**Transactions of the ASME, Journal of Heat Transfer**
(ISSN 0022-1481) is published monthly 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 Heat Transfer, 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.


# Journal of Heat Transfer

**Published Monthly by ASME**

**VOLUME 133 • NUMBER 1 • JANUARY 2011**

***Special Issue on Heat and Mass transfer in Biosystems***

## GUEST EDITORIAL



## 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.

**Journal of Heat Transfer** **Volume 133, Number 1** **JANUARY 2011**



The ASME Journal of Heat Transfer is abstracted and indexed in the following:

*Applied Science and Technology Index, Chemical Abstracts, Chemical Engineering and Biotechnology Abstracts (Electronic equivalent of Process and Chemical Engineering), Civil Engineering Abstracts, Compendex (The electronic equivalent of Engineering Index), Corrosion Abstracts, Current Contents, E & P Health, Safety, and Environment, Ei EncompassLit, Engineered Materials Abstracts, Engineering Index, Enviroline (The electronic equivalent of Environment Abstracts), Environment Abstracts, Environmental Engineering Abstracts, Environmental Science and Pollution Management, Fluidex, Fuel and Energy Abstracts, Index to Scientific Reviews, INSPEC, International Building Services Abstracts, Mechanical & Transportation Engineering Abstracts, Mechanical Engineering Abstracts, METADEX (The electronic equivalent of Metals Abstracts and Alloys Index), Petroleum Abstracts, Process and Chemical Engineering, Referativnyi Zhurnal, Science Citation Index, SciSearch (The electronic equivalent of Science Citation Index), Theoretical Chemical Engineering*

Journal of Heat Transfer

# Guest Editorial

## Special Issue on Heat and Mass Transfer in Biosystems

The involvement of the heat transfer community in the research related to heat and mass transfer in biosystems increases rapidly and covers a wide variety of applications. While the past involved mainly topics of cryogenics and other related fields the current involvement is substantially wider and covers diverse biological fields and applications. The present special issue dedicated to heat and mass transfer in biosystems captures only a portion of this wide topical potential for biosystems research. The objective of this special issue is to report the state of the art on some of the research conducted in this field and motivate the heat transfer community, which is uniquely qualified to make a valuable contribution in providing fundamental understanding of transport of heat, mass, and momentum occurring in biological and biologically inspired systems, to become more involved in this field of research. In particular, applying the engineering way of thinking can be especially useful in developing the fundamental understanding, the governing principles, mechanistic explanations, and predictive modeling for functioning of biological systems. Application of the engineering point of view (through identification, formulation, and solution to bioengineering problems as well as designing machines and systems based on fundamental conservation laws) can usefully complement many biological research efforts and lead to new breakthroughs in this area. Among the topics covered in this special issue are targeted drug delivery, heat generation in nanoparticles with potential application to cancer treatment, tissue engineering by using irreversible electroporation to minimize tissue damage due to Joules heating, temperature distribution in cryosurgery, facilitating transdermal drug delivery via skin electroporation, modeling of microorganism growth including the lag phase, human eye response to thermal disturbances, modeling of bioheat transfer, formation of organelle traps in an axon providing a mechanistic explanation of the onset of Alzheimers disease, the use of micro-cantilever biosensors for detection of microorganisms, and the application of the constructal theory to biophysical systems.

The issue starts with a general topic of applications of the constructal law (this law postulates that the direction of systems evolution is such that it obtains easier access to currents that flow through it) in a paper by Bejan and Lorente. The paper considers applications of the constructal law to various biophysical systems. A novel method of targeted drug delivery (of an anti-cancer drug for example) to lung sites via utilization of a new smart inhaler system is then introduced by Kleinstreuer and Zhang. A correlation between the heat generation rate in nanoparticles (due to applying an alternative current at a specific frequency to a water-cooled coil, which thus generates an alternating magnetic field) and the density of the micro-CT images is presented by Attaluri et al. This correlation is important for understanding basic issues of using nanoparticles for cancer treatment. Evidence that parameters for irreversible electroporation can be chosen in such a way that minimizes the damage due to Joule heating to the tissue is demonstrated with the help of numerical modeling by Phillips et al.. Shitzer presents a comprehensive review of analytical and numerical studies on temperature distributions in cryosurgery; in vitro and in vivo experimental data used for validating these results are also discussed. The situations involving both surface application as well as insertion of cryoprobes are analyzed. Results of analytical modeling of biological media or heat management devices with a nonuniform geometry as a combination of convergent, uniform, and/or divergent configurations is presented by Mahjoob and Vafai. Kuznetsov develops a new model for the formation of organelle traps in an axon with regions where microtubule polarity has been reversed by the formation of microtubule swirls. Biological data link the formation of such microtubule swirl regions to traffic jams in axons that lead to various neurological disorders, such as Alzheimers disease. The developed theory demonstrates how the reversal of microtubule polarity results in the formation of organelle traps for both newly synthesized and used organelles, thus providing a mechanistic explanation of the onset of Alzheimers disease. Vadasz and Vadasz present a generic theoretical mechanism for microbial growth, including the lag phase that reveals the profound impact of the metabolic process on the growth. The theoretical model provides excellent matching to experimental data of microbial growth. Systematical reviews of four existing models of a human eye response to thermal disturbances are presented by Shafahi and Vafai. The physical phenomena included in the models, strengths and limitations of the models in simulating eye thermal response to various ambient conditions, blood temperature, and various thermophysical and biological parameters are discussed. Different models of bioheat transfer are introduced by Wang and Fan. Different approaches for obtaining constitutive relations for the heat flux, including Fourier law, Cattaneo −Vernotte theory, and dual-phase lagging theory are presented. The strengths and limitations of various bioheat models, including Pennes, Wulff, Klinger, Chen and Holmes, dual-phase-lagging, as well as some other models are discussed. A comprehensive overview of existing mathematical models of skin electroporation, which is a promising technique of causing a structural alternation in the uppermost skin layer, the stratum corneum, with the aim of facilitating transdermal drug delivery is presented by Becker. Special attention is given to developing a thermodynamic model of electroporation, which is based on the idea that a large current density through the pore causes lipids composing the stratum corneum to undergo a phase transition. Application of microcantilever biosensors for detection of microorganism is investigated experimentally by Tzeng et al. They explored the use of specific carbohydrate receptors for the functionalization of nanoparticles and demonstrated their binding specificities and their ability to mediate aggregations of targeted bacteria.

We would like to extend our special thanks to the authors of the special issue and to the reviewers for helping to elevate the quality of the papers. We are also thankful to the Editor, Professor Yogesh Jaluria, for providing this forum to discuss this emerging field. Special thanks to the Editorial Assistant, Shefali Patel, for the help in organizing this special issue.

**Peter Vadasz**

**Northern Arizona University,
Department of Mechanical Engineering,
P. O. Box 15600,
Flagstaff, Arizona 86011**

**Andrey V. Kuznetsov
North Carolina State University,
Department of Mechanical and Aerospace
Engineering, Campus Box 7910,
Raleigh, North Carolina 27695**

# The Constructal Law and the Design of the Biosphere: Nature and Globalization

**Adrian Bejan**[1]
Department of Mechanical Engineering and Materials Science,
Duke University,
Durham, NC 27708-0300
e-mail: abejan@duke.edu

**Sylvie Lorente**
Department of Civil Engineering,
Laboratory of Materials and Durability of Constructions,
University of Toulouse, INSA,
135 Avenue de Rangueil,
31077 Toulouse, France

*"Design in nature" is a topic of growing interest throughout science. The constructal law is the physics law of design generation and evolution in nature: "for a flow system to persist in time (to live), it must evolve such that it provides easier and easier access to its currents." In this paper, we show how the constructal law accounts for the main features of the design of the biosphere: global movement of mass as the action of constructal engines (geophysical, animal, and human made) that dissipate their power into brakes, animal locomotion, vision, cognition, and hierarchy. The architecture and hierarchy of vegetation results from the constructal tendency to generate designs that facilitate the flow of water and "the flow of stresses" (i.e., mechanical strength per unit volume). Natural porous media have multiple scales because their flows are also configured as trees. The paper concludes with the oneness of design in nature, global design, and science and technology evolution—all as manifestations of the natural tendency captured by the constructal law and unified constructal theory of evolution.*
[DOI: 10.1115/1.4002223]

*Keywords: constructal law, evolution, biosphere, locomotion, vision, cognition, hierarchy, vegetation, globalization*

## 1 The Constructal Law

Nature looks complicated but it is in fact a tapestry weaved in a very simple loom. The designs consist of many flow types and sizes, all governed by a simple law of physics: the constructal law,

> "For a finite-size flow system to persist in time (to live) it must evolve in such a way that it provides easier and easier access to the currents that flow through it" [1].

Even better, the tapestry itself is constituted (woven and sewn) according to the same law. All the designs fit, the animate and the inanimate, the small and the large, and the human made and the not human made [2–5]. They do not fit perfectly and never will. But, they exhibit a universal tendency in spite of their immense diversity. This, coupled with the fact that all the things that flow and move are free to morph, means that every thread and motif of the tapestry will morph so that the whole flows better.

The more we rise above the details, the simpler the tapestry design looks. Taking a bird's eye view is a very good medicine for those sickened by the poison that nature is complicated and random. In this essay, we sketch how the constructal law underpins design in nature and how it can guide our engineering and global design.

## 2 Winds, Animals, and Vehicles: Engines Connected to Brakes

Here is how the design principle reveals itself if we take a bird's eye view of the whole, which is the whole earth. Begin by imaging the Earth today in a steady state that covers a period longer than one human life. The simplest earth model is a ball with the same temperature all over (Fig. 1). We can predict important features of design in nature even with this extremely simple model.

The sun shoots streams of energy in all directions, radially. Some of these streams are intercepted and absorbed by the earth because the earth is opaque. All together, they represent one current of energy that flows out of the sun and into the earth. This current flows from sun to earth because the sun temperature is higher than the earth temperature. This flow direction (from high to low, always) is demanded by the second law of thermodynamics.

Next, the earth is warmer than the sky and this means that a current of energy must flow from earth to sky. Because the earth in this model is a closed system in steady state, the earth does not store or lose energy. This means that the sun → earth energy current must equal the earth → sky current. From this equality follows the conclusion that the earth must have a particular temperature (somewhere between the sun and sky temperatures) and that the earth temperature must be steady. The earth is just an intermediate node along the path of the current from sun to sky.

We see right away that the earth temperature is predictable and steady (i.e., it does not get out of control) as long as the sun → earth energy current is steady and known. The conclusion then is that the earth is endowed with a property we call "climate" and that climate is predictable if the simple model is correct.

We can predict much more about climate (temperature zones, winds, and day-night variation), because the earth is not a rigid ball [6]. The entire earth is flowing, especially in its spherical shells that house the designs that we observe—the hydrosphere, atmosphere, lithosphere, and biosphere. According to the constructal law, all these flowing things live (i.e., they survive and persist) by generating configurations and by evolving their configurations in time.

How do these flowing things fit inside the big system? The simple model in Fig. 1 is good for answering this question as well. The biggest system is the globe as an intermediate stop for a train of useful energy from the hot sun to the cold sky. The common animal too is an intermediate stop for its own train of useful energy (exergy and food) from meal to excrement.

Every thing that moves on earth does so because it is driven [1,2,7]. The driving is done by very subtle engines, one engine for each flow: muscles for animals, engines for vehicles, and im-

[1]Corresponding author.
Manuscript received March 18, 2010; final manuscript received May 11, 2010; published online September 27, 2010. Assoc. Editor: Peter Vadasz.

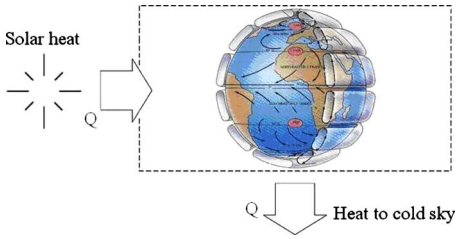



**Fig. 1 The solar heat current (Q) that hits the earth and ultimately sinks into the cold universe. The earth temperature settles at a steady level between the sun temperature and the sky temperature. It is as if the current Q passes through the earth as it flows from high temperature (sun) to low temperature (sky).**

mense wheels for the circuit executed by water in nature, which constantly puts fresh water upstream of the river basin. No matter how numerous and diverse, all these engines are driven by useful energy (fuel) that comes from high temperature (the sun) in the form of the heat current (Q) intercepted by the earth.

In Fig. 2, we imagine that all these engines are represented by one engine, which uses the heat input Q in order to produce the work W that is needed for driving the movement of all the things on earth. The difference between heat flowing in and work flowing out is Q-W, which is rejected as heat to the cold sky.

This completes the first part of the view, the subtle part, because we do not see "engines" in what moves around us. The winds and the rivers flow *by themselves*, we think. We are used to referring to these flows as "free convection" and "natural convection." These flows are thought to be free and natural because we do not pay for moving them. Yet, they really move and this means that engines drive them.

The second part of the design (animal locomotion, animal organs, and human vehicles) is the movement, which occurs against resistances that constantly try to stop this movement. Without such resistances the masses driven by the work W would be accelerating forever and spin out of control. This is not how nature works. All the driving W is dissipated in the brakes that form between the moving masses and their immediate surroundings. Nature flows in a balanced way: power generation is matched by power dissipation, anywhere and any time.

Key is the observation that all the work W is dissipated into heat (called $Q_{diss}$, and equal to W), and that $Q_{diss}$ is also rejected to the cold sky. All together, the earth rejects heat to the sky from the engines (Q-W) and from the brakes ($Q_{diss}$). The sum of these two

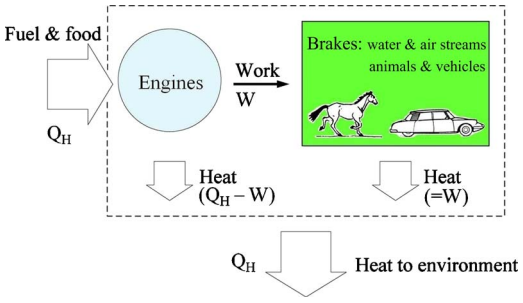



**Fig. 2 The biosphere on the move: the two phenomena that the solar heat drives as it passes through the earth [2,7]. First, Q drives flows (natural mechanisms with moving parts) that function as "engines" and produce work W. Second, the work dissipated in the "brakes" that form between these flows and their immediate environments (neighbors). Seen as a whole, the flowing earth (engines+brakes) receives Q from the sun and rejects Q completely to the sky. This is in agreement with the simpler model of Fig. 1.**

heat currents is the same as Q, because $Q_{diss}=W$. The total heat current rejected to the sky (Q) is the same as the heat current received from the sun.

This comples the bird's eye view of the second model (Fig. 2) and confirms the main feature of the first model (Fig. 1): the continuity of the heat current (Q) through the earth, from the sun to the sky. The material effect of this heat flow is that inside the engine+brake system (Fig. 2), mass has moved to a distance L such that the driving work W always scales as the weight moved (Mg) times the travel (L).

## 3 The Constructal Law of Evolution

We do not know whether the engines that drive the earth's flows are relatively efficient or not. It does not even matter. They are all imperfect, plagued by irreversibilities. What we know is that all the flows have the tendency to evolve in time (to generate designs) so that they flow more easily. This tendency is the constructal law and it can be restated in different ways depending on which side of the engines+brakes model we discuss.

First, if the flows and moving parts of an engine morph in time so that they move more easily, the engine design has the natural tendency to evolve in time toward producing more work (W) from the fixed heat input (Q). This is the direction of improvements in efficiencies—animal designs that are better fit for moving more animal mass on earth and geophysical currents (river basins and oceanic and atmospheric currents) that move more water and air mass more quickly and farther. The constructal evolution of the engines is in accord with statements we hear in all the evolutionary sciences, in biology and engineering. Improvements can be described thermodynamically as a procession of configurations that offer progressively less dissipation in the engines part, i.e., lower rates of entropy generation and higher efficiencies.

Second, on the brakes side of the earth design, the evolution toward more W from the fixed Q means an evolution toward more $Q_{diss}$. The brakes evolve toward more dissipation and higher rates of entropy generation. This conclusion sounds confusing (after reading the preceding paragraph), yet it is a statement heard nowadays in geophysics. What geophysicists say (higher dissipation) is the complete opposite of what animal design and engineering scientists say (lower dissipation). The conflict between the two camps is real [8,9] but it is put to rest by the engines+brakes design of Fig. 2. The two camps refer to different parts of the grand design. Both tendencies—toward lower and higher dissipation—are manifestations of the constructal law, which unifies them. Without the evolution of flow configurations in time, there is no evolution toward less dissipation in some flow systems and more dissipation in others.

In sum, the constructal law and the global design sketched in Figs. 1 and 2 constitute a unified theory of evolution. This view is unifying because it explains and predicts evolution in all the diverse domains in which evolutionary phenomena are observed, recorded, and studied scientifically: animal design, river basins, turbulent flow, animal movement, athletics, technology evolution, and global design [1–10]. Some of the most common animate and inanimate examples are sketched in Fig. 3.

Evolution means design modifications, in time. How these changes are happening are *mechanisms* and should not be confused with the constructal law or the constructal theory. In the evolution of biological design, the mechanism is mutations, biological selection, and survival. In geophysical design, the mechanism is soil erosion, rock dynamics, water-vegetation interaction, and wind drag. In sports evolution, the mechanism is training, recruitment, mentoring, selection, and rewards. In technology evolution, the mechanism is liberty, freedom to question, innovation, education, trade, and emigration.

What flows through a design that evolves is not nearly as special in physics as how the flow system generates and improves its configuration in time. The "how" is the physics principle—the constructal law. The "what" are the mechanisms and they are as

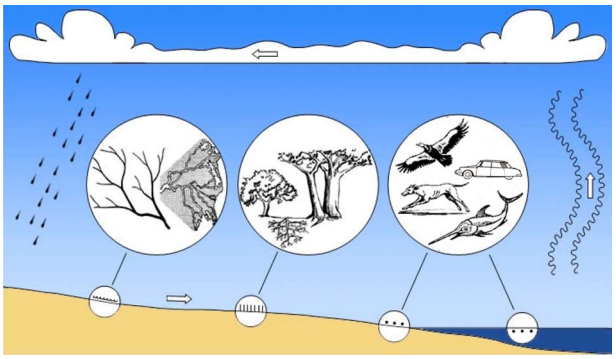


**Fig. 3 The physics phenomenon of design generation and evolution facilitates the circuit executed by water in nature: tree-shaped river basins and deltas, vegetation, and all the features of animal design and human and machine species [10]**

diverse as the flow systems themselves. The what are many and the how is one. Hierarchy more simple than this does not exist.

Having "impact" on the environment is synonymous with having design in nature. There is no part of nature that does not resist the flows and movements that attempt to penetrate it. Movement means penetration and its name differs depending on the direction from which the phenomenon is observed. To the observer of river basins, the phenomenon is the emergence and evolution of the dendritic vasculature. To the observer of the landscape, the phenomenon is erosion and the reshaping of the earth's crust. No flow on earth is more responsible for the evolution (shaping) of the landscape than the flow of water in rivers.

This mental viewing of design generation and environmental impact as a unitary design in nature is universally applicable. Think of the migration paths for animals, versus the riverlike paths and burrows dug into the ground. Think of the migration of elephants versus the toppling of trees. The same holds for all the movements that define our societal existence. The patterns of social dynamics go hand-in-glove with the impact on the environment.

The effect of life is *measurable* in terms of the mass moved over distances during the life time of the flow system [7]. The work required to move any mass on earth (vehicle, river water, and animal mass) scales as the weight of that mass times the distance to which it is moved horizontally, on the landscape. It is this way with the life of the river basin and the animal and it is the same with the life of man, family, country, and empire. The economic activity of a country is all this movement of mass (people and goods) to distances. Because each movement is proportional to the amount of fuel burned in order to drive it, the entire economic activity on a territory must be proportional to the amount of fuel consumed on that territory. This finally explains the observation [11] that the annual Gross National Product (GNP) of a country is roughly proportional to the amount of fuel burned (i.e., the useful energy generated and dissipated) in that country.

## 4 Locomotion, Vision, and Cognition

In politics, history, and sociology, one observes and speaks of the increasing speed of everything—faster modes of transportation and communication, the acceleration of technology, social change, and the pace of life. People feel that they are running out of time, even though the technological changes generate more free time for everybody. More time to do what?

In geography, economics, and urbanism, one observes and speaks of humanity generating more space. This continuing phenomenon is known as expansion, globalization, and the spreading of city living in three dimensions, on the landscape and the vertical direction, upward and downward. People complain about lack

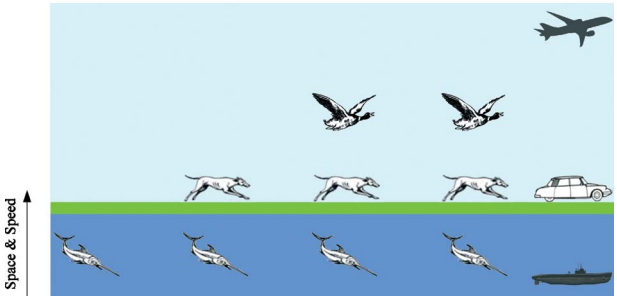



**Fig. 4 Space, speed, and time: The evolution of the biosphere from prehistory to today [12]. Animal flow has been spreading vertically in space and toward higher speeds, longer ranges, and better vision. This montage fits on the left plane of Fig. 5.**

of space because of all the construction sites, even though the construction technology is generating space for people. More space for what?

Better and better language, writing and science also give us more time to think. To think about what? To think about more activity, more movement, more flow of us on the surface of the earth. The same answer holds for the questions of why more time and more space.

