K. D., 1992, "Analysis of the Ideal Phase-Doppler System," Atomization Sprays, **2**, pp. 319–366.
[24] Bolzso, C. D., Mauzey, J. L., McDonell, V G., and Nakamura, S., 2005, "Experimental Investigation of Liquid Fuel Atomization, Mixing and Pollutant Emissions for a 30 kW Gas Turbine Engine," *Proceedings, 18th ILASS-Americas Conference*, Irvine, CA, May.
[25] Lorenzetto, G. E., and Lefebvre, A. H., 1977, "Meausrments of Drop Size on a Plain Jet Airblast Atomizer," AIAA J., **15**(7), pp. 1006–1010.
[26] Benjamin, M. J., McDonell, V. G., and Samuelsen, G. S., 1997, "Effect of Fue/Air Ratio on Air Blast Simplex Nozzle Performance," ASME Paper No. 87-GT-150.

A Hybrid Prognostic Model Formulation and Health Estimation of Auxiliary Power Units

Pradeep Shetty
Honeywell Technology Solutions Laboratory (HTSL),
Bangalore 560076, India
e-mail: pradeep.shetty@honeywell.com

Dinkar Mylaraswamy
Honeywell Laboratories,
Minneapolis, MN 55448
e-mail: dmylaras@htc.honeywell.com

Thirumaran Ekambaram
Honeywell Technology Solutions Laboratory (HTSL),
Bangalore 560076, India
e-mail: thirumaran.ekambaram@honeywell.com

Prognostic health monitoring is an important element of condition-based maintenance and logistics support. The accuracy of prediction and the associated confidence in prediction greatly influence overall performance and subsequent actions either for maintenance or logistics support. Accuracy of prognosis is directly dependent on how closely one can capture the system and component interactions. Traditionally, such models assume a constant and univariate prognostic formulation—that is, components degrade at a constant rate and are independent of each other. Our objective in this paper is to model the degrading system as a collection of prognostic states (health vectors) that evolve continuously over time. The proposed model includes an age dependent deterioration distribution, component interactions, as well as effects of discrete events arising from line maintenance actions and/or abrupt faults. Mathematically, the proposed model can be summarized as a continuously evolving dynamic model, driven by non-Gaussian input and switches according to the discrete events in the system. We develop this model for aircraft auxiliary power units, but it can be generalized to other progressive deteriorating systems. The system identification and recursive state estimation scheme for the developed non-Gaussian model under a partially specified distribution framework has been deduced. The diagnostic/prognostic capabilities of our model and algorithms have been demonstrated using simulated and field data. [DOI: 10.1115/1.2795761]

1 Introduction

An auxiliary power unit (APU) is a small turbo engine used primarily for starting the propulsion engines, provide bleed air for the environmental control system, and meet various electrical loads in an aircraft while on the ground. An APU is considered inoperable if it fails to meet any of the listed functions. In addition, an APU can be rendered inoperable because it cannot be started. Such inoperability can be caused by faults within the APU and/or improper operating conditions. The primary objective of a prognostic system is to predict with a high degree of confidence when such inoperability may occur. In this document, the APU is modeled as a collection of prognostic states, and these states are then estimated using the observed data. By monitoring the dynamic evolution of these states, one can *predict* consequences and, hence, influence intelligent decision making. The more accurately one can describe this prognostic state, the higher will be the prognostic confidence level, and uncertainty about maintenance or corrective action will be less.

Our decision to model the APU as a collection of prognostic states evolving in a hybrid state space is based on the work done by Khalak and Hess [1] under the Joint Strike Fighter program and some initial work done by our group in Honeywell Laboratories [2]. Underlying this formulation is the notion of a health state, denoted by $\mathbf{x}$. In general, $\mathbf{x} \in \Re^n$ indicates the multifaceted component of the health vector. For example x_1, x_2 could represent the health of the power and load sections of an APU. Irrespective of its dimensionality, $\mathbf{x}$ must satisfy the following two properties to ensure physical significance.

1. There exists a vector $\boldsymbol{\delta}$, such that

$$P(\text{APU in operation} \,|\mathbf{x}_i \leqslant \delta_i) \to 1 \quad \forall\, i = 1,2,\ldots,n \tag{1}$$

where $P(b|a)$ indicates the probability of event b given that the event a has occurred. This condition provides significance to the health state with respect to its consequence. Whenever the state vector falls below a predefined threshold, the probability of an APU failing to operate is very high. In order to meet this condition, we retain only those important states that can render the APU inoperable.

2. There exists a subset ϕ_i of failure modes such that

$$P(\phi_i|x_i \leqslant \delta_i) \to 1 \quad \forall\, i = 1,2,\ldots,n \tag{2}$$

This condition provides significance to the health state with respect to underlying root causes. Whenever the state vector falls below a predefined threshold, the probability of a failure mode is very high. This is an important condition for fault localization.

Several factors contribute to the evolution of this health vector:

1. intrinsic damage accumulation or aging of the components
2. interactions between components
3. deviations from design operating envelopes
4. influence of discrete events resulting from abrupt faults or line maintenance actions

As stated earlier, the feasibility of prognosis depends on how accurately one captures the above listed factors. Simplifying assumptions like constant deterioration, noninteracting components increase the uncertainty in prediction. To a large extent, if human intervention is possible, this uncertainty is managed by a customer service engineer. As the decision making process becomes more complex, the need for automation increases [1]. To this end, we propose to formulate a hybrid state-space model for prognostics. This formulation is described in Sec. 2.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 9, 2005; final manuscript received August 3, 2007; published online January 22, 2008. Review conducted by William Rhoden.

As the model becomes complex, one is faced with the problem of extracting values of parameters based on a series of observations made. The problem of parameter estimation goes hand in hand with model formulation. A complex model, albeit closer to reality, is impractical if we cannot estimate its parameters. The problem of estimator design involves both system identification and health vector estimation. We have developed a maximum likelihood (ML)-based system identification and recursive Bayesian health estimation scheme based on partially specified (PS) distributions, which is explained in Sec. 3. The outcome of the formulations and the estimation scheme is studied in detail and presented in Sec. 4.

2 Hybrid Model Formulation

For prognostics, we are interested in the rate of change of the health vector. As part of its normal operation, an APU goes through a series of cycles. Each *cycle* is defined as the period between a start and a shutdown. Within each cycle, the APU operates through a series of discrete modes. Let t denote the cycle, and $p \in \mathcal{M}$ denote the finite set of operating modes. Let there be m modes of operation of APU.

The rate of change can be calculated within each cycle as well as from cycle to cycle. Clearly, there is a dependency. If $\mathbf{x}(t)$ denotes the state vector at the end of the tth cycle, the rate of change of $\mathbf{x}(t)$ depends on how the APU behaved at each mode in this cycle and the state of the APU at the beginning of the tth cycle. In addition, the rate of change will also be influenced by the intrinsic damage accumulation due to aging that occurred in the tth cycle as well as any discrete event that occurred during the tth cycle.

Mathematically, this is expressed as

$$\frac{d\mathbf{x}(t)}{dt} = \boldsymbol{\mathcal{F}}(\mathbf{x}(t), \mathbf{w}^1(t), \mathbf{w}^2(t), \ldots, \mathbf{w}^k(t), \theta(t), \mathbf{v}(t)) \tag{3}$$

$\boldsymbol{\mathcal{F}}$ is a nonlinear function vector. $\theta \in \Re^n$ denotes the intrinsic damage accumulation within the tth cycle. $\mathbf{v} \in N^r$ denotes discrete events within the tth cycle.[1] $\mathbf{w}^k(t)$ denotes the state at the kth mode within the tth cycle.

The state within each cycle, $\mathbf{w}(t)$, evolves as follows:

$$\frac{d\mathbf{w}}{d\tau} = \mathcal{H}(\mathbf{w}(\tau), \mathbf{u}(\tau)) \quad 0 \leqslant \tau \leqslant \Delta_t \tag{4}$$

where $\mathbf{u} \in \Re^p$ is an input during the tth cycle and Δ_t is the duration of the tth cycle.

The straightforward relationship between $\mathbf{x}$ and $\mathbf{w}$ is given below,

$$\mathbf{w}(\tau = 0) = \mathbf{x}(t-1)$$

$$\mathbf{w}(\tau = \Delta_t) = \mathbf{x}(t) \tag{5}$$

Discrete events impacting the prognostic health state can result from line maintenance actions and/or abrupt faults within the system. A fault is defined as *abrupt* if the period between the initiation and manifestation of an abrupt fault is an order of magnitude smaller than $\min \Delta_t$. In simple terms, the rate at which this fault starts and manifests itself is much, much smaller than the duration of the cycle. A *progressive* fault, on the other hand, evolves over several cycles. That is, the period between initiation and manifestation of a progressive fault is much greater than the duration of the cycle. In this respect, one can consider $\mathbf{x}$ as modeling progressive faults. The variable $\mathbf{v}$ is typically an *input* to the prognostic model. The numerical value of $\mathbf{v}$ at any given cycle t can be observed directly or calculated.

To summarize the formulation thus far, there are three main contributors to the rate at which $\mathbf{x}$ evolves from one cycle to another. These are (a) the behavior of the APU at each operating mode $(\mathbf{w}^1, \mathbf{w}^2, \mathbf{w}^3, \ldots)$, (b) the damage accumulation within each cycle (θ) and (c) the discrete interactions within each cycle $(\mathbf{v})$. At this point, we make the following simplifying assumptions.

1. We will assume that the factors that influence the health state are independent and additive to simplify Eq. (3) as follows:

$$\dot{\mathbf{x}}(t) = \mathcal{F}_a(\mathbf{x}(t)) + \sum_{k=1}^{m} \mathcal{F}_b^k(\mathbf{w}^k(t)) + \mathcal{F}_c(\theta(t)) + \mathcal{F}_d(\mathbf{v}(t)) \tag{6}$$

   where the *dot* notation denotes differentiation with respect to flight cycle time t.
2. We assume a quasisteady state behavior within each cycle, similar to that described in Ref. [1]. Introducing this assumption in Eqs. (4) and (5), one can replace $\mathbf{w}^k$ with $\mathbf{u}^k$ for all values of t. This simplifies the prognostic model as follows:

$$\dot{\mathbf{x}}(t) = \mathcal{F}_a(\mathbf{x}(t)) + \sum_{k=1}^{m} \mathcal{F}_b^k(\mathbf{u}^k(t)) + \mathcal{F}_c(\theta(t)) + \mathcal{F}_d(\mathbf{v}(t)) \tag{7}$$

3. We assume that the discrete events that influence the health state are binary and uncorrelated. That is,

$$v_j(t) = \begin{cases} 1 & j\text{th discrete event was present within the } t\text{th cycle} \\ 0 & j\text{th discrete event was absent within the } t\text{th cycle} \end{cases} \tag{8}$$

The uncorrelated condition is expressed best in a probabilistic sense. That is, given the value of v_j at cycle $(t-1)$ sets up no expectation for its value at cycle t. In other words,

$$P(v_j(t) | v_j(t-1)) = P(v_j(t)) \quad j = 1, 2, \ldots, r \tag{9}$$

The next step involves defining the observation vector. A first-order hidden Markov observation model is assumed for the state vector. That is,

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon} \tag{10}$$

In simple terms, Eq. (10) assumes that one can observe noisy values of the health state. It makes no assumption about *how* these observations are made. The observation vector $\mathbf{y}$ could be directly measured or calculated using measurements (transformation of measurements). It therefore suffices to assume that a series of measurements, $\mathbf{z}(t), \mathbf{z}(t-1), \ldots$, are available and a transformation $\mathcal{G}$ provides $\mathbf{y}(t)$ from measurements made during the tth cycle,

$$\mathbf{y}(t) = \mathcal{G}(\mathbf{z}^1(t), \mathbf{z}^2(t), \mathbf{z}^3(t), \ldots) \tag{11}$$

where $\mathbf{z}^k(t)$ denotes the measurements made at the kth mode within the tth cycle. The measurement vector $\mathbf{z}^k(t)$ includes inputs as well. That is,

$$\text{inputs } \mathbf{u}^k(t) \subset \mathbf{z}^k(t) \text{ measurement vector} \tag{12}$$

In nonmathematical terms, our prognostic model can be expressed as follows: The prognostic state at each cycle t evolves as a function of operating conditions, intrinsic aging, and discrete maintenance actions/faults that occurred within this cycle; noisy values of this state are given to us as observations.

At this point, we make several simplifying assumptions, primarily driven by a practical estimator constraint.

1. The health state evolves as per linear dynamics. Using this assumption, our basic model for APU prognostics is given by

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \sum_{k=1}^{m} \mathbf{B}^k \mathbf{u}^k(t) + \text{diag}(\boldsymbol{\beta})\theta(t) + \mathbf{C}\mathbf{v}(t)$$

[1] N stands for a set of natural numbers.

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\epsilon} \tag{13}$$

Here, $\mathbf{A}$ is an $n \times n$ matrix, $\mathbf{B}^k$ is an $n \times p$ matrix, diag($\boldsymbol{\beta}$) is an $n \times n$ diagonal matrix, $\mathbf{C}$ is an $n \times r$ matrix.

2. Intrinsic deterioration is modeled as a Poisson process with constant properties. This simplification was motivated by the Palmgren-Miner equations,

$$P(\theta = t) = \lambda e^{-\lambda}/t!$$

$$\text{diag}(\beta) = \text{const} \tag{14}$$

3. We assume that the contributions due to deviations at each operating mode within the tth cycle are equal. Further, each contribution is proportional to the deviation of the kth operating mode with respect to design conditions. That is,

$$\mathbf{B}^k = \mathbf{B} \quad \text{and} \sum_{k=1}^{m} \mathbf{B}^k \mathbf{u}^k = \mathbf{B} \sum_{k=1}^{m} (\mathbf{u}^k - \mathbf{u}^{k,0}) \tag{15}$$

where $\mathbf{u}^{k,0}$ is the input to the APU operating at the kth mode under design conditions.

4. The observation noise $\boldsymbol{\epsilon}$ is modeled as a zero mean Gaussian, with constant variance/covariance. Hence,

$$E[\boldsymbol{\epsilon}] = \mathbf{0} \quad E[\boldsymbol{\epsilon}^T \quad \boldsymbol{\epsilon}] = \mathbf{R} \tag{16}$$

The following equations summarize the proposed prognostic model as well as the problem statement.

Proposed model:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B} \sum_{k=1}^{m} (\mathbf{u}^k(t) - \mathbf{u}^{k,0}(t)) + \text{diag}(\boldsymbol{\beta}_t)\theta(t) + \mathbf{C}\mathbf{v}(t)$$

$$\mathbf{y}(t) = \mathbf{x}(t) + \boldsymbol{\epsilon}(t)$$

$$\theta \sim P(\lambda) \quad \boldsymbol{\epsilon} \sim N(0,R) \tag{17}$$

Problem statement:

$$\text{Given: } \{\mathbf{y}(t), \mathbf{u}^k(t), \mathbf{v}(t)\}, \{\mathbf{y}(t-1), \mathbf{u}^k(t-1), \mathbf{v}(t-1)\}, \ldots$$

$$\text{Settings: } \lambda, R, \beta_t, \mathbf{u}^{k,0}$$

$$\text{Estimate: } \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{x}(t), \mathbf{x}(t + \delta_t) \tag{18}$$

Note that $(t-1)$ represents the previous cycle, $(t-2)$ represents the cycles in the past, and $\mathbf{x}(t+\delta_t)$ denotes the prediction of the health vector with a prediction window of δ_t, given $\{\mathbf{y}(t), \mathbf{u}^k(t), \mathbf{v}(t)\}$. The period between two cycles can be nonuniform.

3 Estimation of Auxiliary Power Unit Health Vector

In this section, we explain our methodology for identifying the system matrices $(\mathbf{A}, \mathbf{B})$ and estimating the health vectors $\mathbf{x}(t)$ based on continuous-time observations. Due to the use of digital techniques to sample and store the continuous-time wave form, we have an equivalent problem of extracting the desired parameter values from a discrete-time wave form or data set. Mathematically, the n-point data set $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ depends on an unknown parameter θ. We wish to determine θ based on the data, which is done using an *estimator*, defined as

$$\hat{\theta} = g(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n) \tag{19}$$

where g is a function defined in $\mathbf{L}^2$ space. This is the problem of estimation. Note that $\mathbf{y}_k$ denotes the discrete value of observation at time index k. The choice of an estimator that will perform well for a particular application depends on many considerations. Of primary concern is the selection of a good data model. The model should be complex enough to describe the principal features of the data, but at the same time simple enough to allow an estimator that is optimal and easily implementable [3]. The major issues in designing an estimator are determination of the existence of an efficient estimator, implementation feasibility, and computational cost [4]. The system under consideration is hybrid; it considers both the continuous evolution of the states and the discrete jumps in the states. Although the model is linear, dealing with the jumps is not straightforward. Added to this, the health state is driven by a *Poisson process*: hence, the mathematical formulation of the estimator is not easy and the standard solutions (which are formulated mostly for Gaussian processes) are not available [5]. In the next section, we will explain a ML-based estimator for system identification and recursive Bayesian state estimation in a PS distribution framework [6].

3.1 Maximum Likelihood Identification and State Estimation Using Partially Specified Distributions.

The discretized state and observation equation is given by

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B} \sum_{k=1}^{m} \mathbf{U}_t^k + \mathbf{C}\mathbf{v}_t + \text{diag}(\beta_t)\theta_t$$

$$\mathbf{y}_t = \mathbf{x}_t + \boldsymbol{\epsilon}_t \tag{20}$$

$\mathbf{U}_t^k$ is the $(\mathbf{u}^k(t) - \mathbf{u}^{k,0}(t))$. In our model, the random variable θ follows a Poisson distribution with parameter λ. Measurement noise $\boldsymbol{\epsilon}$ follows a normal distribution. That is,

$$\theta \sim P(\lambda) \quad \boldsymbol{\epsilon} \sim N(0,R) \tag{21}$$

β_{ii} is a constant scaling factor for the random variable θ, which relates the intrinsic deterioration to the health vector. Hence, $[\text{diag}(\beta_t)\theta_t]$ can be considered another random variable ω. The duration of the flight cycle is assumed to be constant.

In this approach, we make simplifying assumptions about the underlying true probability density function (PDF) using PS distributions. That is, we are not interested in the exact form of the distribution, but in its moments. Distribution of any random variable p can be approximated with another distribution q, such that

$$\mu_1'(p) = \mu_1'(q) \text{ and } \mu_2'(p) = \mu_2'(q) \tag{22}$$

where μ_n' denotes the nth raw moment of a distribution. Note that this does not impose any constraints on $\mu_n', n \geq 3$. Such approximations are called PS distributions [7] through the first two moments. In our case, ω, which follows a Poisson distribution, will be PS. This implies, $\omega \sim PS(\nu, Q)$. From system knowledge, we deduce these parameters as

$$\nu_i = \beta_{ii}\lambda_i, \quad Q = \text{diag}(\beta_{ii}^2 \lambda) \tag{23}$$

The measurement noise is $\boldsymbol{\epsilon}$ and the probability of damage accumulation is ω; both are modeled as *independent* random variables. In other words,

$$E[\boldsymbol{\epsilon}\omega] = E[\omega\boldsymbol{\epsilon}] = 0 \tag{24}$$

In nonmathematical terms, this equation implies that the noise in measuring the state vector (introduced due to sensor, electrical interference, etc.) does not depend on the probability of damage accumulation. These two are independent of each other. This assumption is necessary to obtain the estimator in closed form.

Once we make this simplifying assumption, the simplified prognostic model is

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B} \sum_{k=1}^{m} \mathbf{U}_t^k + \mathbf{C}\mathbf{v}_t + \omega_t \quad \omega \sim \text{PS}(\nu, Q)$$

$$\mathbf{y}_t = \mathbf{x}_t + \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon} \sim N(0,R) \tag{25}$$

Define the observation vector $\mathbf{Y}_m = [\mathbf{y}_1, \mathbf{y}_t, \ldots, \mathbf{y}_m]$ as a collection of all observations made until time m. Define $\mathbf{x}_n^m = E[\mathbf{x}_n | \mathbf{Y}_m]$, which is the expected value of the state vector at time n, given the

measurements until time m.

Given the estimate of previous time instant $\mathbf{x}_{t-1}^{t-1}$, and P_{t-1}^{t-1} (which, in the case of $t=1$, are initial conditions), the following recursions can be established:

$$E[\mathbf{x}_t|\mathbf{Y}_{t-1}] = E\left[\mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\sum_{k=1}^{m}\mathbf{U}_t^k + \mathbf{C}\mathbf{v}_t + \omega_t|\mathbf{Y}_{t-1}\right]$$

$$\mathbf{x}_t^{t-1} = \mathbf{A}\mathbf{x}_{t-1}^{t-1} + \mathbf{B}\sum_{k=1}^{m}\mathbf{U}_t^k + \mathbf{C}\mathbf{v}_t + \nu \tag{26}$$

Equation (26) expresses the expected value of the state at the tth time interval, given the estimate of the state at the previous time step and the inputs exciting the system at the tth time step. Note that the input $\mathbf{U}_t^k$ is assumed to be available at the state prediction stage. Also, we assume that the discrete event occurrence $\mathbf{v}_t$ is detected using independent observers and supplied in the prediction stage. If these inputs are not available, we rely on the information gathered in the previous flight cycle. Being constants, the expected values turn out to be the same: hence, $\mathbf{x}_t^{t-1}$ can be deduced. The discrete event has been incorporated in the first moment of the health vector. If an event occurs, it will switch the state, which will be estimated accordingly.

The prediction for the state covariance matrix at time t can be defined as

$$P_t^{t-1} = E[(\mathbf{x}_t - E[\mathbf{x}_t|\mathbf{Y}_{t-1}])(\mathbf{x}_t - E[\mathbf{x}_t|\mathbf{Y}_{t-1}])'] \tag{27}$$

P_t^t is a measure of the error between the *actual* value of the state ($\mathbf{x}$) and the expected value of the state based on all observations collected up to that point. Using the uncorrelated properties of ω, this is deduced to be

$$P_t^{t-1} = \mathbf{A}P_{t-1}^{t-1}\mathbf{A}' + Q + \nu\nu' \tag{28}$$

With the new observation $\mathbf{y}_t$, we can obtain the innovation as [5]

$$e_t = \mathbf{y}_t - E[\mathbf{y}_t|\mathbf{Y}_{t-1}] = \mathbf{y}_t - \mathbf{x}_t^{t-1}$$

It follows from the definition of e_t that the expected value of e_t is zero. The variance of the innovation is given by

$$E[e_t e_t|\mathbf{Y}_{t-1}] = E[(\mathbf{y}_t - \mathbf{x}_t^{t-1})|(\mathbf{y}_t - \mathbf{x}_t^{t-1})'] = P_t^{t-1} + R$$

Similarly, the covariance between the state and the innovation can be calculated as follows:

$$\mathrm{Cov}(\mathbf{x}_t, e_t|\mathbf{Y}_{t-1}) = E[(\mathbf{x}_t - E[\mathbf{x}_t|\mathbf{Y}_{t-1}])(e_t - E[e_t|\mathbf{Y}_{t-1}])'] = P_t^{t-1}$$

At this point, we can conclude that the joint distribution of the state variable $\mathbf{x}_t|Y_{t-1}$ and the innovation vector e_t at any given point t, can be PS using measurements $\mathbf{Y}_{t-1}$ as follows:

$$\begin{bmatrix}\mathbf{x}_t|Y_{t-1}\\ e_t\end{bmatrix} = \mathrm{PS}\left(\begin{bmatrix}\mathbf{x}_t^{t-1}\\ 0\end{bmatrix} \quad \begin{bmatrix}P_t^{t-1} & P_t^{t-1}\\ P_t^{t-1} & P_t^{t-1}+R\end{bmatrix}\right) \tag{29}$$

Note that the observation sequence is Gaussian; therefore, we can easily assume the joint PDF of the state and innovation to be PS in the first two moments. Using the conditional independence property of $\mathbf{x}_t|Y_{t-1}$ and e_t, under a PS framework [6], we get

$$\mathbf{x}_t^t = E[\mathbf{x}_t\mathbf{Y}_{t-1}, e_t] = \mathbf{x}_t^{t-1} + \mathbf{K}_t e_t \tag{30}$$

The most common criterion for $\mathbf{K}_t$, the Kalman gain at the ith time step, is to minimize the weighted sum of the diagonal elements of the state covariance matrix P_t^t, which is deduced as

$$\mathbf{K}_t = P_t^{t-1}[P_t^{t-1} + R]^{-1} \tag{31}$$

$$P_t^t = [\mathbf{I} - \mathbf{K}_t]P_t^{t-1} \tag{32}$$

Thus, we have found a recursive way to obtain the prior and posterior PDF of the health vector under a PS distribution framework. Having found the recursive state estimation the identification of system matrices is straightforward under the ML framework. In our methodology, we attempt to maximize the likelihood of the innovation sequence as given in Ref. [5].

3.2 Algorithm for System Identification. The algorithm for obtaining the system matrices and health vectors having a batch of data $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_n$ is summarized below.

1. Initialize the system matrices $\mathbf{A}, \mathbf{B}$ (We incorporate least squares regression (LSR) based methodology as explained in the Appendixes. Assume initial values for $\mathbf{x}_0^0$ and P_0^0.
2. For $t>0$, predict the first two moments of the state variable using Eqs. (26) and (28).
3. Observe the process at each time instant and obtain the innovation.
4. Update the state estimate and the state covariance estimate using Eqs. (30)–(32).
5. Repeat steps 2–4 for $t=1,2,\ldots,n$. Collect all the innovation sequence $e_1, e_2, \ldots, e_n$.
6. Assume that the innovations come from a multivariate Gaussian distribution, the mean of which is zero and the covariance is given in Eq. (29). Obtain the log likelihood function for the innovation sequence, which is

$$L = \sum_{t=1}^{n}\log|P_t^{t-1} + R| + \sum_{t=1}^{n} e_t'(P_t^{t-1} + R)^{-1}e_t \tag{33}$$

7. Obtain the system matrices using

$$[\hat{\mathbf{A}}, \hat{\mathbf{B}}] = \min_{\mathbf{A},\mathbf{B}} L \tag{34}$$

After obtaining the system matrices, the health vectors can be estimated using Eqs. (26), (28), and (30)–(32).

4 Results and Discussion

4.1 Simulated Model Results. We demonstrated the efficacy of the state estimation and system identification scheme using both simulated and field data. The data were simulated considering all aspects of the random process. A striking advantage of the simulated experiments is the availability of the true value of the parameters, which can be used for analyzing the estimator efficiency [4]. We used a simple three-state simulation model for generating the data, giving a 3×3 system matrix, which provides the necessary complexity without incurring computational penalties [8]. The degradation effect due to the deviation of the measurements from the designed value is ignored. That is, the APU is assumed to operate under designed conditions; hence, $(\mathbf{u}_t^k - \mathbf{u}_t^{k,0}) = 0$. We introduced the discrete event $\mathbf{v}_t$ to the process at a known point of time. Realistic values for the state covariance (Q) and observation covariance (R) were assumed, and a data set was generated. The general steps for the generation of the data have been given in the Appendixes.

Using the described method, let $\mathbf{A}$ be the true value of the system matrix and $\hat{\mathbf{A}}$ be the estimate of this matrix. Then, define the error matrix $\zeta = \mathbf{A} - \hat{\mathbf{A}}$. Traditionally, the norm of this error matrix is used as a measure of estimator accuracy. Commonly used norms are 1-norm, 2-norm, Frobenius norm, and infinity norm.

We incorporate the Monte Carlo (MC) methodology for understanding the estimator efficacy in identifying the system. In the MC method, we need to generate the stochastic process for a predetermined number of times (generally taken as 5000 trials), performing an estimation at each step. Thus, we obtain a series of error matrices $\zeta_1, \zeta_2, \ldots, \zeta_{5000}$, which is the difference between the actual and estimated values. We can obtain confidence bounds for the error matrix and use them to evaluate the efficacy of the underlying estimation scheme. Although one can consider the spread of *any suitable matrix norm* as an indirect measure of identification accuracy, the 2-norm is the most commonly used. Thus, given

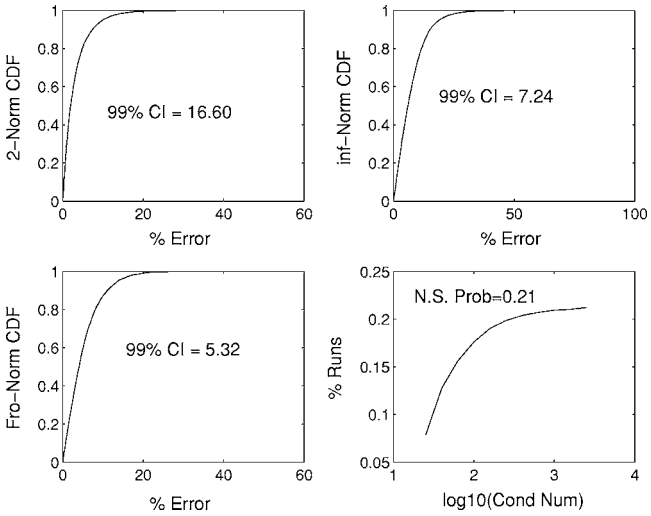


Fig. 1 Error bounds for estimated system matrix

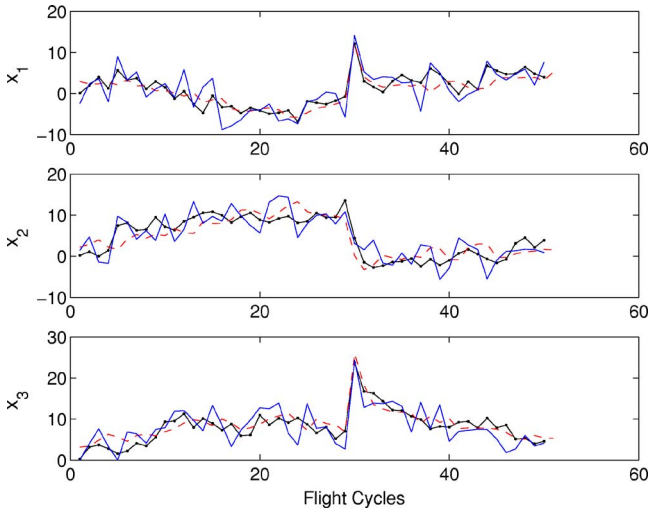


Fig. 2 Estimation of health vector x for simulated data

a series of ζ_i, we can calculate the mean and variance from MC realizations. We can then provide error bounds for our estimator accuracy at the desired level of confidence. For example, if $\mu_{\zeta_p}, \sigma_{\zeta_p}$ denotes the mean and the variance of the pth norm, then we can calculate 99% error bounds as

$$\|\mathbf{A} - \hat{\mathbf{A}}\|_p \leq \mu_{\zeta_p} + 1.96\sigma_{\zeta_p}$$

An estimator is considered better if it possesses minimum variance[2] with less computational complexity. The condition number is another commonly used metric for assessing the numerical stability of the estimator. The condition number is the ratio of the largest to the smallest singular value of ζ. The efficiency of the estimator in identifying the system parameters using above measures under the MC framework of 5000 trials is shown in Fig. 1, wherein the matrix-2-norm, the infinity norm, the Frobenius norm, and the condition number is displayed. Note that these parameters are deduced for system matrix ($\mathbf{A}$). As indicated by the first subplot with 99% confidence, we can conclude that the difference between the 2-norm of the *true* value $\mathbf{A}$ and its estimate will be less than or equal to 16.6. Stated differently, if we estimate the system matrix as $\hat{\mathbf{A}}$, then

$$\|\mathbf{A}\|_2 \leq \|\hat{\mathbf{A}}\|_2 + 16.6 \quad 99\,\% \text{ confidence level} \tag{35}$$

In other words, the true matrix will be *contained* within a spheroid of radius 16.6. Similar numbers for the Frobenius and the infinity norms are shown in Fig. 1.

The last subplot is the condition number distribution. To understand the sensitivity of the method, we perturbed the $\mathbf{A}$ matrix with 2% Gaussian noise while generating the data. The maximum condition number for $\mathbf{A}$ was 10. We counted the number of $\hat{\mathbf{A}}$ such that $\text{cond}(\hat{\mathbf{A}}) > 20$. We found this number in 21% of the trials; that is, one of every five estimations could result in a system matrix with a high condition number. However, this does not impact the health estimation performance. The efficacy of the algorithm in estimating the health vectors with discrete events are shown in Fig. 2. In this figure, the solid lines represent the observed data and the dashed lines represent the estimated state and the true state is indicated by *dash-dotted* lines. We introduced the discrete events (during data generation) and attempted to estimate the states based on the observation. This results in a dc shift in the health vectors. In this work, we assume that the information about the event is supplied by external observers. Also, the event sensitivity $\mathbf{C}$ is assumed to be known. In the Appendixes, we briefly describe the procedure of obtaining $\mathbf{C}$. In Fig. 2, discrete events were introduced at time index 29. The magnitude is determined by the event sensitivity ($\mathbf{C}$) to the state vectors. As explained earlier, the discrete event could be an abrupt fault or a line maintenance action. The effect of an abrupt fault is to pull the health vector down; one can see a sharp negative jump in the health vectors. The opposite is true for line maintenance actions. Note that the time information of the scheduled maintenance action is always known. If the sensitivity is also known, the time can be incorporated in the model. However, the abrupt failures need to be detected using external observers (pattern recognition algorithms).

The estimated health vector was found to follow the true state closely; therefore, the performance of the algorithm in estimating/tracking health is reasonably good. The estimator was also effective in tracking the sudden jump of the state vector due to a discrete event at time stamp 29. Thus, one can conclude that the hybrid estimator was effective in tracking the non-Gaussian state vector as well as the discrete event in the state space.

4.2 Flight Trial Data Results. The APU unit is considered the combination of two broad subsystems: the hot section and the load section. To estimate the health of these two sections, we have a two-dimensional health vector of hot section health and load section health. As the model suggests, the observation vector $\mathbf{y}_t$ is also two dimensional. The observation vector, also called margins, is deduced from measurements (generally, in engineering units) obtained in the load and hot sections. Based on domain knowledge, we normalize the margins to suit the current probabilistic modeling.

Bleed pressure (P2) and exhaust gas temperature (T2) are input $\mathbf{u}$. Thus, the dimension of the $\mathbf{B}$ matrix is 2×2. We consider the event of turbine replacement for the health evaluation. The turbine replacement event is deterministic; its effect is to enhance health.

The process noise (Q) and the observation noise (R) are unknown and must be estimated using data. Observations (also called margins) were found to be nonstationary, with time-varying mean. Hence, we incorporate a detrending methodology for obtaining R (explained in the Appendixes). To obtain the Q matrix, we use the *reliability* information. The reliability tests and field information reveal that the APU has an average life (λ) of 5000 h. The average rate at which observations deteriorate are $\beta = [-1/2700; -0.05/100]$, for the hot and load sections,

[2] *Minimun* is a relative quantity. This indicates the minimum variance among the estimators under consideration. This should not be confused with a Crammer–Rao lower bound. If any estimator attains this bound, then it is the best estimator.

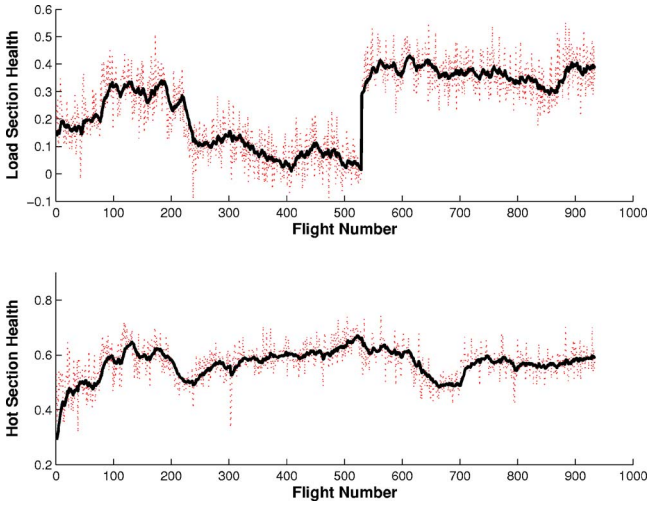


Fig. 3 Health estimation of the APU using a hybrid model

respectively.[3] The process noise can be computed as

$$Q = \begin{bmatrix} \beta_{\text{hot}}^2 \lambda & 0 \\ 0 & \beta_{\text{load}}^2 \lambda \end{bmatrix} = \begin{bmatrix} 0.0007 & 0 \\ 0 & 0.0013 \end{bmatrix}$$

Note that the process noise is assumed to be constant with respect to time.

The initialization of state vector $\mathbf{x}_0$ and associated variance P_0 is important in obtaining accurate health vectors. Because observations are the noisy version of the health vector, we initialize $\mathbf{x}_0 = 1$ to indicate that we are starting with a brand new APU. The variance P_0 is set to be equal to Q. The variance P stabilizes to a fixed value within few iterations of state estimation. Hence, an approximate value for this variable will suffice for our purpose.

With these settings, we perform the estimation task, wherein we identify the system with the first 100 data points. A sample estimate of the system parameters is given below,

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.9182 & -0.0500 \\ -0.1116 & 0.9800 \end{bmatrix}$$

$$\hat{\mathbf{B}} = \begin{bmatrix} -0.019 & 0.004 \\ -0.018 & 0.011 \end{bmatrix}$$

$$C = \begin{bmatrix} 0.32 \\ 0.06 \end{bmatrix}$$

The estimated parameters were found to be within the anticipated engineering limits. It is possible to map these estimates to the physical system. This is important, as the decisions based on these estimates directly contribute to the efficiency of the health management. The off diagonal elements of the system matrix ($\mathbf{A}$) capture the interaction of the subsystems (in our case, the hot and load sections). This means that aging of one subsystem is interdependent with the other subsystems. This information is important in obtaining efficient health estimates.

The performance of the algorithm in estimating the health of an in-service APU is shown in Fig. 3. The first subplot indicates the health of the load section and second subplot indicates the health of the hot section of an APU. The solid lines represent the estimated health vector and the dotted lines indicate the observation (or normalized margins). The estimated health tracks the data closely, which is in accordance with the model (refer to the observation model). The tracking performance of the algorithm seems to be fairly good.

In Eqs. (1) and (2), we defined the conditions for an abstract quantity $\mathbf{x}$ to indicate health. Domain knowledge reveals that if the threshold (δ_i, where $i=1,2$) reaches zero, then the useful life of the asset is zero. So, in Fig. 3, the useful life of an APU diminishes around flight cycle 530. During a scheduled maintenance, the turbine was replaced with a new one. This resulted in an upward jump of the health vector, indicating that the useful life has been enhanced.

System dynamics change with respect to time; hence, we need to regularly update the system parameters. To this end, we performed intermittent system identification and state estimation. The frequency of identification is determined by the field conditions. Ideally, we would like to identify the systems whenever the model changes. Such model changes can be inferred from the innovation sequences, which are ideally zero mean Gaussian processes. However, anomaly can be detected in the characteristics of the innovation sequence if the model varies. This may trigger the system identification. The state estimation results displayed in Fig. 3 are based on an algorithm that intermittently updates the system matrices. The innovation sequence for this case is given in Fig. 7 in the Appendixes.

The tracking performance of the health vectors were found to be reasonably good. A discrete event (turbine replacement) was detected at time stamp 530. In this scenario, the estimator was quick to adjust the health vectors.[4] Note that these tracking performances are important in diagnosing the asset. The inference on the health and the proposed maintenance actions depend on the magnitude of the health vectors.

Conventional state-space models fail to capture the discrete jumps in the data. The output of the conventional Kalman filter using the above data is shown in Fig. 6 in the Appendixes. After the discrete event at flight cycle 530, the state vectors follow the observation vector after a considerable time delay (nearly 25 flight cycles), which is undesirable for an efficient decision making. As indicated earlier, the innovation sequence indicates the efficiency of the model in representing the data. This is a zero mean identical and independently distributed (iid) process for the correct model. The above properties do not hold if the model fails to represent the data. Figures 7 and 8 demonstrate the innovations for our model and general state-space model. It was found that our method is more accurate in representing the data compared to the conventional models.

In the present setup, the availability of external observers to indicate the occurrence of the event is assumed; the efficacy of these observers are critical to the performance of the model. The events are generally interdependent with the states and are not considered here. Estimators capable of automatically estimating and detecting the interdependent health vector and events are necessary. In this direction, we are working on automatic event detection and health estimation procedures.

Obtaining prediction estimates under a linear state-space framework is simple and straightforward. Unavailability of the new observations forces the innovations to be zero; hence, one can recurse the state prediction (Eqs. (26) and (28)) for obtaining the future estimates. The order of the recursion depends on the prognostic window under consideration. The uncertainty bounds also known as the confidence bounds can be obtained using the state covariance matrix [5]. However, in the current model, one needs the future input deviations and events for predicting the health vectors. To tackle this issue, we assume the absence of any event (both deterministic and random). Also, the deviation of the input from the nominal values are assumed to be constant and equal to that of the most recent value. The resulting prediction estimate

[3]These coefficients are estimated using the average life and magnitude of margins for APUs.

[4]Note that the stamp of the event is assumed to be known. In the case of line maintenance action, the event time stamp is deterministic; however, this is random for an abrupt fault. We assume the availability of external observers to detect such events.

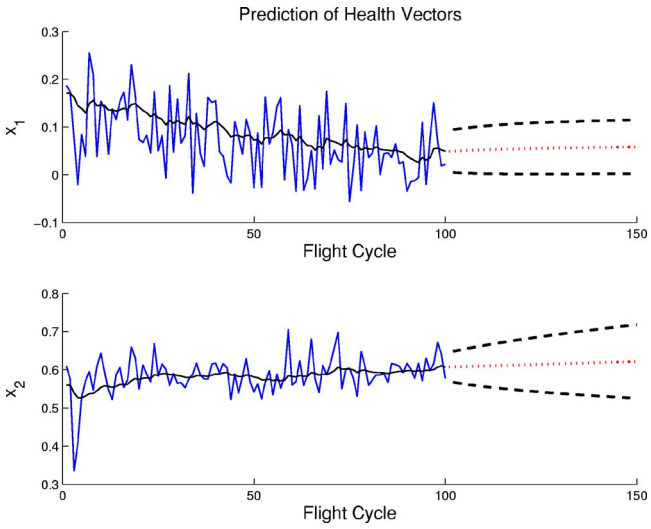


Fig. 4 Prediction of the health vectors

and the associated confidence bounds for a particular APU, with these assumptions, are shown in Figs. 4 and 5. Note that prediction estimates are nothing but the linear trends projected into the future. In the current model, the system matrix ($\mathbf{A}$) determines the trend. Hence, it is required to reidentify the system matrices before predicting the health vectors. In Fig. 4, first 100 data samples are used for estimating the system matrices. Then, the health of the APU for the next 50 flight cycles was predicted. Similar predictions are shown in Fig. 5. A discrete event occurred at a flight cycle 30, introducing a positive dc shift in the observation. Our model and the estimators tackle the discrete jumps in the data and estimates the future trends accurately, which will aid in correct prognostic decisions.

5 Conclusions

Conclusions from this work can be summarized as follows

1. The hybrid prognostic model formulated in this report extends current practices and relaxes constraining assumptions made by leading academic and industrial researchers. It takes into account the interaction between subsystems and uses a multivariate state-space model. It explicitly takes into account the influence of discrete events such as abrupt faults and maintenance actions. This provides a common framework for decision making. It assumes an age-dependent damage accumulation model.

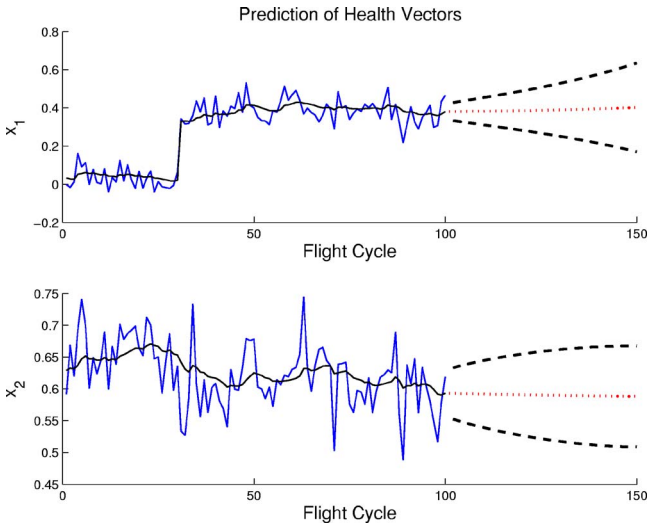


Fig. 5 Prediction of health vectors with events

The partially specified, distribution-based estimation approximates the damage accumulation model (Poisson process) using a partially specified distribution. "Partial" refers to the first two moments. There is no assumption of distributional form for the health vectors; therefore, this method can be used in non-Gaussian state space models. The partially specified, distribution-based estimation procedure is convergent. However, benchmark simulations reveal possible numerical instability. The condition number of some estimates may be abnormally large, but this had no significant impact on the health estimation. The performance of the method of partially specified distribution with the real APU data was reasonably good. The identified system parameters were found to be reasonable and within the engineering limits. Health estimation performance was found to be reasonably good. Our method was found to be efficient in handling the events compared to conventional state-space models.

Thus, we have formulated a scheme for a realistic prognosis of APUs. The formulation is generic and can be extended to many other deteriorating assets.

Nomenclature

$\mathbf{x}$ = health vector
δ_i = threshold for the ith subcomponent of the asset
ϕ = failure mode
$\mathcal{M}$ = set of operating modes
t = APU cycle counter
$\mathbf{w}^k$ = state at the kth mode
n = number of prognostic health states
$\mathbf{A} \in \Re^{n \times n}$ = system matrix
$\mathbf{B} \in \Re^{n \times l}$ = sensitivity of input deviation to the health
l = number of input parameters under consideration
m = number of operating modes within each operational cycle
$\mathbf{u} \in \Re^l$ = measured inputs
$\mathbf{u}^0 \in \Re^l$ = designed inputs
r = number of discrete events within each operational cycle
$\mathbf{C} \in \Re^{n \times r}$ = sensitivity of event to the health
$\mathbf{v} \in \Re^r$ = vector of discrete events
β = scaling factor
θ = intrinsic deterioration rate
$\mathbf{y} \in \Re^n$ = noisy observation of the health state
$\mathbf{z}$ = measurements
Q = covariance of the process noise
R = covariance of the observation noise
ϵ = observation noise
λ = Poisson distribution parameter

Appendix A: Least Sqaures Regression Based Initialization of System Matrices

The initialization of the system parameters is very important in getting good estimates. In this regard, a least squares based scheme is explained here.

The best estimate of the state vector ($\mathbf{x}_t$) can be easily deduced from Eq. (17), which is given by [9]

$$E[\mathbf{x}_t|\mathbf{y}_t] = \mathbf{y}_t$$

we define the error (ϵ_t) as

$$\epsilon_t = \mathbf{x}_t - E[\mathbf{x}_t|\mathbf{y}_t] = \mathbf{x}_t - \mathbf{y}_t$$

Total error is expressed as

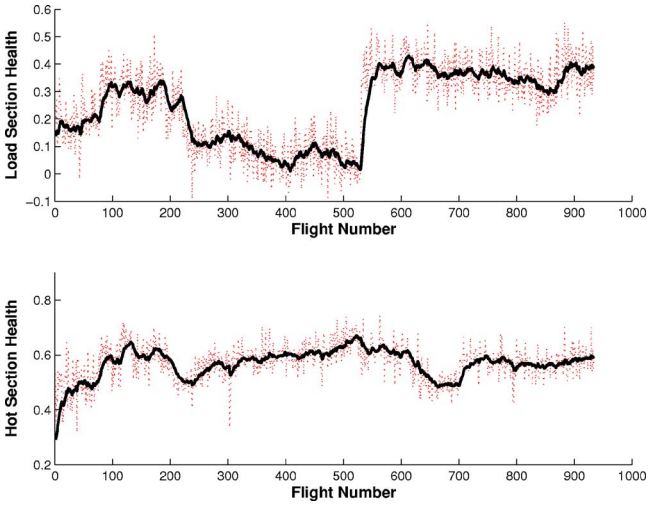


Fig. 6 Health estimation of the APU using conventional Kalman model

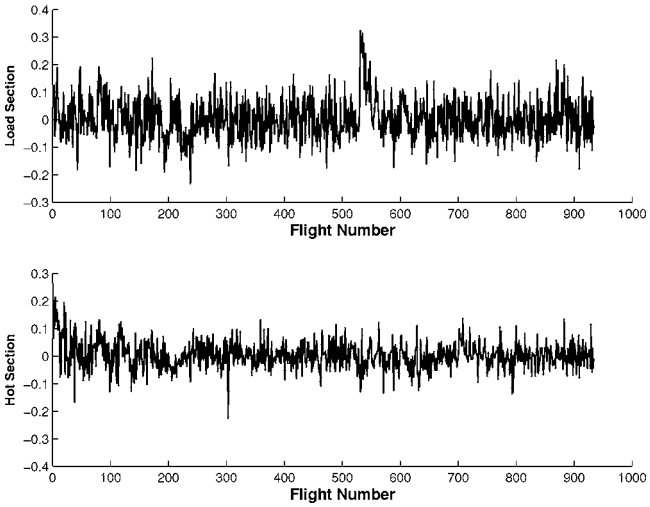


Fig. 8 Innovation sequence using a conventional state-space model

$$\mathcal{E}_n = \sum_{t=1}^{n} (\mathbf{x}_t - \mathbf{y}_t)^T (\mathbf{x}_t - \mathbf{y}_t)$$

$$= \sum_{t=1}^{n} \left(\mathbf{A}\mathbf{x}_{t-1} + \mathbf{B} \sum_{k=1}^{m} \mathbf{U}_t^k + N_t - \mathbf{y}_t \right)^T$$

$$\times \left(\mathbf{A}\mathbf{x}_{t-1} + \mathbf{B} \sum_{k=1}^{m} \mathbf{U}_t^k + N_t - \mathbf{y}_t \right)$$

where n is the number of observations. The system matrices can be obtained by minimizing $\mathcal{E}_n$, which is given by

$$A, B = \arg\min_{A,B} \mathcal{E}_n \tag{36}$$

Appendix B: Observation Noise Estimation

The observations (margins) were observed to be nonstationary with varying mean. Hence, the observation noise needs to be updated over time. For this, we use the detrending method, which is possible because we have a Gaussian observation process ($\mathbf{y}_t$) [5]. The algorithm for obtaining R using the detrending method is explained below.

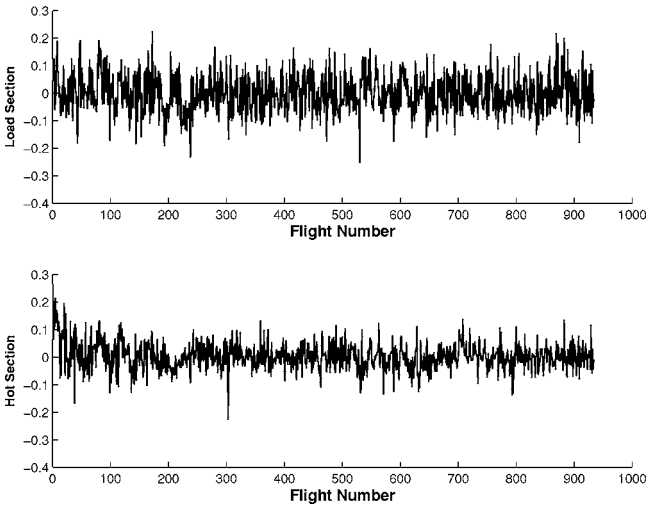


Fig. 7 Innovation sequence using our model

1. Fit a polynomial curve ($\hat{y}$) of *suitable* order (we fit a fourth-order model).
2. Obtain the detrended component of the observation process using $y_{dt} = (\mathbf{y} - \hat{y})$.
3. Estimate R using the detrended data y_{dt} using a conventional variance estimation method.

A sample value for the estimated R matrix is

$$R = \begin{bmatrix} 0.0047 & 0 \\ 0 & 0.0022 \end{bmatrix}$$

Appendix C: Event Sensitivity Calculation

1. Normalize the margins and detect the turbine replacement.
2. Denoise the data using a Wiener filter or wavelet denoising method.
3. Calculate the dc shift in the margins (both EGT and PT) due to the event.
4. Fix the sensitivity (C) equal to the shift.

Appendix D: Data Generation and Montecarlo Experiments

We simulated the data considering all the aspects of the random process. The developed algorithms were tested with these data. We found a distinct advantage of synthetic data if simulated *realistically*—the actual value of the parameters, which are not available in the APU data. These parameters can be used for testing the two crucial properties (performance index) *bias* and *variance* of the estimator by undertaking MC experiments [4]. Details of data generation and MC experimentation are given below.

D.1 Data Generation.

1. Fix the parameters of the model.
2. Fix the initial conditions for state vector.
3. Generate N independent $\mathcal{P}(\lambda)$ random variate (ω).
4. Generate N independent $\mathcal{N}(0, \sigma^2)$ random variate (ν).
5. For $t = 1:N$: Having the parameters of the model and ω, generate the state sequence $\mathbf{x}(t)$ using Eq. (20). For a predetermined time (t_e), add event terms by adding sensitivity coefficients to the state. Generate the observation sequence $\mathbf{y}(t)$ using $\mathbf{x}(t)$ and $\boldsymbol{\nu}(t)$. (Use the second part of equation (20).)

D.2 Montecarlo Experiment. For ind$=1:M$:[5]

1. Generate $\mathbf{y}(t)_{t=1:N}$, as explained in Sec. D.1.
2. Initialize the system matrices using the LSR technique.
3. Obtain the system matrices and the health vector.
4. Obtain the condition number, eigenvalues, and matrix-2-norm.

Obtain the mean and variance of the above parameters and assess the performance of the algorithm.

Appendix E: Comparison With Conventional Kalman Filter

Figure 6 demonstrates the tracking efficiency of the conventional state-space model. It was found that the model fails to capture the event at flight cycle 530. A time delay of 25 flight cycles was needed for the estimates to adjust and follow the observation, whereas our model was found to be accurate. The model insufficiency can also be checked using innovation sequence. The innovation is defined as the difference between the predicted value and the observed value. This is iid with zero mean. Any anomaly from the above properties will indicate the change in system dynamics or insufficiency of the model. Innovation sequence for our scheme has been shown in Fig. 7, whereas Fig. 8 demonstrates the innovations for conventional models. As can be seen, our model is closes to the desired quality compared to that of conventional models [10].

[5] M is the total number of MC experiments, which is taken as 5000.

References

[1] Khalak, A., and Hess, A., 2004, "Modeling for Failure Prognosis in Coupled Systems," *Proceedings of Society of Machine Failure and Prevention Technology*, 58, pp. 449–458.
[2] Mylaraswamy, D., Mukherjee, J., Shetty, P., and Thirumaran, E., 2005, "System and Apparatus for Predicting Removal and Replacement of Deteriorating Products," Report Nos. H0008916–3136 and SBE 3136.
[3] Box, G., Jenkins, G., and Reinsel, G., 1999, *Time Series Analysis: Forecasting and Control*, Pearson Education Inc..
[4] Kay, S. M., 1998, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, Eglewood Cliffs, NJ.
[5] Schumway, R. H., and Stoffer, D. S., 2000, *Time Series Analysis and Its Applications*, Springer-Verlag, New York.
[6] West, M., and Harrison, J., 1997, *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York.
[7] Klein, B. M., 2003, *State Space Model for Exponential Family Data*, Department of Statistics, University of Southern Denmark.
[8] Gelb, A., 2001, *Applied Optimal Estimation*, MIT, Cambridge, MA.
[9] Ljung, L., 1999, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ.
[10] Box, G. E. P., and Jenkins, G. M., 1990, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Derivation of Diagnostic Requirements for a Distributed UAV Turbofan Engine Control System

Olof Hannius
e-mail: olof.hannius@volvo.com

Dan Ring
e-mail: dan.ring@volvo.com

Performance and Control Systems,
Volvo Aero Corporation,
461 81 Trollhättan, Sweden

Johan Karlsson
Department of Computer Science and Engineering,
Chalmers University of Technology,
412 96 Göteborg, Sweden
e-mail: johan@ce.chalmers.se

This paper presents a method for deriving requirements for the efficiency of diagnostic functions in distributed electronic turbofan engine control systems. Distributed engine control systems consist of sensor, actuator, and control unit nodes that exchange data over a communication network. The method is applicable to engine control systems that are partially redundant. Traditionally, turbofan engine control systems use dual channel solutions in which all units are duplicated. Our method is intended for analyzing the diagnostic requirements for systems in which a subset of the sensors and the actuators is nonredundant. Such systems rely on intelligent monitoring and analytical redundancy to detect and tolerate failures in the nonredundant units. These techniques cannot provide perfect diagnostic coverage and, hence, our method focuses on analyzing the impact of nonperfect diagnostic coverage on the reliability and safety of distributed engine control systems. The method is based on a probabilistic analysis that combines fault trees and Markov chains. The input parameters for these models include failure rates as well as several coverage factors that characterize the performance of the diagnostic functions. Since the use of intelligent monitoring can cause false alarms, i.e., an error is falsely indicated by a diagnostic function, the parameters also include a false alarm rate. The method was used to derive the diagnostic requirements for a hypothetical unmanned aerial vehicle engine control system. Given the requirement that an engine failure due to the control system is not allowed to occur more than ten times per million hours, the diagnostic functions in a node must achieve 99% error coverage for transient faults and 90–99% error coverage for permanent faults. The system-level diagnosis must achieve 90–95% detection coverage for node failures, which are not detected by the nodes themselves. These results are based on the assumption that transient faults are 100 times more frequent than permanent faults. It is important to have a method for deriving probabilistic requirements on diagnostic functions for engine control systems that rely on analytical redundancy as a means to reduce the hardware redundancy. The proposed method allows us to do this using an existing tool (FAULTTREE+) for safety and reliability analysis.
[DOI: 10.1115/1.2795776]

Keywords: control system, diagnostic, distributed, gas turbine, fault tree, Markov

1 Introduction

Safety critical control systems often use dual or triple modular hardware redundancy [1]. The use of replicated hardware units is a conceptually simple and efficient approach to deal with random hardware faults in control systems. Control systems for turbofan engines have traditionally been implemented using two control channels [2]. In contrast to manned aircraft, it is common that unmanned aerial vehicles (UAVs) only have one engine. The engine and control system in a single-engine aircraft are often developed based on engines used in twin-engine aircraft. These engines require special adaptation to improve safety. Examples of aircraft with only one engine are the F16 and JAS 39 Gripen. Recently, the Korean T-50 (Golden Eagle) entered service. This is the first single-engine aircraft with a dual redundant electronic engine control system without hydromechanical backup [3]. From a life cycle and maintenance perspective, duplicating all sensors, actuators, and electronic units may not provide an optimal solution to ensure system safety. Using a *partially redundant* engine control system, in which only the most critical hardware units are duplicated, is both feasible and cost effective for many types of UAVs. Sensors do not need to be duplicated in an engine control system since the sensor readings can be replaced by values calculated from other sensor readings by the use of analytical redundancy. Some actuators do not need to be duplicated because they can assume a fail-safe position without leading to a catastrophic failure. There are several advantages of avoiding replication of hardware units: It reduces weight, power consumption, the overall failure rate, equipment cost, and maintenance. The main disadvantage is that error detection becomes more demanding. In systems where all units are duplicated, errors can easily be detected by comparing duplicated data. Errors in nonduplicated data must be detected by intelligent monitoring techniques that assess the correctness of the data based on knowledge about the current system state and the general characteristics of the data. One advantage of intelligent monitoring is its ability to identify a faulty unit, which cannot be done by comparing duplicated data. Identifying a faulty unit is essential in systems that use reconfiguration and graceful degradation to ensure safety. A drawback of intelligent monitoring is that some errors may pass undetected while in some cases the error detection mechanism may flag an error for data that is correct. Hence, it is important to assess the impact of undetected

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 14, 2006; final manuscript received July 6, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal.

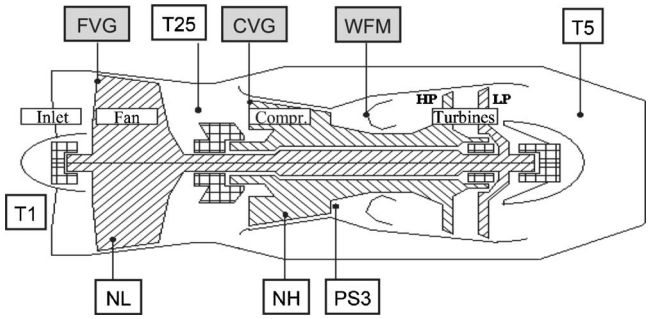


Fig. 1 Proposed turbofan UAV engine

errors and false error detections on safety and reliability.

In this paper, we present a method for deriving and analyzing requirements for the efficiency of the diagnostic functions in a distributed full authority digital engine control (FADEC) system. The method allows us to assess the impact of the diagnostic efficiency on system safety and reliability. It relies on a probabilistic analysis that combines fault trees and Markov models. Engine safety and reliability requirements are used as inputs and diagnostic requirements are derived and expressed in terms of probabilities for successful error detection and recovery, and false alarms. The method is intended for analysis of distributed systems using intelligent sensors and actuators. We consider distributed systems because they have several desirable properties compared to traditional centralized systems. One is the flexibility to add or delete functions with a minimal impact on the rest of the system. Distributed systems are also particularly suited to employ partial hardware redundancy due to their provision of distributed processing

The proposed method is applied to a hypothetical UAV engine control system consisting of seven simplex sensors, one simplex and two duplex actuators. The simplex actuator controls guide vanes in the fan. This actuator does not need to be replicated because a safe engine thrust can be obtained with a fixed fail-safe position of the actuator. The actuators controlling the guide vanes in the compressor and the fuel flow are duplicated. The output from the analysis is a quantitative measure on how well the diagnostic functions must perform. Diagnostic performance is expressed as error coverage factors. A coverage factor expresses the probability that an error is masked or recovered appropriately. The assumption is that both transient and permanent faults will occur in the control system. We assume that permanent faults occur less frequently than transient faults and that the coverage factors for errors caused by transient and permanent faults are different. Permanent hardware faults occur as a result of weaknesses and aging of electronic circuits. Examples of common hardware failure mechanisms are electromigration and gate oxide breakdowns. Transient faults are caused by disturbances that change the logic state of latches and memory elements. These disturbances can be generated by electromagnetic interference or high energy particle radiation. High energy particle radiation is present both at flight altitudes and at ground level [4–6]. Other types of faults are specification mistakes and slipups during software coding. These kinds of faults are not considered in this paper.

The remainder of the paper is organized as follows: Basic turbofan safety requirements are defined in Sec. 2. The distributed control system architecture and requirements are described in Sec. 3. Section 4 contains examples of diagnostic methods and methods to obtain fault tolerance in a control system. The proposed method to determine diagnostic requirements applied on a UAV turbofan engine control system is described in Sec. 5. The quantitative part of the analysis is performed in Sec. 6 and the result and conclusions are presented in Sec. 7.

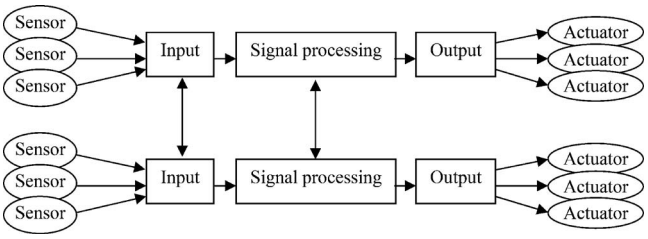


Fig. 2 A traditional FADEC system

2 Safety Requirements

The mission time for a UAV can vary a lot and depends on the type of UAV and the purpose of the mission. There are, for example, both short-range and long-duration high-altitude aerial reconnaissance types of vehicles. We have chosen a mission time of 1 h. The following undesired events with associated quantitative requirements (for a 1 h mission) constitute the starting point for our analysis.

(1) The probability for improper engine thrust shall be less than 10^{-5}.
(2) The probability for mission abort shall be less than 10^{-3}.
(3) The probability for an unscheduled maintenance action after a mission shall be less than 10^{-3}.
(4) The probability that a single failure in the control system causes improper thrust shall be less than 10^{-7}.

In this study, we consider UAVs that can operate together with civil air traffic. For such UAVs, the basic safety requirement is not to cause deaths or injure humans in any way. Human injury and death would be the result from collision with another aircraft or a crash in inhabited areas. The probability to collide with another aircraft must be extremely remote and is avoided by navigation and anticollision systems. Improper thrust (undesired event No. 1) during a sufficiently long time is anticipated to be catastrophic but a discussion of how likely it is to cause injury/death by a failed UAV engine is beyond the scope of this paper. A mission must be aborted (undesired event No. 2) if one node has failed permanently. Another node failure has potential to cause a catastrophic failure and this risk is minimized if the mission is aborted. Failures that impact the ability to pursue advanced flight maneuvers, i.e., maneuvers that may need a sequence of different thrust levels with quick response, are also considered to cause an aborted mission. The control system complexity (and failure rate) must be kept low so that the risk for unscheduled maintenance is low (undesired event No. 3). Undesired event No. 4 concerns the control system architecture and redundancy. The probability that a single failure in the control system becomes catastrophic shall be very improbable.

The proposed UAV engine for which the requirements shall be applied to is a turbofan gas turbine, as shown in Fig. 1. It is a fairly complex engine with variable fan and compressor guide vanes. The rectangular blocks represent the sensors and actuators. The shaded blocks are servo systems with both sensor and an actuating device, and is here simply denoted *actuator*.

3 Control System Architecture

For a long period of time, there has been a desire to decrease the number of redundant channels both for flight control and engine control systems. Examples are Desai et al. [7], published 1979, and Forsberg [8] in 2003. The challenge is to design a control system that has sufficient fault tolerance with a minimum of hardware redundancy. As mentioned in the Introduction, most FADEC systems of today are dual redundant. They are located on the engine in a ruggedized box with cooling and protection against electromagnetic interference (EMI). For a UAV application, it can

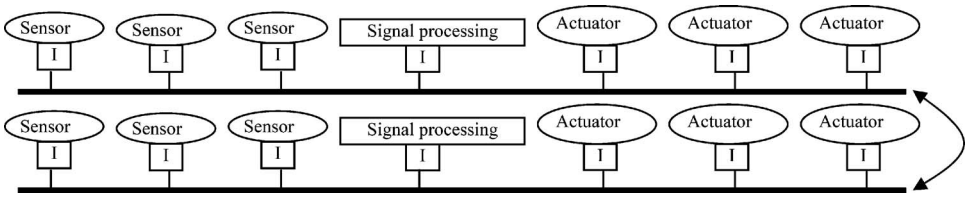


Fig. 3 A distributed FADEC system

be beneficial to take one step further to reduce the redundancy level because less hardware leads to a lower overall failure rate for the electronic equipment. In a single-engine application, there are several mechanical components that can fail and cause a catastrophic event (single point of failure), e.g., turbine disks, bearings, etc. We want the electronic system to be just as robust as the mechanical components using as little hardware redundancy as possible. One way to reduce hardware redundancy is to divide the traditional FADEC system into smaller functions and allow these functions to be reconfigurable, see Figs. 2 and 3. The traditional FADEC system in Fig. 2 has redundant control channels. A sensor failure can be recovered by using the correct sensor signal from the other channel. A failure of the input function can be recovered by using the communication channel between the signal processing blocks. However, any additional failure in the sensors, input, or signal processing blocks will lead to a system failure.

Figure 3 shows a distributed system where each function is connected to a network with an interface (square denoted "I"). The complexity of the distributed system is similar to the traditional system since both are dual redundant. A failure of any function such as sensor, signal processing, or actuator can be recovered. A system failure only occurs if two identical blocks such as the same two actuators in each redundant channel fail. The above example shows that the distributed system has the potential to be safer than the traditional FADEC system. If we are not interested in better safety, we could maintain the same safety as the traditional FADEC system while reducing the hardware redundancy.

The control system architecture in Fig. 4 will be used for describing our method for deriving diagnostic requirements. The white boxes in the upper part of the figure are sensors and the shaded boxes are actuators. The lower part shows power supplies and control units. Compared to the distributed system example in Fig. 3, this system has less redundancy. None of the sensors are dual redundant. Any failed sensor measurement can be calculated from other sensor measurements by the use of analytical redundancy [9]. Also, one of the three actuators, fan variable geometry (FVG), is designed to return to a closed position after a failure and can therefore be of simplex configuration. The engine performance with a closed FVG actuator provides sufficient thrust for safe operation (but the mission must be aborted). The compressor variable geometry (CVG) and fuel flow (WFM) actuators are dual redundant because they must always be operational to ensure proper engine thrust throughout the flight envelope. Even though the control system has a reduced level of hardware redundancy, it can tolerate any single node failure and still control the engine so that a safe thrust level is maintained.

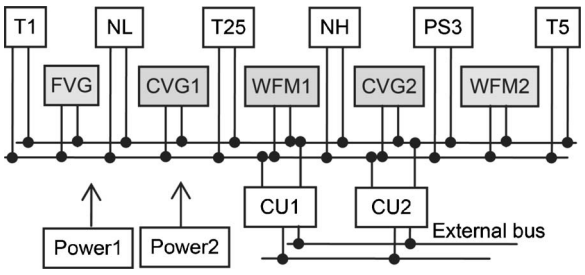


Fig. 4 Hardware architecture

Each box in Fig. 4 is a node in the distributed control system. The nodes can be physically distributed or just a modularized part of a circuit board. It is anticipated that the nodes have computational resources and built-in self-diagnostic capability. The communication between nodes is via a digital network. The network is of bus topology, simply denoted *bus*. The bus is anticipated to be very straightforward in its construction (copper wires). Each node is powered from two sources, Power1 and Power2. The engine control laws are executed by the control unit nodes (CU1 and CU2). The control unit nodes are also connected to a central UAV bus (external bus) for information exchange.

It is anticipated that each actuator has one sensor and one positioning device working in a closed loop. We anticipate that the driver circuit in a positioning device is monitored by an independent electronic function, a watchdog which checks that the driver responds correctly to demands from the control unit nodes. The purpose of the watchdog is to prevent the occurrence of single points of failure. It does so by disconnecting the driver from the positioning device if the actual output current deviates too much from demanded current. Table 1 shows examples of control system failure modes that violate the four safety requirements listed in the beginning of Sec. 2.

The control system in this case study is anticipated to be built of commercial electric/electronic components without need for individual screening test and qualification. We believe that it is more cost efficient to discard tests on individual components and only run delivery tests on complete nodes. The risk for having to scrap a complete node due to a component failure is low in view of the high reliability that modern electronic components have. The risk for poor quality of nodes is avoided by letting multiple suppliers manufacture standardized nodes. This will result in cost efficient and reliable solutions by commercial reasons.

Table 1 Control system failure modes

Requirement		Examples of failure modes leading to an undesired event
No.	Undesired event	
1	Improper thrust	Erroneous output from the CVG and WFM actuators. Unsafe output from the FVG actuator.
2	Mission abort	Loss of any redundant unit or loss of thrust modulation (multiple transient node failures).
3	Maintenance	A permanent fault in any node requires a maintenance action.
4	System failure caused by a single fault	A faulty driver in an actuator is not disconnected. A node acts as a babbling idiot on the bus and disturbs messages sent by other nodes and thereby prevents communication.

4 Fault Tolerance and Diagnostics

Since the introduction of microprocessors in control systems for aircraft and jet engines, effort has been made to improve fault tolerance through better system architectures and new diagnostic methods. Diagnostics can be applied both on the electric/electronic systems and the mechanical engine components. Diagnostics or diagnostic functions can in its simplest form be a range check of a measured quantity, but it can also constitute an advanced model that uses several inputs and requires a lot of computational resources to produce its output. Engine diagnostics can be performed either on board or on ground between missions to reduce or eliminate the need for manual inspection. The purpose with onboard health monitoring is to reduce fault latency and react to detected faults promptly in order to improve safety and engine operability. On ground systems is used to improve maintenance by performing mission independent life prognostics. Good prognostics can reduce unnecessary maintenance actions and thereby lower the cost. In this study, we focus on diagnostics for the electrical and electronic components of the engine control system. Failures from a controls engineering point of view are the result from measurement errors, deviation between models, and physical systems and malfunction in the control object. Computer engineers often focus on the actual hardware architecture of the control system, i.e., microprocessor, memories, compiler, software, and techniques to handle faults inside the control system.

Within controls engineering, the methods used to perform diagnostics are often denoted FDI(A), *failure (or fault) detection, isolation, identification*, and *accommodation* [10,11]. A great variety of diagnostic methods are based on mathematical models of the control object. These are denoted model-based methods. Examples of model-based approaches to solve the FDIA problem are *parameter identification* [12], *parity equations* [13], and *diagnosis observer* [14]. A design procedure for FDI of a gas turbine was presented by Simani [15]. The procedure uses analytical redundancy and a bank of dynamic observers and Kalman filters. The challenge is to develop FDI methods that work well without false alarms even though the model deviates from the actual process plant. Other methods such as *neural networks* [16] and *fuzzy logic* [17] have also been used to recognize nonlinear and transient relationships.

The research within dependable computing has also resulted in a variety of methods to handle errors in a computer system. Examples are *code reordering* to minimize the lifetime of variables [18], duplicate [19], or triplicate [20] variables to detect data errors. *Double or triple execution* [21] can be used to detect transient errors. There are different types of recovery methods to use once an error has been detected. It can be either *rollback* or *rollforward*. Rollback recovery requires that checkpoints are stored at regular intervals so that the operations between the last checkpoint and the detected erroneous state can be executed again. To store the system state may require quite much time and storage overhead. Variants such as *best effort recovery* [22] can be used to reduce the overhead. In rollforward recovery, the system state is corrected to be able to move forward. Application specific techniques use *executable assertions* to check that variables are within specified limits. The previously mentioned controls engineering methods are examples of executable assertions.

In this paper, we use the terminology and basic concepts defined for secure and dependable computing systems in Ref. [23]. This terminology gives precise meanings to the words fault, error, and failure. It defines *failure* as a synonym, or short form, for *service failure*. A failure occurs when the service delivered by a system deviates from the correct service. The concept of a system is generic; it can mean different things depending on the view taken. For example, a system could be an entire computer system, such as an engine control system, a processing node in a distributed system, or a component of a node such as a sensor element. Thus, we can talk about a computer system failure, a node failure, and a sensor failure when these units do not deliver the service we expect from them. A failure is caused by an error, which is an incorrect internal or external state of the system. A fault is the adjudged or hypothesized cause of an error. Examples of faults are a physical defect in hardware and a software flaw. The persistence of a fault can be permanent or transient. Open and short circuits are examples of permanent hardware faults. The passage of an ionizing particle that causes a bit flip in a latch in a VLSI circuit is an example of a transient fault. (Bit flips in VLSI circuits caused by transient hardware faults are often called *soft errors*, since the error can be removed by reloading the correct bit value into the hardware element that stores the bit.) The purpose with control system diagnostics is to improve the fault tolerance. Well designed diagnostic functions can reveal a fault in a system by detection of an abnormal system state, i.e., an error. When an error is detected, it can be prevented from affecting the system outputs beyond specified limits (prevent a service failure). An important factor that must be considered when developing diagnostic functions is the risk of false alarms (i.e., when an error detection mechanism falsely flags an error for a correct system state) and the consequences when the system performs an unnecessary recovery as a result of a false alarm. Other useful terminology that is specific to fly-by-wire architectures is presented in Ref. [8].

5 Method Description

This section describes the proposed method for deriving requirements for the diagnostic functions in a distributed engine control system. The requirements are expressed in terms of coverage factors for node-level and system-level diagnostic functions. A coverage factor is the conditional probability that a set of diagnostic functions detects (or removes) an error given that a fault has caused an error in a certain node. We assume that the nodes are self-checking (to a certain degree) and therefore equipped with mechanisms that can detect or remove errors caused by faults occurring in the node itself. These are the node-level diagnostic functions. We also assume that the nodes have system-level diagnostic functions that are used to monitor the behavior of other nodes. Hence, the system-level functions allow the system to detect errors that are not detected by the node-level functions. For the node-level functions, we use different coverage factors for transient and permanent faults, while we use a single coverage factor for the system-level functions.

To model the impact of errors in each node, we use a Markov chain model, see Fig. 6, containing five states that describe the status of the node: *OK (correct operation), recovering from transient fault, recovering from false alarm, detected node failure*, and *undetected node failure*. The input parameters to the Markov chain models are the coverage factors for transient and permanent faults, the rates of permanent and transient faults, and the rate of false alarms. From the Markov chain model, we derive the state probabilities as a function of time. We then derive approximations of these state probability functions, which we use as inputs in two different fault-tree models. We use one fault tree for calculating the *probability of mission abort* and another one for calculating the *probability of system failure*.

Another method for modeling the impact of diagnostic performance was presented by Åslund et al. in Ref. [24], in which the analysis is a pure fault-tree analysis. When events such as the occurrence of false alarms and undetected faults are included in the tree structure it tends to become complex and difficult to grasp. Also, there is no straightforward approach to model the effects of transient faults in a fault tree. The state diagram of a Markov chain model is well suited for modeling the effects of transient as well as permanent faults, and the impact of error recovery and false alarms.

Errors caused by transient faults can have various effects on the system. They can be overwritten and disappear, detected, and masked by the error handling or they can remain undetected and cause a severe disturbance so that the complete system fails. We distinguish between *covered* and *noncovered faults*. A covered

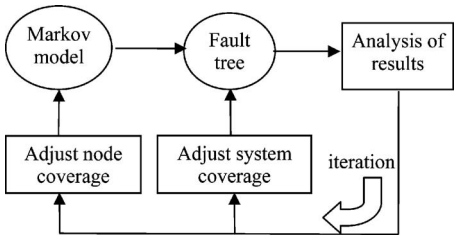


Fig. 5 Iterative process of diagnostic requirement derivation

fault is one that causes an error that is detected while the error from a noncovered fault remains undetected and leads to a failure of a node or of the complete system. We assume that all detected errors are recovered. Any undetected errors that have no or a negligible effect on the system are considered to be covered faults.

The probability factors that represent the diagnostic function's ability to recover from errors are expressed in terms of error coverage factors. Three different types of error coverage factors are used for the control system.

(1) Node-level coverage of transient faults c_t
(2) Node-level coverage of permanent faults c_p
(3) System-level coverage of a node failure c_s

The diagnostic functions in each node are assumed to detect and recover from errors caused by transient faults with the probability c_t. Consequently, they will fail to detect and recover from transient faults and thereby produce a node output failure with the probability $(1\text{-}c_t)$. One example of a transient fault that can cause a node output failure is a bit flip in a register that holds a constant, for example, a gain factor. If the constant is not updated during the mission and the diagnostic functions are unable to detect the error, it will persist throughout the mission causing the node to produce bad data continuously. A permanent hardware fault in a node is assumed to be detected by the node itself with probability c_p. A node that produces erroneous output data is declared as failed by the other nodes of the control system with the probability c_s (system-level coverage).

Figure 5 illustrates the process to derive the diagnostic requirements. Markov models are used to describe the functional state of each node and a fault tree combines the effect of nodes being in different functional states into a top event. The top event expresses which node faults and failures that must occur before an undesired event occurs. We use the process on two undesired events; improper thrust and mission abort (the top event for improper thrust is denoted *system failure* in the fault tree). The coverage factors c_t and c_p for transient and permanent node faults are parameters in the Markov chain model. An assignment of these parameters is done in the process block *adjust node coverage* in Fig. 5. The result from the Markov analysis is then fed into the basic events of the fault tree. The fault tree includes the coverage factor c_s, which can be changed in *adjust system coverage*.

The *analysis of results* consists of comparing the quantitative result from the fault tree analysis with the stated requirements. If the result from the fault tree is higher than the probability for the undesired event stated in the requirement, the assumed error coverage is not sufficient. Then, the error coverage figures need to be adjusted and the process reiterated. The process should continue until the requirements match the analysis result. We have used the FAULTTREE+ version 11 software to develop the Markov chain and fault tree models and to perform the quantiative evaluation.

5.1 A Markov Chain Model of Node Failures. A Markov chain model consists of a number of states and events. A state is a functional state of a node such as "OK" or "Failed." An event is represented by a transition between one state to another. A transition can be defined in continuous or discrete time. Here, continuous time is used and a transition is expressed as a constant failure rate. A Markov chain can be either analytically calculated or numerically evaluated in a simulation tool. The result is the probability that the node is in a certain state after a defined time. The probability for being in different states is then inserted in the fault tree as basic events. This is described in Sec. 5.2.

All the nodes in the control system are described with the same Markov chain model, see Fig. 6. The OK state is the normal error-free state. If the node fails to produce correct data, either a transient fault or a permanent failure has occurred. Any fault or failure changes the state of a node and this is modeled as a transfer to one of the four error states in the Markov chain model. The states *recovering from transient fault* and the *recovering from false alarm* are intermittent and the states *detected node failure* and *undetected node failure* are absorbing. A state is absorbing if the probability to leave the state is zero. The error states and transitions are described in Table 2.

The probability that a node is in an error state can be calculated analytically for this relatively simple Markov chain model. The probability that the system is in the state detected node failure is

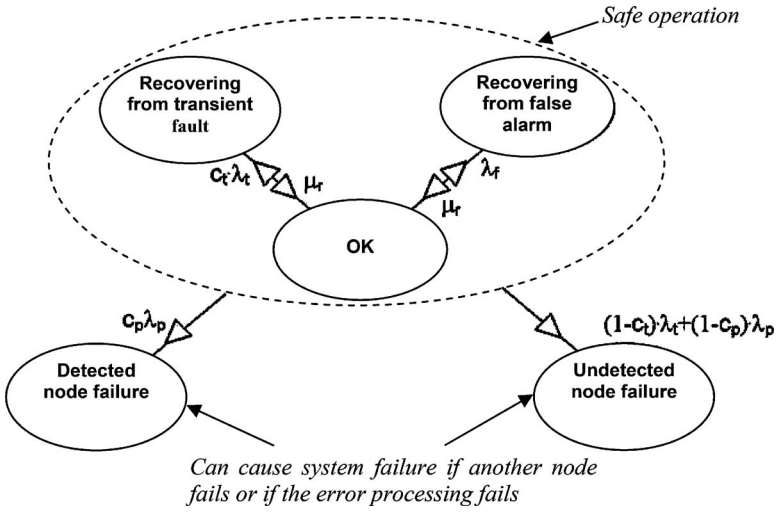


Fig. 6 Markov chain model of a node

Table 2 Description of the error states of a node

Error state	Transition to/from the state
Recovering from transient fault	The node is subjected to a transient fault that can be recovered. It is anticipated that the transition to this state occurs with the rate λ_t multiplied with the error coverage factor c_t. A transition back to the OK state occurs with the recovery rate μ_r.
Recovering from false alarm	The error detection algorithms alert that an error is present when there is none (transition rate λ_f). The node recovers (a transition back to the OK state) with the rate μ_r.
Detected node failure	A detected hardware fault of a node. A transition to this state occurs with the rate $\lambda_p c_p$.
Undetected node failure	Transition to this state occurs if a transient fault is not recovered and manifests itself as a permanent output failure of the node. It also occurs after a permanent error that is not detected and the node continues to send erroneous output data. The transition rate to this state is $(1-c_t)\lambda_t+(1-c_p)\lambda_p$.

$$P_D(t) = \frac{c_p\lambda_p}{(1-c_t)\lambda_t+\lambda_p}(1-e^{-((1-c_t)\lambda_t+\lambda_p)t})$$

and undetected node failure,

$$P_U(t) = \frac{(1-c_t)\lambda_t+(1-c_p)\lambda_p}{(1-c_t)\lambda_t+\lambda_p}(1-e^{-((1-c_t)\lambda_t+\lambda_p)t})$$

The two intermittent states recovering from transient fault and false alarm recover with the same rate μ_r and we can simplify the calculation by merging them into one state (*recovering*). The probability P_R that the node is in this state at time t is

$$P_R(t) = \frac{c_t\lambda_t+\lambda_f}{c_t\lambda_t+\lambda_f+\mu_r}(1-e^{-(c_t\lambda_t+\lambda_f+\mu_r)t})(1-P_D(t)-P_U(t))$$

$$= \frac{c_t\lambda_t+\lambda_f}{c_t\lambda_t+\lambda_f+\mu_r}(1-e^{-(c_t\lambda_t+\lambda_f+\mu_r)t})e^{-((1-c_t)\lambda_t+\lambda_p)t}$$

When a node is recovering, the other nodes are able to analytically calculate the missing parameter and compensate for the loss of data. If two or more nodes are recovering from transient or falsely detected errors, it is assumed that the control system outputs can be limited and the engine kept running (safe operation) during the recovery time. However, the thrust response may be delayed and it is therefore also assumed that the mission must be aborted.

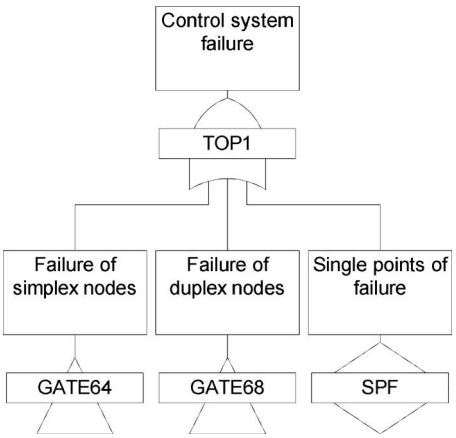


Fig. 7 Top of fault tree

5.2 Fault-Tree Model for System Failure. Fault trees are used to show what combination of subsystem events that must occur before the top event occurs. The top of the fault tree that summarizes the causes for a failure of the control system is shown in Fig. 7.

The top event in Fig. 7 is caused by three major events.

(1) Failure in the nodes that have a simplex configuration (sensors and FVG actuator).
(2) Failure of the duplicated components, i.e., CVG and WFM actuators, control nodes, power supply, or bus.
(3) The event *single points of failure*, (SPF), described in Table 1. The event SPF in the fault tree in Fig. 7 could be assigned to a low number representing a realistic assumption, but is instead assigned to 0 to facilitate the comparison of results.

The fault tree for a control system failure caused by simplex nodes is shown in Fig. 8. It is divided in two subtrees: *covered failure of simplex nodes* and *malicious failure of simplex nodes*. The first subtree models those permanent failures that are covered (as defined in Sec. 5). In Sec. 5.1, we denoted this type of failure detected node failure, and the same name is used for all basic events in Fig. 8. A complete system failure occurs when two or more nodes fail. This is modeled by a threshold gate with a threshold of 2 (represented by "2" in Gate 2).

A quantitative evaluation of a fault tree can be very complicated to perform manually and it is best done with software dedicated for fault-tree analysis. There are different ways of expressing the probability for events. A failure probability model with constant failure rate (exponential density function) and other types of models can be used. Additional properties of events that can be included are repair rate and dormant failure (a failure that is undetected until a certain time has elapsed). In the Markov chain analysis shown in the previous section, the probability for a detected permanent node failure was

$$P_D(t) = \frac{c_p\lambda_p}{(1-c_t)\lambda_t+\lambda_p}(1-e^{-((1-c_t)\lambda_t+\lambda_p)t})$$

Since $((1-c_t)\lambda_t+\lambda_p)t$ is small for the time period and failure rates we are analyzing (with failure rates according to Sec. 6, Table 3), it is convenient to simplify the expression by using the McLaurin expansion

$$e^x = 1+x+\frac{x^2}{2}+\cdots+\frac{x^n}{n!}+\theta(n+1)$$

of the exponential function. If we use the first order approximation and exchange x with $-((1-c_t)\lambda_t+\lambda_p)t$, we will get

$$P_D(t) \approx \frac{c_p\lambda_p}{(1-c_t)\lambda_t+\lambda_p}(1-(1-((1-c_t)\lambda_t+\lambda_p)t)) = c_p\lambda_p t$$

The unavailability of a constant failure rate model with rate $c_p\lambda_p$ is $Q(t)=1-e^{-c_p\lambda_p t}$, which again according to the McLaurin

Table 3 Fault rate assumptions for control system nodes

Node	Permanent fault rate, λ_p	Transient fault rate, λ_t
Sensor	10^{-5} per hour	10^{-3} per hour
Servo	10^{-5} per hour	10^{-3} per hour
CU	10^{-5} per hour	10^{-3} per hour
Power	2×10^{-5} per hour	0
Bus	5×10^{-6} per hour	0

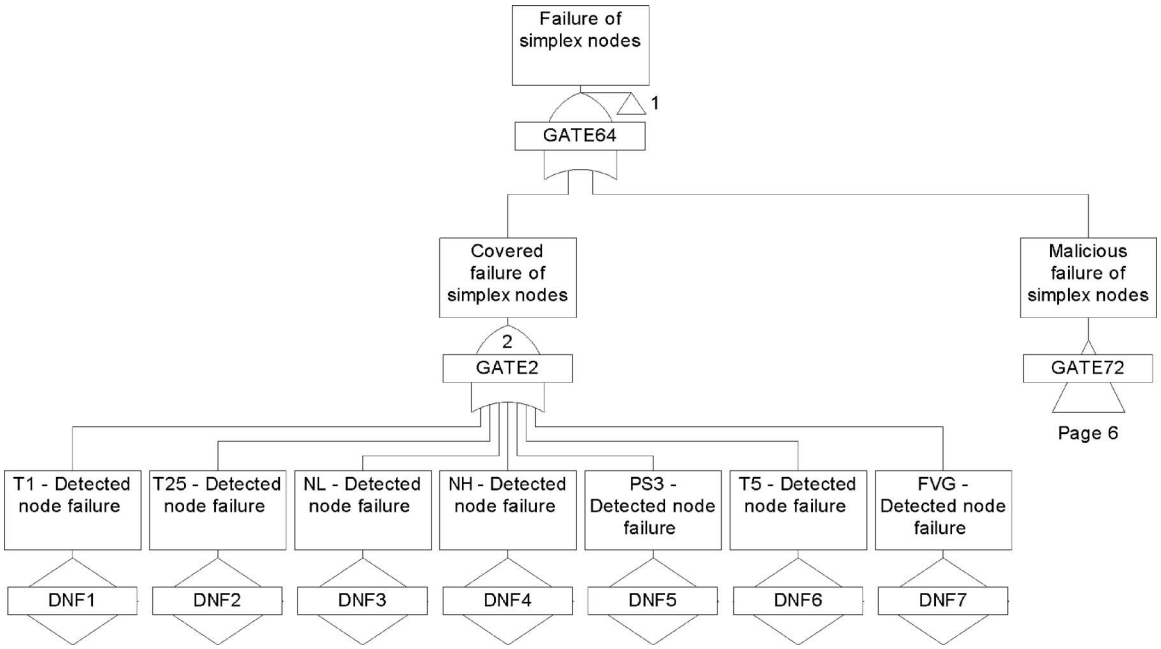


Fig. 8 Failure of simplex nodes

expansion is approximately $c_p\lambda_p t$ for small values. The constant failure rate model with rate $c_p\lambda_p$ can therefore be used for all DNFx events in Fig. 8.

The second subtree in Fig. 8, malicious failure of simplex nodes, is for noncovered faults where output data from a simplex node are wrong and remain undetected, see Fig. 9.

The malicious failure mode is a combination of the failure undetected node failure in Fig. 6 and the inability to detect the failure on system level. The inability to detect a system-level failure is modeled with the SYS_COVERAGE event in Fig. 9 followed

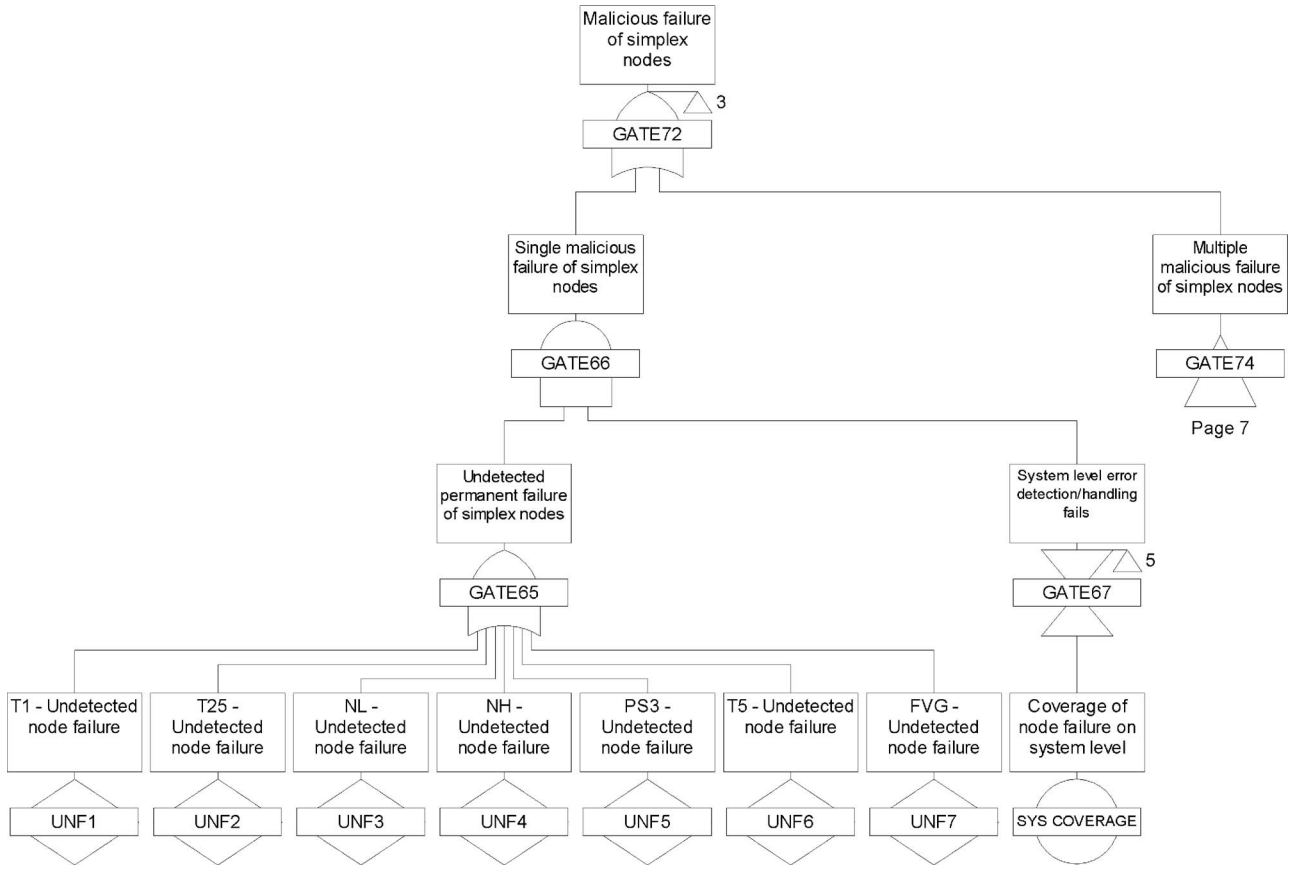


Fig. 9 Malicious failure of simplex nodes

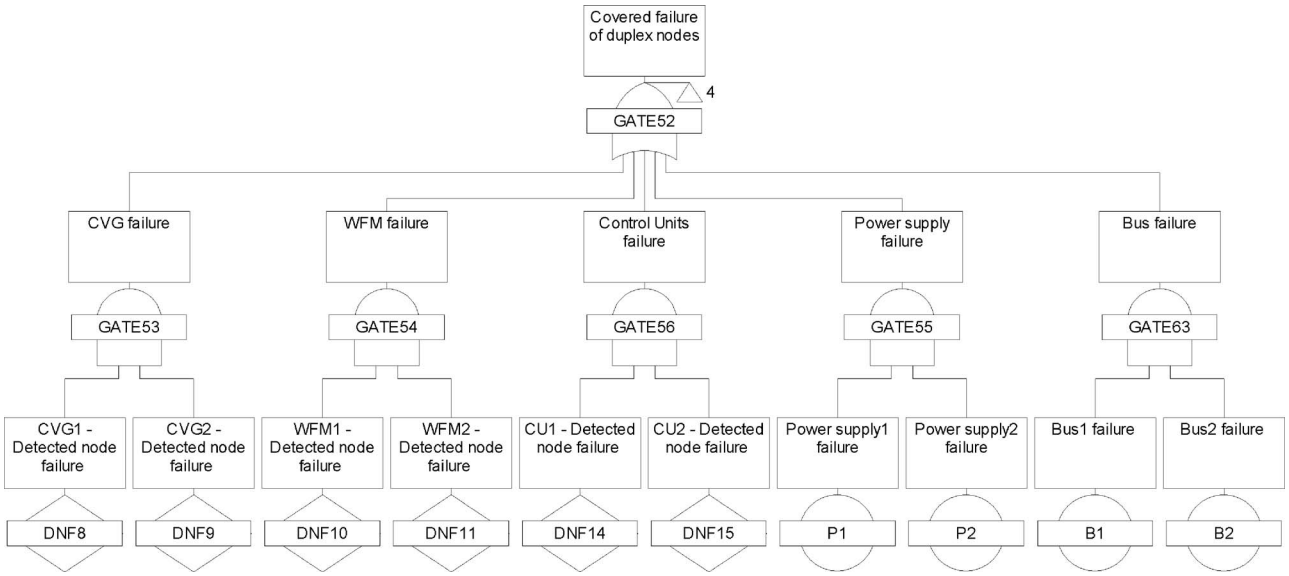


Fig. 10 Covered failure of duplex nodes

by an inverting gate with the textual description *system-level error detection/handling fails*. The gate *undetected permanent failure* of simplex nodes is the OR function of all undetected node failures from simplex nodes. If undetected permanent failures occur in two or more nodes, the system fails irrespective of the ability to detect errors. This is modeled by the gate *multiple malicious failure* of simplex nodes. The probability for an undetected failure was in Sec. 5.1 expressed as

$$P_U(t) = \frac{(1-c_t)\lambda_t + (1-c_p)\lambda_p}{(1-c_t)\lambda_t + \lambda_p}(1 - e^{-((1-c_t)\lambda_t + \lambda_p)t})$$

Using the same assumptions as for P_D, P_U can be approximated to $P_U(t) \approx ((1-c_t)\lambda_t + (1-c_p)\lambda_p)t$. Even if transient faults are assumed to occur more frequently than permanent faults, a constant failure rate model is very accurate for describing the failure probability. The values for failure rates are found in Sec. 6, Table 3.

The part of the fault tree modeling the duplex nodes located immediately below the top gate *control system failure* is not shown in a figure but has a similar structure as the simplex nodes shown in Figs. 8 and 9. It also has two branches: *covered failure of duplex nodes* and *malicious failure of duplex nodes*. The difference is how the covered failures are modeled, and this is shown in Fig. 10. Both nodes in a duplex configuration must fail before the system fails. If, for example, one CVG, one WFM, and one power supply node fail, the system is still OK. A failure in the power supply or bus is assumed to be detected and persists during the complete mission. (The failure rate for these nodes is defined in Table 3.)

5.3 Fault-Tree Model for Mission Abort. A mission must be aborted if one node fails permanently, since this brings the system into a high risk state in which an additional node failure may lead to a complete loss of the UAV. A mission is not aborted if a single node suffers a transient fault. However, a mission abort is necessary if transient faults occur in two or more nodes simultaneously. In this case, there are not enough valid data to calculate correct control signals for the engine and the engine must be kept running at constant thrust during the recovery time. There are situations when it is important to modulate thrust and a failure to do so prevents the ability to perform a successful mission. Even if the transient fault rate is as high as defined in Table 3 in Sec. 6 (1 failure per 1000 h), the risk is very low for the occurrence of simultaneous transient faults. When the system identifies such an event, it is likely that the transient fault rate has increased for some reasons. The increase can, for example, depend on reduced ability to tolerate disturbances due to degraded EMI protection or it can depend on deliberate jamming from an external source. Regardless of the cause of the multiple fault, it is assumed that the mission must be aborted. We have defined the two criteria for mission abort, as shown in the fault tree in Fig. 11.

The branch *permanent failure of one node* is not explicitly shown but is the OR function of all detected permanent node failures. Simultaneous transient node failures can occur for both simplex and duplex nodes and the fault-tree structure is the same as the previously described covered failures shown in Figs. 8 and 10. The difference is that the basic events are taken from the states of the Markov chain that can be recovered instead of the absorbing state for a detected permanent error and that the probabilistic

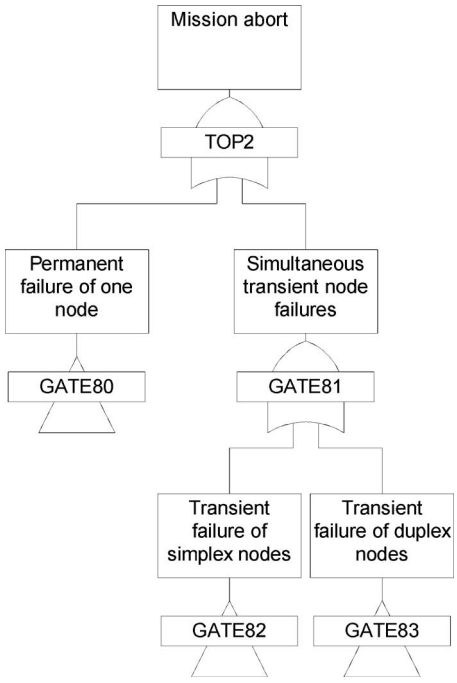


Fig. 11 Mission abort top event

Table 4 Result from quantitative analysis

Category	c_t	c_p	c_s	Failure prob ($\times 10^{-6}$/h)	Comments
A	1	1	don't care	0.003	Ideal system, lower failure bound
B	0.99	0.99	0.999	0.136	Excellent coverage
C	0.99	0.99	0.99	1.32	Very good coverage on all levels
	0.99	0.9	0.99	1.44	
	0.9	0.9	0.999	1.56	
D	0.5	0.5	1	6.12	An ideal system-level coverage compensates for bad node-level coverage
	0.99	0.99	0.95	6.57	
E	0.99	0.99	0.9	13.1	The coverage of transient faults and system coverage are most important
	0.9	0.99	0.99	13.3	
	0.9	0.9	0.99	13.4	
	0.99	0.9	0.9	14.3	
F	0	0	1	24.5	Hypothetical versus realistic
	0.95	0.95	0.95	32.9	
G	0.99	0.99	0.5	65.7	See E
	0.99	0.9	0.5	71.5	
	0.5	0.5	0.99	71.8	
H	0.9	0.99	0.9	130	Even though the node-level diagnostics is excellent, the overall failure probability is high if system-level error diagnostics is missing
	0.99	0.99	0	131	
	0.9	0.9	0.9	132	
	0.99	0.9	0	143	
	0	0	0.99	156	
I	0.9	0.99	0.5	651	See E
	0.9	0.9	0.5	657	
	0.5	0.5	0.9	662	
J	0.9	0.99	0	1300	Just for comparison
	0.9	0.9	0	1310	
	0	0	0.9	1340	
	0.5	0.5	0	6,540	
	0	0	0	13,000	

model for basic events also includes the repair rate.

The unavailability for a basic event with failure rate λ and repair rate μ is $Q(t)=\lambda/(\lambda+\mu)(1-e^{-(\lambda+\mu)t})$ and in Sec. 5.1, the probability $P_R(t)$ that a node is recovering was shown to be

$$\frac{c_t\lambda_t+\lambda_f}{c_t\lambda_t+\lambda_f+\mu_r}(1-e^{-(c_t\lambda_t+\lambda_f+\mu_r)t})e^{-((1-c_t)\lambda_t+\lambda_p)t}$$

If we set the factor $e^{-((1-c_t)\lambda_t+\lambda_p)t}$ to 1, we will get the upper bound

$$\frac{c_t\lambda_t+\lambda_f}{c_t\lambda_t+\lambda_f+\mu_r}(1-e^{-(c_t\lambda_t+\lambda_f+\mu_r)t})$$

of $P_R(t)$. This is a good approximation for the mission time and failure rates we are analyzing. This is the same expression as the constant failure and repair rate model if λ is substituted with $c_t\lambda_t+\lambda_f$ and μ with μ_r.

6 Analysis of the Control System

In this section, we derive the requirements for the control system diagnostics from engine level requirements for the hypothetical UAV engine control system shown in Fig. 4. The process for deriving the diagnostic requirements presented in the previous section (Sec. 5) is used to determine the coverage factors c_t, c_p, and c_s.

Assumptions of transient and permanent fault rates for different nodes are summarized in Table 3. The fault rate of a sensor and a servo node is assumed to be equal even though a servo node contains both input and output circuitry. The motive is that an actuator contains reliable electric components such as linear variable differential transformer (LVDT) and torque motor and that a sensor node contains pressure sensors or temperature sensors that are more error prone due to physical properties of the sensors and the harsh environment. The dominating number used for transient faults is due to the concern about increased sensitivity to external disturbances such as cosmic particle radiation and EMI for new integrated circuit technologies, as these use less charge than previous technologies to store data. Note that transient faults only are assigned to sensor, servo, and control unit nodes and not for power and bus nodes. Transient faults can of course occur in power supplies and buses as well, but is here set to zero because they do not contain components that are sensitive to single event upsets (SEUs).

6.1 Quantitative Analysis. We will start the quantitative analysis by checking that the mission abort probability for a 1 h mission is below the stated requirement 10^{-3}. The calculated result is presented from the fault-tree program as figures on system unavailability, unreliability, failure frequency, etc. For the calculation of mission abort, it is the unreliability that reflects the probability that the top event has occurred at any time during the mission. The exact number for mission abort depends on the assumed coverage factors, but is largest and equal to 1.96×10^{-4} if the coverage is perfect, i.e., $c_t=c_p=1$. A permanent node failure also requires a maintenance action and the mean time is slightly more than 5500 h. Hence, both requirement Nos. 2 and 3 of the *engine requirements* in Sec. 2 are fulfilled.

When calculating the probability for a system failure, we use the method described in Sec. 5 and iteratively perform calculations with different assumptions on coverage. The result is shown and commented in Table 4 and has been categorized based on the probability for a catastrophic event to occur. When ideal figures on error coverage are used ($c_t=c_p=1$ and c_s can be anything between 0 and 1), the chance to get improper engine thrust due to a control system failure is very improbable (Category A in Table 4). However, these systems do not exist in reality. Even if the coverage figures are very close to 1, like Categories B and C, the failure probability rises dramatically. Categories D and E correspond well to the stated safety requirements. Even if the coverage is as high as 0.95 (in Category F), the overall safety requirement is not met.

Let us assume that we design the system so that transient faults

are covered with 99% certainty (c_t=0.99), permanent node failures are covered with 90–99% certainty (c_p=0.9–0.99), and that undetected node failures are captured at system level with 90–95% probability (c_s=0.9–0.95). With these assumptions, the probability for a catastrophic event is between 6.6 and 14 failures per million hours, which is close to the requirement No. 1 (10 failures per million hours) stated in Sec. 2.

7 Summary

A method to derive diagnostic requirements in terms of error coverage factors is presented. The method is used on a distributed turbofan engine control system. Both permanent and transient errors are considered and are included in a Markov chain model for each node. The error states of the Markov chain model are mapped to a fault tree that combines all failure modes of the control system. This method brings two advantages. Firstly, the accuracy of the quantitative result is improved by including coverage factors instead of neglecting them. Secondly, the coverage factors can be used to define requirements on diagnostic performance for each component of the system. These requirements can then be used during development and validation of diagnostic functions.

The proposed method to derive diagnostic requirements for a distributed control system is here summarized and consists of the following steps.

(1) Create probabilistic reliability models for the different types of nodes in the system and introduce states (failure modes) which describe different health status. The states can, for example, be *no error, recovering*, or *permanent failure*. A Markov chain model is well suited for this purpose. Include error coverage factors in the model and derive how the result is to be transferred to the fault tree.
(2) Build a fault tree for the critical events that are of concern and use the failure modes defined in Step 1.
(3) Include system-level coverage in the fault tree and perform iterative simulations with different assumptions on error coverage until it can be shown that the stated requirements are met.
(4) The obtained error coverage factors represent the top level requirements for diagnostic performance both on node and system level. These are used as requirements for hardware and software design and later on when verification activities begin.

The described method with iterative quantitative calculations has been performed in such a way that calculations and mapping between the Markov chain model and the fault tree require manual intervention. The method has also been simplified by treating all nodes equal in terms of failure rate and error coverage. A real system will probably have slightly different figures. The next step to reduce the required effort of manual work and to minimize the risk for using wrong data by mistake could be achieved by automating the calculation process. This would also enhance the usability of the suggested method since different assumptions on architecture, diagnostic performance, etc., more easily can be quantitatively compared to each other.

The quantitative result of the fault tree shows the complete control system failure probability and can easily be compared to stated requirements. The analysis shows that high error coverage, above 95% on an average, on both node and system level is necessary to fulfill requirements. The high number required on error coverage is strongly dependent on the assumed rate of transient faults. However, transient faults are a reality that must be taken into account during analysis. The actual number can of course vary depending on node architecture and the environment for the equipment.

The probability of a critical control system failure for a system with ideal error coverage is almost seven orders of magnitude lower than for a system with no coverage at all and this shows how important the error coverage is. In a real system, the error coverage is closer to 1 than zero. Typically, there is no effect at all for the majority of the faults caused by SEUs [22] even if only simple executable assertions (diagnostic functions) are used. However, there is a quite big gap from detecting *most* errors to detect *all* errors and much effort should be put into the development of diagnostic functions to show that they can detect and recover from errors with high probability. The method presented in this report will help to define requirements for which the diagnostic functions can be validated.

Acknowledgment

We would like to thank the National Flight Research Programme, NFFP, and Volvo Aero for financial support.

Nomenclature

COTS = commercial off the shelf (here, ready-made electronic hardware)
CVG = compressor variable geometry position actuator
EMI = electromagnetic Interference
FADEC = full authority digital engine control
FVG = fan variable geometry position actuator
LVDT = linear variable differential transformer (position sensor)
NH = compressor rotational speed
NL = fan rotational speed
PS3 = combustor static pressure
SEU = single event upset. The content of a memory cell is altered due to highly energetic particles
T1 = total engine inlet temperature
T25 = total compressor inlet temperature
T5 = total turbine exhaust gas temperature
UAV = unmanned aerial vehicle or unmanned air vehicle
VLSI = very-large-scale integration (the process of creating integrated circuits using thousands/millions of transistors on a single chip)
WFM = actuator for fuel metering (W symbols the flow)

References

[1] Yeh, Y. C., 1996, "Triple-Triple Redundant 777 Primary Flight Computer," Proceedings of the IEEE Aerospace Applications Conference, Aspen, CO, pp. 293–307.
[2] Lala, J. H., and Harper, R. E., 1994, "Architectural Principles for Safety-Critical Real-Time Applications," Proc. IEEE, **82**(1), pp. 25–40.
[3] Conigho, S., 2004, "Combat Aircraft Propulsion Revisited," Military Technology, **28**(7), pp. 56–63 (more details on the control system can be found at Lockheed Martin's web site: http://www.lockheedmartin.com/wms/findPage.do?dsp=fec&ci=11564&rsbci=0&fti=112&ti=0&sc=400&prfr=true, 2005).
[4] Normand, E., 1993, "Single Event Upsets in Avionics," IEEE Trans. Nucl. Sci., **40**(2), pp. 120–126.
[5] Normand, E., 1996, "Single Event Effects in Avionics," IEEE Trans. Nucl. Sci., **43**(2), pp. 461-474.
[6] Normand, E., 1996, "Single Event Upsets at Ground Level," IEEE Trans. Nucl. Sci., **43**(6), pp. 2742–2750.
[7] Desai, M. N., Deckert, J. C., and Deyst, J. J., Jr., 1979, "Dual-Sensor Failure Identification Using Analytic Redundancy," J. Guid. Control Dyn., **2**(3), pp. 213–220.
[8] Forsberg, K., 2003, "Design Principles of Fly-By-Wire Architectures," Chalmers, University of Technology Report No. 3D with Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden.
[9] Merrill, W. C., 1985, "Sensor Failure Detection for Jet Engines Using Analytical Redundancy," J. Guid. Control Dyn., **8**(6), pp. 673–682.
[10] Diao, Y., and Passino, K. M., 2004, "Fault Diagnosis for a Turbine Engine," Control Eng. Pract., **12**, pp. 1151–1165.
[11] Gertler, J. J., 1988, "Survey of Model-Based Failure Detection and Isolation in Complex Plants," IEEE Control Syst. Mag., **8**(6), pp. 3–11.
[12] Willsky, S., 1976, "A Survey of Design Methods for Failure Detection in Dynamic Systems," Automatica **12**, pp. 601–611.
[13] Schneider, S., Weinhold, N., Ding, S. X., and Rehm, A., 2005, "Parity Space Based FDI-Scheme for Vehicle Lateral Dynamics," *Proceedings of the IEEE Conference on Control Applications*, pp. 1409–1414.

[14] Frank, P. M., 1990, "Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge Based Redundancy: A Survey and Some New Results," Automatica, **26**, pp. 459–474.
[15] Simani, S., 2005, "Identification and Fault Diagnosis of a Simulated Model of an Industrial Gas Turbine," IEEE Transactions on Industrial Informatics, **1**(3), pp. 202—216.
[16] Kyusung, K., Ball, C., and Nwadiogbu, E., 2004, "Fault Diagnosis in Turbine Engines Using Unsupervised Neural Networks Technique," Proc. SPIE, **5439**, pp. 150–158.
[17] Gayme, D., Menon, S., Ball, C., Mukavetz, D., and Nwadiogbu, E., 2003, "Fault Diagnosis in Gas Turbine Engines Using Fuzzy Logic," *SMC'03 Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Vol. 4, pp. 3756–62.
[18] Benso, A., Chiusano, S., Prinetto, P., and Tagliaferri, L., 2000, "A C/C++ Source-To-Source Compiler for Dependable Applications," *Proceedings of the International Conference on Dependable Systems and Networks*, New York.
[19] Nicolescu, B., and Velazco, R., 2003, "Detecting Soft Errors by a Purely Software Approach: Method, Tools and Experimental Results," *Proceedings of the Design, Automation and Test in Europe Conference and Exhibition*, pp. 57—62 suppl.
[20] Hocenski, Z., and Martinovic, G., 1999, "Influence of Software on Fault-Tolerant Microprocessor Control System Dependability," *IEEE International Symposium on Industrial Electronics ISIE'99*, Vol. 3, pp. 0_1–0_3.
[21] Arlat, J., Crouzet, Y., Karlsson, J., Folkesson, P., Fuchs, E., and Leber, G. H., 2003, "Comparison of Physical and Software-Implemented Fault Injection Techniques," IEEE Trans. Comput., **52**(9), pp. 1115–1133.
[22] Vinter, J., Aidemark, J., Folkesson, P., and Karlsson, J., 2001, "Reducing Critical Failures for Control Algorithms Using Executable Assertions and Best Effort Recovery," *Proceedings of the International Conference on Dependable Systems and Networks (DSN-2001)*, Göteborg, Sweden.
[23] Avizienis, A., Laprie, J. C. C., Randell, B., and Landwehr, C., 2004, "Basic Concepts and Taxonomy of Dependable and Secure Computing," IEEE Transactions on Dependable and Secure Computing, **1**(1), pp. 11–33.
[24] Åslund, J., Biteus, J., Frisk, E., Krysander, M., and Nielsen, L., 2005, "A Systematic Inclusion of Diagnosis Performance in Fault Tree Analysis," *16th IFAC World Congress*, Prague.

Unsteady Entropy Measurements in a High-Speed Radial Compressor

M. Mansour
N. Chokani
A. I. Kalfas[1]
R. S. Abhari

LSM, Turbomachinery Laboratory,
Department of Mechanical and Process
Engineering, ETH Zürich,
8092 Zürich, Switzerland

The time-dependent relative entropy field at the impeller exit of a centrifugal compressor is measured. This study is part of a broader effort to develop comprehensive measurement techniques that can be applied in the harsh environment of turbomachines. A miniature unsteady entropy probe (diameter of 1.8 mm) is designed and constructed in the present study. The unsteady entropy probe has two components: a one-sensor fast-response aerodynamic probe and a pair of thin-film gauges. The time-dependent total pressure and total temperature are measured from the fast-response aerodynamic probe and pair of thin-film gauges, respectively. The time-dependent relative entropy derived from these two measurements has a bandwidth of 40 kHz and an uncertainty of ±2 J/kg. The measurements show that for operating Condition A, $\varphi=0.059$ and $\psi=0.478$, the impeller exit flowfield is highly three dimensional. Adjacent to the shroud there are high levels of relative entropy and at the midspan there are low and moderate levels. Independent measurements made with a two-sensor aerodynamic probe show that the high velocity of the flow relative to the casing is responsible for the high relative entropy levels at the shroud. On the other hand, at the midspan, a loss free, jet flow region and a channel wake flow of moderate mixing characterize the flowfield. At both the shroud and midspan, there are strong circumferential variations in the relative entropy. These circumferential variations are much reduced when the centrifugal compressor is operated at operating Condition B, $\varphi=0.0365$ and $\psi=0.54$, near the onset of stall. In this condition, the impeller exit flowfield is less highly skewed; however, the time-averaged relative entropy is higher than at the operating Condition A. The relative entropy measurements with the unsteady entropy probe are thus complementary to other measurements, and more clearly document the losses in the centrifugal compressor. [DOI: 10.1115/1.2799525]

Introduction

A primary goal in the design of turbomachines is to have higher efficiencies and wider operating ranges. Thus, a substantial effort has been made to try and understand the loss mechanisms and their origins in the various components of turbomachines. Although much progress has been made in understanding loss mechanisms, Denton [1], our understanding is still incomplete. The focus of the present work is the structure of the impeller exit flow in a centrifugal compressor. Centrifugal compressors are used in automotive, marine turbocharging and distributed power applications because of their compact design and high stage pressure ratio. At the exit of the centrifugal compressor's impeller, the flow is unsteady, three dimensional, and turbulent. The structure of this complex flowfield is affected by the tip clearance, which in turn affects the compressor efficiency.

The impeller exit flowfield has been examined using various measurement techniques including hot wires, Inoue and Cumpsty [2], pneumatic and fast-response aerodynamic probes, Roduner et al. [3], and laser velocimetery, Strahlecker and Gyarmathy [4], Schleer and Abhari [5]. All measurement techniques have their respective advantages and disadvantages, and thus are complementary. However, it is incontrovertible that only a measurement of entropy can provide a rational measure of loss, Denton [1]. A difficulty, however, is that entropy cannot be measured directly but only inferred from other properties

$$\Delta s = c_p \ln \frac{T}{T_{\text{ref}}} - R \ln \frac{p}{p_{\text{ref}}} \tag{1}$$

This difficulty has limited progress in experimental studies of the loss mechanisms in turbomachines.

The present paper demonstrates the use of a newly designed unsteady entropy probe that is both miniature and robust, and can provide high frequency measurements in the harsh environment of a centrifugal compressor. The measurements made with this new probe provide new insight into the structure of the impeller exit flow. In the next section, the design of the unsteady entropy probe is presented. Then the test facility used for this study is described. A discussion of the measurements then follows. Finally, the paper concludes with a summary and perspective of the ongoing probe development.

Unsteady Entropy Probe

The principal components of the unsteady entropy probe are a fast-response aerodynamic probe (FRAP) and an unsteady total temperature probe, Fig. 1.

The FRAP is a well-established technology in the Turbomachinery Laboratory at the Swiss Federal Institute of Technology Zürich, and thus only a few of its salient details are discussed below. More complete details, including design, construction, and calibration, of the unsteady total temperature probe are presented.

Fast-Response Aerodynamic Probe. The FRAP employs a miniature silicon piezoresistive chip, Fig. 2, that is glued beneath a pressure tap on the probe tip. A constant excitation current I_e is

[1]Current address: Department of Mechanical Engineering, Aristotle University of Tessaloniki, GR-54124 Greece.

Contributed by the International Gas Turbine Institute of ASME for publication in the ASME JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 27, 2007; final manuscript received August 14, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007.

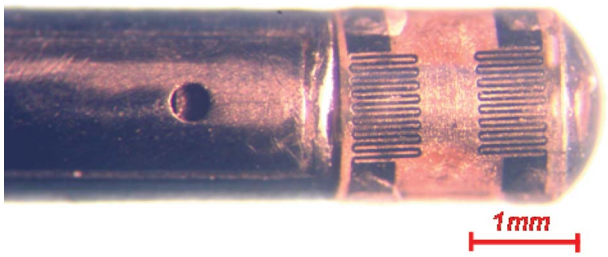


Fig. 1 Photograph of the tip of the unsteady entropy probe

provided to the Wheatstone bridge, which induces an excitation voltage U_e and an output signal voltage U. The output signal is strongly dependent on the pressure differential across the chip and is only weakly dependent on temperature; on the other hand, the excitation voltage is weakly dependent on pressure and more strongly dependent on the temperature. Thus, two-dimensional polynomials, in terms of the excitation and output voltages, are used to yield the sensor calibration relationships for the pressure and temperature. Whereas the frequency response for the pressure is 70 kHz—the limiting factor being the eigenfrequency of the pneumatic cavity that is located between the pressure tap and the diaphragm—an accurate measurement of total temperature is limited to no more than 10 Hz due to the thermal inertia of the chip. More complete details on the FRAP technology are provided by Pfau et al. [6].

Operating Principal: Unsteady Total Temperature Probe. The basic elements of the unsteady total temperature probe are a pair of thin-film gauges that are deposited onto a substrate, Fig. 3, and used as resistance thermometers. The two thin films are mounted azimuthally within an angle of ± 25 deg with respect to the stagnation line, as the Nusselt number is essentially constant over this range [7]. As the heat transfer coefficient α is the same for the two thin films, when the thin-film gauges are electrically heated to two different temperatures T_{f1} and T_{f2},

$$\dot{q}''_{\text{conv1}} = \alpha(T_t - T_{f1})$$

$$\dot{q}''_{\text{conv2}} = \alpha(T_t - T_{f2}) \tag{2}$$

the total temperature of the flow T_t can be determined from

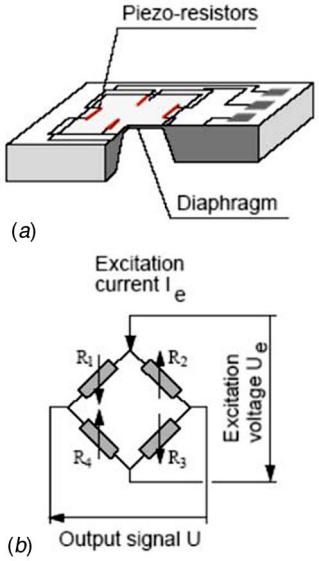


Fig. 2 FRAP: (*a*) miniature silicon piezoresistive pressure sensor chip and (*b*) schematic of Wheatstone bridge

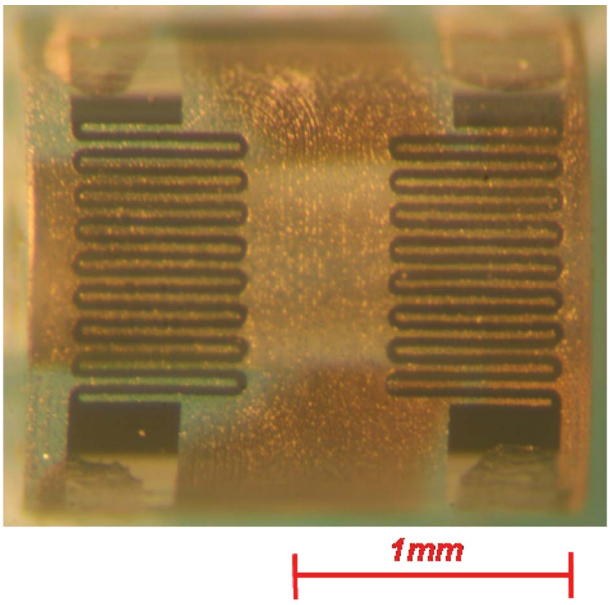


Fig. 3 Unsteady total temperature probe (before integration with the FRAP)

$$T_t = T_{f1} + \frac{\dot{q}''_{\text{conv1}}(T_{f2} - T_{f1})}{\dot{q}''_{\text{conv1}} - \dot{q}''_{\text{conv2}}} \tag{3}$$

This measured temperature is independent of the Reynolds number, Mach number, and the turbulence intensity. Furthermore, the frequency response of the total temperature measurement can be readily made the same as that of the total pressure measured with the FRAP.

The substrate material that is used is fused quartz; its nominal properties are a density of 2203 kg/m^3 and temperature-dependent specific heat and conductivity that are given in Refs. [8,9]. In terms of the thermal and mechanical properties, and adhesive technology, quartz was determined to be the most suitable substrate material. The thin-film gauges are made from 200-nm-thick nickel. In order to assure good adhesion, a 10-nm-thick film of chromium is deposited onto the substrate, prior to sputtering the nickel. As the substrate is a semicylindrical body, the sensor is made by removing nickel using a 20 μm diameter pulsed neodymium-doped yttrium aluminum garnet (Nd:YAG) laser. The resulting serpentine shaped sensor is 30 μm wide and covers a rectangular area of 0.55×0.85 mm^2. Silver leads are deposited at the end of the sensors and run along the flat surface of the semicylinder; then gold wires are wedge bonded to copper wires that are connected to the amplifier.

The diameter of the cylindrical probe is 1.8 mm, and the offset between the thin-film gauges and the pressure sensor is 2.25 mm. Thus although the nondeterministic variation of entropy cannot be measured, by phase locking the deterministic variation in entropy can be quantified.

Calibration: Unsteady Total Temperature Probe. A static calibration, made in an oven, is used to derive the thin-film temperatures of the thin-film gauges that are operated as resistance thermometers. In the calibration, the total temperature probe is exposed to a low speed (5 m/s), isothermal air stream within a channel; the air stream avoids the self-heating of the thin-film gauges and ensures that there is a homogeneous temperature field within the channel. Two thermocouples that are placed in close proximity to the probe are used to measure the temperature of the air stream. The temperature is varied over a range of 20–90°C in steps of 10°C, which corresponds to the temperature range of the present application. The resistances of the thin-film gauges are

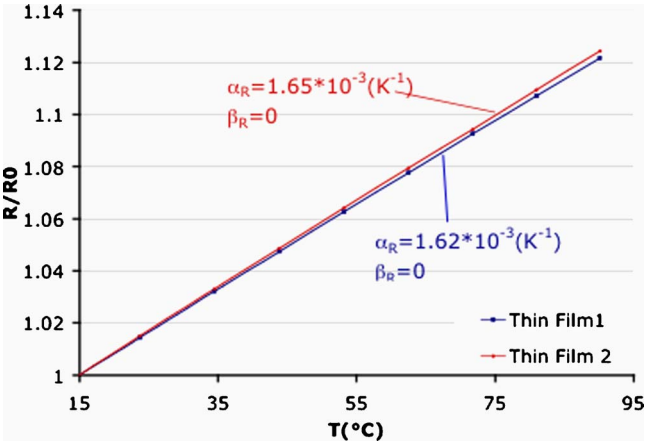


Fig. 4 Static calibration results for unsteady total temperature probe

measured after each increment when thermal equilibrium is achieved. The general form of the temperature-resistance relation is given in Lomas [10]

$$\frac{R}{R_{\text{ref}}} = 1 + \alpha_R(T - T_{\text{ref}}) + \beta_R(T - T_{\text{ref}})^2 \tag{4}$$

Typical calibration results are shown in Fig. 4. It can be seen that the slope is linear for the two thin-film gauges; thus the quadratic coefficient β_R can be neglected. The temperature coefficient of resistance α_R is approximately the same for both thin-film gauges, indicating that the different steps of the manufacturing process are both well controlled and repeatable.

Data Reduction: Unsteady Total Temperature Probe. During the measurements, the thin films are operated at two different constant feeding currents, the Joulean heating of the thin-film gauges is balanced by the conductive and convective heat losses. Thus, the convective heat transfer is given by

$$\dot{q}''_{\text{conv}} = VI - \dot{q}''_{\text{cond}} \tag{5}$$

An electrical circuit is used to pass small constant currents through the thin-film gauges and to measure the corresponding voltages across the thin-film gauges. Thus, the first term on the right-hand side of Eq. (5) is determined from the measured electrical values. On the other hand, the second term on the right-hand side of Eq. (5) is estimated using an unsteady, semi-infinite, heat conduction model; this one-dimensional model is based on the work of Skinner [11], Schultz and Jones [12], Oldfield et al. [13], and Epstein et al. [14]. In this model, the conductive heat transfer rates are solved for using an electrical analog that is composed of a discrete, finite length, series *RC* network. Thus, in terms of the electrical analog, the thermal diffusion equation is written as

$$RC\frac{\partial v}{\partial t} = \frac{\partial^2 v}{\partial x^2} \tag{6}$$

where the voltage v is analogous to temperature T, the capacitance C is analogous to c/ρ, and the resistance R is analogous to $1/k$. The series resistances and capacitances are weighted logarithmically in a manner that is analogous to the logarithmic spacing between nodes that are "placed" across the thermal penetration depth within the substrate. Therefore, an order-of-magnitude fewer series resistances and capacitances can be used to accurately predict the conductive losses than if a linear weighting is used. The discretized set of equations

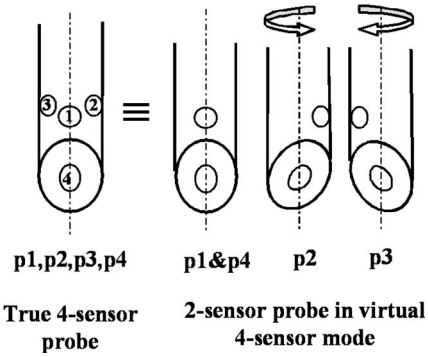


Fig. 5 Measurement concept of a FRAP two sensor in a virtual four-sensor mode

$$\frac{\partial v_i}{\partial t} = \frac{1}{RC}\left[\frac{v_{i-1}}{\Delta x_i \Delta x_{i-1}} - \frac{v_i}{\Delta x_i}\left(\frac{1}{\Delta x_i} + \frac{1}{\Delta x_{i+1}}\right) + \frac{v_{i-1}}{\Delta x_i \Delta x_{i+1}}\right] \tag{7}$$

are solved using a fourth-order, Runge-Kutta method. The boundary conditions for Eq. (6) are derived from the temperatures of the thin-film gauges at the surface and the assumption of a zero heat flux at the depth of thermal penetration. A linear temperature distribution across the thermal penetration depth is assumed in order to provide an initial condition for Eq. (6). The frequency response of the conduction model determines the frequency response of the total temperature measurements, since the electrical circuit has a bandwidth of 120 kHz. The bandwidth of the model used in the present work covers 0.43 Hz–120 kHz with a phase shift of less than 5 deg, which ensures that frequency response of the measured total temperatures is the same as that of the total pressures.

Two-Sensor Fast-Response Aerodynamic Probe

A two-sensor FRAP (2S-FRAP) was used to measure the unsteady velocity field. Pfau et al. [6] describe in detail the use of the virtual four-sensor FRAP, but for sake of completeness brief details are provided below. The 2S-FRAP is calibrated in a virtual four-sensor mode by turning probe into three positions similar to a four-sensor probe, as shown in Fig. 5. A set of four dimensionless calibration coefficients, given in Eq. (8), is determined from this calibration.

$$K_\varphi = \frac{p_2 - p_3}{p_1 - p_m} \quad K_\gamma = \frac{p_1 - p_4}{p_1 - p_m}$$

$$K_t = \frac{p_{\text{tot}} - p_1}{p_1 - p_m} \quad K_s = \frac{p_1 - p_{\text{stat}}}{p_1 - p_m} \tag{8}$$

where $p_m = (p_2 + p_3)/2$. During the measurements, the probe is used in a virtual four-sensor mode. The flow angles are then determined from the yaw (K_φ) and pitch (K_γ) coefficients, and the velocity from the total pressure (K_t) and static pressure (K_s) coefficients. Thus, the three velocity components can be determined.

Error Analysis: Unsteady Total Entropy. Since entropy can only be inferred from other properties, the measurement uncertainty of the entropy probe can only be evaluated in terms of the errors in these other properties. Four properties, the measured total pressure and total temperature, and their respective reference values, are used to determine the entropy, Eq. (1).

The uncertainty in the total pressure is in the range ± 100 Pa, Pfau et al. [6]. The steady total temperature derived from the FRAP has an uncertainty of ± 0.5 K, Kupferschmied [15]. Several sources contribute to the total temperature derived from the unsteady total temperature probe. The propagation error technique is used to combine the uncertainties in these sources into a composite value. The sources (and relative errors) considered here are the output voltage of the constant current electrical circuit (0.1%),

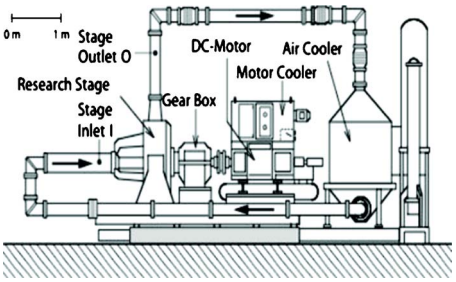


Fig. 6 System layout of the compressor facility "Rigi"

temperature-resistance calibration constants (0.2%), and the materials properties that are used in the conduction model (1.8%). These yield a relative error of 2.5% in the total temperature. The absolute errors on the test rig reference conditions define by Hunziker and Gyrmathy [16] are ±120 Pa for p_{ref} and 0.2 K for T_{ref}.

The relative error in the entropy is expected to be of 2.51%, which is equivalent to an absolute error in the range of ±2 J/kg. This estimated uncertainty is consistent with the reported values for measurements made with the principal components of the unsteady entropy probe.

Test Rig and Measurement Configuration

The test facility is the single stage, centrifugal compressor system, "Rigi," which is located in the Turbomachinery Laboratory. The facility is described in detail in Hunzinker and Gyrmathy [16]. The layout of the system is shown in Fig. 6. It operates in a closed loop with air delivered at a design volume flow rate and a pressure ratio of 3.5 m^3/s and 2.8, respectively.

For the present tests, the system is equipped with a centrifugal impeller followed by a vaneless diffuser. Schleer [17] describes these two components in detail. The impeller is typical of that found in small-scale distributed power generation and automotive turbocharging applications. This scaled-up model matches the main design criteria and nondimensional parameters that are typical of the small-scale devices, and also generates flow structures that are representative of those in small-scale compressors, Schleer et al. [18].

The impeller has seven pairs of full and splitter blades, as shown in Fig. 7. Its outer diameter is 400 mm, and at the impeller exit the blades are swept back by an angle of 30 deg with respect of the radial direction. Immediately downstream of the impeller is a parallel vaneless diffuser with an exit diameter of 580 mm; the diffuser height is 15.7 mm. The clearance ratio, which is defined as the ratio of the tip gap width 0.7 mm and the diffuser height 15.7 mm, is 4.5%. An objective of the present work is to use the newly designed unsteady entropy probe to experimentally quantify details of associated loss generation mechanisms.

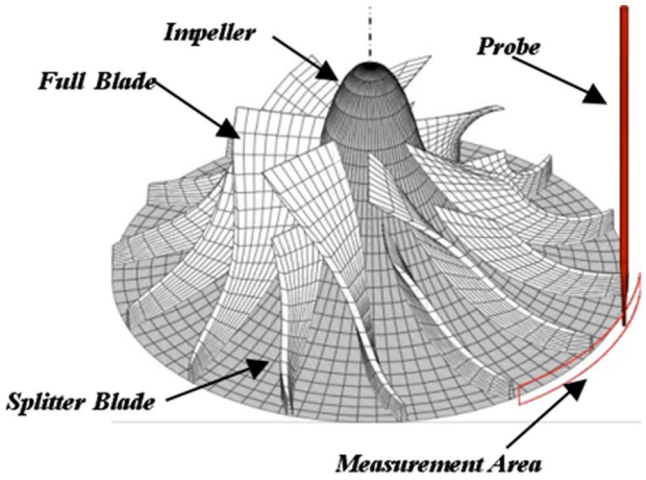


Fig. 7 3D rendering of impeller, showing also location of the unsteady entropy and two-sensor probes and measurement area

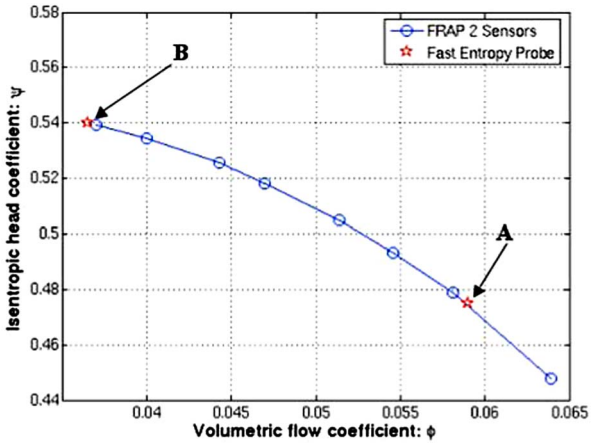


Fig. 8 Operating line for the compressor. A and B are the two conditions at which time-resolved flow measurements are made with the unsteady entropy probe.

The operating line for the compressor at a tip Mach number Ma=0.8 is shown in Fig. 8. The stage inlet reference condition is kept constant to p_{ref}=95,830 Pa and T_{ref}=297 K. Also shown in the plot are the two operating conditions at which unsteady entropy measurements are made. Condition A, φ=0.059 and ψ=0.478, is close to the point of 2 deg flow incidence at the blade's leading edge. Condition B, φ=0.0365 and ψ=0.54, is an operating condition that is close to the onset of stall. The unsteady entropy probe measurements are made at a radial position of R_d/R_{d2}=105% and as illustrated in Fig. 7 are phase locked, using an optical trigger on the shaft, over a full blade passage. A 2S-FRAP is also used to measure the three-dimensional velocity flowfield and flow angularity. Pfau et al. [6] describe the principle of the 2S-FRAP.

The spatial resolution of the measurements using both probes is 0.75 mm in the axial direction, and 0.39 deg azimuthally. At each measurement point, a time series of 400,000 data points is acquired at a sampling rate of 200 kHz using a NI PCI 4452 data acquisition board. All data processing is subsequently done on a laboratory personal computer (PC).

Results and Discussion

As discussed in Ref. [7], the Nusselt number is constant when the thin-film gauges are located on circumferential angles that are within ±25 deg of the stagnation point. Similarly, the calibration of the FRAP is made over a yaw angle range of −30–+30 deg. The circumferentially averaged profiles and phase-locked relative flow angles α_{rel} are shown in Figs. 9 and 10, respectively. The relative flow angle is defined here as the angle between the absolute flow angle (as determined from two-sensor FRAP measurements) and the absolute yaw angle of the unsteady entropy probe. The results in Figs. 9 and 10 show that over the range of application the temperature and pressure measurements are made within the range of validity.

The circumferentially averaged measurements of pressure, temperature, and entropy are presented in Fig. 9 for the two operating conditions. Although the previous laser doppler anemometer (LDA) measurements of Schleer and Abhari [5] show that there are significant flowfield variations in the circumferential direction, the comparison of the circumferentially averaged temperature pro-

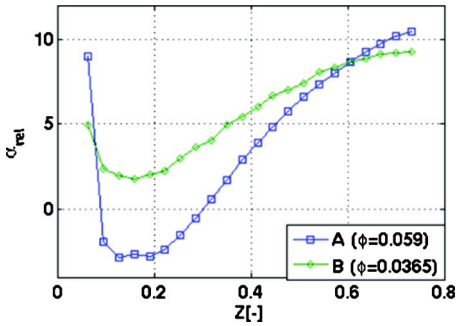


Fig. 9 Circumferentially averaged profiles of relative flow angle α_{rel} at operating Points A (ϕ=0.059) and B (ϕ=0.0365)

files from the piezoresitive chip of the FRAP and thin-film gauges provides one level of assessment of the entropy probe measurements.

As can be seen in Fig. 11, the profiles of the measured temperatures agree very well over the diffuser height. Quantitatively, the differences between the two measurements are no more than 0.8%; this excellent agreement is a validation of the entropy probe's design. Also shown in Fig. 11 is the total pressure distribution that is derived from the FRAP component of the entropy probe. This distribution shows that there are lower total pressures near the shroud wall compared to the hub wall; this suggests that there are more losses in the flow that is adjacent to the shroud. The measured entropy profiles, also shown in Fig. 11, confirm this observation. It should be noted that there are two relative entropy profiles at each operating point, one each derived from the total temperatures measured using the piezoresistive sensor of the FRAP and the thin-film gauges. Both profiles show excellent agreement with each other.

The highly three-dimensional nature of the impeller exit flow is seen in Fig. 12, where the contours of the axial, radial, and tangential velocities are plotted for operating Point A (ϕ=0.059). In

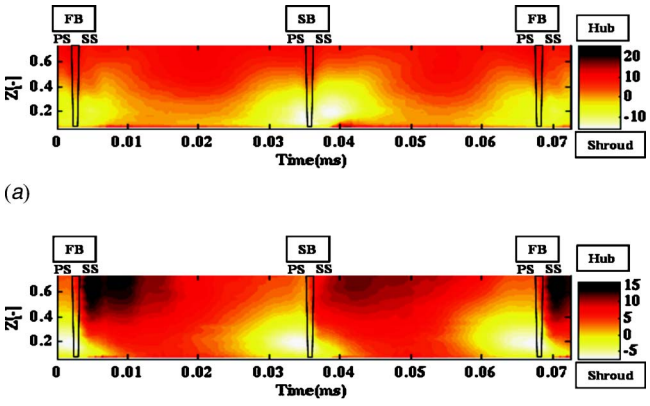


Fig. 10 Time-resolved distribution of relative flow angle α_{rel} (*a*) operating Point A: ϕ=0.059 and (*b*) operating Point B: ϕ=0.0365

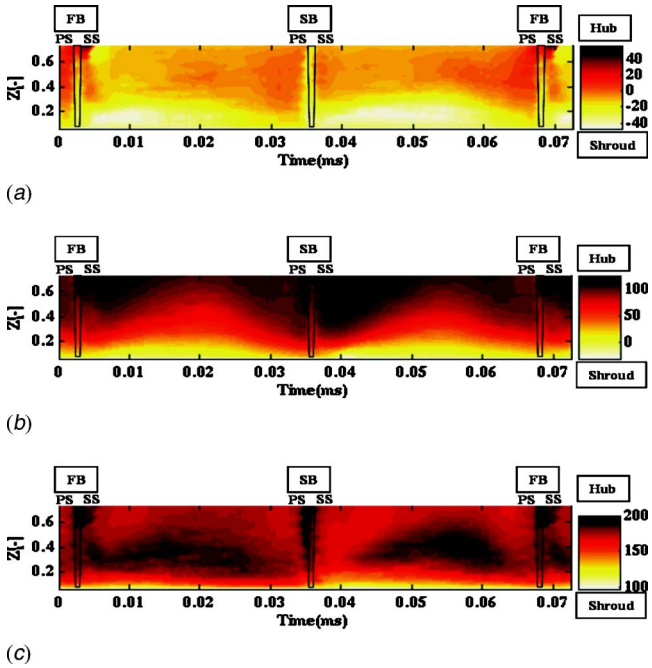


Fig. 12 Time-resolved contours of (*a*) axial velocity, (*b*) radial velocity, and (*c*) tangential velocity, at operating Point A (ϕ=0.059)

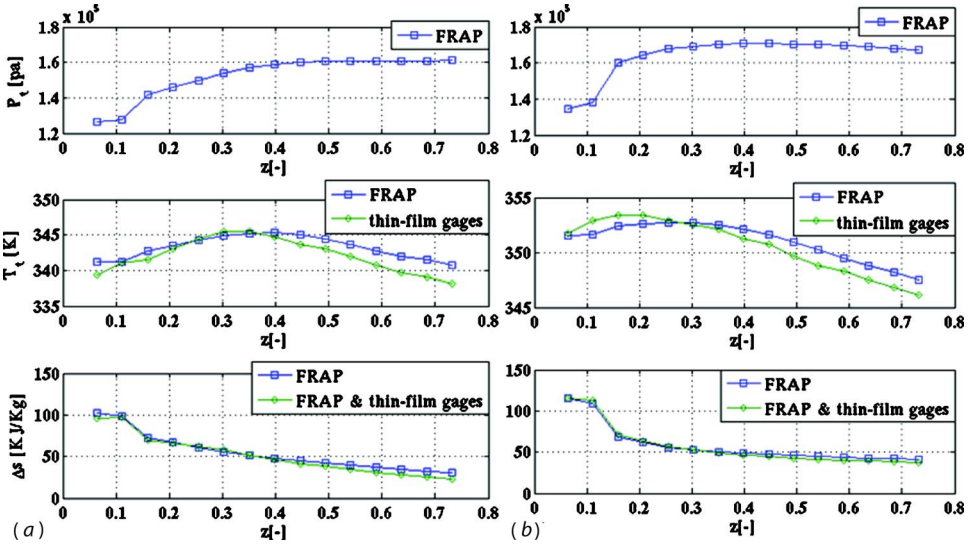


Fig. 11 Circumferentially averaged profiles of total pressure, total temperature, and entropy: (*a*) operating Point A: ϕ=0.059 and (*b*) operating Point B: ϕ=0.0365

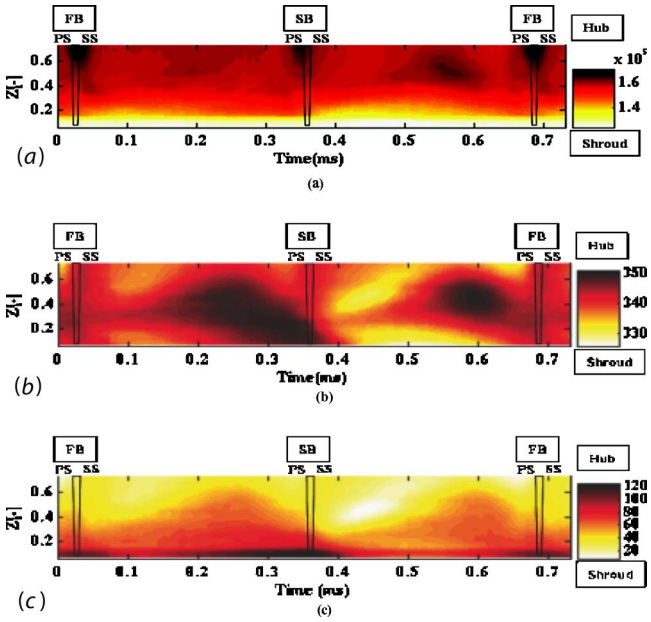


Fig. 13 Time-resolved contours of (*a*) total pressure, (*b*) total temperature, and (*c*) relative entropy, at operating Point A (ϕ =0.059)

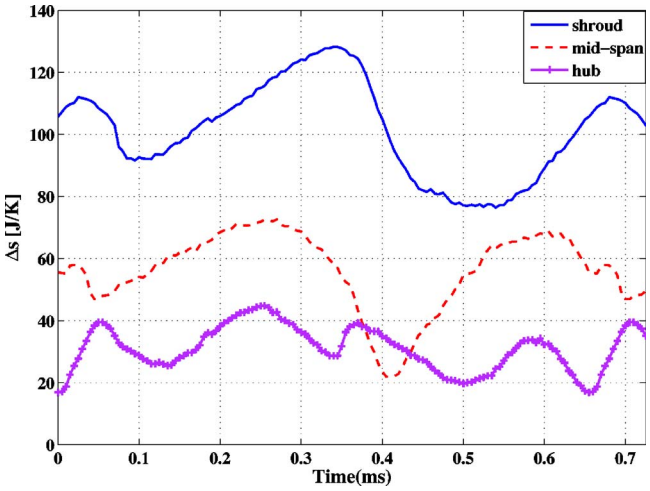


Fig. 14 Circumferential variation of the relative entropy at operating Point A (ϕ=0.059) for three axial positions: *Z*=10% (shroud), *Z*=40% (midspan), and *Z*=70% (hub)

the plots, the blade positions, which are determined from the shaft's encoder signal, are shown. Since the measurements are made downstream of the impeller exit, there is an apparent shift of the flow features relative to the blades. The tangential velocities are relatively high across the height of the diffuser, and relatively low (~100 m/s) tangential velocities are seen only in the near vicinity of the shroud. In this region, adjacent to the shroud, the radial velocities are relatively small in magnitude (±20 m/s) and are alternatively positive and negative along the circumferential direction. The negative radial velocities are seen on the suction side of the blades close to the blade tip, whereas positive velocities are seen on the pressure side. Near the shroud wall, along the blade suction sides, the negative axial velocities are evidence of secondary flows. These secondary flows are more pronounced adjacent to the splitter blade compared to the full blade. Near the hub, the circumferential variation differs from that along the shroud, as relatively high radial velocities occur near the suction side and somewhat smaller radial velocities near the pressure side. Schleer and Abhari [5] have identified this region of high radial and intermediate tangential velocities as a jet flow structure. The intermediate radial velocity/high tangential velocity region that occurs at midspan in the middle of the passage is identified as a channel wake flow. These measurements confirm the highly three-dimensional nature of the impeller's exit flowfield and the profound effect of the tip gap clearance.

The corresponding time-resolved total pressure, total temperature, and relative entropy fields are shown in Fig. 13. The measured total pressures are generally lowered adjacent to the shroud relative to the hub, as suggested by the circumferentially averaged profile shown in Fig. 11(*a*). Figure 13 also shows that there are no strong circumferential variations of the total pressure across the diffuser height. On the other hand, the total temperature field shows that there are gradients both circumferentially and across the diffuser height. The variations are indicative of the skewed nature of the impeller's exit flow. Near the hub on the pressure side of the splitter blade, there is an elongated region of elevated total temperature; on the full blade's pressure side, the region of elevated total temperature is more confined. Skewed regions of relatively low total temperature are seen in the tip clearance region, and this region of low temperature is more pronounced for the splitter blade than for the full blade. These regions of relatively low total temperature are a consequence of upward motion of fluid away from the relative cool shroud wall casing. The heat transfer is especially pronounced in those regions due to the strong secondary flow that is identified in Fig. 12. The contours of the relative entropy field, derived from the total pressure and total temperature measurement, show marked circumferential variations, as a consequence of the total temperature variations. The losses are substantially greater adjacent to the shroud where the high levels of relative entropy are a result of the high velocity of the flow relative to the casing. In the previously identified regions of relatively low total temperature, the levels of relative entropy are not as high due to the effects of the strong secondary flows. These effects are more pronounced at the splitter blade tip. At midspan of the diffuser height, regions of intermediate relative entropy can be seen, and coincide with the channel wake flow that was identified in Fig. 13. On the other hand, the jet flow region has only low level of relative entropy, suggesting that there is little mixing in this region. The circumferential variations of the relative entropy are examined for three axial positions, *Z*=10% (shroud), *Z*=40% (midspan), and *Z*=70% (hub) in Fig. 14. Although the abscissa shows the relative entropy, it is meaningful in this plot to compare the mean and the peak-to-peak variation of three cases. The mean relative entropy increases from the hub toward shroud, as discussed above. It is interesting to note also that while the peak-to-peak range is small (~20 J/kg) at the hub, it is larger and approximately the same (~50 J/kg) at both midspan and shroud positions.

The operating Point B (ϕ=0.0365) that is close to the onset of stall is examined in Figs. 15–17. Similar to the design operating point, examined previously in Fig. 12, the radial velocities in Fig. 15 show that there is a region of reversed flow near the shroud that is caused by the presence of the tip gap.

However, in contrast to the design operating point, the radial velocities in the region between midspan and hub are larger on the pressure side of the blades compared to the velocities on the suction side. Furthermore, the tangential velocities are higher along the suction side compared to the pressure side, which also differs from the design operating point. The classical jet-wake model, first described by Dean and Senoo [19], can be seen in the pattern of radial and tangential velocities.

On the pressure side of the blades, the region of highest radial velocity covers approximately one-third of the passage width and is identified as the jet region. The wake region covers the remaining two-thirds of the passage width, and close to the blades' suc-

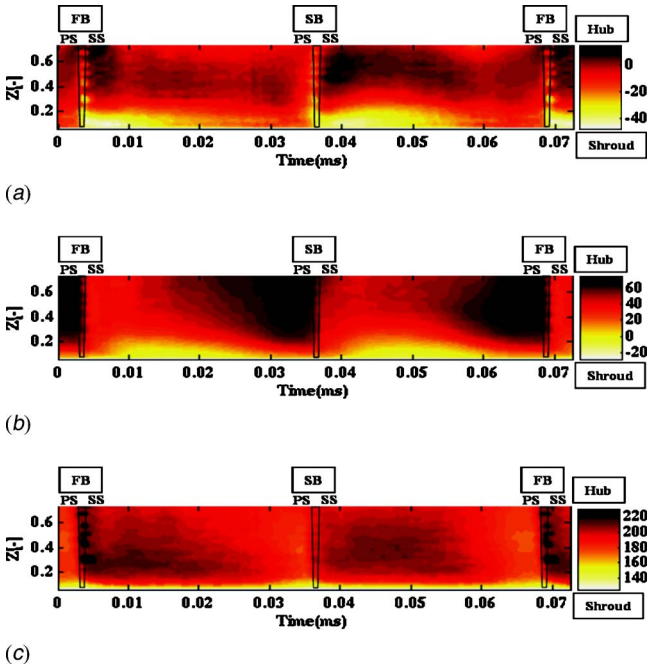


Fig. 15 Time-resolved contours of (*a*) axial velocity, (*b*) radial velocity, and (*c*) tangential velocity, at operating Point B (ϕ =0.0365)

tion side has lower radial velocities and high tangential velocities. The contours of axial velocity show that there are secondary flows along the blades' suction side near the shroud wall. Although the spatial extent of the secondary flow is smaller than that at operating Condition A, shown in Fig. 12, the effect of the flow past the tip of the splitter blade is more pronounced than on the full blade.

Nevertheless, the impeller exit flow is still highly three dimensional, as can be seen from the total temperature field shown in

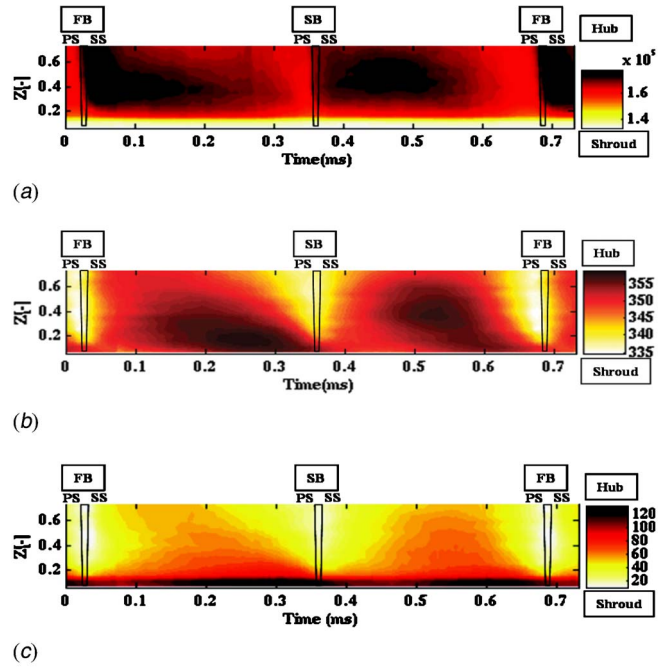


Fig. 16 Time-resolved contours of (*a*) total pressure, (*b*) total temperature, and (*c*) relative entropy, at operating Point B (ϕ =0.0365)

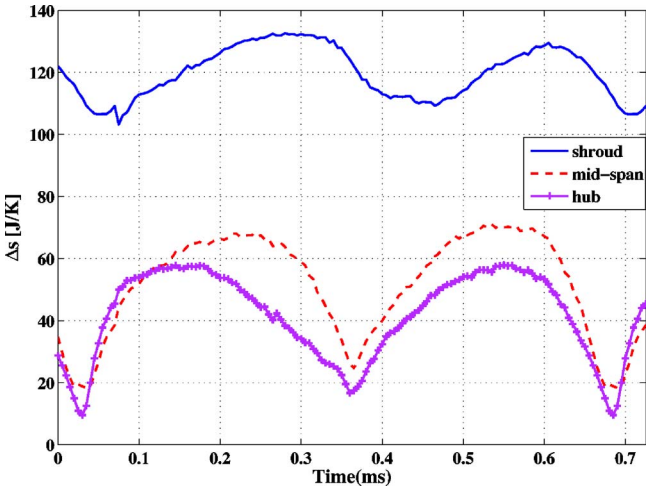


Fig. 17 Circumferential variation of the relative entropy at operating Point B (ϕ=0.0365) for three axial positions: Z=10% (shroud), Z=40% (midspan), and Z=70% (hub)

Fig. 16. Although the flowfield is much less skewed, it is seen that the region of elevated temperature ahead of the full blade is larger compared to that ahead of the splitter blade. The effect of the tip gap on the secondary flow can also be seen in the total temperature distribution. Specifically near the shroud wall there are regions of relatively low total temperature downstream of the blades' suction side. This effect is less pronounced than for operating Point A. In fact, overall, both the total temperature and total pressure flowfields are more uniform for Condition B than those of operating Point A. The time-resolved relative entropy is shown in Fig. 16(*c*). Similar to the design operating point the largest losses occur in the flow that is adjacent to the shroud due to the high flow velocity relative to the casing. The wake region has intermediate levels of relative entropy and the jet region is nearly loss free.

The circumferential distributions of relative entropy at three positions are shown in Fig. 17. It can be seen that the mean values of the relative entropy are similar at the hub and midspan positions. The similar values of relative entropy at midspan and hub confirm that the jet-wake flow spans a large portion of the diffuser height. On the other hand, a higher mean value of the relative entropy is found at the shroud, but the peak to peak fluctuation is less pronounced than for the operating Condition A. This indicates that the tip gap has a smaller effect on the exit flow at the operating Point B. However, the time-averaged relative entropy is 68.2 J/kg, for the Condition B compared to 62 J/kg for the operating condition A. This is due to the fact that the mean values of this relative entropy are in general higher for operating Condition B.

Concluding Remarks

An unsteady entropy probe has been designed, built, and then used to make measurements in a centrifugal compressor. The principal components of the miniature (diameter 1.8 mm) unsteady entropy probe are a FRAP and an unsteady total temperature probe. The FRAP derives total pressure measurements from the electrical output of a miniature silicon piezoresistive chip that is beneath a pressure tap on the probe tip. The total temperature is measured from a pair of thin-film gauges, which are operated as resistance thermometers. An unsteady, semi-infinite, heat conduction model is used to correct the voltages across the thin-film gauges. Thus, high frequency (up to 70 kHz) measurements of the deterministic relative entropy variations of the flow in centrifugal compressor can be made.

The centrifugal compressor is typical of that used in small-scale distributed power generation and automotive turbocharging applications. The impeller has seven pairs of full and splitter blades and is followed immediately downstream by a parallel vaneless diffuser. Phased-locked measurements of the impeller exit flow are made at a fixed radial position. Two operating conditions, one that is near the design point and a second that is close to the onset of stall, are examined. The total pressure, total temperature, and relative entropy measurements of the unsteady entropy probe are complemented with three-dimensional velocities and flow angularity measurements from a 2S-FRAP. At the operating Condition A, the measurements of total pressure and total temperature show that the impeller exit flowfield is highly three dimensional. The total pressures are relatively low at the shroud and relatively high near the hub, on the suction side and pressure side of the blades. On the other hand, the total temperature field is skewed and has gradients both circumferentially and across the diffuser height. Near to the hub on the pressure side of the splitter blade, an elongated region of elevated total temperature is seen and a more confined region of elevated total temperature is seen for the full blade. Therefore, a highly three-dimensional relative entropy field is measured. Adjacent to the shroud high levels of relative entropy are measured, and at the midspan there are regions of low relative entropy interspersed with regions of moderate relative entropy. The independent measurements of the two-sensor aerodynamic probe show that the high velocity of the flow relative to the casing is responsible for the high relative entropy levels at the shroud. On the other hand, at the midspan, the loss free region is associated with a jet flow and the channel wake flow, which has moderate mixing, characterizes the regions of moderate relative entropy. When the centrifugal compressor is operated near the onset of stall, a classical jet-wake model type flow is observed. The wake region, which is close to the blades' suction side, covers approximately two-thirds of the passage width, and has intermediate levels of relative entropy. The jet region, which has the highest radial velocity, is nearly loss free. Similar to the design operating point, the largest relative entropy, and therefore largest losses, occurs in the flow that is adjacent to the shroud; this is due to the high flow velocity relative to the casing. However, although the spatial extent across the diffuser height of the large loss region is smaller at operating Point B than that at operating Condition A, the time-averaged relative entropy is higher for the operating Point B compared to the operating Point A, due to overall higher levels of relative entropy, especially near the hub, where the channel wake flow is still found.

The data from the unsteady entropy probe reveal detailed features of the highly three-dimensional flowfield at the impeller exit of a centrifugal compressor. The steady-state measurements are in very good agreement with the more established FRAP technology. The time-resolved relative entropy measurements are consistent with the velocity component measurements made independently with a 2S-FRAP. The present design of the unsteady entropy probe comprises a single pressure sensor and a pair of thin-film gauges; thus, the velocities must be measured separately from the relative entropy, and analysis is limited to the deterministic fluctuations. A subsequent probe design shall include two pressure sensors with the pair of thin-film gauges; this new design shall enable the simultaneous measurement of the velocity components together with the relative entropy, such that the stochastic fluctuations can be inferred.

Acknowledgment

The authors acknowledge the support of Albert Kammerer in helping to conduct the measurements with the unsteady entropy and 2S-FRAPs. The authors also thank Cornel Reshef for his work in developing the electronic instrumentation and data acquisition system for the unsteady entropy probe, and gratefully acknowledge his continuous support for the development of FRAP and other fast-response measurement techniques at ETH Zürich over the last 20 years.

Nomenclature

c_p = specific heat of the substrate material
D_2 = diameter of impeller tip
FB = full blade
h = enthalpy
I = current
Ma = Mach number at impeller tip $= U_2/\sqrt{\gamma R T_{t,\text{in}}}$
Nu = Nusselt number
p = pressure
PS = pressure side
q = surface heat transfer
Re = Reynolds number
R = gas constant, resistance
Rd = radial position across the impeller
Δs = relative entropy
SB = splitter blade
SS = suction side
t = time
T = temperature
U = voltage signal
U_2 = impeller tip velocity
V, v = voltage
V_0 = inlet velocity
x = distance
Z = axial position measured from shroud to tip

Greek

α = convective heat transfer coefficient
α_R = temperature coefficient of resistance (linear term)
α_{rel} = angle between flow absolute angle and probe zero yaw angle
β_R = temperature coefficient of resistance (quadratic term)
ϕ = specific flow coefficient, $V_0/(D_2^2 U_2)$
ψ = specific work coefficient, $\Delta h_{ts}/U_2^2$

Superscripts

$''$ = per unit of area

Subscripts

2 = impeller outlet
coduction = conduction
convection = convection
$f1$, $f2$ = thin films 1 and 2
rad = radial
θ = tangential
ref = reference condition
s = static
t = total
x = axial

References

[1] Denton, J. D., 1993, "Loss Mechanisms in Turbomachines," ASME J. Turbomach., **115**, pp. 521–652.
[2] Inoue, M., and Cumpsty, N. A., 1984, "Experimental Study of a Centrifugal Impeller Discharge Flow in a Vaneless and Vaned Diffusers," ASME J. Eng. Gas Turbines Power, **106**, pp. 455–467.
[3] Roduner, C., Köppel, P., Kupferschmied, P., and Gyarmathy, G., 1999, "Comparison of Measurement Data at the Impeller Exit of a Centrifugal Compressor Measured With Both Pneumatic and Fast-Response Probes," ASME J. Turbomach., **121**, pp. 609–618.
[4] Strahlecker, D., and Gyarmathy, G., 1998, "Investigations of Turbulent Flow in a Centrifugal Compressor Vaned Diffuser by 3-Component Laser Velocimetry," ASME Paper No. 98-GT-300.
[5] Schleer, M., and Abhari, R. S., 2006, "Clearance Effects on the Evolution of the Flow in the Vaneless Diffuser of a Centrifugal Compressor at Part Load Condition," ASME Paper No. GT2006–90083.

[6] Pfau, A., Schlienger, J., Kalfas, A. I., and Abhari, R. S., 2003, "Unsteady 3-Dimensional Flow Measurement Using a Miniature Virtual 4 Sensor Fast Response Aerodynamic Probe (FRAP)," ASME Paper No. GT2003–38128.
[7] Schmidt, E., and Werner, K., 1941, "Wärmeabgabe über den Umfang eines Angebblasenen Geheizten Zylinders," Forsch. Ingenieurwes., **12**, pp. 65–73.
[8] Hartunian, R. A., and Varwig, R. I., 1962, "On Thin-Film Heat-Transfer Measurements in Shock Tubes and Shock Tunnels," Phys. Fluids, **5**, pp. 169–174.
[9] 1970, *Thermophysical Properties of Matter*, The TPRC Data Series: Thermal Conductivity Nonmetallic Solids Vol. 2, Y. S. Touloukian and C. Y. Ho, eds., IFI/Plenum, New York; Specific Heat Nonmetallic Solids, Vol. 5.
[10] Lomas, C. G., 1986, *Fundamentals of Hot Wire Anemometry*, Cambridge University Press, London.
[11] Skinner, G. T., 1960, "Analogue Network to Convert Surface Temperature to Heat Flux," American Rocket Society Journal, **30**, pp. 569–570.
[12] Schultz, D. L., and Jones, T. V., 1973, "Heat Transfer Measurements in Short Duration Facility Hypersonic Facilities," Agardograph No. 165.
[13] Oldfield, M. L. G., Burd, H. J., and Doe, N. G., 1982, "Design of Wide Bandwidth Analogue Circuits for Heat Transfer Instrumentation in Transient Wind Tunnels," *16th Symposium of International Center for Heat and Mass Transfer*, Hemisphere, Washington, DC, pp. 233–257.
[14] Epstein, A. H., Guenette, G. R., Norton, R. J. G., and Yuzhang, C., 1986, "High-Frequency Response Heat-Flux Gauge," Rev. Sci. Instrum., **57**, pp. 639–649.
[15] Kupferschmied, P., 1998, Zur Methodik Zeitaufgelöstter Messungen mit Strömungssonden in Verdichtern und Turbinen, Diss ETH No. 12474.
[16] Hunziker, R., and Gyrmathy, G., 1993, "The Operational Stability of a Centrifugal Compressor and Its Dependence on the Characteristics of the Subcomponents," ASME Paper 93-GT-284.
[17] Schleer, M. W., 2006, "Flow Structure and Stability of a Turbocharger Centrifugal Compressor," Diss ETH No. 16605.
[18] Schleer, M., Mokulys, T., and Abhari, R. S., 2003, "Design of a High Pressure-Ratio Centrifugal Compressor for Studying Reynolds Number Effects," *International Conference on Compressors and their Systems*, London, pp. 391–404.
[19] Dean, R. C., and Senoo, Y., 1960, "Rotating Wakes in Vaneless Diffusers," ASME J. Basic Eng., **82**, pp. 563–574.

Empirical Tuning of an On-Board Gas Turbine Engine Model for Real-Time Module Performance Estimation

Al Volponi
Pratt & Whitney,
400 Main Street,
East Hartford, CT 06108

Tom Brotherton

Rob Luppold

Intelligent Automation, Inc.,
13029 Danielson Street, Suite 200,
Poway, CA 92064

A practical consideration for implementing a real-time on-board engine component performance tracking system is the development of high fidelity engine models capable of providing a reference level from which performance changes can be trended. Real-time engine models made their advent as state variable models in the mid-1980s, which utilized a piecewise linear model that granted a reasonable representation of the engine during steady state operation and mild transients. Increased processor speeds over the next decade allowed more complex models to be considered, that were a combination of linear and nonlinear physics-based elements. While the latter provided greater fidelity over both transient operation and the engine operational flight envelope, these models could be further improved to provide the high level of accuracy required for long-term performance tracking, as well as address the issue of engine-to-engine variation. Over time, these models may deviate enough from the actual engine being monitored, as a result of improvements made during an engine's life cycle such as hardware modifications, bleed and stator vane schedule alterations, cooling flow adjustments, and the like, that the module performance estimations are inaccurate and often misleading. The process described in this paper will address these shortcomings while maintaining the execution speed required for real-time implementation. [DOI: 10.1115/1.2799527]

Introduction

The current evolution in on-board propulsion health management (PHM) systems, aimed at performing real-time engine diagnostics and prognostics, has placed a greater demand on model accuracy and implementation speed. The complexities in assembling accurate physics-based models for real-time operation has placed greater focus on the use of *hybrid* engine models employing some form of empirical modeling. Such an approach to engine modeling has previously been introduced [1,2] wherein a *hybrid* engine model architecture incorporating both physics-based and empirical components is utilized. This architecture provides a means to automatically tune the engine model to a particular configuration as the engine evolves over the course of its life and furthermore, aligns the model to the particular engine being monitored to insure accurate performance tracking and mitigate the effects of engine-to-engine variation. This paper will discuss the issues and practical considerations for implementing such a self-adapting *hybrid* engine model architecture.

Background

In performance tracking diagnostic systems, an engine model is typically coupled to a performance tracking element that estimates module component performance changes (typically efficiencies and flow parameters) on the basis of gas path measurement parameter residuals that are computed by comparing observed values to those generated by the engine model, which serves as a reference point or baseline. Systems of this type have been in use for several decades and have undergone a tremendous evolution in terms of algorithm development, data management, and user interface. Starting in the 1970s with the work of L. A. Urban [3,4], many developments ensued within the academic and government communities and industry groups. The use of Kalman filters [5–8], neural networks [9,10], fuzzy logic [11], and genetic algorithms [12,13] have been explored for use in the performance fault isolation and tracking process.

Many of these efforts have been focused on (off-board) ground processing of the engine data, where time and memory are manageable commodities. Migration to an on-board system poses real-time constraints and implementation issues not found in off-board systems. The introduction of the full authority digital engine control (FADEC) in the mid-1980s and dedicated PHM electronic boxes in the early 1990s paved the way for insertion of on-board systems [14]. Although there were variants in the individual approaches, the first attempts utilized a simple real-time engine model, typically a state variable model (SVM), coupled with a Kalman filter (KF) observer that together provided an adaptable engine model capable of tracking the monitored engine during steady state and transient operation. The KF observer would act upon the residuals formed by the output of the SVM and the actual observed measurements to provide a set of *tuners* that would adapt the SVM to match the actual observations (hence driving the residuals to zero, on the average). The *tuners* consist of a set of engine module performance parameters such as efficiencies and flow parameters that would allow the engine states and output parameters to be adjusted in order to provide a more faithful match to the actual engine. A typical architecture for such a model is depicted in Fig. 1.

The benefits derived from systems such as this are twofold: (1) Engine parameter synthesis (both measured and nonmeasured) from the engine model, and (2) engine module performance tracking, as provided by the *tuners*. The former being useful for FADEC fault detection and accommodation logic and control law implementation, while the latter has the *potential* to provide valuable performance deterioration information to support engine maintenance operations and logistics as well as providing a means

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 1, 2007; final manuscript received June 4, 2007; published online February 29, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea, and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007.

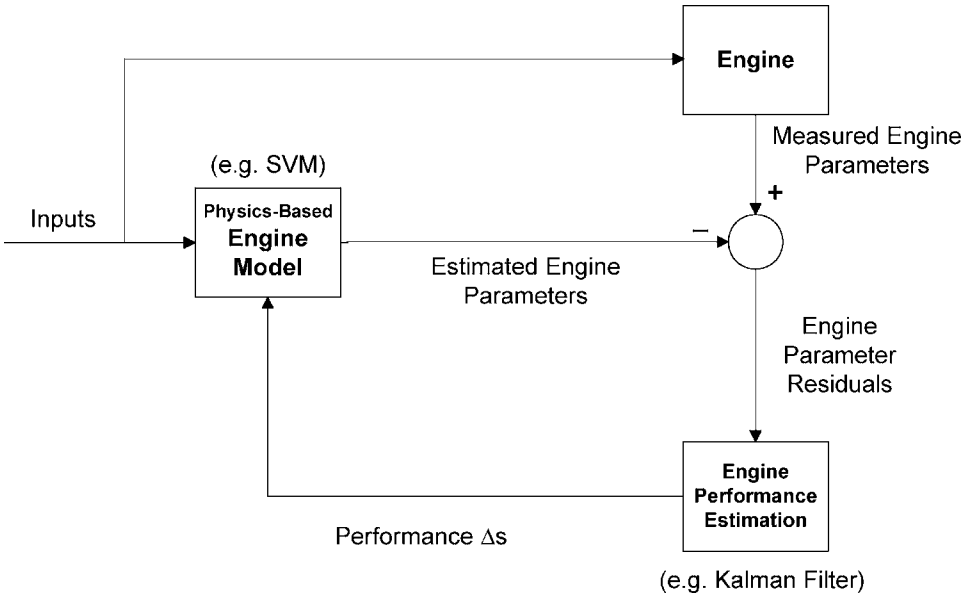


Fig. 1 Typical on-board engine model architecture

to detect performance anomalies on board to enhance flight safety. The word *potential* in the preceding sentence is intentional. A necessary condition to achieve the performance tracking capability is that the (physics-based) engine model provides a faithful representation of the actual monitored engine. When the engine model does *not* provide an accurate representation of the monitored engine, then the residual difference will be interpreted (for the most part erroneously) as changes in module performance and, thus, the *tuners* become a mathematical artifact to close between the model and engine. Figure 2 depicts such a situation. The tuners are module efficiencies from an on-board system configured as illustrated in Fig. 1. The data represents a 3000 sec flight segment of a nominal commercial high bypass commercial turbofan engine. The segment consists of a level flight at 36 K altitude, followed by a gradual climb to 38 K, and level flight thereafter. The ordinate axis represents % Δ efficiency. The large magnitude and erratic behavior of these Δs is a result of an engine model to engine mismatch, and as a result, have little to no value for true performance tracking purposes. For more detail on this phenomenon, the reader is referred to Brotherton et al. [1] and Volponi and Brotherton [15].