With the constructal law, these seemingly unrelated and paradoxical tendencies constitute a universal phenomenon of generation of greater flow access and the evolution of design in nature. They are predictable. They are to be expected because they have been an integral part of design in nature forever.

Increasing speed and spatial expansion are a unitary phenomenon. Animals have been expanding in space, in this unmistakable time direction: from sea to land and later, from land to air (Fig. 4) [12]. The human and machine species are continuing this trend by invading the upper atmosphere, the ocean depths, and the outer space. The same movie (because this is what the occurrence and evolution of design is, a time sequence of images) means that speeds have been increasing in time and will continue to increase. For the same body mass, the runners are faster than the swimmers and the fliers are faster than the runners (Fig. 5). This movie is the same as the evolution of inanimate mass flows, for example, the river basins. Under the persisting rain, all the channels morph constantly to flow more easily.

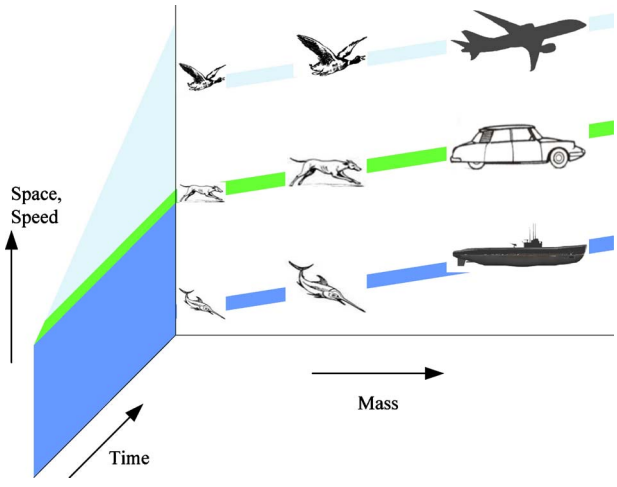



**Fig. 5 Space, speed, mass, and time: The third dimension (mass) of the diversity in the flow of animal mass [12]. At any point in time, the biosphere churns itself with a huge *diversity* of animate bodies organized according to a *pattern*. The larger bodies tend to have higher speeds lower body frequencies and larger forces.**

Design is the speed governor of nature. None of the changes observed in politics, history, sociology, animal speed and river speed are spinning out of control. None of the expansions feared in geography, economics, and urbanism are slamming into a brick wall.

In accord with the constructal law, all the flow systems evolve their configurations in time toward *equilibrium flow structures* [13] that flow more and more easily through their finite-size (i.e., surviving) environments. Animate design together with inanimate design is a unifying story where increasing speeds represent the same evolution phenomenon as increasing access to space. In the river basins, the slower tributaries move more and more water mass into the faster and bigger channels. The same holds true for animal design. If we imagine body sizes that are roughly the same, then, broadly speaking (compare Fig. 5), runners are faster than swimmers and flyers are faster than swimmers and runners

This is in accord with the constructal law because it is aligned with the time arrow of how life has spread on earth, sea→land →air, and not the other way around. The time direction of this evolution has been toward higher speeds and it is shown qualitatively in Fig. 4, which is a detail of the side plane of Fig. 5. More movement and more mixing of the earth (toward more space) have always been aligned with time, more speed, and more space traveled per unit of animal mass and useful energy consumed.

The big jump in the perfecting of the animal locomotion design was the emergence of the organ for vision (the eye). This has created *guided locomotion*—a flow of animal mass that is much more efficient, faster, and enduring because with vision and cognition the flow of animal mass selects for itself ceaselessly better channels to flow: straighter and safer with fewer obstacles and predators. The time direction of this change in the animal locomotion design is in accord with the constructal law, toward more space, speed, and mixing of the earth's crust. The animal design with vision and cognition came after the animals without vision and cognition, not the other way around.

Vision and cognition are one and they are demanded by the constructal law applied at the scale of the earth. Animal mass and water mass (in river basins) are flow systems with configurations and rhythms that facilitate the flow of mass on earth. From this physics view of the constructal law followed all the scaling rules of animal locomotion [14] and the evolution of speed limits in athletics—running and swimming [15]. The constructal law predicted that larger animals must have higher speeds calculated with complete formulas (slope and intercept) in which speeds are proportional to their body masses raised to the power 1/6 and body movement frequencies (stride, flapping, and fishtailing) are in proportion to body masses raised to $-1/6$. These predictions agree with all the known speed-mass data for flyers, land animals, and swimmers. This is suggested qualitatively in the back plane of Fig. 5. The actual speed-mass data are available in Refs. [2,8,14].

Locomotion design is a manifestation of the constructal law and it has been perfecting itself throughout the history of biological forms and flow systems on earth (i.e., throughout "big history"). Another phenomenon that illustrates this tendency is the human preference for rectangular shapes with aspect ratios that resemble the golden ratio [12]. These facilitate the scanning of images and their transmission through vision organs to the brain. The speeding up of this flow goes hand-in-hand with all the other configurations that facilitate the same movement, for example, the dendritic architectures of the nervous system in the eye and the brain. Dendrites maximize the rate of point-volume flow of information inside finite volumes and the rate at which new point-volume connections can occur naturally in the brain. The name of this constructal evolution of brain architecture, every minute and every moment, is cognition—the phenomenon of thinking, knowing, and thinking again, better. "Getting smarter" is the constructal law in action.

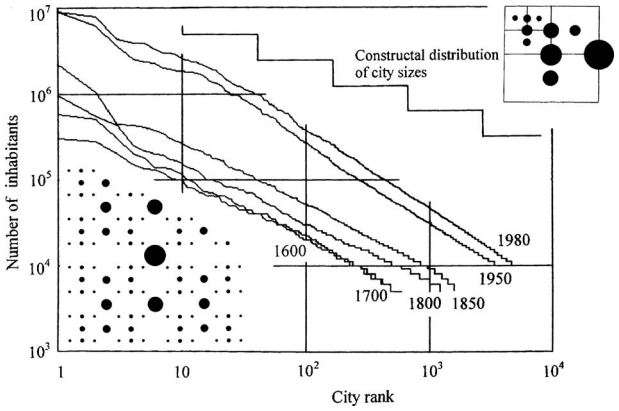



**Fig. 6 City sizes versus city rank in Europe during 1600–1980 and the Zipf distribution predicted with the constructal law [17]**

## 5 Natural Hierarchy

Age matters in evolutionary design and it is good for performance. The river basin positions its channels better and better and the channels stay in place. The channels have hierarchy: a few large channels flow in harmony with the many small channels. A sudden downpour is served well by the "memory" built into the old river beds.

From the mental viewing provided by the constructal law, the hierarchies that are visible in all the flow systems that cover the globe can finally be deduced. These architectures form a multiscale weave of point-area and area-point tree flows, all superimposed and all sustaining everything that flows (i.e., everything that lives) on earth. One example is the hierarchy of channel sizes and numbers in all the river basins that have been cataloged. From the constructal law, we deduced that the number of tributaries that feed a larger channel should be 4 [16]. This prediction is in very good agreement with Horton's empirical correlation of river numbers, which states that the observed number of tributaries falls in the range between 3 and 5.

Another constructal hierarchy is the distribution of city sizes and numbers (of cities of the same size) on large areas such as a continent, Fig. 6 [17]. The distribution is linear when plotted log-log. This line with slope in the range between $-1/2$ and $-1$ is known as a Zipf distribution and it is found empirically in virtually all the natural flow systems that connect discrete points with finite areas or volumes.

The descending line in Fig. 6 was predicted by recognizing the flow access between two populations that live on each area construct (small and large) that covers the landscape. This is suggested by the inset in the upper-right corner of the figure [17]. On every area construct (white) there are two populations that exchange flows: those who live on the area and those who live in a settlement (village, town, and city), which is shown as a black dot. The constructal law also predicts that the straight line must shift upward in time while remaining parallel to itself because of technology improvement and those who live on the area can achieve flow equilibria with larger and larger numbers of people living in the settlement. This too is in agreement with the history of the size-rank distribution of cities over the last 4 centuries.

Another natural hierarchy anticipated with the constructal law is the ranking of tree sizes and numbers in forests (Fig. 7) [10]. The descending bands of size versus rank data were deduced by arranging tree canopies of many sizes on the forest floor such that the entire floor facilitates the flow of water, from the ground to the blowing wind. Two examples of such arrangements (triangular and square) are shown in the upper-right corner. The slope and intercept of the size-rank line is insensitive to the type of arrangement. Important is how the multiscale canopies fill the forest floor area such that the water flow rate from the whole area is greater.

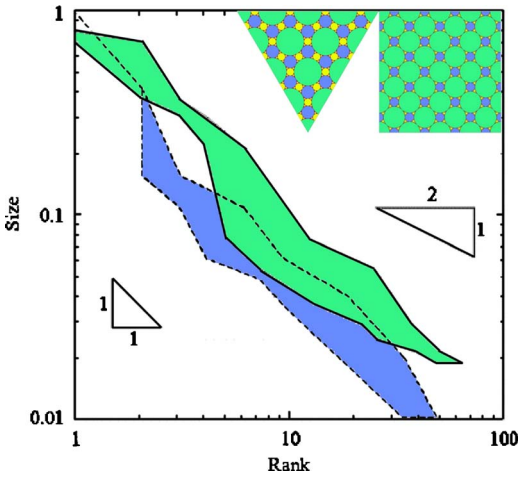



**Fig. 7 Distribution of tree canopy sizes versus rank in the constructal design of the forest floor [10]. The Zipf distribution is insensitive to the pattern (e.g., triangular versus square) in which the multiscale tree canopies are arranged on the forest floor.**

From this holistic view of design generation come the numerous and seemingly random scales of trees in the forest and the Zipf-type alignment of the size versus rank data.

What the constructal law predicted for multiscale river basins, demography, and forests also applies to the design of societal flow. Science and higher education flow through a natural tissue of universities, each university being connected to the entire globe. The older universities have dug the first channels, which are now some of the largest channels that irrigate the student landscape. "Largest" does not mean largest number of bodies moving in and out of classrooms. Largest are the streams of the most creative, i.e., the channels that attract the *individuals* who generate new ideas and who develop disciples who produce and carry new ideas farther on the globe and into the future. The swelling student population is served well by the "memory" built into the education flow architecture. From this theoretical view followed the prediction that the hierarchy of universities should not change in significant ways [18]. This hierarchy is as permanent as the hierarchy of the channels in a river basin. It is natural because it is demanded by the entire flow system (the globe) in which huge numbers of individuals want the same thing (knowledge).

## 6 The Flow of Stresses: Vegetation Design

Trees are flow architectures that emerge during a complex evolutionary process. The generation of the tree architecture is driven by many competing demands. The tree must catch sunlight, absorb $CO_2$, and put water into the atmosphere while competing for all these flows with its neighbors. The tree must survive droughts and resist pests. It must adapt, morph, and grow toward the open space. The tree must be self-healing to survive strong winds, ice accumulation on branches, and animal damage. It must have the ability to bulk up in places where stresses are higher. It must be able to distribute its stresses as uniformly as possible so that all its fibers work hard toward the continued survival of the mechanical structure.

On the background of this complexity in demands and functionality, two flows stand out. The tree must facilitate the flow of water and the flow of stresses (i.e., it must be strong mechanically). The demand to pass water is made abundantly clear by the geographical correlation between the presence of trees and the rate of rainfall (Fig. 3). This correlation is meant "broadly speaking" because the broad view is the key to discovering the pattern, in spite of the obvious and overwhelming diversity (random effects, local geographical variation, winds, and atmospheric humidity). The demand to pass water is also made clear by the dendritic architecture, which according to the constructal law facilitates flow access between one point and a finite-size volume. The demand to be strong mechanically is made clear by features such as the tapered trunks and limbs with round cross section and other designlike features identified in this article. These features of design in solid structures facilitate the *flow of stresses* [4,10]. Avoiding strangulations in the flow of stresses is synonymous with the natural phenomenon of generation of mechanical strength.

We relied solely on the constructal law in order to discover all the main features of vegetation, from root and canopy to forest [4,10]. We took an *integrative* approach to trees as live flow systems that evolve as components of the larger whole (the environment, Fig. 3). We treated the plant and the forest as physical flow architectures that evolve together toward greater mechanical strength against the wind and greater access for the water flowing through the plant. Theoretical features derived from the constructal law are the tapered shape of the root and longitudinally uniform diameter and density of internal flow tubes, the near-conical shape of tree trunks and branches, the proportionality between tree length and wood mass raised to the power 1/3, the proportionality between total water mass flow rate and tree length, the proportionality between the tree flow conductance and the tree length scale raised to a power between 1 and 2, and the existence of forest floor plans that facilitates ground-air flow access, e.g., Fig. 7. The constructal law also predicted that there must exist a characteristic ratio of leaf volume divided by total tree volume, trees of the same size must have a larger wood volume fraction in windy climates, and larger trees must pack more wood per unit of tree volume than smaller trees.

## 7 Alternating Trees: Multiscale Porous Media

Natural porous flow structures also exhibit multiple length scales that are distributed nonuniformly through the available space. Such multiscale flow structures are anticipated based on the constructal law [19]. We showed this by exploring the flow properties of the dendritic flow architecture proposed in Fig. 8. The idea is to connect two parallel lines (or two parallel planes) with trees that alternate with upside down trees. The resulting dendritic pattern connects the bottom boundary of the flow domain with the top boundary.

This alternating sequence of point-to-line trees constitutes a vasculature between the two parallel boundaries of the porous body. The fluid flows in the same direction through all the trees, e.g., upward in Fig. 8. The flow access between the points of one line and the points of a parallel line can be viewed as a sequence of point-to-line flow access structures. The building block with which Fig. 8 is constructed was proposed by Lorente et al. [20], where it was based on optimally shaped rectangular areas, each area allocated to one channel. We then compared the tree structure (Fig. 8) with a reference architecture: an array of N equidistant parallel tubes perpendicular to the two lines, each tube with diameter D. This reference structure carries the same total flow rate in the same total flow volume and over the same area. The structure has one degree of freedom, the tube diameter D, or the number of parallel tubes. The pressure drop along the reference structure $(\Delta P_{\text{ref}}/\dot{m})$ is derivable analytically. When the d spacing is the same as the spacing between parallel channels the two global flow resistances form the ratio $\Delta P_{\text{trees}}/\Delta P_{\text{ref}} \cong 14/2^n$. When the number of branching levels is four or larger, the tree-shaped architecture offers greater access to the flow that permeates through the porous structure. The superiority of the tree design increases fast as n increases: when $n=7$, the ratio $\Delta P/\Delta P_{\text{ref}}$ is as low as 1/10.

When the available flow scales are sufficiently small (d), the flow architecture should have trees, not parallel channels, i.e., not a single scale. From this follows the prediction of the *multiscale* and *nonuniform* character of natural porous media: large numbers of small pores and few large pores (known as "pipes" in hill slope

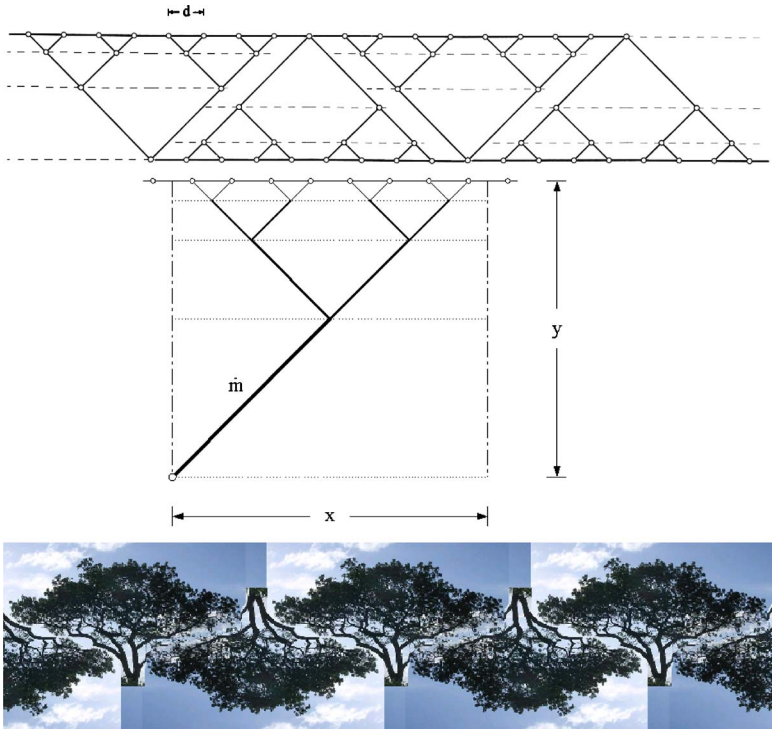



**Fig. 8 The origin of multiscale porous media. The flow from plane to plane encounters considerably less resistance through tree-shaped structures than the flow through parallel channels [19].**

hydrology). This prediction is crucial because it means that the apparent "diversity" and "randomness" are consistent with and predictable from the constructal law.

The observer who looks from the outside at the porous medium (e.g., from the top of Fig. 8) sees a few large pores surrounded by many small pores. From this viewing position, the porous medium appears to have only two main scales, which is why natural porous media are also described as "bidisperse." The design features uncovered based on the constructal law are qualitatively the same as those of natural porous materials (e.g., the soil of the hill slope in a river drainage basin where two scales dominate: fine porous soil with seepage and larger pores (pipes) embedded in the fine structure). The constructal law also predicts that "vascular" designs must occur in nature and that they must be stepwise more complex as they become larger [21].

## 8 Globalization Design: Distributed Energy Systems

Seen through the lens of the design law presented in this essay, the future design of globalization is clear. The bulk of future research in the energy area will be devoted to four fronts:

1. the development of fuel resources
2. energy conversion methods
3. the improvement of devices based on current fuels and conversion methods, and
4. the global design of fuel consumption

Most important but least recognized is the fourth front because it resonates in the daily debates about globalization, sustainability, and environmental impact. These global-scale phenomena are important, and they are hotly debated because they happen and because they threaten to happen. Work on the fourth front adds a fundamental component to any energy initiative in government, which has enormous impact on science, education, and industry.

The global energy design is a complex flow system driven by fuels that are consumed. The generated power moves things on the surface of the earth (goods, people, and information), compare Fig. 2. The flow system is a tapestry of nodes of production of power embedded in areas populated by users and environment, all linked and all sweeping the earth with their movement. Constructal design [4] is showing that the whole basin is flowing better (with fewer obstacles globally) when the production nodes and the channels are allocated in certain ways to the covered areas (the environment). This is how the globe becomes a live system—a living tissue—and why its best future can be designed based on principle. With the constructal law, this design can be pursued predictively.

The distribution, allocation, and consumption of power should be considered together on fronts 1–4, as equal partners. This holistic view includes fields, such as housing and transportation, building materials, heating and air conditioning (i.e., the "energy design" of building), lighting, water distribution, etc., (Fig. 9). In the university, it serves as a healthy unifier of mechanical, civil, and electrical engineering with environmental science.

Taken together, all these concerns allow the global design to emerge with balance, or harmony, between the fuel streams that contribute to the global power stream that drives all of our society on earth. How much fossil fuel versus renewables will be a natural feature of the global design, such as the size of the organ on the

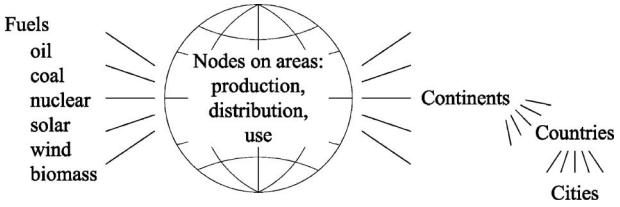



**Fig. 9 The global design as a multitude of distributed energy systems of all scales, designed and interwoven in accordance with the constructal law**

animal—a result, not assumed, or taken off the shelf. The global design will emerge as a construct from elements (homes) to continents (Fig. 9).

## 9 Why We Need Engineering Science

In this essay, we showed that with the constructal law, we can explain and predict design in nature—animate, inanimate, and engineered. The law unites.

Why is this important to know? Here are two answers.

First, we can design better vehicles—a more efficient movement of humanity on earth (e.g., Figs. 2 and 3). We can do this more directly, faster, better, and with greater confidence because now we know the law that captures the design evolution, which otherwise would require enormous numbers of designers (chance events in natural evolution and entire armies and industries that develop technologies), this in addition to time and money.

Second, with engineering science, we predict the future. Figure 5 is just one example of how.

(a) The evolution of the flow of animal mass has taken the animal design toward performance levels that persist in time.
(b) The evolution of every vehicle is pointing toward the designs that look more and more animal-like.

Viewed in time, the design of every technology constitutes a movie scripted according to (a) and (b). The movie tape runs in one direction, toward shapes and rhythms that improve in time and make the movement easier and the solid structures lighter, easier to carry.

With the science of engineering, we do not have to labor to see the entire movie. We "get it" from the first few frames and our minds leap to design images that otherwise would have taken an eternity to discover by trial-and-error. With engineering science, we design the future with us in it. The name of this grand evolutionary design is civilization, science, and education.

## Acknowledgment

This research activity was supported by the National Science Foundation and the Air Force Office of Scientific Research.

## References

[1] Bejan, A., 1997, *Advanced Engineering Thermodynamics*, 2nd ed., Wiley, New York.
[2] Bejan, A., and Lorente, S., 2006, "Constructal Theory of Configuration Generation in Nature and Engineering," J. Appl. Phys., **100**, p. 041301.
[3] Reis, A. H., 2006, "Constructal Theory: From Engineering to Physics, and How Flow Systems Develop Shape and Structure," Appl. Mech. Rev., **59**, pp. 269–282.
[4] Bejan, A., and Lorente, S., 2008, *Design With Constructal Theory*, Wiley, Hoboken, NJ.
[5] Bejan, A., Lorente, S., Miguel, A. F., and Reis, A. H., 2009, *Constructal Human Dynamics, Security, and Sustainability*, IOS, Amsterdam, The Netherlands.
[6] Reis, A. H., and Bejan, A., 2006, "Constructal Theory of Global Circulation and Climate," Int. J. Heat Mass Transfer, **49**, pp. 1857–1875.
[7] Lorente, S., and Bejan, A., 2010, "Few Large and Many Small: Hierarchy in Movement on Earth," Int. J. Design & Nature and Ecodynamics, **5**, pp. 254–267.
[8] Bejan, A., and Marden, J. H., 2009, "The Constructal Unification of Biological and Geophysical Design," Phys. Life. Rev., **6**, pp. 85–102.
[9] Bejan, A., and Lorente, S., 2010, "The Constructal Law of Design Generation and Evolution in Nature," Philos. Trans. R. Soc. London, Ser. B, **365**(1545), pp. 1335–1347.
[10] Bejan, A., Lorente, S., and Lee, J., 2008, "Unifying Constructal Theory of Roots, Canopies, and Forests," J. Theor. Biol., **254**, pp. 529–540.
[11] Ayres, R. U., Ayres, L. W., and Warr, B., 2003, "Exergy, Power, and Work in the US Economy, 1900–1998," Energy, **28**, pp. 219–273.
[12] Bejan, A., 2009, "The Golden Ratio Predicted: Vision, Cognition, and Locomotion as a Single Design in Nature," Int. J. Design & Nature and Ecodynamics, **4**, pp. 97–104.
[13] Bejan, A., and Lorente, S., 2004, "The Constructal Law and the Thermodynamics of Flow Systems With Configuration," Int. J. Heat Mass Transfer, **47**, pp. 3203–3214.
[14] Bejan, A., and Marden, J. H., 2006, "Unifying Constructal Theory for Scale Effects in Running, Swimming, and Flying," J. Exp. Biol., **209**, pp. 238–248.
[15] Charles, J. D., and Bejan, A., 2009, "The Evolution of Speed, Size, and Shape in Modern Athletics," J. Exp. Biol., **212**, pp. 2419–2425.
[16] Bejan, A., Lorente, S., Miguel, A. F., and Reis, A. H., 2006, "Constructal Theory of Distribution of River Sizes," *Advanced Engineering Thermodynamics*, 3rd ed., A. Bejan, ed., Wiley, Hoboken, NJ.
[17] Bejan, A., Lorente, S., Miguel, A. F., and Reis, A. H., 2006, "Constructal Theory of Distribution of City Sizes," *Advanced Engineering Thermodynamics*, 3rd ed., A. Bejan, ed., Wiley, Hoboken, NJ.
[18] Bejan, A., 2007, "Why University Rankings Do Not Change: Education as a Natural Hierarchical Flow Architecture," Int. J. Design & Nature and Ecodynamics, **4**, pp. 319–327.
[19] Lorente, S., and Bejan, A., 2006, "Heterogeneous Porous Media as Multiscale Structures for Maximum Flow Access," J. Appl. Phys., **100**, p. 114909.
[20] Lorente, S., Wechsatol, W., and Bejan, A., 2002, "Tree-Shaped Flow Structures Designed by Minimizing Path Lengths," Int. J. Heat Mass Transfer, **45**, pp. 3299–3312.
[21] Kim, S., Lorente, S., Bejan, A., Miller, W., and Morse, J., 2008, "The Emergence of Vascular Design in Three Dimensions," J. Appl. Phys., **103**, p. 123511.

# Optimal Drug-Aerosol Delivery to Predetermined Lung Sites

**Clement Kleinstreuer**[1]
Department of Mechanical and Aerospace Engineering and Department of Biomedical Engineering,
North Carolina State University,
Raleigh, NC 27695;
University of North Carolina,
Chapel Hill, NC 27599
e-mail: ck@eos.ncsu.edu

**Zhe Zhang**
Department of Mechanical and Aerospace Engineering,
North Carolina State University,
Raleigh, NC 27695

*This review summarizes computer simulation methodologies of air-particle flow, results of drug-aerosol transport/deposition in models of the human respiratory system, as well as aspects of drug-aerosol targeting and associated inhalation devices. After a brief introduction to drug delivery systems in general, the required modeling and simulation steps for optimal drug-aerosol delivery in the lung are outlined. Starting with medical imaging and file conversion of patient-specific lung-airway morphologies, the air-particle transport phenomena are numerically solved for a representative inhalation flow rate of* $Q_{total} = 30$ *l/min. Focusing on microspheres and droplets, the complex airflow and particle dynamics, as well as the droplet heat and mass transfer are illustrated. With this foundation as the background, an overview of present inhaler devices is presented, followed by a discussion of the methodology and features of a new smart inhaler system (SIS). With the SIS, inhaled drug-aerosols can be directly delivered to any predetermined target area in the human lung.* [DOI: 10.1115/1.4002224]

*Keywords: subject-specific lung airways, air-particle flow, computer simulation results, inhaler devices, optimal drug-aerosol targeting, smart inhaler system*

## 1 Introduction and Overview

Drug delivery systems (DDSs) for optimal targeting of pathological sites, such as tumors, are in the midst of a strong development, testing, and application cycles. They may substantially mitigate side-effects by not depositing aggressive drugs onto healthy tissue and reduce cost by lowering the necessary dosage of expensive medicine. Novel devices rely on image-guided, magnetic, or mechanical delivery systems. The development of such drug delivery methodologies and subsequently, device prototyping require computational analysis of complex transport phenomena as well as experimental and clinical device testing. Of the numerous DDS applications, an optimally targeted drug-aerosol delivery via inhalation to predetermined lung areas is becoming a very desirable treatment option not only for combating respiratory disorders but also for systemic diseases. Respiratory treatment applications include chronic obstructive pulmonary disease, asthma, cystic fibrosis, bacterial infection, and lung tumors. Because of the rapid drug absorption from the deep lung region into the systemic system and the ease-of-administration, the inhalation of medical aerosols is also been used for the treatment of diabetes and cancer.

Common oral drug delivery devices include metered-dose inhalers, dry-powder inhalers (DPIs), and nebulizers (see Refs. [1–4] among others). Most common for drug-aerosol delivery is the pressurized metered-dose inhaler (pMDI) where modern versions use a hydrofluoroalkane (HFA) propellant and dose counter. DPIs are breath-actuated, single, or multidose devices. Nebulizers convert solutions into liquid aerosols of suitable size for nasal inhalation or deeper lung penetration. All inhalers generate rather low particle deposition efficiencies in the lung and are *nondirectional*, i.e., drug-aerosol landing sites cannot be controlled. For example, Kleinstreuer et al. [5] analyzed the popular pMDI with a spacer, modified nozzle, and HFA propellant. They demonstrated that up to 46.6% of the inhaled polydisperse droplets may reach the lungs, assuming a steady flow rate of $Q_{in}$ equal to 30 l/min. While present inhalers may deliver some of the medicine to the central lung region, most treatments require a high-percentage drug-aerosol deposition at specific lung sites, such as tumors, or in desired lung regions, e.g., inflamed airways in asthma patients. Some drugs are so aggressive and/or expensive that targeted aerosol delivery is imperative especially for children with asthma to avoid drug delivery to noninflamed lung areas. Clearly, targeted drug-aerosol delivery to desired lung areas required for the treatment of a specific disease is becoming a more prevalent treatment option because it is fast, convenient, and with reduced topical side-effects.

"Targeting" is understood here as a physical goal to bring drug-aerosols from the inhaler exit to a predetermined landing area in the respiratory system for maximum medical effectiveness, needed either locally or systemically [6]. In order to achieve that goal, future inhaler devices have to operate based on a targeting methodology having the following five requirements:

(1) optimal physical microparticle characteristics, i.e., diameter and density
(2) ideal inhalation waveform, i.e., steady, laminar flow rate
(3) optimal particle-release position
(4) sufficient drug dosage for effective treatment
(5) ease-of-use

This mechanistic approach differs from image-guided drug delivery (IGDD), which is more of a visualization approach rather than a drug targeting methodology. IGDD is achieved via real-time visualization of the injected microscale or nanoscale drug carriers, which cruise in the blood stream, with special design of molecular, carrier, and controlled-release properties when reaching the target (e.g., a tumor). Such a drug release can be triggered by high-frequency ultrasound and/or biochemical activation, causing selective distribution characteristics and pharmacokinetic behavior possibly leading to improved therapy [7–10]. A more tumor direct delivery of drugs can be achieved with magnetic nanoparticles, which are intra-arterially injected and then, guided toward the target and concentrated via a high-gradient external magnetic field. However, magnetic targeting is more effective for tumors that are near the body surface and in regions of slow fluid flow. In any case, the implementation of targeted drug delivery and optimal operational conditions require the combination of modern image processing techniques, computational fluid-particle dynamics simulations, and innovative mechanical designs.

The following sections provide a review of the state-of-art computer simulation methodologies as well as results of drug-aerosol

[1]Corresponding author.
Manuscript received March 31, 2010; final manuscript received April 28, 2010; published online September 27, 2010. Assoc. Editor: Peter Vadasz.

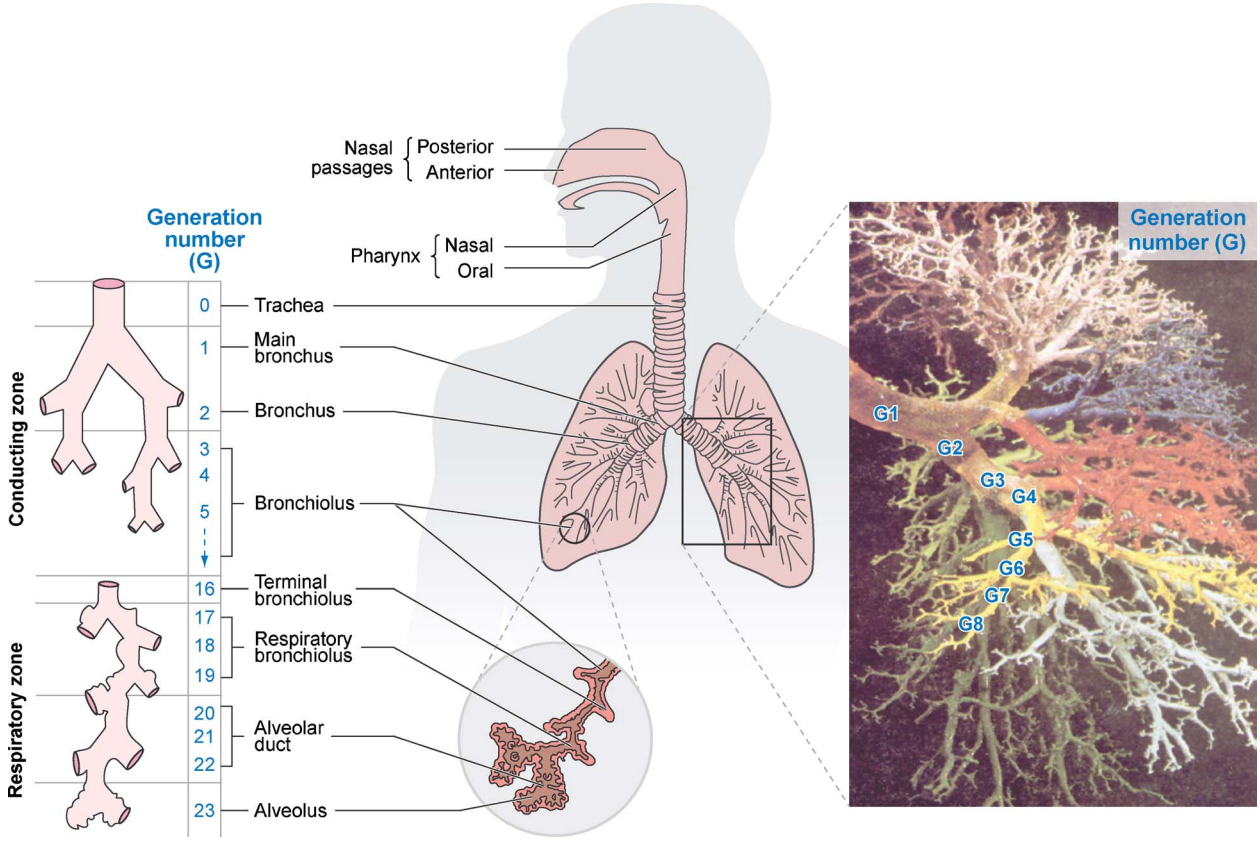



**Fig. 1 Schematics of the human respiratory system [6]**

transport and deposition in the human respiratory system, drug-aerosol targeting methodology, and inhalation devices.

## 2 Airway Models and Transport Phenomena

**2.1 Subject-Specific Airway Models.** In optimal drug-aerosol delivery, once an experimentally validated computer simulation model has been developed, operational parameters for, say, a smart inhaler system (SIS) can be computed to conduct patient-specific therapy. Hence, accurate and realistic airway models are the necessary precursor for experimental or computational airflow and particle transport/deposition analyses (see Fig. 1).

There are two modern approaches for generating geometric lung models for computational/experimental analyses, i.e., either via algorithms, which specify inhaled, volume-based rules for the relationships between parent and daughter airways (e.g., Refs. [11,12]), or lung replicas digitally reconstructed from computerized tomography (CT) scans or magnetic resonance imaging (MRI) data (e.g., Refs. [13,14]). Using the first approach, Tebeckhorst et al. [15] also included in their computer model lung-airway morphogenesis due to development in children or triggered by lung diseases.

As alluded to, the ultimate goal is *patient-specific modeling*, which one day will be applicable to relevant problem areas in biomedical engineering. Specifically, through employing computer simulations supported by experimental data sets, patient-specific modeling would gain quantitative results and valuable physical insight for best possible medical management (e.g., Ref. [16]), including optimal drug delivery. The first step is to generate patient-specific airway geometry data files. Computer models of realistic, patient-specific respiratory segments, i.e., mouth and nasal airways, throat and trachea, and upper lung-airway models (i.e., presently of high resolution up to generation 6, see Fig. 1) can be exactly reconstructed using an image processing computer software converting CT/MRI images (i.e., DICOM files) obtained from a radiologist or surgeon. Specifically, image processing software, e.g., SCANIP and SCANFE, from SimpleWare (Exeter, UK) and Mimics/Geomagic (Materialise, Belgium) can be used to read a stack of two-dimensional medical images, such as CT scans and/or MRI, to reconstruct *subject-specific* respiratory airway models [17]. Reconstructed three-dimensional model geometries can be exported as stereolithography (STL) format files for mesh generation as needed for computer modeling and simulations. Usually, the STL file describes an unstructured triangulated surface by the unit normal and vertices of the triangles using a three-dimensional Cartesian coordinate system.

### 2.2 Airflow, Heat, and Mass Transfer Simulations

*2.2.1 Airflow.* The airflow dynamics of the respiratory tract is always unsteady and driven by the pressure differences under the action of the cyclic breathing process. The incompressible Navier–Stokes equations can be employed to characterize airflow primarily in the upper region. At normal breathing rates, the airflow through the oral airway region and the first few generations is incipient turbulent, becoming laminar at about the sixth to eighth generations and remaining so thereafter [18]. Accurate prediction of the airflow field at respiration-onset is a prime challenge in developing/executing a computational fluid dynamics (CFD) transport/deposition model. Specifically, the transient multiscale vortical velocity vector field induced while in transit through the mouth, pharynx, throat, and upper tracheobronchial branches will persist well into the lower branches and hence, represents a dominant flow feature. Therefore, the physics model must provide autoprediction of "turbulent" flow initiation but must be equally accurate for unsteady pressure-driven laminar flow prediction. In this case, large eddy simulation (LES) and time-averaged Reynolds tensor models (RANS) with relatively better performances (e.g., shear stress transport (SST) transition model) can be employed [19]. For LES, wall-adapted local eddy-viscosity (WALE)

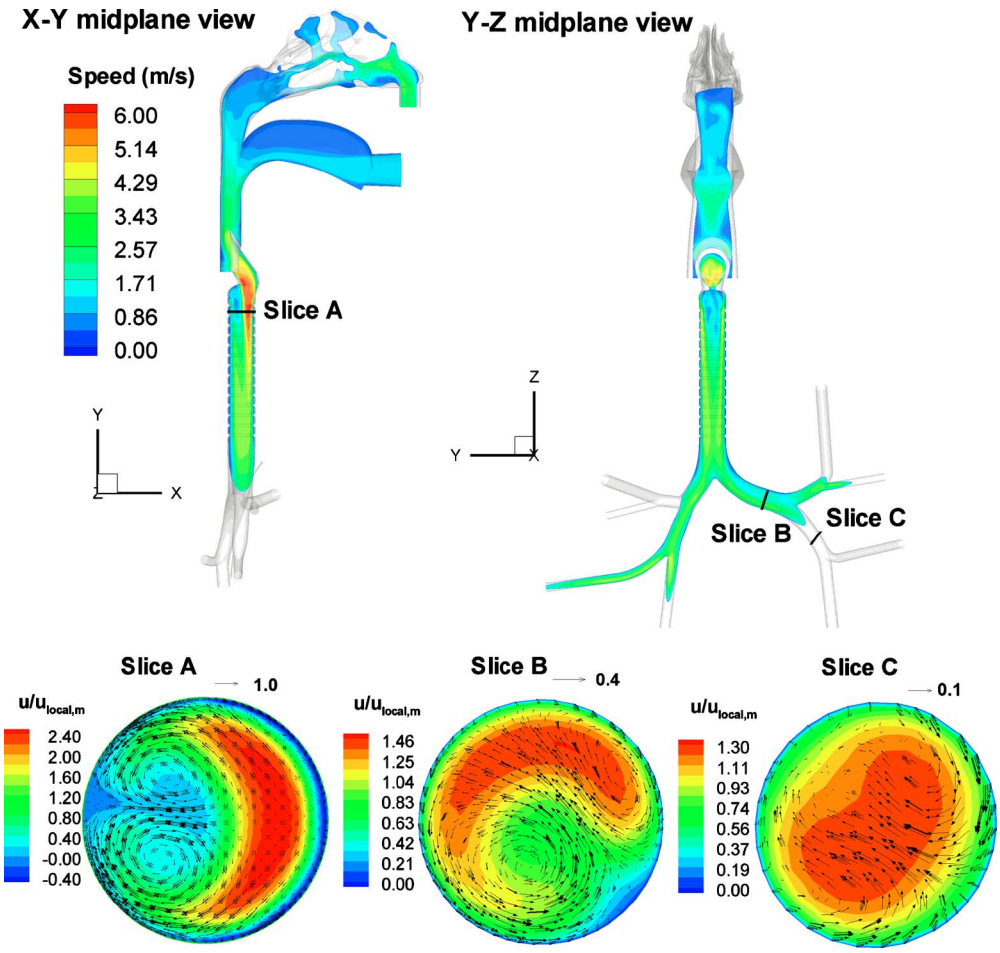



**Fig. 2 Example of airflow structures in a combined human nasal-oral-tracheobronchial airway model at a constant inspiratory flow rate of $Q_{nasal,in}$=15 l/min and $Q_{oral,in}$=15 l/min. The local mean velocities ($u_{local,m}$) of slices A, B, and C are 2.5 m/s, 2.18 m/s, and 2.66 m/s, respectively.**

model may be employed to provide the subgrid-scale (SGS) viscosity. The LES-WALE model is an algebraic model, such as the Smagorinsky model, but it overcomes some known deficiencies of the Smagorinsky model [20]. Zhang and Kleinstreuer [19] compared different RANS and LES models and found that the overall performance of LES, the low Reynolds number (LRN) k-$\omega$ model and the SST transition model, do not have measurable differences in predicting laminar flows and transition to turbulent flow while the SST transition model may give a better prediction of turbulence kinetic energy profiles in some cases. LES can provide instantaneous velocity fluctuations, which may be significant for turbulent microdroplet transport/deposition in human upper airways. However, it requires hundred-fold more computational time than RANS turbulence models.

The essential airflow structures in the human respiratory system were reviewed in Kleinstreuer and Zhang [21]. Figure 2 shows an example of airflow structures in a combined nasal-oral-tracheobronchial airway models with simultaneous nasal and oral breathing ($Q_{nasal,in}$=15 l/min and $Q_{oral,in}$=15 l/min). The main flow features in the human upper airway tract can be summarized as follows. It should be noted that these observations hold for actual transient airflow as well.

- Skewed velocity profiles generated by the centrifugal force are observed in the curved geometric portions, e.g., from the oral/nasal cavity to the pharynx/larynx and from the parent tube to the daughter tubes in the bronchial airways.
- Recirculating flows are created in the regions with abrupt geometric changes (e.g., contraction and expansion). Such regions include the tongue and upper portion of the oral/nasal cavity, the narrow portion of the pharynx, the epiglottis, the entrance to the trachea, as well as some bifurcation areas.
- Asymmetric laryngeal jets and impinging flows are formed due to the acceleration around the structures of the larynx.
- Secondary motions are set up due to the upstream flow histories as well as the centrifugal force induced pressure gradient when the flow turns in bends. The structures and magnitudes of secondary flow fields depend on flow direction, Reynolds number, geometric features, axial location with respect to curved portions, as well as the Womersley number for oscillatory flows. Different distinct secondary vortices may be observed in different locations (Fig. 2). In some cases, two distinct secondary vortices (typical Dean flow) appear in the tube (e.g., slice A). However, in bronchial airways, subsequent bifurcations may constantly interrupt the development of secondary flows (see slices B and C), generating different secondary vortex structures with further development from upstream.