To help mitigate these effects, a *hybrid* engine model approach [15] has been suggested wherein the *differences* (or mismatch) between the physics-based model and the monitored engine are empirically modeled. One can think of this as a model initializa-

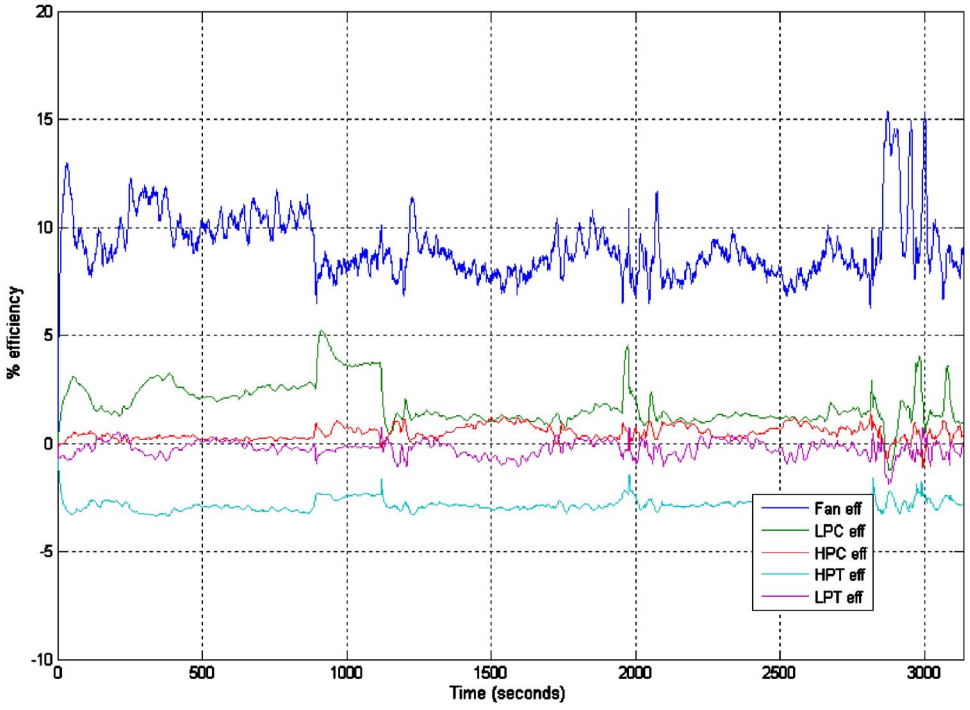


Fig. 2 Example of module performance corruption due to engine model mismatch

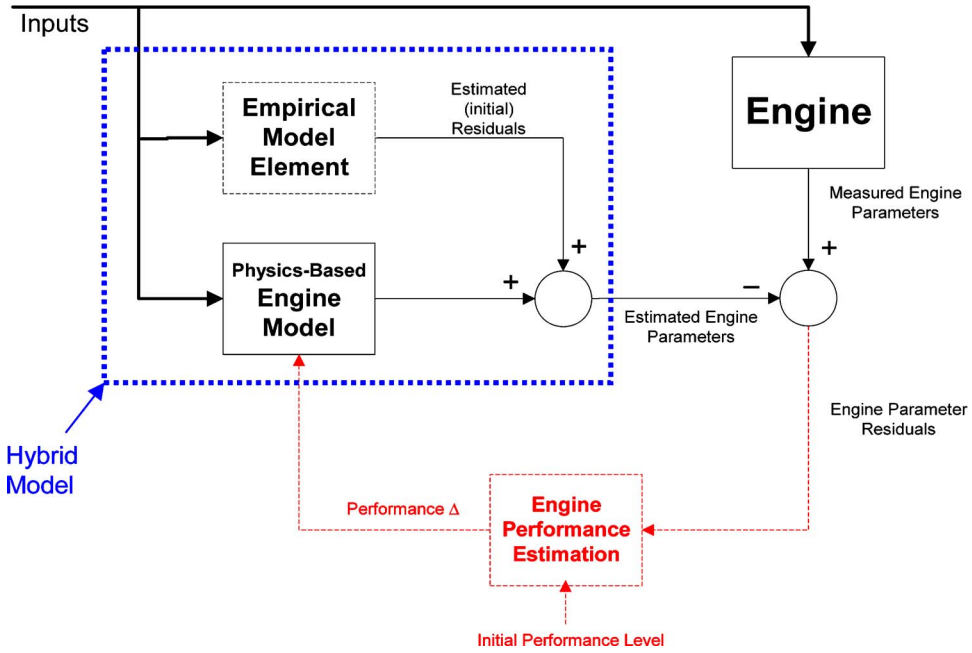


Fig. 3 On-board hybrid engine model architecture

tion process where the resultant model is initialized to represent the particular engine being monitored. This will mitigate the model error effects due to engine-to-engine variation, as well as deficiencies in the physics model due to its simplicity (for real-time operation) as well as model maintenance. The hybrid model configuration is depicted in Fig. 3.

The empirical model element, depicted in Fig. 3, is developed from residuals derived from measured engine outputs and the (physics-based) model predicted outputs. Fusing this empirical model with the physics-based model forms a unique *hybrid* model of the engine. The empirical element can take many forms, including regression models, artificial neural network (ANN) models, etc., however, to be effective, the implementation should be performed *on board* in *real-time* during actual engine operation and flight. This is a crucial enabling implementation requirement in that it forgoes the need to have an extensive data infrastructure in place to handle data transfers (for ground processing) across a fleet of engines, not to mention the logistics of data maintenance and latency in incorporating the final model on an engine-to-engine basis. The primary challenge in developing a *hybrid* model building strategy, therefore, becomes one of developing a methodology and process for capturing the physics model to engine differences within the constraints of limited memory and CPU throughput found in present day FADECs and/or diagnostic monitoring hardware boxes. The remainder of this paper is devoted to describing a particular empirical model-building process that achieves the desired goal and conforms to these constraints.

Empirical Model Building Process Overview

A practical consideration for implementing the proposed *hybrid* engine model involves the application of some form of *sequential* model building for the construction and specification of the empirical model element. This is necessary since the storage and retention of engine and flight input data over a series of flights until such a time that sufficient flight and engine regime data were collected to complete the empirical model would impose an unrealistic requirement in terms of storage capacity for an on-board system. A methodology for incrementally building an empirical model that circumvented the need for storing the *original* engine flight data was reported by Volponi and Brotherton [15]. This process used a flight envelope partition (called *cells*) to create smaller submodels (to decrease the data burden) and introduced a process to incrementally build the empirical submodels in a sequential fashion as flight data became available using computer generated pseudodata. While this methodology decreased the on-board storage requirements, it did not adequately address the CPU throughput issues. The present discussion provides an alternate methodology for creating these empirical submodels (tuning elements) without the use of computer-generated pseudodata. The process, described below, provides a computation (and storage) reduction in excess of 1000:1 over the previous methodology. This savings is essential for a successful real-time implementation of the modeling process.

To represent the empirical model element, we have chosen to use an ANN referred to as a multilayer perceptron (MLP) network [16]. These are standard feed forward (back propagation trained) neural networks. They are commonly used in many engineering applications, since they are known to be universal approximators. Traditionally, MLPs are trained in a batch mode, that is to say, all of the applicable data are collected and the entire data ensemble presented to the neural network for training. In order to effect an incremental model building using these type of networks, the (traditional) training method had to be modified to accommodate this situation and will be described in the sequel.

To facilitate this modeling process, some form of data compression is required. This is essential, since the on-board storage of a flight's worth of data (captured at 10–20 Hz) would impose an unacceptable storage requirement, and the processing time to train an MLP using such a mass of data would be prohibitive. Thus, the approach that was adopted consisted of a two-stage empirical modeling process. The *first* stage is performed in real time during flight operation. This stage performs the flight data reduction necessary to encapsulate a reduced order representation of the residual differences that are to be modeled in the *second* stage. This reduction is accomplished in real time in a sequential clustering process that leverages flight continuity to achieve drastic reductions in time and memory over conventional (generic) clustering algorithms, to create a compressed representation in the form of Gaussian mixture models (GMMs) [17]. The *second* stage is performed on board but off line after the stage 1 processing has been completed. This will typically be performed at the end of the flight and could be triggered by a weight-on-wheels (WOW) indicator.

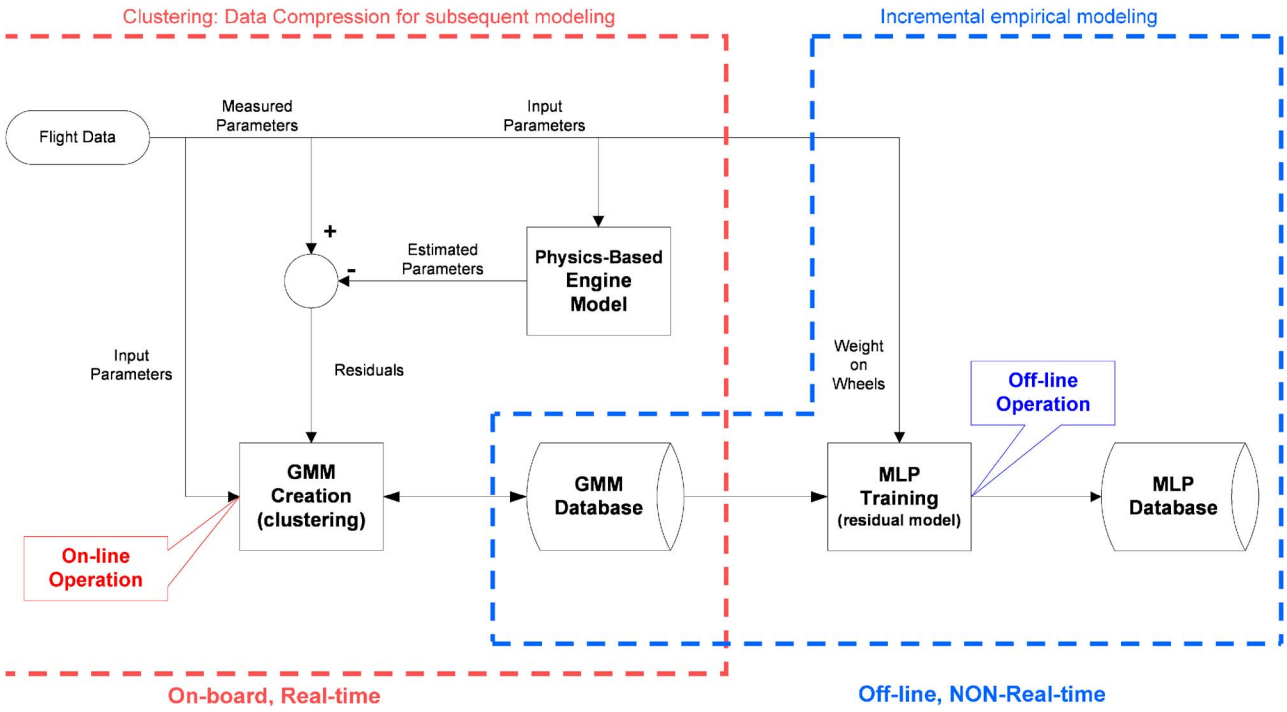


Fig. 4 Two-stage empirical modeling approach overview

This second stage will use the cumulative collection of compact data determined during stage 1 processing from the current and previous flights to determine the residual empirical model (EM). This latter model becomes the empirical element in the *hybrid* model. Figure 4 illustrates the two-stage modeling approach.

The first stage of the process is performed in real time. It is the process of forming the GMMs. In essence, this can be thought of as a data compression phase, wherein clusters of input and residual parameter data are formed in real time during flight and temporarily stored for subsequent processing in the second (off-line) stage of the model building procedure. In addition to forming the (compressed) data clusters, the first stage of this process also performs a myriad of other operations not shown in the overview. These include operations such as (a) regime recognition, i.e., knowing what part of the flight envelope the current data resides (which is important if flight envelope partitioning is used to utilize a collection of submodels in lieu of a large single model [15], and (b) whether the current data point should be part of a new cluster or is indeed already covered by an existing cluster from a previous flight (in which case the current data point is already modeled and the configuration in Fig. 3 can be executed using existing MLPs to track performance changes). Figure 4 depicts only the overview process of GMM creation (real time) and the sequential MLP training (off line, but on board) and does not indicate the intricacies involved in steps (a) and (b) as well as the general control of the *hybrid* model process. A discussion of these issues is beyond the intended scope of this paper and will be deferred to future publications.

The GMMs consist of n-dimensional vectors of engine input and residual parameter averages and standard deviations and associated sample sizes. The input parameters are variables such as altitude, Mach number, ambient temperature and pressure, fuel demand, handling and service engine bleed demands, etc., while the residual parameters are the % differences between the physics-based engine model and the actual engine measurements of parameters in the engine's gas path (interstage temperatures, pressures, spool speeds, etc.). These form a compressed representation of the continuously monitored data. The goal is to define a real-time compression process such that an MLP network subsequently trained on the compressed data would yield essentially the same results as what would be obtained if all of the original data were stored and used to train the MLP network in a traditional batch mode.

Relative to Fig. 4, the steps enclosed within the red dashed lines form the first stage of the process that is performed on board in real time. As each data point is received (typically at 10–20 Hz), its *cell* classification is determined using altitude and Mach number. This defines where in the flight envelope the data resides and what submodel is in effect. Using the *input* parameters (of this data point), a determination is made as to whether the data point should be processed by an existing MLP submodel or whether the current data point represents an area in the flight regime that has *not* been previously modeled. This determination is made by establishing whether or not the n-dimensional input vector falls within a *neighborhood* of an already existing GMM residing in the GMM database. A nearest neighbor criterion, to be described later, is used for this determination. If the data point falls outside any already established GMM, it is a candidate for the formation of a new GMM. The formation of a new GMM is essentially a data clustering process that is performed in real time.

The steps enclosed in the blue dashed lines in Fig. 4 are performed off line (i.e., nonreal time), and are performed on board at the end of the flight. This can be launched by a WOW signal. This is the second stage of the incremental model building process. The GMM database is perused to determine if new GMMs have been added to the database. In this case, on a cell-by-cell basis, new MLPs are generated for the residual parameters, for each *cell* (or submodel) that has new GMMs, and become the current incremental model in the MLP database. Relative to a traditional MLP training methodology that would train directly on stored flight data, the proposed method trains on the GMMs directly. The effect (on the residual) model is essentially the same! Figure 5 highlights the departure from traditional batch training.

GMM Development

Gaussian mixture modeling concerns modeling a statistical distribution by a mixture (or weighted sum) of individual Gaussian

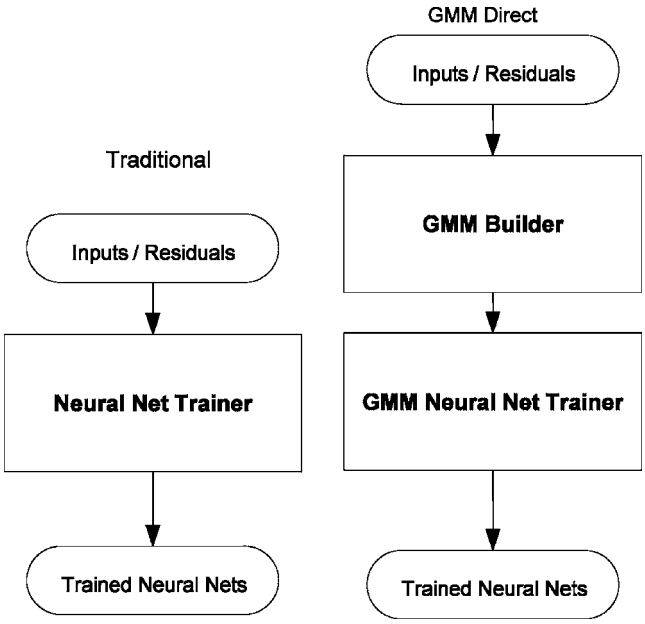


Fig. 5 MLP training overview

distributions. GMM model development has several steps: (1) Determine the number of clusters to break the input data into, (2) determine the cluster centers and which points are associated with the center and (3) determine cluster statistics.

Traditional clustering algorithms, for example *k-means* [18], assume the points to be clustered are independent. Those algorithms are time consuming. The method developed here, an iterative clustering algorithm, takes advantage of the fact that points sequential in time tend to be in the same cluster. This assumption is true due to the physics of how the aircraft and engine traverse through a flight envelope. Figure 6 shows a high level flow diagram for the algorithm. To determine if a new point falls within the current cluster, a hypothesis test is performed. Under the Gaussian assumption, the hypothesis test simply computes the *Mahalanobis* distance (the standard deviation weighted distance) to the cluster center. If that distance is "close" then the test sample is included with the current cluster of points. Incremental estimates of the cluster multidimensional mean and standard deviation are updated as required. To reduce computational complexity and storage, only the main diagonal of the data covariance matrix is stored.

If the new time point falls outside of an existing cluster, it is the first point of a new, *to be determined*, cluster and a new cluster of points is started. The new cluster is assumed to contain some minimum number of points, so that the first $N_{\min}$ points are used to form an initial estimate of the cluster multidimensional mean and standard deviation. $N_{\min}$ is set to different values depending on whether the engine (and aircraft) is in a steady state or transient condition. For steady state, $N_{\min}$ is large while for a transient condition it is relatively small. The procedure now continues with the next new time sample. Smoothing is performed on the various estimates so that outlier points do not cause spurious clusters to be performed.

MLP Neural Network Training

Neural network training is performed by adjusting the various parameters in the neural network to optimize some performance functional. For standard training, this is simply the sum-of-squares error (SSE) defined in Eq. (1) as

$$\mathrm{SSE} = \sum_{k=1}^{N} |r(k) - F(u(k))|^2 \tag{1}$$

where $r(k)$ is the kth point of the residual sequence we are trying to predict, $u(k)$ is the kth point of the multidimensional input data set, and $F(\bullet)$ is the mapping function approximated by the (MLP) neural net. There are N points in the training data set.

Assume that the GMM model is made up of C clusters. Let $M.n(i)$ be the number of points contained in cluster i. Let $M.c(i)$ represent the center of cluster i with associated standard deviation $M.s(i)$. Taking into account these cluster parameters, an *equivalent* SSE error function to be used for training can be derived as given in Eq. (2) below

$$\mathrm{SSE} = \frac{1}{N}\sum_{i=1}^{C} M.n(i) \left| \frac{M.c_r(i) - \hat{F}(M.c_u(i))}{M.s_r(i)} \right|^2 \tag{2}$$

where

$$N = \sum_{i=1}^{C} M.n(i) = \text{total number of points}$$

and the "r" and "u" subscripts refer to the components in the ith cluster that correspond to the residual and inputs respectively, i.e.,

$$M.c(i) = \{M.c_u(i), M.c_r(i)\}$$

This is the only change required in the trainer. Using the GMM method, data compression in excess of 1000:1 can be achieved with a 500× reduction in training time with no appreciable loss in statistical performance.

Example

Returning to the example of Fig. 2, we consider a high bypass commercial turbofan cruising at 36 k altitude, climbing to 38 k altitude, and then cruising at this new altitude condition. Streaming data captured at 20 Hz is used in this example. The flight altitude and Mach number are depicted in Fig. 7.

The input parameters (u) for this flight segment consisted of a total of eight variables; altitude, Mach, low rotor speed (N1), variable compressor vane angle, several engine bleed commands (B25, B8), active (turbine) case clearance control, and an air-oil cooling command. These are depicted in Fig. 8. (Note: The ordinate axis has been removed in this plot to preserve the proprietary nature of the engine data.) The residual parameters for eight gas path outputs (rotor speeds and interstage temperatures and pressures) were calculated from a comparison of the engine measurements and a real-time SVM of the engine and recorded at 20 Hz. The %Δ residuals appear in Fig. 9.

Processing this data through the system depicted in Fig. 1 yields the module performance tracking observed in Fig. 2. Since this engine was known to be of nominal performance, it is clear that it is an engine-model mismatch contributing to the nonzero residuals and attendant module performance Δ behavior. Applying the methodology described in the preceding sections, a set of GMMs was formed. The original 3000+ sec data set consisted of 62,647 points (N) (per parameter) collected at 20 Hz. The number of GMMs (C) generated to represent this data in compressed form was 38. This represents approximately a 1600:1 reduction. Figure 10 depicts the location of the original data (relative to altitude and Mach) with the superimposed GMM centers. There are eight dimensions for the original data inputs and GMM centers; however, only two are displayed here for convenience. The green dots represent the original data points while the large red dots and magenta squares represent the GMM centers (the latter having in excess of 1000 points in the cluster). The attendant 1 sigma ellipses about the centers are superimposed for reference.

The appearance of the GMM distribution relative to the actual data might appear strange at first glance in that they seem to be

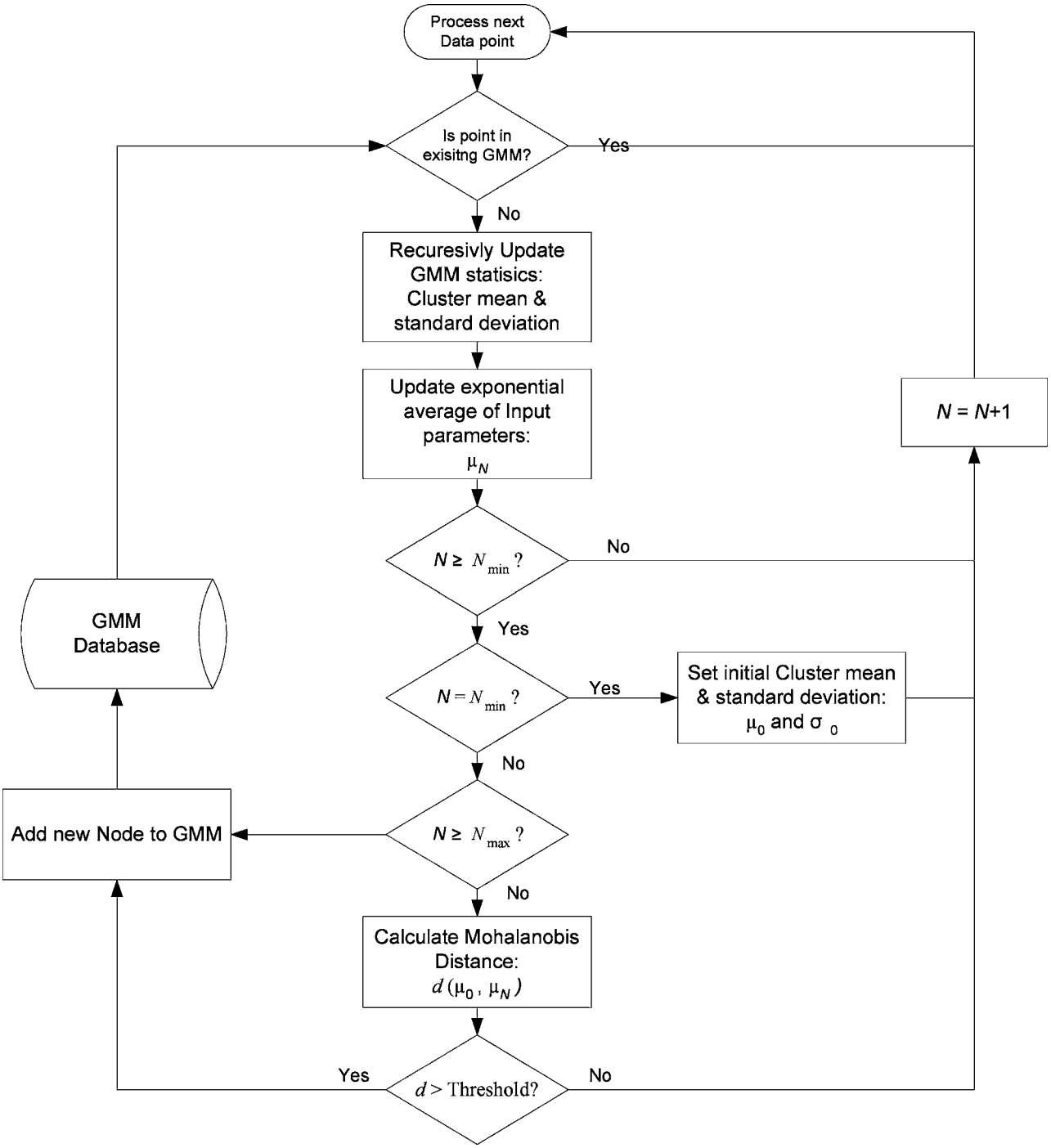


Fig. 6 GMM algorithm overview

clustered together in certain regions and sparse in others. The reason for this is that the GMM centers are multidimensional vectors and only altitude and Mach are depicted. The (seemingly) dense distribution in altitude and Mach is really due to the other six input parameters not shown. In actuality the GMM centers are more nearly uniformly dispersed throughout the eight-dimensional input space.

The 1600:1 compression represented in this segment is rather dramatic, and it would be natural to question the suitability of the GMM formulation and whether or not training an MLP on 38 GMMs (in the second stage of the process) would yield essentially the same result as training on the original 62,647 data points directly. The efficacy of the GMM compression can be tested by running the original data back through the *hybrid* model system depicted in Fig. 3, where the MLP has been trained using the 38 GMMs. The effect on the performance Δs is depicted in Fig. 11.

Although the performance Δs are not identically zero, they exhibit a more reasonable behavior expected from a *nominal* performance engine (compared to Fig. 2). The engine-model mismatch has all but been removed for this flight segment. As the aircraft flies other trajectories, the GMM database would be extended (via the algorithm described above) and the MLP neural networks would be retrained to capture the new regime. Over a short period

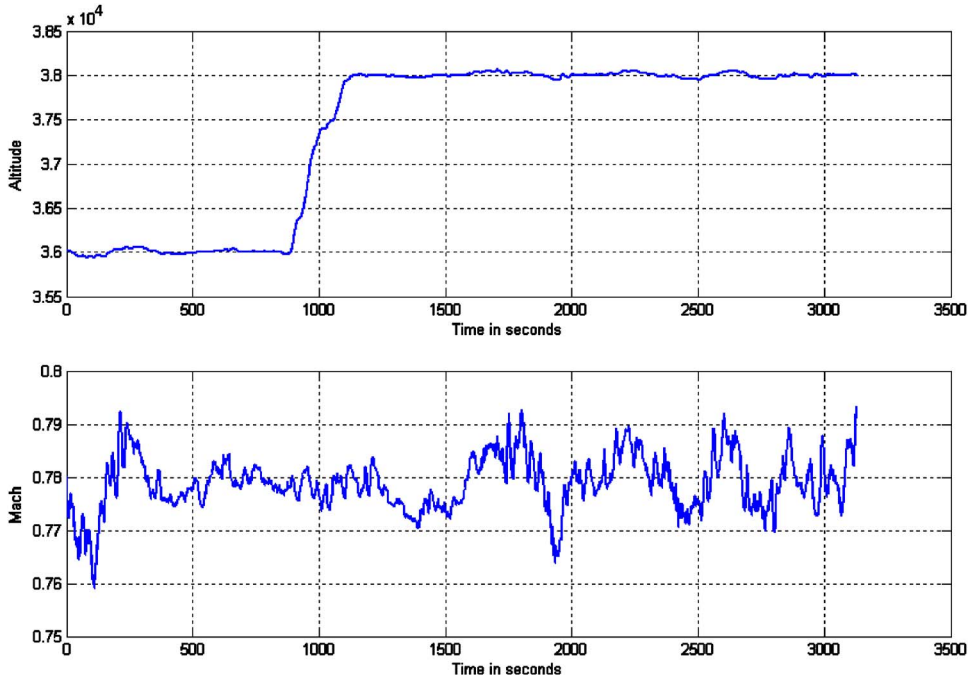


Fig. 7 Example cruise data

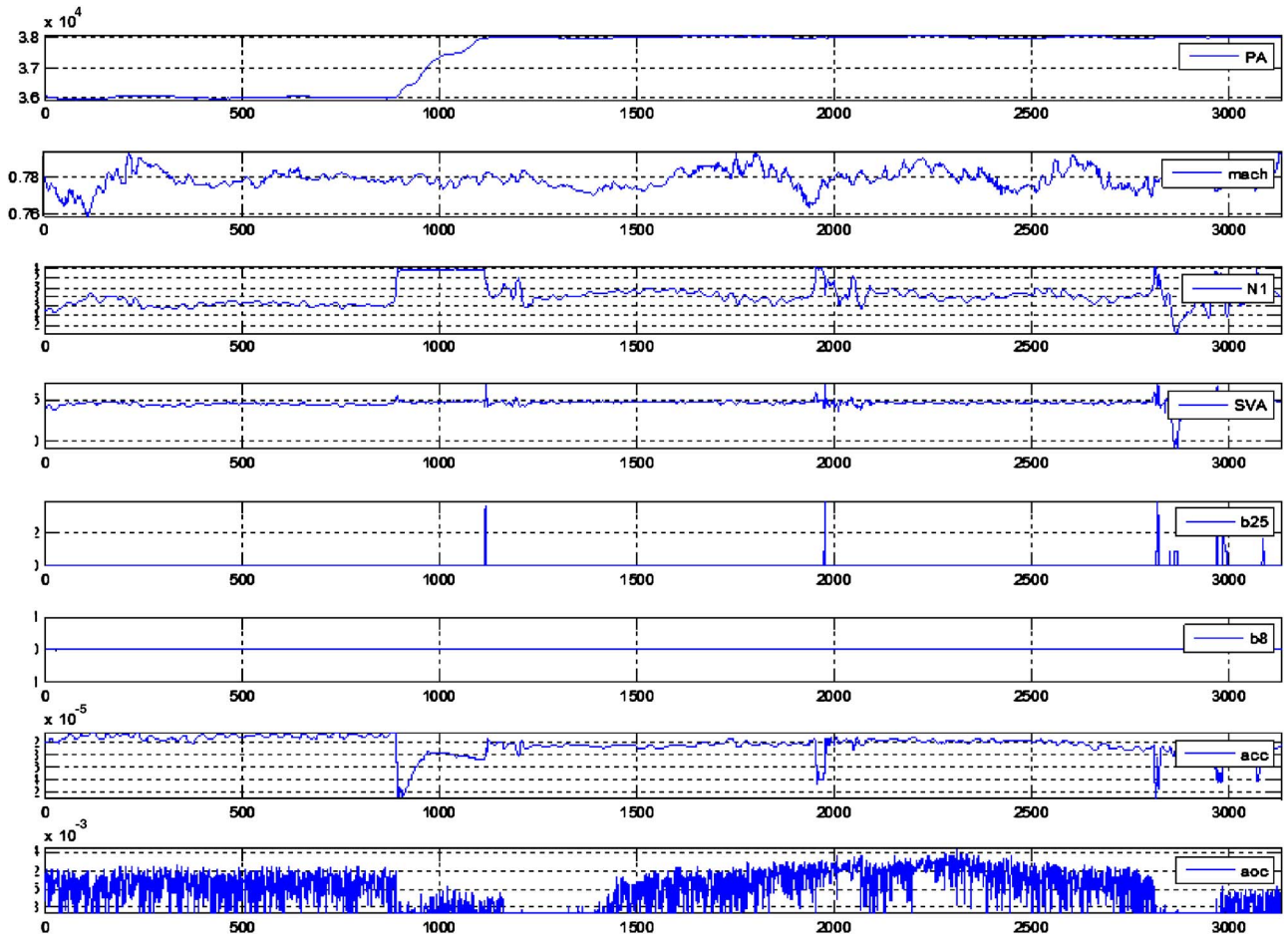


Fig. 8 Example cruise data input parameters

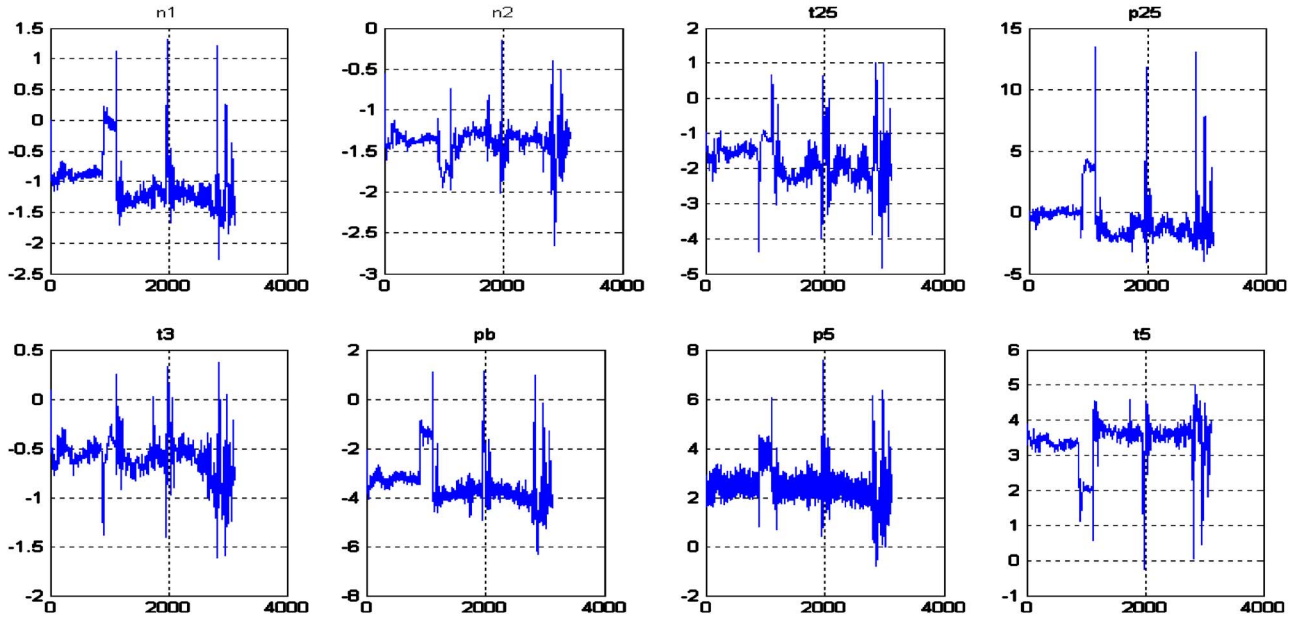


Fig. 9 Example cruise data %Δ residual parameters

of time, the typical flight trajectory of a commercial aeroengine would be captured and modeled and the *hybrid* model would then be complete.

What is significant is the autonomous nature of the process. What needs to be modeled and when the modeling ceases is driven by the data itself and is not predetermined. This allows the process to be deployed with a large variety of aircraft/aeroengine combinations. The process also makes no assumptions relative to the accuracy of the underlying physics-based model whose departure from the monitored engine is being captured and empirically modeled. It should be understood, however, that the empirical model is intended as a *fine-tuning* element in the *hybrid* model strategy and is intended to *augment* the physics-based model, *not* to replace it. A reasonably accurate physics model should always be the primary modeling element. The empirical fine tuning compensates for (physics-based) model deficiencies, engine-to-engine variation, and supports model maintenance over the engine's life cycle.

One of the primary drivers for considering the *hybrid* model approach was to mitigate the effect of model-engine mismatch on performance tracking. The reduction of Fig. 2 to that of Fig. 11 supports this goal. In Fig. 12 we depict a simulated faulty condition to illustrate the point.

The data processing in the example cited above was accom-

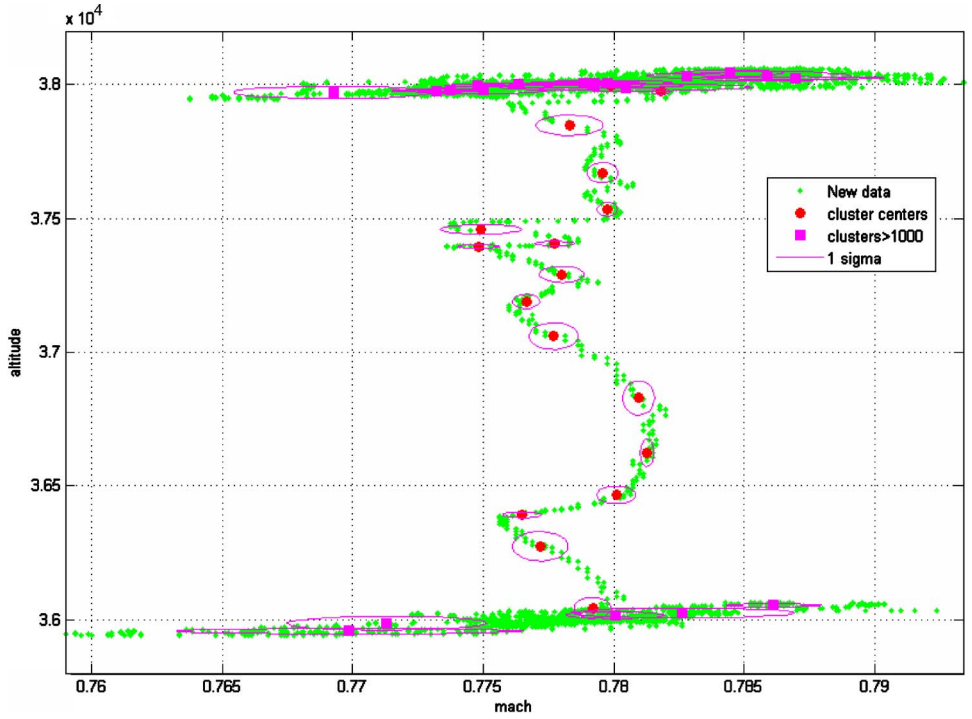


Fig. 10 GMM locations in flight trajectory

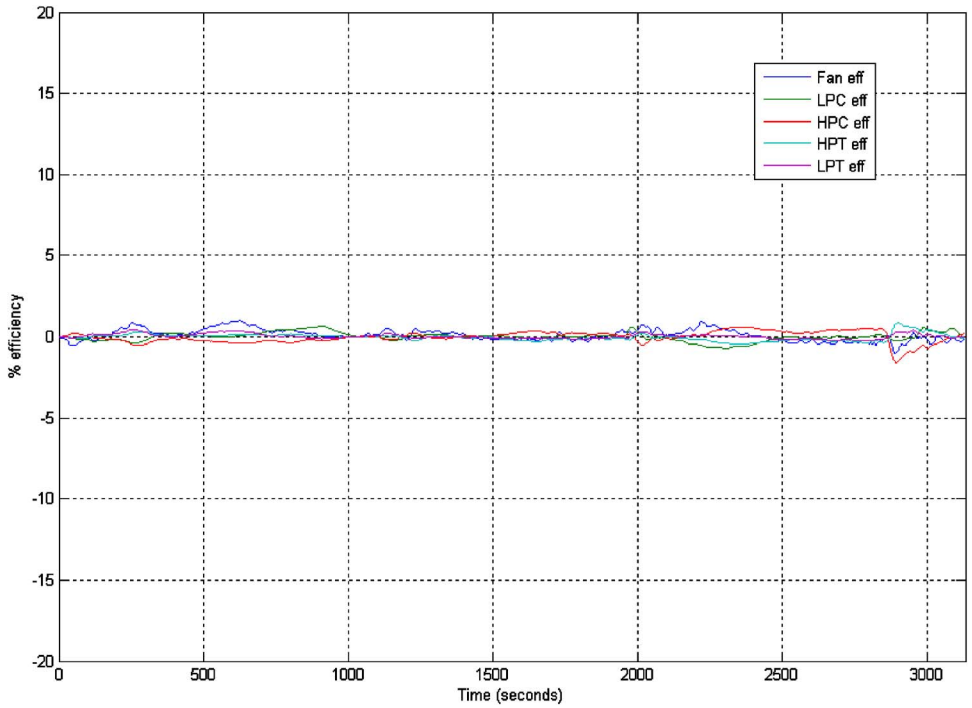


Fig. 11 Performance Δs using *hybrid* model trained with 38 GMMs

plished off board on a test bench using a modification of an existing electronics box that is currently used for aircraft/engine vibration monitoring. Streaming ARINC 429 data was captured beforehand and stored for the 3000 sec flight segment of the commercial high bypass turbofan engine. The hardware, depicted in Fig. 13, has two PC-104 cards that support an ARINC 429 interface card and a CPU/Ethernet card, respectively. The ARINC 429 interface card was used for reading the stored data from a simulated ARINC 429 bus. The CPU card forms the computation platform within the box. This card includes an Ethernet port for real-time output of the *hybrid* modeling calculations and results for user monitoring and graphics. The computational core is supplied by a 233 MHz Pentium II compatible processor. Although the processing was constrained to run in real time at the 20 Hz data rate, benchmark testing indicated that the calculations described in

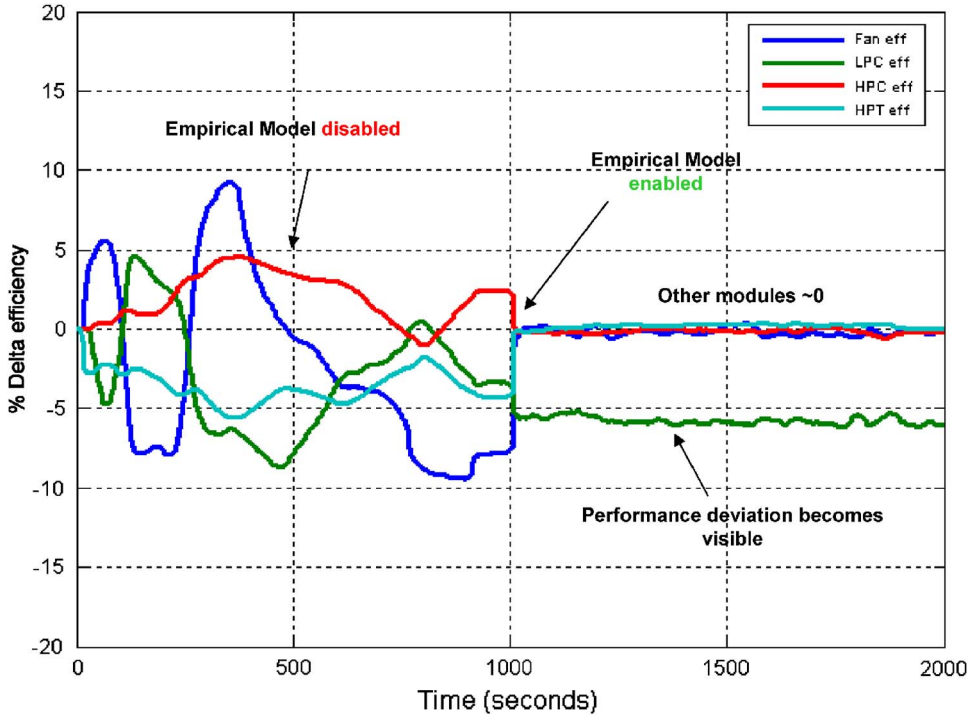


Fig. 12 Hybrid model effect on performance tracking

Front

Technical Specification Overview

Temperature Range: -40C to +55C
Input Power: 6-40VDC, MIL-STD-704D
Power: <18 W
Dimensions: 3" x 5" x 9" w.o. mounting ears
Weight: 4lbs 5oz
Mounting: via mounting plate
Interconnection: MIL-STD 38999 connectors
Remote front panel interface

Processor: Pentium compatible 233 MHz
SDRAM: 128Mbyte
DiskOnChip: 96Mbyte
Compatibility: PC/AT compatibility, AVA
Com. Channels: Ethernet, ARINC 429

Fig. 13 Hardware system summary

this paper, if unconstrained, would run approximately 50× real time in this hardware configuration.

Summary

The concept of a *hybrid* engine model has been introduced and discussed. A pivotal requirement for successful implementation of such systems (containing empirical model elements) is the ability to perform the empirical modeling task in a manner that imposes little additional burden in terms of data manipulation, infrastructure, memory, and computation. A two-stage on-board empirical modeling strategy has been presented (in overview) that supports autonomous real-time model derivation and subsequent model operation. *Hybrid* models show great promise in providing increased accuracy for parameter synthesis and module performance tracking.

Acknowledgment

NASA Glenn Research Center, and Pratt & Whitney are gratefully acknowledged for their funding support for this work. In particular, we wish to thank the COTR, Don Simon (GRC) for his continued interest, support, and program flexibility over the years as this technology evolved.

Nomenclature

ANN = artificial neural network
Cells = flight envelope partition
EM = empirical model
FADEC = full authority digital engine control
GMM = Gaussian mixture model
MPA = module performance analysis
MLP = multilayer perceptron neural network
PHM = prognostic health management
WOW = weight on wheels
SVM = state variable model

References

[1] Brotherton, T., Volponi, A., Luppold, R., and Simon, D., 2003, "eSTORM: Enhanced Self Tuning On-Board Real-Time Engine Model," Proceedings of the 2003 IEEE Aerospace Conference, Big Sky, MT, March, Paper No. 1023.
[2] Volponi, A., 2005, "Use of Hybrid Engine Modeling for On-Board Module Performance Tracking," ASME Paper No. GT2005-68169.
[3] Urban, L. A., 1972, "Gas Path Analysis Applied to Turbine Engine Conditioning Monitoring," AIAA/SAE Paper No. 72-1082.
[4] Urban, L. A., 1974, "Parameter Selection for Multiple Fault Diagnostics of Gas Turbine Engines," *AGARD Conference Proceedings No. 165*.
[5] Volponi, A. J., and Urban, L. A., 1992, "Mathematical Methods of Relative Engine Performance Diagnostics," SAE Tech. Pap. Ser., **101**, 922048.
[6] Doel, D. L., 1992, "TEMPER - A Gas Path Analysis Tool for Commercial Jet Engines," ASME Paper No. 92-GT-315.
[7] Doel, D. L., 1993, "An Assessment of Weighted-Least-Squares Based Gas Path Analysis," ASME Paper No. 93-GT-119.
[8] Kobayashi, T., and Simon, D. L., 2003, "Application of a Bank of Kalman Filters for Aircraft Engine Fault Diagnostics," ASME Paper No. GT-2003-38550.
[9] Kanelopoulos, K., Stamatis, A., and Mathioudakis, K., 1997, "Incorporating Neural Networks Into Gas Turbine Performance Diagnostics," ASME Paper No. 97-GT-035.
[10] Romessis, C., Stamatis, A., and Mathioudakis, K., 2001, "A Parametric Investigation of the Diagnostic Ability of Probabilistic Neural Networks on Turbofan Engines," ASME Paper No. 2001-GT-0011.
[11] Ganguli, R., 2003, "Application of Fuzzy Logic for Fault Isolation of Jet Engines," ASME J. Eng. Gas Turbines Power, **125**(3), pp. 617–623.
[12] Sampath, S., Galuti, A., and Singh, R., 2002, "Fault Diagnostics Using Genetic Algorithm for Advanced Cycle Gas Turbine," ASME Paper No. GT-2002-30021.
[13] Singh, R., and Sampath, S., 2004, "An Integrated Fault Diagnostics Model Using Genetic Algorithm and Neural Networks," ASME Paper No. GT2004-53914.
[14] Luppold, R. H., Gallops, G., Roman, J., and Kerr, L., 1989, "Estimating In-Flight Engine Performance Variations Using Kalman Filter Concepts," AIAA Paper No. 89-2584.
[15] Volponi, A., and Brotherton, T., 2005, "A Bootstrap Data Methodology for Sequential Hybrid Engine Model Building," *2005 IEEE Aerospace Conference Proceedings*, Big Sky, MT, March.
[16] Haykin, S., 1999, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J.
[17] Bisop, C. M., 1995, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, U.K.
[18] Duda, R. O., Hart, P. E., and Stork, D. G., 2001, *Pattern Classification*, 2nd ed., John Wiley & Sons, Inc., New York.

A Sensor-Fault-Tolerant Diagnosis Tool Based on a Quadratic Programming Approach

S. Borguet
e-mail: s.borguet@ulg.ac.be

O. Léonard
e-mail: o.leonard@ulg.ac.be

Turbomachinery Group,
University of Liège,
Chemin des chevreuils 1,
4000 Liège, Belgium

Kalman filters are widely used in the turbine engine community for health monitoring purpose. This algorithm gives a good estimate of the engine condition provided that the discrepancies between the model prediction and the measurements are zero-mean, white random variables. However, this assumption is not verified when instrumentation (sensor) faults occur. As a result, the identified health parameters tend to diverge from their actual values, which strongly deteriorates the diagnosis. The purpose of this contribution is to blend robustness against sensor faults into a tool for performance monitoring of jet engines. To this end, a robust estimation approach is considered and a sensor-fault detection and isolation module is derived. It relies on a quadratic program to estimate the sensor faults and is integrated easily with the original diagnosis tool. The improvements brought by this robust estimation approach are highlighted through a series of typical test cases that may be encountered on current turbine engines. [DOI: 10.1115/1.2772637]

Keywords: engine health monitoring, robust estimation, sensor faults

Introduction

In the last four decades, many research efforts have been conducted to develop efficient fault detection and isolation techniques for gas turbine engines. Generating a precise information about the health condition of the engine leads to improved safety and reliability while reducing operating costs.

The diagnosis tool considered herein is a *gas path analysis* method whose purpose is to assess the changes in engine module performance, described by so-called health parameters, on the basis of measurements collected along the gas path of the engine [1]. Typically, the health parameters are correcting factors on the efficiency and the flow capacity of the components (fan, lpc, hpc, hpt, lpt, nozzle) while the measurements are intercomponent temperatures, pressures, and shaft speeds.

Besides component faults, sensor (instrumentation) faults may occur. Generally speaking, a sensor fault is data generated by a sensor whose behavior no longer respect the manufacturer characteristics. Instrumentation faults can be classified as systematic errors, such as biases and drifts, or as impulsive noise, which are random "spikes" in the measured data.

These instrumentation faults can spoil the estimation of the engine condition. It is therefore mandatory to bring robustness against sensor faults in diagnosis tools, which has been the subject of an increasing number of contributions. A short, nonexhaustive review of various techniques investigated in the jet engine community is given below:

1. *Data cleaning/filtering*. It is actually a preprocessing of the measurements, the purpose of which is to remove aberrant values from the samples before they are used in the monitoring algorithm [2,3]. While very simple to understand, such a technique may be difficult to implement in a real-time environment to process transient data.
2. *Instrumental variables*: This second solution consists in treating sensor faults symmetrically to component faults. To this end, parameters intended to model sensor faults are introduced in the diagnosis problem as additional unknowns to be estimated together with the health parameters [1,4]. This technique, however, makes the estimation problem underdetermined and achieves a poor sensor-fault isolation.
3. *Banks of Kalman filters* have also been applied to cope with sensor faults: Each filter uses all sensors but one to estimate the health parameters. The filter that does not rely on the faulty sensor is the only one achieving an accurate estimation [5,6]. The major drawback of this approach resides in its combinatorial nature, allowing the isolation of a predefined set of fault events.
4. *Robust estimation techniques*: The concept is to include robustness against sensor faults within the estimation algorithm itself. The robustness is obtained by replacing the least-squares criterion by another one (to be defined later), less sensitive to outliers [7–9]. In short, an outlier can be defined as an observation that lies outside the overall pattern of a given distribution.

In the present contribution, the development of a new algorithm for performance monitoring based on *robust estimation* is reported. The innovative aspect is that the robust estimation problem, which is basically a nonlinear one, can be transformed into a quadratic programming problem [10] for which efficient solvers are available. The benefit in terms of stability and accuracy brought by the robust diagnosis tool is illustrated for several sensor faults that may be encountered on a jet engine.

Description of the Method

The scope of this section is to provide a short description of our diagnosis tool (see Ref. 11 for more details), to derive the sensor-fault-detection and isolation (SFDI) module, and to present its integration within the existing diagnosis algorithm.

Diagnosis Tool. As part of model-based techniques, our diagnosis tool requires a simulation model of the turbine engine. In the framework of gas path analysis, these are basically nonlinear aerothermodynamic models based on mass, energy, and momentum conservation laws applied to the engine.

The diagnosis tool is able to process *transient data*. Given the nonlinearity of the system model, the *extended Kalman filter* [12]

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 4, 2007; final manuscript received May 7, 2007; published online March 5, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), 2007, Montreal, Quebec, Canada, May 14-17, 2007, Paper No. GT2007-27324.

is used instead of the generic, linear Kalman filter. The simulation model of the engine is written in the state-space form, namely,

$$x_k = \mathcal{F}(u_k, v_k, w_k, x_{k-1}) + \boldsymbol{\nu}_k \tag{1}$$

$$y_k = \mathcal{G}(u_k, v_k, w_k, x_k) + \boldsymbol{\epsilon}_k \tag{2}$$

where k is a discrete time index, u_k are the command parameters (e.g., fuel flow), v_k are the measurable exogenous inputs (e.g., ambient temperature and pressure, Mach number), w_k are the aforementioned health parameters, and x_k are the state variables. The state variables are associated with the transient phenomena taking place in the gas path of the engine. Generally speaking, three types of transient effects are distinguished: the heat transfers between the gas path and the components of the engine, the shaft inertia, and the fluid transport delays. Equation (1) is named the state transition equation and gathers the deterministic model of the engine dynamics $\mathcal{F}(\cdot)$ and a random variable $\boldsymbol{\nu}_k$, which represents model inaccuracies. Similarly, Eq. (2) is called the measurement equation and gathers the deterministic simulation model $\mathcal{G}(\cdot)$ and a random variable $\boldsymbol{\epsilon}_k$, which represents sensor inaccuracies. To complete the description of the system, a third equation describing the temporal evolution of the health parameters must be supplied. A random walk model is generally adopted (see Ref. 13 for further details).

Both $\boldsymbol{\nu}_k$ and $\boldsymbol{\epsilon}_k$ are assumed to be zero-mean, white and Gaussian random variables,[1] which is denoted by

$$\boldsymbol{\epsilon}_k = \mathcal{N}(0, R_y) \text{ and } \boldsymbol{\nu}_k = \mathcal{N}(0, R_x) \tag{3}$$

The state variables are either measured, thus noisy, or nonmeasurable. Hence, they must be estimated together with the health parameters from the same sequence of measurements y_k. A marginal estimation approach is selected to solve this dual estimation problem. The marginal estimation approach is presented in Ref. 14 and basically relies on two extended Kalman filters running concurrently, one for the health parameters and the other one for the state variables. Once the former filter has updated the health parameters, the current value is used by the latter to update the corresponding state variables. So, the state variable estimation improves as the identified health parameters are getting closer to their actual values.

Provided that a prior value for the health parameters and the state variables is available, the basic step consists in observing the discrepancies between the model outputs, denoted $\hat{y}_k$, and the observed measurements y_k. These discrepancies, also called residuals and denoted r_k, are processed by the marginal estimation Kalman filter, which recursively updates the health parameters and the state variables so that the residuals are driven to zero on average. This process is illustrated in Fig. 1.

For sake of brevity, elements of the estimation technique are recalled for the health parameter identification side of the dual estimation problem. All developments may be transposed to the state variable estimation problem by making adequate substitutions.

The Kalman filter can be seen as a recursive maximum a posteriori approach to parameter identification. Both the health parameters and the measurements are considered as random variables following a normal distribution. Within this framework, the estimated health parameters are obtained by minimizing the following objective function:

$$\mathcal{J}(w_k) = \frac{1}{2}(w_k - \hat{w}_k^-)^T (P_{w,k}^-)^{-1}(w_k - \hat{w}_k^-) + \frac{1}{2} r_k^T R_y^{-1} r_k \tag{4}$$

The first term in the right-hand side of Eq. (4) forces the identified health parameters to lie in a neighborhood of the prior value $\hat{w}_k^-$. The prior covariance matrix $P_{w,k}^-$ specifies the shape of this region

[1]In addition, $\boldsymbol{\nu}_k$ and $\boldsymbol{\epsilon}_k$ are assumed uncorrelated.

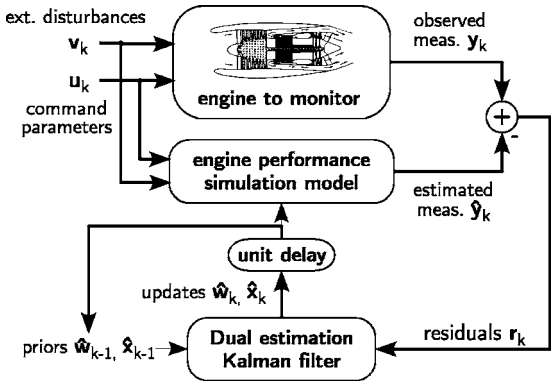


Fig. 1 Health parameter and state variable update mechanism using a DEKF

and summarizes the information contained in the measurement sequence up to time $k-1$. The second term reflects the weighted least-squares criterion. The interested reader may consult Ref. [13] for an extensive derivation.

Considering an update mechanism based on an extended Kalman filter, the residuals are approximated with

$$r_k = y_k - \hat{y}_k = y_k - [\hat{y}_k^- + G_k(w_k - \hat{w}_k^-)] = \hat{r}_k^- - G_k(w_k - \hat{w}_k^-) \tag{5}$$

where

$$\hat{y}_k^- = \mathcal{G}(u_k, v_k, \hat{w}_k^-, \hat{x}_k^-)$$

$$G_k = \left. \frac{\partial}{\partial w_k} \mathcal{G}(u_k, v_k, w_k, \hat{x}_k^-) \right|_{w_k = \hat{w}_k^-}$$

are, respectively, the a priori prediction of the measurements and the Jacobian matrix of the measurement equation around the prior values of the states and health parameters.

Canceling out the first order derivatives of Eq. (4) leads to the update relation for the health parameters:

$$(P_{w,k}^-)^{-1}(w_k - \hat{w}_k^-) - G_k^T R_y^{-1} r_k = 0 \tag{6}$$

Making the Algorithm Sensor-Fault Tolerant. As mentioned in the Introduction, a sensor fault can be defined as data generated by a sensor that no longer follows the manufacturer's characteristics. Mathematically speaking, it means that the measurement noise associated with a faulty sensor cannot be described by a zero-mean Gaussian random variable, which is one of the assumptions of the Kalman filter as seen in Eq. (3). In this section, a robust estimation technique is presented that lowers the sensitivity of the Kalman filter with respect to outliers.

The quadratic penalization of the residuals in the objective function (4), which derives directly from the assumption of a Gaussian measurement noise, makes the algorithm very sensitive to large residuals. Consequently, even a small amount of outliers can strongly deteriorate the quality of the estimation.

The aim of robust estimation techniques is to lower the sensitivity with respect to large residuals by replacing the Gaussian probability density function by another noise distribution prone to outliers. Among the many candidate distributions, the so-called δ-contaminated function (also known as Huber's function) has received much attention in the literature and is selected in this contribution. A detailed description of this function is beyond the scope of this paper, but can be found in Ref. [15]. Basically, Huber's function consists in a Gaussian random variable contaminated by a fraction δ of outliers.

Practically, the residuals r_k in Eq. (6) are replaced with a function $\psi(r_k)$ intended to deemphasize the influence of large residuals:

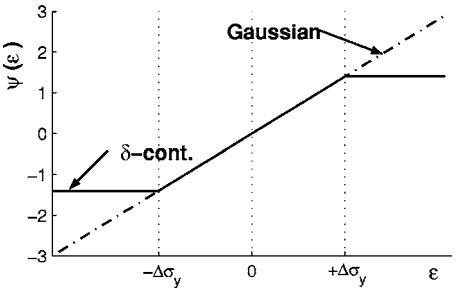


Fig. 2 Huber's weighting function for a scalar random variable

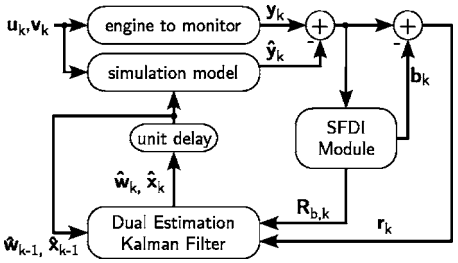


Fig. 3 Integration of the SFDI module

$$(P_{w,k}^{-})^{-1}(w_k - \hat{w}_k^{-}) - G_k^T R_y^{-1} \psi(r_k) = 0 \tag{7}$$

Huber's weighting function is presented in Fig. 2 for a scalar random variable ϵ. Mathematically, it can be written as

$$\psi(\epsilon) = \max[-\Delta\sigma, \min(\epsilon, \Delta\sigma)] \tag{8}$$

where σ is the standard deviation of the "clean" Gaussian variable and the scalar Δ is a threshold depending on the contamination level δ (e.g, $\Delta = 1.399$ for $\delta = 0.05$).

Adopting Huber's function as the penalization function of the residuals leads to robustness against sensor faults but turns the simple linear function (6) for parameter update into a nonlinear one. No explicit update formula can be obtained and relation (7) must be solved numerically, with Newton's method, for instance. This might be cumbersome for on-line applications.

Nonetheless, it is shown in Ref. [10] that the nonlinear program resulting from the choice of a noise distribution following Huber's law can be transformed into a quadratic program for which efficient and fast solvers are available [16]. Reported below are the milestones of this transformation.

As a first step, the vector $\mathbf{b}_k$ is introduced in the measurement equation (2) to model the outliers:

$$y_k = \mathcal{G}(u_k, v_k, w_k, x_k) + \mathbf{b}_k + \boldsymbol{\epsilon}_k \tag{9}$$

Then, the objective function of the robust estimation problem becomes[2]

$$\mathcal{J}(w_k, \mathbf{b}_k) = \frac{1}{2}(w_k - \hat{w}_k^{-})^T (P_{w,k}^{-})^{-1}(w_k - \hat{w}_k^{-}) + \frac{1}{2} r_k^T R_y^{-1} r_k + \Delta^{-1}\sigma^{-T}|\mathbf{b}_k| \tag{10}$$

The equivalence between solving Eq. (7) for w_k with Huber's ψ function and minimizing the objective function (10) with respect to w_k and $\mathbf{b}_k$ is established in Ref. [10]. It should be pointed out that the criterion (10) is convex and therefore admits a unique optimum. In Ref. [17], a robust diagnosis tool for steady-state data is obtained by minimizing equation (10) alternatively with respect to w_k and $\mathbf{b}_k$.

To transform the objective function (10) into a quadratic program, the quantities $\mathbf{b}_k$ and $|\mathbf{b}_k|$ are replaced with their positive and negative parts:

$$\begin{aligned} &\mathbf{b}_k = \mathbf{b}_k^{+} - \mathbf{b}_k^{-} \\ &|\mathbf{b}_k| = \mathbf{b}_k^{+} + \mathbf{b}_k^{-} \end{aligned} \quad \text{with} \quad \begin{cases} \mathbf{b}_k^{+} = \max(\mathbf{b}_k, 0) \\ \mathbf{b}_k^{-} = -\min(\mathbf{b}_k, 0) \end{cases} \tag{11}$$

With this change of variables, the objective function writes down

$$\mathcal{J}(w_k, \mathbf{b}_k^{+}, \mathbf{b}_k^{-}) = \frac{1}{2}(w_k - \hat{w}_k^{-})^T (P_{w,k}^{-})^{-1}(w_k - \hat{w}_k^{-}) + \frac{1}{2} r_k^T R_y^{-1} r_k + \Delta^{-1}\sigma^{-T}(\mathbf{b}_k^{+} + \mathbf{b}_k^{-}) \tag{12}$$

At this point, it may be interesting to compare the presented approach and the one based on instrumental variables [1,4]. Both of them introduce a vector of additional unknowns, namely, the vector $\mathbf{b}_k$, to model the sensor faults. The major difference lies in the fact that the instrumental variable approach treats symmetrically the health parameters and the bias vector $\mathbf{b}_k$ while the robust one adds a linear penalizing term on $\mathbf{b}_k$—the third term in the right-hand side of Eq. (10).

Unlike the Kalman filter, the quadratic programming formulation does not supply an explicit relation for the update of the covariance matrix of the health parameters, which is certainly a drawback. Yet, in the realm of performance monitoring, the degradation undergone by the engine evolves relatively slowly and is continuously tracked so that the prior value $\hat{w}_k^{-}$ will generally be close to the true value. Under this assumption, a simplification is applied in the robust estimation problem: the estimations of the bias $\mathbf{b}_k$ and the health parameters are decoupled. First, the sensor-fault vector is estimated by minimizing (10) in which the health parameters are frozen to their prior value. This results in the following quasiunconstrained quadratic program:

$$\min_{\mathbf{b}_k^{+}, \mathbf{b}_k^{-}} \mathcal{J}(\hat{w}_k^{-}, \mathbf{b}_k^{+}, \mathbf{b}_k^{-}) \quad \text{subject to} \begin{cases} \mathbf{b}_k^{+} \geqslant 0 \\ \mathbf{b}_k^{-} \geqslant 0 \end{cases} \tag{13}$$

Secondly, a regular extended Kalman filter comes into play to update the health parameters and their covariance matrix. The effect of sensor biases is accounted doubly: The residual processed by the Kalman filter is corrected with the vector $\mathbf{b}_k$ and the covariance matrix of the measurement noise is adapted accordingly:

$$r_k = y_k - \hat{y}_k \rightarrow r_k = y_k - \hat{y}_k - \mathbf{b}_k \tag{14}$$

$$R_y \rightarrow R_{b,k} = R_y + \mathbf{b}_k \mathbf{b}_k^T \tag{15}$$

From a marginal estimation standpoint, splitting the estimation of the bias and of the health parameters constitutes an advantage given that an objective function analogous to Eq. (10) describes the problem of robust state estimation. Yet, with the present technique, only one call to the quadratic programming solver is needed to determine the sensor biases. This is highlighted in the following section.

Modification of the Diagnosis Algorithm. The block diagram of the modified diagnosis algorithm is shown in Fig. 3. A short description of the procedure is given in the sequel.

Similar to the original procedure detailed in Fig. 1, the previous estimates of the state variables $\hat{x}_{k-1}$ and health parameters $\hat{w}_{k-1}$ are used together with the current inputs u_k and v_k by the engine performance model to generate an a priori estimation of the measurements. The a priori residuals between the measurements and their estimation are input to the SFDI module, which computes the sensor-fault vector $\mathbf{b}_k$ by solving Eq. (13). The residual r_k is formed by subtracting the sensor-fault vector $\mathbf{b}_k$ to the a priori residuals and is fed into the original dual estimation Kalman filter (DEKF). The covariance matrix of the measurement noise R_y is replaced with $R_{b,k}$ to further lower the contribution of the faulty sensor to the state and parameter updates.

A further operation, which is not sketched in Fig. 3, has been

[2]Note that the residuals r_k are also a linear function of w_k and $\mathbf{b}_k$.

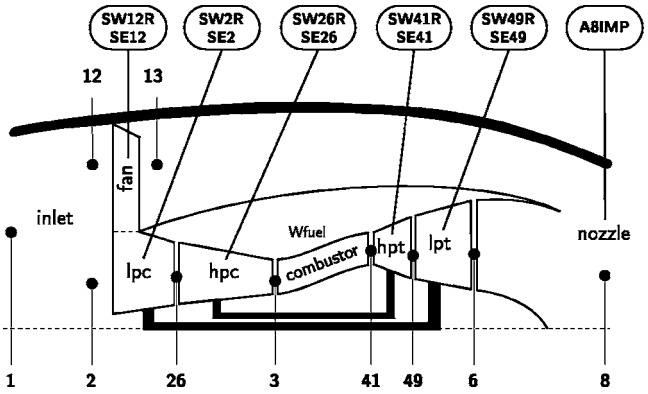


Fig. 4 Turbofan layout with station numbering and health parameter's location

implemented in the robust diagnosis tool. Besides its direct use in Eqs. (14) and (15), the sensor-fault vector $\mathbf{b}_k$ is also processed with an exponentially weighted moving average filter. As soon as one of the components of the filtered $\mathbf{b}_k$ exceeds a predefined threshold, the related sensor is discarded in the DEKF. In that way, the DEKF becomes totally immunized against a detected sensor bias or drift by processing an incomplete residual vector. Reference [18] explains the modifications to apply to the Kalman filter in that case.

The impact of the SFDI module on the global computational burden is very limited (less than 10% increase) due to the existence of very efficient solvers for quasiunconstrained quadratic problems. Moreover, most of the CPU time is spent in the computation of the Jacobian matrices needed by the extended Kalman filters.

Application of the Method

Engine Layout. The application used as a test case is a large bypass ratio mixed-flow turbofan. Due to the nonavailability of real data, the robust diagnosis tool has been tested on simulated data only. The engine performance model has been developed in the frame of the OBIDICOTE[3] project and is detailed in Ref. [19]. A schematic of the engine is sketched in Fig. 4 where the location of the 11 health parameters and the station numbering are also indicated. One command variable, which is the fuel flow rate fed in the combustor, is considered in the following.

A dynamic model is available in the state-space form specified by relations (1) and (2). The modeled engine dynamics account for the shaft inertia and for the heat transfers in the hpc, the combustor, and the hpt, which results in the set of seven state variables listed in Table 1.

The sensor suite selected for tracking the performance degradation is representative of the instrumentation available on-board contemporary turbofan engines and is detailed in Table 2 where the nominal accuracy of each sensor is also reported. In this configuration, the number of health parameters to estimate exceeds the number of sensors. Therefore, if a sensor fault occurs, it is likely that several combinations of the health parameter deviations exactly reproduce the sensor fault, thus leading to a wrong diagnosis of the engine condition. To allow the simultaneous detection of sensor and component faults, the redundancy must consequently be increased.

In the present application, the redundancy is extended by taking into account the fact that, on modern engines, the instrumentation is generally "dual channeled:" Each of the eight probes is connected to two independent lanes (sensing element and signal processing hardware). As a result, 16 measurements are available, which provides the necessary redundancy to perform the robust estimation problem. It must be pointed out that this trick does not modify the observability of the health parameters, which is linked to the variety in the sensor suite.

[3]A Brite/Euram project for On-Board Identification, Diagnosis and Control of Turbofan Engine.

Table 1 State variables modeling the transient effects (following the OBIDICOTE nomenclature)

Label	Description
N_{lp}	Low pressure spool rotational speed
N_{hp}	High pressure spool rotational speed
T_{m3b}	High pressure compressor blade temperature
T_{m3c}	High pressure compressor casing temperature
T_{m4b}	Combustion chamber casing temperature
T_{m42b}	High pressure turbine blade temperature
T_{m42c}	High pressure turbine casing temperature

Definition of the Test Cases. A series of test cases has been designed to assess the efficiency of the new robust estimation technique. The operating conditions are set to sea-level static, standard day conditions. The evolution of the fuel flow versus time is open-loop scheduled, as depicted in Fig. 5. The sequence is 800 s long and is made of two successive power sweeps between ground idle and maximum continuous regimes, followed by a slam between ground idle and take-off ratings.

The engine wear is simulated from the component fault case proposed in Ref. [20]. It consists in a drift of nearly all health parameters, starting from a healthy engine (all parameters at their nominal values) at $t=0$ s and with the following degradation at the end of the sequence ($t=800$ s): −0.5% on SW12R, −0.5% on SE12, −0.4% on SW2R, −0.5% on SE2, −1.0% on SW26R, −0.7% on SE26, +0.4% on SW41R, −0.8% on SE41, −0.5% on SE49.

Gaussian noise, whose magnitude is specified in Table 2, is added to the clean simulated measurements in order to make them closer to real ones. The data are processed at a sampling frequency of 50 Hz, which is a sufficient value to capture the aforementioned engine dynamics.

Three types of sensor faults are investigated in the test cases: impulsive noise, sensor bias, and sensor drift. The sign of the sensor fault must also be accounted for; this results in six types of

Table 2 Selected sensor suite (uncertainty is three times the standard deviation)

Label	Uncertainty	Label	Uncertainty
p_{13}^0	±100 Pa	T_{26}^0	±2 K
p_{26}^0	±500 Pa	T_3^0	±2 K
p_3^0	±5000 Pa	T_6^0	±2 K
N_{lp}	±6 rpm	N_{hp}	±12 rpm

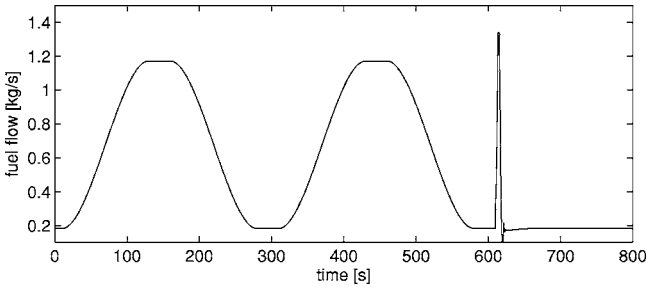


Fig. 5 Fuel flow profile

Table 3 Description of the sensor faults

Case	Description	Magnitude
a	5% impulsive noise	$+10\sigma$
b	5% impulsive noise	-10σ
c	Sensor bias	$+5\sigma$
d	Sensor bias	-5σ
e	Sensor drift	$+10\sigma$
f	Sensor drift	-10σ

sensor faults, which are described in Table 3.

The level of impulsive noise per sensor is set to 5%; it means that on average, 5 out of 100 samples are aberrant. The magnitude is set to ± 10 times the standard deviation. For the biases, only one sensor is simulated faulty at a time. The bias starts at $t = 100$ s and does not evolve until the end of the sequence. Considering the drifts, only one sensor is simulated faulty at a time as well. The drift starts at time $t = 100$ s with a magnitude of zero and reaches a magnitude of ± 10 standard deviation at $t = 800$ s. The magnitudes of the sensor faults have been selected according to contributions from members of the OBIDICOTE project.

Definition of a Figure of Merit. The efficiency of the estimation performed by the original diagnosis tool and the robust one is assessed in terms of the maximum root mean square error (RMSE) over the whole sequence:

$$e_{rms} = \max\left[\sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\frac{w_k - \hat{w}_k}{w^{hl}}\right)^2}\right] \tag{16}$$

where w^{hl} are the nominal values of the health parameters.

Given the stochastic character of the measurement noise, each test case has been run 20 times and the RMSEs reported in Tables 4–6 are the average over the 20 runs in order to guarantee that they are statistically representative. A test case characterized by an averaged maximum RMSE below 0.25% is declared as successful, which is indicated by a checkmark. This threshold corresponds to three times the standard deviation of the identified health parameters (i.e., the square root of the diagonal terms of the covariance matrix $P_{w,k}$).

Table 4 Comparison of the estimation errors obtained with the original and with the robust diagnosis tool in the case of no sensor fault (nosf) and impulsive noise

Fault case	Regular DEKF		Robust DEKF	
nosf	0.07%	√	0.07%	√
a	0.67%	—	0.07%	√
b	0.68%	—	0.06%	√

Table 5 Comparison of the estimation errors obtained with the original and with the robust diagnosis tool in the case of sensor biases

	Regular DEKF				Robust DEKF			
	Case c		Case d		Case c		Case d	
p_{13}^0	0.78%	—	1.15%	—	0.08%	√	0.08%	√
p_{26}^0	1.14%	—	1.12%	—	0.07%	√	0.07%	√
T_{26}^0	3.95%	—	4.51%	—	0.08%	√	0.07%	√
p_3^0	0.42%	—	0.50%	—	0.08%	√	0.08%	√
T_3^0	0.57%	—	0.60%	—	0.07%	√	0.07%	√
N_{lp}	0.47%	—	0.58%	—	0.07%	√	0.07%	√
N_{hp}	0.35%	—	0.44%	—	0.07%	√	0.07%	√
T_6^0	0.29%	—	0.39%	—	0.06%	√	0.06%	√

Table 6 Comparison of the estimation errors obtained with the original and with the robust diagnosis tool in the case of sensor drifts

	Regular DEKF				Robust DEKF			
	Case e		Case f		Case e		Case f	
p_{13}^0	1.13%	—	3.21%	—	0.08%	√	0.08%	√
p_{26}^0	1.67%	—	1.64%	—	0.07%	√	0.08%	√
T_{26}^0	5.82%	—	6.37%	—	**0.24**%	√	**0.24**%	√
p_3^0	0.58%	—	0.67%	—	0.07%	√	0.09%	√
T_3^0	0.74%	—	0.80%	—	0.09%	√	0.07%	√
N_{lp}	0.61%	—	0.74%	—	0.11%	√	0.09%	√
N_{hp}	0.41%	—	0.51%	—	**0.31**%	—	0.08%	√
T_6^0	0.34%	—	0.43%	—	0.06%	√	**0.44**%	—

Results: Impulsive Noise. Table 4 reports the figure of merit defined above for the test cases involving no sensor fault (i.e., Gaussian noise only) and impulsive noise (fault types a and b). It can be seen that the robust DEKF achieves the same efficiency in the estimation of the health parameters as the regular DEKF when no sensor fault is present. The figure of merit obtained by the standard DEKF when facing impulsive noise underlines the pronounced sensitivity of the Kalman filter with respect to outliers. On the contrary, the robust DEKF keeps the same level of accuracy when processing data contaminated with "spikes."

Results: Sensor Biases. The scores obtained by both diagnosis tools in the case of one biased sensor are given in Table 5. Whatever the faulty sensor and the sign of the bias, the basic DEKF provides a spoiled diagnosis of the engine condition. It can be concluded that only 1 measurement out of 16 violating the *zero-mean* Gaussian assumption suffices to put the Kalman filter into trouble. This is another proof of the high sensitivity of least-squares based methods to outliers. As far as the robust DEKF is concerned, it seems to be simply unaffected by the sensor bias as the figure of merit is roughly the same as for the no-sensor-fault case.

For illustrative purpose, Fig. 6 depicts the estimation made by the SFDI module of a bias affecting one of the hpc discharge pressure channels.

The bias is introduced at $t = 100$ s and nearly immediately detected by the SFDI module. This rather fast detection of the sensor fault is due to the classification nature of the algorithm supporting the SFDI task (basically relying on Huber's weighting function (8)). It can be seen that a fair estimation of the magnitude of the sensor fault is achieved. Note that analogous comments apply to the other fault cases involving a sensor bias.

Results: Sensor Drifts. The performance of the diagnosis tools when the data are contaminated with a drifting reading on one sensor is summarized in Table 6. As for the sensor bias case, the regular DEKF is unable to cope with the sensor drift and generates a very poor information about the engine condition. In most cases, the robust DEKF manages to capture the instrumentation fault and to keep an accurate tracking of the engine wear.

Nonetheless, it can be seen in Table 6 that the robust estimation

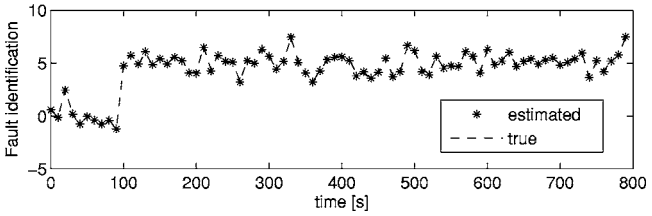


Fig. 6 Identification of the bias on p_3^0

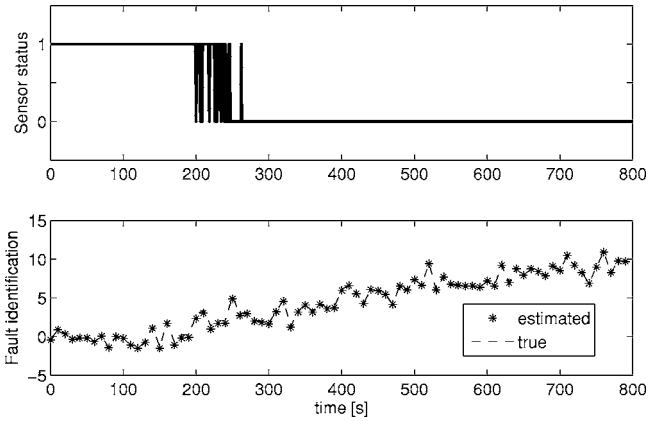


Fig. 7 Identification of a positive drift on T_{26}^0

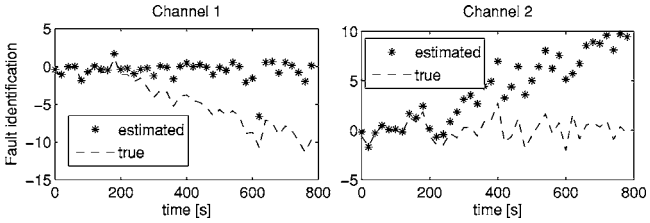


Fig. 9 Misidentification of the sensor drift on T_6^0

algorithm is not able to deal with a positive drift on the hp spool speed and with a negative drift on T_6^0. A drift (positive or negative) on the hpc inlet temperature leads to a higher RMSE too, which is highlighted by bold values in Table 6. Two of these cases are studied deeper hereafter.

First, the case of a positive drift on T_{26}^0 is investigated. In Fig. 7, the identification of the sensor fault is plotted. The upper figure indicates the status of the considered sensor, 1 being healthy and 0 faulty. It can be seen that the sensor is really declared faulty from $t=250$ s whereas the drift begins at $t=100$ s. Between $t=100$ s and $t=200$ s, the sensor is seen as healthy by the SFDI module and between $t=200$ s and $t=250$ s, the algorithm hesitates about the status of the sensor.

This fuzzy behavior can be explained by the very nature of the implemented sensor fault. Indeed, the drift starts with a zero magnitude and linearly increases up to +10 times the standard deviation, as can be seen in the bottom figure. Hence, in the first dozens of seconds after its appearance, the drift has on average a lower magnitude than the threshold used by the SFDI module to classify the sensor fault (which is, as a reminder, $\Delta\sigma$). Superimposed to the low magnitude of the drift, the Gaussian noise makes the fault detection even tougher.

The evolution of the identified health parameters is depicted in Fig. 8. If attention is only paid to the estimated values in the last part of the test, the accuracy is of good quality. When looking overall, this statement must be qualified, however. Indeed, the estimation of the health parameters related to the turbines and the nozzle is fairly good throughout the sequence. A similar comment applies to the identified values for the fan and the hpc. On the contrary, the evolution of the health parameters associated with the lpc, namely, SW2R and SE2, exhibits a bump, which extends roughly between $t=100$ s and $t=400$ s, reaching its maximum around $t=250$ s. This hump, responsible for the increase in the RMSE (see Table 6), is directly linked to the aforementioned difficulty for the SFDI module to capture a rising drift. The bump goes down as soon as the sensor is declared faulty, the rate of decrease being controlled by the covariance of the health parameters $P_{w,k}$.

The second detailed investigation is carried for the case of a negative drift on T_6^0 (exhaust gas temperature). Figure 9 sketches the information generated by the SFDI module for both electron gas temperature (EGT) channels. The fault is actually implemented on channel one, as shown in the left graph, but the SFDI module assigns the fault, after some 250 s, to channel 2 as is indicated in the two pictures. Clearly, in this case, the sensor fault is wrongly diagnosed and this translates into a totally spoiled estimation of the engine condition.

To prove this statement, the identified health parameters are plotted versus time in Fig. 10. Up to $t=250$ s, the robust DEKF is able to track accurately the engine degradation because the magnitude of the sensor drift is too small to perturb the estimator. Afterwards, the SFDI module diagnoses the wrong EGT channel as faulty and discards its use by the DEKF. The health parameters are then adapted to catch not only the engine wear but also the drift on the remaining EGT channel (which is believed nonfaulty by the algorithm). This leads to an erratic estimation of the engine condition. The health parameters that are most affected by the wrong sensor are SW2R, SW26R, SW41R, and SW49R with up to 2% variation for the hpt capacity at the end of the sequence. This is not surprising as it is known that measurements are more sensitive to flow capacity corrections than to efficiency corrections [21].

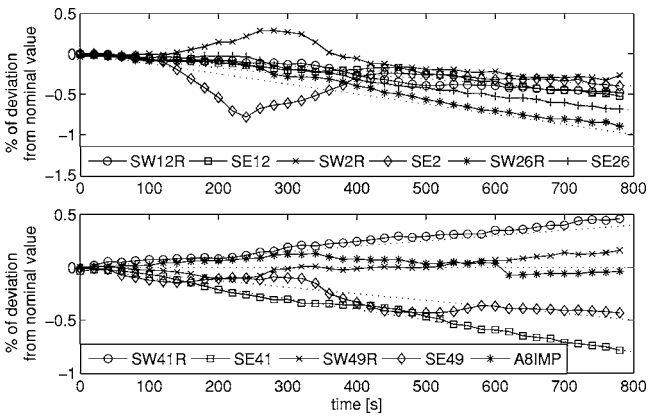


Fig. 8 Identification of the health parameters with a positive sensor drift on T_{26}^0. Dotted lines show actual parameter values.

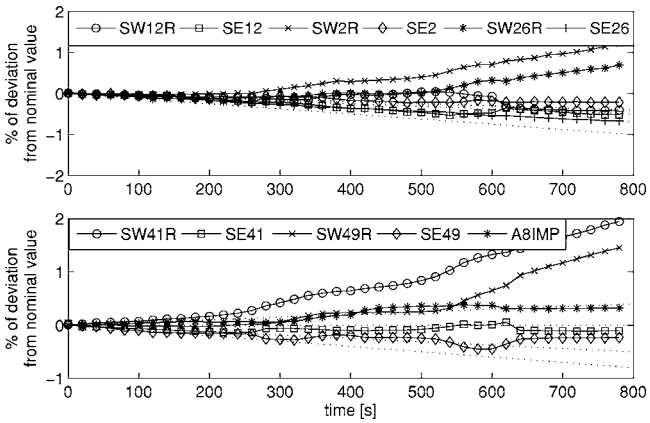


Fig. 10 Identification of the health parameters with a sensor drift on T_6^0. Dotted lines show actual parameter values.

Discussion

To complete the analysis of the results, some issues that may lead to further developments of the proposed methodology are discussed in the following.

The first question is related to the rejection of sensor biases. The SFDI module has managed to detect a bias with a magnitude of +5 times the standard deviation. It is clear that a higher sensor bias will also be correctly diagnosed by the classification logic, but the determination of the lowest bias for each sensor that can be caught by the SFDI module may be of practical interest.

A second interesting axis of research consists in making the present robust DEKF, which is primarily dedicated to performance monitoring, capable to track abrupt variation of the engine condition (due, for instance, to foreign object damage). This problem is quite challenging since the abrupt component fault may be interpreted by the robust diagnosis tool as sensor faults.

Finally, in the light of the quadratic programming approach presented herein, the application of moving horizon estimation techniques to the problem of robust estimation may be tempting. Indeed, moving horizon estimation (MHE) possesses some very appealing features such as the recursive minimization of a quadratic objective function over a finite time interval or the natural inclusion of constraints on the estimated variables [22].

Conclusion

In this contribution, a sensor-fault-tolerant tool for performance monitoring has been developed in the framework of robust estimation. The advantages of this approach such as direct integration in the estimation procedure of the robustness issue or ability to handle a variety of fault scenarios have been pointed out. The improvements in the accuracy of the estimation of the engine condition achieved by the robust diagnosis tool with respect to the classical one are demonstrated for typical sensor-fault cases that may occur on a turbofan engine.

The robust estimation tool can cope correctly with impulsive noise and sensor biases. Considering the estimation of sensor drifts, which is by far more complicated than the two other types, promising results are obtained. They tend to show that protection against sensor drifts may be achieved at the price of a relatively high hardware redundancy (variety of the on-board instrumentation).

Nomenclature

- $\hat{a}$ = estimation of an unknown variable a
- A8IMP = nozzle exit area (nominal value: 1.4147 m^2)
- $\mathbf{b}_k$ = the sensor-fault vector
- DEKF = dual estimation Kalman filter
- hpc = high pressure compressor
- hpt = high pressure turbine
- k = discrete time index
- lpc = low pressure compressor
- lpt = low pressure turbine
- p_i^0 = total pressure at station i
- SEi = efficiency degradation of the component whose entry is located at section i (nominal value: 1.0)
- SFDI = sensor-fault detection and isolation
- SWiR = flow capacity of the component whose entry is located at section i (nominal value: 1.0)
- T_i^0 = total temperature at station i
- u_k = actual command parameters
- v_k = actual external disturbances
- w_k = actual but unknown health parameters
- x_k = actual but unknown state variables
- y_k = observed measurements
- $\boldsymbol{\epsilon}_k$ = measurement noise vector
- $\boldsymbol{\nu}_k$ = process noise vector
- $\boldsymbol{\sigma}$ = vector of the standard deviations of the sensors
- $\mathcal{N}(m,\mathbf{R})$ = a Gaussian probability density function with mean m and covariance matrix $\mathbf{R}$

References

[1] Volponi, A. J., 2003, *Foundation of Gas Path Analysis (Part I and II)*, von Karman Institute Lecture Series: Gas Turbine Condition Monitoring and Fault Diagnosis, (2003-01).

[2] Liu, H., Shah, S., and Jiang, W., 2004, "On-Line Outlier Detection and Data Cleaning," Comput. Chem. Eng. **28**, 1635–1647.

[3] Surrender, V., and Ganguli, R., 2004, "Adaptive Myriad Filter for Improved Gas Turbine Condition Monitoring Using Transient Data," ASME Paper No. GT2004-53080.

[4] Volponi, A. J., DePold, H., Ganguli, R., and Daguang, C., 2003, "The Use of Kalman Filter and Neural Network Methodologies in Gas Turbine Performance Diagnostics: A Comparative Study," ASME J. Eng. Gas Turbines Power **125**, 917–924.

[5] Kobayashi, T., and Simon, D. L., 2003, "Application of a Bank of Kalman Filters for Aircraft Engine Fault Diagnostics," ASME Paper No. GT2003-38550.

[6] Kobayashi, T., and Simon, D. L., 2004, "Evaluation of an Enhanced Bank of Kalman Filters for In-Flight Aircraft Engine Sensor Fault Diagnostics," ASME Paper No. GT2004-53640.

[7] Dewallef, P., and Léonard, O., 2001, "Robust Measurement Validation on Jet Engines," *Proceedings of the Fourth European Conference on Turbomachinery*.

[8] Grodent, M., and Navez, A., 2001, "Engine Physical Diagnosis using a Robust Parameter Estimation Method," 37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference.

[9] Dewallef, P., and Léonard, O., 2005, "On-Line Performance Monitoring and Engine Diagnostic Using Robust Kalman Filtering Techniques," ASME Paper No. GT2003-38379.

[10] Fuchs, J. J., 1999, "An Inverse Approach to Robust Linear Regression," 14th IFAC World Congress.

[11] Borguet, S., Dewallef, P., and Léonard, O., 2005, "On-Line Transient Engine Diagnostics in a Kalman Filtering Framework," ASME Paper No. GT2005-68013.

[12] Haykin, S., 2001, *Kalman Filtering and Neural Networks* (Wiley Series on Adaptive and Learning Systems for Signal Processing, Communication and Control), Wiley, New York.

[13] Dewallef, P., 2005, "Application of the Kalman Filter to Health Monitoring of Gas Turbine Engines: A Sequential Approach to Robust Diagnosis," PhD. thesis, University of Liège.

[14] Nelson, A. T., 2000, "Nonlinear Estimation and Modeling of Noisy Time Series by Dual Kalman Filtering Methods," PhD. thesis, Oregon Graduate Institute of Technology.

[15] Huber, P. J., 1996, *Robust Statistical Procedures*, CBMS-NSF Regional Conf. Series in Applied Mathematics No. 68, SIAM, Philadelphia.

[16] Fletcher, R., 2000, *Practical Methods of Optimization*, Wiley, New York.

[17] Dewallef, P., 2006, "An Algorithm for Aircraft Engine Performance Monitoring and Sensor Fault Diagnosis," Technical Report, Turbomachinery Group, University of Liège, http://www.ulg.ac. be/turbo/research/paper/tr-pd06.pdf

[18] Cipra, T., and Romera, R., 1997, "Kalman Filter With Outliers and Missing Observations, " T.E.S.T. Journal, **6**(2), pp. 379–395.

[19] Stamatis, A., Mathioudakis, K., Ruiz, J., and Curnock, B., 2001, "Real-Time Engine Model Implementation for Adaptive Control and Performance Monitoring of Large Civil Turbofans," ASME Paper No. 2001-GT-0362.

[20] Curnock, B., 2000, "Obidicote Project—Work Package 4: Steady-State Test Cases," Rolls-Royce Plc, Technical Report DNS62433.

[21] Roth, B. A., Doel, D. L., and Cissell, J. J., 2005, "Probabilistic Matching of Turbofan Performance Models to Test Data," ASME Paper No. GT2005-68201.

[22] Haseltine, E. L., and Rawlings, J. B., 2003, "A Critical Evaluation of Extended Kalman Filtering and Moving Horizon Estimation," Technical Report No. 2002-03, Texas-Wisconsin Modeling and Control Consortium, University of Wisconsin–Madison, http://www.che.wisc.edu/jbr-group/tech-reports/twmcc-2002-03.pdf

Performance Assessment of Turbocharged Pem Fuel Cell Systems for Civil Aircraft Onboard Power Production

Stefano Campanari
Associate Professor
e-mail: stefano.campanari@polimi.it

Giampaolo Manzolini
Research Engineer
e-mail: giampaolo.manzolini@polimi.it

Andrea Beretti
Research Engineer

Dipartimento di Energetica,
Politecnico di Milano,
Piazza Leonardo da Vinci 32,
20133 Milano, Italy

Uwe Wollrab
Systems Engineering,
Fuel Cell Systems Development,
AIRBUS Deutschland Gmbh,
21129Hamburg, Germany
e-mail: uwe.wollrab@airbus.com

In recent years, civil aircraft projects are showing a continuous increase in the demand of onboard electrical power, both for the partial substitution of hydraulic or pneumatic controls and drives with electrical ones, and for the consumption of new auxiliary systems developed in response to flight safety and environmental control issues. Aiming to generate onboard power with low emissions and better efficiency, several manufacturers and research groups are considering the possibility to produce a relevant fraction of the electrical power required by the aircraft by a fuel cell system. The first step would be to replace the conventional auxiliary power unit (based on a small gas turbine) with a polymer electrolyte membrane (PEM) fuel cell type, which today is favored with respect to other fuel cell types; thanks to its higher power density and faster startup. The PEM fuel cell can be fed with a hydrogen rich gas coming from a fuel reformer, operating with the same jet fuel used by the aircraft, or relying on a dedicated hydrogen storage onboard. The cell requires also an air compression unit, where the temperature, pressure, and humidity of the air stream feeding the PEM unit during land and in-flight operation strongly influence the performance and the physical integrity of the fuel cell. In this work we consider different system architectures, where the air compression system may exploit an electrically driven compressor or a turbocharger unit. The compressor type and the system pressure level are optimized according to a fuel cell simulation model, which calculates the cell voltage and efficiency as a function of temperature and pressure, calibrated over the performances of real PEM cell components. The system performances are discussed under different operating conditions, covering ground operation, and intermediate and high altitude cruise conditions. The optimized configuration is selected, presenting energy balances and a complete thermodynamic analysis.
[DOI: 10.1115/1.2772636]

1 Introduction

New projects of civil aircrafts are in these years frequently influenced by a development strategy focusing on more electric aircraft (MEA) or even all electric aircraft (AEA) concepts. The partial substitution of conventional hydraulic or pneumatic controls and drives with electrical ones (see Table 1), and the introduction of new auxiliary systems bring about an increase in onboard electric power consumption, reaching values around 560 kW for airplanes such as the B777 or A330/A340, and going toward 1.3–1.5 MW for next generation aircrafts [1–5].

Furthermore, the possible elimination of power offtakes from the main engines would increase the nominal power output required by a separated onboard electricity generator.

It is well known that a fraction of electric power is generated on-board civil aircrafts by small turbine units, acting as auxiliary power units (APUs). Such machines operate with simple cycle, uncooled operation, low turbine inlet temperature (TIT), and pressure ratio, generally with single stage radial compressor and a power output in the range of several tenth kilowatts and up to the hundred kilowatt scale. Their advantages include low weight, rapid startup, and robustness; disadvantages are primarily the low efficiency (15–18%) and significant NO_x and CO emissions. The possible removal of power offtakes from the main engines would increase the nominal power of the APU system, making more important to look for higher efficiency and lower pollution devices.

Polymer electrolyte membrane fuel cells (PEM FCs) on their own are widely experimented in prototypes and generally recognized very attractive for future application in the automotive field; thanks to their ability to generate electricity with high efficiency (e.g., 50–55%) in tenth-kilowatt scale systems fed with hydrogen and operating at low temperatures (60–70°C). They also show a rather high power density (in terms of kW/kg and kW/dm^3), fast startup, and negligible emissions. The quick development of this technology has suggested to consider their application also in the aeronautic field, within the perspective of a step by step development, which could also represent a new possible market for the beginning of their commercialization.

The technology roadmap for this development includes several steps, where the first should be introducing a pure hydrogen PEM system with minor aircraft changes, aiming to provide a fraction of electric power (well below the potential power requirements shown in Table 1) for auxiliary loads and emergency systems actually sustained by APUs and other generators (for instance, the air turbine that is used to drive the pumps of the emergency hydraulic circuit). Subsequent steps could involve the use of onboard reformers [3] as well as different FC types with increasing power output.

The concept of integrating a PEM unit onboard civil aircrafts has been already introduced in several works [2,6,7], where the fuel cell has been generally considered as a device with assigned

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 28, 2007; final manuscript received May 9, 2007; published online February 29, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007, Paper No. GT2007-27658.

Table 1 Actual and future electric consumptions, active in-flight or during ground operation, for large aircrafts. Values with parentheses are related to systems active in emergency only

Main electrical consumers	System architecture			System power (kW)
	Conventional	MEA	AEA	
Emergency pumps	(X)	(X)	(X)	(30)
Aux. hydraulic pumps	X	X	X	60
Fuel pumps	X	X	X	10
Ice and rain protection	X	X	X	10
Lighting	X	X	X	15
Commercial loads	X	X	X	25
Avionics	X	X	X	10
Galley	X	X	X	75
Cargo doors	—	X	X	15
Flight control	—	—	X	80
Landing gear	—	—	X	25
Engine starter	—	X	X	350
Wing antiice	—	—	X	200
Environmental and cooling system (ECS)	—	X	X	400
Total demand (kW)	205 (235)	970 (1000)	1275 (1305)	
In-flight demand (kW)	205	605	910	

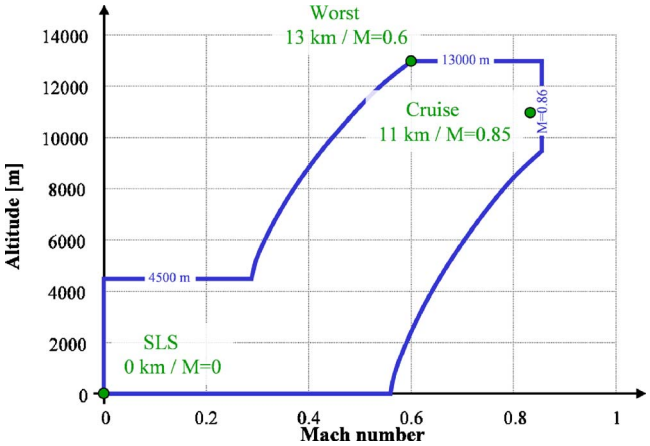


Fig. 1 Operating range that must be satisfied by an APU

energy balances. The scope of this paper is to discuss with better detail the possible plant configurations and the optimization of the FC operating conditions, using a model which takes into account the effects of variable fuel cell performances.

The PEM fuel cell requires to be fed with a high purity hydrogen fuel and with air; both reactants shall be properly humidified to allow an efficient and durable operation of the cell membrane, which works efficiently when wet. By the point of view of the cell operating conditions, pressurization increases the cell efficiency and power output (for a given fuel consumption), but is energetically expensive because the air flow has to be compressed. If such compression is carried out without recovering useful power from the cell pressurized exhausts, the compressor power consumption is higher than the power gain by the FC, so the final effect on the system is negative. For this reason, PEM systems usually work at pressure between 1.2 bars and 1.5 bars.

Conversely, when operating at high altitude and low ambient temperature, a certain degree of pressurization is necessary to allow the presence of liquid water at the cell typical operating temperatures. In such conditions, it becomes important to exploit the expansion of the cell exhaust gases with a turbine, leading to a turbocharged cell design. In all cases, it is therefore necessary to define the optimal pressure level of the fuel cell system and the most suitable system arrangement, which can be used to obtain the cell pressurization.

In the first section of the paper, a model for predicting the performances of PEM fuel cell, with variable pressure operation and hydrogen feeding, is discussed. Calculations are performed based on state-of-the-art performances of small PEM fuel cell stacks. Subsequently, two system arrangements are investigated, where the PEM system is pressurized by (i) an electrically driven compressor or (ii) a turbocharger unit, with (iii) an eventual burner ahead the turbine. This work presents a detailed analysis of the fuel cell operating parameters and of the other component characteristics (pressure drops and efficiencies), together with their effects on the system efficiency. The performances of such a system are predicted, evidencing the most suitable solution.

The option of integrating the PEM system with an eventual fuel reformer—capable of producing onboard the hydrogen required by the fuel cell, thus relying only on a jet fuel supply—is not addressed in this work, where we prefer to focus on the issues of integrating the PEM and the air compression unit. However, the option of storing hydrogen onboard is frequently regarded as possible in the mid- and long-term future, a time scale appropriate also for the eventual use of such kind of fuel cell systems.

2 Operating Conditions

The onboard PEM system shall work under variable ambient conditions during the aircraft flight. A typical operating range is shown in Fig. 1. We consider here three representative conditions (see dots in figure): ground operation, cruise conditions, and high altitude, low speed flight (the last called "worst" condition).

The following table (Table 2) shows the ambient conditions in the three cases. As already mentioned, the pressure level of the PEM system shall allow the presence of liquid water (avoiding the risk of boiling). Within a hypothesis of maximum cell local temperature of 80–85°C, the minimum tolerable pressure would be close to 0.57 bar according to water saturation tables; considering possible pressure oscillations, startup, and dynamic load variation issues, the design minimum pressure level can be set at 0.8 bar for the PEM system under flight conditions.

In all cases, we consider to design the fuel cell system for achieving a 60 kW net electric power output at cruise conditions, a value considered by preliminary projects at Airbus [4,8]; however, calculations may be easily extended to the case of a different power output.

3 Calculation Model

3.1 Fuel Cell Model. We make here reference to a PEM stack, which reflects the features of Nuvera fuel cell technology, with operating temperature of 70°C, which has been used for prototypes testing in the range of few kilowatt output as well as for hundred kilowatt-scale units [9].

The FC can be humidified by direct water injection (DWI) at the cathode, without requiring additional humidification at the anode side, because the amount of water used maintains the whole membrane humid, avoiding dehydration problems. The water stream can also act as the cell cooler by sensible heat exchange

Table 2 Selected operating conditions

	1 Ground	2 Cruise	3 Worst
Altitude (m)	0	11000	13000
Mach number	0	0.85	0.6
Static pressure (bar)	1.013	0.226	0.165
Pressure after dynamic inducer (bar)	0.974	0.268	0.182
Static temperature (K)	288.15	216.65	243.77
Temperature after dynamic inducer (K)	288.15	247.96	261.32

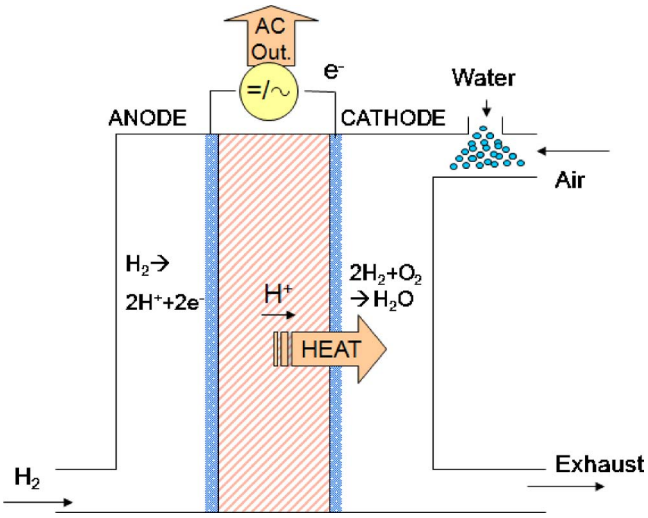


Fig. 2 PEM module operating principles and schematic layout

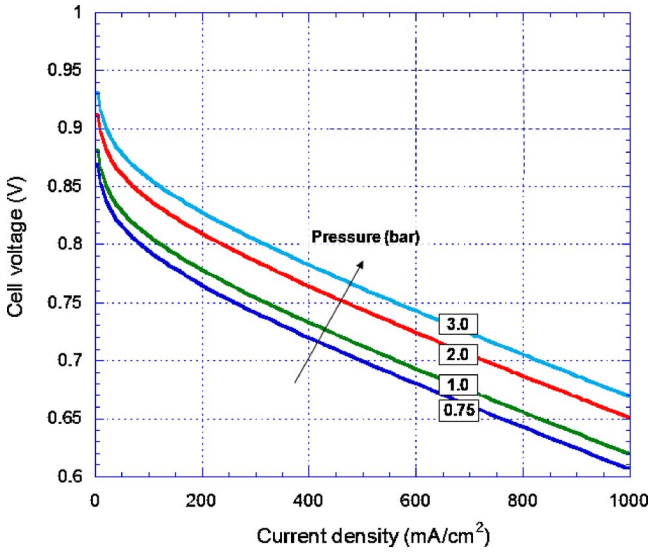


Fig. 3 Cell polarization curves at variable pressure (data for GORE® Primea Series 56 membranes, $T=75°C$ Ref. [16]).

and partial evaporation, with the latter being the dominant effect [10]. Alternatively, the primary cooling effect can be carried out with a separated water loop, a solution preferred, for instance, for automotive applications; in this case, a water-glycol mixture circulates into cooling channels, which are inserted between adjoining cell groups; the coolant is then circulated in an air-cooled heat exchanger before returning to the FC [11,12]. The two options do not affect significantly the system energy balances discussed in this work; however, when applied to aircraft operation, both solutions have their drawbacks: (i) the DWI concept may suffer the pressure variations, which are experienced by the cell stack during real operation, requiring large variations of the amount of water sprayed at the cathode to obtain the cooling effect; moreover, evaporation becomes progressively inefficient when the stack pressure is increased above 1.5–2 bars, given that the cell average temperature cannot be increased above 75–80°C with actual membrane materials, so that heat exchange is primarily carried out by the liquid stream which increases its mass flow rate; (ii) the solution with a closed circuit cooling requires to adopt a saturator on the inlet air flow, which complicates the system layout; the saturator could be avoided during ground operation, but it becomes mandatory when operating with very dry inlet conditions, as typical when the FC is fed with compressed air during high altitude flight.

In this paper, we consider the first solution (cell cooling with water injection), presenting in some cases results of the second alternative. The PEM is fed by an oxidizer at the cathode (as mentioned, air mixed with water sprayed for humidification) and a fuel (Fig. 2). The two fluxes flow in contact with cathode and anode porous surfaces, separated by a solid membrane electrolyte, which is a good H^+ ion conductor. The ionization of molecular hydrogen to H^+ takes place at the anode, thanks to the effect of proper catalysts (generally platinum); hydrogen is then oxidized to steam at the cathode.

If the fuel at the anode inlet is pure hydrogen, as assumed in this paper, the PEM can work in a "dead-end" arrangement: the anode has only an entrance side and no exit (except for periodical purging of accumulated inerts and pressure regulation), and all the hydrogen is used in the fuel cell. If the fuel at the anode inlet is a mixture of hydrogen and other components, hydrogen can be electrochemically oxidized only up to a maximum fuel utilization factor, to avoid the large voltage losses due to reactant concentration gradients and limited gas diffusivity near the electrodes active area. The same consideration applies to the air flow; air and fuel utilization factors (see Nomenclature) quantify the consumed fractions.

The FC generates direct current electrical energy with an efficiency proportional to its voltage. The efficiency is positively affected by pressurized operation, with a gain proportional to the operating pressure, while it decreases by increasing the cell current density. The model calculates the cell electric power production, heat generation, and efficiency together with the thermodynamic properties and chemical composition of the inlet and outlet flows, as a function of the cell operating conditions (temperature, pressure, current density, fuel, and oxidizer composition and utilization factors).

The amount of hydrogen that permeates and reacts with oxygen is calculated by multiplying the mass stream at the anode inlet with the fuel utilization factor. In the same way, the inlet air flow is estimated based on the air utilization factor. The anode outlet mass flow is computed as the difference between the inlet mass flow and the hydrogen permeated; the cathode outlet stream is the sum of the cathode inlet air, the water for cooling, and the hydrogen permeated.

The cell voltage V_c is estimated based on the cell current density i_c by means of the following equation [13]:

$$V_c = A + B \log_{10} i_c + C i_c \tag{1}$$

This formula does not consider the concentration losses, because in this study we assume that the fuel cell operating conditions always fall in the linear portion of the polarization curve, where the dominant effect is a linear resistance loss. The parameters A, B, and C are function of temperature and pressure and they can be expressed by the following equations [14,15]:

$$A = a + b \log_{10} P \tag{2}$$

$$B = c + dT \tag{3}$$

Table 3 Constants in Eqs. (1)–(4)

a	0.779
b	0.104
c	0.0324
d	-2.21×10^{-4}
e	−1.50
f	3.85×10^{-3}

Table 4 PEM model assumptions

$\Delta p/p$ air/fuel side	1%
Heat loss	1%
dc-ac efficiency	94%
Fuel utilization factor U_f	100%
Air utilization factor U_a	50%
Current density i_c	0.4 A/cm^2
Fuel composition	100% H_2

$$C = e + fT \tag{4}$$

The constants $a-f$ of the equations are calibrated by interpolation of experimental data for state-of-the-art PEM membranes (Fig. 3). The final equation, whose constants are reported in Table 3, allows us to reproduce experimental results with an average voltage error below 1% in the regions of interest.

Based on the cell voltage, the stack electric power is calculated by

$$P_{\text{el}} = 2 \times F \times N_{\text{H}_{2,p}} \times V_c \tag{5}$$

where $N_{\text{H}_{2,p}}$ (mol/s) are the moles of hydrogen permeated from anode to cathode. It is then possible to find the cell active area, A_c, as the ratio between power output (P_{el}) and the product of current density and cell voltage ($i_c V_c$).

The amount of water injected for cooling is calculated iteratively with to respect the energy balance of the system, reaching an assigned value of the cathode and anode outlet temperature (75°C). Compositions are calculated based on the quantity of hydrogen permeated.

The following table (Table 4) shows other fuel cell simulation assumptions. Fuel is assumed here to come from pressurized tanks, so that no fuel compressor is required.

3.2 Compressor and Other Components. The system layout includes a compressor, with the function of ralsing the cell pressure up to a desired value, and may include a turbine. These components work with maximum pressure ratios in the range 3.5–4.5, similar to those of radial turbomachines used for gas microturbines [17–19]. However, the air mass flow rate handled in the PEM system considered here is rather low (below 0.1 kg/s, i.e., that of a few kilowatt microturbine), suggesting the necessity to adopt very high speed components [20,21], comparable to those used for the turbochargers of the car industry. The assumptions used for their simulation, together with those of other plant components (heat exchangers), are shown in Table 5 [22–24].

Based on the aircraft operating conditions considered above, the compressor system has to deal with a large variation of pressure ratio (1.2–4.5) and an extremely wide variation of inlet volumetric air flow rate, the latter yielding a sixfold variation of corrected mass flow at compressor inlet (Eq. (6)) going from ground operation to high altitude flight.

$$m_r = m_{\text{in}} \frac{\sqrt{T/T_{\text{ref}}}}{p/p_{\text{ref}}} \tag{6}$$

The typical operating range for centrifugal compressors allows a corrected mass flow variation around 1:3 between the minimum and maximum pressure ratios requested here [22]; this consideration suggests that the compressor system has to be designed with two parallel units, the first operating at ground and low altitudes, and the second added at high altitude when the inlet air volumetric flow rate becomes too large.

4 Plant Configurations and Thermodynamic Results

The proposed plant configuration is shown in Figs. 4(*a*)–4(*c*) with the corresponding energy balances and with the thermodynamic conditions of all the relevant cycle points. The analysis of the complete power cycle is made with a modular simulation code (GS) already described in previous works and tested on a wide variety of gas turbine and fuel cell cycles [25–27].

Table 5 Compressor and other component model assumptions (cases A–C are presented in Sec. 4)

Compression system	
Design conditions (cruise)	
Pressure ratio	3.0
Mass flow at compressor inlet (kg/s)	0.07
Corrected mass flow (Eq. (6), kg/s)	0.274
Compressor isentropic efficiency [6]	0.76
Turbine isentropic efficiency [1]	0.82
Operating range	
Pressure ratio	1.2–4.4
Mass flow at compressor inlet (kg/s)	0.06–0.09
Corrected mass flow (Eq. (6), kg/s)	0.07–0.46
Combustor	
Combustor $\Delta p/p$ (air side, Case C)	3%
Combustion efficiency (Case C)	0.96
Mechanic and electric losses	
Organic efficiency	0.92 (A,B)–0.97 (C)
Electric motor efficiency	0.85 (A)–0.81 (B)
Heat exchangers	
Minimum ΔT (°C)	10
$\Delta p/p$ hot/cold side	1%
Heat loss	1%

The fuel cell is designed to work at a given current density (0.4 A/cm^2, Table 4) at cruise conditions, where the system air mass flow rate is adjusted to achieve the desired power output. The remaining operating conditions are calculated within the hypothesis of keeping the same cell active surface A_c and changing the current density (influencing the cell voltage according to Eq. (1)) to reach the same power output.

4.1 Base Case. The base case arrangement (A) is shown in Fig. 4(*a*). The air flow enters the system from the intake (1), where speed is turned into pressure by a diffuser, and then it is compressed up to the pressure required by the FC. After compression, the air flow (3) shall be cooled to the temperature required by the FC with a heat exchanger, for instance, going from 95°C to 100°C down to 80°C. At the inlet of the cathode (5), air is mixed with water at about 55°C (6) for cell humidification, resulting in a further reduction of temperature. Hydrogen (4) feeds the PEM anode after being depressurized to the same pressure of the air flow[1]. The water required for humidification is separated by condensation from the cathode exhaust stream and recycled; the remaining exhausts are vented in the atmosphere (7). Thermodynamic properties and composition of the main streams are showed in Table 6.

Preliminary optimization of the cell operating pressure has shown, with all the operating conditions of Table 2, that there is no practical gain to operate the fuel cell at higher pressure when the external ambient pressure is lower than the minimum pressure level considered above (0.8 bar): It always happens that the power required to drive the compressor is much higher than the gain obtained with a higher cell voltage and cell power output [24]. Consequently, cell pressure is set at a minimum of 0.8 bar during flight, while at higher ambient pressure conditions (i.e., ground conditions) the system is only slightly pressurized at 1.2 bars to sustain internal pressure drops. The system energy balances are

[1]Fuel preheating at 15°C is necessary when the tank temperature is too low. It is accomplished with a small heat exchanger (not shown in figures for simplicity), recovering heat from the cell cooling loop.

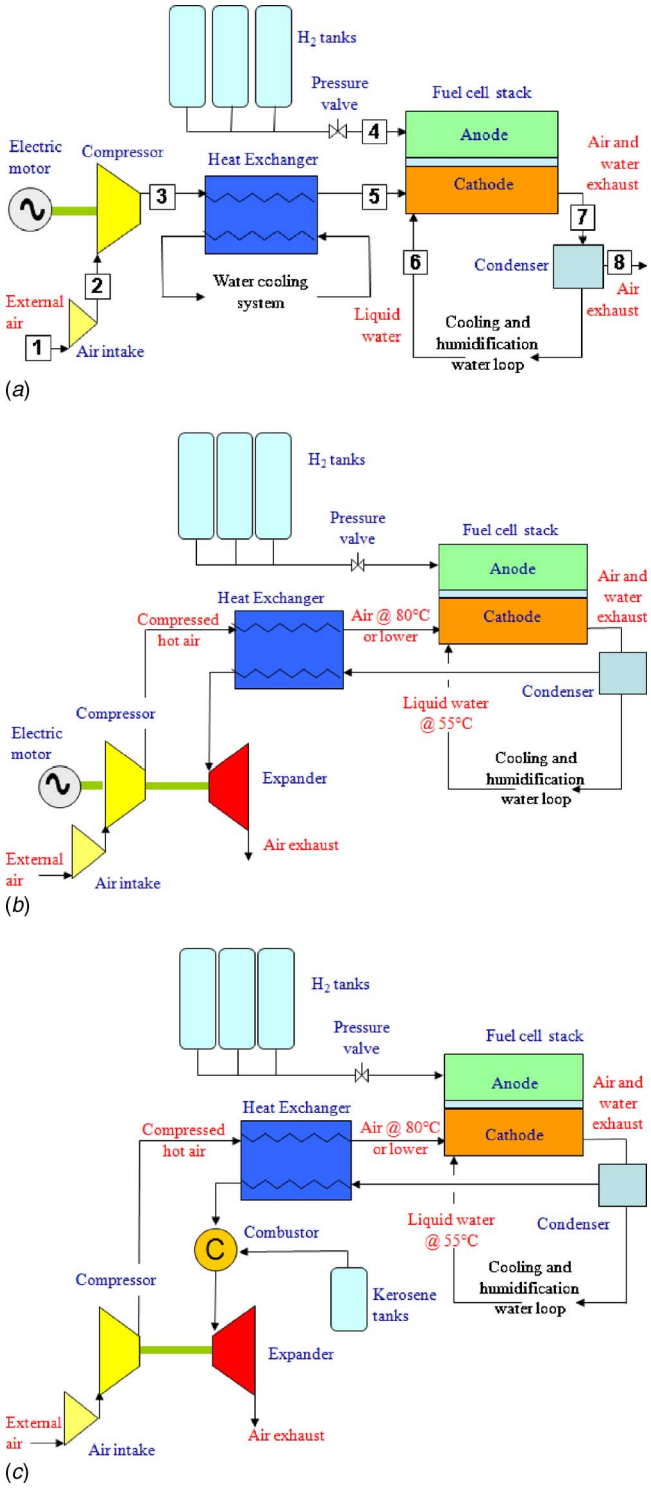


Fig. 4 (*a*) Schematic layout of the base case. (*b*) Schematic layout of the case with turbocharger. (*c*) Schematic layout of the case with combustor and turbochanger.

presented in Table 7. They indicate a design efficiency of 43.5% at cruise conditions, while the efficiency is 52.5% during ground operation, where the compressor consumption becomes negligible.

A possible improvement of this configuration would be to exploit the expansion of the vented gases (point 8 in Fig. 4(*a*)) in a turboexpander; however, direct expansion is not feasible because the low turbine inlet temperature would easily lead to frost formation in the expander. It could be possible to stop the turbine expansion at a higher pressure than atmospheric in order to avoid this problem, but the power gain would be too low to justify the complexity added to the system. The solution of adopting a turboexpander becomes feasible only changing the system layout as presented in the following cases.

4.2 Case With Turbocharger. In the second case (B), shown in Fig. 4(*b*), the cooling of compressed air is done with the cell exhaust gases. This solution allows us to avoid introducing an additional cooling flow, and the temperature of the gases becomes acceptable for a turbine expansion. The resulting TIT is anyway very low, so that expansion power remains lower than compression power; therefore, an electric motor is required to cooperate in driving the compressor. However, the electric power consumption of the motor is significantly lower than in Case A, giving the possibility to achieve a higher system efficiency. The system net electrical efficiency reaches 50.5% at cruise conditions (Table 8), with an operating pressure higher than in Case A.

The system optimal operating pressure is set to maximize the efficiency, as shown in Fig. 5 for the case of cruise conditions, where the optimized pressure of 1.2 bars corresponds to a compressor pressure ratio $\beta=4.5$.

At the other two operating conditions, the effect of pressure is different: Efficiency decreases with pressure in the worst case, due to the extremely low initial pressure at compressor inlet, then the PEM is operated at 0.8 bar (and $\beta\sim4.5$). At ground conditions, pressurization gives instead some advantages, with the efficiency reaching a flat maximum at about 2.5 bars ($\beta\cong2.5$)[2]

Although the electrical efficiency is very high, the system layout is complicated by the concept of powering the compressor partly by the turbine and partly by the electric motor, a solution which would probably face significant control issues under the very different operating points required by this application. However, it should be noted that such arrangement is, in principle, similar (except for the direction of the power flow) to the one commonly used in gas turbines, where the turbine drives the compressor and the alternator; moreover, the electric motor could be used during startup and other transient operation procedures.

4.3 Case With Combustor and Turbocharger. Case C is similar to the previous one, with a combustor added to make the turbocharger always self-sustained (Fig. 4(*c*)). The system layout avoids using the electric motor[3], but the combustor consumes additional fuel; aiming to reduce the hydrogen consumption, we have assumed here to use the same jet fuel already onboard the aircraft to feed this combustor.

In all cases, the system efficiency remains between that of Cases A and B, reaching 48.7% under cruise conditions (Table 9). The cell active area is about 6% lower than in Case B and 20% lower than in Case A, allowing significant savings on the PEM cost, because the fuel cell does not have to sustain the consumption of the electric motor.

The effects of cell pressure on electrical efficiency and the resulting TIT are shown in Figs. 6 and 7. At cruise conditions, results are similar to what has been found in Case B. The required TIT is always very low (i.e., below 550 K), so the combustor should be operated with an extremely large excess air, with the possibility of achieving low NO_x emissions.

At ground conditions, the efficiency rises up to $\beta\cong4$, but the majority of the gain is already achieved at $\beta=2-3$; setting the operating point within this range allows us also to minimize the TIT.

On the other hand, the very low TIT remarks that the gas cycle,

[2]Above this pressure level pressurization would require a substantial increase of the water spray necessary to sustain the DWI cooling concept, as discussed at Sec. 3.1.

[3]This yields also an increase of mechanical efficiency thanks to the absence of a gearbox.

Table 6 Stream data for Case A (Fig. 4(*a*)), cruise conditions

	T (°C)	P (bar)	m (g/s)	Molar fractions (%)					
				Ar	CO_2	H_2	H_2O	N_2	O_2
1	−56.5	0.23	79.2	0.92	0.03	—	1.03	77.28	20.73
2	−25.2	0.27	79.2	0.92	0.03	—	1.03	77.28	20.73
3	95.4	0.81	79.2	0.92	0.03	—	1.03	77.28	20.73
4	15	0.81	1.15	—	—	100	—	—	—
5	80	0.80	79.2	0.92	0.03	—	1.03	77.28	20.73
6	55.4	0.80	25.1	—	—	—	100	—	—
7	75	0.79	105.5	0.57	0.018	—	44.99	47.98	6.44
8	53.8	0.75	80.4	0.83	0.03	—	19.72	70.02	9.39

Table 7 Energy balances for Case A (PEM A_c=27.5 m^2)

	Cruise	Worst	Ground
PEM pressure (bar)	0.8	0.8	1.2
Pressure ratio β	3.0	4.4	1.2
Cell voltage V_c (V)	0.700	0.685	0.722
Cell current density (A/cm^2)	0.400	0.461	0.333
m_{air} (kg/s)	0.079	0.091	0.066
PEM fuel heat input ($m_{fuel,PEM}\times$LHV) (kW)	137.7	158.7	114.7
$P_{el,PEM}$ (kW)	72.36	81.52	62.15
$P_{el,compressor}$ (kW)	12.36	21.52	2.15
Electric efficiency η_{el} (%)	43.57	37.82	52.31

Table 8 Energy balances for Case B (PEM A_c=23.7 m^2)

	Cruise	Worst	Ground
PEM pressure (bar)	1.2	0.8	2.5
Pressure ratio β	4.5	4.4	2.6
Cell voltage V_c (V)	0.719	0.695	0.757
Cell current density (A/cm^2)	0.400	0.421	0.381
m_{air} (kg/s)	0.068	0.072	0.065
PEM fuel heat input ($m_{fuel,PEM}\times$LHV) (kW)	118.7	125.1	113.1
$P_{el,PEM}$ (kW)	64.00	65.20	64.21
$P_{el,motor}$ (kW)	4.00	5.20	4.21
$P_{compressor}$ (kW)	12.11	13.26	7.89
$P_{turbine}$ (kW)	9.13	9.39	4.76
Losses (mech., el.) (kW)	1.02	1.33	1.08
Electric efficiency η_{el}	50.53	47.98	53.04

although self-sustained, has poor thermodynamic quality; this turns into the efficiency decrease—in all operating points—with respect to Case B.

4.4 Comparisons. The energy analysis shows that both Cases B and C, relying on a turbocharger to drive the compressor, achieve better electrical efficiency and reduce the electric power output required to the PEM with respect to Case A. The system layout of Cases B and C is of course more complicated, but given that the specific costs of the PEM are much higher than those of the expander, it is possible that these solutions are preferable also from an economic point of view[4]

Among the two favored cases, plant B reaches maximum efficiency, also featuring zero NO_x and CO emissions, thanks to the absence of any conventional combustor.

[4]A detailed economic analysis, which goes beyond the scope of this paper, should be carried out to investigate this aspect.

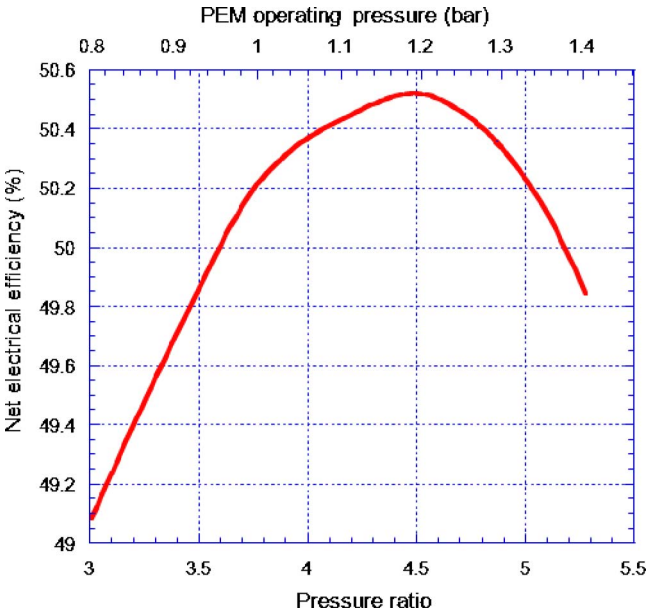


Fig. 5 Electrical efficiency of Case B as a function of pressure at cruise conditions

Table 9 Energy balances for Case C (PEM A_c=22.2 m^2)

	Cruise	Worst	Ground
PEM pressure (bar)	1.2	0.8	2.5
Pressure ratio β	4.5	4.4	2.6
Cell voltage V_c (V)	0.719	0.697	0.757
Cell current density (A/cm^2)	0.400	0.412	0.380
m_{air} (kg/s)	0.064	0.066	0.061
GT TIT (K)	546	593	692
PEM fuel heat input ($m_{fuel,PEM}\times$LHV) (kW)	111.3	114.7	105.7
GT fuel heat input ($m_{fuel,GT}\times$LHV) (kW)	11.8	15.5	21.6
$P_{el,PEM}$ (kW)	60.0	60.0	60.0
$P_{compressor}$ (kW)	11.36	12.16	7.37
$P_{turbine}$ (kW)	11.71	12.54	7.60
Losses (mech.) (kW)	0.35	0.38	0.23
Electric efficiency η_{el}	48.73	46.09	47.12

It is interesting to compare the best PEM configurations (Cases B and C) with a conventional APU. Table 10 shows the energy balances of the plants: both PEM cases have of course a much better efficiency, which allows significant advantages in terms of

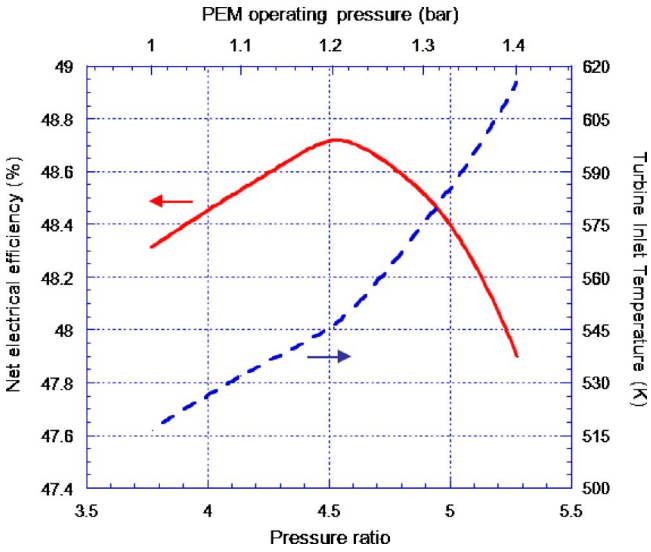


Fig. 6 Electrical efficiency and TIT of Case C as a function of pressure for cruise conditions

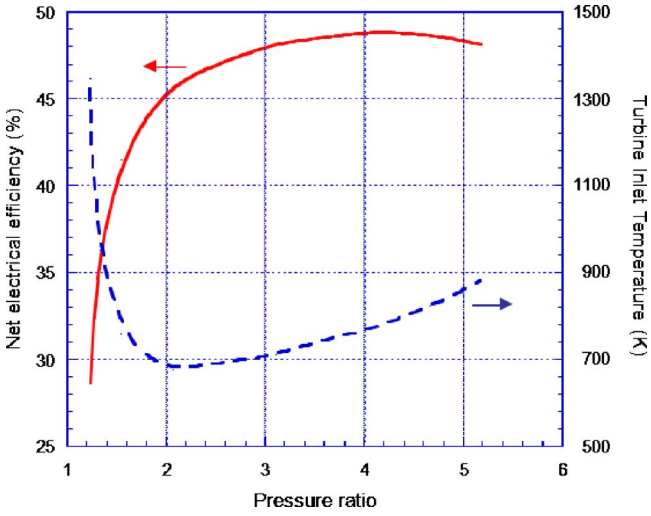


Fig. 7 Electrical efficiency and TIT of Case C as a function of pressure for ground conditions

reducing the weight of the fuel consumed during flight[5]

By the point of view of water consumption, the amount of water sprayed at the PEM cathode for cell humidification and cooling (by partial evaporation and sensible heat exchange) ranges between 375 g/s and 353 g/s for Cases B and C at cruise conditions (Table 10), and can be entirely recovered by condensation of the system exhausts.

For comparison, the alternative of cooling with a separate water circuit would require 624.5 g/s for scenario B and 588.3 g/s for scenario C (with heat exchange only through sensible heat, under the same temperature differences), while reducing the amount of water to be evaporated for the saturation of the cathode flow at about 8.3 and 7.8 g/s, respectively, for Cases B and C.

6 Conclusions

This work has considered the issue of integrating a PEM fuel cell onboard an aircraft for generating a fraction of auxiliary systems electric consumption. Calculations have been performed with a model that takes into account the effects of temperature and pressure on the cell efficiency, optimizing the fuel cell operating point at variable flight conditions. Different system architectures have been considered, where the FC is fed with hydrogen as fuel and compressed air taken from the external surroundings as oxidizer: the first (Case A) with the air fed to the FC by an electrically driven compressor, the other (Cases B and C) relying on a turbocharger to drive the compressor, reaching better electrical efficiency and reducing the size of the PEM. Optimized pressure ratios have been found, with the cell operating pressure ranging from 0.8 bar to 2.5 bars. In all cases, the proposed system yields relevant fuel savings with respect to a conventional gas-turbine APU, with the distinctive advantage of achieving zero NO_x emissions in the most efficient configuration, where net electrical efficiency exceeds 50% at aircraft cruise conditions.

[5]Further analysis would be required to compare the systems in terms of weight at takeoff; for instance, the PEM generator is expected to be significantly—but not extremely—heavier than a conventional APU.

Table 10 Energy balances for a conventional turbine APU (β=3, TIT=1073 K) and for the PEM Cases B and C (cruise conditions)

Energy balances	PEM B	PEM C	GT
System net power output (kW)	60.0	60.0	60.0
PEM fuel heat input ($m_{\text{fuel,PEM}} \times$ LHV) (kW)	118.7	111.3	0.0
GT fuel heat input ($m_{\text{fuel,GT}} \times$ LHV) (kW)	—	11.8	349.3
Hydrogen consumption (kg/h)	3.56	3.34	—
Jet fuel consumption (kg/h)	—	0.99	21.43
m_{air} (kg/s)	0.068	0.064	0.30
Stack outlet temperature (K)	280	391	836
Electric efficiency η_{el}	50.53	48.73	17.75

Acknowledgment

The reported research was conducted at the Department of Energy of Politecnico di Milano within the framework of a program agreement for M.Sc. stages active with Airbus. The authors wish to thank Professor A. Coghe of Politecnico di Milano and Dr. Guenter Walper of Airbus Deutschland for their cooperation.

Nomenclature

A_c = cell active area (cm^2)
F = Faraday's constant (96 439 C/mol of electrons)
i_c = cell current density (A/cm^2)
m = mass flow rate (kg/s)
p = pressure (Pa)
P_{el} = electric power (MW)
Q_{th} = thermal power (MW)
T = temperature (°C or K)
U_a = cell air utilization factor: $U_a = O_{2,\text{consumed}}/O_{2,\text{inlet}}$
U_f = cell fuel utilization factor: $U_f = H_{2,\text{consumed}}/H_{2,\text{inlet}}$
V_c = cell voltage (V)
β = pressure ratio
η_{el} = electric efficiency (LHV base)

Acronyms

dc/ac = direct/alternating current
FC = fuel cell
GT = gas turbine
LHV = lower heating value (kJ/kg)
PEM = polymer electrolyte membrane fuel cell
TIT = turbine inlet temperature

References

[1] Eelman, S., Pozo del Poza, I., and Kreig, T., 2004, "Fuel Cell APU'S in Commercial Aircraft-An Assessment of SOFC and PEMFC Concepts," 24th International Congress of Aeronautical Sciences (ICAS).

[2] Daggett, D. L., Eelman, S., and Kristiansson, G., 2003, "Fuel Cell APU for Commercial Aircraft," AIAA/ICAS International Air and Space Symposium and Exposition, OH, July.

[3] Anon, 2005, "6th European Framework Program, PROPAIR: Processing of Kerosene for PEM Fuel Cell Applications for Aircraft," Airbus Internal Document, Part B, July.

[4] Anon, 2005, "6th European Framework Program, CELINA: Fuel Cell Integration in a new Configured Aircraft," Description of Work-Airbus Internal Documents, February.

[5] Anon, 2006, "787 Special-Electric Dream," Flight International, pp. 58–59; http://www.flightglobal.com/articles

[6] Barchewitz, L. P., and Seume, J. R., 2006, "Conceptual Analysis of Air Supply Systems for In-Flight PEM-FC," ASME Paper No. 2006-GT-90688.

[7] Cunningham, J. M., Hoffman, M. A., and Friedman, D. J., 2001, "A Comparison of High-Pressure and Low-Pressure Operation of PEM Fuel Cell Systems," University of California, SAE Engineering Congress.

[8] Anon, 2005, "Fuel Cell Systems for Emergency Power Supply Onboard of an Aircraft," European Consortium for Advanced Training in Aerospace, ECATA, Airbus Internal Document, Final Report.

[9] Anon, 2006, "Avanti and Forza PEM Modules," Nuvera website: http://www.nuvera.com

[10] Nuvera Europe Technical Management and Experimental Laboratories, 2006 (private communication).

[11] Anon, 2006, *Fuel Cell Handbook*, 7th ed. U.S. Department of Energy, Morgantown, WV.

[12] S., Campanari, E., Macchi E., and A., Silva, 2005, *Micro-Cogeneration With Natural Gas*, Polipress, Milan, Italy (in Italian).

[13] Larminie, J., and Dicks, A., 2003, *Fuel Cell Systems Explained*, Wiley, New York.

[14] Kulp, G. W., 2001, "A Comparison of Two Air Compressors for PEM Fuel Cell Systems," MS thesis, Mechanical Engineering, Virginia.

[15] Pratt, J., Brouwer, J., and Samuelsen, G., 2003, *Experimental Evaluation and Computer Simulation of an Air-Breathing PEM Fuel Cell at Aircraft Flight Altitudes*, National Fuel Cell Research Centre, Irvine, CA.

[16] Anon, 2006, "Primea Series 56 Membranes," Website: www.gore.com.

[17] Campanari, S., 2000, "Full Load and Part-Load Performance Prediction for Integrated SOFC and Microturbine Systems," ASME J. Eng. Gas Turbines Power, **122**, pp. 239–246.

[18] O'Brien, P., 1998, "Development of a 50-Kw, Low-Emission Turbo-Generator for Hybrid Electric Vehicles," ASME Paper No. 98-GT-400.

[19] Jones, A. C., 1994, "Design and Test of a Small, High Pressure Ratio Radial Turbine," ASME Paper No. 94-GT-135.

[20] Rodgers, C., "25 to 5 kWe Microturbine Design Aspects," ASME Paper No. 2000-GT-0626.

[21] Gerendas, M., and Pfister, R., 2000, "Development of a Very Small Aeroengine," ASME Paper No. 2000-GT-536.

[22] Balje, O. E., 1981, *Turbomachines-A Guide to Design, Selection and Theory*, Wiley, New York.

[23] Perdichizzi, A., and Lozza, G., 1987, "Design Criteria and Efficiency Prediction for Radial Inflow Turbines," ASME Paper No. 87-6T-231.

[24] Beretti, A., 2006, "Analysis of Turbocharged PEM Fuel Cell Systems for Civil Aircraft Onboard Power Production," MS thesis, Politecnico di Milano.

[25] Consonni, S., Lozza, G., Macchi, E., Chiesa, P., and Bombarda, P., 1991, "Gas-Turbine-Based Advanced Cycles for Power Generation-Part A: Calculation Model," *International Gas Turbine Conference*, Vol. III, Yokohama, pp. 201–210.

[26] Campanari, S., and Macchi, E., 1998, "Thermodynamic Analysis of Advanced Power Cycles Based Upon Solid Oxide Fuel Cells, Gas Turbines and Rankine Bottoming Cycles," ASME Paper No. 98-GT-585.

[27] Chiesa, P., and Macchi, E., 2004, "A Thermodynamic Analysis of Different Options to Break 60% Electric Efficiency in Combined Cycle Power Plants," ASME J. Eng. Gas Turbines Power, **126**, pp.770–784.

Time Characterization of the Anodic Loop of a Pressurized Solid Oxide Fuel Cell System

A. Traverso
e-mail: alberto.traverso@unige.it

F. Trasino
e-mail: francesco.trasino@unige.it

L. Magistri
e-mail: loredana.magistri@unige.it

A. F. Massardo
e-mail: massardo@unige.it

Thermochemical Power Group,
Dipartimento di Macchine Sistemi Energetici e Trasporti,
via Montallegro 1,
Università di Genova,
1614S Genova, Italy

A dynamic solid oxide fuel cell (SOFC) model was integrated with other system components (i.e., reformer, anodic off-gas burner, anodic ejector) to build a system model that can simulate the time response of the anode side of an integrated 250 kW *pressurized SOFC hybrid system. After model description and data on previous validation work, this paper describes the results obtained for the dynamic analysis of the anodic loop, taking into account two different conditions for the fuel flow input: in the first case (I), the fuel flow follows with no delay the value provided by the control system, while in the second case (II), the flow is delayed by a volume between the regulating valve and the anode ejector, this being a more realistic case. The step analysis was used to obtain information about the time scales of the investigated phenomena: such characteristic times were successfully correlated to the results of the subsequent frequency analysis. This is expected to provide useful indications for designing robust anodic loop controllers. In the frequency analysis, most phase values remained in the* 0–180 deg *range, thus showing the expected delay-dominated behavior in the anodic loop response to the input variations in the fuel and current. In Case I, a threshold frequency of* 5 Hz *for the pressure and steam to carbon ratio and a threshold frequency of* 31 Hz *for the anodic flow were obtained. In the more realistic Case II, natural gas pipe delay dominates, and a threshold frequency of* 1.2 Hz *was identified, after which property oscillations start to decrease toward null values.* [DOI: 10.1115/1.2772638]

Introduction

High temperature fuel cells, both molten carbonate fuel cells (MCFCs) and solid oxide fuel cells (SOFCs), are expected to play an important role in the production of highly efficient distributed power generation in the near future.

It is germane to quickly summarize the reasons that make these systems attractive:

- clean and environmentally friendly power generation (almost zero emissions of NO_x, SO_x, CO, and UHC)
- high performance (efficiency well over 60% in a midterm perspective)
- possible application in distributed power generation (efficiency is almost independent of size)
- efficient exploitation of hydrogen as a fuel

The integration of turbomachinery is necessary to increase fuel cell efficiency, to reduce plant dimensions, and possibly to lower specific capital costs as well. The resulting hybrid system certainly constitutes a concentration of hazardous components that need to be managed in a synergetic way under all the main operating conditions, which are startup, idle operation (i.e., zero net power), power generation, and shutdown. Furthermore, the fuel cell itself may require particular operational steps, such as the initial conditioning period, but these are more pertinent to the fuel cell stack than to the entire hybrid system. Extensive modeling and performance analyses have previously been carried out by the authors [1,2]. The present challenge is to define and assess the control strategy for the entire hybrid system: this implies both an off-design analysis to determine the operational envelope in which the system can be safely operated at part load and a dynamic analysis in order to understand and control the main time-dependent phenomena occurring during transition from one regime to the next. The present work is concerned with the first step toward a complete control-oriented time characterization of a pressurized SOFC hybrid system: this step is the analysis of the anode loop.

The system analyzed in this paper is shown in Fig. 1 and includes an anodic ejector, a reformer, and a SOFC stack. A turbomachinery and air management system (recuperator, ejector, etc.) is not included in the model, because it is primarily intended to control the average fuel cell stack temperature [3], thus having an impact on the relatively long-term response of the system. The anodic loop, by contrast, is supposed to show a quicker response due to the relatively small volume it occupies (actually, this is strongly dependent on the SOFC stack technology); moreover, the quick response is also a requirement for best fitting variations in external electrical load. Therefore, a detailed analysis of the anodic loop may highlight the limitations in load following the SOFC generation package, as well as providing useful information for the development of a control system [4]. In the anodic loop analysis, the conditions of the cathodic side were fixed, representing a sort of boundary condition for the system: it is clear that in an actual system, the cathodic and anodic sides are strongly interdependent, but this creates strong interrelationships between anode and cathode sides [3] and does not allow a straightforward interpretation of the anodic loop free response. The approach of the present study, instead, separates the anodic loop behavior from the rest of the plant, thus providing information on the time-dependent characteristics of this part of the system.

Integrated Reformer Solid Oxide Fuel Cell Stack Model

The detailed integrated SOFC model (detailed stack model) was developed in a MATLAB-SIMULINK® environment through a specific mex-S function. It considers a detailed geometry of the pressurized planar Rolls-Royce fuel cell system (RRFCS) SOFC stack

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 5, 2007; final manuscript received May 9, 2007; published online February 29, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007, Paper No. GT2007-27135.

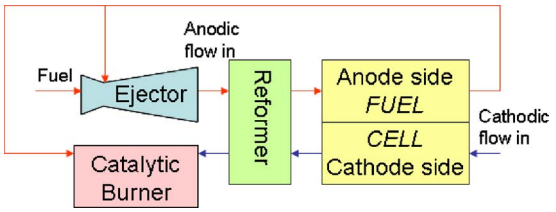


Fig. 1 Layout of SOFC system anode loop

[5], shown in Fig. 2, including a reforming unit [6].

Due to the high level of modeling detail, the computational time for such a model proved to be too high for real-time applications. Hence, the development of a new multi-control-volume SOFC stack model (dynamic stack model), still in the MATLAB-SIMULINK® environment, retaining the basic dynamic information but avoiding the requirement of an in-depth knowledge of the internal geometry of the stack, is effective in speeding up the time-dependent simulations to almost real-time calculations. The previous detailed model was used as a verification tool for both the on-design and off-design analyses of the dynamic model.

The dynamic model integrates stack and reformer. The main assumptions underlying the model are the following.

- The cell is considered adiabatic and the current density is uniform.
- All the chemical reactions are considered at equilibrium.
- The electrochemical reaction of H_2 is considered while the reaction of CO is neglected.
- All the thermophysical properties of the gas streams are evaluated on the basis of the average stream composition, temperature, and pressure in each control volume.
- The internal distribution of the anodic and cathodic flows approximates the actual distribution of the detailed model.
- Electric response of fuel cell is instantaneous; no electrical delay (accumulation of positive/negative charges) is considered.
- Leakages of oxidant or fuel through the electrolyte are included.

The SOFC stack is modeled with a discrete number of control volumes: the energy (1) and mass balances (2) of all the streams are calculated between the inlet and outlet of each control volume while the solid thermal transient is calculated through Eq. (3).

$$T^{\text{out}} = \frac{\sum_i N_i^{\text{in}} C_{p,i} T^{\text{in}} + Sl \sum_k r_k (-\Delta H_k) - W_{\text{el}}}{\sum_i N_i^{\text{out}} C_{p,i}} \tag{1}$$

$$N_i^{\text{out}} = N_i^{\text{in}} + Sl \sum_k \nu_{i,k} r_k \tag{2}$$

$$\frac{d(T_{\text{solid}})_i}{dt} = \frac{(q_{\text{solid}})_i}{(C_{v\_\text{solid}})_i} \tag{3}$$

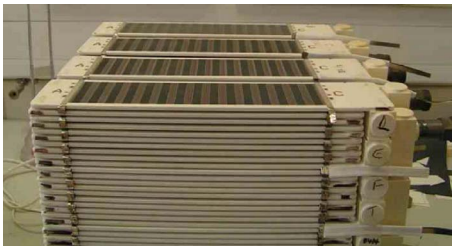


Fig. 2 A block of RRFCS fuel cell strips [5]

Table 1 Design-point conditions

Property	Value
Fuel temperature	400 °C
Ejector design recirculation factor (secondary over primary flow)	7
Steam to carbon ratio (reformer inlet)	2.4
Anodic side pressure	7.0×10^5 Pa
Reformer inlet temperature	800 °C
Anodic side outlet temperature	900 °C
Fuel cell power	approximately 250 kW

The number of control volumes (4 for the whole stack) is kept to a minimum in order to minimize model complexity without compromising the accuracy of the results. This model simplification greatly improved calculation time without affecting the overall accuracy of system results.

The operating voltages are calculated by subtracting the local values for activation, Ohmic, and diffusion losses from the local Nernst potential: related equations are briefly reported in a simplified version in Eqs. (4)–(7).

$$E_{\text{Nernst}} = -\frac{\Delta G(T)}{nF} + \frac{R_u T}{nF} \ln\left(\frac{P_{\text{H}_2} P_{\text{O}_2}^{1/2}}{P_{\text{H}_2\text{O}}}\right) \tag{4}$$

$$\Omega_{\text{cat act}} = \frac{R_u T}{F} \frac{1}{K_{\text{O}_2} P_{\text{O}_2}^{0.25} e^{k_{\text{O}_2}/TR_u}} \tag{5}$$

$$\Omega_{\text{an act}} = \frac{R_u T}{F} \frac{1}{K_{\text{H}_2} P_{\text{H}_2} P_{\text{H}_2\text{O}} e^{k_{\text{H}_2}/TR_u}} \tag{6}$$

$$\rho_i = A_i e^{-B_i/T} \tag{7}$$

Steam reforming, water-gas shift, and hydrogen electrochemical reaction are simultaneously solved for each control volume; due to the high operating temperature of the stack, steam reforming is considered to be at a fixed approach (>95% overall) to equilibrium. The reaction rates for the reforming and shifting are calculated through Eqs. (8) and (9), which provide the equilibrium of the reforming and shifting reactions.

$$K_{p,\text{ref}} = \frac{P_{\text{CO}} P_{\text{H}_2}^3}{P_{\text{CH}_4} P_{\text{H}_2\text{O}}} \tag{8}$$

$$K_{p,\text{shift}} = \frac{P_{\text{CO}_2} P_{\text{H}_2}}{P_{\text{CO}} P_{\text{H}_2\text{O}}} \tag{9}$$

The reforming unit is also modeled as a multinode reformer (ten nodes) because this solution provides an insight of the temperature profile inside the reformer, when compared to a single-node reformer model: the average temperature of the reformer is still the same, but the temperature of the anodic stream entering the stack is higher in the second case. This allowed a better simulation of the overall thermal balance of the stack without significantly increasing the computational effort.

The SOFC technology, as well as the internal stack arrangement, is the proprietary planar pressurized SOFC stack of RRFCS [6]. For this reason, at this stage, no kinetic model has been included because comprehensive data on the catalyst used are not available yet. Nevertheless, it is believed that the introduction of the kinetic model, which will be the subject of further investigation, will change the amplitudes but not the characteristic times of the phenomena.

The new dynamic model was verified against results from the detailed model at steady state, both at design point and at part-load conditions. Table 1 compares some reference data of the design point for both models.

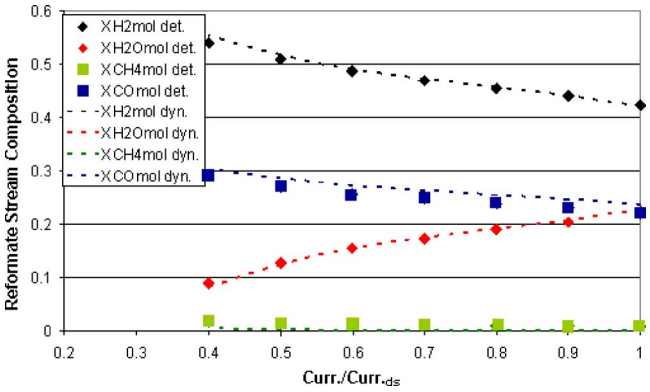


Fig. 3 Reformate stream composition under part-load conditions

Figures 3 and 4 show the behavior of some operating parameters under part-load conditions: the off-design operating conditions of the SOFC system were extracted from the study in Ref. [6], where changes in air flow, utilization factor, stack temperatures, etc., are scheduled according to a complex procedure. Voltage data were nondimensionalized over the relative design-point data.

Figure 3 presents the reformate stream composition for both the detailed model (dots) and the dynamic model (dotted lines) at the reformer outlet: in particular, the dynamic model seemed to overestimate the reforming and underestimate the shifting. Overall, the agreement looks satisfactory for a wide range of operating conditions; this should ensure that the dynamic model can depict the thermal behavior of the reformer with an acceptable degree of accuracy.

Figure 4 presents the voltage and SOFC efficiency under different load conditions, showing good agreement between the dynamic model and the detailed model results. Further results of model comparison can be found in Ref. [7].

Anode Ejector Model

Two dynamic models have been developed and validated at the thermochemical power group (TPG) laboratories: the 1D source-term model and the so-called "lumped-volume model." Extensive description and experimental validation may be found for both models in Refs. [8–10]. The latter model was used for the purposes of the present work, because it allows significantly faster calculation time with proven accuracy in representing the overall ejector behavior, despite the fact that no description of the ejector's internal properties is provided.

The ejector lumped-volume model was implemented within the MATLAB®-SIMULINK® environment, as a component of the

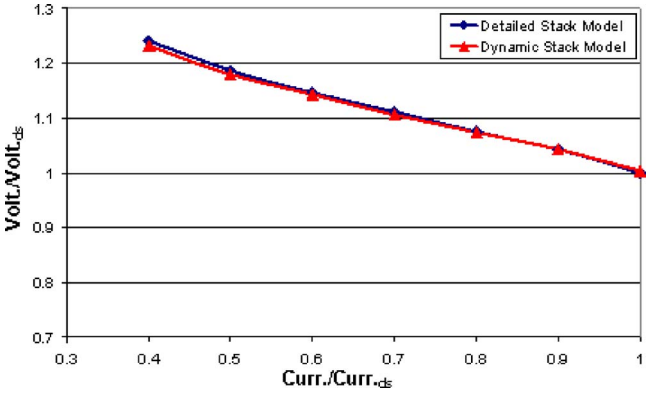


Fig. 4 Operating voltage under part-load conditions

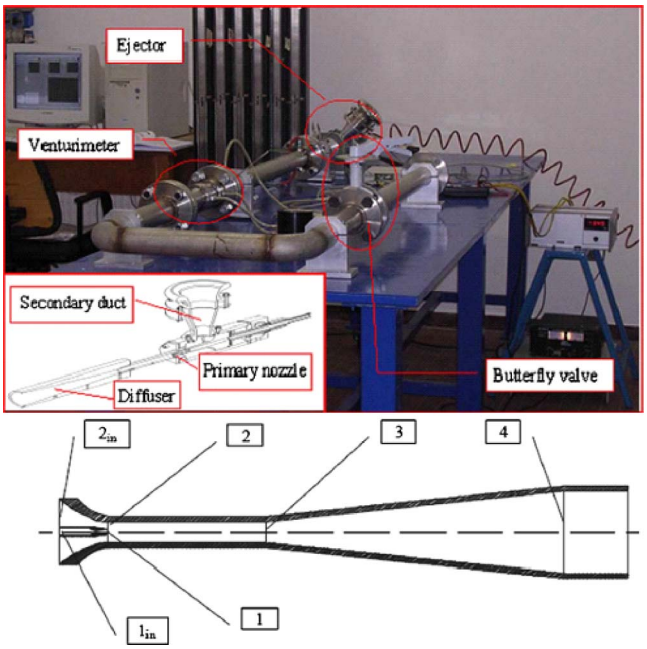


Fig. 5 Ejector scheme and TPG ejector test rig in open loop configuration [9]

TRANSEO tool. The "lumped-volume" approach [8] is a simplified scheme for constant mixing chamber ejectors, developed with the objective of achieving the reduction of calculation time, which is a necessary requirement when simulations are carried out at plant level. The model is based on mass, momentum, and energy steady-state equations, taking into account the primary momentum loss with a constant coefficient, the viscous pressure losses into the mixing chamber (10), and the diffuser (11) (see Fig. 5 for the subscripts).

$$(\Delta P_f)_{\text{mix}} = f_{\text{mix}} \frac{L_{\text{mix}}}{D_{\text{mix}}} (\rho_{\text{av}})_{\text{mix}} \frac{[(c_{\text{av}})_{\text{mix}}]^2}{2} \tag{10}$$

$$(\Delta P_f)_{\text{diff}} = \frac{C_f}{4\alpha}\left[1 - \left(\frac{A_3}{A_4}\right)^2\right] + \alpha\left(1 - \frac{A_3}{A_4}\right)^2 (\rho_{\text{av}})_{\text{diff}} \frac{[(c_{\text{av}})_{\text{diff}}]^2}{2} \tag{11}$$

A pressure loss has been considered by an aerodynamic loss (12) to take into account the effect of the primary duct thickness.

$$(\Delta P)_{\text{th}} = C_a \rho_2 \frac{c_2^2}{2} \tag{12}$$

The time-dependent behavior is evaluated through the momentum and energy equations, also taking into account the heat exchange between walls and environment.

The test rig shown in Fig. 5 was used both in open loop and closed loop configurations to compare simulations with experiments: in the closed loop case, a pressure vessel was introduced to physically simulate, on a reduced scale, the anodic volume of the actual SOFC anode loop. Results of transient validation are reported in Ref. [10] and summarized in Fig. 6.

Dynamic Analysis

The dynamic reformer-SOFC model was integrated with the ejector model according to the layout in Fig. 1, in order to investigate the time-dependent behavior of the anodic loop. The system response was investigated for two different variations: a step variation and a sinusoidal variation at different frequencies in fuel and current inputs.

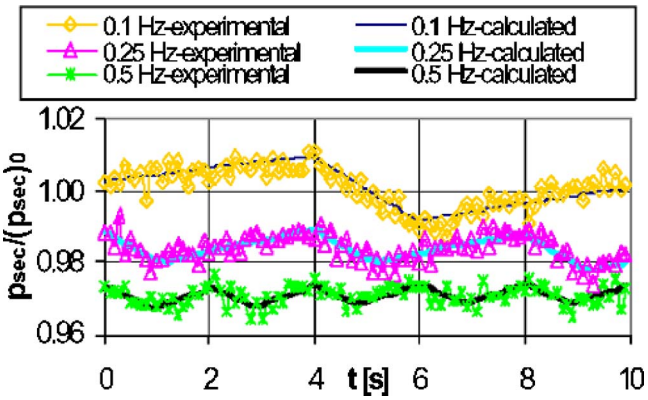


Fig. 6 Ejector model validation under unsteady conditions in closed loop configuration [10]

Fuel and current inputs to the model were proportionally varied in order to keep the ratio constant. Even if this control strategy, which is at the constant utilization factor, may not be feasible over the entire range studied (100–70% load) [3], the time response of the anode loop is not likely to change significantly if such a factor is changed.

The fuel cell current is assumed to perfectly follow the current input from the controller, as no delay is introduced here; instead, regarding the natural gas flow at ejector inlet, two cases were studied for both step and frequency analyses.

Case I. The fuel pipe volume is not included, that is, no delay occurs between the fuel flow input of the controller and the fuel flow occurring at ejector primary nozzle.

Case II. The fuel pipe volume is included.

Case I allows better highlighting of the dynamic phenomena occurring within the SOFC anode loop, which makes it interesting although less realistic than Case II. In the following analyses, Case I is first investigated in detail, then it is compared to Case II: in general terms, the more "nervous" and prompt response in Case I is shown to be mitigated by the pipe volume in Case II.

Dynamic Analysis: Step Variation

In order to characterize the time response of the anodic side of the fuel cell from a time perspective, the response to a step variation in fuel and current was studied. The design point is illustrated in Table 1: from such design operating conditions, fuel and current have been reduced simultaneously to 70% of the original value, so keeping the fuel utilization factor constant. No variation was made to the cathodic side to avoid influencing the free response of the anodic side: it is obvious that in an actual hybrid system, the cathodic side flow conditions have to be varied in order to keep the fuel cell stack temperature under control (e.g., stack inlet and/or outlet air temperature); however, this is out of the scope of the present work.

Table 2 reports the assumed volume and thermal capacitances of the anode loop: some of those properties have been parametrically changed in order to assess their influence on the step response of the system. It is evident that the total thermal capacitance of the fuel cell is so large that its effect is marginal on the relatively short time-scale phenomena under consideration in this step analysis: nevertheless, it was included in the calculations.

Table 2 Geometric and thermal capacitance assumptions

Property	Value
Total anodic volume	0.3 m^3
Total natural gas pipe volume (only Case II)	0.0025 m^3
Total anodic flow path length (including recirculation)	2.5 m
Average anodic flow path section	0.020 m^2
Anodic ejector thermal capacitance	11.3 kJ/K
Total fuel cell thermal capacitance	2000 kJ/K

Table 3 Resulting characteristic time constants

Component	Time constant type	Charact_time (s)
Natural gas pipe (only Case II)	Filling	0.85
	Thermal	0.74
	Flow	0.002
Anodic ejector	Filling	0.012
	Thermal	0.010
	Solid_thermal	55.0
SOFC anodic side	Flow	0.032
	Filling	5.1
	Thermal	4.1

Table 3 reports the characteristic times, calculated as described in Ref. [11]. It is germane to point out that such time constants are calculated as a result of the geometric characteristics of the component, and they are not used in the calculations: they only provide approximate information about the expected time scales of the phenomena under simulation for the software user. With regard to a sample pipe, the meaning of each time constant can be summarized as follows:

Flow. This represents the approximate time for the mass flow to reach about 63% of the new steady-state value after an inlet pressure step is applied.

Filling. This represents the approximate time for the outlet composition (e.g., mass fractions) to reach about 63% of the new steady-state values after an inlet composition step is applied.

Thermal. This represents the approximate time for the outlet temperature to reach about 63% of the new steady-state value after an inlet temperature step is applied; this time constant is only related to the thermal capacitance of the fluid contained in the sample pipe.

Solid_thermal. This represents the approximate time for the pipe wall temperature to reach about 63% of the new steady-state value after an inlet flow temperature step is applied; this time constant is only related to the thermal capacitance of the pipe wall.

With regard to Table 3, natural gas pipe (present only in Case II) shows characteristic times in the order of 1 s, which implies that faster input variations are likely to be filtered by such a volume.

The anode ejector has a very fast time response, which is not likely to interfere with the overall fluid-dynamic behavior of the system. Only the ejector "solid_thermal" characteristic time is quite high: in fact, this is due to the ineffective heat exchange with the evolving flow, but the relatively small heat capacitance (Table 2) makes its contribution almost negligible.

On the other hand, the characteristic time constants of the anodic side of the SOFC are likely to be the driving factor of the anodic side loop transients. The "flow" constant is very small, which causes the anodic flow to quickly update to the new regime: this is expected to be beneficial from the point of view of the control system, because the risk of depleting the available hydrogen on the anodic side should be minimized. The "filling" and "thermal" characteristic time constants are related to the overall volume of the anodic side: their influence is expected to be on the composition and thermal behavior of the whole gaseous loop. Instead, the pressure is proportionally influenced by the same volume, but its characteristic time cannot be generally defined because it depends on the specific case: it will be determined on the basis of the simulation results for the step variation.

The results of a step decrease from 100% to 70% and then back to 100% on the fuel flow and current are reported from Figs. 7–16 for both Cases I and II.

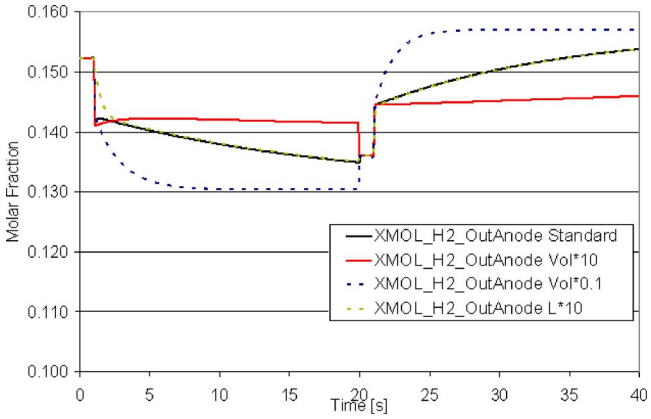


Fig. 7 Molar fraction of residual hydrogen at stack anodic outlet (Case I)

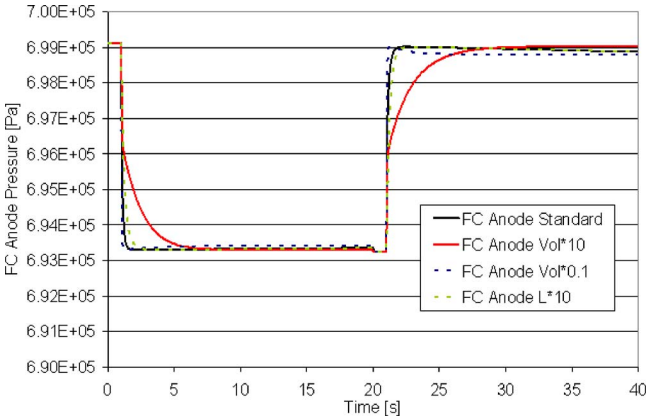


Fig. 10 Anodic side average operating pressure (Case I)

The step analysis had the following pattern in all the simulations.

a. The system was at the steady-state at time "0."
b. After 1 s, the fuel flow and current were decreased to 70% of the initial value.
c. This reduced value was maintained for 19 s.
d. After 20 s, the system was restarted at the final steady-state value (the transient to regime collapsed).
e. After 21 s, the fuel flow and current were restored to the original values.
f. The simulations were stopped after 40 s.

The system was simulated with standard as well as modified geometrical values: the anodic volume ("vol") was increased and decreased by a factor of 10, while the anodic loop length was increased by a factor of 10. No reduction in the anodic loop length was considered because it already showed a very quick response (see Table 3, SOFC anodic side flow characteristic time constant).

A general conclusion may be drawn from Fig. 7, which refers to the hydrogen fraction remaining at SOFC stack outlet in Case I: the anodic volume works out as hydrogen storage, as it has a significant influence on the trajectory between the two steady-state points for hydrogen behavior at the SOFC stack outlet. The transient between the two regimes is actually driven by the anode side volume: the lower fraction of hydrogen at 70% load is due to the

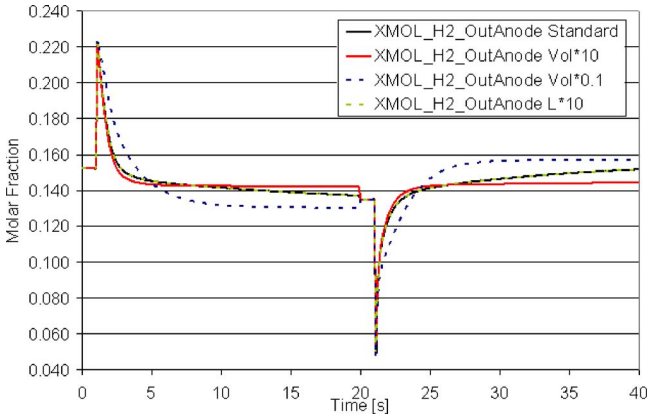


Fig. 8 Molar fraction of residual hydrogen at stack anodic outlet (Case II)

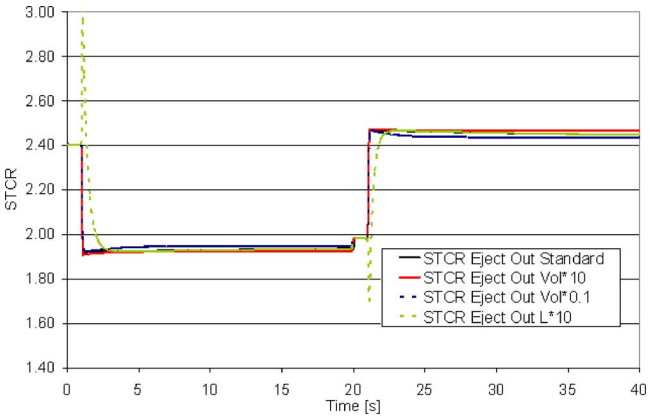


Fig. 11 STCR at anodic ejector outlet (Case I)

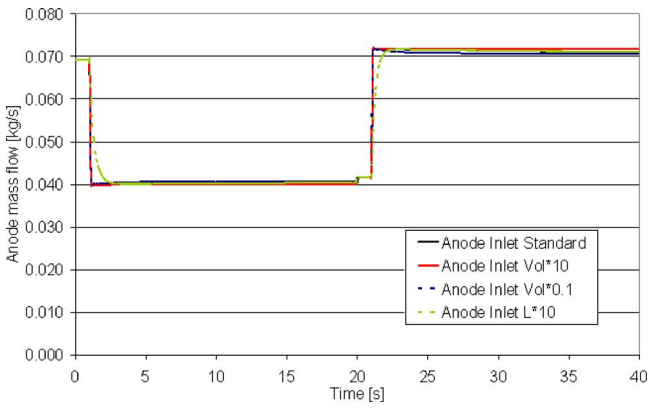


Fig. 9 Anodic side inlet mass flow (Case I)

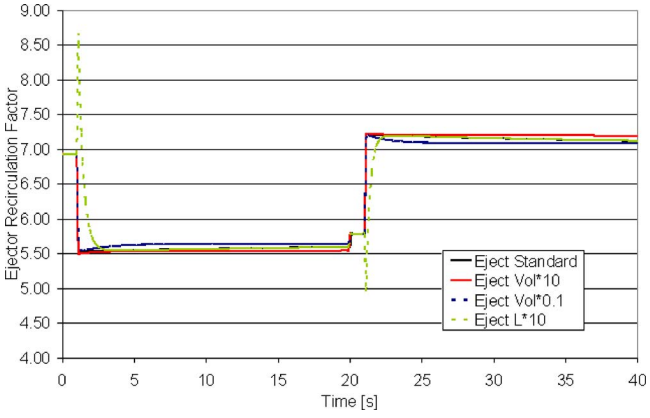


Fig. 12 Anodic ejector recirculation factor (Case I)

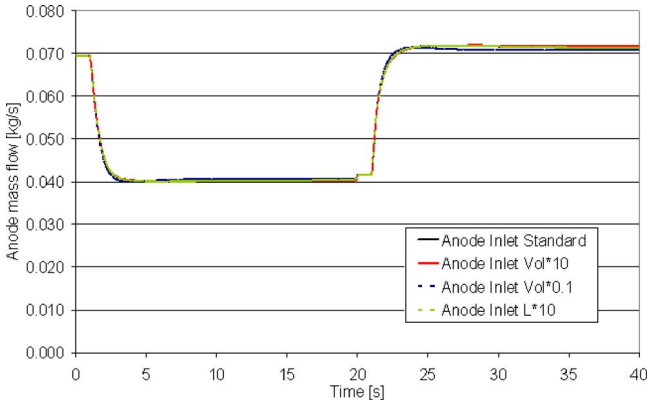


Fig. 13 Anodic side inlet mass flow (Case II)

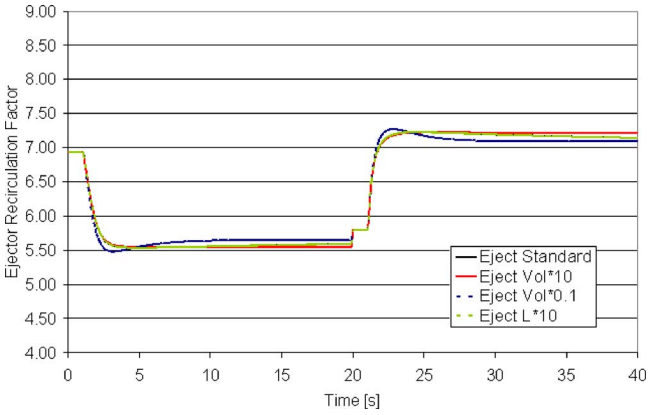


Fig. 16 Anodic ejector recirculation factor (Case II)

lower ejector recirculation factor (Fig. 12) and to the greater relevance of hydrogen leakages. Such a transient is characterized by the filling characteristic time (5.1 s for standard geometry), which is enhanced by the ejector recirculation: in fact, the characteristic time was calculated on the overall anodic outflow, while the actual flow modifying the anode exit composition is constituted by the flow through the electrolyte, which is about one-fourth of anode outlet flow; thus, it follows that the actual filling characteristic time should be about $5.1 \times 4 = 20.4$ s, which can be inferred from Fig. 7 for the standard geometry. Figure 8 refers to the same depleted hydrogen fraction in Case II: it is evident that the delay in natural gas flow, with respect to current, causes such peaking behavior, while conclusions on the actual filling characteristic time remain the same.

Further property behaviors are provided from Figs. 9–12 for Case I, while the same properties are reported for Case II from Figs. 13–16.

Changes in anode mass flow are clearly influenced only by the overall anode side length (Fig. 9), while the pipe volume of Case II (Fig. 13) almost cancels this effect because of its filling time constant (0.85 s), which is larger than the anode side enhanced flow time ($0.032 \times 10 = 0.32$ s).

With regard to pressure, the time scale (steady to steady) is about one-twenty-fifth the calculated filling characteristic time (about 0.2 s, Fig. 10): this result is very useful in the next frequency analysis. Figure 14 for Case II shows that, in this case also, natural gas flow delay dominates, apart from the case of enhanced anodic side volume: in fact, 0.85 s is lower than $0.2 \times 10 = 2$ s.

During load reduction, that is the first step, the momentum delay (Fig. 9) causes the recirculation factor (Fig. 12) to remain above the steady state due to the anodic loop length: this effect is also shown by steam to carbon ratio (STCR) (Fig. 11), which clearly peaks for the anodic side enhanced length case. Peaks of STCR toward low values are dangerous for carbon deposition into the reformer: however, this problem may arise only with very long anodic loop paths, as the standard geometry case shows an almost immediate flow and recirculation factor accommodation to the new regime, and the anodic side enhanced length case shows overshooting periods of no more than 1 s. Figs. 15 and 16 for Case II demonstrate that natural gas pipe delay may be helpful to smooth out "peaky" dynamic behaviors for STCR and recirculation factor.

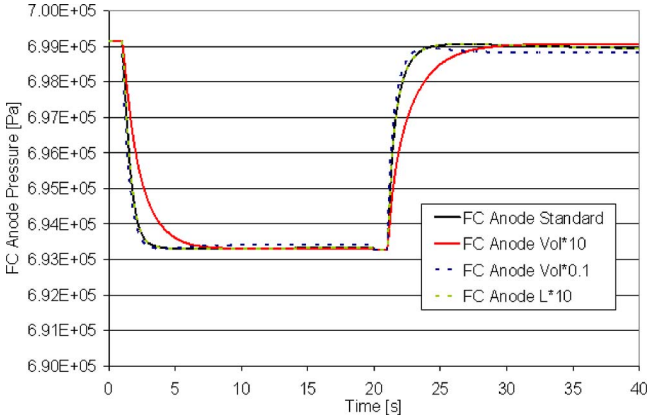


Fig. 14 Anodic side average operating pressure (Case II)

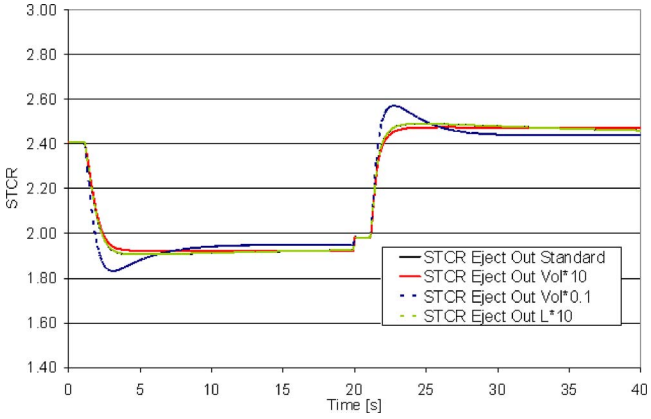


Fig. 15 STCR at anodic ejector outlet (Case II)

Dynamic Analysis: Sinusoidal Variation

Frequency characterization of a component is a necessary step for the development of robust control systems. In general terms, the response of a system to sinusoidal input can be represented in the so-called "Bode" diagram, where the output variable amplitude is recorded against the input frequency. Examples of such diagrams are to be found from Figs. 17–20, where the phase of the output parameter is also reported: phase equal to zero implies oscillations of output in phase with the input. In all these cases, the input parameters are fuel and current, which are varied in parallel between 100% and 70% of their nominal values, according to the sinusoidal law. Cases I and II are represented by dashed and solid lines, respectively. As in the step analysis, Case I is first investigated in detail, then it is compared with Case II.

Depending on the type of system under study, the response of the output variables under fully developed oscillating conditions can show either monotonically decreasing amplitudes or peaks in

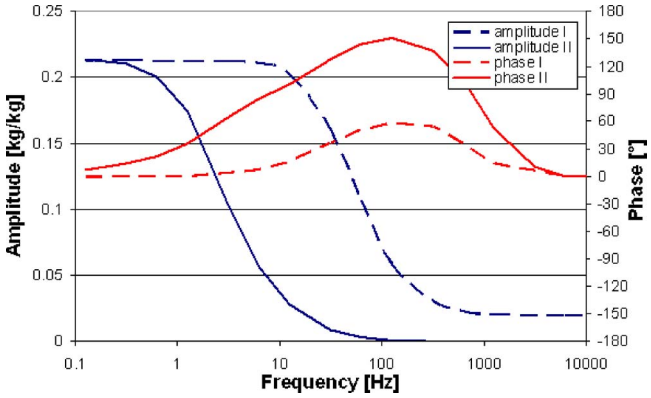


Fig. 17 Bode diagram of the mass flow at the anodic side inlet (amplitudes are nondimensionalized with the reference 100% value of 0.0695 kg/s)

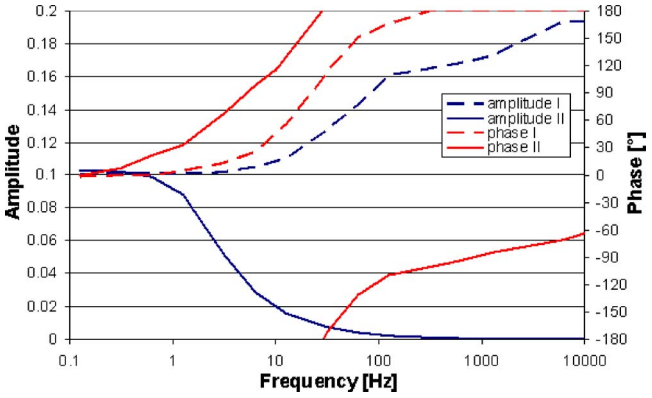


Fig. 20 Bode diagram of the ejector recirculation factor (amplitudes are nondimensionalized with the reference 100% value of 6.9)

proximity to its natural frequencies. Despite the feedback nature of the anodic loop, because of the recirculating flow from the stack outlet to reformer inlet through the anodic ejector, the anodic side never showed such amplification behavior. Instead, the anodic loop seemed to behave as a low-frequency filter, being able to absorb high-frequency disturbances: in fact, such high-frequency disturbances bring the system to a sort of new steady-state regime, which is equivalent to a constant input value equal to the one-period average value of the input (in our case, this is equivalent to 85% of the current and fuel flow).

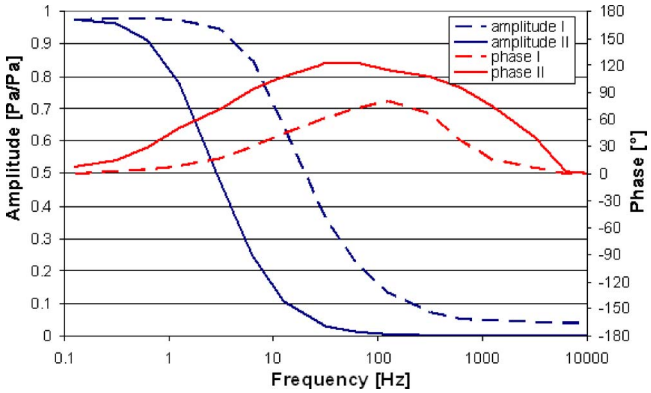


Fig. 18 Bode diagram of the fuel cell stack inlet pressure (amplitudes are nondimensionalized with the reference 100% value of the cathode-anode differential pressure of 3000 Pa)

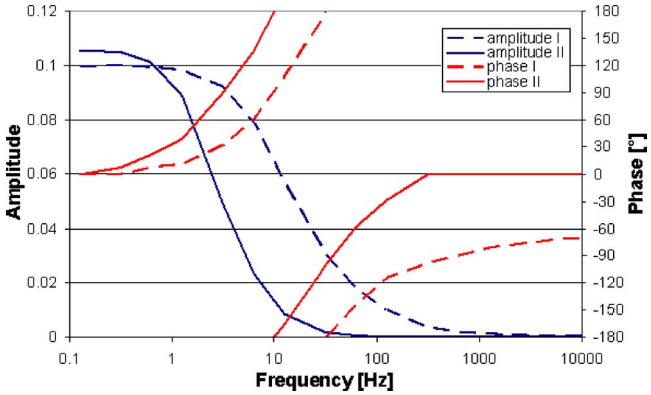


Fig. 19 Bode diagram of the STCR (amplitudes are nondimensionalized with the reference 100% value of 2.4)

With regard to output amplitudes, it is useful to try to correlate the characteristic frequencies, reported in Table 4 and calculated as the inverse of the characteristic times in Table 3, with the variations in amplitudes shown by the system output parameters. First of all, all the analyzed output parameters (i.e., mass flow (Fig. 17), pressure (Fig. 18), STCR (Fig. 19), and recirculation factor (Fig. 20) show an initial (low frequency) and a final (high frequency) constant amplitude. The change between the two occurs smoothly in an intermediate frequency range from about 10 Hz to 100 Hz, for Case I, and from about 1 Hz to 10 Hz for Case II. The initial amplitude is normally larger than the final one.

In Case I, anode mass flow oscillations reduce to very low values (2% of the design mass flow) in the 10–100 Hz range: the cutoff frequency is well represented by the anode side flow characteristic frequency of 31 Hz. In Case II, such a threshold is actually anticipated by the natural gas pipe characteristic frequency of 1.2 Hz: this is the reason why the mass flow oscillations tend to zero in the 1–10 Hz range. Phases of Case II tend to be higher than Case I because of the delay introduced by the natural gas pipe.

Anode pressure and STCR of Case I show a similar dependence on frequency: actually, the STCR follows the pressure because of the ejector behavior, which is mainly driven by volumetric flows. Recalling the inverse of the pressure characteristic time obtained in the previous step analysis (0.2 s), a characteristic frequency of 5 Hz is obtained for pressure: this frequency, followed by the 31 Hz flow characteristic frequency, can actually be used as an approximate threshold beyond which the pressure oscillations begin to reduce to almost nondetectable values. In Case II, natural gas pipe characteristic frequency dominates again; thus, pressure oscillations start decreasing from the 1.2 Hz threshold. Conclusions about the phases are the same as for anode mass flow.

Table 4 Main driving frequencies of the anodic loop (obtained from Table 3)

Component	Time constant type	Charact_frequency (Hz≡s^{-1})
Natural gas pipe (only Case II)	Filling	1.2
	Thermal	1.4
Anodic ejector	Flow	500
	Filling	83
	Thermal	100
	Solid_thermal	0.02
SOFC anodic side	Flow	31
	Filling	0.20
	Thermal	0.25

An exception is constituted by the recirculation factor of Case I: in fact, the higher the frequency, the larger the amplitude. This is due to the way the recirculation factor is calculated: it is the ratio between the secondary and primary flows. The primary flow is one of the two input parameters, the other being the current: the flow characteristic frequency of the anodic loop is dominated by the lowest frequency, which is the SOFC anodic side flow frequency, equal to 31 Hz. Actually, this frequency is the approximate threshold over which the secondary flow is only marginally influenced by primary flow variations, thus resulting in higher recirculation factor amplitudes. Therefore, the secondary flow (almost constant) over primary flow (fluctuating) ratio tends to increase its oscillations with increasing frequency. The same threshold of 31 Hz can be applied as the beginning for oscillation increases. On the other hand, the more realistic Case II shows that the natural gas pipe smoothes out flow input frequencies over 1.2 Hz, thus leading to the "conventional" low filter behavior also for recirculation factor.

Most phase values remained in the 0–180 deg range, thus showing the expected delay-dominated behavior in the anodic loop response to input variations in fuel and current. The only exception is constituted by STCR of Cases I and II and recirculation factor of Case II; however, their amplitudes become almost negligible after the +180 deg phase is reached, having little influence on system behavior.

Conclusions

A dynamic SOFC model was integrated with the simulation modules of other system components (i.e., reformer, anodic off-gas burner, anodic ejector) to construct within the TRANSEO tool a system model that could simulate the time response of the anode loop of an integrated 250 kW pressurized SOFC hybrid system.

Two general cases are studied: Case I, the fuel pipe volume is not included, i.e., no delay occurs between the fuel flow input of the controller and the fuel flow occurring at ejector primary nozzle; Case II, the fuel pipe volume is included, this being a more realistic case.

The step variation tests show that the main parameters affecting the anodic loop dynamic response are mainly two: the anodic loop length and the anodic loop volume. The first determines the time to regime of flows, thus playing an important role in the hydrogen depletion in the anodic channels as well as in the STCR behavior at reformer inlet; the second determines the time-dependent behavior of the hydrogen fraction in the exhaust fuel and pressure of the anodic side. However, when a significant natural gas pipe volume is present, such as in Case II, it tends to dominate the transient behavior and helps in smoothing out "peaky" dynamic behaviors for STCR and recirculation factor.

The frequency analysis shows that the anodic loop response to fuel and current sinusoidal variations is similar to a low-frequency filter: low-frequency amplitudes of output parameters are high and constant at low frequency and low and constant at high frequency. The threshold frequencies that determined the starting point of the reduction in amplitudes can be correlated with the inverse of the characteristic times obtained from the step variation analysis. In Case I, a threshold frequency of 5 Hz for the pressure and STCR, and a threshold frequency of 31 Hz for the anodic flow were obtained. In the more realistic Case II, natural gas pipe dominates again, and 1.2 Hz becomes the threshold frequency after which property oscillations start to decrease toward null values.

Acknowledgment

This work was partially financed by the European Integrated Project FELICITAS and the Italian National Projects FISR 2002 and PRIN 2005.

Nomenclature

A = area (m^2)
A, B = coefficients in Eq. (7) (Ω m, K)
c = velocity (m/s)
C_a = aerodynamic friction coefficient (−)
C_f = diffuser friction coefficient (−)
C_p = specific molar heat (J mol^{-1} K^{-1})
D = diameter (D)
D_1, D_2 = coefficients (A)
d = cell diameter (m)
E = activation energy (J mol^{-1})
E_{Nernst} = thermodynamic electric potential (V)
F = Faraday constant (C mol^{-1})
ΔG = Gibbs free energy change (J mol^{-1})
ΔG^o = standard Gibbs free energy change (J mol^{-1})
H = enthalpy (J mol^{-1})
h = heat transfer coefficient (W m^{-2} K^{-1})
I = electric current (A)
K = coefficients in Eqs. (5) and (6) (m^3 C^{-1} Ω^{-1})
$K_{p,ref}$, $K_{p,shift}$ = equilibrium constants (Pa^2, –)
k = coefficients in Eqs. (5) and (6) (J/mol)
k = thermal conductivity of the solid (W m^{-1} K^{-1})
l = cell length (m)
L = length (m)
N = molar flow rate (mol s^{-1})
P = pressure (Pa)
P_i = partial pressure (Pa)
r = reaction rate (mol m^{-3} s^{-1})
R_u = gas constant (J mol^{-1} K^{-1})
S = cross section (m^2)
s = thickness (m)
SOFC = solid oxide fuel cell
STCR = steam to carbon ratio
T = temperature (K)
V = electric potential of stack (V)
vol = anodic side volume (m^3)
V_p = loss of electric potential (V)
W_{el} = electric power (W)
x = longitudinal cell coordinate (−)

Greek Letters

α = diffuser angle (rad)
Ω = electric resistance (Ω)
ρ = electric resistivity, density (Ω^{-1}, kg/m^3)
ν = stoichiometric coefficient (−)

Subscripts

0 = reference
1 = primary duct outlet
2 = secondary duct outlet
3 = diffuser inlet
4 = diffuser outlet
act = activation
an = anode
av = average
cat = cathode
el = electrolyte
diff = diffuser
f = friction
k = chemical reaction
i = chemical component
mix = mixing duct
Ohm = Ohmic
s = solid structure
th = primary duct thickness
tube = quartz tube

References

[1] Costamagna, P., Magistri, L., and Massardo, A. F., 2001, "Design and Part Load Performance of a Hybrid System Based on a Solid Oxide Fuel Cell Reactor and a Micro Gas Turbine," J. Power Sources, **96**, pp. 352–368.

[2] Magistri, L., Traverso, A., Cerutti, F., Bozzolo, M., Costamagna, P., and Massardo, A. F., 2005, "Modelling of Pressurised Hybrid Systems Based on Integrated Planar Solid Oxide Fuel Cell (IP-SOFC) Technology," *Fuel Cells, Topical Issue Modelling of Fuel Cell Systems*, Vol. 1, Wiley-VCH, Weinheim.

[3] Ferrari, M. L., Traverso, A., and Massardo, A. F., 2005, "Control System for Solid Oxide Fuel Cell Hybrid Systems," ASME Paper No. 2005-GT-68102.

[4] Muller, F., Brouwer, J., Jabbari, F., and Samuelsen, S., "Dynamic Simulation of an Integrated Solid Oxide Fuel Cell System Including Current-Based Fuel Flow Control," ASME Paper No. FUELCELL2005-74107.

[5] Travis, R., Balestrino, C., Hill, R., and Bernardi, D., 2006, "Development of a 1MW SOFC System at Rolls-Royce Fuel Cell Systems," *Fuel Cell Conference*, Lucerne, Switzerland.

[6] Magistri, L., Bozzolo, M., Tarnowski, O., Agnew, G., and Massardo, A. F., 2003, "Design and Off-Design Analysis of a MW Hybrid System Based on Rolls-Royce Integrated Planar SOFC," ASME Paper No. GT-2003-38220.

[7] Traverso, A., Trasino, F., and Magistri, L., 2006, "Dynamic Simulation of the Anodic Side of an Integrated Solid Oxide Fuel Cell System," ASME Paper No. ESDA2006-95770.

[8] Ferrari, M. L., Traverso, A., and Massardo, A. F., 2004, "Transient Analysis of Solid Oxide Fuel Cell Hybrids. Part B: Anode Recirculation Model," ASME Paper No. GT2004-53716.

[9] Ferrari, M. L., Bernardi, D., and Massardo, A. F., 2006, "Design and Testing of Ejectors for High Temperature Fuel Cell Hybrid Systems," ASME J. Fuel Cell Sci. Technol., **3**, pp. 1–8.

[10] Ferrari, M. L., Pascenti, M., and Massardo, A. F., 2006, "Experimental Validation of an Unsteady Ejector Model for Hybrid Systems," ASME Paper No. GT2006-90447.

[11] Traverso, A., 2005, "TRANSEO Code for the Dynamic Performance Simulation of Micro Gas Turbine Cycles," ASME Paper No. GT2005-68101.

An Optimal Maintenance Policy for Compressor of a Gas Turbine Power Plant

P. Naga Srinivasa Rao

V. N. Achutha Naikan

Reliability Engineering Centre,
Indian Institute of Technology,
Kharagpur 721 302, West Bengal, India

This paper proposes a condition based maintenance policy for compressors of industrial gas turbine. Compressor blade fouling contributes a major performance loss in the operation of gas turbine. Water washing is usually done for recovery of the blade fouling problem. There exist two different washing methods, namely, online and offline washings. Many researchers suggested that performing a combined program of regular online washing plus periodic offline washing would give fruitful results with respect to economy. However, such studies are of empirical nature or have considered only deterministic treatment. Considering the rate of fouling as discrete state random process, we propose a condition based maintenance policy with periodic online washing and inspection directed offline washing. According to this policy, the compressor undergoes regular online washes for every $1/\lambda_m$ *operating hours, and also undergoes inspections at constant rate* λ_I*. If the observed condition at an inspection is worse than threshold deterioration state, then perform offline washing. Otherwise, continue with online washing. The proposed algorithm gives optimum schedules for both online washing and inspections considering minimization of total cost per operating hour as objective. It also gives optimum threshold deterioration level for performing offline washing. A comparison of the results for a hypothetical gas turbine compressor is presented as illustration.*
[DOI: 10.1115/1.2795762]

Keywords: gas turbine fouling, preventive maintenance, condition based maintenance, online washing, offline washing

Introduction

With the growing interest in life cycle costs for heavy-duty gas turbines, equipment operators are investigating the trade-off between performance improvements and associated maintenance costs. One of the key factors leading to performance losses during the gas turbine power plant operation is compressor fouling. The performance degradation attributed to compressor fouling is mainly due to deposits formed on the compressor blades by particles carried in by the air. Despite the use of advanced filtering methods and filter maintenance, some of the particles are not large enough (typically, a few microns diameter) to be blocked by these inlet filters. Therefore, the ingestion of substances that can cause fouling cannot be completely suppressed. Depending on the environment, these particles may range from dust and soot particles to water droplets or even insects, causing an increase in surface roughness, changes in the aerofoil shapes and/or aerofoil inlet angles, and a narrowing of the aerofoil-throat aperture. Fouling also takes place when the plant is operating in a sandy or dusty environment, which severely erodes compressor blades. The rate of fouling further increased when the operating environment includes the presence of alkali or similar minerals, which adhere to compressor blades when sulfur is present in the environment. The water quality is also a significant parameter, which directly influences the fouling rate [1].

In general, plant operating environment contains a combination of the above mentioned conditions. Compressor blade fouling reduces the flow capacity, and thereby the pressure ratio of the unit [2–5]. This results in an overall loss in power output and efficiency of a gas turbine. It is one of the commonest causes of performance reduction encountered by the gas turbines [4,6] and can account for 70–85% of the performance loss during operation [7,8]. A deterministic cost estimate is given by Diakunchak [2]. In extreme cases, fouling may also result in surge problems. The fouling rate depends largely on the following five factors:

- site location
- surrounding environment
- layout of the air intake system
- atmospheric parameters
- plant maintenance

The first four factors belong to plant design that management makes decisions on before constructing the plant. Once the plant constructed, the first four factors cannot be influenced during the operation; the plant maintenance is the critical one for preventing extra costs resulting from degraded plant performance.

Fortunately, compressor fouling is a "recoverable" deterioration in that periodic cleaning of the compressor can minimize it. Various methods have been used in the past to clean fouled compressors. At times when heavy-duty gas turbines did not yet possess highly sophisticated cooling schemes and coated compressor blades, cleaning was achieved by abrasion with the injection of solid compounds such as nutshells or rice husks. This had to be replaced by wet cleaning methods (water or solvent based) to protect modern coatings and to keep state-of-the-art cooling systems from blockages. The most effective wet cleaning process is the crank soak or offline wash. For this, the unit has to be shut down and cooled off in order to assure that the cleaning agent reaches all compressor stages and does not evaporate. The cleaning agent is injected into the compressor with the turbine turning at low speed. After a soaking time, the compressor is rinsed with water, which must be drained from the engine. Before the unit can be operated again commercially, it has to be dried. Thus, offline washing reduces the availability of a unit.

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 1, 2006; final manuscript received August 2, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal.

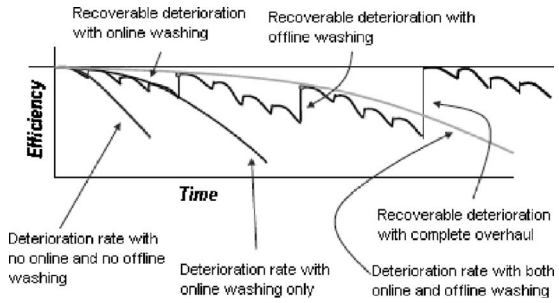


Fig. 1 Effects of washing and overhaul on compressor efficiency

With a growing number of gas turbines being used in combined cycle or combined heat and power applications, there was a need for the development of online washing systems that would provide acceptable performance improvement without shutdown and that would extend the operating period until the next offline washing. These systems are now state of the art in modern heavy-duty gas turbines. In an online wash, demineralized water or a mixture of demineralized water and detergent is injected into the compressor while the gas turbine is running. The fluid droplets impact and thoroughly wet the blades and inner surface of the compressor at high speeds to loosen and partially remove deposits. However, high performance recovery can only be achieved by an offline wash. Therefore, to get the benefits of both higher availability and increased degradation recovery, it is required to perform a combined maintenance policy with online and offline water washes. Proper combination of online and offline washing schedules can give optimum benefits by minimizing the life cycle cost of the plant. The basic objective of offline cleaning is to clean a dirty compressor and to restore the power and efficiency to the maximum extent possible. The primary objectives of online washing are the following:

(a) to maintain the cleanliness of a compressor after offline washing
(b) to maintain power and efficiency by minimizing ongoing losses, and
(c) to extend the period between offline washes by minimizing the buildup of deposits in the compressor, thereby reducing the ongoing incremental power losses.

Offline washing requires shutdowns of the plant and cooling of the components; therefore, it requires more time and money, and it also reduces the throughput of the plant. On the other hand, outages or shutdowns are not required for online cleaning. The effect of the various maintenance tasks (washing and overhaul) on gas turbine engine efficiency is shown in Fig. 1.

Statement of the Problem

With a growing presence of gas turbine technologies and competition among the industries, a stronger focus is being placed on trade-off analysis between performance optimization and operational and maintenance (O&M) costs. As a result, cost/benefit evaluation of performance recovery methods has been at the forefront of these efforts. However, optimization of both compressor offline (crank) and online washing intervals from the standpoints of operating cost and proactive maintenance is of primary interest to the organizations and the focus of this paper. Although the benefits of compressor washing are undoubted, the frequency of online and offline washes and the type of detergents to be used (if at all) are widely debated issues. It is essential to develop maintenance schedules based on the deterioration characteristics of the plant in order to balance the maintenance costs with lost revenue and extra operating costs [9].

The present research is aimed (i) to propose a method of mathematical modeling of the compressor online and offline washings with considering the random nature of the fouling and (ii) to propose an optimum condition based preventive maintenance policy for the compressor of a gas turbine power plant. Many researchers (Meher-Homji and Bromley [10] and Stalder [11]) proved that the fouling rate of gas turbine compressor is random and follows an exponential law. Therefore, finding of the optimum schedules for (i) mean time between successive on line washes, (ii) mean time between successive inspections by considering the fouling rate of compressors as exponential, and also (iii) threshold deterioration level for performing offline washing is considered as objectives of the proposed research. Here, the purpose of the inspections is to know the condition of the compressor in order to make a decision regarding whether or not to perform an offline wash.

Definitions

Failure Type

i. Deterioration: A process where the important parameters of a system gradually worsen. If left unattended, the process leads to deterioration failure (for example, wear, dimensional changes with time, effect of contamination, and deleterious environmental factors, e.g., temperature, voltage, and radiation are relevant examples and sources of component degradation).

   For most of the systems, it is possible to measure the amount of deterioration incurred up to time t either by directly measuring the above mentioned parameters or any other dependent parameters such as vibration, power, performance, etc.

   The deterioration failure may or may not be a physical failure of the system. It is a situation where the equipment performance is unacceptable (due to economy or statutory environmental regulations, or any other). In this paper, we are using the term "deterioration failure" to represent the situation where the compressor performance is unacceptable.
ii. Random failure: A failure whose occurrence is unpredictable except in a probabilistic or statistical sense (for example, failure of an automobile to start, occurrence of fatigue crack on the compressor blade surface).

Maintenance Type

i. Inspection: Inspections at specified schedules are performed to know the operating condition of the system. It involves direct measurement of some related parameters of the system, which directly represent the current operation condition of the system.
ii. Preventive maintenance (PM): A type of restoration or maintenance activity wherein an unfailed system, at preselected intervals, has its accumulated deterioration reduced or eliminated.
iii. Minimal PM: Maintenance of limited effort and effect. For deterioration modeled as discrete stage process, minimal PM activity restores the system to the previous deterioration stage or to deterioration Stage 1, whichever is worse.
iv. Major PM: Maintenance wherein a system is restored to "as good as new" condition.
v. General PM: It restores the system to the previous τ deterioration stages or to deterioration Stage 1, whichever is worse.
vi. Repair: A restoration wherein a failed system is returned to a working condition.
vii. Minimal repair: Repair wherein the system is returned to that operable state in which it was operating just before the failure.
viii. Perfect repair/overhaul/replacement: A restoration action wherein the failed system is returned to as good as new condition.

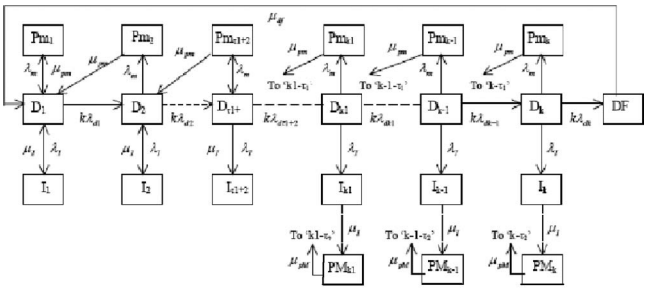


Fig. 2 State-transition diagram for periodic online washing and condition based offline washing policy

Assumptions

The following assumptions are made.

1. The compressor encounters two types of shutdowns. The first one is due to unacceptable level of deterioration, and the other is due to random failure of the compressor's parts (for example, blade fracture).
2. Compressor deterioration is modeled as discrete state process with k acceptable operating stages and one unacceptable (failed) stage. The compressor has a deterioration failure (unacceptable level of deterioration) immediately following the completion of k stages of deterioration.
3. The duration of each deterioration state has an exponential distribution with rate $k\lambda_d$. That gives a mean life of $1/\lambda_d$ between successive overhaul shutdowns.
4. Following deterioration failure (unacceptable level of deterioration), the compressor is overhauled, i.e., restored to as good as new. The overhaul duration is exponentially distributed.
5. Random failure occurs at an increasing rate or intensity, depending on the deterioration stage of the compressor. Minimal repairs are performed. The minimal repair duration is exponentially distributed.
6. The time between successive online washes is exponentially distributed; online washing makes the device τ_1 deterioration states younger. That is, if the deterioration stage before an online washing is n^{th}, then the deterioration stage after the online washing is $(n-\tau_1)^{\text{th}}$.
7. The compressor undergoes inspection at constant rate based on the observed deterioration stage (D_i) of the compressor; an offline wash can take place (if $i \geq k_1$), which makes the compressor τ_2 deterioration stages younger, or do nothing (if $i < k_1$).
8. Mean times to perform online and offline washings are exponentially distributed.
9. There are no transitions between the failure and maintenance states.

Model

Continuous time Markov chain (CTMC) model for the proposed condition based preventive maintenance policy of a gas turbine compressor is shown in Fig. 2. According to this policy, the compressor undergoes periodic online washings, and inspection directed offline washings. States D_1 to D_k are the k operating states in increasing order of deterioration. D_1 is the best operating state and D_k is the worst. D_i represents the state in which the compressor is in the i^{th} deterioration stage and is in working condition. I_i represents the state in which the device is in the i^{th} deterioration stage and under inspection. Pm_i is the state where the device is in the i^{th} deterioration stage and is under online washing. PM_i is the state where the device is in the i^{th} deterioration stage and is under offline washing. RF_i is the state where the device is in the i^{th} deterioration stage and is under repair after a random failure.

In this model, the compressor undergoes periodic online washings with constant rate λ_m with a mean duration of $1/\mu_{\text{pm}}$ that makes the compressor τ_1 deterioration stages younger. The compressor is inspected after a random period that is exponentially distributed with mean $1/\lambda_I$. If the device is found in deterioration stage i such that $i \geq k_1$, then offline washing of the compressor takes place that restores the compressor to $i-\tau_2$ deterioration stage with a mean duration of $1/\mu_{\text{pM}}$. When the device is in the random failure state, minimal repair takes place that brings the device to just before failure condition in operating mode with a mean duration $1/\mu_{\text{DF}}$. When the device is in deterioration failure (DF) state, complete overhaul is carried out to bring the device to as good as new state with a mean duration of $1/\mu_{\text{DF}}$. To increase the clarity of the figure, the random failure states of the compressor are not shown in Fig. 2.

Steady-State Solution

Using the Markov approach associated with Fig. 2, we first obtain the Chapman–Kolmogorov equations (Misra [12]). These are then converted into a set of linear equations (also called system balance equations) representing the steady-state conditions of the process in Fig. 2. Their solution yields the long-term state probabilities.

The state-transition equations for the proposed maintenance policy depend on the real values of the number of deterioration stages k, limit for starting offline washings (k_1), recovered deterioration states by performing online washing (τ_1), and that by offline washing (τ_2). The state-transition equations of the system depend on the following three important quantities:

- $(k-\tau_1)$
- $(k-\tau_2)$
- $(k_1-\tau_2)$

Values of these quantities depend on the values of the system parameters (k, k_1, τ_1, and τ_2). Based on the possible ranges of values of these quantities, we developed different sets of state-transition equations. Any one of them will execute depending on the above parameters. A recursive solution algorithm and software code (in "C") has been developed for solving these linear equations for steady-state probabilities of the different states of the system.

Optimization

Optimal inspection and maintenance schedule can be obtained by minimizing the total cost. The cost components include maintenance, repair, and out-of-service costs. A simple and more appropriate model for the cost analysis is proposed. Let C_I, C_{pm}, C_{pM}, C_{RF}, C_{DF}, C_{dt}, and C_{op} be, respectively, the unit cost components for various outages for inspections, online washings, offline washings, repairs after random and DF, down time, and operation. We first calculate respective cost for a deterioration replacement cycle (CT). CT is the expected time between successive deterioration replacements. From these costs, cost per operating hour or cost per unit operating time is calculated for decision-making purpose.

If CT is the expected time between successive deterioration replacements, then for a CT, there is one deterioration replacement. Therefore,

$$\text{CT}P_{\text{DF}} = 1/\mu_{\text{DF}} \Rightarrow \text{CT} = 1/(\mu_{\text{DF}}P_{\text{DF}}) \tag{1}$$

Cost Model

Expected costs per CT for DF, inspections, preventive maintenance, random failures, down time, and operating costs are calculated as follows:

i. Expected cost of DF during CT
Since for a DF cycle (CT) there exists only one expected replacement due to DF, therefore,

$$E(TC_{\mathrm{DF}}) = C_{\mathrm{DF}} \tag{2}$$

ii. Expected cost for performing inspections during CT,

$$E(TC_I) = \left(\sum_{i=1}^{k} P_{Ii} \right) \mu_I C_{pi} \mathrm{CT} \tag{3}$$

iii. Expected cost for performing online washings during CT,

$$E(TC_{\mathrm{pm}}) = \left(\sum_{i=1}^{k} P_{mi} \right) \mu_{\mathrm{pm}} C_{\mathrm{pm}} \mathrm{CT} \tag{4}$$

iv. Expected cost for performing offline washings during CT,

$$E(TC_{\mathrm{pM}}) = \left(\sum_{i=1}^{k} P_{Mi} \right) \mu_{\mathrm{pM}} C_{\mathrm{pM}} \mathrm{CT} \tag{5}$$

v. Expected cost for random failures during CT,

$$E(TC_{\mathrm{RF}}) = \sum_{i=1}^{k} (P_{fi}\mu) C_{\mathrm{RF}} \mathrm{CT} \tag{6}$$

vi. Expected cost of downtime during CT,

$$E(TC_{\mathrm{dt}}) = \left(\sum_{i=1}^{k} (P_{Ii} + P_{mi} + P_{Mi} + P_{fi}) + P_{\mathrm{DF}} \right) C_{dt} \mathrm{CT} \tag{7}$$

vii. Expected operating cost during CT,

$$E(TC_{\mathrm{op}}) = \left(\sum_{i=1}^{k} (P_i + P_{pmi}) \right) C_{\mathrm{op}} \mathrm{CT} \tag{8}$$

Using (i)–(vii), the expected cost per operating hour is expressed as

$$E(TC) = \frac{(E(TC_{\mathrm{DF}}) + E(TC_I) + E(TC_{\mathrm{pm}}) + E(TC_{\mathrm{pM}}) + E(TC_{\mathrm{RF}}) + E(TC_{\mathrm{op}}) + E(TC_{\mathrm{dt}}))}{\mathrm{CT}} \tag{9}$$

Optimization Procedure

The optimization procedures are as follows.

i. Depending on the accuracy required and/or deterioration replacement rate, choose an appropriate unit for
   (a) number of deterioration states (k)
   (b) time between inspections ($1/\lambda_I$)
   (c) time between online washings ($1/\lambda_m$)
ii. Set $k_1 = 1$
iii. Set $1/\lambda_m$ equal to 1
iv. Set $1/\lambda_I$ equal to 1
v. Calculate different deterioration state probabilities and availability. The complete procedure for evaluation of these quantities is available in Ref. [13].
vi. Using Eqs. (1)–(9), find the expected total costs for various cost components, and cost per operating unit time
vii. Increase $(1/\lambda_I)$ by 1, and repeat the calculations from step v. until $(1/\lambda_I)$ reaches a maximum $(1/\lambda_d)$
viii. Select the PM interval time $(1/\lambda_I)$ that minimizes the cost per operating hour, and note the corresponding k_1, $(1/\lambda_m)$, and $(1/\lambda_I)$ values
ix. Increase $(1/\lambda_m)$ by 1 and repeat steps from iv to x
x. Increase k_1 by 1 and repeat steps from iii to xi
xi. By minimizing the total costs, select a policy from all the selected policies in step x

Illustration

Let us consider a 30 MW gas turbine power plant. The output of the plant drops to 27 MW (10% deterioration in power output) after an average operation of 500 h if it is not washed. At this stage, the performance of the turbine is not economical and also the condition is not safe. Therefore, this state is treated as DF stage; a complete overhaul is required to make the system like new, which takes more downtime and money. Currently, washes are performed on a preventative schedule of 50 h for online washing and 500 h for offline washing. This maintenance policy is adopted by neither engineering assessment of conditional need nor optimal time to perform. In addition to the loss of availability and maintenance time incurred, unnecessary washes generate an environmental impact with the used detergent. Clearly, operating with a module that assesses condition and predicts the time to wash more appropriately would benefit the organization and environment.

In addition to DF, the system also undergoes random failures whose intensity increases with deterioration. After every random failure, only minimal repair is carried out. Therefore, the system condition after repair is the same as that of before failure condition. Each offline wash requires a mean downtime of 50 h and average cost of \$3000. Online wash takes 15 min, requires no downtime, and incurs average cost of \$2000. Take $\tau_1 = 1$ MW, $\tau_2 = 2$ MW. The management wants to decide the optimum schedules for online and offline washes, which reduces total cost by minimizing unnecessary washes.

To get more accurate results, we model the total deterioration (3 MW performance reduction) with large number of deterioration stages (60 deterioration stages) each represents 50 kW deterioration. Therefore, each online washing and offline washing makes the plant, respectively, 20 and 40 deterioration states younger.

The following parameters are deducted from the above data. Let $1/\lambda_d = 500$ h, $1/\mu_{\mathrm{pm}} = 0.25$ h, and $1/\mu_{\mathrm{DF}} = 50$ h. The following data are also considered.

Random failures of the device have a Weibull intensity with scale parameter $\eta = 500$, shape parameter $\beta = 3$, online washing cost $(C_{\mathrm{pm}}) = \$2000$, minimal repair cost at random failure $(C_{\mathrm{RF}}) = \$2000$, and offline washing cost $(C_{\mathrm{DF}}) = \$3000$. The mean time to minimal repair at random failure is 25 h. As the deterioration of the compressor (due to blade fouling) increases, the operating cost of the power plant also increases. It includes fuel and other consumables; its value depends on the deterioration stage of the plant. Now, a day's energy prices (fuel and electricity) change rapidly. This will be taken care by systematically collecting the data on

Table 1 Solution for the illustration

Policy	Time between online washes ($1/\lambda_m$)	Time between inspections ($1/\lambda_I$)	k_1	Time between offline washes	Availability	Total cost per operating hour	Savings per year ($)
Present (online wash for 50 h, and offline wash for 500 h of operation	50	—	—	500	0.906883	1079	0
Time based optimal PM policy	60	—	—	5100	0.984359	1065	120,722
Condition based PM policy	69	15	30	2213	0.972896	1062	144,884

changes in price of these commodities, and finding a relation with time. We can make a forecast for the next year (either on weekly or monthly basis). In the example, we have considered a linear relationship between operating cost and level of deterioration of the compressor. If we consider the varying price, then this function will have a nonlinear nature. This requires collection of data on fluctuation of the fuel price on a monthly or yearly basis, and preparing and updating the washing schedules. In this illustration, we assume that the operating cost at base load operating condition for state (D_1) is \$1000/h. It uniformly increases with deterioration, and at worst operating condition (D_{60}), it is \$1200/h.

Solution for the Illustration. For this problem, if we consider availability maximization as objective, we require to perform continuous online washes of the compressor. However, as mentioned earlier, unnecessary washes require large resources. It will also generate an environmental impact due to the large quantity of detergent used. So, total cost minimization is the better objective for this problem. Following the different steps of the optimization procedure as discussed earlier, we get the optimal solution for total cost minimization. This is shown in Table 1.

By the analysis of the results in Table 1, the following observations are made.

(i) It is estimated that by applying the proposed model, we can reduce the annual operating cost of the plant up to a sum of \$148,920.
(ii) Significant improvement in availability is also achieved with the application of the proposed model (from 90.7% to 97.3%). This increased availability improves the throughput of the plant, and thereby profitability.
(iii) Reduced environmental pollution is also achieved with the proposed model.

Conclusions

Considering the rate of fouling as discrete state continuous time random process, we propose a condition based maintenance policy for compressors of industrial gas turbine with periodic online washing and inspection directed offline washing. Proposed policy improves the availability of the plant by optimizing online and offline washing frequencies. Considerable amount of savings in operating cost, considerable gain in availability, and thereby profitability are also possible with the application of the proposed model.

References

[1] Boyce, M. P., and Gonzales, F., 2005, "A Study of On-Line and Off-Line Turbine Washing to Optimize the Operation of a Gas Turbine," *ASME Turbo Expo 2005*, Reno, NV, June, ASME Paper No. GT2005-69126.
[2] Diakunchak, I. S., 1992, "Performance Deteriorations in Industrial Gas-Turbines," ASME J. Eng. Gas Turbines Power, **114**, pp. 161–168.
[3] Zaita, A. V., Bulet, G., and Karlsons, G., 1998, "Performance Deterioration Modelling in Aircraft Gas-Turbine Engines," ASME J. Eng. Gas Turbines Power, **120**, pp. 344–349.
[4] Lakshminarasimha, A. N., Boyce, M. P., and Meher-Homji, C. B., 1994, "Modelling and Analysis of Gas-Turbine Performance Deterioration," ASME J. Eng. Gas Turbines Power, **116**, pp. 46–52.
[5] Aker, G. F., and Saravanamuttoo, H. I. H., 1989, "Predicting Gas-Turbine Performance Behaviour Due to Compressor Fouling Using Computer-Simulation Techniques," ASME J. Eng. Gas Turbines Power, **111**, pp. 343–350.
[6] Upton, A. W. J., 1974, "Axial-Flow Compressor and Turbine-Blade Fouling: Some Causes, Effects and Cleaning Methods," *First Symposium on Gas-Turbine Operation and Maintenance*, National Research Council of Canada.
[7] Christoph, P. L., Christoph, S., and Lutz, D., 2004, "Performance Benefits Using Siemens Advanced Compressor Cleaning System," ASME J. Eng. Gas Turbines Power, **126**, pp. 763–769.
[8] Ogaji, S. O. T., Sampath, S., Singh, R., and Probert, S. D., 2002, "Parameter Selection for Diagnosing a Gas-Turbine's Performance-Deterioration," Appl. Energy, **73**, pp. 25–46.
[9] Tarabrin, A. P., Schurovsky, V. A., Bodrov, A. I., and Stalder, J. P., 1998, "An Analysis of Axial Compressor Fouling and a Blade Cleaning Method," ASME J. Turbomach., **120**, pp. 256–261.
[10] Meher-Homji, C. B., and Bromley, A. F., 2004, "Gas Turbine Axial Compressor Fouling and Washing," *33rd Turbomachinery Symposium, Turbomachinery Laboratory*, Texas A&M University, Houston, TX, Sept., pp. 163–192.
[11] Stalder, J. P., 2001, "Gas Turbine Compressor Washing State of the Art: Field Experiences," ASME J. Eng. Gas Turbines Power, **123**, pp. 363–370.
[12] Misra, K. B., *Reliability Analysis and Prediction: A Methodology Oriented Treatment*, Elsevier Science, Amsterdam, 1992.
[13] Rao, P. N. S., and Naikan, V. N. A., 2006, "Generalized Condition-Based Preventive Maintenance Policy for Markov Deteriorating Systems," Int. J. Performability Engineering, **2**(2), pp. 175–189.

Analytical Study of the Effects of the Clogging of a Mechanical Precipitator Unit in Air Preheaters in a High-Performance Thermoelectric Power Plant Based on Available Data

J. M. Blanco[1]

F. Peña
e-mail: jesusmaria.blanco@ehu.es

Departamento de Máquinas y Motores Térmicos,
Escuela Técnica Superior de Ingeniería,
Universidad del País Vasco/E.H.U,
Alameda de Urquijo s/n (48013) Bilbao, España

Following a detailed study of two of the mechanical precipitators in the air preheaters of a thermoelectric power plant, a large amount of ash that was deposited on one of the inlet conduits was observed, obstructing the incoming gas flow. A comparison of the available data for the two most recent hopper cleaning operations revealed that, on the one hand, the amount of ash collected by the clogged precipitator (A) was significantly less than that collected by the other (B) and, on the other hand, the temperature of the ash in the former was noticeably lower than in the latter. Prior to the cleaning of the conduits, a certain amount of damage was caused to the boiler dome, which meant that subsequent cleaning required the use of a hydrolazer, where it was noted that inlet pressures were very high. All of this indicated that the cause of the clogging was not physical. This paper provides a comprehensive analytical analysis that explains what happened, as well as resolving the situation. [DOI: 10.1115/1.2795771]

Keywords: thermoelectric power plant, precipitator, exhaust gases, ashes, hoppers, conduit clogging

1 Introduction

This paper provides a theoretical explanation of the causes that led to the clogging by ash of one of the two groups of mechanical precipitators (MPs) in a high-performance thermoelectric power plant equipped with two Ljungström air preheaters (LAPs), arranged in the manner shown in Fig. 1.

The rotary regenerative heat exchanger was patented by the Swedish engineer Fredrik Ljungström in 1920. The basic component is a continuously rotating cylinder, called "the rotor," which is packed with thousands of square thin-walled steel plates, packed in frame baskets of heat transfer surface. During each revolution of the rotor, heat is absorbed by the heating surface passing through the hot gas stream and transferred to the combustion air stream. The basic operating scheme is shown in Fig. 2.

This accumulated heat is released into the incoming air as the same surfaces pass through the other half of the structure. The heat transfer cycle is continuous in periods of about 60 s as the surfaces are alternately exposed to the outgoing gas and incoming air streams [1,2]. Heat energy is captured and transferred to the incoming air for combustion before it is lost to the stack. The result is a substantial saving in fuel that would otherwise be required to bring the air up to combustion temperature [3].

Until now, such regenerators were designed to rotate steadily and continuously being used mainly for gases near atmospheric pressure. They are used in boilers as air preheaters, in flue gas desulfurization plants as gas/gas heaters, and in tail-end selective catalytic reduction NO_x plants as gas preheaters. They are both designed and built to operate over extended periods with durable and uninterrupted service. Simplicity of design also makes them easy and economical to maintain while in operation and at scheduled [4,5]. The majority of air preheaters in use in the power industry today are of the Ljungström type. In 1995, the American Society of Mechanical Engineers (ASME) declared them to be a "Mechanical Engineering Landmark."

When the exhaust gases leave, each one of the air heaters in this installation does not go directly to the stack, but first passes through the respective MPs where the ash and cinders in the gases are retained by means of centrifugal forces created by settling cyclones, and collected later by a deposition tank through a deposition valve, as can be seen in Fig. 3(*a*).

Following a thorough cleaning of the MPs and a subsequent period of operation, in which a study was made of the pressure differences in the gas conduits operating under nominal load, a major obstruction was detected in just one of the precipitators. A subsequent inspection showed that a considerable amount of ash had been deposited on its inlet conduits, as can be seen in Fig. 3(*b*), which was reducing the amount of incoming gas by 60%. This is clearly an atypical circumstance, without precedent at all in the incident records of thermoelectric power plants.

Furthermore, a comparison of the available data for the two most recent hopper cleaning operations showed that, on the one hand, the amount of ash collected by the precipitator that had become clogged was three times less than that collected by the other and, on the other hand, the temperature of the ash held in the hoppers for the blocked one was significantly lower than the temperature of the ashes collected by the other one. This latter circumstance is directly related to the air and gas, inlet and outlet temperatures, as we shall see throughout this paper.

[1]Corresponding author.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 5, 2006; final manuscript received July 20, 2007; published online January 22, 2008. Review conducted by Antonio Peretto.

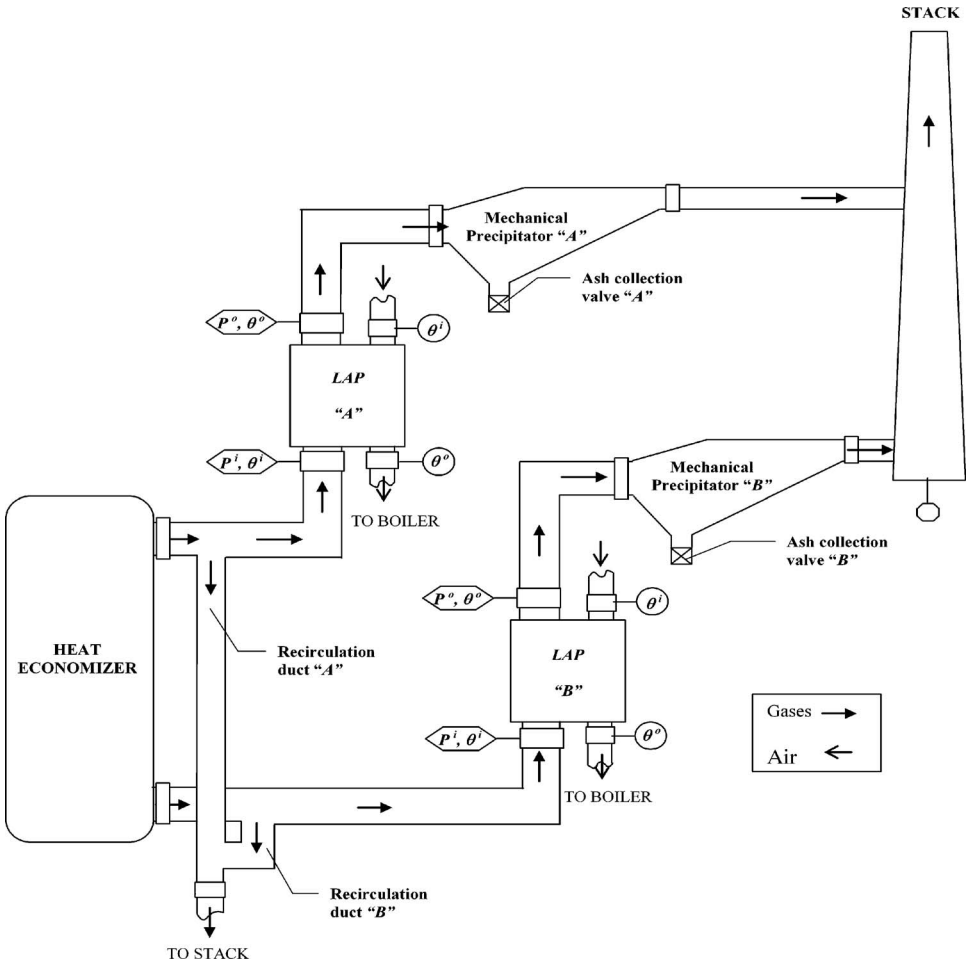


Fig. 1 Simplified layout of the air heaters and MPs in the power plant of reference

2 Aims and Methodology

Once the operating problem described in the preceding section has been roughly identified, the aim of this paper is to carry out a thorough study of the installation's operation in order to explain it in an effective way. Thus, there are sufficient background data to assess that, with the power plant operating at nominal load (480 MW) and the LAP's baskets (heat transfer surfaces) in perfect conditions, the relative pressure in the boiler furnace never should exceed a value of 600 mm of water column, which in turn means heater inlet pressures of below 300 mm of water column, and never exceed 350 mm of water column, with extremely dirty heaters due to the accumulation of waste from the exhaust gases.

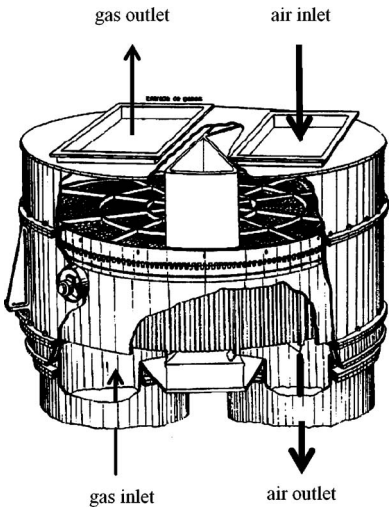
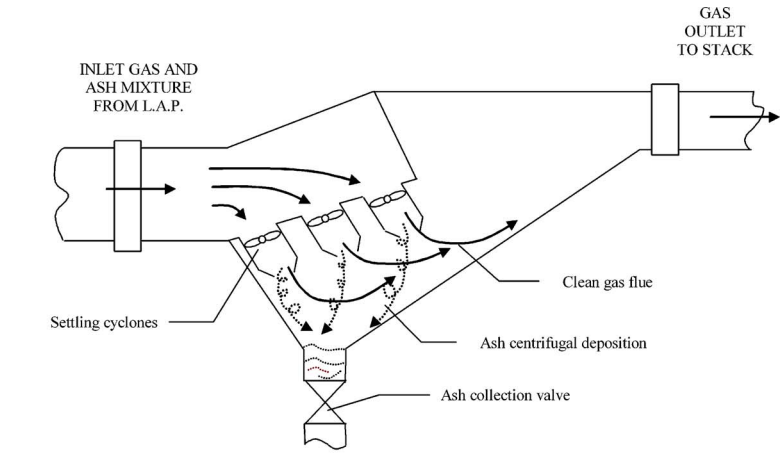


Fig. 2 Description an air heater type "Ljungstrom" (ALSTOM Power Inc. Air Preheater Company)

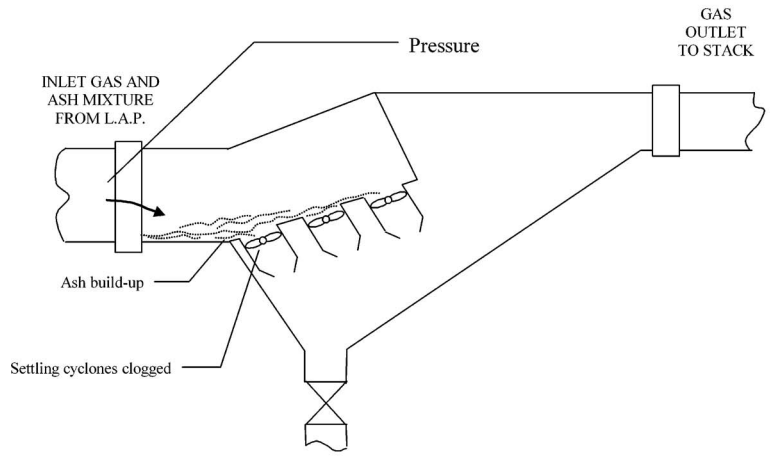
It is important to note that immediately after the conduits were cleaned, normal values for pressure were reached. Anyway as a consequence of these cleaning operations, certain damage was caused to the boiler dome (top structure). This signaled a turning point in the trends for the pressure differences in the gas conduits; as from that moment on, the pressure in the gas inlet on the (LAP "A") reached a value of about 425 mm of water column, together with a furnace pressure of 730 mm of water column, so the power plant's effective output was reduced to 83% of its rated figure.

A more thorough cleaning was subsequently undertaken by means of a hydrolazer system, involving a jet of water and perforation under rotation using bits that turn at between 1500 rpm and 2000 rpm, whereby all deposits were removed without damaging the conduits at all. It was noted that the inlet pressures on both heaters were reduced significantly but clearly exceeded 300 mm of water column so it means that original values were not reestablished in this way.

Even after a short operating period of time, a new and dangerous increase of the inlet pressure was detected in (LAP A), as its MP becomes definitely clogged. A final decision of changing baskets was finally assumed; so as soon as the baskets were finally replaced, the normal values for pressure were again reached. All



a) Nornal conditions L.A.P. (A)



b) Clogging L.A.P. (A)

Fig. 3 A two step description of the clogging process for LAP A

this time dependent process can be clearly seen in Fig. 4 where a historical description of inlet and outlet pressure changes through both Units A and B was depicted.

This led us to suspect that the reason for the clogging was not physical, but due to some other causes. The first consideration related to the loss of power was to take into account the air flow inlet toward exit gases across the seal system of our LAP, which after measured only supposed less than 6% of the total air flow, which was carefully certified for both units according to the historical available data from the power plant, so the conclusion was that it did not reveal much importance. A theoretical study was taken into account in order to explain the phenomena that gave rise to such anomalies above explained. A heat transfer survey, both in the heaters and conduits, was carried out in order to ascertain the reasons for the loss of power already tested in one of them, as it was suspected that its origin might lie with the increase of the heat transfer process inside the heater because of the turbulent flow; a further result of it will address the accumulation of ash in the conduits of the MPs.

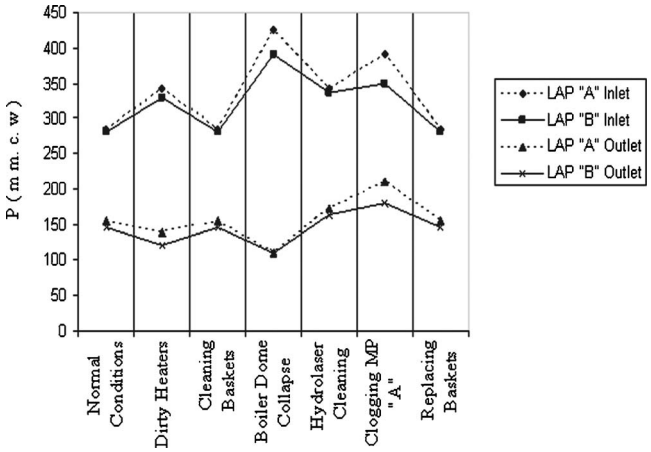


Fig. 4 Historical description of inlet and outlet pressure changes through LAPs A and B.

3 Theoretical Time Dependent Study of the Heat Transfer Process Inside Ljungström Air Preheater A

The heat transfer between the air and gas areas in the LAP responds to the case of "walls undergoing periodic changes in surface temperatures" [6]. The time dependent temperature variations on a surface inside the wall (metal plates) is calculated on the basis of the "equation on one-dimensional conduction in transient state" [7].

$$a\frac{\partial^2\theta}{\partial x^2} = \frac{\partial\theta}{\partial t} \tag{1}$$

It is shown [8] that the solution to the expression is provided by

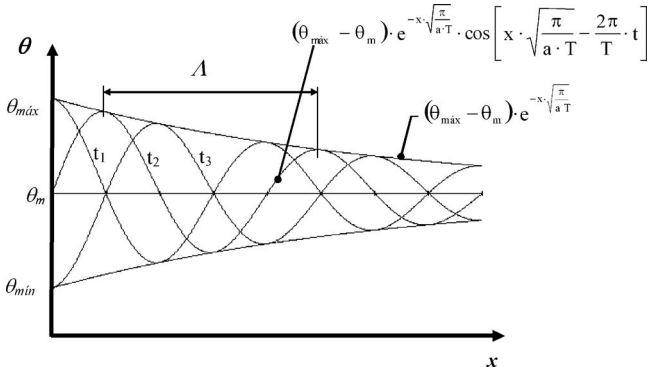


Fig. 5 Decreasing of the time dependent theoretical temperature field over the metallic baskets

$$(\theta - \theta_m) = (\theta_{\max} - \theta_m)e^{-x\sqrt{\pi/aT}} \cos\left[x\sqrt{\frac{\pi}{aT}} - \frac{2\pi}{T}t\right] \quad (2)$$

This equation corresponds to a sinusoidal-type function with exponential decay, whereby we can define the following parameters:

$$A = (\theta_{\max} - \theta_m)e^{-x\sqrt{\pi/aT}} \quad (3)$$

$$F_a = e^{-x\sqrt{\pi/aT}} \quad (4)$$

$$F = \left[x\sqrt{\frac{\pi}{aT}} - \frac{2\pi}{T}t\right] \quad (5)$$

$$\Lambda = 2\sqrt{\pi aT} \quad (6)$$

The function decays, around an average temperature, as x increases, whereby its time dependent graphic representation is shown in Fig. 5.

In practice, the temperature evolution of the baskets does not respond to a sinusoidal-type function, but instead follows a parabolic-type function in heating and a logarithmic-type function in cooling [9,10], according to available data, as shown in Fig. 6.

The mathematical relationships of analytical calculus, which govern the heat transfer between the air and gas areas, are extremely complex in the real case of a LAP; nevertheless, there are approximate calculation methods. The most appropriate is the following expression [11]:

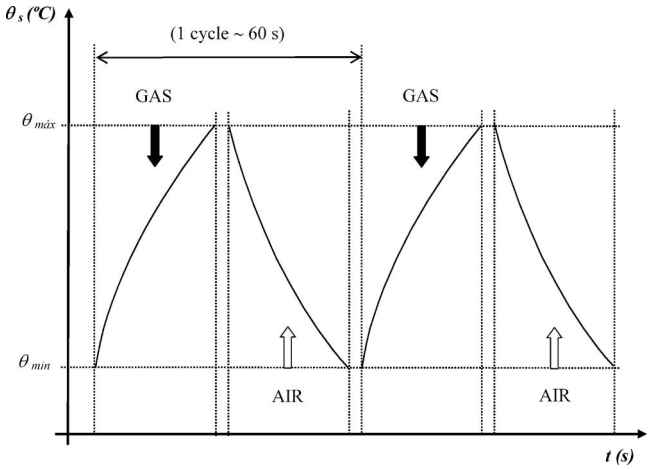


Fig. 6 Time dependent empirical surface temperature field over the metallic baskets

Table 1 Comparative inlet and outlet temperatures of air and gases, respectively, for both air heaters

	Normal conditions		Clogging situation	
θ (°C)	LAP A	LAP B	LAP A	LAP B
Air inlet	34.80	35.90	37	37
Air outlet	**271**	235	**286**	233
Gas inlet	340	340	337	348
Gas outlet	**150**	150	**91**	168

$$\frac{q}{A} = \frac{\Delta\theta m/(\varphi_g - \varphi_a)}{1/h_g \cdot \varphi_g + 1/h_a \cdot \varphi_a + 1/2{,}5 \cdot Cp_s \cdot \rho_s \cdot r_B + r_B/k \cdot (\varphi_g + \varphi_a)} \quad (7)$$

The third term in the denominator considers the accumulation of heat between the plates (thermal inertia) and the fourth term is a correction factor regarding conductivity. When these two latter terms are insignificant, with regard to the first two, this equation is reduced to that of a heat exchanger operating in steady state. Accordingly, the higher the coefficient for film h, the greater the heat transfers by convection [12,13].

The determination of the film coefficients, in the gas and air areas, respectively, in the baskets, responds to the typical case of forced convection along flat surfaces [14] (The Reynolds number, in both the air and gas areas, clearly exceeds the value 5×10^5, which for gases constitute the critical value for the transition from laminar to turbulent flow [15]). Nonetheless, it should be taken into account that there must be a certain area of laminar flow at the beginning of the metallic plates, before the critical Reynolds number is reached, with the onset of the turbulent flow.

With all these in mind, the more appropriate calculation expression [4] could be as follows, provided that the "Prandtl" number does not differ greatly from unity:

$$N_{\mathrm{Nu}} = \frac{hL}{k} = (N_{\mathrm{Pr}})^{1/3} \cdot [(N_{\mathrm{Re}})^{0.8} - 23.100] \quad (8)$$

In order to facilitate the presentation to the reader, it will be accepted that the heat transferred from the heater to the combustion air has the same value as the heat transferred from the gases to the heater, which is tantamount to considering a lossless heater, according to the previous consideration in this way. In such a case, the mass and energy balance in the heater corresponds to the following expression [16,17]:

$$\dot{Q} = \dot{m}_a c_{p(a)}(\theta_a^o - \theta_a^i) = \dot{m}_g c_{p(g)}(\theta_g^o - \theta_g^i) = US\Delta\theta_m \quad (9)$$

Considering a nominal combustion air flow of 1.457 ton/h and an output of reference for the thermoelectric power plant of 480 MW, it was noted that in the last cleaning operations that the temperature of the ashes held in the hoppers of MP A was significantly lower than the temperature of the ashes contained in the hoppers of precipitator B. In these working conditions (normal behavior and clogging situation, respectively), available data measured at inlet and outlet for both heaters in an ambient temperature of 26°C are shown in Table 1.

This drop in temperature at the gas outlet and the respective increase in the air outlet temperature on LAP A for the so called "clogging situation" can only be due to a significant increase in the value of the film coefficients for the gases and air, respectively.

Accordingly, the Prandtl numbers of the fluids, both from the air and gases in both heaters, are also similar, given that they are located between common sources and heat sinks. Consequently, the only reasonable explanation for the increase in film coefficients is that the Reynolds numbers for LAP A, in both gas and air areas, have reached significantly higher values than those corresponding to LAP B, and this can only be due to a simultaneous increase in air and gas flows in the former.

In addition, it seems reasonable to consider that the significant drop in the flow velocity of both air and gases is due to a considerable drop in load [18] between the metal plates on the baskets on LAP A. This may be due to two circumstances whose effects are compounded:

1. to the clogging of the open spaces between plates due to the unexpected destruction of the boiler top (dome)
2. considering the long time elapsed since the last inspection of the heater baskets, to the partial destruction by acid corrosion [19,20] of the metal plates on these baskets (formation of acid dew point, ADP), which could cause the displacement of the plates when cleaning with a hydrolazer, with the subsequent overlapping of some plates on top of each other, deforming the passages between them, which clearly magnified the formation of turbulent flow in those passages [21], with the ensuing increase in loss of power.

Subsequently, and on the basis of this study, the baskets on LAP A were carefully examined, which has indeed been shown that the expectations outlined here have been readily fulfilled, regarding clogging between metal plates, proceeding this way to repair them immediately.

4 Analysis of the Ash Accumulation Inside Ljungström Air Preheater A

When the exhaust gases leave the LAP, they are fed into the MPs, where the ashes are separated out from the gases by the effect of the centrifugal forces they exert, and from here the "cleaned" gases are finally fed to the stack, as it was previously shown in Fig. 3(*a*). We shall now proceed to analyze the accumulation of ash in the area between the gas outlet on LAP A and its corresponding MP, with the subsequent entrainment current and clogging of the precipitator hoppers, as it was previously shown in Fig. 3(*b*). This is why it is deduced that there is a significant loss of power, together with a significant increase in furnace pressure and, therefore, also in the incoming gases on both heaters as it was previously shown in Fig. 4.

Whereas it can be considered in LAP B that the through section is free in practical terms and, therefore, the flow across the metal heat transfer plates may be considered slightly turbulent and inconsistent [22]. Nevertheless, in the case of LAP A, the flow can be considered as turbulent, inconsistent, and clearly chaotic, with the presence of "restrictions" and, consequently, the existence of a certain "lamination."

The laminate phenomenon is an irreversible process that is, in theory, carried out at constant enthalpy. Besides not producing any useful kind of energy, it is characterized by a significant increase in entropy and, consequently, by a sharp drop in "exergy" and, therefore, in the operating capacity of the circulating fluid. Yet, in addition, after the restriction there is turbulent flow in the sections close to said restriction [23,24], with the formation of a large number of eddies or vortexes, and finally returning to turbulent flow, although slightly more consistent.

In practice, lamination with real gases cannot in any way be considered a process at constant enthalpy [25], which means taking the kinetic energy of the circulating fluid to be constant, before and after the restriction that gives rise to the lamination, and it is widely known that this is not the case. It is precisely the formation of the aforementioned turbulent flow after the restriction that renders the process utterly irreversible, and in practice the flow velocity in the section located immediately upstream is very high, and the corresponding kinetic energy will have been reached at the expense of a drop in enthalpy [26]. However, as the fluid flows through the widening of the outlet section, the velocity again drops while the original enthalpy is not reestablished.

The confused, turbulent flow at the gas outlet on the LAP A means that there are a large number of particles in suspension, colliding against the "boundary layer" of the conduits where the flow velocity decreases significantly and, consequently, so does the velocity of the aforementioned particles in suspension. In 1904, Prandtl proposed that the study of the movement of a low viscosity fluid could be likened to that of a perfect fluid [27], except in a layer close to the boundary, of tiny thickness, which was the focus of friction phenomena, which has been called the "boundary layer." Thus, on the outside of this layer, the tangential stresses are insignificant, with the forces of inertia prevailing over those of viscosity, whereas inside the boundary layer, the proximity of the boundary means that the velocity gradient is very high and, therefore, the tangential stress is also significant; the friction forces in this situation can be considered of the same order of magnitude as the forces of inertia.

Thus, the force upholding a submerged solid in a fluid flow is given by Buckingham's "pi" theorem [28]:

$$F_s = S\rho V^2 C_s \tag{10}$$

Defining the lift coefficient as

$$C_s = C\left(\frac{\mu}{L\rho V}\right)^b \left(\frac{E}{\rho V^2}\right)^c \tag{11}$$

When the Mach number is equal to or greater than unity, the lift coefficient may be considered solely a function of the aforementioned Mach number. Simplifying and including the velocity height, the lift force of the suspended particles in the gases produced by combustion that flow through the gas conduits is

$$F_s = C_s\left(\frac{\omega}{g}\right)A_p\left(\frac{V^2}{2}\right) \tag{12}$$

For a particle to remain suspended in a fluid, the condition of equilibrium in a vertical direction must be fulfilled, namely,

$$\sum F_y = 0$$

$$F_s - W = 0 \tag{13}$$

Obviously, when any particle strikes the flow's boundary layer, its velocity is reduced significantly and therefore so is its lift force, and in the end it falls through the effect of gravity. In the case of laminar flow, analytical means can be used to obtain the velocity distribution in a straight section [29]. The velocity profile follows a law of parabolic variation, whereby the velocity can be considered zero on the conduit wall and maximum at the conduit axis, where its value could reach twice the average velocity in this section.

In the case of turbulent flows, it is not possible to obtain the velocity distribution by analytical means. Nonetheless, on the basis of experimental data provided by Nikuradse and other researchers, it is known that the velocity profile is more uniform: Yet as with laminar flows, the minimum velocity is recorded close to the boundary layer and the maximum at the pipe axis. An experimental expression that provides the velocity profile for turbulent flows, on the basis of average velocity, is as follows [30]:

$$\frac{v - V}{V - \sqrt{f}} = 2\log\left(\frac{y}{r_0}\right) + 1'32 \tag{14}$$

This equation makes it clear that it would not be necessary for the particles to directly strike the boundary layer [31,32]; instead, it would be enough for them to get sufficiently close to it, in order to be finally decanted.

It is reasonable to suppose that the turbulent outlet flow of the combustion gases from LAP A will lead to the renewal of suspended particles, either colliding directly with the boundary layer or getting sufficiently close to it, which will also result in a significant increase in the number of particles finally decanted.

At the end, the mass of particles decanted would be entrained by the gas flow toward the entrance to the mouth tools of the MPs on LAP A. This entrainment effect explains why the MPs of this heater become clogged, even after short operating times.

5 Results and Discussion

The presence of high furnace pressure, with a load of 85% and even lower, denotes the existence of a serious problem in the gas conduits. Furthermore, the fact that the MP A registers major clogging after only short operating times confirms the previous hypothesis and, in addition, focuses the problem of this gas conduit.

Furthermore, the temperature differences in the air and gas conduits, before and after the heaters, constitute the main grounds for analysis in this study, which is based on the air-gas heat transfer in these heaters, according to the data previously shown in Table 1.

The specifications of the air and gases are similar in both air heaters, as they both operate between common sources and heat sinks. However, the air leaving LAP A is hotter and the gases are significantly cooler. For this to happen, a significant increase in film h coefficients has to be considered, and the only reasonable explanation for this is that the Reynolds numbers in this heater, in both the air and gas areas, have reached values that are markedly higher than those corresponding to the other one and, in turn, this can only be due to a simultaneous increase in the velocity of air and gas flows in the former.

Accordingly, the fulfillment of the Eq. (9) necessarily requires a decrease in both mass flows, in air and gases alike.

The initial cause of the high loss of power in LAP A is undoubtedly the partial destruction of the boiler top (dome), with its component materials becoming detached and falling onto the heater, as this signals a turning point in the trends for the differential pressures in both heaters. Likewise, the destruction, also partial, by acid corrosion of the metal plates on the heater baskets (formation of ADP), as well as the displacement of the same due to the subsequent cleaning with a hydrolazer, with them overlapping each other, constitutes the ultimate cause of this pressure drop through the formation of a highly chaotic turbulent flow.

Finally, the increase in pressure drop recorded by the gases through LAP A means that a greater number of suspended particles in the gases collide and get sufficiently close to the boundary layer of the gas conduit to lose lift force and become finally decanted. The gas flow then entrains the particle mass to the inlet mouth tools of the MPs, clogging them up in short operating times, as we previously showed.

6 Concluding Remarks

The following conclusions may be drawn from this paper:

1. The relative differences of temperature in outgoing air and combustion gases from LAP A are explained, accepting a significant increase in the flow velocity of both air and gases due to the deformation of the gas flow passages between the metal plates on the heater, with formation of inconsistent turbulent flow inside, and also allowing for laminations due to significant clogging of the spaces between the metal plates on the baskets in the aforementioned heater.
2. On the basis of this study, LAP A was finally dismantled and repaired, fully confirming the expectations held here.
3. Once repaired and under subsequent operation of the thermoelectric power station, it was verified that its load of power was once again within design limits. Naturally enough, the problem of the clogging of MP A with ash was resolved and the load limitation for high furnace pressure was corrected automatically, as can be seen at the end of Fig. 4.
4. This served as a prompt to examine and clean LAP B. At the time of writing, the furnace pressures at full load remain below 600 mm of water column and there is no limitation whatsoever on the load of the thermoelectric power plant, as shown by the readings taken for it.

Acknowledgment

Authors are deeply thankful to all staff of the power plant of reference for their help and facilities on the test experiments carried out.

Nomenclature

A = wave amplitude (μm)
a = thermal diffusivity of the material on the metal plates (m^2/s)
ADP = acid dew point temperature (°C)
A_p = characteristic particle surface (mm^2)
$\Delta\theta m$ = average total logarithmic temperature of air and gas flows (°C)
C = lift constant
c_{ps} = specific heat at constant pressure of the heater's metal plates (kJ/kg K)
C_s = lift coefficient
E = elasticity constant
f = friction factor (Moody diagram)
F = phase
F_a = decay factor
F_s = lift force
g = acceleration of gravity (m/s^2)
h = coefficient of heat transfer by convection (film coefficient) (W/m^2 K)
k = thermal conductivity of the metal plates on the regenerator heater (W/m K)
L = transfer length (mm)
LAP = ljungström air preheater
MP = mechanical precipitator
$\dot{m}$ = mass flow (kg/s)
N_{Nu} = dimensionless Nusselt number
N_{Pr} = dimensionless Prandtl number
N_{Re} = dimensionless Reynolds number
P = pressure (mm of water column)
$\dot{Q}$ = thermal power (W)
r_B = radius of the packing volume, by surface unit (mm)
r_0 = pipe radius (mm)
S = heater transfer area (m^2)
T = period (s^{-1})
t = time (s)
U = overall heat transfer coefficient (conduction in plates and convection in air) (W/m^2 K)
v = fluid velocity profile
V = average fluid velocity (m/s)
W = fluid weight (kg)
x = distance to the surface (mm)
y = depth of the boundary layer (μm)

Greek symbols

Δ = increase
Λ = wavelength (μm)
θ = temperature (°C)
∂ = partial derivative
ρ_s = density of the metal plates on the heater (kg/m^3)
φ = times of gas or air passage, respectively, through the rotary regenerative heat exchanger
μ = viscosity
ω = fluid specific weight

Subscripts

a = relative to the air
amb = relative to the atmosphere
exp = experimental
g = relative to the gas

i = inlet
m = average
max = maximum
min = minimum
o = outlet
s = relative to the surface

Superscripts

b,c = lift coefficients
i = relative to the inlet
o = relative to the outlet

References

[1] Mills, A. F., 1999, *Heat Transfer*, 2nd ed., University of California, Los Angeles.
[2] Cotta, R. M., Mikhailov, M. D., and Özisik, M. N., 1986, "Transient Conjugated Forced Convection in Ducts With Periodically Varying Inlet Temperature," Int. J. Heat Mass Transfer **30**(10), pp. 2073–2082.
[3] Schmidt, F. W., and Willmot, A. J., 1981, *Thermal Energy Storage and Regeneration*, McGraw-Hill, New York.
[4] Park, P. M., Cho, H. C., and Shin, H. D., 2003, "Unsteady Thermal Flow Analysis in a Heat Regenerator With Spherical Particles," Int. J. Energy Res., **27**, pp. 161–172.
[5] Zarrinehkafsh, M. T., and Sadrameli, S. M., 2004, "Simulation of Fixed Bed Regenerative Heat Exchangers for Flue Gas Heat Recovery," Appl. Therm. Eng., **24**, pp. 373–382.
[6] Baukal, C. E., 2000, *Heat Transfer in Industrial Combustion*, CRC LLC, Boca Raton, FL.
[7] Chapman, A. J., 1968, *Heat transfer*, 2nd ed., Ed. Interciencia, Madrid.
[8] Li, W., and Sadik, K., 1991, "Unsteady Thermal Entrance Heat Transfer in Laminar With a Periodic Variation Inlet Temperature," Int. J. Heat Mass Transfer, **34**(10), pp. 2581–2592.
[9] Zhang, H. Y., Ebadian, M. A., and Campo, A., 1980, "Effects of Heat Generation and Axial Heat Conduction in Laminar Flow Inside a Circular Pipe With a Step Change in Wall Temperature," Trans. ASME, Ser. C: J. Heat Transfer, **102**, pp. 58–63.
[10] Zariffeh, E. K., Soliman, H. M., and Trupp, A. C., 1982, "The Combined Effect of Wall and Fluid Axial Conduction on Laminar Heat Transfer in Circular Tubes," *Proc. of 7th International Heat Transfer Conf.*, **4**, pp. 131–135.
[11] Rummel, K., 1931, "The Calculation of the Thermal Characteristics of Regenerators," J. Inst. Fuel **3**, pp. 160–174.
[12] Rafidi, N. E., 2005, "Thermodynamic Aspects and Heat Transfer Characteristics of HiTAC Furnaces With Regenerators," Doctoral dissertation, Royal Institute of Technology, School of Industrial Engineering and Management, Stockholm.
[13] Yang, W., and Blasiak, W., 2002, "Combustion and Heat Flux Evaluation in a Furnace Equipped With a High-Cycle Regenerative System," *Proceedings of Challenges in Reheating Furnaces*, London.
[14] Blasiak, W., Yang, W., and Rafidi, N., 2004, "Physical Properties of a LPG Flame With High-Temperature Air on a Regenerative Burner," Combust. Flame, **136**, pp. 567–569.
[15] Hausen, H., *Heat Transfer in Counter Flow, Parallel Flow and Cross Flow*, McGraw-Hill, New York.
[16] Athey, D. R., and Chew, P. E., 1986, "Calculation of High Temperature Regenerative Heat Exchangers," *Proceedings of the International Center for Heat and Mass Transfer*, A. E. Sheindlin, ed., Hemisphere, Washington, DC, pp. 73–113.
[17] Heggs, P. J., 1986, "Calculation of High Temperature Regenerative Heat Exchangers," *Proceedings of the International Center for Heat and Mass Transfer*, ed., A. E. Sheindlin, Hemisphere, Washington, DC, pp. 115–149.
[18] Klein, H., and Eigenberger, G., 2001, "Approximate Solution for Metallic Regenerative Heat Exchangers," Int. J. Heat Mass Transfer, **44**, pp. 3553–3563.
[19] Ganapathy, V., 1989, "Cold-End Corrosion: Causes and Cures," Hydrocarbon Process., **68**, p. 57–59.
[20] Blanco, J. M., Mendía, F., and Peña, F., 2006, "Comparative Analysis of CO_2 and SO_2 Emissions Between Combined and Conventional Cycles With Natural Gas and Fuel Oil Consumption Over the Spanish Thermal Power Plants," Fuel, **85**, pp. 1280–1285.
[21] Bejan, A., Tsatsaronis, G., and Moran, M., 1996, "Thermal Design and Optimization," Wiley, New York.
[22] Vincent, A., and Meneguzzi, M., 1991, "The Spatial Structure and Statistical Properties of Homogeneous Turbulence," J. Fluid Mech., **225**, pp. 1–20.
[23] Incropera, F. P., and Dewitt, D. P., 2002, *Fundamentals of Heat and Mass Transfer*, 5th ed., Wiley, New York.
[24] Liou, T. M., Hwang, J. J., and Chen, S. H., 1992, "Turbulent Transport Phenomena in a Channel With Periodic Rib Turbulators," J. Thermophys. Heat Transfer, **6**, pp. 513–521.
[25] Sewall, E. A., and Tafti, D. K., 2005, "Large Eddy Simulation of Flow and Heat Transfer in the 180° Bend Region of a Stationary Ribbed Gas Turbine Internal Cooling Duct," *ASME Turbo Expo*, paper No. GT 2005-68518.
[26] Murata, A., and Mochizuki, S., 2001, "Comparison Between Laminar and Turbulent Heat Transfer in Stationary Square Ducts With Transverse or Angled Rib Turbulators," Int. J. Heat Mass Transfer, **44**, pp. 1127–1141.
[27] Ekkad, V., and Han, J. C., 1997, "Detailed Heat Transfer Distributions in Tow-Pass Square Channels With Rib Turbulators," Int. J. Heat Mass Transfer, **40**, pp. 2525–2537.
[28] Curtis, W. D., Logan, J. D. and Parker, W. A., 1982, "Dimensional Analysis and the pi Theorem," Linear Algebr. Appl., **47**, pp. 117–126.
[29] Naylor, D., and Tarasuk, J. O., 1993, "Natural Convective Heat Transfer in a Divided Vertical Channel: Part I-Numerical Study," ASME J. Heat Transfer, **115**, pp. 377–387.
[30] Churchill, S. W., and Usagi, R., 1972, "A General Expression for the Correlation of Rates of Transfer and Other Phenomena," AIChE J., **18**, pp. 1121–1128.
[31] Antonia, R. A., and Krogstad, P. A., 2001, "Turbulence Structure in Boundary Layers Over Different Types of Surface Roughness," Fluid Dyn. Res., **28**, pp. 139–157.
[32] Grass, A. J., Stuart, R. J., and Mansour-Tehrani, M., 1993, "Common Vortical Structure of Turbulent Flows Over Smooth and Rough Boundaries," AIAA J., **331**, pp. 437–447.

Inlet Air Cooling Applied to Combined Cycle Power Plants: Influence of Site Climate and Thermal Storage Systems

Nicola Palestra
e-mail: nicola.palestra@unibg.it

Giovanna Barigozzi
e-mail: giovanna.barigozzi@unibg.it

Antonio Perdichizzi
e-mail: antonio.perdichizzi@unibg.it

Dipartimento di Ingegneria Industriale,
Università degli Studi di Bergamo,
Viale Marconi 5,
24044 Dalmine, BG Italy

The paper presents the results of an investigation on inlet air cooling systems based on cool thermal storage, applied to combined cycle power plants. Such systems provide a significant increase of electric energy production in the peak hours; the charge of the cool thermal storage is performed instead during the night time. The inlet air cooling system also allows the plant to reduce power output dependence on ambient conditions. A 127 MW *combined cycle power plant operating in the Italian scenario is the object of this investigation. Two different technologies for cool thermal storage have been considered: ice harvester and stratified chilled water. To evaluate the performance of the combined cycle under different operating conditions, inlet cooling systems have been simulated with an in-house developed computational code. An economical analysis has been then performed. Different plant location sites have been considered, with the purpose to weigh up the influence of climatic conditions. Finally, a parametric analysis has been carried out in order to investigate how a variation of the thermal storage size affects the combined cycle performances and the investment profitability. It was found that both cool thermal storage technologies considered perform similarly in terms of gross extra production of energy. Despite this, the ice harvester shows higher parasitic load due to chillers consumptions. Warmer climates of the plant site resulted in a greater increase in the amount of operational hours than power output augmentation; investment profitability is different as well. Results of parametric analysis showed how important the size of inlet cooling storage may be for economical results.* [DOI: 10.1115/1.2771570]

Introduction

It is well known that combined cycle power output strongly decreases, as gas turbine inlet air temperature increases. In the summer time, average daily energy production may diminish even by 10%, compared to ISO (1.01325 bar, 15°C, and 60% relative humidity) conditions. In the last years peak energy cost during summer raised considerably; this was related to the continuous increase of energy consumption due to air conditioning and to the introduction in 2004 of the new Italian Electric Stock Exchange [1,2]. Thus, there is an increased interest in maintaining a low gas turbine inlet air temperature, and keeping the power output constant during peak hours. One option is given by inlet air cooling systems based on cool thermal storage: this allows the increase of electric energy produced during summer daytime when electricity price is relevant, and shifting mechanical chillers consumption during night time when energy is cheaper. In general inlet air cooling technologies may be classified into two main systems [3–5]:

- Water evaporation systems: evaporative cooling and inlet fogging; and
- Direct heat transfer systems: by direct mechanical chilling, direct adsorption chilling, and indirect mechanical chilling through cool thermal storage.

In water evaporation systems, a certain amount of demineralized water is mixed with air and starts to evaporate, decreasing the stream temperature. With such a technology, ambient wet bulb temperature is the minimum achievable temperature. Better performances are assured by direct heat transfer systems based on a chiller, to lower inlet gas turbine air temperature. The lower obtainable temperature now only depends on heat exchange surface and coolant temperature (for water typically 4°C).

The DOE carried out a detailed analysis [6] of different inlet cooling systems applied to gas turbine based power plants. With economical assumptions valid for the United States market, the best solution was found to be inlet air cooling by mechanical chilling coupled with cool thermal storage. A similar work [7] applied to the Spanish scenario showed how different design criteria, ambient conditions, and economical assumptions may arrive at a different result. In this case a small-size direct cooling system performed better than thermal storage.

Cool thermal storage, in addition to standard direct cooling system components, requires a tank; as cooling media, both chilled water and ice are commonly used. If water is chosen, the amount of stored energy depends on the difference between the warm water coming from the air coils and the cold water produced by the chillers. Among the different technologies available to keep separate warm and cold water, the most used is stratified chilled water storage. Such an application is simple and, compared to standard cooling systems, does not require any extra component but the storage tank. However, a huge volume is necessary for the tank, especially if the difference in temperature between chilled water and return water is small. Stratified chilled water systems applied to combined cycle power plants have been analyzed in a previous work [8]; in this paper their performances will be compared with the ice harvester storage system. To store cool energy, this second option uses the ice latent heat of fusion which is stored in a tank in two phase equilibrium with liquid water. Ice harvest-

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 26, 2007; final manuscript received May 10, 2007; published online February 29, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007, Paper No. GT2007-27046.

ers produce ice on the external surface of an evaporator; when a certain thickness is reached, ice is harvested by heating the evaporator and leaving the sheets of ice falling by gravity into the storage tank. The heat required by the harvesting process is generally provided by the hot refrigerant gas taken from the chiller compressor. An ice harvester system applied to a simple cycle power plant has been described by Al Bassam and Al Said [9]: in Saudi Arabia, thanks to the favorable climatic conditions, power output increment exceeded 25%. A wider analysis on the potential of other inlet cooling technologies applied to a power plant located in the same environment has been done by Alhazmy et al. [10]. A comparison between the ice harvester and stratified chilled water for power plants located in Iran was carried out by Ameri et al. [11]. They found that in such a condition the larger power output increase, given by ice thermal storage, was not enough to compensate for the highest capital cost. Another work by Zurigat et al. [12] applied to the Oman scenario concluded that, for a thermal storage system applied to a simple cycle gas turbine, economical results are generally negative. This review of previous investigations has proven that profitability of inlet air cooling systems applied to gas turbine power plants is strongly dependent on design parameters, as well as on climatic and economical conditions. The present paper reports a comparison of the results obtained for a chilled water stratified storage system (CW) and that of an ice harvester. The simulation code, already presented in [8], has been implemented with the ice harvester (IH) model. This code first provides the automatic design of the inlet cooling system, by defining size and performance of all relevant components; then it carries out the simulation throughout the year of the whole plant (i.e., the combined cycle and the cooling system). The code also provides the profitability of such an investment; this is made through the calculation of the revenue on the basis of a year due to the extra power production, and the estimation of the capital cost of the inlet cooling system.

Reference Power Plant and Inlet Air Cooling System Layout

A 127 MW combined cycle power plant is used as a reference. It is based on two Siemens SGT-800 gas turbines and a two pressure level bottoming steam cycle (Fig. 1). The cycle is rated with an efficiency of 54% and it is built in a 2×1 configuration with each gas turbine equipped with its own heat recovery steam generator (HRSG). Reference plant relevant performance data and operating conditions are reported in Table 1. Several air cooling system configurations can be applied to cool inlet air flow in a combined cycle. The one considered for the present analysis (Fig. 2) is based on centrifugal compressor chillers driven by ac motors, operating during the off-peak hours to accumulate coolant capacity into the storage system. Two different storage systems have been considered: the first one based on a stratified chilled water tank and the second one on an ice harvester system with the ice maker installed above the chilled water tank. In the last case the chillers shall operate with much lower temperatures at the evaporator, as they have to be coupled with the ice harvester machines. Hence, chillers coefficient of performance (COP) (defined as the ratio of the useful cooling energy to the electric energy input to the compressor) for ice making will be smaller than for the previous case. For an ambient wet bulb temperature of 19.5°C, the COP for the cw system is rated at 5.5, while for ice making it decreases to 4. In both configurations, when the inlet cooling system is turned on, the chilled water is pumped into a battery of cross flow heat exchangers installed in the inlet casing just upstream the compressor inlet. Water design inlet temperature is 4°C (as required by thermal stratification) for CW; it is around 1°C for IH. A limit value of 10°C for the coils air exit temperature has been assumed in the heat exchangers design. The water to cooled air stream pinch-point has been set to 2°C for both solutions. Chillers have been supposed to be cooled by evaporative towers, so their performance will be partially affected by ambient conditions.

Table 1 Combined cycle specifications

CC performances in ISO conditions	
GT power	2×43 MW
GT efficiency	36%
GT compression ratio	19:1
Exhaust gas temp.	540°C
ST power	41 MW
HP steam mass flow	31.2 kg/s
HP steam temperature	520°C
HP steam pressure	80 bar
LP steam mass flow	7.4 kg/s
LP steam temperature	220°C
LP steam pressure	7.7 bar
Condensing pressure	0.06 bar
CC power	127 MW
Gas consumption	235 MW t
CC efficiency	54%

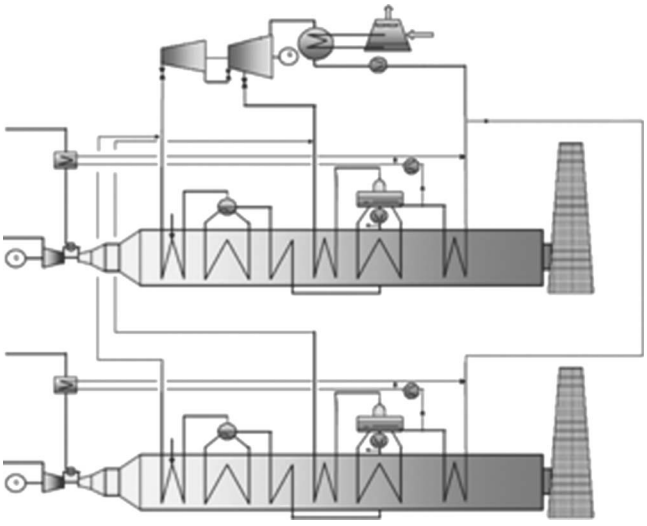


Fig. 1 Layout of the 130 MW reference CC

Simulation Code

A MATLAB computer code has been developed to model and analyze the performance of the whole plant (i.e., the combined cycle and the inlet air cooling system) for any possible operating conditions. A detailed description of the code has already been presented in [8]. For this analysis the code has been implemented in order to also simulate IH based inlet air cooling systems. The structure of the program is presented in Fig. 3.

Inlet Cooling System Design. The first step is the automatic design of the inlet air cooling system through the definition of all components, depending on the gas turbine model provided and power plant location. For the heat exchanger, frontal area, surfaces, ranks, fins typology, and position are automatically determined on the basis of the assumed temperature values for air cooling in the most severe ambient conditions occurring in the site. Storage tank volume is selected by considering the cooling demand during on-peak hours, while chiller is sized in order to restore the stored energy completely during off-peak hours. Optimization routines have been used whenever different design options were possible to satisfy imposed design conditions. An example may be given by storage tank design: while its volume depends on the cooling demand, its size proportions are automati-

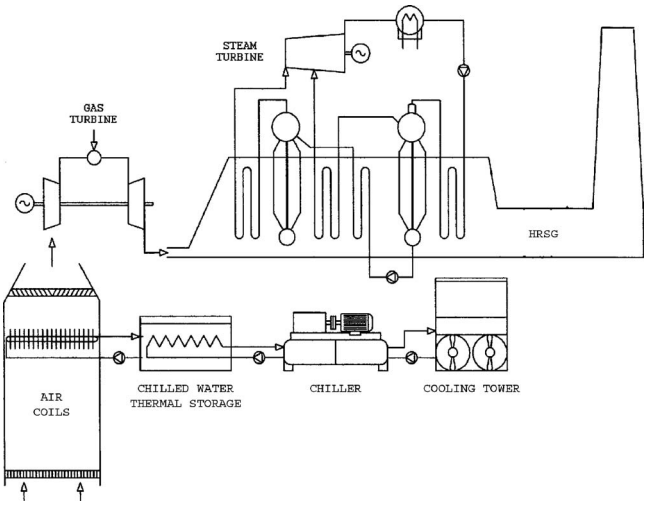


Fig. 2 Layout considered for cool thermal storage inlet air cooling system

cally chosen in order to minimize construction costs. The final choice of heat exchanger layout is performed by minimizing not only its cost, but also the inlet pressure losses to gas turbine. During the design procedure, the user may decide to undersize the cooling storage in order to limit installation costs.

Combined Cycle Performance Data Base. To carry out the simulation of the whole plant throughout the year on a 1 h basis, a detailed data base of combined cycle performance for all possible operating conditions has to be carried out previously and included in the code. Climatic data occurring in the site over 1 year (see data of Figs. 4 and 5) and inlet air conditions provided by the cooling system (air temperature, humidity, and inlet pressure losses due to air coils) have been considered. Such a data base was obtained by using a power plant simulation software (THERMOFLEX) tuned with performance data provided by the manufacturer.

Plant Control Strategy. During night hours (i.e., when the electricity price is low) chillers will be supposed to charge the storage system. An air cooling system will instead be operated during the hottest hours of the following day up to complete consumption of stored cooling capacity. Therefore a control strategy (i.e., turn on/off timing of both chillers and cooling air system) has to be defined before starting the simulation.

Through user interface, plant design and control data, like location site, climatic conditions, control strategies, and economical assumptions (electric energy prices and capital costs) will be given. In order to show the relevance of different climates on the results, two different sites, located in northern and southern Italy, respectively, have been considered in the present study. All the historical daily climatic data for the period 1994–2003 are available from NOAA [13]. Starting from this information, a method [14] to extrapolate hourly temperature and humidity from daily data was applied. In such a way it was possible to derive a typical year for each location site (Figs. 4 and 5).

Plant Simulation. When all input data have been defined, the code performs a 1 year simulation of the whole plant, determining for each hour the operating conditions of both combined cycle and inlet air cooling system. If, in a single day, the storage tank could not accomplish the requirements of cooling, an optimization is carried out. This consists of the modification of the control strategy, by reducing the number of operational hours or the inlet air temperature (by modifying the cold water flow rate) with the objective of maximizing the income of the day. At the end of the simulation, all relevant parameters are available for every hour of every day of the year. So, day by day the code computes the larger amount of electric energy produced by the combined cycle with the inlet air cooling system, the related natural gas consumption increase, and the chillers energy consumption during night time. By assuming an energy price corresponding to the value (Fig. 6) occurred in the Italian market between April 1, 2004 and March 31, 2005 (Unique National Price of the Italian Electric Power Exchange [1,2]), the code also provides the net year revenue produced by the inlet air cooling system. A natural gas cost of 0.18 €/Sm^3 and operating and maintenance (O&M) expenses equal to 1.5% of the total installation cost have been assumed. Capital costs have also been estimated on the basis of data found in the literature [15] or directly provided by cooling system manufacturers [16]. From these data, assuming the investment over 15 years against a discount rate of 6%, with an availability of the power plant of 8000 h/year, indicators of profitability like payback time, net present value of the investment (NPV), and internal rate of return are determined.

Thermodynamic Results

The first step of this investigation was to obtain the estimation of energetic performances induced by IH and CW systems. Some of the assumptions used for the sizing process for different plant locations and thermal storage technologies are resumed in Table 2.

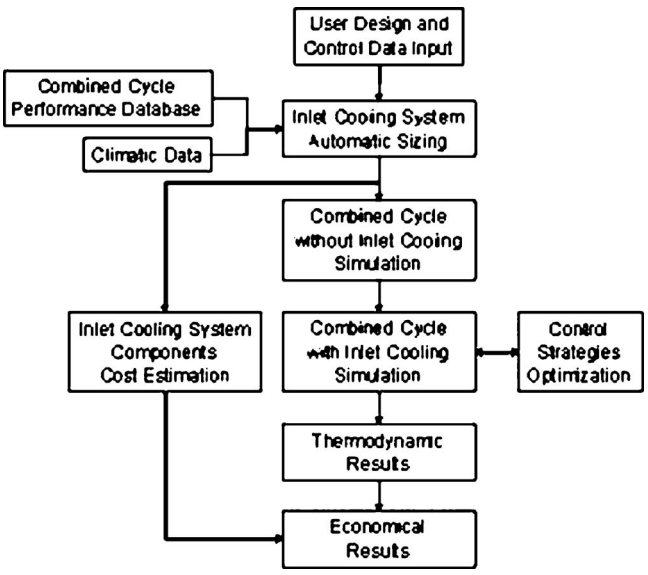


Fig. 3 Simulation program structure

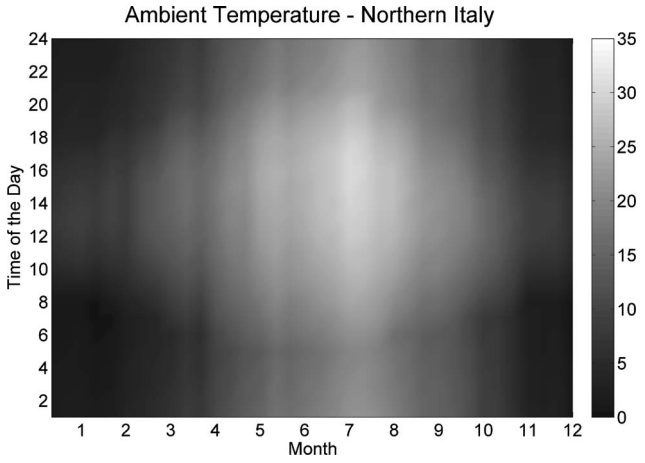


Fig. 4 Ambient temperature for a typical year in Northern Italy

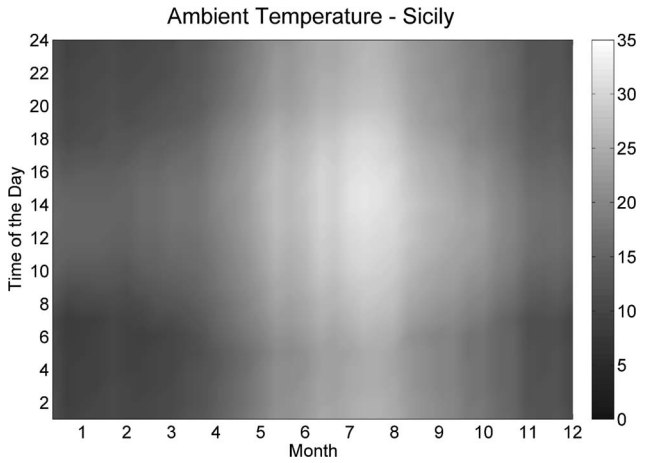


Fig. 5 Ambient temperature for a typical year in Southern Italy (Sicily)

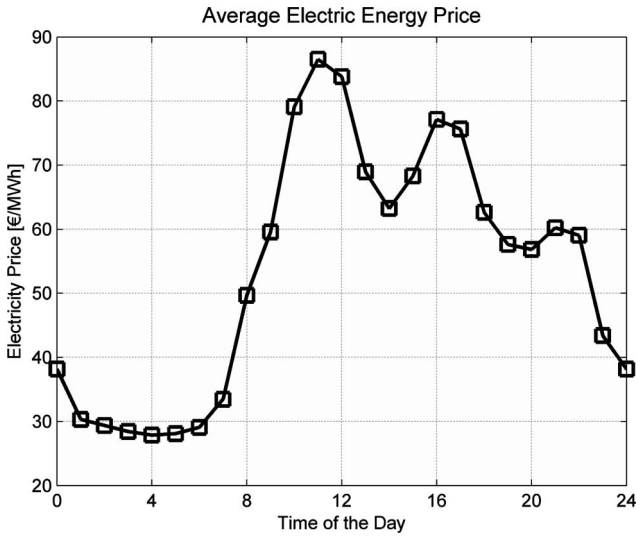


Fig. 6 Average unique national price

As design inlet air temperature and water return temperature are the same for all cases, considered total heat exchange area resulted in being wider for air coils applied to stratified CW systems (Table 3). As a consequence, total pressure losses on the air side are slightly larger for CW, if compared to IH. Cooling systems have been sized in order to guarantee, during the hottest day of the year, 50% of the cooling demand.

The main features of storage tank and chiller for all the cases considered are reported in Table 4. It is important to emphasize that thermal storage capacity is estimated "a priori" by only considering design condition operations. As one can note, the storage volume in the case of CW is more than ten times larger than for IH. Plant location climate is shown to have a significant influence on size and performances of all cooling system components.

Figures 7 and 8 show the actual values of gas turbine inlet air temperature for CW during plant operation on four typical days. Figures 9 and 10 report the corresponding power output augmentations.

As shown in the figures, the inlet air cooling temperature is different for each day, depending on ambient temperature and cooling system capacity. An optimization routine indeed provides the hourly inlet air temperature control that maximizes daily revenues. Typically, during the hottest days, a slight increment of inlet air temperature may allow the inlet air cooling system to remain operational all along the peak period. In contrast, if a lower inlet air temperature would have been chosen, thermal storage would have been ended before the time and inlet cooling should have been turned off. Figure 11 shows the ice storage level during the same typical days. One should note that during most of the summer days cold energy is completely consumed. This is due to the design assumption that thermal storage could guarantee only 50% of maximum cooling demand: value that is generally lower than cooling energy required during summer. When stored ice is completely exhausted, it is still possible to use the sensible heat capacity of cooled water contained in the tank: since ice is no longer present, thermal storage average temperature is going to increase dramatically. As a consequence air coils, designed to work with an inlet cooling water temperature of 1 °C, are no longer able to maintain compressor inlet air temperature at the design value. Figure 12 shows this behavior.

In the attempt to compensate the air temperature rise, the cold water mass flow rate to air coils is increased up to the maximum value.

Figure 13 presents the chilled water level for Case 1. As should be noted, during summer days, about 80% of total stored chilled water is consumed. This value of effectiveness is typical of stratified chilled water systems, where useful thermal stored energy is never bigger than 80–90% of the total. Thus, similarly to IH, stored cooling energy is consumed completely because the tank size only guarantees 50% of maximum demand.

Figure 14 shows the power output augmentation due to inlet air temperature reduction, for all the year hours when the air cooling system is on. The power value is net, i.e., after deductions for auxiliaries and inlet losses due to air coils. The negative values refer to off-peak hours when chillers are operating to recharge the thermal storage; they also include the gas turbine power reduction due to the presence of air coils. One can see that power augmentation for CW at high ambient temperatures reaches 10–12 MW, roughly corresponding to 8–9% of the combined cycle power at ISO conditions; during night time, power decrease is never larger than 1.2 MW.

In order to evaluate how efficiently peak energy is produced by this system, a marginal electric efficiency parameter is introduced

Table 2 Design data for different simulation cases

Combined cycle Location	Design data			
	Siemens SGT-800			
	Northern Italy		Southern Italy	
Storage technology	CW	IH	CW	IH
Comp. inlet air temp. (°C)	10	10	10	10
Chilled water temp. (°C)	4	1	4	1
Storage water return temp	8	8	8	8
Operational hours	9	9	9	9
Storage tank sizing (%)	50	50	50	50
Simulation case	1	2	3	4

Table 3 Air coils data

Simulation case	Air coils			
	1	2	3	4
Number of batteries	12	18	12	18
Number of ranks	8	6	8	8
Single battery size (m)	3.68*2.04	2.98*1.68	3.61*2.04	2.92*1.68
Front area (m^2)	90.09	90.12	88.37	88.30
Heat exch. area (m^2)	13,853	10,390	13,603	10,764
Head losses (Pa)	129	100	129	102

Table 4 Cooling system data

Simulation case	Storage tank 1	2	3	4
Storage volume (m^3)	9500	840	13,000	1135
Design storage capacity (kWh)	38,900	38,900	52,600	52,600
	Cooling system			
Chiller capacity (MWf)	4200	3450	5800	4750
Chiller size (MWe)	820	920	1220	1360
Chiller average COP	5.12	3.75	4.75	3.49
Evaporative towers size (MWf)	5050	4350	7000	6150

$$\eta_{\mathrm{MARG,CC}} = \frac{\Delta P_{\mathrm{CC}}}{\Delta Q_{\mathrm{CC}}} \tag{1}$$

where ΔP_{CC} is the net power output augmentation and ΔQ_{CC} is the extra fuel input. $\eta_{\mathrm{MARG,CC}}$ is an important parameter as it gives, for every peak hour, the reference marginal cost of generated peak energy by means of inlet air cooling. Thus it is possible to compare $\eta_{\mathrm{MARG,CC}}$ with the average efficiency of other typical solutions for generating peak energy (inlet fogging, supplementary firing, simple-cycle gas turbines, etc.).

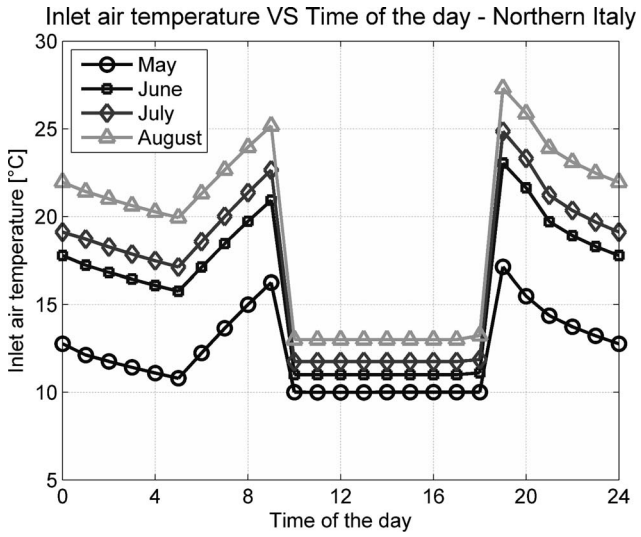


Fig. 7 Inlet air cooling temperature—Case 1

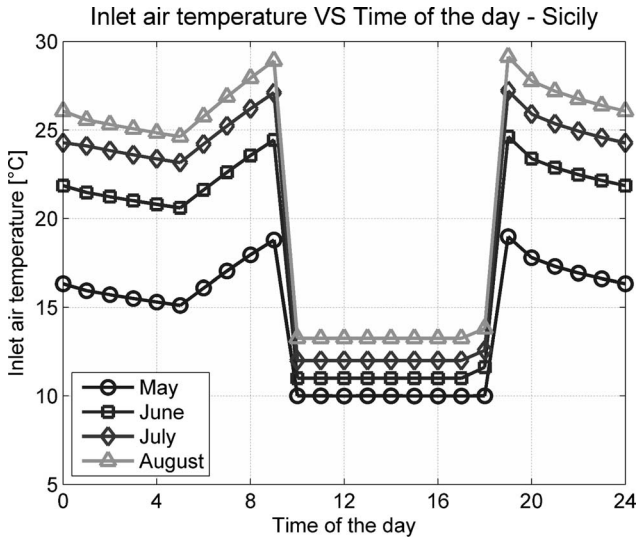


Fig. 8 Inlet air cooling temperature—Case 3

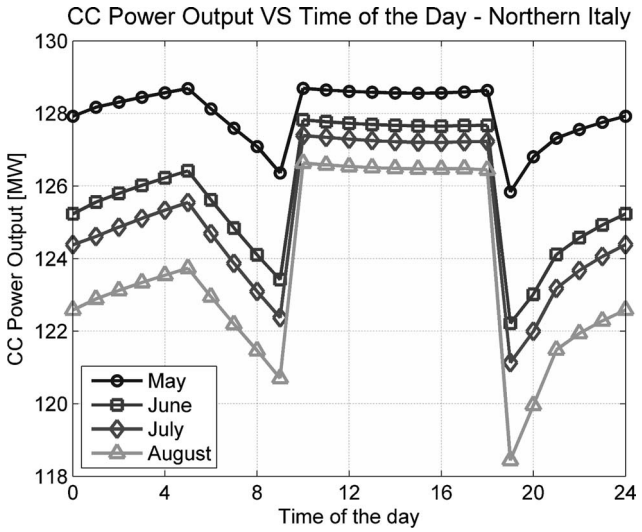


Fig. 9 CC power output—Case 1

Figure 15 shows the marginal electric efficiency for Case 1, as a function of ambient temperature. The following remarks can be made:

1. When ambient temperature is low, power output increment is small and efficiency is penalized by power losses due to air coils and auxiliaries. As a consequence, the net benefit in terms of overproduction of electric energy is obtained with a low overall efficiency, i.e., about 40–45%; and
2. For higher ambient temperatures, $\eta_{\mathrm{MARG,CC}}$ increases considerably, approaching values not far from the combined cycle net electric efficiency at ISO condition, i.e., 54%.

Gross electric energy overproduction all over the year is shown in Fig. 16, for all reference cases. From the presented results, the following features can be drawn:

1. Inlet air cooling system is operational all along the year only for southern Italy site (Cases 3 and 4);
2. Maximum power is obtained in July for the northern Italy

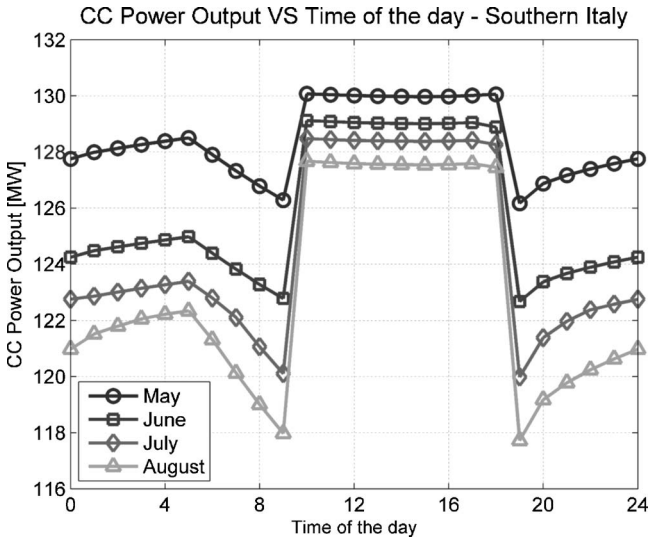


Fig. 10 CC power output—Case 3

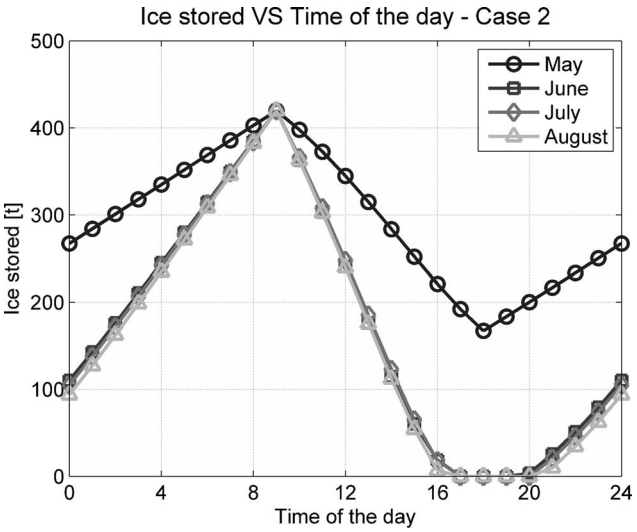


Fig. 11 Ice stored into the tank—Case 2

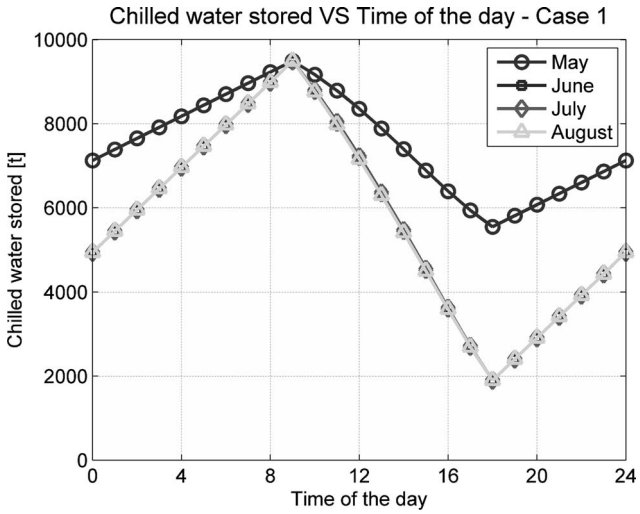


Fig. 13 Chilled water stored into the tank—Case 1

site (Cases 1 and 2) while in August (Cases 3 and 4) for the southern Italy site. Generally, the gap between these two sites is around 0.5–1 GWh/month.

3. CW generally performs better in the summer time, because the increase in inlet air temperature value implies a larger availability of cooling storing capacity per unit volume, due to the increase in temperature difference across the tank. The rise of cooling energy availability allows the increase of gross electric energy overproduction. On the contrary IH storing capacity per unit volume remains almost constant; and
4. IH performs better than CW during the winter and middle seasons, when the inlet cooling system is working at design inlet temperature (10°C) or is not operational, because of smaller head losses due to air coils.

Auxiliaries' consumptions are presented in Fig. 17. A similar seasonal trend to the gross electric energy overproduction can be noticed. Most of the auxiliaries depend on chiller electric energy requirements:

1. The southern Italy hotter climate affects chiller consumptions because cooling demand is higher and because chillers COP is lower. When wet-bulb temperature is higher, cooling evaporative towers efficiency is reduced, and consequently chiller condenser temperature rises driving a reduction of the overall COP; and
2. Chillers applied to CW perform more efficiently as evaporator temperature is higher, thus night time consumption of such a technology is lower than that of IH by about 22–23%.

In Table 5 the thermodynamic performances, averaged all over the year, are finally resumed, together with the chiller system total operational hours. The most significant parameters are reported in order to evaluate not only power augmentation and energy overproduction, but also the efficiency at which this energy is generated. $\eta_{MARG,GROSS}$ and $\eta_{MARG,NET}$ are calculated as the ratio between gross and net electric energy overproduction, respectively, and the fuel extra consumptions. From the results obtained it is possible to infer that:

1. Hotter climatic conditions give the opportunity to keep IC system operational for a different amount of hours: in southern Italy it remains in use all along the year, while in northern Italy does not;
2. Maximum and average power output augmentations during

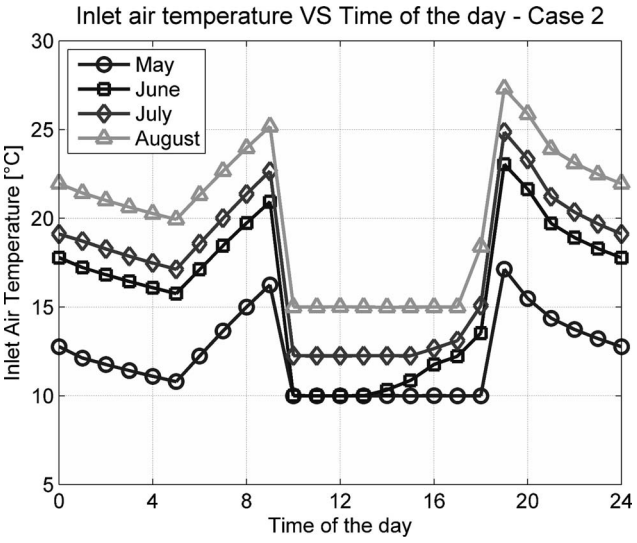


Fig. 12 Inlet air cooling temperature—Case 2

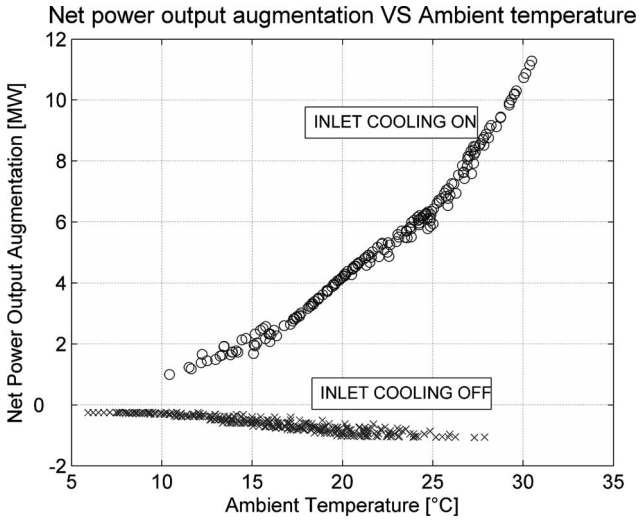


Fig. 14 Net power output augmentation—Case 1

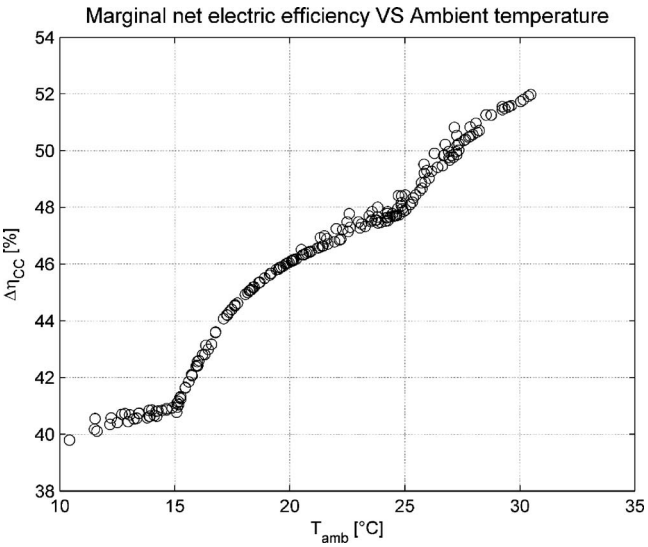


Fig. 15 Marginal net electric efficiency—Case 1

peak hours are also different: larger values are obtained in southern Italy with a maximum of 10.3% for Case 3; the best average value is 4.54 MW for Case 4;

3. Gross electric energy overproduction is slightly larger for IH, even if net overproduction is larger for CW in both the sites considered because of auxiliaries;
4. Marginal efficiency of gross energy overproduction is around 47% for all reference cases: as already noted, this is generally better than the average efficiency of other technologies. As an example, if peak energy is produced with a simple-cycle gas turbine it would present an efficiency of 36–40% depending on the model, while supplementary firing would operate with efficiency at around 27%; and
5. If energy consumed by auxiliaries overnight is considered, the net marginal efficiency decreases to 36% for IH and to 39% for CW. However, it has to be taken into account that peak energy is generated in day hours with an efficiency of 47% (with a lower marginal cost), while the 8–10% penalty for auxiliaries is due to energy consumed by night, i.e., when electric energy has a lower value.

Economic Results

Economic evaluations of the investment have been investigated by using a data base of capital costs of the main items. Every component of the inlet air cooling system cost has been related to a certain number of physical parameters by using correlations found in the literature [15] or provided by manufacturers [16]. As an example, estimated installation cost of a CW storage tank is reported in Fig. 18 as a function of its volume. Different soil occupation costs have been considered in this evaluation because required terrain may be of property or not. As a reference a cost of 100 €/m^2 was estimated and included in the storage tank cost.

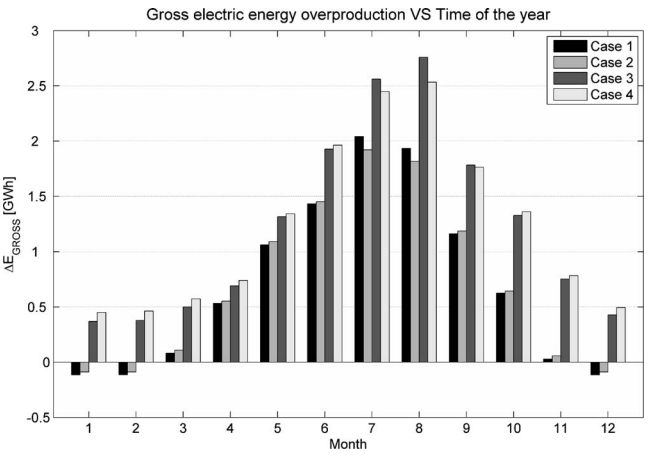


Fig. 16 Gross electric energy overproduction

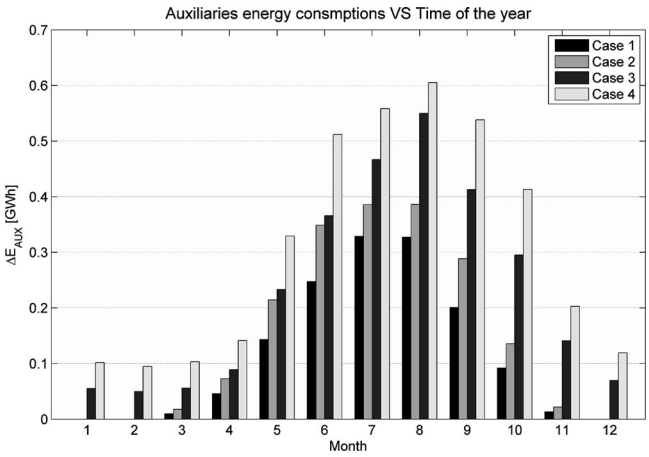


Fig. 17 Auxiliaries' energy consumptions

Economic results given by the investment for all reference cases are summarized in Table 6. For the CW system, about 50% of the total capital cost depends on the storage tank: this happens because of the huge volume required. In contrast, the IH storage tank costs seven times less, but expensive ice generators are needed together with special chillers, generally more expensive than standard models. The selection of 10°C for the design inlet air temperature plays against CW systems: these become cheaper as far as inlet air temperature is increased and, therefore, tank storing capacity per unit volume increases. In such a way storage tank volume can be reduced and, consequently, its installation cost. Both solutions perform similarly in terms of cash flows; a net positive annual revenue of about 220 € for northern Italy and of 370 € for southern Italy have been found. This difference is explained only by the site climate that is obviously more favorable for Cases 3 and 4. Also total capital cost is influenced by plant location, but even if a plant located in southern Italy would cost about 30% more than in northern Italy, it would generate an annual cash flow that is 70% greater. As expected, profitability indexes are more dependent on site location than on adopted inlet cooling technology: to install an IC system in southern Italy gives a NPV (at 15 years) of 1.7–1.8 M€ with a payback time that is only 6 years.

Table 5 Thermodynamic performances

	Thermodynamic performances			
Simulation case	1	2	3	4
Total operation hours	2100	2100	3300	3300
Max ΔP_{CC} (MW)	11.3	10.6	13.1	12.2
Max ΔP_{CC} (%)	8.9	8.3	10.3	9.6
Average ΔP_{CC} (MW)	4.1	4.1	4.48	4.54
Average ΔP_{CC} (%)	3.2	3..2	3.5	3.6
ΔE_{GROSS} (GWh)	8.6	8.6	14.8	15
ΔE_{AUX} (GWh)	1.4	1.9	2.8	3.7
ΔE_{NET} (GWh)	7.2	6.7	12	11.3
ΔQ_{TOT} (GWh)	18.35	18.19	31.2	31.3
$\eta_{MARG,GROSS}$ (%)	46.9	46.9	47.4	47.9
$\eta_{MARG,NET}$ (%)	39.2	36.8	38.5	36.1

Parametric Analysis

As design choices heavily influence investment profitability, a parametric analysis, by varying thermal storage size, has been carried out. The total cooling energy that can be stored and used daily is directly influenced by tank volume reduction or increase. This choice determines the characteristics of main components like chillers, ice generators, and evaporative towers. It also influences the annual revenues, because extra production of electricity is someway proportional to total consumed cooling energy. Therefore thermal storage volume size has been varied for all reference cases and the consequent variation of both thermodynamic and economical performances have been analyzed. Figure 19 shows in a logarithmic scale the relationship between gross electric energy overproduction and storage volume; reference cases have been highlighted on every line with a black mark.

An increment of thermal storage size allows for a bigger energy production, but one can note that the first derivative of such a function is negative, thus a continued growth of tank volume is going to give less benefits in terms of electric energy production. Especially for CW, when a volume about 4000–6000 m^3 is reached, the energy overproduction increases slightly. This result may be translated in terms of profitability: Figure 20 reports the investment payback time as a function of thermal storage size. Reference cases were not optimized: by carefully sizing the IC system PBT may decrease from 7–8 years to 5, for northern Italy, and from 5–6 years to 3.5 for southern Italy.

Figure 21 shows net present value NPV at 15 years as a function of total capital cost of the investment for all the reference cases. Maximum NPV is obtained for both location sites using CW technology: by reducing IC size and, therefore, installation cost, it is possible to achieve an NPV of 2.5 M€ for southern Italy and of 1.2 M€ for northern Italy. If these values are compared with reference cases, a 75% improvement for southern Italy and a 42% one for northern Italy become possible. Such an approach may be repeated even for many other design parameters (like, for example, design inlet air temperature) and control strategies, with the objective to further increase profitability of the investment. This goal could be obtained if a multivariable optimization procedure would be applied. This approach would lead to automatically size all the relevant design parameters in order to get the best economical results. This kind of investigation will be carried out in the future.

Table 6 Economic parameters of the investment

Simulation case	1	2	3	4
	Capital costs			
Air coils (k€)	261	239	281	303
Storage tank (k€)	774	107	1023	134
Chillers (k€)	278	317	382	439
Ice generators (k€)	0	411	0	569
Cooling towers (k€)	79	75	108	102
Piping (k€)	38	32	41	35
Pumps (k€)	60	59	63	61
Electric auxiliaries (k€)	43	35	59	49
TOT (k€)	**1533**	**1275**	**1957**	**1692**
Cost per incremental kW (€/kW)	**136**	**120**	**149**	**138**
	Cash flows			
Extra production of electricity (k€)	594	571	999	965
Extra consumption of fuel (k€)	−344	−340	−584	−586
O&M costs (k€)	−23	−19	−29	−25
TOT (k€)	227	212	386	354
	Profitability			
NPV 15 (k€)	680	780	1796	1743
IRR 15 (%)	12.3	14.4	18.2	19.5
PBT (years)	8.87	7.70	6.22	5.80

Conclusions

Two different cool thermal storage technologies, applied to air inlet cooling of combined cycle power plant, have been analyzed and compared for two sites located, respectively, in northern and southern Italy. An in-house code previously developed was extended and adapted in order to also simulate the ice harvester. By using the same design parameters and climatic data both CW and IH guaranteed roughly the same amount of power augmentation and gross energy overproduction. Only parasitic loads, during off-peak hours are different: the average COP of IH is considerably smaller than CW, thus energy consumptions due to chillers will be higher. Marginal efficiency of peak energy overproduced is similar for all reference cases and it is better than typical values of other peaking technologies. Climatic conditions of site locations proved to be more influent: thus, in southern Italy it is possible to obtain almost double the gross electric energy overproduction as in northern Italy. The economy of the inlet cooling system strongly depends on design parameters: in reference cases IH

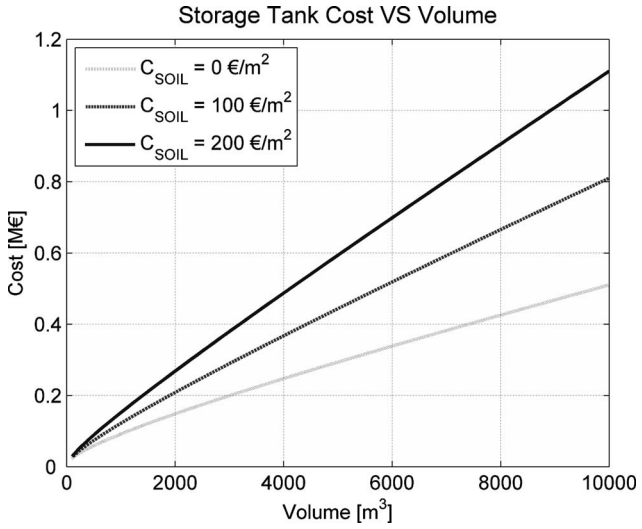


Fig. 18 Storage tank cost as a function of its volume

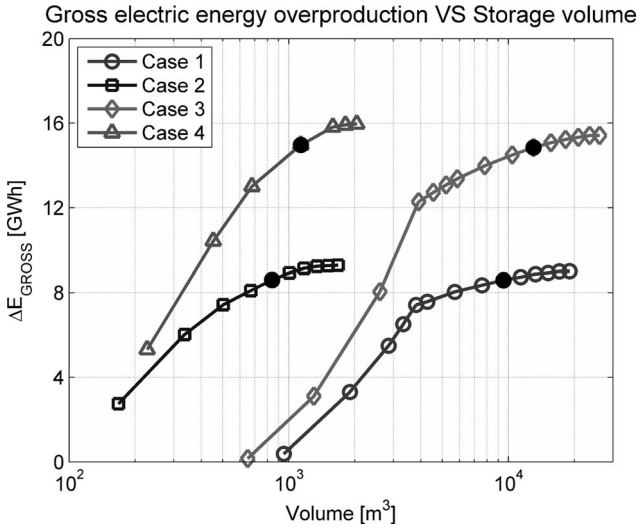


Fig. 19 Gross electric energy overproduction as a function of thermal storage volume

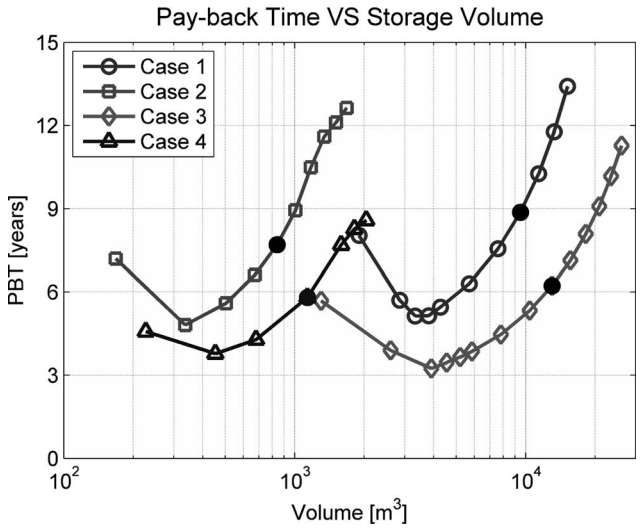


Fig. 20 Payback time of the investment as a function of thermal storage volume

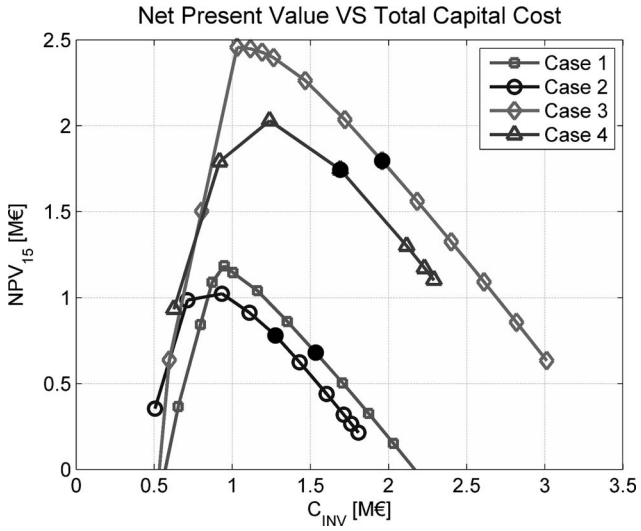


Fig. 21 Net present value as function of the total capital cost

proved to be cheaper in terms of installation costs and slightly better in terms of profitability of the investment. A sensitivity analysis showed that, by optimizing the thermal storage size, better results with CW technology can be achieved. Southern Italy locations, if compared to northern Italy, are particularly advantageous for inlet air cooling adoption; net present value of the investment is almost double and payback time may be reduced to 3 years, with an appropriate sizing of CW storage.

Nomenclature

CC = combined cycle
CW = stratified chilled water
GT = gas turbine
HP = high pressure
IH = ice harvester
IC = inlet cooling
LP = low pressure
ST = steam turbine
ΔE_{AUX} = tearly electric energy consumed by inlet cooling
ΔE_{GROSS} = yearly gross electric energy overproduced
ΔE_{NET} = yearly net electric energy overproduced
$\eta_{MARG,CC}$ = marginal electric efficiency
$\eta_{MARG,GROSS}$ = average gross marginal electric efficiency
$\eta_{MARG,NET}$ = Average net marginal electric efficiency
ΔP_{CC} = power output augmentation
ΔQ_{CC} = fuel extra consumptions
ΔQ_{TOT} = yearly extra consumptions of fuel thermal energy

References

[1] GME, 2004, "Rapporto Annuale Mercato Elettrico," www.mercatoelettrico.org
[2] GME, 2005, "Rapporto Annuale Mercato Elettrico," www.mercatoelettrico.org
[3] Omidvar, B., 2001, "Gas Turbine Inlet Air Cooling System," *Proceedings of the 3rd Annual Australian Gas Turbine Conference*.
[4] Cortes, C., and Willems, D., 2003, "Gas Turbine Inlet Air Cooling Techniques, An Overview of Current Technologies," *Proceedings of PowerGen 2003*, Las Vegas, NV December 9–11.
[5] Kakaras, E., Doukelis, A., Prelipceanu, A., and Karellas, S., 2004, "Inlet Air Cooling Methods for Gas Turbine Based Power Plants," ASME Paper No. GT-2004–53765.
[6] Brown, D. R., Katipamula, S., and Koynenbelt, J. H., 1996, "A Comparative Assessment of Alternative Combustion Turbine Cooling Systems," DOE Report No. PNNL-10966.
[7] Gareta, R., Romeo, L. M., and Gil, A., 2004, "Methodology for the Economic Evaluation of Gas Turbine Air Cooling Systems in Combined Cycle Applications," Energy, **29** pp. 1805–1818.
[8] Palestra, N., Barigozzi, G., and Perdichizzi, A., 2006, "Inlet Air Cooling Through Thermal Storage Systems for Combined Cycle Power Plants," *Proceedings ATI/ASME Conference 2006, Energy Production, Conservation and Distribution*.
[9] Al Bassam, A., and Al Said, Y. M., 2001, "Qassim Central Power Plant Inlet Air Cooling System," ASME Paper No. GT-2001–53765.
[10] Alhazmy, M. M., Jassim, R. K., and Zaki, G. M., 2007, "Performance Enhancement of Gas Turbine by Inlet Air-Cooling in Hot and Humid Climates," International Journal of Energy Researh, in press.
[11] Ameri, M., Hejazi, S. H., and Montesar, K., 2005, "Performance and Economic of the Thermal Energy Storage Systems to Enhance the Peaking Capacity of Gas Turbines," Soryushiron Kenkyu, **25** pp. 241–251.
[12] Zurigat, Y. H., Dawoud, B., and Bortmany, J., 2006, "On the Technical Feasibility of Gas Turbine Inlet Air Cooling Utilizing Thermal Energy Storage," Int. J. Energy Res. **30** pp. 291–305.
[13] NOAA, 2004, "Surface Data, Global Summary of day," www4.ncdc.noaa.gov
[14] Cesaraccio, C., Spano, D., Duce, P., and Snyder, R. L., 2001, "An Improved Model for Determining Degree-day Values from Daily Temperature Data," Int. J. Biometeorol **45**, pp. 161–169.
[15] ASHRAE, 1994, *Design Guide For Cool Thermal Storage*, ASHRAE, Atlanta, GA.
[16] Trane, 2006, "Quick Reference for Efficient Chiller System Design," www.trane.com/commercial/

Feasibility Study of an Intercooled-Cycle Marine Gas Turbine

Wen Xue-you

Xiao Dong-ming

Harbin Marine Boiler and Turbine Research Institute,
No. 108 Hongqi Avenue Harbin China,
Harbine, Heilongjiang 150036, China

From the perspective of an overall entity we analyzed the performance obtainable from the adoption of an intercooled-cycle gas turbine under different typical cycle parameters of a gas turbine. On this basis, a study was conducted of the conversion of a high-power simple-cycle marine gas turbine (MGT-33) into a type of intercooled-cycle marine gas turbine. The precondition of the conversion is to keep the flow path and the majority of the structure of the original engine gas generator unchanged in order to inherit the reliability of the prototype machine. The results of the study indicate that after the adoption of an intercooled cycle under the precondition of performing minimum structure modifications and maintaining the compactness of the engine as a whole, there is still a significant enhancement of the gas turbine overall performance with its power output and efficiency being increased by about 34% and 4.1%, respectively, demonstrating the merits of the engineering conversion under discussion. [DOI: 10.1115/1.2799526]

Foreword

Beginning in the late 1980s and with the continually rising demand of main propulsion plants for surface naval vessels by navies around the world, high-power marine gas turbines (above 25 MW) have become the main current of development for the navies of various countries. For details, see Table 1.

Figure 1 shows the development tendency of high-power marine gas turbine performance for the last 15 years. With a gradual increase in single-machine capacity, the maximum power is currently about 36 MW (50,000 hp at ISO conditions). There is also a gradual enhancement in efficiency with the efficiency of a simple cycle attaining 40% and that of a complex cycle 42%.

It is generally recognized that marine gas turbines of 30–36 MW (40,000–50,000 hp) can meet the main propulsion plant requirements of high-power gas turbines for large- and medium-sized surface vessels by the navies of various countries in the forthcoming 10–15 years.

There are two approaches for enhancing the performance of marine gas turbines: (1) On the basis of a traditional simple cycle by increasing the pressure ratio and turbine inlet temperature and by improving component efficiency, it is possible to attain the desired new aim. In this connection, LM2500+ can serve as a representative engine; (2) by employing a complex cycle and through the introduction of cycle improvements, a higher performance can be realized. In this case, WR-21 engine with the use of a complex cycle featuring intercooling and recuperation can be cited as a representative machine.

On the basis of an existing engine with its core components remaining maximally unchanged, the authors have explored the possibility of developing a high-power marine gas turbine through the use of only a relatively simple intercooled cycle (IC).

Intercooled Cycle Analysis

First, the relationship between the performance of an intercooled-cycle gas turbine and main cycle parameters is to be quantitatively examined. The engine is composed of a dual-shaft gas generator (including an intercooler) and a power turbine (see Fig. 2). The assumed component performance parameters are listed in Table 2.

Under the condition of different total pressure ratios (PR $=12:42$), different turbine inlet temperatures (TIT $=1100-1400\,^{\circ}\mathrm{C}$) and different intercooler efficiencies ($\varepsilon=0-0.85$) [2] the calculation results for engine overall performance are shown, respectively, in Figs. 3–6.

Given below are the results of a simple analysis of an intercooled cycle. When a practical study is conducted of an existing engine involving its development to form a prototype engine of intercooled cycle gas turbine, one can discover some factors favorable to the enhancement of efficiency. Such factors include a moderate increase of turbine inlet temperature, an increase in working point efficiency of some components, and the changes in corrected speed of some components after their renewed matching, etc. The following makes it possible to realize a simultaneous enhancement of engine power and engine cycle efficiency:

(1) From Fig. 3, one can perceive the impact of adopting intercooling on engine efficiency. Under the parameters of modern marine gas turbines (turbine inlet temperature at about $1300\,^{\circ}\mathrm{C}$, pressure ratio 20:25), the use of intecooling has very little influence on engine cycle efficiency. Moreover, with an increase in intercooler efficiency, there will be a slight decrease in cycle efficiency. This is because with the use of intercooling though the power consumption of the compressor being reduced and effective power output increased, there is a reduction in compressor outlet temperature. To make the gas achieve the predetermined turbine inlet temperature, one has to consume more fuel. Only under the condition of a high-pressure ratio will the impact of a compressor power consumption decrease surpass the influence of fuel increase caused by the introduction of intercooling. In this case, its efficiency will be higher than that of a simple cycle.

(2) Figure 4 shows the impact of intercooling on engine specific work. It is obvious that the use of intercooling will lead to a significant increase in gas turbine specific work. The greater the intercooler efficiency, the greater the specific work. In the meantime, the pressure ratio value corresponding to a maximum specific work also increases dramatically. Under the parameters of modern marine gas

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 28, 2007; final manuscript received June 6, 2007; published online March 4, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea, and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007.

Table 1 General situation of the development of high-power marine gas turbines(up to the year of 2005) [1]

Model	LM2500	FT8	UGT-25000	WR-21	LM2500+	MT-30
First year available	1969	1990	1993	1997	1998	2001
ISO MAX(hp)	33,600	36,860	42,400	33,850	40,500	48,275
Cycle efficiency	0.372	0.389	0.381	0.421	0.391	0.398
SFC(kg/kW hr)	0.227	0.217	0.221	0.200	0.215	0.212
Pressure ratio	19.3	18.8	21	16.2	22.2	24
Flow(kg/s)	70.4	83.3	87.6	73.1	85.8	116.7
Manufacturer	GE	P&W	Zorya-Mashproekt	R.R	GE	R.R

turbines, the specific work after the use of intercooling can be increased by 22–30%.

(3) Figure 5 shows the superimposition of the overall performance curves of a simple cycle and an intercooled one. It can be seen clearly that after the use of intercooling, both the pressure ratio value corresponding to a maximum specific work or that corresponding to an optimum efficiency have all increased very significantly.

(4) Figure 6 shows the impact of intercooling on compressor outlet temperature. This signifies that the temperature of the cooling air entering the high temperature turbine blades has decreased. When the metal surface temperature of the high temperature turbine blades is kept unchanged, it is allowed to moderately enhance the turbine inlet temperature. In addition, when the compressor corrected speed remains unchanged, the physical speed of the high pressure rotor will markedly decrease, thereby significantly reducing its working stress.

(5) To sum up, under the parameters of modern marine gas turbines, the use of intercooling can significantly enhance specific work and increase the power output of the engine. The higher the intercooler efficiency, the greater the effective power output. The use of intercooling exercises a relatively weak influence on cycle efficiency. The greater the intercooler efficiency, the lower the cycle efficiency. However, in general, the margin of efficiency change is relatively small.

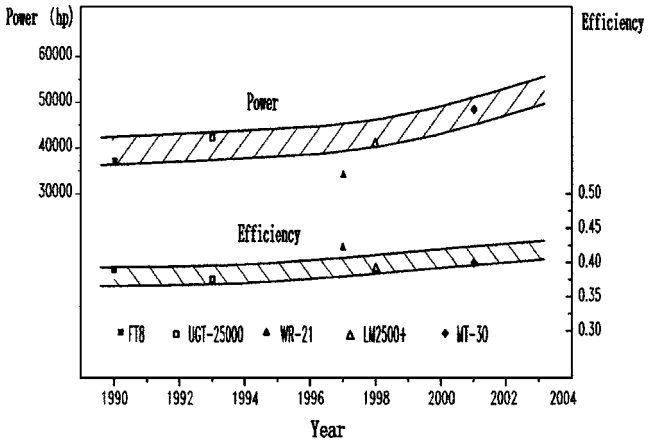


Fig. 1 Trend of the change in power output and efficiency of high-power marine gas turbines

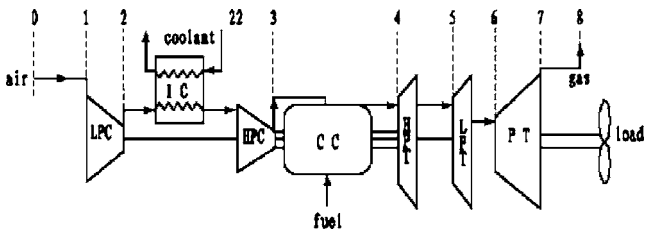


Fig. 2 IC cycle schematic drawing

Some Specific Cases of Intercooled Cycle Schemes

It is determined to perform a study by selecting a certain model of engine (MGT-33) to serve as a prototype engine and develop it to form an IC cycle. The aim of the study is to determine the technical feasibility of the selected scheme. For the main performance of simple cycle gas turbine MGT-33, see Table 3.

The main principles governing the development of gas turbine MGT-33 to CGT-IC cycle gas turbine can be given as follows:

(1) At ISO conditions, CGT-IC has the same turbine inlet temperature as the prototype engine or an increase of the inlet temperature not exceeding 30°C;
(2) At ISO conditions, the corrected speed of the CGT-IC low-pressure compressor is the same as that of the prototype engine;
(3) Adopt to a maximum degree the flow path of the gas generator of the prototype engine;
(4) The cooling structure of the high- and low-pressure turbine blades and also the high-pressure turbine rotating blades will remain unchanged;
(5) Inherit to a maximum degree the reliability of the prototype engine components.

The implementation of the above-mentioned principles has provided the basic preconditions for enhancing the power output and cycle efficiency of the engine after its modification. Moreover, under the precondition of making minimum changes to the structure of the gas generator of the prototype engine and by taking

Table 2 Calculation condition

Parameter	Unit	Value
P_0	bar	1.013
T_0	°C	15
T_{cool}	°C	20
ΔP_{in}	mm H_2O	100
ΔP_{ex}	mm H_2O	400
PR_{LPC}/PR_{HPC}		1
η_{LPC}		0.88
η_{HPC}		0.90
η_{CC}		0.99
η_{HPT} [a]		0.87
η_{LPT} [a]		0.89
η_{PT}		0.92
ξ_{IC}		0.05
ξ_{CC}		0.05
ξ_{ID}		0.01
LHV	kJ/kg	42700

[a]With 1200°C serving as a benchmark, an increase (or decrease) of turbine inlet temperature by every 100°C will lead to a decrease (or increase) in turbine efficiency by 1.25%.

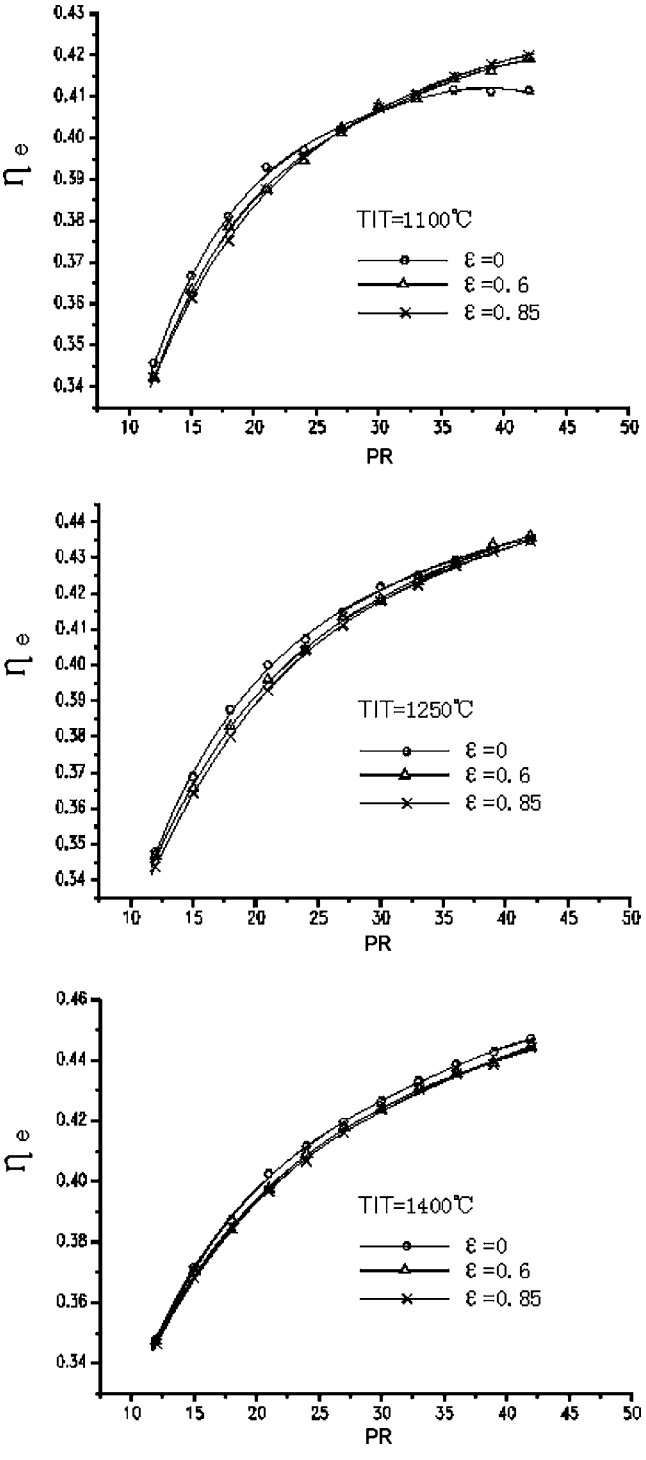


Fig. 3 Impact of intercooling on engine efficiency

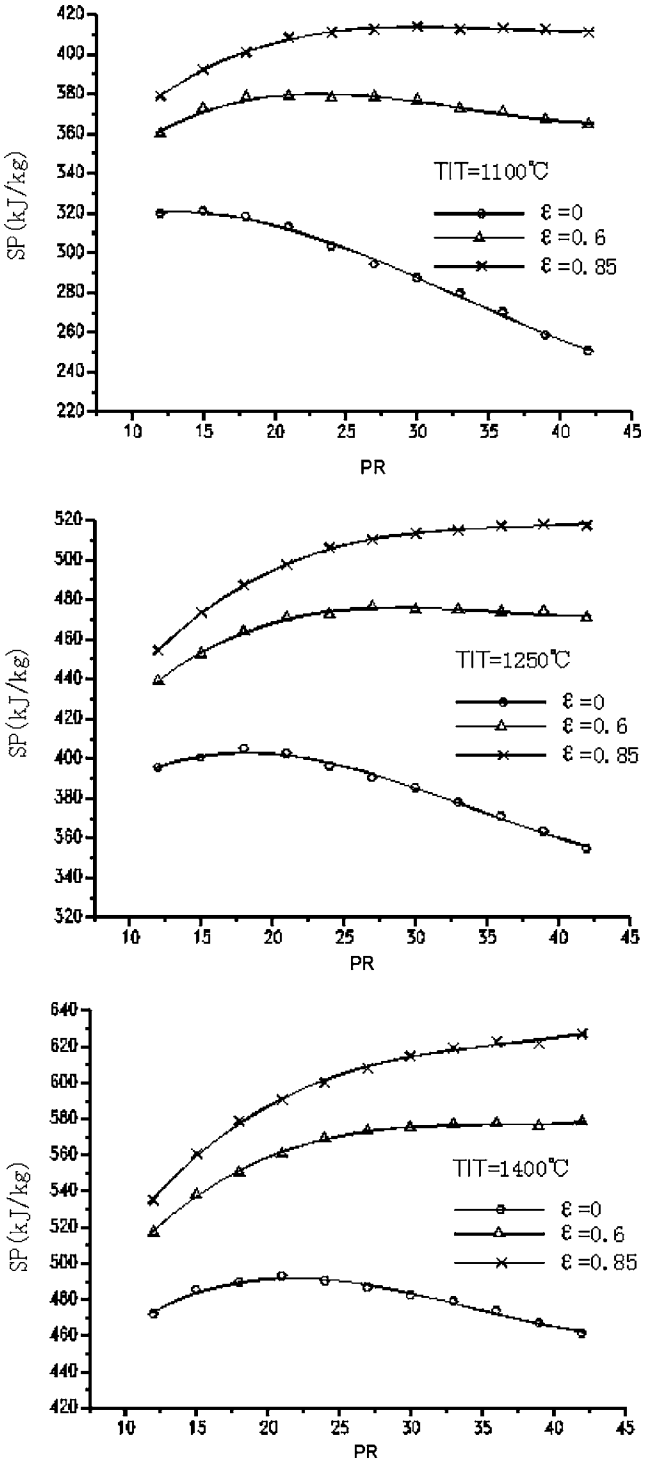


Fig. 4 Impact of intercooling on engine specific work

advantage of the maximum versatility of the flow path, the inherent high reliability can be readily taken over in its entirety.

For the cycle performance after optimization calculations, see Table 4. For the schematic drawings of MGT-33 and CGT-IC, see Fig. 7.

The technical highlights of the scheme (in our case, Scheme II) can be given as follows:

(1) Compressor. From the low-pressure compressor, the last two stages are to be removed with the new last stage guide vanes and straightening blades being designed anew. This enables the working points of the low-pressure compressor after stage removal and under the condition of keeping the reduced flow rate and speed unchanged to have the outlet pressure lowered and to balance the influence of the decrease of a reduced flow rate at the high-pressure compressor inlet caused by the use of intercooling. The pressure ratio of the low-pressure compressor after adjustments will be reduced by 19.1% and the efficiency enhanced by 1.82%. With the high-pressure compressor being kept unchanged, the corrected speed after a matching process will be enhanced by 6.2%, physical speed decreased by 11.5%,

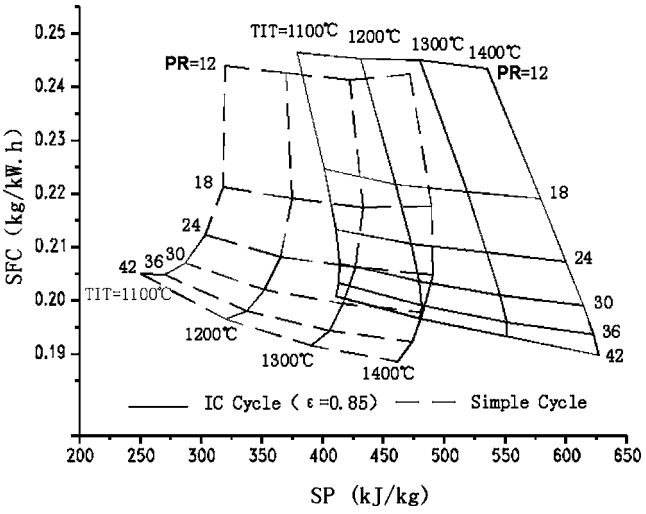


Fig. 5 Overall performance curves of a simple cycle and an intercooled one

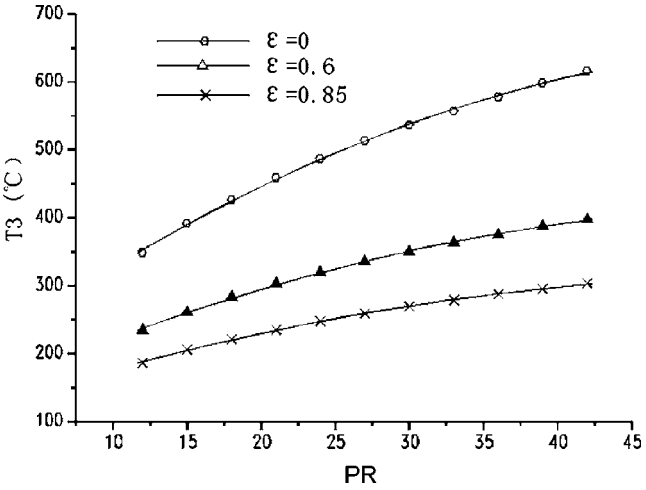


Fig. 6 Impact of intercooling on compressor outlet temperature

Table 3 Gas turbine MGT-33 performance at ISO conditions

Ser. Number	Parameter	Symbol	Unit	
1	Power	Ne	kW	28,500
2	Cycle efficiency	ηe	%	37
3	Air flow rate	Ga	kg/s	90
4	Turbine inlet temperature	TIT	K	1543
5	Total pressure ratio	PR		22.2

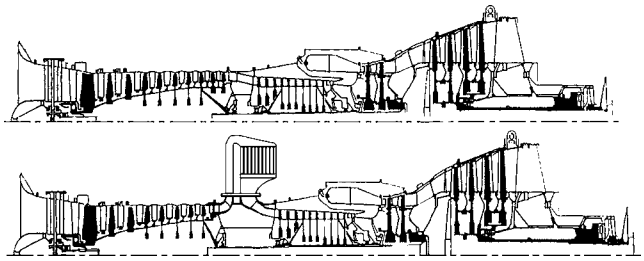


Fig. 7 Schematic drawings of MGT-33 and CGT-IC

pressure ratio increased by 20.5%, and efficiency decreased by 1.4%.

(2) Intercooler. The intercooler is a key component for raising power output by a large margin. The insertion of an intercooler between two compressors enables air to be precooled before entering the high pressure compressor, thereby decreasing the power consumption of the high-pressure compressor and effectively enhancing the power output of the engine. To realize an intercooler with high compactness and low flow resistance, a modularized structure has been adopted. With an intercooler efficiency of 0.85, the total pressure loss coefficient of the intercooler is 0.05.

(3) High and low-pressure turbine. Owing to the specific features of the prototype engine, such as a relatively high turbine inlet temperature, a complicated blade cooling structure of the high- and low-pressure turbine, the cooling structure of rotating blades of the high- and low-pressure turbine have been kept unchanged in order to lower the technical risk to a minimum. After a matching process, the expansion ratio of the high-pressure turbine has been reduced greatly by the power consumption decrease of high-pressure compressor, but the efficiency of the high-pressure turbine decreases by 1.8% for the increase of the stage loading coefficient. The removal of the last two stages of the low-pressure compressor has led to a reduction of the expansion ratio of the low-pressure turbine by 16.3%, with its efficiency increased by 0.9%. Moreover, owing to the decrease of the high-pressure turbine expansion ratio and the increase of turbine inlet temperature, the temperature of the gas entering the low-pressure turbine has increased by 5.7%. Due to a significant decrease in the temperature of the cooling air entering the high- and low-pressure turbine blades, the metallic surface temperature of the blades will not exceed the corresponding temperature of the prototype engine. Moreover, the lowering of the physical speed of the high-pressure rotor has led to a decrease in centrifugal force by 21.7%, but the increase of thermal stresses in blades should be taken into account.

(4) Power turbine. The use of intercooling and the removal of the last two stages from the low-pressure compressor have resulted in the decrease of power consumption by the high- and low-pressure compressor, an increase in power turbine inlet temperature by 9.2%, an enhancement of expansion ratio by 36.2%, a significant increase in power turbine power output, and a need to design the power turbine anew.

Table 4 Performance of gas turbine CGT-IC at ISO conditions

Ser. Number	Parameter	Symbol	Unit	Scheme I	Scheme II
1	Power	Ne	kW	36,400	38,200
2	Cycle efficiency	ηe	%	38.1	38.5
3	Air flow rate	Ga	kg/s	89.6	89.6
4	Turbine inlet temperature	TIT	K	1543	1572
5	Total pressure ratio	PR		22.3	21.2

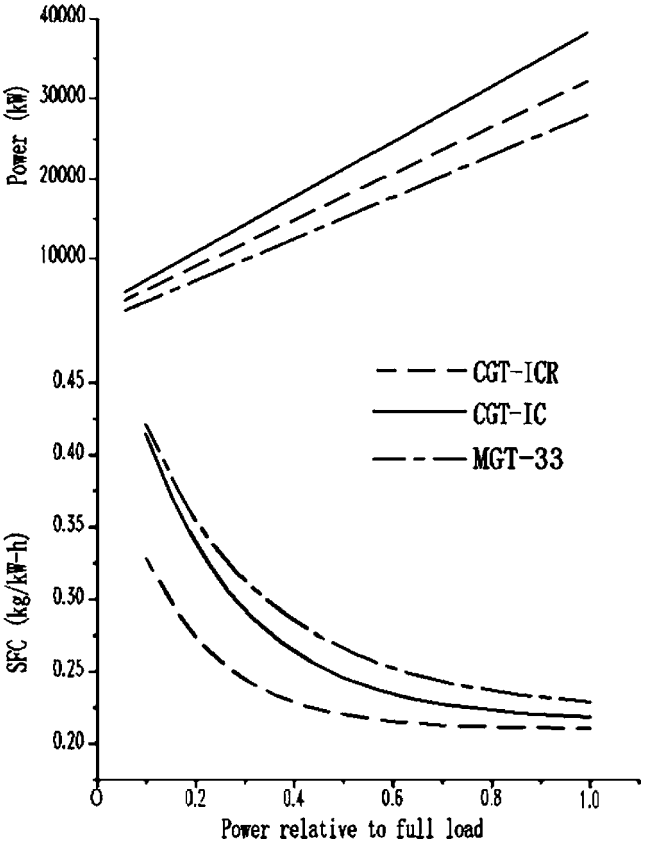


Fig. 8 Power output and SFC performance

The efficiency of the power turbine designed anew has increased by 1%, leading finally to a power output increase amounting to 34%. As the efficiency of the power turbine is by far higher than that of the high- and low-pressure turbine, a considerable portion of energy has been transferred to a high-effective zone for expansion and doing work. This has played a helpful role in improving the efficiency of the engine as a whole. Hence, so far as an IC cycle is concerned, it is especially important to design a high-efficiency power turbine.

(5) Critical speed of the low-pressure rotor. Due to the necessity of inserting an intercooler between the high-pressure and low-pressure compressor of the prototype engine, it is inevitable to cause an increase of the axial dimensions of the low-pressure rotor, leading to a change in the critical speed of the low-pressure rotor. One of the reasons for the removal of the last two stages from the low-pressure compressor consists of decreasing the influence of the increase in axial dimensions of the low-pressure rotor caused by the introduction of an intercooler, and also in controlling the engine total length.

(6) Combustor and fuel oil system. The use of intercooling will cause a significant decrease of the temperature of air entering a combustor, but the turbine inlet temperature will not change. Hence, fuel consumption will increase (relative to a simple cycle, the increase being assessed at 28.3%). As for fuel oil pumps, fuel oil spray nozzles, and combustor and fuel oil regulation system there arises a need for performing pertinent adjustments or a renewed design.

Performance Comparison

For CGT-IC gas turbine power output and oil consumption performance curves, see Fig. 8. For comparison purposes, also given are the prototype engine performance and calculated performance for developing MGT-33 to CGT-ICR (used only for conducting comparisons).

The IC cycle has the highest power output. The use of simple intercooling technology can avoid the piping pressure losses from high-pressure compressor outlet to combustor by way of a recuperator, and also the piping pressure losses from power turbine outlet to the atmosphere by way of a recuperator. These losses will arise due to the presence of the recuperator. Because of the above circumstances, the IC cycle occupies a more advantaged position in enhancing gas turbine power output as compared with ICR cycle. The reduction of pressure losses can, to a certain extent, compensate for the loss in cycle efficiency due to the absence of recuperation.

The ICR cycle has the highest efficiency. The use of recuperation and inlet guide vane variable geometry technology for the power turbine makes it possible for the ICR to have a relatively high efficiency, not only at the design point, but also when operating at partial loads.

Conclusions

(1) As viewed from the analysis of ordinary cycles, the use of an intercooling cycle can significantly increase the specific work of gas turbines, but with respect to cycle efficiency, the adoption of intercooling may usually bring about a slightly negative influence.

(2) An analysis of the IC cycle modification scheme implemented on a specific engine indicates that it is possible to simultaneously enhance its power output and efficiency. This is because there exist the following factors available for taking advantage of:

- The physical speed of the high pressure rotor decreases significantly. It is possible to properly adjust its corrected speed. Its centrifugal stress also markedly decreases.
- Owing to the significant decrease of the air temperature at the compressor outlet, under the precondition of ensuring the blade metal surface temperature of the high temperature turbine not higher than the stipulated value, there is still room for realizing an increase in the turbine inlet temperature.
- Due to the decrease of power consumption in the gas generator, a greater portion of energy transfer is completed in the power turbine portion, which has a higher efficiency.
- In the renewed matching process of the gas turbine, the parameters of some components, such as efficiency, pressure ratio, expansion ratio and temperature, etc. will change in a more favorable direction.

The results of an analysis indicate that the power output has increased from 28,500 kW to 38,200 kW (ΔNe=34%) and the thermal efficiency from 37% to 38.5% ($\Delta\eta_e$=4.1%). The effectiveness of engine overall performance enhancement is significant.

(3) The main merits of the CGT-IC are:

(1) The changes made to the core component structure of the prototype engine gas generator have been kept to a minimum with the flow path characterized by its maximum versatility. As a result, the high reliability of the prototype engine can be inherited in its entirety.

(2) Technical difficult points concerning the reliability of such components as recuperator and power turbine adjustable guide vanes have been sidestepped, resulting in a relatively simple system. All these factors can contribute to shortening development cycle, decreasing development expenses, and also enhancing the overall performance of the engine.

(3) The performance of the CGT-IC occupies an intermedi-

ate position between ICR and a simple cycle. For a developing country that lacks a fully developed aero industry, the scheme described above may be regarded as a reliable and realistic compromise.

Nomenclature

- Ga = flow rate
- LHV = fuel lower heating value
- Ne = power output
- P = pressure
- PR_{LPC} = pressure ratio of low-pressure compressor
- PR_{HPC} = pressure ratio of high-pressure compressor
- PR = total pressure ratio
- T = temperature
- TIT = turbine inlet temperature
- SFC = specific fuel consumption
- SP = specific power output
- η = polytropic efficiency
- ξ = pressure loss coefficient, $\xi = \Delta P / P$
- ε = intercooler efficiency, $\varepsilon = (T_2 - T_{22}) / (T_2 - T_{cool})$
- η_e = cycle efficiency
- ΔP = pressure loss

Subscripts

- in = inlet
- LPC = low pressure compressor
- IC = intercooler
- HPC = high-pressure compressor
- CC = combustion chamber
- HPT = high-pressure turbine
- LPT = low-pressure turbine
- PT = power turbine
- ID = intermediate diffuser between low-pressure turbine and power turbine
- ex = exhaust
- cool = coolant
- 0,1,2,3… = calculation points, see Fig. 2

References

[1] "Gas Turbine World 2005 Performance Specs," Vol. 34 No. 6, 2005.

[2] Anthony J. Crisalli, "Overview of WR-21 Intercooled Recuperated Gas Turbine Engine System, A Modern Engine for A Modern Fleet," 93-GT-231.

Robust Strategies for Forced Response Reduction of Bladed Disks Based on Large Mistuning Concept

M. Nikolic
e-mail: marija.nikolic@imperial.ac.uk

E. P. Petrov
e-mail: y.petrov@imperial.ac.uk

D. J. Ewins
e-mail: d.ewins@imperial.ac.uk

Mechanical Engineering Department,
Centre of Vibration Engineering,
Imperial College London,
South Kensington Campus,
London, SW7 2AZ, UK

In this paper, robust maximum forced response reduction strategies based on a "large mistuning" concept are introduced, including both (i) random and (ii) deterministic approaches. An industrial bladed fan disk serves as an application example for a reliability assessment of the aforementioned strategies using two well-established tools for uncertainty analysis: (i) statistics and (ii) sensitivity and robustness. The feasibility and other practical aspects of implementing large mistuning as a means of preventing excessive forced response levels caused by random mistuning and ensuring the predictability of the response are discussed. [DOI: 10.1115/1.2799524]

Introduction

Since the major problem surrounding mistuned bladed disks is related to uncertainty in the blades' forced response amplitudes, which might lead to excessive vibration levels in one or more blades, one of the key issues that current industry practitioners face remains the development of efficient ways of mitigating the hazardous forced response levels and preventing the worst-case scenarios. Various strategies have been generated for the past 40 years to address this problem; some of which seek to establish the most favorable mistuning patterns from the forced response perspective. They involve the introduction of deliberate, systematically controlled variations in blade properties to obtain the desired response characteristics, commonly referred to as "intentional mistuning." Different types of intentional mistuning patterns have been adopted in the past, such as (i) "alternate" mistuning, by alternating high and low frequency blades [1]; (ii) periodic mistuning [2,3], (iii) harmonic mistuning [4–6], and, most recently, (iv) linear mistuning [7,8]. The advantage of implementing intentional mistuning lies in the ability to predict and control the vibration levels at the design stage, although, as blade properties change with time, a chosen mistuning pattern strategy might be appropriate for a short time scale only. For this reason, most of the above-mentioned studies identified the robustness of a selected pattern as one of the crucial factors to be considered when developing an intentional mistuning strategy, and this resulted in the assessment of sensitivity of the selected intentional mistuning to the additional "unintentional" random mistuning. From the papers surveyed, it can be concluded that in the past, researchers have exploited the effects of intentional mistuning for relatively small blade frequency mistuning ranges, typically of the order of $\pm 0.5\%$ to $\pm 5\%$. Nevertheless, studies have shown that there is a threshold of frequency mistuning beyond which the maximum forced response levels off, or even drops, as the degree of mistuning is increased further, as depicted in Fig. 1.

From Fig. 1, it is evident that there are two possible intentional mistuning strategies for reducing the forced response amplification factor. The first one is to introduce a very small amount of mistuning and, subsequently, endeavor to maintain a nearly tuned system, which is an expensive procedure in terms of manufacturing costs and may not be successful due to blade property changes with time, so that the initial system may depart significantly from the nearly tuned state. On the other hand, the tendency of the forced response amplification factor to decrease with blade mistuning, after the initial rise, suggests that there might be a possibility of exploiting the effects of larger frequency mistuning ranges, perhaps of the order of $\pm 15/\pm 20\%$ or higher, as a means of reducing the problems arising from uncertainty in the forced response levels of bladed disks. This novel intentional mistuning idea will be further referred to as "large mistuning" (LM), implicating large frequency mistuning scatter.

The aims of the study reported in this paper are (i) to assess effectiveness and reliability of a LM technique for minimizing consequences of blade mistuning on forced response of bladed disks, and (ii) to establish whether current trends to relax manufacturing tolerances that might have apparent financial incentives and advantages could be used to control the vibration levels. This work will attempt to obtain the answers to the following questions.

- Could a substantial increase in frequency mistuning range contribute to a reduction of maximum forced response levels?
- How could the maximum forced response be controlled efficiently?

This paper consists of two parts: (i) an assessment of "random" LM strategy and (ii) utilization of the LM concept in a regulated "deterministic" manner, involving commonly accepted intentional mistuning patterns of unconventionally large strength of mistuning. Both parts include the forced response results of an industrial 26-bladed disk obtained with a high-fidelity in-house prediction tool. In the first part, statistical characterization of forced response is obtained numerically via extensive Monte Carlo (MC) simulations of mistuning patterns drawn randomly from a uniform distribution for different blade frequency mistuning, spanning from small to large ranges. Subsequently, goodness-of-fit tests based on the Kolmogorov–Smirnov and the chi-square (Pearson's) criteria are performed to identify the theoretical distribution functions that represent the best match to the empirical distributions of forced responses. In the second part of the paper, the effectiveness of several deterministic mistuning patterns based on LM is examined and their sensitivity and robustness to small changes in blade properties are tested in an attempt to determine the most beneficial

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 26, 2007; final manuscript received June 8, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007.

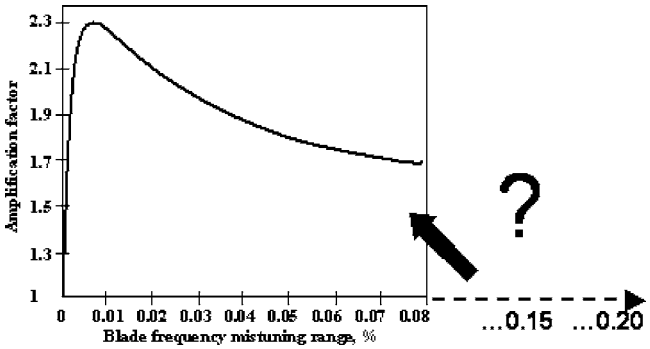


Fig. 1 Forced response amplification factor as a function of blade frequency mistuning range

and successful LM intentional patterns. Finally, the random and deterministic LM intentional mistuning strategies are compared in terms of their practicality and overall success.

Random Large Mistuning Strategy for Forced Response Reduction

Model Description. A finite element model of a realistic bladed fan disk, depicted in Fig. 2, which contains 26 blades and consists of over 120,000 degrees of freedom (DOFs) per sector, contributing to more than 3,000,000 DOFs of a full structure, has been used in the analysis.

The model consists of flexible blades, which are weakly coupled through a relatively stiff disk. Free tuned vibration characteristics shown in Fig. 3 reveal that the bladed disk assembly has modes of vibration with rather close natural frequencies for most of the mode families.

Mistuning Type. Mistuning expressed in terms of deviations of blade frequencies from nominal values, which are the consequence of mass and stiffness variations, was chosen for the study. Random blade frequency values were generated from a uniform distribution over a predetermined frequency mistuning range, as depicted in Fig. 4.

The specific frequency mistuning ranges generated from the MC simulations were (i) $\pm 20\%$, 30%, and 40% of nominal blade frequency, representing LM, (ii) $\pm 5\%$, 10%, and 15%, and (iii) $\pm 0.5\%$, 1%, 1.5%, and 2%, accounting for "moderate" and "small" frequency mistuning ranges, respectively. The latter are included in this study for the sake of comparison with the LM results.

Forced Response Considerations. Forced response calculations for different frequency mistuning ranges were performed using the MISTRES in-house prediction code based on a new method for dynamic analysis of mistuned bladed disks developed in Ref. [9]. The method is based on an exact relationship between tuned and mistuned systems, which allows use of large finite element models, since only one sector is needed to represent the tuned and mistuned systems, while the computational cost is independent of the size of the original single blade segment. The program permits computation of the forced response at any selected so-called "active" DOFs. In the current analysis, the amplitudes of forced response were obtained for all bladed disk sectors at four chosen nodes where the maximum displacements were anticipated. These nodes are shown as black circles in Fig. 5, while the gray circles indicate the "passive" nodes where the uniformly distributed loads are applied.

A conventional engine order excitation by 3, 6, and 13 engine orders (EOs) was considered in the analysis over an excitation frequency range corresponding to the predominantly first flapwise vibration mode (1F), as shown in Fig. 3(*b*) by a rectangular area constrained between the two dashed horizontal lines. Forced response amplitudes of the mistuned system were normalized with respect to the maximum amplitudes of the equivalent tuned bladed disk under the same excitation conditions. The damping loss factor was set to 0.003.