*2.2.2 The Heat and Mass Transfer*. Some pharmaceutical droplets or dry powders may undergo size changes due to evaporation or condensation (labeled hygroscopicity) when being inhaled through the respiratory tract. Many studies have shown that the functional dependence of deposition fraction on particle size shifts toward smaller size particles with increasing hygroscopicity and relative humidity (see Refs. [22–25] among others). For example, hygroscopicity can substantially increase the size and

hence, the airway deposition fraction of drug particles with initial dry sizes of about 0.5–2 $\mu$m [24]. Accompanied by droplet shrinkage or growth is usually the conditioning process of the inspired air, which may affect droplet transport and deposition. Hence, the investigation of heat and moisture transport in human airways is very important in clinical studies for determining drug-aerosol trajectories. There has been a number of reports published attempting to quantify heat and water vapor transfer in the human airways. For example, Johnson et al. [26] obtained a convective heat and mass transfer coefficient $h_c$ and $h_m$, averaged over three bronchial generations based on in vivo measured airway temperatures and water vapor contents of several mechanically respired mongrel dogs. However, the correlation extrapolated from the canine measurements should be carefully examined before it is interpreted for humans. Nuckols et al. [27] carried out a series of convective heat and water vapor transfer experiments with a cast replica of human upper airways extending from the mouth and nose to the trachea as well as with a mechanical bronchial airway model representing generations G0 to G2 after Weibel [28]. Considering inspiration only, he expressed his results in terms of Nusselt and Sherwood numbers averaged over the entire cast from the mouth to trachea and the bronchial airway model. Both Johnson et al. [26] and Nuckols et al. [27] did not derive inspiratory convective heat and mass transfer coefficients for each individual generation of the upper bronchial airways. Daviskas et al. [29] developed a time-dependent model based on a single differential equation with an analytical solution as well as experimental data to predict the intra-airway temperature and water vapor contents as a function of air residence time at each location. Tawhai and Hunter [30] modeled the water vapor and heat transfer in the normal and the incubated airways employing unsteady, one-dimensional (axial) transport equations where a power-law method was used to describe the radial distribution of temperature, water vapor, and air velocity. The airway geometries were approximated as cylindrical segments. However, most abovementioned mathematical models and numerical simulations employed empirically derived heat and mass transfer coefficients to estimate transfer processes between resident air and airway walls or focused only on 1D and 2D water vapor and temperature distributions in the human airways. More recently, Zhang and Kleinstreuer [31], Zhang et al. [32], and Elad et al. [33] computationally simulated 3D heat transfer as well as coupled water vapor transport phenomena in the human upper airways.

For calculating the temperature distributions, the energy equation can be written as

$$\frac{\partial(\rho T)}{\partial t} + \frac{\partial}{\partial x_j}(\rho u_j T) = \frac{\partial}{\partial x_j}\left[\left(\frac{\lambda}{c_g} + \frac{\rho \nu_T}{\sigma_T}\right)\frac{\partial T}{\partial x_j}\right] \tag{1}$$

where $u_j$ and $x_j$ are the velocity and space coordinates, respectively, $\rho$ is the density, $c_g$ is the specific heat of the gas, $T$ is the temperature, $\lambda$ is the thermal conductivity of the gas phase, and $\sigma_T$ is the turbulent Prandtl number taken as 0.9.

The mass transfer equation for water vapor in the airway stream can be written as

$$\frac{\partial Y}{\partial t} + \frac{\partial}{\partial x_j}(u_j Y) = \frac{\partial}{\partial x_j}\left[\left(\tilde{D} + \frac{\nu_T}{\sigma_Y}\right)\frac{\partial Y}{\partial x_j}\right] \tag{2}$$

where $Y$ is the mass fraction, $u_j$ is the velocity, $\nu_T$ is the turbulent viscosity, $\sigma_Y = 0.9$ is the turbulence Schmidt number for $Y$, and $\tilde{D}$ is the molecular diffusivity of water vapor in air. The effect of temperature change on diffusivity is expressed by the following semi-empirical correlation [32]:

$$\tilde{D} = 2.16 \times 10^{-5}(T/273.15)^{1.8} \quad (\mathrm{m^2/s}) \tag{3}$$

where $T$ is the air temperature. The water vapor concentration ($Y$) can also be expressed in terms of the relative humidity ($RH$). The relative humidity at an air-water interface is defined as the actual water vapor concentration divided by that under saturation conditions, i.e.,

$$RH = \frac{\rho_v}{\rho_{\mathrm{v,sat}}} = \frac{Y\rho_a}{(1-Y)\rho_{\mathrm{v,sat}}} \tag{4}$$

where $\rho_a$ is the density of dry air and $\rho_{\mathrm{v,sat}}$ is the density of saturated water vapor, which can be obtained from Ref. [34]

$$\rho_{\mathrm{v,sat}} = 3.638 \times 10^5 \exp(-4943/T) \tag{5}$$

Regional variations in wall temperature were employed as the thermal boundary condition for air at the air-wall interface. The wall temperatures in human airways can be described as [29]

$$T_w = T_m + (37 - T_m)[1 - e^{-(0.00564\dot{V}_E + 0.336)L/\dot{V}}] \tag{6}$$

where $T_m$ is the temperature at the lips, which is close to the temperature at the end of expiration, $L$ is the cumulative length at the middle of each airway, $\dot{V}$ is the airflow rate in the corresponding airway, and $\dot{V}_E$ is the ventilation rate.

The mass fraction of water vapor at the wall ($Y_w$) was modeled assuming variable saturation conditions at $T_w$, i.e.,

$$Y_w = \frac{RH_{\mathrm{wall}} \cdot \rho_{\mathrm{v,sat}}(T_w)}{RH_{\mathrm{wall}} \cdot \rho_{\mathrm{v,sat}}(T_w) + \rho_a(T_w)} \tag{7}$$

Considering the vapor pressure at the wall in equilibrium with the blood serum, $RH_{\mathrm{wall}}$ is assumed to be 99.5% [29].

Figure 3 presents an example of temperature distributions in the human upper airways. Clearly, the axial velocity profile is a good indicator for temperature distributions. Specifically, a higher temperature is observed near flow separation as well as stagnation and low velocity zones because convection heat transfer is relatively low in these regions and heat cannot be dissipated easily. The air warms up more rapidly in the nasal cavity than in the oral cavities due to the enhanced heat transfer with their more complicated geometric features and larger surface areas. For example, the air may almost reach the body temperature in the bifurcation airway G3. Both inhalation flow rate and inlet air temperature may influence the air temperature variations in the upper airways [31].

The distributions of water vapor concentrations are similar to those of temperature, which were described by Zhang et al. [32]. An example of the quantitative variations of water vapor concentrations in the upper airway model (mouth to end of trachea) under different inspiratory flow conditions is displayed in Fig. 4 in terms of cross-sectional, area-averaged relative humidity ($RH$). Inhalation flow rate as well as air temperature and relative humidity at the mouth inlet are three important factors influencing the variations of water vapor concentrations in human airways [32]. With the transport of water vapor from the airway wall to the unsaturated air stream, the $RH$ increases gradually from the mouth to the trachea and the upper bronchial tree. The lower the inlet values ($RH_{\mathrm{in}}$), the higher the increase rate for the $RH$ due to the larger vapor gradients and wall vapor fluxes (see Fig. 4). The $RH$ variation is larger with higher inlet air temperature because the saturation vapor pressure at the liquid/gas interface increases with temperature. The averaged $RH$ in the airway conduits with a lower inlet air temperature ($T_{\mathrm{in}}$) increases axially much more rapidly than at a higher $T_{\mathrm{in}}$ because of the lower saturated vapor pressure corresponding to the lower temperature (see Fig. 4). The $RH$ values after generation G3 are in the range of 81–100% when the inlet conditions are $20\% \leq RH_{\mathrm{in}} \leq 80\%$, $Q_{\mathrm{in}} \leq 30$ l/min, and $T_{\mathrm{in}} \leq 303$ K. Of interest is that the maximum $RH$ value can reach about 104% (i.e., supersaturation) in the pharynx/larynx region when starting with an initial air temperature of 283 K and an $RH$ of 80%. Supersaturation was also reported in the theoretical work by Ferron et al. [35] for nasal breathing at a low inlet air temperature but at a higher inlet $RH$. Longest et al. [36] reported that the supersaturated conditions can be achieved in the respiratory tract through the inhalation of warm ($T > T_{\mathrm{body}}$) saturated air. The satu-

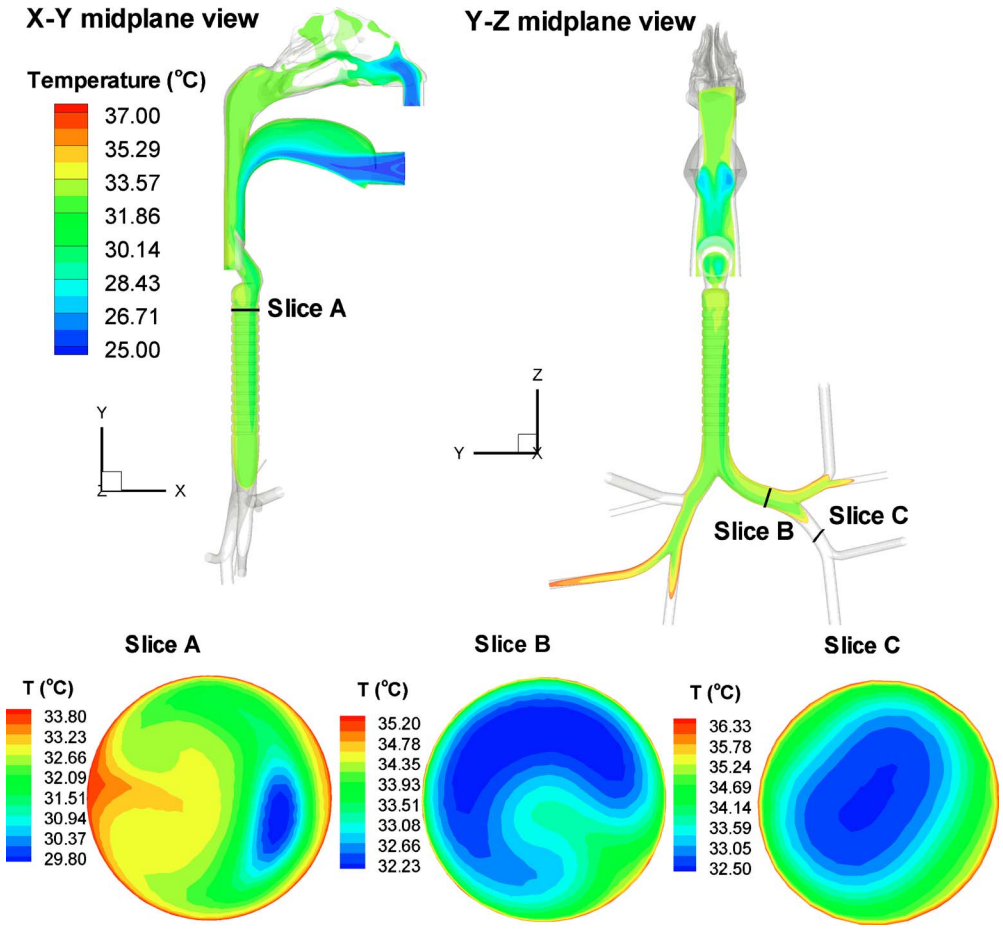



**Fig. 3 Example of air temperature distributions in a combined human nasal-oral-tracheobronchial airway model at an inlet temperature of 25°C and a constant inspiratory flow rate of $Q_{nasal,in}$=15 l/min and $Q_{oral,in}$=15 l/min**

rated or supersaturated water vapor in the upper airways may enhance condensational growth of submicrodrug-aerosols and then, change their delivery sites and efficiencies [36].

### 2.3 Solid Particle and Droplet Transport and Deposition

*2.3.1 Solid Particle or Droplet Transport.* For any given inlet concentration of typically noninteracting, effectively spherical drug-aerosols, a Lagrangian frame of reference can be employed for the trajectory computations of solid particles or droplets. In light of the large particle-to-air density ratio, negligible rotation, and small thermophoretic forces, drag and Brownian motion for microparticles and submicroparticles are the dominant point forces away from the walls [37]

$$\frac{d}{dt}(m_p u_i^p) = \frac{f_e}{8}\pi\rho d_p^2 C_{Dp}(u_i - u_i^p)|u_j - u_j^p| + F_i^{\text{Brownian}} + F_i^{\text{gravity}} + F_i^{\text{lift}} \tag{8}$$

Here, $\rho$ is the gas (air) density, $u_i^P$ and $m_p$ are the velocity and mass of the droplet, respectively, $f_e$ is the effective drag force correction for nonspherical droplets, $C_{DP}$ is the drag force coefficient, and $u_i$ is the instantaneous fluid velocity with $u_i=\bar{u}_i+u_i'$ where $\bar{u}_i$ is the time-averaged or bulk velocity of the fluid and $u_i'$ is its fluctuating component. $F_i^{\text{gravity}}$ is the gravity and $F_i^{\text{Brownian}}$ is the interaction force due to Brownian motion that results from collision of air molecules with the droplets of $d_p<1$ $\mu$m. The Brownian force can be modeled as a Gaussian white noise random process [38]. Shear induced particle lift $F_i^{\text{lift}}$ can be considered based on the three-dimensional numerical expression described in Longest et al. [39]. The effect of the lubrication force or near-wall drag modifications are expected to be neglected for the aerosol system of interest here [39].

*2.3.2 Multicomponent Droplet Evaporation or Condensation Growth of Droplets or Dry Powders.* In light of the complex liquid phase aerosols, this analysis assumes that the temperature and composition are maintained spatially uniform inside each droplet but evolve with time and the droplet remains spherical during vaporization [40]. Considering the convective heat and mass transfer over the whole surface of the spherical droplets, the change in droplet mass can be expressed as [32]

$$\frac{dm_p}{dt} = -\sum_{s=1}^{m}\int_{\text{surf}} n_s dA \approx -\sum_{s=1}^{m}(\bar{n}_s \cdot A) \tag{9}$$

where $m_p$ is the droplet mass and $n_s$ is the mass flux of each chemical species at the droplet surface, which can be calculated as [41,42]

$$\bar{n}_s = \frac{\rho_g \text{Sh}\tilde{D}_s}{d_p} C_m \, 1\frac{1 - Y_{s,\infty}}{1 - Y_{s,\text{surf}}} \tag{10}$$

where $\tilde{D}$ is the vapor diffusivity in the air, Sh is the Sherwood number $\text{Sh}=(1+\text{Re}_p\,\text{Sc})^{1/3}\max[1,\text{Re}_p^{0.077}]$, and $Y_{s,\text{surf}}$ and $Y_{s,\infty}$ are the gas phase mass fractions of each component on the droplet surface and far from the droplet, respectively. Clearly, evaporation

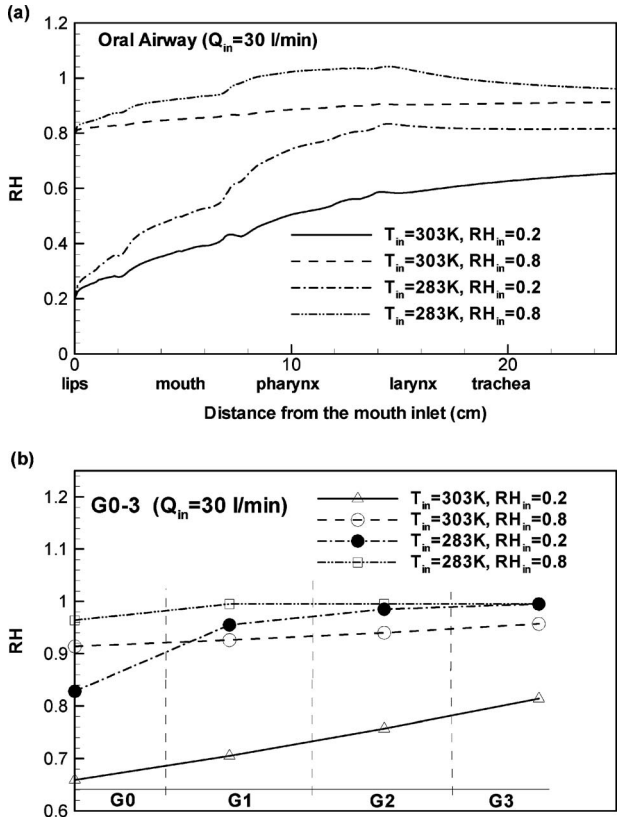



**Fig. 4 Variations of cross-sectional averaged relative humidity with $Q_{in}$=30 l/min in (*a*) the oral airway model and (*b*) the bifurcation airway models G0–G3. Adapted from Zhang et al. [32].**

occurs when $Y_{s,surf}$ is greater than $Y_{s,\infty}$ while condensation growth happens if $Y_{s,surf}$ is less than $Y_{s,\infty}$.

The coupled heat transfer equation for liquid droplets reads

$$m_p c_{pd}(dT/dt) = \pi d_p k_g C_T \mathrm{Nu}(T_a - T_p) - \sum_{s=1}^{m} \int_{\mathrm{surf}} n_s \gamma_s dA \quad (11)$$

where $c_{pd}$ is the liquid specific heat, $d_p$ is the droplet diameter, $k_g$ is the thermal conductivity of gas mixture, $T_p$ and $T_a$ are the temperatures of droplet and surrounding air, respectively, $\gamma_s$ is the latent heat of vaporization, Nu is the Nusselt number $\mathrm{Nu}=(1+\mathrm{Re}_p\,\mathrm{Pr})^{1/3}\max[1,\mathrm{Re}_p^{0.077}]$, $\gamma$ is the latent heat of vaporization, and $C_m$ and $C_T$ are the correction factors (i.e., Knudsen number corrections) considering the noncontinuum effects, specifically [34,43]

$$C_m = \frac{1+\mathrm{Kn}}{1+\left(\frac{4}{3\alpha_m}+0.377\right)\mathrm{Kn}+\frac{4}{3\alpha_m}\mathrm{Kn}^2} \quad (12)$$

$$C_T = \frac{1+\mathrm{Kn}}{1+\left(\frac{4}{3\alpha_T}+0.377\right)\mathrm{Kn}+\frac{4}{3\alpha_T}\mathrm{Kn}^2} \quad (13)$$

Here, Kn is the Knudsen number defined as $\mathrm{Kn}=2\lambda/d_p$ with $\lambda$ being the mean free path of the gas surrounding the droplets and $\alpha_m$ and $\alpha_T$ are the mass and thermal accommodation coefficients, respectively. Values of $\alpha_m=\alpha_T=1$ are usually used [44].

Whatever is being released by the droplets (see Eqs. (10) and (11)) becomes a heat and vapor source for the air, possibly affecting the air temperature and vapor concentration [25]. In addition to the vapor emitted by the droplets, the ambient water vapor level may change as well due to varying humidity between the incoming air and the (undisturbed) internal air.

The transport of solid particles in human respiratory system was reviewed by Kleinstreuer and Zhang [21] while the respiratory droplet vaporization and transport were recently reported by Zhang et al. [32], Zhang et al. [45], and Longest and Xi [46] among others.

## 3 Inhalers for Drug-Aerosol Delivery

**3.1 Overview.** The emergence of new types of drugs, e.g., proteins, peptides, and deoxyribonucleic acid (DNA)-based therapeutics, required innovative (targeting) delivery technology, necessarily coupling inhalation mode, as well as device and drug developments. Major areas for improvements include high, controlled efficacy of drug delivery, sustained drug release and rapid absorption, ease-of-use of the device, and operational independence from health care workers. While pulmonary delivery is very attractive because of the large absorption area and avoidance of first-pass metabolism in the liver, goals, such as device effectiveness, ease-of-use and low cost as well as dosage reproducibility, sustained drug release, rapid tissue uptake, optimal targeting, and reduced side-effects, are still rather elusive. Nevertheless, in light of the fast growing and very profitable drug delivery market, the pharmaceutical industry and medical device companies seek technical solutions to these problems in order to meet the demands of investors, stockholders, healthcare providers, and patients.

Several book chapters in Bissgard et al. [47], Finlay [34], Gradon and Marijnissen [48], and Hickey [49] discussed the pros and cons of inhaler devices, i.e., mainly pMDIs and DPIs as well as conceptual nebulizer systems (NSs) and spray devices for the treatment of asthma and chronic obstructive pulmonary disease (COPD) symptoms. The prescription for pMDIs (with HFA as the propellant) remain dominant at least in the USA and UK because they are easier to use (at least for most patients), more reliable, and cheaper per dose. However, the advent of multidose blister packs, i.e., new formulations and manufacturing technologies, has vastly improved DPI performance and ease-of-use [50]. Parallel to improved inhaler designs, laboratory models are being refined to estimate pulmonary absorption rates of inhaled medicine and to provide local aerosol deposition data for computer model validations [51].

Drug-aerosol deposition depends on several factors of which the inhalation flow rate $Q_{\mathrm{in}}(t)$, i.e., the patient's breathing pattern, particle size, and airway geometry are the most important ones. However, particle shape, density, thermodynamic state, and surface characteristics including roughness and charge may be influential as well. In any case, for inhaler device design, only breathing pattern and particle characteristics can be traditionally manipulated [52,53].

It is evident that nebulizer systems including nasal spray devices, metered-dose inhalers, and dry-powder inhalers are nondirectional, i.e., they cannot target drug-aerosols to predetermined sites in a controlled repeatable fashion. Furthermore, ideal size monodisperse aerosols and the best inhalation flow waveform synchronized with particle transport cannot be generated. In order to somewhat improve the situation, R&D efforts in academia and industry focused mainly on trained breathing, e.g., bolus inhalation [52], as well as combined propellant formulation and device improvements for pMDIs and DPIs. New NS concepts have been pursued as well. Concerning the *pulmonary* pathway, the goals are to produce drug-aerosol and delivery characteristics, which lead to deeper and/or peripheral lung penetration at higher deposition concentrations, mainly for asthma and COPD patients [1,53–62]. As a result, inhaler products on the market, in terms of new formulations used in modified pMDI and DPI devices, exhibit remarkable performance improvements. Specifically, Clark [56] showed that prior to 1995, the *average lung delivery efficiencies* for both pMDIs and DPIs hovered around 20% while newly reported values were 40–60% and conceptual "soft mist inhalers"

reaching up to 80%. Again, these high efficiencies only indicate that most of the inhaled drug-aerosols did not get stuck in the throat, i.e., their pulmonary depositions are still nonfocused.

In addition to improvements in formulation, packaging, and device components, researchers also considered the manipulation of the patient's mode of inhalation. For example, Denyer et al. [63] described an "adaptive aerosol nebulizer" where only during the first part of the inhalation cycle, a finite drug amount is pulsed into the air stream. This patient-specific timing of the aerosol pulse minimizes waste of medicine, which is of major concern for NSs. In contrast, Bondesson et al. [64] suggested "time point delivery" of dry-powder late in the inhalation cycle in order to achieve stronger lung penetration. Heyder [52] argued that effective targeting of the bronchial region can only be achieved with bolus inhalation. Bennett et al. [65] reviewed inhalation techniques and best particle sizes for targeting serial and especially parallel lung regions. They concluded that shallow aerosol boluses or very slow inhalation may selectively target the conducting airways while large volume inhalation with breath-hold allow for easy aerosol delivery to the alveolar region. Another example of manipulating the breathing pattern is the enforcement of "biphasic inhalation for DPIs" [66] where first, a small volume (e.g., a 6.5 $\mu$m droplet bolus of 300–900 ml) is slowly inhaled at 8 l/min followed 10 s later by a rapid inspiratory flow rate of 30 l/min. That procedure imposed on ten healthy adults achieved deposition efficiencies $>70\%$ in the whole lung and $>50\%$ in the lung periphery.

**3.2 Pressurized Metered-Dose Inhaler.** A pMDI is a compact, easy-to-use droplet spray dispenser for multidose treatment of pulmonary diseases, the most common being asthma. Typically, 200 high-pressure ($P_{\text{canister}} > 10$ atm) actuations containing $20-5000$ $\mu$g of drug in metered volumes of $25-100$ $\mu$l can be generated and inhaled (see for example Refs. [67,68]). A key component of the device is the actuator spray nozzle (with orifice diameters of 0.25–0.5 mm), which controls the atomization process and resultant spray formation in terms of spray angle and droplet size distribution. The formulation in pMDIs consists of a therapeutic material and a propellant as well as surfactants and other excipients, which dissuade drug-aerosol coagulation and lubricate moving parts of the metering valve. While pressurized chlorofluorocarbon (CFC) substances are being phased out worldwide because released chlorine atoms deplete the earth's ozone layer, hydrofluoroalkane (e.g., HFA-134a and HFA-227) have become replacement propellants. However, due to the different physical and chemical properties of CFC and HFA, both the formulation in pMDIs and their structural elements had to be modified [69]. In any case, 70–80% of the discharged formulation is liquid with nozzle exit velocities in the range of 150–225 m/s.

The overall goal for pMDIs is to deliver most of the medicine to a patient's broncho-pulmonary tree and lower airways, i.e., avoiding depositions inside the actuator and the oral airways. Numerous in vitro and in vivo tests with pMDIs have been conducted to approach that goal where companies usually employ a geometric model of the upper airways, i.e., the United States pharmacopoeia (USP) throat and the Andersen cascade impactor, to estimate particle size from pMDIs and to estimate the deposition pattern using gamma scintigraphy. For example, Leach et al. [70] compared CFC-driven and HFA-driven pMDI performances. While most of the drug-aerosols from CFC-pMDIs remained in the oropharyngeal region (i.e., 80–90%), HFA-pMDIs delivered up to 60% of the inhaled medicine into the central and intermediate lung zones. Such an increase in drug delivery with newer HFA-pMDIs can be contributed to the improved formulation characteristics and actuator design as well as the use of spacers [71,72], resulting in smaller drug-aerosols. In vitro tests were carried out by a number of researchers [73–75]. They argued that the use of a realistic human upper airway replica is preferable to the USP throat and compared CFC-pMDI versus HFA-pMDI performances, considering the holding chamber and consistent drug dosing, the effect of temperature, the influence of spacers, and the impact of orifice diameter on spray force.

In summary, if properly used [76], HFA-pMDIs may deliver a significant portion of the inhaled medicine to the central lung region; however, therapeutic droplet deposition is nondirectional and most treatments require high-percentage drug-aerosol deposition at specific lung sites, such as the peripheral airways, tumors, or deeper lung regions, e.g., alveolated ducts.

**3.3 Dry-Powder Inhaler.** Various DPIs are available, all breath-actuated, i.e., propellant-free, thereby, eliminating the coordination problem sometimes encountered with pMDIs. Coates et al. [77] analyzed computationally and experimentally the influence of airflow rate of a DPI on particle dispersion and throat deposition. An optimal inhaler performance was found at $Q = 65$ l/min. The dry-powder formulation is a critical component of a DPI (see Refs. [1,78] among others). Most formulations are sensitive to moisture during manufacturing, storage, and usage, which may affect the ultimate delivery of one to five microsize, stable, and reproducible drug-aerosols. However, new packaging technologies including multidose blister packs, the use of nonirritating excipients, and improved device designs have made DPIs more acceptable [79].

## 4 A SIS

**4.1 SIS Methodology.** In cases where expensive and/or biochemically aggressive drug-aerosols have to reach a specific site in the lung (e.g., to attack a tumor) or a specific deep lung region (e.g., to achieve rapid drug absorption of, say, insulin), *targeted drug-aerosol delivery* is most desirable. This can be accomplished with a new, patient-specific "controlled air-particle stream" methodology. Thus, a new SIS has been proposed (patent pending), which allows for optimal drug-aerosol targeting to patient-specific airway locations [80]. Specifically, the underlying SIS methodology has been tested, relying on experimentally validated computer simulations and laboratory experiments (see Figs. 5 and 6). Specifically, in Fig. 5, the high drug deposition percentage in the oral cavity and throat using an off-the-shelf inhaler [5] (see Fig. 5(*a*)) is contrasted with the direct targeting of a semispherical tumor (see Fig. 5(*b*)). The computational/experimental proof-of-concept of optimal drug-aerosol targeting is depicted in Fig. 6 where the effect of three different inhaler exit (or mouth inlet) drug release points are illustrated, all exiting 100% at generation 3.1 (Fig. 6(*a*)) or the trachea (Fig. 6(*b*)). This is accomplished as follows. Given *any* respiratory system, theoretically, 100% of the inhaled drug-aerosols can be delivered to specific lung sites (presently up to generation 6 due to still limited MRI/U.S./CT scan resolution) or regions (see Hyun et al. [17] for details). In any case, once the patient-specific, CAD-like geometry files have been obtained, the computational tasks to be accomplished for each patient or a representative group of similar patients are the following.

- determination of optimal diameter and density of (drug) microspheres, i.e., solid or liquid
- determination of suitable inhalation flow rates, which should be laminar and steady for best air drug stream control from inhaler outlet/mouth inlet to a predetermined target site/region
- determination of optimal (radial) release position of the drug-aerosols from the (flexible) nozzle tip
- determination of the drug loading rate to assure deposition of a prescribed dosage onto the lung target area

These tasks are accomplished via *realistic, experimentally validated* computer simulations (see Kleinstreuer [81] and Kleinstreuer and Zhang [21] for details) as well as computer experiments, finding best microsphere characteristics, inhalation flow rates, and drug dosage. Most important is the patient-specific (or group specific) drug-aerosol release position (i.e., microprocessor-

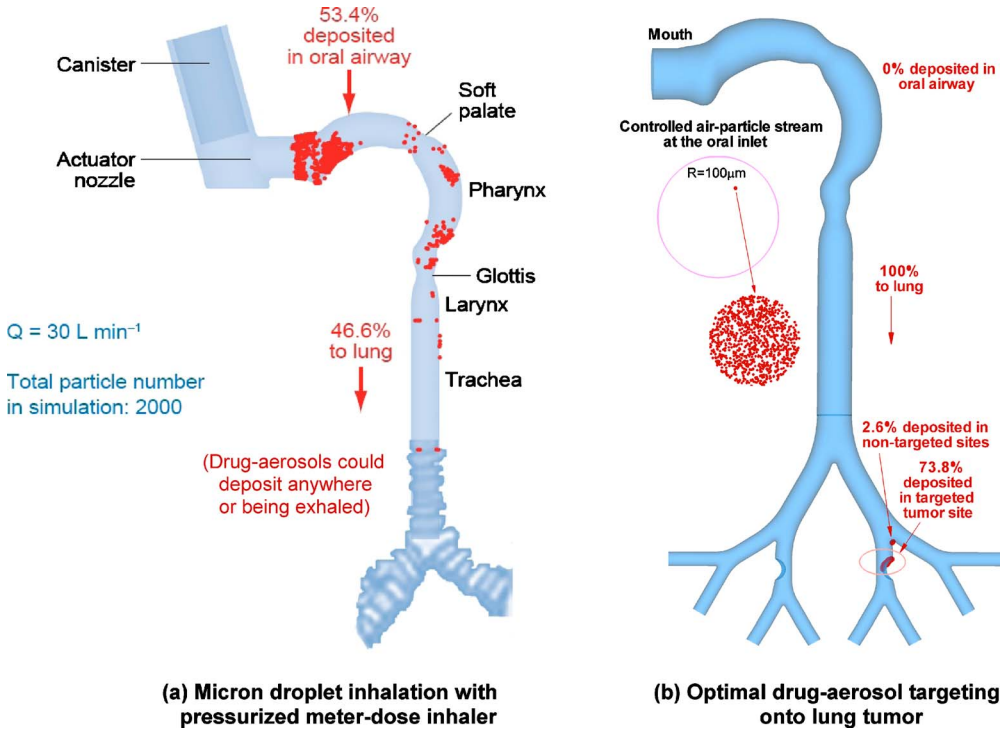



**Fig. 5 Comparison of nondirectional (*a*) and more targeted (*b*) drug-aerosol delivery**

controlled nozzle tip location in the SIS) for optimal targeting. This is obtained via "backtracking," i.e., a random particle load (distributed across the inhaler exit tube area) is released and those particles depositing on the target site (or in the target region) are back-tracked. That generates a drug-aerosol release map, which correlates specific release zones with specific target sites/regions (see Fig. 7).

**4.2 SIS Features and Operation.** In order to implement the methodology outlined in Sec. 4.1, the SIS has to feature three key elements, i.e., a liquid (or powder) drug reservoir with aerosol generator, an inhaled flow waveform modulator to achieve laminar pseudosteady airflow, and a flexible nozzle with controlled radial positioning for optimal drug-aerosol release. It should be noted that the SIS consist from the computational angle only of a 2 cm inhaler tube with an air-particle supply, i.e., a modulated constant airstream and an optimally positioned nozzle from which five to ten microparticles will be released.

Clearly, the radial nozzle tip location is computer-determined, as described in the previous paragraph. For the inhalation waveform modulation, computer experiments provided transient area parameter values for actuator-controlled flow resistance and air bleeding area changes (see Fig. 8). The result is a conversion of a patient's actual transient air flow/pressure profile into a partially steady inhalation flow rate during which the drug-aerosols are released.

So far, we have shown proof-of-concept of the SIS technology in vitro. For most treatments/applications, it is desirable to target not only *spot* locations in the lung but to selectively cover predetermined target *areas*. In order to accomplish that, the SIS features a flexible nozzle, which can be adjusted via two sets of three shape memory alloy (SMA) wires. With the first set, any predetermined drug release position can be obtained while the second set of wires will align the nozzle tip parallel to the air co-flow to avoid premature mixing of aerosols due to shear layer effects or even premature wall deposition. The overall device is designed with snap-in panels to allow interior access for precise and reproducible wire mounting and assembly. Electrical leads are embedded into the structure and connected to an external power supply as well as a control and data acquisition system at the bottom with a special minimal footprint connection to allow for simplified mounting and flawless operation (see Fig. 9).

The nozzle-positioning SMA wires' multifunctionality (e.g., actuation/positioning, sensing) can be extended to include flow rate sensing capabilities similar to a conventional wire anemometer.

## 5 Conclusions and Future Work

This review summarizes computer simulation methodologies of air-particle flow, results of drug-aerosol transport/deposition in models of the human respiratory system, as well as aspects of drug-aerosol targeting and associated inhalation devices. DDSs for optimal targeting of pathological sites, such as tumors, are in the midst of a strong development, testing, and application cycle. Of the numerous DDS applications, optimally targeted drug-aerosol delivery via inhalation to predetermined lung areas is becoming a very desirable treatment option not only for combating respiratory disorders but also for systemic diseases. Respiratory treatment applications include chronic obstructive pulmonary disease, asthma, cystic fibrosis, bacterial infection, and lung tumors. Because of the rapid drug absorption from the deep lung region into the systemic system and ease-of-administration, inhalation of medical aerosols is also been used for the treatment of diabetes and cancer.

After a brief introduction to drug delivery systems in general, the required modeling and simulation steps for optimal drug-aerosol delivery in the lung are outlined. Starting with medical imaging and file conversion of patient-specific lung-airway morphologies, the air-particle transport phenomena are numerically solved for a representative inhalation flow rate of $Q_{\text{total}} = 30$ l/min. Focusing on microspheres and droplets, the complex airflow and particle dynamics as well as the droplet heat and mass transfer are illustrated. With this foundation as the background, an overview of common inhaler devices, such as the pressurized metered-dose inhaler and the dry-powder inhaler, is presented. Clearly, employing the power and flexibility of realistic and accu-

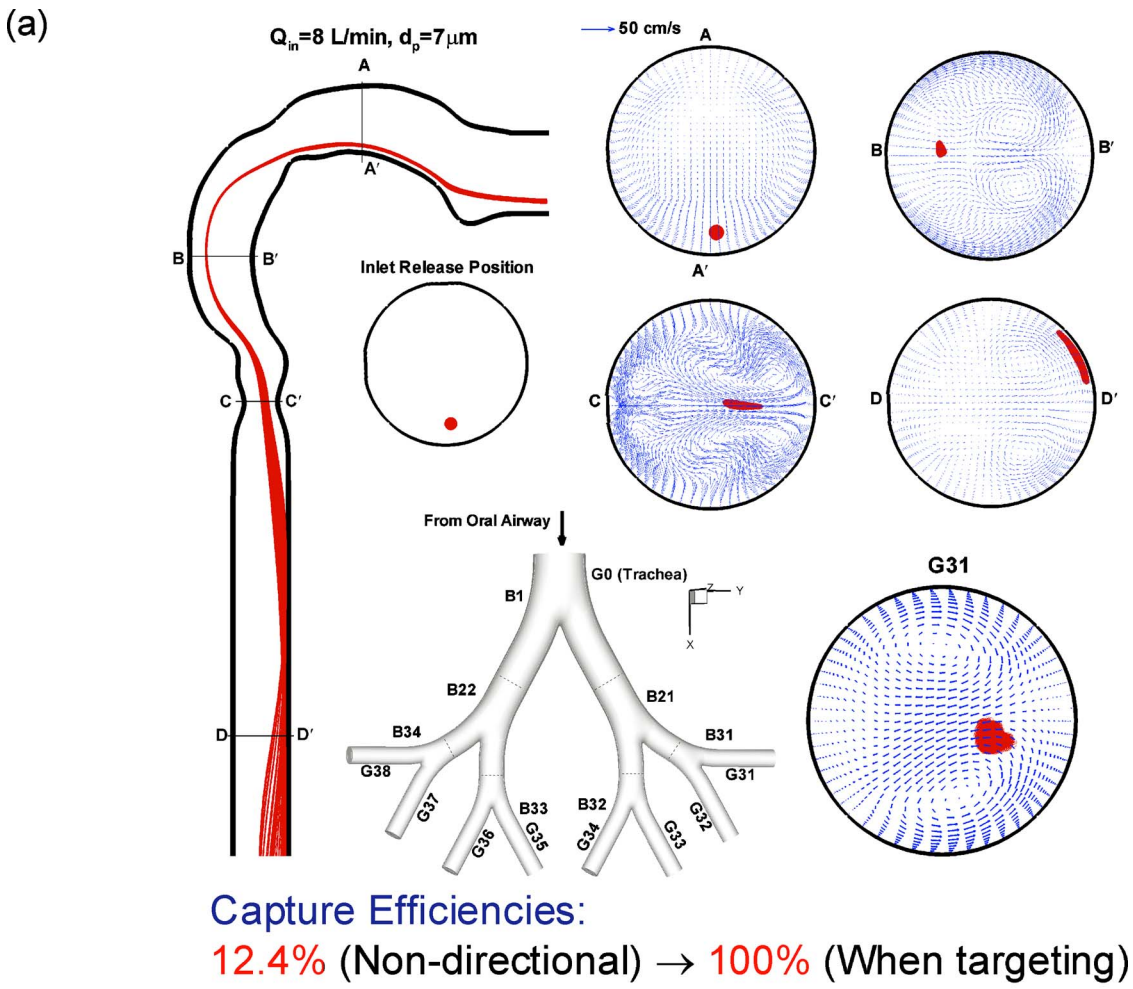



**Fig. 6** (*a*) **Controllable particle distributions in the upper airway models when releasing from a given position at mouth inlet with $Q_{in}$=8 l/min and $d_p$=7 μm. The capture efficiency of particles after G3.1 increases from 12.4% (normal, nondirectional inhalation) to 100% (targeted inhalation). (*b*) Comparison between computational and experimental measurements of particle paths through the trachea as a function of release point.**

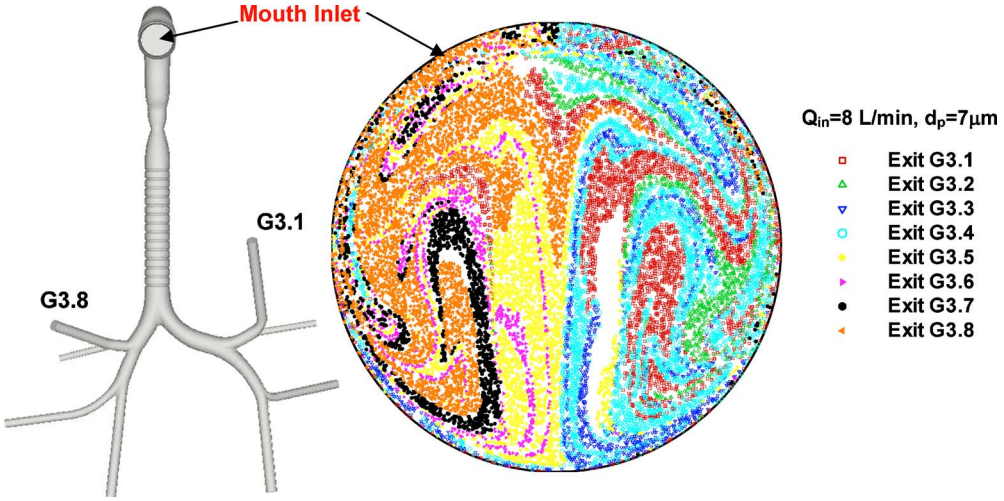



**Fig. 7 Regions of optimal particle-release positions for targeting at the mouth inlet (or inhaler exit)**

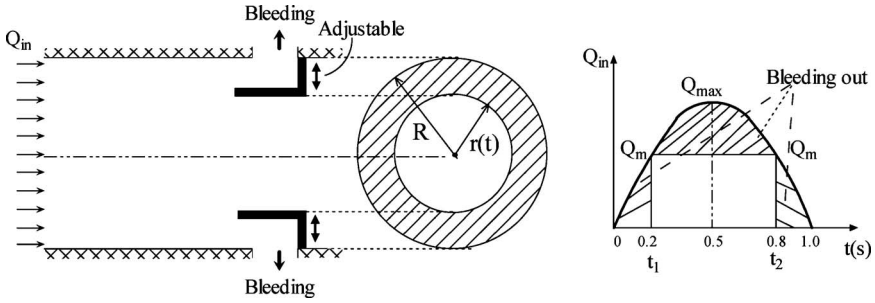



**Fig. 8 Schematic of inhalation flow control system (left panel) and modulated inhalation flow waveform (right panel $Q_m$ should be less than 8 l/min to maintain laminar flows)**

rate computer simulations, novel inhaler devices can be developed and clinically tested in order to achieve optimal drug-aerosol delivery to predetermined lung sites.

This goal is discussed in terms of the suitable underlying methodology and features of a new SIS. It is demonstrated that with the SIS, inhaled drug-aerosols can be directly delivered to any predetermined target area in the human lung. Future work will focus on SIS prototyping as well as laboratory and clinical testing.