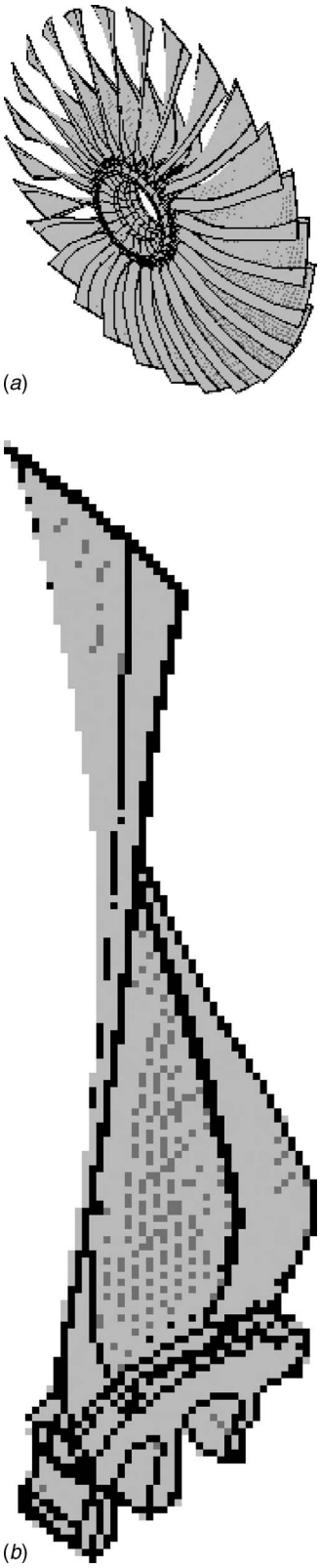


Fig. 2 Bladed fan disk: (*a*) full model and (*b*) its cyclic sector

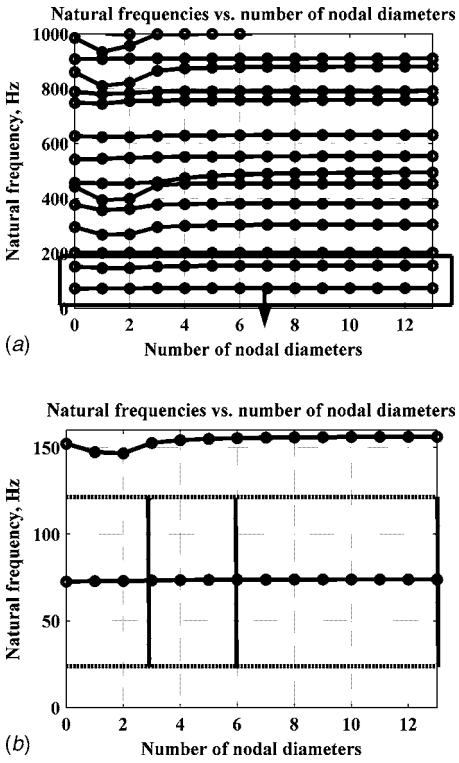


Fig. 3 Natural frequencies of a tuned bladed fan disk, the excitation frequency range (rectangular area between two dashed horizontal lines), and analyzed excitation EOs (vertical lines)

Dependency of Forced Response on Frequency Mistuning Range. The MC simulations for 1000 randomly generated mistuning configurations from the uniform distribution yielded normalized forced response amplitudes for selected ±40%, 15%, and 0.5% frequency mistuning ranges under 3, 6, and 13 EO excitations, as shown in Fig. 6.

It is evident that the worst-case forced response (i) decreases as the frequency mistuning range is increased, and (ii) varies depending on the EO of excitation. Furthermore, Fig. 7 demonstrates that

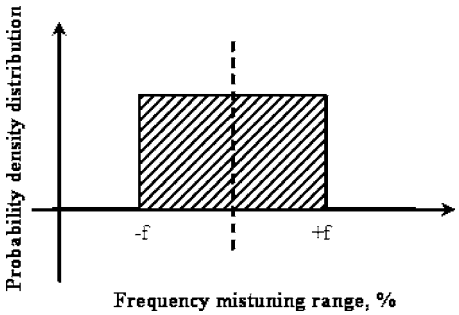


Fig. 4 Uniform PDF of frequency mistuning range

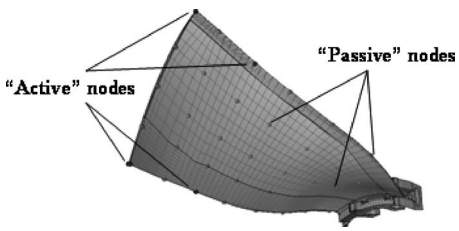


Fig. 5 "Active" and "passive" nodes

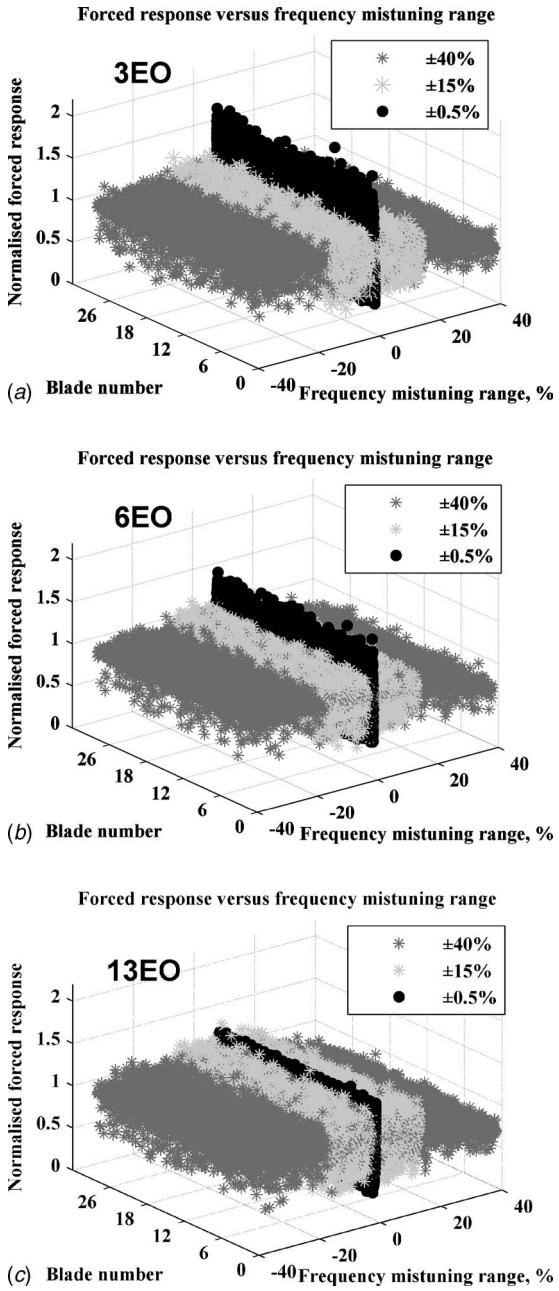
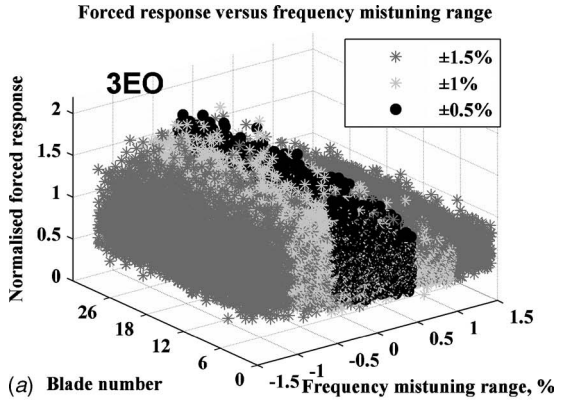


Fig. 6 Forced response results for ±40%, 15%, and 0.5% frequency mistuning ranges obtained for (*a*) 3 EO, (*b*) 6 EO and (*c*) 13 EO

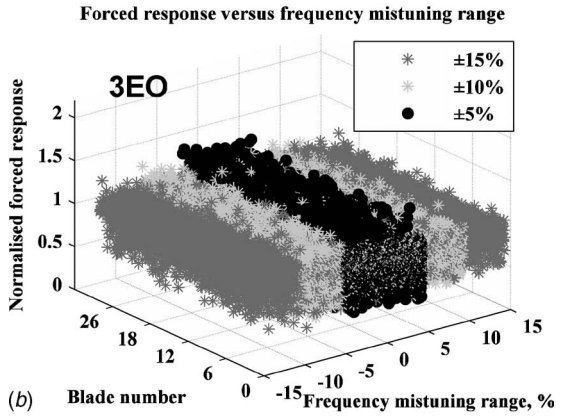
(iii) the spread of forced response decreases as the frequency mistuning range is increased and that (iv) there is no clear correspondence between blade natural frequency characteristics and its forced response level.

The maximum, mean, and minimum forced response amplitudes for each of the analyzed random mistuning patterns are shown in Fig. 8 for the cases of ±0.5% and 40% frequency mistuning ranges. The spread of maximum forced response reduces with the frequency range, while that of minimum forced response increases under the same conditions.

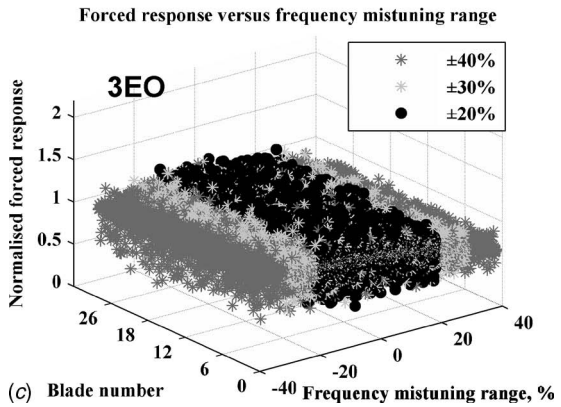
Results obtained from this statistical study are finally summarized in Fig. 9 and Table 1, which depict the forced response dependency upon frequency mistuning range for maximum, mean, and minimum values under 3, 6, and 13 EO excitations.



(a)



(b)

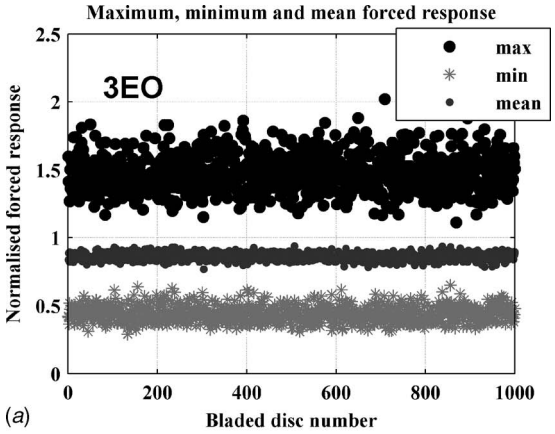


(c)

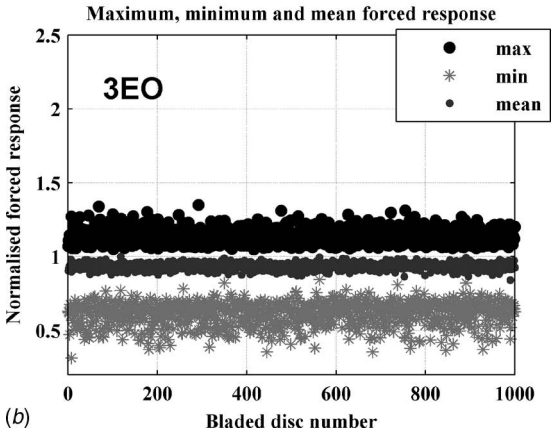
Fig. 7 Forced response results obtained under 3 EO for (*a*) ±1.5%, 1%, and 0.5%, (*b*) ±15%, 10%, and 5%, and (*c*) ±40%, 30%, and 20% frequency mistuning ranges

The results obtained indicate a reduction in the maximum forced response of approximately 33% from 2.02 at ±0.5% frequency mistuning range to 1.35 at ±40% frequency mistuning range for 3 EO excitation. The maximum forced response decreases for 6 EO and 13 EO are 22% and 16%, respectively.

It is evident that although a precise amount of maximum forced response reduction is dependent on the EO excitation considered, there is nevertheless a clear trend of a consistent response reduction with the increase of frequency mistuning range after the initial rise. The observed tendency of the maximum forced response to decrease could be explained by considering plots in Fig. 10, which demonstrate the frequency response functions (FRFs) calculated for all blades for representatives of very small (±0.1%), small (±2%), and large (±20%) frequency mistuning ranges.



(a)



(b)

Fig. 8 Maximum, minimum, and mean forced response results for (*a*) ±0.5% and (*b*) ±40% frequency mistuning ranges obtained for 3 EO

It can be seen that there is very little effect of interactions between the blades for very small (Fig. 10(*a*)) and large (Fig. 10(*c*)) frequency mistuning ranges, and subsequently, low energy transfer between the blades, whereas the opposite is true for ±2%, small frequency mistuning case (Fig. 10(*b*)). The amount of interaction between the blades and the energy transfer between them could be associated with the maximum forced response amplification levels, which are high in the ±2% frequency mistuning case, as this leads to a localized response of one blade that is

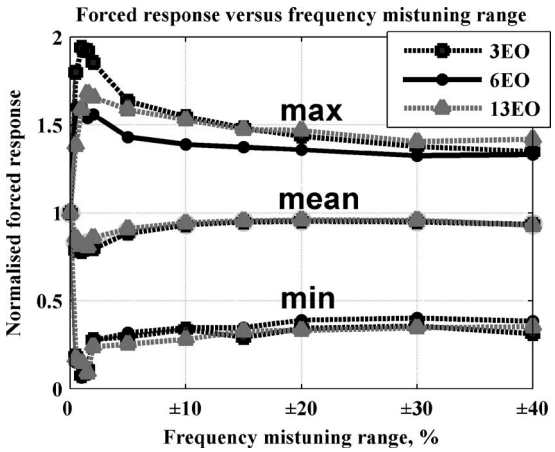


Fig. 9 Maximum, minimum, and mean forced response results for 3, 6, and 13 EO

Table 1 Maximum, minimum, and mean forced response results for 3, 6, and 13 EO

Frequency mistuning range	3 EO			6 EO			13 EO		
	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
±0.5%	2.02	0.29	0.86	1.67	0.31	0.85	1.44	0.30	0.88
±1%	1.99	0.25	0.76	1.70	0.29	0.81	1.62	0.24	0.85
±1.5%	1.94	0.28	0.77	1.61	0.29	0.81	1.68	0.24	0.85
±2%	1.92	0.28	0.80	1.56	0.27	0.83	1.66	0.24	0.86
±5%	1.64	0.29	0.88	1.43	0.32	0.89	1.59	0.25	0.91
±10%	1.55	0.34	0.93	1.39	0.35	0.94	1.53	0.28	0.94
±15%	1.48	0.29	0.95	1.37	0.35	0.95	1.47	0.33	0.95
±20%	1.44	0.34	0.95	1.36	0.39	0.96	1.47	0.33	0.96
±30%	1.38	0.36	0.95	1.33	0.40	0.95	1.42	0.35	0.96
±40%	1.35	0.31	0.93	1.32	0.38	0.94	1.41	0.35	0.93
% reduction (from ±0.5% to ±40%)	33.25%			22.12% (from ±1% to ±40%)			16.30% (from ±1.5% to ±40%)		

contributed by the accumulated responses of all the other blades. The maximum forced response levels are substantially lower in the nearly tuned and heavily mistuned systems as low energy transfer between the blades does not contribute substantially to formation of the localized response.

The evidence presented in this section suggests that random LM can be a promising approach that, if practically feasible, could be employed as a potentially beneficial forced response diminution strategy.

Practical Issues of Large Mistuning Implementation. LM levels for few blades in the assembly are not uncommon in bladed disk practice. Examples include (i) single-crystal turbine blades (known for their unmatched resistance to high-temperature creep and fatigue) and (ii) damaged blades, which can reach ±35% deviation in the first natural frequency compared with the corresponding value of the tuned blade [10]. Practically, if the implementation of LM blade types in the bladed disk assembly does not introduce any adverse effects, then the advantage is seen in extending design acceptance limits as a way to reduce the manufacturing costs. Moreover, a study of forced response characteristics of few blades with uncommonly large natural frequency discrepancies provides a means of assessment of the damage tolerance of the assembly, which may be of importance in the engine industry.

Statistical Characterization of Forced Response. The aim of the statistical analysis presented in this paper is to characterize the probability distributions of forced response levels obtained in the previous section for different blade frequency mistuning ranges with an aim to ultimately gain some insight into the bladed disk worst-case response likelihood due to impact of the proposed LM strategy.

Outline of the Statistical Methods. Initially, a histogram-based method is applied in order to yield an estimate of the empirical distribution function of the normalized forced response population: namely, Sturges' rule, which estimates the number of bins in the histogram k as a function of the number of observed data n:

$$k = 1 + 3.3 \log_{10} n \tag{1}$$

Descriptive measures, or the "first four moments" of the distribution, are then determined, including the expected value or mean, variance, skewness, and kurtosis, respectively:

$$\tilde{m}_x = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

$$\tilde{\sigma}_x^2 = \tilde{D}_x = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \tilde{m}_x)^2 \tag{3}$$

$$\tilde{a}_x = \frac{1}{(n-1)\tilde{\sigma}_x^3}\sum_{i=1}^{n} (x_i - \tilde{m}_x)^3 \tag{4}$$

$$\tilde{e}_x = \frac{1}{(n-1)\tilde{\sigma}_x^4}\sum_{i=1}^{n} (x_i - \tilde{m}_x)^4 - 3 \tag{5}$$

where x_i, $i=1,2,\ldots,n$, are the values for the n data points, and $\tilde{\sigma}_x = \sqrt{\tilde{D}_x}$.

Subsequently, the confidence intervals for the general parameters are calculated using the obtained empirical distribution parameters. The confidence interval for the precision of the estimate of the mean value is calculated as

$$\tilde{m}_x - \frac{\tilde{\sigma}_x t_{1-(p/2)}(f)}{\sqrt{n}} \leq m_x \leq \tilde{m}_x + \frac{\tilde{\sigma}_x t_{1-(p/2)}(f)}{\sqrt{n}} \tag{6}$$

where $\tilde{m}_x$ and $\tilde{\sigma}_x$ are the estimated mean and standard deviation, and $t_p(f)$ is Student's t-distribution quantile, obtained from tables, corresponding to p—the statistical significance level—from which the confidence interval is given as $(1-p)$.

The corresponding confidence interval for the variance is found from

$$\frac{f\tilde{D}_x}{\chi^2_{1-(p/2)}(f)} \leq D_x \leq \frac{f\tilde{D}_x}{\chi^2_{(p/2)}(f)} \tag{7}$$

where $\tilde{D}_x$ is the estimated variance, $\chi_p^2(f)$ is Pearson's χ^2-distribution quantile corresponding to p (obtained from tables), and f is the number of DOFs.

Similarly, the confidence intervals for the skewness and kurtosis are obtained using the Chebyshev inequality as

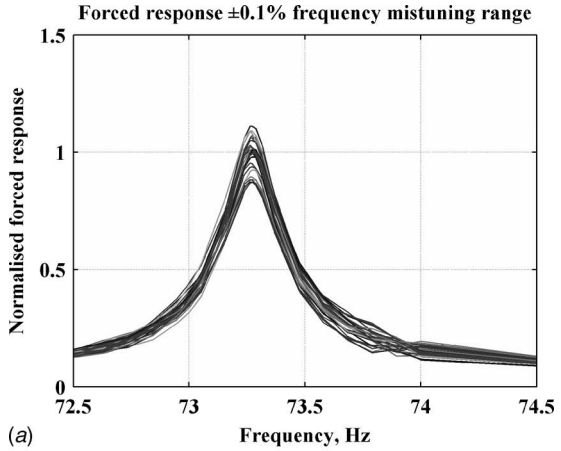
$$\tilde{a}_x - \sqrt{\frac{D_a}{p}} \leq a_x \leq \tilde{a}_x + \sqrt{\frac{D_a}{p}} \tag{8}$$

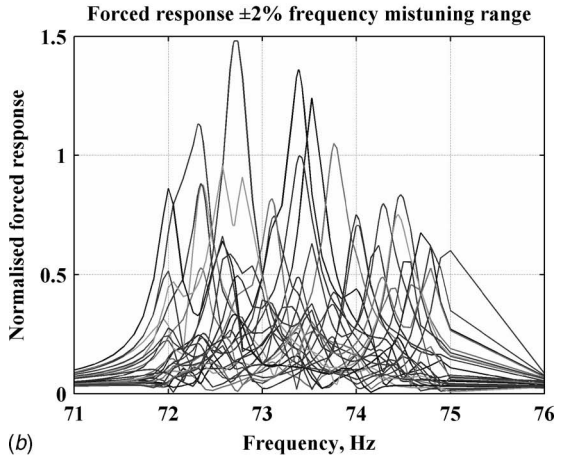
$$\tilde{e}_x - \sqrt{\frac{D_e}{p}} \leq e_x \leq \tilde{e}_x + \sqrt{\frac{D_e}{p}} \tag{9}$$

where D_a and D_e are the variances of the estimated skewness and kurtosis:

$$D_a = \frac{6(n-1)}{(n+1)(n+3)} \tag{10}$$

$$D_e = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)} \tag{11}$$





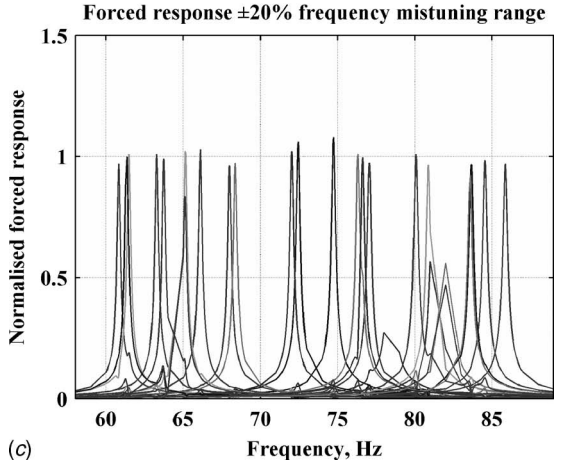


Fig. 10 FRFs for all blades under 3 EO excitation for (*a*) ±0.1%, (*b*) ±2%, and (*c*) ±20% frequency mistuning ranges

The next phase of the statistical characterization of forced response involved (i) the choice of the theoretical probability density function (PDF) and cumulative density function (CDF) to describe the empirical data, and (ii) the estimation of their parameters: the location, scale, and shape of the distribution. For predictive and controlling purposes, it is often desirable to calculate accurately the attributes of an underlying distribution of population. To determine this distribution, it is common to fit hypothetical theoretical distributions to an empirical distribution and, subsequently, to assess the degree of correlation between the two. Thus, theoretical distributions, including Gaussian, uniform, Rayleigh, beta, gamma, Weibull, and extreme value distributions, are fitted to the empirical distribution, where their parameters are estimated using the principle of maximum likelihood.

By performing two of the standard hypothesis tests, namely, the chi-square (Pearson's) and the Kolmogorov–Smirnov, the "degree of fit" of the theoretical distributions with the empirical is assessed at a given significance level. The idea of the chi-square test is to divide the range of distribution into bins and to compare the observed number in each bin to the number that would be expected if the assumed distribution were true. The hypothesis that the data are from a population with the specified distribution is rejected if the following relationship for the chi-square test holds:

$$\sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j} \leqslant \chi_\alpha^2(k-c) \tag{12}$$

where n_j is the empirical observed number in each bin, np_j is the theoretical observed number in each bin, $p_j = F_t(b_j) - F_t(a_j)$, $F_t(x)$ is the theoretical CDF, a_j and b_j are the lower and upper bin limits, $\chi_\alpha^2(k-c)$ is the chi-square test statistic at a significance level α (obtained from tables), and j is the jth bin.

The Kolmogorov–Smirnov test is based on the largest difference in absolute value between the empirical distribution function and its hypothesized counterpart:

$$D = \max_x |F_x(x) - \tilde{F}_x(x)| \tag{13}$$

where $F_x(x)$ is the theoretical cumulative distribution, and $\tilde{F}_x(x)$ is the empirical distribution.

Using D, the value of $\lambda = D\sqrt{n}$ is evaluated, which is subsequently compared to the Kolmogorov–Smirnov test statistic, λ_α, obtained from the tables: If $\lambda \leqslant \lambda_\alpha$, then the null hypothesis can be accepted at a given significance level; if $\lambda > \lambda_\alpha$, the theoretical distribution hypothesized is rejected.

Statistical Analysis of Forced Response Results Obtained for All Blades for Over 1000 Bladed Disks. The statistical results for forced response for all 26 blades over 1000 bladed disks were obtained for all frequency mistuning ranges under three ED excitations. Figures 11–13 demonstrate selected results for ±0.5%, 15%, and 40% frequency mistuning ranges, representing small, moderate, and large ranges, respectively, which include (a) the calculated empirical distribution functions for a sample size of 26,000 and (b) the best-fit theoretical CDF and the greatest vertical distance between the two CDFs used in the Kolmogorov–Smirnov hypothesis test.

Figure 11 illustrates the statistical results for forced response obtained for ±0.5% frequency mistuning range. It can be seen that the forced response empirical distribution characteristics are dependent on the ED of the excitation, as also shown in Table 2. While the first two moments of the distributions are reasonably close in all three cases, the third and the fourth moments, the skewness and the kurtosis, describing the asymmetry and the peakedness properties, are dissimilar, accounting for the observable differences in the histogram shapes. Subsequently, the theoretical closest fit distribution functions are different: The gamma CDF is found to be the most suitable in representing the empirical data in the cases of 3 EO and 6 EO, while the Weibull CDF is the best-fit theoretical approximation to the MC simulation (empirical) results for 13 EO. It is noted here that the theoretical distribution parameters were obtained using the principle of maximum likelihood with 95% confidence intervals and were tested according to the Kolmogorov–Smirnov and the chi-square hypothesis tests under a 0.01 (or 1%) significance level.

The empirical distribution descriptive measures of the forced response obtained for ±15% frequency mistuning range are presented in Table 3 and Fig. 12.

As for the ±0.5% range, the skewness and the kurtosis of the

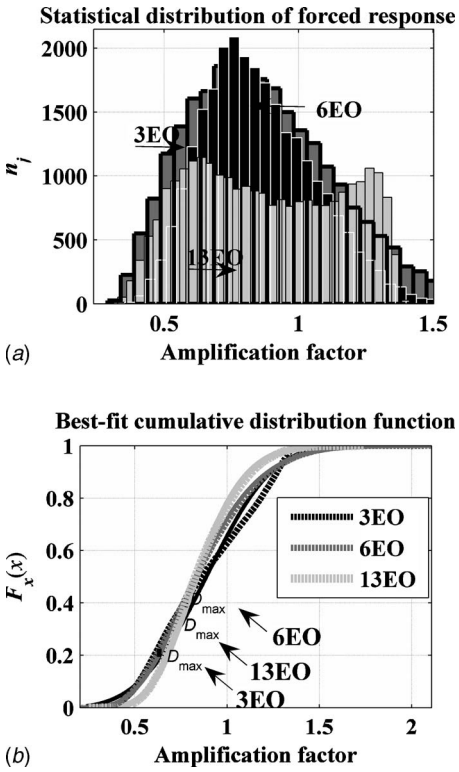


Fig. 11 Forced response statistical results for ±0.5% frequency mistuning range. Different solid-line colors denote the hypothetical theoretical distribution functions.

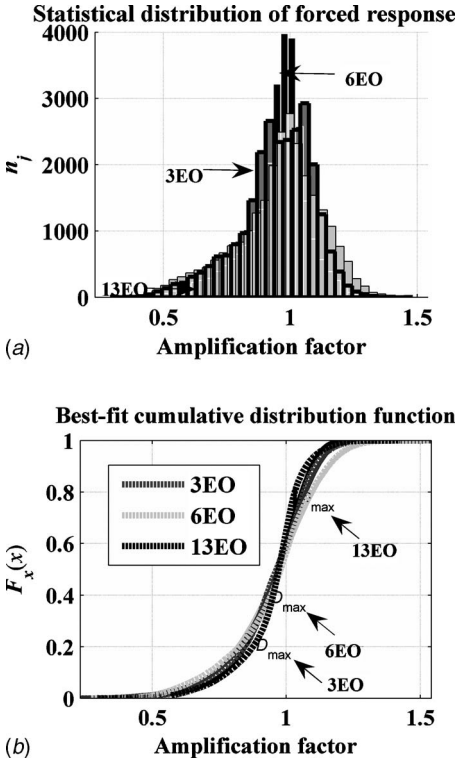


Fig. 12 Forced response statistical results for ±15% frequency mistuning range

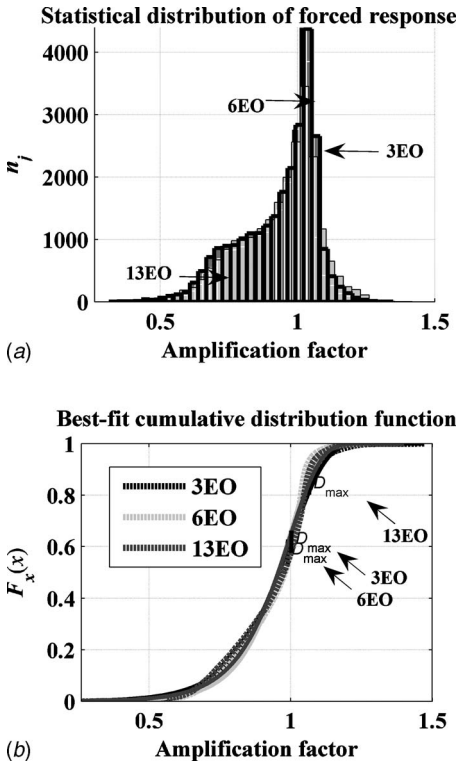


Fig. 13 Forced response statistical results for ±40% frequency mistuning range

distributions vary for different EOs for the ±15% frequency mistuning range. The best-fit CDF for 3 EO and 6 EO is the extreme value distribution, whereas that corresponding to 6 EO is the Weibull distribution, as shown in Fig. 12(*b*).

Table 4 summarizes the moments of the obtained distributions for ±40% frequency mistuning range. It can be seen that the third and fourth moments are distinct, but to a lesser degree than for the ±0.5% and ±15% frequency mistuning ranges. As a consequence, the CDF that presents the best fit to all empirical distributions is the extreme value distribution, as shown in Fig. 13(*b*) and Table 5, which demonstrates selected statistical hypothesis testing results. The significance level for the Kolmogorov–Smirnov and the chi-square tests adopted was 0.01 (or 1%), which means that there is a 99% chance that the selected theoretical distribution is a genuine probability distribution of the response data, but also that there is a 1% chance that the difference between the theoretical and the empirical distributions is true.

The results from Table 5 suggest the closest theoretical CDF fits to the empirical distributions. However, for all of the frequency mistuning range results considered, the indicated hypotheses were not accepted at the prescribed significance level. This could signify that none of the tested theoretical distributions model the empirical data accurately enough. On the other hand, the statistical studies indicate that there is an effect of the sample size on the result of a goodness-of-fit test: "If one employs a large enough sample size, almost any goodness-of-fit test will result in rejection of the null hypothesis" [11]. In order to conclude on the basis of a goodness-of-fit test that data conform to a specific distribution, the data should be reasonably close to the specifications of the distribution. Thus, in some cases, when a large sample size is involved, such as 26,000 forced responses considered here, the null hypothesis may be rejected. Yet in spite of the latter, if the sample data are reasonably close to the hypothesized distribution, one can operate on the assumption that the sample data provide an adequate fit for the hypothesized distribution. Table 6 demonstrates an ef-

Table 2 Forced response empirical data characteristics for ±0.5% frequency mistuning range

	x_{min}	x_{max}	$\tilde{m}_x$	$\tilde{\sigma}_x$	$\tilde{D}_x$	$\tilde{a}_x$	$\tilde{e}_x$
3 EO	0.29	2.02	0.86	0.27	0.07	0.50	−0.15
6 EO	0.32	1.67	0.85	0.21	0.04	0.47	−0.14
13 EO	0.30	1.45	0.88	0.28	0.08	0.07	−1.22

fect of sample size on statistical hypothesis test results, from which one can see that reduction of the sample size from 26,000 to 1000 results in the acceptance (A) of the hypothesized theoretical distribution.

From the mistuning literature, it was found that the forced response data for all blades have been represented in the past using the Weibull or the extreme value distributions (Refs. [4,12,13,8]), which is in agreement with the results obtained in this study for the frequency mistuning ranges in the excess of ±5%, and demonstrates that using such a distribution could provide savings in the computational cost of the MC simulations. The study in Ref. [14] showed that there was no unique probability distribution that could describe the forced response under every possible EO excitation, which is also supported by evidence presented here for cases of small frequency mistuning ranges.

Several conclusions can be drawn from the statistical analysis of the forced response for all blades:

(i) The forced response empirical distribution characteristics are dependent on the EO excitation for all frequency mistuning ranges: In particular, the third and the fourth moments of the distributions, the skewness and the kurtosis, describing the asymmetry and the peakedness properties, are dissimilar, accounting for the observable differences in the histogram shapes. As the frequency mistuning range increases, the PDFs of the forced response become narrower and shift toward the lower values, which indicates benefits of the LM.

(ii) The hypothesis tests suggest that the best-fit theoretical CDFs are the Weibull and the extreme value distributions for the frequency mistuning ranges above ±5%, whereas the closest statistical models to describe the empirical data obtained for small frequency mistuning ranges are the gamma or the Gaussian CDF, depending on the EO excitation.

Table 3 Forced response empirical data characteristics for ±15% frequency mistuning range

	x_{min}	x_{max}	$\tilde{m}_x$	$\tilde{\sigma}_x$	$\tilde{D}_x$	$\tilde{a}_x$	$\tilde{e}_x$
3 EO	0.29	1.48	0.95	0.15	0.02	−0.73	0.64
6 EO	0.35	1.37	0.95	0.12	0.02	−0.99	1.73
13 EO	0.33	1.47	0.96	0.17	0.03	−0.60	0.30

Table 4 Forced response empirical data characteristics for ±40% frequency mistuning range

	x_{min}	x_{max}	$\tilde{m}_x$	$\tilde{\sigma}_x$	$\tilde{D}_x$	$\tilde{a}_x$	$\tilde{e}_x$
3 EO	0.31	1.35	0.93	0.14	0.02	−0.81	0.08
6 EO	0.38	1.33	0.94	0.13	0.02	−0.94	0.45
13 EO	0.35	1.42	0.94	0.14	0.02	−0.68	0.26

Table 5 Statistical hypothesis test results for forced response characteristics obtained for all blades over 1000 bladed disks.

Frequency mistuning range and EO	Goodness-of-fit test analysis 26,000 samples				
	Best-fit distribution	Test statistic	Cutoff	Conclusion	Test
±0.5%, 3 EO	Gamma	2.62	1.63	NA[a]	K-S
	"	399.00	53.40	NA	Chi
±0.5%, 6 EO	Gamma	3.12	1.63	NA	K-S
	"	301.30	42.60	NA	Chi
±0.5%, 13 EO	Weibull	10.90	1.63	NA	K-S
	"	6060.00	53.50	NA	Chi
±15.0%, 3 EO	Extreme	4.37	1.63	NA	K-S
	"	1580.00	46.90	NA	Chi
±15.0%, 6 EO	Extreme	12.50	1.63	NA	K-S
	"	5370.00	48.30	NA	Chi
±15.0%, 13 EO	Weibull	6.97	1.63	NA	K-S
	"	1060.00	52.20	NA	Chi
±0.5%, 3 EO	Extreme	11.40	1.63	NA	K-S
	"	4650.00	49.60	NA	Chi
±40.0%, 6 EO	Extreme	13.80	1.63	NA	K-S
	"	5730.00	48.30	NA	Chi
±40.0%, 13 EO	Extreme	7.94	1.63	NA	K-S
	"	2990.00	46.90	NA	Chi

[a]"NA" means that the null hypothesis is not accepted under a prescribed 0.01% or 1% significance level.

Table 6 Effect of sample size on statistical hypothesis test results

Frequency mistuning range and EO	Goodness-of-fit test analysis, 0.01% or 1% significance level					
	Sample size	Best-fit distribution	Test statistic	Cutoff	Conclusion	Test
±0.5%, 3 EO	26,000	Gamma	2.62	1.63	NA	K-S
	”	”	399.00	53.40	NA	Chi
±0.5%,3 EO	1000	Gamma	1.06	1.63	A	K-S
	”	”	46.10	36.20	NA	Chi
±0.5%, 6 EO	26,000	Gamma	3.12	1.63	NA	K-S
		”	301.30	42.60	NA	Chi
±0.5%, 6 EO	1000	Gamma	0.65	1.63	A	K-S
	”	”	26.90	37.60	A	Chi

Deterministic Large Mistuning Strategy for Forced Response Reduction

Deterministic Mistuning Patterns Based on Large Mistuning. The first part of this paper introduced the concept of LM and determined its effectiveness by controlling the mistuning scatter as a means of alleviating the severity of the maximum forced response. In this part, several deterministic intentional mistuning patterns based on LM, including (i) alternate, (ii) harmonic, (iii) linear mistuning patterns, and (iv) those with “damaged” blades (i.e., few blades with LM frequencies), are exploited and their sensitivity and robustness to small unavoidable random mistuning due to manufacturing tolerances and wear are assessed. Selected forced response data obtained by using the above-mentioned deterministic mistuning patterns of different mistuning degrees are shown in Fig. 14 and Table 7.

The simplest form of mistuning from the practical viewpoint—alternate mistuning—was investigated initially by incorporating combinations of small and LM degrees, such as 0.5%/−0.5%, 20%/−20%, etc. Compared with the “reference” random mistuning cases reported in Table 1, it can be seen that introduction of regulated mistuning decreases appreciably the maximum amplification factors. Moreover, it can be observed that an increase in the magnitude of alternate mistuning decreases the maximum forced response levels by a small amount.

Analysis of the harmonically defined mistuning patterns showed that, depending on particular EO excitation, the increase in frequency mistuning magnitude may result in either an increase or a decrease of the maximum forced response. Capacity of “linear” mistuning patterns to reduce the maximum forced response below the tuned level was suggested in Ref. [8], and was investigated further in this paper. Selected results are illustrated in Fig. 14(*c*) and Table 7, from which one can observe an occasional small adverse influence of increasing the frequency mistuning range on maximum forced response levels for some EO excitations.

In order to determine the influence of a few blades with LM, or the damaged blades, the forced response characteristics have been calculated for a chosen “randomly” selected mistuned pattern from ±5% frequency mistuning range with one, two, or three damaged blades. Interestingly, the damaged blades, i.e., those with LM, do not themselves exhibit the largest forced response levels, as indicated in Fig. 14(*d*). Furthermore, it was found that increasing the mistuning degree of such blades reduces marginally the forced response.

Robustness Assessment. Deterministic analysis of the aforementioned systematically regulated mistuning patterns provided some signs of their effectiveness as a forced response control strategy. Since the maximum amplitude magnification represents a local maximum in the forced response, it should therefore be sufficiently robust, so that the sensitivity of the system to additional random mistuning about this point should be small.

The blade amplitudes increase with mistuning strength up to a certain critical level, usually exhibiting a peak forced response at low degrees of mistuning, beyond which a further increase in mistuning causes the forced response to drop and flatten off. This implies that the bladed disk system is highly sensitive to mistuning around the tuned condition and that increasing the mistuning

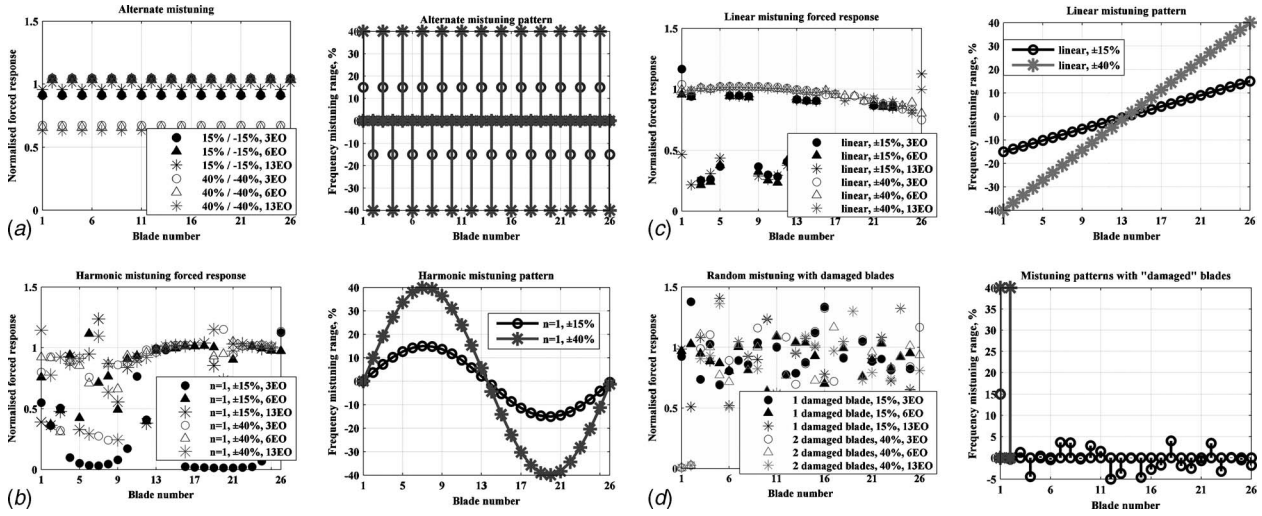


Fig. 14 Forced response results for deterministically defined (*a*) alternate, (*b*) harmonic, (*c*) linear mistuning patterns, and (*d*) those with few “damaged” blades

Table 7 Forced response results for deterministically defined mistuning patterns

Deterministic mistuning pattern	Normalized forced response		
	3 EO	6 EO	13 EO
Alternate 0.5%/−0.5%	1.204	1.042	0.179
Alternate 15%/−15%	1.045	1.037	1.011
Alternate 20%/−20%	1.046	1.042	1.024
Alternate 40%/−40%	1.042	1.044	1.038
Harmonic $n=1$, ±0.5%	1.020	1.230	1.140
Harmonic $n=1$, ±15%	1.130	1.120	1.240
Harmonic $n=1$, ±20%	1.170	1.050	1.210
Harmonic $n=1$, ±40%	1.150	1.050	1.150
Linear ±0.5%	1.250	1.110	1.210
Linear ±15%	1.170	0.960	1.130
Linear ±20%	1.130	0.990	0.980
Linear ±40%	1.040	1.020	1.010
One damaged blade, 15%	1.380	1.090	1.410
One damaged blade, 40%	1.363	1.164	1.405
Three damaged blades, 15%	1.330	1.150	1.320
Three damaged blades, 40%	1.326	1.165	1.311

significantly may result in a decrease in the sensitivity and robustness of forced response, which also justifies the choice of LM as a viable intentional mistuning strategy.

The robustness of the few selected intentionally mistuned patterns was investigated by examining the variation in the forced response as an additional unintentional random mistuning of ±0.5% is incorporated, which could occur in practice as a result of manufacturing imperfections and/or wear in operation. A further 100 MC simulations were performed in which small random variations in frequency mistuning were added onto the original intentionally mistuned system. The results for the main intentional mistuning patterns considered with (w) and without (w/o) random mistuning are given in Table 8.

In the case of alternate mistuning, the effect of increasing the mistuning degree of high frequency blades is highly dependent on the EO excitation, and may result in either an increase or a decrease of the maximum forced response levels. It can also be seen that the alternate mistuning strategy is not robust for most of the analyzed combinations. In particular, for small mistuning patterns, the average difference between the original and the perturbed maximum forced response levels ranges from approximately 36% to 61% depending on the excitation, whereas for the LM patterns, the corresponding difference oscillates between 46% and 92%. Based on the evidence presented for all mistuning ranges considered, it can be concluded that alternate mistuning may not be a robust strategy for maximum forced response control.

For harmonic mistuning patterns, the increase of the mistuning scatter range reduces significantly the maximum forced response levels. As the mistuning range is amplified, the robustness to additional mistuning variations is improved three to seven times depending on the excitation. Overall, the LM is beneficial for most of the calculated harmonic mistuning cases in mitigating the damaging effects of random mistuning.

A maximum forced response reduction of 60% from 1.73 to 1.08 and a tenfold robustness improvement are achieved for linear mistuning patterns by increasing the frequency mistuning range from ±0.5% to ±40% for 3 EO excitation. A corresponding drop in the maximum forced response was found for 6 and 13 EO cases of approximately 43% and 30%, respectively. It is also observed that the difference between the original and the perturbed maximum forced response levels is of the order of 2–8%, whereas a significantly poorer robustness is achieved for smaller frequency mistuning ranges: for example, 51% increase after 100 perturbations for ±5% range for 3 EO case.

Finally, the robustness of the bladed disk assemblies with a few largely mistuned, or damaged blades, is determined. The effect of increasing the number of damaged blades from 1 to 3 does not seem to introduce any considerable changes in the maximum forced response. However, the mistuning pattern with one damaged blade is marginally more robust to small additional mistuning variations of ±0.5%.

Discussion of Large Mistuning Based Forced Response Reduction Strategies

The results obtained in this study provide an indication of the effectiveness of a concept of LM as a means of maximum forced response reduction and control. An example of an industrial bladed fan disk was presented, which demonstrates that both of the proposed approaches based on LM—random and deterministic—result in a substantial reduction of the peak amplification factor and an improvement of the sensitivity and robustness of the forced response. Although increasing the frequency mistuning range decreases the maximum forced response, the exact amount of this reduction varies depending on the EO excitation considered. The largest forced response reduction obtained was approximately 33% for 3 EO excitation. Additionally, statistical analysis revealed that as the frequency mistuning range in-

Table 8 Robustness assessment of selected intentional mistuning patterns to ±0.5% additional unintentional mistuning

Mistuning pattern type	Normalized forced response								
	3 EO			6 EO			13 EO		
	w rand	w/o rand	% diff	w rand	w o rand	% diff	w rand	w/o rand	% diff
Alternate 0.5%/−0.5%	1.938	1.204	60.96%	1.441	1.042	38.29%	1.607	1.179	36.30%
Alternate 15%/−15%	1.566	1.045	49.77%	1.446	1.037	41.36%	1.866	1.011	84.61%
Alternate 20%/−20%	1.532	1.046	46.47%	1.481	1.042	42.17%	1.932	1.024	88.68%
Alternate 40%/−40%	1.637	1.042	57.10%	1.594	1.044	52.68%	1.996	1.038	92.29%
Harmonic $n=1$, ±0.5%	1.436	1.020	40.78%	1.531	1.230	24.44%	1.367	1.140	19.91%
Harmonic $n=1$, ±15%	1.221	1.080	13.06%	1.232	1.290	4.50%	1.373	1.280	7.27%
Harmonic $n=1$, ±20%	1.180	1.170	0.85%	1.230	1.050	17.14%	1.390	1.210	14.88%
Harmonic $n=1$, ±40%	1.215	1.150	5.65%	1.151	1.050	9.62%	1.218	1.150	5.91%
Linear, ±0.5%	1.725	1.250	38.02%	1.488	1.150	29.39%	1.390	1.210	14.88%
Linear, ±15%	1.221	1.170	4.39%	1.019	0.960	6.17%	1.156	1.130	2.26%
Linear, ±20%	1.150	1.130	1.77%	1.070	0.990	8.08%	1.225	0.980	25.00%
Linear, ±40%	1.083	1.040	4.09%	1.042	1.020	2.20%	1.075	1.010	6.46%
One damaged blade, 15%	1.513	1.380	9.64%	1.249	1.090	14.59%	1.469	1.410	4.16%
One damaged blade, 40%	1.505	1.363	10.42%	1.248	1.164	7.23%	1.469	1.405	4.56%
Three damaged blade, 15%	1.546	1.330	16.24%	1.242	1.150	8.00%	1.456	1.320	10.30%
Three damaged blades, 40%	1.546	1.326	16.59%	1.242	1.165	6.61%	1.456	1.313	10.89%

creases, the distribution of the forced response becomes narrower and shifts toward the lower values, which further show benefits of implementing the LM.

The robustness of the maximum forced response of selected deterministic LM intentional mistuning patterns to additional unintentional random mistuning, inevitable in practice due to manufacturing imperfections and wear, was investigated in the second part of this study. Fundamentally, a high sensitivity of nearly tuned bladed disks was demonstrated, followed by a relative insensitivity to additional random mistuning after reaching the peak forced response levels, all of which points out prospective advantages from introducing large frequency mistuning. It was shown that it is possible to reduce the maximum forced response levels appreciably and to improve the robustness similarly by using the LM concept for some of the considered cases. In particular, linear mistuning patterns based on LM exhibited a significant robustness advantage over other mistuning patterns, in addition to a substantial forced response reduction, which, in combination with the desirable insensitivity, provides predictability and controllability of the maximum forced response. An attempt was also made to assess the "damage tolerance" of the bladed disk forced response with a few damaged blades included, and found that the precise extent of the damaged blade frequency deviation has no major influence on the maximum forced response levels or its robustness.

The maximum forced response levels achieved using some examples of the deterministic approach are lower than those corresponding to the random strategy, although the latter is believed to be more practical. As for the deterministic patterns, there are certain ways to accomplish their wider applicability in practice: Examples may include use of so-called "pseudoharmonic" or "pseudolinear" mistuning, which could include less blade types required to reach the desirable forced response characteristics. These practically oriented intentional mistuning patterns have been incorporated in the past [4], and in combination with the LM might provide the effective means of the maximum forced response control.

Conclusions

In this study, robust maximum forced response reduction strategies based on LM have been explored using an industrial bladed fan disk. Focusing on statistical analyses, the random approach has been assessed to reveal advantages of relaxing the manufacturing tolerances and of extending the typical design acceptance limits. Subsequently, the deterministic intentional mistuning method demonstrated a substantial reduction in maximum forced response and significant improvement in the robustness to the additional unintentional random mistuning by introducing unconventionally LM degrees for bladed disks.

The results from this study show ways of implementing the LM concept into the design of bladed disks as a means of avoiding large forced response levels caused by random mistuning and of ensuring the predictability of the response.

References

[1] Griffin, J. H., and Hoosac, T. M., 1984, "Model Development and Statistical Investigation of Turbine Blade Mistuning," ASME J. Vib., Acoust., Stress, Reliab. Des., **106**, pp. 204–210.

[2] Imregun, M., and Ewins, D. J., 1984, "Aeroelastic Vibration Analysis of Tuned and Mistuned Bladed Systems," *Proceedings of the Second Symposium on Unsteady Aerodynamics of Turbomachines and Propellers*, Cambridge, UK.

[3] Rzadkowski, R., 1993, "The General Model of Free Vibrations of Mistuned Bladed Discs, Part I: Theory, Part II: Numerical Results," J. Sound Vib., **173**(3), pp. 377–393.

[4] Castanier, M. P., and Pierre, C., 1997, "Consideration on the Benefits of Intentional Blade Mistuning for the Forced Response of Turbomachinery Rotors," *Proceedings of the ASME Aerospace Division*, AD-Vol. 55, pp. 419–425.

[5] Kenyon, J. A., and Griffin, J. H., 2000, "Intentional Harmonic Mistuning for Robust Forced Response of Bladed Disks," *Fifth National Turbine Engine HCF Conference*, Chandler, AZ, 2000.

[6] Mignolet, M. P., Hu, W., and Jadic, I., 2000, "On the Forced Response of Harmonically and Partially Mistuned Bladed Disks, Part I: Harmonic Mistuning; Part II: Partial Mistuning and Applications," Int. J. Rotating Mach., **6**(1), pp. 29–56.

[7] Brown, J. M., and Beachkofski, B., 2000, "A Phenomenological Investigation of Sequence Effects on Mistuned Rotor Response," *36th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Huntsville, AL.

[8] Jones, W. J., and Cross, C. J., 2002, "Reducing Mistuned Bladed Disk Forced Response Below Tuned Resonant Amplitudes," *Seventh National Turbine Engine High Cycle Fatigue Conference*, Palm Beach, FL.

[9] Petrov, E. P., Sanliturk, K. Y., and Ewins, D. J., 2002, "A New Method for Dynamic Analysis of Mistuned Bladed Disks Based on the Exact Relationship Between Tuned and Mistuned Systems," Trans. ASME: J. Eng. Gas Turbines Power, **124**, pp. 586–597.

[10] Petrov, E. P., Sanliturk, K. Y., Ewins, D. J., and Elliot, R., 2000, "Quantitative Prediction of the Effects of Mistuning Arrangement on Resonant Response of a Practical Turbine Bladed Disc," *Fifth National Turbine Engine High Cycle Fatigue Conference*, Chandler, AZ.

[11] Sheskin, D. J., 2000, *Handbook of Parametric and Nonparametric Statistical Procedures*, 2nd ed., Chapman and Hall, London.

[12] Bladh, R., Castanier, M., and Pierre, C., 1998, "Reduced Order Modeling and Efficient Forced Response Statistics Prediction for Mistuned Bladed Disks," *Third National Turbine Engine High Cycle Fatigue Conference*, San Antonio, TX.

[13] Myhre, M., Moyroud, F., and Fransson, T. H., 2003, "Numerical Investigation of the Sensitivity of Forced Response Characteristics of Bladed Disks to Mistuning," Paper No. 2003-GT-38007.

[14] Sanliturk, K. Y., Imregun, M., and Ewins, D. J., 1992, "Statistical Analysis of Random Mistuning of Bladed Assemblies," *IMechE Conference on Vibrations in Rotating Machinery*, Bath, pp. 51–58.

Explicit Finite Element Models of Friction Dampers in Forced Response Analysis of Bladed Disks

E. P. Petrov
Centre of Vibration Engineering,
Mechanical Engineering Department,
Imperial College London,
South Kensington Campus,
London SW7 2AZ, UK
e-mail: y.petrov@imperial.ac.uk

A generic method for analysis of nonlinear forced response for bladed disks with friction dampers of different designs has been developed. The method uses explicit finite element modeling of dampers, which allows accurate description of flexibility and, for the first time, dynamic properties of dampers of different designs in multiharmonic analysis of bladed disks. Large-scale finite element damper and bladed disk models containing $10^4 - 10^6$ *degrees of freedom can be used. These models, together with detailed description of contact interactions over contact interface areas, allow for any level of refinement required for modeling of elastic damper bodies and for modeling of friction contact interactions. Numerical studies of realistic bladed disks have been performed with three different types of underplatform dampers: (i) a "cottage-roof" (also called "wedge") damper, (ii) seal wire damper, and (iii) a strip damper. Effects of contact interface parameters and excitation levels on damping properties of the dampers and forced response are extensively explored.* [DOI: 10.1115/1.2772633]

Introduction

Due to the high modal density of natural frequencies of realistic bladed disks and of a broad spectrum of aerodynamic excitation forces, complete prevention of the occurrence of resonance regimes is not feasible. Hence, special devices that dissipate energy of vibrations and reduce resonance amplitudes to acceptable levels are commonly used in practical bladed disks. One of the most effective types of damping devices is a friction damper. This is generally a small piece of metal that can be fitted under platforms or shrouds of adjacent blades. When relative motion of adjacent blades becomes large enough, and micro- or, in some cases, even macroslip starts at damper contacting surfaces, the rubbing dissipates vibrational energy. Moreover, dampers stiffen a structure, which can significantly affect resonance frequencies of a bladed disk. The total damper stiffening effect depends on the stiffness of the damper body and the stiffness of the blade-damper friction contact interfaces. The stiffness properties of damper contact interfaces are dependent on the relative motion of pairing contact surfaces, and the dynamic stiffness of the damper body varies with vibration frequency, which can have significant effect, especially for highly flexible dampers.

There is more than 30 years of history of development of models for dynamic analysis of bladed disks with friction dampers. Yet, underplatform damper (UPD) models have been developed to date only for dampers of simple geometric shape (for the so-called wedge or cottage-roof dampers). In real conditions, such underplatform friction dampers have two interfaces at which they interact with adjacent blades. In order to allow for prediction of the forced response in conjunction with multi-degree-of-freedom (multi-DOF) blade models, several "kinematic" models for cottage-roof dampers have been developed, e.g., in Refs. [1–5]. These models are based on some kinematic hypotheses about the relative motion of a wedge-shaped damper and blade platform, and, hence, the damper's dynamics is ignored and inertia forces are neglected. As a result, the friction forces are expressed, in these models, through motion of the blade platforms. In order to describe microslip at the contact interfaces, two approaches were explored: (i) representation of friction contact by an array of spring-slider elements with different parameters (see Ref. [1]) and (ii) use of an analytical solution for friction forces between a bar and a foundation supporting this bar (see Refs. [2,3]). Characterization of such UPD models and analysis of the effects of damper contact stiffness are performed in Refs. [6–8]. Assessment of dampers' dissipation and stiffness effects, in reducing the forced response, is made in Ref. [9]. UPDs of semicylindrical and wedge shapes were studied in Ref. [10] with allowance for damper stiffness and inertia forces due to rigid-body motion.

Dynamic UPD models, in conjunction with multiharmonic balance method, have recently been developed for cottage-roof and for two-piece "split" dampers (see Ref. [11]). The first attempt to use the friction contact modeling capabilities and multiharmonic analysis code developed in Ref. [12] was made in Ref. [13] to explore possibilities of forced response analysis of a flexible finite element (FE) damper using a two-blade system. A system composed of two blades and a flexible model of a cottage-roof damper have been analyzed in Ref. [14] using simplified, monoharmonic representation for nonlinear forced response.

However, simplifying assumptions introduced in existing damper models, such as the neglect of damper inertia or other dynamic properties, including the damper's vibration modes, and neglect of inhomogeneity of stress distribution over contact interfaces and contact-separation transitions, significantly restrict the accuracy and area of applicability of these models.

Moreover, there is a tendency to introduce new damper designs in gas-turbine industry. Dampers, such as seal wires or shell-like dampers with very thin walls (see, e.g., Ref. [15]), are very flexible. Their own natural frequencies can be close to and even lower than the natural frequencies of blades. Adequate modeling cannot, therefore, be done without allowing for flexible vibrational modes of dampers and, neither existing kinematic damper models nor even damper models including damper inertia and elasticity, but allowing only rigid-body motion, cannot be used for such highly flexible dampers.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 27, 2007; final manuscript received May 24, 2007; published online February 25, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), Monteral, Quebec, Canada, May 14–17, 2007. Paper No. GT2007-27980.

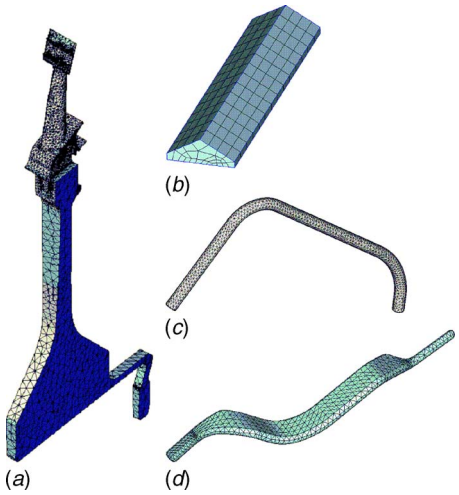


Fig. 1 FE models of (*a*) a bladed disk sector, (*b*) a cottage-roof damper, (*c*) a seal wire damper, and (*d*) a seal strip damper

Thus, predictive numerical studies could not be performed with acceptable accuracy for bladed disks with many existing and forthcoming types of friction damper design.

In this paper, a generic method for modeling underplatform and other types of friction damper has been developed. The method is based on explicit modeling of friction dampers in the analysis of nonlinear forced response of bladed disks using large-scale FE models. Such high-fidelity models allow accurate description of realistic geometric shapes and dynamic properties of all components of bladed disks, including dampers. As a result, any damper design and geometry can be analyzed, and, moreover, dynamic properties of the friction dampers and blade-damper contact interactions are accurately described without any simplified assumptions, which have been unavoidable hitherto.

Numerical investigations of bladed disks with flexible dampers of different designs are performed to demonstrate and validate the capabilities of the new models. Realistic large-scale FE models of bladed disks and friction dampers containing hundreds of thousands of DOFs are used in the calculations. An analysis of the influence of damper parameters on the forced response of bladed disks is included.

Method for Forced Response Analysis of Bladed Disks With Explicit Damper Models

Modeling of Dampers of Different Design and Bladed Disks. Large-scale FE models are used for the analysis of bladed disks and dampers. The models can frequently comprise millions of DOFs for bladed disks and tens or hundreds of thousands of DOFs for a damper elastic body. Examples of FE models for a bladed disk sector and for dampers of different design are shown in Fig. 1.

For the case of a tuned assembly, a single sector model is used for forced response calculations. A method for the analysis of nonlinear dynamics of structures with cyclic symmetry previously reported in Ref. [16] is used for this case. It allows significant reduction of computational expense while ensuring completeness and accuracy of calculation even for highly nonlinear structures. A method developed in Refs. [17,18] can be applied to a mistuned structure.

Dynamic models of a bladed disk and of a damper can be created separately when dampers are manufactured separately from a bladed disk and are not elastically connected to it. Models of dampers and bladed disks, considered in this section, describe the dynamic properties of elastic bodies of dampers and bladed disks. Because of this, the models are linear, which facilitates determination of their modal characteristics and allows accurate condensation of the models. Nonlinear friction contact and unilateral contact interaction between damper elastic bodies and blades are taken into account at a later stage, described in this paper in the following section.

For a case of steady-state vibrations, equations of motion for a bladed disk and for a damper can be written in the following form:

$$[(\mathbf{K}_e^{B,D} + \mathbf{K}_g^{B,D} - \Omega^2 \mathbf{M}_\Omega^{B,D}) + i\omega \mathbf{C}^{B,D} - \omega^2 \mathbf{M}^{B,D}]\hat{\mathbf{Q}}^{B,D} = \hat{\mathbf{P}}^{B,D} \tag{1}$$

where $\mathbf{K}_e^{B,D}$ is the conventional elastic stiffness matrix of a bladed disk (superscript B) or of a damper (superscript D), $\mathbf{K}_g^{B,D}(\Omega)$ is the so-called geometric stiffness matrix reflecting the stiffening effects of the centrifugal forces, $\mathbf{M}_\Omega^{B,D}$ is the spin-softening matrix describing the stiffness softening due to the changing direction of the centrifugal forces under vibration; $\mathbf{C}^{B,D}$ and $\mathbf{M}^{B,D}$ are damping and mass matrices, respectively, $i=\sqrt{-1}$, and Ω and ω are the rotation speed and an excitation frequency, respectively. $\hat{\mathbf{Q}}^{B,D}$ is the vector of complex amplitudes of displacements for all DOFs in a FE model of a bladed disk or a damper, and $\hat{\mathbf{P}}^{B,D}$ is the vector of complex amplitudes of excitation forces.

For a tuned bladed disk a single sector can be used for the exact calculation of both free vibration and forced responses of a whole structure. This is achieved by imposing special cyclic symmetry conditions, which describe the interaction of a sector with other sectors of a bladed disk. The cyclic symmetry conditions establish a relationship between DOFs located on the left and right boundaries of a sector, i.e.,

$$\hat{\mathbf{Q}}_{\text{left}}^{B} = e^{i(2\pi/N_B)m}\hat{\mathbf{Q}}_{\text{right}}^{B} \tag{2}$$

where N_B is the total number of sectors in a bladed disk and m is the number of deformation waves considered.

In order to allow efficient calculation of nonlinear forced response for bladed disks with dampers interacting at friction contact interfaces, the models provided by Eq. (1) are condensed. The DOFs at the nodes where there are no contact are excluded from the equations; full accuracy of modeling is preserved and forced response for excluded DOFs can still be determined.

The condensation process results in the (i) dynamic compliance, frequency response function (FRF) matrices for DOFs, where contact interactions are expected to occur, and (ii) complex amplitudes of displacements, which are evoked by the excitation forces at the contact nodes.

FRF matrices for a tuned bladed disk without dampers, $\mathbf{A}_{m_j}^{B}(m_j\omega)$, and for the damper(s) are generated from natural frequencies, mode shapes, and modal damping factors:

$$\mathbf{A}_m^B(\omega) = \sum_{r=1}^{N_m^B} \frac{\phi_{rm}^B(\Omega)[\phi_{rm}^B(\Omega)]^*}{[1 - i\eta_{rm}^B(\Omega)][\omega_{rm}^B(\Omega)]^2 - \omega^2} \tag{3}$$

$$\mathbf{A}^D(\omega) = \sum_{r=1}^{N^D} \frac{\phi_r^D(\Omega)[\phi_r^D(\Omega)]^T}{[1 - i\eta_r^D(\Omega)][\omega_r^D(\Omega)]^2 - \omega^2} \tag{4}$$

where $j=\overline{1,n}$, and $\omega_{rm}^B(\Omega)$, $\phi_{rm}^B(\Omega)$, and $\eta_{rm}^B(\Omega)$ are the rth natural frequency, mode shape, and modal damping factor, respectively, for a bladed disk without dampers corresponding to a family of modes with m nodal diameters. $\omega_r^D(\omega)$, $\phi_r^D(\omega)$, and $\eta_r^D(\omega)$ are the rth natural frequency, mode shape, and modal damping factor, respectively, for an individual damper. N_m^B and N^D are the numbers of modes used for the generation of FRF matrices of a bladed disk and a damper, respectively. Superscript $*$ indicates a Hermitian conjugate and superscript T indicates transposition here.

Vectors of mode shapes $\phi_{rm}^B(\omega)$ and $\phi_r^D(\omega)$ contain modal displacements only for nonlinear DOFs. Their components are simply selected from mass-normalized vectors $\hat{\phi}_{rm}^B(\Omega)$ and $\hat{\phi}_r^D(\Omega)$ obtained from the solution of eigenproblem resulting from Eq. (1)

by omitting terms $i\omega \mathbf{C}^{B,D}$ and $\hat{\mathbf{P}}^{B,D}$.

Dependency of all modal characteristics on the rotation speed Ω is allowed here in order to describe the effects of the rotation speed on the modal properties. Modal characteristics are determined for a number of traveling deformation waves m considered in Eq. (1). Mode shapes are calculated using a sector model for a tuned system, and these mode shapes take complex values for all nodal diameter numbers different from 0 and $N_B/2$.

Complex amplitudes of displacements caused by excitation forces distributed over the blade and damper nodes can be very efficiently calculated at the contact nodes as proposed in Ref. [16]:

$$^{\text{lin}}\mathbf{Q}_m^B(\omega) = \sum_{r=1}^{N_B} \frac{[\hat{\boldsymbol{\phi}}_{rm}^B(\Omega)]^* \hat{\mathbf{P}}^B}{[1 - i\eta_{rm}^B(\Omega)][\omega_{rm}(\Omega)]^2 - \omega^2} \phi_{rm}^B \tag{5}$$

$$^{\text{lin}}\mathbf{Q}^D(\omega) = \sum_{r=1}^{N_D} \frac{[\hat{\boldsymbol{\phi}}_r^D(\Omega)]^T \hat{\mathbf{P}}^D}{[1 - i\eta_r^D(\Omega)][\omega_r(\Omega)]^2 - \omega^2} \phi_r^D \tag{6}$$

Multiplication of large vectors, $(\hat{\boldsymbol{\phi}}_{rm}^B)^*\ \hat{\mathbf{P}}^B$ and $(\hat{\boldsymbol{\phi}}_r^D)^T\ \hat{\mathbf{P}}^D$, involving all sector DOFs provides a single coefficient for each mode shape, which is usually called "a modal force." These modal forces take into account arbitrary distribution of the excitation forces over all nodes of the bladed disk sector model and of the damper.

Modeling of Blade-Damper Friction Contact Interaction. The forces acting at friction contact interfaces are generally strongly nonlinear. The nonlinearity of the interaction is due to (i) unilateral contact of interaction along directions normal to contact surfaces, when compression normal stresses can be acted at these surfaces but tension stresses are not allowed; (ii) variation of contact areas during the vibration period, including closing and opening clearances and interferences, and contact-separation transitions over a whole interface surface or some parts of t; (iii) friction forces whose magnitude and stick-slip transitions are affected by contact separation and normal stress variation.

The friction contact interface models and friction contact elements developed in papers Refs. [19,20] take into account all the above phenomena and allow analytic calculation of interaction force vectors and stiffness matrices for the case of multiharmonic vibration, accurately and very fast. These elements allow modeling of friction contact interactions, which can be localized at a point or distributed over a line or an area. It should be noted that the actual contact conditions experienced by each contact node (e.g., slip, stick, separation, and full contact) are determined as a result of calculation and, therefore, there is no requirement to know nonlinear DOFs a priori. A reasonable guess about DOFs where nonlinear interactions appear simply helps us to reduce the number of DOFs kept in the nonlinear equations. Moreover, actual contact area, which can vary during a vibration period, and energy dissipated by each area friction contact element during the vibration period are also calculated.

Examples of applications of area and line friction contact element are shown in Fig. 2. For a cottage-roof damper (Fig. 2(a)), area contact elements are spread over both contact surfaces of the damper and, accordingly, over both pairing contact surfaces of the blade platform. For a seal wire damper (Fig. 2(b)), which has a circular cross section, contact is expected to be along lines and, hence, line friction contact elements are distributed over contact interfaces. For a strip damper (Fig. 2(c)), area contact elements are applied over one contact surfaces and line contact elements are distributed over an edge of the damper.

The friction interface elements allow calculation of the multiharmonic components of the contact interaction forces, $\mathbf{f}_e^{\text{loc}}(\mathbf{u}_e^{\text{loc}}) = \{\mathbf{f}_\tau, \mathbf{f}_\nu, \mathbf{f}_\zeta\}^T$, at each eth node of the FE mesh for a given multiharmonic component of relative displacements, $\mathbf{u}_e^{\text{loc}} = \{\mathbf{u}_\tau, \mathbf{u}_\nu, \mathbf{u}_\zeta\}^T$, and they also allow calculation of a stiffness matrix of the contact interface: $\mathbf{k}_e^{\text{loc}}(\mathbf{u}_e^{\text{loc}}) = \partial \mathbf{f}_e^{\text{loc}}(\mathbf{u}_e^{\text{loc}}) / \partial \mathbf{u}_e^{\text{loc}}$.

Due to the different nature of the contact forces acting in the plane of the contact and along a normal direction, the contact forces have to be calculated in a local coordinate system, τ, ν, ζ, with the ζ axis directed along the normal to the contact surface and the two other axes of this local Cartesian system, τ and ν, are located in a plane tangent to this contact surface. Each coordinate component of multiharmonic contact force and displacement here is a vector comprising $2n$ harmonic coefficients of an expansion of the forces and displacement, i.e.,

$$\mathbf{f}_\gamma = \{\tilde{f}_1^\gamma, \tilde{\bar{f}}_1^\gamma, \ldots, \tilde{f}_n^\gamma, \tilde{\bar{f}}_n^\gamma\}^T, \quad \mathbf{u}_\gamma = \{\tilde{u}_1^\gamma, \tilde{\bar{u}}_1^\gamma, \ldots, \tilde{u}_n^\gamma, \tilde{\bar{u}}_n^\gamma\}^T \tag{7}$$

where $\gamma = \tau,\ \nu,\ \zeta$ and n is the number of harmonics used in the multiharmonic representation of displacements.

Both the contact interaction forces and the contact stiffness matrix are generally strongly nonlinear and are dependent on the relative displacements. The analytical derivation of the expressions for $\mathbf{f}_e^{\text{loc}}$ $(\mathbf{u}_e^{\text{loc}})$ and $\mathbf{k}_e^{\text{loc}}$ $(\mathbf{u}_e^{\text{loc}})$ in Ref. [19] allows their exact and extremely fast calculation. In order to simulate the effects of three-dimensional friction contact interactions, two friction contact elements developed in Ref. [19] are applied in parallel for each contact node with directions of tangential motion for each of these two elements selected to be mutually perpendicular.

Relative displacements in the local coordinates are determined from vectors of displacements of blades, $\mathbf{u}_e^B = \{\mathbf{u}_x^B, \mathbf{u}_y^B, \mathbf{u}_z^B\}^T$, and dampers, $\mathbf{u}_e^D = \{\mathbf{u}_x^D, \mathbf{u}_y^D, \mathbf{u}_z^D\}^T$, at pairing nodes of the contact interface:

$$\mathbf{u}_e^{\text{loc}} = \mathbf{R}_e(\mathbf{u}_e^B - \mathbf{u}_e^D) \tag{8}$$

where $\mathbf{R}_e(6n \times 6n)$ is a matrix of transformation for the eth contact node from the global coordinate system to the local one.

In order to use the contact forces and stiffness matrices in the calculation of the whole bladed-disk-UPD system, they are transformed into the global coordinate system, which is used for the construction of the FE model for the bladed disk:

$$\mathbf{f}_e = \mathbf{R}_e^T \mathbf{f}_e^{\text{loc}}, \quad \mathbf{k}_e = \mathbf{R}_e^T \mathbf{k}_e^{\text{loc}} \mathbf{R}_e \tag{9}$$

Multiharmonic Balance Equation of Motion. In the method developed, a bladed disk and dampers are considered as elastic bodies, which are modeled separately by FE method, and then these models are condensed and combined in a coupled system using friction contact elements described in the preceding section.

The equation of motion for a coupled system bladed disk and dampers takes the following form:

$$\mathbf{K}\mathbf{q}(t) + \mathbf{C}\dot{\mathbf{q}}(t) + \mathbf{M}\ddot{\mathbf{q}}(t) + \mathbf{f}[\mathbf{q}(t)] - \mathbf{p}(t) = \mathbf{0} \tag{10}$$

where $\mathbf{K}$, $\mathbf{C}$, and $\mathbf{M}$ are stiffness, damping, and mass matrices of the combined system, respectively, $\mathbf{f}[\mathbf{q}(t)]$ is a vector of nonlinear interface forces, and $\mathbf{p}(t)$ is a vector of excitation forces.

The steady-state periodic vibration response is calculated, and therefore, calculation can be efficiently done in the frequency domain using the multiharmonic balance method. In accordance with this method, displacements are represented by a restricted Fourier series:

$$\mathbf{q}(t) = \sum_{j=1}^{n} (\bar{\mathbf{q}}_j \cos m_j \omega t + \bar{\bar{\mathbf{q}}}_j \sin m_j \omega t) \tag{11}$$

where $\bar{\mathbf{q}}_j$ and $\bar{\bar{\mathbf{q}}}_j$ $(j = 1, \ldots, n)$ are vectors of cosine and sine harmonic coefficients of the multiharmonic expansion (marked here and further by single and double bars above a symbol), and m_j $(j = 1, \ldots, n)$ are specific numbers of harmonics that are kept in the displacement representation. Choice of the total number of harmonics n kept in the multiharmonic expansion and values selected for these harmonics m_j are determined by the solution accuracy required and by the calculated spectrum of the forced response: harmonics that can contribute significantly to the forced response

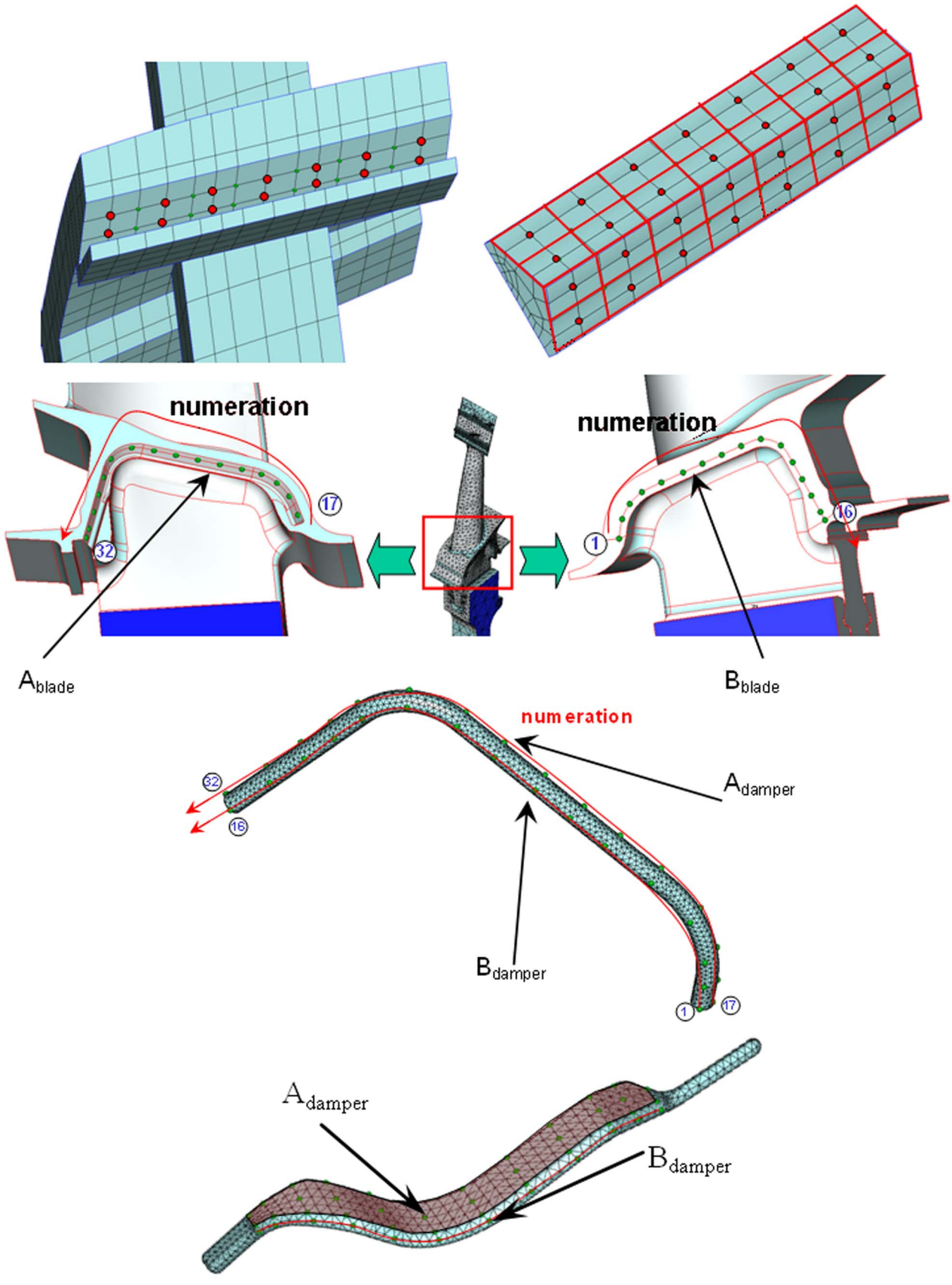


Fig. 2 Application of area and line friction contact interface elements: (*a*) a bladed disk with cottage-roof dampers, (*b*) a bladed disk with seal wires, and (*c*) a strip damper

should be included. For the search of major and superharmonic resonances, m_j is an integer number, and for the search of subharmonic and combination resonances, m_j can be a fractional number, appropriately selected. As a result, multiharmonic representation of the forced response allows the determination of periodic forced response even for strongly nonlinear structures including (i) major resonances, (ii) superharmonic resonances, (iii) subharmonic resonances, and (iv) combination resonances with any accuracy required.

Application of the multiharmonic balance method and use of the condensed matrix expressions of Eqs. (3)–(6) provide condensed equations of motion in the frequency domain. These equations are formulated exclusively for nonlinear DOFs, i.e., with respect to harmonic coefficients of the multiharmonic expansion for a set of DOFs located at contact interfaces. These equations take the form

$$\mathbf{R}_j = \begin{Bmatrix} \mathbf{Q}_j^B \\ \mathbf{Q}_j^D \end{Bmatrix} - \begin{Bmatrix} {}^{\text{lin}}\mathbf{Q}_j^B \\ {}^{\text{lin}}\mathbf{Q}_j^D \end{Bmatrix} + \begin{bmatrix} \mathbf{A}_j^B & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_j^D \end{bmatrix} \begin{Bmatrix} \mathbf{F}_j(\mathbf{Q}^B, \mathbf{Q}^D) \\ -\mathbf{F}_j(\mathbf{Q}^B, \mathbf{Q}^D) \end{Bmatrix} = \mathbf{0} \tag{12}$$

where $j = \overline{1,n}$, and $\mathbf{Q}_j^B = \bar{\mathbf{Q}}_j^B + i\bar{\bar{\mathbf{Q}}}_j^B$ and $\mathbf{Q}_j^D = \bar{\mathbf{Q}}_j^D + i\bar{\bar{\mathbf{Q}}}_j^D$ are sought for vectors of complex amplitudes for the bladed disk and the damper, respectively, for the jth harmonic. Subscript j is added to ${}^{\text{lin}}\mathbf{Q}_j^{B,D}$ and $\mathbf{A}_j^{B,D}$ to indicate that these vectors and matrices are evaluated for frequency $m_j\omega$ and m_j numbers of traveling waves. $\mathbf{F}_j(\mathbf{Q}^B, \mathbf{Q}^D)$ is a vector of complex amplitudes of the jth harmonic of the interaction forces acting at the contact interface. This vector is dependent on vectors $\mathbf{Q}^B = \{\mathbf{Q}_1^B, \mathbf{Q}_2^B \ldots, \mathbf{Q}_n^B\}^T$ and $\mathbf{Q}^D = \{\mathbf{Q}_1^D, \mathbf{Q}_2^D, \ldots, \mathbf{Q}_n^D\}^T$, which include amplitudes of all harmonics for the bladed disk and the damper, and, therefore, the equations of motion given by Eq. (12) for all harmonics are coupled.

Calculation of the Multiharmonic Amplitudes. Equation (12) represents a nonlinear set of equations with respect to vector of multiharmonic amplitudes $\mathbf{Q} = \{\mathbf{Q}^B, \mathbf{Q}^D\}^T$. The Newton–Raphson method together with schemes of solution continuation/tracing is applied to solve this equation. An iterative Newton–Raphson solution process is expressed by the following formula:

$$\mathbf{Q}^{(k+1)} = \mathbf{Q}^{(k)} - \mathbf{J}^{-1}(\mathbf{Q}^{(k)})\mathbf{R}(\mathbf{Q}^{(k)}) \tag{13}$$

where superscript (k) indicates the number of the iteration and the real residuals vector $\mathbf{R}(\mathbf{Q})$ is formed from residual vectors obtained for each harmonic in Eq. (12):

$$\mathbf{R}(\mathbf{Q}) = \{\text{Re}(\mathbf{R}_1), \text{Im}(\mathbf{R}_1), \ldots, \text{Re}(\mathbf{R}_n), \text{Im}(\mathbf{R}_n)\}^T \tag{14}$$

The iterative solution is terminated when a required accuracy is achieved, i.e., $\|\mathbf{R}(\mathbf{Q})\| < \varepsilon$. A Jacobian of Eq. (12) is determined as

$$\mathbf{J}(\mathbf{Q}) = \partial\mathbf{R}(\mathbf{Q})/\partial\mathbf{Q} = \mathbf{I} + \mathbf{A}(\omega)\mathbf{K}_{\text{nln}}(\mathbf{Q}) \tag{15}$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{A}(\omega)$ is a real matrix formed from real and imaginary parts of the FRF matrices, and $\mathbf{K}_{\text{nln}}(\mathbf{Q}) = \partial\mathbf{F}(\mathbf{Q})/\partial\mathbf{Q}$ is a stiffness matrix of the friction interface.

For the solution of Eq. (12) using the scheme given by Eq. (13), the nonlinear friction contact forces acting at blade-damper interfaces, $\mathbf{F}(\mathbf{Q})$, and the stiffness matrix of the friction contact interface, $\mathbf{K}_{\text{nln}}(\mathbf{Q})$, have to be calculated. These are formed by summing up vectors of nonlinear forces and stiffness matrices at all friction contact interfaces, i.e.,

$$\mathbf{F}(\mathbf{Q}) = \bigcup_{e=1}^{N_e} \mathbf{f}_e(\mathbf{Q}), \quad \mathbf{K}_{\text{nln}}(\mathbf{Q}) = \bigcup_{e=1}^{N_e} \mathbf{k}_e(\mathbf{Q}) \tag{16}$$

where N_e is the total number of contact elements. Since an analytical formulation is derived for the friction contact elements, these matrices are calculated very fast and exactly.

Test Case Studies

The methodology developed has been applied for the analysis of three different bladed disks with UPDs (see Fig. 2) of different types.

ADTurbII Blisk With Cottage-Roof Dampers. Forced response of a tuned ADTurbII blisk (see Refs. [11,21]) with cottage-roof UPDs is analyzed using explicit FE modeling of the damper. The blisk consists of 24 blades, and the FE model of one sector of the blisk used in the calculation contains 21,555 DOFs and the FE model of the cottage-roof damper contains 6146 DOFs.

For generation of the FRF matrices of the blisk, the first 32 natural frequencies and mass-normalized mode shapes were determined for each nodal diameter (cyclic index) involved in the forced response calculation. For generation of the FRF matrices of the cottage-roof damper, the first 24 natural frequencies and mass-normalized mode shapes were used.

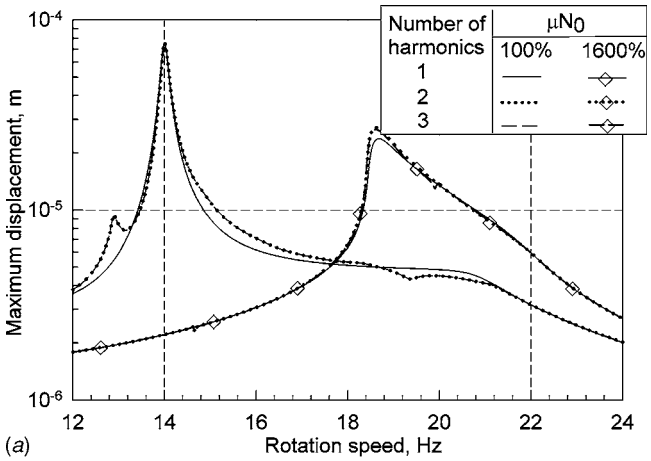
The natural frequencies–nodal diameters diagram for this bladed disk is given in paper Ref. [18]. The background damping due to damping in the blisk material is assumed to be low: $\eta = 7.5 \times 10^{-5}$. Dissipation of vibration energy due to the friction forces produced by UPDs is calculated together with the forced response calculation. A 19th engine order traveling wave excitation pattern is studied.

Blade-damper interaction was modeled by 28 area friction contact elements distributed evenly on both sides of the cottage-roof damper. Area contact elements and nodes, which are used for determination of the contact stresses at the UPD, are shown in Fig. 2(*a*). Forced response level was determined at the blade tip.

Choice of Number of Harmonics Kept in a Solution. At the beginning of the numerical studies, the total number of harmonics required to be included in the multiharmonic forced response has to be determined.

To do this, the forced response was calculated with different numbers of harmonics kept in the multiharmonic expansion of the displacements, namely, 1, 2, and 3. In the multiharmonic expansion, first odd harmonics were used, namely, 19, 57, and 95 (which are the first, third, and fifth harmonics corresponding to the case of 19EO excitation analyzed). In order to be sure that the choice of number of harmonics will provide accurate results for the range of damper parameters to be explored, two cases were checked: (i) a case of nominal damper mass, 100%, and (ii) a heavy damper with mass value 1600% of its nominal value. The maximum displacements found over the vibration period and the amplitudes of each harmonic component included in the analysis are shown in Figs. 3(*a*) and 3(*b*), respectively. One can see that use of two first odd harmonics gives the forced response level practically identical to the case when three harmonics are included. Therefore, in further calculations, only two harmonics, 19th and 57th, are kept.

Effects of Damper Parameters on Forced Response. Results of the forced response calculation, obtained for different levels of static normal stresses, are shown in Fig. 4. This variation of static normal stresses can be achieved in practice by a change in the damper mass. Cases of the normal stresses corresponding to a damper mass from 50% of the nominal value to 12,800% were studied. One can see that for lower damper mass values (from 50% to 100%), the resonance frequency for the blisk with dampers is close to the resonance frequency of the 1F mode of the blisk without dampers. For higher mass values (from 6400% to 12,800%), this resonance frequency is close to the resonance frequency of the blisk with fully stuck dampers. For mass values from 200% to 1600%, the resonance frequency of the blisk with dampers takes intermediate values between resonance frequencies of linear systems: (i) the blisk without dampers and (ii) the blisk with stuck dampers. For low values of the static normal stresses (i.e., 50% and 100%), separation of contact surfaces has been observed over a large part of the contact interface area, which



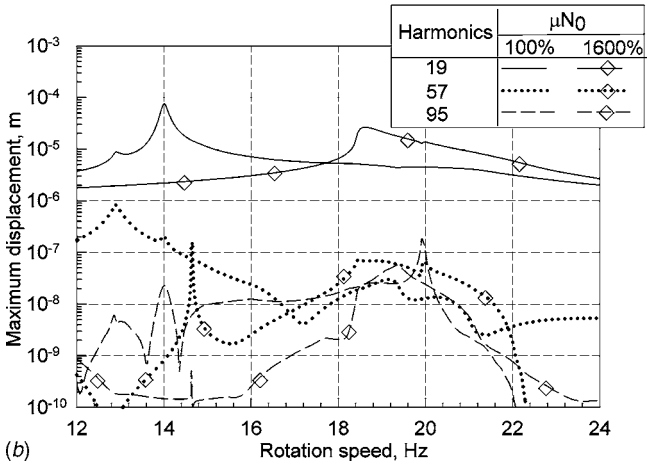


Fig. 3 Forced response levels calculated with different numbers of harmonics: (*a*) maximum displacement and (*b*) amplitudes of harmonic components

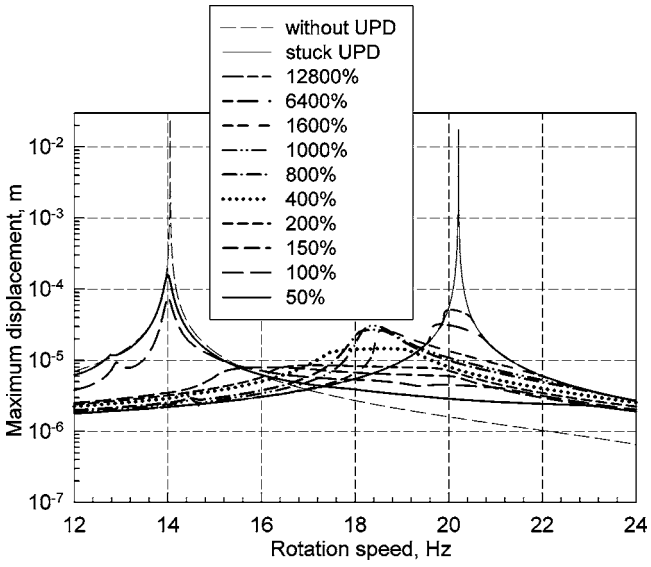


Fig. 4 Forced response of the blisk: effects of levels of the normal contact stresses/damper mass

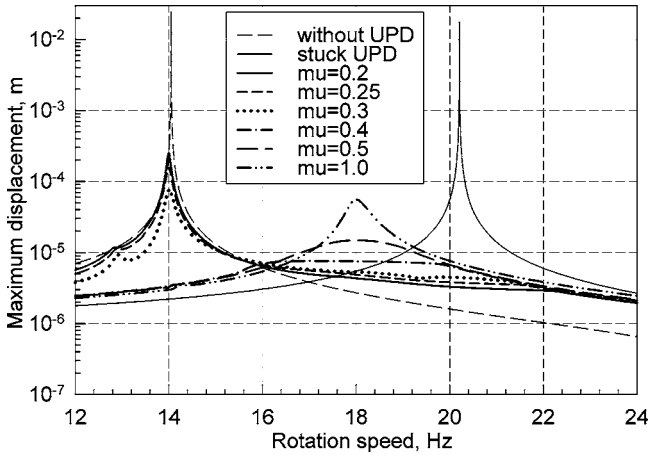


Fig. 5 Forced response of the blisk: effect of friction coefficient value

indicates possibility of the so-called damper clapping.

Forced response levels calculated for the nominal 100% damper mass value, but with different friction coefficients, are shown in Fig. 5.

Dependency of resonance frequency and resonance amplitude on parameters of the contact interfaces, such as (i) friction coefficient value and (ii) normal load level, is plotted in Fig. 6. One can see that the resonance frequency for friction coefficient values less than 0.3 is close to the resonance frequency of the blisk without dampers (resonance rotation speed of 14 Hz). When the friction coefficient varies from 0.3 to 0.5, the resonance frequency changes significantly; for higher friction coefficient values from 0.5 to 1.0, the resonance frequency is almost constant (resonance rotation speed of 18 Hz). Both the normal load and the friction coefficient significantly affect the resonance response level, which can be reduced for this blisk by factor of 10^3 by appropriate choice of friction coefficient or normal load.

Bladed Disk With Seal Wire Dampers. Another structure analyzed is a high pressure turbine bladed disk which comprises 114 blades with seal wire dampers (see Fig. 1(*b*)) fitted in special cavities underneath the blade platform. The seal wires are in contact with adjacent blades and they dissipate vibration energy similar to conventional cottage-roof UPDs. FE models contain (i) 314,187 DOFs for a single bladed disk sector and (ii) 37,785

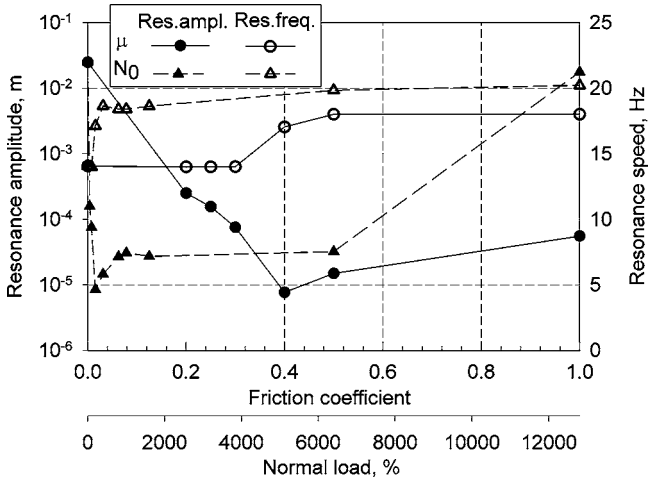


Fig. 6 Dependency of the resonance rotation speed and amplitude on contact parameters: (i) friction coefficient and (ii) normal load

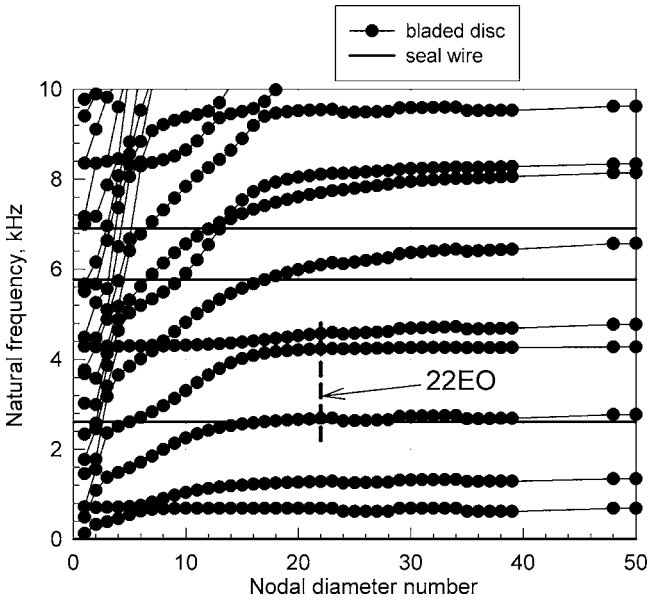


Fig. 7 Motion of the cottage-roof damper and blade platforms over vibration period

DOFs for a seal wire damper. Excitation by 22EO excitation was considered, and the damping loss factor, which is due to material energy dissipation, is assumed to be 0.005.

For generation of the FRF matrices for the blisk, the first 48 natural frequencies and mass-normalized mode shapes were determined for each nodal diameter involved in the forced response calculation. For generation of the FRF matrices of the seal wire, the first 48 natural frequencies and mass-normalized mode shapes were calculated. The lower natural frequencies of the bladed disk and of the damper are plotted in Fig. 7. In this figure, the frequency range analyzed is also shown. Due to its high flexibility, the seal wire natural frequencies are close to the natural frequencies of the bladed disk. In the frequency range of interest, there are natural frequencies of the bladed disk and of the seal wire and, hence, the seal wire deforms significantly when vibrating in this frequency range. Therefore, the seal wire cannot be modeled as a rigid body, and allowing for its dynamic properties is essential for accurate forced response analysis.

The number of harmonics necessary for accurate forced response calculation was decided by comparison of the forced re-

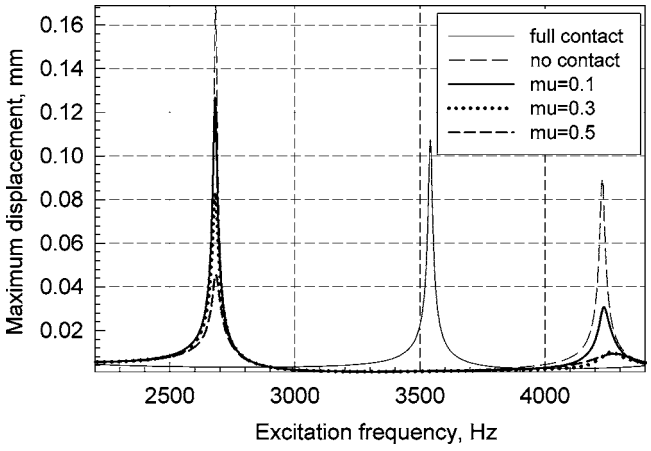


Fig. 8 Forced responses under 100% excitation and normal load levels

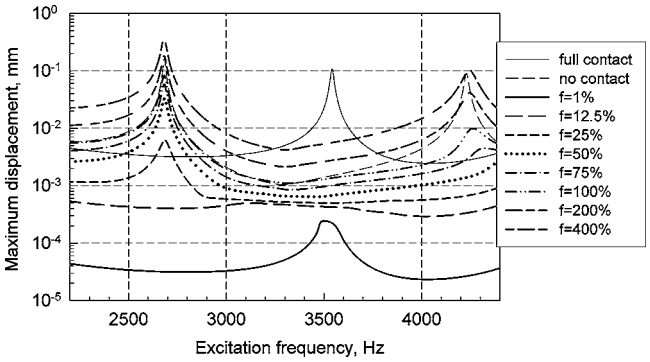


Fig. 9 Forced response levels for different excitation levels (μ=0.3 and 100% normal load level)

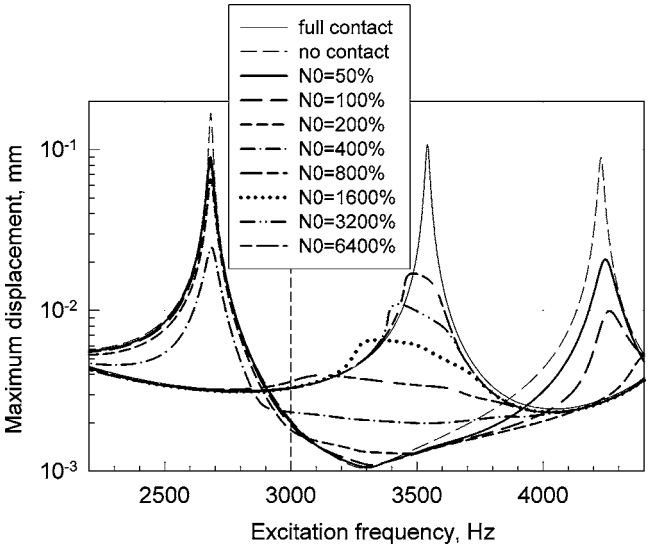


Fig. 10 Forced response for different normal loads applied at contact nodes (μ=0.3 and 100% excitation level)

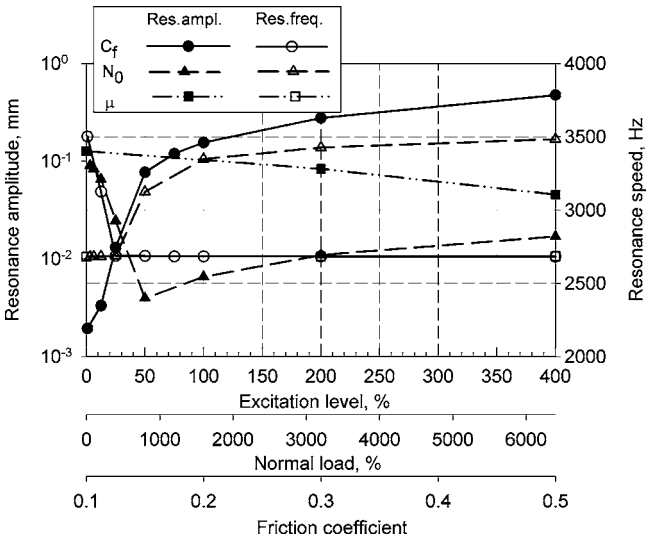


Fig. 11 Dependency of the resonance rotation speed and amplitude on (i) excitation level, (ii) normal load, and (iii) friction coefficient

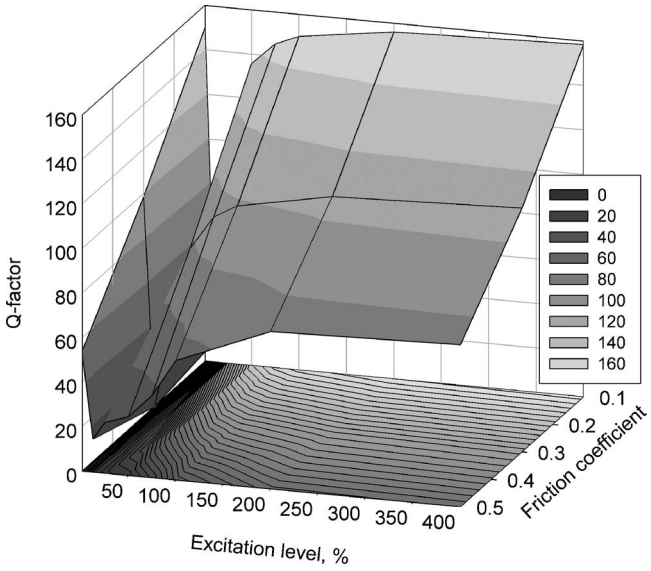


Fig. 12 Dependency of the *Q* factor on the excitation level and friction coefficient (a case of N_0=100%)

sponses obtained with one, two, and three first odd harmonics of the 22EO excitation. Since for all three cases considered the results are very close, in further analysis, only one harmonic was used.

Effects of the friction coefficient values, excitation levels, and normal load levels on the forced response characteristics are shown in Figs. 8–10. In all these figures, forced responses of the bladed disk for two limiting cases are also plotted for comparison: (i) a forced response of the bladed disk without dampers (a solid black line) and (ii) a forced response of the bladed disk with fully stuck dampers (a dashed black line). 100% excitation level was considered for both these limiting cases.

Figure 8 demonstrates significant influence of the friction coefficient on resonance forced response. Noticeable variation of the resonance frequencies can also be observed. From Fig. 9, one can see that the damping effect of the seal wire is evident for all excitation force levels. Even for an excitation level of 1% of its nominal value, the forced response of a structure with dampers is significantly lower than that for the case when dampers are absent or when they are fully stuck. For a 1% excitation level, the first resonance frequency of the bladed disk with dampers is close to the first resonance frequency of a bladed disk with fully stuck dampers. For the other excitation levels considered here, the first resonance frequency is close to the resonance frequency of a bladed disk without dampers.

Results displayed in Fig. 10 show that only for extremely high levels of normal loads at contact interfaces (from 1600% to 6400%) does the resonance frequency approach the resonance frequency of the bladed disk with fully stuck dampers. For all realistic values, from 50% to 800%, they are close to the resonance frequency of the bladed disk without dampers.

Resonance frequency and amplitude are plotted in Fig. 11 as functions of friction coefficient, excitation level, and level of normal load. The results show high sensitivity of the resonance response levels to damper parameters: a small variation of any of the parameters considered can change forced response level by several times and the resonance frequency can be changed by up to 40% in some ranges of parameter variation. However, the resonance frequency appears to be insensitive, for the considered case, to variation of friction coefficient from 0.1 to 0.5.

The *Q* factor was extracted from the forced response calculations, and its dependency on the friction coefficient and the excitation level is shown in Fig. 12 in the form of 3D mesh and contour plots.

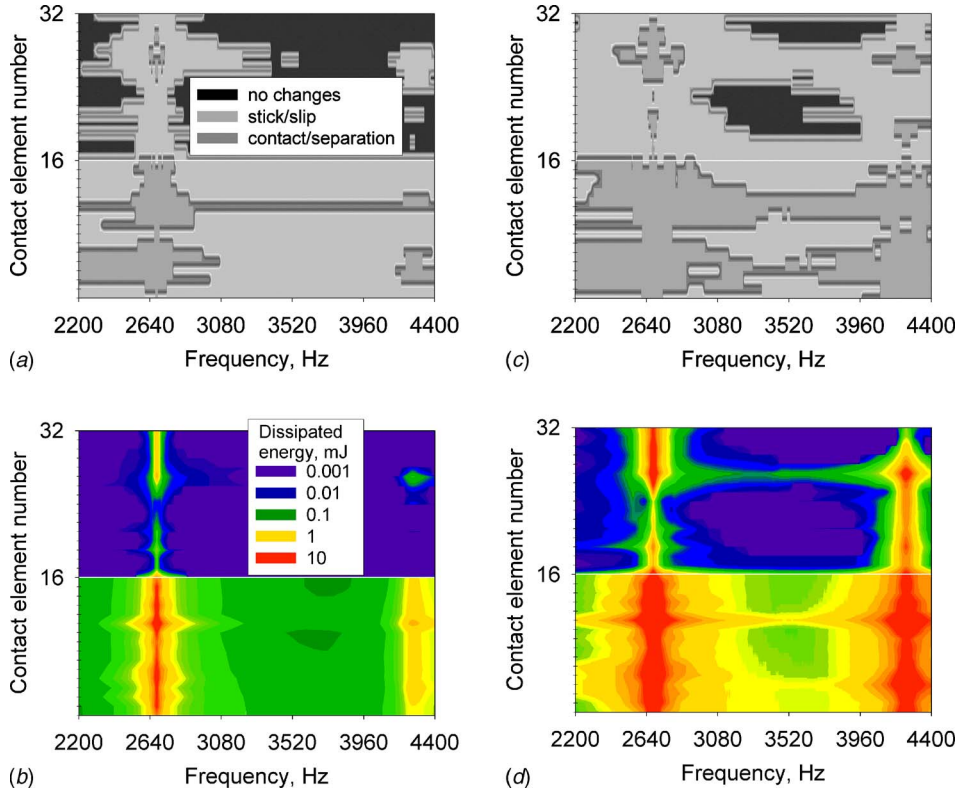


Fig. 13 Contact conditions and energy dissipated (mJ) by each of the 32 friction contact elements for two excitation levels: (*a*) 100% and (*b*) 400%

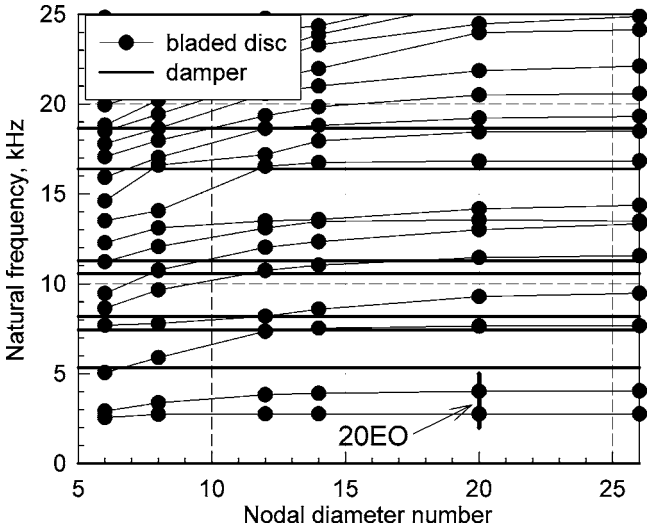


Fig. 14 Natural frequencies of the bladed disk and of the UPD

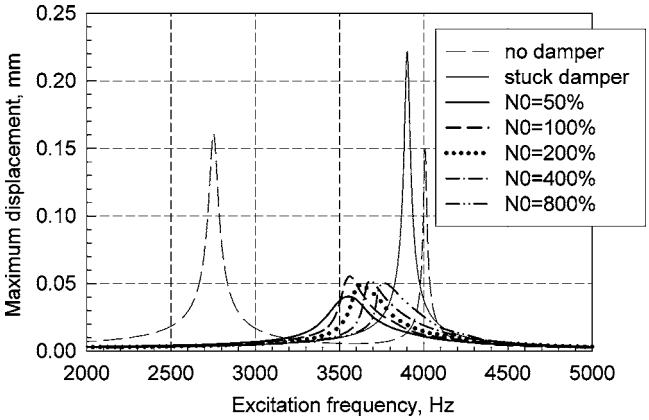


Fig. 16 Forced response levels for different normal loads at contact interfaces

The friction contact interface elements allow the determination, for each element, of the (i) contact conditions and (ii) energy dissipated over the area covered by this element.

The contact conditions are determined over the period of vibration for each excitation frequency, and the following conditions are indicated for each contact interface element:

(i) no change in contact conditions, i.e., the pairing nodes at both sides of a contact interface are in contact and remain stuck over the vibration period or the pairing nodes are separated and remain separated over the whole vibration period
(ii) stick-slip transition, i.e., the pairing nodes are always in contact and a slip-stick transition occurs at least once during the vibration period
(iii) contact-separation transition, i.e., contact-to-separation or separation-to-contact transition of the contact state occurs at least once over the vibration period. Slip-stick transitions can also occur during intervals when nodes are in contact.

Energy dissipated over a vibration period is also calculated for each friction contact element. Plots indicating contact condition and energy dissipated by each of the 32 friction contact elements over the whole frequency range analyzed are shown in Fig. 13.

Three possible contact conditions are indicated by different colors: (i) a case without contact condition changes is colored blue, (ii) a case of slip-stick transitions without contact separation is green, and (iii) a case of contact-separation transition is red. Contour lines of the dissipated energy are plotted in a logarithmic scale. The legend explaining the color scheme used for indication of energy dissipation levels is also given in Fig. 13. One can see that, for a case of 100% excitation level, majority of contact nodes are slipping or separating, although at contact interface A (see Fig. 2(*b*)), there is still a large area that is fully stuck over a wide excitation frequency range. Increase of the excitation level to 400% increases the number of contact elements where contact separation occurs. In all cases, most energy is dissipated at contact interface B.

Bladed Disk With Strip Dampers. The third bladed disk analyzed has 54 blades. A novel damper design was applied in this assembly: seal damper strips. The FE model of this damper is shown in Fig. 1(*c*). A FE model of a bladed disk sector used in the analysis contains 309,990 DOFs, and a FE model of the damper strip contains 24,753 DOFs. 45 friction contact elements were distributed over two contact interfaces of the damper and, accordingly, over pairing surfaces of the blade platforms: 32 area contact elements distributed over a contact surface marked by letter A in Fig. 1(*c*) and 13 line contact elements at an interface marked by letter B.

For generation of the FRF matrices of the bladed disk, the first

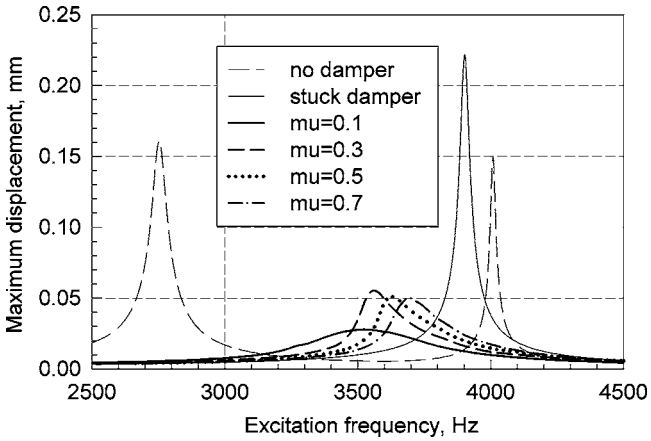


Fig. 15 Forced responses for different friction coefficient values

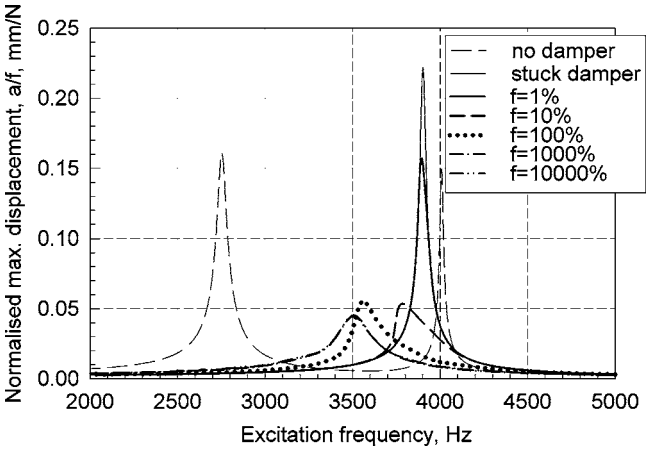


Fig. 17 Normalized forced response under different excitation levels

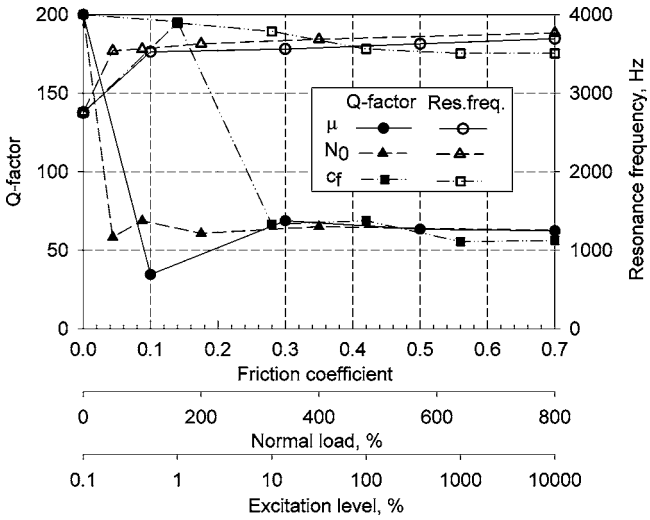


Fig. 18 Dependency of the resonance frequency and *Q* factor on (i) excitation level, (ii) normal load, and (iii) friction coefficient value

48 natural frequencies and mass-normalized mode shapes were determined for each nodal diameter involved in the forced response calculation. For generation of the FRF matrices of the damper strip, the first 48 natural frequencies and mass-normalized mode shapes were used. Excitation by 20EO is considered and the inherent damping loss factor, assumed for a case when dampers are not fitted, is 0.005.

The lower natural frequencies of the bladed disk and of the damper strip are plotted in Fig. 14. One can see that the strip damper has natural frequencies that are close to the natural frequencies of the bladed disk, and they are located within frequency ranges analyzed; allowance for flexibility of the damper is therefore essential for accurate calculations.

Analysis of the number of harmonics necessary showed that use of only one harmonic can incur 30% error in the determination of forced response levels. The results obtained with two harmonics were very close to the results obtained with 3 and higher number of harmonics, and therefore, two first odd harmonics of 20EO were used in the calculations: 20th and 60th.

Forced response levels calculated for different friction coefficient values are shown in Fig. 15.

For comparison, forced responses of two limiting linear structures are also plotted in this figure: (i) forced response of the bladed disk without dampers and (ii) forced response of the bladed disk with fully stuck dampers. One can see that the friction coefficient value significantly affects the resonance level and the resonance frequency. The resonance frequency of the bladed disks with strip dampers is located between resonance frequencies of the limiting linear cases. For higher friction coefficient values, it is closer to the resonance frequency of the bladed disk with fully stuck dampers; for $\mu=0.1$, the resonance frequency is almost centered between resonance frequencies of the bladed disk without and with stuck dampers. The lowest level of resonant forced response is achieved for $\mu=0.1$.

Forced response levels calculated for different static normal stresses are shown in Fig. 16.

Normalized forced response levels calculated for different ex-

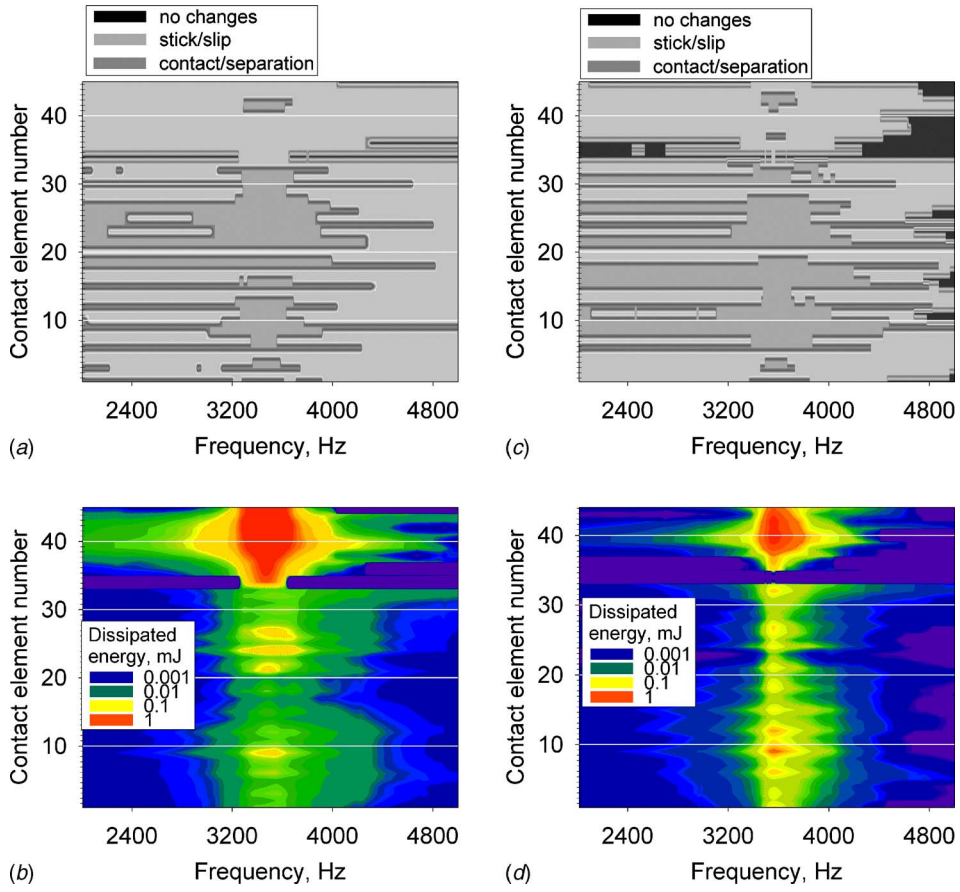


Fig. 19 Contact conditions for each of the 45 friction contact elements: (*a*) $\mu=0.1$ and (*b*) $\mu=0.3$

citation levels are shown in Fig. 17. The normalization is performed by dividing the forced response by the excitation level coefficient. This normalization allows comparison of nonlinear effects of dampers on forced response: for the case of linear vibrations, all curves would be identical. One can see that for excitation levels varying from 1% to 1000%, normalized amplitude levels and resonance frequencies are different. However, for high excitation levels, from 1000% to 10,000%, normalized forced response curves are almost identical, which indicates smaller nonlinear effects for high level of excitation.

Effects of all parameters analyzed on the resonance frequency and Q factor are summarized in Fig. 18. It is important to notice the high robustness of the damping properties of this damper design: for wide ranges of variation of damper parameters and excitation levels the damper exhibits stable, almost constant values for the resonance frequency and damping characteristics.

Examples of contact condition and energy dissipation plots for each of the 45 friction contact elements, over the whole frequency range analyzed, are shown in Fig. 19.

One can see that the number of contact nodes that experienced temporary separation is larger for higher values of the friction coefficient. In all cases analyzed, the dominant part of the vibration energy is dissipated at the contact interface marked by letter B in Fig. 1(c).

Conclusions

A generic method for the analysis of nonlinear forced response for bladed disks with friction damping devices of different designs has been developed. The method allows the use of explicit FE models of dampers, which can accurately describe any complex geometric shape, flexibility, and, for the first time, dynamic properties of dampers in multiharmonic analysis of bladed disks.

Steady-state, strongly nonlinear forced response is calculated by a multiharmonic balance method in conjunction with an original analytical formulation for the extended Jacobian of the nonlinear equations and for evaluation of nonlinear contact forces and contact stiffnesses.

Large-scale FE damper and bladed disk models containing 10^4-10^6 DOFs can be used in order to describe the complex geometric shape of modern dampers and bladed disks of different designs with any required level of refinement.

Detailed description of contact interactions over area and line contact interfaces between damper and blades is achieved by especially developed friction contact elements. These elements allow for friction, stick slip, and contact-separation contact state transitions. Effects of normal load variation not only on tangential friction force level but also on time instant of stick slip are included in the friction modeling.

Numerical studies of realistic bladed disks with three different types of UPDs have been performed: (i) a cottage-roof damper, (ii) a seal wire damper, and (iii) a strip damper. Effects of contact interface parameters and excitation levels on damping properties of the dampers on forced response have been extensively explored.

For the first time, dampers of highly flexible design, such as seal wire dampers and strip damper, have been studied, and, moreover, distributions over contact interface areas for energy dissipated and for contact conditions have been calculated.

Acknowledgment

The author is grateful to Rolls-Royce plc. for providing the financial support for this project and for giving permission to publish this work.

References

[1] Sanliturk, K. Y., Ewins, D. J., and Stanbridge, A. B., 2001, "Underplatform Dampers for Turbine Blades: Theoretical Modelling, Analysis and Comparison With Experimental Data," Trans. ASME: J. Eng. Gas Turbines Power, **123**, pp. 919–929.

[2] Csaba, G., 1998, "Forced Response Analysis in Time and Frequency Domains of a Tuned Bladed Disk With Friction Dampers," J. Sound Vib., **214**(3), pp. 395–412.

[3] Jareland, M. H., 2001, "A Parametric Study of a Cottage Roof Damper and Comparison With Experimental Results," ASME Paper No. 2001-GT-0275.

[4] Yang, B.-D., and Menq, C.-H., 1998, "Characterization of Contact Kinematics and Application to Design of Wedge Dampers in Turbomachinery Blading," Trans. ASME: J. Eng. Gas Turbines Power, **119**, pp. 410–423.

[5] Zucca, S., Borrajo, J., and Gola, M. M., 2006, "Forced Response of Bladed Disks in Cyclic Symmetry With Underplatform Dampers," ASME Paper No. GT2006–90785.

[6] Szwedowicz, J., Kissel, M., Ravindra, B., and Kellerer, R., 2001, "Estimation of Contact Stiffness and Its Role in the Design of a Friction Damper," ASME Paper No. GT-2001-0290.

[7] Koh, K. H., Griffin, J. H., Filippi, S., and Akay, A., 2005, "Characterization of Turbine Blade Friction Dampers," ASME J. Eng. Gas Turbines Power, **127**(4), pp. 856–862.

[8] Koh, K.-H., and Griffin, J. H., 2006, "Dynamic Behaviour of Spherical Friction Dampers and Its Implication to Damper Contact Stiffness," *Proceedings of ASME Turbo Expo*, Barcelona, Spain, May 8–11.

[9] D'Ambrosio, F., Chatelet, E., and Jacquet, G., 2005, "Influence of Contact States on the Dynamic Behavior of Rubbing Structures," ASME Paper No. GT2005-68560.

[10] Panning, L., Sextro, W., and Popp, K., 2003, "Spatial Dynamics of Ttuned and Mistuned Bladed Disks With Cylindrical and Wedge-Shaped Friction Dampers," Int. J. Rotating Mach., **9**(3), pp. 219–228.

[11] Petrov, E. P., and Ewins, D. J., 2007, "Advanced Modelling of Underplatform Friction Dampers for Analysis of Bladed Disc Vibration," ASME J. Turbomach., **129**, pp. 143–150.

[12] Petrov, E. P., and Ewins, D. J., 2004, "State-of-the-Art Dynamic Analysis for Nonlinear Gas Turbine Structures," Proc. IMechE: J. Aerosp. Eng., **218**(G3), pp. 199–211.

[13] Yeo, S., and Kielb, J., 2002, "The Application of a Numerical Code to the Optimisation of Inter-Platform Damping Technology," *Proceedings of the Seventh HCF Conference*, FL, May 14–17.

[14] Firrone, C. M., Botto, D., and Gola, M. M., 2002, "Modelling a Friction Damper: Analysis of the Experimental Data and Comparison With Numerical Results," *Proceedings of ESDA2006, Eighth Biennial ASME Conference on Engineering Systems Design and Analysis*, Torino, Italy, July 4–7.

[15] Szwedowicz, J., Gibert, C., Sommer, T. P., and Kellerer, R., 2006, "Numerical and Experimental Damping Assessment of a Thin-Walled Friction Damper in the Rotating Set-Up With High Pressure Turbine Blades," ASME Paper No. GT2006–90951.

[16] Petrov, E. P., 2003, "A Method for Use of Cyclic Symmetry Properties in Analysis of Nonlinear Multiharmonic Vibrations of Bladed Discs," ASME Paper No. GT-2003-38480.

[17] Petrov, E. P., and Ewins, D. J., 2005, "Method for Analysis of Nonlinear Multiharmonic Vibrations of Mistuned Bladed Discs With Scatter of Contact Interface Characteristics," ASME J. Turbomach., **127**, pp. 128–136.

[18] Petrov, E. P., and Ewins, D. J., 2005, "Mistuning Effects on Forced Response of Bladed Discs With Friction Dampers," *Proceedings of NATO Symposium: "Evaluation, Control and Prevention of High Cycle Fatigue in Gas Turbine Engines for Land, Sea and Air Vehicles*," Granada, Spain, Oct. 3–5, Paper No. 38–1.

[19] Petrov, E. P., and Ewins, D. J., 2003, "Analytical Formulation of Friction Interface Elements for Analysis of Nonlinear Multi-Harmonic Vibrations of Bladed Discs," ASME J. Turbomach., **125**, pp. 364–371.

[20] Petrov, E. P., and Ewins, D. J., 2006, "Effects of Damping and Varying Contact Area at Blade-Disc Joints in Forced Response Analysis of Bladed Disc Assemblies," ASME J. Turbomach., **128**, pp. 403–410.

[21] Elliott, R., Green, J. S., and Seinturier, E., 2005, "Aeroelastic Design of Turbine Blades-ADTurB II Overview," *Sixth European Turbomachinery Conference*, Lille, France, March 7–11, Paper No. AMP-105-01/62.

A Sensitivity-Based Method for Direct Stochastic Analysis of Nonlinear Forced Response for Bladed Disks With Friction Interfaces

E. P. Petrov
Imperial College London,
Centre of Vibration Engineering,
Mechanical Engineering Department,
South Kensington Campus,
London SW7 2AZ, UK
e-mail: y.petrov@imperial.ac.uk

An efficient method is developed to calculate stochastic and uncertainty characteristics of forced response for nonlinear vibrations of bladed disks with friction and gap contact interfaces. Uncertainty ranges, statistical characteristics, and probability density functions for forced response levels are determined directly without any sampling procedure. The method uses approximations of the forced response level based on derived analytically and calculated extremely fast and accurately sensitivity coefficients of forced response with respect to friction contact interface parameters. The method effectiveness allows analysis of strongly nonlinear vibration of bladed disks using realistic large-scale finite element models. The method is implemented in a program code developed at Imperial College and numerical examples of application of the method for stochastic analysis of a realistic blisc with underplatform dampers are provided.
[DOI: 10.1115/1.2772634]

Introduction

Forced response of bladed disks with friction contact interfaces is usually strongly nonlinear due to the action of friction forces, gap closing and opening, variations of the contact area, unilateral and impact interactions, etc. The nonlinear forced response depends significantly on the values of key parameters of these contact interfaces, such as friction coefficient, gaps, interferences, stiffness properties of contact surface roughness, normal stress distribution, and others. The contact interface parameters always have some degree of uncertainty in their values, and this is due to (i) an inevitable scatter in the design parameters which can differ from their nominal values within some ranges determined by manufacturing tolerances; (ii) variation of the design parameters during lifetime and uncertainty of effects of operating conditions on design parameters, e.g., effects of wear, oxidation, aging, variation of temperature fields, etc.; (iii) variation of the contact parameters during variable operation conditions; and (iv) uncertainty in experimentally determined values of contact parameters such as friction coefficient and contact stiffness coefficients, and from other causes. Moreover, there is an uncertainty in the excitation forces applied to a structure.

Because of this, there is a pressing practical need to estimate the effects of uncertainty in contact interface parameters and excitation on the forced response of strongly nonlinear structures. Different aspects of uncertainty analysis in structural dynamics and in other technical fields are presented, for example, in Refs. [1–3]. In Refs. 4 and 5 the effect of manufacturing variations on the modal characteristics and resonance response was investigated and sensitivity coefficients of modal properties to blade geometry variation were used in conjunction with Monte Carlo simulations. In Refs. [6–10] different approximations are used to facilitate statistical analysis of effects of blade mistuning on forced response. A closed-form probability function of forced response for a mistuned bladed disk is derived in Ref. [11]. A review of tools for probabilistic confidence interval estimation and an original sampling procedure aimed at reduction of the confidence intervals are presented in Ref. [12]. A review of some modern approximation techniques used in stochastic analysis is given in Ref. [13].

The existing techniques for calculation of estimates for the stochastic characteristics of the forced response with uncertain design parameters and excitation levels are, in most cases, based on a multitude of numerical experiments, such as Monte Carlo simulation and methods derived from it. Therefore, they are often very numerically inefficient, which makes their use prohibitively expensive in practical applications where models containing a large number of degrees of freedom (DOFs) are customarily used. The computational expense becomes especially large for the case of strongly nonlinear vibrations and when estimates for low probability cases corresponding to "tails" in probability density functions have to be obtained.

There were no methods allowing investigations of strongly nonlinear, nonsmooth dynamics of gas-turbine structures with gaps, interferences, and unilateral and friction contact interactions.

In this paper an efficient method is developed to obtain reliable estimates of uncertainty characteristics for nonlinear forced response predictions of bladed disks with friction contact interfaces, caused by uncertainty of the contact interface parameters.

The method is aimed at the analysis of large-scale finite models of components of gas-turbine engines, such as bladed disks, containing, possibly, millions of DOFs. It is based on the determination of sensitivity of the forced response to variations of the parameters analyzed. The sensitivity coefficients of first and second order are calculated and used for constructing explicit expressions for the forced response levels as functions of bladed disk design parameters.

The design parameters analyzed can include gap value, contact stiffness, and friction coefficients of the contact interfaces, underplatform damper mass, damper geometrical characteristics, and others. Although explicit expressions based on the use of the sensitivity coefficients are approximated, they usually provide quite

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 27, 2007; final manuscript received May 24, 2007; published online February 29, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007. Paper No. GT2007-27981.

acceptable accuracy for practical, usually relatively small, ranges of such design parameter uncertainty. These explicit expressions allow calculation of statistical characteristics of the forced response directly as functions of statistical characteristics of the design parameters and excitation.

Probability density functions of forced response are derived analytically for combinations of different types of probability density functions of design parameters simultaneously affecting the structural response.

As a result the methodology proposed avoids any sampling procedure, such as customarily used in Monte Carlo based statistical simulation methods and which requires very large computational times, since the number of numerical experiments required is usually very high, especially when high accuracy of probabilistic distribution is required.

Calculation of the sensitivity-based approximations, statistical characteristics, and probability density functions for forced response does not increase noticeably the computation time required for a single nonlinear forced response calculation. Numerical studies illustrating accuracy and efficiency of the approach are presented for a representative set of problems for a bladed disk with friction underplatform dampers.

Formulation of the Problem

The equation of motion for a structure with nonlinear interfaces can be written in the following form:

$$\boldsymbol{K}q(t)+\boldsymbol{C}\dot{q}(t)+\boldsymbol{M}\ddot{q}(t)+\boldsymbol{f}(\boldsymbol{q}(t),\boldsymbol{b})=\boldsymbol{p}(t) \tag{1}$$

where $\boldsymbol{q}(t)$ is a vector of displacements for all DOFs in the structure considered; $\boldsymbol{K}$, $\boldsymbol{C}$, and $\boldsymbol{M}$ are stiffness, viscous damping, and mass matrices used for the description of linear forces; and $\boldsymbol{p}(t)$ is a vector of periodic excitation forces, which have a period, T, and a corresponding to this period primary frequency, $\omega=2\pi/T$.

When a structure has to be considered in a noninertial coordinate system, as it is usually a case for bladed disks and other structures mounted on a rotating rotor, the stiffness matrix includes terms accounting for the rotation effects. These rotation effect terms comprise a geometric stiffness matrix describing stiffening effects of the centrifugal forces; $\boldsymbol{K}_g(\Omega)$, and a spin-softening matrix describing stiffness softening due to the changing direction of the centrifugal forces under vibration, $\boldsymbol{M}_\omega$. As a result we have $\boldsymbol{K}(\Omega)=\boldsymbol{K}_e+\boldsymbol{K}_g(\Omega)-\Omega^2\boldsymbol{M}_\omega$, where $\boldsymbol{K}_e$ is a conventional elastic stiffness matrix and Ω is the rotation speed. In Eq. (1) $\boldsymbol{f}(\boldsymbol{q}(t),\boldsymbol{b})$ is a vector of nonlinear forces applied at friction contact interfaces of a bladed disk. This vector depends on displacements at the interface nodes, $\boldsymbol{q}(t)$, and on a vector of parameters, $\boldsymbol{b}$, which characterize all friction contact interfaces included in the structure. The contact interface parameters can be gap, friction coefficient, contact stiffness coefficients, normal stresses which are due to the action of the centrifugal forces and temperature fields, and others. In modern industrial problems large-scale finite element models containing 10^5-10^7 DOFs are routinely used. An example of a practical bladed disk with friction contact interfaces is shown in Fig. 1.

In studies of bladed disks steady-state periodic forced response analysis is of primary interest. A maximum displacement, $\boldsymbol{a}(\omega,\boldsymbol{b})$, determined for a primary excitation frequency, ω, and corresponding to a steady-state solution, $\boldsymbol{q}^*(t,\omega,\boldsymbol{b})$, found from the nonlinear equation of motion, Eq. (1), is dependent on values of the contact parameters, i.e.,

$$\boldsymbol{a}(\omega,\boldsymbol{b})=\max_{t\in[0,T]}\boldsymbol{q}^*(t,\omega,\boldsymbol{b}) \tag{2}$$

There is always some degree of uncertainty in values of the friction contact interfaces and other design parameters of practical applications, which produces uncertainty in the forced response. In some cases these parameter uncertainties can be quantified by probability density functions, by mean, variance, and covariance values, and by other statistical characteristics. In other cases only estimates of ranges within which the parameters can take values can be provided.

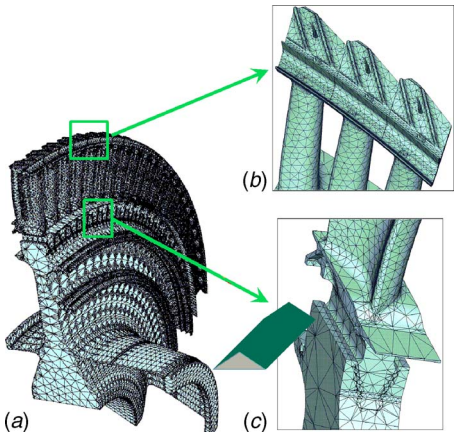


Fig. 1 An example of a structure with friction contact interfaces: (*a*) a bladed-disk assembly, (*b*) friction contact between blade shrouds, and (*c*) friction underplatform dampers

In all these cases, there is a need to have an efficient method allowing calculation of the stochastic characteristics and ranges of uncertainty in the forced response as a function of characteristics of design parameter uncertainty (see Fig. 2).

Analysis of Stochastic and Uncertainty Characteristics for Forced Response

Sensitivity-Based Approximations. If the sensitivity of the forced response to variation of the contact parameters are known, then it is possible to write expressions of the forced response levels directly as functions of the design parameters using Taylor series approximation. For brevity, the dependency of the response on the primary frequency ω is assumed implicitly in further expressions. The first-order approximation for the forced response takes the following form:

$$\underset{(N\times 1)}{\boldsymbol{a}_{\mathrm{I}}}(\boldsymbol{b})=\underset{(N\times 1)}{\boldsymbol{a}}(\boldsymbol{b}_0)+\left[\underset{(N\times n)}{\frac{\partial\boldsymbol{a}(\boldsymbol{b}_0)}{\partial\boldsymbol{b}}}\right]\underset{(n\times 1)}{\Delta\boldsymbol{b}} \tag{3}$$

and the second-order approximation takes the form

$$\underset{(N\times 1)}{\boldsymbol{a}_{\mathrm{II}}}(\boldsymbol{b})=\underset{(N\times 1)}{\boldsymbol{a}}(\boldsymbol{b}_0)+\left[\underset{(N\times n)}{\frac{\partial\boldsymbol{a}(\boldsymbol{b}_0)}{\partial\boldsymbol{b}}}\right]\underset{(n\times 1)}{\Delta\boldsymbol{b}}+\frac{1}{2}\underset{(n\times 1)}{\Delta\boldsymbol{b}}^{T}\left[\underset{(n\times N\times n)}{\frac{\partial^2\boldsymbol{a}(\boldsymbol{b}_0)}{\partial\boldsymbol{b}^2}}\right]\underset{(n\times 1)}{\Delta\boldsymbol{b}} \tag{4}$$

where $\Delta\boldsymbol{b}=(\boldsymbol{b}-\boldsymbol{b}_0)$, N is the total number of DOFs in the structure, n is the number of design parameters, $\partial\boldsymbol{a}(\boldsymbol{b}_0)/\partial\boldsymbol{b}$ and $\partial^2\boldsymbol{a}(\boldsymbol{b}_0)/\partial\boldsymbol{b}^2$ are the first-order and second-order sensitivity coefficients of the forced response to parameter variation, and $\boldsymbol{b}_0$ is the vector of

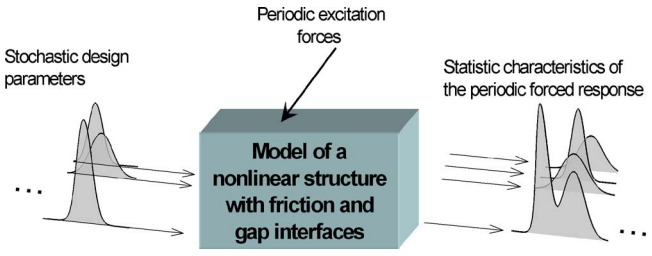


Fig. 2 Establishing dependencies of the statistic characteristics of the forced response on those of the design parameters

nominal parameter values for which the forced response and sensitivity analysis are performed.

It should be noted that in many cases these approximations can be sufficiently accurate when applied in relatively small parameter variation ranges resulting from the technological imperfections and operational uncertainties discussed in the previous section.

In practical applications, some subset of DOFs is usually selected for stochastic analysis, e.g., those DOFs which are crucial for assessment of forced response and stress levels and, hence, high cycle fatigue and durability of a structure. For simplicity and without loss of generality the approach developed for the assessment of stochastic characteristics of the forced response is demonstrated further for an example of forced response, $a(\omega,\mathbf{b})$, obtained for one DOF arbitrarily selected from vector $\mathbf{a}(\omega,\mathbf{b})$. The expressions derived can be applied simultaneously to a multitude of DOFs and the total number of DOFs, for which the analysis is performed, is not restricted but determined by practical needs.

Dimensionless Sensitivity-Based Forced Response Approximations. The friction contact parameters are measured in different units and their magnitudes can differ significantly. Moreover, the sensitivity coefficients are dependent on the forced response level and it is often necessary to assess the sensitivity of the forced response for different levels of the forced response. Hence, there is some difficulty in comparing absolute values of the sensitivity coefficients. It is often more convenient to calculate dimensionless sensitivity parameters in order to characterize the sensitivity of the forced response. The dimensionless first- and second-order sensitivity coefficients, $s_j^{\%}=\partial a_{\%}/\partial b_j^{\%}$ and $s_{ij}^{\%}=\partial^2 a_{\%}/\partial b_i^{\%}\partial b_i^{\%}$, are expressed through absolute values of the sensitivity coefficients, $s_j=\partial a/\partial b_j$ and $s_{ij}=\partial^2 a/\partial b_j\partial b_j$, in the form

$$s_j^{\%}=(b_j^{(0)}/a(\boldsymbol{b}_0))s_j \quad \text{and} \quad s_{ij}^{\%}=(10^{-2}b_i^{(0)}b_j^{(0)}/a(\boldsymbol{b}_0))s_{ij} \tag{5}$$

The dimensionless sensitivity coefficients allow constructing approximations for relative variation of the forced response, $\Delta a^{\%}(\boldsymbol{b})=100[a(\boldsymbol{b})-a(\boldsymbol{b}_0)]/a(\boldsymbol{b}_0)$ for cases of first-order and second-order approximations take the forms

$$\Delta a_{\mathrm{I}}^{\%}(\boldsymbol{b})=\sum_{j=1}^{n}s_j^{\%}\Delta b_j^{\%} \tag{6}$$

$$\Delta a_{\mathrm{II}}^{\%}(\boldsymbol{b})=\sum_{j=1}^{n}s_j^{\%}\Delta b_j^{\%}+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}s_{ij}^{\%}\Delta b_i^{\%}\Delta b_j^{\%} \tag{7}$$

where $\Delta b_j^{\%}=100(b_j-b_j^{(0)})/b_j^{(0)}$ are relative parameter variations, and the subscripts I and II indicate here and further quantities related to first- and second-order approximations, respectively.

Calculation of Forced Response Uncertainty Ranges. For this kind of problem, only estimates for ranges of variation of the design parameters are known, which are given by minimum, $\boldsymbol{b}^-$, and maximum, $\boldsymbol{b}^+$, possible values, i.e.,

$$\boldsymbol{b}^-\leqslant\boldsymbol{b}\leqslant\boldsymbol{b}^+=0 \tag{8}$$

To characterize uncertainty in the forced response its possible minimum, a^-, and maximum, a^+, values have to be estimated.

Effects of Individual Parameter Uncertainty. When uncertainty of each of the design parameters is independent of each other, the high and low limits of uncertainty of the forced response, which are due to uncertainty of each of these design parameters, $b_j^{\pm}$, can be assessed directly using the linear or quadratic approximations developed above.

The absolute values of the maximum and minimum forced response levels can be obtained from Eqs. (3) and (4) in the form

$$a_{\mathrm{I}}^{\pm}=a_{\mathrm{I}}(\boldsymbol{b}_0)+s_j\Delta b_j^{\pm} \quad a_{\mathrm{II}}^{\pm}=a_{\mathrm{I}}(\boldsymbol{b}_0)+(s_j+0.5\Delta b_j^{\pm}s_{jj})\Delta b_j^{\pm} \tag{9}$$

The relative uncertainty of the forced response is obtained from Eqs. (6) and (7) as follows:

$$a_{\mathrm{I}}^{\%\pm}=s_j^{\%}\Delta b_j^{\%\pm} \quad a_{\mathrm{II}}^{\%\pm}=(s_j^{\%}+0.5\Delta b_j^{\%\pm}s_{jj}^{\%})\Delta b_j^{\%\pm} \tag{10}$$

Optimization Formulation for Calculation of Overall Forced Response Uncertainty. The overall effect of uncertainties of all design parameters on the forced response uncertainty can be found by minimizing and maximizing accordingly the forced response

$$a^-=a_{(\mathrm{I\ or\ II})}(\boldsymbol{b})\rightarrow\min \quad a^+=a_{(\mathrm{I\ or\ II})}(\boldsymbol{b})\rightarrow\max \tag{11}$$

while satisfying constraints given by Eq. (8). This problem is an optimization problem and its solution can be efficiently made using approximations given by Eqs. (3) and (4).

For a case of the first-order approximation, $a_{\mathrm{I}}(\boldsymbol{b})$, the optimization problem formulated here is a linear programming problem, and for a case of the second-order approximation, $a_{\mathrm{II}}(\boldsymbol{b})$, we have a quadratic programming problem. In both cases the optimization problem has simple constraints and values of the global maximum, a^+, and the global minimum, a^-, can always be efficiently calculated, and in a limited number of iterations, by readily available optimization methods and software.

Calculation of Statistical Characteristics. Probability density functions (PDFs), $p(b_j)$, allow the most full description of the stochastic properties for each jth design parameter. However, in many cases the stochasticity is characterized by a set of some integral characteristics of the probability distributions, especially when the PDFs are not available, as it frequently happens in practical applications. When statistical characteristics of the design parameters are known, then statistical characteristics of the forced response can be explicitly expressed through characteristics of the design parameters and the sensitivity coefficients of the forced response. For example, let us assume that all design parameters are stochastically independent and their expected values, $\mu_j=\langle b_j\rangle=\int_{-\infty}^{-\infty}b_jp(b_j)dp_j$, together with their standard deviations, $\sigma_j^2=\langle(b_j-\mu_j)^2\rangle=\int_{-\infty}^{-\infty}(b_j-\mu_j)^2p(b_j)dp_j$, are known for each jth parameter. It is convenient to construct the approximations for the forced response in the vicinity of the expected values of the parameters, i.e., $\boldsymbol{b}_0=\boldsymbol{\mu}=\{\mu_1,\mu_2,\ldots,\mu_n\}^T$. Then, corresponding statistical characteristics for the forced response can be obtained for the expected value of the forced response in the form

$$\mu_{\mathrm{I}}^a=\langle a_{\mathrm{I}}\rangle=a(\boldsymbol{\mu}) \tag{12}$$

$$\mu_{\mathrm{II}}^a=\langle a_{\mathrm{II}}\rangle=a(\boldsymbol{\mu})+\frac{1}{2}\sum_{j=1}^{n}s_{jj}\sigma_j^2 \tag{13}$$

and for standard deviation of the forced response,

$$(\sigma_{\mathrm{I}}^a)^2=\langle(a_{\mathrm{I}}-\mu_{\mathrm{I}}^a)^2\rangle=\sum_{j=1}^{n}(s_j\sigma_j)^2 \tag{14}$$

$$(\sigma_{\mathrm{II}}^a)^2=\sum_{j=1}^{n}\left[s_j^2\sigma_j^2+s_js_{jj}\langle(b_j-\mu_j)^3\rangle+\frac{s_{jj}^2}{4}\langle(b_j-\mu_j)^4\rangle\right]-\frac{1}{2}\sum_{j=1}^{n}s_{jj}^2\sigma_j^4+\frac{1}{4}\sum_{i,j=1}^{n}s_{ij}^2\sigma_i^2\sigma_j^2 \tag{15}$$

Determination of the Probability Distribution for the Forced Response Levels. In some cases there are enough statistical data to obtain functions of the PDFs for the stochastic design parameters, $p_j(b_j)$. A function of the probability that the forced response level does not exceed some required value, α, takes the following form:

$$P(a(\boldsymbol{b}) < \alpha) = \int_{-\infty, a<\alpha}^{\infty} p_1(b_1)p_2(b_2)\cdots p_n(b_n)db_1db_2\cdots db_n \tag{16}$$

where the design parameters are assumed to be statistically independent.

Then this function, $P(a(\boldsymbol{b})<\alpha)$, and its PDF, $p(a(\boldsymbol{b})<\alpha)$, can be derived analytically through the PDFs of design parameters by the way described below.

It is evident that the condition $a(\boldsymbol{b})<\alpha$ can be satisfied by imposing special constraints on design parameters. Selection of design parameters can be made arbitrarily and these constraints are derived using the sensitivity-based approximations. For example, the linear sensitivity-based approximation for $a(\boldsymbol{b})$ given by Eq. (3) can be resolved with respect to parameter b_1. Then, the following constraint can be derived, which allows condition $a(\boldsymbol{b})<\alpha$ to be satisfied:

$$b_1(\alpha, b_2, \ldots, b_n) = b_1^0 + \left(\alpha - a_0 - \sum_{j=2}^{n} s_j(b_j - b_j^0)\right)/s_1 > \alpha \tag{17}$$

This inequality constraint ensures that condition $a(\boldsymbol{b})<\alpha$ is satisfied for a case of $s_1>0$. For a case of $s_1<0$, the inequality sign $>$ in Eq. (17) should be alternated by $<$. Therefore, Eq. (16) can be written in the following form:

$$P(\alpha) = \int_{-\infty}^{\infty} p_n(b_n)\cdots\int_{-\infty}^{b_1(\alpha,b_2,\ldots,b_n)} p_1(b_1)db_1\cdots db_n \quad \text{for } s_1 > \alpha \tag{18}$$

$$P(\alpha) = \int_{-\infty}^{\infty} p_n(b_n)\cdots\int_{b_1(\alpha,b_2,\ldots,b_n)}^{\infty} p_1(b_1)db_1\cdots db_n \quad \text{for } s_1 < \alpha \tag{19}$$

The PDF for the forced response level, which does not exceed value, α, can then be determined as

$$p(\alpha) = |s_1|^{-1}\int_{-\infty}^{\infty} p_n(b_n)\cdots\int_{-\infty}^{\infty} p_1(b_1(\alpha, b_2, \ldots, b_n))db_2\cdots db_n \tag{20}$$

When the number of the random parameters involved in Eq. (20) is not large, it is possible to evaluate such integrals analytically.

Case of One Stochastic Parameter. For a case of when there is only one stochastic parameter, a PDF of the forced response level can be expressed through a PDF of the stochastic parameter in the form

$$p(\alpha) = |\partial a/\partial b_1|^{-1}p_1(b_1^0 + (\alpha - a_0)/s_1) \tag{21}$$

Case of Two Stochastic Parameters. For a case of two stochastic design parameters, the analytical expression for the forced response PDF can also be derived analytically for a large variety of combinations of most commonly used stochastic distributions of parameters. As a result the PDFs of forced response can be defined by explicit formulas. This facilitates significantly analysis of a structure by avoiding completely a need of very time-consuming Monte Carlo simulations. Some examples of probability forced response density functions derived for given probability distributions of design parameters are shown below.

(1) A case when the first parameter has the normal distribution and the second parameter is uniformly distributed, i.e., their PDFs are represented as follows:

$$p_1(b_1) = (\sqrt{2\pi}\sigma_1)^{-1}e^{-(b_1-\mu_1)^2/2\sigma_1^2} \tag{22}$$

$$p_2(b_2) = \begin{cases} 1/(2\sqrt{3}\sigma_2) & \text{for } \mu_2 - \sqrt{3}\sigma_2 \leqslant b_2 \leqslant \mu_2 + \sqrt{3}\sigma_2 \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

then we derive an expression for a PDF of the forced response, which takes the following form:

$$p(\alpha) = \frac{\sqrt{3}s_1}{12\tilde{\sigma}_2|s_1|}\left(\text{erf}\left(\frac{a_0 + \sqrt{3}\tilde{\sigma}_2 - \alpha}{\sqrt{2}\tilde{\sigma}_1}\right) - \text{erf}\left(\frac{a_0 - \sqrt{3}\tilde{\sigma}_2 - \alpha}{\sqrt{2}\tilde{\sigma}_1}\right)\right) \tag{24}$$

where $\tilde{\sigma}_j = s_j\sigma_j$.

(2) A case, when the first parameter has normal distribution, given by Eq. (22), and the second parameter has Laplace distribution, i.e.,

$$p_2(b_2) = \frac{1}{\sqrt{2}\sigma_2}e^{-\sqrt{2}|b_2-\mu_2|/\sigma_2} \tag{25}$$

It can be shown that the PDF of the forced response takes the form

$$p(\alpha) = \frac{s_1 e^{2(\tilde{\sigma}_1^2 - \sqrt{2}\tilde{\sigma}_2(a_0+\alpha)/\tilde{\sigma}_2^2)}}{2\sqrt{2}\tilde{\sigma}_2|s_1|}(e^{2\sqrt{2}\alpha/\tilde{\sigma}_2}\vartheta_+ + e^{2\sqrt{2}\alpha/\tilde{\sigma}_2}\vartheta_-) \tag{26}$$

where $\vartheta_\pm = \text{sgn}(s_1 s_2) \mp \text{erf}((\sqrt{2}\tilde{\sigma}_2(\alpha - a_0) \pm 2\tilde{\sigma}_1^2)/\tilde{\sigma}_1\tilde{\sigma}_2)$.

Case of Larger Numbers of Stochastic Parameters. In cases when probability distribution for all design parameters is Gaussian normal, the forced response PDF is also Gaussian. The expected value and the standard deviation of this Gaussian normal distribution can be calculated by the formulas given by Eqs. (12)–(15).

In the cases when the analytical derivation is difficult, a numerical evaluation of the PDF can be made by generating random numbers with specified probability distribution functions for values of all stochastic design parameters. For evaluation of the forced response corresponding to these random values the approximation expressions of Eqs. (3) and (4) can be used. Therefore, these calculations do not require much computational expense even when high-accuracy determination of resulting forced response probability functions is required and, therefore, the forced response needs to be evaluated for 10^5–10^7 different parameter values for each excitation frequency of interest.

Sensitivity Analysis of Nonlinear Forced Response

The crucial point in the analysis of the characteristics of uncertainty of the forced response is a capability of fast and accurate calculation of the sensitivity of the forced response to variation of the friction contact interface parameters. A method providing this capability is discussed below.

Multiharmonic Balance Solution of the Nonlinear Equations of Motion. For a search of the periodic vibration response, the displacements are represented as a restricted Fourier series, which contain whichever harmonic components are necessary to approximate the sought solution, i.e.,

$$\boldsymbol{q}(t) = \boldsymbol{Q}_0 + \sum_{j=1}^{n}(\boldsymbol{Q}_j^{(c)}\cos m_j\omega t + \boldsymbol{Q}_j^{(s)}\sin m_j\omega t) \tag{27}$$

where $\boldsymbol{Q}_j^{(c)}$ and $\boldsymbol{Q}_j^{(s)}$ $(j=1,\ldots,n)$ are vectors of cosine and sine harmonic coefficients marked by superscripts $^{(c)}$ and $^{(s)}$ accordingly; $\boldsymbol{Q}_0$ is a vector of constant components of the displacements and m_j $(j=1,\ldots,n)$ are specific numbers of harmonics that are kept in the displacement expansion in addition to the constant component. Equation (27) and similar expansions obtained for the nonlinear forces, $\boldsymbol{f}(\boldsymbol{q}(t),\lambda)$, and for the periodic excitation forces,

$\boldsymbol{p}(t)$, are substituted into the equation of motion, Eq. (1), and equating coefficients from the same harmonics gives a nonlinear multiharmonic equation of motion, which contains all harmonic coefficients for all DOFs included in the model. Using methods developed in Refs. [14,15] the equation of motion can be obtained in the following form:

$$\boldsymbol{R}(\boldsymbol{Q},\boldsymbol{b}) = \boldsymbol{Q}(\boldsymbol{b}) - \tilde{\boldsymbol{Q}} + \boldsymbol{A}(\omega)\boldsymbol{F}(\boldsymbol{Q},\boldsymbol{b}) = 0 \tag{28}$$

where $\boldsymbol{Q}=\{\boldsymbol{Q}_0,\boldsymbol{Q}_1^{(c)},\boldsymbol{Q}_1^{(s)},\ldots,\boldsymbol{Q}_n^{(s)}\}^T$ is a sought vector of harmonic coefficients for DOFs where nonlinear forces are applied and $\tilde{\boldsymbol{Q}}=\{\tilde{\boldsymbol{Q}}_0,\tilde{\boldsymbol{Q}}_1^{(c)},\tilde{\boldsymbol{Q}}_1^{(s)},\ldots,\tilde{\boldsymbol{Q}}_n^{(s)}\}^T$ is a vector of harmonic coefficients of the forced response calculated for a structure with all nonlinear contact interface forces assumed to be zero. $\boldsymbol{A}(\omega)=\mathrm{diag}[\boldsymbol{A}(0),\boldsymbol{A}_1(m_1\omega),\ldots,\boldsymbol{A}_n(m_n\omega)]$ is a multiharmonic frequency response function (FRF) matrix of the linear structure, which is combined from FRF matrices for all harmonics included into the multiharmonic expansion. $\boldsymbol{F}(\boldsymbol{Q},\boldsymbol{b})=\{\boldsymbol{F}_0,\boldsymbol{F}_1^{(c)},\boldsymbol{F}_1^{(s)},\ldots,\boldsymbol{F}_n^{(s)}\}^T$ is a vector of nonlinear contact forces, which is dependent on the displacements, $\boldsymbol{Q}$, and on a vector of friction contact parameters. It is important to note that all DOFs, which do not have nonlinear interaction forces, are excluded here without loss of the model accuracy.

Solution of Eq. (28) is performed using the Newton-Raphson method. For efficient calculation of the solutions over some range of parameter variation (such as excitation frequency, gaps, friction coefficient, etc.) so-called "continuation," solution tracing methods are applied (see, e.g., Ref. [15]). These methods provide an efficient choice of the step size for the parameter variation, together with the calculation of solution approximations for each step, and ensure tracing of the solution belonging to the same solution branch from manifold of solutions that are usually inherent for an essentially nonlinear structure. In accordance with the Newton-Raphson method the solution vector, $\boldsymbol{Q}^*$, is obtained from the following iterative process:

$$\boldsymbol{J}(\boldsymbol{Q}^{(k)})\Delta\boldsymbol{Q} = \boldsymbol{R}(\boldsymbol{Q}^{(k)}) \quad \boldsymbol{Q}^{(k+1)} = \boldsymbol{Q}^{(k)} + \Delta\boldsymbol{Q} \tag{29}$$

where $\boldsymbol{J}=\partial\boldsymbol{R}/\partial\boldsymbol{Q}=\boldsymbol{I}+\boldsymbol{A}(\omega)\partial\boldsymbol{F}/\partial\boldsymbol{Q}$ is the Jacobian of the nonlinear equation (28), $\boldsymbol{I}$ is the identity matrix, and superscript (k) indicates the iteration number.

Analytical Expressions for First- and Second-Order Sensitivity Coefficients for Multiharmonic Forced Response. First- and second-order sensitivity coefficients with respect to the contact parameters, b_j, i.e., $\partial\boldsymbol{Q}/\partial b_j$ and $\partial^2\boldsymbol{Q}/\partial b_i\partial b_j$, can be obtained analytically by differentiating Eq. (28). Resulting expressions for the determination of the sensitivity coefficients take the form

$$\boldsymbol{J}\frac{\partial\boldsymbol{Q}}{\partial b_j} = -\boldsymbol{A}(\omega)\frac{\partial\boldsymbol{F}}{\partial b_j} \tag{30}$$

$$\boldsymbol{J}\frac{\partial^2\boldsymbol{Q}}{\partial b_i\partial b_j} = -\boldsymbol{A}(\omega)\left(\frac{\partial^2\boldsymbol{F}}{\partial b_i\partial b_j} + \frac{\partial}{\partial b_i}\frac{\partial\boldsymbol{F}}{\partial\boldsymbol{Q}}\frac{\partial\boldsymbol{Q}}{\partial b_j}\right) - \frac{\partial\boldsymbol{J}}{\partial b_i}\frac{\partial\boldsymbol{Q}_n}{\partial b_j} \tag{31}$$

where

$$\frac{\partial\boldsymbol{J}}{\partial b_j} = \boldsymbol{A}(\omega)\left[\frac{\partial}{\partial b_j}\frac{\partial\boldsymbol{F}}{\partial\boldsymbol{Q}} + \frac{\partial}{\partial\boldsymbol{Q}}\frac{\partial\boldsymbol{F}}{\partial\boldsymbol{Q}}\frac{\partial\boldsymbol{Q}}{\partial b_j}\right] \tag{32}$$

The Jacobian $\boldsymbol{J}$ needed for the sensitivity evaluation is obtained as a by-product of the forced response analysis and, moreover, LR factorization of the Jacobian is calculated during the last Newton-Raphson iteration. Owing to this, the calculation of the sensitivity coefficients does not incur a significant computation cost since this calculation requires the solution of only one additional linear algebraic equation with already factorized matrix. The right-hand part of Eq. (30) contains derivatives of the friction contact interface forces and stiffness matrices. Their accurate and effective calculation is a very important problem, which is discussed in the following section.

Table 1 Forces of the friction contact interaction

Status	Condition of the contact state alternation	Tangential force, f_x	Normal force, f_y
Stick	$f_x(\tau)=\pm\mu f_y(\tau)$	$k_t(x-x(\tau_j))\pm\mu f_y(\tau_j)$	$N_0+k_n y$
Slip	$\pm k_t\dot{x}(\tau)=\mu k_n\dot{y}(\tau)$	$\pm\mu f_y$	
Separation	$N_0+k_n y(\tau)=0$	0	0

Nonlinear Friction and Unilateral Contact Modeling Allowing Exact Sensitivity Coefficient Calculation. Calculation of the first-order sensitivity of the forced response requires determination of the sensitivity of the nonlinear contact interaction forces to variation of the friction contact parameters, $\partial\boldsymbol{F}/\partial\lambda$, together with the tangent stiffness matrix, $\partial\boldsymbol{F}/\partial\boldsymbol{Q}_n$, and nonlinear contact interaction forces, $\boldsymbol{F}(\boldsymbol{Q}_n)$. All these expressions have been derived analytically in Refs. [15–17] for a general case of a friction contact with variable normal load. Clearances and interferences are also included into the contact models developed there and unilateral character of the interaction forces acting along the normal direction to the contact surface is accounted for. Parameters of the friction contact interfaces considered in those papers include (i) a clearance/interference value, g; (ii) normal load/stress, N_0; (iii) friction coefficient, μ; (iv) normal stiffness of the contact surface, k_n; and (v) tangential stiffness of the contact surface, k_t.

Forces occurring at the friction contact interface are expressed in terms of the relative displacement along a direction tangential to the contact surface, $x(\tau)$, and a normal relative displacement, $y(\tau)$. Expressions for tangential and normal components of the interaction forces for all possible contact conditions are shown in Table 1.

Conditions for alternation of the contact conditions are given in the second column of Table 1. These conditions determine time instants, τ_j, when the contact condition changes from stick to slip or from being in contact to separation and back.

For the determination of the second-order sensitivity coefficients a vector of second-order derivatives of the nonlinear interface forces, $\partial^2\boldsymbol{F}/\partial b_j^2$, matrix of derivatives of the nonlinear interface tangent stiffness matrix with respect to the parameter analyzed, $\partial(\partial\boldsymbol{F}/\partial\boldsymbol{Q}_n)/\partial b_j$, and a matrix of derivatives of the product of the tangent stiffness matrix and first-order sensitivity coefficients, $\partial[\partial\boldsymbol{F}/\partial\boldsymbol{Q}_n\partial\boldsymbol{Q}_n/\partial b_j]/\partial\boldsymbol{Q}_n$, with respect to the harmonic coefficients of the displacements have to be determined.

These vectors and matrices have been derived analytically and obtained in an explicit form in Ref. [17]. Owing to the analytical derivation for all matrices necessary for the calculation of first- and second-order sensitivity coefficients, they are calculated extremely fast and with very high accuracy. As a result the time needed for the calculation of the right-hand parts of Eqs. (30) and (31) is negligible compared with the computational expense necessary for forced response calculation, and in practical calculations the sensitivity analysis does not cause a noticeable increase of calculation time.

Test Case Studies

As an example, forced response of an ADTurbII blisc (see Ref. [18]) with cottage-roof (CR) underplatform dampers (UPDs) is analyzed. The blisc consists of 24 blades, and a finite element model of its one sector contains 21,555 DOFs. The UPD is modeled by a damper model developed in Ref. [19]. Finite element (FE) blisc and UPD models are shown in Fig. 3. In the numerical studies presented here a cyclically symmetric bladed disk is considered and it is assumed that the stochastic variation of UPD parameter keeps the structure cyclically symmetric. As a result of this assumption, specific effects caused by blade and UPD mistuning (e.g., see Ref. [20]) are not considered here.

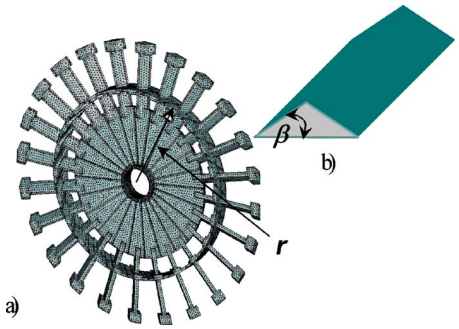


Fig. 3 FE model of a blisc with a CR damper

The UPD model is formulated for a general case of multiharmonic forced response analysis. It allows for inertia forces occurring due to damper vibration and effects of normal load variation on stick-slip transitions at the contact interfaces, including contact separation. The UPD is characterized by the following parameters: (i) damper mass, $m_{\mathrm{UPD}}=2.6g$; (ii) friction coefficient, $\mu=0.3$; (iii) distance of the UPD inertia center from the rotation axis, $r=137$ mm; and (iv) damper angle, $\beta=30$ deg (see Figs. 3(*a*) and 3(*b*)).

These UPD parameters allow determination of inertial forces: static centrifugal and dynamic forces, and, moreover, static and dynamic friction contact interaction forces at damper contact surfaces the effect of variation of the centrifugal forces with variation of the rotor rotation speed is also included in the analysis.

The natural frequencies–nodal diameters diagram for this bladed disk is given in Ref. [20]. The background damping due to damping in the blisc material is assumed to be low with the damping loss factor: $\eta=7.5\times10^{-5}$. A traveling wave excitation pattern of a 19th engine order is studied and first five odd harmonics are used for the multiharmonic representation of the forced response, i.e., 19th, 57th, 95th, 133rd, and 171st.

Sensitivity of Forced Response to Contact Parameters. Forced response levels calculated for a node located at the blade tip are plotted in Fig. 4 for cases of three different UPD mass values. The damper mass of 2.6 g is used here as a 100% value. For comparison, forced response of two limiting cases is also plotted here: (i) forced response of a blisc without dampers and (ii) forced response of a blisc with fully stuck (or welded) dampers.

Dimensionless first-order sensitivity coefficients, $s_j^{\%}$, of forced response levels with respect to all major UPD parameters are shown in Fig. 5. A case of 100% damper mass is considered here. One can see that the forced response is the most sensitive to all

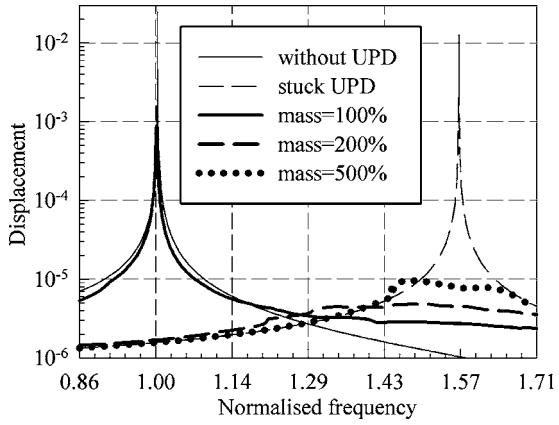


Fig. 4 Forced response of the blisc with UPDs

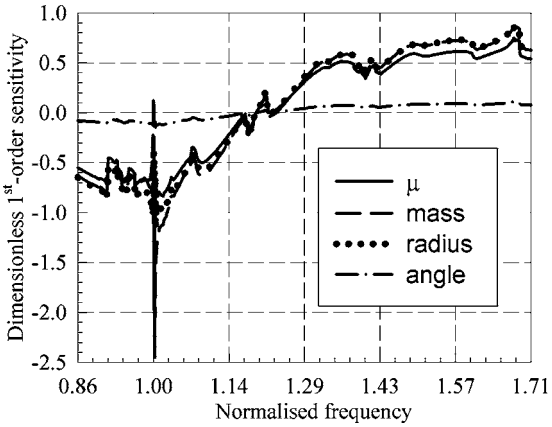


Fig. 5 Dimensionless sensitivity of the forced response to UPD parameters

parameters at the resonance peak. The most influential parameter is the damper mass value: its 1% increase can result in 2.5% reduction of the resonance response levels. 1% increase of friction coefficient can reduce response level by 2.2%, while for damper radius and angle change the response level by 1% and 0.14%, accordingly.

Validation of the Sensitivity-Based Approximations. Accuracy of the sensitivity-based approximations (SBAs) of the forced response has been investigated. Examples illustrating high accuracy of these approximations are shown in Figs. 6 and 7. Forced response curves obtained by two ways are compared here: (i) forced response and sensitivity coefficients are calculated for nominal parameter values and then first-order SBAs given by Eq. (3) are used to obtain forced response curves for new parameter values and (ii) forced response is calculated directly for new pa-

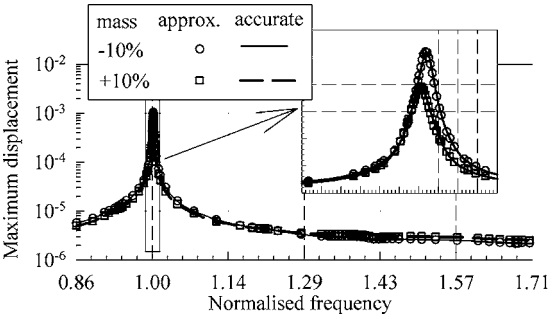


Fig. 6 Comparison of forced responses obtained by SBAs with accurate ones: a case of damper mass value variation

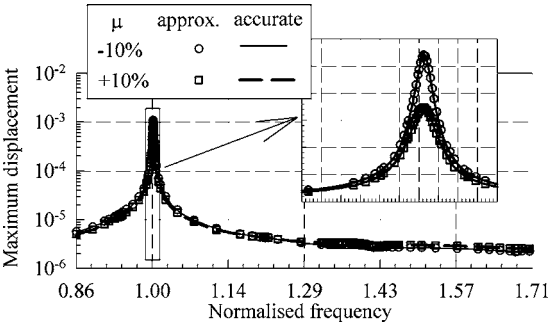


Fig. 7 Comparison of forced responses obtained by SBAs with accurate ones: a case of friction coefficient variation

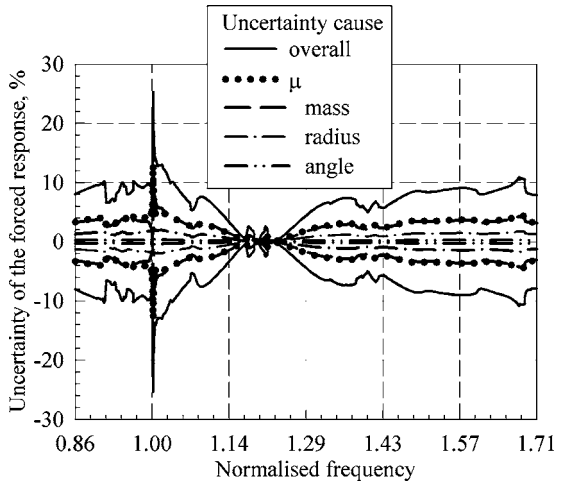


Fig. 8 Uncertainty of forced response caused by damper parameter uncertainties: a case of 100% damper mass

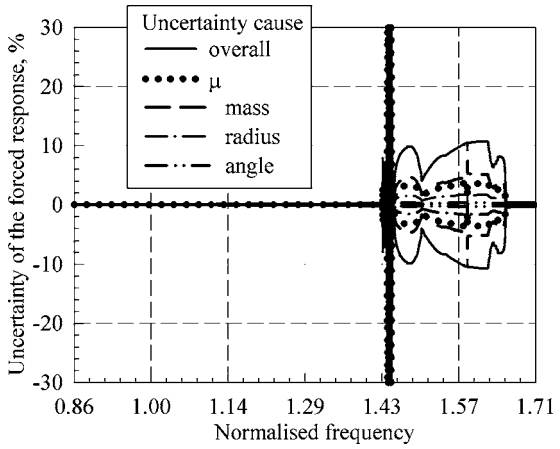


Fig. 10 Uncertainty of forced response caused by damper parameter uncertainties: a case of 500% damper mass

rameter values (see, e.g., Ref. [16]).

The cases of two most influential UPD parameters are considered here: (i) damper mass (Fig. 6) and (ii) friction coefficient (Fig. 7). A range of parameter variation is chosen in both cases to be 10%, which is expected to cover possible scatter/uncertainty of damper parameters for the blisc analyzed for most cases. One can see that the results obtained by first-order SBA are practically indistinguishable from accurate results obtained by explicit calculation for new parameter values over the whole frequency range analyzed and, therefore, these approximations can be used for this bladed disk and they have been applied for calculations of statistical characteristics.

Ranges of Forced Response Uncertainty. The methodology developed above for calculation of estimates of uncertainty for forced response caused by uncertainty of parameters of structure has been applied to calculate uncertainty of the forced response when uncertainty in (i) damper mass is ±6%, (ii) friction coefficient: ±5%, (iii) distance from rotation axis: ±2%, and (iv) damper angle: ±3%. Uncertainty of forced response caused by each of these parameters individually and forced response uncertainty when all these parameters are uncertain simultaneously are shown in Figs. 8–10 for cases of 100%, 200%, and 500% damper mass, respectively.

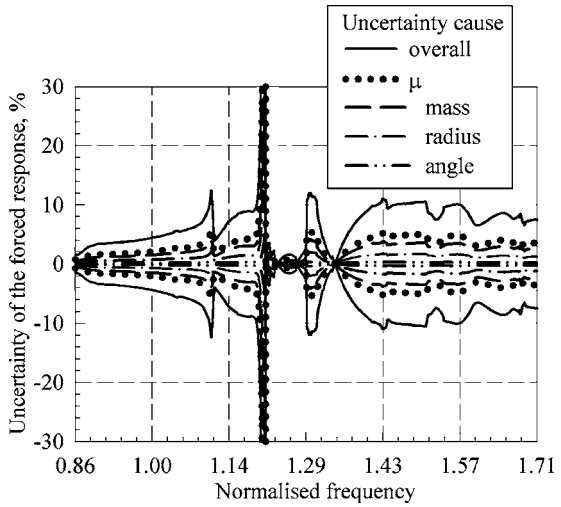


Fig. 9 Uncertainty of forced response caused by damper parameter uncertainties: a case of 200% damper mass

The overall uncertainty was calculated using the optimization formulation developed in this paper, and linear programming problem was solved for each excitation frequency for which the forced response and the sensitivity coefficients are evaluated within the frequency range analyzed.

The forced response uncertainty is plotted as a percentage to forced response level achieved for a current excitation frequency. It is evident that the overall forced response uncertainty is affected by many damper parameters and is significantly larger than the uncertainty caused by each of the parameters individually.

For the case of 100% mass, the largest overall forced response uncertainty of 26% is observed at the resonance frequency. Over the rest of the frequency range analyzed the overall forced response uncertainty is lower than 10%. There is a frequency range (from 1.14 to 1.23) where forced response uncertainty is much smaller than the uncertainty of the damper parameters, which is due to the stuck damper condition observed mostly within this frequency range.

For a case of 200% damper mass, the uncertainty of the forced response is contained within 10% for practically all frequencies including a resonance and a wide near-resonance frequency range observed for this damper mass. The large uncertainty peak observed at a frequency of 1.20 is due to an abrupt change of contact conditions of the damper: at this frequency the damper begins to alternate contact and separation from blade platform during each vibration period. This effect is also reflected by the shape of the forced response curve in Fig. 4, which has an angular shape in the vicinity of this frequency.

For a case of heavier, 500% damper mass uncertainties of the damper parameters in the assumed above ranges do not affect forced response, and provide 0% forced response uncertainty in frequency ranges from 0.86 to 1.43 and from 1.64 to 1.71. This is due to the fact that in these frequency ranges this damper is fully stuck and damper parameter variation in the ranges studied does not affect forced response level. At the normalized frequency, 1.43, first slip-stick transitions occur and a large uncertainty peak is observed. For excitation frequencies higher than 1.43, the overall uncertainty of forced response levels is less than 10%.

Characteristics of Variance of the Forced Response. Standard deviation and coefficient of variance of the forced response were calculated as a function of excitation frequency assuming that all damper parameters have a value 5% for the coefficient of variance. In Fig. 11 calculated coefficient of variance, for the forced response level, σ_a/μ_a, is plotted, which results from damper parameter scatter.

Similar to the uncertainty ranges, calculated in the previous

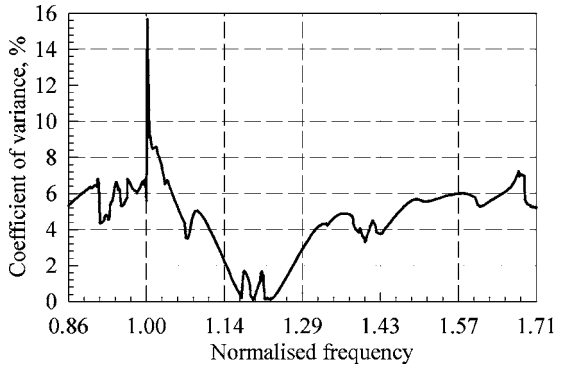


Fig. 11 Coefficient of variance of forced response

section, the forced response coefficient of variance is calculated simultaneously with a calculation of forced response levels and without a noticeable increase of computational time.

Probability Density Functions of the Forced Response. The derived analytical expressions for PDFs of forced response (see Eqs. (24) and (26)), when PDFs of two of the damper parameters are provided, were validated by comparison with PDFs obtained by Monte Carlo simulation. Examples of such comparison are shown in Fig. 12, where PDFs are compared for forced response determined at two different excitation frequencies. The results of Monte Carlo simulation are shown by vertical bar charts (in blue for $f=1.0$, and in red for $f=0.86$) and analytically derived PDFs are plotted as solid and dashed curves.

Two of the most significant damper parameters are assumed to be stochastic: (i) the damper mass value has a uniform PDF within the range of [0.24, 0.36] and has the coefficient of variance 2.9% and (ii) the friction coefficient has a normal PDF with the coefficient of variance 2.0%. For Monte Carlo simulation 10^6 trial points obeying to chosen PDFs were generated and then forced response and the PDF histogram were calculated. In order to facilitate the comparison of the PDFs calculated at different excitation frequencies, and therefore having very different response levels, the normalized forced response and normalized PDFs are plotted here. The normalization is made by dividing the forced response by a mean value of the forced response observed at the excitation frequency considered. Accordingly, the PDFs are multiplied by this forced response mean value.

One can see that the analytically derived curves are very accurate envelopes of the Monte Carlo bar charts and, hence, results of Monte Carlo simulation and results obtained with the analytical formula derived are identical. Again, the calculation of PDFs using the analytical formulas based on approximations using sensitivity coefficient does not require any significant computation expense.

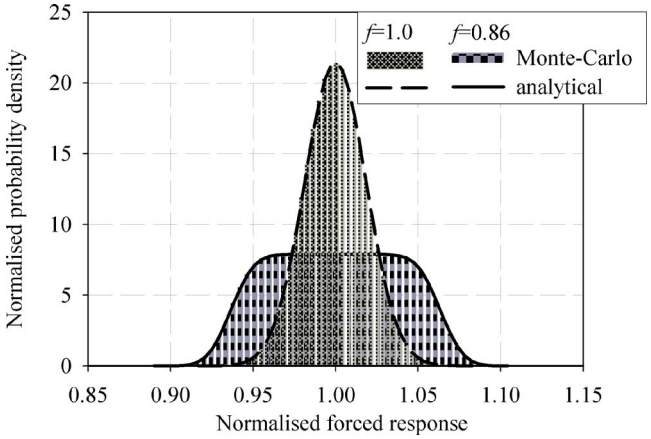


Fig. 12 Comparison of the PDFs determined (i) by the Monte Carlo simulation (bar charts) and (ii) by the derived analytical expressions (solid curves)

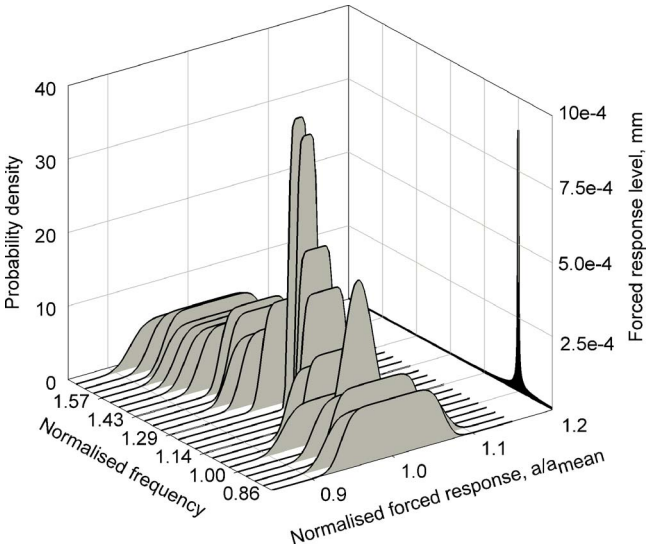


Fig. 13 PDF for different excitation frequency values: normal distribution for friction coefficient and uniform for mass

In Figs. 13 and 14 PDFs of the forced response are plotted for 24 different excitation frequencies uniformly distributed over the frequency range analyzed. Forced response plots are also shown here on the right.

The results plotted in Fig. 13 correspond to damper parameters with the same statistical distributions as in the example above. The results shown in Fig. 14 are obtained for a case when both parameters, damper mass and friction coefficient, are uniformly distributed with a coefficient of variance 2.9%. It is evident that the PDFs of forced response change significantly with variation of excitation frequency.

In Fig. 15 a case when all four parameters of the damper have normal PDF with a coefficient of variance 5%. In contrast to the

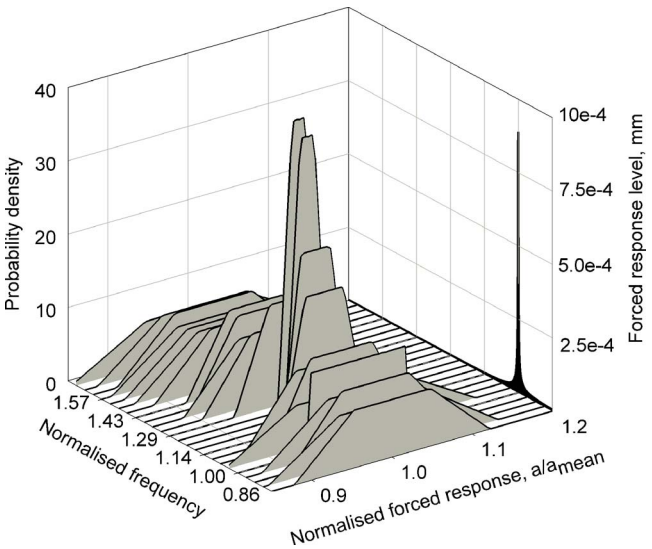


Fig. 14 PDF for different excitation frequency values: uniform distribution for friction coefficient and uniform for mass

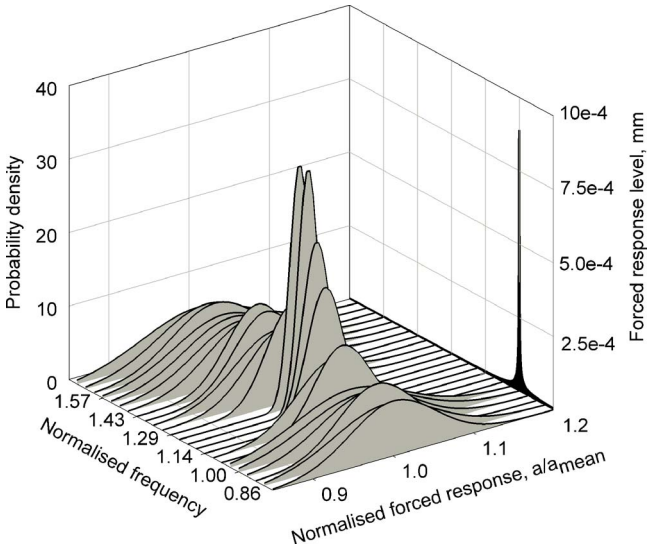


Fig. 15 PDF for different excitation frequency values: normal distribution for all parameters

cases discussed previously, for this case PDF of forced response stays normal, as PDFs of damper parameters are but the standard deviation is varied.

Conclusions

An efficient method has been developed to calculate stochastic and uncertainty characteristics of forced response for essentially nonlinear vibrations of bladed disks. The stochastic and uncertain parameters of friction and gap contact interface are considered.

Uncertainty ranges, coefficients of variance, and PDFs for forced response levels are determined directly without any sampling procedure for a wide frequency range analyzed as functions of friction contact interface parameters for the first time.

The method is based on SBAs of the forced response level constructed over a whole frequency range analyzed. The first- and second-order sensitivity coefficients with respect to friction contact parameters such as friction coefficient, gap, normal stresses, and contact stiffness are derived analytically, which provides extremely fast and accurate calculation of the sensitivity coefficients for strongly nonlinear vibration of large-scale FE models of realistic structures.

The effectiveness of the method allows uncertainty and statistical characteristics of the forced response to be calculated accurately and simultaneously with the calculation of the forced response without a significant increase of the computational effort.

Numerical studies of a blisc with underplatform CR dampers have been performed. Accuracy of the SBA of the forced response for practical ranges of uncertainty of damper parameters has been confirmed. Analysis of overall and individual effects of uncertainty of damper parameters on the uncertainty of forced response levels has been carried out over a large frequency range.

Analytical formulas derived for PDFs of forced response have been validated compared with the results obtained by Monte Carlo simulation and applied for the analysis of forced response PDFs.

Acknowledgment

The author is grateful to Rolls-Royce for providing the financial support for this project and for giving permission to publish this work.

References

[1] Satelli, A., and Scott, M., 1997, "The Role of Sensitivity Analysis in the Corroboration of Models and Its Link to Model Structural and Parametric Uncertainty," Reliab. Eng. Syst. Saf., **57**, pp. 1–4.
[2] Helton, J. C., 1993, "Uncertainty and Sensitivity Analysis Techniques for Use in Performance Assessment for Radioactive Waste Disposal," Reliab. Eng. Syst. Saf., **42**, pp. 327–367.
[3] Ibrahim, R. A., and Pettit, C. L., 2005, "Uncertainties and Dynamic Problems of Bolted Joints and Other Fasteners," J. Sound Vib., **279**, pp. 857–936.
[4] Brown, J. M., Slater, J., and Grandhi, R. V., 2006, "Probabilistic Analysis of Geometric Uncertainty Effects on Blade Modal Response," ASME Paper No. GT2006-90146.
[5] Brown, J. M., and Grandhi, R. V., 2004, "Probabilistic Analysis of Geometric Uncertainty Effects on Blade-Alone Forced Response," ASME Paper No. GT2004-53959.
[6] Sinha, A., 1986, "Calculating the Statistics of Forced Response of a Mistuned Bladed Disk Assembly," AIAA J., **24**(11), pp. 1797—1801.
[7] Sinha, A., and Chen, S., 1989, "A Higher Order Technique to Calculate the Statistics of Forced Response of a Mistuned Bladed Disk Assembly," J. Sound Vib., **130**(2), pp. 207–221.
[8] Mignolet, M. P., and Hu, W., 1998, "Direct Prediction of the Effects of Mistuning on Forced Response of Bladed Disks," ASME J. Turbomach., **120**, pp 626–634.
[9] Sextro, W., Panning, L., Götting, F., and Popp, K., 2002, "Fast Calculation of the Statistics of the Forced Response of Mistuned Bladed Disk Assemblies With Friction Contacts," ASME Paper No. GT2002-30427.
[10] Sinha, A., 2003, "Computation of the Statistics of Forced Response of a Mistuned Bladed Disk Assembly via Polynomial Chaos," ASME Paper No. GT2003-38961.
[11] Mignolet, M. P., and Lin, C.-C., 1993, "The Closed Form-Perturbation Approach to the Analysis of Mistuned Bladed Disks," ASME J. Turbomach., **115**, pp. 771–780.
[12] Beachkofski, B. K., and Grandhi, R. V., 2005, "Tools for Minimizing Probabilistic Confidence Intervals," ASME Paper No. GT2005-68872.
[13] Ghiocel, D. M., and Wang, L., 2005, "Advances in Computational Stochastic Mechanics for HCF Risk Predictions," *Proceedings of NATO AVT-121 Symposium on Evaluation, Control and Prevention of High Cycle Fatigue in Gas Turbine Engines*, Granada, Spain, Oct. 3–7.
[14] Petrov, E. P., 2004, "A Method for Use of Cyclic Symmetry Properties in Analysis of Nonlinear Multiharmonic Vibrations of Bladed Discs," ASME J. Turbomach., **126**, pp. 175–183.
[15] Petrov, E. P., 2004, "Method for Direct Parametric Analysis of Nonlinear Forced Response of Bladed Discs With Friction Contact Interfaces," ASME J. Turbomach., **126**, pp. 184–192.
[16] Petrov, E. P., and Ewins, D. J., 2003, "Analytical Formulation of Friction Interface Elements for Analysis of Nonlinear Multi-Harmonic Vibrations of Bladed Discs," ASME J. Turbomach., **125**, pp. 364–371.
[17] Petrov, E. P., 2005, "Sensitivity Analysis of Nonlinear Forced Response for Bladed Discs With Friction Contact Interfaces," ASME Paper No. GT2005-68935.
[18] Elliott, R., Green, J. S., and Seinturier, E., 2005, "Aeroelastic Design of Turbine Blades—ADTurB II Overview," *Sixth European Turbomachinery Conference*, Lille, France, Mar. 7–11, Paper No. AMP-105_01/62.
[19] Petrov, E. P., and Ewins, D. J., 2006, "Advanced Modelling of Underplatform Friction Dampers for Analysis of Bladed Disc Vibration," ASME Paper No. GT2006-90146.
[20] Petrov, E. P., and Ewins, E. P., 2005, "Method for Analysis of Nonlinear Multiharmonic Vibrations of Mistuned Bladed Discs With Scatter of Contact Interface Characteristics," ASME J. Turbomach., **127**, pp. 128–136.

Identification of Rotordynamic Forces in a Flexible Rotor System Using Magnetic Bearings

Zachary S. Zutavern
Raytheon Co.,
Dallas, TX 75243
e-mail: zzutavern@raytheon.com

Dara W. Childs
Texas A&M University,
College Station, TX 77843
e-mail: dchilds@turbo-lab.tamu.edu

A method is presented for parameter identification of an annular gas seal on a flexible-rotor test rig. Dynamic loads are applied by magnetic bearings (MBs) that support the rotor. MB forces are measured using fiber-optic strain gauges that are bonded to the poles of the MBs. In addition to force and position measurements, a finite element rotor model is required for the identification algorithm. The FE rotor model matches free-free characteristics of the test rotor. The addition of smooth air sealed to the system introduces stiffness and damping terms for identification that are representative of reaction forces in turbomachines. Tests are performed to experimentally determine seal stiffness and damping coefficients for different running speeds and preswirl conditions. Stiffness and damping coefficients are determined using a frequency domain identification method. This method uses an iterative approach to minimize error between theoretical and experimental transfer functions. Test results produce seal coefficients with low uncertainties. [DOI: 10.1115/1.2799529]

Introduction

Historical attempts at force measurement in magnetic bearings (MBs) have experienced limited success. methods using magnetic flux sensors, load cells, and empirical current and position formulas have produced results with relatively large uncertainties. fiber-optic strain gauge (FOSG) technology substantially improves the accuracy of the force measurements [1].

Traxler and Schweitzer [2] mounted piezoelectric load cells between the magnetic bearing housings and the test platform to measure reaction forces. Accelerations of the bearing housings produce forces that affect the reaction-force measurements. The housing forces are calculated using accelerometer measurements and are then subtracted from the measured total force to calculate the actual force applied to the rotor. The signal-to-noise ratio is low, and the inertial forces are large at higher frequencies.

Lee et al. [3] used a similar approach to perform system identification. They also experienced large uncertainties. In general, this approach is problematic because the load cells must be sized to properly secure the bearings and must have an acceptable sensitivity for forces generated at high frequencies. This causes the low frequency results to have a poor signal-to-noise ratio. The high frequency results are suspect because the bearing motion increases with frequency, creating extraneous force measurements.

Matros et al. [4] used an empirical formula relating the bearing currents and the rotor position to the applied force. Their formulas neglect eddy current loss, hysteresis, and magnetic saturation effects. Matros et al. modeled hysteresis and saturation properties in an effort to improve results. The force calculations are used to determine bearing and seal coefficients. In specific cases, stiffness is overpredicted by 8%.

Fittro et al. [5] measured forces on a static test rig, varying eccentricity and force amplitude. They found that eccentricity changes contribute to most of the uncertainty in the results. The mean error distribution and standard deviation were 1% and 4% of the bearing load capacity, respectively.

Gahler [6] used hall sensors to measure the magnetic flux from the bearing poles. The rotor position and magnetic flux are related to the force with an empirical formula. A correction algorithm is implemented to correct for eddy currents, hysteresis, and saturation. Dynamic forces are applied at frequencies from 20 Hz to 200 Hz with constant amplitude, and the force error is $\pm 11\%$ of load capacity. Dynamic forces are then applied at 120 Hz for various amplitudes, and the force error is reduced to $\pm 2\%$ of load capacity.

Knopf and Nordmann [7] used flux measurements to identify dynamic properties of hydrodynamic bearings. Uncertainties are around 1% of load capacity for static measurements, but they deteriorate to 5% with increasing eccentricity and rotor speed.

Pottie et al. [8] used several methods to determine forces of magnetic bearings. A current and position dependent force model is attempted, and considerable time and effort are spent trying to map the model coefficients. Hall sensors are also attempted. A third method is accomplished by supporting the poles (not the bearing housings) with load cells. However, load cells with sufficient sensitivity for accurate measurements allow the MB poles to move noticeably. Accelerometers are installed to compensate for the pole inertial forces. Unfortunately, the load cells introduce new vibration modes and resonances. All of these methods are unable to significantly improve uncertainties over previous methods.

Raymer and Childs [9] used FOSGs to measure dynamic forces applied by an external exciter. This method results in dramatic improvements in uncertainty. The uncertainty is 4 N or .1% of the bearing load capacity. In 2002, Pavesi [10] used an empirical formula based on current and position to calibrate the FOSGs at low frequencies. The formula is believed to be sufficiently accurate at low frequencies because of the high repeatability of the results. This method encounters difficulties associated with parameter errors in the empirical formula, and the resulting uncertainties are not as low as in the method of Raymer and Childs.

Zutavern and Childs [11] developed a dynamic calibration for the FOSGs. In this method, the FOSGs are calibrated using rotor inertial force. The rotor inertial force is generated by exciting the rotor with frequencies below the first bending mode. The calibrated FOSGs are then used to characterize the split natural frequencies resulting from gyroscopic effects. The calibration has an uncertainty of 9 N (2 lb) or .2% of the bearing load capacity, a reduction by a factor of 10 from Gahler's result.

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 1, 2007; final manuscript received June 8, 2007; published online March 3, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea, and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007.

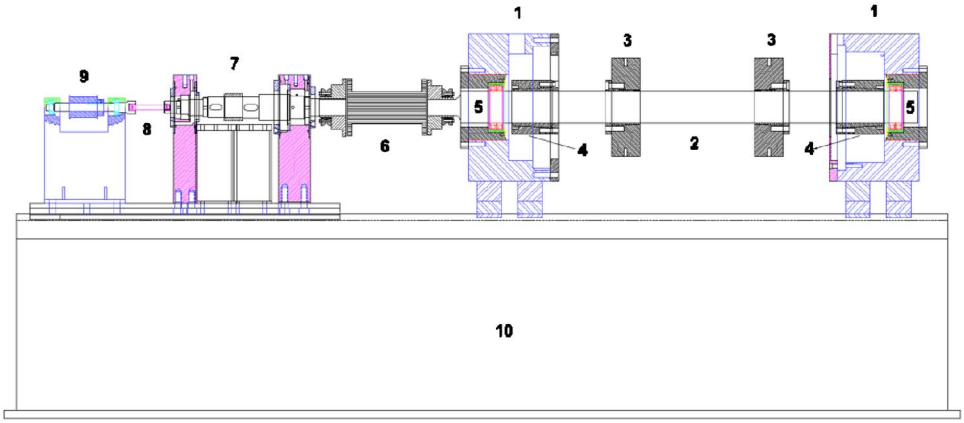


Fig. 1 Magnetic bearing test rig for rotordynamic testing. The test rig is located at the Turbomachinery Laboratory, Texas A&M University, College Station, TX. It is currently used to develop new rotordynamic force identification methodologies.

FOSGs have produced results with the lowest uncertainties to date. Previous FOSG results provide justification for applying this technology to improve the knowledge of real turbomachinery. Using force measurements from FOSGs for parameter identification in the current test configuration provides an excellent foundation for identification of rotordynamic forces in real turbomachinery.

Theory of Operation

Test Rig Description. The MBs (1) (see Fig. 1) have a load capacity of 3560 N and support a steel rotor (2) weighing 2130 N. Disks (3) provide substantial rotational inertia and, accordingly, gyroscopic coupling. The laminated sleeves (4) are the surface on which the magnetic force is exerted. Auxiliary bearings (5) support the rotor when it is not levitated. The brakes (7) can be used to rapidly decelerate the system if the rotor delevitates. The pulley (9), coupling (6), and quill shaft (8) transfer torque from the motor. The test stand base (10) is constructed of 19 mm steel plates with a 76 mm steel top.

The test rig has four FOSGs bonded to the poles of each MB. For both bearings, the FOSGs are oriented as shown in Fig. 2. The FOSGs are offset from the centerline of the pole groupings; however, poles within a group act in unison, and, accordingly, FOSG measurements represent the force applied by each pole. Forces are measured at both ends of the rotor. The combination of force and position measurements provides the necessary information for parameter identification.

Air Seal Design and Theory. The test rig, described previously, has been modified by adding back-to-back annular gas seals at the axial center of the rotor to produce suitable forces for identification. An air seal with a radial clearance of 305 μm and a length of 102 mm is selected based on the FE analysis of seal force characteristics.

The seal assembly shown in Fig. 3 is located at the rotor midspan. The smooth air seal halves (1,2) are clamped by the lower housing (3) and the upper housing (4). The end seal halves (5,6) have exhaust ports that divert the air away from the test rig. The base plate (7) supports the stands (8) that bolt to the lower housing. The adapters (9) are used to center the seals about the rotor. Swirl rings (10) are pressed into the air seals to control the preswirl conditions.

Historical and current test results for smooth air seals indicate that the model from Eq. (1) is representative of smooth seal forces for an eccentricity ratio of less than 0.5 [12]. Cross-coupled damping terms are excluded as they are not historically observed in this seal configuration. The radial forces f_x and f_y applied to the rotor are modeled as stiffness and damping forces associated with radial displacements and velocities or the rotor at the seal. Note that the diagonal terms are equal, and the off-diagonal terms are equal and opposite. In addition, the cross-coupled damping forces are considerably smaller than the other forces, and are neglected for the purposes of identification. This limits the number of parameters for identification to 3. The focus of the present research is on identifying these parameters

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = -\begin{bmatrix} C & c \\ -c & C \end{bmatrix}\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} - \begin{bmatrix} K & k \\ -k & K \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \tag{1}$$

Experiment

System Model. The identification method described later requires an accurate model of the rotor. The rotor is modeled using finite elements described in [13]. At constant running speed Ω, Eq. (2) displays the general linear equation of motion for the rotor. The mass, gyroscopic, and stiffness matrices are developed from the rotor geometry and mechanical properties [14]

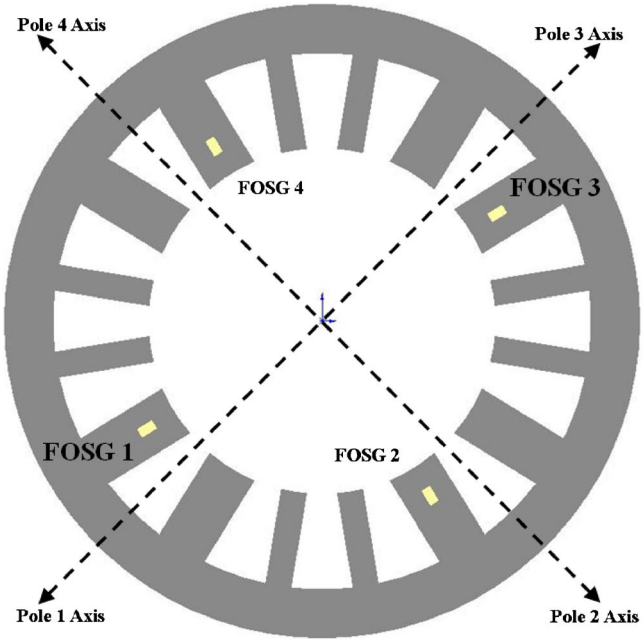


Fig. 2 Axial view of magnetic bearing poles, axes, and fiber-optic strain gauge locations

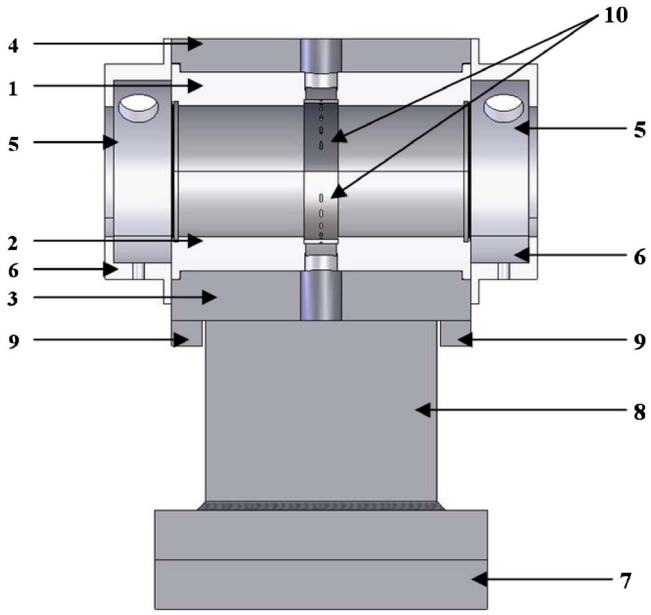


Fig. 3 Air seal assembly installed at the axial center of the test rotor on the magnetic bearing test rig. The seal is added to produce representation rotordynamic forces for identification.

$$\mathbf{M}\ddot{q} - \Omega\mathbf{G}\dot{q} + \mathbf{K}q = F \tag{2}$$

The vector of generalized coordinates q includes 4 degrees-of-freedom at each station along the length of the rotor. Radial displacements of the rotor along two orthogonal axes are included as are rotations about those axes. Equation (3) shows the arrangement of the displacement and rotation variables within the generalized coordinate vector

$$q = [x_1\ y_1\ \beta_{x_1}\ \beta_{y_1}\ \cdots\ x_n\ y_n\ \beta_{x_n}\ \beta_{y_n}]^T \tag{3}$$

The model is tuned to match the natural frequencies and mode shapes for the first three free-free bending modes. The resulting model-experimental agreement ensures that the model accurately captures necessary dynamics of the rotor.

The seal coefficients contained in the seal damping and stiffness matrices C_s and K_s, respectively, can be readily added to the rotor model. For the parameter identification process, the rotor model matrices are known (constants), and the seal coefficient matrices are treated as unknown (variables)

$$F_s = -C_s\dot{q} - K_s q$$

$$M\ddot{q} + (C_s - \Omega G)\dot{q} + (K_s + K)q = F$$

$$C_s = \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & C & 0 & \\ & & 0 & C & \\ & & & & \ddots \end{bmatrix}$$

$$K_s = \begin{bmatrix} 0 & & & & \\ & \ddots & & & \\ & & K & k & \\ & & -k & K & \\ & & & & \ddots \end{bmatrix} \tag{4}$$

Frequency Domain Identification. Frequency domain identification is accomplished by experimentally measuring dynamic flexibility transfer functions (DFTFs) using the MBs, and then matching the experimental results with modeled DFTFs. In Eq. (5), the DFTF matrix $\mathbf{H}$ is a function of excitation frequency ω. It describes the relationship between the force applied at the MBs and rotor position at the MBs

$$X_{\mathrm{MB}} = \mathbf{H}(j\omega) \times F_{\mathrm{MB}} \tag{5}$$

MB forces are represented by a horizontal (x) and a vertical (y) component at each MB. The DFTF matrix in Eq. (6) contains 16 complex valued frequency dependent terms to describe the rotor response at each bearing

$$\begin{bmatrix} x_{\mathrm{MB1}} \\ y_{\mathrm{MB1}} \\ x_{\mathrm{MB2}} \\ y_{\mathrm{MB2}} \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} & H_{14} \\ H_{21} & H_{22} & H_{23} & H_{24} \\ H_{31} & H_{32} & H_{33} & H_{34} \\ H_{41} & H_{42} & H_{43} & H_{44} \end{bmatrix} \begin{bmatrix} F_x^{\mathrm{MB1}} \\ F_y^{\mathrm{MB1}} \\ F_x^{\mathrm{MB2}} \\ F_y^{\mathrm{MB2}} \end{bmatrix} \tag{6}$$

Experimental DFTFs are computed using spectral analysis. DFTF computation requires four independent excitations at each test frequency. The rotor is excited using sequential, single frequency excitation with frequencies ranging about the rotor natural frequency of 140 Hz.

Modeled DFTFs are computed from the finite element model (FEM) (Eq. (4)), and then fitted to the experimental DFTFs using a standard nonderivative search minimization algorithm. The seal parameters are identified as coefficients that minimize the theoretical versus experimental error in the least-squares sense. Other frequency domain identification approaches are outlined in the literature by Maslen et al. [15], Wang and Maslen [16], and Wang et al. [17].

Tests are conducted at 0 rpm and 7700 rpm. For the 0 rpm cases, the tests are repeated ten times to obtain uncertainties at each frequency. Uncertainties are used to estimate seal coefficient uncertainties, assuming random experimental errors. The process of determining the seal coefficients is detailed below:

1. The experimental DFTFs and uncertainties are measured. DFTF uncertainties are characterized by calculating the standard deviations in the test results for repeated tests at 0 rpm.
2. The FEM DFTFs are computed using an initial set of seal coefficient values (predictions from a seal code) for the seal coefficients. The rotor model (excluding seal terms) is known, and is not altered during the identification process.
3. The errors between the measured DFTFs and FEM DFTFs are computed for each test i. The total error vector e is a column vector of all the errors for each transfer function and each test frequency

$$\mathbf{E}_i = \mathbf{H}_i^{\mathrm{measured}} - \mathbf{H}_i^{\mathrm{predicted}}$$

$$\mathbf{E}_i = [E_{i,1} \quad E_{i,2} \quad E_{i,3} \quad E_{i,4}]$$

$$e_i = \begin{bmatrix} E_{i,1} \\ E_{i,2} \\ E_{i,3} \\ E_{i,4} \end{bmatrix}$$

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{7}$$

4. The sum of the error squares J is computed. The error terms were initially weighted inversely to the corresponding variances to obtain a maximum likelihood solution. However, test rig and sensing limitations cause an unweighted error calculation to produce superior results

$$J = e^T e \tag{8}$$

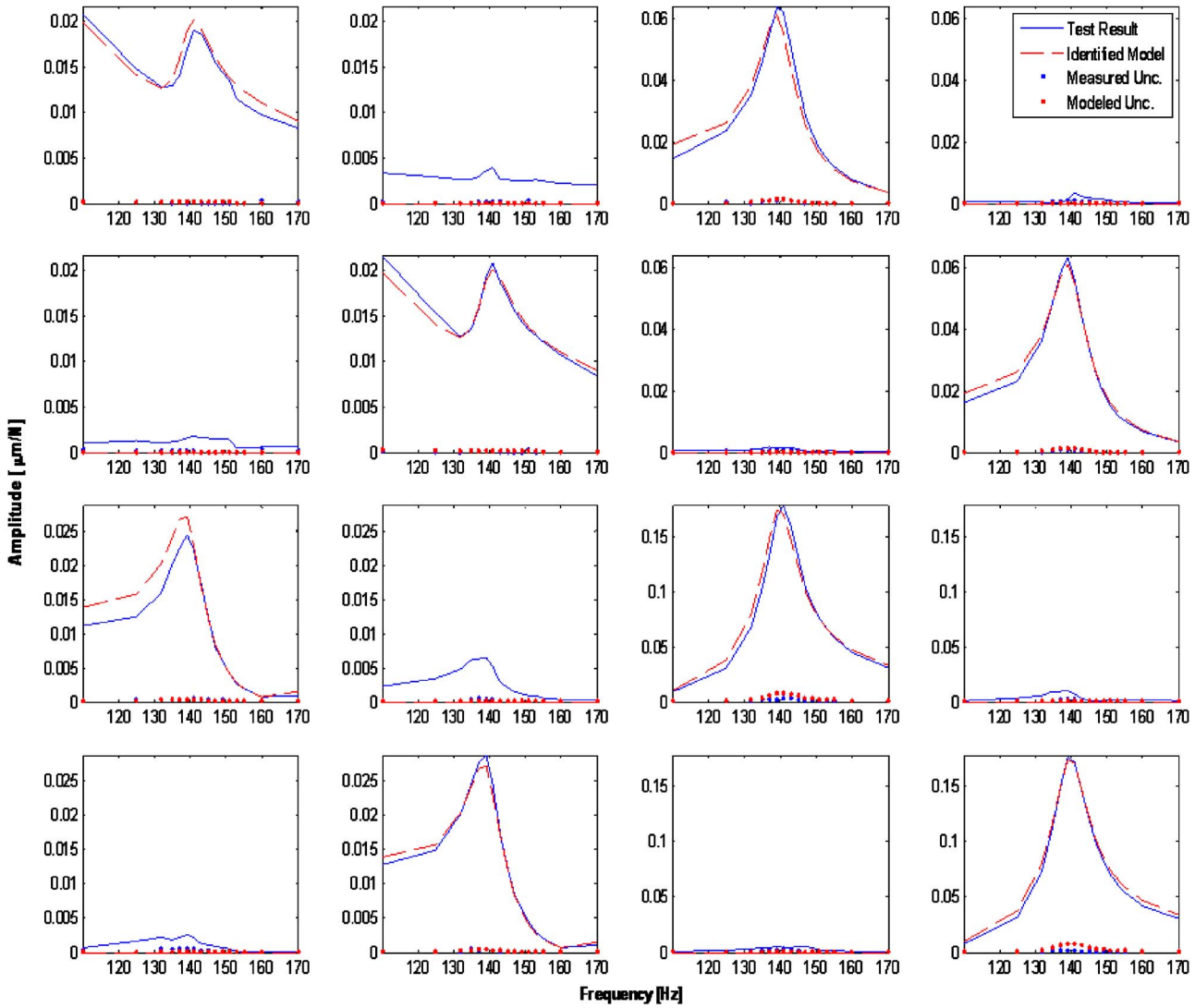


Fig. 4 Amplitudes of DFTFs for experimental test results and the identified FEM model. This test case occurs at 0 rpm with no seal inlet preswirl. Subplot locations correspond to H_{ij} locations within the DFTF matrix.

5. A multidimensional unconstrained nonlinear minimization algorithm known as the Nelder-Mead simplex method is used to iterate (repeating steps 2 through 4) with new FEM seal coefficients to find the local minimum of J [18].
6. Once the seal coefficients are identified, a Monte Carlo uncertainty analysis is performed. The experimental DFTF values are perturbed with errors based on test measurement uncertainty. The identification process is repeated for 10 sets of perturbed experimental DFTFs. A statistical analysis of the seal coefficient identification results yields standard deviations for each of the identified seal coefficients.

The parameter values that minimize J solve the local least-squares minimization problem. If convergence is achieved, the parameters are, therefore, guaranteed (locally) to achieve the best fit in the least-squares sense.

Results and Conclusions

The frequency domain identification results display a remarkably accurate identification. Figures 4 and 5 (Annex A) display the amplitudes of the experimental DFTFs and identified FEM DFTFs. The former figure shows results for a nonrotating test with no seal inlet preswirl, and the latter displays results at 7700 rpm with preswirl. The subplot locations indicate the $\mathbf{H}_{ij}$ position within the DFTF matrix from Eq. (6).

The identified seal coefficients are displayed in Table 1. All test cases are repeated for comparison. In some cases, the results from the repeated test fall within the standard deviations. In other cases this is not true; i.e., direct stiffness K for the tests with preswirl at 0 rpm.

This demonstrates that there are changes in the results that are not explained by random errors. Some possible explanations include change in seal location, operating conditions, and seal roughness. Regarding the variation in direct stiffness, a comparison of the 0 rpm preswirl DFTFs reveals a shift in the natural frequency observed in the repeated test results. The natural frequency shift indicates the influence of extraneous experimental conditions rather than random error. Additionally, note that the difference in force levels for the identified direct stiffness coefficients from the two tests is only 3.4 N (.8 lb), which is .1% of the bearing load capacity.

These results are consistent with uncertainty claims from the previous tests programs on this test rig [1,9–11]. The conclusion is that the uncertainties for the frequency domain identification, at most, correspond with force levels of .1% of load capacity. In

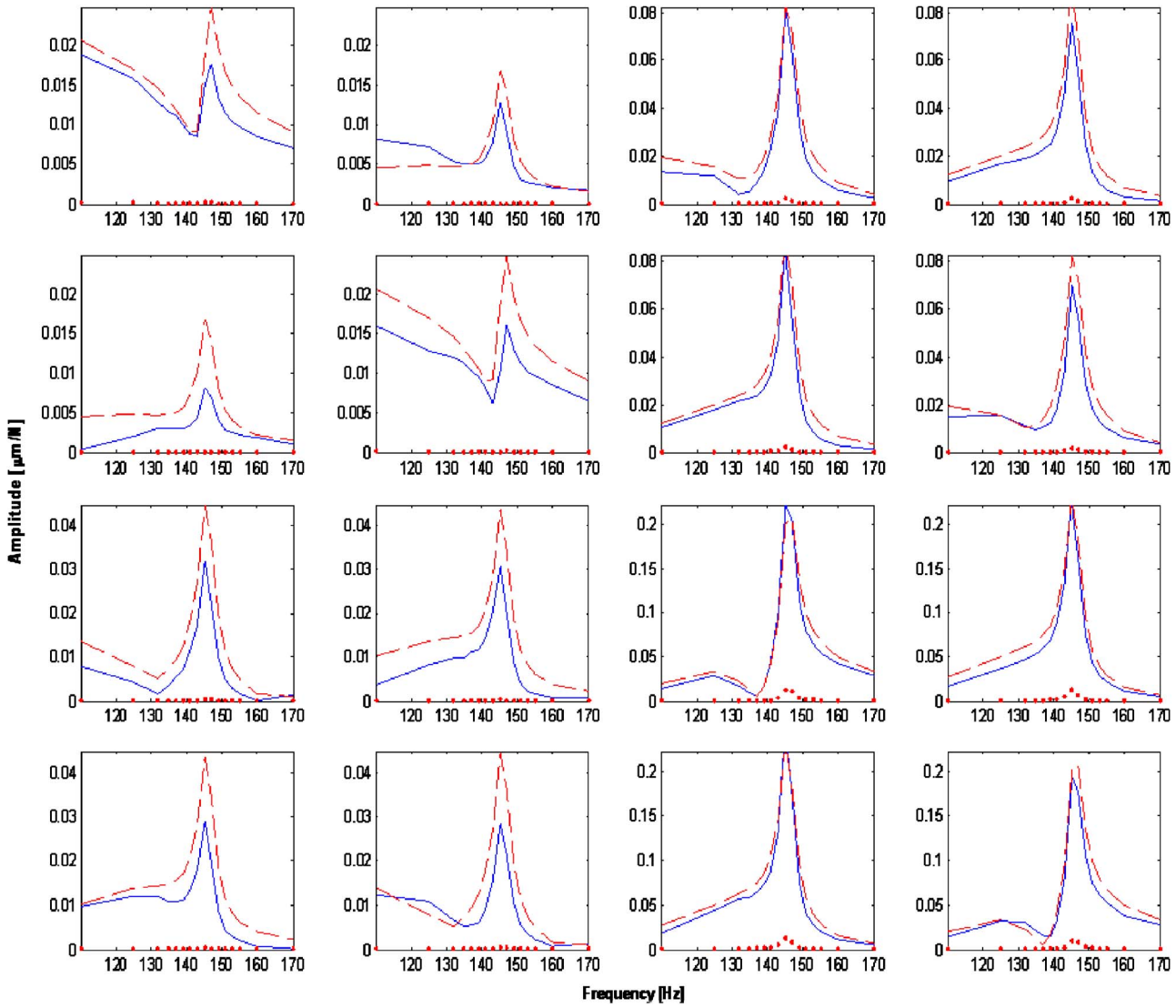


Fig. 5 Amplitudes of the DFTFs for experimental test results and the identified FEM model. This test case occurs at 7700 rpm with seal inlet preswirl. Subplot locations correspond to H_{ij} locations within the DFTF matrix.

some cases, smaller uncertainties appear reasonable. The statistical uncertainty analysis indicates that, in the absence of all but random errors, uncertainties of .01% to .05% of load capacity are achievable.

Extensions

Impeller coefficient testing has been a primary motivation for the present research, and there are significant commonalities worthy of mention. Impeller forces likely produce significant changes in the system response versus the free rotor behavior. In particular, destabilizing impeller forces, potentially the cause of many historic cases of compressor instability, will be readily observable from DFTFs. The availability of compressors with MBs provides not only the ability to use the MBs for testing, but also a reduction in the system complexity from an identification standpoint. The MBs are not part of the identification system model; the MB

Table 1 Air seal coefficient identification results

Preswirl (-)	Ω (rpm)	K (N/m)	K Standard Deviation	C (N s/m)	C Standard Deviation	k (N/m)	k Standard Deviation
N	0	-1.93×10^{6}	4.17×10^{4}	7.26×10^{3}	8.77×10^{1}	—	—
N	0	-1.96×10^{6}	3.69×10^{4}	7.49×10^{3}	7.71×10^{1}	—	—
N	8000	-2.00×10^{6}	2.26×10^{4}	8.97×10^{3}	3.37×10^{1}	1.74×10^{6}	2.84×10^{4}
N	7700	-2.09×10^{6}	2.22×10^{4}	1.07×10^{4}	4.94×10^{1}	2.20×10^{6}	2.96×10^{4}
Y	0	-1.70×10^{6}	2.63×10^{4}	7.85×10^{3}	2.34×10^{2}	8.18×10^{6}	1.96×10^{5}
Y	0	-2.04×10^{6}	2.05×10^{4}	8.11×10^{3}	2.87×10^{2}	7.88×10^{6}	2.65×10^{5}
Y	7700	-2.97×10^{6}	2.57×10^{4}	1.01×10^{4}	8.12×10^{1}	1.16×10^{7}	8.83×10^{4}
Y	7700	-2.94×10^{6}	2.00×10^{4}	1.14×10^{4}	2.39×10^{5}	1.21×10^{7}	2.26×10^{5}

forces act as external excitations to the system. Using MBs as calibrated exciters, therefore, creates a unique ability to limit the model requirements to the rotor characteristics and the impeller, seal, or other rotordynamic forces. This identification method provides the opportunity to accurately identify a variety of rotordynamic forces in flexible rotor systems.

References

[1] Zutavern, Z., 2004, "Fiber Optic Strain Gauge Calibration and Dynamic Flexibility Transfer Function Identification in Magnetic Bearings," M.S. thesis, Texas A&M University, College Station, TX.

[2] Traxler, A., and Schweitzer, G., 1984, *Measurement of the Force Characteristics of a Contactless Electromagnetic Rotor Bearing*, Institute for Mechanics, ETH Zurich, Switzerland.

[3] Lee, C., Ha, Y., and Kim, C., 1994, "Identification of Active Magnetic Bearing System Using Magnetic Force Measurement," *Proc. 4th International Symposium on Magnetic Bearings*, Zurich, Switzerland, pp. 305–309.

[4] Matros, M., Sobotzik, J., and Nordmann, R.,1996, "A New Model-Based Method for the Accurate Measurement of Magnetic Bearing Forces," *Proc. 4th International Symposium on Magnetic Suspension Technology*, Gifu City, Japan, pp. 239–248.

[5] Fittro, R., Baun, D., Maslen, E., and Allaire, P., 1997, "Calibration of an 8-Pole Planar Radial Magnetic Actuator," *Proc. International Gas Turbine & Aeroengine Congress & Exhibition*, Orlando, FL., ASME Paper No. 97-GT-18.

[6] Gahler, C., 1998, "Rotor Dynamic Testing and Control with Active Magnetic Bearings," Ph.D. thesis, ETH, Zurich, Switzerland.

[7] Knopf, E., and Nordmann, R., 2000, "*Identification of the Dynamic Characteristics of Turbulent Journal Bearings Using Active Magnetic Bearings*," IMechE Paper No. C576/110/2000.

[8] Pottie, K., Matthijssen, J., Norbart, C., and Gielen, L., 1999, "Modal Parameter Estimation of Rotation Machinery," IMechE Paper No. C556/005/99.

[9] Raymer, S., and Childs, D., 2001, "Force Measurements in Magnetic Bearings Using Fiber Optic Strain Gauges," *Proc. International Gas Turbine and Aeroengine Congress and Exhibition*, New Orleans, LA, ASME Paper No. 2001-GT-0027.

[10] Pavesi, L., 2002, "Force Measurement in Magnetic Bearings Using Fiber Optic Strain Gauges," M.S. thesis, Texas A&M University, College Station, TX.

[11] Zutavern, Z., and Childs, D., 2005, "Fiber-Optic Strain Gauge Calibration and Dynamic Flexibility Transfer Function Identification in Magnetic Bearings," *Proc. ASME TurboExpo 2005*, Paper No. GT2005-68484.

[12] Kerr, B., 2005, "Experimental and Theoretical Rotordynamic Coefficients and Leakage of Straight Smooth Annular Gas Seals," M.S. thesis, Texas A&M University, College Station, TX.

[13] Childs, D., 1993, *Turbomachinery Rotordynamics: Phenomena, Modeling, & Analysis*, John Wiley & Sons Inc., New York.

[14] Nelson, H., and McVaugh, J., 1976, "The Dynamics of Rotor-Bearing Systems Using Finite Elements," J. Eng. Ind., **98**, pp. 593–600.

[15] Maslen, E., Vazquez, J., and Sortore, C., 2002, "Reconciliation of Rotordynamic Models with Experimental Data," ASME J. Eng. Gas Turbines Power, **124**, pp. 351–356.

[16] Wang, Q., and Maslen, E., 2005, "Identification of Frequency Dependent Parameters in a Flexible Rotor System," *Proc. ASME Turbo Expo 2005*, Paper No. GT2005-68577.

[17] Wang, Q., Maslen, E., and Ahn, H., 2003, "Incorporating Data Uncertainty in Rotordynamic Model Reconciliation," *Proc. AMSE Turbo Expo 2003*, Paper No. GT2003-38595.

[18] Lagarias, J., Reeds, J., Wright, M., and Wright, P., 1998, "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," SIAM J. Optim., **9**(1), pp. 112–147.

New Steps to Improve Rotordynamic Stability Predictions of Centrifugal Compressors

Manoj K. Gupta
e-mail: manoj_k_gupta@dresser-rand.com

Thomas A. Soulas
e-mail: TSoulas@dresser-rand.com

Dresser-Rand Company,
1200 W. Sam Houston Pkwy N.,
Houston, TX 77043

Dara W. Childs
Turbomachinery Laboratory,
Turbo Expo 2006,
Texas A&M University,
College Station, TX 77843
e-mail: dchilds@mengr.tamu.edu

Improved rotordynamic stability is desired by end users, and centrifugal compressor manufacturers are expected to meet, if not exceed, this expectation. Compressor manufacturers are required to design and build machines that are rotordynamically stable on the test stand and in the field. Confidence has been established in predicting the excitation forces from seals and bearings, but impeller aerodynamic excitation forces continue to be a challenge. While much attention is paid to impellers from an aerodynamic performance point of view, more efforts are needed from a rotordynamic standpoint. A high-pressure, reinjection centrifugal compressor is analyzed in order to predict rotordynamic stability using the best available resources for seals and bearings. Impeller shroud forces are predicted using the bulk-flow model developed by Gupta and Childs (Gupta, M., and Childs, D., Proc. of ASME Turbo Expo 2000, Power for Land, Sea, and Air). Each impeller stage is analyzed and an attempt is made to improve the estimation of impeller aerodynamic excitation forces. Logarithmic decrement (log dec) predictions for the full rotor model consisting of all the stages and seals are compared to the full-load full-pressure test measured values using a magnetic bearing exciter. A good correlation is obtained between the measured test results and analytical predictions.
[DOI: 10.1115/1.2799531]

Keywords: rotordynamic coefficients, rotor stability, impeller-shroud forces, magnetic-bearing exciter

Introduction

Reinjection compressors, used to inject natural gas into oil wells at pressures ranging from 100 to 700 bar, have traditionally created rotordynamic challenges due to high pressure and density. As a rule, the instabilities are load versus speed dependent. Specifically, in the past, above a limiting load condition (head rise), the rotors could display subsynchronous unstable motion at the rotor's first natural frequency. Rotordynamicists over the years have successfully followed the dual approach of excitation source elimination and vibrations absorption.

The "Kaybob" compressor instability was eliminated by stiffening the rotor (Smith [1] and Fowlie and Miles [2]). "Ekofisk" injection compressor instabilities were eliminated by several changes, including shortening the bearing span by 140 mm (5.5 in.), modifying the bearings, increasing the "back-wall" clearances on both sides of the impellers, using a squeeze-film damper at the coupling-end bearing, and building a new rotor (Geary et al. [3], and Cochrane [4]). Several other improvements, such as reducing the bearing span, removing vane diffusers, reducing labyrinth diameters, providing shunt hole injection, building swirl brakes, and using hole pattern damper seals have greatly improved rotordynamic stability. While most of those enhancements have been towards bearings and seals, very little effort has been made in studying impellers and their influence on rotordynamic stability. Although several validated commercial codes are available for predicting rotordynamic characteristics of seals and bearings, none exists for impellers. Over the years, the prediction tools for seals and bearings have gained several validations and have been greatly improved. Unfortunately, this is not true for impellers from a rotordynamic standpoint. Predicting accurate impeller forces continues to be a challenge; thus, a complete rotordynamic analysis is difficult to perform.

The only modeling tool widely used in predicting the impeller aeroexcitation comes from the Wachel and von Nimitz [5] purely destabilizing empirical model (Eq. (1)) for the aerodynamic forces (in. lb sec units) and is given in Eq. (2)

$$-\begin{Bmatrix} f_X \\ f_Y \end{Bmatrix} = \begin{bmatrix} 0 & Q \\ -Q & 0 \end{bmatrix} \begin{Bmatrix} X \\ Y \end{Bmatrix} \tag{1}$$

$$Q = \left(\frac{6300 \times \mathrm{HP} \times \mathrm{MW}}{D \times h_t \times \mathrm{RPM}} \right) \times \frac{\rho_D}{\rho_S} \tag{2}$$

where HP is the horsepower, MW is the molecular weight of the gas, D is the impeller outside diameter, h_t is the impeller tip opening at discharge, and ρ_D and ρ_S are the fluid densities at discharge and suction, respectively. The current API standard uses MW = 30 for the calculation of cross-coupled stiffness, and is a modified Wachel's formula. It combines all possible stabilizing actions caused by damping from the impellers and all destabilizing force actions from the impellers into one purely destabilizing element to be applied at the center of the rotor model. Some analysts have tried to further refine the Wachel model's application by calculating labyrinth seal forces and then using Wachel's model separately to account for the unknown forces that are required to explain observed instabilities.

Memmott [6,7] introduced the modal predicted aerocross coupling (MPACC) number. This number is defined as

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 2, 2007; final manuscript received July 6, 2007; published online March 3, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Land, Sea, and Air (GT2007), Montreal, Quebec, Canada, May 14–17, 2007.

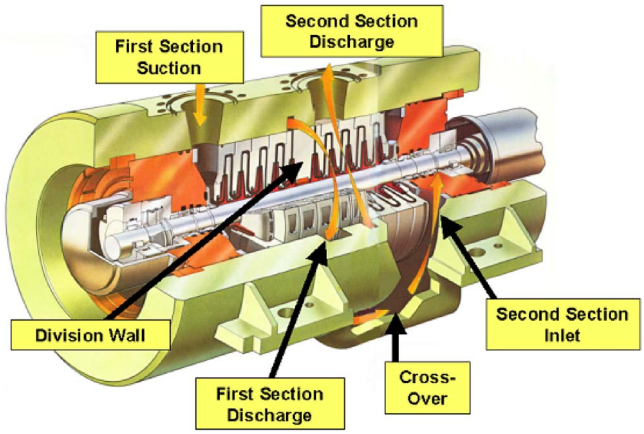


Fig. 1 Sketch of high pressure, back-to-back compressor

$$\text{MPACC} = 189{,}000 \times \sum_{j=1}^{\text{NI}} \frac{\text{HP}_j}{N \times D_j \times h_j} \left(\frac{\rho_d}{\rho_s}\right)_j x_j^2 \tag{3}$$

where x_j is the modal coordinate of the stage. By taking a modal sum based on the first forward whirling mode shape, an effective aerocross-coupling coefficient (Kxy) is calculated and applied at the midspan of the rotor. This has been benchmarked on numerous test cases operating with different mole weight gases, and has been successfully used by the author's company.

Since the introduction of Wachel's destabilizing aeroexcitation formula, research interest has slowly arisen in the study of impeller dynamics and rotordynamic characteristics. Bolleter et al. [8] presented rotordynamic-coefficient data for several pump impellers. Childs [9] then compared Bolleter's test data with reasonable success to his bulk flow model for pump impellers. Yoshida et al. [10] made flow and pressure measurements in the back shroud/casing clearance of an inclined precessing centrifugal impeller and integrated the unsteady pressure distribution to obtain the fluid moment on the precessing impeller shroud, showing good correlation between bulk flow model and test measurements. However, they never made force measurements on the shrouded impeller.

Gupta and Childs [11] presented a bulk-flow model for the annular flow paths between the impeller shroud and its adjacent housing based on Childs' [9] earlier work for pump impellers. They predicted both the force and moment coefficients for a compressor impeller shroud surface using the reaction-force/moment model of the form given in Eq. (4), where (X,Y) and (α_Y,α_X) are components of the impeller's displacement and rotation vectors, and (f_X,f_Y) and (M_Y,M_X) are components of the impeller reaction force and moments.

$$-\begin{Bmatrix} f_X \\ f_Y \\ M_Y \\ M_X \end{Bmatrix} = \begin{bmatrix} K & k & K_{\varepsilon\alpha} & -k_{\varepsilon\alpha} \\ -k & K & -k_{\varepsilon\alpha} & -K_{\varepsilon\alpha} \\ K_{\alpha\varepsilon} & k_{\alpha\varepsilon} & K_\alpha & -k_\alpha \\ k_{\alpha\varepsilon} & -K_{\alpha\varepsilon} & k_\alpha & K_\alpha \end{bmatrix} \begin{Bmatrix} X \\ Y \\ \alpha_Y \\ \alpha_X \end{Bmatrix} + \begin{bmatrix} C & c & C_{\varepsilon\alpha} & -c_{\varepsilon\alpha} \\ -c & C & -c_{\varepsilon\alpha} & -C_{\varepsilon\alpha} \\ C_{\alpha\varepsilon} & c_{\alpha\varepsilon} & C_\alpha & -c_\alpha \\ c_{\alpha\varepsilon} & -C_{\alpha\varepsilon} & c_\alpha & C_\alpha \end{bmatrix} \begin{Bmatrix} \dot{X} \\ \dot{Y} \\ \dot{\alpha}_Y \\ \dot{\alpha}_X \end{Bmatrix} + \begin{bmatrix} M & m & M_{\varepsilon\alpha} & -m_{\varepsilon\alpha} \\ -m & M & -m_{\varepsilon\alpha} & -M_{\varepsilon\alpha} \\ M_{\alpha\varepsilon} & m_{\alpha\varepsilon} & M_\alpha & -m_\alpha \\ m_{\alpha\varepsilon} & -M_{\alpha\varepsilon} & m_\alpha & M_\alpha \end{bmatrix} \begin{Bmatrix} \ddot{X} \\ \ddot{Y} \\ \ddot{\alpha}_Y \\ \ddot{\alpha}_X \end{Bmatrix} \tag{4}$$

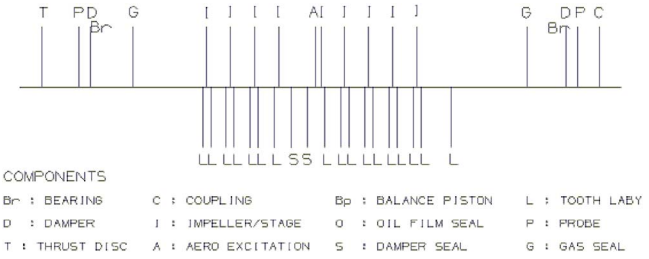


Fig. 2 Schematic of the rotor model

The above model accounts for shroud forces, but does not consider potential impeller-diffuser interaction forces. Gupta and Childs [11] reported a more stable stage than that predicted using Wachel's model. This was in direct agreement with Memmott [6] for a large centrifugal compressor. Memmott found poor correlation between predictions and test and field experience using the specific Wachel's formula. While Wachel's model predicted an unstable compressor for Memmott's large centrifugal compressor, the compressor was stable in operation. Qualitatively, the impeller-shroud model of Eq. (4) did a good job in predicting the frequency characteristics of the measurements and an adequate job in predicting the cross-coupled stiffness k and direct damping C.

To date, there has not been a clear validated rotordynamic program for impeller-shroud force predictions. Furthermore, there has been no comparison between rotor stability predictions using the impeller bulk-flow model and test results. Using the log dec measurement technique described by Moore et al. [12] and Moore and Soulas [13], log dec for a reinjection machine is measured and comparisons are made between the test results and the predictions.

Reinjection Compressor Description

The compressor used in this study is a nine-stage, nonintercooled, back-to-back centrifugal compressor. A sketch of a typical back-to-back centrifugal compressor is shown in Fig. 1. The center division wall seal is a damper seal (hole pattern seal). The compressor has tilting-pad radial journal bearings in series with squeeze film dampers, a tilting-pad thrust bearing, and dry gas seals. This unit has a 7200 psi case pressure rating. The advantages of using a back-to-back machine for high-pressure reinjection applications have been mainly better thrust balance ability, especially at off-design conditions, elimination of large diameter balance piston resulting in less leakage, and higher damping because of the optimum location of hole pattern seal.

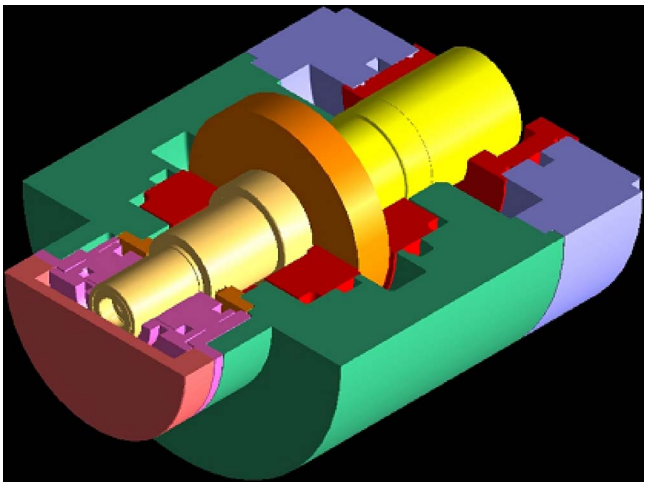


Fig. 3 Solid model of the magnetic bearing exciter

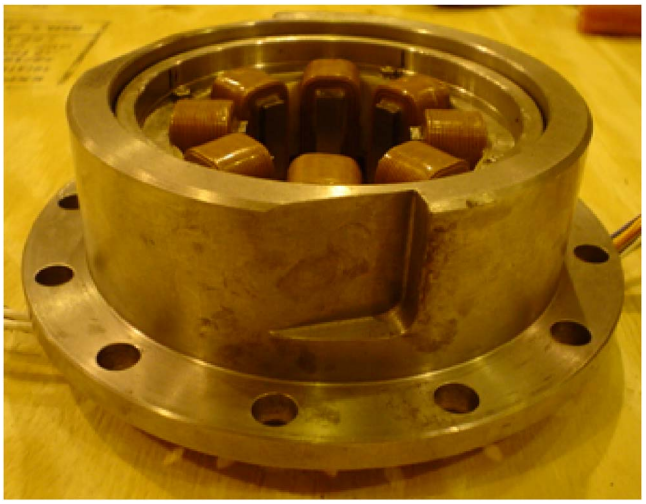


Fig. 4 Magnetic bearing used for the excitation

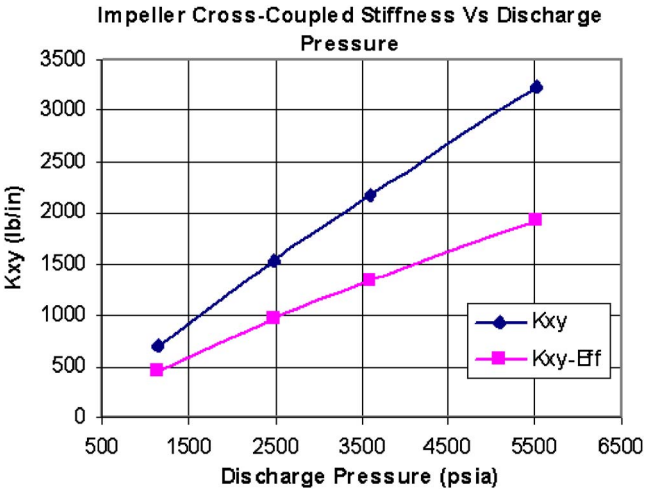


Fig. 5 Bulk-flow impeller code predicted cross-coupled stiffness versus discharge pressure

Analytical Modeling

Rotordynamic modeling of this nine-stage, back-to-back, high-pressure reinjection compressor is done by considering only the reaction forces and no moments. A model of the rotor is built using a rotordynamic analysis software developed by Ramesh [14]. A schematic of the rotor model is shown in Fig. 2. For the complete lateral analysis of this compressor, the following components are modeled: Tilt-pad journal bearings in series with squeeze film dampers, hole pattern division wall seal, impeller eye and interstage stationary tooth labyrinth seals, second-section gas balance stationary tooth labyrinth seal, and all nine impeller stages.

Tilt-pad bearing coefficients are computed using the work of Nicholas et al. [15]. Hole pattern damper seal coefficients are determined using the ISOTSEAL program developed by Kleynhans and Childs [16]. The toothed labyrinth seals are modeled by the program of Kirk [17]. Impeller coefficients are calculated using both the modified form of Wachel number (MPACC) and the bulk-flow impeller code developed by Gupta and Childs [11]. The impeller bulk-flow code solves the turbulent bulk-flow continuity and momentum equations to obtain full force and moment reaction matrices. Therefore, a reduced reaction force only matrix of the form given in Eq. (5) is used for the analysis

$$-\begin{Bmatrix} F_X \\ F_Y \end{Bmatrix} = \begin{bmatrix} K & k \\ -k & K \end{bmatrix} \begin{Bmatrix} X \\ Y \end{Bmatrix} + \begin{bmatrix} C & c \\ -c & C \end{bmatrix} \begin{Bmatrix} \dot{X} \\ \dot{Y} \end{Bmatrix} + \begin{bmatrix} M & m \\ -m & M \end{bmatrix} \begin{Bmatrix} \ddot{X} \\ \ddot{Y} \end{Bmatrix} \tag{5}$$

For seals and bearings, the mass matrix is negligible and is often eliminated in modeling.

The current approach to compute the total impeller aeroexcitation involves computing the aeroexcitation for each stage and then taking the modal sum by squaring the normalized deflection of the first fundamental frequency. A similar approach is followed when using the impeller bulk-flow code. The bulk-flow code produces both cross-coupled stiffness *Kxy* and direct damping *C*; therefore, effective cross-coupled stiffness defined in Eq. (6) is used to compute the net aerocross-coupled stiffness

$$(k_{xy})_{\text{Eff}} = k_{xy} - C\omega_d \tag{6}$$

The first forward damped natural frequency ω_d in Eq. (6) is taken as the unloaded damped frequency. In this particular study, the loaded damped natural frequency differed only slightly from the unloaded damped natural frequency.

Magnetic Bearing Exciter Test Setup

The Type 1 test in accordance with ASME PTC-10 test specification consisted of eight head-capacity points (1–8), and three additional data points (9–11) at the end of the test; all eleven points in coordination with magnetic bearing exciter sweeps. The three last test points were taken during the evacuation of the gas from the test loop to obtain data at decreasing density profiles across the compressor. These three additional points provide further information about the effect of density on the stability of the system. The magnetic bearing exciter was attached to the free end of the rotor, and an asynchronous force was injected into the rotor system to excite the first forward whirling mode. This technique measures the rotor's log dec as described by Moore et al. [12]. A solid model assembly with the magnetic bearing exciter installed on the shaft is shown in Fig. 3, and the magnetic bearing exciter used is shown in Fig. 4.

Comparison Between Test Data and Analytical Results

Figure 5 shows predicted impeller cross-coupled stiffness as a function of the discharge pressure for constant speed data points 6 and 9–11. The bulk-flow impeller code predicts the net aerocross coupling forces to increase with the increase in discharge pressure.

Comparison between measurement and predictions of the rotor log dec as a function of discharge pressure is shown in Fig. 6 for

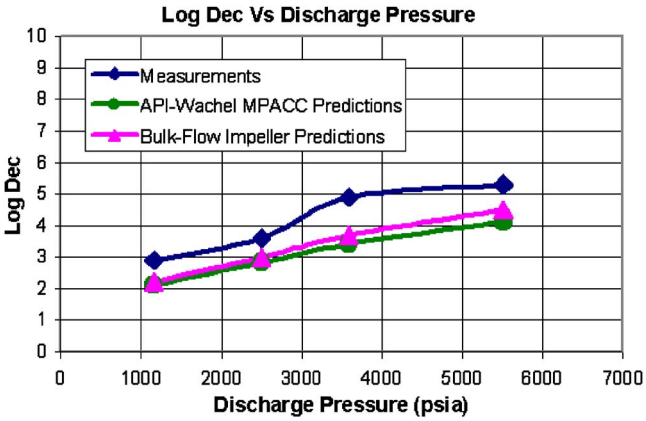


Fig. 6 Test and predicted rotor log dec versus discharge pressure

Table 1 Test operating conditions and log dec values

Point number	Inlet pressure psia	Discharge pressure psia	Discharge density Lbs/ft^3	Gas power HP	Speed rpm	Measured log dec δ	API-Wachel MPACC log dec δ	Bulk-Flow Flow log dec δ
1	2327.98	5586.58	20.08	11,633	7234	4.8	5.5	6.1
2	2313.80	5528.14	20.49	7993	6494	6.2	6.3	6.8
3	2333.81	5559.78	20.53	5875	6153	6.8	7.3	7.8
4	1756.30	5352.03	13.11	14,704	9495	5.8	3.9	4.4
5	2010.38	5526.35	13.57	14,313	9150	5.2	3.8	4.2
6	2016.08	5514.50	13.68	10,487	8486	5.3	4.1	4.5
7	2050.88	5574.98	13.68	8276	8199	5.7	4.3	4.5
8	1469.51	5408.53	12.27	10,464	9492	5.3	4.3	4.6
9	1205.14	3590.58	9.84	6808	8480	4.9	3.4	3.7
10	831.70	2494.42	7.19	4701	8471	3.6	2.8	3
11	402.59	1156.04	3.35	2149	8504	2.9	2.1	2.2

the constant speed data points used in Fig. 5. Figure 6 illustrates that system stability increases with increasing discharge pressure. This result can be attributed mainly to the "dominance" of the hole pattern seal at higher discharge pressures. Thus, even though the impeller generated aerocross coupling force increases with discharge pressure, hole pattern seals provide net positive effective damping at a faster rate than the excitations.

A good correlation is obtained between the measurement and predictions, especially considering the stability trend with discharge pressure. The small underprediction in the log dec can be attributed to various uncertainties and the conservative nature of predictions.

There is a small improvement in the predictions using the bulk-flow impeller model. The impeller force excitation does not have a significant impact on the overall system stability for this high-pressure reinjection machine. The impeller bulk-flow code correctly predicted the increased stability and provided a slightly better match than API-Wachel based MPACC, but the impeller coefficients are significantly small when compared to the hole pattern seal. Complete predictions of all the test points covering a wide range of operating conditions are given in Table 1.

Table 1 clearly shows that system stability is maintained, even when running at off-design conditions. Points 1 and 2, although operating at similar discharge pressure and gas density, differ in running speed. This change in speed produces less damping for point 1, which is close to the overload limit. Small overprediction occurs at points 1, 2, and 3, as shown in Table 1. Note also that the high log dec measured at low discharge pressure is due to external squeeze film dampers in series with the tilt-pad journal bearings.

Summary, Discussion, and Conclusions

A complete analysis of the rotor model is done using the state-of-the-art tools available, including the newly developed impeller bulk-flow code. Stability predictions were made for a wide range of conditions to cover both the head changes and density influence. Predicted results using a bulk-flow model and the MPACC numbers were compared to the measured test data results. A good correlation is obtained between measurement and predictions. Actual impeller force coefficients were predicted by integrating the dynamic pressure and shear stress field in the shroud-casing clearance, thus providing a reasonable estimate of the impeller contribution to the overall rotordynamic stability. This new model has helped in rational estimation of the impeller shroud forces. The results show that increasing gas density yields increased stability when hole pattern seals are used. The measurements provide further validation in the analytical tools and have helped in further validating the bulk-flow code predictions. The results presented demonstrate the low impact of impeller produced aeroexcitation on the rotordynamic stability when hole pattern seals are used at high discharge pressure. Clearly, hole pattern seals are the most dominant element at high discharge pressure, and knowing the exact running clearances in operation, although difficult to achieve, could further improve the predictions.

It has been shown that the compressor is stable for a wide range, and can operate satisfactorily under off-design conditions from a rotordynamic stability standpoint.

Acknowledgment

The authors would like to thank the Dresser-Rand Test Department and in particular, Gary Colby and Rick Antle, as well as Dr. Oscar de Santiago, for his contribution during the test. We also would like to thank Jay Koch, Jason Kopko, Robert Kunselman, and Dr. Krish Ramesh for their technical help and suggestions. Finally, the authors want to thank Dresser-Rand for its support and permission to publish this work.

Nomenclature

- MPACC = modal predicted aero cross-coupling
- NI = number of impellers
- ω_d = unloaded first damped natural frequency (rad/s)
- δ = logarithmic decrement
- x_j = modal co-ordinate of the stage
- $(k_{xy})_{Eff}$ = effective aero-cross coupled stiffness
- HP_j = horsepower of impeller j (hp)

References

[1] Smith, K., 1974, "An Operational History of Fractional Frequency Whirl," *Proceedings, 4th Turbomachinery Symposium*, pp. 115–125.
[2] Fowlie, D., and Miles, D., 1975, "Vibration Problems with High Pressure Centrifugal Compressors," Petroleum Mechanical Engineering Conference, Tulsa, OK, ASME Paper No. 75-PET-28.
[3] Geary, C., Damratowsky, L., and Seyer, C., 1976, "Design and Operation of the World's Highest Pressure Gas Injection Centrifugal Compressors," Offshore Technology Conference, Houston, TX, Paper No. OTC 2485.
[4] Cochrane, W., 1976, "New-Generation Compressor Injecting Gas at Ekofisk," Oil & Gas J., pp. 63–70.
[5] Wachel, J. C., and von Nimitz, W. W., 1981, "Ensuring the Reliability of Offshore Gas Compressor Systems," JPT, J. Pet. Technol., **8**, pp. 2252–2260.
[6] Memmott, E. A., 2000, "The Lateral Stability Analysis of a Large Centrifugal Compressor in Propane Service at an LNG Plant," *ImechE, Proceedings of the 7th International Conference on Vibrations in Rotating Machinery*, Nottingham, England, Sept. 12–14, pp. 187–198.
[7] Memmott, E. A., 2000, "Empirical Estimation of a Load Related Cross-Coupled Stiffness and the Lateral Stability of Centrifugal Compressors," *CMVA, Proceedings of the 18th Machinery Dynamics Seminar*, Halifax, Canada, April 26–28, pp. 9–20.
[8] Bolleter, U., Leibundgut, E., Sturchler, R., and McCloskey, T., 1989, "Hydraulic Interaction and Excitation Forces of High Head Pump Impellers," *Pumping Machinery 1989, Proceedings of the Third Joint ASCE/ASME Mechanics Conference*, La Jolla, CA, pp. 187–194.
[9] Childs, D., 1989, "Fluid Structure Interaction Forces at Pump-Impeller-Shroud Surfaces for Rotordynamic Calculations," ASME J. Vib., Acoust., Stress, Reliab. Des., **111**, pp. 216–225.
[10] Yoshida, Y., Saito, A., Ishizaki, S., Tsujimoto, Y., and Ohashi, H., 1996, "Measurement of the Flow in the Backshroud/Casing Clearance of a Precessing

Centrifugal Impeller," *Proceedings of the 6th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery*, Vol. 2, pp. 151–160.
[11] Gupta, M., and Childs, D., 2006, "Rotordynamic Stability Predictions for Centrifugal Compressors Using a Bulk-Flow Model to Predict Impeller Shroud Force and Moment Coefficients," *Proceedings of ASME Turbo Expo 2006, Power for Land, Sea, and Air*, Montreal, Canada, May 8–11.
[12] Moore, J. J., Walker, S. T., and Kuzdzal, M. J., 2002, "Rotordynamic Stability Measurement During Full-Load, Full-Pressure Testing of a 6000 Psi Re-Injection Centrifugal Compressor," *Proceedings of the Thirty-first Turbomachinery Symposium, Turbomachinery Laboratory*, Department of Mechanical Engineering, Texas A&M University, College Station, TX.
[13] Moore, J. J., and Soulas, T., 2003, "Damper Seal Comparison in a High-Pressure Re-Injection Centrifugal Compressor during Full-Load, Full-Pressure Factory Testing Using Direct Rotordynamic Stability Measurement," *Proceedings of DETC'03, ASME 2003 Design Engineering Technical Conferences and Computers and Information in Engineering Conferences*, Chicago, IL, September 2–6.
[14] Ramesh, K., 2002, "State-of-the-art Rotor Dynamic Analysis Program," 9th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery (ISROMAC), Honolulu, HI, Feb. 10–14.
[15] Nicholas, J. C., Gunter, E. J., and Allaire, P. E., 1979, "Stiffness and Damping Coefficients for Five Pad Tilting Pad Bearing," ASLE Trans., **22**(2), pp. 113–124.
[16] Kleynhans, G. F., and Childs, D. W., 1996, "The Acoustic Influence of Cell Depth on the Rotordynamic Characteristics of Smooth-Rotor/Honeycomb-Stator Annular Gas Seals," ASME International Gas Turbine and Aeroengine Congress and Exposition, Brimingham, U.K., June 10–13.
[17] Kirk, R. G., 1990, "User's Manual for the Program DYNPC28—A Program for the Analysis of Labyrinth Seals," Negavib Research and Consulting Group, Virginia Tech, Blacksburg, VA.

Jerzy T. Sawicki[1]
Center for Rotating Machinery Dynamics and Control,
Cleveland State University,
Cleveland, OH 44115-2425
e-mail: j.sawicki@csuohio.edu

Samuel A. Johansson
Philips Medical Systems,
Cleveland, OH 44143-2131
e-mail: andreas.johansson@philips.com

John H. Rumbarger
Rolling Bearing Consultant,
Nazareth, PA 18064
e-mail: jhrinc@aol.com

Ronald B. Sharpless
Philips Medical Systems,
Cleveland, OH 44143-2131
e-mail: ron.sharpless@philips.com

Fatigue Life Prediction for Large-Diameter Elastically Constrained Ball Bearings

The application of large-diameter bearing rings and the thereof inherited low stiffness make them susceptible to local distortions caused by their surrounding structures, which are often under heavy loads. The standard accepted design criteria for these bearings are based on the estimation of the internal load distribution of the bearing, under the assumption of rigid circular and flat supporting structures, that keep the bearing inner and outer races in circular, flat, i.e., not deformed shapes. However, in the presence of structural distortions, the element load distribution can be severely altered and cannot be predicted via the standard design criteria. Therefore, the application of large-diameter ball and roller bearing rings as the critical components in rotating machines becomes more of a design task than making a catalog selection. The analytical and finite element approach for fatigue life prediction of such a bearing application is presented. The undertaken approach and the results are illustrated based on the analysis and fatigue life simulation of the computed tomography scanner's main rotor bearing. It has been demonstrated that flexibility of the rings can significantly reduce the fatigue life of the ball bearing. [DOI: 10.1115/1.2772632]

Introduction

A computed tomography (CT) scanner's main bearing must be capable to handle radial, axial as well as overturning moment. The four-point contact ball bearing design with one row of balls is capable of carrying simultaneous radial, thrust, and overturning moment loads. The four-point contact ball bearing has two ball paths in each race. The two race path curvature centers are offset forming a "gothic arch" between the two curvatures.

The use of rigid (nonseparable) or captured four-point contact ball bearings is not recent. Rumbarger [1] and Pritts and Jones [2] described their uses in radar antennas, construction equipment, cranes, and bucket wheel excavators. Uneven internal distribution of ball loads due to structural hard points in crane and excavator frameworks was recognized by Pritts [3]. Filetti and Rumbarger [4] showed a method for predicting the influence of structural supports on the ball or roller bearing performance. Sague and Rumbarger [5] described the core crushing type of failure in large-diameter, case-hardened four-point contact ball bearings. The more recent use of four-point contact ball bearings is for wind turbine pitch and yaw bearings [6] and X-ray scanner bearings (both medical and baggage scanners). The design and rating of four-point contact ball bearings are given in detail (along with examples) by Rumbarger [6].

There are five design criteria, all of which must be satisfied, in order to accomplish a satisfactory and reliable four-point contact ball bearing application [5,6]:

- sufficient bearing fatigue spalling L_{10} life: operating loads
- sufficient bearing static capacity: maximum or survival loads
- prevention of surface failures: satisfactory lubrication and sealing
- prevention of core crushing failures: adequate case depth and core hardness
- nonrolling items (i.e., cages, ball spacers, external bolting, integral seals) and friction torque

ISO International Standard 281 [7] is used to rate fatigue life of ball and roller bearings. The L_{10} *rating life* is commonly used to denote the fatigue life that will be met or exceeded by 90% of a group of apparently identical bearings, and where life is determined by the first macroscopic evidence of spalling of either the ball or the ball path. The statistical concept of fatigue and development of the basic equations were first presented by Lundberg and Palmgren [8]. Correlating computerized rolling bearing analysis results to the ISO Std. 281 is described in detail by Rumbarger and Poplawski [9].

This paper describes the mathematical development on modeling of four-point contact large-diameter and elastically constrained ball bearing toward its fatigue life prediction. The developed procedure has been implemented in a computer program and the generated results are presented for the case study of the CT-scanner main bearing.

Mathematical Modeling of Four-Point Contact Ball Bearing

The mathematical analysis of the four-point contact ball bearing closely follows the pioneer work of Jones [10] and the private communications for four-point contact ball bearing analysis (in three degrees of freedom; radial, axial, and tilt) by Jones [11]. This paper introduces improvements to these early papers and describes the development of a computer code for the analysis of the bearing in five degrees of freedom.

The elastic displacements of the inner race and outer race supporting structures are used as movements of the inner and outer race curvature centers (C_i, $i=1,\ldots,4$, see Fig. 1) and are included in the analysis. The actual number of balls in the bearing is modeled to correspond to the number of circumferential, equally spaced nodes in the finite element model, as described by Rumbarger [12].

The four-point contact ball bearing in a right-hand Cartesian coordinate system is shown in Fig. 1. The outer race is assumed

[1]Corresponding author.

Contributed by the International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received April 26, 2007; final manuscript received May 10, 2007; published online March 5, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Turbo Expo 2007: Power for Land, Sea and Air, Montreal, Canada, May 14–17, 2007. Paper No. GT2007-27564.

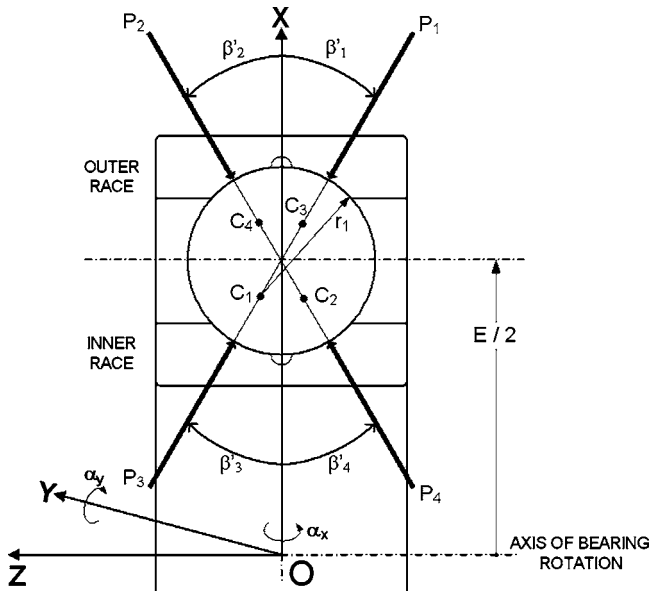


Fig. 1 Four-point contact ball bearing load lines and initial contact angles (forces shown acting on ball)

fixed in this system and the inner race is free to displace in translation and rotation. All forces, moments, translations, and rotations are positive in the positive sense of the coordinate system. Each individual load path is defined by a line of action between the curvature center and the ball center. Race curvature centers for contact path 1–3 (similar description can be provided for contact path 2–4) and ball center movements V_{1n} and V_{2n} of the nth ball

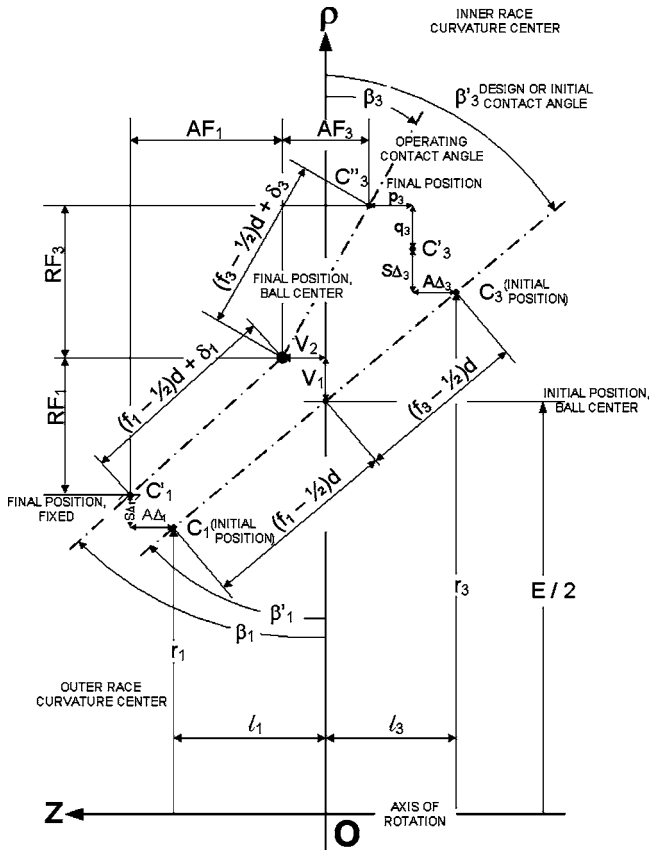


Fig. 2 Race curvature and ball center movements; 1–3 path

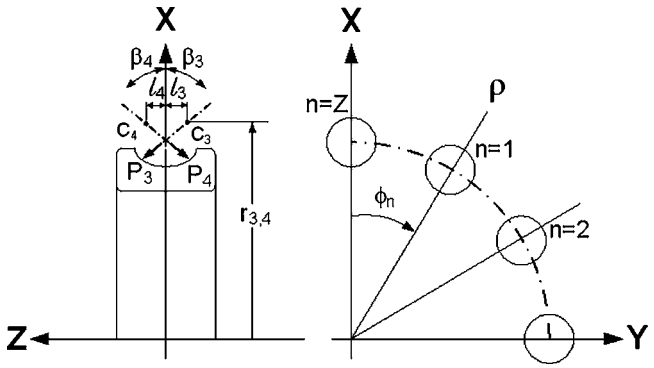


Fig. 3 Force and moments, acting on inner race (shaft)

are described in Fig. 2. The inner race curvature centers C_3 and C_4 of Fig. 2 all are displaced when the inner race is displaced.

In Fig. 2, the outer race and inner race curvature centers are allowed the initial movements $S\Delta_{in}$ and $A\Delta_{in}$, which are the *out-of-roundness* ($S\Delta_{in}$) and *out-of-flatness* ($A\Delta_{in}$) displacements taken from the finite element model. The outer race curvature centers (C_1 and C_2) are then fixed in the coordinate system. The inner race and its curvature centers are displaced by means of initial guesses along with an initial guess for each ball center movement. The geometries, contact angles, and ball compressions (ball load) of Fig. 2 are then fully defined.

Forces and movements acting on the inner race and shaft, as well as the angular location of the nth ball are described in Fig. 3.

Method of Solution. The method of solution is to displace the inner race and its curvature centers (C_3 and C_4) by means of initial displacements of the inner race support (shaft) in five degrees of freedom (i.e., along X, Y, Z, and about X and Y). The analysis of an application with the inner race fixed and the outer race free to displace is solved by reversing the algebraic signs of the forces and moments applied to the outer race and then applying these forces and moments as applied to the inner race.

An initial displacement (computer generated) for each ball then allows for the balls to be solved for force equilibrium (two degrees of freedom) with all four contacts by Newton-Raphson iteration (inner ball loop). The ball forces acting on the inner race (Fig. 3) are then summed to provide force and moment reactions on the inner race. The differences between the reaction loads and moments, and the applied loads and moments, form the error functions (five degrees of freedom), and equilibrium is obtained by Newton-Raphson iteration (outer inner race loop).

The bearing internal geometry as described in Figs. 1–3 is represented by

$$\begin{aligned} AF_{1n} &= (f_1 - 1/2)d \sin \beta'_{1n} - V_{2n} + A\Delta_{1n} \\ AF_{2n} &= (f_2 - 1/2)d \sin \beta'_{2n} + V_{2n} - A\Delta_{2n} \\ AF_{3n} &= (f_3 - 1/2)d \sin \beta'_{3n} + V_{2n} - p_{3n} - A\Delta_{3n} \\ AF_{4n} &= (f_4 - 1/2)d \sin \beta'_{4n} - V_{2n} + p_{4n} + A\Delta_{4n} \end{aligned} \tag{1}$$

$$\begin{aligned} RF_{1n} &= (f_1 - 1/2)d \cos \beta'_{1n} + V_{1n} - S\Delta_{1n} \\ RF_{2n} &= (f_2 - 1/2)d \cos \beta'_{2n} + V_{1n} - S\Delta_{2n} \\ RF_{3n} &= (f_3 - 1/2)d \cos \beta'_{3n} - V_{1n} + q_{3n} + S\Delta_{3n} \\ RF_{4n} &= (f_4 - 1/2)d \cos \beta'_{4n} - V_{1n} + q_{4n} + S\Delta_{4n} \end{aligned} \tag{2}$$

$$l_i = (f_i - 1/2)d \sin \beta_i' \quad i = 1,2,3,4 \tag{3}$$

where l_i is the axial distance from the curvature center to the center of the bearing, and subscript i specifies the specific ball-race contact number. The radial distance r_i to the respective curvature center is

$$r_i = \frac{E}{2} - k_i(f_i - 1/2)d \cos(\beta_i') \quad \begin{matrix} k_1 = k_2 = +1 \\ k_3 = k_4 = -1 \end{matrix} \tag{4}$$

Since outer race contacts (1 and 2) are fixed in the coordinate system, only initial structural shape descriptions ($S\Delta_{\text{in}}$ and $A\Delta_{\text{in}}$) change these original curvature centers.

The inner race is free to move in five degrees of freedom in the coordinate system. Rotation about Z (α_z) is bearing rotation and does not affect curvature center displacements. Inner race movement displaces each inner race curvature center (contacts 3 and 4) as described by p_{in} and q_{in} seen in Figs. 2 and 3.

$$p_{1n} = p_{2n} = q_{1n} = q_{2n} = 0 \tag{5}$$

$$p_{3n} = z - \alpha_y r_3 \cos \phi_n + \alpha_x r_3 \sin \phi_n \tag{6}$$

$$p_{4n} = z - \alpha_y r_4 \cos \phi_n + \alpha_x r_4 \sin \phi_n \tag{7}$$

$$q_{3n} = x \cos \phi_n + y \sin \phi_n + \alpha_x l_3 \sin \phi_n - \alpha_y l_3 \cos \phi_n - \frac{Pd}{2} \tag{8}$$

$$q_{4n} = x \cos \phi_n + y \sin \phi_n - \alpha_x l_4 \sin \phi_n + \alpha_y l_4 \cos \phi_n - \frac{Pd}{2} \tag{9}$$

where

$$\phi_n = \frac{2\pi n}{Z} \tag{10}$$

is the nth ball angular position and Z is the total number of balls (see Fig. 3).

Equilibriums of forces and moments acting on the shaft are

$$F_x - \sum_{n=1}^{Z} P_{3n} \cos \beta_{3n} \cos \phi_n - \sum_{n=1}^{Z} P_{4n} \cos \beta_{4n} \cos \phi_n \neq 0 = \varepsilon_1 \tag{11}$$

$$F_y - \sum_{n=1}^{Z} P_{3n} \cos \beta_{3n} \sin \phi_n - \sum_{n=1}^{Z} P_{4n} \cos \beta_{4n} \sin \phi_n \neq 0 = \varepsilon_2 \tag{12}$$

$$F_z + \sum_{n=1}^{Z} P_{3n} \sin \beta_{3n} - \sum_{n=1}^{Z} P_{4n} \sin \beta_{4n} \neq 0 = \varepsilon_3 \tag{13}$$

$$M_x + \sum_{n=1}^{Z} P_{3n}(r_3 \sin \beta_{3n} - l_3 \cos \beta_{3n}) \sin \phi_n - \sum_{n=1}^{Z} P_{4n}(r_4 \sin \beta_{4n} - l_4 \cos \beta_{4n}) \sin \phi_n \neq 0 = \varepsilon_4 \tag{14}$$

$$M_y - \sum_{n=1}^{Z} P_{3n}(r_3 \sin \beta_{3n} - l_3 \cos \beta_{3n}) \cos \phi_n + \sum_{n=1}^{Z} P_{4n}(r_4 \sin \beta_{4n} - l_4 \cos \beta_{4n}) \cos \phi_n \neq 0 = \varepsilon_5 \tag{15}$$

The individual ball load, P_{in}, is expressed by

$$P_{\text{in}} = K_i \delta_{\text{in}}^{3/2} \geqslant 0 \tag{16}$$

where δ_{in} is the compression of the ball-race contact measured between the race curvature center and the ball center. The angle with respect to the plane (X-Y) of the bearing is the contact angle β_{in} defined by the race curvature center and the ball center, as shown in Figs. 2 and 3.

$$\delta_{\text{in}} = [(RF_{\text{in}})^2 + (AF_{\text{in}})^2]^{1/2} - (f_i - 1/2)d \geqslant 0 \tag{17}$$

It is convenient to use the same number of balls as nodes or gap elements used in the finite element structural model, which represents the ball bearing. Then, the structural displacements at race centroids can be directly converted to $A\Delta_{\text{in}}$ (out-of-flatness) and $S\Delta_{\text{in}}$ (out-of-roundness) curvature center movements as the angular spacing (ϕ_n) is the same. The summation of ball loads $\Sigma_{n=1}^{Z}$ can be treated as a line integral and adjusted to the sum over the reduced number of balls (Z^*) by $(Z/Z^* \Sigma_{n=1}^{Z^*})$ [12].

Let us define

$$B = Z/Z^* \tag{18}$$

and introduce ball to race deflections, and introduce subscript notation jk for the five degrees of freedom of loads and displacements. Then, substituting formulas in Eqs. (16) and (18) into Eqs. (11)–(15) yields

$$F_1 - B\sum_{n=1}^{Z^*} K_3 \delta_{3n}^{3/2} \cos \beta_{3n} \cos \phi_n - B\sum_{n=1}^{Z^*} K_4 \delta_{4n}^{3/2} \cos \beta_{4n} \cos \phi_n \neq 0 = \varepsilon_1 \tag{19}$$

$$F_2 - B\sum_{n=1}^{Z^*} K_3 \delta_{3n}^{3/2} \cos \beta_{3n} \sin \phi_n - B\sum_{n=1}^{Z^*} K_4 \delta_{4n}^{3/2} \cos \beta_{4n} \sin \phi_n \neq 0 = \varepsilon_2 \tag{20}$$

$$F_3 + B\sum_{n=1}^{Z^*} K_3 \delta_{3n}^{3/2} \sin \beta_{3n} - B\sum_{n=1}^{Z^*} K_4 \delta_{4n}^{3/2} \sin \beta_{4n} \neq 0 = \varepsilon_3 \tag{21}$$

$$M_x + B\sum_{n=1}^{Z^*} K_3 \delta_{3n}^{3/2}(r_3 \sin \beta_{3n} - l_3 \cos \beta_{3n}) \sin \phi_n - B\sum_{n=1}^{Z^*} K_4 \delta_{4n}^{3/2}(r_4 \sin \beta_{4n} - l_4 \cos \beta_{4n}) \sin \phi_n \neq 0 = \varepsilon_4 \tag{22}$$

$$M_y - B\sum_{n=1}^{Z^*} K_3 \delta_{3n}^{3/2}(r_3 \sin \beta_{3n} - l_3 \cos \beta_{3n}) \cos \phi_n + B\sum_{n=1}^{Z^*} K_4 \delta_{4n}^{3/2}(r_4 \sin \beta_{4n} - l_4 \cos \beta_{4n}) \cos \phi_n \neq 0 = \varepsilon_5 \tag{23}$$

These equations can be linearized and solved iteratively by the Newton-Raphson method.

Let x_k^0, $k = 1, \ldots, 5$, be initial estimates of the displacement variables X, Y, Z, α_x, and α_y in that sequence in the equilibrium equations. Improved values are

$$x_k' = x_k^0 - [a_{jk}]^{-1} \varepsilon_j \quad k,j = 1, \ldots, 5 \tag{24}$$

where ε_j are error functions (residual forces and moments).

The elements a_{jk} of the stiffness matrix are

$$a_{jk} = \frac{\partial F_j}{\partial x_k} \quad k,j = 1, \ldots, 5 \tag{25}$$

and they are determined by differentiation of Eqs. (19)–(23).

The ball center movements V_1 and V_2 (see Fig. 2) must be determined by an inner loop iteration in order to obtain the derivatives $\partial V_1/\partial x_k$ and $\partial V_2/\partial x_k$ for the outer loop iteration.

Following the nomenclature shown in Fig. 1, and substituting the initial (design) contact angles β_i' by operating contact angles β_i, one can write the equations for the nth ball equilibrium as the summation of forces along the Z direction

$$P_{1n}\sin\beta_{1n} - P_{2n}\sin\beta_{2n} - P_{3n}\sin\beta_{3n} - P_{4n}\sin\beta_{4n} \neq 0 = \Gamma_{1n} \tag{26}$$

and along the X direction

$$P_{1n}\cos\beta_{1n} - P_{2n}\cos\beta_{2n} - P_{3n}\cos\beta_{3n} - P_{4n}\cos\beta_{4n} \neq 0 = \Gamma_{2n} \tag{27}$$

where Γ_{un}, $u=1,2$ are the error functions for the nth ball equilibrium.

The Newton-Raphson matrix inversion is applied to get the ball equilibrium: let V_m^0, $m=1$ or 2, be initial estimates of the ball center displacements V_1 and V_2. Dropping the subscript n for convenience, one can solve for ball location as

$$V_m' = V_m^0 - [b_{um}]^{-1}\Gamma_u \tag{28}$$

The elements $\mathbf{b}_{um}$ of the matrix are

$$b_{um} = \frac{\partial\Gamma_u}{\partial V_m} \quad u,m = 1,2 \tag{29}$$

Substituting Eq. (16) into Eqs. (26) and (27) yields

$$K_1\delta_{1n}^{3/2}\sin\beta_{1n} - K_2\delta_{2n}^{3/2}\sin\beta_{2n} - K_3\delta_{3n}^{3/2}\sin\beta_{3n} + K_4\delta_{4n}^{3/2}\sin\beta_{4n} \neq 0 = \Gamma_{1n} \tag{30}$$

$$-K_1\delta_{1n}^{3/2}\cos\beta_{1n} - K_2\delta_{2n}^{3/2}\cos\beta_{2n} + K_3\delta_{3n}^{3/2}\cos\beta_{3n} + K_4\delta_{4n}^{3/2}\cos\beta_{4n} \neq 0 = \Gamma_{2n} \tag{31}$$

and the derivatives of Eq. (29) are calculated based on the above equations.

The partial derivatives $\partial V_m/\partial x_k$, $m=1,2$; $k=1,\ldots,5$, needed for the outer loop iteration of stiffness matrix elements are obtained by solving the two equations for ball equilibrium, Eqs. (26) and (27), until Γ_{1n} and $\Gamma_{2n}\cong 0$, after the inner loop ball iteration (Eq. (28)) is completed.

Case Study: Philips Computed Tomography Scanner Main Bearing

Background. The four-point contact bearing case study is based on a computed axial tomography scanner also known as CAT, CT scanner, or body section roentgenography. CT is a medical imaging method utilizing tomography which, with the aid of digital geometry processing, can create 3D images of the internals of an object from a large series of two-dimensional X-ray images, which are taken around a single axis of rotation. The analyzed scanner, shown in Fig. 4, is a third generation machine where X-ray slice data are generated using an X-ray tube (XRT) that rotates around the patient, while sensors (DMS) are positioned on the opposite side to the XRT. This sets high demands on the manufacturing tolerances and the internal play of the bearing. Data scans are progressively taken as the object is gradually passed through the gantry.

The current bearing design has a criterion of an L_{10} of 420×10^6 revolutions, which corresponds to 6.6 years of continuous use at 120 rpm. Because the expected life span of the CT scanner is ten years, the calculated L_{10} life is compared to a ten year continuous operation at a new design speed of 143 rpm, or an L_{10} life of 752×10^6 revolutions. The goal of the conducted study was to validate if the current design, assembly, and the operational loads could affect bearing reliability and ultimately lead to premature failure. The analysis is performed without the assumptions of rigid raceways for the four-point contact bearing.

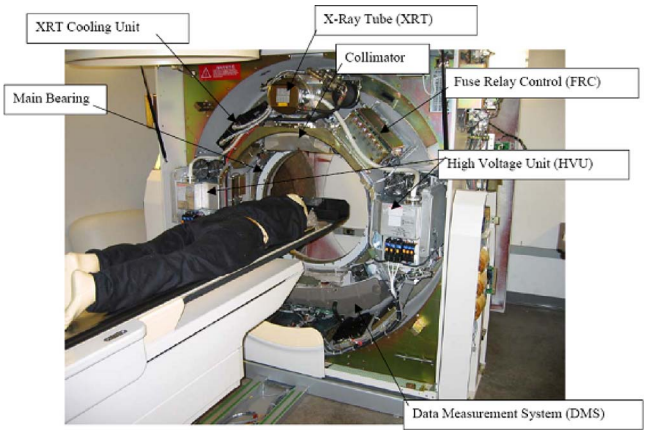


Fig. 4 CT scanner, cover removed (courtesy of Philips Medical Systems, 2005)

Description of Computed Tomography Scanner Rotor-Bearing System. The main bearing for CT scanner of Philips Medical Systems (PMS) is a four-point contact bearing made of SAE 52100 steel. The bearing is through hardened and its inner race is bolted to the stator while an outer race rotates with the rotor. The rotor is machined to the specific dimensions from an aluminum casting. The complete system with all components is shown in Fig. 4, where the components of the most significant masses are labeled. These are modeled as lumped masses in the finite element model.

During the rotation, the attached to the rotor, heavy masses contribute to significant distortions of the thin section bearing and its supporting structure [13], i.e., the aluminum rotor. The complicated geometry of the rotor creates highly nonuniform stiffness distribution and subsequently affects operational distortions. In order to account for all dynamic effects, a full scale finite element analysis was required.

During the scanning operation, the rotor system experiences various rotational speeds depending on the specific function performed, in a range from 0 rpm to a maximum of 143 rpm. To simulate the worst load case scenario, the presented analysis has been conducted for maximum rotational speed. In addition, the rotor system operates with three different tilt positions, where the most common is the 0 deg position, and occurs when the rotational axis of the system is parallel to the supporting floor. The two other operational positions are either +30 deg or −30 deg tilts. The rotor-bearing system has a center of gravity which is offset from the bearing center in the axial direction, and therefore

Table 1 Bearing system input parameters

Parameter name	Data
Pitch diameter	31.000 in. (0.7874 m)
Ball diameter	0.500 in. (0.0127 m)
Number of balls	98
Model no. of balls	96
Conformity ratio	0.52
Contact angle	30 deg
Poisson's ratio	0.25
Modulus of elasticity	29 Mpsi (200 MPa)
Friction start	0.006
Friction run	0.003
Diametral clearance	0.0005 in. (12.7 μm)
Rotor speed	143 rpm

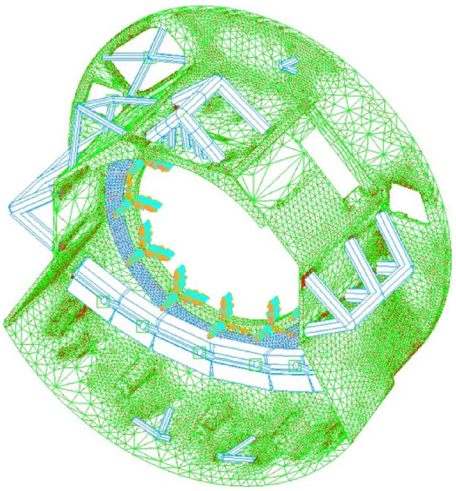


Fig. 5 Rotor and bearing finite element model

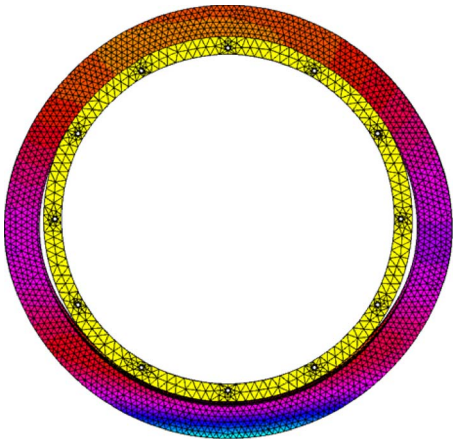


Fig. 6 Bearing distortions—inner and outer race

creates an overturning moment. The total mass of the rotor with all its components is around 1600 lbs (720 kg). The parameters of the bearing are provided in Table 1.

Results and Discussion

Background. The bearing is modeled with nonlinear springs to simulate the stiffness of balls and the rotor is modeled using tetrahedral elements, where lumped masses are used to represent the most critical and heaviest components. The finite element analysis accounts for gravity and a rotational speed of 143 rpm. The displacements of the inner and outer races of the bearing are taken from the finite element modeling results and are used to calculate movements of the ball-race curvature centers, as described in previous sections. An analysis with perfect raceways, i.e., no initial structural/dynamic distortions, was also performed. The two analyses are compared to see the impact of the distorted races compared to the perfect (rigid) races with no distortions. The following sections summarize the obtained results.

Finite Element Model. The rotor and bearing's inner and outer races have been modeled using finite element modeling in I-DEAS. The rotor's model includes lumped masses representing heavy components such as XRT, DMS (see Fig. 4), etc., which are connected to the rotor by bar elements. The bearing's balls are modeled with 96×2 nonlinear springs representing the ball-race contact in compression and "tension" of each ball. The precalculated spring rate in compression was 10.2×10^6 lb/in. (1.78 GN/m). In order to satisfy the finite element analysis (FEA) simulation constraints and avoid singularities, a small spring rate of 1.0 lb/in. (17 N/m) was assigned for tension rates. Thus, the nonlinear load deflection relationship, valid only for the compression, is the following:

$$P = K\delta^{3/2} = 10.2 \times 10^6 \delta^{3/2} \tag{32}$$

Figure 5 illustrates the finite element model of the rotor with lumped masses (green squares) and corresponding connector bars (blue and white). The inner race (green) is fixed at its bolting positions to reduce the number of degrees of freedom and avoid the rigid body motion. The outer race (blue) shares nodes with the rotor at contacting surfaces and at bolting locations. The bearing and the rotor were assigned their respective material properties. The model consists of approximately 250,000 tetrahedral elements and the solution time was about 1 h using a Pentium 4 personal computer with 4 Gbyte RAM. The application of nonlinear springs increases the solution time and the required computational power.

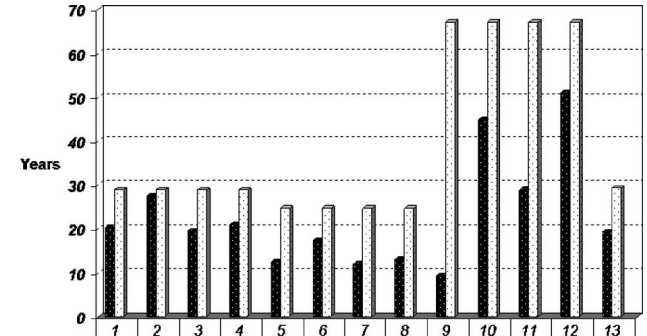
The distortion of bearing rings is shown in Fig. 6. The rotor system rotates at 143 rpm and the rotor gravity vector is at 6 o'clock, i.e., in the negative X direction according to the coordinate system seen in Fig. 3. A clearance created between the inner and the outer race at 3 and 9 o'clock positions is due to centrifugal forces caused by the lumped masses representing the HVU units (see Fig. 4), which are located at just these locations on the rotor.

Ball Bearing Analysis. The input parameters for the bearing system include load profiles and displacements from the FEA of ball bearing load distribution. These have been used in the computer program based on the development presented earlier. The bearing parameters are listed in Table 1. All load cases run at 143 rpm and are single load cases, i.e., each case has a duty cycle time of 100%. Positive tilt is in the sense of positive rotation around the Y axis, as illustrated in Fig. 3. Gravity and rotational speed of 143 rpm are applied to the model.

The gravity vector was superpositioned to simulate 4×3 different rotational positions of the rotor. The summary of load cases is provided in Table 2. The four analyzed positions of the rotor were with gravity vector at 3, 6, 9, and 12 o'clock positions and with three tilt angles, 0 deg, −30 deg, and +30 deg. The solution is sought in terms of displacements of the spring nodes at each curvature center of the bearing races, which are then used as an input

Table 2 Load cases—forces and moments

Load case no.	Radial (lb)	Axial (lb)	Moment (lb in)	Tilt angle (deg)	Case no. 13 (duty time) (%)
1–4	1600 (7117 N)	0 (0 kg)	7800 (881 Nm)	0	60
5–8	1386 (6165 N)	800 (3559 N)	6755 (763 Nm)	+30	20
9–12	1386 (6165 N)	−800 (−3559 N)	6755 (763 Nm)	−30	20



	1	2	3	4	5	6	7	8	9	10	11	12	13
■ OoF/R	20.2	27.4	19.5	20.9	12.5	17.4	12.0	13.1	9.30	44.9	29.0	51.1	19.3
□ No OoF/R	28.8	28.8	28.8	28.8	24.7	24.7	24.7	24.7	67.2	67.2	67.2	67.2	29.3
a3	0.7	0.95	0.68	0.73	0.51	0.7	0.49	0.53	0.14	0.67	0.43	0.76	0.66

Fig. 7 Load cases with corresponding L_{10} and a_3

of out of flatness and out of roundness to computer code.

An additional load case, load case no. 13, combines all load cases 1–12 to simulate a full rotation and with duty cycle time accordingly to the specific tilt cases. Load case no. 13 consists of load cases 1–4 at 60% of time, load cases 5–8 at 20% of time, and load cases 9–12 at 20% of time.

All load cases with corresponding structural data gathered from the FEA were used as an input for the final nonlinear ball load distribution analysis.

Results. The results of the fatigue life for the analyzed bearing are presented in Fig. 7. The black bars in Fig. 7, labeled "OoF/R" (out-of-flatness/roundness), represent bearing B_{10} life with raceway distortions extracted from the FEA, and white bars, labeled "No OoF/R" (No out-of-flatness/roundness), represent B_{10} life in a case of no raceway distortions, i.e., perfect raceways. The presented results show that the shortest bearing life exists for the load cases at the 3 and 9 o'clock positions, with the lowest B_{10} value of 9.3 years in case no. 9. This is (the only one case) below the required B_{10} of ten years, as specified by PMS. The life adjustment factor, a_3, see Ref. [13], is calculated as a ratio between the life for undistorted raceways and life of distorted raceways, and is used as an indicator of how much the flexibility of the raceways actually impacts the life of the bearing.

The B_{10} life for case no. 13 is over 19 years, which is almost twice as much as the desired B_{10} by PMS, which is set to 10 years. The ball load distribution for the worst load case no. 9 (i.e., the lowest B_{10}), for contacts 3 and 4, is shown in Fig. 8, where the maximum ball load is observed at contact 4 and is 137.5 lbs (611.6 N).

Case no. 12 has the largest B_{10} life, which is derived from the load distribution seen in Fig. 9. The maximum ball load is observed at contact 3 and is 57.1 lbs (254 N). In this case, the load distribution is quite different in both shape and magnitude, as compared to case no. 9.

The case of perfect raceways (no distortions) is illustrated in Fig. 10. The resulting ball load distribution is much smoother as compared with both case nos. 9 and 12. The corresponding B_{10} is 67 years with the resulting maximum loads being 50.7 lbs (225.5 N) for contact 4 and 28.5 lbs (126.77 N) for contact 3.

Conclusions and Recommendations

The comprehensive mathematical modeling for the ball load distribution for the four-point contact ball bearing has been revisited. The details of mathematical modeling along with formulas have been provided. This treatment might be useful for anyone who wishes to investigate fatigue life of large-diameter bearings with raceways considered as flexible.

The full finite element model has been developed to determine bearing housing distortions in order to calculate the actual bearing

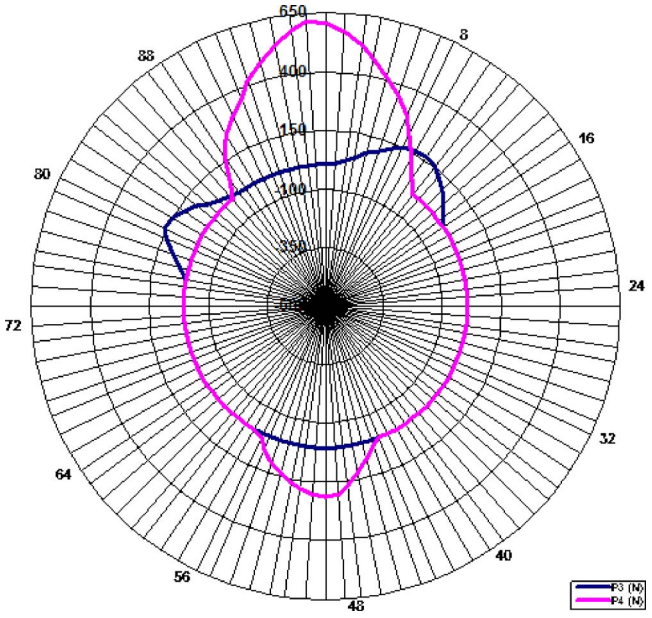


Fig. 8 Bearing load distribution for load case no. 9

loads. The model is limited in its ability to allow deformations on the inner race due to applied boundary conditions at bolting locations. This should, however, not significantly affect the predicted L_{10} life, since the stator, which the inner race is bolted on, is fairly stiff and should not affect the resulting race curvature distortions.

In this case study, a significant difference between the fatigue life of the perfect bearing with no raceway distortions and the flexible bearing ring with distortions was observed. The bearing with flexible raceways has a life reduction factor, a_3, ranging from 0.14 to 0.95, as compared to the case of perfect and rigid raceways. However, the analysis results indicate only one case which has the life of 9.3 years, which is shorter than the specified requirement of 10 years. It should be noted that all analyzed cases were independent of each other and had an assumed rotor speed of 143 rpm, 24/7, in the particular position of the rotor gravity vector, and thus they are quite different than the actual operational

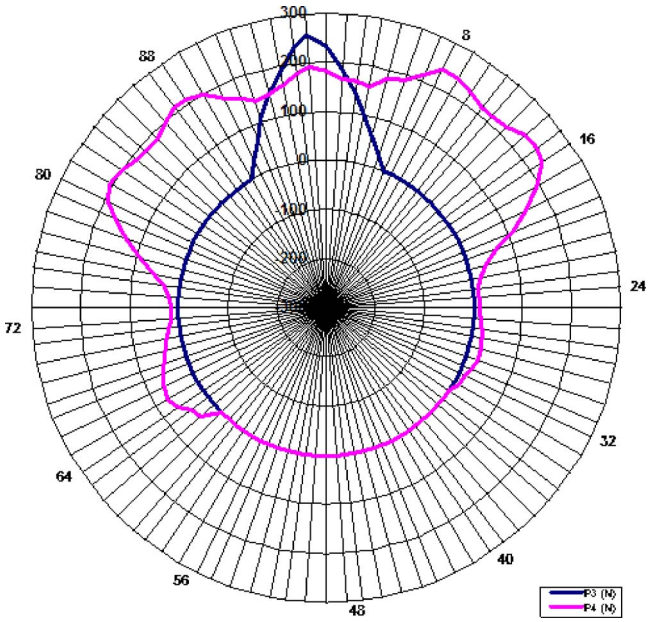


Fig. 9 Bearing load distribution for load case no. 12

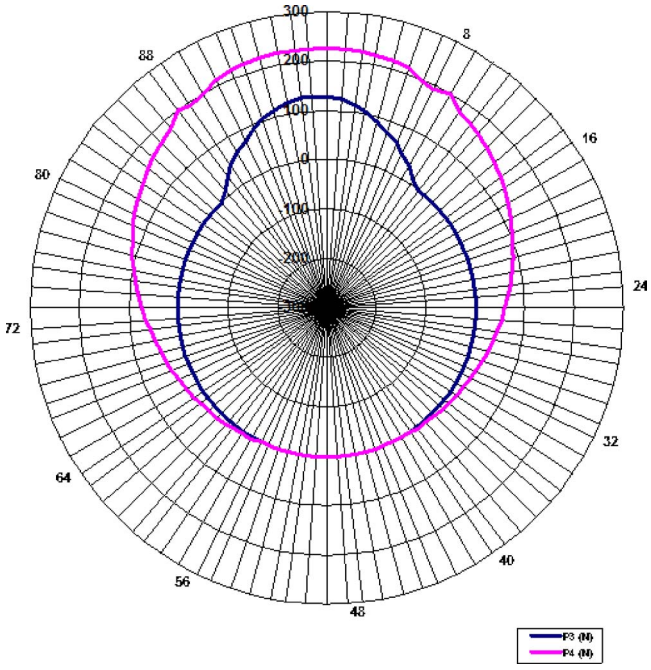


Fig. 10 Bearing load distribution for load case nos. 9 and 12 with perfect raceways

scenarios. With all load cases combined (i.e., load case no. 13) to more realistically simulate the actual use of the scanner, with correct duty cycle time for the different tilting conditions, the calculated life of 19 years was almost twice the required life. The overall summary of the predicted bearing fatigue life for all considered cases is presented in Table 3.

Based on the analysis, and the generated results thereof, the following conclusions can be drawn.

1. The FEA and the bearing load distribution analysis result in B_{10} life which is equal or exceeds the CT-scanner design criteria (B_{10} of ten years) in all load cases but one (case no. 9).
2. When combining load case nos. 1–12 to simulate the actual load profile and duty cycle, a bearing life B_{10} of 19 years is calculated, which is twice as large as the design criteria.
3. The analyzed bearing is adequate in its design based on the operational conditions of the studied CT scanner and the predicted fatigue life.

Table 3 Operational cases and corresponding bearing fatigue life prediction

No.	Gravity vector	L_{10} ($\times 10^6$ rev)	B_{10} (h)	B_{10} (yr)
1	3 o'clock	1527	1.78*E*+05	20.27
2	6 o'clock	2067	2.41*E*+05	27.48
3	9 o'clock	1475	1.72*E*+05	19.59
4	12 o'clock	1578	1.84*E*+05	20.99
5	3 o'clock	943.8	1.10*E*+05	12.55
6	6 o'clock	1312	1.53*E*+05	17.40
7	9 o'clock	909.4	1.06*E*+05	12.06
8	12 o'clock	986.7	1.15*E*+05	13.13
9	3 o'clock	699.2	8.15*E*+04	**9.30**
10	6 o'clock	3380	3.94*E*+05	44.96
11	9 o'clock	2187	2.55*E*+05	29.03
12	12 o'clock	3843	4.48*E*+05	51.11
13		**1458**	**1.70*E*+05**	**19.36**

Nomenclature

AF = axial distance between centers, mm
$A\Delta$ = out-of-flatness of curvature center, mm
B = ratio of actual and model number of balls, Z/Z^*
B_{10} = fatigue life that 90% of bearings endure, h
C_i = curvature center of raceway, initial position
C_i' = curvature center of raceway, displaced position
C_i'' = curvature center of raceway, final position
E = pitch diameter of ball train, mm
F = applied force, N
L_{10} = fatigue life that 90% of bearings endure, rev. $\times 10^6$
K = ball to race stiffness constant, N/mm
M = applied overturning moment, Nm
O = origin or center of coordinates
P_i = ball load at contact i, N
Pd = bearing internal diametral clearance, mm
RF = radial distance between centers, mm
$S\Delta$ = out-of-roundness of curvature center, mm
V_m = ball center movements, mm
X,Y,Z = inertial coordinate system
Z = actual number of balls per row in bearing
Z^* = modeled number of balls in bearing
a_3 = life factor=L_{10} (with FEA)/L_{10} (no FEA)
$\mathbf{a}_{jk}$ = stiffness matrix, N/m
$\mathbf{b}_{um}$ = stiffness matrix, N/m
d = ball diameter, mm
f_i = raceway curvature factor
l_i = axial distance from origin to C_i, mm
mr = millions of revolutions
n = ball counter
p,q = inner race curvature center movements, mm
r_i = radius from bearing center to curvature center, mm

Greek Symbols

$\alpha_x,\alpha_y,\alpha_z$ = rotation about X,Y,Z axis, rad
β_i' = initial or design contact angle, deg
β_i = operating contact angle under load, deg
Γ_i = error functions for ball equilibrium, N
δ = ball to race deflection, mm
ε = error functions for shaft equilibrium, N or Nm
ϕ = ball location angle in the X-Y plane, rad

Subscripts

i = contact number, 1,…,4
j,k = stiffness coefficients 1,2,…,5
u,m = ball iteration coefficients 1,2
n = ball number $1,2,\ldots,Z$

References

[1] Rumbarger, J. H., 1959, "Lubrication of Large Diameter Rolling Bearings," *Proceedings of Conference on Lubrication of Large Bearing and Drive Gearing Systems*, Lincoln Laboratory, MIT, May 13–14.
[2] Pritts, B. A., and Jones, J. I., 1967, "Applying Large Multiload Bearings to Rotating Structures," SAE Paper No. 670107.
[3] Pritts, B. A., 1973, "Swing Bearing Systems for Cranes and Excavators," SAE Paper No. 730738.
[4] Filetti, E. G., and Rumbarger, J. H., 1970, "A General Method for Predicting the Influence of Structural Support Upon Rolling Bearing Performance," ASME J. Lubr. Technol., **92**(1), pp. 121–128.
[5] Sague, J. E., and Rumbarger, J. H., 1977, "Design Criteria to Prevent Core Crushing Failure in Large Diameter, Case Hardened, Ball and Rolling Bearings," ASME Design Engineering Division ASME Paper No. 77-DE-39.
[6] Rumbarger, J. H., 2000, "Design Guideline DG03—Wind Turbine Design, Yaw and Pitch Bearing Life," National Renewable Energy Research Laboratory Report No. DG03_990610.
[7] 1990, "Rolling Bearings—Dynamic Load Ratings and Rating Life," ISO Standard ISO, Geneva, Switzerland, p. 281.
[8] Lundberg, G., and Palmgren, A., 1947, "Dynamic Capacity for Rolling Bearings," Acta Polytech. Scand., Mech. Eng. Ser., **1**(3), pp. 1–50.

[9] Rumbarger, J. H., and Poplawski, J. V., 1994, "Correlating Computerized Rolling Bearing Analysis Techniques to the ISO Standards on Load Rating and Life," STLE Tribol. Trans., **37**(4), pp. 793–801.
[10] Jones, A. B., 1960, "A General Theory for Elastically Constrained Ball and Radial Roller Bearings Under Arbitrary Load and Speed Conditions," ASME J. Basic Eng., **82**, pp. 309–320.
[11] Jones, A. B., 1959, "A Method for Determining the Displacements, Internal Load Distribution and Fatigue Life of Four-Point Contact Ball Bearings Under Thrust, Radial and Moment Loads," Messinger Bearings, Inc., Private Consultant Report.
[12] Rumberger, J. H., 2004, "Analysis of Rolling Bearings With Reduced Numbers of Balls or Rollers," ASME J. Tribol., **126**, pp. 407–411.
[13] Zaretsky, E. V., ed., 1992, *STLE Life Factors for Rolling Bearings*, Chap. 4, STLE Publication SP-34, Park Ridge, IL.

Hydrogen-Enhanced Gasoline Stratified Combustion in SI-DI Engines

Enrico Conte
e-mail: enrico.conte@alumni.ethz.ch

Konstantinos Boulouchos

Aerothermochemistry and Combustion Systems Laboratory,
ETH,
Zurich 8092, Switzerland

Experimental investigations were carried out to assess the use of hydrogen in a gasoline direct injection (GDI) engine. Injection of small amounts of hydrogen (up to 27% on energy basis) in the intake port creates a reactive homogeneous background for the direct injection of gasoline in the cylinder. In this way, it is possible to operate the engine with high exhaust gas recirculation (EGR) rates and, in certain conditions, to delay the ignition timing as compared to standard GDI operation, in order to reduce NO_x and HC emissions to very low levels and possibly soot emissions. The results confirmed that high EGR rates can be achieved and NO_x and HC emissions reduced, showed significant advantage in terms of combustion efficiency and gave unexpected results relative to the delaying of ignition, which only partly confirmed the expected behavior. A realistic application would make use of hydrogen-containing reformer gas produced on board the vehicle, but safety restrictions did not allow using carbon monoxide in the test facility. Thus, pure hydrogen was used for a best-case investigation. The expected difference in the use of the two gases is briefly discussed. [DOI: 10.1115/1.2795764]

Introduction

Transportation systems with near-zero emission potential require innovative technologies to address the important issue of minimization of locally/regionally active pollutants such as NO_x, unburnt hydrocarbons, soot, etc., along with a significant reduction of specific greenhouse gas emissions. In this context, use of hydrogen for automotive power trains, involving either advanced internal combustion engines or fuel cells, is one of the promising future paths to be explored. However, infrastructures as well as technologies for production and storage of hydrogen efficiently and at a large scale are still not available and might be decades away. Hydrogen has very special characteristics that make it at the same time an extremely promising and challenging fuel for internal combustion engines. It is "clean" burning (the major combustion product being water—but emissions must be accounted for at the site of production), is easily ignited, and has wide flammability limits. Nevertheless, some important issues arise, such as onboard storage, safety concern, preignition and backflash, combustion control, emission of NO_x, unburned H_2 and H_2O_2, power density, and some more, not least lack of infrastructure for distribution. In the midterm time frame, combustion of mixtures of hydrogen-containing gas and gasoline appears to be a good solution to combine the major advantages given by both fuels. This would avoid many problems, especially if small amounts of hydrogen are produced on-board directly from gasoline by means of a reformer. Use of hydrogen and gasoline blends seems to be especially suitable for part load operation and reduction of emissions during cold start.

Interest about addition of hydrogen or reformer gas to conventional fuels has been growing over the last years. From large gas engines for stationary applications to passenger car engines, the investigation into this concept has been so far concentrated on spark-ignited homogeneous combustion engines. Our previous works gave a comprehensive characterization of the global engine behavior and of the combustion process through detailed heat release analysis [1], of the capabilities of this concept in the homologation cycles [2], and of the influence of reformer gas on laminar and turbulent premixed flame speed in engine conditions [3].

Our own work stemmed from earlier investigations, which assessed from a global point of view the performance of a real prototype of reformer and showed promising results in terms of emissions [4–6].

Other investigations also showed the combustion characteristics of hydrogen or reformer gas blends with iso-octane or other primary reference fuels in research burners or with simulation tools, mainly at standard conditions (Ref. [7], for example).

During our investigations into premixed combustion, the idea arose of looking at whether the concept was applicable also to the stratified combustion of gasoline in a direct-injection spark-ignition engine. The idea is that small amounts of hydrogen can be injected into the intake port to create a reactive background for the combustion of the gasoline injected into the cylinder. The expected effects were as follows:

- allow delaying the ignition timing as compared to standard gasoline operation, still having stable combustion
- allow the recirculation of large amounts of exhaust gas, still having stable combustion

Both measures are important to reduce the tendency of gasoline direct injection (GDI) engines to produce high quantities of NO_x. In fact, they both make the combustion temperature decreases, the first by giving gasoline more time to evaporate and form a local mixture whose stoichiometry is leaner than the one giving the highest NO_x formation ($\lambda \sim 1.1$), and the second by limiting the temperature increase through the introduction of high heat capacity inert gas to absorb the combustion heat as well as by reducing the oxygen concentration in the oxidizer stream.

In normal gasoline operation, the possible ignition delay and the exhast gas recirculation (EGR) rate are very limited, because they cause ignition and combustion stability to deteriorate. This results in a sudden increase of partial combustion products (unburnt hydrocarbons) and poor engine drivability. In principle, due to the wide flammability limits of hydrogen, both EGR rate and ignition delay can be increased while the combustion process re-

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 18, 2006; final manuscript received September 10, 2007; published online January 22, 2008. Review conducted by Margaret Wooldridge. Paper presented at the 2006 Spring Conference of the ASME Internal Combustion Engine Division (ICSE2006), Aachen, Germany, May 6–10, 2006.

Table 1 Engine specifications

Engine	Daimler Benz Basis M 111 One-cylinder Stroke: 86.6 mm, Bore: 89.9 mm Compression ratio: 10:1 Two intake valves, One exhaust valve
Injection	Intake channel injection from Bosch Direct injection from Siemens: (hollow-cone injector); Rail pressure: 80 bars
Ignition system	Coil from BMW, Transistor from Bosch Spark plug from NGK
Brake	ELIN, asynchronous motor/generator
EGR	Water-cooled, one-way valve, water trap
Oil and water conditioning	Oil and water pump: external actuation Electrical heating with water heat exchanger Monitoring elements for temperatures and pressures

mains stable. NO_x emissions can therefore be reduced. Unburnt hydrocarbons are also significantly reduced and the brake efficiency increased. As a side effect, also a reduction of the soot emissions can be expected, but this was not measured in our experiments.

To the authors' knowledge, the extension of the hydrogen/gasoline blending concept to DI engines was investigated before only in Ref. [8]. In that work, methane gas was injected in a lean hydrogen-air mixture. Results from that research showed that the combustion system could achieve lower hydrocarbon exhaust emission and higher thermal efficiency compared to methane DI spark-ignition combustion. However, the premixed hydrogen increased the cooling losses to combustion chamber wall. The combustion system of methane DI stratified combustion in hydrogen lean mixture showed to be effective especially in lean conditions.

Experimental Setup

In this investigation, a single-cylinder research engine was used. The engine is a modification of a Mercedes Benz M111, whose specifications are listed in Table 1. The cylinder head (Fig. 1) was originally equipped with four valves. For this application, one exhaust valve seat was shut. The spark plug was moved from the central position to a side position in the space freed by the shut valve, close to the gasoline injector, which was placed in the central position. The hollow-cone injector is from Siemens. The system is spray guided. The spark plug is NGK iridium. This was chosen for its better performance as opposed to a conventional Bosch plug. The very thin electrode of the NGK plug is believed to offer a smaller surface for soot to deposit and compromise the performance of the ignition source.

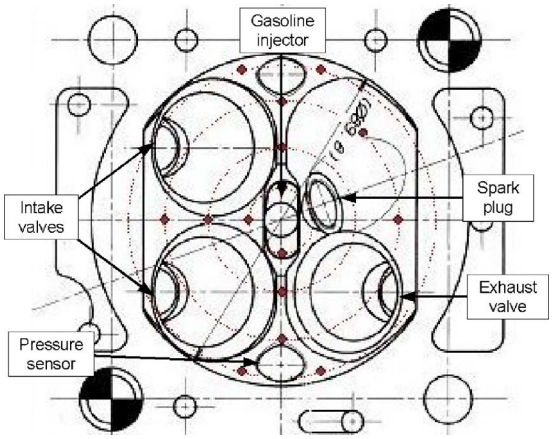


Fig. 1 Cylinder head

Gasoline is injected at high pressure (80 bars) through a common rail injection system. For the injection of hydrogen in the intake port, two special injectors from Bosch and suitable drivers were used. Each injector was placed in one of the two channels leading to the two intake valves.

One engine speed (2000 rpm), two loads (3 bar and 5 bar indicated mean effective pressure (IMEP)—corresponding to slightly more than 2 bar and 4 bar brake mean effective pressure (BMEP)) and different combinations of EGR rate and hydrogen additions were investigated. Measurements were carried out at pure gasoline operation, 17% and 27% hydrogen contents (energy fraction). At 3 bar IMEP, use of EGR was also investigated, up to 30% at 27% H_2. The engine was operated unthrottled. Airflow measurements were not available on the test bench.

Every measurement consisted of a sweep of the ignition timing with fixed injection end. The injection duration of gasoline and hydrogen was varied according to the desired load and the relative fraction of the two fuels. In particular, for every combination of hydrogen enrichment and EGR, the total amount of fuel was determined as to have always the same IMEP at the best ignition timing for IMEP. The amount of the two fuels was then kept constant during the ignition-timing sweep. The injection timing of hydrogen in the intake port was set at $-320°$CA (crank angle) (during the intake stroke). The injection timing of gasoline was set to end injection always at about $-55.6°$CA (electronic). The start of injection was varied depending on the amount of fuel to be injected.

The pressure indication was averaged over 290 cycles. The heat release rate was calculated using our internal code, which is based on a detailed two-zone model. The calculation is based upon measured quantities (cylinder pressure, fuel consumption, EGR rate, and others). The code takes into account the mass transfer in the piston ring gap and the mass losses due to blow-by. For the determination of the state-dependent gas properties c_v, c_p, etc., both dissociation and real gas properties are taken into account. The wall heat losses are estimated through the Woschni model.

In what follows, the amount of hydrogen is always expressed as percent fraction of the total energy yielded by hydrogen. The coefficient of variation (COV) of the IMEP is a global indication of the engine stability:

$$\mathrm{COV}_{\mathrm{IMEP}} = \frac{\sigma_{\mathrm{IMEP}}}{\mathrm{IMEP}} \times 100[\%]$$

Here, σ_{IMEP} is the standard deviation of the indicated mean effective pressure, calculated over the whole set of single engine cycles. A value of about 10% is normally considered to be the ultimate drivability limit [9] but it is rather high if compared to what is normally accepted as standard for car engines. The present engine exhibited in many operating conditions a COV of about 10% or higher. It should, however, be taken into account that this is a single-cylinder engine and that it was neither originally designed nor optimized for direct injection operation.

Results

3 Bar Indicated Mean Effective Pressure Without Exhaust Gas Recirculation. Figure 2 illustrates engine stability and power output versus ignition timing at pure gasoline operation and at 17% and 27% hydrogen additions. In general, hydrogen additions improves the stability of the combustion process. When the fraction of hydrogen in the fuel blend becomes larger, the optimal ignition point for best stability (lowest COV) lies earlier in the cycle. As expected, hydrogen addition widens the ignition window, but the effect is clearly stronger for early ignition timings than for late ignition. In all cases, the deterioration of the combustion stability is very sharp for too early ignition and more gradual for late ignition.

It is, in general, known that addition of hydrogen to gasoline increases the flame speed and widens the flammability limits, but

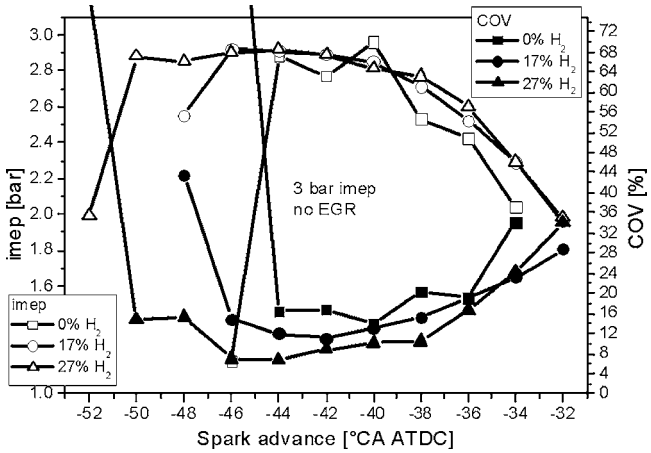


Fig. 2 IMEP and COV_{IMEP} versus ignition timing at gasoline operation, 17% and 27% hydrogen enrichments, without EGR

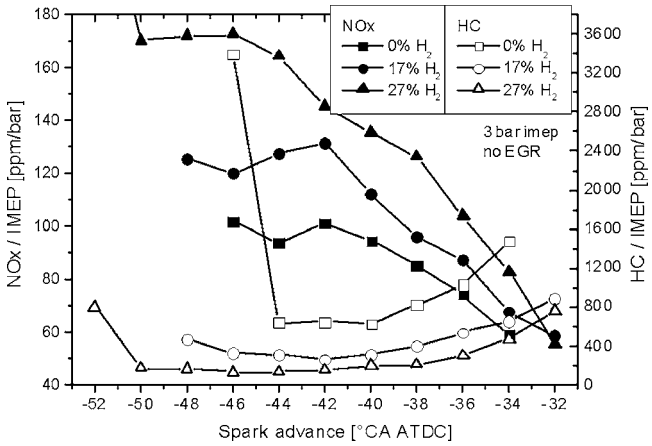


Fig. 4 NO_x and HC emissions versus ignition timing at 3 bar IMEP, without EGR

the effect is much stronger in rich mixtures than in lean mixtures. This means that rich mixtures between hydrogen and gasoline are more reactive and easily ignitable than lean mixtures. Figure 3 illustrates in a qualitative way how the mixture stoichiometry changes with time at the spark plug location during injection and evaporation of the gasoline spray. This kind of stoichiometry evolution was demonstrated through CFD simulation of a gasoline spray in the same gasoline direct injection (GDI) combustion chamber and the optimal ignition window was confirmed by the measured engine stability [10]. A gasoline/air mixture can be ignited in a narrow window around the optimal point, which is when $\phi \approx 1$ at the ignition source. The addition of hydrogen widens the stoichiometry window in which the mixture can be ignited, but mainly on the rich side. This is reflected in a widening of the ignition window more to earlier timing than to late timing. This agrees well with the results in Fig. 2. In general, the engine load changes the stoichiometry profile at the spark plug, because the curve tends to a higher value for large t, due to the larger amount of fuel for roughly the same air. This effect changes significantly the result in terms of extension of the ignition window, as it will be shown later.

Figure 4 illustrates how NO_x and HC emissions vary with the ignition timing at pure gasoline operation and at two different degrees of enrichment. The emissions are normalized by the IMEP. HC emissions are generally lowered by hydrogen enrichment in the whole ignition window. This is due to better combustion stability and to the lower fraction of gasoline in the fuel mixture. Instead, NO_x emissions increase in the whole range, due to the generally higher combustion temperatures produced by hydrogen. The experience from homogeneous combustion of gasoline-reformer gas blends [1] suggests that use of reformer gas would give same or slightly lower NO_x formation than at pure gasoline operation, because the reformer gas contains large amounts of inert N_2, whose thermal capacity limits the temperature increase due to hydrogen. The increase of NO_x is particularly evident for early ignition, when a richer mixture is burnt. This confirms the previous observations regarding the stronger effect of hydrogen in rich mixtures. Also for late ignition timings, hydrogen addition increases NO_x formation as compared to pure gasoline operation. The little extension of the ignition window to late timings is not enough to bring down NO_x to lower values than at gasoline operation. However, the HC-NO_x trade-off (Fig. 5) is clearly better than at pure gasoline operation. In fact, a significant reduction of HC emission can be achieved with H_2 enrichment at same NO_x levels as at pure gasoline operation.

Figure 6 shows NO_x emissions versus engine stability, expressed as COV_{IMEP}. This allows comparing emissions when using hydrogen enrichment at the same level of stability as at pure gasoline operation. Although the present system is not a state-of-the-art GDI engine, it is clear that hydrogen addition improves the quality and repeatability of combustion, but at the cost of higher NO_x emissions. This is true also if the ignition timing is delayed as much as possible, having as constraint that the same engine

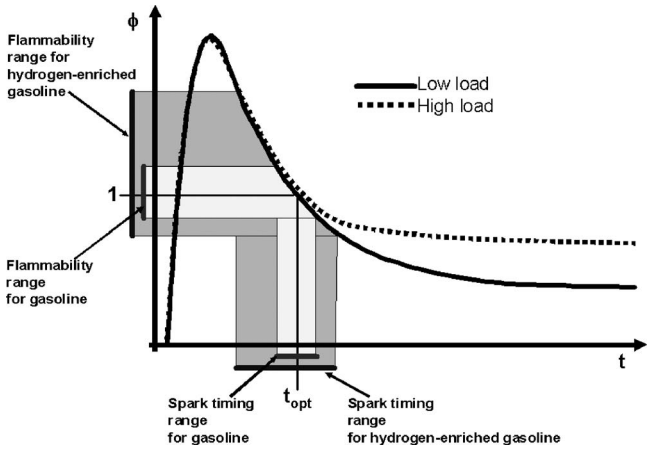


Fig. 3 Qualitative evolution of the mixture stoichiometry with time at the location of the ignition source, for gasoline and hydrogen additions, at low load

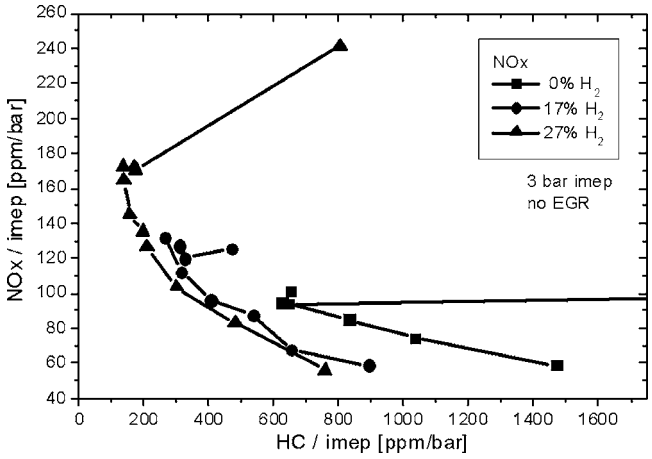


Fig. 5 NO_x versus HC emissions at 3 bar IMEP, without EGR

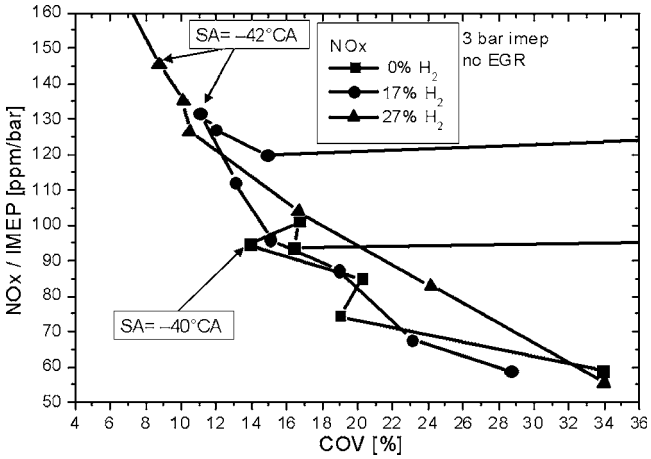


Fig. 6 NO_x emissions versus engine stability (COV_{IMEP}) at 3 bar IMEP, without EGR

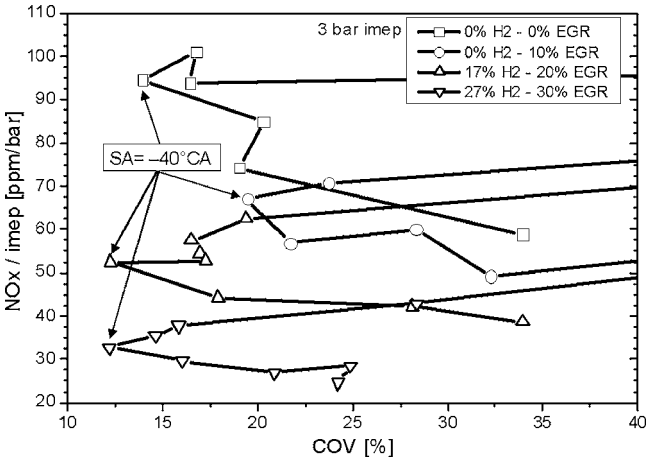


Fig. 8 NO_x–COV trade-off for some combinations of hydrogen enrichment and EGR rates

stability as at pure gasoline operation is met.

In synthesis, hydrogen enrichment per se increases NO_x emissions. The widening of the ignition window is larger for early ignition than for late ignition. Late ignition is not enough, alone, to reduce NO_x production from pure gasoline levels. HC emissions are lowered.

Exhaust Gas Recirculation at 3 Bar Indicated Mean Effective Pressure. Hydrogen enrichment allows using higher EGR rates than at pure gasoline operation for the same engine stability. At pure gasoline operation, already a 10% EGR rate causes the combustion stability to deteriorate, but the power output is slightly higher, because of better efficiency. At 17% hydrogen operation, the deterioration of the combustion stability caused by EGR is smaller, and less so for even higher amount of hydrogen. This allowed reaching 20% EGR at 17% hydrogen operation and 30% EGR at 27% hydrogen operation. HC emissions are only slightly affected by the EGR rate increase, unless very high EGR rates (30% at 27% hydrogen) are used. On the other hand, 10% EGR is already very effective in reducing NO_x formation and 30% EGR at 27% hydrogen operation can abate NO_x by a factor of 5 as compared to pure gasoline without EGR. In Fig. 7, the NO_x emissions for all the strategies illustrated so far are plotted. Once again, it is clear that, at same EGR rate, hydrogen addition increases NO_x formation as compared to pure gasoline operation and that the later ignition allowed by its use is never enough to compensate the higher NO_x formation. On the other hand, the more hydrogen is used, the higher EGR rates can be achieved, with significant abatement of NO_x production.

Figure 8 illustrates the trade-off between NO_x and combustion stability, for some combinations of hydrogen enrichment and EGR rates. If the same engine stability as in the most stable point at pure gasoline operation is taken as target, NO_x emissions can be reduced by adding hydrogen and at same time increasing EGR by a factor of 3 in the best case, which is 27% hydrogen enrichment with 30% EGR.

The trade-off between NO_x and HC emissions (Fig. 9) shows that hydrogen addition and high EGR rates can effectively abate the emissions of both pollutants. It is interesting to notice that the lowest HC emissions can be reached already with 17% hydrogen addition.

Figure 10 shows the trade-off between indicated efficiency and NO_x emissions. The indicated efficiency, which is calculated as ratio between IMEP and fuel energy, increases noticeably with hydrogen addition. Also, EGR increases the engine efficiency because the lower combustion temperatures reduce the thermal losses. When the 27% H_2–30% EGR operation point is compared to the standard gasoline operation without EGR, the advantage in terms of both efficiency and NO_x is considerable.

The extent of the efficiency increase given by hydrogen addition is certainly astonishing and requires some more discussion. At this regard, measurements were carried out with homogeneous

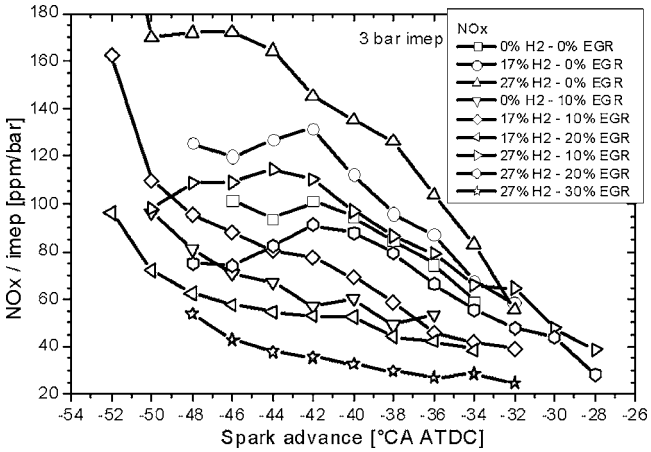


Fig. 7 Comparison of NO_x emission for different EGR rates and hydrogen enrichment degrees

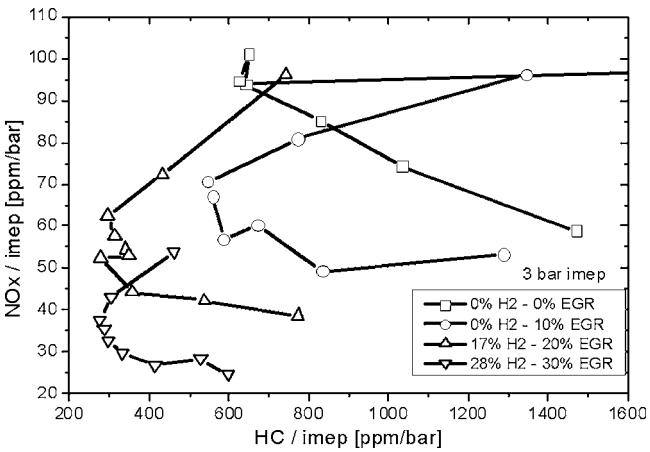


Fig. 9 NO_x–HC trade-off for some combinations of hydrogen enrichment and EGR rates

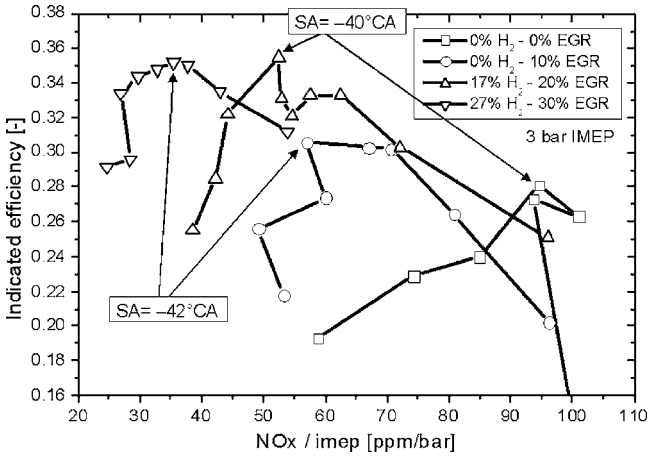


Fig. 10 Trade-off between indicated efficiency and NO_x for some combinations of hydrogen enrichment and EGR rates

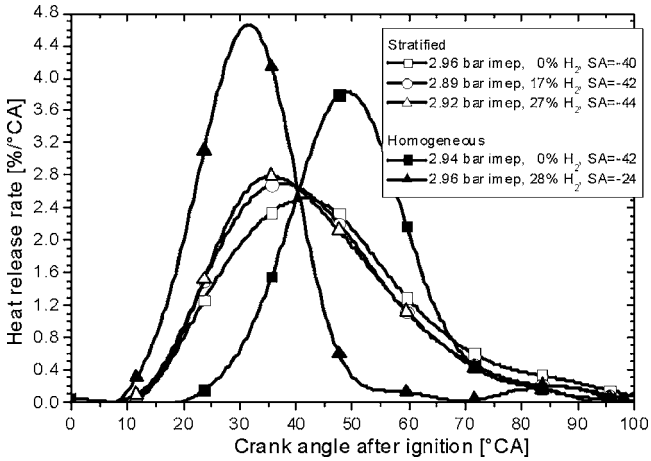


Fig. 11 Heat release rate for stratified operation, without EGR. For comparison, also the heat release rate at homogeneous operation is plotted

charge (throttled, $\lambda = 1$), at same load and level of hydrogen addition, to compare the efficiency increase. Table 2 summarizes the most important results. First of all, it appears that, at homogeneous operation, hydrogen addition gives a much smaller efficiency increase, stemming from reduced pumping losses and shorter combustion, which might be compensated by higher heat losses due to higher combustion temperatures. The second important aspect is that the engine efficiency at pure gasoline homogeneous operation is higher than that at pure gasoline stratified operation, although the pumping losses are four times higher. This leads to the conclusion that the combustion quality of this system is not optimal at stratified operation.

The efficiency increase given by hydrogen addition must therefore be ascribed to improved combustion quality to the largest extent. If one considers the ratio between HC emissions (measured as C_1) and CO_2 emissions, it appears that the fraction of fuel, which is not or only partially burnt, is rather large and can account for the missing energy, more so if one considers that part of the unburnt hydrocarbons is also oxidized in the exhaust tract and does not reach the gas analysis. The ratio between HC and CO_2 decreases with hydrogen addition and reaches a level comparable to homogeneous combustion. If one now compares stratified and homogeneous combustion with same hydrogen addition, the difference in efficiency is, as expected, directly linked to the decreased pumping losses (pumping mean effective pressure (PMEP)), as the combustion quality is now roughly comparable. In conclusion, this analysis reveals that the observed increase of combustion efficiency is to some extent specific to the GDI system used here and should not be generalized. Nevertheless, it is clear that, besides the effect on NO_x—which is the main focus of this investigation—small amounts of hydrogen have a striking effect on combustion quality and also in a modern engine could help to reduce HC and soot emissions remarkably. Moreover, when looking at the efficiency increase produced by the hydrogen addition, it should not be forgotten that, in a real application, reformer gas (more likely than pure hydrogen) would be obtained by partially oxidizing gasoline through a reformer. The efficiency of a gasoline reformer for internal combustion engine applications can be as high as 85–90% (on a lower heating value (LHV) basis) [11–13]. Our investigations into the combustion properties of homogeneous reformer gas/gasoline blends showed that the efficiency benefit could be large enough to compensate and even overcome the energy loss in the reformer [1].

3 Bar Indicated Mean Effective Pressure: Combustion Analysis. Figure 11 illustrates the rate of heat release at 3 bar IMEP, without EGR, for the ignition timing giving best stability. Also, the heat release rate at homogeneous operation for the ignition timing giving best stability is plotted for comparison. All the curves are shifted to account for the different ignition timings. In the homogeneous case, hydrogen addition gives a faster start of combustion as compared to combustion of pure gasoline. In stratified combustion instead, the heat release rate in the first phase of combustion appears to be similar for all hydrogen enrichment degrees. This is because in stratified combustion the limiting factor is the evaporation of gasoline and the formation of an ignitable mixture with air and hydrogen. After combustion has started, the heat release rate is higher and combustion slightly faster with increasing H_2 enrichment. As it will be shown later, at higher load this behavior is different, because hydrogen alone is already in an ignitable stoichiometry with air; therefore, the evaporation and mixing of gasoline is less a limiting factor.

Figures 12–14 show the duration in °CA of the three main combustion phases, 0–5%, 5–50%, and 50–90% of the total energy release (integral of the heat release rate, IHRR), at pure gasoline operation, 17% and 27% hydrogen enrichments, without EGR. In all cases, delaying the ignition timing has a major effect

Table 2 Comparison of stratified and homogeneous combustion, without EGR. For each strategy, the ignition timing giving the highest IMEP is selected.

Type	H_2 (%)	PMEP (bar)	PMEP/IMEP (%)	COV (%)	Indicated efficiency	HC/CO_2 (%)	CO/CO_2 (%)
Strat.	0	–0.134	4.5	13.9	0.280	6.4	6.6
Strat.	17	–0.207	7.2	11.09	0.326	2.9	4.6
Strat.	27	–0.225	7.7	6.75	0.359	1.7	3.4
Homog.	0	–0.643	21.9	2.5	0.296	0.5	5.2
Homog.	28	–0.55	18.6	2.06	0.318	0.3	2.4

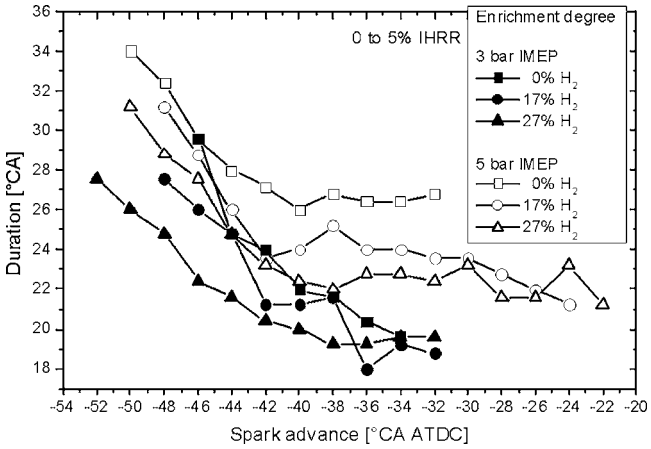


Fig. 12 Duration of the 0–5% combustion phase at different degrees of hydrogen enrichment, without EGR

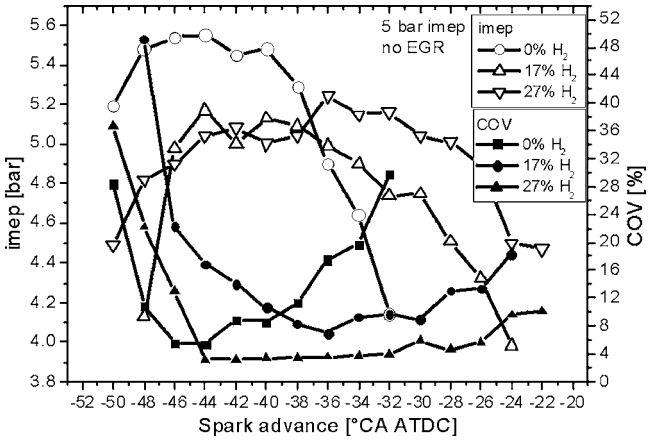


Fig. 15 IMEP and COV versus ignition timing at gasoline operation, 17 and 27% hydrogen additions, without EGR, 5 bar IMEP

on the first phase of combustion. The more the ignition is delayed, the more the liquid fuel has time to evaporate and mix with air. When the mixture is finally ignited, the combustion rate is higher, especially in the region closer to the ignition source, where the first 5% is burnt. If one compares how the duration of combustion changes with the hydrogen amount, it results that for increasing quantities of hydrogen the 0–5%, the 5–50% and the 50–90% become faster in the whole range of investigated spark timings, but the duration decrease is not large, especially from 0% to 17% H_2-enrichment degree.

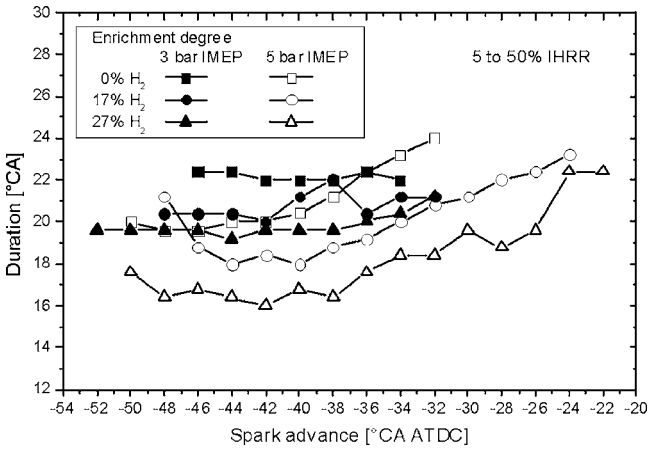


Fig. 13 Duration of the 5–50% combustion phase at different degrees of hydrogen enrichment, without EGR

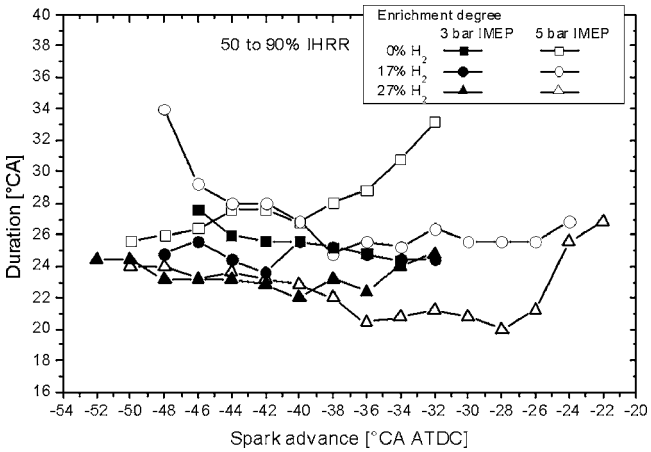


Fig. 14 Duration of the 50–90% combustion phase at different degrees of hydrogen enrichment, without EGR

When comparing the duration of the 0–5% energy release from pure gasoline to 27% hydrogen enrichment (Fig. 12), an interesting behavior appears: The higher the hydrogen enrichment degree, the less the combustion duration seems to be affected by the delaying of the ignition timing. For late ignition timing at 27% hydrogen, the curve tends to be flat. A possible explanation is that at higher hydrogen enrichment degrees, the quantity of injected gasoline is smaller and ignition can be delayed more. This means that the smaller amount of gasoline in the vicinity of the spark plug has more time to evaporate and mix up with air. The combustion process tends therefore to shift from less to more homogeneous. For early ignition, the evaporation process is believed to be the controlling mechanism that determines the duration of combustion in the first 5% energy release, whereas for very late ignition, the combustion speed is believed to become the main controlling factor, which is not influenced by the ignition timing.

5 Bar Indicated Mean Effective Pressure, Without Exhaust Gas Recirculation. Figure 15 illustrates how power output and engine stability vary with the spark advance at higher load (5 bar IMEP), without EGR. The measurement with pure gasoline was taken at a slightly higher load (~10% higher than the target, 5 bar IMEP). This explains why the curve is slightly shifted to earlier ignition timings than the curves at 17% and 27% hydrogen. Taking this into account, it is possible to notice that, as already showed at lower load, hydrogen addition increases the engine stability at early ignition timings, but in this case the most striking effect is the stability improvement at late ignition timings, which was not observed at lower load. Figure 16 helps explaining the possible reason for this difference. At higher load, the global fuel-to-air ratio ϕ is higher and this changes the profile of the stoichiometry at the spark plug location; in fact, ϕ tends to a higher value for large t. The more hydrogen is used, the more the flammability range is extended, mainly on the rich side. For large amounts of hydrogen, the extension on the lean side though tends to encompass completely the stoichiometry profile at the spark plug for large t, making possible to delay ignition to very late timings. In fact, when carrying out the experiments, it was clear that at higher load, as opposed to lower load, combustion of hydrogen was taking place also without injection of gasoline (although in this case no net power was delivered by the engine) because the stoichiometry was close or within the flammability limit of the hydrogen/air mixture.

For what concerns emissions (Fig. 17), at higher load as at

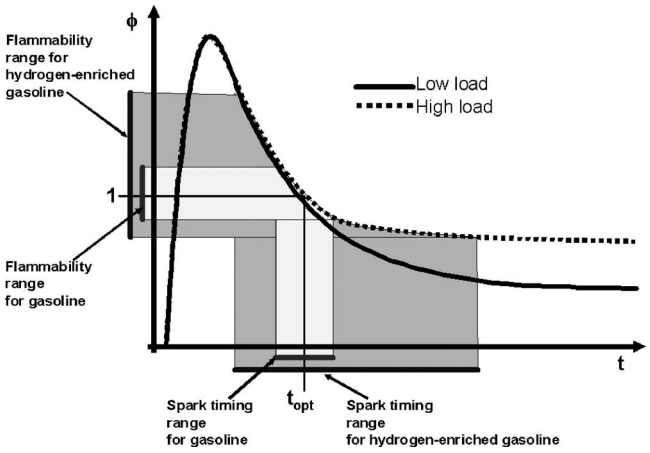


Fig. 16 Qualitative evolution of the mixture stoichiometry with time at the location of the ignition source, for gasoline and hydrogen additions, at high load

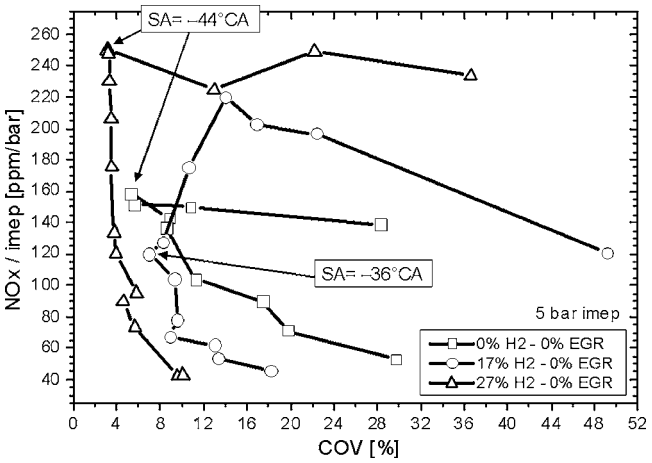


Fig. 18 NO_x–COV trade-off at high load (5 bar IMEP)

lower load hydrogen enrichment significantly reduces HC emissions. In this case, when 27% hydrogen is used, HC emissions remain very low even if the ignition is set to very late timings. In general, as at lower load, NO_x emissions increase when hydrogen is used, but in this case the large ignition delay allowed by hydrogen addition makes possible to reduce NO_x formation and keep it at same level as at pure gasoline operation. If one compares NO_x emissions at same engine stability (Fig. 18), it is clear that at high load a remarkable NO_x reduction (~50% from the most stable point at pure gasoline operation) is possible even without using EGR. The trade-off between NO_x and HC emissions is shown in Fig. 19. By delaying the ignition timing, very low emissions of both pollutants can be achieved, even without EGR. Use of EGR at high load can be reasonably expected to bring an even more striking reduction.

The efficiency increase given by hydrogen is evident in the whole range of ignition timings. For very early or very late ignition, it is due to better stability. Also for the best ignition timing, the increase of efficiency is remarkable. For instance, at –42°CA ignition, the indicated efficiency increases from 34% at pure gasoline operation to 36% at 17% hydrogen enrichment to 41% at 27% hydrogen enrichment. This is the result of better combustion stability, more complete combustion (as previously discussed regarding operation at low load) and shorter combustion, which in the considered point decreases from 74°CA duration at pure gasoline operation to 62°CA at 27% hydrogen operation. Figure 20 illustrates the trade-off between indicated efficiency and NO_x emissions. At same efficiency as in the best gasoline operation point, NO_x emissions can be curbed by a factor of 4 at 27% hydrogen operation from pure gasoline operation, just by delaying ignition, without using EGR.

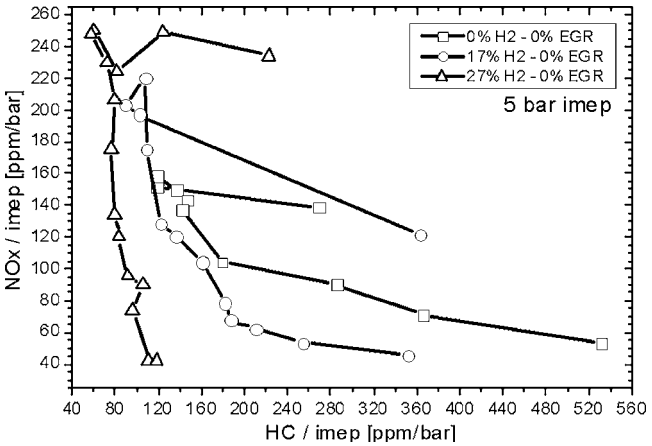


Fig. 19 NO_x–HC trade-off at high load (5 bar IMEP)

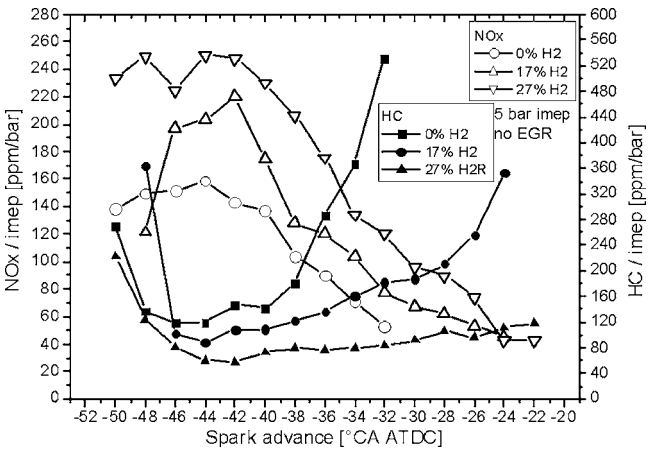


Fig. 17 NO_x and HC emissions, 5 bar IMEP, without EGR

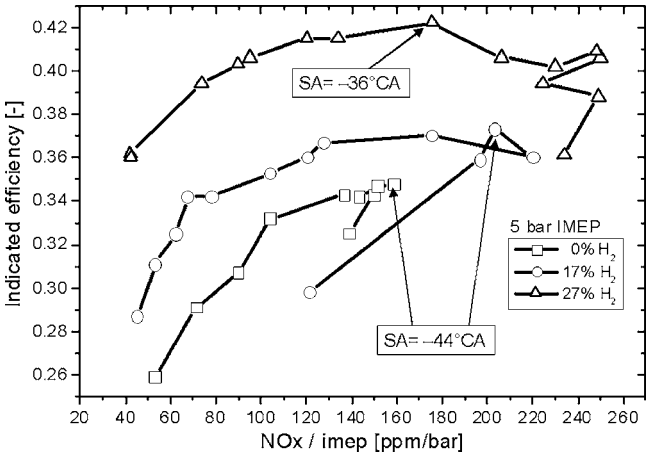


Fig. 20 Indicated efficiency versus NO_x emissions, for increasing hydrogen enrichment degree. 5 bar IMEP, without EGR

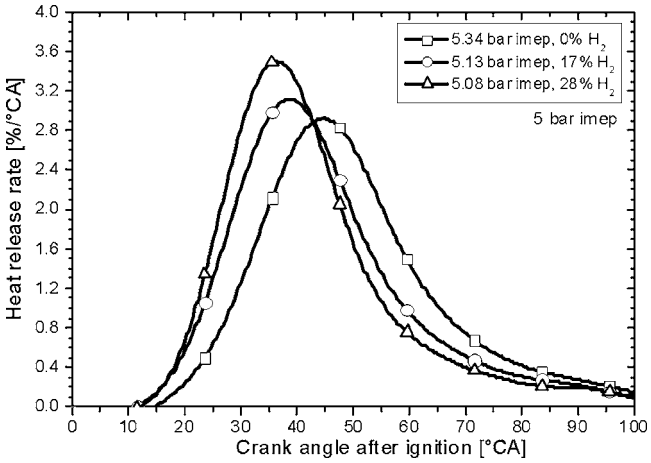


Fig. 21 Heat release rate at 5 bar IMEP, without EGR

5 Bar Indicated Mean Effective Pressure: Combustion Analysis. Figure 21 illustrates the rate of heat release. The curves are shifted to take into account the different ignition timings. A different behavior than at lower load can be observed in the way energy is released. In this case, hydrogen addition produces a higher combustion rate since the beginning, because at higher load the global ϕ is higher and the hydrogen/air mixture is already ignitable or very close to it; therefore, the evaporation of gasoline is less a limiting factor for the start of the combustion process. Figures 12–14 illustrate the duration of the 0–5%, 5–50%, and 50–90% phases. Similar trends as at lower load can be observed. In general, hydrogen enrichment shortens the duration of the whole combustion process. In the 0–5% phase, as at lower load, delaying ignition makes the onset phase faster. As already explained, at high load it is possible to delay ignition significantly more than at low load. For very late ignition, the duration of the 0–5% phase seems to be less affected by the spark timing, if at all. This is due to the fact that the fuel has reached a higher degree of mixing with air; therefore, the development of the combustion process is less limited by the spray evaporation. In the 5–50% phase, extreme ignition delays make the combustion duration increase, because the mixture has leaned out, causing slower flame propagation. In the final phase, 50–90%, where the mixture is anyhow very lean, the combustion duration for the three hydrogen enrichment degrees seems to be not much affected by the spark timing. The durations of the completion phase for early ignition at 17% hydrogen and for late ignition at 0% hydrogen are an exception, whose reason is not completely clear. The comparison of the combustion durations between 3 and 5 bar IMEP can be interpreted as follows: In general combustion in the 0–5% phase is rich. At high load, the mixture is richer than at low load, giving slower flame speeds and longer duration. Combustion in the 5–50% phase is probably stoichiometric or lean, but at low load the mixture is leaner than at high load, giving slower flame speeds and longer duration. Finally, when the mixture is anyhow very lean, as in the 50–90% phase, there is no large difference in the flame speeds at low or high load, and the combustion durations lie in the same range.

Conclusions

The different combinations of global and local stoichiometries at the spark plug location give different results in terms of ignitability and NO_x formation, depending on H_2 fraction, EGR rate, and spark delay. Accounting for all the concurring effects is not trivial but can shed light into interesting phenomena.

At lower load (lower global ϕ) the small spark delay allowed by H_2 addition alone is not enough to compensate the higher NO_x production. On the other hand, at lower load H_2 addition allows reaching EGR rates high enough to significantly curb NO_x emissions.