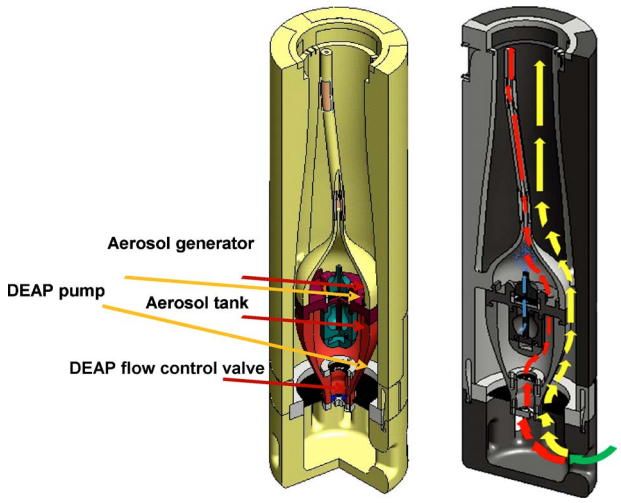



**Fig. 9 Illustration of autonomous SIS design (patent pending): slice view with components (left) and air flow pattern with co-flow splitting for aerosolizer supply (right)**

## Acknowledgment

The authors would like to thank Prof. Seelecke, MAE Department, NC State University, for the supply of Fig. 9. This effort was sponsored by the NSF Grant No. CBET-0834054 (Dr. Marc S. Ingber, Program Director) and the Air Force Office of Scientific Research, Air Force Material Command, USAF, under Grant No. FA9550-07-1-0461 (Dr. Walt Kozumbo, Program Manager). The use of both CFX software from ANSYS Inc. (Canonsburg, PA) and IBM LINUX CLUSTER software at the High Performance Computing Center at North Carolina State University (Raleigh, NC) are gratefully acknowledged as well. The authors also thank Professor Andrew Pollard and Mr. Christopher Ball of Queen's University at Kingston for supplying their idealized geometry of the oral airway (see Figs. 2 and 3).

## References

[1] Edwards, D. A., and Dunbar, C., 2002, "Bioengineeiring of Therapeutic Aerosols," Annu. Rev. Biomed. Eng., **4**, pp. 93–107.
[2] O'Riordan, T. G., 2005, "Aerosol Delivery Devices and Obstructive Airway Disease," Expert Rev. Med. Devices, **2**(2), pp. 197–203.
[3] Telko, M. J., and Hickey, A. J., 2005, "Dry Powder Inhaler Formulation," Respir. Care, **50**, pp. 1209–1227.
[4] Hess, D. R., 2008, "Aerosol Delivery Devices in the Treatment of Asthma," Respir. Care, **53**(6), pp. 699–723.
[5] Kleinstreuer, C., Shi, H. W., and Zhang, Z., 2007, "Computational Analyses of a Pressurized Metered Dose Inhaler and a New Drug-Aerosol Targeting Methodology," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **20**(3), pp. 294–309.
[6] Kleinstreuer, C., Zhang, Z., and Donohue, J. F., 2008, "Targeted Drug-Aerosol

Delivery in the Human Respiratory System," Annu. Rev. Biomed. Eng., **10**, pp. 195–220.
[7] Sahoo, S. K., Parveen, S., and Panda, J. J., 2007, "The Present and Future of Nanotechnology in Human Health Care," Nanomedicine, **3**(1), pp. 20–31.
[8] Kaparissides, C., Alexandridou, S., Kotti, K., and Chaitidou, S., 2006, "Recent Advances in Novel Drug Delivery Systems," Journal of Nanotechnology Online, **2**, pp. 1–11.
[9] Ciofani, G., Raffa, V., Menciassi, A., Micera, S., and Dario, P., 2007, "A Drug Delivery System Based on Alginate Microspheres: Mass-Transport Test and In Vitro Validation," Biomed. Microdevices, **9**, pp. 395–403.
[10] Dandamudi, S., and Campbell, R. B., 2007, "The Drug Loading, Cytotoxicty and Tumor Vascular Targeting Characteristics of Magnetite in Magnetic Drug Targeting," Biomaterials, **28**, pp. 4673–4683.
[11] Burrowes, K. S., Swan, A. J., Warren, N. J., and Tawhai, M. H., 2008, "Towards a Virtual Lung: Multi-Scale, Multi-Physics Modelling of the Pulmonary System," Philos. Trans. R. Soc. London, Ser. A, **366**(1879), pp. 3247–3263.
[12] Ma, B., and Lutchen, K., 2009, "CFD Simulation of Aerosol Deposition in an Anatomically Based Human Large-Medium Airway Model," Ann. Biomed. Eng., **37**(2), pp. 271–285.
[13] Van Ertbruggen, C., Corieri, P., Theunissen, R., Riethmuller, M. L., and Darquenne, C., 2008, "Validation of CFD Predictions of Flow in a 3D Alveolated Bend With Experimental Data," J. Biomech., **41**(2), pp. 399–405.
[14] Luo, H. Y., and Liu, Y., 2008, "Modeling the Bifurcating Flow in a CT-Scanned Human Lung Airway," J. Biomech., **41**(12), pp. 2681–2688.
[15] Tebeckhorst, S., Lee, D., Wexler, A. S., and Oldham, M. J., 20072007, "Interaction of Epithelium With Mesenchyne Affects Global Features of Lung Architecture: A Computer Model of Development, J. Appl. Physiol., **102**(1), pp. 294–305.
[16] Taylor, C. A., and Draney, M. T., 2004, "Experimental and Computational Methods in Cardiovascular Fluid Mechanics," Annu. Rev. Fluid Mech., **36**, pp. 197–231.
[17] Hyun, S., Kleinstreuer, N., Welch, B., Ryans, J., Birchard, K., Zhang, Z., and Kleinstreuer, C., 2010, "Variations in Tracheobronchial Airway Morphology and Fractal Analysis of Patient-Specific Models," J. Aerosol Sci., unpublished.
[18] Zhang, Z., Kleinstreuer, C., and Kim, C. S., 2008, "Airflow and Nanoparticle Deposition in a 16-Generation Tracheobronchial Airway Model," Ann. Biomed. Eng., **36**(12), pp. 2095–2110.
[19] Zhang, Z., and Kleinstreuer, C., 2010, "Laminar-to-Turbulent Fluid-Nanoparticle Dynamics Simulations: Model Comparisons and Nanoparticle-Deposition Applications," Phys. Fluids, in press.
[20] Nicoud, F., and Ducros, F., 1999, "Subgrid-Scale Stress Modelling Based on the Square of the Velocity Gradient Tensor," Flow, Turbul. Combust., **62**(3), pp. 183–200.
[21] Kleinstreuer, C., and Zhang, Z., 2010, "Airflow and Particle Transport in the Human Respiratory System," Annu. Rev. Fluid Mech., **42**, pp. 301–334.
[22] Broday, D. M., and Georgopoulos, P. G., 2001, "Growth and Deposition of Hygroscopic Particulate Matter in the Human Lungs," Aerosol Sci. Technol., **34**(1), pp. 144–159.
[23] Ferron, G. A., Kreyling, W. G., and Haider, B., 1988, "Inhalation of Salt Aerosol-Particles. 2. Growth and Deposition in the Human Respiratory-Tract," J. Aerosol Sci., **19**(5), pp. 611–631.
[24] Ferron, G. A., Oberdörster, G., and Henneberg, R., 1989, "Estimation of the Deposition of Aerosolized Drugs in the Human Respiratory Tract Due to Hygroscopic Growth," J. Aerosol Med., **2**, pp. 271–284.
[25] Finlay, W. H., and Stapleton, K. W., 1995, "The Effect on Regional Lung Deposition of Coupled Heat and Mass-Transfer Between Hygroscopic Droplets and Their Surrounding Phase," J. Aerosol Sci., **26**(4), pp. 655–670.
[26] Johnson, C. E., Linderoth, L. S., and Nuckols, M. L., 1977, "Analysis of Sensible Respiratory Heat-Exchange During Inspiration Under Environmental-Conditions of Deep Diving," Mech. Eng., **99**(5), p. 94.
[27] Nuckols, M. L., Zumrick, J. L., and Johnson, C. E., 1983, "Heat and Water-Vapor Transport in the Human Upper Airways at Hyperbaric Conditions," ASME J. Biomech. Eng., **105**(1), pp. 24–30.
[28] Weibel, E. R., 1963, *Morphometry of the Human Lung*, Academic, New York.
[29] Daviskas, E., Gonda, I., and Anderson, S. D., 1990, "Mathematical-Modeling of Heat and Water Transport in Human Respiratory-Tract," J. Appl. Physiol., **69**(1), pp. 362–372.
[30] Tawhai, M. H., and Hunter, P. J., 2004, "Modeling Water Vapor and Heat Transfer in the Normal and the Intubated Airways," Ann. Biomed. Eng., **32**(4), pp. 609–622.
[31] Zhang, Z., and Kleinstreuer, C., 2003, "Species Heat and Mass Transfer in a Human Upper Airway Model," Int. J. Heat Mass Transfer, **46**(25), pp. 4755–4768.
[32] Zhang, Z., Kleinstreuer, C., and Kim, C. S., 2006, "Water Vapor Transport and its Effects on the Deposition of Hygroscopic Droplets in a Human Upper Airway Model," Aerosol Sci. Technol., **40**(1), pp. 1–16.
[33] Elad, D., Wolf, M., and Keck, T., 2008, "Air-Conditioning in the Human Nasal Cavity," Respir. Physiol. Neurbiol., **163**(1–3), pp. 121–127.
[34] Finlay, W. H., 2001, *The Mechanics of Inhaled Pharmaceutical Aerosols: An Introduction*, Academic, London.
[35] Ferron, G. A., Haider, B., and Kreyling, W. G., 1985, "A Method for the Approximation of the Relative-Humidity in the Upper Human Airways," Bull. Math. Biol., **47**(4), pp. 565–589.
[36] Longest, P. W., Hindle, M., and Tian, G., 2009, "Improved Inhalation Drug Delivery of Submicrometer Aerosols Through Enhanced Condensational Growth: Numerical and Experimental Proof of Concept Studies," AAAR's 28th Annual Conference, Minneapolis, MN.
[37] Kleinstreuer, C., 2003, *Two-Phase Flow: Theory and Applications*, Taylor & Francis, New York.
[38] Li, A., and Ahmadi, G., 1992, "Dispersion and Deposition of Spherical-Particles From Point Sources in a Turbulent Channel Flow," Aerosol Sci. Technol., **16**(4), pp. 209–226.
[39] Longest, P. W., Kleinstreuer, C., and Buchanan, J. R., 2004, "Efficient Computation of Micro-Particle Dynamics Including Wall Effects," Comput. Fluids, **33**(4), pp. 577–601.
[40] Clift, R., Grace, J. R., and Weber, M. E., 1978, *Bubbles, Drops, and Particles*, Academic, New York.
[41] Benaissa, A., Gauthier, J. E. D., Bardoni, M. F., and Laviolette, M., 2002, "Modelling Evaporation of Multicomponent Fuel Droplets Under Ambient Temperature Conditions," J. Inst. Energy, **75**(502), pp. 19–26.
[42] Berlemont, A., Grancher, M. S., and Gouesbet, G., 1995, "Heat and Mass-Transfer Coupling Between Vaporizing Droplets and Turbulence Using a Lagrangian Approach," Int. J. Heat Mass Transfer, **38**(16), pp. 3023–3034.
[43] Vesala, T., Kulmala, M., Rudolf, R., Vrtala, A., and Wagner, P. E., 1997, "Models for Condensational Growth and Evaporation of Binary Aerosol Particles," J. Aerosol Sci., **28**(4), pp. 565–598.
[44] Ferron, G. A., Haider, B., and Kreyling, W. G., 1988, "Inhalation of Salt Aerosol-Particles. 1. Estimation of the Temperature and Relative-Humidity of the Air in the Human Upper Airways," J. Aerosol Sci., **19**(3), pp. 343–363.
[45] Zhang, Z., Kleinstreuer, C., and Kim, C. S., 2006, "Isotonic and Hypertonic Saline Droplet Deposition in a Human Upper Airway Model," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **19**(2), pp. 184–198.
[46] Longest, P. W., and Xi, J., 2008, "Condensational Growth May Contribute to the Enhanced Deposition of Cigarette Smoke Particles in the Upper Respiratory Tract," Aerosol Sci. Technol., **42**(8), pp. 579–602.
[47] Bissgard, H., O'Callaghan, C., and Smaldone, G., 2002, *Drug Delivery to the Lungs*, Marcel Dekker, New York.
[48] Gradon, L., and Marijnissen, J., 2003, *Optimization of Aerosol Drug Delivery*, Kluwer Academic, Norwell, MA.
[49] Hickey, A. J., 2004, *Pharmaceutical Inhalation Aerosol Technology*, Marcel Dekker, New York.
[50] O'Connor, B. J., 2004, "The Ideal Inhaler: Design and Characteristics to Improve Outcomes," Respir. Med., **98**, pp. S10–S16.
[51] Sakagami, M., 2006, "In Vivo, In Vitro and Ex Vivo Models to Assess Pulmonary Absorption and Disposition of Inhaled Therapeutics for Systemic Delivery," Adv. Drug Delivery Rev., **58**(9–10), pp. 1030–1060.
[52] Heyder, J., 2004, "Deposition of Inhaled Particles in the Human Respiratory Tract and Consequences for Regional Targeting in Respiratory Drug Delivery," Proc. Am. Thorac. Soc., **1**, pp. 315–320.
[53] Shoyele, S. A., and Cawthome, S., 2006, "Particle Engineering Techniques for Inhaled Biopharmaceuticals," Adv. Drug Delivery Rev., **58**(9–10), pp. 1009–1029.
[54] Chow, A. H. L., Tong, H. H. Y., Chattopadhyay, P., and Shekunov, B. Y., 2007, "Particle Engineering for Pulmonary Drug Delivery," Pharm. Res., **24**(3), pp. 411–437.
[55] Smyth, H. D. C., and Leach, C. L., 2005, "Alternative Propellant Aerosol Delivery Systems," Crit. Rev. Ther. Drug Carrier Syst., **22**(6), pp. 493–534.
[56] Clark, A. R., 2004, "Pulmonary Delivery Technology: Recent Advances and Potential for the New Millennium," *Pharmaceutical Inhalation Aerosol Technology*, A. J. Hickey, ed., Marcel Dekker, New York, pp. 571–591.
[57] Tobyn, M., Staniforth, J. N., Morton, D., Harmer, Q., and Newton, M. E., 2004, "Active and Intelligent Inhaler Device Development," Int. J. Pharm., **277**(1–2), pp. 31–37.
[58] Zanen, P., 2003, "Targetting Aerosols to Disease Areas," *Optimization of Aerosol Drug Delivery*, L. Gradon and J. Marijnissen, eds., Kluwer Academic, Norwell, MA, pp. 243–276.
[59] Dalby, R., Spallek, M., and Voshaar, T., 2004, "A Review of the Development of Respimat((R)) Soft Mist (TM) Inhaler," Int. J. Pharm., **283**(1–2), pp. 1–9.
[60] Musante, C. J., Schroeter, J. D., Rosati, J. A., Crowder, T. M., Hickey, A. J., and Martonen, T. B., 2002, "Factors Affecting the Deposition of Inhaled Porous Drug Particles," J. Pharm. Sci., **91**(7), pp. 1590–1600.
[61] Edwards, D. A., 1995, "The Macrotransport of Aerosol-Particles in the Lung-Aerosol Deposition Phenomena," J. Aerosol Sci., **26**(2), pp. 293–317.
[62] Fu, J., Fiegel, J., Krauland, E., and Hanes, J., 2002, "New Polymeric Carriers for Controlled Drug Delivery Following Inhalation or Injection," Biomaterials, **23**(22), pp. 4425–4433.
[63] Denyer, J., Nikander, K., and Smith, N. J., 2004, "Adaptive Aerosol Delivery (AADR) Technology," Expert Opin. Drug Deliv., **1**(1), pp. 165–176.
[64] Bondesson, E., Bengtsson, T., Borgstrom, L., Nilsson, L. E., Norrgren, K., Olsson, B., Svensson, M., and Wollmer, P., 2005, "Dose Delivery Late in the Breath Can Increase Dry Powder Aerosol Penetration Into the Lungs," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **18**(1), pp. 23–33.
[65] Bennett, W. D., Brown, J. S., Zeman, K. L., Hu, S. C., Scheuch, G., and Sommerer, K., 2002, "Targeting Delivery of Aerosols to Different Lung Regions," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **15**(2), pp. 179–188.
[66] Clark, A. R., Chambers, C. B., Muir, D., Newhouse, M. T., Paboojian, S., and Schuler, C., 2007, "The Effect of Biphasic Inhalation Profiles on the Deposition and Clearance of Coarse (6.5 m $\mu$m) Bolus Aerosols," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **20**(1), pp. 75–82.
[67] Dunbar, C. A., 1997, "Atomization Mechanisms of the Pressurized Metered Dose Inhaler," Part. Sci. Technol., **15**(3–4), pp. 253–271.

[68] Dunbar, C. A., Watkins, A. P., and Miller, J. F., 1997, "An Experimental Investigation of the Spray Issued From a pMDI Using Laser Diagnostic Techniques," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **10**(4), pp. 351–368.
[69] Brown, B. A., 2003, "5 Myths About MDIs," Drug Delivery Technology.
[70] Leach, C. L., Davidson, P. J., and Boudreau, R. J., 1998, "Improved Airway Targeting With the CFC-Free HFA-Beclomethasone Metered-Dose Inhaler Compared With CFC-Beclomethasone," Eur. Respir. J., **12**(6), pp. 1346–1353.
[71] Kwok, P. C. L., Collins, R., and Chan, H. K., 2006, "Effect of Spacers on the Electrostatic Charge Properties of Metered Dose Inhaler Aerosols," J. Aerosol Sci., **37**(12), pp. 1671–1682.
[72] Newman, S. P., 2004, "Spacer Devices for Metered Dose Inhalers," Clin. Pharmacokinet, **43**(6), pp. 349–360.
[73] Cheng, Y. S., Fu, C. S., Yazzie, D., and Zhou, Y., 2001, "Respiratory Deposition Patterns of Salbutamol pMDI With CFC and HFA-134a Formulations in a Human Airway Replica," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **14**(2), pp. 255–266.
[74] Rahmatalla, M. F., Zuberbuhler, P. C., Lange, C. F., and Finlay, W. H., 2002, "In Vitro Effect of a Holding Chamber on the Mouth-Throat Deposition of QVAR((R)) (Hydrofluoroalkane-Beclomethasone Dipropionate)," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **15**(4), pp. 379–385.
[75] Asmus, M. J., Coowanitwong, I., Kwon, S. H., Khorsand, N., and Hochhaus, G., 2003, "In Vitro Performance of Two Common Valved Holding Chambers With a Chlorofluorocarbon-Free Beclomethasone Metered-Dose Inhaler," Pharmacotherapy, **23**(12), pp. 1538–1544.
[76] Canadian Medical Association, 2005, "Inhalation Devices," Can. Med. Assoc. J., **173**(6), pp. S39–S45.
[77] Coates, M. S., Chan, H. K., Fletcher, D. F., and Raper, J. A., 2005, "Influence of Air Flow on the Performance of a Dry Powder Inhaler Using Computational and Experimental Analyses," Pharm. Res., **22**(9), pp. 1445–1453.
[78] Chan, H. K., 2006, "Dry Powder Aerosol Delivery Systems: Current and Future Research Directions," Journal of Aerosol Medicine-Deposition Clearance and Effects in the Lung, **19**(1), pp. 21–27.
[79] Dolovich, M. B., Ahrens, R. C., Hess, D. R., Anderson, P., Dhand, R., Rau, J. L., Smaldone, G. C., and Guyatt, G., 2005, "Device Selection and Outcomes of Aerosol Therapy: Evidence-Based Guidelines," Chest, **127**(1), pp. 335–371.
[80] Kleinstreuer, C., Zhang, Z., Li, Z., Roberts, W. L., and Rojas, C., 2008, "A New Methodology for Targeting Drug-Aerosols in the Human Respiratory System," Int. J. Heat Mass Transfer, **51**, pp. 5578–5589.
[81] Kleinstreuer, C., 2006, *Biofluid Dynamics: Principles and Selected Applications*, CRC, Boca Raton, FL.

# Using MicroCT Imaging Technique to Quantify Heat Generation Distribution Induced by Magnetic Nanoparticles for Cancer Treatments

**Anilchandra Attaluri**

**Ronghui Ma**

**Liang Zhu**[1]
Associate Professor
e-mail: zliang@umbc.edu

Department of Mechanical Engineering,
University of Maryland,
Baltimore County, 1000 Hilltop Circle,
Baltimore, MD 21250

*Magnetic nanoparticles have been used in clinical and animal studies to generate localized heating for tumor treatments when the particles are subject to an external alternating magnetic field. Currently, since most tissue is opaque, the detailed information of the nanoparticle spreading in the tissue after injections cannot be visualized directly and is often quantified by indirect methods, such as temperature measurements, to inversely determine the particle distribution. In this study, we use a high resolution microcomputed tomography (microCT) imaging system to investigate nanoparticle concentration distribution in a tissue-equivalent agarose gel. The local density variations induced by the nanoparticles in the vicinity of the injection site can be detected and analyzed by the microCT system. Heating experiments are performed to measure the initial temperature rise rate to determine the nanoparticle-induced volumetric heat generation rates (or specific absorption rate (SAR $W/m^3$)) at various gel locations. A linear relationship between the measured SARs and their corresponding microCT pixel index numbers is established. The results suggest that the microCT pixel index number can be used to represent the nanoparticle concentration in the media since the SAR is proportional to the local nanoparticle concentration. Experiments are also performed to study how the injection amount, gel concentration, and nanoparticle concentration in the nanofluid affect the nanoparticle spreading in the gel. The nanoparticle transport pattern in gels suggests that convection and diffusion are important mechanisms in particle transport in the gel. Although the particle spreading patterns in the gel may not be directly applied to real tissue, we believe that the current study lays the foundation to use microCT imaging systems to quantitatively study nanoparticle distribution in opaque tumor.*
[DOI: 10.1115/1.4002225]

*Keywords: magnetic nanoparticles, hyperthermia, cancer, heating, temperature, microCT imaging*

## 1 Introduction

Magnetic nanoparticles or microparticles have been proposed to heat tumor tissue in cancer treatment since the 1950s. In this method, superparamagnetic nanoparticles can deliver adequate heating to irregular and/or deep-seated tumors when exposed to a relatively low magnetic field and frequency. The heat generated by these particles when exposed to an external alternating magnetic field is mainly due to the Néel relaxation mechanism and/or the Brownian motion of particles [1,2]. Previous studies showed that iron oxides magnetite $Fe_3O_4$ and maghemite $\gamma$-$Fe_2O_3$ nanoparticles are biocompatible in the human tissue [3]. In the past decade, there have been renewed interests in using magnetic nanoparticles in cancer treatments [4–7]. Due to technical advancement in manufacturing nanosized magnetic particles, nanoparticle hyperthermia has emerged as an attractive alternative to costly and risky surgical procedures because of its few associated complications and targeted delivery of thermal energy to the tumor.