At higher load (higher global ϕ) the large spark delay allowed by H_2 enrichment is able alone to limit substantially NO_x production. Use of EGR at higher load is expected to give even better results.

The indicated efficiency increases with H_2 addition and EGR operation. This is due to the concurring effects of better engine stability, lower HC emissions, and some faster combustion. In all conditions, HC emissions are substantially lowered by hydrogen addition.

In general, H_2 addition gives extraordinary good results in terms of NO_x, HC, efficiency, and stability when considering the trade-offs among these quantities. At low load and with 30% EGR, NO_x can be reduced by a factor of 3 at same engine stability as at pure gasoline operation.

Use of reformer gas instead of pure hydrogen can be expected to give lower NO_x emissions, as the inert content would limit the temperature increase caused by hydrogen combustion. On the other hand, perhaps lower EGR rates could be achieved when using reformer gas instead of pure hydrogen. Also, use of reformer gas would produce a higher global ϕ, because of the inert gas displacing air; therefore, an effect on ignition delay similar to what observed at higher load with pure hydrogen could be expected.

Nomenclature

ATDC = After top dead centre
B = bore (mm)
BMEP = brake mean effective pressure (bar)
°CA = crank angle (deg)
COV = coefficient of variation of the IMEP (%)
EGR = exhaust gas recirculation (%)
ϕ = fuel-to-air relative mass ratio, $\phi = 1/\lambda$
HRR = heat release rate (%/°CA)
IHRR = integral of the heat release rate (%)
λ = air-to-fuel relative mass ratio: $(\text{air/fuel})/(\text{air/fuel})_{\text{stoichiometric}} = 1/\phi$
IMEP = indicated mean effective pressure (bar)
LHV = lower heating value (MJ/kg)
MBT = maximum brake torque (N m)
PMEP = pumping mean effective pressure (bar)
S = stroke (mm)
SA = spark advance (°CA)

References

[1] Conte, E., and Boulouchos, K., 2004, "Influence of Hydrogen-Rich Gas Addition on Combustion, Pollutant Formation and Efficiency of an IC-SI Engine," SAE Paper No. 2004-01-0972.
[2] Allgeier, T., Klenk, M., Landenfeld, T., Conte, E., Boulouchos, K., and Czerwinski, J., 2004, "Advanced Emission and Fuel Economy Concept Using Combined Injection of Gasoline and Hydrogen in SI-Engines," SAE Paper No. 2004-01-1270.
[3] Conte, E., and Boulouchos, K., 2005, "A Quasi-Dimensional Model for Estimating the Influence of Hydrogen-Rich Gas Addition on Turbulent Flame Speed and Flame Front Propagation in IC-SI Engines," SAE Paper No. 2005-01-0232.
[4] Kirwan, J. E., Quader, A. A., and Grieve, M. J., 1999, "Advanced Engine Management Using On-Board Gasoline Partial Oxidation Reforming for Meeting Super-ULEV (SULEV) Emission Standards," SAE Paper No. 1999-01-2927.
[5] Kirwan, J. E., Quader, A. A., and Grieve, J. M., 2002, "Fast Start-Up On-Board Gasoline Reformer for Near Zero Emissions in Spark-Ignition Engines," SAE Paper No. 2002-01-1011.
[6] Quader, A. A., Kirwan, J. E., and Grieve, J. M., 2003, "Engine Performance and Emissions Near the Dilute Limit With Hydrogen Enrichment Using and On-Board Reforming Strategy," SAE Paper No. 2003-01-1356.
[7] Huang, Y., Sung, C. J., and Eng, J. A., 2004, "Laminar Flame Speeds of Primary Reference Fuels and Reformer Gas Mixtures," Combust. Flame, **139**(3), pp. 239–251.

[8] Shudo, T., and Tsuga, K., 2001, "Analysis of Direct Injection SI Stratified Combustion in Hydrogen Lean Mixture," Int. J. Automotive Technology, **2**(3), pp. 85–91.
[9] Heywood, J. B., 1988, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.
[10] Schanzlin, K., Koch, T., Tzannis, A. P., and Boulouchos, K., 2001, "Characterization of Mixture Formation in a Direct-Injected Spark Ignition Engine," SAE Paper No. 2001-01-1902.
[11] Bromberg, L., Cohn, D. R., Rabinovich, A., and Heywood, J. B., 2000, *Emissions Reduction Using Hydrogen From Plasmatron Fuel Converters*, MIT, Cambridge, MA.
[12] Houseman, J., and Cerini, D. J., 1974, "On-Board Hydrogen Generator for a Partial Hydrogen Injection Internal Combustion Engine," SAE Paper No. 740600.
[13] Bromberg, L., Cohn, D. R., Rabinovich, A., and Alexeev, N., 2000, "Experimental Evaluation of SI Engine Operation Supplemented by Hydrogen Rich Gas From Compact Plasma Boosted Reformer," SAE Paper No. 2000-01-2206.

Effect of Piston Friction on the Performance of SI Engine: A New Thermodynamic Approach

E. Abu-Nada
I. Al-Hinti
A. Al-Sarkhi
B. Akash

Department of Mechanical Engineering,
Hashemite University,
Zarqa 13115, Jordan

This paper presents thermodynamic analysis of piston friction in spark-ignition internal combustion engines. The general effect of piston friction on engine performance was examined during cold starting and normal working conditions. Considerations were made using temperature-dependent specific heat model in order to make the analysis more realistic. A parametric study was performed covering wide range of dependent variables such as engine speed, taking into consideration piston friction combined with the variation of the specific heat with temperature, and heat loss from the cylinder. The results are presented for skirt friction only, and then for total piston friction (skirt and rings). The effect of oil viscosity is investigated over a wide range of engine speeds and oil temperatures. In general, it is found that oils with higher viscosities result in lower efficiency values. Using high viscosity oil can reduce the efficiency by more than 50% at cold oil temperatures. The efficiency maps for SAE 10, SAE 30, and SAE 50 are reported. The results of this model can be practically utilized to obtain optimized efficiency results either by selecting the optimum operating speed for a given oil type (viscosity) and temperature or by selecting the optimum oil type for a given operating speed and temperature. The effect of different piston ring configurations on the efficiency is also presented. Finally, the oil film thickness on the engine performance is studied in this paper. [DOI: 10.1115/1.2795777]

Introduction

In most previous studies on air-standard power cycles, friction is generally neglected for simplicity of the analysis [1–4]. However, due to the high revolution speed of the engine, this assumption becomes less realistic where a large percentage of engine power is dissipated into friction. Although air-standard power cycle analysis gives only approximation to the actual conditions and outputs [5], it would be very useful to study the cycle by including the effect of friction.

It is well known that reduction of engine mechanical friction increases the engine efficiency. Previous studies modeled the behavior of Otto, Diesel, and dual cycles including friction [6–8]. However, they dealt with average values of piston coefficient of friction and neglected type of lubricants and engine configuration details such as engine skirt and rings. Also, the dependence of friction on oil temperature was neglected in these studies. On the other hand, some studies have correlated the friction encountered in engines using empirical formulas [9,10]. In fact, these formulas are written in terms of engine speed and such correlations neglected engine operational details such as oil type, skirt, and ring configurations.

It is a fact that a large percentage of the mechanical friction loss in engines occur on the lubricated surfaces between the skirt and the cylinder liner as well as between the cylinder rings and cylinder liner [11]. The lubrication between rings and cylinder liner is effected by oil viscosity, oil film thickness, piston ring configuration, and the operational specifications of the engine. Besides, the friction between the piston skirt and the cylinder liner is affected by the clearance, piston tilt, piston skirt design, and surface roughness [12]. By considering energy consumption within the engine, it is found that friction loss contributes the major portion of the energy consumption developed in an engine. About two-thirds of it is caused by piston skirt friction, piston rings, and bearings, and the other third is due to the valve train, crankshaft, transmission, and gears [13]. Similar figures were reported by Kim et al. [14], who were successful in reducing engine friction through liner rotation. Piston ring lubrication models were used for prediction of engine cylinder friction. An accurate representation of lubrication conditions at the piston ring-cylinder liner interface is required for the estimation of frictional losses. Examples of the work conducted on modeling and analysis of the frictional losses at the piston ring-cylinder liner contact are given in Refs. [15–27]. Moreover, Xu et al. [28] presented two theoretical models that predict friction for piston ring and cylinder liner. Their work was compared to experimental findings. Also, Xu et al. proposed an inter-ring gas flow model by considering the effect of orifice flow through the ring end gap and ring side clearance [29].

It was found that piston ring friction force was higher than previously predicted. Also, it was pointed that fuel economy improvements exceeding 4% may result from combined application of reducing lubricant viscosity and proper surface treatment [30]. Their work was based on numerical findings using four different simulation methodologies, such as RINGPAK. Other studies have used RINGPAK tool in their engine friction simulation [31,32]. Other models are also available in literature [33,34].

The effect of viscosity on the oil film thickness and on the total friction force is significant. The oil viscosity is temperature dependent and under different engine conditions during cold start, normal operation, or severe operation condition, the viscosity of the lubricating oil changes significantly. A model for the prediction of the engine friction, including the viscosity effects, has been presented by Taylor [35]. The results include simulations for fully warmed-up and cold-start conditions, where the total engine friction is investigated. Accordingly, the total engine friction immediately after a cold start is about five times higher than the warmed-up conditions [35]. Further studies have been made on the oil film thickness of piston rings and effect of load and speed on oil film thickness [36]. Recently, more research has focused on the effect of oil film temperature on its thickness of piston rings [37–40]. Their results indicated that the oil film thickness could be

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received July 19, 2006; final manuscript received August 27, 2007; published online January 22, 2008. Review conducted by Thomas W. Ryan III. Category: Internal Combustion Engines.

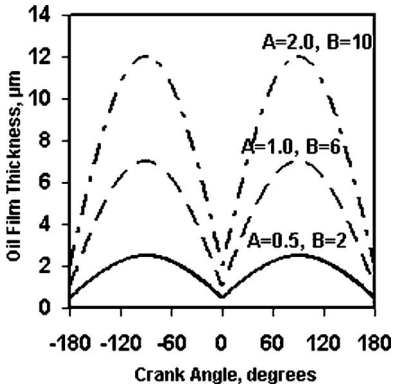


Fig. 1 Different distribution models for the ring oil film thickness

calculated by using the viscosity estimated from the oil film temperature [38]. It was reported that oil film thickness between the ring and liner increases with increasing oil viscosity [39]. Therefore, it can be seen that the temperature dependence of viscosity, engine operational conditions, and ring configuration are very essential in analyzing friction encountered in piston skirt and piston rings.

In the present work, a spark-ignition (SI) engine is analyzed by including the effect of skirt and ring friction on its performance. Different ring configurations and oil types are examined. Furthermore, the effect of oil film temperature on the skirt and ring friction is investigated. Moreover, this study evaluates the contribution of piston rings and skirt on the overall efficiency of the cycle and on the overall friction mean effective pressure (fmep). In addition, a temperature-dependent specific heat is considered.

Thermodynamic Analysis

Piston Friction. The piston friction work, in the combustion chamber, consists of two major parts, which are the skirt friction and pressure ring friction. By using Newton's law of viscosity, friction work is defined as

$$\delta W_{\text{irrev}} = \left(\mu \frac{du}{dy}\right)_{\text{skirt}} L_{\text{skirt}} \pi D \Delta x + \left(\mu \frac{du}{dy}\right)_{\text{ring}} L_{\text{ring}} \pi D \Delta x \quad (1)$$

The friction work (irreversible work) can be expressed in terms of instantaneous piston speed as

$$\delta W_{\text{irrev}} = \mu \frac{U_p(\theta)}{C} L_{\text{skirt}} \pi D \Delta x + \mu \frac{U_p(\theta)}{\varepsilon} L_{\text{ring}} \pi D \Delta x \quad (2)$$

where C is the skirt clearance and ε is the clearance between the liner and the pressure ring. In the present study, the value of C is taken as a constant to represent the average clearance between the cylinder liner and the skirt of the piston. However, ε is taken as the oil film thickness between the ring and cylinder liner. This thickness reaches a minimum value at the bottom dead center (BDC) and top dead center (TDC) and has higher values between them [36,38]. The distribution of oil film thickness with crank angle was reported [38,41]. The shape of the oil film thickness can be approximated by a trigonometric function where minimum values at BDC and TDC and higher values in between. In the current study, the following distribution is assumed that agrees with the trend of previous published work:

$$\varepsilon(\theta) = A + B|\sin(\theta)| \quad (3)$$

where A and B are constants. Figure 1 gives a representation of three different distributions of ring oil film thickness obtained by using different combinations of A and B. These values are chosen to agree with distributions reported in literature [36,38,41].

The instantaneous piston speed, in Eq. (2), can be expressed in terms of average piston speed as [11]

$$U_p(\theta) = \overline{U_p} \frac{\pi}{2} \sin(\theta) \left(1 + \frac{\cos(\theta)}{\left(\left(\frac{\ell}{R}\right) - \sin^2(\theta)\right)^{1/2}}\right) \quad (4)$$

Thus, the friction work in Eq. (2) is written as

$$\delta W_{\text{irrev}} = \mu \pi D \Delta x \overline{U_p} \frac{\pi}{2} \sin(\theta) \left(1 + \frac{\cos(\theta)}{\left(\left(\frac{\ell}{R}\right) - \sin^2(\theta)\right)^{1/2}}\right) \times \left(\frac{L_{\text{skirt}}}{C} + \frac{L_{\text{ring}}}{\varepsilon}\right) \quad (5)$$

Thermodynamic properties of air. In most air-standard power cycle models, air is assumed to behave as an ideal gas with constant specific heats. The values of specific heats are usually used as cold properties. However, this assumption can be valid only for small temperature differences. Therefore, the assumption would produce greater error in all air-standard power cycles. In order to account for the large temperature difference encountered in air-standard power cycles, constant average values of specific heats and specific heat ratios are sometimes used. These average values are evaluated using the extreme temperatures of the cycle, and are believed to yield better results. Obviously, this remains a rough simplification and can result in significant deviations from reality. Thus, the incorporation of variable specific heats in air-standard power cycle models can improve their predictions and bring them closer to reality.

An equation that can be used, for the temperature range 300–3500 K, is obtained from Sonntag et al. [42]. It is based on the assumption that air is an ideal gas mixture containing 78.1% nitrogen, 20.95% oxygen, 0.92% argon, and 0.03% carbon dioxide (on mole basis). It is presented in the following equation:

$$C_p = 2.506 \times 10^{-11} T_g^2 + 1.454 \times 10^{-7} T_g^{1.5} - 4.246 \times 10^{-7} T_g + 3.162 \times 10^{-5} T_g^{0.5} + 1.3303 - 1.512 \times 10^4 T_g^{-1.5} + 3.063 \times 10^5 T_g^{-2} - 2.212 \times 10^7 T_g^{-3} \quad (6)$$

The results obtained from the above equation are in agreement with those reported in literature [43]. It is found from Eq. (6) that specific heat at constant pressure increases with temperature from about 1.0 kJ/kg K at 300 K to about 1.3 kJ/kg K at about 3000 K. Surely, such difference should be taken into consideration. Similarly, the specific heat ratio k decreases from 1.40 to about 1.28 within the same temperature range.

Thermodynamic analysis. For a closed system and a small change of the process, the first law of thermodynamics is simply written as

$$\delta Q - \delta W = dU \quad (7)$$

Therefore, by using the definition of work, the first law can be expressed as

$$\delta Q_{\text{in}} - \delta Q_{\text{loss}} - (PdV - \delta W_{\text{irrev}}) = dU \quad (8)$$

where the irreversible work is mainly due to friction work.

For an ideal gas, the equation of state is expressed as

$$PV = mR_g T_g \quad (9)$$

By differentiating Eq. (9), we can get

$$PdV + VdP = mR_g dT_g \quad (10)$$

Also, for an ideal gas with constant specific heats, the change in internal energy is expressed as

$$dU = mC_v dT_g \quad (11)$$

By substituting Eq. (11) into Eq. (10), then

$$dU = \frac{C_v}{R_g}(PdV + VdP) \tag{12}$$

By substituting Eq. (12) into Eq. (8), the following equation is obtained:

$$\delta Q_{\text{in}} - \delta Q_{\text{loss}} - (PdV - \delta W_{\text{irrev}}) = \frac{C_v}{R_g}(PdV + VdP) \tag{13}$$

where the δW_{irrev} is given by Eq. (5). The total amount of heat input to the cylinder by combustion of fuel in one cycle is

$$\delta Q_{\text{in}} = m_f \text{LHV} \tag{14}$$

The total heat added from the fuel to the system until the crank position reaches angle θ is given as

$$\delta Q(\theta) = \delta Q_{\text{in}} x_b \tag{15}$$

where x_b is the Weibe function that is used to determine the combustion rate of the fuel and is expressed as [11]

$$x_b = 1 - \exp\left(-a\left(\frac{\theta - \theta_s}{\Delta\theta}\right)^n\right) \tag{16}$$

where a and n equal to 5 and 3, respectively. Also, the total amount of heat loss from the system when the crank moves an increment of $d\theta$ is given as

$$\delta Q_{\text{loss}} = \frac{h_{\text{cg}} A_h}{\omega}(T_g - T_w)d\theta \tag{17}$$

By substituting Eqs. (14)–(17) into Eq. (13) followed by differentiation with respect to crank angle (θ), the following equation is obtained:

$$\frac{dP}{d\theta} = \frac{k-1}{V}\left(\delta Q_{\text{in}}\frac{dx_b}{d\theta} - \frac{h_{\text{cg}}A_h}{\omega}(T_g - T_w)\frac{\pi}{180}\right) - k\frac{P}{V}\frac{dV}{d\theta} + \frac{k-1}{V}(\delta W_{\text{irrev}})\frac{dx}{d\theta} \tag{18}$$

Equation (18) can be solved by using explicit finite difference technique with second order accurate differentiation. The result is given as

$$P(\theta) = \frac{4}{3}P(\theta - \Delta\theta) - \frac{1}{3}P(\theta - 2\Delta\theta) + \frac{k-1}{3V}\delta Q_{\text{in}}(3x_b(\theta) - 4x_b(\theta - \Delta\theta) + x_b(\theta - 2\Delta\theta)) + \frac{2}{3}\frac{(k-1)}{3V} \times (h_{\text{cg}}A_h(\theta)(T_g - T_w))\frac{1}{\omega} - \frac{2kP(\theta - \Delta\theta)}{3V(\theta)} \times \left(\frac{V(\theta + \Delta\theta) - V(\theta + \Delta\theta)}{2\Delta\theta}\right) + \frac{2}{3}\frac{(k-1)}{3V}\delta W_{\text{irev}}\frac{dx}{d\theta}\Delta\theta \tag{19}$$

where $dP/d\theta$ is expressed as

$$\frac{dP}{d\theta} = \frac{3P(\theta) - 4P(\theta - \Delta\theta) + P(\theta - 2\Delta\theta)}{2\Delta\theta} \tag{20}$$

The instantaneous cylinder volume, area, and displacement are given by the slider crank model as [11]

$$V(\theta) = V_c + \frac{\pi D^2}{4}x(\theta) \tag{21}$$

$$A_h(\theta) = \frac{\pi D^2}{4} + \frac{\pi DS}{2}(R + 1 - \cos(\theta) + (R^2 - \sin^2(\theta))^{1/2}) \tag{22}$$

Table 1 Engine and operational specifications used in simulation

Fuel	C_8H_{18}
Compression ratio	8.3
Cylinder bore (m)	0.0864
Stroke (m)	0.0674
Skirt length (m)	0.0674
Skirt clearance (m)	2.3×10^{-5}
Connecting rod length (m)	0.13
Crank radius (m)	0.0337
Clearance volume (m^3)	5.41×10^{-5}
Swept volume (m^3)	3.95×10^{-4}
Inlet pressure (bar)	1
Inlet temperature (K)	300
Equivalence ratio	1
Ignition timing	−25 deg BTDC
Duration of combustion	70 deg
Wall temperature (K)	400

$$x(\theta) = (\ell + R) - (R\cos(\theta) + (\ell^2 - \sin^2(\theta))^{1/2}) \tag{23}$$

Once the pressure is calculated, the temperature of the gases in the cylinder can be calculated using the equation of state as

$$T_g = \frac{P(\theta)V(\theta)}{mR_g} \tag{24}$$

The convective heat transfer coefficient in Eq. (17) h_{cg} is given by the Woschni model as [44–46]

$$h_{\text{cg}} = 3.26D^{-0.2}P^{0.8}T_g^{-0.55}w^{0.8} \tag{25}$$

where w is the velocity of the burned gas and is given as

$$w(\theta) = \left(C_1\overline{U_p} + C_2\frac{V_d T_{\text{gr}}}{P_r V_r}(P(\theta) - P_m)\right) \tag{26}$$

In the above equation, the displacement volume is V_d. However, V_r, T_{gr}, and P_r are reference state properties at closing of inlet valve and P_m is the pressure at the same position to obtain P without combustion (pressure values in cranking). The values of C_1 and C_2 are given as follows: For compression, $C_1 = 2.28$, $C_2 = 0$ and for combustion and expansion, $C_1 = 2.28$, $C_2 = 0.00324$.

Solution Methodology

Equation (19) is solved for each crank angle for $-180 \leq \theta \leq 180$ using a step size $\Delta\theta = 1$ deg. The values of $\theta = \pm 180$ correspond to BDC whereas the value of $\theta = 0$ corresponds to TDC. The heat addition in Eq. (19) is only valid for $\theta_s < \theta < (\theta_s + \Delta\theta)$, i.e., during the period of combustion. In solving Eq. (19), notice that k, P, T, and h_{cg} are coupled, i.e., solution of one of these variables depends on the solution of others. The solution procedure is as follows: By knowing the pressure of the gases at the BDC, the initial temperature of the gases is first calculated using Eq. (24). Then, the value of the pressure, after an increment of $\Delta\theta$, is determined using Eq. (18). Once the value of the pressure is obtained, the temperature-dependent properties $C_p(T)$ is calculated by using Eq. (6). The value of $C_v(T)$ is then determined from the relation $[C_v(T) = C_p(T) - R]$. Thereafter, the value of k is calculated as $k(T) = C_p(T)/C_v(T)$. Finally, the heat transfer coefficient is calculated using the Woschni model given by Eq. (25). The above mentioned procedure is repeated for each value of θ many times until the change between two successive iterations for all variables (T, P, k, and h_{cg}) is less than 10^{-4}. After solving for the pressure in the cylinder, the total friction work for skirt and rings was calculated by integrating Eq. (5) for -180 deg $\leq \theta \leq 180$ deg. Then, the fmep is calculated as

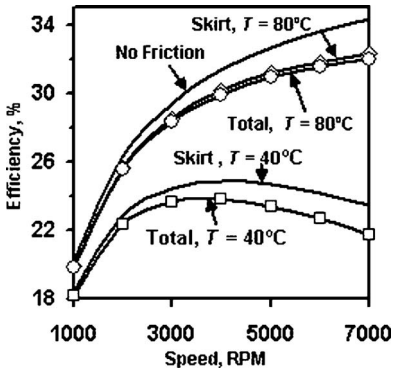


Fig. 2 Efficiency versus engine speed for skirt and total friction contribution, SAE 30, $A=1$, $B=6$, number of rings=2 (each 1.5 mm thick), $C=23$ μm

$$\text{fmep} = \frac{W_{\text{friction}}}{V_d} \tag{27}$$

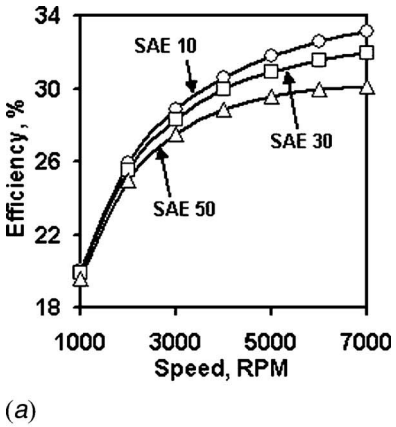
The indicated work is calculated by integrating the PV diagram as

$$W_i = \int_{\theta=-180}^{\theta=180} pdV \tag{28}$$

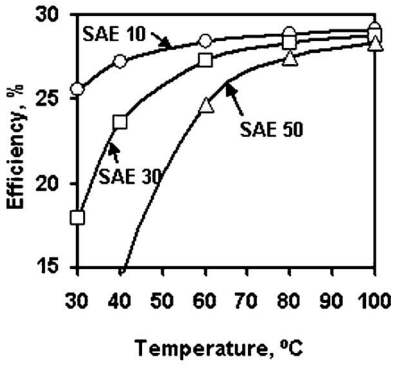
The brake work is calculated as

$$W_b = W_i - W_f \tag{29}$$

Finally, the thermal efficiency is calculated as

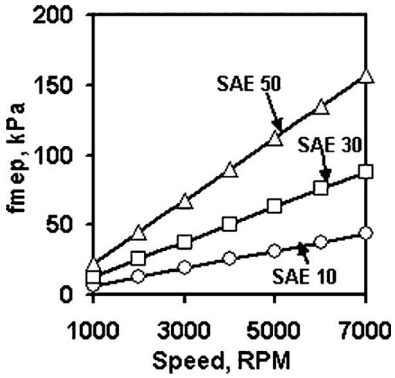


(*a*)

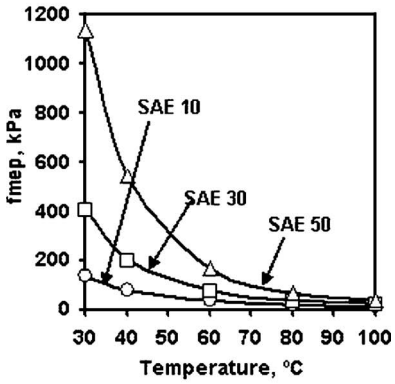


(*b*)

Fig. 3 (*a*) Efficiency versus engine speed for various oil types at 80°C, (*b*) Efficiency versus oil temperature for various oil types at 3000 rpm. (Both (*a*) and (*b*) have $A=1$, $B=6$, number of rings=2 (each 1.5 mm thick), $C=23$ μm.)



(*a*)



(*b*)

Fig. 4 (a) fmep versus engine speed for various oil types at 80°C, (b) fmep versus oil temperature for various oil types and 3000 rpm. (Both (*a*) and (*b*) have $A=1$, $B=6$, number of rings =2 (each 1.5 mm thick), $C=23$ μm.)

$$\eta_{\text{th}} = \frac{W_b}{m_f \text{LHV}} \tag{30}$$

A parametric study has been performed based on the numerical solution of Eq. (19). The study covers wide range of dependent variables such as engine speed, taking into consideration piston friction combined with the variation of temperature-dependent specific heat, and heat loss from the cylinder. Engine specifications, dimensions, and other constants used in the parametric study are listed in Table 1.

Results and Discussion

In order to examine the general effect of piston friction on engine performance, Fig. 2 is presented. It shows engine's thermal efficiency with two oil temperatures: 40°C and 80°C; one representing cold-start conditions, while the other representing normal operating conditions. For each condition, two cases were considered; first, considering skirt friction only, and second considering total piston friction (i.e., skirt and rings). They were compared to frictionless piston, which is usually assumed in most thermodynamic analyses that do not consider piston friction. An observation can be drawn from this figure in which the inclusion of the piston friction in efficiency evaluation can reduce the calculated values. This becomes particularly significant in the high speed range and at low oil temperatures. Also, the minimal effect of ring friction in comparison to skirt friction is clearly illustrated. Although the contribution of ring friction becomes more significant at low oil temperatures and high engine speeds, skirt friction remains the dominant factor in piston friction.

The effect of oil viscosity is investigated in Fig. 3 over a wide

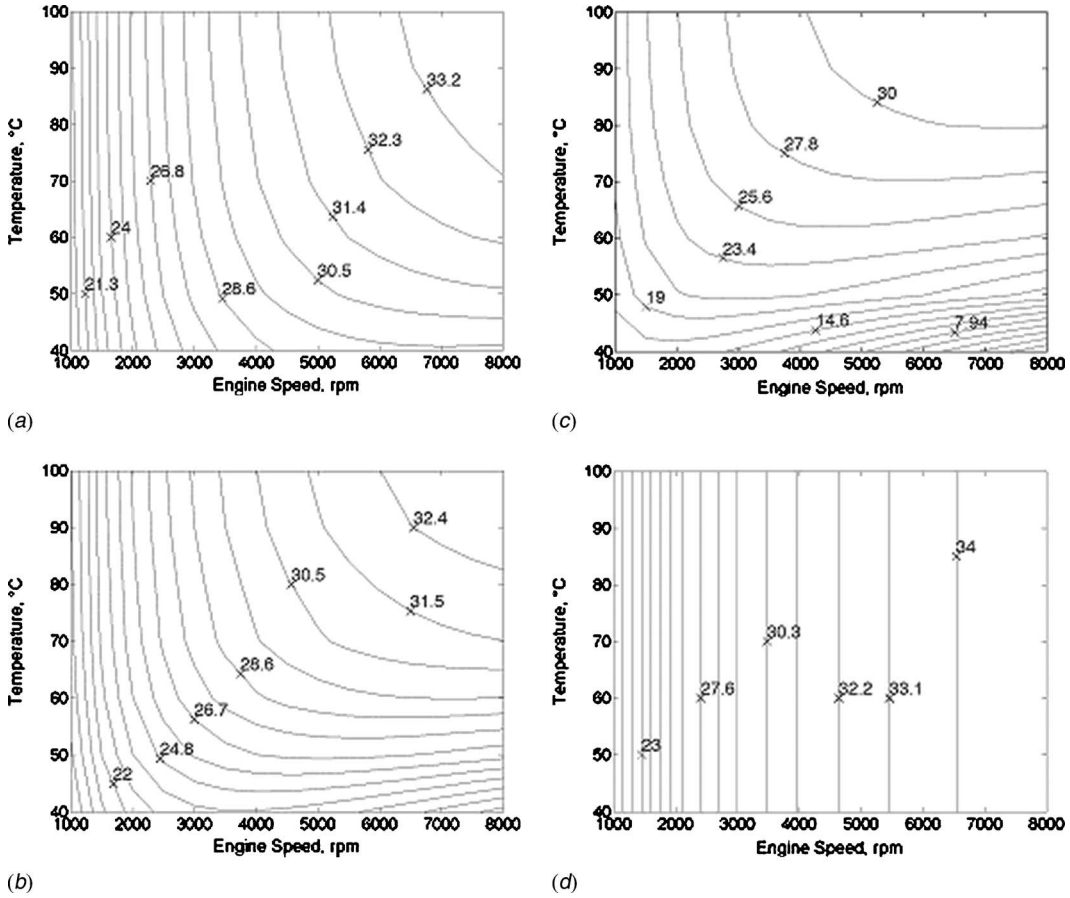


Fig. 5 Efficiency contours $A=1$, $B=6$, number of rings=2 (each 1.5 mm thick), $C=23$ μm: (*a*) SAE 10, (*b*) SAE 30, (*c*) SAE 50, and (*d*), no friction case

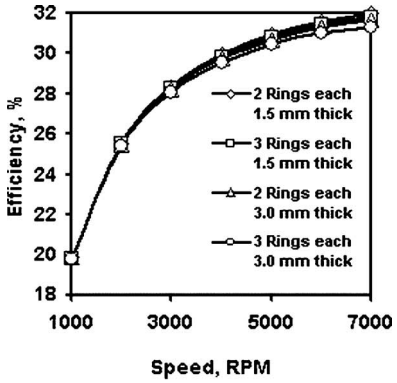
range of engine speeds and oil temperatures, as shown in parts (*a*) and (*b*), respectively. In general, oils with higher SAE numbers result in lower efficiency values due to their higher viscosities. Under normal operating conditions (i.e., relatively high oil temperatures), the difference in performance with the application of different oils is minimal at low speeds and only starts to become more pronounced at high speeds. However, as demonstrated in Fig. 3(*b*), this changes completely at lower oil temperatures, which could be encountered at cold starts. Using high viscosity oil can reduce the efficiency by over 50% at cold oil temperatures.

In order to analyze how the friction irreversible work is affected by the speed and the oil temperature, Fig. 4 is presented. As it can be seen from Fig. 4(*a*), the fmep increases linearly with engine speed. This trend is expected from the examination of Eq. (5) where the piston friction work is linearly related to the piston speed. However, Fig. 4(*b*) reveals a strong dependence of the fmep on the temperature. This is due to the strong dependence of the viscosity of lubricating oils on temperature over the range of temperatures, practically, encountered during engine operations. About 50°C drop in oil temperature can increase its viscosity by an order of magnitude.

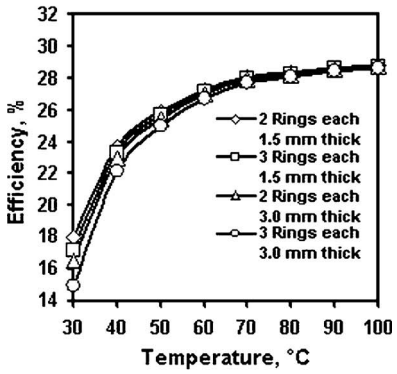
The efficiency maps for SAE 10, SAE 30, and SAE 50 are given in Fig. 5. It illustrates how the results of this model can be practically utilized to obtain optimized efficiency results either by selecting the optimum operating speed for a given oil type (i.e., viscosity), temperature, and engine speed. For example, the efficiency at a given speed is less dependent on the oil temperature for SAE 10 oil as compared to SAE 50. Therefore, it is much easier to obtain high efficiency values at high speeds with low oil temperatures when SAE 10 is used in comparison with SAE 50. On the other hand, for a given oil (e.g., SAE 30), in order to achieve the efficiency obtained at a low-medium range speed with high oil temperatures, higher engine speeds are required with low oil temperatures. Figure 5(*d*) presents the resulting map without the inclusion of piston friction in the analysis. Similar maps could also be obtained when empirical formulas relating the engine friction to the engine speed are used [9,47]. Obviously, such map is rather theoretical and totally oblivious to the effect of oil type or temperature and thus it is of limited use in practical considerations.

Figure 6 demonstrates the effect of different piston ring configurations on the efficiency. As expected, increasing the number or thickness of piston rings increases piston friction and thus reduces the brake power and the efficiency. However, the effect of changing piston ring configuration is limited in comparison to other parameters. This is due to dominance of skirt friction over ring friction, which was illustrated in Fig. 2. As explained earlier and shown in Fig. 1, the oil film thickness between the ring and the cylinder liner varies with crank angle and can be approximated with the trigonometric function given in Eq. (3).

In order to examine the effect of oil film thickness on engine performance, Fig. 7 is presented. It shows the effect of speed and temperature on efficiency for three different values of constants A and B in Eq. (3) yielding three different oil film distributions. Although decreasing oil film thickness results in a drop in efficiency, this effect becomes noticeable only at low temperatures and at high engine speeds. It is also interesting to note that the two distributions assume maximum film thicknesses as of 7 μm and 12 μm resulted in nearly identical efficiency curves. However, the third distribution that assumes a maximum film thickness of 2.5 μm resulted in a somewhat different curve. This suggests that

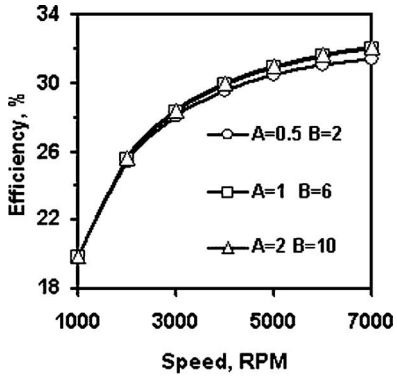


(*a*)

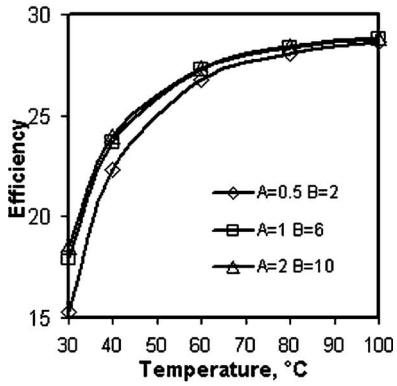


(*b*)

Fig. 6 (*a*) Efficiency versus engine speed for different ring configuration, SAE 30, T=80°C, A=1, B=6, C=23 μm. (b) Efficiency versus oil temperature for different ring configuration for SAE 30, N=3000°C, A=1, B=6, C=23 μm.



(*a*)



(*b*)

Fig. 7 (a) Efficiency versus engine speed for various oil film thickness, SAE 30, T=80°C, number of rings=2 (each 1.5 mm thick), C=23 μm. (b) Efficiency versus oil temperature for various oil film thickness, SAE 30, N=3000, A=1, B=6, number of rings=2 (each 1.5 mm thick), C=23 μm.

there exists a threshold value in the order of 1 μm for the oil film thickness, below which ring friction can start to play significant role in piston friction.

Conclusions

The performance of a SI engine has been analyzed by including the effect of skirt and ring frictions. Different ring configurations and oil types were examined. The effect of oil film temperature on the skirt and the ring friction has been investigated. A parametric study has been performed covering wide range of dependent variables such as engine speed, by considering piston friction combined with temperature-dependent specific heat, and heat loss from the cylinder. Moreover, this study has evaluated the contribution of piston skirt and rings on the overall cycle efficiency. Cases of cold-start and warm-up conditions were presented; first considering skirt friction only, and second with the consideration of total piston friction (skirt and rings). It was found that the inclusion of the piston friction in efficiency evaluation can reduce the calculated values especially in high speed range and at low oil temperatures. Although the contribution of ring friction becomes more significant at low oil temperatures and high engine speeds, skirt friction remains the dominant factor in piston friction.

The effect of oil viscosity over a wide range of engine speeds and oil temperatures implies that oils with higher viscosities result in lower efficiency values. Alternatively, under normal operation conditions, the difference in performance with the application of different oils is minimal at low speeds and only starts to become more pronounced at high speeds. However, using high viscosity oil can reduce the efficiency by more than 50% at cold oil temperatures. Also, the fmep increases linearly with engine speed. It has strong dependence on the temperature and viscosity. A drop in oil temperature by 50°C can result in an increase of its viscosity by an order of magnitude.

The efficiency maps for SAE 10, SAE 30, and SAE 50 are developed. They illustrate how the results of this model can be practically utilized in efficiency calculations. The oil film thickness between the ring and the cylinder liner varies with the position of the crank angle. Although it can be generally stated that decreasing the oil film thickness results in a drop in the efficiency, this effect becomes noticeable only at low temperatures and at high engine speeds. It is concluded that the two distributions that assume maximum film thicknesses of 7 μm and 12 μm resulted in nearly identical efficiency curves. However, the third distribution that assumes a maximum film thickness of 2.5 μm resulted in a somewhat different curve. This suggests that there exists a threshold value in the order of 1 μm for the oil film thickness, below which ring friction can start to play a significant role in piston friction.

Nomenclature

A_h = heat transfer area (m^2)
A = coefficient in oil film thickness equation (m)
a = constant in Weibe function
B = coefficient in oil film thickness equation (m)
C = skirt clearance (m)
C_p = constant pressure specific heat (kJ/kg K)
C_v = constant volume specific heat (kJ/kg K)
D = cylinder diameter (m)

h_{cg} = heat transfer coefficient for gases in the cylinder ($W/m^2 K$)
k = specific heat ratio, dimensionless
L_{ring} = ring thickness
L_{skirt} = skirt length
LHV = lower heating value (kJ/kg)
ℓ = connecting rod length (m)
m = mass of cylinder contents (kg)
m_f = mass of burned fuel (kg)
n = constant in Weibe function
P = pressure inside cylinder (Pa)
P_i = inlet pressure (Pa)
Q = heat transfer (kJ)
Q_{in} = heat added from burning fuel (kJ)
R = crank radius (m)
R_g = gas constant (kJ/kg K)
T = oil temperature in the cylinder (°C)
T_g = gas temperature in the cylinder (K)
T_i = inlet temperature (K)
T_w = cylinder temperature (K)
U = internal energy (kJ)
$\underline{U_p}$ = piston speed (m/s)
$\overline{U_p}$ = average piston speed (m/s)
V = cylinder volume (m^3)
V_c = clearance volume (m^3)
V_d = displacement volume (m^3)
x = distance from top dead center (m)
x_b = burning rate of the fuel, dimensionless
W_i = indicated work (J)
W_b = brake work (J)
$W_{friction}$ = friction work (J)
W_{irrev} = irreversible work (J)
w = average cylinder gas velocity (m/s)
α = thermal diffusivity (m^2/s)
ε = oil film thickness (m)
θ = angle (deg)
θ_s = start of combustion or heat addition (deg)
$\Delta\theta$ = duration of combustion (deg)
μ = oil dynamic viscosity ($N\,s/m^2$)
η_{th} = thermal efficiency (%)

Abbreviations

fmep = friction mean effective pressure (kPa)
bmep = brake mean effective pressure (kPa)
imep = indicated mean effective pressure (kPa)
irrev = irreversible
TDC = top dead center
BDC = bottom dead center

References

[1] Al-Sarkhi, A., Jaber, J., Abu-Qudais, M., and Probert, S., 2006, "Effects of Friction and Temperature-Dependent Specific-Heat of the Working Fluid on the Performance of a Diesel-Engine," Appl. Energy, **83**(2), pp. 153–165.
[2] Ge, Y., Chen, L., Sun, F., and Wu, C., 2005, "Thermodynamic Simulation of Performance of Otto Cycle With Heat Transfer and Variable Specific Heats of Working Fluid," Int. J. Therm. Sci., **44**(5), pp. 506–511.
[3] Jafari, A., and Hannani, S., 2006, "Effect of Fuel and Engine Operational Characteristics on the Heat Loss From Combustion Chamber Surfaces of SI Engines," Int. Commun. Heat Mass Transfer, **33**, pp. 122–124.
[4] Ozsoysal, O., 2006, "Heat Loss as a Percentage of Fuel's Energy in Air Standard Otto and Diesel Cycles," Energy Convers. Manage., **47**, pp. 1051–1062.
[5] Pulkrabek, W., 2004, *Engineering Fundamentals of the Internal Combustion Engine*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.
[6] Angulo-Brown, F., Fernandez-Betanzos, J., and Diaz-Pico, C. A., 1994, "Compression Ratio of an Optimized Otto-Cycle Model," Eur. J. Phys., **15**(1), pp. 38–42.
[7] Chen, L., Lin, J., Luo, J., Sun, F., and Wu, C., 2002, "Friction Effects on the Characteristic Performance of Diesel Engines," Int. J. Energy Res., **26**(10), pp. 965–971.
[8] Wang, W., Chen, L., Sun, F., and Wu, C., 2002, "The Effects of Friction on the Performance of an Air Standard Dual Cycle," Exergy, **2**(4), pp. 340–344.
[9] Patton, K. J., Nitschke, R. G., and Heywood, J. B., 1989, "Development and Evaluation of a Friction Model for Spark Ignition Engine," SAE Paper No. 890836.
[10] Bishop, I. N., 1973, "Effect of Design Variables on Friction and Economy," SAE Paper No. 640807.
[11] Ferguson, C., and Kirkpatrick, A., 2001, *Internal Combustion Engines: Applied Thermosciences*, Wiley, New York.
[12] Röhrle, M. D., 1995, *Pistons for Internal Combustion Engines: Fundamentals of Piston Technology*, MAHLE GmbH, Verlag Moderne Industrie, Landsberg/Lech, Germany, p. 70.
[13] Tung, S. C., and McMillan, M. L., 2004, "Automotive Tribology Overview of Current Advances and Challenges for the Future," Tribol. Int., **37**, pp. 517–536.
[14] Kim, M., Dardalis, D., Matthews, R. D., and Kiehne, T. M., 2005, "Engine Friction Reduction Through Liner Rotation," SAE Paper No. 2005-01-1652.
[15] Rohde, S. M., Whitaker, K. W., and McAllister, G. T., 1980, "A Mixed Friction Model for Dynamically Loaded Contacts With Application to Piston Ring Lubrication, Surface Roughness Effects in Hydrodynamic and Mixed Lubrication," *ASME Winter Annual Meeting*, pp. 19–50.
[16] Dowson, D., Economou, P. N., Ruddy, B. L., Strachan, P. J., and Baker, A. J., 1979, "Piston Ring Lubrication, Part II-Theoretical Analysis of a Single Ring and a Complete Ring Pack, Energy Conservation Through Fluid Film Lubrication Technology, Frontiers in Research and Design," *ASME Winter Annual Meeting*, pp. 23–52.
[17] Aoyama, S., 1994, "Numerical Simulation of Piston Ring in Mixed Lubrication: A Non-Axisymmetrical Analysis," ASME J. Tribol., **116**, pp. 470–478.
[18] Yang, Q., and Keith, T. G., 1995, "An Elasto-Hydrodynamic Cavitation Algorithm for Piston Ring Lubrication," STLE Tribol. Trans., **38**, pp. 97–107.
[19] Sawicki, J., and Yu, B., 2000, "Analytical Solution of Piston Ring Lubrication Using Mass Conserving Cavitation Algorithm," STLE Tribol. Trans., **43**, pp. 587–594.
[20] Akalin, O., and Newaz, G. M., 2001, "Piston Ring-Cylinder Bore Friction Modeling in Mixed Lubrication Regime, Part I-Analytical Results," ASME J. Tribol., **123**, pp. 211–218.
[21] Akalin, O., and Newaz, G. M., 2001, "Piston Ring-Cylinder Bore Friction Modeling in Mixed Lubrication Regime: Part II-Correlation With Bench Test Data," ASME J. Tribol., **123**, pp. 219–223.
[22] Priest, M., Dowson, D., and Taylor, C. M., 2000, "Theoretical Modeling of Cavitation in Piston Ring Lubrication," Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci., **214**, pp. 435–447.
[23] Ting, L. L., 1993, "Development of a Reciprocating Test Rig for Tribological Studies of Piston Engine Moving Components: Part 1—Rig Design and Piston Ring Friction Coefficients Measuring Method," SAE Paper No. 930685.
[24] Ting, L. L., 1993, "Development of a Reciprocating Test Rig for Tribological Studies of Piston Engine Moving Components: Part 2—Measurements of Piston Ring Coefficients and Rig Test Confirmation," SAE Paper No. 930686.
[25] Dearlove, J., and Cheng, W. K., 1995, "Simultaneous Piston Ring Friction and Oil Film Thickness Measurements in a Reciprocating Test Rig," SAE Paper No. 952470.
[26] Arcoumanis, C., Duszynski, M., Flora, H., and Ostovar, P., 1995, "Development of a Piston-Ring Lubrication Test-Rig and Investigation of Boundary Conditions for Modeling Lubrication Film Properties," SAE Paper No. 952468.
[27] Bolander, N. W., Steenwyk, B. D., Kumar, A., and Sadeghi, F., 2004, "Film Thickness and Friction Measurement of Piston Ring Cylinder Liner Contact With Corresponding Modeling Including Mixed Lubrication," *ASME-ICED Fall Technical Conference Proceedings*, ASME Paper No. ICEF2004-903.
[28] Xu, H., Bryant, M. D., Matthews, R. D., Kiehne, T. M., Steenwyk, B. D., Bolander, N. W., and Sadeghi, F., 2004, "Friction Prediction for Piston Ring-Cylinder Liner Lubrication," *presented at the ASME Internal Combustion Engine Division Conference*, Long Beach, CA, Oct. 2004, ASME Paper No. ICEF2004-885; also in *Proceedings of the ASME Internal Combustion Engine Division: 2004 Fall Technical Conference*.
[29] Xu, H., Kim, M., Dardalis, D., Bryant, M. D., Matthews, R. D., and Kiehne, T. M., 2005, "Numerical and Experimental Investigation of Piston Ring Friction," *presented at the ASME Internal Combustion Engine Division Conference*, Chicago, IL, Apr. 5–7, ASME Paper No. ICES 2005-1086; also in *Proceedings of the ASME Internal Combustion Engine Division: 2005 Spring Technical Conference*.
[30] Fox, I. E., 2005, "Numerical Evaluation of the Potential for Fuel Economy Improvement Due to Boundary Friction Reduction Within Heavy-Duty Diesel Engines," Tribol. Int., **38**, pp. 265–275.
[31] Gulwadi, S. D., 1998, "A Mixed Lubrication and Oil Transport Model for Piston Rings Using a Mass Conservation Algorithm," ASME J. Eng. Gas Turbines Power, **120**, pp. 199–208.
[32] Piao, Y., and Gulwadi, S. D., 2003, "Numerical Investigation of the Effects of Axial Cylinder Bore Profiles on Piston Ring Radial Dynamics," ASME J. Eng. Gas Turbines Power, **125**, pp. 1081–1089.
[33] Haywood, J. B., 2003, "An Improved Friction Model for Spark-Ignition Engines," SAE Paper No. 2003-01-0725.
[34] Livanos, G., and Kyrtatos, N. P., 2006, "A Model of the Friction Losses in Diesel Engines," SAE Paper No. 2006-01-0888.
[35] Taylor, R., 1997, "Engine Friction: The Influence of Lubricant Rheology," Proc. Inst. Mech. Eng., Part J: J. Eng. Tribol., **211**(3), pp. 235–246.
[36] Allen, D. G., Dudely, B. R., Middletown, J., and Panka, D. A., 1976, "Prediction of Piston Ring Cylinder Bore Oil Film Thickness in Two Particular Engines and Correlation With Experimental Evidences," *Piston Ring Scuffing*, Mechanical Engineering Publication Ltd., London, p. 107.

[37] Harigaya, Y., Akagi, J., and Suzuki, M., 2000, "Prediction of Temperature Viscosity and Thickness in Oil Film Between Ring and Liner of Internal Combustion Engines," *CEC and SAE Spring Fuels and Lubricants Meeting*, Paris, France, Jun. 2000, Paper No. 200011790.
[38] Harigaya, Y., Suzuki, M., and Takiguchi, M., 2003, "Analysis of Oil Film Thickness on a Piston Ring of Diesel Engine: Effect of Oil Film Temperature," ASME J. Eng. Gas Turbines Power, **125**, pp. 596–603.
[39] Harigaya, Y., Suzuki, M., Toda, F., and Takiguchi, M., 2006, "Analysis of Oil Film Thickness and Heat Transfer on a Piston Ring of a Diesel Engine: Effect of Lubricant Viscosity," ASME J. Eng. Gas Turbines Power, **128**, pp. 685–693.
[40] Takiguchi, M., Sasaki, R., Takahashi, I., Ishibashi, F., Furuhama, F., Kai, R., and Sato, M., 2000, "Oil Film Measurement and Analysis of a Three Ring Pack in an Operating Diesel Engine," *CEC and SAE Spring Fuels and Lubricants Meeting*, Paris, France, Jun. 2000, Paper No. 200011787.
[41] Tamminen, J., Sanddtröm, C., and Andersson, P., 2006, "Influence of Load on the Triboligical Conditions in Piston Ring and Cylinder Liner Contacts in a Medium-Speed Diesel Engine," Tribol. Int., **39**, pp. 1643–1652.
[42] Sonntag, R., Borgnakke, C., and Van Wylen, G., 1998, *Fundamentals of Thermodynamics*, 5th ed., Wiley, New York.
[43] Burcat, A., and Ruscic, B., 2005, "Third Millennium Ideal Gas and Condensed Phase Thermochemical Database for Combustion With Updates from Active Thermochemical Tables," Argonne National Laboratory Report No. ANL-05/20, http://www.chem.leeds.ac.uk/combustion/combustion.html.
[44] Woschni, G., 1967, "Universally Applicable Equation for the Instantaneous Heat Transfer Coefficient in Internal Combustion Engine," SAE Paper No. 670931.
[45] Karamangil, M., Kaynakli, O., and Surmen, A., 2006, "Parametric Investigation of Cylinder and Jacket Side Convective Heat Transfer Coefficients of Gasoline Engines," Energy Convers. Manage., **47**, pp. 800–816.
[46] Stone, R., 1999, *Introduction to Internal Combustion Engines*, 3rd ed., Palgrave, New York.
[47] Branes-Moss, H. W., 1975, "A Designer's Viewpoint in Passenger Car Engines," *Conference Proceedings*, Institution of Mechanical Engineers, London, pp. 133–145.

Development of a Torsional Behavior Powertrain Model for Multiple Misfire Detection

Fabrizio Ponti
DIEM,
University of Bologna,
Bologna 40136, Italy

Many methodologies have been developed in the past for misfire detection purposes based on the analysis of the instantaneous engine speed. The missing combustion is usually detected, thanks to the sudden engine speed decrease that takes place after a misfire event. Misfire detection and, in particular, cylinder isolation are nevertheless still a challenging issue for engines with a high number of cylinders, for engine operating conditions at low load or high engine speed, and for multiple misfire events. When a misfire event takes place, a torsional vibration is excited and shows up in the instantaneous engine speed wave form. If a multiple misfire occurs, this torsional vibration is excited more than once in a very short time interval. The interaction between these successive vibrations can generate false alarms or misdetection, and an increased complexity when dealing with cylinder isolation. This paper presents the development of a powertrain torsional behavior model in order to identify the effects of a misfire event on the instantaneous engine speed signal. The identified wave form has then been used to filter out the torsional vibration effects in order to enlighten the missing combustions even in the case of multiple misfire events. The model response is also used to speed up the setup process for the detection algorithm employed, thus evaluating, before running specific experimental tests on a test bench facility, the values for the threshold and the optimal setup of the procedure. The proposed algorithm is developed in this paper for an SI L4 engine; its application to other engine configurations is possible, as is also discussed in this paper. [DOI: 10.1115/1.2770486]

Introduction

Ever since onboard diagnostics (OBD) regulations enforced onboard misfire detection [1], many successful approaches have been proposed to address the issue [2–6]; the great majority of them making use of the instantaneous engine speed [6–13], since this signal is strictly related to the lack of torque associated with a misfire event. While isolated misfire detection can be considered a solved problem for engines with a low number of cylinders (up to 4), it is still a challenging issue for engines with a high number of cylinders (especially at low load and high engine speed, and mainly for cylinder isolation problems). In addition, multiple misfire detection (i.e., two or more misfires within the same engine cycle) is still challenging for all engine architectures, due to the interactions between the misfire events and the torsional behavior of the system.

A misfire event can be considered as an impulsive excitation on the crankshaft, related to the lack of torque resulting from the missing combustion. This excitation causes a sudden engine speed decrease followed by an oscillation with characteristics (amplitude and frequency) associated with the torsional behavior of the engine-load system.

An example of such behavior can be observed in Fig. 1 where the engine speed wave form for an isolated misfire event has been reported, while the engine was running at approximately 3000 rpm engine speed and 49 N m (5.13 bar brake mean effective pressure (BMEP)) load.

Figure 1 has been obtained running an L4 SI engine (whose characteristics are reported in Table 1) mounted on a test bench. As can be observed, a torsional oscillation at a frequency equal to 30 Hz arises after the misfire event and damps out after few engine cycles. The same behavior can also be observed for other engines or if the engine is mounted onboard a vehicle. Obviously, if the engine-load configuration is different, the torsional vibration that arises after a misfire event will show a different frequency and amplitude, being those values associated with the modes of the powertrain system.

In multiple misfire events, the missing combustions that closely follow the first one can happen while the torsional vibration is still strongly excited. The interaction between the oscillation caused by the first missing combustion and those that follow can be such that the engine speed decrease associated with each misfire can be, in some cases, hidden or overemphasized. In these cases, misfire diagnosis can fail both in the detection of the right number of misfiring cylinders and in their correct cylinder isolation, as will be shown in the following section of the paper.

In order to better understand the described interaction and therefore avoid incorrect multiple misfire detections (false alarms or misdetections), it is important to develop an engine-load model capable of reproducing the torsional behavior of the system. The development of this model is described in the first section of the paper, and the possibility of developing such a model for other possible engine-load configurations is also discussed. The following section presents the misfire detection algorithm employed in this paper, together with the way in which the algorithm can be modified in order to account for the torsional behavior of the system, and, therefore, avoid misdetections or false alarms even in multiple misfire events. These modifications have been done based on the model response to a misfire event, with the aim of filtering out the torsional vibration effects from the engine speed wave form, thus highlighting a sudden engine speed decrease after each misfire event. Finally, some experimental results will follow in order to validate the proposed algorithm and to evaluate its detection capabilities.

Contributed by the Internal Combustion Engine Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received September 20, 2005; final manuscript received July 13, 2007; published online February 25, 2008. Review conducted by Jim Cowart. Paper presented at the 2005 Spring Conference of the ASME Internal Combustion Engine Division (ICES2005), McCormick Place Convention Center, Chicago, IL, April 5–7, 2005.

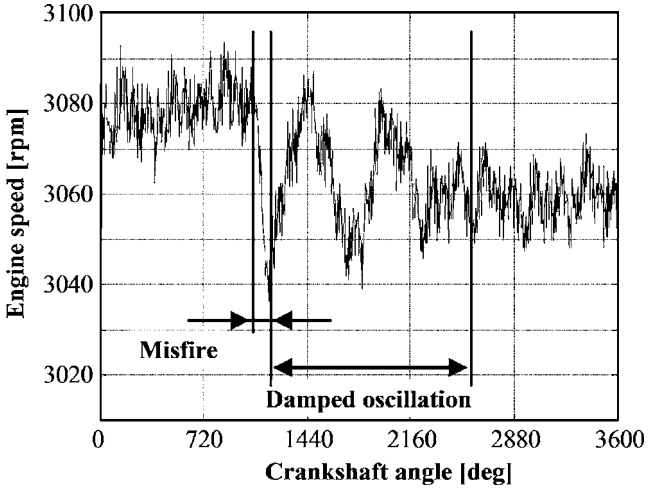


Fig. 1 Engine speed trend with Cylinder 1 misfiring (TDC of the misfiring cylinder is at 1080 deg crankshaft angle). The engine is running at 3000 rpm and 5.13 bar BMEP.

Torsional Vibration Model

The engine used throughout this study is an L4 SI engine, whose characteristics are reported in the following table.

The approach presented can also be applied to other engines (even mounted onboard a vehicle) by following the same steps described in this study, as will be reviewed at the end of this paper. The main difference when applying this approach to an engine mounted on a vehicle is that the same steps have to be followed for each gear inserted, and therefore, a slightly different model is required and has to be developed for each gear.

The first step followed for the development of the torsional model is to generate a computer aided design (CAD) representation of the whole engine-load system (see Fig. 2).

The engine-load system has been divided into seven inertias: distribution, four inertias for the crankshaft (one equivalent inertia for each crank-slider mechanism), flywheel, and load (in our case, an eddy-current brake). The inertias are connected with stiffnesses and dampings (see Fig. 3).

Using a modern CAD software, it has been possible to derive from the geometric representation the inertias and stiffnesses of the different portions of the system, with the exception of distribution and eddy-current brake inertias (no drawings available), and the stiffnesses of the connection between distribution and Cylinder 1, and the connection between the flywheel and the brake. Other unknowns within the model are all the dampings, that have therefore to be identified on an experimental basis, together with the other unknown parameters (they are, as already mentioned, distribution and eddy-current brake inertias, and the stiffness of their connections to the engine).

Table 1 Main engine characteristics

Total displacement	1242 cm^3
Architecture	L4
Injection system	Sequential speed-density multipoint
Bore	70.8 mm
Stroke	78.86 mm
Connecting rod length	129 mm
Compression ratio	9.8 ± 0.2
Number of valves	4 per cylinder
Idle speed	850 rpm
Maximum power	54 kw at 6000 rpm
Maximum torque	105 nm at 4000 rpm
Firing sequence	1-3-4-2
Flywheel teeth number	116

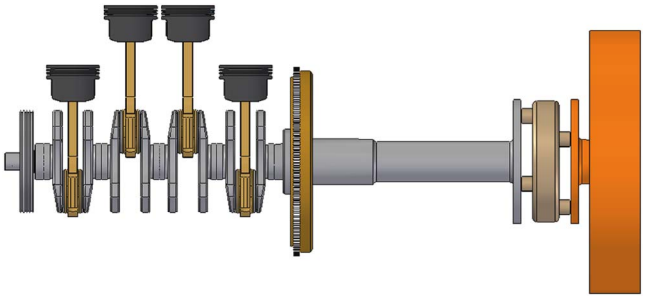


Fig. 2 Complete drawing of the engine-load configuration considered in the paper

The following step is to model the torques that are applied to the system.

Indicated Torque. Indicated torque is the torque acting on the crankshaft due to the in-cylinder pressure. If the in-cylinder pressure is measured, then it is possible to evaluate the indicated torque for the mth cylinder as

$$T_{\mathrm{ind}m}(\theta_m) = A_p r p_m(\theta_m) f(\theta_m) \tag{1}$$

The indicated torque evaluated from Eq. (1) when the engine is fired will be referred to as the firing indicated torque. If the engine is running without combustion into the mth cylinder, at the same engine speed and manifold pressure, then the corresponding motoring indicated torque can be evaluated as

$$T_{\mathrm{indmot}_m}(\theta_m) = A_p r p_{\mathrm{mot}m}(\theta_m) f(\theta_m) \tag{2}$$

The difference between firing and motoring indicated torques will be referred to as combustion indicated torque.

$$T_{\mathrm{indcomb}_m}(\theta_m) = T_{\mathrm{ind}m}(\theta_m) - T_{\mathrm{indmot}_m}(\theta_m) \tag{3}$$

Figure 4 reports firing, motoring, and combustion indicated torques calculated from in-cylinder pressure while the engine is running at 3000 rpm engine speed and 49 N m load (5.13 bar BMEP).

Reciprocating Torque. The reciprocating torque is the torque acting on the crankshaft due to the inertia of the reciprocating masses. It can be evaluated for the mth cylinder [14,15] as

$$T_{rm}(\theta_m) = M_{\mathrm{eq}} r^2 f(\theta_m) \left[f(\theta_m) \ddot{\theta}_m + \frac{df(\theta_m)}{d\theta_m} \dot{\theta}_m^2 \right] \tag{4}$$

A simplification is usually made to Eq. (4), neglecting the term proportional to $\ddot{\theta}_m$, with minor approximation [14,15]

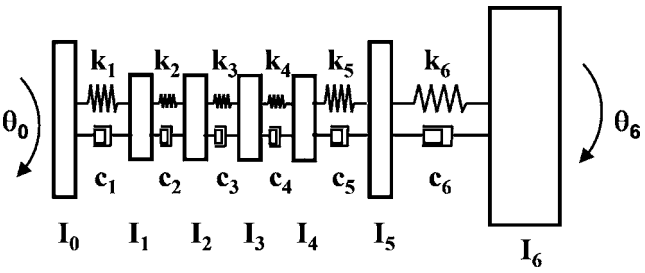


Fig. 3 Complete scheme of the engine-load model

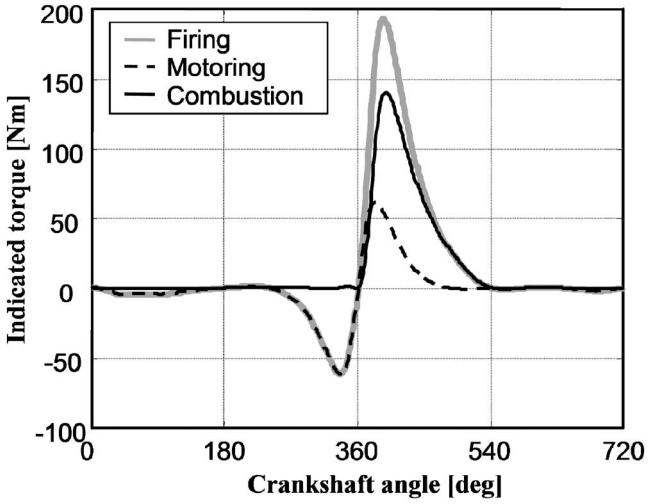


Fig. 4 Firing, motoring, and combustion indicated torques

$$T_{rm}(\theta_m) = M_{\text{eq}} r^2 f(\theta_m) \frac{df(\theta_m)}{d\theta_m} \dot{\theta}_m^2 \tag{5}$$

Figure 5 reports the reciprocating torque at 3000 rpm engine speed for Cylinder 1. The approximation adopted using Eq. (5) gives a difference with respect to Eq. (4) lower than ± 0.1 N m ($\pm 0.2\%$).

Friction Torque. Many models of the friction torque can be found in the literature; some of which are very complex when taking into account the different contributions to friction (i.e., piston-cylinder contact, valvetrain distribution friction, etc.). The one chosen for the torsional model simply considers friction contribution as a polynomial function of engine speed [15]. In this way, friction can be expressed as

$$T_{fm} = f_{m1} + f_{m2}\dot{\theta}_m + f_{m3}\dot{\theta}_m^2 \tag{6}$$

Load Torque. This is the torque applied by the dyno in order to simulate the load given by road and aerodynamic effects when the engine is mounted onboard a vehicle. This quantity can be measured on the eddy-current brake and therefore be considered as an external input for the torsional model. It must be stressed that usually, due to the low frequency content of the load torque, it can be considered constant over an engine cycle.

Table 2 Measured signals and sensors employed

In-cylinder pressures	Piezoelectric sensor in the spark plug
Flywheel speed (inertia 5 in Fig. 3)	Magnetic pickup
Brake speed (inertia 6 in Fig. 3)	Magnetic pickup
Load torque	Load cell
Manifold pressure	Piezoresistive sensor
Throttle opening	Potentiometer

The equation of the model can now be written. Using matrix notation, it yields

$$[\mathbf{I}]\{\ddot{\boldsymbol{\theta}}\} + [\mathbf{c}]\{\dot{\boldsymbol{\theta}}\} + [\mathbf{k}]\{\boldsymbol{\theta}\} = \{\mathbf{T}\} \tag{7}$$

The complete expression of the vectors and matrices is reported in the Appendix. It is important to recall here that the vector $\{\mathbf{T}\}$ represents the torques applied to each inertia; for the crank-slider mechanism inertias, the torques applied can be expressed by $T_{\text{indm}}(\theta_m) - T_{rm}(\theta_m)$. Equation (7) still contains some unknowns (brake and distribution inertias I_0, I_6, the stiffnesses k_1, k_6, and all the internal dampings $c_1, \ldots, c_6$) that will be identified using properly performed experimental tests.

The tests designed for model identification have been performed accelerating the engine from approximately 1000 rpm to approximately 4500 rpm, and then decelerating back to 1000 rpm, while maintaining the throttle to a constant value. These tests have been possible correctly controlling the torque absorbed by the brake. The signals measured during the tests have been reported in Table 2 together with the sensors employed.

The flywheel speed wave form obtained has been reported in Fig. 6. By performing an order analysis on that signal, the diagrams reported in Fig. 7 can be obtained. They show, respectively, the amplitudes and the phases of the torsional oscillation on the flywheel, for each engine order, as a function of engine speed. For each order, the associated frequency can be obtained using

$$f = \frac{\text{ord}n}{2\pi} \tag{8}$$

In this way, amplitudes and phases can also be represented as a function of frequency. Figure 8, for example, reports the amplitudes and phases of the torsional oscillation on the flywheel, for each engine order, as a function of the frequency.

Some observations arise from Figs. 7 and 8. Looking at Fig. 8, for example, it is possible to note that the system presents one

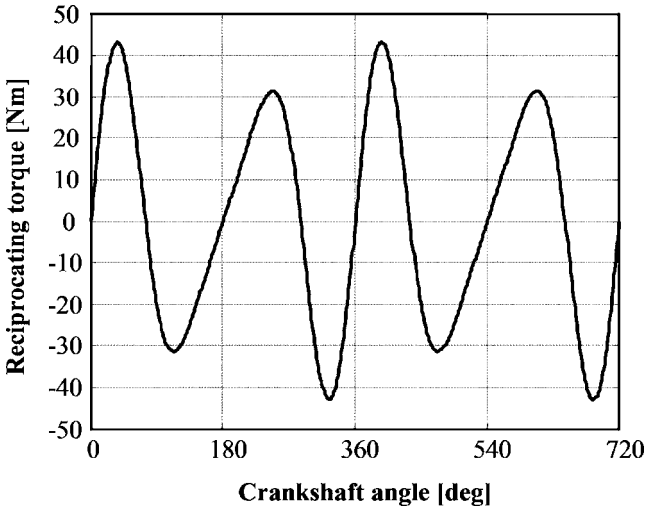


Fig. 5 Reciprocating torque at 3000 rpm

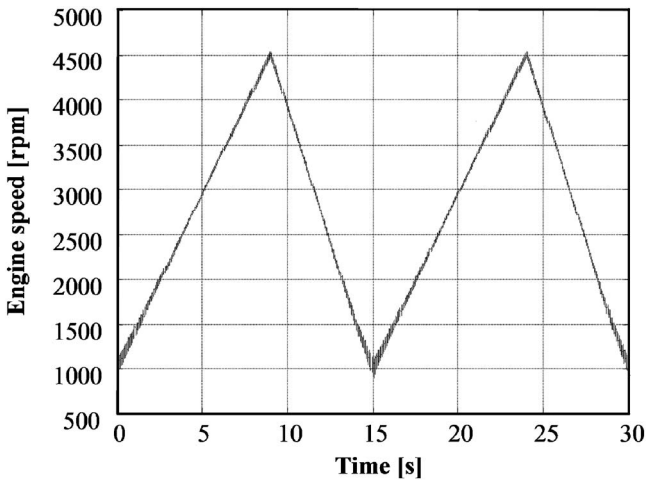


Fig. 6 Engine speed variation during experimental test

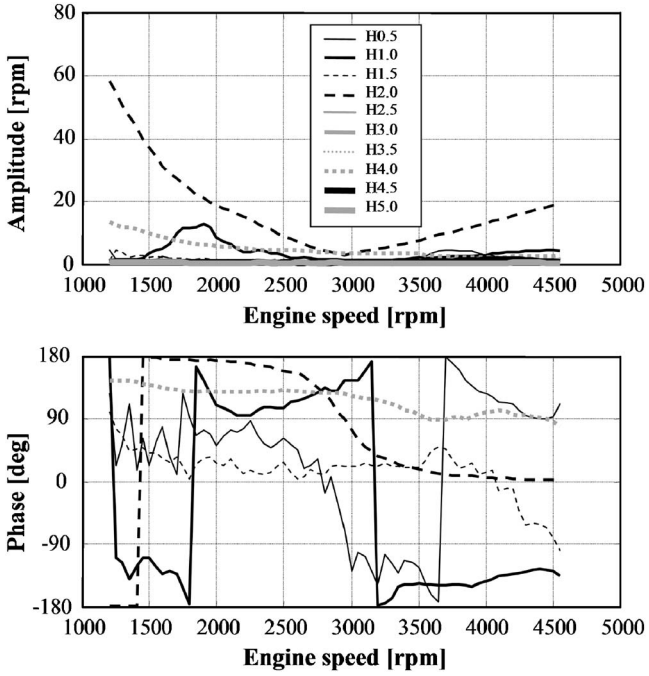


Fig. 7 Engine speed oscillation amplitude and phase

main resonance at approximately 30 Hz, while all the other resonances seem to be at frequencies at least higher than 200 Hz, since any resonance peak is present at any order over 30 Hz.

The amplitude of the engine speed order 2 component (Figs. 7 and 8) is high at low and at high engine speed, while it has an amplitude approximately equal to zero near 2800 rpm. The reason for such behavior can be found in that the amplitude of the engine speed order 2 component is mainly related to the corresponding indicated and reciprocating torque components, whose amplitudes and phases are reported in Figs. 9 and 10 as a function of the

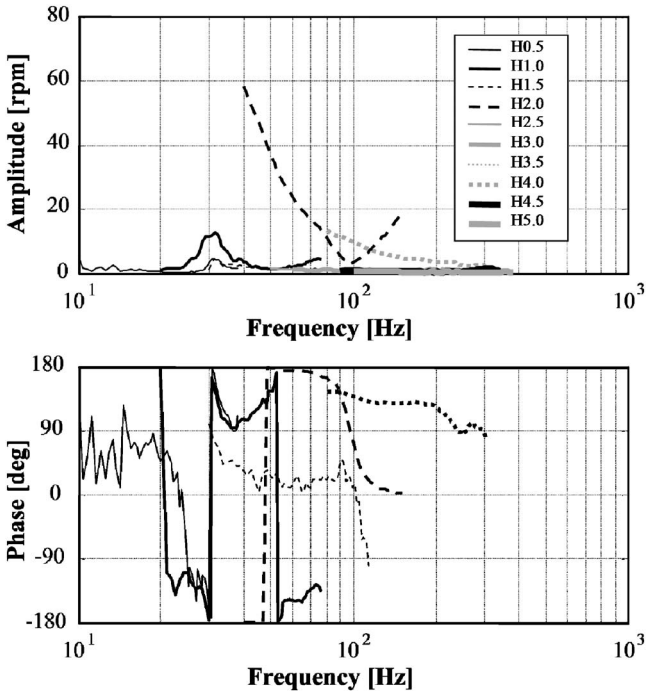


Fig. 8 Engine speed oscillation amplitude and phase

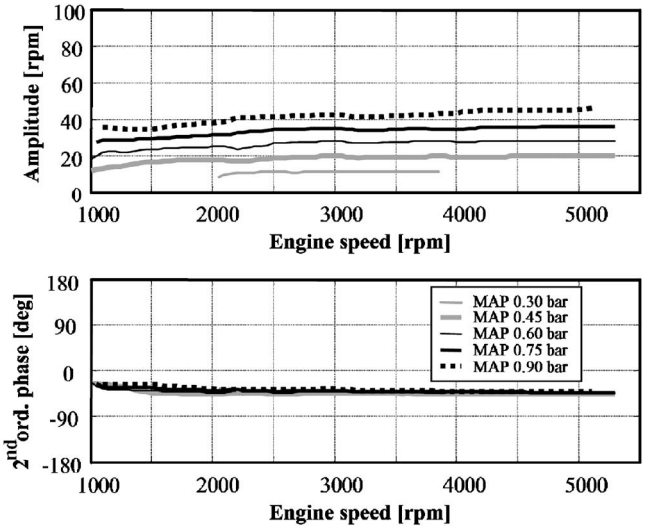


Fig. 9 Amplitude and phase of the indicated torque second order component

engine operating condition.

As can be seen in Figs. 9 and 10, the torque components have approximately the same phase, and while the reciprocating torque amplitude is proportional to the square of the engine speed, the amplitude of the indicated torque component at each load is approximately constant. In addition, it is to be noted that they appear in Eq. (7) as torques applied to crank-slider mechanism inertias with opposite sign. Therefore, at low engine speed, the indicated torque component prevails over that of the reciprocating torque one, while at high engine speed, the reciprocating torque component dominates. At medium engine speed, the two torques have similar amplitudes and, therefore, they compensate each other resulting in an engine speed component very close to zero.

The unknown values of the torsional model are now identified trying to impose, for the model, the same torsional behavior of the real system described in Figs. 7 and 8. A multivariable optimization approach has been used, the simplex search method [16], a direct search method that does not use numerical or analytic gradients. Once the identification process has been run, the engine-load model is complete, and the total set of parameter values can

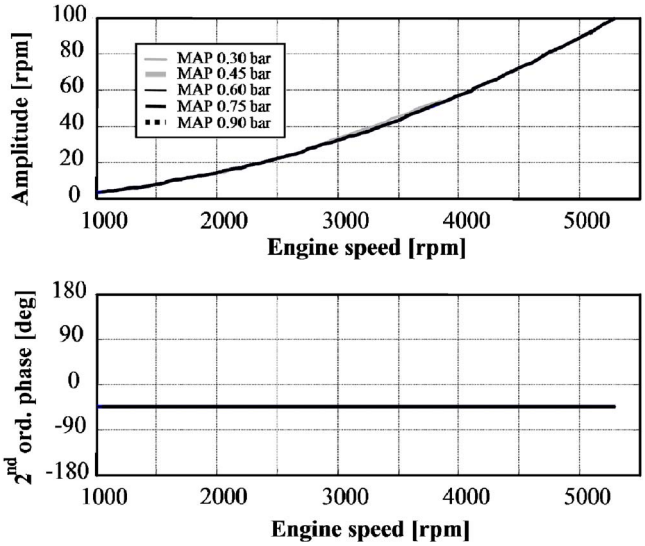


Fig. 10 Amplitude and phase of the reciprocating torque second order component

Table 3 Total set of parameter values

Inertia values ($kg\,m^2$)	Stiffnesses (N m/rad)	Internal dampings (N m s/rad)	Friction coefficients
$\mathbf{I_0}=12.2\times10^{-3}$	$\mathbf{k_1}=81.7\times10^3$	$\mathbf{c_1}=15.6$	$\mathbf{f_1}=9.77$
$I_1=4.22\times10^{-3}$	$k_2=433\times10^3$	$\mathbf{c_2}=0.6$	N m
$I_2=4.22\times10^{-3}$	$k_3=433\times10^3$	$\mathbf{c_3}=0.6$	$\mathbf{f_2}=2.45\times10^{-3}$
$I_3=4.22\times10^{-3}$	$k_4=433\times10^3$	$\mathbf{c_4}=0.6$	N m s/rad
$I_4=4.22\times10^{-3}$	$k_5=586\times10^3$	$\mathbf{c_5}=0.6$	$\mathbf{f_3}=1.44\times10^{-5}$
$I_5=80.2\times10^{-3}$	$\mathbf{k_6}=2.15\times10^3$	$\mathbf{c_6}=2.3$	N m s^2/rad^2
$\mathbf{I_6}=173\times10^{-3}$			
	$M_{eq}=0.412$ (kg)		

be reported in Table 3 (identified values in bold).

The torsional behavior estimated by the model is compared in Fig. 11 with the real system behavior experimentally evaluated.

Figure 11 shows that the model developed is able to reproduce correctly the engine-load torsional behavior. It has been noted already that the system seems to have a natural frequency at approximately 30 Hz, while no other natural frequency is present under 200 Hz. This can be confirmed by the eigenvalue evaluation of the system

$$[\mathbf{I}]\{\ddot{\boldsymbol{\theta}}\}+[\mathbf{k}]\{\boldsymbol{\theta}\}=0 \tag{9}$$

that allows for determining the following natural frequencies:

$$f_{\lambda1}=30.3\ \text{Hz}$$

$$f_{\lambda2}=323\ \text{Hz}$$

$$f_{\lambda3}=812\ \text{Hz}$$

$$f_{\lambda4}=1760\ \text{Hz} \tag{10}$$

The eigenvector evaluation allows for determining the modal shape for all the natural frequencies (see Fig. 12 for the first three natural frequencies).

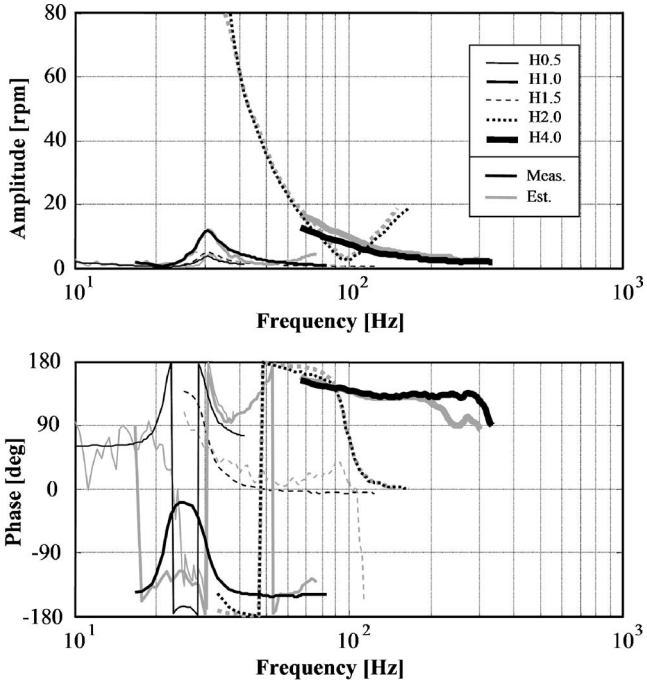


Fig. 11 Experimental (black lines) and simulated (gray lines) order analysis

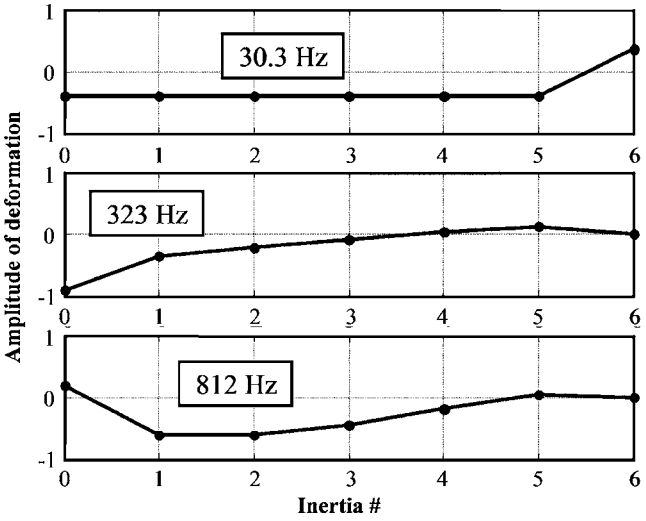


Fig. 12 Modal shape for the first three natural frequencies

As can be seen for the system first mode, the deformation is mainly concentrated on the connection between the engine and the load, while the engine crankshaft can be considered as a rigid body. The lower mode, where the crankshaft cannot be considered rigid, is the second, at a frequency equal to approximately 300 Hz. Therefore, if we are interested in engine-load torsional behavior at frequencies lower than 300 Hz, we could consider the engine crankshaft as a rigid body. This observation is very important since it could allow for strongly simplifying the engine-load model, lowering the number of inertias to be considered, if interest is focused on the engine-load torsional behavior at frequencies lower than 300 Hz.

Using the complete model, whose parameter values are reported in Table 3, it is now possible to reconstruct the engine speed wave form for each engine operating condition. This has been done using SIMULINK as environment to solve the model dynamic equations. In particular, a fixed step discrete representation of the model has been employed. The results have been obtained, for example, for three steady-state operating conditions at 2000 rpm and 3.67 bar BMEP in the first, at 3000 rpm and 6.70 bar BMEP in the second, and at 4000 rpm and 6.60 bar BMEP in the third. The comparison between the measured engine speed and the corresponding estimated ones is reported in Fig. 13.

It should be noted that for the tests at 2000 rpm and 4000 rpm, the engine speed wave forms show opposite phases. This is due to the fact that indicated and reciprocating torques order 2 components have similar phases but that they compare with opposite sign in Eq. (7) and that at a lower engine speed, the indicated torque prevails, while at a higher engine speed, the reciprocating one dominates, as previously explained.

The model also allows for estimating the engine speed wave form in the case of misfire events. It is sufficient, in fact, to apply to the misfiring cylinder, in the misfiring engine cycle, the motoring indicated torque instead of the firing indicated torque. As an example, measured and estimated engine speed wave forms when an isolated misfire is present in Cylinder 1 are reported in Fig. 14 for the same engine operating conditions considered in Fig. 13.

The linear model proposed in Eq. (7) is, therefore, shown to be effective in reproducing the system behavior. It is then possible to conclude that the system shows a linear behavior, since it can be well described using a linear model. In order to isolate the misfire effect, under the hypothesis of a system linear behavior, it is sufficient to compute the difference between the engine speed wave forms reported in Figs. 13 and 14. The results are reported in Fig.

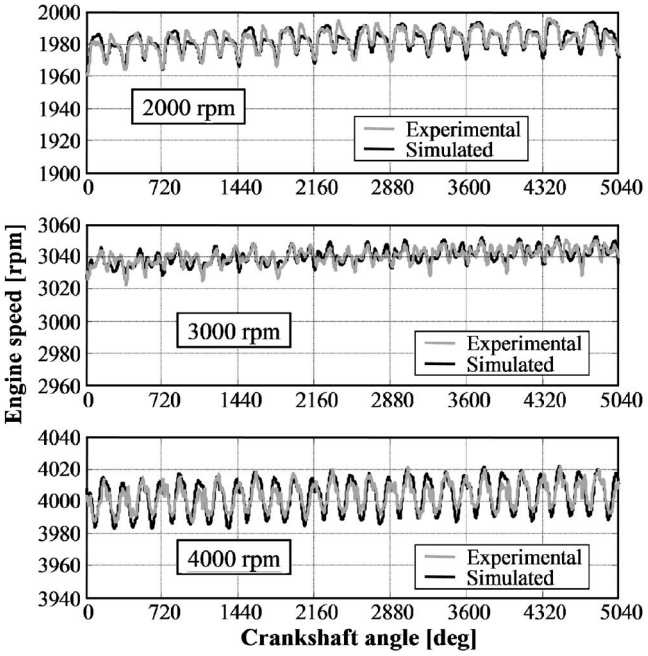


Fig. 13 Experimental (gray line) and simulated (black line) engine speed for steady-state operating conditions

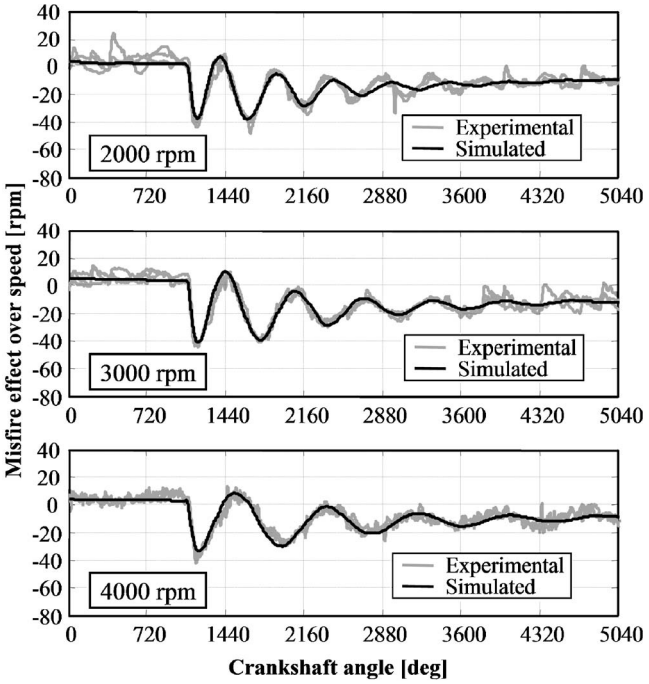


Fig. 15 Misfire effect over engine speed

15: the wave forms start approximately from a 0 value since the computed difference is approximately null until at least one misfire event occurs (not exactly null due to the engine cyclic variability).

The wave forms of Fig. 15 have been normalized by dividing by the manifold absolute pressure (MAP) (since MAP has been considered as a load indicator) and reporting them as a function of time. The obtained result can be seen in Fig. 16.

As can be noted in Fig. 16, the normalized wave forms are quite well overlapped to one another, meaning that the system response to a misfire event could be represented by a unique wave form over time, normalized with respect to the load. This consideration is consistent with the linear hypothesis and will be further developed in the next section.

If a multiple misfire occurs, the expected system response will be the sum of two or more single misfire responses, each of them with a phase that depends on the instant the corresponding misfire occurs. Figure 17, for example, reports for an engine operating condition at 4000 rpm and 6.60 bar BMEP, the experimental and simulated engine speed wave forms for three multiple misfire patterns (1-3, 1-4, and 1-2). This means that when a 1-3 misfire pattern is employed, the combustion into Cylinders 1 and 3 is inhibited only for one engine cycle within the same cycle. Since the engine firing order is 1342, the patterns chosen in the tests (1-3, 1-4, and 1-2) are amenable to investigation of all the possible multiple misfire configurations:

— multiple misfire 1-3 will result in two adjacent misfires,

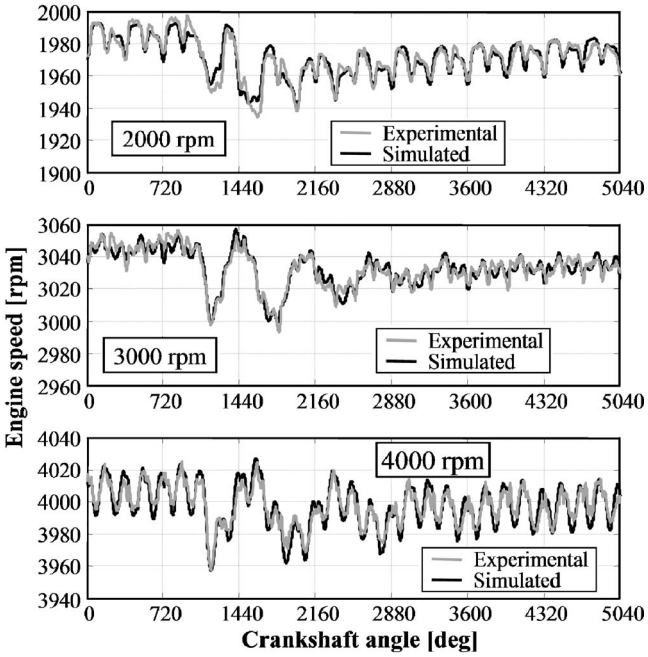


Fig. 14 Experimental (gray line) and simulated (black line) engine speed for misfiring operating conditions

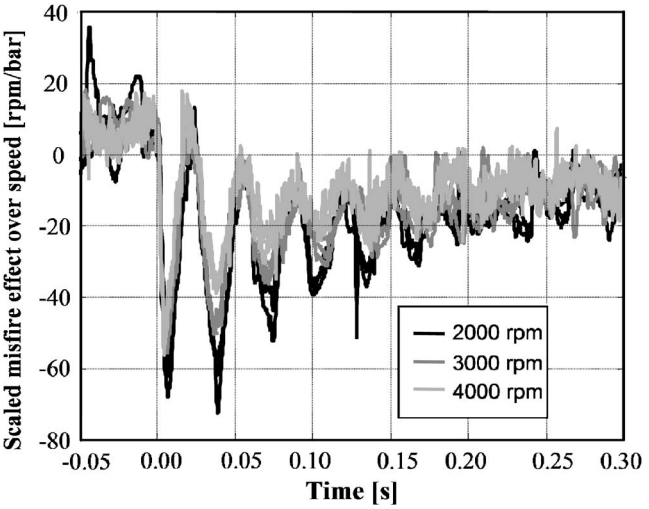


Fig. 16 Misfire effect over engine speed wave form

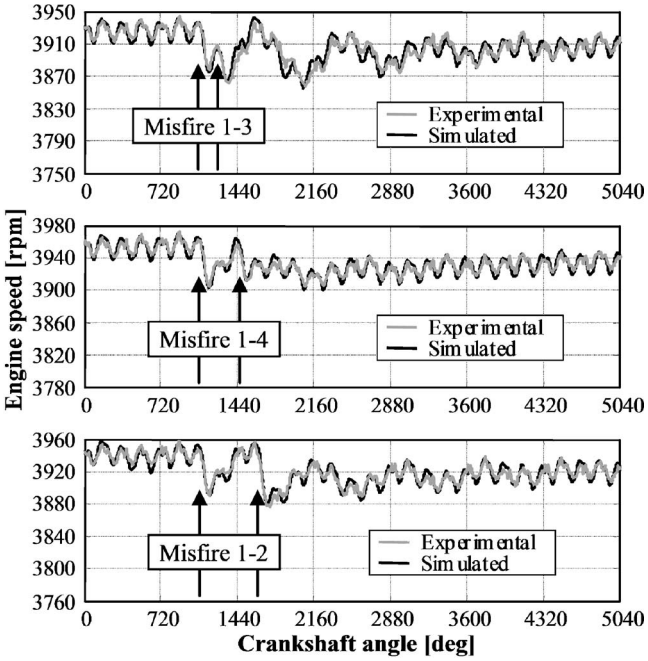


Fig. 17 Experimental (gray line) and simulated (black line) engine speed for multiple misfiring operating conditions (misfire patterns 1-3, 1-4, and 1-2)

— multiple misfire 1-4 gives rise to a normal combustion between two misfires, and
— multiple misfire 1-2 will result in two normal combustions between the misfires.

Also, it is possible, in this case, to isolate the misfire effect (see Fig. 18) in a similar way to what was done to produce Fig. 15

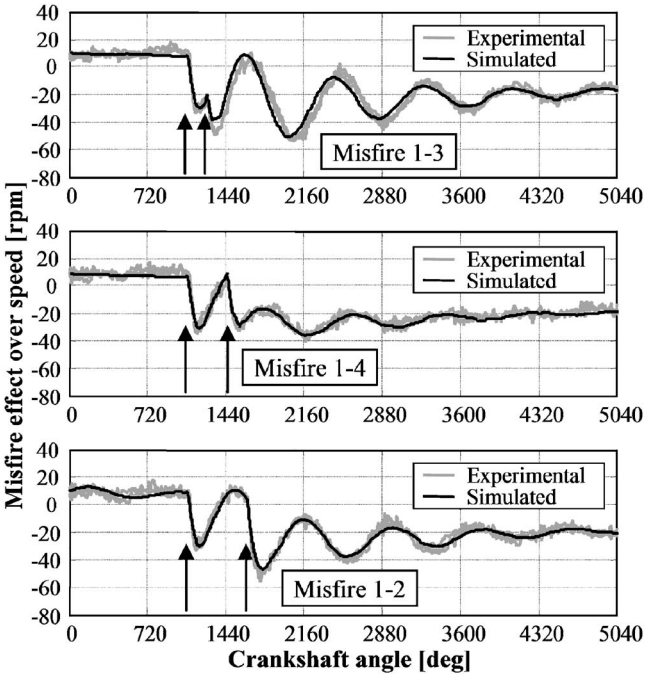


Fig. 18 Misfire effect for different misfire patterns (misfire patterns 1-3, 1-4, and 1-2)

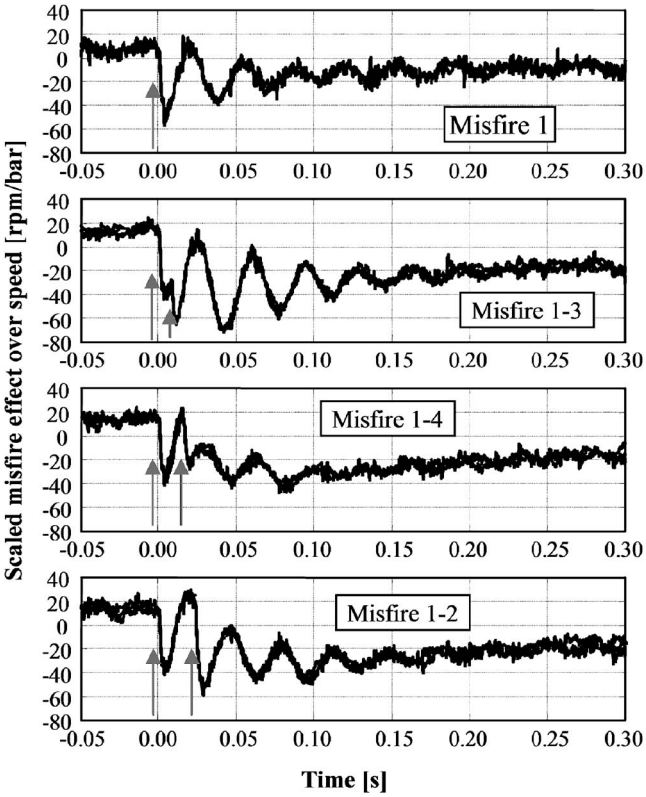


Fig. 19 Misfire effect for different misfire patterns

(i.e., computing the difference between the wave forms reported in Fig. 17 and that for the test at 4000 rpm, reported in the last diagram in Fig. 13, under the hypothesis of a linear system behavior).

It can be seen in Fig. 18 that for the case of a misfire pattern 1-3, the detection of the second misfire is quite difficult since the sudden engine speed decrease usually associated with a misfire event is not present. This is due to the fact that while this second misfire occurs, the system is still oscillating (and, in particular, accelerating) due to the effects of the first misfire. The presence of the torsional vibration due to the first misfire event tends, in fact, to hide the presence of the second. In order to avoid this possible misdetection, it is important to filter out the torsional vibration that follows each misfire event.

In order to confirm that the response to a multiple misfire is the superposition of the responses to each misfire event, the following figures have been reported. Figure 19 presents all the responses to the multiple misfire patterns also considered in Fig. 18, together with the response to a single misfire in Cylinder 1.

The important aspect to be underlined is that up to the instant the second misfire occurs, the engine speed wave forms are very similar and could overlap (Fig. 19). This happens, for example, when comparing the Misfire 1 case with the Misfire 1-3 case approximately up to the time of 0.01 s (i.e., before the second misfire occurs in the 1-3 pattern case). If the difference between the multiple misfire responses and the single one in Cylinder 1 is now computed, it is possible to determine the separate response to the second misfire of the patterns imposed. Figure 20 reports the difference between the wave forms obtained for a multiple misfire event and that obtained for a misfire in Cylinder 1. The computed difference, under the hypothesis of a linear system behavior, allows the isolation of second misfire event response in a multiple misfire case. Comparing the isolated responses to the second misfire event for the multiple misfire cases (misfires in Cylinders 3, 4, and 2, respectively) with the single misfire case in Cylinder 1 (first plot of Fig. 20), it is possible to observe that the wave forms are

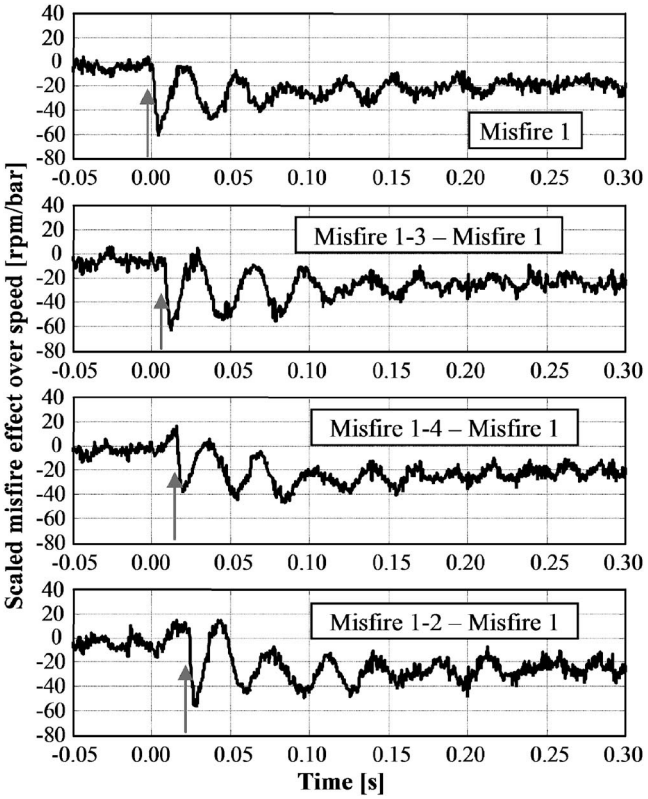


Fig. 20 Misfire effect of the second misfire of the multiple pattern

quite similar, being the main difference a phase shift. This is a further confirmation of the linearity of the system response and of the applicability to the system of the superposition principle.

Misfire Detection Algorithm

The algorithm developed for single and multiple misfire detection and cylinder isolation is based on the well known parameter [12,13]

$$LU_i = \frac{\tau_{i+1} - \tau_i}{\tau_i^3} \tag{11}$$

τ_i is the so called combustion period and can be evaluated as the summation of the tooth times comprised into a defined angular sector $\Delta\theta$.

$$\tau_i = \sum_k t_k \tag{12}$$

It is possible to demonstrate that LU_i is proportional to the kinetic energy variation during the engine crankshaft rotation $\Delta\theta$.

$$LU_i = \frac{\tau_{i+1} - \tau_i}{\tau_i^3} \cong \frac{\tau_{i+1}^2 - \tau_i^2}{2\tau_i^2\tau_{i+1}^2} = \frac{1}{2\tau_i^2} - \frac{1}{2\tau_{i+1}^2} = \left(\frac{\dot\theta_i^2}{2} - \frac{\dot\theta_{i+1}^2}{2}\right) \Big/ \Delta\theta^2 \tag{13}$$

Since the combustion periods are evaluated from the tooth times via a moving average calculation, the frequency content of the combustion period signal (and therefore that of the LU_i one) is lower than that associated with the tooth time signal. For example, if the angular sector $\Delta\theta$ for the combustion period evaluation is equal to 180 deg, the higher frequency component of the combustion period signal has, at 6000 rpm, a frequency equal to 200 Hz. Therefore, for LU_i evaluation, all the high frequency engine-load dynamics can be neglected. For this reason, the model previously developed can be strongly simplified, while still correctly represented all the engine-load dynamic behavior that can affect LU_i wave forms. Since only the first natural frequency needs to be modeled (being the frequency range of interest limited to approximately 200 Hz), the simplified model presents only two inertias: engine and load (Fig. 21).

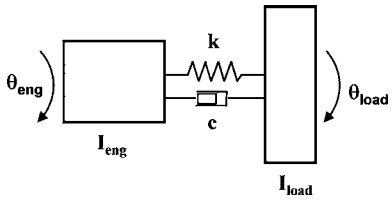


Fig. 21 Simplified model configuration

The equations that can be written for the simplified model are

$$I_{\text{eng}}\ddot\theta_{\text{eng}} = T_{\text{ind}} - T_r - T_f + k(\theta_{\text{load}} - \theta_{\text{eng}}) + c(\dot\theta_{\text{load}} - \dot\theta_{\text{eng}})$$

$$I_{\text{load}}\ddot\theta_{\text{load}} = -T_L + k(\theta_{\text{eng}} - \theta_{\text{load}}) + c(\dot\theta_{\text{eng}} - \dot\theta_{\text{load}}) \tag{14}$$

Using this simplified model, it is easier to demonstrate that the system response to a misfire event can be well represented using a unique wave form over time, scaled with respect to the load as previously proposed. To do this, the Laplace transform representation of Eq. (14) is considered.

$$(I_{\text{eng}}s^2 + cs + k)\theta_{\text{eng}} - (cs + k)\theta_{\text{load}} = T_{\text{ind}} - T_r - T_f$$

$$-(cs + k)\theta_{\text{eng}} + (I_{\text{load}}s^2 + cs + k)\theta_{\text{load}} = -T_L \tag{15}$$

Solving the equations with respect to θ_{eng}, it is possible to obtain

$$\theta_{\text{eng}} = \frac{1}{s^2}\frac{I_{\text{load}}s^2 + cs + k}{I_{\text{eng}}I_{\text{load}}s^2 + (I_{\text{eng}} + I_{\text{load}})(cs + k)}(T_{\text{ind}} - T_r - T_f) - \frac{1}{s^2}\frac{cs + k}{I_{\text{eng}}I_{\text{load}}s^2 + (I_{\text{eng}} + I_{\text{load}})(cs + k)}T_L \tag{16}$$

or similarly

$$\dot\theta_{\text{eng}} = \frac{1}{s}\frac{I_{\text{load}}s^2 + cs + k}{I_{\text{eng}}I_{\text{load}}s^2 + (I_{\text{eng}} + I_{\text{load}})(cs + k)}(T_{\text{ind}} - T_r - T_f) - \frac{1}{s}\frac{cs + k}{I_{\text{eng}}I_{\text{load}}s^2 + (I_{\text{eng}} + I_{\text{load}})(cs + k)}T_L \tag{17}$$

Equation (17) can now be used to determine the system response to a misfire event. As already mentioned, a misfire event can be schematized as an impulsive excitation due to the combustion indicated lack of torque in the misfiring cylinder. The effect of a misfire on the system and, in particular, the effect over the engine speed can therefore be determined considering an excitation equal to $-T_{\text{ind comb}_m}$ on the engine.

$$\dot\theta_{\text{eng}} = -\frac{1}{s}\frac{I_{\text{load}}s^2 + cs + k}{I_{\text{eng}}I_{\text{load}}s^2 + (I_{\text{eng}} + I_{\text{load}})(cs + k)}T_{\text{indcomb}_m} \tag{18}$$

The engine speed wave forms for the engine operating conditions considered in Fig. 16 have been evaluated, computing the antitransform of Eq. (18), and reported in Fig. 22, divided by the MAP value (i.e., normalized with respect to the load).

The difference between the three wave forms reported in Fig. 22 is due to the difference that exists between the term $-T_{\text{ind comb}_m}/\text{MAP}$ for the three operating conditions investigated. These wave forms can now be used to theoretically address two different tasks needed to set up the detection algorithm proposed.

1. Definition of the threshold values for LU_i that discriminate misfire events from the normal operating condition.
2. Definition of the LU_i wave form to be used to compensate

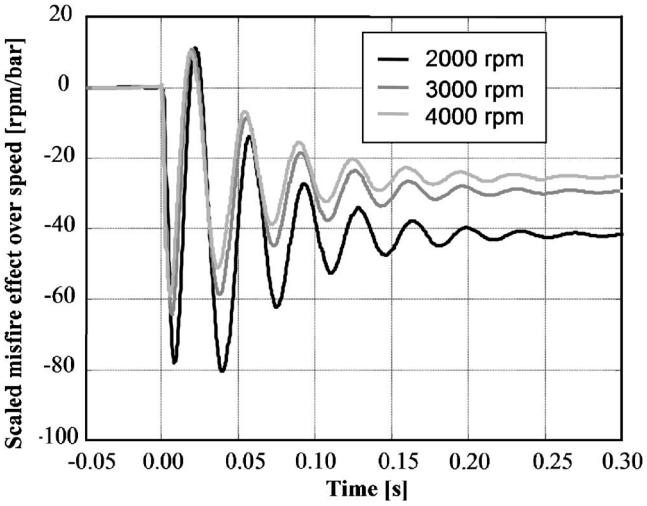


Fig. 22 Misfire effects over engine speed wave form at different speeds and loads

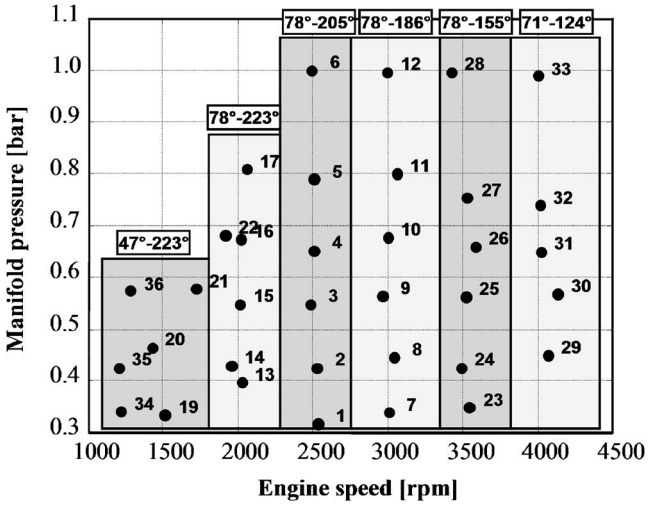


Fig. 23 Best sector definition (the angular values are reported with respect to the TDC of the corresponding cylinder)

for the torsional vibration effects and therefore enhance multiple misfire detection and reduce the risk of false alarms or misdetections. It is possible to consider a different wave form for each engine operating condition (electronic control unit (ECU) resource consuming solution) or to determine a unique averaged wave form to be applied over the entire engine operating range (preferred solution in this study).

While the first task is useful to improve and quicken the algorithm setup process, the second also allows a remarkable improvement in terms of detection capabilities, as will be shown in the following.

Threshold Definition. Repeating the analysis for the whole engine operating range, it is possible to determine a theoretical response to a misfire event for each engine operating condition. The obtained engine speed wave forms can be used to evaluate LU_i. To do this, the angular sectors over which the combustion periods are evaluated have to be chosen.

The choice of the best angular sector is usually a compromise between a good signal sensitiveness (higher if the sectors are smaller) and the necessity of reducing noise influence (better if the sectors are larger). The final choice is usually experimentally determined as the one that guarantees the best signal to noise ratio, and can be a function of the engine operating condition. For example, the evaluation of the best sector positions and dimensions has given the results reported in Fig. 23 for the engine considered in this study. The best sectors have been determined trying to maximize the detection capability of the misfire detection algorithm on a properly designed set of tests.

This choice allows for calculating the LU_i values for each engine speed wave form representing the system response to a misfire event. LU_i, evaluated for the misfiring cylinder, represents a first try setup of the threshold expected to be used for the detection process. This result is very important in the setup process of the procedure, since each different choice in terms of angular sectors for combustion period evaluation requires its threshold value. Mapping, for example, the values obtained over the engine operating range for angular sectors of 78–186 deg, it is possible to determine the expected threshold (see Fig. 24).

The optimal threshold, experimentally determined (in case that angular sectors of 78–186 deg are applied), is also reported in the same figure, showing that the developed model enables a good first try evaluation of the expected threshold over the whole engine operating range. This has been done for one angular sector choice but is obviously possible for any other choice. Also in this case, the optimal threshold values have been determined trying to maximize the detection capability of the misfire detection algorithm on a properly designed set of tests.

Torsional Vibration Compensation. In order to study the torsional behavior effects on LU_i, we should obtain from the engine speed response wave forms (see, for example, Fig. 22) the corresponding combustion periods and then apply Eq. (11). Combustion period determination is equivalent to a moving average evaluation, as we already noticed in the previous section, whose result strongly depends on the amplitude and position of the angular sectors chosen. Therefore, even the corresponding LU_i wave form will depend on the angular sector chosen. As an example, LU_i calculation results divided by manifold pressure (i.e., normalized with respect to the load) are reported in Fig. 25 for the same tests of Fig. 22 and for three different angular sector choices: 78–223 deg after top dead center (ATDC) of the corresponding cylinder, 78–186 deg ATDC, and 71–124 deg ATDC.

Looking at the expected LU_i wave form after a misfire, it is easy to understand that the torsional vibration that follows a misfire event can become a source of false alarms due to the presence of a second LU_i peak. In addition, in the case of a multiple misfire event, the LU_i resulting wave form will be given by the interaction between two wave forms similar to those reported in Fig. 25.

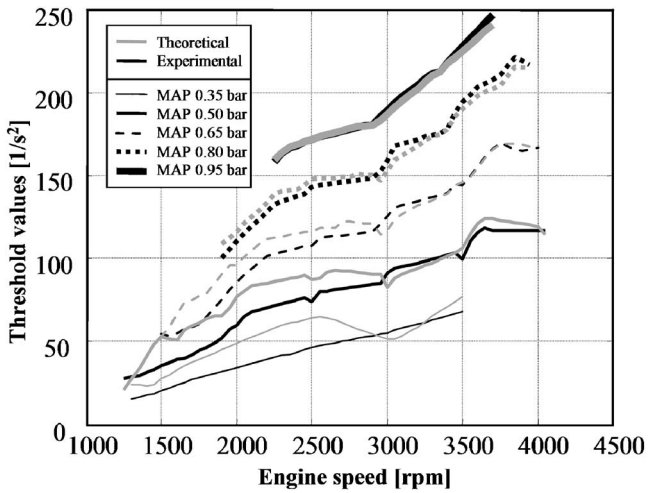


Fig. 24 Theoretical and experimental threshold definition

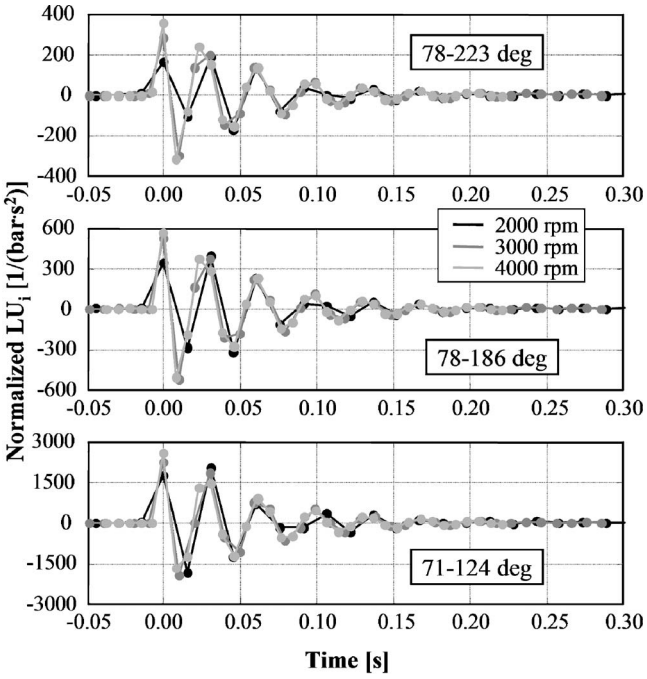


Fig. 25 Misfire effects over LU_i wave form

This can easily generate false alarms or misdetections. Therefore, it is important to filter out the effect of the torsional vibration once a misfire presence is detected. To do this, a specific wave form has to be stored in memory and used to compensate the LU_i values that follow the misfire detection. Assuming that the effect of the misfire can be neglected after 0.15 s the misfire event occurred, the number of LU_i values to be compensated is equal to

$$p = 2\frac{0.15n}{60} \tag{19}$$

which gives a maximum number of 30 points for a 6000 rpm engine operating condition.

It can be observed from Fig. 25 that the load normalized LU_i wave form can be considered unique over the engine operating range, but that it also depends on the angular sector choice. This means that for each different angular sector choice, it is necessary to store in memory a different load normalized LU_i wave form to be used for torsional behavior compensation.

Experimental Tests

A set of experimental tests has been performed running the Fire 1200 engine on a test bench. During the tests, all the main engine-load actuation parameters were externally controlled in real time using a virtual ECU (VECU) developed by this group of research [17] that completely substitutes the original ECU. This device allows for controlling the engine actuators (ignition coils and injectors) in all the cylinders, independent of the original ECU. The engine was equipped with in-cylinder pressure sensors, in order to evaluate the real torque production of the single cylinders. Two magnetic pickups were also used, one mounted in front of the camshaft pulley to generate a cycle reference signal, and another one facing the 116 flywheel teeth. Acquisition of the data was performed on a crankshaft angle basis, and the signal coming from the flywheel pickup was used as analog-to-digital (A/D) conversion clock. Also, the time intervals between subsequent teeth were acquired.

Since all the main engine actuation parameters were externally controlled, several misfire (missing ignition) and misfuel (missing injection) patterns were induced to analyze their effects on engine speed. The investigated engine operating range is shown in Fig. 26, taking as a load reference the intake manifold pressure value. The numbered dots and the gray area represent the engine operating conditions at which the missing combustion patterns reported in Table 4 have been induced during the experimental tests.

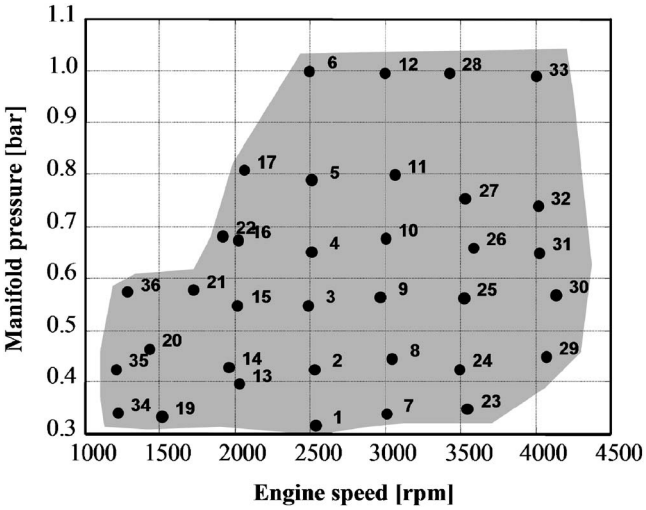


Fig. 26 Engine operating conditions investigated

Finally, a steady-state test was also performed at each engine operating condition, with nominal injection and ignition actuations on all cylinders.

For the sake of simplicity, the word "misfire" will be used in the following to define a generic missing combustion, either due to a missing injection or ignition.

The patterns chosen in the tests allow for the investigation of all the possible multiple misfire configurations, as already mentioned in the previous section.

- Multiple Misfire 1–3 will result in two adjacent misfires.
- Multiple Misfire 1–4 gives rise to a normal combustion between two misfires.
- Multiple Misfire 1–2 will result in two normal combustions between the misfires.

Results

The results obtained are reported as an example only for the engine operating condition at 4000 rpm and 6.70 bar BMEP already discussed in the previous sections. The threshold for misfire detection is approximately equal to 1000 l/s^2 for this operating condition and an optimal angular sector choice is 71–124 deg. The comparison between the LU_i values and the threshold should allow for the distinction of normal combustions from misfiring ones. This unfortunately does not happen if no torsional vibration compensation takes place, as can be seen in Fig. 27, where the results for four different tests at 4000 rpm and 6.70 bar BMEP and different misfiring patterns (respectively, Cylinder 1, the first test, 1–3 the second one, 1–4 the third, and 1–2 the last one) are given.

Table 4 Tests performed

Single misfuel on Cylinder 1 every 20 engine cycles
Single misfuel on Cylinder 2 every 20 engine cycles
Single misfuel on Cylinder 3 every 20 engine cycles
Single misfuel on Cylinder 4 every 20 engine cycles
Single misfire on Cylinder 1 every 20 engine cycles
Multiple misfuel on Cylinders 1 and 2 every 20 engine cycles
Multiple misfuel on Cylinders 1 and 3 every 20 engine cycles
Multiple misfuel on Cylinders 1 and 4 every 20 engine cycles

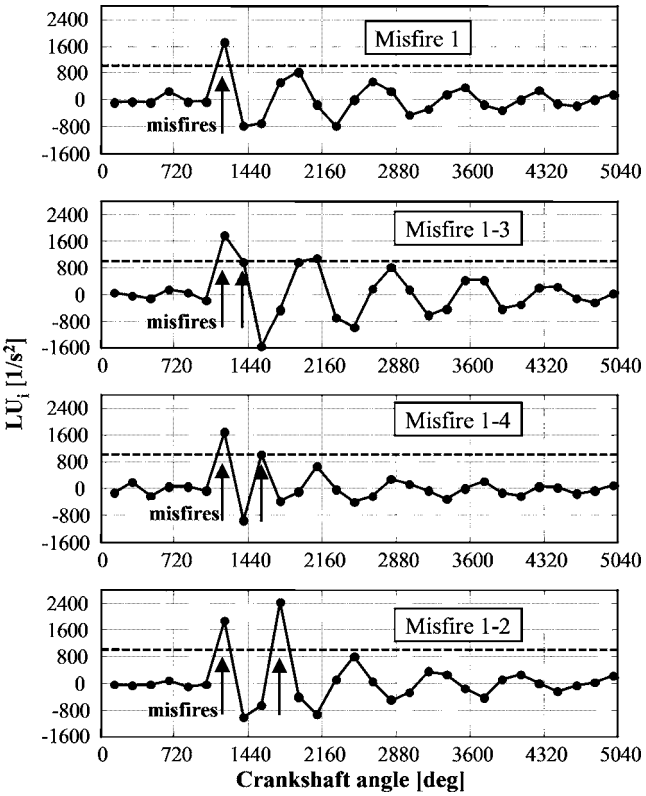


Fig. 27 ***LU**$_i$* **evaluated for four different tests at 4000 rpm and 6.70 bar BMEP with different misfire patterns**

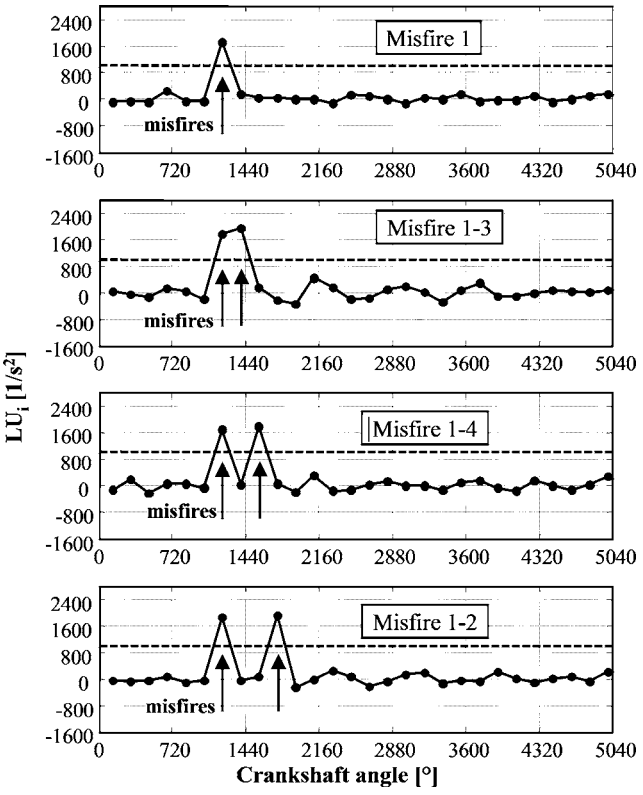


Fig. 28 ***LU**$_i$* **evaluated for four different tests at 4000 rpm and 6.70 bar BMEP with different misfire patterns applying the torsional vibration compensation**

As can be seen, some misfires could not be detected, and some false alarms arise. Applying the torsional vibration compensation to the same tests, the misfire detection is correct without any misdetection or false alarm (see Fig. 28). In order to apply the compensation after each misfire detection, the following steps have to be followed. The normalized LU_i wave form, corresponding to the angular sectors chosen, has to be first selected and then multiplied by the actual manifold pressure value. From the wave form (stored as a function of time), the values for the following combustions have to be evaluated, taking into account the actual engine speed and, therefore, the time distance between two adjacent LU_i evaluations; the value obtained has to be subtracted from the actual evaluated LU_i, in order to apply the compensation. This operation has to be performed for a number of combustions following the identified misfiring given by Eq. (19). The results are reported for the same tests of Fig. 27 in Fig. 28.

The misfire events after the torsional vibration compensation are now more clearly identified and the risk of misdetections or false alarms is much lower, enhancing the detection capability of the developed methodology. Analogous results have also been obtained for the other engine operating conditions investigated.

Figure 29 summarizes the results obtained in terms of detection capabilities over the whole engine operating range with and without the torsional vibration compensation, where the detection capability has been evaluated following Eq. (20):

$$dc = \frac{tm - md - fa}{tm} 100 \tag{20}$$

As can be seen, the difference in terms of detection capabilities is quite strong. This is due also to the fact that the majority of the tests conducted is of the multiple misfire type that strongly require the presence of the torsional vibration compensation.

Conclusions

The torsional vibration that arises after a misfire event can tend to hide the presence of other closely following missing combustions. The developed multiple misfire detection algorithm, compensating the torsional vibration effects after each detected misfire, proved to be very effective both for misfire detection and cylinder isolation. The methodology has been developed using both a torsional vibration model for the engine-load configuration taken into account and experimental data. The model proved to be

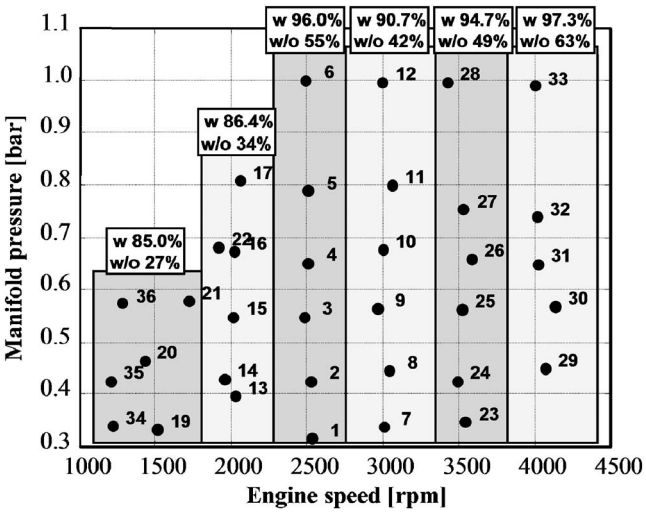


Fig. 29 Detection capability obtained over the whole engine operating range with (w) and without (w/o) torsional vibration compensation.

effective in describing the behavior of the system both in normal and misfiring operating conditions and allowed a quicker setup of the whole methodology.

The application of the methodology to other engine-load configurations (i.e., onboard a vehicle, an engine with a different number of cylinders) is possible, even though the wave forms identified for torsional vibration compensation, the threshold values, and the engine-load model will obviously be different. In addition, for onboard application, the system modifications due to the different gears inserted have to be taken into account. This will cause a different system response for each gear inserted, and for this reason, a different model is required (and therefore a different LU_i wave form for torsional vibration compensation has to be stored in memory) for each gear. In any case, following the path described in the paper, the setup of the methodology is also possible in these cases.

Nomenclature

θ = crank angle (rad)
$\theta_{\mathrm{TDC}m}$ = mth cylinder TDC crank angle position (rad)
θ_m = mth cylinder crank angle position (rad)
r = crank-slider mechanism radius (m)
l = connecting rod length (m)
A_p = piston area (m^2)
$\lambda = r/l$ = crank-slider radius/connecting rod length ratio
$p_m(\theta_m)$ = relative pressure inside mth cylinder (Pa)
$f(\theta_m)$ = $\sin(\theta-\theta_{\mathrm{TDC}m})+\lambda\sin[2(\theta-\theta_{\mathrm{TDC}m})]/2\sqrt{1-\lambda^2\sin^2(\theta-\theta_{\mathrm{TDC}m})}$ crank-slider mechanism function for the mth cylinder
p_{mot_m} = relative pressure inside mth cylinder for motored operating conditions (Pa)
T_{ind_m} = indicated torque for mth cylinder (N m)
$T_{\mathrm{ind\ mot}_m}$ = motoring indicated torque for mth cylinder (N m)
$T_{\mathrm{ind\ comb}_m}$ = combustion indicated torque for mth cylinder (N m)
M_{eq} = equivalent mass of the reciprocating masses (kg)
T_{r_m} = reciprocating torque for mth cylinder (N m)
T_{f_m} = friction torque for mth cylinder (N m)
f_m = external damping coefficient for mth cylinder (N m s/rad)
f = frequency (Hz)
ord = engine order
n = engine speed (rad/s)
$[\mathbf{I}]$ = inertia matrix (kg m^2)
$[\mathbf{k}]$ = stiffness matrix (N m/rad)
$[\mathbf{c}]$ = damping matrix (N m s/rad)
$\{\boldsymbol{\theta}\}$ = inertia angular position vector (rad)
$\{\mathbf{T}\}$ = torque applied to inertia vector (N m)
$f_{\lambda i}$ = natural frequency of the ith mode (Hz)
τ_i = combustion period relative to the ith combustion (s)
LU_i = misfire detection parameter relative to the ith combustion ($1/\mathrm{s}^2$)
t_k = time between consecutive teeth on the flywheel relative to the kth tooth (s)
k = generic tooth
$\Delta\theta$ = angular sector over which LU_i is evaluated (rad)
θ_{eng} = engine angular position (rad)
θ_{load} = load angular position (rad)
I_{eng} = engine inertia (kg m^2)
I_{load} = load inertia (kg m^2)
k = engine-load connection stiffness (N m/rad)
c = engine-load connection damping (N m s/rad)
p = number of LU_i values to be compensated from the torsional vibration effects
dc = detection capability of the algorithm
tm = total number of misfires performed
md = misdetections
fa = false alarms
H = harmonic order
T_L = load torque (N m)
BMEP = brake mean effective pressure
OBD = onboard diagnostics
TDC = top dead center
SI = spark ignited
CAD = computer aided design
MAP = manifold absolute pressure
ECU = electronic control unit
ATDC = after top dead center

Appendix

The terms that compare in Eq. (7) are here fully reported:

$$[\mathbf{I}]=\begin{bmatrix} I_0 & & & & & \\ & I_1 & & & & \\ & & \ddots & & & \\ & & & I_5 & & \\ & & & & I_6 & \end{bmatrix}$$

$$[\mathbf{k}]=\begin{bmatrix} k_1 & -k_1 & & & & \\ -k_1 & k_1+k_2 & \ddots & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & k_5+k_6 & -k_6 & \\ & & & -k_6 & k_6 & \end{bmatrix}$$

$$[\mathbf{c}]=\begin{bmatrix} c_1+f_0 & -c_1 & & & & \\ -c_1 & c_1+c_2+f_1 & \ddots & & & \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & c_5+c_6+f_5 & -c_6 & \\ & & & -c_6 & c_6+f_6 & \end{bmatrix}$$

$$\{\boldsymbol{\theta}\}=\begin{Bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{Bmatrix} \quad \{\mathbf{T}\}=\begin{Bmatrix} 0 \\ T_{\mathrm{ind1}}-T_{r1} \\ \vdots \\ T_{\mathrm{ind4}}-T_{r4} \\ 0 \\ -T_L \end{Bmatrix}$$

References

[1] California Air Resources Board, 1991, "Technical Status Update and Proposed Revisions to Malfunction and Diagnostic System Requirements Applicable to 1994 and Subsequent California Passenger Cars, Light-Duty Trucks, and Medium-Duty Vehicles—(OBD II)," CARB Staff Report No. 91-37.
[2] Ceccarani, M., Rebottini, C., and Bettini, R., 1998, "Engine Misfire Monitoring for a V12 Engine by Exhaust Gas Pressure Analysis," SAE Technical Paper No. 980518.
[3] Willimowski, M., and Isermann, R., 2000, "A Time Domain Based Diagnostic System for Misfire Detection in Spark-Ignition Engines by Exhaust Gas Pressure Analysis," SAE Technical Paper No. 2000-01-0366.
[4] Lee, A., and Pyko, J. S., 1995, "Engine Misfire Detection by Ionization Current Monitoring," SAE Technical Paper No. 950003.
[5] Chung, Y., Bae, C., Choi, S., and Yoon, K., 1999, "Application of a Wide Range Oxygen Sensor for the Misfire Detection," SAE Technical Paper No. 1999-01-1485.
[6] Azzoni, P. M., Cantoni, G., Ceccarani, M., Mazzetti, S., Minelli, G., Moro, D., and Rizzoni, G., 1995, "Measurement of Engine Misfire in a Lamborghini 533 V-12 Engine Using Crankshaft Speed Fluctuations," SAE Technical Paper No. 950837.
[7] Ribbens, W. W., and Park, J., 1994, "Road Tests of a Misfire Detection System," SAE Technical Paper No. 940975.

[8] Iida, K., Akishino, K., and Kido, K., 1990, "IMEP Estimation From Instantaneous Crankshaft Torque Variation," SAE Technical Paper No. 900617.
[9] Williams, J., 1996, "An Overview of Misfiring Cylinder Engine Diagnostic Techniques Based on Crankshaft Angular Velocity Measurements," SAE Technical Paper No. 960039.
[10] Moskwa, J. J., Wang, W., and Bucheger, D. J., 2001, "A New Methodology for Engine Diagnostics and Control, Utilizing "Synthetic" Engine Variables: Theoretical and Experimental Results," ASME J. Dyn. Syst., Meas., Control, **123**(3), pp. 528–534.
[11] Forster, J., Lohmann, A., Mezger, M., and Ries-Muller, K., 1997, "Advanced Engine Misfire Detection for SI-Engines," SAE Technical Paper No. 970855.
[12] Klenk, M., Moser, W., Muller, W., and Wimmer, W., 1993, "Misfire Detection by Evaluating Crankshaft Speed: A Means to Comply With OBD II," SAE Paper No. 930399.
[13] Plapp, G., Klenk, M., and Moser, W., 1990, "Methods of On-Board Misfire Detection," SAE Technical Paper No. 900232.
[14] Heywood, J. B., 1989, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York.
[15] Ligier, J. L., and Baron, E., 2002, *Acyclisme et Vibrations*, Technip, Paris.
[16] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E., 1998, "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," SIAM J. Optim., **9**(1), pp. 112–147.
[17] Cavina, N., Corti, E., and Ponti, F., 2001, "Setup of a Rapid Control Prototyping System for New Engine Control and Diagnostic Algorithms Testing," *Control and Diagnostics in Automotive Applications*, Genova, Italy, July, Paper No. 01A3040, pp. 1–9.

Systematic Assessment of Rigid Internal Combustion Engine Dynamic Coupling

D. Geoff Rideout[1]
Assistant Professor
Department of Mechanical Engineering,
S.J. Carew Building, Memorial University,
St. John's, NL, A1B 3X5, Canada
e-mail: grideout@engr.mun.ca

Jeffrey L. Stein
Professor
Automated Modeling Laboratory,
Department of Mechanical Engineering,
University of Michigan,
G029 Auto Lab, 1231 Beal Avenue,
Ann Arbor, MI 48109-2121
e-mail: stein@umich.edu

Loucas S. Louca
Lecturer
Department of Mechanical and Manufacturing Engineering,
University of Cyprus,
75 Kallipoleos Street,
P.O. Box 20537,
1678 Nicosia, Cyprus
e-mail: lslouca@ucy.ac.cy

Accurate estimation of engine vibrations is essential in the design of new engines, engine mounts, and the vehicle frames to which they are attached. Mount force prediction has traditionally been simplified by assuming that the reciprocating dynamics of the engine can be decoupled from the three-dimensional motion of the block. The accuracy of the resulting one-way coupled models decreases as engine imbalance and cylinder-to-cylinder variations increase. Further, the form of the one-way coupled model must be assumed a priori, and there is no mechanism for generating an intermediate-complexity model if the one-way coupled model has insufficient fidelity. In this paper, a new dynamic system model decoupling algorithm is applied to a Detroit Diesel Series 60 in-line six-cylinder engine model to test one-way coupling assumptions and to automate generation of a proper model for mount force prediction. The algorithm, which identifies and removes unnecessary constraint equation terms, is reviewed with the aid of an illustrative example. A fully coupled, balanced rigid body model with no cylinder-to-cylinder variations is then constructed, from which x, y, and z force components at the left-rear, right-rear, and front engine mounts are predicted. The decoupling algorithm is then applied to automatically generate a reduced model in which reciprocating dynamics and gross block motion are decoupled. The amplitudes of the varying components of the force time series are predicted to within 8%, with computation time reduced by 55%. The combustion pressure profile in one cylinder is then changed to represent a misfire that creates imbalance. The decoupled model generated by the algorithm is significantly more robust to imbalance than the traditional one-way coupled models in the literature; however, the vertical component of the front mount force is poorly predicted. Reapplication of the algorithm identifies constraint equation terms that must be reinstated. A new, nondecoupled model is generated that accurately predicts all mount components in the presence of the misfire, with computation time reduced by 39%. The algorithm can be easily reapplied, and a new model generated, whenever engine speed or individual cylinder parameters are changed. [DOI: 10.1115/1.2795770]

1 Introduction

Accurate prediction of engine vibration is important for the design of engine mounts, for prediction of loads transmitted to the vehicle frame, and for assessing whether vibration affects customer perceptions of quality.

Reciprocating engine components move within a block that has six degrees of freedom on its compliant mounts. Angular velocity of the block creates gyroscopic forces and torques on the reciprocating components within. These gyroscopic effects subsequently affect the forces and torques back on the block from the reciprocating components, the resulting motion of the block, and the mount forces transmitted to the vehicle frame. A "fully coupled" model, in which the reciprocation occurs within a block moving in three dimensions, can be computationally unwieldy [1]. To reduce model complexity and simulation time, mount forces have traditionally been predicted with a decoupled model as follows. First, reciprocating dynamics are simulated with a model in which the block is stationary. The required constraint forces and torques to keep the block stationary are then applied to a constant-inertia nonrunning model of the engine on its mounts.

Use of one-way coupled models is described in Refs. [1–3], where the modeler is cautioned against the use of "constant-inertia" models for predicting shaft speed and acceleration for all but low-speed or low-bandwidth control applications. Single- and six-cylinder models were constructed in Ref. [4] to show that the one-way coupled model should be restricted to well-balanced or low-speed engines, and only when the engine block is not subject to large pitch or yaw excitation by the vehicle.

Proper modeling [5] has been formally defined as the systematic determination of the model of minimal complexity that (a) satisfies the modeling objectives and (b) retains physically meaningful parameters and variables for design. Algorithms have been developed to help automate the production of proper models of dynamic systems. Recent work [6,7] has focused on systematically determining if partitions exist in a model, so that the use of one-way coupled models can be validated. Using an energy-based metric, an arbitrary, fully coupled model can be assessed to determine whether or not partitioning is possible without reliance on an assumed form of a one-way coupled model. Decoupled submodels, where appropriate, are automatically generated. If the system parameters or external inputs change, the intensity of the decoupling can be monitored and models of intermediate complexity can be automatically generated. If, for example, cylinder-to-cylinder variations are introduced in a low-speed balanced engine, or if the engine is simulated in a moving vehicle, the efficacy of a one-way coupled model can be monitored. If two-way coupling is required, the algorithm identifies the specific constraint equations in which the decoupling breaks down, thus increasing physical insight into the system dynamics. Thus, it is the hypothesis of this

[1]Corresponding author.

Contributed by International Gas Turbine Institute of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received June 28, 2006; final manuscript received July 11, 2007; published online February 29, 2008. Review conducted by Christopher J. Rutland. Paper presented at the 2004 ASME International Mechanical Engineering Congress (IMECE2004), Anaheim, CA, November 13–19, 2004.

Table 1 Generalized bond graph quantities

Variable	General	Translation	Rotation
Effort	$e(t)$	Force	Torque
Flow	$f(t)$	Velocity	Angular velocity
Momentum	$p=\int e\,dt$	Linear momentum	Angular momentum
Displacement	$q=\int f\,dt$	Displacement	Angular displacement
Energy	$E(p)=\int^p f\,dp$ $E(q)=\int^q e\,dq$	Kinetic potential	Kinetic potential

work that application of these existing decoupling algorithms can lead to systematic decoupled engine models for specific user-defined conditions.

The paper is organized as follows. Section 2 describes the decoupling search and model partitioning algorithm, and the bond graph formalism that facilitates its execution, with the aid of an illustrative example. Section 3 describes the in-line six-cylinder engine fully coupled model (FCM), and Sec. 4 shows the results of applying the partitioning algorithm to a balanced and an unbalanced engine. Discussion and conclusions follow in Secs. 5 and 6.

2 Decoupling Search and Model Partitioning

As described in Ref. [6], partitions are defined as one-way coupled groups of dynamic elements that are created when negligible constraint equation terms are removed from a model. Individual constraint terms are assessed by examining their "activity" [8] relative to the maximum term activity in the equation. Activity over a time interval t_1 to t_2 is defined as

$$A=\int_{t_1}^{t_2}|P|dt \tag{1}$$

where P is the instantaneous power, the product of generalized "effort" (e.g., force, torque, voltage, pressure) and "flow" (e.g., velocity, angular velocity, current, volume flow rate). Activity is always positive, and either constant or monotonically increasing, over a given time interval. Because of the use of the power-based activity metric, and the need to systematically search all model constraint equations for inactive terms, bond graphs are a convenient (but not required) formalism to generate the engine model prior to application of the partitioning algorithm.

2.1 Bond Graph Modeling Formalism. In bond graphs [9], generalized inertias and capacitances store energy as a function of the system state variables, sources provide inputs from the environment, and generalized resistors remove energy from the system. The state variables are generalized momentum and displacement for inertias and capacitances, respectively. The time derivatives of generalized momentum p and displacement q are generalized effort e and flow f. Table 1 expresses the generalized power (effort and flow) variables and energy (momentum and displacement) variables in the terminology of common engineering disciplines.

Power-conserving elements allow changes of state to take place. Such elements include power-continuous generalized transformer (TF) and gyrator (GY) elements that algebraically relate elements of the effort and flow vectors into and out of the element. In certain cases, such as large motion of rigid bodies in which coordinate transformations are functions of the geometric state, the constitutive laws of these power-conserving elements can be state modulated. Dynamic force equilibrium and velocity summations in rigid body systems are represented by power-conserving elements called 1 and 0 junctions, respectively.

Sources represent ports through which the system interacts with its environment. The power-conserving bond graph elements—TF, GY, 1 junctions, 0 junctions, and the bonds that connect them—are collectively referred to as "junction structure." Table 2 defines the symbols and constitutive laws of sources, storage and dissipative elements, and power-conserving elements in scalar form. Bond graphs may also be constructed with the constitutive laws and junction structure in matrix-vector form, in which case the bond is indicated by a double line.

Table 2 Bond graph elements

	Symbol	Constitutive law (Linear)	Causality constraints
		Sources	
Flow	Sf→	$f=f(t)$	Fixed flow out
Effort	Se→	$e=e(t)$	Fixed effort out
		Energetic elements	
Inertia	→I	$f=\frac{1}{I}\int e\,dt$	Preferred integral
	→I	$e=I\frac{df}{dt}$	
Capacitor	→C	$e=\frac{1}{C}\int f\,dt$	Preferred integral
	→C	$f=C\frac{de}{dt}$	
Resistor	→R	$e=Rf$	None
	→R	$f=\frac{1}{R}e$	
		Port elements	
Transformer	$\overset{1}{}\ \overset{2}{}$ →TF→	$e_2=ne_1$ $f_1=nf_2$	Effort in-effort out or flow in-flow out
Modulated transformer	$\overset{n}{\downarrow\theta}$ →MTF→ $n(\theta)$	$e_2=n(\theta)e_1$ $f_1=n(\theta)f_2$	
Gyrator	$\overset{1}{}\ \overset{2}{}$ →GY→	$e_2=nf_1$ $e_1=nf_2$	Flow in-effort out or effort in-flow out
Modulated gyrator	$\overset{n}{\downarrow\theta}$ →MGY→ $n(\theta)$	$e_2=n(\theta)f_1$ $e_1=n(\theta)f_2$	
		Constraint nodes	
1 junction	$\overset{1}{}\ \overset{2}{}$ →1→ $\overset{\swarrow}{3}$	$e_2=e_1-e_3$ $f_1=f_2$ $f_3=f_2$	One flow input
0 junction	$\overset{1}{}\ \overset{2}{}$ →0→ $\overset{\swarrow}{3}$	$f_2=f_1-f_3$ $e_1=e_2$ $e_3=e_2$	One effort input

Power bonds contain a half-arrow that indicates the direction of algebraically positive power flow, and a causal stroke normal to the bond that indicates whether the effort or flow variable is the input or output from the constitutive law of the connected elements. The constitutive laws in Table 2 are consistent with the placement of the causal strokes. Full arrows are reserved for modulating signals that represent powerless information flow such as orientation angles that determine the transformation matrix between a body-fixed and inertial reference frame.

The engine bond graphs that follow will contain bonds and elements with both scalar and vector governing equations. The reader is referred to Ref. [9] for a more thorough development of bond graphs.

Once a bond graph model of a system has been constructed, "conditioning" and "partitioning" algorithms are applied as summarized below with the aid of an illustrative example.

2.2 Model Conditioning and Partition Search Using Relative Activity. The algorithm (described in detail in Ref. [6]) consists of the following general steps, as depicted in Figs. 1 and 2.

Step 0. Construct a bond graph model of the system that can be considered the "full" model inasmuch as its complexity captures

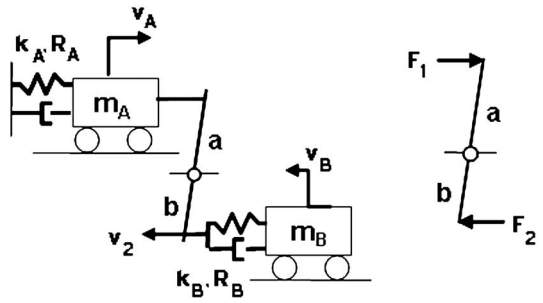


Fig. 1 Illustrative mechanical system example

the dynamics of interest. Record the outputs of interest from the full model for later comparison with reduced or partitioned models.

Step 1: Relative Activity Calculation. Calculate the activity of each bond in the graph, and compare, at each junction, the activity of each connected bond to the junction maximum. Low relative activity of bond i at a junction implies that

i. for a 0 junction with n bonds, the flow f_i can be neglected in the flow constraint equation

$$\sum_{j=1}^{n} f_j = 0$$

ii. for a 1 junction with n bonds, the effort e_i can be neglected in the effort equation

$$\sum_{j=1}^{n} e_j = 0$$

Locally inactive bonds are defined as those with an activity ratio falling below a user-defined threshold:

$$\frac{A_i}{\max(A_i)} < \varepsilon$$

Step 2: Conditioning. "Condition" the bond graph by converting the bonds with negligible activity to modulated sources. Table 3 gives examples of the conversion implications for bonds depending on the junctions or elements to which they are attached. The variables f_i and e_i represent the flow and effort of the bond with label i. For the internal bond case, if the activity of Bond 1 is low compared to the other bonds at the 0 junction, but is on the order of the other bond activities at the 1 junction, then the flow input to the 0 junction can be eliminated. The effort e_1 input to the 1 junction is delivered by a modulated effort source. Conditioning the bond into the 0 junction removes a term from the flow constraint equation as shown. The presence of the modulated source does not affect the 1-junction equations in either example. If the bond activity is instead locally inactive at the 1 junction, then the bond is converted to a modulated flow source imposing f_1 on the 0 junction.

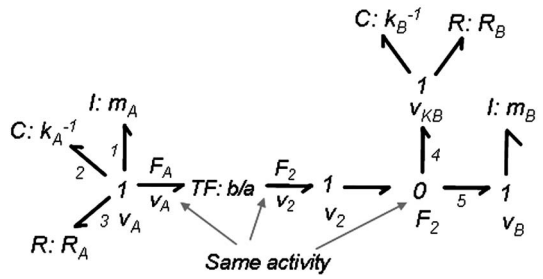


Fig. 2 Example system bond graph

Table 3 Bond conversion

Original	Conditioned	
Internal Bond	$A_{1a} << A_{m+1,\ldots,n}$	
0-junction equations	**reduced 0-junction eq'ns**	**1-junction eq'ns (unaffected)**
$e_{m+2,\ldots n} = e_{m+1}$ $e_1 = e_{m+1}$ $f_{m+1} = -(f_{m+2} + \ldots f_n) - f_1$	$e_{m+2,\ldots n} = e_{m+1}$ $e_1 = e_{m+1}$ $f_{m+1} = -(f_{m+2} + \ldots f_n)$	$f_{3,\ldots m} = f_2$ $f_1 = f_2$ $e_2 = -(e_3 + \ldots + e_m) + e_1$
Transformer	$A_{1a} << A_{m+1,\ldots,n}$	**Equivalent**
Gyrator	$A_{1a} << A_{m+1,\ldots,n}$	**Equivalent**

Table also shows examples of TF and GY bond conversion. As power-conserving elements, TF and GY elements have the same activities for Bonds 1a and 1b.

Step 3: Subgraph Identification. Identify bond subgraphs (collections of bond graph elements) that are connected by modulating signals instead of power bonds. Whereas a power bond is a conduit for two-way flow of information, the one-way modulating signals carry information from a locally "driving" element to a locally "driven." If removing modulating signals from the model results in two or more separate bond graphs, then subgraphs have resulted from bond conditioning and the most important prerequisite for partitioning has been met.

Step 4: Partitioning. Identify driving and driven partitions—subgraphs between which all modulating signals carry information in the same direction (i.e., from subgraph i to subgraph j).

If desired, the analyst can reduce the model as follows. If the only outputs of interest are in driving partitions, eliminate the driven partitions. If an output of interest is associated with a driven partition element, replace its driving partition(s) with the necessary inputs (time histories) to excite the driven partition.

2.3 Illustrative Example. Figure 1 shows two mass-spring-damper subsystems connected by a lever (power-conserving TF). The lever provides both a velocity and a force constraint:

$$v_2 = \frac{b}{a} v_1 \quad F_1 = \frac{b}{a} F_2 \tag{2}$$

Writing Newton's second law for m_A, and recognizing that the relative velocity of the endpoints of k_A is equal to the mass velocity v_A gives

$$m_A \dot{v}_A + R_A v_A + k_A x_A + F_1 = 0 \tag{3}$$

The preceding equation contains four force terms that are associated with common velocity v_A, suggesting a bond graph 1 junction with four bonds.

The following equation defines the relative velocity v_{k_B} of the endpoints of the spring k_B (and parallel damper R_B):

$$v_{k_B} = v_2 - v_B \tag{4}$$

These three velocities are associated with the combined spring and damper force F_2, suggesting a 0 junction with three bonds. Finally, we can associate the following effort equation terms with the spring velocity v_{k_B}:

$$F_2 = k_B \int v_{k_B} dt + R_B v_{k_B} \tag{5}$$

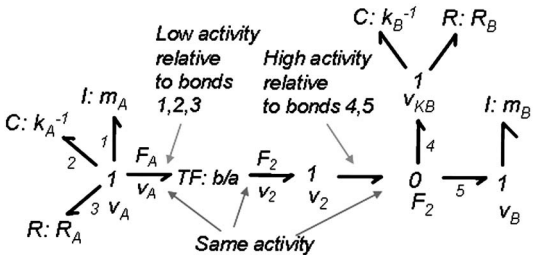


Fig. 3 Activity relations for decoupled system

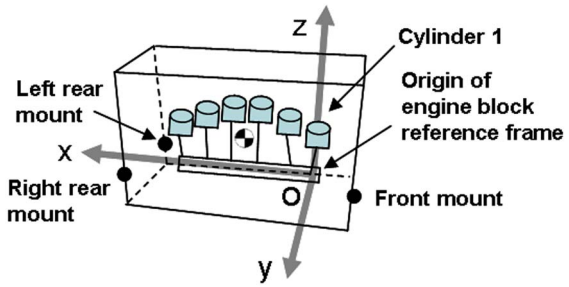


Fig. 5 Engine schematic

Figure 2 shows the system bond graph, with the preceding equations represented by bonds at the 1 and 0 junctions.

The above equations can be combined and rearranged to give the state equations in conventional first-order form

$$\dot{v}_A = -\frac{R_A}{m_A}v_A - \frac{k_A}{m_A}x_A - \frac{1}{m_A}F_1$$

$$\dot{x}_A = v_A$$

$$\dot{v}_B = -\frac{R_B}{m_B}\left(\frac{b}{a}\right)v_A - \frac{k_B}{m_B}\left(\frac{b}{a}\right)x_A + \frac{R_B}{m_B}v_B + \frac{k_B}{m_B}x_B \tag{6}$$

$$\dot{x}_B = v_B$$

where

$$F_1 = \frac{b}{a}\left[k_B\left(\frac{b}{a}x_A - x_B\right) + R_B\left(\frac{b}{a}v_A - v_B\right)\right]$$

Suppose the relative values of k_B, R_B, k_A, R_A, and the lever ratio b/a render the lever force on m_A negligible compared to the inertial, spring, and damper forces. This would result in low relative activity of the $-F_1$ term in Eq. (3). Suppose further that the transformed velocity v_A is essential for defining the velocities of k_B and R_B and thus the dynamic response of m_B. The term v_2 would be locally active in Eq. (4). The implications for bond activity are described in Fig. 3.

A one-way coupled model would result, in which k_A, R_A, and m_A could be simulated in isolation to predict the response of m_A and to generate v_A. The velocity v_2 $(=b/av_A)$ could then be extracted from the m_A-k_A-R_A subsystem to excite the dynamics of m_B-k_B-R_B. Figure 4 shows the conditioned bond graph and partitions resulting from low activity of the bond associated with the $-F_1$ term in Eq. (3).

The partitioned state equations can be derived from the bond graph, and are given below: Driven,

$$\begin{Bmatrix} \dot{v}_A \\ \dot{x}_A \end{Bmatrix} = \begin{bmatrix} -\frac{R_A}{m_A} & -\frac{k_A}{m_A} \\ 1 & 0 \end{bmatrix} \begin{Bmatrix} v_A \\ x_A \end{Bmatrix}$$

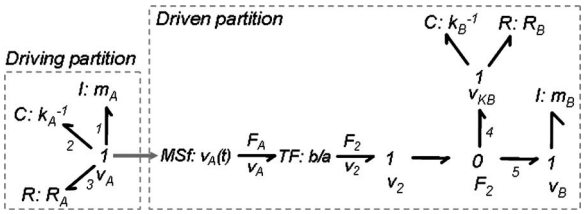


Fig. 4 Conditioned bond graph showing partitions

$$\begin{Bmatrix} v_A \\ x_A \end{Bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} v_A \\ x_A \end{Bmatrix} \tag{7}$$

Driving,

$$\begin{Bmatrix} \dot{v}_B \\ \dot{x}_B \end{Bmatrix} = \begin{bmatrix} \frac{R_B}{m_B} & \frac{k_B}{m_B} \\ 1 & 0 \end{bmatrix} \begin{Bmatrix} v_B \\ x_B \end{Bmatrix} + \begin{bmatrix} -\left(\frac{b}{a}\right)\frac{R_B}{m_B} & -\left(\frac{b}{a}\right)\frac{k_B}{m_B} \\ 0 & 0 \end{bmatrix} \begin{Bmatrix} v_A \\ x_A \end{Bmatrix} \tag{8}$$

The parameters b/a, m_B, k_B, and R_B do not appear in Eq. (7). The driving output v_A is extracted from Eq. (7) and input to the driving subsystem in Eq. (8).

The partitions can be simulated sequentially or in a parallel computing environment, with each partition having a lower-dimension state space than the original model.

3 Engine Model Development

A Detroit Diesel Series 60 in-line six-cylinder four-stroke direct injection engine was modeled by Hoffman and Dowling in Ref. [4] to study the ability of a one-way coupled model to predict engine mount forces during steady-state operation, compared to a FCM. The one-way coupled model (OWCM) mount forces in Ref. [4] were generated by (1) simulating the engine with the block rigidly constrained to the inertial frame, (2) measuring the rigid constraint forces and torques, and (3) applying the constraint forces and torques to the nonrunning engine on its mounts. This application is revisited here as a test case for the conditioning and partitioning algorithms, given that a set of realistic engine parameters and measured combustion forces is available. The engine is run at low speed as in Ref. [4], without a large effective crankshaft inertia to represent the vehicle mass and gearing. Such as scenario would be useful for the study of engine and body vibrations in an idling or parked running vehicle, e.g., at a rest stop. An overall rotary damping rate was used to achieve the desired steady-state speed.

The in-line six-cylinder architecture is inherently balanced for the first and second engine rotational speed harmonics. The dominant component of the engine mount forces for an engine on a test stand is therefore third order. The OWCM in Ref. [4] showed single-digit percentage errors, compared to a FCM, in predicting the third-order force component for an engine with no cylinder-to-cylinder variation in mechanical parameters or combustion forces. As imbalance was added or actual measured combustion forces were used in Ref. [4], model predictive ability was reduced.

In order to predict mount forces without relying on assumption and intuition about the feasibility of decoupling, a new FCM of the engine was developed by the authors using the bond graph formalism, and was subjected to the conditioning and partitioning

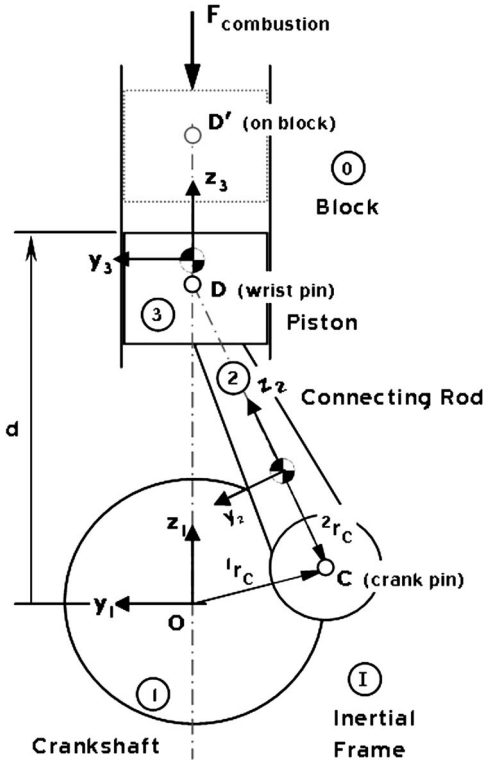


Fig. 6 Individual cylinder slider crank

algorithms to

(1) verify that decoupling could be systematically found without assuming the form of the OWCM
(2) compare the resulting partitioned model with the OWCM in Ref. [4], and demonstrate the physical insight gained
(3) quantify and locate the sites where decoupling became invalidated by a significant cylinder-to-cylinder variation that would create imbalance.

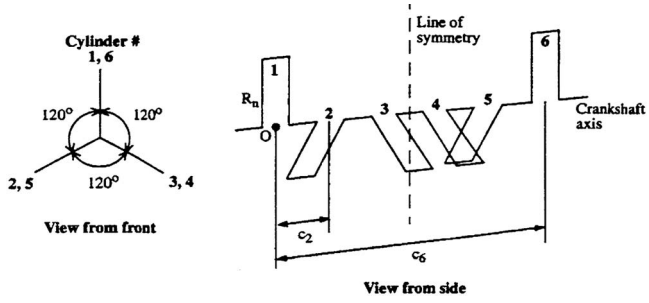


Fig. 7 Crankshaft schematic [4]

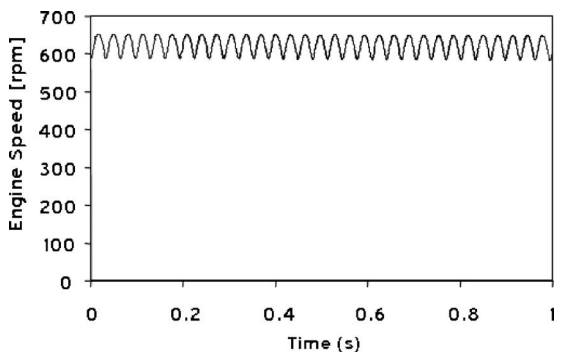


Fig. 8 Balanced engine speed

The vector bond graph formulation for multibody systems [10] was chosen, in which velocity vectors are defined with respect to coordinate frames affixed to individual bodies. Using this formalism, a model based on the full model of Hoffman and Dowling [4] to represent the engine vibrations was constructed. The basic building blocks of this model are given in Ref. [11]. Figures 5–7 show schematics of the engine, individual cylinder slider-crank mechanisms, and the crankshaft. The respective body-fixed coordinate system origins and orientations are as shown. The block and crankshaft origins were both located at point O, and their respective coordinate systems were initially coincident, with the z axis upward through the bore of Cylinder 1. The "front" mount is located at the front of the engine (near Cylinder 1) along the vertical centerline, and the "left-rear" and "right-rear" mounts are located to the left and right of the block-fixed x-z plane beyond Cylinder 6. The mounts are modeled as three orthogonal linear springs with directions parallel to the block-fixed system.

These forces, when applied to the engine with an internal damping value of 30 Nm s/rad at the main bearing, produced a steady-state speed of approximately 65 rad/s (620 rpm), as shown in Fig. 8.

The model parameters and combustion forces are from Ref. [12]. For the present case study, the Cylinder 1 combustion force versus crank angle curve was used for all cylinders, with the curves for Cylinders 2–6 shifted along the crank angle axis to mimic the firing order.

Parasitic elements (stiff spring and damper elements in parallel) were used in the model at all pin joints. This is a standard procedure to allow a small violation of the ideal rotational joint constraint such that an explicit set of ordinary differential equations can be derived [13]. Parasitic stiffnesses were tuned to suppress constraint violation displacements below 0.025 mm. In the description that follows, the left superscript indicates the reference frame number according to Fig. 6. The position or velocity of a body's center of gravity is indicated by a numerical subscript, e.g., ${}^1\mathbf{r}_1$ locates the crankshaft center of gravity with respect to point O, in crankshaft-fixed coordinates. The position or velocity of an arbitrary point is indicated by a letter subscript, e.g., ${}^1\mathbf{r}_C$ in Fig. 6.

Numerical simulations were performed within the 20SIM [14] bond graph simulation environment on a Pentium IV computer using the Vode–Adams variable-step stiff system integrator. Absolute and relative integration tolerances were set at 10^{-6}. The Appendix lists model parameters.

4 Simulation Results for Balanced Engine

After applying the conditioning algorithm (Sec. 2.2, Steps 1 and 2) to determine local activities, partitions in the balanced engine model were identified for a local activity threshold of 0.3% of the maximum equation activity.

4.1 Balanced Engine Partitions. The partitions are summarized below, are shown schematically in Figs. 9 and 10, and are consistent with the notion of reciprocation about a fixed crankshaft axis as per the OWCM in Ref. [4].

"Reciprocating" (driving partition) contains the following:

- crankshaft spin axis degree of freedom and velocity ${}^1\omega_{1x}$
- internal friction
- crankpin translational velocity in the crankshaft y-z plane
- coordinate transformations between crankshaft, connecting rods, and pistons in the y-z reciprocation plane
- connecting rod spin axis degree of freedom
- connecting rod inertial forces in the y-z plane
- piston motion and inertial forces in the z direction
- combustion forces

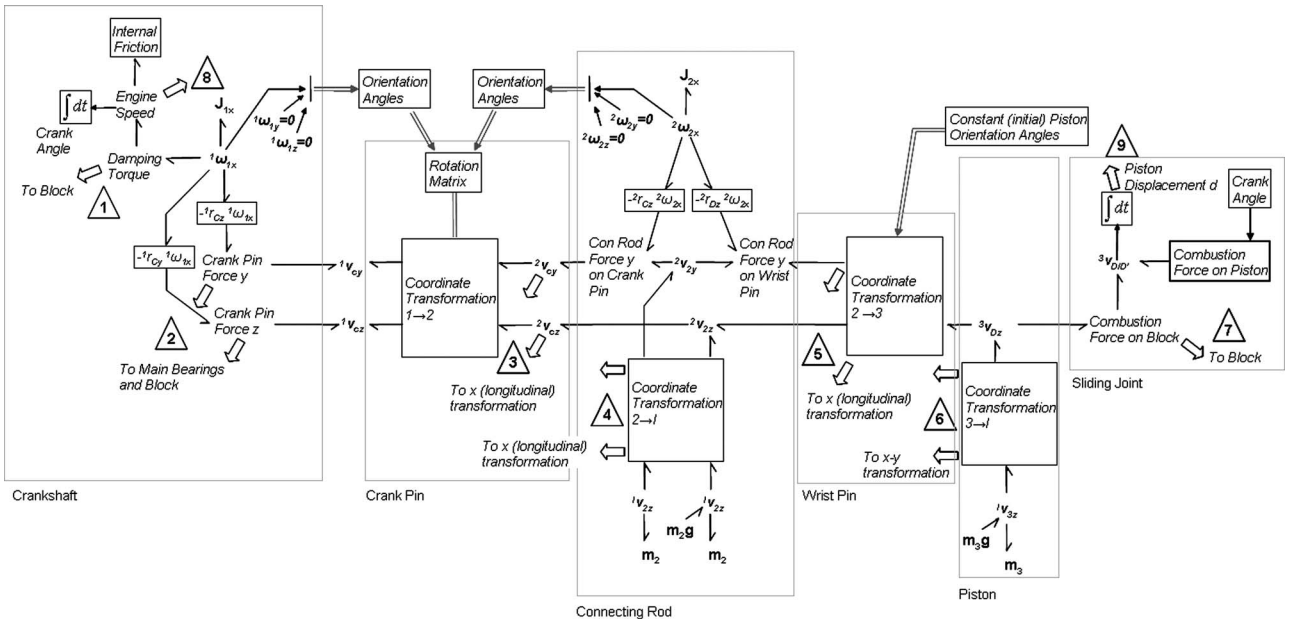


Fig. 9 Driving partition power flow schematic

The absence of the y and z components of the crankshaft angular velocity implies that the crankshaft spin axis (and thus the connecting rod y-z reciprocation planes) may be considered fixed.

"Extended block" (driven partition) contains the following:

- complete engine block and main bearings from FCM
- crankshaft translational velocity; mass and inertial forces in the x, y, z directions
- crankshaft ω_y and ω_z degrees of freedom
- connecting rod ω_y, ω_z, v_x degrees of freedom; mass and inertial force in the x direction
- piston angular motion; piston mass and inertial forces in the x and y directions
- sliding joint forces and moments on block in the x and y directions as a function ofpiston vertical displacement

95 driving outputs (and thus driven partition inputs) result from

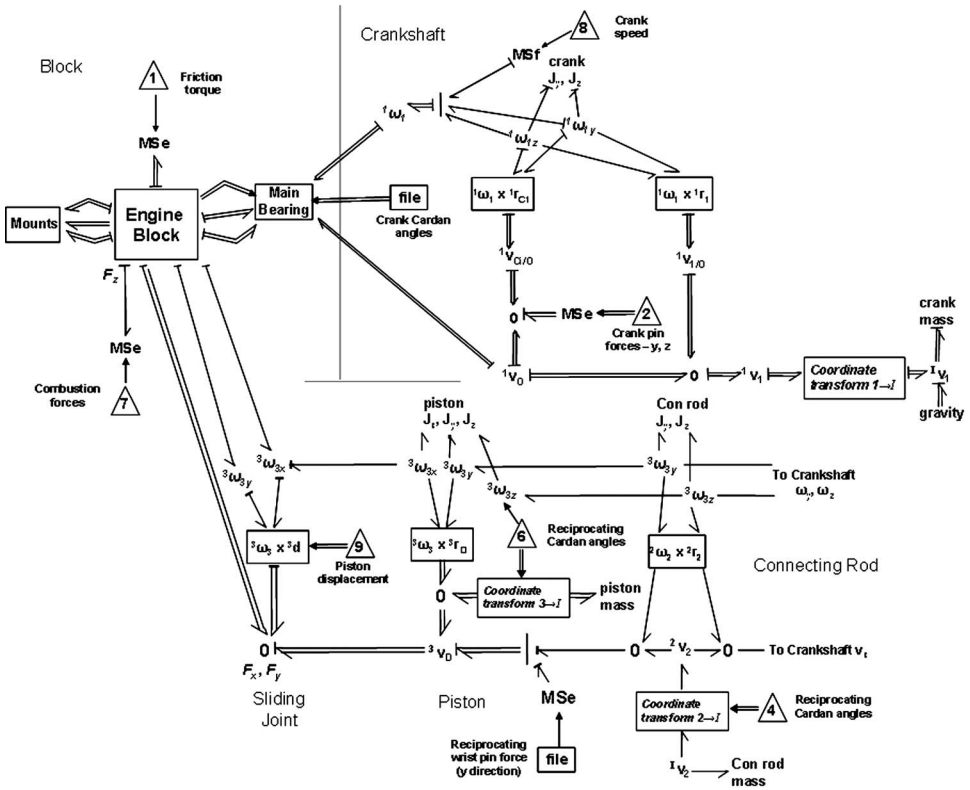


Fig. 10 Driven partition power flow schematic

Table 4 Description of negligible constraint terms

Markers from Figure 9	Modulating Signal	Constraint Equation	Remarks
1	Internal friction.	$\left({}^0\omega_{1/0}\right)_x \cong {}^0\omega_{1x} - {}^0\cancel{\omega}_{0x}$	Block has negligible spin about crank axis.
2	Crank pin force transmitted to main bearings.	${}^1v_{Cy} \cong {}^1\cancel{v}_{1y} - {}^1r_{Cy}\,{}^1\omega_{1x}$ ${}^1v_{Cz} \cong {}^1\cancel{v}_{1z} - {}^1r_{Cz}\,{}^1\omega_{1x}$	Total crank pin velocity ≈ relative velocity w.r.t. crankshaft.
3	Crank pin y and z velocity w.r.t. frame 2.	${}^2\begin{Bmatrix} F_{Cy} \\ F_{Cz} \end{Bmatrix} \cong \begin{bmatrix} \cancel{r}_{12} & r_{22} & r_{32} \\ \cancel{r}_{13} & r_{23} & r_{33} \end{bmatrix} {}^1\begin{Bmatrix} F_{Cx} \\ F_{Cy} \\ F_{Cz} \end{Bmatrix}$	Crank pin force in vertical reciprocating plane is not affected by main bearing axial force.
4	Connecting rod y and z force and velocity components.	Remove first row and column from transformation matrix.	Con rod force/velocity: longitudinal components insignificant.
5	Wrist pin force transmitted to connecting rod.	${}^2\begin{Bmatrix} v_{Dy} \\ v_{Dz} \end{Bmatrix} \cong \begin{bmatrix} \cancel{r}_{22} & r_{23} \\ \cancel{r}_{32} & r_{33} \end{bmatrix} {}^3v_{Dz}$	Connecting rod motion defines piston velocity along fixed vertical axis.
6	Piston inertial z velocity and transformed inertial force.	Remove first and second rows and columns from transformation matrix.	Piston velocity in local z frame ≈ total piston velocity in inertial frame.
7	Combustion force on block.	$\dot{d} \cong {}^0v_D - {}^0\cancel{v}_{D'}$	Total piston velocity ≈ relative velocity between piston and cylinder.
8	Crankshaft spin speed	$J_{1x}\dot{\omega}_{1x} \cong -R\omega_{1x} + \sum_{i=1}^{6} r_{zi}F_{Cyi} - \sum_{i=2}^{5} r_{yi}F_{Czi}$	Negligible moment due to local-z forces from crank pins 1 and 6

conversion of power bonds to modulating signals.

Figure 9 shows a "word bond graph" of the driving partition. One of the cylinder submodels is shown in its entirety, but the omitted cylinders are identical. Two-way power bonds have half-arrows, and one-way signals have full arrowheads. The block arrows represent signals from conditioned bonds, and carry the power variables associated with eliminated constraint terms. These are the outputs of the driving partition that are used as inputs to the driven partition. These outputs are typically written to a data file during simulation of the reciprocating dynamics, and then can be used later as inputs to the driven block partition.

The dominant power flow path is from the combustion force source, through the piston-fixed z axis (${}^3\mathbf{v}_{Dz}$), to translation of the connecting rod in its y-z plane (${}^3\mathbf{v}_{2y}, {}^2\mathbf{v}_{2z}$) and rotation about its x axis (${}^2\boldsymbol{\omega}_{2x}$), to rotation of the crankshaft about its longitudinal axis (${}^1\boldsymbol{\omega}_{1x}$), and transmission of forces to the block y-z plane ("crank pin force y,z").

The crankshaft and connecting rod orientation angles can be generated by integrating an angular velocity vector containing the spin component only. The physical interpretation is that the piston and connecting rod move in a reciprocating body-fixed y-z plane that can be assumed fixed to the inertial y-z frame.

The power flow paths within the driven partition are shown schematically via the partial bond graph of Fig. 10. Some individual elements are removed for clarity, including parasitic constraining springs and some coordinate transformation junction structure. Again, double-lined power bonds are vector bonds with which velocity and force/torque vectors are associated, while single-lined bonds carry scalar power variables. Some causal strokes are shown to indicate where forces and torques are being imposed upon the block and crankshaft. Given that some bond graph structure is not included, the reader is cautioned against assigning causal strokes to all bonds of the graph.

The engine block, main bearing, and engine mounts submodels are unchanged from the FCM. The driven partition differs from the OWCM in Ref. [4] in that it is not a constant-inertia model of the block with a stationary crankshaft, connecting rods, and pistons. For example, the piston displacement d is a required input to

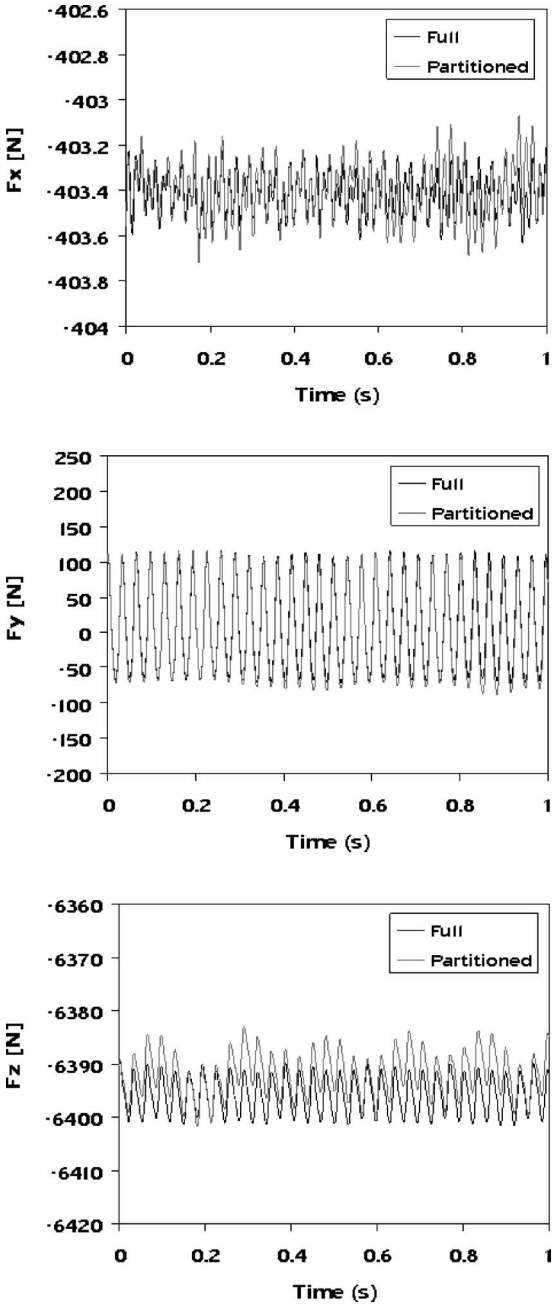


Fig. 11 Partitioned model: front mount forces

the driven partition. The varying piston height significantly changes the moments on the block about its x and y axes. In a constant-inertia driven OWCM, the piston displacements with respect to the block would remain constant.