Different methods exist to deliver nanoparticles to tumor: either systemically if the blood vessels of the diseased organ are well known or by directly injecting the nanoparticles in the extracellular space of the tumor. The systemic delivery of the nanoparticles relies on the coating on the particles and its strong affinity to receptors on the tumor cells. Recently, a numerical study [8] proposed a method combining both heat and mass transfer for targeted drug delivery. In the past years, our group [9–11] investigated the feasibility of elevating the temperature of the entire tumor above a certain threshold using direct injection of nanofluid at multiple injection sites while preserving the surrounding healthy tissue from thermal damage. The direct injection method is the focus of this study and its advantage is that multiple-site injections can be exploited to cover the entire target region in the case of an irregularly shaped tumor. The success of this approach depends strongly on the nanoparticle distribution after injection, which is affected by the injection strategies including injection rate and injection amount of the nanofluid.

Previous investigations have demonstrated that particle size, particle coating, and magnetic field strength and frequency determine its heating capacity defined as the specific loss power (SLP) [1,2,12–14]. However, once the nanoparticles are manufactured, it is the spatial distribution of the particles dispersed in tissue that dominates the spatial temperature elevations. Previous experimental data have suggested that the particle concentration is not uniform after the injection and is quite different from the ferrofluid concentration [12,15–17]. Since tissue is opaque, nanoparticle distribution is usually indirectly quantified via temperature distribu-

[1]Corresponding author.
Manuscript received April 6, 2010; final manuscript received April 21, 2010; published online September 27, 2010. Assoc. Editor: Andrey Kuznetsov.

tion measured in the tumor using inverse heat transfer analyses. There lacks a direct imaging technique to study the particle distribution.

X-ray tomography nowadays is a well-established standard diagnostic tool in medical fields for inspection and testing. X-ray microcomputed tomography (microCT) offers a nondestructive way to obtain the complete high resolution three-dimensional spatial morphology of a small specimen (~cm). Recently, iron based nanoparticles have gained prominence in medical imaging [18,19]. The present work is intended to show the feasibility of using microCT technology to quantify the nanoparticle distribution in a tissue-equivalent agarose gel after direct injection. Although heat transfer in biological media can be complicated compare with agarose gel, agarose gel has similar convection/diffusion properties as that of tumor [21]. The gel has been used in the past to measure specific absorption rate (SAR) in hyperthermia studies. We hypothesize that a high resolution microCT system is sufficiently sensitive to image the density variations induced by the injected nanoparticles in the gel. Although the microCT does not allow direct visualization of individual nanoparticles, the accumulation of nanoparticles in the gel would result in a region with a much higher density than the rest of the gel area and the density distribution can be detected by the microCT system. Several previous studies [6,7] have utilized computed tomography (CT) images to study ferrofluid distribution, however, those studies are limited by their spatial resolution (~mm) and temperature resolution (~1°C). In this study, we injected a commercially available ferrofluid containing magnetic nanoparticles. The high resolution (~$\mu$m) three-dimensional images of the density distribution were quantified to establish a correlation with the measured SAR values at various gel locations. The images were later analyzed for different injection parameters including the injection amount (0.1 cc or 0.2 cc), gel concentration (0.5%, 1%, 2%, or 4%), and ferrofluid concentration (3.9% or 5.8%) by volume of particle concentration. The spreading of the ferrofluid in the gel after injection was discussed based on the microCT pixel index variations in the gel.

## 2 Methods

**2.1 Nanoparticle Injection and Agarose Gel.** Water based ferrofluids (EMG705 series, Ferrotec (USA), Corporation, Nashua, NH) with a concentration of 3.9% or 5.8% by volume and a particle size of 10 nm were injected in an agarose gel at different infusion flow rates. The agarose gel was prepared by dispersing agarose powder (Sigma-Aldrich, Saint Louis, MO) in a 10% buffer solution (Tris-borate EDTA, Gibco BRL, Rockville, MD). The mixture was then heated until the agarose was completely dissolved. After being cooled at room temperature to 60°C, the solution was loaded into a transparent container and was cooled further to room temperature (25°C) until solidification. The nanofluid distribution was studied in four agarose concentrations (0.5%, 1%, 2%, and 4%). Previous experimental studies showed that higher concentration agarose gels have a microstructure similar to hard tissue [20] while lower concentration gels have a porosity similar to soft tissue such as brain.

The ferrofluid was loaded on a syringe pump (Genie Plus, Kent Scientific, Inc. Torrington, CT) that enables a precise control of the flow rate and the amount of injection. The injection amount was selected as 0.1 cc or 0.2 cc [8,9] and the injection rate was adjusted so that the region containing the nanoparticle and the ferrofluid, which appears dark in the agarose gel, is approximately having a spherical shape. Figure 1 illustrates a semitransparent agarose gel with the injected ferrofluid represented by the dark color. Our previous experiments have suggested using a very low injection rate to provide gentle pressure to the gel [8]. Flow rates of 1 $\mu$l/min, 1.25 $\mu$l/min, 2.5 $\mu$l/min, and 3 $\mu$l/min were tested to give spherical shapes in the 4%, 2%, 1%, and 0.5% gel concentrations, respectively. The injection amount of 0.1 cc or 0.2 cc were selected instead of 0.3 cc used in the previous study [8] due to the consideration of the maximum specimen size allowed in the scan chamber of the imaging system.

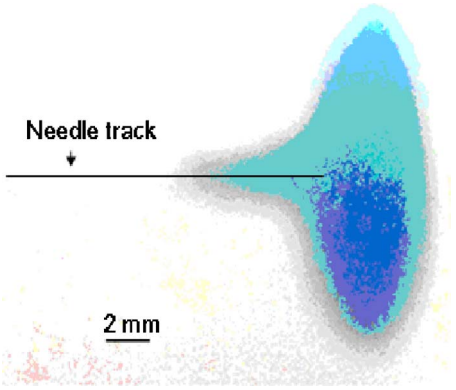



**Fig. 1 Ferrofluid infused in a semitransparent agarose gel. There is limited back flow of the ferrofluid along the needle track.**

**2.2 MicroCT Imaging.** Our previous experience on nanoparticle spreading has suggested that there is very little change in the nanoparticle spreading after injection within 24 h of the injection. In this study, we used a high resolution microCT imaging system (Skyscan 1172, Microphotonics, PA) to image the density of the agarose gel with nanoparticles. The gel block was mounted on the platform of the imaging system with the help of a stage holder. The gel block was placed in a low-density styrofoam, which would absorb little X-ray during the scan. A medium resolution scan of 17 $\mu$m (pixel size) was done at 100 kV, 100 $\mu$A without a filter. The total CT scan time was approximately less than 20 min. The images acquired from the microCT scan were reconstructed using the NRECON® software package provided by Microphotonics. Scan parameters and reconstruction parameters were kept the same for all the gel blocks in this study, therefore, the final pixel index number in the images is correlated with the same density of the sample. Each slice of the reconstructed images has a thickness of 17 $\mu$m. Maximum intensity projection images or pseudo 3D projections were generated from the reconstructed images. The density distribution was analyzed using the analysis software CTAN® (Microphotonics, PA). The change in the pixel density was analyzed in both the axial direction parallel to the needle and the radial direction.

**2.3 Specific Absorption Rate Measurements.** Nanoparticle heating was carried out via placing the gel block in a two-turn water cooled coil (20 cm in diameter and 7 cm in height), which is connected to a radio-frequency generator (Hotshot 2, Ameritherm, Inc., Rochester, NY). An alternative current of up to 384 A at a frequency of 183 kHz was generated through the coil and an alternating magnetic field was induced. Prior to each heating experiment, the gel compartment was kept at room temperature, so that a uniform temperature distribution can be established throughout the gel. The interference between the magnetic field and the thermocouple was found to be negligible. In the absence of any heat conduction in the gel, the volumetric heat generation rate or the SAR can be evaluated by the initial value of the transient term in the heat conduction equation [9,22–24]

$$\mathrm{SAR} = \rho c_p \left.\frac{\partial T}{\partial t}\right|_{t=0} \tag{1}$$

where $T$ is the temperature measurements at each gel location before and after the heating was turned on and $\rho$ and $c_p$ are the density and the specific heat of the gel, respectively [9]. The initial slope of each curve was obtained by a linear fitting of the first four measurements of the heating curve. A minimum value of 96% for

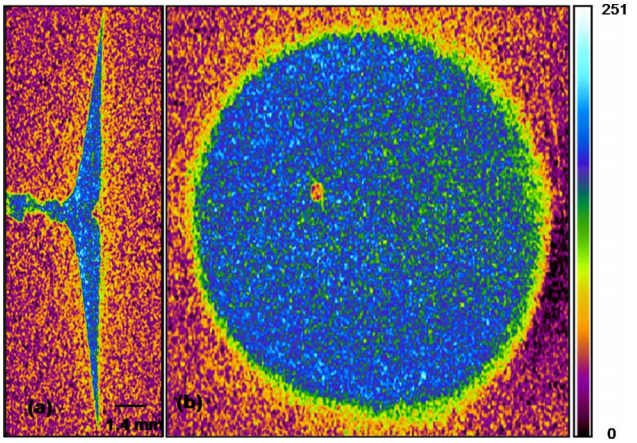



**Fig. 2 Side view (*a*) and top view (*b*) of the intensity index distribution of the disk pattern from the center slice**

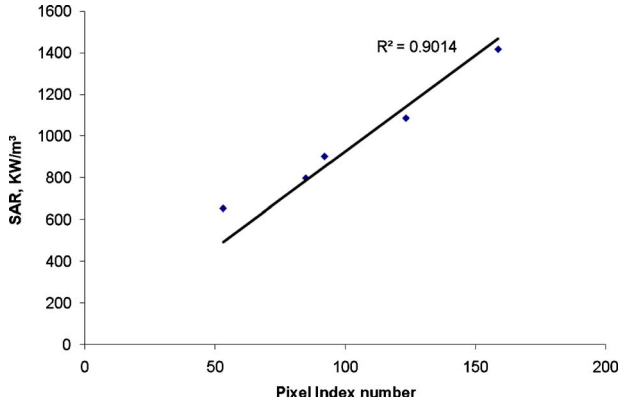



**Fig. 3 Linear correlations between the SAR measurements and the pixel index numbers. Lines represent curve fitted curve and symbols denote experimental measurements.**

the $R^2$ value was obtained for the curve fittings.

Guided by the X-ray images of the gel with ferrofluid, several fine thermocouples were inserted into the gel at locations having various pixel intensities, which represent different nanoparticle concentrations. The gel was then placed in the coil to measure the initial temperature rises at various locations and later to determine the slope of the initial temperature rising curves. All temperatures were measured and recorded by a PC based LABVIEW® program.

## 3 Results

A nanofluid spreading pattern was observed from the scanned images. Figure 2 provides typical pseudo three-dimensional microCT images of the density variations in the vicinity of the injection site where a disk shaped region containing the nanofluid can be seen. The top view (Fig. 2(*b*)) gives a nearly perfect circular region. The side view (Fig. 2(*a*)) illustrates the feature of the disk as well as the needle track. Back flow along the needle path is evident in the figure. The pixel index numbers shown in the microCT images seem relatively uniform in the disk region. Figure 2 also shows that infiltration of nanoparticles occurs around the edge of the high density disk.

Measuring the SAR values at various gel locations can provide a quantitative comparison to the pixel index number in the microCT images. In this experimental study, we use the microCT imaging system to identify a gel location representing a specific density (denoted by the pixel index number). Therefore, the pixel index number can be considered proportional to the nanoparticle concentration. The microCT images are used to guide the insertion of a thermocouple to that location to measure its SAR value based on Eq. (1). The relationship between the SAR magnitude and the pixel index number is illustrated in Fig. 3 where symbols represent experimental data of the SAR. A linear line is used to fit the scattered data and the fitting $R^2$ value is larger than 0.9. This confirms our original hypothesis that the pixel index number can be used to represent the concentration distribution of the nanoparticles since the ferrofluids used in the experiments contain the same nanoparticle concentration (3.9%). In principle, the SAR magnitude is proportional to the nanoparticle concentration in the tissue.

In this study, we also evaluated how the injection amount of the ferrofluid affects the particle spreading in the gel. As shown in Fig. 4, the lines represent the distribution of the pixel index number in the vicinity of the injection site for the nanoparticle dispersion pattern shown in Fig. 2. If we assume that the pixel index number is truly proportional to the nanoparticle concentration, the nanoparticles are dispersed in the gel quite uniformly. A few nanoparticles are seen outside the disk region. When the amount of the ferrofluid is doubled from 0.1 cc to 0.2 cc, the average pixel index number in the disk decreases by 20% while the radius of the disk region increases by almost 40%.

The relationship between the nanoparticle dispersion and the gel concentration is illustrated in Fig. 5. Smaller gel concentrations should result in higher porosity and less flow resistance to the ferrofluid and the nanoparticles. As shown in the measured disk span in the figure, nanoparticles are spreading farther away

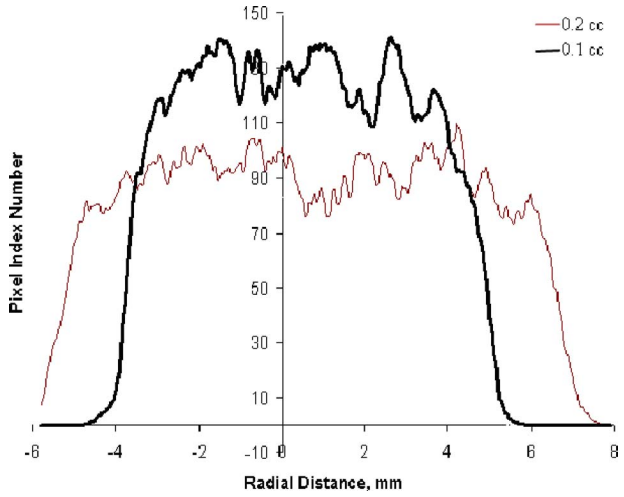



**Fig. 4 Pixel density profile along the radial distance for the particle spreading pattern in Fig. 3**

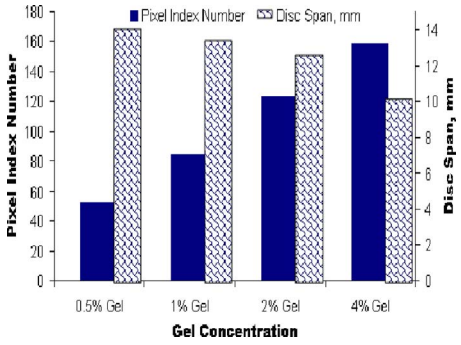



**Fig. 5 Variations of the average pixel index number inside the disk and the disk diameter as a function of the gel concentration. The left *y*-axis represents the pixel index and the right *y*-axis represents the disk diameter.**

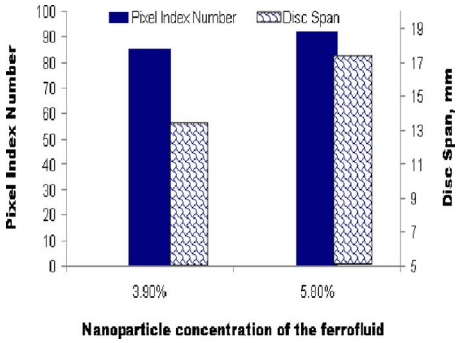



**Fig. 6 Variations of the pixel index number and the disk span affected by the magnetic particle concentration in the injected ferrofluid**

from the injection site when the gel concentration is low. The diameter of the disk increases from 10 mm to 12.8 mm when the gel concentration decreases by half (from 4% to 2%). More than 36% and 42% increase in the disk diameter (13.6 mm and 14.2 mm versus 10 mm) is observed when the gel concentration further decreases to 1% and 0.5%, respectively. It is clear that from the solid bars representing the pixel index number shown in Fig. 5, the dependence of the pixel index number on the gel concentration is reversed. The nanoparticle concentration inside the disk decreases by 25% when the gel concentration is lowered from 4% to 2%. Similarly, when the gel concentration is lowered from 4% to 0.5%, the decrease in the concentration of the nanoparticles is more than 63%.

Figure 6 illustrates the effect of nanoparticle concentration of the ferrofluid. Two nanoparticle concentrations of the ferrofluid (3.9% and 5.8%) are commercially available from the company. Using the same injection amount and injection rate, the ferrofluid with a higher magnetic particle concentration (5.8%) should contain more nanoparticles in the injected solution. It is not a surprise to see that both the pixel index number and the disk diameter are bigger in the images using the 5.8% ferrofluid than that using the 3.9% ferrofluid. The pixel index number only changes slightly from one ferrofluid concentration to another (84 versus 92), implying more concentrated nanoparticles surrounding the injection site when higher ferrofluid concentration is used. Due to increased number of nanoparticles in the injected amount with the 5.8%, the disk span also increases from 13.3 mm using the 3.9% ferrofluid to 17.7 mm with the 5.8% ferrofluid.

## 4 Discussion and Conclusions

Nanoparticle transport in tissue is a complicated process that involves nanofluid flow in the extracellular matrix of tissues, convection and diffusion of the particles in the fluid phase, deposition of the nanoparticles on the solid structure, and particle agglomeration. The complex heterogeneous structure of tissues and the deformation and breakage of a tissue induced by injection add additional difficulty in understanding the nanoparticle transport behavior during an injection. Due to the opaque nature of tissue and the insufficient techniques for characterization of nanoparticle concentration distribution, the understanding of nanoparticle transport in tissue remains limited.

Tissue equivalent phantom gels have been used in the past to study drug delivery in a tissue, considering that a gel is composed of a solid matrix with pores. The gel concentration and solidification temperature can be adjusted to mimic transport properties of various tissues, including tumors. One limitation of using gel to study nanoparticle transport in tumor is the homogeneous structure in comparison to the complicated tumor morphology heterogeneity and the different thresholds for the occurrence of pressure-induced breakage. Another limitation is the absence of vasculature in gels. However, gels are semitransparent porous materials and can provide visualization of drug spreading in drug delivery study. In the current investigation, using gel provides an easy way to identify the locations of the temperature sensors used in the SAR measurements. In principle, the local average SAR distribution depends on the number of nanoparticles per unit volume. Therefore, the comparison between the heating experiments and the microCT quantification can be used to establish a correlation between the SAR magnitude and the local nanoparticle concentration. Although CT has been used to visualize ferrofluid in the tissue in previous studies [6,7], the CT images were only a qualitative method to verify the total volume of the ferrofluid after injections. To the best of our knowledge, the current study is the first to establish quantitative relationship between the SAR and the microCT pixel index number, which is proportional to the local nanoparticle density.

The distinctive disk shape of ferrofluid spreading observed in the obtained 3D images suggested that they may be due to the delicate nature of the gel. Unlike real tissue, which can sustain sufficiently high pressure elevation without breakage, gels typically have a very low tolerance to pressure rise. It is highly possible that the disk shape obtained in experiments is a direct result of a tearing created by the injection needle. Once a crack is formed, the ferrofluid driven by the injection pressure tends to move along the vertical plane to further create a disk shaped cavity with its main axis along the vertical direction, as shown in Fig. 2. This explains the observation of a nearly uniform nanoparticle concentration within the disk. We also consider that the nanoparticles do not simply fill up the crack, rather, particle convection and diffusion have appreciable contribution to the spreading of ferrofluid in gels. Driven by the elevated infusion pressure, convection, diffusion, and deposition of the nanoparticles also occur in the porous gel structure adjacent to the cavity. Our observation and analysis were consistent with a previous work done by Nicholson [25] and Nicholson and Sykova [26] in the tissue environment. It suggested that the ferrofluid forms a cavity in the tissue that expands with further injected amounts and infiltrates in the extracellular space. Our results agree with the above prediction except for the time taken for the diffusion to occur, which may be attributed to the presence of blood perfusion in the tissue environment in their study.

The combined mechanism of nanoparticle transport by formation of a cavity at the injection site and particle convection and diffusion on the boundary can explain the ferrofluid spreading patterns observed in the experiment. For example, when doubling the injection volume of the ferrofluid, the increased injection amount causes formation of a larger cavity near the injection site and extended injection duration. The larger boundary area and longer injection time may facilitate particle convection and diffusion, thereby, leading to a larger disk span and a reduced average particle concentration, as shown in Fig. 4. Another observation is that the particle concentration in the disk region increases with the gel concentration while the disk span exhibits opposite trend. This can be explained by the deformation properties of gels of different concentrations. Typically, gels of low concentrations are easier to break and are more prone to deformation under the infusion pressure. They also have larger pores that not only impose less resistance to ferrofluid but also make particle penetration easier. With these properties, injection in softer gels tends to create larger cavity and the effect of convection and diffusion are more pronounced. The enhanced infiltration of the particles in the gels causes a large ferrofluid spreading region and reduces the average particle concentration.