The crankshaft translates with the block in the driven partition, as indicated by the vector bonds from the crankshaft velocity vector node (${}^1\mathbf{v}_1$) to the mass and gravity force source. The crankshaft angular velocity components in the body-fixed y and z directions (${}^1\boldsymbol{\omega}_{1y}, {}^1\boldsymbol{\omega}_{1z}$), are present in the driven partition, and create relative velocity contributions to the crankshaft and main bearings through the ${}^1\boldsymbol{\omega}_1 x {}^1\mathbf{r}_1$ block.

Table 4 describes the inactive terms according to the numbers inside triangles next to the block arrows in Fig. 9, and gives the physical interpretation of the resulting reduced constraint equation. Negligible terms are struck out in gray in the table. When interpreting Table 4, recall that an inactive bond at a velocity node

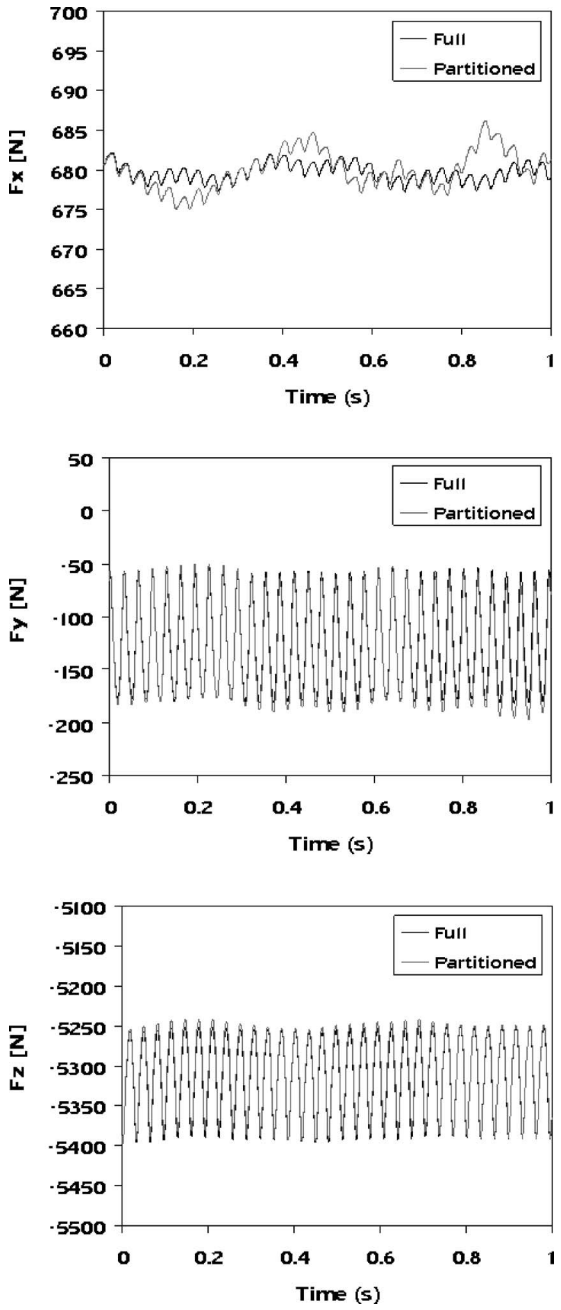


Fig. 12 Partitioned model: left-rear mount forces

represents a negligible term in a force/torque equation and vice versa. Bond conditioning removes the force/torque term from the equation and moves it to the driven dynamics. The node velocity then modulates the driven dynamics.

Figures 11 and 12 compare the front- and left-rear steady-state mount force predictions of the fully coupled and partitioned models. The partitioned model forces were generated by running the driving partition, saving the required driving signals (see Fig. 9) to a file, and using the file to excite the driven partition containing the block and mounts. The right-rear mount forces are not shown, but are qualitatively similar to the left rear.

The front mount is characterized by lower vibration amplitudes due to its location in the x-z plane. The partitioned model predicts all the major vibrating components accurately, and also matches the dc offset of all components. Table 5 shows single-digit percentage differences between the fully coupled and partitioned models in predicting the magnitude of the dominant third-order vibration amplitudes.

Table 5 Balanced engine mount force components: partitioned model

Mount force component		Freq. (Hz)/Magnitude (N)		% error (Mag.)
		Full model	Partitioned	
Front	x	31/0.055	31/0.058	5.5
	y	31/78	31/80	2.5
	z	31/4.2	31/4.25	1.9
		63/0.69	62/0.69	3.5
Left rear	x	31/0.81	31/0.87	7.9
	y	31/53	31/55	3.0
	z	31/60	31/61	1.6
Right rear	x	31/0.90	31/0.97	8.0
	y	31/53	31/55	3.0
	z	31/62	31/63	1.7

4.2 Misfire Mount Force Prediction. Next, an imbalance-inducing combustion event is simulated by setting the Cylinder 1 combustion force to zero at all crank angles. Figure 13 shows the resulting engine speed variation. As a result of the misfire, the engine block motions are severe enough to affect the reciprocating forces and moments in the block-fixed frame and compromise decoupling. Relative activity calculation and bond graph conditioning, using the same 0.3% threshold as for the balanced engine, show that creation of two completely decoupled partitions is no longer possible. For illustrative purposes, Figs. 14 and 15 are provided to show that the original partitioned model nonetheless predicts several mount force components accurately. However, significant deviations in the predictions of the vertical (z) component of the front mount force arise. The dc component is closely predicted, but the variation about the mean is not. As expected, given the nature of the imbalance, first-order vibration components dominate the response as shown in the figures.

The relative activity measure also heightens physical insight by allowing the analyst to determine systematically which new constraint terms are required to predict the front mount z component effectively in the face of misfire. Consider the crankshaft, where the equations defining the velocities of the crank pins are partitioned for the balanced engine. Expanding the vector equation

$$^1v_C = {}^1v_1 + {}^1\omega_1 \times {}^1r_C \tag{9}$$

in scalar form, eliminating negligible terms, and partitioning, the components of crank pin velocity $\mathbf{v}_C$ are defined by
Driving,

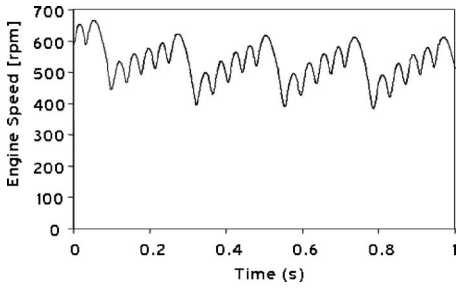


Fig. 13 Engine speed with misfire

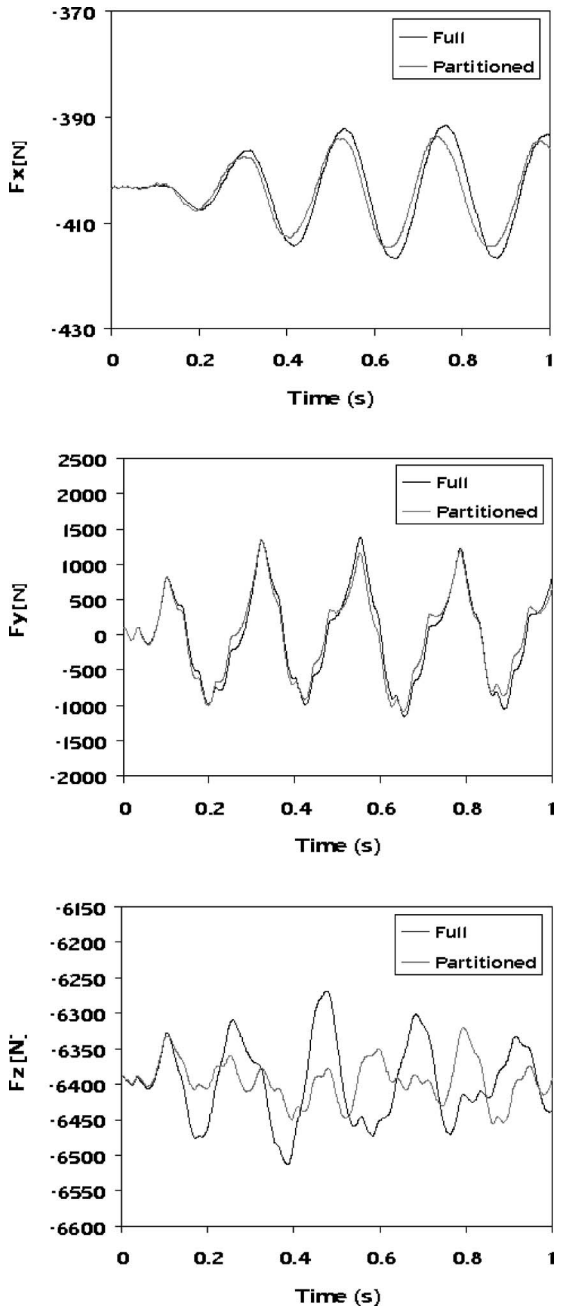


Fig. 14 Partitioned model: front mount forces with misfire

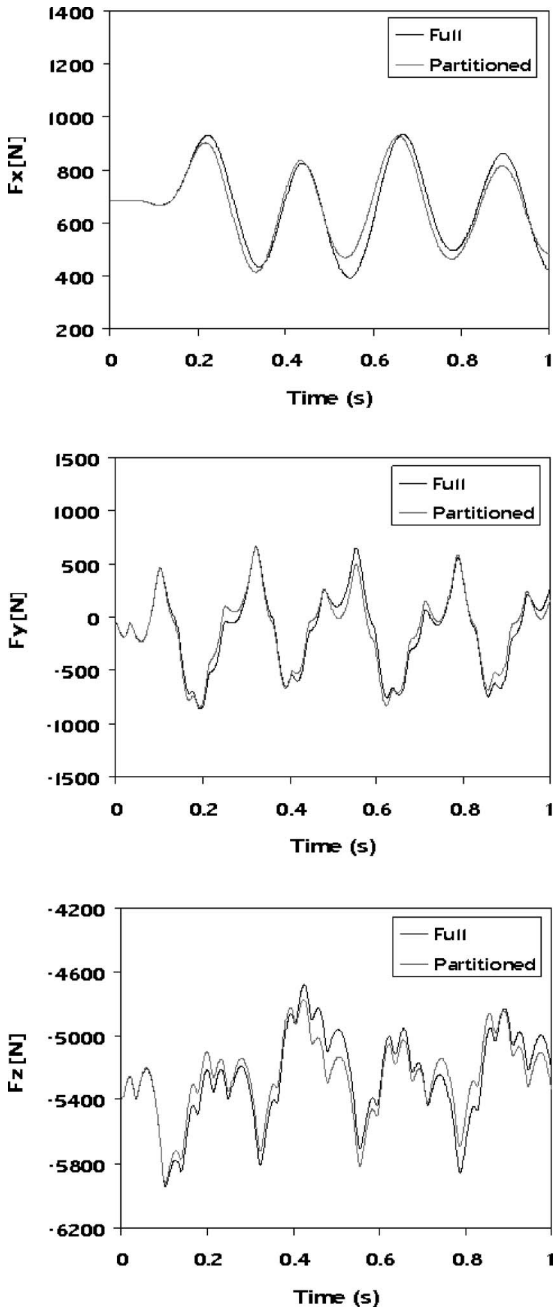


Fig. 15 Partitioned model: left-rear mount forces with misfire

$$ {}^{1}\begin{Bmatrix} v_{Cy} \\ v_{Cz} \end{Bmatrix} = {}^{1}\begin{bmatrix} -r_{Cz} \\ r_{Cy} \end{bmatrix}{}^{1}\omega_{1x} \tag{10} $$

Driven,

$$ {}^{1}v_{Cx} = {}^{1}v_{1x} + {}^{1}[r_{Cz} \; -r_{Cy}] \begin{Bmatrix} \omega_{1y} \\ \omega_{1z} \end{Bmatrix} \tag{11} $$

Note again the elimination of the crankshaft translational velocity components in Eq. (10). Introduction of the misfire requires that crankshaft translation terms be reinstated to predict the crank pin absolute velocities. Equation (10) expands to

$$ {}^{1}\begin{Bmatrix} v_{Cy} \\ v_{Cz} \end{Bmatrix} = {}^{1}\begin{Bmatrix} v_{1y} \\ v_{1z} \end{Bmatrix} + {}^{1}\begin{bmatrix} -r_{Cz} \\ r_{Cy} \end{bmatrix}{}^{1}\omega_{1x} \tag{12} $$

as the relative activities of the translation terms increase from an average of 0.21% to 1.16%.

Local activity also shows that in the presence of misfire, the motion of the piston relative to the block ($\dot{\mathbf{d}}$ in the local z direction) becomes coupled to off-axis velocity components to a greater extent in Cylinder 1 than in other cylinders. The front mount is most affected given its proximity to Cylinder 1.

Note that *conditioning* the misfire model (i.e., removing unnecessary constraint equation terms without trying to separate the model into separate, decoupled partitions) using the original 0.3%

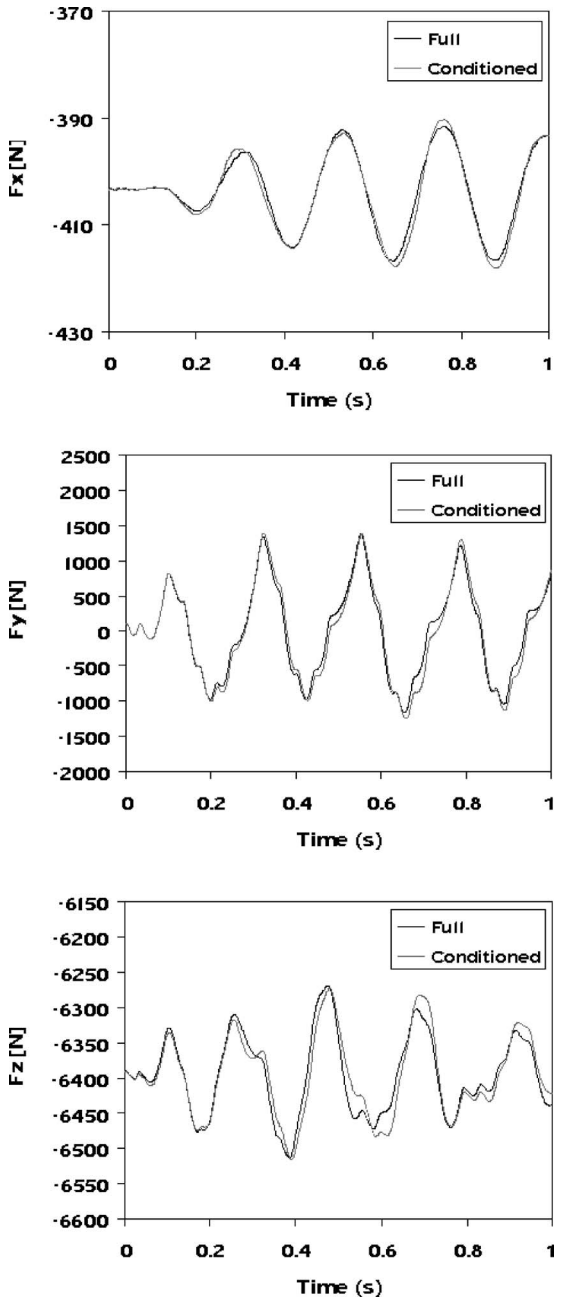


Fig. 16 Conditioned misfire model: front mount forces

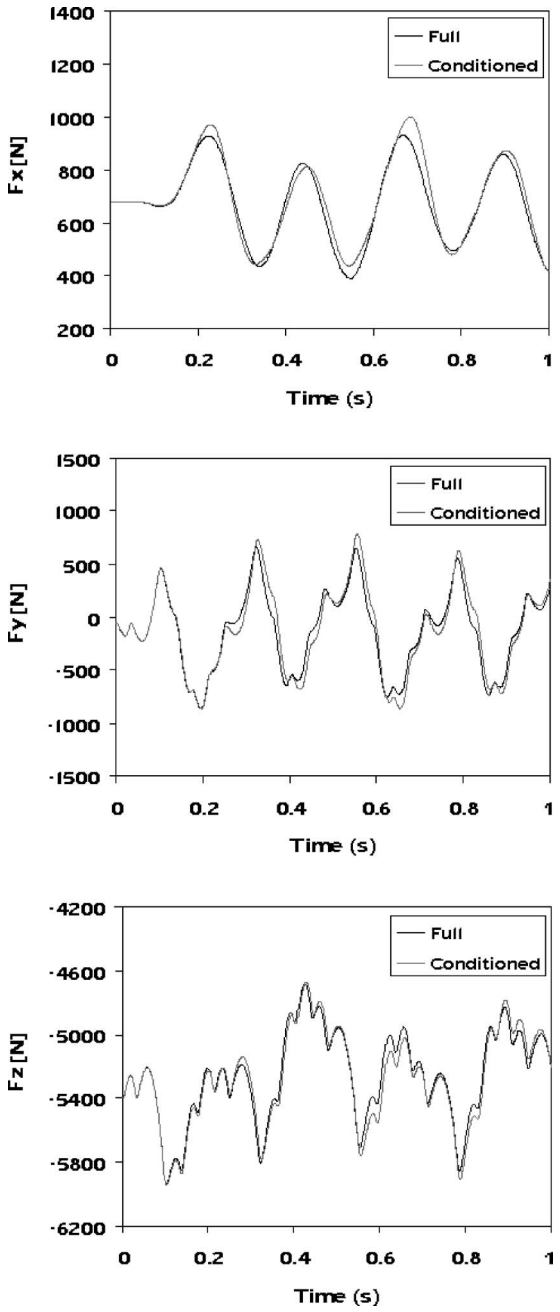


Fig. 17 Conditioned misfire model: left-rear mount forces

threshold can improve predictions of the unbalanced mount forces. Obviously, fewer bonds can be conditioned than in the balanced model, and complete partitioning is not possible. Figures 16 and 17 show the front- and left-rear mount force predictions for the misfire engine, using a model conditioned with a 0.3% threshold. Note the improvement over Figs. 14 and 15. While two-way coupling exists, many sites of local coupling are removed by elimination of negligible constraint terms. Model size is reduced and computational savings result, as described in Sec. 4.3

4.3 Model Size and Computation Time Comparison. Table 6 compares the model size and number of computation steps for the full, conditioned, and partitioned models of the balanced engine. Figure 18 shows the relative computation times. For the partitioned model, the individual times for sequential simulation of the driving and driven partitions sum to approximately one-half the computation time of the FCM. The small computation time for the driving partition is highlighted in the figure. The Vode–Adams absolute and relative integration tolerances were 10^{-6} for all runs.

Table 6 Model size comparison, balanced engine

Model	Equations	Variables	States	Vode–Adams integration steps
Fully coupled	5682	6501	279	989,618
Conditioned	4881	5863	294	908,492
Partitioned driving	1688	1999	77	60,973
Partitioned driven	3452	4111	197	854,084

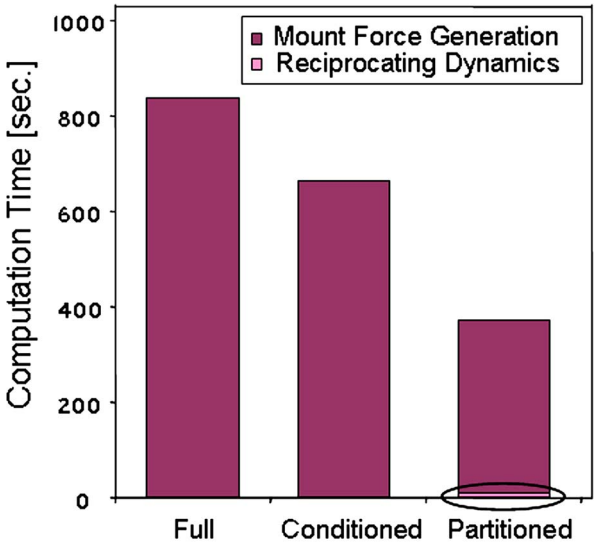


Fig. 18 Computation time, balanced engine

The times include the effort required to write the driving partition outputs to a file every 0.0001 s, and the mount forces to a data file every 0.001 s. Computation steps and time are for 1 s of simulated engine motion.

Identification of partitions reduces the size of individual simulation models, which will typically bring computational gains as shown. Less obvious, but potentially very useful, is the fact that simply conditioning the model can break algebraic loops, reduce state dependencies, and improve computation time even without partition identification and elimination. The misfire mount forces of Figs. 16 and 17, generated with a conditioned model, reduced computation time by 39% compared to a FCM with misfire.

5 Discussion

The primary contention of this paper is that a recently developed partitioning algorithm can be used in the context of engine mount force prediction to yield, systematically, insight about the relative importance of different physical phenomena contained in a conventional vehicle engine model. It has been shown, based on a 0.3% relative activity threshold, that a decoupled model for a balanced engine predicts similar engine forces as the traditional OWCM in Ref. [4]. Though the systematically generated model is slightly more complicated than the traditional OWCM, it appears to be more robust to imbalance due to the larger number of inputs to the driven subsystem. For example, providing piston displacement to the driven partition allows forces on the cylinder walls to be calculated more accurately than in a nonrunning engine model with pistons constrained to their initial positions.

The existence of partitions depends on the arbitrary threshold value used. Loosening this tolerance (making the threshold numerically higher) would necessarily result in a more completely partitioned model. Setting a local activity threshold a priori can be difficult, meaning that some iteration may be necessary to find an acceptable trade-off between activity threshold and accuracy of the partitioned model. Unfortunately, an analytical determination of this threshold is unknown at this time. Future work will combine the conditioning algorithm with a quantitative model accuracy algorithm [15], with the goal of threshold setting based on tolerances on points of engineering relevance such as overshoot and rise time that can be identified from the full model output plots.

A major point of this paper is that the algorithms allow the decoupling to be easily rechecked for whatever scenario is of particular interest to the modeler. Given any set of engine parameters and inputs of interest, one can simply rerun the model and recalculate the bond relative activities. Had the engine been connected to a rolling vehicle model through a gearbox with internal friction, instead of to an idling vehicle or test stand, the engine speed variations of Figs. 8 and 13 would be lessened. Reduction of engine speed variation and filtering out of the effects of staggered combustion events would likely reduce coupling between reciprocating dynamics and block motion even further. Running the engine at higher speeds, with more severe combustion events, while filtering out engine speed variations with a large effective crankshaft inertia and damping, may create higher local activities in some of the partition boundary bonds. As in the misfire scenario, continued existence of partitions, and locations where decoupling breaks down, can be easily checked.

Computational savings are reported in the case study. Computational savings become especially significant for design processes in which an optimization algorithm may execute the model thousands of times. Given that as the engine parameters and running condition change, the partition boundary bonds may become active; the user may wish to recheck a reduced model periodically by resimulating the full system. Aggregate computation time can still be significant even if an occasional run must be made with a model that is needlessly complex, but is used to adjust the reduced design model.

The physical insight gained by the proposed easy-to-use systematic partitioning algorithm is also a valuable attribute of the technique. For engine vibration, the specific constraint equation terms related to partitionability, or lack thereof, can be identified. A spectrum of models of intermediate complexity is suggested by "reinstated" power bonds as imbalance and/or speed increase, and the modeler can get a physical sense of where decoupling erodes. In the traditional OWCM which has an a priori assumed form, the jump from full to reduced model is made with nothing in between. Intermediate models which exhibit some two-way coupling, but which have many conditioned bonds (eliminated constraint terms), are shown in the case study to be useful and efficient.

Future engine vibration case studies can now be done with the partitioning method, for example, a study of the effect of vehicle motion on the coupling between engine reciprocating dynamics and block motion. The authors in Ref. [4] note the potential for vehicle pitch motion to cause gyroscopic coupling in engines with heavy crankshafts that are aligned with the vehicle's longitudinal axis.

The algorithm was developed using the bond graph modeling language. While facilitating the calculation and comparison of activity, elimination of negligible terms, reformulation of the model, and identification of partitions, the use of graphical modeling formalisms is not a theoretical prerequisite for applying the algorithm.

For the large model in question, equation formulation by the 20SIM processor was easier when parasitic elements (stiff springs) were used to create explicit ordinary differential equations. Accuracy of the model with parasitic elements was ensured by tuning the stiffnesses so that no joint constraint violation exceeded 0.025 mm. As reported in Ref. [16], parasitic springs can be tuned by suppressing their relative activity below a tight threshold so that the elements do not participate noticeably in the system dynamics. Bonds to these elements would obviously not be subject to conditioning. Individually tuning the elements for maximum compliance will minimize numerical stiffness and help reduce computation time. Note, however, that the relative activity partitioning method can just as easily be used (and has been used) for models with ideal joint constraints that yield a set of differential-algebraic equations.

6 Summary and Conclusions

In response to the ongoing need to systematically generate accurate models with minimal complexity for simulation-based engineering design of engines and other vehicle systems, an energy-based partitioning algorithm has been applied to the problem of determining whether or not reciprocating dynamics can be decoupled from engine block motion in predicting mount forces.

An in-line six-cylinder Diesel engine model was constructed, first with perfect balance and then with a simulated misfire. Activity, an aggregate measure of power flow, was used to identify negligible constraint equation terms within the model. The bond graph formalism facilitated the removal of negligible terms, the subsequent creation of local one-way coupling sites, and the search for separate partitions between which one-way information flow occurred. For the balanced engine, a partitioned model was generated that was able to predict mount forces accurately and with significant computational savings. The partitioned model was qualitatively similar to traditional ad hoc OWCMs; however, its varying inertia made it more robust for mount force predictions in the face of imbalance. Using the partitioned model to predict unbalanced engine mount forces gave better results than the ad hoc models. The algorithm suggested that partitions no longer existed based on the original constraint activity threshold, and indeed some mount force component predictions were significantly degraded. The algorithm pinpointed specific locations within the equation model where decoupling broke down. Running the algorithm after system inputs were changed automatically generated a model of intermediate complexity compared to the fully coupled and partitioned models. The intermediate model was more computationally efficient than the FCM despite the fact that partitions were not created.

The partition search method described herein facilitates monitoring of decoupling strength throughout the design process as parameters and inputs change. Physical insight into the system dynamics is heightened, and computation time can be reduced even when system dynamics cannot be entirely decoupled.

Appendix: Model Parameters

Table 7 Mass properties

Body	Mass (kg)	Inertia (kg m^2) I_{XX}, I_{YY}, I_{ZZ}
Engine block	1207	122.3, 238.6, 233.1
Crankshaft	300	4.23, 70.38, 70.38
Connecting rod	6.32	0.0888, 0.0865, 0.0070
Piston	5.08	0.0120, 0.0120, 0.0095

Table 8 Mount parameters

Mount	Stiffness (kN/m) K_x, K_y, K_z	Damping (N s/m) R_x, R_y, R_z	Location rel. to o (m) $[x, y, z]$
Front	193.1, 193.1, 1410	2340, 2340, 6340	$[-0.318, 0, -0.197]$
Left Rear	314.5, 314.5, 157.0	2990, 2990, 2110	$[1.000, -0.318, 0]$
Right Rear	314.5, 314.5, 157.0	2990, 2990, 2110	[1.000, 0.318, 0]

Table 9 Geometric parameters

Body	Point	Location rel to o (m) $[x, y, z]$
Block	A (main bearing)	$[-0.0830, 0, 0]$
	Center of mass	$[0.4430, -0.0088, 0.2830]$
Crankshaft	A (main bearing)	$[-0.0830, 0, 0]$
	Center of mass	[0.4920, 0, 0]
	C_1	[0, 0, 0.0800]
	C_2	$[0.1651, 0.0693, -0.0400]$
	C_3	$[0.3302, -0.0693, -0.0400]$
	C_4	$[0.4953, -0.0693, -0.0400]$
	C_5	$[0.6604, 0.0693, -0.0400]$
	C_6	[0.8255, 0, 0.0800]
Connecting rod i	C_i (rel. to c.g.)	$[0, 0, -0.111]$
	D_i (rel. to c.g.)	[0, 0, 0.158]
Piston i	D_i (rel. to c.g.)	$[0, 0, -0.050]$

References

[1] Norling, R. L., 1978, "Continuous Time Simulation of Forces and Motion Within an Automotive Engine," Society of Automotive Engineers SAE Paper No. 780665, Warrendale, PA.

[2] Suh, C.-H., and Smith, C. G., 1997, "Dynamic Simulation of Engine-Mount Systems," Society of Automotive Engineers SAE Paper No. 971940, Warrendale, PA.

[3] Shaio, Y.-J., Pan, C.-H., and Moskwa, J. J., 1994, "Advanced Dynamic Spark Ignition Engine Modeling for Diagnostics and Control," Int. J. Veh. Des., **15**, pp. 578–596.

[4] Hoffman, D. M. W., and Dowling, D. R., 2001, "Fully Coupled Rigid Internal Combustion Engine Dynamics and Vibration—Part I: Model Development," ASME J. Eng. Gas Turbines Power, **123**, pp. 677–684.

[5] Stein, J. L., and Wilson, B. H., 1995, "An Algorithm for Obtaining Proper Models of Distributed and Discrete Systems," ASME J. Dyn. Syst., Meas., Control, **117**(4), pp. 534–540.

[6] Rideout, D. G., Stein, J. L., and Louca, L. S., 2004, "System Partitioning and Physical-Domain Model Reduction Through Assessment of Bond Graph Junction Structure," *Proceedings of the IMAACA'04*, International Mediterranean Modeling Multiconference, Genoa, Italy.

[7] Rideout, D. G., Stein, J. L., and Louca, L. S., 2005, "System Partitioning and Improved Bond Graph Model Reduction Using Junction Structure Power Flow," *Proceedings of the International Conference on Bond Graph Modeling ICBGM'05*, New Orleans, LA, Society for Computer Simulation, San Diego, CA, pp. 43–50.

[8] Louca, L. S., Stein, J. L., Hulbert, G. M., and Sprague, J., 1997, "Proper Model Generation: An Energy-Based Methodology," *Proceedings of the International Conference on Bond Graph Modeling ICGBM'97*, Phoenix, AZ, Society for Computer Simulation, San Diego, CA.

[9] Karnopp, D. C., Margolis, D. L., and Rosenberg, R. C., 1990, *System Dynamics: A Unified Approach*, Wiley, New York.

[10] Tiernego, M. J., and Bos, A. M., 1985, "Modelling the Dynamics and Kinematics of Mechanical Systems With Multibond Graphs," J. Franklin Inst., **319**, pp. 37–50.

[11] Rideout, D. G., 2004, "System Partitioning and Physical-Domain Proper Modeling Through Assessment of Power-Conserving Model Structure," Ph.D. thesis, University of Michigan, Ann Arbor, MI.

[12] Hoffman, D. M. W., 1999, "In-Line Internal Combustion Engine Dynamics and Vibration," Ph.D. thesis, University of Michigan, Ann Arbor, MI.

[13] Karnopp, D. C., and Margolis, D. L., 1979, "Analysis and Simulation of Planar Mechanism Systems Using Bond Graphs," ASME J. Mech. Des., **101**, pp. 187–191.

[14] 20SIM, Version 3.5, 2005, Controllab Products b. v., Enschede, Netherlands.

[15] Sendur, P., Stein, J. L., Peng, H., and Louca, L. S., 2002, "A Model Accuracy and Validation Algorithm," *Proceedings of the 2002 ASME International Mechanical Engineering Conference and Exhibition*, New Orleans, LA, American Society of Mechanical Engineers, New York, NY.

[16] Rideout, D. G., and Stein, J. L., 2003, "An Energy-Based Approach to Parameterizing Parasitic Elements for Eliminating Derivative Causality," *Proceedings of the International Conference on Bond Graph Modeling ICBGM'03*, Orlando, FL, Society for Computer Simulation, San Diego, CA, pp. 121–127.

High-Temperature Air Combustion Phenomena and Its Thermodynamics

Nabil Rafidi
e-mail: rafidi@kth.se

Wlodzimierz Blasiak
e-mail: blasiak@mse.kth.se

KTH,
Royal Institute of Technology,
10044 Stockholm, Sweden

Ashwani K. Gupta
Department of Mechanical Engineering,
University of Maryland,
College Park, MD 20742
e-mail: akgupta@eng.umd.edu

The fundamentals and thermodynamic analysis of high-temperature air combustion (HiTAC) technology is presented. The HiTAC is characterized by high temperature of combustion air having low oxygen concentration. This study provides a theoretical analysis of HiTAC process from the thermodynamic point of view. The results demonstrate the possibilities of reducing thermodynamic irreversibility of combustion by considering an oxygen-deficient combustion process that utilizes both gas and heat recirculations. HiTAC conditions reduce irreversibility. Furthermore, combustion with the use of oxygen (in place of air) is also analyzed. The results showed that a system, which utilizes oxygen as an oxidizer, results in higher first and second law efficiencies as compared to the case with air as the oxidizer. The entropy generation for an adiabatic combustion process is reduced by more than 60% due to the effect of either preheating or oxygen enrichment. This study is aimed at providing technical guidance to further improve efficiency of a combustion process, which shows very small temperature increases due to mild chemical reactions. [DOI: 10.1115/1.2795757]

1 Introduction

High temperature air combustion (HiTAC) technology is now well recognized as the most advanced combustion technique for the combustion of fossil or derived fuels with simultaneous achievements of significant energy savings (about 25%) and pollution reduction (about 25%) [1,2]. Flameless oxidation of fuels has also been observed under HiTAC conditions. HiTAC is particularly attractive for processes that require uniform thermal field, higher and uniform heat flux distribution in the combustion zone, and compact size of the equipment [3,4]. Controlled flame behavior results in uniform thermal field in the entire flame zone [2] (less than about 50 K), significantly reduced pollutants emission (about 25%) and combustion noise [5] (more than 7 dB), and compact size of the furnace or increased material throughput for same furnace size (about 25%) [1–4]. Much of the experience with HiTAC has been on combustors for use in furnaces and boilers [1–15]. The flow dynamics is much different under HiTAC conditions as compared to normal air case or normal air preheated to high temperatures [5]. Recently, high temperatures have also been used for the gasification of wastes [6,7]. Some of the benefits of HiTAC are given below.

2 Benefits of Combustion With High-Temperature Air

2.1 Energy Savings. In all combustion systems, significant loss occurs during the discharge of waste gases to the environment. In the HiTAC technology, thermal energy from the waste gases in the combustion products is extracted using regenerative type heat exchangers. Ceramic honeycomb or balls have been demonstrated to provide thermal energy storage media in heat exchangers. Honeycomb type heat exchanger is more effective than ceramic ball type because of the larger surface area, low-pressure drop, and high efficiency. Exhaust gases from industrial furnaces and processes represent one of the major energy losses from the system. The regenerative media used in the HiTAC devices (both ceramic honeycomb and ball type) recover much of the thermal energy from the exhaust gases and transport it back to the combustion zone. The amount of recovered energy translates directly into fuel savings and reduction of CO_2 emission to the environment. The goal of "High Performance Industrial Furnace Development" project [1] was to simultaneously demonstrate energy savings (about 30%) using regenerative combustion, reduce pollutant's emission, including CO_2, (about 30%), and reduce equipment size (about 25%). Fuel energy savings in the range of 10–60% have been successfully demonstrated in different types of furnaces along with uniform thermal field and pollution reduction. Field tests on 167 furnaces demonstrated significant fuel savings and the real benefits of HiTAC [1,2].

2.2 CO_2 Reduction. The role of CO_2 emission to the environment from the combustion of carbon based fuels is of great concern as increased levels of CO_2 result in global warming. The demands for reducing CO_2 emission are higher than ever before as we continue to use more energy per capita worldwide and increased population. All hydrocarbon fuels contain carbon, which generate CO_2 as by-product during the combustion of fuels, so that a reduction in energy consumption directly translates to reduction of CO_2 emission. The CO_2 reduction with HiTAC technology is estimated to be about 25% [1].

2.3 NO_x Reduction. Emission of NO_x is now known to be responsible for the destruction of ozone layer in the upper atmosphere. NO_x (NO, NO_2, N_2O, N_2O_4, N_2O_5, etc.) involves the complicated reaction mechanisms, which results in accelerating the ozone depletion in the oxygen cycle on earth. Therefore, combustion engineers have focused their attention to develop various strategies to reduce NO_x emission and improve the combustion process. HiTAC is one of the most advanced techniques because of low levels of NO_x formation and emission, significant energy savings, and uniform thermal properties in the entire combustion space [1–5,8–10]. The reason for low NO_x is due to uniform and low overall temperatures. The thermal field uniformity has been demonstrated to be much less than 50 K in the entire combustion zone using HiTAC technology so that HiTAC combustion can also be called isothermal combustion process. Under certain conditions, the flame has been shown to be colorless. This mode of combustion is called flameless oxidation of fuel.

2.4 Reduction of Equipment Size. Higher and far uniform

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received September 30, 2005; final manuscript received July 15, 2006; published online January 22, 2008. Review conducted by Kenneth Mark Bryden.

thermal field and heat flux distribution is obtained with HiTAC technology as compared to any other known method for combustion, such as recuperative combustion or combustion with room temperature or moderate temperature air [1,2,8,9]. The heat flux can be as high as 350–400 kW/m^2 as compared to the ordinary combustion with heat flux of the order of 75–125 kW/m^2. This means that the application of HiTAC can essentially eliminate the convective section of the boiler that is of much larger in size as compared to the radiative section. The ignition delay time with HiTAC flames is higher with high-temperature and low oxygen concentration combustion as compared to that obtained using high-temperature air [10]. However, the combustion in HiTAC condition is analogous to distributed combustion so that the combustion occurs in the entire volume of the combustion chamber with the result that the actual size of the combustion chamber is much smaller. This means that industrial combustion equipment has significantly reduced size to result in materials conservation and better product quality from the uniform thermal field with HiTAC [1]. Simultaneous realization of the above benefits was considered impossible before. The unique flame features associated with HiTAC assist in utilizing this technology for wider range of applications than initially envisioned [1].

3 Basic Principle of High Temperature Air Combustion Technology

3.1 Thermodynamic Consideration of Combustion Process. Flame temperature is one of the important factors for considering combustion efficiency and energy conversion efficiency. Weinberg [11] provided initial concept of excess enthalpy combustion for the combustion of low heating value fuels. In his study, limitations on combustion temperature were discussed, including both positive and negative factors associated with combustion temperatures in a certain range. Heat circulation from the exhaust gases, using high-efficiency heat exchanger, was adopted in order to increase the combustion temperature thus allowing the combustion of low heating value fuels that could otherwise not be combusted without using auxiliary fuel. However, combustion engineers have to pay attention to the upper limit of combustion temperature because of materials constraints used in the equipment and/or pollution formation at higher temperatures, in particular, NO_x. From the economic point of view, it is of course better to use thermal energy generated by combustion process itself to heat up the oxidant or fuel, which via other means. The amount of combustion-generated energy circulated into the combustion process is given as [1]

$$\int_{T_0}^{T_t} C_p dT = Q_c + Q_a = H_t - H_0 \tag{1}$$

where T_f is the final temperature, T_0 is the initial temperature, Q_c is the energy released by chemical energy conversion, Q_a is the energy added, H_t and H_0 are the enthalpy at two states. The circulation part of thermal energy from combustion-generated products will increase the combustion temperature so that enthalpy of the reaction zone will be above the conventional combustion level. This has resulted in the use of term called "Excess Enthalpy Combustion."

Increase in thermodynamic efficiency must be coupled with other desirable characteristics, such as low NO_x formation, reliability of the equipment and refractory material, and uniform thermal field in the combustion zone. Nonetheless, the heat circulation and excess enthalpy methods provide new light on next generation of advanced energy conversion technology and combustion chamber design. The method provides new ideas to control temperature in the combustion zone. The method is independent of the fuel composition and can simultaneously satisfy the demands of high combustion intensity and reduced pollutant formation from fuels, including low-grade fuels.

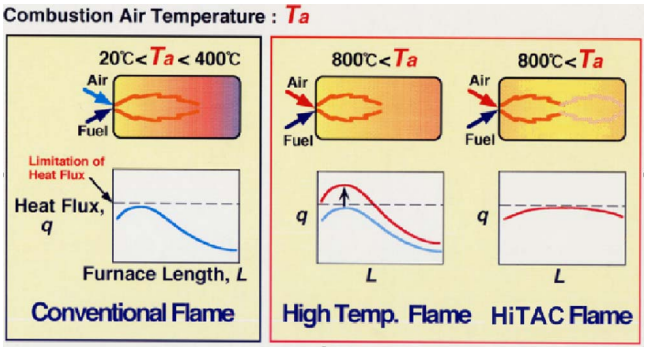


Fig. 1 A schematic diagram of flame and heat flux distribution in a furnace with low-temperature combustion air (left), high-temperature air (middle), and high-temperature and low oxygen concentration combustion air (HiTAC) condition (right)

3.2 Development of Excess Enthalpy Combustion to High-Temperature Air Combustion. Air preheats add enthalpy to the combustion zone. From the point of pollutant formation, especially NO_x, very high combustion temperatures result in high levels of NO_x emission. A method to maintain the same temperature in the combustion zone is to dilute the incoming combustion air with hot combustion products to result in low oxygen concentration combustion air. Partial recirculation of hot products into the incoming combustion air dilutes the inlet concentration of oxygen in the air. This diluted air will lower the peak flame temperatures and also lower the oxidation of N_2. Further examples of this are given in Refs. [1,2,12].

In the North American design, ceramic balls have usually been used in the regenerators to preheat the combustion air with hot gases exiting the furnace, while in Japan, both ceramic balls and ceramic honeycomb regenerators have been used. The use of ceramic balls in a packed bed provided higher combustion air temperatures; however, the regenerator efficiency is not so high as compared to honeycomb type regenerator due to very high heat exchange surface area. Nonetheless, the air preheat temperatures were much higher than those achieved previously with the use of recuperators. The honeycomb regenerator is much more compact than a bed packed with ceramic balls, has high specific surface area and low thermal inertia, and provides very low-pressure drop [2]. In HiTAC, the increase of temperature of the combustion gases in the furnace is small (only about 50–100°C above the incoming HiTAC). The oxygen concentration in the combustion air is very low (only about 2–5% by volume, depending on the application) to result in flameless oxidation (or colorless flame). Under these conditions, the thermal field in the combustion zone is very uniform [1,8,12]. The heat flux from the flame with high-temperature combustion air is also very high [1,2,12]. However, the flame fluctuations are very low and thermal field uniformity is very high in HiTAC flames as compared to the normal flames [13–16]. A schematic diagram of flame and heat flux distribution with normal and high-temperature combustion air is shown in Fig. 1.

3.3 Flame Stability and NO_x Emission Characteristics. The flame characteristics of propane fuel and high-temperature combustion air have been examined in a test furnace facility using several advanced diagnostics. The flame stability limits as a function of air-preheat temperature and oxygen concentration have been found to be extremely wide [1]. The flame stability limits increase significantly at high air preheats. The high temperature is referred to as the temperature above the autoignition temperature of the fuel. It is to be noted that very wide flame stability limit occurs even with low oxygen concentration air. Under HiTAC conditions, the flame stability is essentially infinite. The results also suggest that it is possible to use exhaust gases from a neigh-

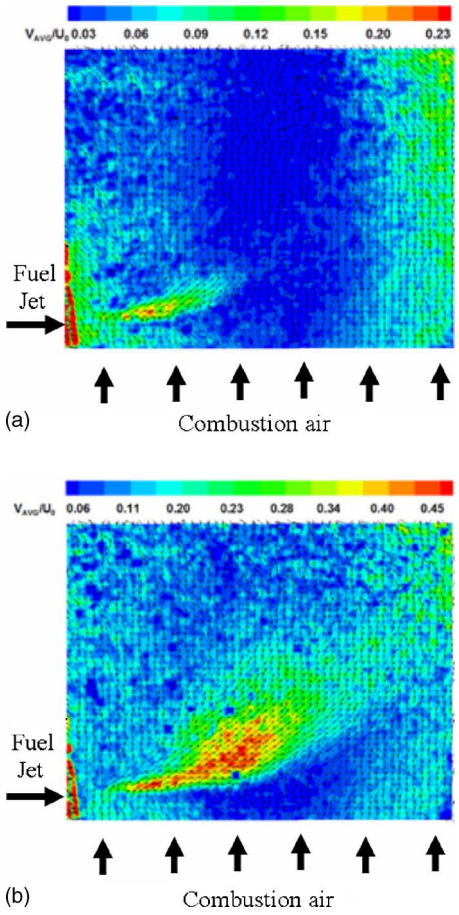


Fig. 2 Fuel jet impingement into a cross flow of normal temperature air (left) and high-temperature air (right)

boring furnace as oxidant without further dilution or gas recirculation since these gases often contain several percent of oxygen. Very low NO_x emission levels have been obtained under high-temperature and low oxygen concentration combustion conditions. NO_x emission at air-preheat temperature of 1150°C decreased from 2800 ppm at 21% O_2 to 40 ppm at 2% O_2.

3.4 Flow Dynamics of High-Temperature Air Combustion Flames. Figures 2(*a*) and 2(*b*) show the mean flow velocity and direction for the nonburning and burning cases, respectively, using propane as the fuel [5]. The fuel jet causes strong impingement into the cross flow for the high-temperature air case with the same momentum between nonburning and burning cases. For the normal air temperature case, the jet deflects gradually into the cross flow with a smoother trajectory. However, for the high-temperature case, the jet has a much larger velocity distribution in a small well defined region before its deflection. This suggests greater dependency of the combustion air temperature and properties on the jet mixing and to provide a large influence on the jet impingement. Further downstream, the velocity distribution is smaller where most of the combustion occurs. The results also showed that combustion decreases the turbulence in regions where the jet deflects into the cross flow. Overall, the turbulence levels for the high-temperature case were low as compared to the normal air case. A very small local region of higher turbulence was observed in the zone of high vorticity and axial strain. This region provides local mixing of the fuel jet with the cross flow.

4 Analysis of High-Temperature Air Combustion

Further innovation of this combustion phenomena have been continued with pure oxygen in place of air as the oxidizer [17]. This is referred to as flameless oxy-fuel combustion. To reach the same level of low oxygen concentration, the gas recirculation rate when mixing oxygen and combustion gases is larger compared to that when mixing air and combustion gases. The gas recirculation rate R is defined as the ratio of the mass flow rate of the recirculated combustion gases to that of the oxidizer. The extra advantages of using oxygen are near zero NO_x emission, a favorable furnace atmosphere for more gas radiation and further reduced furnace size. In addition, preheating of air is not required because the high recirculation ratio is capable to guarantee the high temperature required for the stable combustion. As a result, a simpler burner without regenerator or recuperator can be used to achieve the oxygen-deficient combustion using oxygen as an oxidizer.

In the literature, many terms have been used to refer to this technology, see, for example, Cavaliere et al. [18]. They refer to it as mild combustion because the temperature increase due to the reaction is lower or milder than that of the preheated reactants. However, the most common term used by many researchers including us is HiTAC [1]. Flameless combustion [19], high-temperature combustion technology (HiCOT), and excess enthalpy combustion are few other examples. However, the common thing in all of the above cases is the *low oxygen concentration* of the reactants. Since we consider here different cases of temperatures, oxidizers, and gas recirculation rate, "oxygen-deficient combustion" (ODC) will be used hereinafter.

Most thermodynamic analysis interest in ODC has focused on temperature issues to ensure reduced peak temperatures and temperature variations. Very limited attention has been given on the global thermodynamics or exergy analysis of ODC. The reason is probably because this technology is relatively new and most of the research was focused on exploitation of the technology in industrial heating in which second law analysis is used only sparingly [20]. In contrast, there is always an interest in improving the availability of fuel energy after combustion to provide useful work. In fact, approximately 1/3 of the fuel chemical energy becomes unavailable for a typical atmospheric combustion [21,22]. Most of this inherent thermodynamic irreversibility is associated with internal heat transfer between products and reactants.

Detailed thermodynamic analysis is provided in the HiTAC book by Tsuji et al. [1]. They show that second law efficiency increases due to preheating and is higher in the case of isothermal combustion than in the case of adiabatic combustion. However, they did not study the effect of gas recirculation. Instead, they did analysis at different equivalence rations ϕ from 0.1 to 1. In fact, ODC can be best represented by R or the reactant oxygen level and not by ϕ. Daw et al. [23] showed the possibilities of reducing thermodynamic irreversibility of combustion by considering a conceptual isobaric combustion process that utilizes controlled preheating to promote near equilibrium combustion. However, that was only conceptual and not very practical but helps one to illustrate the trend for less irreversible combustion.

In the present study, global first and second law analyses were used on ODC to examine the possibilities for reducing irreversibility. Exergy and energy flows and losses are evaluated at every stream and stage in the combustion processes. A comparison is made between a conventional and ODC using either air or oxygen.

5 Thermodynamic Aspects of Oxygen-Deficient Combustion

The ODC is a system comprising three components (see areas under dashed line in Fig. 3) performing the processes of heat recuperating of combustion product by the oxidizer, mixing of preheated oxidizer and combustion products to yield a low oxygen concentration reactant, and finally the combustion of these reactants with the fuel. Although the low oxygen concentration is achieved in practice by internal flue gas recirculation, the recirculation and mixing process is hypothesized in a separate mixer

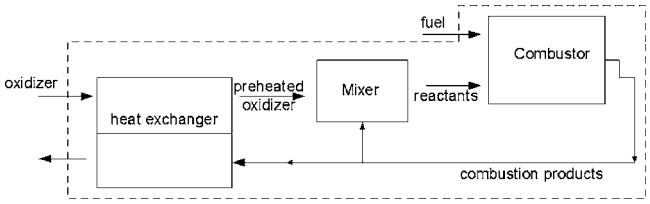


Fig. 3 The oxygen-deficient system comprising three components

located between the heat exchanger and the combustor, as shown in Fig. 3. The following assumptions have been used for the analysis:

(i) Environment is dry air at 298.15 K and 101.325 kpa.
(ii) All processes are steady state steady flow.
(iii) Kinetic and potential energies are neglected.
(iv) The oxidizer and the fuel enter the system at 298.15 K.
(v) Heat recuperation is a counterflow heat exchanger having effectiveness of 0.9.
(vi) Dilution of oxidants is only due to gas recirculation (mixing of product gases with preheated or nonpreheated oxidizers).
(vii) The fuel is 100% propane.
(viii) Oxidizer/fuel ratio was set to correspond to 3% oxygen in the combustion product gases (for complete combustion).
(ix) Combustion processes are adiabatic and isobaric.
(x) Assumed combustion product species are O_2, CO_2, H_2O, N_2, O, H, OH, H_2, CO, and NO.
(xi) Combustion gases entering the mixer or the heat exchanger are assumed to be a product of complete combustion containing only O_2, CO_2, H2O, and N_2.
(xii) Irreversibility due to thermal interaction with environment is neglected in the mixer and the combustor. However, energy loss to the environment from a heat exchanger was accounted for based on some experimental data.
(xiii) Pressure loss is accounted for only in the heat exchanger.
(xiv) Constant pressure process in the mixer and combustor.
(xv) Thermodynamic gas properties obtained from data tables [21,22,24], and
(xvi) GASEQ code [25] used to calculate equilibrium states and temperatures.

Cases of oxidizer as air or oxygen are studied. For air-ODC system, a combined heat and gas recirculation was considered. However, only gas recirculation was considered in the case of oxygen-ODC system. Table 1 shows the examined cases for different gas recirculation rate R values (R is defined as the ratio of the mass flow rate of the recirculated combustion gases to that of the oxidizer) and the corresponding oxygen mass fraction (ω_{O_2}) of the reactants. Cases of oxygen-OCD system with a combined gas and heat recirculation will be shown later and compared with cases of only gas recirculation.

5.1 Enthalpy: Temperature Diagram of Combustion Process. Conventional thermodynamics analysis was made by constructing the H-T diagram shown in Fig. 4(a). The H_R curve is the enthalpy of the unburned reactants as a function of temperature for a mixture of fuel and air (ω_{O_2}=0.23). The other curves, underneath and parallel to it, are also the enthalpies of unburned reactants but for diluted air and fuel. The H_P curve is the enthalpy of chemically equilibrated burned product. It is independent of oxygen concentration because the products of combustion at equilibrium and at certain temperature are the same for all cases if dilution is made by flue gas recirculation as assumed in this study. Since the two curves (H_R and H_P) are converging due to the thermal dissociation and the increase in specific heat of the product, heat of combustion becomes less until it becomes zero, which corresponds to the adiabatic limit temperature T_{alf}. It is clear from the figure that T_{alf} decreases with increase in dilution. T_{alf} decreases from 3233 K to 2520 K when the oxygen mass fraction in the reactant is reduced from 23% to 5.6%. In fact, this is not the only theoretical limiting temperature. Since we are considering preheats by heat recirculation and gas recirculation, the maximum preheated temperature of the reactants remains lower than that of T_{alf}. This is because of the finite temperature difference in the heat exchanger (preheater). This maximum limiting temperature of the reactants due to mixing and preheating T_{Limp} can be determined by knowing the effectiveness of the heat exchanger η and R as follow:

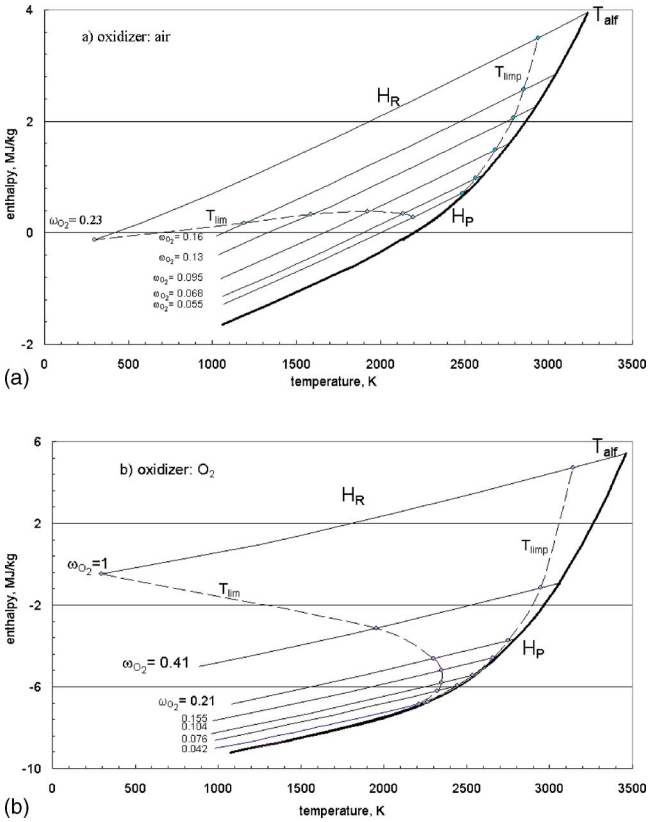


Fig. 4 *H-T* diagram for C_3H_8/air or O_2 mixture (top diagram, air; bottom diagram, O_2)

Table 1 *R* values and the corresponding oxygen mass fraction in the cases examined

Air	R	0	0.5	1	2	4	6		100
	ϖ_{O_2}	0.233	0.16	0.013	0.095	0.068	0.056		0.028
Oxygen	R	0	1.5	4	6	10	15	35	100
	ω_{O_2}	1	0.41	0.21	0.155	0.104	0.076	0.042	0.024

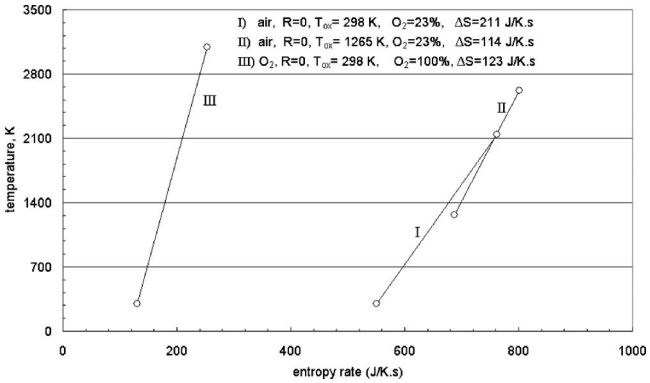


Fig. 5 Effect of preheating of the reactants and oxygen enrichment on the entropy rate generation in an adiabatic isobaric combustion process

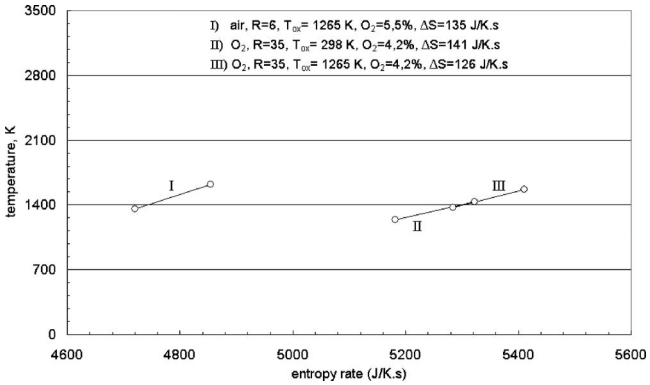


Fig. 6 Effect of dilution by flue gas recirculation on the entropy rate generation in an adiabatic isobaric combustion process

$$T_{\text{Limp}} = \frac{T_{\text{alf}}(\eta c_{p,\text{ox},o} + R c_{p,\text{alf}}) + (1-\eta) c_{p,\text{ox},o} T_{\text{ox},i}}{(1+R) c_{p,\text{Limp}}} \tag{2}$$

where the subscripts ox,i and ox,o refer to the condition of the oxidizer before and after the heat exchanger, respectively. T_{Limp} is shown in Fig. 4(*a*) as a dashed line. The intersection between H_P and a straight horizontal line drawn from T_{Limp} toward H_P is the maximum temperature of the products corresponding to T_{Limp} if combustion is adiabatic. In contrast, if the preheating is with gas recirculation without using a heat exchanger, limiting maximum temperature of the reactants will be less than T_{Limp}, which is shown by another dashed lined marked with T_{Lim} in the same figure.

$$T_{\text{Lim}} = \frac{c_{p,\text{ox}} T_{\text{ox}} + c_{p,\text{alf}} T_{\text{alf}}}{c_{p,\text{Lim}}(1+R)} \tag{3}$$

In fact, operation without heat recirculation is only realistic in practice when using oxygen as the oxidizer and with gas recirculation. Figure 4(*b*) shows the same *H-T* diagram but in the case of oxygen-ODC system.

5.2 Entropy Generation. Depending on the value of R, the ODC involves variations of mass flow rates for the same energy input of the fuel. Therefore, it is more convenient to use the absolute entropy rate $\dot{S}$ (J/K s) to compare ODC with other combustion process. By comparing Curve I with Curves II and III in the *T-S* diagram (Fig. 5), the entropy generation for an adiabatic combustion process is reduced by more than 60% due to the effect of either preheating or oxygen enrichment. These calculations are based on 200 kW energy input from propane. In the case of preheating the reactants, the temperature increase of the product gases is less than that of the reactant gases because of the increase in specific heat caused by thermal dissociation of product. As a result, less entropy generation is due to preheating. In the case of oxygen combustion, although ΔT is higher than that of air combustion, entropy generation of oxygen combustion is minimized because N_2 dilution is reduced and this reduces the heat capacity of both reactants and products [22].

Figure 6 shows the effect of the gas recirculation on entropy generation for both cases of oxygen- and air-ODC combustion systems. In both cases, R were set to 35 and 6, respectively, by which O_2 concentration is reduced to about 5%. By comparing Curves I and II from Fig. 6 with Curves II and III from Fig. 5, one can conclude that entropy generation is increased by about 14–18% due to gas recirculation. However, in the case of oxygen, preheating oxygen before mixing further will minimize ΔS (comparing Curves II and III in Fig. 5). Figures 5 and 6 are to same scale to facilitate comparison showing that gas recirculation favors maximum temperature of the products.

5.3 Irreversibility of Oxygen-Deficient Combustion Process. For and adiabatic combustion process, the irreversibility rate $\dot{I}$ can be calculated directly as

$$\dot{I} = T_o \dot{\Pi} \tag{4}$$

where the entropy generation $\dot{\Pi}$ is equivalent to $\Delta\dot{S}$ in this case. $\dot{I}$ can also be found from the exergy balance of the combustor.

$$\sum_{\text{in}} \dot{E} - \sum_{\text{out}} \dot{E} = \dot{I} \tag{5}$$

or $\dot{E}_R + E_F - E_P = \dot{I}$, where the subscripts R, F, and P refer to reactants, fuel, and products, respectively. The exergy rate of a flue gas stream of m components is given by

$$\dot{E} = \sum_j^m \dot{n}_j \varepsilon_{\text{ch},j} + \sum_j^m \dot{n}_i \varepsilon_{\text{ph},j} \tag{6}$$

The specific molar chemical and physical exergy are

$$\tilde{\varepsilon}_{\text{ch}} = \sum_j^m x_j \tilde{\varepsilon}_{\text{ch},j} + \tilde{R} T_o \sum_j^m x_j \ln(x_j) \tag{7}$$

$$\varepsilon_{\text{ph}} = (T - T_o) \sum_j^m x_j \tilde{c}_{p,j}^{\varepsilon} + R T_o \ln(P/P_o) \tag{8}$$

where P is total pressure of the mixture and $\tilde{c}_p^{\varepsilon}$ is the mean molar isobaric exergy capacity defined as

$$\tilde{c}_p^{\varepsilon} = \frac{1}{T - T_o} \left[\int_{T_o}^{T} \tilde{c}_p dT - T_o \int_{T_o}^{T} \frac{\tilde{c}_p dT}{T} \right] \tag{9}$$

Since P equal P_o, the second term of the right hand side of Eq. (8) for the isobaric combustion process is zero. It is useful to normalize $\dot{I}$. Thus, the fraction that is lost from the total exergy input $(\dot{E}_{\text{ox}} + \dot{E}_F)$ through irreversibility in the combustor is called the efficiency defect δ.

Figure 7 shows the relation between the efficiency defect due to combustion δ_{comb} and the temperature of the combustion gases products. It is clear that one can get higher second law efficiency if oxygen concentration in the reactants is reduced and/or if preheating raises the combustion products temperature. Thus, the combination between heat and gas recirculation produces least irreversibility.

Although there is no heat exchanger (oxidizer preheater) in the case of oxygen-ODC system (Fig. 7(*b*), bottom diagram) in which mixing only performs the preheating of the reactants, the efficiency defect is slightly lower in the case of oxygen-ODC system

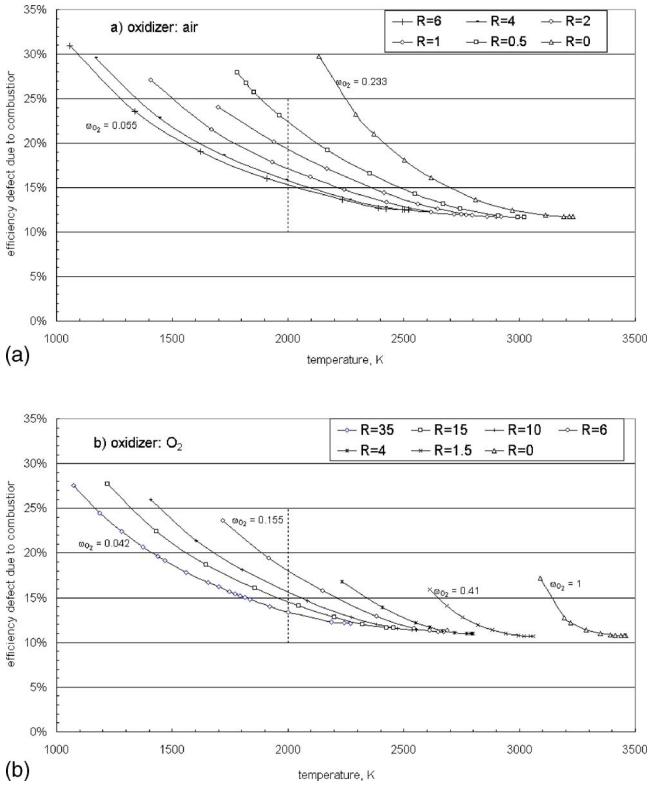


Fig. 7 Efficiency defect due to combustion as a function of equilibrated temperature of the products for various oxygen concentration levels

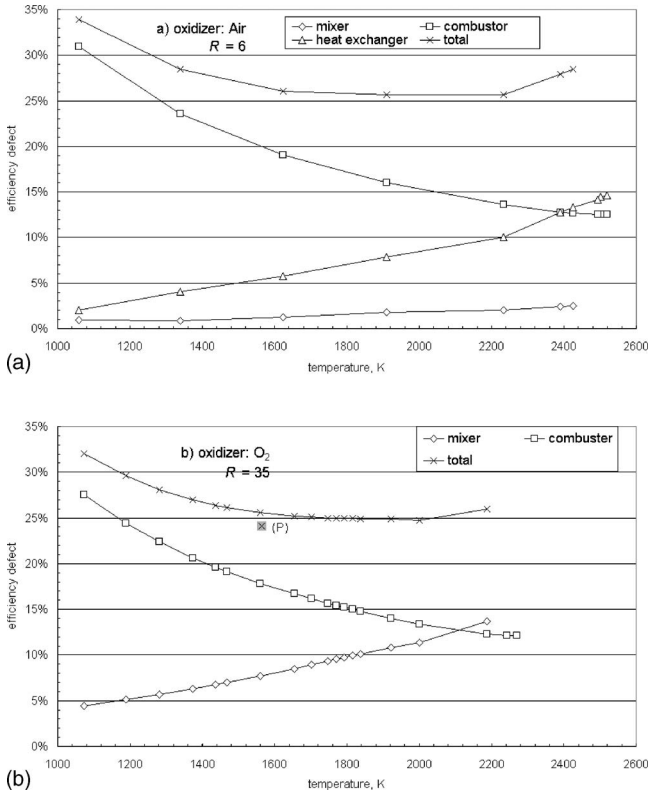


Fig. 8 The efficiency defect of every component of the ODC system as a function of equilibrated temperature of the product combustion gases

than that for air-ODC system if we keep the product temperature constant (see the dotted vertical lines at 2000 K in Fig. 7(*a*), top diagram and Fig. 7(*b*), bottom diagram).

5.4 Irreversibility of the Whole Oxygen-Deficient Combustion System. A more comprehensive global thermodynamic analysis of the oxygen-deficient system is presented in this section and includes the other two components (the mixer and the heat exchanger). The irreversibilities from the components added to reduce irreversibility of the ODC may or may not exceed that of an ordinary combustion. In order to examine this, $\dot{I}$ for both mixing and heat exchange processes were calculated. In the calculation of $\dot{I}$ for the heat exchanger, experimental values for pressure loss and heat loss to the environment were used and generalized for all cases. The total irreversibility is thus due to the entropy change, Eq. (10), and the thermal interaction with the environment, I^Q in Eq. (11). The later is equivalent to the exergy loss to the environment, E^Q.

$$\dot{I} = T_o\left[\dot{m}_{\text{ox}}\sum_{\text{ox}}\left(\int_{T_{\text{ax},i}}^{T_{\text{ox},o}} x_i c_{p_i}\frac{dT}{T} + R_{\text{ox}}T_o\ln\left(\frac{P_{\text{ox},i}}{P_{\text{ox},o}}\right)\right) - \dot{m}_{fg}\sum_{fg}\left(\int_{T_{fg,o}}^{T_{fg,i}} x_i c_{p_{fg}}\frac{dT}{T} + R_{fg}T_o\ln\left(\frac{P_{fg,i}}{P_{fg,o}}\right)\right)\right] \tag{10}$$

$$\dot{I}^Q = Q_s\frac{T_s - T_o}{T_s} = \dot{E}^Q \tag{11}$$

The irreversibility or the efficiency defect due to the heat exchanger can also be found from the exergy balance as

$$\delta = 1 - \frac{\dot{E}_{fg,o} - \dot{E}_{fg,i}}{\dot{E}_{\text{ox},i} - \dot{E}_{\text{ox},o}} \tag{12}$$

On the other hand, $\dot{I}$ for the mixing

$$\dot{I} = T_o[\dot{m}_{\text{ox}}c_{p_R}\ln(T_R/T_{\text{ox}}) + \dot{m}_{fg}c_{p_R}\ln(T_R/T_{fg})] \tag{13}$$

For such multicomponent system, one can calculate the second law efficiency of the whole ODC system ψ as given by

$$100\,\% = \Psi + \delta_{\text{HX}} + \delta_{\text{mix}} + \delta_{\text{comb}} \tag{14}$$

where the last three terms represent the efficiency defect due to the heat exchanger, the mixer, and the combustor, respectively.

Figure 8 below shows the particular efficiency defect of every component of the system at relatively high gas-recirculation rates. It should be noted again that the heat exchanger is absent in the oxygen-ODC system and therefore maximum temperatures are not as high as for air-ODC system. Since mixing takes place at lower temperature differences for the case of air than for oxygen case. δ_{mix}, in the oxygen case is much higher than for the air cases and it can be even higher than both efficiency defects δ_{HX} and δ_{mix} in the case of air. Therefore, the total efficiency of both systems is almost comparable. Point (P) in Fig. 8(*b*), on the other hand, represents the total efficiency defect of an ODC when both gas and heat recirculation are used in the case of oxygen-ODC system. Therefore, adding heat recirculation increases the total efficiency of the system.

Figure 9 shows a comparison between an ordinary combustion process (the first column) and oxygen deficient cases at low oxygen concentrations (Bars 2, 3, and 4). The second column is for an air-ODC system while Columns 3 and 4 are for oxygen-ODC system. In Column 3, only gas recirculation takes place, while in Column 4, both gas and heat recirculation take place. These comparisons clearly show that oxygen system is the most efficient as compared to others, especially when adding the heat exchanger. The irreversibilities due to the heat exchanger and the mixer together (Column 4) are less than that of the mixer alone (Column

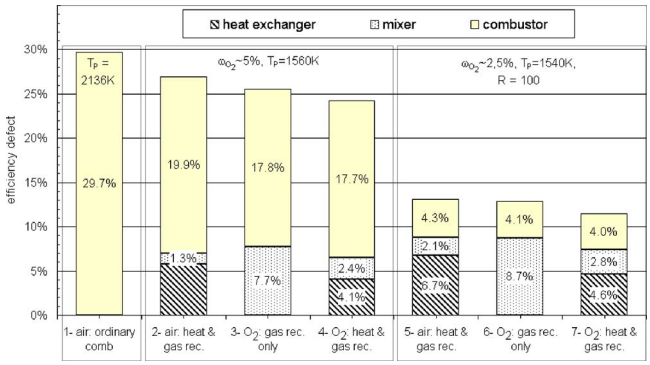


Fig. 9 Comparison between an ordinary combustion and various cases of ODC

3) when mixing performs the preheating. This additional preheating using the heat exchanger (heat recirculation) increases the reactant temperatures and thus further reduces the irreversibility of the combustion process. The overall increase of efficiency is about 8% when adding the heat exchanger to the ODC-oxygen system.

The last three columns (5, 6, and 7) in Fig. 9 are for extreme gas recirculation ($R=100$) to show the trend for high second law efficiency of the ODC process. It is clear that δ_{comb} is reduced with increase in gas and heat recirculation. Although δ_{mix} and δ_{HX} increase when ω_{O_2} of the reactants decreases, the second law efficiency of the whole ODC system increases and is higher if oxygen is used as an oxidizer. In fact, if the same trend is followed, not only the second law efficiency increases but also the first law efficiency increases. Flue gas energy loss of oxygen-ODC system is less than 3% of the total energy in the fuel if recuperator effectiveness of 0.9 is used to preheat oxygen (the cases of Columns 4 and 7). This loss is ~10% for an air-OCD system (Columns 2 and 5).

6 Conclusions

HiTAC has proven benefits of significant energy savings, reduced pollution, and decrease in size of the equipment and uniform thermal field. No other technology has allowed one to achieve such significant simultaneous benefits.

The first and second law efficiencies are higher if oxygen concentration in the reactants is lowered and/or if preheating is used to raise the maximum temperature. Thus, the ODC which is a combination of heat and gas recirculation, produces higher thermodynamic efficiencies. Maximum thermodynamic temperature of an ODC system is mild and less than that of ordinary combustion. It is directly proportional to the heat recirculation and inversely proportional to the gas-recirculation rate and hence is lower in the case of oxygen used as an oxidizer. The exergy analyses in this study can be used as a technical tool to guide further efficiency-improvements in combustion processes.

Nomenclature

C_p = specific heat (J/kg K)
$\tilde{c}_p^{\varepsilon}$ = mean molar isobaric exergy capacity (J/mol K)
E = exergy rate (W)
H = enthalpy (J/kg)
$\dot{I}$ = irreversibility rate (W)
$\dot{m}$ = mass flow rate (kg/s)
$\dot{n}$ = rate of number of moles (mol/s)
P = pressure (Pa)
Q = energy (J/kg)
R = gas recirculation rate
$\dot{S}$ = absolute entropy rate (J/K s)
T = temperature (K)
T_{alf} = the adiabatic limit temperature (K)
T_{lim} = limiting maximum temperature of the reactants
T_{limp} = maximum limiting temperature of the reactants due to mixing and preheating (K)
x = molar fraction
δ = second law efficiency defect
$\tilde{\varepsilon}$ = specific molar exergy (W/mol)
ϕ = equivalence ration
η = effectiveness of the heat exchanger
$\dot{\Pi}$ = entropy generation (J/K s)
ω = mass fraction
ψ = second law efficiency

Subscripts

a = added
ch = chemical
comb = combustion
f = final
F = fuel
ph = physical
HX = heat exchanger
i = input
mix = mixer
o = output, initial, environmental condition
ox = oxidizer
P = products
R = reactants
s = surface
$\sim$ = molar

References

[1] Tsuji, H., Gupta, A., Hasegawa, T., Katsuki, M., Kishimoto, K., and Morita, M., 2003, *High Temperature Air Combustion; From Energy Conservation to Pollution Reduction*, CRC, Boca Raton, FL.
[2] Hasegawa, T., Mochida, S., and Gupta, A. K., 2002, "Development of Advanced Industrial Furnace Using Highly Preheated Combustion Air," J. Propul. Power, **18**(2), 233–239.
[3] Gupta, A. K., Bolz, S., and Hasegawa, T., 1999, "Effect of Air Preheat and Oxygen Concentration on Flame Structure and Emission," ASME J. Energy Resour. Technol., **121**, pp. 209–216.
[4] Gupta, A. K., 2004, "Thermal Characteristics of Gaseous Fuel Flames Using High Temperature Air," ASME J. Eng. Gas Turbines Power, **126**(1), pp. 9–19.
[5] Mortberg, M., Gupta, A. K., and Blasiak, W., 2004, "Flow Phenomena of Normal and Low Calorific Value Fuels in High Temperature Air Combustion Conditions," *AFRC/JFRC Joint Fall Symposium*, Maui, HI, Oct. 10–14.
[6] Jangsawang, W., Klimanek, A., and Gupta, A. K., 2005, "Experiments for Enhanced Yield of Hydrogen From Wastes Using High Temperature Steam Gasification," *24th International Conference on Incineration and Thermal Treatment Technologies (IT3)*, Galveston, TX, May 9–13; see also Jangsawang, W., Klimanek, A., and Gupta, A. K., 2006, "Enhanced Yield of Hydrogen From Wastes Using High Temperature Steam Gasification," ASME J. Energy Resour. Technol., **128**(3), pp. 179–185.
[7] Kalisz, S., Abeyweera, R., Szewczyk, D., Jansson, A., Lucas, C., and Blasiak, W., 2004, "Energy Balance of the High Temperature Air/Steam Gasification of Biomass in Up-Draft Fixed Bed Type Gasifier," *23rd International Conference on Incineration and Thermal Treatment Technologies (IT3)*, Phoenix, AZ, May 10–14.
[8] Konishi, N., Kitagawa, K., Arai, N., and Gupta, A. K., 2002, "Two-Dimensional Spectroscopic Analysis of Spontaneous Emission From a Flame Using Highly Preheated Air Combustion," J. Propul. Power, **18**(1), pp. 199–204.
[9] Kitagawa, K., Konishi, N., Arai, N., and Gupta, A. K., 2003, "Temporally Resolved 2-D Spectroscopic Study on the Effect of Highly Preheated and Low Oxygen Concentration Air on Combustion," ASME J. Eng. Gas Turbines Power, **125**, pp. 326–331.
[10] Gupta, A. K., 2004, "Flame Length and Ignition Delay During the Combustion of Acetylene in High Temperature Air (in Japanese)," Industrial Heating Journal, JIFMA, **41**(5), pp. 44–52.
[11] Weinberg, F., 1971, "Combustion Temperature," Nature (London), **233**, pp. 239–241.
[12] Gupta, A. K., 2001, "Technological Evolution, Challenges and Future Prospects for the Application of HiTAC to HiCOT," *Invited Lecture at the High Temperature Air Combustion Technology Symposium*, Tokyo, Japan, Oct. 11.
[13] Shimada, T., Akiyama, T., Fukushima, S., Mitsui, K., Jinno, M., Kitagawa, K., Arai, N., and Gupta, A. K., 2005, "Time Resolved Temperature Profiling of Flames With Highly Preheated/Low Oxygen Concentration Air in an Industrial

Size Furnace," ASME J. Eng. Gas Turbines Power, **127**(3), pp. 464–471.
[14] Mortberg, M., Blasiak, W., and Gupta, A. K., 2006, "Combustion of Low Calorific Value Fuels in High Temperature and Oxygen Deficient Environment," Combust. Sci. Technol., **178**, pp. 1345–1372.
[15] Shimada, T., Akiyama, T., Fukushima, S., Kitagawa, K., Arai, N., Konishi, N., Itoh, S., Terabayashi, T., Ohkuboto, Y., and Gupta, A. K., 2004, "Spectroscopic Observation of Heavy Oil Luminous Flames in an Industrial Regenerative Furnace," J. Propul. Power, **20**(5), pp. 919–926.
[16] Konishi, N., Kitagawa, K., Arai, N., and Gupta, A. K., 2002, "Two-Dimensional Spectroscopic Analysis of Spontaneous Emission From a Flame Using Highly Preheated Air Combustion," J. Propul. Power, **18**(1), pp. 199–204.
[17] Krishnamurthy, N., Blasiak, W., and Lugnet, A., 2004, "Development of High Temperature Air and Oxy-Fuel Combustion Technologies for Minimized CO2 and NOx Emission in Industrial Heating," *The Joint International Conference on Sustainable Energy and Environment (SEE)*, Hua Hin, Thailand, Dec. 1–3.
[18] Cavaliere, A., and De Joannon, A., 2004, "Mild Combustion," *Progress in Energy and Combustion Science*, Elsevier, New York, Vol. 30, pp. 329–366.
[19] Wünning, J. A., and Wünning, J. G., 1997, "Flameless Oxidation to Reduce Thermal NO-Formation," *Progress Energy and Combustion Science*, Elsevier, New York, Vol. 23, pp. 81–94.
[20] Rosen, M. A., 2002, "Does Industry Embrace Exergy?," Exergy, an International Journal, **2**, pp. 221–223.
[21] Szargut, J., Morris, D., and Steward, F., 1988, "Exergy Analysis of Thermal, Chemical, and Metallurgical Processes," Hemisphere, New York.
[22] Kotas, T. J., 1995, "The Exergy Methods of Thermal Plant Analysis," Krieger Publishers, Melbourne, FL.
[23] Daw, S., Chakravarthy, K., Conklin, J., and Graves, R. A., 2004, "Generic Concept for 'Flameless' Combustion With Higher Thermodynamic Availability," *Proceedings of the Joint International Combustion Symposium of the American and Japanese Flame Research Committee*, Maui Oct. 10–13.
[24] Van Wylen, G., and Sonntag, R., 1991, *Fundamentals of Thermodynamics*, 3rd ed., Wiley, New York.
[25] Morley, C., 2005, "Chemical Equilibria in Perfect Gases," GASEQ V. 0.79 software, www.gaseq.co.uk

Combined Effects of Overheating and Soot-Blower Erosion on Reheater Tubing in a Gas-Fired Large Capacity Boiler

Nasr M. Hosny
Higher Institute of Engineering,
Thebes Academy,
P.O. Box 2075,
Cairo 11511, Egypt

In a large capacity tangentially fired boiler, the final reheater tubing sustained abnormal oxidation and localized excessive metal wastage in a short time of the unit operation. The root causes of the problem are identified by test data analysis. The test data indicated that the reheater tubing metal temperatures in the affected areas exceeded the recommended limit of the metal oxidation temperature due to higher than expected local gas temperatures and velocities. A soot-blower facing the overheated portion of the reheater leading tubes accelerated the process of metal wastage by periodically removing the oxide layer. The configuration of the boiler internals upstream of the reheater section is found to be the main cause of the localized overheating. Side-to-side gas flow/temperature stratification due to tangential firing contributed to a lesser degree to the problem. The results and conclusions presented in this paper should be a beneficial guide to the designer of large capacity boilers. [DOI: 10.1115/1.2795769]

1 Introduction

Reheater tube failure is one of the leading causes of forced outages of fossil-fired utility boilers. The main reasons responsible for reheater tube failure are creep rupture, caused by long-term overheating, and fly ash erosion. However, abnormal oxidation of tube outer surface (caused by short-term overheating) combined with erosion (caused by frequent use of a soot-blower) can seriously reduce the tube-wall thickness and lead to premature failure.

The formation of oxide scale on ferritic alloys is a diffusion-controlled reaction that follows a parabolic growth rate law [1]. The rate of oxide formation is initially high but decreases as the layer thickness increases and becomes self-limiting/protective. Also, the rate is a function of temperature; too high a temperature can lead to too rapid oxidation. Soot blowers can cause erosion to boiler tubes in the neighborhood of their operation. The basic mechanism is the removal of the protective oxide layer. The exposed clean tube metal in the high temperature gas atmosphere is reoxidized, consuming the tube metal each cycle [2]. By scale formation and removal, tube-metal thickness is severely reduced to the point of failure.

This paper presents such a case where leading tubes of the final reheater assemblies, in a large capacity boiler, sustained severe metal wastage after 31,000 h in operation. The boiler is a controlled-circulation unit rated for 578 kg/s of steam at 17.2 MPa and 541°C. The reheater was designed for 4.2 MPa at 541°C, and consists of two sections: primary and secondary (final). The final reheater consists of 192 pendant tube assemblies (panels) placed across the convection zone of the boiler; each assembly contains five tube loops in a row. Figure 1 shows a side view of the upper section of the boiler where the final reheater is placed downstream of the final superheater, and a screen of water tubes is installed between them. The final superheater consists of 192 pendent (vertical) tube assemblies; each assembly contains six tube loops in a row. Each tube loop is composed of two portions: the front loop and the rear loop. The rear loop is short, leaving a wide gap between the bottom of that portion of the superheater and the floor of the boiler convection zone.

The boiler is equipped with a number of soot blowers. One of the blowers is a retractable type installed on a track to move horizontally back and forth across the boiler width and facing the lower portion of the reheater tubing at a distance of 206 mm from the front tube row. The estimated temperature of blowing steam, just before tube impingement, is 400°C.

The boiler is tangentially fired in a pressurized divided furnace. The boiler has provisions to burn natural gas or oil, but it was firing clean natural gas all of the time. Hence, fuel-ash corrosion or erosion is not expected to contribute to the problem under investigation.

In a tangentially fired boiler, there are two known factors that contribute to superheater/reheater tube overheating: gas flow and temperature imbalances across their perspective sections due to the rotating swirl of flue gas and the imbalance of steam flow distribution within each section [3]. However, improper tube assemblies' configuration or placement in the convection pass can strongly aggravate the gas flow/temperature imbalance, causing unexpected localized regions of tube overheating. If these conditions are associated with an erosion enhancing factor, severe damage can be reflected on tube assemblies placed in these regions in a short time of unit operation.

In order to determine the extent and severity of the reheater tubing damage, a survey of tube thickness across the boiler width was conducted and tube samples were taken to measure oxide scale thickness and for microstructural examination. Test measurements of tubing metal temperature at selected locations, flue-gas temperature, and element outlet steam temperature were performed in order to explore the root causes of the problem.

2 Reheater Visual Examination

A visual examination of the reheater tube assemblies indicated that the front five tubes of the assemblies close to the sidewalls of the convection pass and the first two tubes of the middle assemblies had a thick external oxide scale. The scale on the front side of the leading tubes was thicker and partially exfoliated. The effective tube-wall thickness was reduced from 4.18 mm to 1.6 mm at some points. Tube thinning was particularly severe on leading

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received May 28, 2006; final manuscript received June 10, 2007; published online January 22, 2008. Review conducted by Kenneth Mark Bryden.

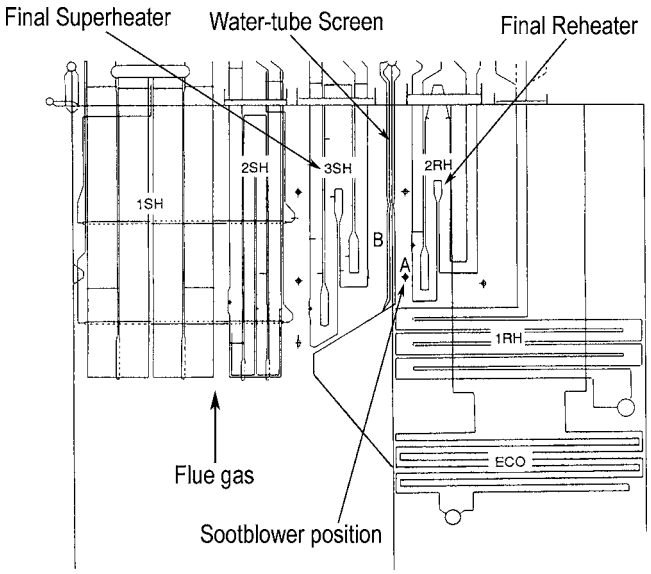
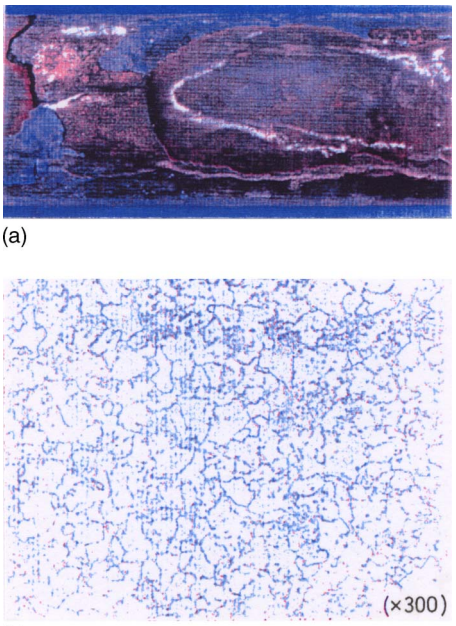


Fig. 1 A side view of the upper section of the boiler shows the final reheater and upstream installations: a soot blower, water-tube screen, and the final superheater

tubes close to the sidewalls at the bottom of the inlet to the reheater section. The affected tube length was concentrated around the soot-blower elevation (±300 mm).

3 Tube Sample Examination

Tube samples removed from the affected areas of the reheater front row exhibited a very thick brown colored scale of metal oxide. Figure 2(*a*) shows the external appearance of a tube sample removed from the leading tube of assembly 181 at the soot-blower level. The scale had grown in layers over the outer surface of the tube. The thickness of the outside scale was as much as 1.37 mm



(a)

(b)

Fig. 2 (*a*) Front-side external appearance of a tube sample removed from the leading tube of assembly 181 at the soot-blower level. (*b*) Microstructure of the same tube sample as appeared on the transverse section of the tube-wall front side.

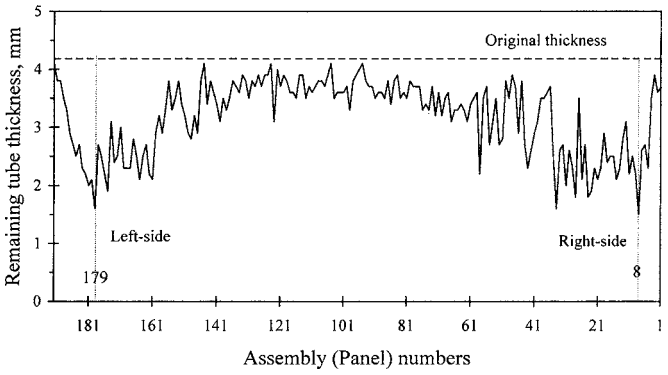


Fig. 3 Leading tube-thickness readings across the boiler width at the soot-blower elevation

and the front side external scale was thicker than the rear side. The internal surface of the tube sample was coated with a layer of magnetite (Fe_3O_4). The thickness of that layer was 0.48 mm.

The microstructure of the tube sample revealed a spheroidization feature of the carbide in ferrite, as shown in Fig. 2(*b*). The spheroidization indicated exposure of the tube material (ferritic steel, $2\frac{1}{4}$Cr–1Mo) to high metal temperatures above 579°C, which is the recommended limit of the oxidation temperature for this material [4].

4 Tube Thickness Survey

As indicated before, a visual examination of the reheater section revealed that tube-wall thinning was particularly prominent at the front side of the leading tube of each assembly. Therefore, the remaining thickness of all leading tubes, at the soot-blower elevation, was measured using an ultrasound test (nondestructive evaluation (NDE)). The profile of wall-thickness readings is shown in Fig. 3. The profile illustrates that the tube-wall loss was severe at the sides of the unit and tapered off toward the center. The left and right side 35 assemblies sustained the most severe metal loss, and the most significant reduction in wall thickness was recorded at assemblies 8 and 179 (from right to left) where the wall loss exceeded 50%. Tube thickness readings of the second tube row of the outboard assemblies indicated less exfoliation, and the minimum wall thickness is located on the tube sides rather than on the front side of the tube.

In addition, tube-wall thickness was measured along the most affected tubes above and below the soot-blower elevation. Figure 4 illustrates the measurement results. The figure indicates that tube thinning was most severe at the soot-blower elevation and decreased quickly above and below this level.

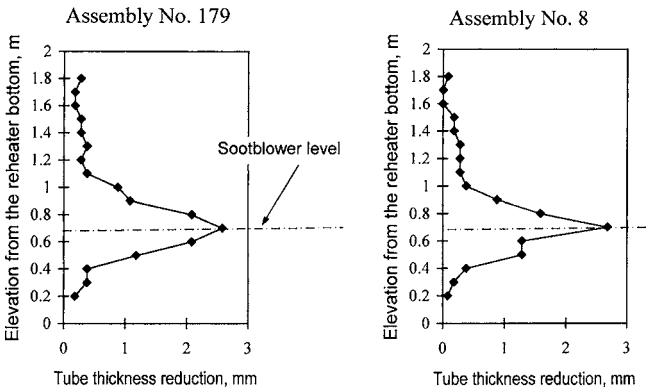


Fig. 4 Reduction in tube thickness along the leading tube of each of the assemblies 8 and 179

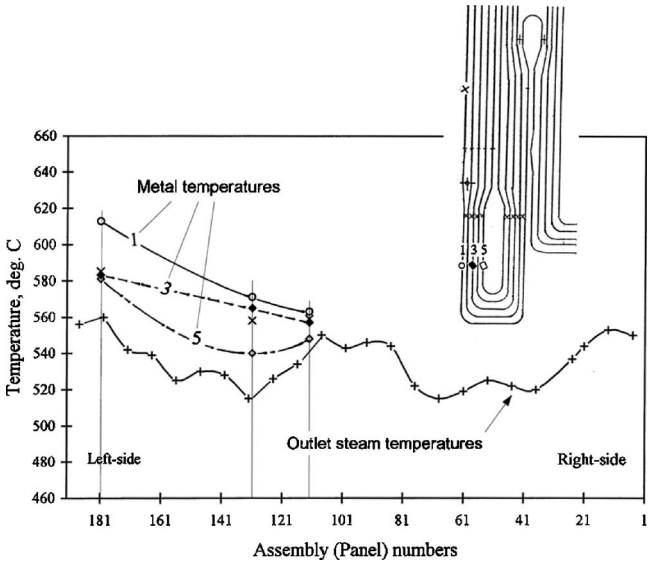


Fig. 5 Tube-metal temperatures and the front-tube outlet-steam temperatures

5 Temperature Measurements

In order to identify the causes of the problem and to determine an appropriate solution, test data were collected at a boiler steaming rate of 566 kg/h (98% of the boiler maximum rating), which corresponds to the unit full load. The test data included element outlet steam temperature measurements, flue-gas temperature measurements, and tubing metal temperature measurements at selected locations. The evaluation of metal temperatures for the reheater tubing was conducted at the unit full load rather than at intermediate loads because initial measurements of temperature indicated that the metal temperatures are highest at the full load. Temperature readings were collected during 2.5 h of the unit stable operation. The collected readings were taken at 10 min intervals. The calculated standard deviation of the collected readings is $\pm 0.8°C$ (68.3% certainty).

The outlet steam temperature profile for the front tube loop, across the boiler width, is shown in Fig. 5. The profile pattern is typical for a tangentially fired divided furnace, where element steam high temperatures are found at the sides and center of the units [3]. This pattern of temperature profile occurs, partially, due to side-to-side gas velocity/temperature stratification associated with this type of firing. However, this effect does not provide a full explanation for the localized tube overheating at the bottom of the reheater section.

Outside surface metal temperatures of the gas-touched reheater tubing were measured by specially installed chordal thermocouples at selected locations. Assemblies 181, 131, and 112 were selected for this purpose in order to establish the tube-metal temperature profiles in and around the affected area of tube damage. The location of the thermocouples on each assembly is indicated in Fig. 5. Metal temperature data are shown in the same figure. The peak of the metal temperature readings is in the leading tube row of the side assemblies close to the wall of the gas pass where thinning rate of the reheater tubing is the highest, and the temperature decreases toward the center. This compares well with the tube-wall thickness profile in Fig. 3. Also, as indicated in Fig. 5, the highest metal temperature along the front tube is in the lower portion of the tube at the soot-blower level indicating localized overheating at this level.

Temperatures of the flue gas entering the bottom of the reheater section (at location A, Fig. 1) and the gas leaving the front superheater section (at location B, Fig. 1) were measured using a water-cooled moving probe. Location B is at a higher elevation than location A, but in a vertical plane upstream of the water-tube screen located ahead of the reheater section. The gas temperature partial profiles are shown in Fig. 6.

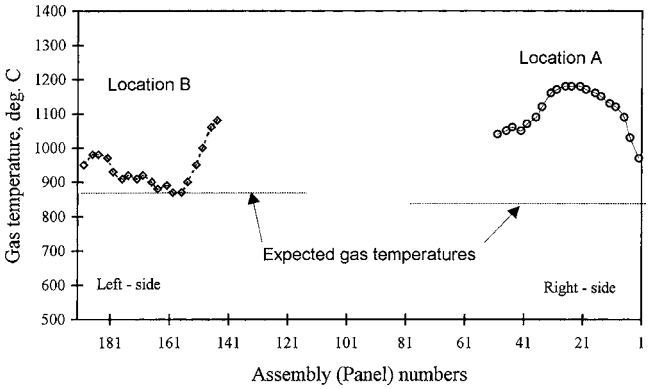


Fig. 6 Flue-gas temperature partial profiles

6 Results and Discussion

The wall-thickness profile of the front tube row, shown in Fig. 3, indicates severe metal wastage close to the sidewalls of the unit decreasing gradually toward the center. The metal wastage was concentrated at the lower portion of the reheater section around the soot-blower level, as indicated in Fig. 4. Also, the second tube row sustained less tube damage than the front row, and tubing external exfoliation decreased in the downstream direction.

Metal temperature readings, shown in Fig. 5, compare well with the wall-thickness profiles (the cross-correlation coefficient is −0.93083). The peak readings coincide with the areas of maximum metal wastage in the lower portion of the front row close to the unit sidewalls. The temperature decreases toward the center of the unit and in the subsequent tube rows. This indicates that localized overheating was the primary reason for the abnormal external oxidation and consequent serious thinning of the final reheater tubing. The maximum recorded metal temperature was 613°C, which exceeded the oxidation limit of the tubing material in the affected areas by 34°C.

The partial profile of the flue-gas temperature entering the lower portion of the reheater (Fig. 6, location A) has the same trends as the metal temperature profiles. The profile has a peak temperature of 1180°C near the sidewall. The average measured gas temperatures is 1108°C, while the manufacturer's design value of the gas temperature at this location was 840°C. This indicates that the localized tubing overheating was due, in part, to the higher than expected gas temperature entering the bottom of the reheater section.

The average of the measured gas temperatures at location B (Fig. 6) is 940°C. Thus, the average gas temperatures entering the bottom of the final reheater section is higher than the average gas temperatures leaving the upstream final superheater by 168°C. Apparently, this is because the gas stream entering the bottom of the final reheater section was subjected to less cooling surface area due to the short length of the rear pendant loop in the upstream superheater. However, the average temperature at location B is higher than the expected gas temperature by 70°C, indicating a gas temperature gradient having a peak near the bottom of the gas path and decreasing toward the boiler roof. This temperature gradient is due to the boiler upper section configuration and internal arrangement, and it seems to be prominent near the sidewalls because of the relatively poor gas mixing.

Calculated steam-side flow imbalance in the final reheater is found to be insignificant, approximately 4%. In a reheater of a uniform steam-side flow distribution, element outlet steam temperature distribution is a reflection of heat absorption distribution in each tubing circuit. In turn, the heat absorption distribution is

directly influenced by gas flow/temperature imbalance. The element outlet steam temperature distribution, in Fig. 5, exhibits peak temperatures at the sides and the center of the unit. This type of pattern is typical for a tangentially fired divided furnace. On the other hand, the peak readings at the sides coincide with the tubing of maximum metal wastage, but this is not the case at the center of the unit. Therefore, the gas flow/temperature imbalance due to tangential firing may be the prime reason of the outlet steam peak temperatures, but is not the main cause of the localized tube overheating at the bottom of the reheater.

A computer model of the upper section of the boiler was developed in order to calculate the gas flow distribution at the reheater section. The model utilized a finite-volume method in solving the governing equations of the gas flow. The modeling was carried out in three typical cross sections; each section is in the flow direction and parallel to the side wall of the boiler. The sections are spaced apart from each other to contain assemblies 181, 131, and 112 of the reheater. Based on the model input of the measured operational data and the boiler dimensional data, the calculated gas velocity entering the bottom of the reheater tubing is in the order of 22 m/s, while that velocity near the boiler roof is in the order of 10 m/s.

At the furnace outlet, the gas accelerates along with the boiler deflection arch. The accelerated gas stream encounters less resistance at the bottom of the final superheater due to the short length of its rear loop. Also, as indicated before, this gas stream is hotter than the rest of the gas leaving the superheater section. An installed baffle wall on the water-tube screen was directing the accelerated hot gas stream toward the area of the reheater where excessive tubing metal wastage was found. Therefore, configuration of the boiler internals and their placement with respect to each other contribute largely to the much higher than expected gas velocities and temperatures entering the lower portion of the reheater section. These conditions are combined with side-to-side gas flow/temperature stratification, associated with tangential firing, to cause unexpected localized tubing overheating.

Initial oxide scale on a low alloy steel, as that of the reheater tubing, grows stress free by a parabolic growth law (rapid initial formation). Later, a multilamination oxide structure is formed [5]. The scale formed is susceptible to exfoliation once it is cracked by application of locally excessive strains. Sources of strain from operation include cooling strain caused by application of soot blowing steam jets and subsequent cooling of the oxide structures from one uniform temperature to a lower temperature. The cooling strain is developed because of the difference in the coefficient of thermal expansion between the base material and the oxide, and also because of the differential thermal contraction of the oxide layered structure. The strain tolerance decreases with increasing oxide thickness and, therefore, decreases with operating time and higher temperatures [6].

Based on the data of Rehn et al. [7] and the approximated equation of French [8], a typical set of oxide growth curves for the measured metal temperatures (on the leading tube of each of the assemblies 181, 131, and 112) is shown in Fig. 7. The calculated scale thickness is based on the assumption of constant outside surface tube metal temperature in each case. For the metal temperature of 613°C (the maximum measured tubing temperature) and operating time of 31,000 h, the calculated scale thickness is 0.565 mm, which is approximately equivalent to 0.283 mm in tube-wall reduction. The actual measured reduction in that tube wall is 2.18 mm, which is 7.7 times the expected value. This ratio is 6.1 and 5.2 for the measured tube temperature of 571°C and 563°C, respectively. This indicates that the soot blower played an effective role in tube thinning, but the effectiveness of that role decreases with the decrease in the metal temperature toward the center of the boiler. Therefore, the root cause of the problem is the localized overheating of the reheater tubing, and the soot-blower accelerated the process of metal-tube wastage.

The result of this investigation could be applicable to other gas-firing boilers of comparable size and the same design configuration. In oil-firing and coal-firing boilers, the effects of erosion and corrosion, caused by fuel ash and other flue-gas corrosion oxides, are of additional considerations.

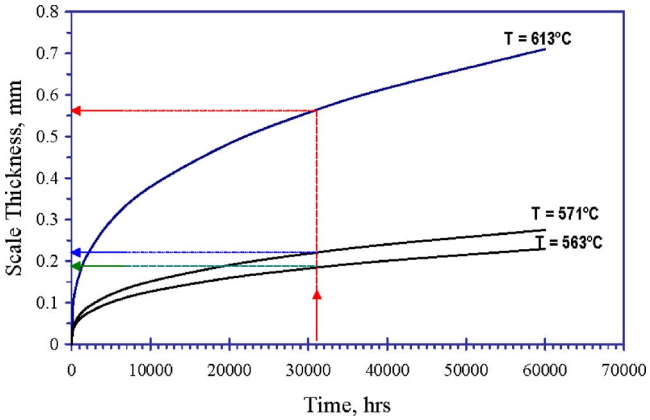


Fig. 7 Calculated scale thickness as a function of time for the measured tube-metal temperatures of 563°C, 571°C, and 613°C (based on the approximated equation of French [8])

7 Conclusion

The test data analysis and results have indicated that the excessive metal wastage of the reheater tubing is due to localized overheating as a result of higher than expected gas temperatures and velocities at the bottom and the sides of the inlet to the reheater section. The soot blower accelerated the process of metal wastage by periodically removing the oxide layer from the leading edge of the reheater tubes. The higher than expected gas temperature and velocity was mainly due to the combined effects of the short rear pendant loop in the upstream-installed tube assemblies and the design configuration of the upper section of the boiler. Side-to-side stratification in gas velocity/temperature at the furnace exit, associated with tangential firing, is another factor that contributed to the localized tube overheating.

In such a case, proper design configuration and placement of the boiler internals might decrease or eliminate localized tubing overheating. An accurate three-dimensional computer model of the gas path could be a useful tool to guide the boiler designer in this regard.

Wherever tubing overheating is suspected, higher alloys of tube materials should be considered and soot-blowing should not be performed on a regular basis but only as needed. Additional considerations to the effects of fuel-ash erosion and corrosion in oil-fired and coal-fired boilers are required.

References

[1] Davies, M. H., Simnad, M. T., and Birchenall, C. E., 1951, "On the Mechanism and Kinetics of the Scaling of Iron," J. Met., **3**, pp. 889–896.

[2] French, D. N., 1982, *Metallurgical Failures in Fossil Fired Boilers*, Wiley-Interscience, New York, p. 153.

[3] Llinares, V., and Schmidt, S., 1993, *Peak Temperature Reduction in Superheater and Reheater Assemblies: Approaches and Solutions*, ABB C-E Services, Inc., Windsor, CT., Technical paper No. TIS 8599.

[4] French, D. N., 1982, *Metallurgical Failures in Fossil Fired Boilers*, Wiley-Interscience, New York, p. 263.

[5] Dooley, R. B., and McNaughton, W. P., 1996, *Boiler Tube Failures: Theory and Practice, Volume 1: Boiler Tube Fundamentals*, Electric Power Research Institute, Palo Alto, CA, p. 215.

[6] Armitt, J. R., Holmes, M. I., Manning, D. B., and Metcalfe, E., 1978, "The Spalling of Steam-Grown Oxide From Superheater and Reheater Tube Steels," Technical Planning Study 76-655, Electric Power Research Institute, Palo Alto, CA, Final Report No. FP-686.

[7] Rehn, I. M., Apblett, W. R., Jr., and Stringer, J., 1981, "Controlling Steam-Side Oxide Exfoliation in Utility Boiler Superheaters and Reheaters," *Materials Performance*, Vol. 20, pp. 27–31.

[8] French, D. N., 1982, *Metallurgical Failures in Fossil Fired Boilers*, Wiley-Interscience, New York, p. 249.

Deviations in Predicted Condenser Performance for Power Plants Using HEI Correction Factors: A Case Study

Komandur S. Sunder Raj
President
Power & Energy Systems Services,
880 Ridgewood Avenue,
Oradell, New Jersey 07649

The Heat Exchange Institute (HEI) Standards for Steam Surface Condensers are used to design and predict the performance of condensers for power plant applications. Since their inception, the Standards have undergone numerous changes to incorporate technological advances and revisions to various factors based on testing and operating experiences. Admiralty and copper-nickel (CuNi) tubes were very popular until the 1970s. Subsequently, increasing concerns with the use of copper-based alloys in nuclear power plants as well as other factors led to specification and use of stainless steel (SS) and titanium. The first condenser designed with titanium tubes was put into service in 1977. In 1978, the HEI published the seventh edition of the HEI Standards for Steam Surface Condensers. The eighth edition was issued in 1984 followed by Addendum 1 in 1989. The ninth edition was issued in 1995 and Addendum 1 to the ninth edition was published in 2002. Notable differences between the ninth and seventh editions include higher circulating water inlet temperature correction factors below 70.0°F*; for Admiralty, higher tube material and gauge correction factors for tube wall gauge below 16 Birmingham wire gauge (BWG) and lower values above 20 BWG; for* 90/10 *CuNi and 304 SS, higher tube material and gauge correction factors for tube wall gauge between 12 BWG and 24 BWG; and for titanium, higher tube material and gauge correction factors for tube wall gauge above 18 BWG. Depending upon the tube diameter, material, wall gauge, and the correction factors used for a specific condenser application and its operating range, there could be substantial deviations in predicted condenser performance and associated impact on output. Using a case study, this paper examines the use of the correction factors from the seventh and ninth editions in power plant condenser performance predictions. It provides recommendations for developing proper benchmarks and for ensuring optimum condenser performance.* [DOI: 10.1115/1.2795779]

Keywords: HEI, condenser, correction factors

Background

In accordance with the Heat Exchange Institute (HEI) Standards for Steam Surface Condensers, condenser performance guarantees are provided at a single point, namely, the design condenser pressure. The choice of design circulating water inlet temperature corresponding to the design pressure is a matter of economic evaluation. The higher the design circulating water inlet temperature, the larger is the condenser and the greater the capital cost. The first cost has to be balanced against the economic benefits of performance gains from the lower condenser pressures. Design circulating water inlet temperatures for power plants using once through cooling may range from 50°F to 81°F. For plants using wet cooling towers, the values may be higher.

The HEI Standards clearly state that the performance of the condenser cannot be exactly predicted under varying operating conditions. The performance curves and tabulations provided in the Standards are approximate excepting at the design point.

The HEI published in 1978 the seventh edition of the Standards. In 1984, the eighth edition was issued and this was followed by Addendum 1 in 1989. The ninth edition was published in 1995 followed by Addendum 1 in 2002 [1–3].

Contributed by the International Gas Turbine Institute (IGTI) of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 14, 2006; Final manuscript received August 21, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Power 2006 (PWR2006), Atlanta, GA, May 2, 2006–04, 2006.

For the currently operating power plants that went into service during the period 1969–1983, the design of their condensers preceded the eighth and ninth editions of the HEI Standards.

During the 1970s, use of Admiralty and CuNi alloys for the tube materials was quite common. In the 1980s, adverse effects of copper-bearing alloys on nuclear steam generators and other factors led to use of stainless steel (SS) and titanium. The seventh edition of the HEI Standards contained limited data on titanium.

Heat Exchange Institute Correction Factors

The HEI Standards for Steam Surface Condensers provide overall tube bundle heat transfer rates for different tube diameters, wall thicknesses (Birmingham wire gauge (BWG)), tube materials, circulating water inlet temperatures, and average water velocities. The information was developed using, as the reference, clean, 18 BWG Admiralty tubes at a circulating water inlet temperature of 70°F.

Uncorrected heat transfer coefficients U_1 are provided for different tube diameters and average water velocities, assuming clean tubes. The service or expected heat transfer coefficient U is obtained by multiplying the uncorrected heat transfer coefficient U_1 with three correction factors, namely, F_W for circulating water inlet temperature correction, F_M for tube material and gauge correction, and F_C for tube cleanliness correction.

$$U = U_1 \times F_W \times F_M \times F_C \quad (1)$$

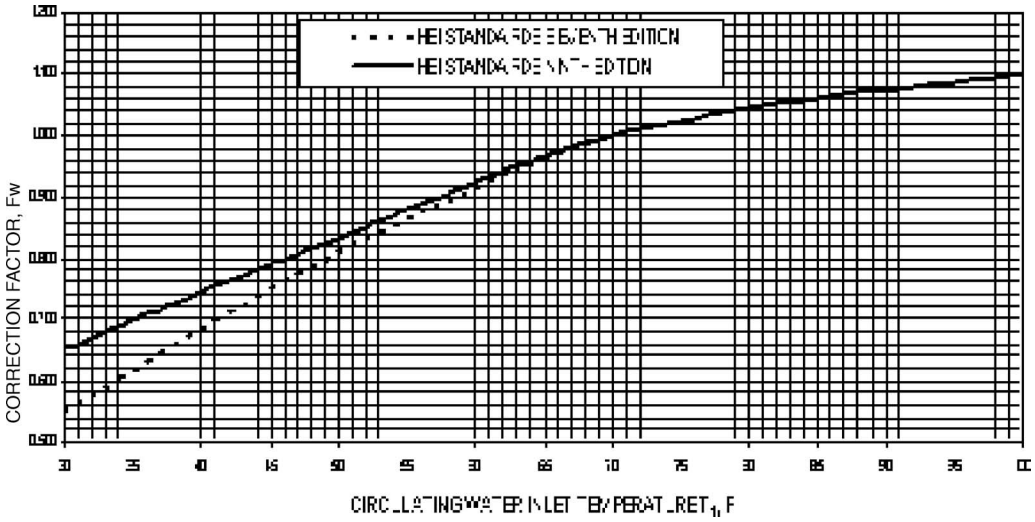


Fig. 1 Comparison of circulating water inlet temperature correction factor F_W

The uncorrected heat transfer coefficient U_1 is determined from the following:

$$U_1 = C \times \sqrt{V} \tag{2}$$

C is a constant and V is the average water velocity in the tubes. C has a value of 267 for tube outside diameters (ODs) of 5/8 in. and 3/4 in., 263 for tube ODs of 7/8 in. and 1.0 in., and 259 for tube ODs of 11/8 in. and 11/4 in.

Table 1 Comparison of temperature correction factor F_W

Circulating water inlet temperature T_1, °F	Correction factor F_W			Increase in heat transfer rate, %
	HEI 7th edition	HEI 9th edition	Change	
35	0.615	0.696	0.081	13
40	0.683	0.743	0.060	9
50	0.810	0.834	0.024	3
60	0.915	0.923	0.008	1
70	1.000	1.000	0.000	0

While U_1, F_W, and F_M are obtained from the HEI Standards, selection of the design tube cleanliness correction factor F_C depends upon the tube material, cooling water characteristics, operating conditions, etc. Commonly used design values are 0.85 for Admiralty and 90/10 CuNi, 0.90 for 304 SS, and 0.90–0.95 for titanium.

For Admiralty and 90/10 CuNi, 18 BWG for tube wall thickness is common. For 304 SS and titanium, to compensate for the reduced heat transfer characteristics, thinner wall thicknesses of 22 BWG or 24 BWG are common. For the purposes of this paper, discussion is limited to wall thicknesses of 18 BWG and 22 BWG.

The ninth edition of the HEI Standards was a complete rewrite of the earlier standards. It incorporated technological advances as well as revisions to the correction factors F_W and F_M, based on knowledge gained from testing and operating experiences.

Figure 1 shows a comparison of the correction factor F_W from the seventh and ninth editions as a function of the circulating water inlet temperature T_1.

Table 2 Comparison of tube material and gauge correction factor F_M

Tube material	18 BWG				22 BWG			
	HEI 7th edition	HEI 9th edition	Change	Increase in heat transfer rate, %	HEI 7th edition	HEI 9th edition	Change	Increase in heat transfer rate, %
Admiralty	1.00	1.00	0.00	0	1.04	1.02	−0.02	−2
90/10 CuNi	0.90	0.93	0.03	3	0.97	0.98	0.01	1
304 SS	0.69	0.75	0.06	9	0.79	0.86	0.07	9
Titanium	0.71	0.83	0.12	17	0.81	0.91	0.10	12

Table 3 Comparison of both F_W and F_M for 18 BWG Admiralty

Circulating water inlet temperature T_1, °F	Admiralty 18 BWG						
	HEI 7th edition			HEI 9th edition			Change in
	F_W	F_M	$F_W \times F_M$	F_W	F_M	$F_W \times F_M$	$F_W \times F_M$
35	0.615	1.00	0.615	0.696	1.00	0.696	13
40	0.683	1.00	0.683	0.743	1.00	0.743	9
50	0.810	1.00	0.810	0.834	1.00	0.834	3
60	0.915	1.00	0.915	0.923	1.00	0.923	1
70	1.000	1.00	1.000	1.000	1.00	1.000	0
80	1.045	1.00	1.045	1.045	1.00	1.045	0
90	1.075	1.00	1.075	1.075	1.00	1.075	0

Table 4 Comparison of both F_W and F_M for 18 BWG 90/10 Cuni (B=bowl; BP=Booster pump)

Circulating water inlet temperature	90-10 Copper-Nickel 18 BWG						
	HEI 7th edition			HEI 9th edition			Change in
T_1, °F	F_W	F_M	$F_W \times F_M$	F_W	F_M	$F_W \times F_M$	$F_W \times F_M$
35	0.615	0.90	0.554	0.696	0.93	0.647	17%
40	0.683	0.90	0.615	0.743	0.93	0.691	12%
50	0.810	0.90	0.729	0.834	0.93	0.776	6%
60	0.915	0.90	0.824	0.923	0.93	0.858	4%
70	1.000	0.90	0.900	1.000	0.93	0.930	3%
80	1.045	0.90	0.941	1.045	0.93	0.972	3%
90	1.075	0.90	0.968	1.075	0.93	1.000	3%

At temperatures above 70°F, there is no change in F_W between the two editions. However, at temperatures below 70°F, F_W for the ninth edition becomes progressively greater.

Table 1 shows a comparison of the circulating water inlet temperature correction factor F_W at various temperatures below 70°F. It can be noted that the increase in heat transfer coefficient in the ninth edition varies from 1% at 60°F to about 13% at 35°F.

Similarly, Table 2 shows a comparison of the tube material and gauge correction factor F_M for 18 BWG and 22 BWG for Admiralty, 90/10 CuNi, 304 SS, and titanium. The largest increase in heat transfer coefficient is for titanium followed by 304 SS, 90/10 CuNi, and Admiralty.

To keep the analysis simple, the effect of F_M was not considered separately. The combined effects of F_W and F_M from Tables 1 and 2, respectively, at different circulating water inlet temperatures are shown in Tables 3–6. At 35°F, the largest increase in heat transfer coefficient is for titanium (27%) followed by 304 SS (23%), 90/10 CuNi (17%), and Admiralty (13%). At 90°F, the corresponding increases are 12%, 9%, 3%, and 0%, respectively.

Case Study: 800 MWe Boiling Water Reactor Power Plant

To study the impact of the HEI correction factors on condenser performance and output, we will examine a case study for a BWR nuclear power plant with a nominal rating of 800 MWe. The methodology and discussion that follow are equally applicable to fossil power plants.

The condenser was designed to produce a pressure of 3.7 in. Hg (absolute) at a circulating water inlet temperature of 77.0°F when supplied with 352,600 gal/min of cooling water from a lake. The condenser was equipped with 7/8 in. OD, 18 BWG, Admiralty tubes, 43.82 m effective tube length and surface area of 370,000 sq ft. The design tube cleanliness factor (CF) was 85% (F_C=0.85). The correction factors F_W and F_M from the seventh edition of the HEI Standards were used to specify and design the condenser [1].

Using published procedures [4], design data and a performance modeling tool, Fig. 2 shows the heat balance generated for the case study at the licensed reactor thermal power of 2536 MWt and design circulating water inlet temperature of 77.0°F. The calculated condenser pressure is 3.70 in. Hg (absolute) and the calculated generator output approximately 864 MWe.

Admiralty Tubes: Changes in F_W. Table 7 shows comparison of the results between the seventh and ninth editions of the HEI Standards for the effect of changes in F_W at different circulating water inlet temperatures.

The following may be noted:

- The condenser pressure using F_W from the ninth edition is

Table 5 Comparison of both F_W and F_M for 22 BWG titanium

Circulating water inlet temperature	Titanium 22 BWG						
	HEI 7th edition			HEI 9th edition			Change in
T_1, °F	F_W	F_M	$F_W \times F_M$	F_W	F_M	$F_W \times F_M$	$F_W \times F_M$
35	0.615	0.81	0.498	0.696	0.91	0.633	27%
40	0.683	0.81	0.553	0.743	0.91	0.676	22%
50	0.810	0.81	0.656	0.834	0.91	0.759	16%
60	0.915	0.81	0.741	0.923	0.91	0.840	13%
70	1.000	0.81	0.810	1.000	0.91	0.910	12%
80	1.045	0.81	0.846	1.045	0.91	0.951	12%
90	1.075	0.81	0.871	1.075	0.91	0.978	12%

Table 6 Comparison of both F_W and F_M for 22 BWG 304 SS

Circulating water inlet temperature	304 Stainless Steel 22 BWG						
	HEI 7th edition			HEI 9th edition			Change in
T_1, °F	F_W	F_M	$F_W \times F_M$	F_W	F_M	$F_W \times F_M$	$F_W \times F_M$
35	0.615	0.79	0.486	0.696	0.86	0.599	23%
40	0.683	0.79	0.540	0.743	0.86	0.639	18%
50	0.810	0.79	0.640	0.834	0.86	0.717	12%
60	0.915	0.79	0.723	0.923	0.86	0.794	10%
70	1.000	0.79	0.790	1.000	0.86	0.860	9%
80	1.045	0.79	0.826	1.045	0.86	0.899	9%
90	1.075	0.79	0.849	1.075	0.86	0.925	9%

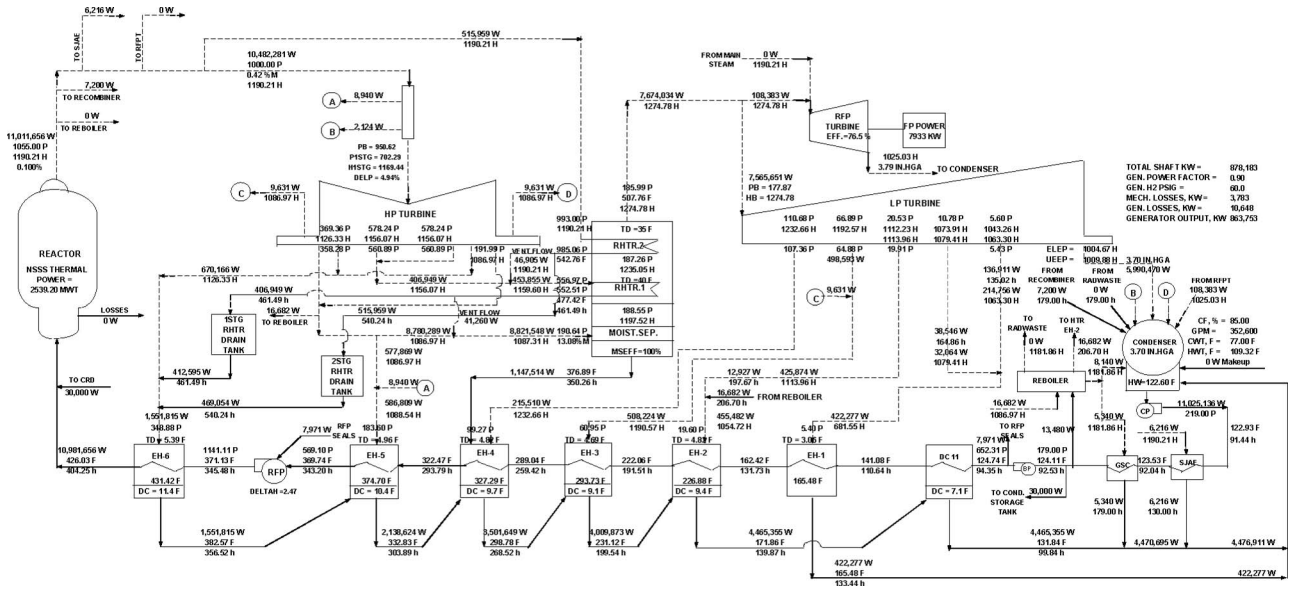


Fig. 2 Heat balance for case study at design circulating water inlet temperature

lower by approximately 0.02 in. Hg (absolute) (0.7%) at 60.0°F and by about 0.24 in. Hg (absolute) (13.7%) at 35.0°F.

- The generator output is virtually unchanged.

Using F_W from the ninth edition results in lower condenser pressures at circulating water inlet temperatures below 70.0°F. Had the condenser for the case study been designed using F_W from the ninth edition in lieu of the seventh edition, the service heat transfer coefficients would have been higher at circulating water inlet temperatures between 35.0°F and 60.0°F, as shown in Table 7. Also, if we assume actual condenser pressures to correspond to values calculated for the seventh edition as shown in Table 7, this would be equivalent to computing new values of CF using the following equation:

$$\mathrm{CF}_{\mathrm{new}} = \mathrm{CF}_{\mathrm{design}}(F_{W7\mathrm{thed}}/F_{W9\mathrm{thed}}) \tag{3}$$

At the circulating water inlet temperature of 35.0°F, the above equation would yield a new CF of 75% (=85×0.615/0.696) for a reduction of about 10 percentage points. At 40.0°F, the new CF would be 78%, for reduction of about 7 percentage points and at 50.0°F the new CF would be 82%, for a reduction of about 3 percentage points.

The impact of F_W on CF lessens as circulating water inlet temperature increase above 35.0°F. The analysis indicates that there could be significant changes in the calculated CF values, depending upon the circulating water inlet temperatures, actual condenser pressures, and the correction factors used.

Admiralty Tubes: Changes in F_W and F_M. Since there is no change in F_M between the seventh and ninth editions for 18 BWG Admiralty, the effect of changes in F_W and F_M is the same as that for F_W alone shown in Table 7.

90/10 CuNi Tubes: Changes in F_W. Using 90/10 CuNi, Table 8 shows comparison of the results between the seventh and ninth editions of the HEI Standards for the effect of changes in F_W at different circulating water inlet temperatures.

The following may be noted:

- The condenser pressure using F_W from the ninth edition is lower by 0.02 in. Hg (absolute) (0.75%) at 60.0°F and by 0.30 in. Hg (absolute) (15.1%) at 35.0°F.
- The increase in generator output varies from insignificant at 60.0°F to about 0.7 MWe (0.08%) at 35.0°F.

Using F_W for 90/10 CuNi from the ninth edition results in lower condenser pressures at circulating water inlet temperatures below 70.0°F. Similar to the analysis for Admiralty, these would

Table 7 Comparison of results for Admiralty—Changes in F_w

Circ. Water Inlet Temp. T_1, °F	Water Vel., ft/s	Constant C	Uncorr. Heat Transfer Coefficient U_1, Btu hr Sq Ft F $U_1 = C \times \sqrt{V}$	Circ. water Inlet temp. Corr. Factor. F_w 7th ed	F_w 9th ed	Tube Material & Gauge Corr Factor. F_w 7th ed	Tube cleanliness corr. factor, F_C	Heat transfer coefficient U. Btu hr SQ FT F 7th ed	9th ed	Cond. press., IN.HGA 7th ed	9th ed	Generator output. MW 7th ed	9th ed	ΔMW
35.00	6.47	263	669.1	0.615	0.696	1.00	0.85	349.8	395.8	1.724	1.487	885.927	885.833	0.094
40.00	6.47	263	669.1	0.683	0.743	1.00	0.85	388.4	422.6	1.772	1.617	885.858	885.985	0.127
50.00	6.47	263	669.1	0.810	0.834	1.00	0.85	460.7	474.3	2.016	1.964	885.118	885.309	0.191
60.00	6.47	263	669.1	0.915	0.923	1.00	0.85	520.4	524.9	2.452	2.435	882.215	882.363	0.148
70.00	6.47	263	669.1	1.000	1.000	1.00	0.85	568.7	568.7	3.083	3.083	874.268	874.268	0.000
77.00	6.47	263	669.1	1.033	1.033	1.00	0.85	587.5	587.5	3.698	3.698	863.935	863.935	0.000
85.00	6.47	263	669.1	1.060	1.060	1.00	0.85	602.9	602.9	4.577	4.577	847.299	847.299	0.000

Table 8 Comparison of results for 90/10 Cuni: Changes in F_w

Circ. water inlet temp. T_1, °F	Water vel., ft/s	Constant C	Uncorr. heat transfer coefficient U_1, Btu hr sq ft f $U_1 = C \times \sqrt{V}$	Circ. water inlet temp. corr. factor. F_w 7th ed	Circ. water inlet temp. corr. factor. F_w 9th ed	Tube material & gauge corr factor. F_w 7th ed	Tube cleanliness corr. factor, F_C	Heat transfer coefficient U. btu hr SQ FT F 7th ed	Heat transfer coefficient U. btu hr SQ FT F 9th ed	Cond. press., In. HGA 7th ed	Cond. press., In. HGA 9th ed	Generator output. MW 7th ed	Generator output. MW 9th ed	ΔMW
35.00	6.47	263	669.1	0.615	0.696	0.90	0.85	314.8	356.3	1.985	1.685	885.242	885.951	0.709
40.00	6.47	263	669.1	0.683	0.743	0.90	0.85	349.6	380.3	2.009	1.815	85.147	885.770	0.623
50.00	6.47	263	669.1	0.810	0.834	0.90	0.85	414.6	426.9	2.235	2.169	883.932	84.343	0.411
60.00	6.47	263	669.1	0.915	0.923	0.90	0.85	468.4	472.5	2.678	2.658	879.847	880.007	0.160
70.00	6.47	263	669.1	1.000	1.000	0.90	0.85	511.9	511.9	3.336	3.336	870.214	870.214	0.000
77.00	6.47	263	669.1	1.033	1.033	0.90	0.85	528.8	528.8	3.988	3.988	885.580	858.580	0.000
80.00	6.47	263	669.1	1.045	1.045	0.90	0.85	534.9	534.9	4.311	4.311	852.465	852.465	0.000

translate to changes in calculated CF of about 10 percentage points at 35.0°F, 7 percentage points at 40.0°F, and 3 percentage points at 50.0°F.

90/10 CuNi Tubes: Changes in F_W and F_M. Table 9 shows comparison of the results between the seventh and ninth editions of the HEI Standards for the effect of changes in both F_W and F_M at different circulating water inlet temperatures.

The following may be noted from Table 9:

- The condenser pressure using both F_W and F_M from the ninth edition is lower by approximately 0.1 in. Hg (absolute) above 60.0°F and by about 0.37 in. Hg (absolute) (18.5%) at 35.0°F.
- The increase in generator output varies from about 0.7 MWe (0.08%) at 35.0°F to about 2.0 MWe (0.2%) at 80.0°F.

The combined effect of using F_W and F_M for 90/10 CuNi from the ninth edition results in lower condenser pressures at all circulating water inlet temperatures. These would equate to changes in calculated CF of about 12 percentage points at 35.0°F, 9 percentage points at 40.0°F, 5 percentage points at 50.0°F, and 3 percentage points above 60.0°F.

304 Stainless Steel Tubes: Changes in F_W. Using 304 SS, Table 10 shows comparison of the results between the seventh and ninth editions of the HEI Standards for the effect of changes in F_W at different circulating water inlet temperatures.

The following may be noted:

- The condenser pressure using F_W from the ninth edition is lower by approximately 0.03 in. Hg (absolute) (0.9%) at 60.0°F and by about 0.41 in. Hg (absolute) (16.9%) at 35.0°F.

Table 9 Comparison of results for 90/10 Cuni: Changes in F_W and F_M

Circ. water inlet temp. T_1, °F	Water vel., ft/s	Constant C	Uncorr. heat transfer coefficient U_1, Btu hr sq ft f $U_1 = C \times \sqrt{V}$	Circ. water inlet temp. corr. factor. F_w 7th ed	Circ. water inlet temp. corr. factor. F_w 9th ed	Tube material & gauge corr factor. F_w 7th ed	Tube material & gauge corr factor. F_w 9th ed	Tube cleanliness corr. factor, F_C	Heat transfer coefficient U. btu hr SQ FT F 7th ed	Heat transfer coefficient U. btu hr SQ FT F 9th ed	Cond. press., In.HGA 7th ed	Cond. press., In.HGA 9th ed	Generator output. MW 7th ed	Generator output. MW 9th ed	ΔMW
35.00	6.47	263	669.1	0.615	0.696	0.90	0.93	0.85	314.8	368.1	1.985	1.618	885.242	885.979	0.737
40.00	6.47	263	669.1	0.683	0.743	0.90	0.93	0.85	349.6	393.0	2.009	1.749	885.147	885.896	0.749
50.00	6.47	263	669.1	0.810	0.834	0.90	0.93	0.85	414.6	441.1	2.235	2.101	883.932	884.723	0.791
60.00	6.47	263	669.1	0.915	0.923	0.90	0.93	0.85	468.4	488.2	2.678	2.583	879.847	880.830	0.983
70.00	6.47	263	669.1	1.000	1.000	0.90	0.93	0.85	511.9	528.9	3.336	3.257	870.214	871.674	1.460
77.00	6.47	263	669.1	1.033	1.033	0.90	0.93	0.85	528.8	546.4	3.988	3.891	585.580	860.369	1.816
80.00	6.47	263	669.1	1.045	1.045	0.90	0.93	0.85	537.9	552.7	4.311	4.208	852.465	854.429	1.964

Table 10 Comparison of results for 304 SS: Changes in F_W

Circ. water inlet temp. T_1, °F	Water vel., ft/s	Constant C	Uncorr. heat transfer coefficient U_1, Btu/hr sq ft f $U_1 = C \times \sqrt{V}$	Circ. water inlet temp. corr. factor, F_W 7th ed	Circ. water inlet temp. corr. factor, F_W 9th ed	Tube material and gauge corr. factor, F_M 7th ed	Tube cleanliness corr. factor, F_C	Heat transfer coefficient U, Btu/hr SQ FT F 7th ed	Heat transfer coefficient U, Btu/hr SQ FT F 9th ed	Cond. press., In Hga (absolute) 7th ed	Cond. press., In Hga (absolute) 9th ed	Generator output, MW 7th ed	Generator output, MW 9th ed	ΔMW
35.00	5.83	263	634.8	0.615	0.696	0.79	0.90	277.6	314.1	2.396	1.990	882.693	885.223	2.530
40.00	5.83	263	634.8	0.683	0.743	0.79	0.90	308.3	335.3	2.376	2.117	882.865	884.627	1.762
50.00	5.83	263	634.8	0.810	0.834	0.79	0.90	365.6	376.4	2.566	2.480	881.079	881.946	0.867
60.00	5.83	263	634.8	0.915	0.923	0.79	0.90	413.0	416.6	3.019	2.992	875.308	875.699	0.391
70.00	5.83	263	634.8	1.000	1.000	0.79	0.90	451.3	451.3	3.715	3.715	863.562	863.562	0.000
77.00	5.83	263	634.8	1.033	1.033	0.79	0.90	466.2	466.2	4.422	4.422	850.344	850.344	0.000
80.00	5.83	263	637.8	1.045	1.045	0.79	0.90	471.6	471.6	4.772	4.772	843.643	843.643	0.000

Table 11 Comparison of results for 304 SS: Changes in F_W and F_M

Circ. water inlet temp. T_1, °F	Water vel., ft/s	Constant C	Uncorr. heat transfer coefficient U_1, Btu/hr sq ft F $U_1=C\times\sqrt{V}$	Circ. water inlet temp. corr. factor, F_W		Tube material and gauge corr. factor, F_M		Tube cleanliness corr. factor, F_C	Heat transfer coefficient U, Btu/hr sq ft F		Cond. press., IN.HGA		Generator output. MW		
				7th ed	9th ed	7th ed	9th ed		7th ed	9th ed	7th ed	9th ed	7th ed	9th ed	ΔMW
35.00	5.83	263	634.8	0.615	0.696	0.79	0.86	0.90	277.6	342.0	2.396	1.774	882.693	885.840	3.147
40.00	5.83	263	634.8	0.683	0.743	0.79	0.86	0.90	308.3	365.1	2.376	1.903	882.865	885.533	2.667
50.00	5.83	263	634.8	0.810	0.834	0.79	0.86	0.90	365.6	409.8	2.566	2.260	881.079	883.757	2.678
60.00	5.83	263	634.8	0.915	0.923	0.79	0.86	0.90	413.0	453.5	3.019	2.756	875.308	878.906	3.598
70.00	5.83	263	634.8	1.000	1.000	0.79	0.86	0.90	451.3	491.3	3.715	3.448	863.562	868.314	4.752
77.00	5.83	263	634.8	1.033	1.033	0.79	0.86	0.90	466.2	507.5	4.422	4.116	850.344	856.170	5.826
80.00	5.83	263	634.8	1.045	1.045	0.79	0.86	0.90	471.6	513.4	4.772	4.447	843.643	849.865	6.222

- The increase in generator output varies from insignificant at 60.0°F to 2.5 MWe (0.3%) at 35.0°F.

Using F_W for 304 SS from the ninth edition results in lower condenser pressures at circulating water inlet temperatures below 70.0°F. Similar to Admiralty and 90/10 CuNi, these would translate to change in calculated CF of about 10 percentage points at 35.0°F, 7 percentage points at 40.0°F, and 3 percentage points at 50.0°F.

304 Stainless Steel Tubes: Changes in F_W and F_M. Table 11 shows comparison of the results between the seventh and ninth editions of the HEI Standards for the effect of changes in both F_W and F_M at different circulating water inlet temperatures.

The following should be noted:

- The condenser pressure using both F_W and F_M from the ninth edition is lower by about 0.33 in. Hg (absolute) (6.8%) at 80.0°F, 0.26 in. Hg (absolute) (8.7%) at 60.0°F and by about 0.62 in. Hg (absolute) (26.0%) at 35.0°F.
- The increase in generator output varies from about 3.0 MWe (0.36%) at 35.0°F to about 6.0 MWe (0.74%) at 80.0°F.

The combined effect of using F_W and F_M for 304 SS from the ninth edition results in lower condenser pressures at all circulating water inlet temperatures. These would equate to changes in calculated CF of about 17 percentage points at 35.0°F, 14 percentage points at 40.0°F, 10 percentage points at 50.0°F, and 8 percentage points above 60.0°F.

Titanium Tubes: Changes in F_W. Using titanium, Table 12 shows comparison of the results between the seventh and ninth editions of the HEI Standards for the effect of changes in F_W at different circulating water inlet temperatures.

The following may be noted:

- The condenser pressure is lower by about 0.02 in. Hg (absolute) (0.8%) at 60.0°F and by about 0.33 in. Hg (absolute) (15.6%) at 35.0°F.
- The increase in generator output varies from insignificant at 60.0°F to 1.2 MWe (0.14%) at 35.0°F.

Using F_W for titanium from the ninth edition results in lower condenser pressures at circulating water inlet temperatures below 70.0°F. Similar to Admiralty, 90/10 CuNi, and 304 SS, these would translate to changes in calculated CF of about 10 percentage points at 35.0°F, 7 percentage points at 40.0°F, and 3 percentage points at 50.0°F.

Titanium Tubes: Changes in F_W and F_M. Table 13 shows comparison of the results between the seventh and ninth editions of the HEI Standards for the effect of changes in both F_W and F_M at different circulating water inlet temperatures.

The following should be noted:

- The condenser pressure using both F_W and F_M from the ninth edition is lower by about 0.37 in. Hg (absolute) (8.3%) at 80.0°F, 0.29 in. Hg (absolute) (10.3%) at 60.0°F, and by about 0.57 in. Hg (absolute) (27.0%) at 35.0°F.
- The increase in generator output varies from about 1.3 MWe (0.15%) at 35.0°F to 7.0 MWe (0.83%) at 80.0°F.

The combined effect of using F_W and F_M for titanium from the ninth edition results in lower condenser pressures at all circulating water inlet temperatures. These would equate to changes in calculated CF of about 20 percentage points at 35.0°F, 18 percentage points at 40.0°F, 13 percentage points at 50.0°F, and 11 percentage points above 60.0°F.

Table 12 Comparison of results for titanium: Changes in F_W

Circ. water inlet temp. T_1, °F	Water vel., ft/s	Constant C	Uncorr. heat transfer coefficient U_1, Btu/hr sq ft F $U_1=C\times\sqrt{V}$	Circ. water inlet temp. corr. factor. F_W		Tube material and gauge corr. factor, F_M 7th ed	Tube cleanliness corr. factor, F_C	Heat transfer coefficient U, btu/hr sq ft F		Cond. press., IN.HGA		Generator output. MW		
				7th ed	9th ed			7th ed	9th ed	7th ed	9th ed	7th ed	9th ed	ΔMW
35.00	5.83	263	634.8	0.615	0.696	0.81	0.95	300.4	340.0	2.122	1.790	884.605	885.814	1.209
40.00	5.83	263	634.8	0.683	0.743	0.81	0.95	333.6	362.9	2.132	1.917	884.551	885.489	0.938
50.00	5.83	263	634.8	0.810	0.834	0.81	0.95	395.7	407.4	2.346	2.274	883.118	883.660	0.542
60.00	5.83	263	634.8	0.915	0.923	0.81	0.95	447.0	450.9	2.793	2.771	878.428	878.719	0.291
70.00	5.83	263	634.8	1.000	1.000	0.81	0.95	488.5	488.5	3.465	3.465	868.081	868.081	0.000
77.00	5.83	263	634.8	1.033	1.033	0.81	0.95	504.6	504.6	4.135	4.135	855.802	855.802	0.000
80.00	5.83	263	634.8	1.045	1.045	0.81	0.95	510.5	510.5	4.468	4.468	849.470	849.470	0.000

Table 13 Comparison of results for titanium: Changes in F_W and F_M

Circ. water inlet temp. T_1, °F	Water vel., ft/s	Constant C	Uncorr. heat transfer coefficient U_1, Btu/hr sq ft F $U_1 = C \times \sqrt{V}$	Circ. water inlet temp. corr. factor, F_W		Tube material and gauge corr. factor, F_M		Tube cleanliness corr. factor, F_C	Heat transfer coefficient U, Btu/hr sq ft F		Cond. press., IN.HGA		Generator output. MW		
				7th ed	9th ed	7th ed	9th ed		7th ed	9th ed	7th ed	9th ed	7th ed	9th ed	ΔMW
35.00	5.83	263	634.8	0.615	0.696	0.81	0.91	0.95	300.4	381.9	2.122	1.548	884.605	885.912	1.307
40.00	5.83	263	634.8	0.683	0.743	0.81	0.91	0.95	333.6	407.7	2.132	1.678	884.551	885.970	1.419
50.00	5.83	263	634.8	0.810	0.834	0.81	0.91	0.95	395.7	457.7	2.346	2.027	883.118	885.071	1.953
60.00	5.83	263	634.8	0.915	0.923	0.81	0.91	0.95	447.0	506.5	2.793	2.504	878.428	881.640	3.212
70.00	5.83	263	634.8	1.000	1.000	0.81	0.91	0.95	488.5	548.8	3.465	3.162	868.081	873.115	5.034
77.00	5.83	263	634.8	1.033	1.033	0.81	0.91	0.95	504.6	566.9	4.135	3.788	855.802	862.302	6.500
80.00	5.83	263	634.8	1.045	1.045	0.81	0.91	0.95	510.5	573.5	4.468	4.099	849.470	856.509	7.039

Conclusions and Recommendations

Using a case study, the paper has discussed the difference in condenser performance and associated impact upon output due to the changes in correction factors between the seventh and ninth editions of the HEI Standards for Steam Surface Condensers.

The case study examined four tube materials, namely, Admiralty, 90/10 CuNi, 304 SS, and titanium with wall thicknesses of 18 BWG and 22 BWG. The design CF values used were 85% for Admiralty and 90/10 CuNi, 90% for 304 SS, and 95% for titanium.

The most significant conclusions for the case study are as follows:

- For the tube materials examined, the use of F_W alone from the ninth edition results in a decrease in condenser pressure of 0.02–0.03 in. Hg (absolute) at 60.0°F and 0.2–0.4 in. Hg (absolute) at 35.0°F.
- When F_W is combined with F_M from the ninth edition, the decrease in condenser pressures is 0.4–0.6 in. Hg (absolute) at 35.0°F and 0.1–0.4 in. Hg (absolute) at 80.0°F.
- For all tube materials studied, the use of F_W alone from the ninth edition at temperatures below 70.0°F results in lower condenser pressures. As noted in the analysis, this would be equivalent to changes in calculated CF, ranging from 1 percentage point at 60.0°F to 10 percentage points at 35.0°F.
- When F_W is combined with F_M, the equivalent changes in calculated CF range from 3 percentage points above 60.0°F to about 12 percentage points at 35.0°F for 90/10 CuNi. For 304 SS, the changes are 8 percentage points and 17 percentage points, and for titanium, 11 percentage points and 20 percentage points, respectively.
- At 35.0°F, the gain in output due to use of F_W alone from the ninth edition ranges from insignificant up to 3 MWe.
- At 35.0°F, when F_W is combined with F_M, the output gain is about 1 MWe for 90/10 CuNi and titanium, and 3 MWe for 304 SS.
- At 80.0°F, when F_W is combined with F_M, the output gain is about 2 MWe for 90/10 CuNi, 6 MWe for 304 SS, and 7 MWe for titanium.

It should be noted that the actual gain in output for a specific application depends upon the condenser design, its operating range in relation to the LP turbine last stage design, and the end loading [5].

Analysis of various power plants shows that the last stage may reach maximum output (choked flow conditions) at circulating water inlet temperatures of 70.0°F and below. For these plants, there would be virtually no gain in output below 70.0°F, regardless of the edition used to calculate condenser pressures.

While the methodology in this paper is applicable to any size unit, both fossil and nuclear, a case-by-case analysis should be undertaken to determine individual results.

As discussed in this paper, condensers originally designed using the seventh edition of the HEI Standards may see different results between predicted and actual performance when the ninth edition is used. Actual condenser pressures might be significantly different from predicted values and calculated tube cleanliness factors may not be consistent with design or predicted values. Compared to the seventh edition, the correction factors in the ninth edition reflect technological advances as well as additional testing and operating experiences.

To ensure proper benchmarking, schedule optimal tube cleanings, and optimize performance, the following recommendations are proposed:

- For new condensers or those plants that have retubed or modified their condensers, acceptance tests or accurate performance tests should be conducted to validate the HEI correction factors. The tests should be conducted at or close to the design point and, in addition, at other points in the operating range.
- Operating plants should conduct accurate performance tests on their condensers before and after condenser tube cleanings.
- Tube cleanliness factors should be calculated using results of the performance tests. The values should help to validate the appropriateness of the HEI correction factors.
- Performance predictions should be developed for the entire operating range by all plants using condenser design data and the appropriate HEI correction factors.
- The performance predictions may be used to schedule optimal condenser cleanings as well as to monitor changes in performance.
- Operating plants should provide feedback from their tests and operating experiences to the HEI for incorporation into future editions of the Standards.
- The HEI should provide guidance to the power industry regarding use of the appropriate correction factors. A position paper would be beneficial.

Nomenclature

C = constant for calculating uncorrected heat transfer coefficient U_1
CF = cleanliness factor, %
Δ = enthalpy rise, Btu/lb
F_C = tube cleanliness correction factor
F_M = tube material and gauge correction factor
F_W = circulating water inlet temperature correction factor
T = temperature, °F
T_1 = inlet temperature, °F
U_1 = uncorrected heat transfer coefficient, Btu/h sq ft °F

U = service heat transfer coefficient, Btu/h sq ft °F
$= U_1 \times F_W \times F_M \times F_C$

References

[1] 1978, Heat Exchange Institute Standards for Steam Surface Condensers, Seventh Edition.

[2] 1995, Heat Exchange Institute Standards for Steam Surface Condensers, Ninth Edition.

[3] 2002, Addendum 1 Standards for Steam Surface Condensers, Ninth Edition.

[4] Baily, F. G., Cotton, K. C., and Spencer, R. C., 1967, "Predicting The Performance of Large Steam Turbine-Generators Operating With Saturated And Superheated Steam Conditions," *GER-2454A Presented at the American Power Conference*, Chicago, Illinois, Apr. 25–27.

[5] Sunder Raj, K. S., 2006, "Last Stage Performance Considerations in Low-Pressure Turbines of Power Plants—A Case Study," ASME Paper No. PWR2006-88196.

Last Stage Performance Considerations in Low-Pressure Turbines of Power Plants: A Case Study

Komandur S. Sunder Raj
President
Power & Energy Systems Services,
880 Ridgewood Avenue,
Oradell, NJ 07649

The last stage blades (LSBs) of low-pressure (LP) turbine power plants have been historically specified and designed on the basis of optimization studies by matching the turbine to the condenser/cooling system for a specified unit rating. LSB sizes for U.S. nuclear applications currently range from 38 in. to 52 in. for unit ratings of 600 MWe to 1200 MWe. LP turbine arrangements usually consist of two or three double-flow sections in parallel. Last stage end loadings (last stage mass flow divided by the last stage annulus area) vary from approximately 8000 lb/h sq ft to 14,000 lb/h sq ft, with corresponding unit loadings (electrical output in megawatts divided by last stage annulus area) of 1.1 MWe/sq ft to 2.1 MWe/sq ft. Several power plants have been upgrading/replacing their LP turbines. Considerations include efficiency, reliability, power uprates, operating license renewals (nuclear), aging, inspection, and maintenance. In some cases, LP turbine rotors are being replaced with new rotors, blading, and steam path. Others are replacing LP turbines with new and advanced designs incorporating improved technology, better materials, optimized steam paths, more efficient blading, longer LSB sizes, redesigned exhaust hoods, etc. Unlike the other stages in the LP turbine, the last stage performance is affected by both the upstream (load) and downstream (condenser) conditions. While the LP turbines are being upgraded or replaced, no major modifications or upgrades are being made to the condensers. To address vibration effects due to increased flows and velocities from power uprates, the condenser tubes may be staked. Circulating water pumps may or may not be upgraded depending upon the particular application. Consequently, while improvements in LP turbines lead to more efficient utilization of the available energy and higher output, the last stage performance may be out of synchronization with the existing condenser/cooling system. Undersized or oversized LSB sizes in relation to the unit rating and end loading may result in less than optimum performance depending upon the design and operating range of the condenser/cooling system. This paper examines the various factors that affect last stage performance of LP turbines. Using a case study, it discusses the relationships between the last stage, the unit rating, the end loading, and the operating range of the condenser/cooling system. It examines different last stage exhaust loss curves and provides recommendations for selection of LSB sizes for optimum performance. [DOI: 10.1115/1.2795781]

Keywords: low-pressure turbine, last stage

General

The stage pressures decrease and the volumetric flows increase as steam expands through various stages of a low-pressure (LP) turbine. To maintain blade stresses at acceptable levels, the blade sizes and flow areas (annulus areas) also increase. The last stage, consequently, has the largest blade size and annulus area and the highest annulus velocity. For last stage blade (LSB) sizes from 38 in. to 52 in., the corresponding pitch diameters (diameter based on mean height of the blades) vary from 10 ft to 13 ft. At 1800 rpm, the blade velocities range from about 1000 ft/s to 1200 ft/s with corresponding tip speeds of approximately 1300–1600 ft/s.

For a given load, the work output of the last stage is a function of the pressure ratio (ratio of downstream to upstream pressure) across the last stage. A decrease in the downstream (condenser) pressure will be accompanied by an increase in volumetric flow, annulus velocity, and work output from the last stage. When the critical pressure ratio is reached, the last stage output is maximum and the last stage is under choked flow conditions. For given upstream conditions and load, any further reduction in condenser pressure will not be acknowledged by the last stage and there will be no further increase in the last stage output [1].

Conversely, as the condenser pressure increases, the pressure ratio across the last stage also increases, the volumetric flow and annulus velocity decrease and there is a reduction in last stage output. Although the available energy to the last stage has decreased, the last stage must continue to spin at a high velocity.

The energy in the steam leaving the last stage varies as one-half of the square of the last stage annulus velocity and is not recoverable. Since the last stage has the highest annulus velocity, axial Mach numbers for this stage may be as high as 1.6. Blading losses are a function of the blade design and the Mach number. In order to keep the leaving losses low, several advances have been made in design of the last stage blades [2,3].

The LP turbine exhaust hood facilitates conveyance of the exhaust steam from the last stage to the condenser. The LP turbine

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 14, 2006; final manuscript received August 20, 2007; published online January 22, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Power 2006 (PWR2006), Atlanta, GA, May 2–4, 2006.

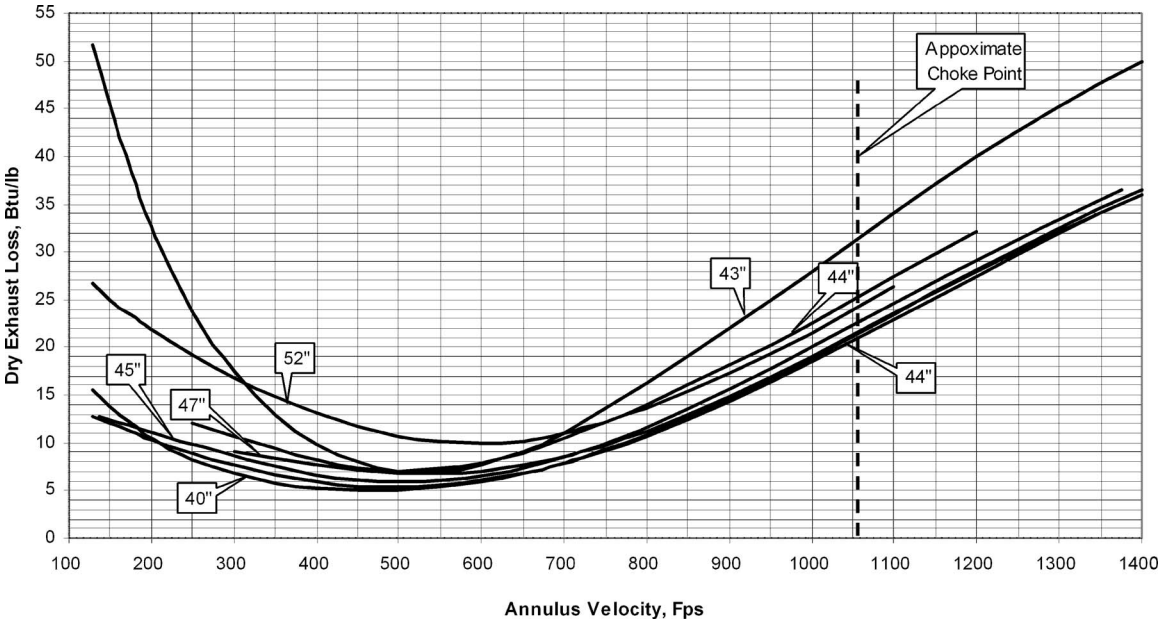


Fig. 1 Exhaust loss curves for various LSB sizes

performance is based on the exhaust pressure measured adjacent to the flange connection between the exhaust hood and the condenser neck. The hood loss is the isentropic change in enthalpy corresponding to the static pressures at the last stage exit and at the flange. Consequently, the pressure drop in the exhaust hood becomes a critical consideration in performance of the last stage [4,5].

Thus, the exhaust loss curve for the last stage is a composite of three components: first, the leaving loss that is a function of the kinetic energy in the exhaust steam; second, the pressure loss occurring in the exhaust hood from the exit of the LSB to the LP turbine exhaust flange, where the exhaust pressure is measured; third, the turn-up loss that occurs at low annulus velocities [1]. Figure 1 shows the exhaust loss curves for various LSB sizes for nuclear applications.

High-performance LP elements with diffusing exhaust hoods generally choke at axial Mach numbers of approximately 0.87–0.90 (last stage annulus velocity of about 1058 fps).

Figure 2 shows the annulus velocities at different end loadings and last stage exhaust pressures. Shown on the figure are the expected velocity ranges for various LSB sizes for a nuclear unit on once-through cooling equipped with 43 in. LSB.

At full load, for the condenser operating range of

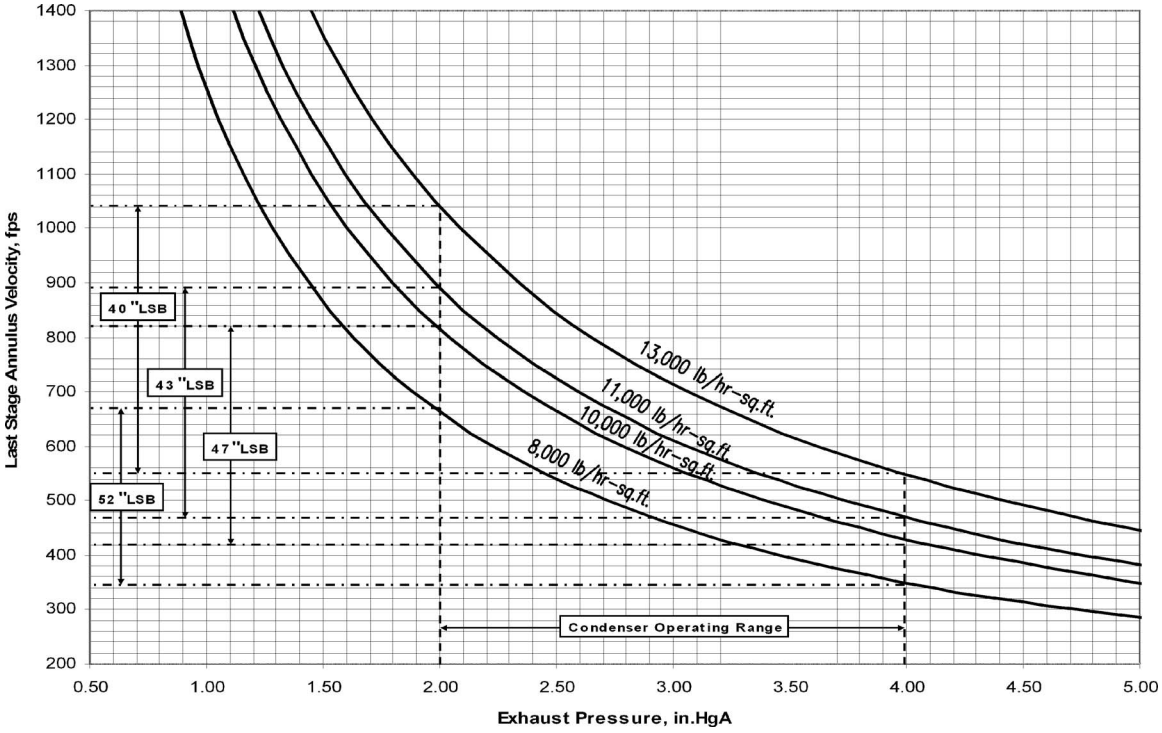


Fig. 2 Last stage annulus velocities for different end loadings, exhaust pressures

Table 1 Last stage performance for 43 in. LSB

in. Cond. press. Hg (absolute)	Last stage bowl press. P_B, psi (absolute)	Last stage shell press. P_{Sh}, psi (absolute)	Last stage press. ratio, P_{Sh}/P_B	Last stage bowl enthalpy H_B, Btu/lb	Last stage bowl entropy S_B, Btu/lb	Isentropic enthalpy H_{isen}, Btu/lb	Last stage available energy AE, Btu/lb	Expansion line end point ELEP, Btu/lb	Last stage exhaust flow W, Lb/h
(1)	(2)	(3)	(4) =(3)/(2)	(5)	(6)	(7)	(8)=(5)-(7)	(9)	(10)
2.0	4.85	0.98	0.203	1044.2	1.70763	953.0	91.2	967.9	8,093,233
2.5	4.89	1.23	0.251	1044.5	1.70726	964.7	79.8	977.3	8,154,128
3.0	4.92	1.47	0.299	1044.7	1.70702	974.4	70.3	985.2	8,205,506
3.5	4.95	1.72	0.347	1044.9	1.70675	982.8	62.2	991.9	8,250,172
4.0	4.98	1.96	0.394	1045.1	1.70643	990.0	55.0	997.9	8,289,834

Last stage specific volume v, cft/lb	Last stage volumetric flow Wv, cft/h	Last stage annulus velocity V_{an}, ft/s	Dry exhaust loss EL, Btu/lb	Used energy end point UEEP, Btu/lb	Last stage UEEP efficiency, %	Last stage output, MW	Generator output, MWe	Last stage output, % of gen. output	End loading, lb/h sq ft
(11)	(12)=(10) ×(11)	(13)[a]	(14)	(15)	(16)={(5) − (15)}/(8)	(17)[b]	(18)	(19) ={(17)/(18)} ×100	(20) =(10)/742.8
2.94.206	2.381E+09	890	21.1	982.5	67.70%	146.4	1128.851	12.97%	10.896
239.937	1.956E+09	732	12.5	986.0	73.26%	139.7	1121.771	12.46%	10.978
203.160	1.667E+09	623	8.5	991.1	76.20%	128.8	1110.617	11.60%	11.047
176.526	1.456E+09	545	7.1	997.0	77.13%	115.9	1097.568	10.56%	11.107
156.312	1.296E+09	485	7.2	1003.0	76.47%	102.3	1083.781	9.44%	11.160

[a]Annulus area per last stage is 123.8 sq ft. For six exhaust ends, total last stage annulus area is 742.8 sq ft. Annulus velocity, fps=Last stage volumetric flow/(742.8 ×3600).
[b]Last stage output, MW=Last stage flow×(Last stage bowl enthalpy−Used energy end point)/3412141.63.

2.0–4.0 in. Hg (absolute) and a last stage end loading of 11,000 lb/h sq ft, the annulus velocity ranges from about 460 ft/s to 890 ft/s. If the same unit were equipped with 52 in. LSB, the end loading would drop to approximately 8000 lb/h sq ft and the annulus velocity range shifts downward to between 340 ft/s and 660 ft/s.

If we superimpose these velocity ranges on the 43 in. LSB and 52 in. LSB exhaust loss curves, respectively, in Fig. 1, the dry exhaust loss for the 43 in. LSB would vary from 8 Btu/lb to 22 Btu/lb and for the 52 in. LSB, from 15 Btu/lb to 10 btu/lb. Obviously, selection of the optimum LSB size entails careful balance of capital and operating costs based on the condenser/cooling system design and operating range, end loading, annulus velocity, and exhaust losses.

It is customary in turbine performance work to charge the last stage with the exhaust loss and to consider separately the effects of exhaust loss on LP turbine performance. The LP turbine expansion line provided by the turbine manufacturer terminates at a fictitious end point (enthalpy) on the Mollier Chart known as the expansion line end point (ELEP). Using the exhaust loss determined from the exhaust loss curve provided by the manufacturer, the ELEP is corrected to yield the used energy end point (UEEP) or turbine end point (TEP). This end point (enthalpy) represents the actual state of steam at the LP turbine exhaust pressure and is used in determination of the work output of the last stage [6,7].

Case Study: 1100 MWe Nuclear Unit

Performance With 43 in. Last Stage Blade. Let us examine a case study for a nuclear unit originally rated for 1100 MWe at 1.5 in. Hg (absolute). The unit was designed with three, double-flow, LP turbine sections arranged in parallel with 43 in. last stage blades (TC6F-43 in. LSB). A single-pressure condenser with three shells was provided and designed to produce 2.3 in.Hg (absolute) at a circulating water inlet temperature of 52°F. The source of cooling water for the condenser is a lake.

At full load conditions, for lake temperatures between 35°F and 80°F, the condenser operating range is 2.0–4.0 in. Hg (absolute). Using the thermal kit (turbine-generator performance data) provided by the OEM, the performance of the last stage with the 43 in. blades was determined for the condenser operating range using a performance modeling tool. The tool uses the first law of thermodynamics to achieve mass and energy balance throughout the turbine cycle.

Table 1 shows the predicted performance. The following should be noted:

- The last stage mass flow increases as the condenser pressure and the last stage pressure ratio increase. For the operating range of 2.0–4.0 in. Hg (absolute), the increase is a little over 2%.
- The last stage volumetric flow and the annulus velocity decrease as the condenser pressure and last stage pressure ratio increase. The reduction is about 46% for the operating range.
- The dry exhaust loss reaches a minimum at 3.5 in. Hg (absolute) and increases thereafter.
- The last stage efficiency based on the UEEP reaches a maximum value at 3.5 in. Hg (absolute) and declines thereafter.
- The reduction in generator output for the condenser operating range is about 45 MW and mirrors the reduction in last stage output of about 44 MW.
- At 2.0 in. Hg (absolute), the last stage produces about 13% of the total generator output. It decreases to about 9% at 4.0 in. Hg (absolute).
- The last stage end loading is approximately 11,000 lb/h sq ft.

Performance With 52 in. Last Stage Blade. If the same unit is equipped with 52 in. LSB, the predicted last stage performance for the same condenser operating range of 2.0–4.0 in. Hg (absolute) is shown in Table 2.

Compared to Table 1, the following may be noted for Table 2:

- The percentage increase in the last stage mass flow as the

Table 2 Last stage performance for 52 in. LSB

Cond. press in. Hg (absolute)	Last stage bowl press. P_B, psi (absolute)	Last stage shell press. P_{Sh}, psi (absolute)	Last stage press. ratio, P_{Sh}/P_B	Last stage bowl enthalpy H_B, Btu/lb	Last stage bowl entropy S_B, Btu/lb	Isentropic enthalpy H_{sen}, Btu/lb	Last stage available energy AE, Btu/lb	Expansion line end point ELEP, Btu/lb	Last stage exhaust flow W, lb/h
(1)	(2)	(3)	(4) =(3)/(2)	(5)	(6)	(7)	(8)=(5)–(7)	(9)	(10)
2.0	4.85	0.98	0.203	1044.1	1.70753	953.0	91.2	967.9	8,093,667
2.5	4.89	1.23	0.251	1044.4	1.70716	964.6	79.8	977.3	8,153,834
3.0	4.92	1.47	0.299	1044.7	1.70695	974.4	70.3	985.2	8,204,912
3.5	4.95	1.72	0.347	1044.9	1.70670	982.7	62.2	991.9	8,249,484
4.0	4.98	1.96	0.394	1045.1	1.70643	990.0	55.0	997.9	8,289,181

Last stage specific volume v, cft/lb	Last stage volumetric flow Wv, cft/h	Last stage annulus velocity V_{an}, ft/s	Dry exhaust loss EL, Btu/lb	Used energy end point UEEP, Btu/lb	Last stage UEEP efficiency, %	Last stage output, MW	Generator output, MWe	Last stage output, % of gen. output	End loading, lb/h sq. ft
(11)	(12)=(10) ×(11)	(13)[a]	(14)	(15)	(16)={(5) – (15)}/(8)	(17)[b]	(18)	(19) ={(17)/(18)} ×100	(20) =(10)/1055
294.190	2.381E+09	627	10.0	974.7	76.10%	164.6	1146.978	14.35%	7,672
239.924	1.956E+09	515	10.5	984.6	75.00%	143.0	1125.284	12.71%	7,729
203.151	1.667E+09	439	12.0	993.6	72.68%	122.8	1104.932	11.12%	7,777
176.522	1.456E+09	383	13.7	1001.6	69.67%	104.7	1086.645	9.63%	7,819
156.310	1.296E+09	341	15.2	1008.7	66.18%	88.5	1070.27	8.27%	7,857

[a]Annulus area per last stage is 175.8333 sq. ft. For six exhaust ends, total stage annulus area is 1055 sq. ft. Annulus velocity, fps=Last stage volumetric flow/(1055 ×3600).
[b]Last stage output, MW=Last stage flow×(Last stage bowl enthalpy-Used energy end point)/3412141.63.

condenser pressure and the last stage pressure ratio increase remains unchanged at 2%.
- The percentage decrease in the last stage volumetric flow and annulus velocity as the condenser pressure and last stage pressure ratio increase is unchanged at 46%.
- The dry exhaust loss increases continuously as the condenser pressure increases.
- The last stage efficiency based on the UEEP decreases continuously as the condenser pressure increases.
- The reduction in generator output and the last stage output for the condenser operating range is greater and is about 77 MW compared to 45 MW from Table 1.
- The 52 in. LSB produce about 18 MW more output than the 43 in. LSB at 2.0 in. Hg (absolute) and about 3 MW more at 2.5 in. Hg (absolute). At 3.0 in. Hg (absolute), however, the 52 in. LSB produce 6 MW less and this difference increases to about 14 MW at 4.0 in. Hg (absolute).
- At 2.0 in. Hg (absolute), the last stage produces about 14% of the total generator output. It decreases to about 8% at 4.0 in. Hg (absolute).
- The last stage end loading is approximately 8000 lb/h sq ft.

It is clear from the foregoing comparison that, for much of the condenser operating range above the design condenser pressure of 2.3 in. Hg (absolute), the 52 in. LSB produce progressively less output than the 43 in. LSB and are oversized for the unit rating.

Power Uprate Impact on 43 in. Last Stage Blade Performance. Let us study the impact of a proposed 15% power uprate on performance of the 43 in. LSB. Reference [8] describes recent industry efforts in power uprate programs.

Table 3 shows the results of the power uprate. It is assumed that the existing condenser is capable of accommodating the increases of approximately 18% in steam flows and 16% in heat loads.

Compared to Table 1, we have the following for Table 3:

- The condenser pressure at the lake temperature of 35°F is 2.76 in. Hg (absolute) compared to 2.0 in. Hg (absolute) prior to the uprate.
- At a lake temperature of 80°F, the condenser pressure is 5.0 in. Hg (absolute) compared to 4.0 in. Hg (absolute).
- The end loading is approximately 13,000 lb/h sq ft for an increase of about 18% from the original end loading of 11,000 lb/h sq ft.
- Last stage annulus velocities range from 773 ft/s at 2.76 in. Hg (absolute) to 459 ft/s at 5.0 in. Hg (absolute) compared to 890 ft/s at 2.0 in. Hg (absolute) to 485 ft/s at 4.0 in. Hg (absolute).
- The lower annulus velocities at the uprated condition shift the range to the left on the exhaust loss curve in Table 2 resulting in lower exhaust losses.
- The power uprate results in better utilization of the 43 in. LSB performance characteristics.

Power Uprate Impact on 52 in. Last Sage Blade Performance. Table 4 shows the impact on last stage performance at the uprated power assuming 52 in. LSB.

The following should be noted in comparison to Tables 2 and 3.

- The condenser operating range is essentially the same as in Table 3.
- The end loading is approximately 9000 lb/h sq ft for an increase of about 13% from the original end loading of 8000 lb/h sq ft.
- The range of last stage annulus velocities after the power uprate is lower compared to the values in Table 2 prior to the uprate.
- The lower annulus velocities for the uprated condition shift the range to the left on the exhaust loss curve in Fig. 1 resulting in higher exhaust losses.

The following conclusions may be drawn:

- If power uprate were not a consideration, for the original

Table 3 43 in. LSB performance for 15% power uprate with original condenser

Cond. press in. Hg (absolute)	Last stage bowl press. P_B, psi (absolute)	Last stage shell press. P_{sh}, psi (absolute)	Last stage press. ratio, P_{sh}/P_B	Last stage bowl enthalpy H_B, Btu/lb	Last stage bowl entropy S_B, Btu/lb	Isentropic enthalpy H_{sen}, Btu/lb	Last stage available energy AE, Btu/lb	Expansion line end point ELEP, Btu/lb	Last stage exhaust flow W, Lb/h
(1)	(2)	(3)	(4) =(3)/(2)	(5)	(6)	(7)	(8)=(5)–(7)	(9)	(10)
2.76	6.01	1.36	0.226	1039.5	1.67891	953.8	85.7	968.1	9,576,736
2.85	6.02	1.40	0.233	1039.5	1.67871	955.5	84.1	969.5	9,588,814
3.25	6.07	1.60	0.263	1039.5	1.67783	962.0	77.5	974.7	9,637,186
3.95	6.14	1.94	0.316	1039.4	1.67649	971.9	67.5	982.6	9,711,370
5.00	6.24	2.46	0.394	1039.3	1.67488	984.1	55.2	992.4	9,804,425

Last stage specific volume v, cft/lb	Last stage volumetric flow Wv, cft/h	Last stage annulus velocity V_{an}, ft/s	Dry exhaust loss EL, Btu/lb	Used energy end point UEEP, Btu/lb	Last stage UEEP efficiency, %	Last stage output, MW	Generator output, MWe	Last stage output, % of gen. output	End loading lb/h sq. ft
(11)	(12)=(10) ×(11)	(13)[a]	(14)	(15)	(16)={(5) – (15)}/(8)	(17)[b]	(18)	(19) ={(17)/(18)} ×100	(20) =(10)/742.8
215.802	2.067E+09	773	14.5	978.0	71.79%	172.7	1267.413	13.62%	12,893
209.450	2.008E+09	751	13.4	978.6	72.49%	171.3	1266.769%	13.52%	12,909
185.840	1.791E+09	670	9.9	981.5	74.81%	163.7	1258.675	13.01%	12,974
155.439	1.510E+09	565	7.3	987.7	76.60%	147.2%	1241.382	11.86%	13,074
125.309	1.229E+09	459	7.6	997.7	75.32%	119.6	1212.802	9.86%	13.199

[a]Annulus area per last stage is 123.8 sq. ft. For six exhaust ends, total last stage annulus area is 742.8 sq. ft. Annulus velocity, fps=Last stage volumetric flow/(742.8 ×3600).
[b]Last stage output, MW=Last stage flow×(Last stage bowl enthalpy-Used energy end point)/3412141.63.

condenser operating range of 2.0–4.0 in. Hg (absolute), better performance is obtained at condenser pressures below 2.5 in. Hg (absolute) with the 52 in. LSB compared to the 43 in. LSB. The gain in output is about 18 MW at 2.0 in. Hg (absolute). However, at condenser pressures above 2.5 in. Hg (absolute), the 43 in. LSB blades perform better and the gain in output with this size is about 14 MW at 4.0 in. Hg (absolute).

- If the unit were uprated by 15% retaining the original condenser/cooling system design, the condenser heat loads

Table 4 52 in. LSB performance for 15% power uprate with original condensor

Cond. Press in. Hg (absolute)	Last stage bowl press. P_B, psi (absolute)	Last stage shell press. P_{sh}, psia (absolute)	Last stage press. ratio, P_{sh}/P_B	Last stage bowl enthalpy H_B, Btu/lb	last stage bowl netropy S_B, Btu/lb	Isentropic enthalpy H_{sen}, Btu/lb	Last stage available energy AE, Btu/lb	Expansion line end point ELEP, Btu/lb	Last stage exhaust flow W, lb/h
(1)	(2)	(3)	(4) =(3)/(2)	(5)	(6)	(7)	(8)=(5)–(7)	(9)	(10)
2.74	6.01	1.35	0.224	1039.5	1.67891	953.5	86.1	967.8	9,574,576
2.84	6.02	1.40	0.232	1039.5	1.67873	955.2	84.3	969.3	9,587,724
3.26	6.07	1.60	0.264	1039.5	1.67782	962.1	77.4	974.7	9,637,599
3.98	6.14	1.95	0.318	1039.4	1,67646	972.2	67.1	982.9	9,713,204
5.05	6.24	2.48	0.398	1039.3	1.67483	984.6	54.7	992.8	9,807,062

Last stage specific volume v, cft/lb	Last stage volumetric flow Wv, cft/hr	Last stage annulus velocity V_{an}, ft/s	Dry exhaust loss EL, Btu/lb	Used energy end point UEEP Btu/lb	Last stage UEEP efficiency, %	Last stage output, MW	Generator output, MWe	Last stage output, % of gen. output	End loading lb/h sq. ft
(11)	(12)=(10) ×(11)	(13)[a]	(14)	(15)	(16)={(5) – (15)}/(8)	(17)[b]	(18)	(19) ={(17)/(18)} ×100	(20) =(10)/1055
217.219	2.080E+09	548	10.1	947.7	75.29%	181.8	1277.477	14.23%	9,075
210.288	2.016E+09	531	10.3	976.3	75.04%	177.8	1273.083	13.96%	9,088
185.559	1.788E+09	471	11.2	982.5	73.63%	161.0	1256.005	12.82%	9,135
154.451	1.500E+09	395	13.3	992.1	70.42%	134.6	1229.071	10.95%	9,207
124.251	1.219E+09	321	16.0	1003.9	64.63%	101.7	1195.266	8.51%	9,296

[a]Annulus area per last stage is 175.8333 sq. ft. For six exhaust ends, total last stage annulus area is 1055 sq. ft. Annulus velocity, fps=Last stage volumetric flow/(1055 ×3600).
[b]Last stage output, MW=Last stage flow×(Last stage bowl enthalpy-Used energy end point)/3412141.63.

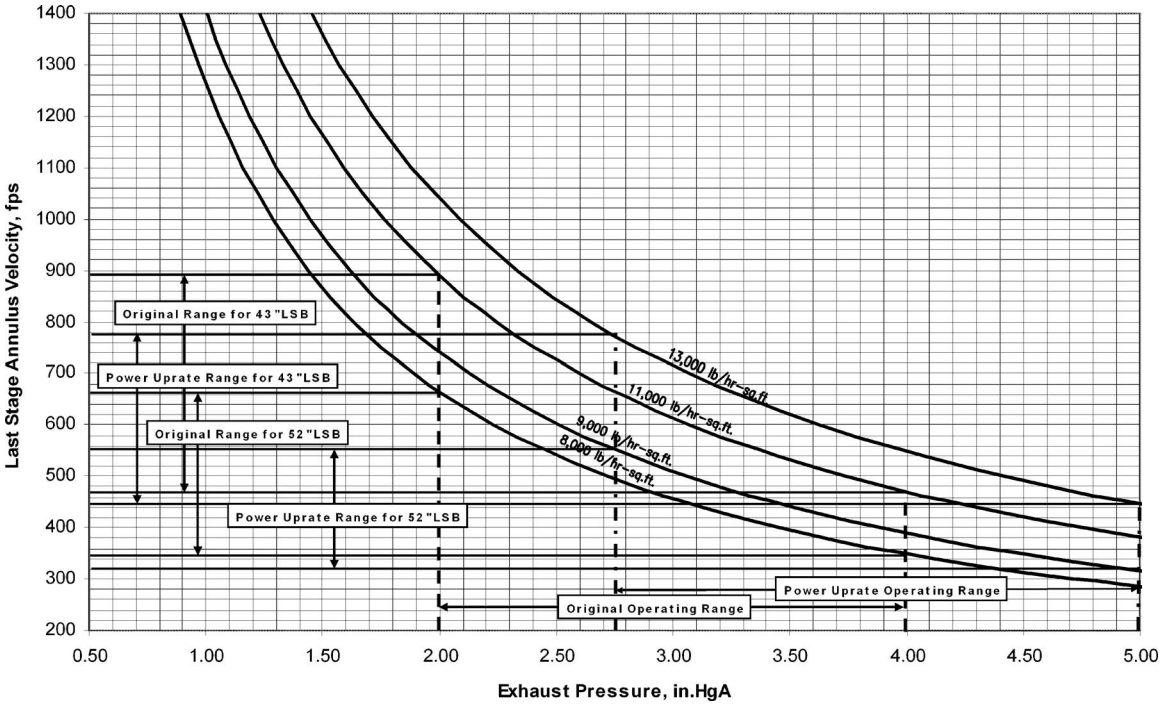


Fig. 3 Performance of 43 in. LSB and 52 in. LSB before and after power uprate

increase by about 16%. The condenser operating range shifts to between 2.7 in. Hg (absolute) and 5.0 in. Hg (absolute). In this range, the 43 in. LSBs are being utilized more efficiently while the 52 in. LSBs perform poorer.

- If the unit were to be uprated by 15% and the 52 in. LSB used, the condenser/cooling system would need to be modified/upgraded to better utilize the larger blades. The desired range would be 2.0–4.0 in. Hg (absolute).

Figure 3 illustrates the impact of the operating ranges before and after the power uprate for the two blade sizes.

Conclusions and Recommendations

The paper has discussed the considerations involved in matching the last stage performance to the condenser/cooling system design and operating range for different end loadings and LSB sizes.

The case study utilized a nuclear unit with a once-through cooling system and examined two different blade sizes to keep the analysis simple. While the analysis shown applies to various last stage blade sizes for both once-through and closed cooling systems, the results shown are specific to the case study. For other units, information specific to those units should be considered in the analysis.

The following recommendations are proposed:

- For optimum last stage performance in existing units, evaluate the performance over the entire operating range of the condenser/cooling system and not merely at the condenser design point. Consider potential upgrades/modifications to the condenser/cooling system if the evaluation reveals mismatch or less than optimum performance.
- When power uprates are being considered, also evaluate potential upgrades/modifications to the condenser/cooling system to ensure that the last stage performance is properly matched to the new condenser operating range for optimum performance.
- Where LP turbine rotors are being replaced with fully bladed rotors including new LSB, evaluate the impact of the new last stage on the entire operating range of the existing condenser/cooling system. Reference [9] discusses the performance considerations in replacement of LP turbine rotors.
- When planning on complete LP turbine replacements with new or advanced designs, evaluate the different LSB offerings over the entire condenser operating range. Consider the impact of the last stage upon the condenser/cooling system and evaluate if potential modifications are warranted to the condenser/cooling system to properly match the LSB size to the condenser/cooling system.
- Although the hood loss is considered from the last stage exhaust to the condenser flange connection, pressure losses in the condenser neck and the associated impact on last stage performance can be significant. Consequently, the condenser neck should be part of the evaluation when LP turbine replacements and/or condenser upgrades/modifications are being planned. References [5,10] provide more details on the LP turbine last stage/condenser neck interface.

Nomenclature

A = area, square feet
P = pressure, psia (absolute)
V = velocity, fps
v = specific volume, cft/lb
W = Mass flow, lb/h

Subscripts

an = annulus
B = bowl
isen = isentropic
Sh = Shell

References

[1] Salisbury, J. K., 1974, *Steam Turbines and their Cycles*, Krieger, Huntington, NY.

[2] Steltz, W. G., Evans, D. H., and Desai, K. J., 1978, "Advances in Low Pressure Steam Turbine Thermal Performance," *Paper Presented at the Westinghouse Steam Turbine-Generator Technology Symposium*, Charlotte, NC, Oct. 4–5.

[3] Jacobsen, G., Oeynhausen, H., and Termuehlen, H., 1991, "Advanced LP Turbine Installation at 1300 MW Power Station Unterweser," *Paper Presented at the American Power Conference*, Chicago, IL, Apr. 29-May 1.

[4] Gray, L., Sandhu, S. S., Davids, J., and Southall, L. R., 1989, "Technical Considerations in Optimizing Blade-Exhaust Hood Performance for Low Pressure Steam Turbines," *ASME Paper Presented at the Joint Power Generation Conference*, Dallas, TX, Oct. 22–26.

[5] Owczarek, J. A., Warnock, A. S., and Malik, P., 1989, "A Low Pressure Turbine Exhaust End Flow Model Study," *ASME Paper Presented at the Joint Power Generation Conference*, Dallas, TX, Oct. 22–26.

[6] Baily, F. G., Cotton, K. C., and Spencer, R. C., 1967, "Predicting The Performance of Large Steam Turbine-Generators Operating With Saturated And Superheated Steam Conditions," *Presented at the American Power Conference*, Chicago, IL, Apr. 25–27.

[7] Baily, F. G., Booth, J. A., Cotton, K. C., and Miller, E. H., 1974, "Predicting The Performance of 1800-Rpm Large Steam Turbine-Generators Operating With Light Water-Cooled Reactors," General Electric Company Paper GET-6020.

[8] Sunder Raj, K. S., 2005, "Performance Considerations in Power Uprates of Nuclear Power Plants," ASME Paper PWR2005-50003.

[9] Sunder Raj, K. S., 2005, "Performance Considerations in Replacement of Low-Pressure Turbine Rotors for Nuclear Power Plants—A Case Study," ASME Paper PWR2005-50005.

[10] General Electric Guidelines No. 222A4446, 1969, "Summary of Condenser Neck Recommendations for Use by Architect-Engineer and Condenser Manufacturer."

A Coal-Fired Power Plant With Zero-Atmospheric Emissions

Joel Martinez-Frias

Salvador M. Aceves
e-mail: saceves@llnl.gov

J. Ray Smith

Lawrence Livermore National Laboratory,
7000 East Avenue, L-644,
Livermore, CA 94551

Harry Brandt
Clean Energy Systems, Inc.,
11330 Sunco Drive, Suite A,
Rancho Cordova, CA 95670
e-mail: hbrandt@cleanenergysystems.com

This paper presents the thermodynamic and cost analysis of a coal-based zero-atmospheric emissions electric power plant. The approach involves an oxygen-blown coal gasification unit. The resulting synthetic gas (syngas) is combusted with oxygen in a gas generator to produce the working fluid for the turbines. The combustion produces a gas mixture composed almost entirely of steam and carbon dioxide. These gases drive multiple turbines to produce electricity. The turbine discharge gases pass to a condenser where water is captured. A stream of carbon dioxide then results that can be used for enhanced oil recovery or for sequestration. The term zero emission steam technology is used to describe this technology. We present the analysis of a 400 MW *electric power plant. The power plant has a net thermal efficiency of 39%. This efficiency is based on the lower heating value of the coal, and includes the energy necessary for coal gasification, air separation, and for carbon dioxide separation and sequestration. This paper also presents an analysis of the cost of electricity and the cost of conditioning carbon dioxide for sequestration. Electricity cost is compared for three different gasification processes (Texaco, Shell, and Koppers-Totzek) and two types of coals (Illinois 6 and Wyodak). COE ranges from* 5.95 ¢/kW h to 6.15 ¢/kW h*, indicating a 3.4% sensitivity to the gasification processes considered and the coal types used.* [DOI: 10.1115/1.2771255]

Introduction

More than half of the electricity consumed in the United States (52%) is generated from coal. This fraction is likely to increase because the U.S. has about 25% (275×10^9 tons) of the world's coal reserves. Pollution from coal-fired power plants is a pressing environmental problem and the emission of carbon dioxide is of increasing concern due to global warming.

Carbon dioxide capture and geologic storage offer a new option for reducing greenhouse gas emissions that can complement the current strategies of improving energy efficiency and increasing the use of nonfossil energy resources.

Production of electric power from coal with zero-atmospheric emissions is a goal of the FutureGen Program of the U.S. Department of Energy (DOE) [1]. A decade ago, such a concept would not have been viable. However, recent research [2–11] has addressed technical and economic issues associated with the concept, making it a viable option.

In this paper, we describe a coal-fired zero emission steam technology (ZEST) power plant concept. This power plant uses a Rankine cycle to drive three turbines connected in series. However, unlike conventional steam power plants, the plant does not use a boiler to generate steam. Use of a boiler presents two disadvantages to the efficiency of the Rankine cycle. First, the maximum cycle temperature is limited by the maximum temperature that boiler components can withstand. Second, 10–15% of the energy in the fuel is lost with the exhaust gases that are vented to the atmosphere.

In this study, the turbine working fluid is produced in a gas generator by the stoichiometric combustion of synthetic gas (syngas) and oxygen. Hence, the maximum operating temperature of the Rankine cycle is no longer controlled by the maximum operating temperature of a boiler. Rather, the maximum operating temperature that the turbines can withstand becomes the efficiency-limiting temperature.

Contributed by the Advanced Energy Systems Division of ASME for the publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received September 15, 2006; final manuscript received May 29, 2007; published online February 21, 2008. Review conducted by Dilip R. Ballal. Paper presented at the 2003 ASME International Mechanical Engineering Congress (IMECE2003), Washington, DC, November 15–21, 2003.

The adiabatic flame temperature of the stoichiometric combustion of syngas is too high for today's turbine technology. Therefore, in the gas generator, water is premixed with the syngas and oxygen before the mixture enters the combustion chamber. In addition, the gas generator [8–11] has several sections in which water is added to the combustion products to bring the gas temperature to a level acceptable to available turbines.

The turbine discharge gases pass to a condenser where water is captured as liquid and gaseous carbon dioxide is pumped from the system. The carbon dioxide can be compressed for enhanced recovery of oil or coal-bed methane, or the compressed carbon dioxide can be injected for sequestration into a subterranean formation. The technology described in this paper is the subject of several U.S. patents [12–21].

The methodology followed in this paper was previously applied to a natural gas power plant by the same authors [11]. In the current analysis, the fuel is coal, processed by a gasifier to produce a synthetic gaseous fuel (syngas). This publication also presents an analysis of the cost of electricity (COE).

Several coal gasification technologies have been developed and applied to integrated gasification-combined-cycle (IGCC) power plants [22–24]. In this paper, gasification technologies by Texaco [25,26], Shell [27], and Koppers-Totzek [28] are analyzed to study their influence on performance and cost of electricity.

Power Plant Configuration

Figure 1 presents the power plant configuration analyzed in this paper. The power plant has four major sections: (1) coal gasification and syngas compression, (2) air separation and oxygen (O_2) compression, (3) power generation, and (4) carbon dioxide (CO_2) separation and sequestration. Each of these sections consists of multiple components, as shown in Fig. 1. For this analysis, the plant is assumed to operate on syngas that is combusted with oxygen. The syngas is produced in a coal gasification plant, and it is compressed to the inlet pressure of the gas generator: 10,200 kPa (1480 lb./in.^{2}), point 22, Fig. 1. Part of the syngas is compressed to a pressure of 2,140 kPa (310 lb./in.^{2}) for Reheater 1 that is installed between the high-pressure turbine and the intermediate-pressure turbine (point 5). The compression system for the syngas consists of four compressors (Compressors 1–4) and three intercoolers (Intercoolers 1–3). Oxygen is generated in

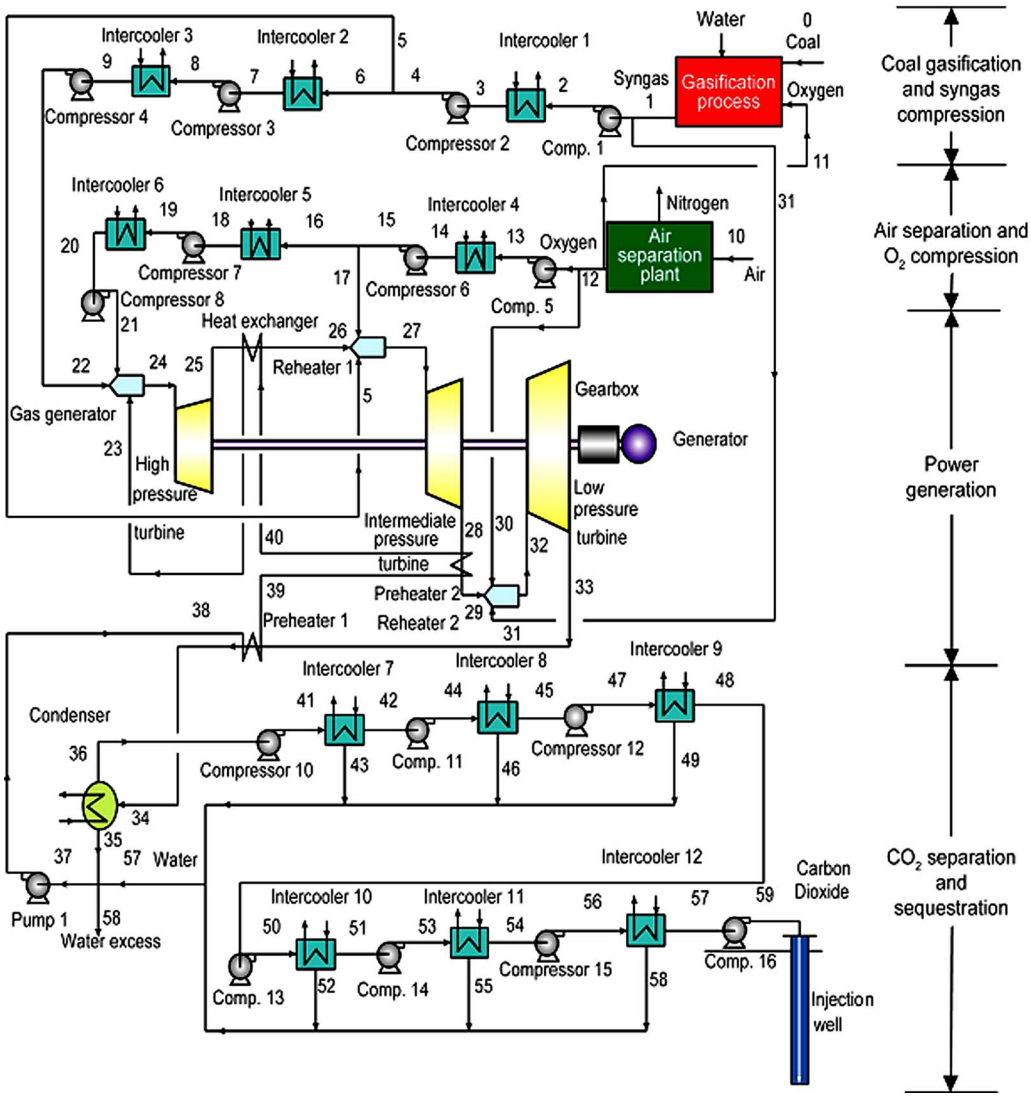


Fig. 1 Schematic diagram of the zero-atmospheric emissions 400 MW coal power plant

an air separation plant and is compressed to feed the gas generator and two reheaters. The oxygen compression system consists of four compressors (Compressors 5–8) and three intercoolers (Intercoolers 4–6).

The water for this cycle is generated by the cycle itself. The gasification process consumes 3.6 kg/s of water, and combustion produces 18.8 kg/s, resulting on a net water surplus of 15.2 kg/s. The water leaving the condenser is heated in Preheaters 1 and 2 before the water is injected into the combustion products in the gas generator. These preheaters increase the efficiency of the cycle by improving heat recovery.

The preheaters are located in the discharge lines of both the intermediate-pressure turbine and the low-pressure turbine. The preheaters heat the water that is routed from the condenser to the gas generator where the water is evaporated to cool the combustion products in the gas generator. If the water were not preheated, a smaller amount of water would be required to cool the gases in the gas generator. However, taking thermal energy out of the discharge of the drive gas from the low-pressure turbine reduces the energy that is delivered to the condenser. As a result, less heat is transferred to the condenser cooling water. This reduced condenser cooling water heat loss increases plant efficiency. The location of the preheaters, the amount of heat removed from the turbine drive gas, and the temperature of the cooling water entering the gas generator all affect cycle efficiency. How this increase in efficiency is obtained is not a priori clear but is determined from optimization studies of the entire cycle.

Combustion products from the gas generator are delivered to the high-pressure turbine (point 24) where the mixture of steam and carbon dioxide expands, thereby producing power in the turbine and electrical generator system. The mixture consists of a 0.92 mass fraction of steam and 0.08 mass fraction of carbon dioxide. After the steam and carbon dioxide mixture leaves the high-pressure turbine, the mixture interchanges heat with the water that comes from Preheater 2 and the enthalpy of the mixture increases. Then the mixture goes to Reheater 1 (point 26). The reheater increases the temperature of the mixture before it enters the intermediate-pressure turbine. Composition at the intermediate-pressure turbine consists of a 0.72 mass fraction of steam and a 0.28 mass fraction of carbon dioxide. After leaving the intermediate-pressure turbine, the mixture goes through Preheater 2 and then it goes to a second reheater before entering the low-pressure turbine for its final expansion (point 32). The exhaust from the low-pressure turbine flows through Preheater 1 to preheat the water that was separated from the turbine working fluid in the condenser.

Most of the water that is generated in the cycle is separated from the turbine working fluid mixture in the condenser. Liquid

water is extracted from the condenser by Pump 1 and is recycled to the system. The water temperature is increased in Preheaters 1 and 2 before the water goes to the gas generator (point 23) to control the temperature of the combustion products.

A mixture consisting primarily of carbon dioxide, but containing a substantial amount of moisture, is extracted from a port (point 36) at the top of the condenser. The carbon dioxide with the remaining moisture from the condenser is then delivered to several compressors and intercoolers to obtain high-pressure carbon dioxide with almost no moisture. The compression-sequestration system consists of seven compressors (Compressors 10–16) and six intercoolers (Intercoolers 7–12).

Analysis

The power plant system consists of an oxygen separation plant, a coal gasification plant, a gas generator, three turbines, two reheaters, a condenser, 15 compressors, a pump to recirculate the water from the condenser to the gas generator, a pump for the condenser cooling water, 12 intercoolers, two preheaters, a heat exchanger, and an electric generator. Energy and mass conservation laws are applied to every system component. The equations used to describe the power plant components (more than 1200 equations) are solved simultaneously in a computer code [29] to analyze plant efficiencies. Individual system components are described next.

Oxygen Separation Plant. For this analysis, the power to operate the oxygen separation plant, 0.22 kW h/kg of oxygen, was obtained from data presented in the literature [30]. This value is similar to the 0.214 kW h/kg given in Ref. [31] (Table 4-1 Case 3A, IGCC with CO_2 removal) for a cryogenic air separation plant. Advances in oxygen separation are expected to reduce this power to 0.16 kW h/kg when ion transport membrane (ITM) technology matures [32].

Coal Gasification Plant. Three power plant configurations are compared. Each configuration uses a different oxygen-blown coal gasifier (Texaco [25,26], Shell [27], and Koppers-Totzek [28]). Efficiency of the power plant and cost of electricity are compared for these three gasification processes operating on Illinois 6 coal and Wyodak coal.

Gas Generator and Reheater. Syngas and oxygen are combusted in the gas generator to produce the turbine working fluid. The temperature of the combustion products of syngas with oxygen is controlled by adding water to the combustion products in the gas generator. The mass flow rate of water into the gas generator depends on the desired inlet temperature of the working fluid for the high-pressure turbine.

A reheater burns syngas with oxygen and mixes the combustion products with the high-pressure turbine exhaust. The hot mixture can then be efficiently expanded in the intermediate-pressure turbine. A second reheater is installed between the intermediate-pressure and low-pressure turbines.

In the gas generator and the reheaters, the absolute enthalpy (including both sensible enthalpy and enthalpy of formation) of the products is equal to the absolute enthalpy of the reactants (assuming an adiabatic process). Complete combustion is considered in the gas generator and in the reheaters.

Turbines. Turbines are modeled by the equation of isentropic efficiency [33]. The turbine isentropic efficiencies for the high-pressure turbine, the intermediate-pressure turbine, and the low-pressure turbine are assumed to be 90%, 91%, and 93%, respectively (see Table 1). The efficiency of the high-pressure turbine takes into account the use of short blades; the efficiency of the intermediate-pressure turbine takes into account the blade cooling losses. These efficiencies compare to values of 93% used by Bannister et al. [34], 85% used by Bolland et al. [35], and 93% by Aoki et al. [36].

Table 1 Parameters used in the simulation of the ZEST plant

System parameters	Value
Preheater effectiveness	0.85
Condenser effectiveness	0.90
Intercooler effectiveness	0.85
Ambient temperature	288 K
Isentropic efficiency of the high-pressure turbine	90%
Isentropic efficiency of the intermediate-pressure turbine	91%
Isentropic efficiency of the low-pressure turbine	93%
Isentropic efficiency of the compressors	85%
Efficiency of the water pump	85%
Efficiency of the electric generator	98%

Heat Exchangers. We use an effectiveness equation to determine the performance of the heat exchangers (intercoolers, preheaters, and condenser). The heat exchanger effectiveness is defined as the ratio of the actual rate of heat transfer in a given heat exchanger to the maximum possible rate of heat exchange.

This analysis assumes an effectiveness of 85% for intercoolers and preheaters (see Table 1) [35]. The temperature of the environment and cooling water is assumed to be 288 K (59 °F) to be consistent with the environment temperature used in the analysis of combined cycle plants.

Compressors. Compressors are modeled by the equation of isentropic efficiency [33] defined as the ratio of power needed to compress gases in an isentropic process and the actual power needed in the compression of the gases. The compressors were assumed to have an isentropic efficiency 85%. Previous researchers [34–36] have used compressor efficiencies in the range of 85–89%.

Water Recirculation Pump. The isentropic efficiency of the water pump is assumed to be 85%. Previous researchers [34–36] have used pump efficiencies in the range of 85–99%.

Computational Assumptions. Complete combustion was assumed in the gas generator. This is justified because the gas generator uses platelet injectors that provide extremely uniform mixing of oxygen, fuel, and water. In addition, bench-scale tests recently made at the University of California at Davis show an absence of hydrocarbons in the exhaust and only minor concentrations of carbon monoxide. These results are in agreement with

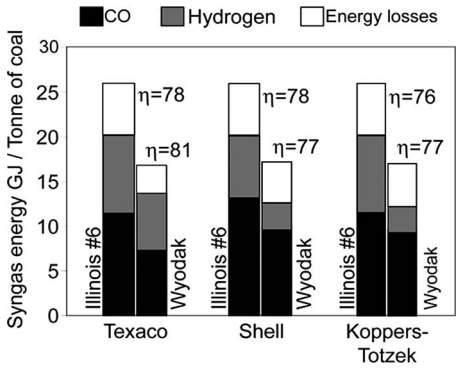


Fig. 2 Syngas energy per ton of coal for three different gasification processes (Texaco, Shell, and Koppers-Totzek) and two types of coal (Illinois and Wyodak). The figure also shows efficiency (η) for each gasification process.

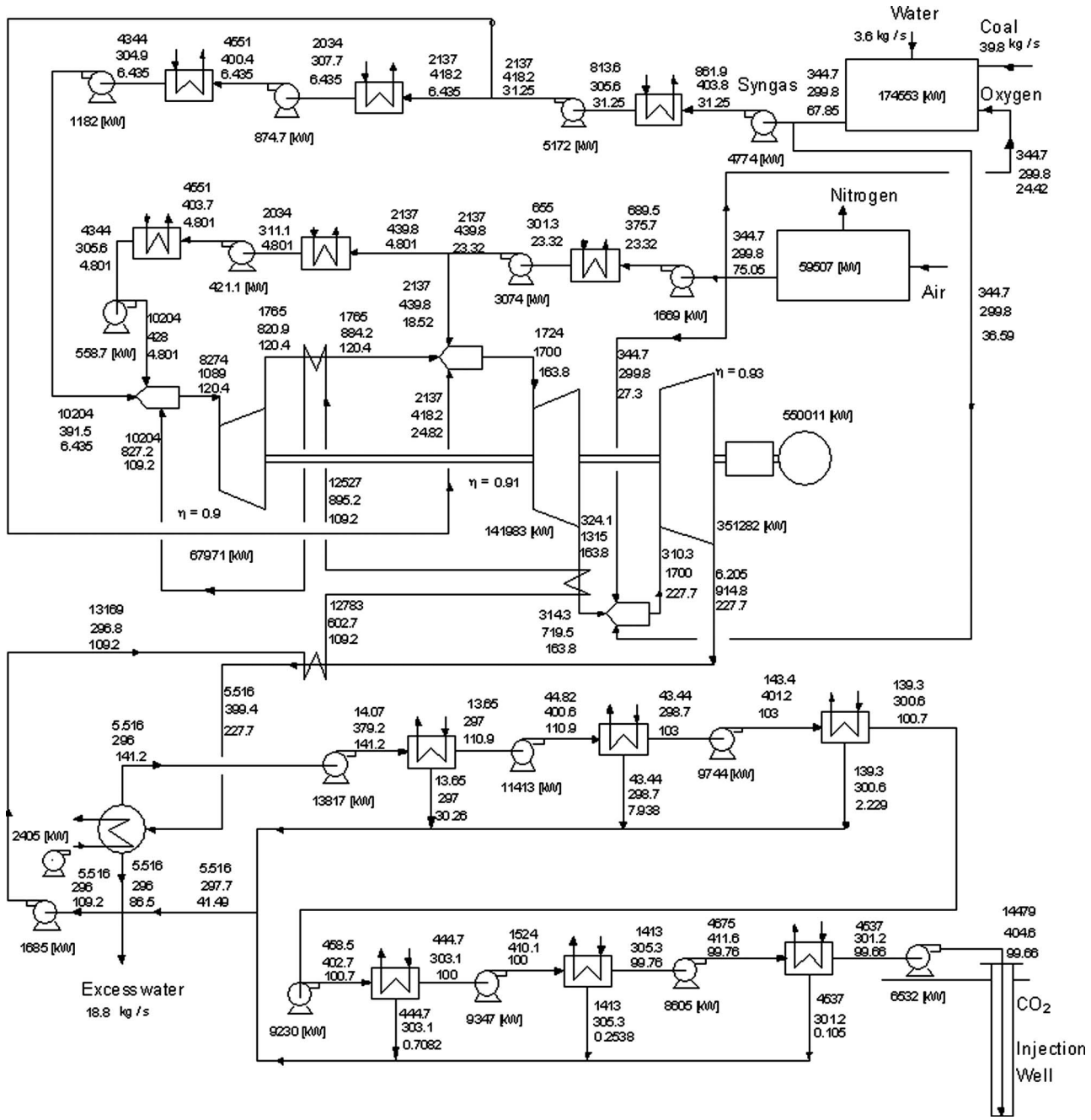


Fig. 3 Zero-atmospheric emissions power plant data for base case analysis. Notation: pressure, kPa (top); temperature, K (middle); mass flow, kg/s (lower). Input power=1,030 MW LHV, thermal energy consumed in the gasification process =175 MW. Electric power generated=550 MW, electric parasitic power of pumps, compressors, and gasification plant =150 MW. Net electric power=400.0 MW. Net LHV thermal efficiency=39%.

predictions based on the use of the chemical kinetics code CHEMKIN-II [37,38].

Pressure drops are considered negligible in all pipelines. Heat transfer losses to the environment from lines connecting plant components are also considered to be negligible. Heat losses to the environment from heat exchangers are neglected. A commercial oxygen separation plant for this type of application would produce an oxygen stream that contains about 1–2% argon. In this analysis, the contribution of argon in the turbine working fluid is neglected. Addition of argon to the working fluid mixture of steam and carbon dioxide makes the iterative computations more complex. Studies show that the noncombustible gas does not change significantly the efficiency calculations, but primarily change the output power due to the change in molecular weight of the working fluid.

The system of equations is solved with an iterative equation solver [29] by using computer-based tables of properties for all the substances involved (water [39], carbon dioxide [40], oxygen [41], carbon monoxide [41], and hydrogen [42]). Table 1 shows the values of the system parameters used in the analysis.

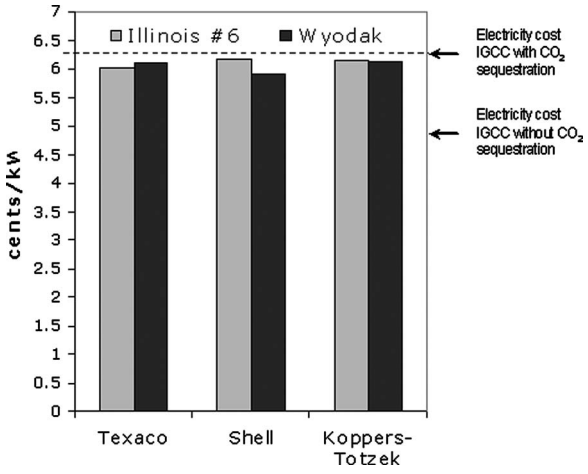


Fig. 4 ZEST electricity cost for three different gasification processes (Texaco, Shell, and Koppers-Totzek) and two types of coal (Illinois 6 and Wyodak). COE for IGCC with and without CO_2 sequestration is also included.

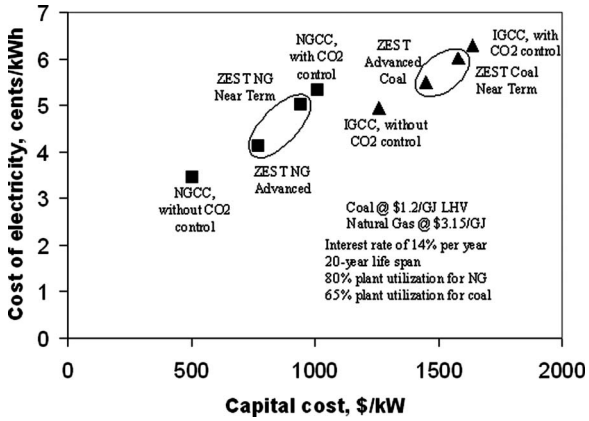


Fig. 5 COE as a function of capital cost for ZEST and other technologies for comparison

Cost Analysis. A method for assessing the economics of a power plant is to calculate the unit cost of electricity (COE) produced by the plant [43]. To determine this cost, the following information is used:

A. unit capital cost ($/kW h)
B. plant net thermal efficiency
C. fuel cost ($/kW h)
D. operating and maintenance cost ($/kW h).

If income from plant by-products is excluded to simplify the calculations, the cost of electricity is given by: COE=A+C+D, where C is a function of B, and where D is conservatively estimated to be $D=0.15\times(A+C)$. Plant capital cost was based on 65% utilization, 20 year life span, and 14% interest rate [31].

Results

Figure 2 shows energy of the syngas per ton of coal for the three gasification processes: Texaco [25,26], Shell [27], and Koppers-Totzek [28], and for the two types of coal: Illinois 6 and Wyodak. Black in every column represents energy from CO, gray represents energy from hydrogen, and white represents energy losses. The three gasification processes considered in this analysis have efficiencies ranging from 76% to 81% [25–28]. The Texaco process using either Illinois 6 or Wyodak coal is the most efficient process.

Figure 3 shows the results for the base case power plant analysis. Figure 3 shows pressures, temperatures, and mass flow rates for this power plant at more than 50 locations. In Fig. 3, power is given in kilowatts, pressure in kilopascals, temperatures in Kelvin, and mass flow rates in kg/s. The base case assumes a high-pressure turbine with an inlet temperature of 1,089 K (1500°F) and an intermediate-pressure and low-pressure turbines running at a 1700 K (2600°F) inlet temperature.

The base case analysis considers the Shell gasification process using Illinois 6 coal [27]. The LHV of the Illinois 6 coal is 25.8 MJ/kg [27]. The dry compositions of the syngas produced in the shell gasification process are 31.6% of H_2, 64% of CO, and 4.4% of CO_2.

The power plant configuration has a net thermal efficiency of 39% and a net electrical output of 400 MW. The net thermal efficiency is based on the lower heating value of coal, and includes the energy required to separate oxygen from air, the thermal energy consumed in the gasification process, and the energy required to compress the carbon dioxide for underground sequestration to a pressure of 14.5 MPa (2100 lb./in.^{2}). This sequestration pressure is sufficient to inject the carbon dioxide either into an oil zone for enhanced oil recovery or into a subterranean aquifer at an approximate depth of 1200 m (3937 ft).

In the current analysis, the power plant and the air separation plant were treated as individual units. By integrating the air separation plant and the power plant and by optimizing the performance of the combined units, a higher overall efficiency and therefore lower cost of electricity could be obtained [44].

Chiesa and co-workers [45–47] studied a zero-atmospheric emissions power plant cycle with oxygen and gasification plants. The turbine working fluid in the cycle of Chiesa and co-workers [45–47] differs from the one presented here because in their power plant CO_2 is recycled rather than H_2O.

Figure 4 shows the COE in ZEST power plants for the three different gasification processes and two types of coal. COE varies from 5.95 ¢/kW h for the Shell process using Illinois 6 to 6.15 ¢/kW h coal for the Shell process using Wyodak coal. This represents a 3.4% variation for three different gasification processes and two types of coal. This is an indication that the gasification process and the type of coal have little influence on the electricity cost.

A comparison of the COE for electric power plants operating on both syngas derived from coal and on natural gas is presented in Fig. 5. Squares indicate COE from natural gas and triangles indicate COE from coal. The analysis done for this figure uses a natural gas price of US$3.15/GJ (US$3.33/MM Btu) and coal price of US$1.2/GJ (US$1.25/MM Btu), capital charges of 14% per year, 80% utilization for natural gas power plants, and 65% utilization for coal power plants. The figure includes electricity costs for plants with and without exhaust gas sequestration.

The numbers for the natural gas combined cycle (NGCC) with and without sequestration and for the IGCC were taken from an EPRI report [31]. Figure 5 compares COE for NGCC and IGCC technologies, with and without sequestration of exhaust gases, with the COE of the ZEST. The lowest electricity cost is US$0.035/kW h for NGCC plant without exhaust gas sequestration (Ref. [31], Case 1C: Base NGCC without CO_2 removal, class F gas turbine). The COE for NGCC with CO_2 removal is US$0.054/kW h (Ref. [31], Case 1A: Base NGCC with CO_2 removal, class F gas turbine). COE for a ZEST natural gas using near term technology is US$0.050/kW h, which is slightly lower than the electricity cost for NGCC with CO_2 removal. An analysis of the COE is also included in Fig. 5 for a ZEST natural gas plant using advanced technology. The analysis includes use of high temperature turbines, membrane oxygen separation, and optimization of the combined power plant and oxygen plant [44]. COE for advanced ZEST natural gas technology is US$0.042/kW h.

ZEST power plants would use transpiration-cooled turbines in which the turbine blades would be cooled with steam in a range from 450 K (350°F) to 480 K (404°F) [48]. Because steam cooling is more efficient than air cooling, the blade surface temperature would be lower than with air cooling. ZEST near term conditions represent turbines operating at drive gas temperatures of 1089 K (1500°F) for the high-pressure turbine and 1700 K (2600°F) for both the intermediate- and high-pressure turbines. The corresponding temperatures for the advanced ZEST plants would be 1089 K (1500°F), 1922 K (3000°F), and 1922 K.

The COE for IGCC without CO_2 removal is US$0.05/kW h (Ref. [31], Case 1B, E-Gas™ IGCC, H class turbine without CO_2 removal). COE for IGCC with CO_2 removal is US$0.063/kW h (Ref. [31], Case 3A; E-Gas™ IGCC, H class turbine with CO_2 removal). COE for ZEST coal near term technology is US$0.060/kW h, which is a slightly lower than the electricity cost for IGCC with CO_2 removal. The advanced case has an estimated COE of US$0.055/kW h [44].

Another advantage of the ZEST technology over combined cycle technology is the lower cost to condition CO_2 for sequestration of US$4.6/ton versus US$17.6/ton [31]. This lower CO_2 conditioning cost could provide additional revenue for ZEST plants where the CO_2 could be used for enhanced oil or coal-bed methane recovery, or could be sold as an industrial by-product.

Conclusions

This paper presents a thermodynamic analysis of a zero-atmospheric emissions power plant. The simulation considers the compression process of syngas and oxygen to feed the gas generator and the reheaters, a Rankine cycle with three turbines, and the carbon dioxide separation and sequestration processes.

The analysis predicts a 39% net thermal efficiency in a 400 MW zero-atmospheric emissions power plant that can be constructed with high temperature turbine technology in use today. The net thermal efficiency is based on the lower heating value of coal, and includes the energies required to separate oxygen from air, to make syngas from coal and the energy required to compress the carbon dioxide for underground sequestration.

The results indicate that the zero emission power plant will produce electricity at lower cost than competing technologies that include carbon sequestration. These results apply for coal-fired as well as for natural gas fired power plants. Further cost reductions may be obtained through integration of the coal gasification plant, the oxygen plant, and the power generation cycle [44].

Acknowledgment

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48. Some of the concepts in this publication are based on the intellectual property of Clean Energy Systems, Inc.

References

[1] DOE, 2003, "FutureGen," U.S. Department of Energy Announcement, February 2, 2003.
[2] Bilger, R. D., 1999, "Zero Release Combustion Technologies and the Oxygen Economy," *Proceedings of the Fifth International Conference on Technologies and Combustion for a Clean Environment*, Lisbon, Portugal, pp. 167–176.
[3] Falk-Pederson, O., and Dannström, H., 1997, "Separation of Carbon Dioxide From Offshore Gas Turbine Exhaust," Energy Convers. Manage., **38**, pp. S81–S86.
[4] Herzog, H., Drake, E., and Adams, E., 1997, "CO_2 Capture, Reuse, and Storage Technologies for Mitigation Global Climate Change," U.S. Department of Energy, Final Report No. DE-AF22–96PC01257.
[5] Stevens, S. H., and Gale, J., 2000, "Geologic CO_2 Sequestration May Benefit Upstream Industry," Oil Gas J., **98**, pp. 40–44.
[6] Wildenborg, T., 2000, "Costs of CO_2 Sequestration by Underground Storage," Greenhouse Issues, International Energy Agency Greenhouse Gas R&D Programme, Report No. 47, pp. 2–4.
[7] Wong, S., and Gunter, B., 1999, "Testing CO_2-Enhanced Coalbed Methane Recovery," Greenhouse Issues, International Energy Agency Greenhouse Gas R&D Programme, Report No. 45, pp. 1–3.
[8] Anderson, R. E., Brandt, H., Mueggenburg, H. H., Taylor, J., and Viteri, F., 1998, "A Power Plant Concept Which Minimizes the Cost of Carbon Dioxide Sequestration and Eliminates the Emission of Atmospheric Pollutants," *Proceedings of the Fourth International Conference on Greenhouse Gas Control Technologies*, P. Reimer, B. Eliasson, and A. Wokaun, eds., Interlaken, Switzerland, Pergamon, London.
[9] Anderson, R. E., Brandt, H., Doyle, S. E., Mueggenburg, H. H., Taylor, J., and Viteri, F., 2000, "A Unique Process for Production of Environmentally Clean Electric Power Using Fossil Fuels," *Proceedings of the Eighth International Symposium Transport Phenomena and Dynamics Rotating Machinery (ISROMAC-8)*, Pacific Center of Thermal-Fluids Engineering, Honolulu, HI.
[10] Smith, J. R., Surles, T., Marais, B., Brandt, H., and Viteri, F., 2000, "Power Production with Zero Atmospheric Emissions for the 21st Century," *Proceedings of the Fifth International Conference on Greenhouse Gas Control Technologies*, Cairns, Queensland, Australia.
[11] Martinez-Frias, J., Aceves, S. M., Smith, J. R., and Brandt, H., 2004, "Thermodynamic Analysis of Zero-Atmospheric Emissions Power Plant," ASME J. Eng. Gas Turbines Power, **126**, pp. 2–8.
[12] Viteri, F., 1993, "Turbomachinery for Modified Ericsson Engines and Other Power/Refrigeration Applications," U.S. Patent No. 5473899.
[13] Viteri, F., 1997, "Turbocharged Reciprocation Engine for Power and Refrigeration Using the Modified Ericsson Cycle," U.S. Patent No. 5590528.
[14] Viteri, F., 1997, "Clean Air Engines for Transportation and Other Power Applications," U.S. Patent No. 5680764.
[15] Beichel, R., 1998, "Reduced Pollution Hydrocarbon Combustion Gas Generator," U.S. Patent No. 5709077.
[16] Beichel, R., 1998, "Reduced Pollution Power Generation System," U.S. Patent No. 5715673.
[17] Beichel, R., 1999, "Reduced Pollution Power Generation System Having Multiple Turbines and Reheater," U.S. Patent No. 5956937.
[18] Beichel, R., 1999, "Reduced Pollution Hydrocarbon Combustion Gas Generator," U.S. Patent No. 5970702.
[19] Viteri, F., Taylor, J. P., Brandt, H., and Anderson, R. E., 2000, "Hydrocarbon Combustion Power Generation System with CO_2 Sequestration," U.S. Patent No. 6170264.
[20] Mueggenburg, H. H., 2001, "Steam Generator Injector," U.S. Patent No. 6206684.
[21] Viteri, F., 2001, "Clean Air Engines for Transportation and Other Power Applications," U.S. Patent No. 6247316.
[22] Cargill, P., DeJonghe, G., Howsley, T., Lawson, B., Leighton, L., and Woodward, M., 2001, "Piñon Pine IGCC Project: Final Technical Report to the Department of Energy," DOE Report No. DE-FC21-92MC29309.
[23] Sundin, U., 1994, "The Puertollano IGCC Project, a 335 MW Demonstration Power Plant for the Electricity Companies in Europe," *Proceedings of the 13th EPRI Conference on Gasification Power Plants*, San Francisco, CA.
[24] Bechtel Power Corporation, and Radian Corporation, 1986, "Cool Water Coal Gasification Program, Fourth Annual Progress Report," EPRI Interim Report No. AP-4832.
[25] Tsatsaronis, G., and Tawfik, T., 1990, "Performance Comparisons of Integrated Gasification-Combined Cycle Power Plants," *Proceedings of the 25th Intersociety Energy Conversion Engineering Conference*, Reno, NV, Vol. 5, pp. 490–494.
[26] Cook, C. S., Corman, J. C., and Todd, D., M., 1995, "System Evaluation and LBTU Fuel Combustion Studies for IGCG Power Generation," ASME J. Eng. Gas Turbines Power, **117**, pp. 673–677.
[27] McCullough, G. R., van der Burgt, M. J., and Waller, J., 1982, "Shell Coal Gasification Process," Shell Development Company Westhollow Research Center, Houston, TX.
[28] Farnsworth, J. F., 1978, "Carbonization of Coal and Gas Making," Marks' *Standard Handbook for Mechanical Engineers*, McGraw-Hill, New York.
[29] Klein, S. A., and Alvarado, F. L., 2002 "Engineering Equation Solver," F-CHART Software, Madison, WI.
[30] Kobayashi, H., and Prasad, R., 1999, "A Review of Oxygen Combustion and Oxygen Production Systems," Praxair Technology, Inc.
[31] Electric Power Research Institute (EPRI), 2000, "Evaluation of Innovative Fossil Fuel Power Plants with CO_2 Removal," EPRI, Palo Alto, CA, U.S. Department of Energy—Office of Fossil Energy, Germantown, MD, and U.S. Department of Energy/NETL, Pittsburgh, PA, p. 1000316.
[32] Armstrong, P., Sorensen, J., and Foster, T., 2003, "ITM Oxygen: An Enabler for IGCC," *Proceedings of the Gasification Technologies Conference*, San Francisco, CA.
[33] Wilson, D. G., and Korakianitis, T., 1998, *The Design of High-Efficiency Turbomachinery and Gas Turbines*, Prentice-Hall, Englewood Cliffs, NJ.
[34] Bannister, R. L., Newby, R. A., and Yang, W. C., 1999, "Final Report on the Development of a Hydrogen-Fueled Combustion Turbine Cycle for Power Generation," ASME J. Eng. Gas Turbines Power, **121**, pp. 38–45.
[35] Bolland, O., Kvamsdal, H. K., and Boden, J. C., 2001, "A Thermodynamic Comparison of the Oxy-Fuel Power Cycles Water-Cycle, Gratz-Cycle and Matiant-Cycle," *Proceedings of the International Conference of Power Generation and Sustainable Development*, Liège, Belgium.
[36] Aoki, S., Uematsu, K., Suenaga, K., Mori, H. H., and Sugishita, H., 1998, "A Study of Hydrogen Combustion Turbines," *Proceedings of the International Gas Turbines and Aeroengines Congress and Exhibition*, Stockholm, Sweden.
[37] Kee, R. J., Rupley, F. M., and Miller, J. A., 1991, "Chemkin-II: A Fortran Chemical Kinetics Package for the Analysis of Gas Chemical Kinetics," Sandia National Laboratories, Report No. SAND89–8009B.

[38] Bowman, C. T., Frenklach, M., Wang, H., Goldberg, M., Smith, G. P., Golden, D. M., Hanson, R. K., Davidson, D. F., Gardiner, W. C., Jr., and Lissianski, V., 1997, "GRI-MECH2.11—An Optimized Detailed Chemical Reaction Mechanism for Natural Gas Combustion and NO Formation and Reburning," *Proceedings of the American Institute Chemical Engineers*, Los Angeles, CA, AIChE, New York.
[39] IAPWS, 1996, "Release on the IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for Generaland Scientific Use," IAPWS Secretariat, http://www.iapws.org/release. htm
[40] Span, R., and Wagner, W., 1996, "A New Equation of State for Carbon Dioxide Covering the Fluid Region from the Triple-Point Temperature to 1100 K at Pressures up to 800 MPa," J. Phys. Chem. Ref. Data, **25**, pp. 1509–1596.
[41] Sonntag, R. E., and Van Wylen, G., J, 1986, *Fundamentals of Classical Thermodynamics*, Wiley, New York.
[42] McCarty, R. D., 1977, *Hydrogen: Its Technology and Implications* (Hydrogen Properties Vol. III) K. E. Cox, and K. D. Williamson, eds., CRC, Cleveland, OH.
[43] Horlock, J. H., 1995, "Combined Power Plants—Past, Present and Future," ASME J. Eng. Gas Turbines Power, **117**, pp. 608–616.
[44] Marin, O., Bourhis, Y., Perrin, N., Di Zanno, P., Viteri, F., and Anderson, R. E, 2003, "High Efficiency, Zero Emission Power Generation Based on a High-Temperature Steam Cycle," *Proceedings of the 28th International Technical Conference on Coal Utilization and Fuel Systems*, Clearwater, FL.
[45] Chiesa, P., and Lozza, G., 1999, "CO_2 Emission Abatement in IGCC Power Plants by Semiclosed Cycles: PartA—with Oxygen-Blown Combusiton," ASME J. Eng. Gas Turbines Power, **121**, pp. 635–641.
[46] Chiesa, P., and Lozza, G., 1999, "CO_2 Emission Abatement in IGCC Power Plants by Semiclosed Cycles: Part B—With Air-Blown Combustion and CO_2 Physical Absorption," ASME J. Eng. Gas Turbines Power, **121**, pp. 642–649.
[47] Chiesa, P., and Consonni, S., 2000, "Natural Gas Fired Combined Cycles With Low CO_2 Emissions," ASME J. Eng. Gas Turbines Power, **122**, pp. 429–436.
[48] Anderson, R. E., Hoffman, L. C., and Viteri, F., 2004, "Integration of Clean Energy System's Technology With Air Separation Units, Gas Turbines and Steam Turbines into Zero-Emission Power Plants," *Natural Gas Technologies Conference II*, Phoenix, AZ.

Performance Trends of an Air-Cooled Steam Condenser Under Windy Conditions

J. A. van Rooyen[1]
e-mail: javr@sun.ac.za

D. G. Kröger
e-mail: dgk@sun.ac.za

Department of Mechanical Engineering,
University of Stellenbosch,
Private Bag X1,
Matieland 7602, South Africa

Air-cooled steam condensers (ACSCs) are increasingly employed to reject heat in modern power plants. Unfortunately, these cooling systems become less effective under windy conditions and when ambient temperatures are high. A better understanding of the fundamental air flow patterns about and through such ACSCs is essential if their performance is to be improved under these conditions. The present numerical study models the air flow field about and through a particular ACSC. The performance of the fans is modeled with the aid of a novel numerical approach known as the "actuator disc model." Distorted air flow patterns that significantly reduce fan performance in certain areas and recirculatory flows that entrain hot plume air are found to be the reasons for poor ACSC performance. It is found that the reduction in fan performance is the main reason for the poor ACSC performance while recirculation of hot plume air only reduces performance by a small amount. [DOI: 10.1115/1.2771567]

Keywords: air-cooled steam condensers, windy conditions, performance, distorted air flow, recirculation

1 Introduction

Large air-cooled heat exchangers (ACHEs) are used in systems to reject heat to the atmosphere, which acts as the heat sink. They are typically found in power and chemical plants. Kröger [1] describes various configurations of ACHEs which are found in practice. In a power plant, steam is the process fluid. Heat is rejected from the turbine exhaust steam by means of an air-cooled steam condenser (ACSC). The performance of an ACSC consisting of 30 A-frame or fan units as shown in Fig. 1 will be determined under windy conditions. In ACSCs, finned tube bundles (heat exchangers) are arranged in the form of an A-frame (Fig. 2(*a*)) with an apex angle of approximately 60 deg, thereby reducing the plot area and ensuring drainage of the condensate. The air acts as a cooling medium, and is forced through the heat exchanger by means of a fan.

Although it is well known that the performance of ACHEs and ACSCs are reduced under windy conditions, very little useful quantitative information concerning this problem is available. A few limited numerical studies have been done to evaluate aspects of the phenomenon.

Van Aarde and Kröger [2] conducted practical experiments on a full-scale ACSC and they observed that wind had a significant influence on the performance of the particular ACSC. The wind affects the air flow pattern and the static pressure at the inlet to each fan. This in turn affects the volumetric flow rate of the fan. It is evident from this study that wind plays an important role in the performance of an ACSC system.

Goldschagg [3] found that windy conditions have a negative effect on ACSC performance. This was observed at the world's largest ACSC (Matimba power plant) where the turbine performance was reduced measurably during certain windy conditions and occasional turbine trips occurred under extremely gusty conditions.

Salta and Kröger [4] conducted laboratory experiments and found that the effectiveness of a single or multiple fan row ACHE decreases exponentially as the fan deck or platform height is lowered. Their experiments also showed that the volumetric effectiveness of the edge or peripheral fans is always lower than that of the inner fans.

The negative impact of wind on an ACHE, according to Duvenhage and Kröger [5], is twofold:

- Hot plume air recirculation occurs when a fraction of the hot buoyant outlet air (plume) is drawn back into the ACHE inlet. This results in an increase in the effective temperature of the cooling air with a corresponding reduction in heat rejection rate; and
- Fan performance is reduced (decrease in air mass flow rate through the system) due to distorted inlet air flow conditions.

It is important to note that these two effects are not always coupled. One or the other is usually dominant, depending on the ACHE geometry, wind speed, and direction.

Bredell et al. [6] numerically investigated the performance of a two-dimensional forced draft air-cooled steam condenser under windless conditions. The numerical flow field showed that the inlet flow distortions have a negative effect on the fan flow rate due to a number of factors. The type of fan was also shown to have an effect on the performance of a forced draft ACSC.

Coetzee and du Toit [7] conducted a numerical study on the air flow through an ACHE. Despite the many simplifications and assumptions, they claim that numerical modeling offers a valuable tool in the performance evaluation of ACHEs.

Based on previous experimental and numerical research it may be concluded that computational fluid dynamics (CFD) can be effectively employed to investigate the performance of ACHEs (or ACSCs) under windy conditions. The objective of this study is to investigate the trend in performance and thermal-flow characteristics about and through the ACSC shown schematically in Fig. 1 under windy conditions with the aid of a commercial CFD code, FLUENT.

[1]Corresponding author.

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 5, 2006; final manuscript received May 14, 2007; published online February 29, 2008. Review conducted by Dilip R. Ballal.

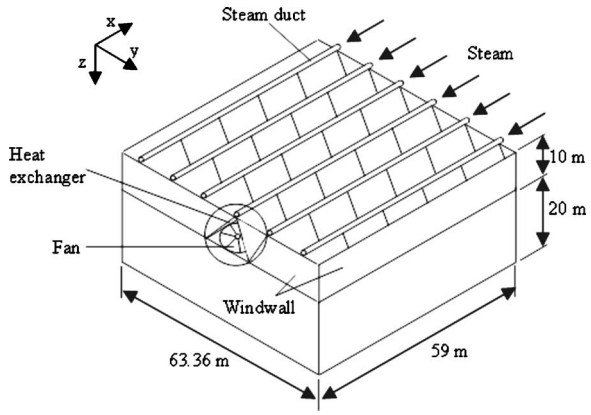


Fig. 1 ACSC plant consisting of 5×6=30 A-frame units

2 Numerical Modeling

As a default, the SIMPLE solution algorithm is implemented in FLUENT, for a steady-state solution. The first-order upwind differencing scheme is used to ensure stability, and turbulence is modeled using the k-ε model, with default model constant settings. It is assumed that the flow is incompressible. Buoyancy effects are taken into account by means of the Boussinesq model.

Due to computational limitations, the detailed air flow through the complete ACSC with all 30 fans will not be solved directly. Instead a simplified global flow field will be solved and data from this model will be exported to each fan to determine its volumetric effectiveness employing the "actuator disk model" approach (Bredell [8]).

2.1 ACSC System Model. In the global flow field analysis the fans in the ACSC are represented by a simple model consisting of a rectangle as shown in Fig. 3 having a uniform inlet and outlet velocity distribution. As buoyancy (non-adiabatic) is taken into account, a constant temperature is also assigned to the uniform velocity boundaries.

The wind speed distribution across a relatively smooth surface is usually expressed as follows

$$v_H = v_{\text{ref}}(H/H_{\text{ref}})^b \tag{1}$$

In the present analysis a value of $b=1/7$ will be assumed. Wind speeds will be specified at a height of 20 m above the ground (fan deck or platform height).

2.2 ACSC Fan Unit Model. Figure 2(*a*) shows the actual ACSC fan unit. Obstacles or appurtances such as supports, beams, inlet screens, electrical fan drives, and ducting were not individually modeled in the simplified model of an ACSC fan unit, as shown in Fig. 2(*b*). The mechanical energy losses (or corresponding pressure drop) resulting from the obstacles, heat exchanger, and outlet losses are taken into account in the heat exchanger model (porous media model in FLUENT) by a viscous and an inertial loss coefficient. Similar to Bredell et al. [8] this simplification can be justified, as the purpose of this study is to investigate fan performance and system performance and not the detailed modeling of the flow in the plenum chamber or at the ACSC outlet. The advantages and disadvantages of various methods of modeling an axial flow fan are discussed by Bredell et al. [8]. The "actuator disk model" of Bredell et al. [8] will be used in this study.

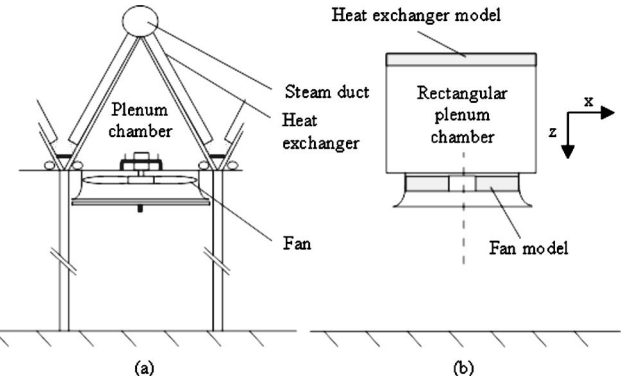


Fig. 2 (*a*) A-frame ACSC fan unit; and (*b*) simplified numerical model of an ACSC fan unit

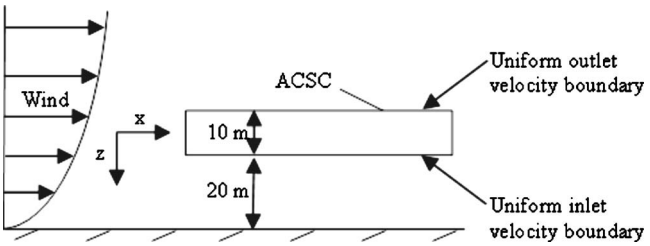


Fig. 3 Schematic of ACSC, side elevation (simplified numerical model)

Refer to Van Rooyen and Kröger [9] for a detailed overview of the numerical modeling.

2.3 System Performance. The procedure followed in determining the volumetric effectiveness of the ACSC is as follows:

- Solve the global flow field about the ACSC with a simplified global flow field numerical model (refer to Fig. 3);
- Use the data from the global flow field numerical model as boundary conditions for the detailed fan unit numerical model (refer to Fig. 2(*b*)), and determine the volume flow rate (V_{fij}) of the specific fan. Each fan has its own numerical model, but only the flow through a selected few fans will be determined (to reduce computational time). The volumetric effectiveness of a fan is defined as the ratio of the actual air volume flow rate through the fan (V_f) divided by the ideal volume flow through the fan (V_{fid}); i.e., that is the case where air flow inlet conditions to the fan are undisturbed;
- By interpolating and extrapolating, the volumetric effectiveness of the fans not modeled is determined approximately;
- The ACSC volumetric effectiveness is determined based on the volumetric effectiveness of the individual fans; and
- The effectiveness of the ACSC is determined based on the mass flow rate through each fan and the air inlet temperature at each fan.

3 Discussion of Results

This section presents the results of the numerical analysis and the flow characteristics of the flow field about an ACSC and its performance when the wind blows in the positive x direction and positive x-y direction (45 deg with respect to the x and y directions) at speeds of 3 m/s, 6 m/s, and 9 m/s, respectively, at an elevation corresponding to the fan platform height above ground level (20 m).

Under non-adiabatic conditions the effectiveness (this includes changes in volume flow and inlet air temperature) of each condenser A-frame or fan unit (Kröger [1]) is

$$e = Q_f/Q_{fid} = 1 - \exp(-UA/m_{afid}c_{pa}) \tag{2}$$

where for this particular case

$$UA/m_{afid}c_{pa} = 44.41/m_{afid}^{0.5443} \tag{3}$$

For the ACSC consisting of thirty fan units operating under ideal conditions (no inlet flow distortions) with an ideal air flow rate of $m_{afid}=717.65$ kg/s per fan unit at ambient pressure of p_a

$=90{,}000\ \mathrm{N/m^2}$ and ambient and steam temperatures of $T_a = 288.75$ K and $T_v = 333.15$ K, respectively, the total heat transfer from the steam to the air is

$$\begin{aligned} Q_{id} &= 30 m_{afid} c_{pa} (T_v - T_a)[1 - \exp(-UA/m_{afid} c_{pa})] \\ &= 30 \times 717.65 \times 1006.609 \times (333.15 - 288.75) \\ &\quad \times [1 - \exp(-44.41/717.650^{0.5443})] \\ &= 683.445 \times 10^6\ \mathrm{W} \end{aligned} \tag{4}$$

During windy periods

$$\begin{aligned} Q &= \sum_{i=1}^{5} \sum_{j=1}^{6} Q_{fij} \\ &= \sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa} (T_v - T_{aij})[1 - \exp(-U_{ij} A/m_{afij} c_{pa})] \\ &= \sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa} (T_v - T_{aij})[1 - \exp(-44.41/m_{afij}^{0.5443})] \end{aligned} \tag{5}$$

where

$$m_{afij} = \rho_a V_{fij}, \quad \mathrm{kg/s} \tag{6}$$

In a steam turbine power generating plant (during steady state or base load operation) the heat that is to be rejected by the ACSC is approximately constant at all ambient temperatures. When the ambient temperature rises or during windy periods when the ACSC effectiveness is reduced there will be a rise in steam temperature and a corresponding turbine back pressure rise in order to reject the same amount of heat. The following equations were employed to determine the rise in steam temperature and corresponding increase in turbine back pressure

$$\begin{aligned} Q = Q_{id} &= 30 m_{afid} c_{pa} (T_v - T_a)[1 - \exp(-UA/m_{afid} c_{pa})] \\ &= \sum_{i=1}^{5} \sum_{j=1}^{6} Q_{fij} \\ &= \sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa} (T_{vw} - T_{aij})[1 - \exp(-U_{ij} A/m_{afij} c_{pa})] \\ &= \sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa} T_{vw}[1 - \exp(-U_{ij} A/m_{afij} c_{pa})] \\ &\quad - \sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa} T_{aij}[1 - \exp(-U_{ij} A/m_{afij} c_{pa})] \end{aligned} \tag{7}$$

or

$$T_{vw} = \frac{30 m_{afid} c_{pa} (T_v - T_a)[1 - \exp(-UA/m_{afid} c_{pa})] + \sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa} T_{aij}[1 - \exp(-U_{ij} A/m_{afij} c_{pa})]}{\sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa}[1 - \exp(-U_{ij} A/m_{afij} c_{pa})]}$$

$$= \frac{30 m_{afid} c_{pa} (T_v - T_a)[1 - \exp(-44.41/m_{afid}^{0.5443})] + \sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa} T_{aij}[1 - \exp(-44.41/m_{afij}^{0.5443})]}{\sum_{i=1}^{5} \sum_{j=1}^{6} m_{afij} c_{pa}[1 - \exp(-44.41/m_{afij}^{0.5443})]}$$

where the temperatures are in °C.

The saturation pressure of the steam corresponding to this temperature is given by (Kröger [1])

$$p_{vw} = 10^z, \ \mathrm{N/m^2} \tag{8}$$

where

$$\begin{aligned} z &= 10.79586(1 - 273.16/T_{vw}) + 5.02808 \log_{10}(273.16/T_{vw}) \\ &\quad + 1.50474 \times 10^{-4}[1 - 10^{-8.29692\{(T_{vw}/273.16)-1\}}] \\ &\quad + 4.2873 \times 10^{-4}[10^{4.76955(1-273.16/T_{vw})} - 1] + 2.786118312 \end{aligned}$$

where T_{vw} is in K.

3.1 ACSC Performance Under Windy Conditions (Wind in the *x* Direction). The following characteristics of the flow field about and through the ACSC are found:

1. At a wind speed of 3 m/s the plume rise angle (angle between plume and vertical, immediately downstream of the ACSC) is approximately 30 deg, 45 deg at 6 m/s and 60 deg at 9 m/s. The latter is shown in Fig. 4.
2. The volumetric effectiveness of fan (4,1) (refer to Fig. 5 for fan location) is reduced due to a low-pressure region and a corresponding distorted air flow pattern under it as is shown in Fig. 6.
3. Figure 7 shows that an increase in the wind speed has a significant effect on the volumetric effectiveness of fan (4,1) and to a lesser extent on fan (6,1). The wind effect on the

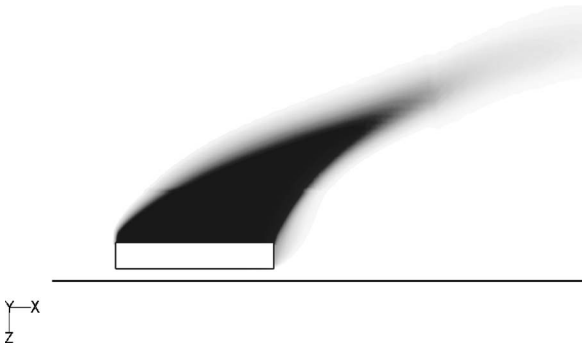


Fig. 4 Plume trajectory at a wind speed of 9 m/s

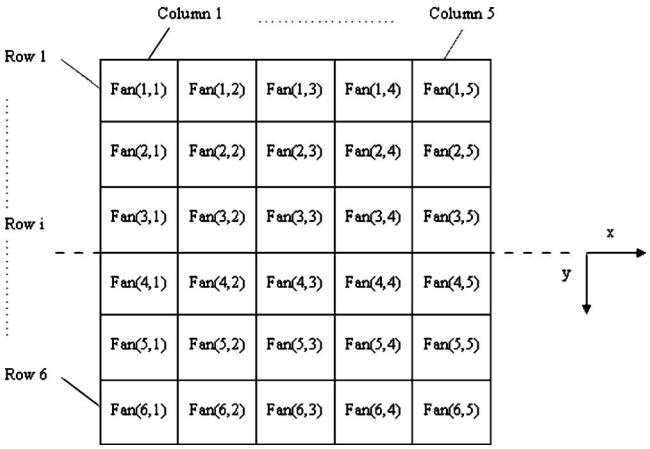


Fig. 5 Numerical model (ACSC), symmetry plane, and fan location

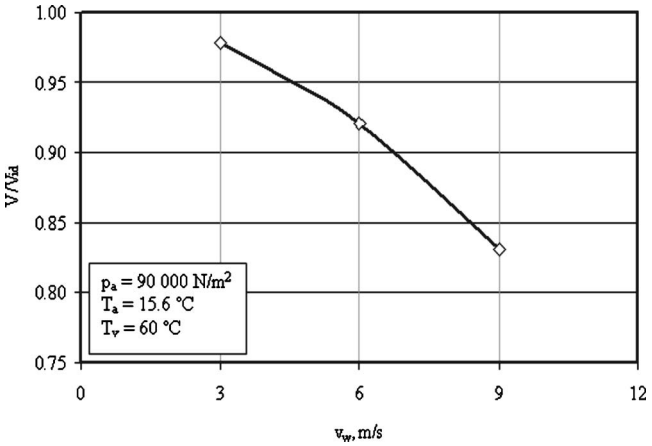


Fig. 8 Volumetric effectiveness of ACSC (wind in the *x* direction)

volumetric effectiveness of the remaining fans is considerably less. The wind may actually improve the volumetric effectiveness of some of the fans. Fan (4,5) is generally not very sensitive to wind and has a volume flow rate that is more than the ideal volume flow rate of 661 m^3/s (corresponding to the ideal mass flow rate of 717.65 kg/s). The volume flow rate through fan (4,3), fan (5,3), and fan (6,5) also exceeds the ideal. This trend may be ascribed to the relatively high momentum and low distortion of the air entering these fans.

4. As shown in Fig. 8, the net volumetric effectiveness of the entire ACSC decreases measurably with increasing wind speed.
5. As shown in Fig. 9, there is a swirling flow (vortex), which increases in strength in the downstream *x* direction and contains entrained hot plume air along the sides of the ACSC. Fan (6,1) hardly experiences any increase in its inlet temperature (essentially ambient temperature) while fan (6,5) experiences a measurable increase in inlet air temperature (approximately 6 °C) due to the strong vortex as is shown in Fig. 10.
6. The effectiveness of the ACSC decreases with increasing wind speed as shown in Fig. 11. The dotted line shows the effectiveness when only reduced fan air mass flow rates are considered (ambient air temperature assumed at the inlet to each fan; i.e., hot air recirculation is ignored). Due to the relatively high wind wall the effect of plume air recirculation is shown to have a small influence on the effectiveness.
7. If the particular ACSC is connected to the outlet of a steam turbine the latter's outlet steam temperature and corresponding back pressure will change due to wind speed and ambi-

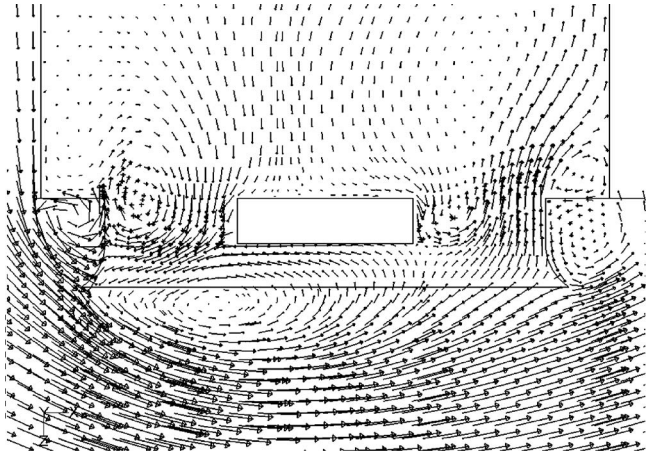


Fig. 6 Vector plot, *v* (m/s), at inlet to fan(4,1) at a wind speed of 9 m/s in the *x* direction

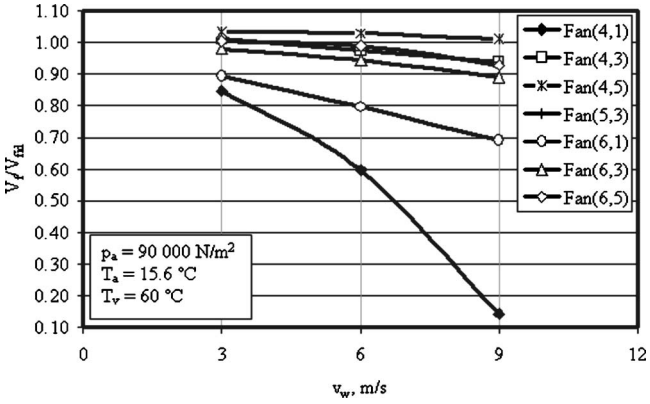


Fig. 7 Volumetric effectiveness of fans

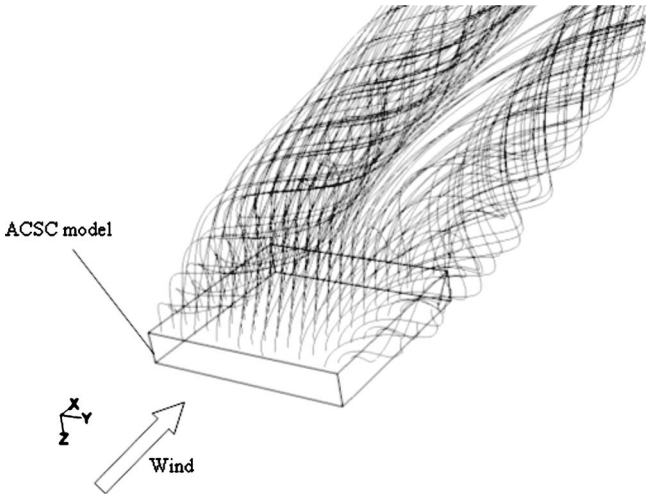


Fig. 9 Streamline plot: global flow field showing plume vortices of increasing magnitude in the downstream direction at a wind speed of 9 m/s in the *x* direction

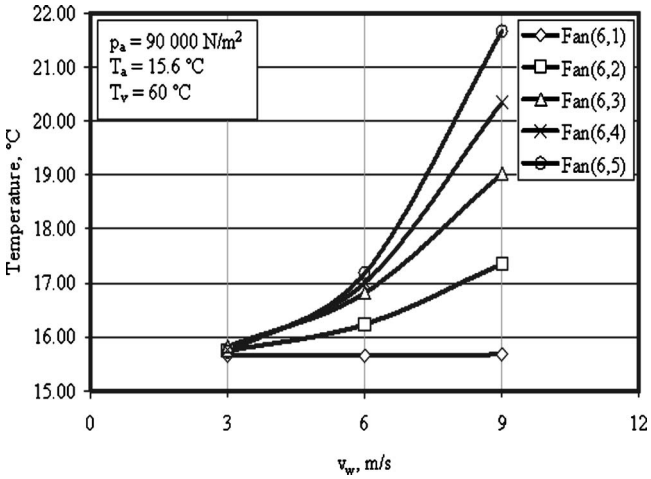


Fig. 10 Air inlet temperature of the fans in row 6, with the wind in the *x* direction

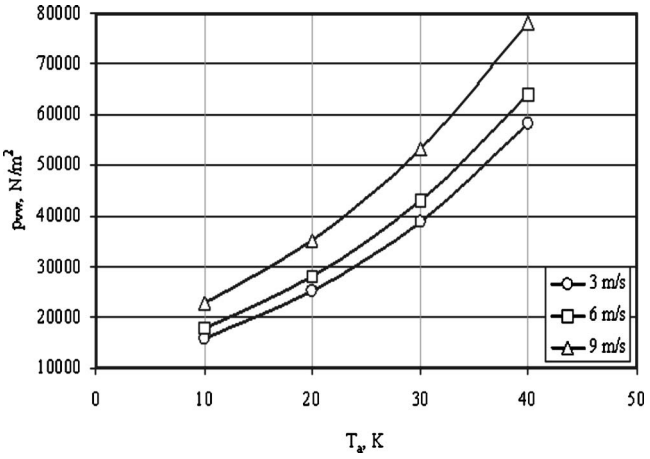


Fig. 13 Ambient temperature and wind effect on turbine back pressure

ent air temperature as shown in Figs. 12 and 13. Approximate empirical fits to the data in Figs. 12 and 13 for wind in the *x* direction are, respectively

$$T_{vw} = 0.013(75.63 + 0.035v_w^{2.49})(45.17 + T_a), \quad ^\circ\text{C} \quad (9)$$

$$p_{vw} = 0.094T_{vw}^3, \quad \text{N/m}^2 \quad (10)$$

where all temperatures are in °C.

3.2 ACSC Performance Under Windy Conditions (Wind at 45 Deg With Respect to *x* Direction). In this case most (wind at 45 deg with respect to *x* direction) of the characteristics of the flow field about and through the ACSC are similar to the previous case (wind in the *x* direction). The differences between this case and the previous case are as follows:

1. In the previous case a low-pressure region existed at the upstream leading edge. In this case the low-pressure region exists on the two leading edges. This leads to a decrease in volumetric effectiveness of fan (3,1), fan (4,1), and fan (1,3). The negative effect of the wind, in this case, on the volumetric effectiveness of the fans is different. Compare Figs. 14 and 17;
2. The swirling flow (vortex) is also present in this case, but it increases in the downstream *x*-*y* direction, not in the downstream *x* direction as in the previous case;
3. As shown in Fig. 15, the volumetric effectiveness of the ACSC decreases measurably with increasing wind speed;
4. The effectiveness of the ACSC decreases with increasing wind speed as shown in Fig. 16. The dotted line shows the effectiveness when only reduced fan air mass flow rates are considered (ambient air temperature assumed at the inlet to each fan; i.e., hot air recirculation is ignored). Due to the

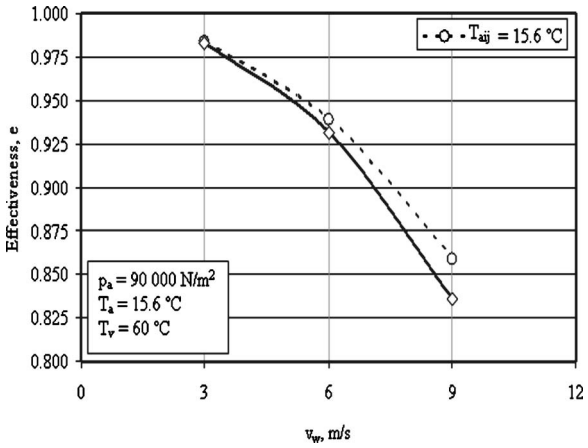


Fig. 11 Effectiveness of ACSC (wind in the *x* direction)

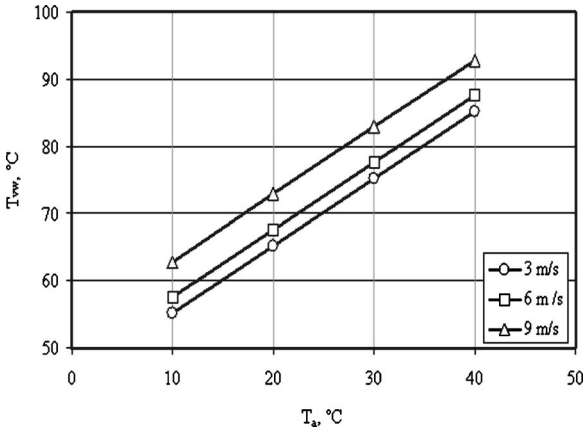


Fig. 12 Ambient temperature and wind effect on saturation temperature of the turbine exhaust steam

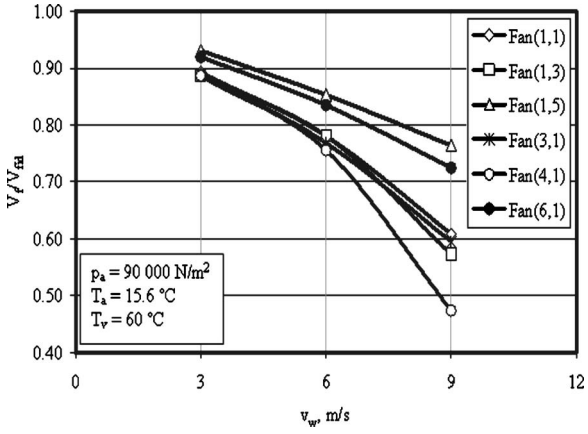


Fig. 14 Volumetric effectiveness of fans (wind in the *x*-*y* direction)

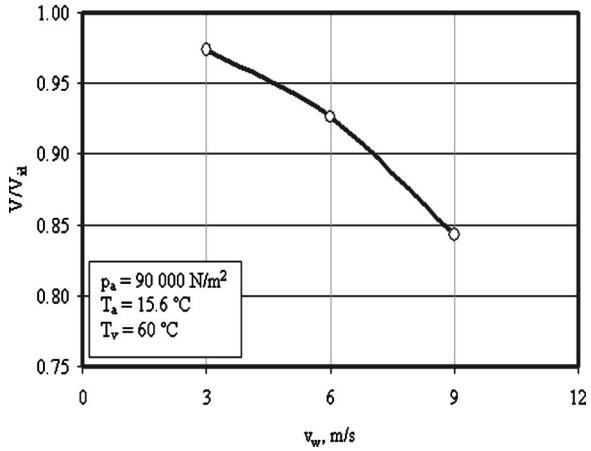


Fig. 15 Volumetric effectiveness of ACSC (wind in the *x-y* direction)

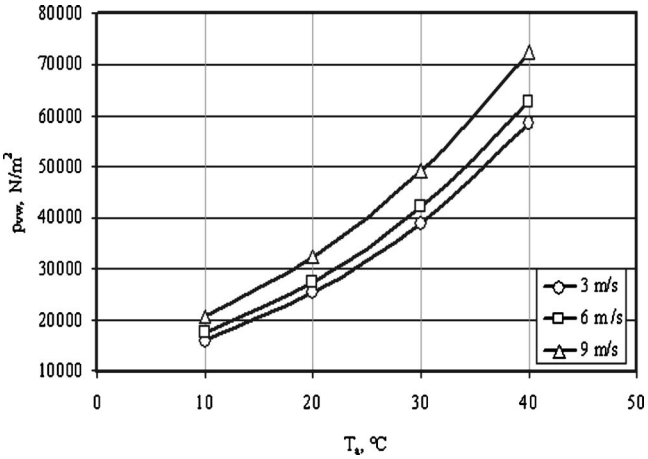


Fig. 18 Ambient temperature and wind effect on turbine back pressure

relatively high wind wall the effect of plume air recirculation is shown to have a small influence on the effectiveness; and

5. The outlet steam temperature and corresponding back pressure is found to change as a function of wind speed and ambient air temperature as shown in Figs. 17 and 18. Approximate empirical fits to the data in Figs. 17 and 18 for wind in the *x-y* direction are, respectively

$$T_{vw} = 0.0133(74.89 + 0.031v_w^{2.4})(45.33 + T_a), \quad ^\circ\text{C} \tag{11}$$

$$p_{vw} = 0.094T_{vw}^3, \quad \text{N/m}^2 \tag{12}$$

where all temperatures are in °C.

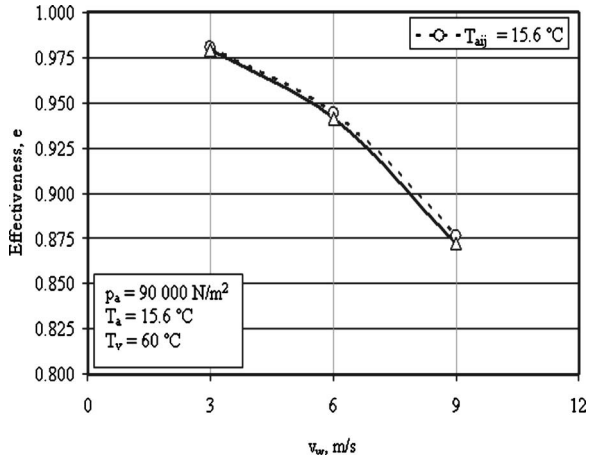


Fig. 16 Effectiveness of ACSC (wind in the *x-y* direction)

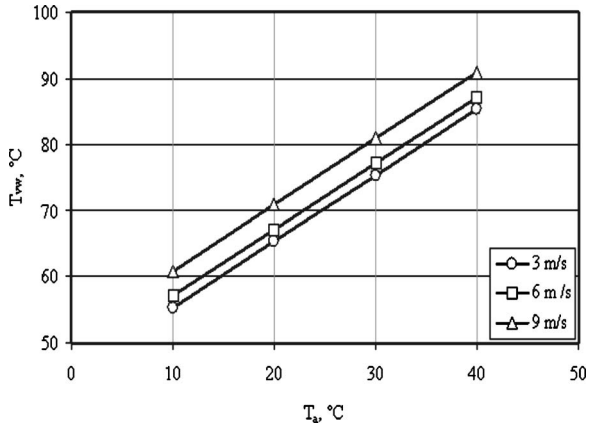


Fig. 17 Ambient temperature and wind effect on saturation temperature of the turbine exhaust steam

4 Conclusion

The objective of this study was to numerically (CFD) determine the performance trend and thermal-flow characteristics about and through the ACSC shown in Fig. 1, under windy conditions.

The model analyzed in this study shows that wind does influence the performance of an ACSC significantly. The flow distortions and corresponding low-pressure region at the upstream edge fans contribute mainly to the net decrease in the ACSC performance as the wind speed is increased, but the wind does have a positive influence on certain fans. The volumetric effectiveness of certain downstream fans is increased to above 100%. Due to the relatively high wind wall, the effect of plume air recirculation is shown to have a small influence on the effectiveness as shown in Figs. 11 and 16.

5 Recommendations

The primary focus of this study was to determine performance trends of an ACSC under windy conditions. The following recommendations are made:

- The performance of other air-cooled condensers having more or fewer fan units differently arranged and at various heights above ground level with different windwall heights should be determined; and
- The method used to analyze the ACSC flow pattern was essentially a manual parallel solving method. To analyze the flow more time efficiently and accurately, a more powerful computer system should be used.

Nomenclature

A = area, m^2
b = constant
c_p = specific heat at constant pressure, J/kg K
e = effectiveness
exp = exponential function
H = height, m
m = mass flow rate, kg/s

p = pressure, N/m^2
Q = heat transfer rate, W
T = temperature, °C or K
U = overall heat transfer coefficient, $W/m^2 K$
V = volumetric flow rate, m^3/s
v = velocity, m/s
x = coordinate
y = coordinate
z = coordinate

Greek Symbols

Δ = differential
ρ = density, kg/m^3
Σ = summation

Subscripts

a = air, or ambient
f = fan
H = height
i = numerical index, 1,2,3…
j = numerical index, 1,2,3…
id = ideal
ref = reference
sc = screen
v = vapor
w = water, or wind

References

[1] Kröger, D. G., 2004, *Air-Cooled Heat Exchanger and Cooling Towers*, Penwell Corp., Tulsa, OK.

[2] Van Aarde, D. J., and Kröger, D. G., 1990, "Vloeiverliese deur 'n A-Raam Vinbuisbundel in 'n Lugverkoelde Kondensator," M.Sc.Eng thesis, University of Stellenbosch Stellenbosch, South Africa.

[3] Goldschagg, H. B., 1993, "Lessons Learned From the World's Largest Air Cooled Condenser," *EPRI Proceedings, International Symposium on Improved Technology for Fossil Power Plants—New and Retrofit Applications*, Washington, D.C.

[4] Salta, C. A., and Kröger, D. G., 1995, "Effect of Inlet Flow Distortions on Fan Performance in Forced Draught Air-Cooled Heat Exchangers," Heat Recovery Syst. CHP, **15**(6), pp. 555–561.

[5] Duvenhage, K., and Kröger, D. G., 1996, "The Influence of Wind on the Performance of Forced Draft Air-Cooled Heat Exchangers," J. Wind. Eng. Ind. Aerodyn., **62**, 259–277.

[6] Bredell, J. R., Kröger, D. G., and Thiart, G. D., 2006, "Numerical Investigation of Fan Performance in a Forced Draft Air-Cooled Steam Condenser," Appl. Therm. Eng., **26**, pp. 846–852.

[7] Coetzee, R. V., and du Toit, C., 2000, "Numerical Simulation of an Air-Cooled Heat Exchanger," B.Eng. thesis, Potchefstroom University for Christian Higher Education, Potchefstroom, South Africa.

[8] Bredell, J. R., Kröger, D. G., and Thiart, G. D., 2005, "Numerical Investigation of Fan Performance in a Forced Draft Air-Cooled Steam Condenser," M.Sc.Eng. thesis, University of Stellenbosch, Stellenbosch, South Africa.

[9] Van Rooyen, J. A., and Kröger, D. G., 2007, "Performance Trends of an Air-Cooled Steam Condenser Under Windy Conditions," M.Sc.Eng. thesis, University of Stellenbosch, Stellenbosch, South Africa.

Estimating HP-IP Midspan Packing Leakage in Combined Cycles

Tsatsu Fiadjoe
Steam Turbine Technical Leader
GE Energy,
Schenectady, NY 12306

When a turbine has combined high-pressure (HP) and intermediate-pressure (IP) sections, there is a steam flow path between the sections. In combined cycle steam turbines, this internal leakage flow rate needs to be determined for the steam turbine performance calculations. However, since the leakage is internal to the turbine, it cannot be measured directly. One method, which has been employed in determining the midpacking leakage flow rate, is the variation of initial and/or reheat temperature method. It involves using the convergence of IP efficiency plots from multiple test runs to estimate the HP-IP leakage flow rate. Although this method has been employed successfully in large steam applications, it has generally not produced consistent results for combined cycle steam turbines. The lack of convergence for combined cycles may be due to the fact that some of the assumptions made in applying the method to large steam applications are not valid for combined cycle applications. Some of the assumptions, which need to be reviewed and modified for combined cycle application, are as follows: (a) constant IP efficiency for all test runs, (b) constant throttle flow during all test runs, (c) constant section pressure ratios for all test runs, and (d) no influence of external cooling or admission flows. This paper reviews the modifications to the traditional initial and/or reheat temperature variation method to make the midpacking leakage calculations more consistent for combined cycle applications. Some data have shown that incorporating these additional changes improves the convergence of midpacking leakage determination.
[DOI: 10.1115/1.2772631]

Introduction

The method commonly used to determine high-pressure–intermediate-pressure (HP-IP) midpacking leakage flow in opposed-flow steam turbines is the variation of initial and/or reheat temperature method. This method, commonly referred to as the temperature inference method, has been applied successfully to steam turbines operating in a fossil type configuration.

In opposed-flow combined cycle configurations, there have been limited successes in applying the temperature inference method to determine HP-IP leakage flow rate. This may be partly due to the manner in which the plant is controlled and the manner in which the data are analyzed. For example, in a fossil plant configuration, heat input to the boiler can be varied by manipulating the boiler controls whereas in a combined cycle plant configuration, the heat input is primarily fixed by the gas turbine exhaust with fewer controls available to obtain the desired steam temperatures. This paper reviews some of the considerations and analysis methods, which will quantify and improve the results of the temperature inference test on combined cycle steam turbines.

The remainder of this paper will outline the (1) overview of current methods for calculating the HP-IP leakage, (2) factors that need to be considered when applying the inference test method to a combined cycle application, (3) test and analysis methodology, and (4) field data comparing the old and new methodologies.

1 Background and Related Work

Booth–Kautzmann Paper. In the paper "Estimating the Leakage From HP to IP Turbine Sections" presented by Booth and Kautzmann [1], there are two methods presented for determining the HP-IP leakage in opposed-flow steam turbine configurations. These are the "blowdown" and the "variation of inlet and/or reheat temperature" methods.

The first method described in this paper is the blowdown method for determining the HP-IP leakage. In large fossil reheat turbines, there is generally a valve and piping system to remove steam from the HP-IP packing during a unit trip or shutdown. The blowdown system can be used to quantify HP-IP leakage by running a test where the HP-IP leakage flow is diverted through the blowdown piping to the condenser. The HP-IP leakage can then be obtained from direct measurement of this blowdown flow or inferred from comparison of the IP efficiency with and without the HP-IP packing flow diverted to the condenser.

The second method involves the variation of initial and/or reheat steam temperature to determine the HP-IP steam leakage. This method is commonly known as the inference method. Steam leakage from the HP to IP section causes an erroneous calculation of IP efficiency (hot reheat conditions to IP exhaust conditions) if the leakage flow is not taken into account. The magnitude of this error is proportional to the difference in enthalpy between the leakage and hot reheat steam, and the relative proportion of the mass flow of the leakage and the mass flow of the hot reheat steam. This method requires a minimum of two test runs at varying reheat and initial steam temperatures. For each test run, a plot of calculated IP efficiency versus assumed leakage flow rate is plotted. The lines of the test runs will intersect at the point of the actual leakage flow rate. The leakage flow rate is generally expressed as a percentage of reheat bowl flow. Often a third test run will be conducted to increase the confidence of the tests. It is important to note that whereas the blowdown method captures only leakage through the HP-IP packing, the inference method captures total leakage into the IP bowl (i.e., packing leakage, horizontal joint leakage, etc.).

Contributed by the Power Division of ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 16, 2006; final manuscript received December 6, 2006; published online March 5, 2008. Review conducted by Dilip R. Ballal. Paper presented at the ASME Power 2006 (PWR2006), Atlanta, GA, May 2–4, 2006.

Table 1 HP-IP leakage test assumptions and consideration

Traditional assumption	Combined cycle assumption/consideration
HP-IP leakage flow is a constant percentage of reheat bowl flow	HP-IP leakage flow calculated for each test run is based on design packing geometry and actual steam conditions
IP efficiency is constant during all test runs	IP efficiency may change between test runs due to changes in pressure ratio and group velocity ratio
Negligible impact of external IP first stage wheel cooling flows on the leakage result	The impact of external cooling flows on IP efficiency needs to be accounted for in the calculations
Large differences in temperature between the initial and reheat steam of up to 50°F are achievable	Turbine hardware and operational constraints limit the maximum achievable temperature differences

2 Application to Combined Cycle Steam Turbines

In the performance analysis of a steam turbine operating in a combined cycle steam configuration, the internal leakage flow from the HP to the IP turbine shell is needed in order to quantify the impact of a change in reheater duty from the design reference basis. In combined cycle steam turbines, there are generally no provisions for using the blowdown test to determine the HP-IP leakage. Therefore, the inference method is most commonly used to determine the HP-IP leakage flow rate. In applying this method to combined cycle steam turbines, a number of the traditional assumptions may need to be modified. Table 1 gives a comparison of the current assumptions made for the inference test and the modifications necessary for combined cycle application.

High-Pressure–Intermediate-Pressure Leakage Flow as a Constant Percentage of Reheat Bowl Flow. In the traditional fossil application, the assumption is made that the HP-IP leakage flow is a constant function of the reheat bowl flow. In a combined cycle application, factors such as cold reheat admission flow and changing section pressure ratios can affect the relationship of the HP-IP leakage flow to reheat bowl flow for each test run. Therefore, the HP-IP leakage flow used in the calculations is determined from the design labyrinth packing geometry and actual steam conditions. The flow is calculated using Martin's formula[1] in Eq. (1). In cases where adequate instrument provisions are not available (e.g., first stage pressure), design relationships may be used to estimate the values of various parameters.

$$Q_L = 25KA\sqrt{\frac{1-(P_2/P_1)^2}{N-\log_c(P_2/P_1)}}\sqrt{\frac{P_1}{v_1}} \tag{1}$$

where Q_L is the leakage flow (lbm/h), A the area ($\Pi \times D \times C$) (in.^{2}), C the clearance (in.), v_1 the upstream specific volume (ft^3/lbm), N the number of throttlings through the packing, K the packing flow coefficient, D the shaft diameter (in.), P_1 the upstream (first stage shell) pressure (psi (absolute)), and P_2 the downstream (IP bowl) pressure (psi (absolute)).

Constant Intermediate-Pressure Efficiency During All Test Runs. The basis of the traditional inference test is that the true IP efficiency in all the test runs is constant. For the combined cycle steam turbine, however, changing pressure ratios due to changing admission and spray flows may result in slight changes in true IP efficiency between the multiple test runs. These changes in IP efficiency can be estimated accurately from the change in the group velocity ratio of the IP stages. If the design relationship between IP efficiency and group velocity ratio is known, a correction to the test data can be made based on the measured group velocity ratio during each test run. Equation (2)[2] shows the parameters that influence the group velocity ratio, and Fig. 1 shows a typical normalized plot of IP efficiency against group velocity ratio.

$$\frac{W}{V_0} = f^n\{\mathrm{AE}, \mathrm{RPM}, \mathrm{PD}\} \tag{2}$$

where W/V_0 is the group velocity ratio, RPM the turbine speed, PD the bucket pitch diameter, and AE the section available isentropic energy.

Impact of External Cooling Flows. On some opposed-flow configurations, there is an external cooling flow from an intermediate stage in the HP to cool the first reheat stage wheel. This cooling flow impacts the calculation of the IP bowl enthalpy used to determine true IP efficiency. If there are no provisions for measuring the cooling flow, design relationships may be used to determine the cooling flow rate and enthalpy. Note that this cooling flow may not be shown on design heat balance diagrams.

Achievable Temperature Differences (Splits) Between Initial and Reheat Steam. Due to the construction manner and operational steam conditions, there are generally turbine engineering limitations placed on the allowable differences between the reheat and HP steam temperatures over the load range. In some turbines, the allowable temperature split with the reheat temperature lower than the inlet steam temperature is often greater than that allowed where the inlet steam temperature is lower than the reheat temperature.

3 Test and Analysis Methodology

Test Measurements. The following measurements need to be taken during the test program:

- throttle pressure
- throttle temperature
- HP bowl pressure
- cold reheat pressure
- cold reheat temperature

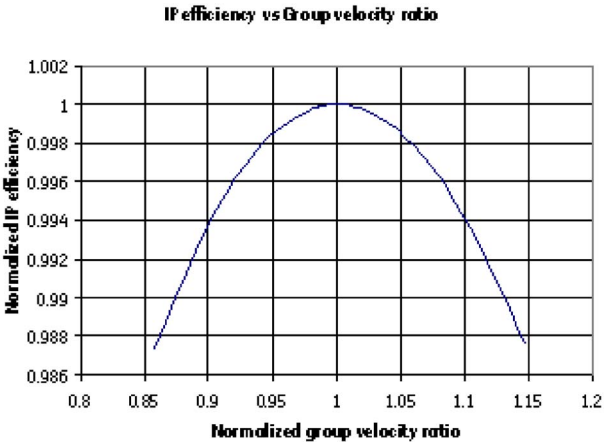


Fig. 1 Normalized IP efficiency versus normalized group velocity ratio

[1]Published by Cotton [2] in "Evaluating and Improving Steam Turbine Performance."

[2]Source: GE Technical Information Series [3].

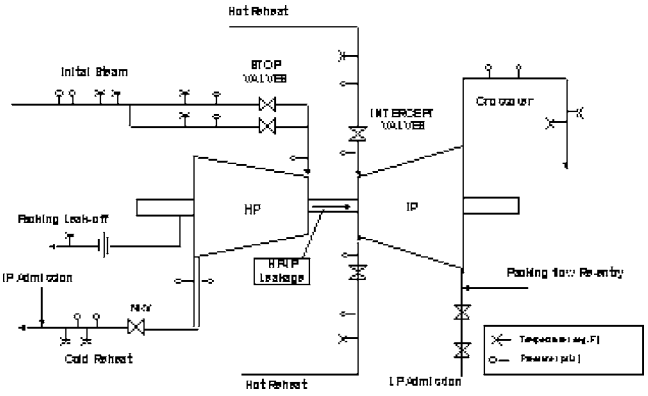


Fig. 2 Typical instrumentation diagram of a combined cycle steam turbine

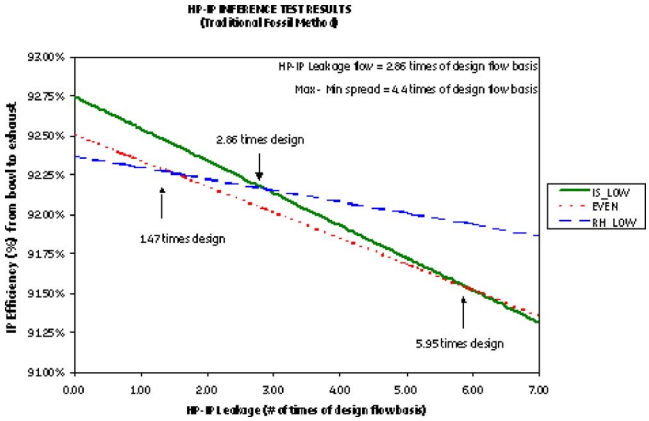


Fig. 3 HP-IP leakage using traditional fossil analysis method

- HP end packing leak-off flow
- hot reheat pressure
- hot reheat temperature
- IP bowl pressure
- IP cooling flow (or valve position)
- IP exhaust/crossover pressure
- IP exhaust/crossover temperature
- turbine control valve position

An illustration of the test measurements is given in Fig. 2.

Test Conditions. The following conditions are recommended for conducting the inference test:

- Thermally stable test conditions (Reference Table 3.1 of ASME PTC 6-1996) [4].
- Plant at approximately 75% load. At this lower load and steam flow, the ability of a fixed amount of attemperation spray flow to reduce the steam temperature will be increased.
- Isolation of the low-pressure admission steam flow to the IP exhaust. Drains downstream of the low-pressure (LP) admission control valve should also be manually isolated for the duration of the test.
- Measurement or quantification of external cooling flows and pertinent reentering packing flows.
- Test run 1. Inlet steam temperature higher than reheat temperature by up to 50°F.[3]
- Test run 2. Reheat steam temperature higher than inlet steam temperature by up to 25°F.
- Test run 3. Inlet and reheat steam temperatures set to sets have the IP efficiency plot bisect those of Test runs 1 and 2.

Testing Methodology. The inference test generally consists of three individual test runs. To minimize the impact of external factors on the IP efficiency, it is recommended that all three tests be conducted consecutively or at least on the same day. The testing order should be selected to obtain the best possible intersection of all three IP efficiency plots. Therefore, it is recommended that the two tests with extreme (i.e., low) inlet and reheat temperatures be conducted first. The overall solution of HP-IP leakage is the intersection of the IP efficiency plots of the first two test runs with the third test run providing a measure of repeatability and acceptability of the test results. Without the third test run, there is no measure of repeatability of the test results since the solution will be only based on one intersection point between two lines. The ideal temperature split (inlet and reheat steam) for the third test is dependent on the temperature splits achieved for the two previous test runs. Since the two previous test runs will have IP efficiency lines with the most divergent slopes, the temperature split for the third test run should be selected to allow the IP efficiency plot to bisect those of the two previous test runs. If the IP efficiency plot from the third test run does not approximately bisect the previous two test runs, then that test run will be biased closer to either one of the two previous test runs and the resulting solution from the intersecting points will be less meaningful. The optimum temperature split for the third test run is given in Eq. (3).

$$S_3 = \frac{1}{2}(S_1 + S_2) \tag{3}$$

where S_1 is the inlet steam minus reheat steam temperature (Test 1), S_2 the inlet steam minus reheat steam temperature (Test 2), and S_3 is the inlet steam minus reheat steam temperature (Test 3).

Calculation Methodology. After a suitable time period has been established for averaging the test data, a straight-line plot of IP efficiency against the assumed leakage flow basis will be developed for each test run. This plot is made by assuming two leakage flow values (as a function of the design flow basis) and drawing a straight line between them. For example, for first point on the plot, the HP-IP leakage is assumed to be equal to the design flow basis (i.e., one times the design), while for the second point the leakage is assumed to be twice the design flow basis (i.e., two times the design). It is important to note that since the leakage value is calculated based on the design packing geometry, the leakage value from each test run will be slightly different due to the differences in steam conditions (temperature, pressure ratio, etc.). For example, the leakage flow (lb/h) calculated for an esti-

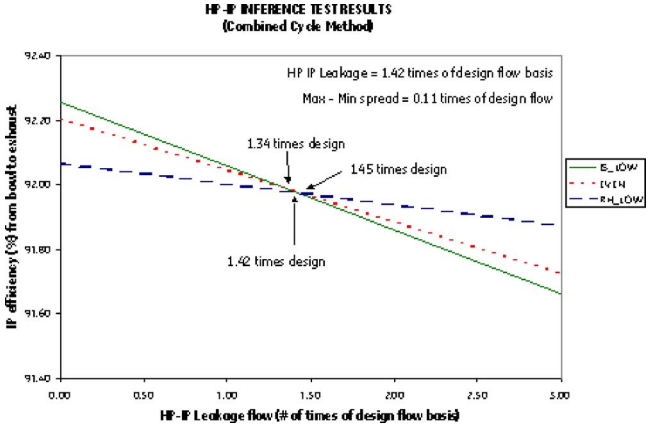


Fig. 4 HP-IP leakage test using combined analysis method

[3]Maximum temperature spreads are subject to the turbine manufacturer's specifications.

mated leakage of two times of the design flow basis for Test 1 will be slightly different from that of Test 2 due to the differences in steam conditions. This is in contrast to the traditional calculation method where the leakage flow is assumed to be constant for each test run. For each leakage flow basis, the actual flow is calculated using Eq. (1).

The following assumptions are made for each test run.

- The origin of the HP-IP leakage is from the shell of the first HP stage.
- The mixed IP bowl enthalpy consists of the HP-IP leakage and joint leakage, external cooling flow, and hot reheat flow.
- As-tested or design bowl flow capacities are used in calculating the HP and IP steam flows.

In calculating the mixed IP bowl enthalpy, it may be necessary to iterate on the HP-IP leakage enthalpy if an IP bowl flow capacity is used for the reheat bowl flow determination.

The intersection point of all the IP efficiency plots indicates the HP-IP leakage value as a fraction of the design flow basis. For the case where three or more test runs do not readily converge at one point, the HP-IP leakage is taken as the intersection of the two extreme test runs (Test runs 1 and 2 discussed previously). The intersection of these two lines is the most probable HP-IP leakage value. The third line (Test run 3) is used as an indication of the quality and repeatability of the test.

4 Field Data

HP-IP inference test data were analyzed for four GE-D11 steam turbines using both the tradition fossil analysis method and the combined cycle analysis method. Figures 3 and 4 illustrate the HP-IP leakage plots from Turbine 1. Table 2 provides comparisons of the HP-IP leakage results from all four turbines. In the table, "Mean" refers to the overall HP-IP leakage value and "Spread" refers to the difference between the two extreme intersection points on the IP efficiency plot. For the cases reviewed in Table 2, application of the combined cycle analysis method generally yields an improvement (reduction) in the spread (scatter) of the HP-IP leakage result. Application of the combined cycle analysis method also yields a more accurate HP-IP leakage flow value due to the accounting of the various flow interactions taking place.

Table 2 Comparison of HP-IP leakage values and spreads

Turbine No.	Traditional method		Combined cycle method		% change in spread
	Mean	Spread	Mean	Spread	
1	2.86	4.4	1.4	0.11	−98%
2	1.52	0.8	1.02	0.6	−25%
3	1.59	0.37	1.37	0.15	−59%
4	1.84	1.32	1.16	0.76	−42%

Conclusion

The use of the methods outlined in this paper should enhance the conduct of the temperature inference test on combined cycle steam turbines because it takes into account combined cycle specific variables to improve the accuracy of the test.

Nomenclature

HP = high-pressure turbine section
IP = intermediate-pressure turbine section

References

[1] Booth, J. A., and Kautzmann, D. E., 2004, "Estimating the Leakage From HP To IP Turbine Sections," presented at the *EPRI Power Plant Performance Monitoring Conference/Workshop*, Oct. 24.
[2] Cotton, K. C., 1998, *Evaluating and Improving Steam Turbine Performance*, Cotton Fact Inc., Rexford, NY.
[3] GE Technical Information Series No. 84STG02.
[4] ASME PTC 6-1996, Performance Test Code 6 on Steam Turbines.

Thermodynamic Performance Assessment of Gas Turbine Trigeneration System for Combined Heat Cold and Power Production

Abdul Khaliq[1]
Department of Mechanical Engineering,
Faculty of Engineering and Technology,
Jamia Millia Islamia,
New Delhi 110025, India
e-mail: abd_khaliq2001@yahoo.co.in

Rajesh Kumar
Mechanical and Automation Engineering Department,
Amity School of Engineering and Technology,
GGSIPU,
New Delhi 110061, India

The thermodynamic performance of the combustion gas turbine trigeneration system has been studied based on first law as well as second law analysis. The effects of overall pressure ratio and process heat pressure on fuel utilization efficiency, electrical to thermal energy ratio, second law efficiency, and exergy destruction in each component are examined. Results for gas turbine cycle, cogeneration cycle, and trigeneration cycle are compared. Thermodynamic analysis indicates that maximum exergy is destroyed during the combustion and steam generation process, which represents over 80% of the total exergy destruction in the overall system. The first law efficiency, electrical to thermal energy ratio, and second law efficiency of trigeneration system, cogeneration system, and gas turbine cycle significantly varies with the change in overall pressure ratio but the change in process heat pressure shows small variations in these parameters. Results clearly show that performance evaluation of the trigeneration system based on first law analysis alone is not adequate and hence more meaningful evaluation must include second law analysis.
[DOI: 10.1115/1.2771565]

Keywords: first-law, second-law, exergy destruction, gas turbine cycle, trigeneration cycle, heat recovery steam generator, cogeneration cycle

Introduction

Trigeneration is an engineering concept involving the production of electricity, heat, and cold in one operation, thereby utilizing fuel more effectively than if the desired products were produced separately. Recent interest in the topic has been primarily aimed at a variety of heat recovery applications. Maidment and Tozer [1] have reviewed a number of trigeneration plants operating in supermarkets. The authors analyzed different schemes of trigeneration plant. Bassols et al. [2] have presented different examples of trigeneration plants in the food industry. Havelsky [3] has analyzed the problem of energetic efficiency evaluation of trigeneration systems. Minciuc et al. [4] presented a method for analyzing trigeneration systems, and established the limits for the best energetic performance of gas turbine trigeneration with an absorption chilling machine from a first law of thermodynamics point of view. The first law of thermodynamics gives a good answer to the expected performance of a cycle and it can certainly lead to an assessment of the overall efficiency of the plant; however, this is concerned only with the conversion of energy, and therefore it cannot show how or where irreversibilities in a system or process occur. Thus while producing the final design result, first law analysis is incapable on its own of locating sources of losses [5,6]. Second law analysis gives much more meaningful evaluation by indicating the association of irreversibilities or exergy destruction with combustion and heat transfer processes and allows thermodynamic evaluation of energy conservation options in power and refrigeration cycles, and thereby provides an indicator that points in the direction in which engineers should concentrate their efforts to improve the performance of thermal power and refrigeration systems [7–9].

An inspection of the literature on trigeneration, however, reveals a need for further studies on trigeneration using the second law of thermodynamics. The present study deals with the combined application of first and second law analysis for the performance evaluation of the trigeneration system. The exergy balances for cycles and its components are presented and are compared to energy balances. The effects of compressor pressure ratio and process heat pressure have been observed on the thermodynamic parameters of the cycle, viz first-law efficiency, second-law efficiency, electrical to thermal energy ratio, and the magnitude of exergy losses in each components of the cycle.

Description of System

Figure 1 shows a schematic diagram of the trigeneration system. Ambient air is compressed from State 1 to State 2 and is then supplied to the combustion chamber (CC) where fuel is burned, producing hot gas at 3. The hot gas is then expanded to 4 in turbine (T) to a lower pressure and temperature. This expanded gas is utilized in the HRSG to generate process heat (Q_P). The stack gas coming out of HRSG (at 5) is sent to the generator of the vapor absorption system. The refrigerant (H_2O) separated from $LiBr/H_2O$ in the generator by means of the heat given by the stack gas.

Thermodynamic Analysis

For a thermodynamic analysis it is necessary to define some performance parameters which will be useful for analyzing a trigeneration plant. The most relevant parameters that can characterize an energy generating facility are the fuel-utilization efficiency, the electrical to thermal energy ratio, and the exergetic efficiency. Their expressions may be listed as follows

$$\eta_{\mathrm{I}} = \frac{(\dot{W}_{\mathrm{el}} + \dot{Q}_P + \dot{Q}_E)}{\dot{Q}_f} \tag{1}$$

$$R_{\mathrm{ET}} = \dot{W}_{\mathrm{el}}/(\dot{Q}_P + \dot{Q}_E) = \frac{\dot{m}_a C_P \left[(1+A) T_{\max} \eta_T \psi_T - \dfrac{T_1 \psi_C}{\eta_C} \right] \eta_g}{\dot{m}_a C_P (1+A)(T_{\max} - \eta_T \psi_T T_{\max} - T_5) + \dot{m}_r (h_9 - h_8)} \tag{2}$$

$$\eta_{\mathrm{II}} = \frac{\dot{W}_{\mathrm{el}} + \dot{E}_P + \dot{E}_E}{\dot{E}_f} \tag{3}$$

Exergy Destruction Model

If the system operates in a steady state, steady flow condition, and all the nonreacting gases are arbitrarily assigned as zero thermomechanical enthalpy, entropy and exergy at the condition of ambient pressure and temperature regardless of their chemical

[1]Corresponding author.

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received December 4, 2006; final manuscript received May 21, 2007; published online February 29, 2008. Review conducted by Dilip R. Ballal.

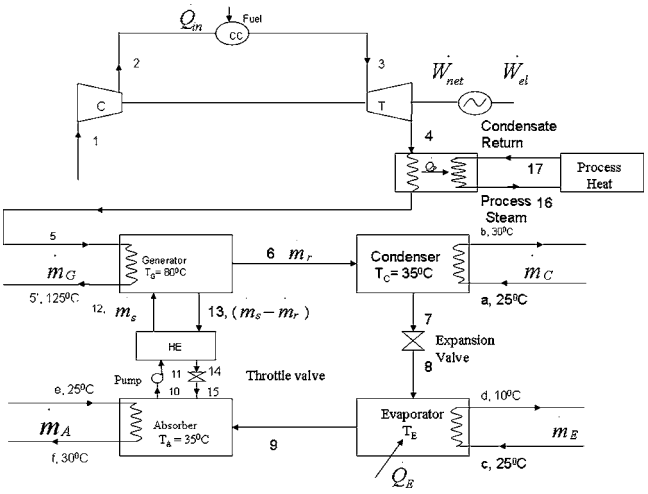


Fig. 1 Schematic diagram of the gas turbine trigeneration system for combined heat cold and power production

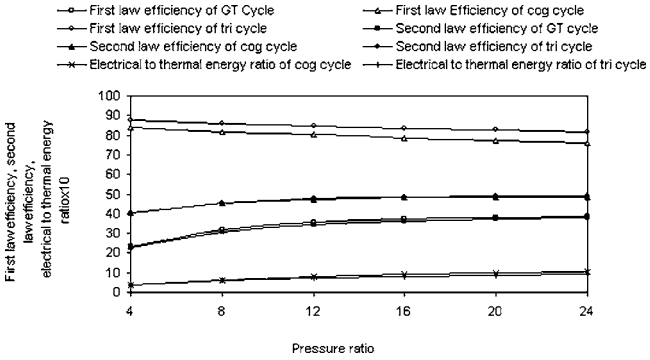


Fig. 2 Effect of variation of pressure ratio on first law efficiency, second law efficiency, and electrical to thermal energy ratio

composition, then the entropy of mixing different gaseous components can be neglected, and the general exergy-balance equation is given by Bejan [8]

$$\dot{E}_W = \sum_{i=1}^{n} (\dot{E}_Q)_i + \sum_{\text{in}} \dot{m}e - \sum_{\text{out}} \dot{m}e - \dot{E}_D \tag{4}$$

Applying the model for exergy destruction after using Eq. (4), the exergy destruction in each component of trigeneration may be obtained after using the following equations

$$\dot{E}_{D,C} = \dot{W}_C + \dot{m}_a(e_2 - e_1) \tag{5}$$

$$E_{D,\text{CC}} = Ae_{f_{\text{CC}}} + \dot{m}_a(e_2 - e_3) \tag{6}$$

$$\dot{E}_{D,T} = \dot{m}_4(e_3 - e_4) - \dot{W}_T \tag{7}$$

$$\dot{E}_{D,\text{HRSG}} = \dot{m}_4(e_4 - e_5) - \dot{m}_w(e_{16} - e_{17}) \tag{8}$$

$$\dot{E}_{D,G} = \dot{m}_r[f\psi_{12} - (f-1)\psi_{13} - \psi_6] + \dot{m}_a(1+A)(e_5 - e_{5'}) \tag{9}$$

$$\dot{E}_{D,\text{Con}} = \dot{m}_C(\psi_a - \psi_b) + \dot{m}_r(\psi_6 - \psi_7) \tag{10}$$

$$\dot{E}_{D,\text{EV}} = \dot{m}_r(\psi_7 - \psi_8) \tag{11}$$

$$\dot{E}_{D,E} = \dot{m}_r(\psi_8 - \psi_9) + \dot{m}_E(\psi_c - \psi_d) \tag{12}$$

$$\dot{E}_{D,A} = \dot{m}_A(\psi_e - \psi_f) + \dot{m}_r[\psi_9 + (f-1)\psi_{15} - f\psi_{10}] \tag{13}$$

$$\dot{E}_{D,\text{HE}} = \dot{m}_r(f-1)(\psi_{13} - \psi_{14}) + \dot{m}_r(\psi_{11} - \psi_{12}) \tag{14}$$

$$\dot{E}_{D,\text{TV}} = \dot{m}_r(f-1)(\psi_{14} - \psi_{15}) \tag{15}$$

Results and Discussion

In this paper, the effects of pressure ratio across the compressor (π_C) and process heat pressure (p_p), on the first law efficiency and electrical to thermal energy ratio (R_{ET}) is obtained by the energy balance approach or the first law analysis of the cycle. However, the second law efficiency of the trigeneration cycle has also been investigated under the exergy balance approach or the second law analysis of the cycle.

Figure 2 shows the variation of first law efficiency (η_{I}), second law efficiency (η_{II}), and electrical to thermal energy ratio (R_{ET}) for cogeneration and trigeneration cycle with a change in compressor pressure ratio (π_C) for a fixed values of (TIT=1500 K, p_P=5 bar, $\Delta p/p$=4%, and T_E=5°C). As the pressure ratio increases the air temperature at the inlet of combustion chamber increases, which results in decreasing the heat added to the cycle. Hence, as π_C increases, the first law efficiency of the gas turbine cycle increases. Figure 2 also shows the variation of second law efficiency which is a more accurate measure of thermodynamic performance, and it is observed that the second law efficiency is slightly lower than the first law efficiency of the gas turbine cycle. The first law efficiency of cogeneration cycle ($\eta_{\text{I,cog}}$) is higher than the first law efficiency of the gas turbine cycle ($\eta_{\text{I,GT}}$) because the gas turbine exhaust is utilized to produce the process heat. The second law efficiency for the cogeneration cycle ($\eta_{\text{II,cog}}$) is higher than the second law efficiency for the gas turbine cycle ($\eta_{\text{II,GT}}$). The first law efficiency of the trigeneration cycle ($\eta_{\text{I,tri}}$) is higher than the first law efficiency of the cogeneration cycle and the first law efficiency of the cogeneration cycle is higher than the first law efficiency of the gas turbine cycle ($\eta_{\text{I,GT}}$). The second law efficiency for the trigeneration cycle ($\eta_{\text{II,tri}}$) is slightly higher than the second law efficiency for the cogeneration cycle ($\eta_{\text{II,cog}}$). The electrical to thermal energy ratio (R_{ET}) for the cogeneration

Table 1 Effect of variation of pressure ratio on exergy destruction in different components of the cycle for TIT=1500 K, $\Delta p/p$ =4%, p_P=5 bar, T_E=5°C, p_{atm}=1 bar, and T_{atm}=298 K.

π_C	$E_{D,C}$ (kW)	$E_{D,\text{CC}}$ (kW)	$E_{D,T}$ (kW)	$E_{D,\text{HRSG}}$ (kW)	$E_{D,G}$ (kW)	$E_{D,\text{Con}}$ (kW)	$E_{D,\text{EV}}$ (kW)	$E_{D,E}$ (kW)	$E_{D,A}$ (kW)	$E_{D,\text{HE}}$ (kW)
4	505.83	11679.23	403.92	5624.81	271.76	38.97	3.3	54.7	62.6	15.8
8	686.14	9540.50	676.58	3549.75	271.76	38.97	3.3	54.7	62.6	15.8
12	774.92	8299.00	858.00	2623.90	271.76	38.97	3.3	54.7	62.6	15.8
16	831.60	7415.26	997.83	2077.80	271.76	38.97	3.3	54.7	62.6	15.8
20	872.18	6737.90	1111.16	1711.60	271.76	38.97	3.3	54.7	62.6	15.8
24	903.60	6175.30	1208.30	1446.45	271.76	38.97	3.3	54.7	62.6	15.8

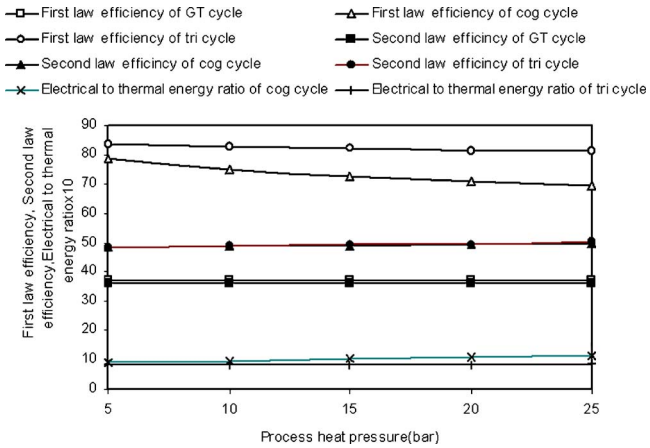


Fig. 3 Effect of variation of process heat pressure on first law efficiency, second law efficiency, and electrical to thermal energy ratio

and trigeneration cycles increases as the net work output increases, process heat decreases, and the cooling load remains the same with increase in pressure ratio.

Table 1 shows the variation of magnitude of exergy destruction in each component of the system with the change in pressure ratio (π_C) for fixed values of ($\Delta p/p=4\%$, TIT=1500 K, $p_P=5$ bar, $T_E=5°C$). It is found that the exergy destruction in the combustion process dominates. It represents over 60% of the total exergy destruction in the overall system. As the pressure ratio increases the exergy destruction in the combustion chamber decreases significantly. It is further shown that as the pressure ratio increases the exergy destruction in HRSG decreases. This is because the higher pressure ratio results in higher exergy of combustion products and lower turbine exhaust exergy, which leads to the higher turbine output. Therefore the exergy destruction decreases with the increase in pressure ratio.

The effect of process heat pressure (p_P) for fixed values of ($\pi_C=16$, $\Delta p/p=4\%$, TIT=1500 K, $T_E=5°C$) on the first law efficiency, electrical to thermal energy ratio, and second law efficiency of cogeneration and trigeneration system is shown in Fig. 3. It is found that the first law efficiency of cogeneration decreases with an increase in the pressure of process heat. The first law efficiency of the trigeneration cycle also decreases with the increase in process heat pressure. Figure 3 also shows the variation of second law efficiency of the system with the increase in process heat pressure. It is found that in cogeneration, the second law efficiency increases with the increase in pressure of process heat if turbine exhaust temperature is relatively high. For trigeneration, the second law efficiency also increases with the increase in process heat pressure, the same as in cogeneration, but the magnitude of second law efficiency for trigeneration is slightly higher than cogeneration because the exergy associated with cold at evaporator temperature is very small as compared to the exergy associated with power output and process heat. Figure 3 further shows that the electrical to thermal energy ratio for cogeneration and trigeneration is almost constant with the increase in process heat pressure.

The effect of process heat pressure on exergy destruction in each component of the system is shown in Table 2. It has been observed that the exergy destruction in all components of gas turbine and cogeneration cycle is more or less independent of the process heat pressure but the increase in process heat pressure causes significant decrease in the exergy destruction in HRSG. The exergy destruction in each component of the refrigeration cycle increases with the increase in process heat pressure. This is expected because higher pressure for process heat results in higher flue gas temperature. Consequently, more heat will be added to the generator and hence mass flow rate of refrigerant increase, and that would result in higher exergy destruction in each component of the vapor absorption system.

Conclusions

This paper has presented a methodology for analyzing the trigeneration system, for combined production of heat, cold, and power. Combined first and second law analysis of the given system leads to the following conclusions.

Maximum exergy is destroyed during the combustion and steam generation process; it represents over 80% of the total exergy destruction in the overall system. The exergy destruction in the combustion chamber and heat recovery steam generator decreases significantly with the increase in pressure ratio. The exergy destruction in HRSG decreases, and in the components of vapor absorption refrigeration increases significantly with the increase in process heat pressure but the exergy destruction in the compressor, combustion chamber, and turbine seem to be constant for all process heat pressures. The first law efficiency of cogeneration and trigeneration decreases with the increase in pressure ratio but the second law efficiency and electrical to thermal energy ratio for these systems increases with the same. The first law efficiency of cogeneration and trigeneration decreases slightly with the increase in process heat pressure but the second law efficiency and electrical to thermal energy ratio increases with the same.

Nomenclature

A = fuel to air ratio
f = solution circulation ratio
ψ = specific exergy (kJ/kg (refrigerant)); $\psi_C = \pi_C^{\alpha} - 1$; $\psi_T = 1 - 1/\pi_T^{\alpha}$
θ = maximum to minimum temperature ratio
η_g = electrical conversion efficiency (%)

Subscripts

P = process
r = refrigerant
f = fuel
$1,2,3,\ldots,a,b,\ldots$ = state points in Fig. 1
HRSG = heat recovery steam generator

Table 2 Effect of variation of process heat pressure on exergy destruction in different components of the cycle for $\pi_C=16$, TIT =1500 K, $\Delta p/p=4\%$, $T_E=5°C$, $p_{atm}=1$ bar, $T_{atm}=298$ K.

p_P (bar)	$E_{D,C}$ (kW)	$E_{D,CC}$ (kW)	$E_{D,T}$ (kW)	$E_{D,HRSG}$ (kW)	$E_{D,G}$ (kW)	$E_{D,Con}$ (kW)	$E_{D,EV}$ (kW)	$E_{D,E}$ (kW)	$E_{D,A}$ (kW)	$E_{D,HE}$ (kW)
5	831.6	7415.26	997.83	2077.80	271.76	38.97	3.3	54.7	62.6	15.8
10	831.6	7415.26	997.83	1629.55	430.43	60.06	5.1	84.0	96.1	24.3
15	831.6	7415.26	997.83	1383.26	469.63	73.58	6.2	103.3	118.2	29.8
20	831.6	7415.26	997.83	1229.08	685.50	84.18	7.1	118.2	135.2	34.1
25	831.6	7415.26	997.83	1093.10	782.74	92.68	7.8	130.1	148.8	37.6

References

[1] Maidment, G. G., and Tozer, R. M., 2002, "Combined Cooling Heat and Power in Supermarket," Appl. Therm. Eng., **22**, pp. 653–665.

[2] Bassols, J., Kuckelkorn, B., Langreck, R., and Schneider, H. V., 2002, "Trigeneration in Food Industry," Appl. Therm. Eng., **22**, pp. 595–603.

[3] Havelsky, V., 1999, "Energetic Efficiency of Cogeneration Systems for Combined Heat, Cold and Power Production," Int. J. Refrig., **22**, pp. 479–485.

[4] Minciuc, E., Corre, O. L., Athanasovici, V., Tazerout, M., and Bitir, I., 2003, "Thermodynamic Analysis of Trigeneration With Absorption Chilling Machine," Appl. Therm. Eng., **23**, pp. 1391–1404.

[5] Kotas, T. J., 1985, *The Exergy Method of Thermal Power Plant Analysis*, Butterworths, London.

[6] Nag, P. K., and De, S., 1998, "Study of Thermodynamic Performance of an Integrated Gasification Combined Cycle Power Plants," Proc. Inst. Mech. Eng., Part A, **212**, pp. 89–95.

[7] Wall, G., 2003, "Exergy Tools," Proc. Inst. Mech. Eng., Part A, **217**, pp. 125–136.

[8] Bejan, A., 2002, "Fundamentals of Exergy Analysis, Entropy Generation Minimization, and the Generation of Flow Architecture," Int. J. Energy Res., **26**, pp. 545–565.

[9] Huang, F. F., 1990, "Performance Evaluation of Selected Combustion Gas Cogeneration Systems Based on First and Second Law Analysis," ASME J. Eng. Gas Turbines Power, **112**, pp. 117–121.

A Wireless Microwave Telemetry Data Transfer Technique for Reciprocating and Rotating Components

Scott A. Miers
Argonne National Laboratory,
9700 South Cass Avenue,
Argonne, IL 60439

Glen L. Barna
IR Telemetrics, Inc.,
Houghton, MI 49931

Carl L. Anderson

Jason R. Blough

M. Koray Inal

Michigan Technological University,
Houghton, MI 49931

Stephen A. Ciatti
Argonne National Laboratory,
Argonne, IL 60439

Wireless microwave telemetry addresses the difficult issue of obtaining transducer outputs from reciprocating and rotating components through the use of advanced electronic components. This eliminates the requirements of a direct link between the transducer and the acquisition system. Accuracy of the transducer signal is maintained through the use of a double frequency modulation technique which provides temperature stability and a 20 point calibration of the complete system. Multiple transmitters can be used for larger applications and multiple antennas can be used to improve the signal strength and reduce the possibility of dropouts. Examples of automotive torque converter and piston temperature measurements are provided, showing the effectiveness of the wireless measuring technique. [DOI: 10.1115/1.2771562]

Keywords: telemetry, torque converter, piston temperature, data acquisition

Introduction

There is an increasing and inherent need to fully understand and measure mechanical systems as federal regulations tighten and consumer demands increase. This understanding begins with successful and efficient acquisition of data from devices such as automotive torque converters and internal combustion engines. The challenge lies in transferring the output from strategically placed transducers, such as pressure, strain, and/or temperature sensors, to suitable data acquisition systems. Continuous wires are not practical due to transducer locations deep within an engine or torque converter. Desired installations include internal combustion engine pistons, engine crankshafts, and automotive torque converters, as shown in Fig. 1 [1].

Two specific wireless microwave telemetry installations with corresponding data are presented, showing the wide range of applicability of the technique as well as the dynamic range and accuracy with which the original signal is reproduced.

Background

There have been several attempts at designing systems to record transducer outputs that are located in hard to reach areas [2–4]. Each has their advantages and disadvantages, as will be discussed in the following subsections.

Contact point or fusible plugs have been used in various forms for several years. Inserting a known material in a particular measurement location and subjecting the component to various temperatures, changes the hardness of the plug material, based on maximum temperature reached. While effective at capturing peak temperatures and being relatively inexpensive, dynamic phenomena are not able to be captured and nondestructive testing is impossible.

L-link systems [5] provide a continuous, solid connection to the transducer for reciprocating components, such as internal combustion engine pistons, connecting rods, and crankshafts. A schematic of a typical L-link system is shown in Fig. 2.

Modification of the existing hardware, i.e., engine, is extensive to accommodate the L-link mechanism, thus adding significant cost combined with increased imbalance in the system. Additionally, the transducer wires are continually flexed during operation and are susceptible to fatigue failure.

Both reciprocating and rotating components can utilize the slip ring configuration, which eliminates the continual flexing of transducer wires. Two similar material disks, one fixed and one rotating, are used to transfer the data from the transducer to the acquisition system. Continuous transducer data can be recorded with the slip ring configuration; however, contact noise can degrade the signal significantly and both size and cost can be an issue. Typically, the solution to noise reduction is expensive contact plates, which wear rather quickly.

Electromagnetic induction systems [6] transmit transducer output at a specified location in the cycle. Reciprocating components can benefit from inductive telemetry systems and are often installed to transmit piston temperatures at the bottom dead center location of the piston. While the system is relatively inexpensive and simple, only temperature measurements at bottom dead center are available. No information about component temperature near top dead center (i.e., the combustion event) is available.

Wireless telemetry addresses the shortfalls of the previous systems by providing the capability to measure transducer output continuously and eliminate any wire flexing. The telemetry transmitter is often located close to the transducer, so some additional weight is added to the dynamic system. However, this usually accounts for less than 10% of the total weight of the original system. Additionally, because there is no sliding or slipping connection, an additional noise source is eliminated.

This paper covers the study of wireless, microwave telemetry data acquisition, identification of its many advantages over existing measurement techniques, addresses its' few shortcomings, and presents current automotive torque converter and piston temperature data.

System Operation

The wireless telemetry system consists of four main components; the transmitter, the power supply, the antenna, and the receiver, as shown in Fig. 3 for a typical internal combustion engine piston installation.

The wireless telemetry component specifications are shown in Table 1. The 16 channels of input are typical and can be varied to accommodate the particular test being performed. Also note the

Submitted to ASME for publication in the JOURNAL OF ENGINEERING FOR GAS TURBINES AND POWER. Manuscript received October 9, 2006; final manuscript received May 23, 2007; published online February 21, 2008. Review conducted by Thomas W. Ryan III. Paper presented at the 2005 Fall Conference of the ASME Internal Combustion Engine Division (ICEF2005), Ottawa, Ontario, Canada, September 11–14, 2005.

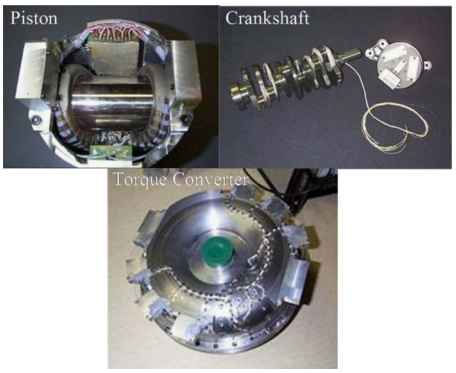


Fig. 1 Wireless telemetry installations

Table 1 Wireless telemetry specifications

Number of data channels per transmitter	16
Maximum bandwidth	10,000 Hz
Maximum operating temperature for electronics and batteries	150°C
Maximum internal loading	3,000 g reciprocating 50,000 g continuous
Transmitter weight	30–50 g
Multiplexing time	20–2000 ms
Acceptable inputs	Pressure, thermistors, strain, thermocouples, accelerometers, proximity
Battery weight	40 g

high level of internal loading able to be sustained by the electronics, thus making the system suitable for piston-type applications.

The transmitter is located in close proximity to the transducer and is responsible for acquiring the signals, multiplexing multiple transducer inputs, converting the low voltage input to a high-frequency signal, and for transmitting the signal to the antenna. Typical transmitter inputs are pressure, strain, and temperature transducers. Two frequency modulations (FMs) are utilized to produce a stable signal over a wide range of operating temperatures. A 16 channel sequential multiplexer exists inside the transmitter housing and controls which transducers signal is converted and transmitted to the antenna. The multiplexer can be designed to read the transducer input for as short a time as 20 ms or as long as 2 s.

The millivolt, emf output from the multiplexer is converted to a frequency-modulated, 10–50 kHz square wave in the first voltage to frequency conversion (V/F). Higher input voltage from the transducer is associated with higher frequencies in the square wave. This step produces a signal that contains the original transducer output, yet does not vary with changes in operating temperature, otherwise known as drift. The square wave is used to modulate a voltage-controlled oscillator (VCO) operating at a nominal frequency of 2.5 GHz. The output from the VCO, for example, could range from 2.49 GHz to 2.51 GHz. This frequency varying signal is transmitted to the environment and picked up by either single or multiple antennas. A schematic representation of the transducer output conversion process inside the transmitter is shown in the upper section of Fig. 4.

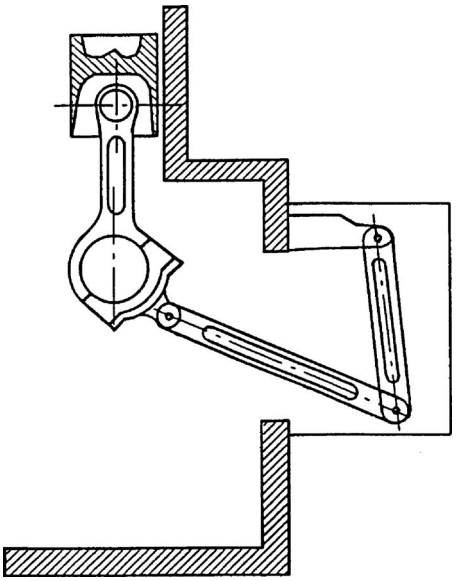


Fig. 2 Typical L-link system schematic

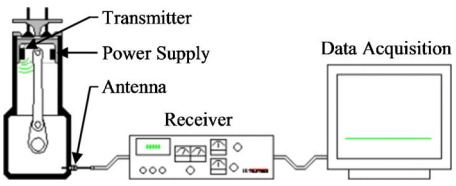


Fig. 3 Telemetry component schematic

The high-frequency signal from the transmitter is received by the antenna which can be installed up to 20 m from the transmitter output; however typical distances are less than 30 cm. The antenna signal is amplified and passes directly to the microwave receiver. The receiver is a superheterodyne unit that first down-converts the signal from 2.5 GHz to approximately 30 MHz. A frequency tuning knob is provided to match the incoming signal with the receiver specifications. Demodulation takes place to remove the 30 MHz signal and reproduce the square wave. A second demodulation reproduces the original transducer output, from the square wave, also known as the F/V section. Finally, the converted signal is output to a high-speed data acquisition system for analysis and post-processing. The lower half of Fig. 4 shows the schematic version of the two conversion steps within the receiver.

During the frequency to voltage conversion, it is possible for some of the square wave to bleed through and show up on the final voltage output of the receiver. Typically, this results in a 10–50 kHz ripple, superimposed on the reproduced transducer output. For this reason, a 2000 Hz, fifth-order, low-pass, Bessel filter is applied to the output of the F/V section, before high-speed acquisition. This also serves as an anti-aliasing filter. The effect of the low-pass filtering will be shown in the "Calibration" section.

There are two systems used to power the transmitter and multiplexer: batteries and inductive coils. Batteries are simple, easy to install, and provide 10–20 h of service when used with a thermal switch inside the transmitter. The g loading on batteries is limited to approximately 3000 rpm in engine applications, above which internal shorting of the battery is likely to occur. Inductive coils are not limited to g loading, but do require additional modifications to the original system. The inductive power unit is similar to an air-gap transformer with a stationary, externally powered coil and a moving coil. During some part of the cyclic operation, the coils couple and transmit power to the transmitter pack. Properly sized capacitors store the necessary power to operate the transmitter until the coils couple on the next cycle. Note that the transmitter and battery components are incorporated as a system to reduce the amount of imbalance to the original hardware.

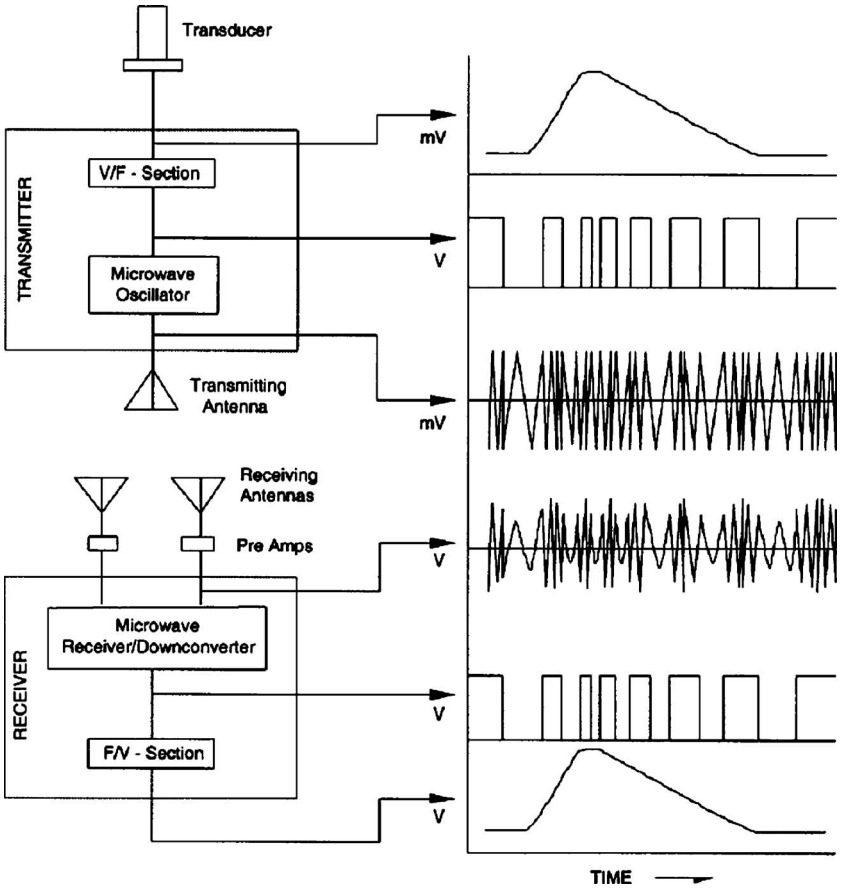


Fig. 4 Signal evolution, from transducer to output of receiver

Calibration

The calibration of the wireless telemetry consists of two separate operations. The first is a determination of the gain and phase of the overall system and the second is a 20 point temperature calibration and ice point verification.

Gain and Phase Determination. The gain and phase characteristics of the wireless telemetry system were obtained through the determination of the frequency response function. A 5–5000 Hz, millivolt burst-chirp signal was used to simulate a thermocouple output, over a wide range of dynamic temperature measurements. The signal was input to the telemetry system, in place of a thermocouple emf, and also input to a high-speed data acquisition system. The output of the telemetry receiver was input to a second channel of the data acquisition system, which allowed determination of the gain and phase characteristic. Connection details are shown in Fig. 5.

As shown in the upper part of Fig. 6, the cutoff frequency, based on 3 dB attenuation, was determined to be 2030.5 Hz. The phase remained linear up to 3500 Hz, as shown in the lower part of Fig. 6.

This is consistent with the fifth-order Bessel filter applied to the output of the last conversion step within the receiver. A 2000 Hz low-pass filter is used to eliminate the 10–50 kHz ripple associated with the frequency to voltage conversion. These results show

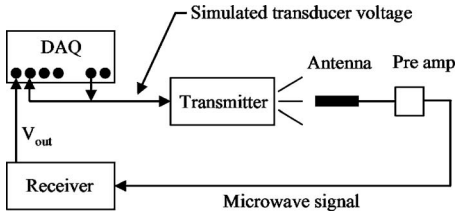


Fig. 5 Electrical connections for determining gain and phase characteristics

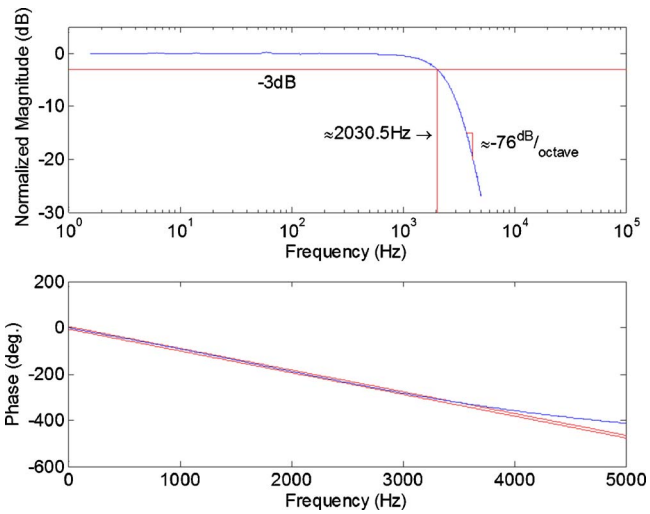


Fig. 6 Telemetry gain and phase characteristics

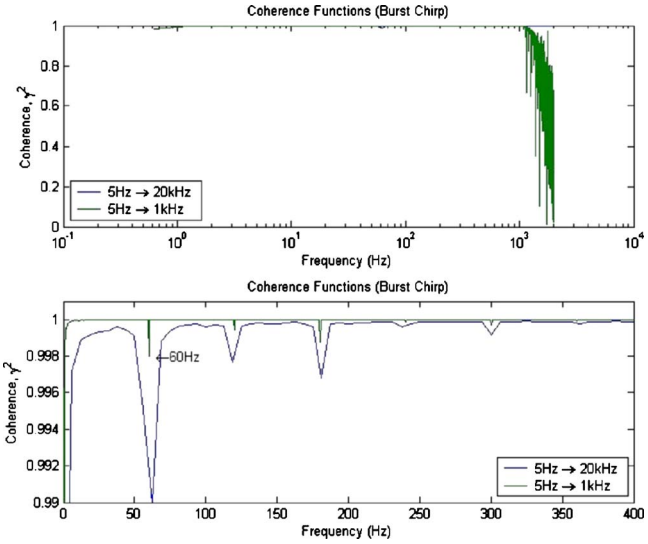


Fig. 7 Coherence function and 60 Hz noise

that the data being transmitted by the wireless telemetry system have not been significantly altered between the frequencies of 0 Hz and approximately 2000 Hz.

During the acquisition of data to determine the gain and phase characteristics, a 120 V ac, 60 Hz power cord was in close proximity to the antenna of the telemetry system. A coherence function calculation identified that the telemetry system recorded the 60 cycle noise and multiples up to 300 Hz, as shown in Fig. 7. Therefore, during an actual test, low-frequency (0–2000 Hz) noise sources should be kept away from the antenna area to avoid introducing noise into the transmitted signal. Also note that most engine noise sources, such as alternators, spark plugs, and coil drivers are well shielded from the transmitter pack due to the pack mounting location. Crankcase and torque converters are seldom large sources of noise and actually act as a shielded environment for the transmitter electronics and antennas.

20 Point Temperature Calibration. The wireless temperature transmitter is placed in Furnace No. 1 with one of its thermocouple inputs leading to an isothermal block in Furnace No. 2, as shown in Fig. 8. The transmitter temperature is monitored by calibrated thermocouple Meter A. During the first calibration test, the wireless transmitter is held at room temperature (20°C) while Furnace No. 2 is increased from room temperature to 350°C. The actual temperature of the transmitter input thermocouple is monitored with another thermocouple connected to calibrated thermocouple Meter B.

The transmitter output is recorded simultaneously with a temperature reading from calibrated thermocouple Meter B at approximately 20°C increments. The transmitter output versus temperature is plotted and a regression analysis technique is used to curve fit the data sets. The temperature of Furnace No. 1 (transmitter) is then increased to 135°C and Furnace No. 2 is operated from room temperature to 350°C, in 20°C increments. The elevated temperature environment data are compared to the room temperature output data and must agree to within 1%. This verifies the proper function of the transmitter electronic ice point. After the ice point function is verified, typically the 135°C calibration curve is used as the official calibration curve for service. If the operational temperature of the transmitter varies significantly from 135°C, then a new official calibration curve is created.

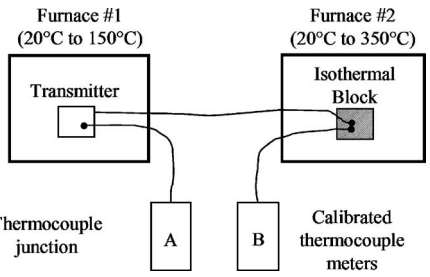


Fig. 8 Temperature calibration schematic

After the wireless transmitter and thermocouples have been installed on a component, such as a piston, the entire assembly is placed in Furnace No. 1 and heated to 150°C. The transmitter output is compared to calibrated thermocouple Meter A and to the official calibration equation. A minimum accuracy between the measured temperature and the actual temperature of 1% should be observed. Note that thermocouple accuracies are typically in the range of ±2%. After the transmitter has been used in an actual test, it is returned to the manufacturer for post-calibration to verify the absence of electronic drift or changes during the test. The post-calibration consists of repeating the two furnace tests outlined above and comparing this data to the original calibration curves.

Experimental Results

Two applications of the wireless microwave telemetry technique are briefly discussed in this section. Measurements of the fluid pressure on the stator blade of a rotating torque converter have been conducted to assist in the validation of computer simulations and promote further understanding of the fluid cavitation process. Additionally, high-speed diesel piston temperatures were recorded under a wide range of operating conditions to identify impingement signatures and provide boundary conditions for heat transfer analysis.

Automotive Torque Converters. Computational fluid dynamics (CFD) models of the torque converter operating in stall condition show that the lowest pressure in the converter typically occurs at the nose of the stator blade [7,8]. At stall condition fluid flows over the stator blade at a high angle of attack leading to flow separation on the suction side of the blade [9,10]. The combination of flow separation and high fluid velocities leads to very low pressures at the nose of the stator. Figure 9 is a cross section of the stator passage showing a CFD generated static pressure map at a location 25% of the way from the converter's outer shell to the converter's core (0.25 *S-C*) at a pump speed of 2000 rpm and a charge pressure of 70 psi.

An experimental stator was designed to measure static pressure at the nose of the stator, using wireless telemetry. Fifteen 0.5-mm-diameter pressure taps were situated in a 3×5 array that wrapped around the nose of the stator at locations shown in Fig. 10.

Each of the pressure manifolds were filled with automatic transmission fluid (ATF) throughout the pressure tap by letting air escape through an evacuation hole that was subsequently sealed with a gasket after all the air was evacuated. The microwave telemetry technique using a single transmitter mounted in the stator hub was used to transmit the 15 multiplexed transducer signals from the stator hub and out of the spinning torque converter. The sampling rate for the data was 12,800 Hz. Figure 11 is a photograph of the experimental stator with the cover removed showing the telemetry electronics mounted and potted in the hub of the stator.

Pressure data were taken at conditions that promoted the onset of cavitation. A cavitation signature was found in the fluctuating component of the static pressure, p'. Ca$'$ is a nondimensional fluctuating pressure that is referred to as the fluctuating cavitation number and is based on the mean square pressure (MSP) value of p', $\sqrt{\text{MSP}}$

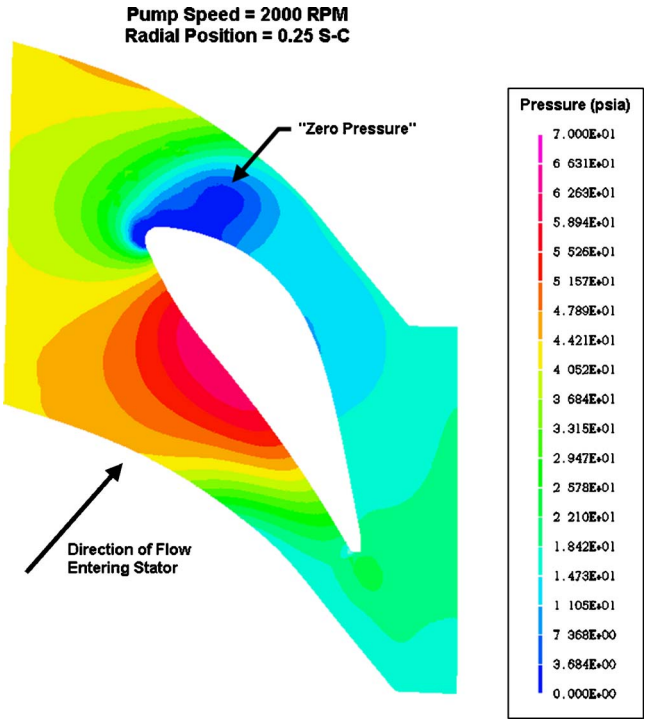


Fig. 9 CFD prediction of static pressures around a stator section at 0.25 *S-C*. The pump speed is 2000 rpm and the charge pressure is 70 psi.

Fig. 11 Stator with cover removed to show telemetry electronics installed in the hub

$$Ca' \equiv (\sqrt{\text{MSP}})/(1/2\rho(D\Omega)^2) \quad (1)$$

$\tilde{\Omega}_p$ is dimensionless pump speed defined as

$$\tilde{\Omega}_p \equiv (D\Omega)/(\sqrt{p_c/\rho}) \quad (2)$$

At a dimensionless pump speed of 1.25, cavitation bubbles start to form away from the blade and begin to damp the pressure fluctuations near the blade. When a dimensionless pump speed of 1.63 is reached, the fluctuating cavitation number is nearly zero and remains there for all higher values, as shown in Fig. 12. A dimensionless pump speed of 1.63 was defined as advanced cavitation and is the point were a three point running average was below 10% of the initial value of fluctuating cavitation number at low dimensionless pump speeds [11].

Ca′ dropped to a value near zero at the dimensionless pump speed marking advanced cavitation. This behavior is associated with a coalescence of the cavity formation zone near the surface where the pressure transducers were mounted. Attenuation of the pressure fluctuations by newly formed and growing cavities marks the inception of cavitation at the stator blades [11].

Piston Temperature Measurement Results. Two separate sets of analysis were performed with nearly the same set of data for the piston temperature measurements. The first was an impingement study on the surface of the piston and the second was a piston thermal loading study.

Figure 13 shows the location of the eight surface mounted thermocouples [1–6,14] and seven embedded thermocouples on a piston crosssection. Figure 14 shows the top view of the piston and corresponding thermocouple locations.

For the impingement study, the placement of the surface ther-

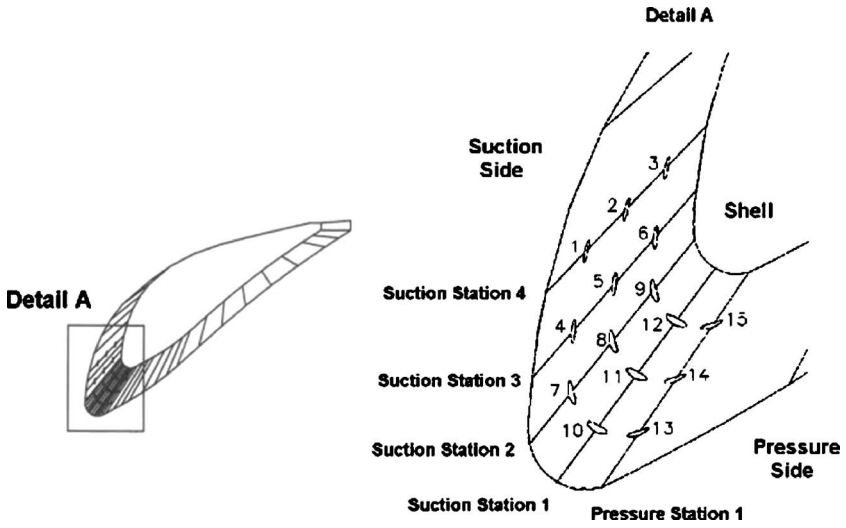


Fig. 10 Location of pressure taps on stator blade

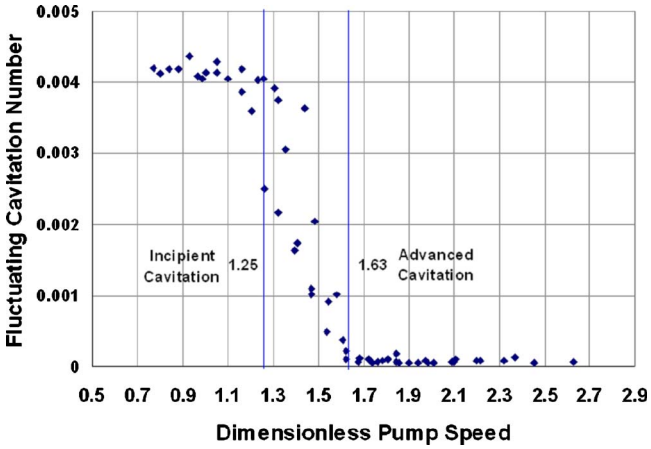


Fig. 12 Fluctuating cavitation number as a function of dimensionless pump speed at SS1, 0.25 S-C

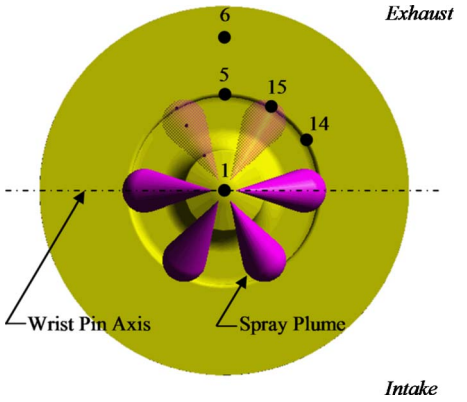


Fig. 14 Surface thermocouple locations from top view of piston (thermocouples 2, 3, and 4 omitted for clarity)

mocouples was chosen at locations on the piston of probable impingement. Thermocouples 5, 14, and 15 are mounted on the piston bowl lip and are circumferentially 30 deg from each other [14].

The fast-response surface thermocouples have an outer diameter of 1.55 mm (0.061 in.), a surface thermal deposition coating of 2 μm, and a time constant of 1 μs, according to the manufacturer (L. Jones, personal communication, March 20, 2002). Surface thermocouple specifications are provided in Table 2.

Figure 15 shows the surface temperature verses crank angle degree for the Thermocouple 5 location at two engine speeds, multiple loads, and 0 deg injector rotation.

0 crank angle degree (CAD) corresponds to the top dead center location of the piston. As load increases, both mean and peak surface temperature is observed to increase, as expected. When the same loads, speeds, and injector orientation are analyzed for Thermocouple 15, a unique signature appears at the higher loads, as shown in Fig. 16.

Thermocouple 15 displays a higher peak temperature and a distinct temperature profile around 0 CAD, compared to Thermocouple 5. Only Thermocouple 15 identified a distinct feature on the piston surface at this injector orientation. The locations on the piston and the operating condition of the engine are critical factors in determining if the signature is present. The change in the temperature profile is believed to be the result of the burning spray impinging on the thermocouple location. The temperature profile change occurs after the initial rise in the cylinder pressure which results from combustion, as shown in Fig. 17.

The rate of temperature change versus crank angle degree was calculated to further improve the identification of the impingement signature. Figure 18 shows the effectiveness of the identification technique. When impingement is present, it is transformed into a "spike" prior to top dead center, before the main combustion event. The peak value of the spike appears to correlate with the magnitude of the impingement event.

The speed/load dependence of the impingement signature was identified by calculating the peak value of $dT/d\theta$ at the impinge-

Table 2 Fast response surface thermocouple specifications

Thermocouple type	Triaxial, Type E (chromel/constantan)
Diameter	1.55 mm (0.061 in.)
Electrical connection	Surface deposited chromium
Connection thickness	2 μm
Time constant	Less than 1 μs
Bandwidth	10,000 Hz

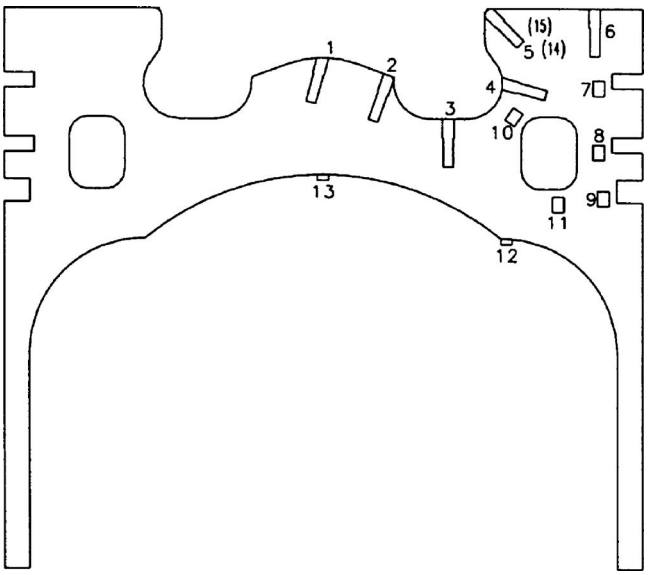


Fig. 13 Thermocouple locations on piston cross section

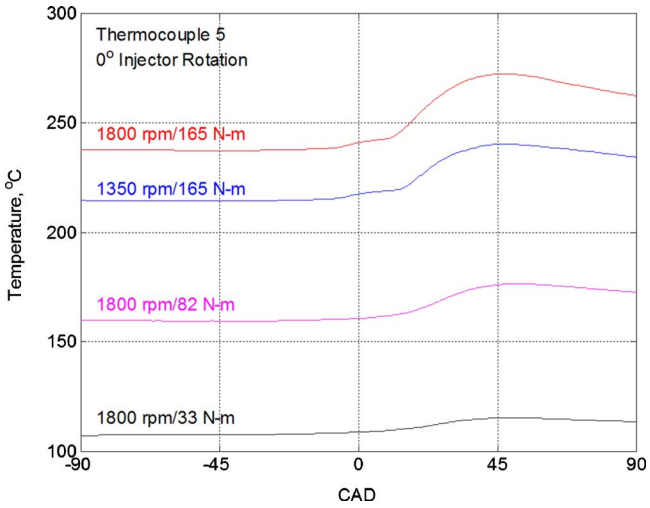


Fig. 15 Typical piston surface temperature traces: Thermocouple 5, 0 deg injector rotation

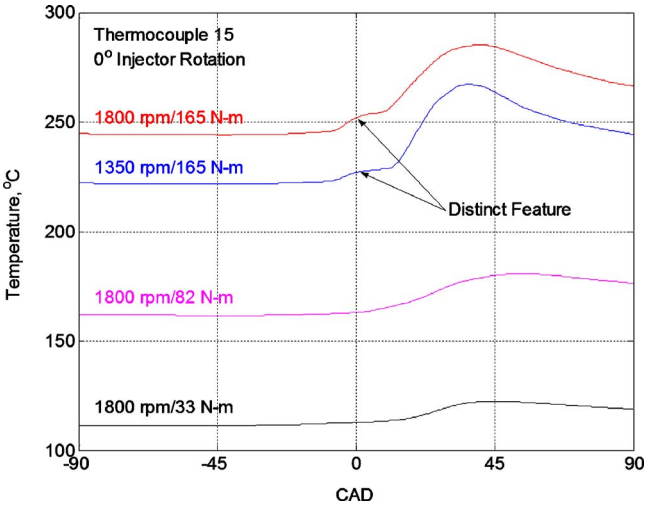


Fig. 16 Distinct feature in temperature trace: Thermocouple 15, 0 deg injector rotation

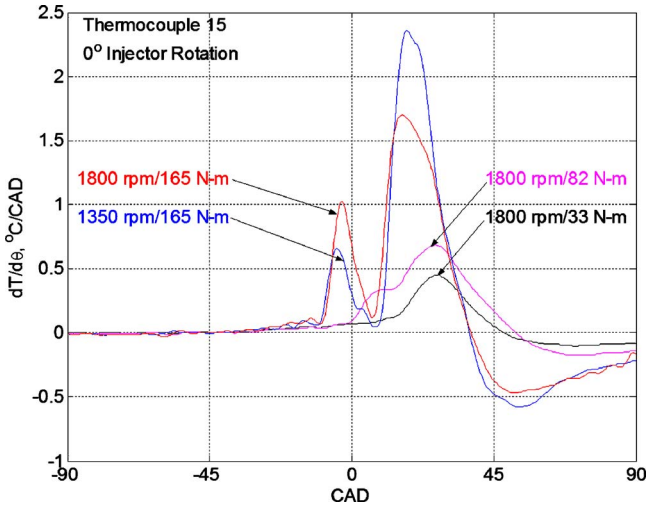


Fig. 18 $dT/d\theta$ identifies impingement at 1800rpm/165 N m and 1350 rpm/165 N m: Thermocouple 15, 0 deg injector rotation

ment location. The peak value of the impingement signature, relative to zero, was recorded for all test conditions. The smallest, measurable value of $dT/d\theta$ peak was 0.1. A threshold value of 0.15 was used to separate the data sets into impinging and nonimpinging conditions. Figure 19 shows a clear speed dependency between impinging and nonimpinging operating conditions [12].

The effect of impingement can be observed in the time domain and displays a unique signature at specific thermocouple locations, which do not exist at other locations on the piston surface. A simple parameter, $dT/d\theta$, effectively identifies the impingement event over a wide range of operating conditions and reveals a speed dependency on the impingement signature.

The instantaneous and steady-state piston thermal loading analysis involved the surface and the imbedded thermocouples to supply the boundary conditions for two-dimensional finite-element and one-dimensional semi-infinite solutions. Dynamic changes of temperature and heat flux exist several millimeters into the piston surface, with the highest temperatures observed at the surface [13]. A transient finite-element model was developed to solve for the temperature and heat transfer distribution throughout the piston cross section. The instantaneous surface heat transfer results from the finite-element model were then compared with the results from the one-dimensional semi-infinite solution. This was done to analyze the validity of one-dimensional semi-infinite behavior of the piston surface, a widely applied piston heat transfer assumption. A steady-state finite-element model, supplied with time-averaged temperature boundary conditions from the wireless telemetry, was also developed to calculate the distribution of heat transfer through different sections of the piston, such as the upper surface, rings, skirt, bottom of the piston, and oil gallery.

Two-dimensional (2D) effects are expected to be more dominant in locations such as the bowl lip compared to the center of the piston. The 1D and 2D results, as shown in Fig. 20 for the center of the piston [tc1], are almost identical. This proves that 2D effects at the center of the bowl do not significantly influence the heat transfer at this location.

However, a measurable difference in the heat transfer was observed for the piston bowl lip location [tc5] as shown in Fig. 21. The 1D semi-infinite solution overpredicts the peak heat flux to the piston lip for this particular case by approximately 14%.

One-dimensional semi-infinite heat transfer requires proper dissipation of the heat through the solid. In locations such as the bowl lip, heat transfer into the surface from both directions may cause the subsurface to reach higher steady periodic temperatures than the expected temperatures from the solution to the one-dimensional semi-infinite solid. Figure 22 shows the temperature

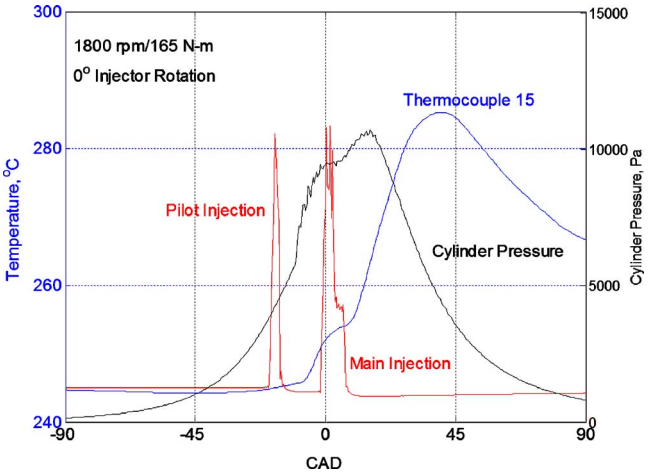


Fig. 17 Thermocouple 15, injection pulse and cylinder pressure relationship

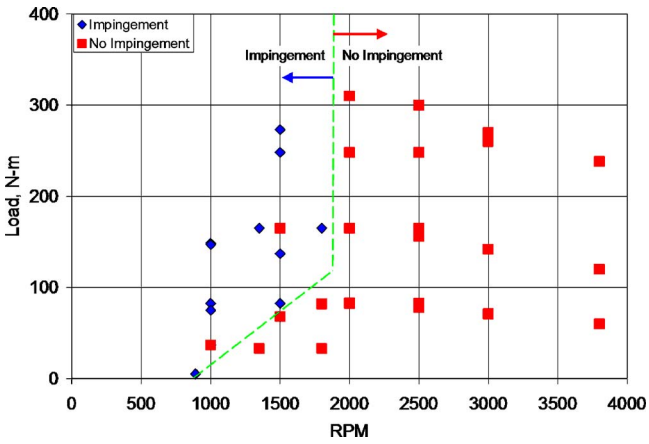


Fig. 19 Load versues rpm for impinging and nonimpinging conditions

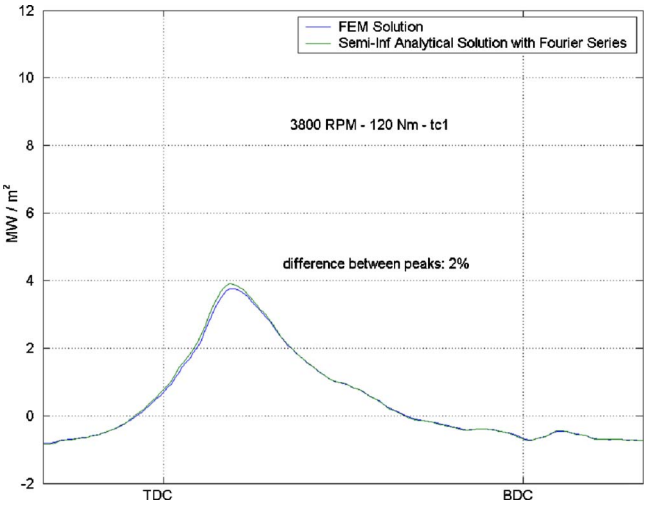


Fig. 20 Comparison of surface heat flux at thermocouple Location 1, for 2D FEM and 1D semi-infinite solutions

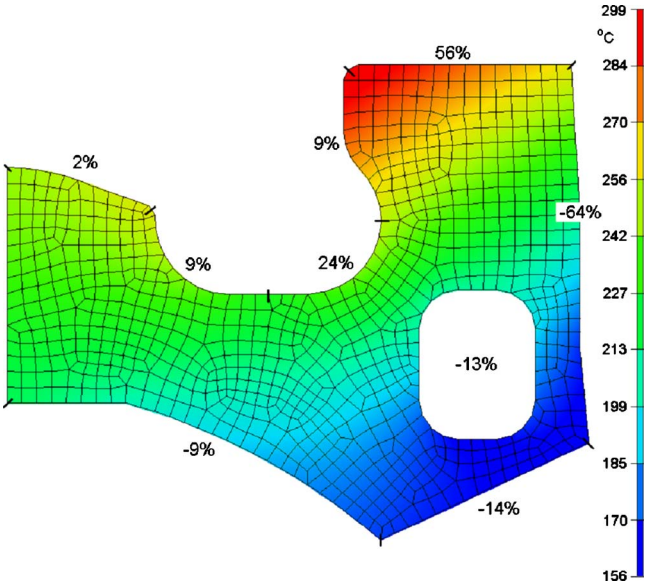


Fig. 22 Steady-state temperature and heat transfer (%) in a piston cross section

gradients from the steady-state, 2D, finite-element model and the distribution of the total steady-state heat transfer, in percent, at different sections of the piston.

Conclusions

1. The wireless microwave telemetry technique is an effective solution for measuring transducers in rotating and reciprocating components;
2. The high level of precision and accuracy with which measurements can be made in challenging locations has been demonstrated;
3. The wireless telemetry technique allows a continuous stream of data from the transducers, increasing component understanding; and
4. The system is extremely flexible for a variety of transducers, in a variety of applications, such as automotive torque converters and internal combustion engine pistons.

Acknowledgment

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science Laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paidup nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. Additional recognition is extended to Michigan Technological University and IR Telemetrics, Inc. for technical support and funding.

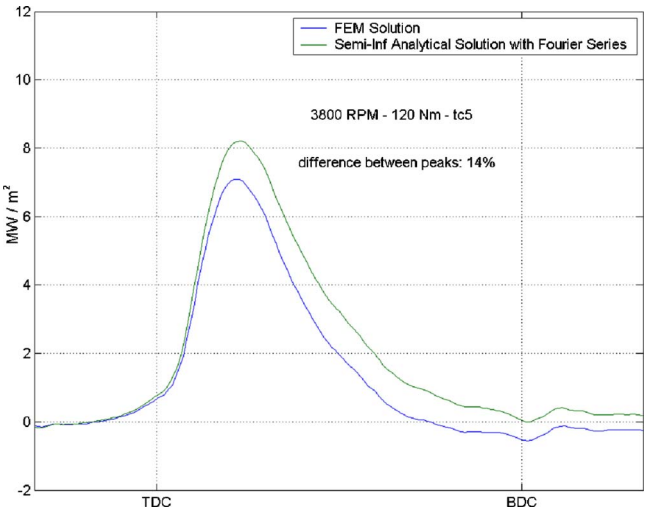


Fig. 21 Comparison of surface heat flux at thermocouple Location 5, for 2D FEM and 1D semi-infinite solutions

Nomenclature

ATF	=	automatic transmission fluid
CAD	=	crank angle degree
CFD	=	computational fluid dynamics
emf	=	electromotive force
FEM	=	finite-element method
FM	=	frequency modulation
F/V	=	frequency to voltage conversion
Hz	=	Hertz (cycles/s)
MSP	=	mean squared pressure
rpm	=	revolutions per minute
VCO	=	voltage controlled oscillator
V/F	=	voltage to frequency conversion

References

[1] Barna, G., 2004, IR Telemetrics, Inc.
[2] Lawrason, G., and Rollwitz, W., 1967, "A Temperature Telemetry Technique for Reciprocating Engines," SAE Technical Paper Series No. 670026, pp. 167–175.
[3] Burrahm, R. W., Davis, J. K., Perry, W. D., and De Los Santos, A., 1992, "Development of a Piston Temperature Telemetry System," SAE Technical Paper Series No. 920232, pp. 1–11.
[4] Barna, G., 1989, "An Infrared Telemetry Technique for Remote Piston Temperature Measurements in Internal Combustion Engines," M.S. thesis, Michigan Technological University, Houghton, MI.
[5] Assanis, D., and Friedmann, F., 1991, "A Telemetry Linkage System for Piston Temperature Measurements in a Diesel Engine," SAE Technical Paper Series No. 910299, pp. 1–12.
[6] Kato, N., Moritsugu, M., Shimura, T., and Matsui, J., 2001, "Piston Temperature Measuring Technology Using Electromagnetic Induction," SAE Technical Paper Series No. 2001-01-2027, pp. 1890–1896.
[7] Dong, Y., Dorivi, V., Attibele, P., and Yuan, Y., 2002, "Torque Converter CFD Engineering Part II: Performance Improvement through Core Leakage Flow and Cavitation Control," SAE Paper No. 020884.
[8] Schweitzer, J., and Gandham, J., 2002, "Computational Fluid Dynamics on Torque Converters—Validation and Application," *Proceedings 9th ISROMAC*.

[9] Ejiri, E., and Kubo, M., 1999, "Performance Analysis of Automotive Torque Converter Elements," ASME J. Fluids Eng., **121**, pp. 266–275.
[10] Watanabe, H., Kurshashi, T., and Kojima, M., 1997, "Flow Visualization and Measurement of Torque Converter Stator Blades Using a Laser Sheet Lighting Method and a Laser Doppler Velocimeter," SAE Paper No. 970680.
[11] Mekkes, J., Anderson, C., and Narain, A., 2004, "Static Pressure Measurements and Cavitation Signatures on the Nose of a Torque Converter's Stator Blades," Paper No. ISROMAC10-2004-035.
[12] Miers, S., 2004, "Impingement Identification and Characterization in a High Speed Diesel Engine using Piston Surface Temperature Measurements," Ph.D. dissertation, Michigan Technological University, Houghton, MI.
[13] Inal, M. K., 2005, "Thermal Loading and Surface Temperature Analysis of the Piston of a Small HSDI Diesel Engine," Ph.D. dissertation, Michigan Technological University, Houghton, MI.
[14] Miers, S., Anderson, C., Blough, J., and Inal, K., 2005, "Impingement Identification in a High Speed Diesel Engine using Piston Surface Temperature Measurements," SAE Trans. J. Mater. Manuf., **114–5**, pp. 845–851.