The experiment with different ferrofluid concentration suggests that the concentration can substantially affect the spreading volume. We consider this observation a strong proof of the substantial contribution of convection and diffusion to nanoparticle transport in gels. In the absence of convection and diffusion, ferrofluids of different concentration should produce the same spreading volume if other injection parameters, such as volume, injection rate,

and gel properties, remain the same. However, the significant increase in the spreading volume for ferrofluid of higher concentration demonstrates that it is convection and diffusion that cause enhanced particle spreading in the gel. This is consistent with the measurements in a previous study [12] that the heat generation rate did not increase by 100% when the particle concentration was doubled.

In conclusion, our studies have shown the feasibility of using a high resolution microCT imaging system to obtain the 3D nanoparticle concentration distribution induced by directly injected nanoparticles in the gel. Although the particle spreading patterns in the gel may not be directly applied to real tissue, we believe that the current study lays the foundation to use microCT imaging systems to quantitatively study nanoparticle distribution in opaque tumor. The quantification of the relationship between the pixel index number and the nanoparticle concentration based on heat transfer experiments in this study will further help understand the distribution of the volumetric heat generation rate induced by nanoparticles. The investigation of nanoparticle transport pattern in gels suggests that convection and diffusion are important mechanisms for nanoparticle transport in porous structures. Our long term goal is to enhance treatment planning for the required thermal dose at each injection site in nanoparticle hyperthermia.

## Acknowledgment

The authors are gratefully acknowledging the financial support from the NSF research award (Grant No. CBET-0828728) and the NSF-MRI award (Grant No. CBET-0821236). The research was performed in partial fulfillmen of the requirements of the Ph.D. degrfee from University of Maryland Baltimore County by Anilchandra Attaluri.

## References

[1] Hergt, R., Andra, W., d'Ambly, C. G., Hilger, I., Kaiser, W. A., Richter, U., and Schmidt, H., 1998, "Physical Limits of Hyperthermia Using Magnetite Fine Particles," IEEE Trans. Magn., **34**, pp. 3745–3754.

[2] Rosensweig, R. E., 2002, "Heating Magnetic Fluid With Alternating Magnetic Field," J. Magn. Magn. Mater., **252**, pp. 370–374.

[3] Moroz, P., Jones, S. K., and Gray, B. N., 2002, "Magnetically Mediated Hyperthermia: Current Status and Future Directions," Int. J. Hyperthermia, **18**(4), pp. 267–284.

[4] Hilger, I., Hergt, R., and Kaiser, W. A., 2005, "Towards Breast Cancer Treatment by Magnetic Heating," J. Magn. Magn. Mater., **293**, pp. 314–319.

[5] Hergt, R., Hiergeist, R., Hilger, I., Kaiser, W. A., Lapatnikov, Y., Margel, S., and Richter, U., 2004, "Maghemite nanoparticles With Very High AC-Losses for Application in RF-Magnetic Hyperthermia," J. Magn. Magn. Mater., **270**, pp. 345–357.

[6] Wust, P., Gneveckow, U., Johannsen, M., Bohmer, D., Henkel, T., Kahmann, F., Sehouli, J., Felix, R., Ricke, J., and Jordan, A., 2006, "Magnetic Nanoparticles for Interstitial Thermotherapy—Feasibility, Tolerance and Achieved Temperatures," Int. J. Hyperthermia, **22**(8), pp. 673–685.

[7] Johannsen, M., Gneveckow, U., Thiesen, B., Taymoorian, K., Cho, C. H., Waldofner, N., Scholz, R., Jordan, A., Loening, S. A., and Wust, P., 2007, "Thermotherapy of Prostate Cancer Using Magnetic Nanoparticles—Feasibility, Imaging, and Three-Dimensional Temperature Distribution," Eur. Urol., **52**(6), pp. 1653–1662.

[8] Zhang, A., Mi, X., Yang, G., and Xu, X. L., 2009, "Numerical Study of Thermally Targeted Liposomal Drug Delivery in Tumor," J. Heat Transfer, **131**(4), p. 043209.

[9] Salloum, M., Ma, R., Weeks, D., and Zhu, L., 2008, "Controlling Nanoparticle Delivery in Hyperthermia for Cancer Treatment: Experimental Study in Agarose Gel," Int. J. Hyperthermia, **24**, pp. 337–345.

[10] Salloum, M., Ma, R., and Zhu, L., 2008, "An In-Vivo Experimental Study of Temperature Elevations in Animal Tissue During Magnetic Nanoparticle Hyperthermia," Int. J. Hyperthermia, **24**, pp. 589–601.

[11] Salloum, M., Ma, R., and Zhu, L., 2009, "Enhancement in Treatment Planning For Magnetic Nanoparticle Hyperthermia: Optimization of the Heat Absorption Pattern," Int. J. Hyperthermia, **25**, pp. 309–321.

[12] Masuko, Y., Tazawa, K., Viroonchatapan, E., Takemori, S., Shimizu, T., Fujimaki, M., Nagae, H., Sato, H., and Horikoshi, I., 1995, "Possibility of Thermosensitive Magnetoliposomes as a New Agent for Electromagnetic Induced Hyperthermia," Biol. Pharm. Bull., **18**, pp. 1802–1804.

[13] Lv, Y.-G., Deng, Z.-S., and Liu, J., 2005, "3-D Numerical Study on the Induced Heating Effects of Embedded Micro/Nanoparticles on Human Body Subject to External Medical Electromagnetic Field," IEEE Trans. Nanobiosci., **4**, pp. 284–294.

[14] Hergt, R., Hiergeist, R., Zeisberger, M., Glockl, G., Weitschies, W., Ramirez, L. P., Hilger, I., and Kaiser, W. A., 2004, "Enhancement of AC-Losses of Magnetic Nanoparticles for Heating Applications," J. Magn. Magn. Mater., **280**, pp. 358–368.

[15] Johannsen, M., Jordan, A., Scholz, R., Koch, M., Lein, M., Deger, S., Roigas, J., Jung, K., and Loening, S. A., 2004, "Evaluation of Magnetic Fluid Hyperthermia in a Standard Rat Model of Prostate Cancer," J. Endourol, **18**(5), pp. 495–500.

[16] Johannsen, M., Thiesen, B., Jordan, A., Taymoorian, K., Gneveckow, U., Waldofner, N., Scholz, R., Koch, M., Lein, M., Jung, K., and Loening, S. A., 2005, "Magnetic Fluid Hyperthermia (MFH) Reduces Prostate Cancer Growth in the Orthotopic Dunning R3327 Rat Model," Prostate, **64**(3), pp. 283–292.

[17] Bruners, P., Braunschweig, T., Hodenius, M., Pietsch, H., Penzkofer, T., Baumann, M., Günther, R. W., Schmitz-Rode, T., and Mahnken, A. H., 2010, "Thermoablation of Malignant Kidney Tumors Using Magnetic Nanoparticles: An In-Vivo Feasibility Study in a Rabbit Model," Cardiovasc. Intervent Radiol., **33**(1), pp. 127–134.

[18] Corot, C., Robert, P., Idée, J. M., and Port, M., 2006, "Recent Advances in Iron Oxide Nanocrystal Technology for Medical Imaging," Adv. Drug Delivery Rev., **58**, pp. 1471–1504.

[19] Yu, Y., and Sun, D., 2010, "Superparamagnetic Iron Oxide Nanoparticle 'Theranostics' for Multimodality Tumor Imaging, Gene Delivery, Targeted Drug and Prodrug Delivery," Expert Review of Clinical Pharmacology, **3**(1), pp. 117–130.

[20] Jain, R. K., 1999, "Transport of Molecules, Particles, and Cells in Solid Tumors," Annu. Rev. Biomed. Eng., **1**, pp. 241–263.

[21] Mahjoob, S., and Vafai, K., 2010, "Analysis of Bioheat Transport Through a Dual Layer Biological Media," J. Heat Transfer, **132**(3), p. 031101.

[22] Chen, Z. J., Broaddus, W. C., Viswanathan, R. R., Raghavan, R., and Gillies, G. T., 2002, "Intraparenchymal Drug Delivery via Positive-Pressure Infusion: Experimental and Modeling Studies of Poroelasticity in Brain Phantom Gels," IEEE Trans. Biomed. Eng., **49**, pp. 85–96.

[23] Zhu, L., Xu, L. X., and Chencinski, N., 1998, "Quantification of the 3-D Electromagnetic Power Absorption Rate in Tissue During Transurethral Prostatic Microwave Thermotherapy Using Heat Transfer Model," IEEE Trans. Biomed. Eng., **45**, pp. 1163–1172.

[24] Wang, X., Gu, H., and Yang, Z., 2005, "The Heating Effect of Magnetic Fluids in an Alternating Magnetic Field," J. Magn. Magn. Mater., **293**, pp. 334–340.

[25] Nicholson, C., 1985, "Diffusion From an Injected Volume of a Substance in Brain Tissue With Arbitrary Volume Fraction and Tortuosity," Brain Res., **333**, pp. 325–329.

[26] Nicholson, C., and Sykova, E., 1998, "Extracellular Space Revealed by Diffusion Analysis," Trends Neurosci., **21**, pp. 207–215.

# Principles of Tissue Engineering With Nonthermal Irreversible Electroporation

**Mary Phillips**
Department of Mechanical Engineering,
University of California, Berkeley,
6124 Etcheverry Hall,
Berkeley, CA 94720
e-mail: mary_phillips@berkeley.edu

**Elad Maor**
Biophysics Graduate Group,
University of California, Berkeley,
Berkeley, CA 94720
e-mail: eladmaor@gmail.com

**Boris Rubinsky**
Department of Mechanical Engineering and
Biophysics Graduate Group,
University of California, Berkeley,
Berkeley, CA 94720
e-mail: brubinsky@gmail.com

*Nonthermal irreversible electroporation (NTIRE) is an emerging tissue ablation modality that may be ideally suited in developing a decellularized tissue graft. NTIRE utilizes short electric pulses that produce nanoscale defects in the cell membrane lipid bilayer. The electric parameters can be chosen in such a way that Joule heating to the tissue is minimized and cell death occurs solely due to loss in cell homeostasis. By coupling NTIRE with the body's response, the cells can be selectively ablated and removed, leaving behind a tissue scaffold. Here, we introduce two different methods for developing a decellularized arterial scaffold. The first uses an electrode clamp that is applied to the outside of a rodent carotid artery and the second applies an endovascular minimally invasive approach to apply electric fields from the inner surface of the blood vessels. Both methods are first modeled using a transient finite element analysis of electric and thermal fields to ensure that the electric parameters used in this study will result in minimal thermal damage. Experimental work demonstrates that both techniques result in not only a decellularized arterial construct but an endothelial regrowth is evident along the lumen 7 days after treatment, indicating that the extracellular matrix was not damaged by electric and thermal fields and is still able to support cell growth.*
[DOI: 10.1115/1.4002301]

*Keywords: irreversible electroporation, bioheat transfer, tissue engineering*

## 1 Introduction

Nonthermal irreversible electroporation (NTIRE) has been developed as a method for controllable cell ablation. Electroporation occurs when an electric field is applied across the cell, destabilizing the electric potential maintained by the cell membrane and resulting in the formation of nanoscale defects in the lipid bilayer. This technology has been utilized over the last 30 years [1], temporarily polarizing the cell membrane and producing reversible pores in the lipid bilayer in order to introduce molecules, such as genes and drugs, into the cells. Changing the electrical parameters can, however, create permanent defects in the cell membrane and result in cell death from irreversible electroporation. An electric field can, by its very nature, create heating due to the Joule effect. It has been shown, however, that irreversible electroporation (IRE) can be isolated from this thermal effect and NTIRE can be used as an independent modality for tissue ablation [1]. NTIRE is unique from other tissue ablation methods. Not only does it avoid thermal damage, but it also produces a well defined region of tissue ablation with sharp, cell-scale borders between affected and unaffected regions [2,3]. NTIRE specifically targets the cell membrane and thus, spares other tissue components such as macromolecules, connective tissue, and tissue scaffold [2].

Recently, the nonthermal controllable cell ablation modality of NTIRE has been harnessed for medical applications such as the treatment of cancer. Miller et al. demonstrated the ability of NTIRE to ablate cancer cells in vitro [4], and in a more recent study, NTIRE was used to successfully ablate the prostrate of a dog, demonstrating that structures, such as the urethra, the vessels, the nerves, and the rectum, were undamaged by the treatment method [5]. NTIRE has also shown success in clinical trials for cancer treatment [6,7]. The effects of NTIRE on the blood vessels have also been examined for treatment of restenosis, indicating that this technology can be used to quickly and effectively ablate vascular smooth muscle cells without causing damage to the extracellular matrix (ECM) [8–10]. The results of this work in cancer treatment and restenosis indicate that NTIRE may also prove successful as a method for developing a decellularized tissue scaffold for use in tissue engineering applications.

A variety of methods have been employed to develop tissue engineered grafts for the replacement of diseased or damaged tissues and organs. Some of these methods have focused on developing a scaffold that is either seeded with cells in vitro or directly implanted and allowed to repopulate in vivo. Although a great deal of research has been aimed at developing biodegradable polymer scaffolds, others have focused on producing natural scaffolds for developing such tissue engineered grafts. Such a natural scaffold can be produced by decellularizing xenographic or human based tissue and repopulating it with the recipient's own cells, eliminating the need for immune-suppressant drugs and reducing the risk of graft rejection. Most tissue decellularization methods typically include some combination of physical, chemical, or enzymatic processes [11–14]. Although these have shown promise, as demonstrated by the work of Ott et al. [15] in developing a decellularized heart, there has, in general, been little long term follow-up [16] and some of the methods commonly employed have been shown to potentially risk damage to the ECM [17].

Here, we examine a new method of tissue decellularization that utilizes both the NTIRE and the body's host response. We have already shown in previous studies that after applying NTIRE, some mechanism acts to remove the ablated cells, resulting in a decellularized scaffold [8,18]. Those earlier studies led us to conceive a new method to obtain a tissue scaffold for tissue engineering by applying NTIRE to the donor tissue, allowing the donor's host response, most likely the immune system, to depopulate the targeted cells and then, harvesting the remaining tissue scaffold. The decellularized construct could then be implanted into the recipient and the cells would be able to repopulate the scaffold in vivo. In this paper, we focus on obtaining a decellularized arterial

Manuscript received February 25, 2010; final manuscript received March 19, 2010; published online September 27, 2010. Assoc. Editor: Andrey Kuznetsov.

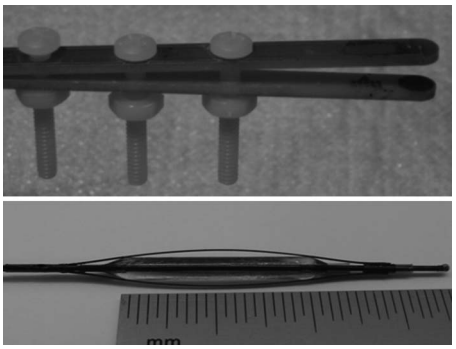


**Fig. 1 Electrodes used to apply NTIRE. The electrode clamp (top image) consists of two printed circuit boards with disk electrodes at the end. When used on the rat carotid artery, the electrodes are held apart by approximately 0.4 mm. The electrode catheter (bottom image) is shown in its inflated state. When in use, the four electrodes are pressed gently against the inner wall of the artery.**

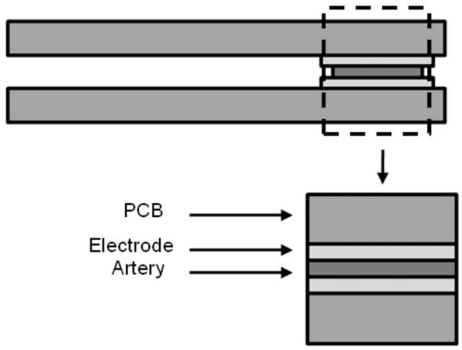



**Fig. 2 Schematic of model geometry for the electrode clamp and carotid artery. The artery is shown here in cross-section pressed between the two electrodes. The artery was modeled as $0.4\times3$ mm$^2$ and the copper electrodes are 0.1 mm thick. The printed circuit boards were modeled as having the material properties of Flame Retardant 4 (FR4) and have the dimensions of $1.6\times3$ mm$^2$. The artery-clamp system was modeled as being surrounded by a $3\times3$ cm$^2$ block of air.**

scaffold using two different methods: applying the electrodes to the outside of the artery and applying the electrical field minimally invasively using endovascular electrodes.

## 2 Finite Element Modeling

**2.1 Two Designs for Applying NTIRE.** We chose to examine the ability to produce a decellularized arterial scaffold using two different NTIRE-treatment methods. The first method used an electrode clamp to apply an electrical pulse to a rodent carotid artery. The electrode clamp used in this study is pictured in Fig. 1. It consists of two printed circuit boards with disk electrodes at the end. The artery is gently pressed between the electrode and a pulse is applied across the artery.

The second method examines the ability to apply NTIRE to an artery in a minimally invasive manner. The endovascular device is pictured in Fig. 1 and consists of four electrodes made of a rectangular nickel titanium wire, an electrically insulated catheter shaft, and a standard polyethylene terephthalate noncompliant balloon. The electrodes are oriented parallel to the catheter shaft and over the balloon and they are spaced out evenly around the circumference of the balloon. The electrodes can be retracted into a flexible tube in order for the device to be maneuvered to the desired artery location. Once in place, the electrodes can be expanded by pushing them forward out of the tube and gently pressed in contact with the inner wall of the artery by balloon inflation.

**2.2 Methods.** One of the key aspects of NTIRE in developing a decellularized tissue scaffold for tissue engineering applications is its ability to selectively damage the cell's membrane. Potential Joule heating from the electric field, however, is bound to occur and cannot be ignored. Although locally induced thermal damage has been utilized with drug delivery for cell ablation applications such as the treatment of cancer [19], such heating can also harm the ECM and thus, must be avoided here. Previous studies have also examined the effect of electroporation and thermal damage on other tissues such as the skin and liver [20–24]. In order to ensure that NTIRE does not cause significant protein denaturation due to Joule heating effects, electric parameters must be carefully designed. By decreasing the pulse length and the pulse frequency, the cell membrane can be targeted without resulting in thermal damage to the rest of the tissue components. In order to choose electrical parameters for experimental use that would not cause extensive heating and damage to the tissue, transient finite element analysis of both electrode devices was performed, modeling the effect of Joule heating on the temperature distribution of the tissue. These results were then examined to determine the accumulated thermal damage over time and to choose electrical parameters that would minimize that damage.

*2.2.1 Theoretical Model of the Clamp Electrode Device.* Using a finite element program (Comsol Multiphysics 3.5a), the temperature distribution throughout the arterial tissue was modeled. Due to the simplicity of the clamp electrode geometry, the artery-clamp system was modeled two-dimensionally, as depicted in Fig. 2. This simplification assumes that the artery and the electrodes are infinite in the axial direction, providing an overestimate of the temperature increase to the artery. The artery's dimensions were based on previous experimental observations and since both the electrode clamp and the artery are held very close to the body during the procedure, the artery-clamp system was modeled as surrounded by air at an elevated temperature of 37°C. The artery's thermal properties were assumed to be both isotropic and homogenous in cross-sections (see Fig. 2).

A solution for a single electroporation pulse was first modeled using the Laplace equation to evaluate the electric potential distribution:

$$\nabla \cdot (\sigma \nabla \phi) = 0 \tag{1}$$

where $\phi$ is the electric potential and $\sigma$ is the electrical conductivity. Equation (1) can be solved for the heat generation per unit volume ($q_{JH}$):

$$q_{JH} = \sigma |\nabla \phi|^2 \tag{2}$$

The electrodes were represented by a fixed voltage (Dirichlet) boundary condition. For the clamp electrode device, the top electrode was set to having a positive potential and the bottom electrode was set to zero:

$$\phi_1 = V_o \tag{3}$$

$$\phi_2 = 0 \tag{4}$$

where $V_o$ is the potential difference applied across the electrodes during the electroporation pulse. The boundaries between the artery and the air were set as electrically insulating.

Since the artery is exposed to the air during the procedure, the temperature was solved using conduction between the arterial tissue, electrodes, printed circuit boards, and air:

**Table 1 Thermal constants used in the simulation. The values obtained for FR4, copper, and air were taken from the COMSOL Multiphysics 3.5a material library.**

| | | | Tissue | FR4 | Copper | Air |
|---|---|---|---|---|---|---|
| Electrical conductivity | $\sigma$ | S/m | 0.6 [25] | 0.004 | $5.998\times10^{7}$ | 0.0001 |
| Heat capacity | $C$ | J/kg K | 3750 [26] | 1369 | 385 | $1.007\times10^{-3}$ |
| Density | $\rho$ | kg/m$^3$ | 1000 [26] | 1900 | 8700 | 1.1614 |
| Thermal conductivity | $k$ | W/m K | 0.5 [26] | 0.3 | 400 | 0.0263 |
| Initial temperature | $T_o$ | °C | 37 | 37 | 37 | 37 |

$$\rho C\frac{\partial T}{\partial t} - \nabla\cdot(k\nabla T) = q_{JH} \tag{5}$$

where $\rho$ is the material density, $C$ is the heat capacity, and $k$ is the thermal conductivity. In this model, $q_{JH}$ is determined from the Joule heating and is given in Eq. (2). In order to solve for the resulting temperature distribution, the entire system was initially held at the body temperature of the arterial tissue $T_o$ (37°C). The internal boundaries between the artery, electrodes, printed circuit board, and air were all defined as thermally continuous and the edges of the air space were maintained at $T_o$, providing a conservative overestimate on the rise in temperature to the artery. Thermal constants used in this evaluation are given in Table 1.

*2.2.2 Theoretical Model of the Catheter Electrode Device*. The endovascular device uses four longitudinal electrodes in contact with the inner surface of the arterial wall. A detailed description of the endovascular device was published elsewhere [9]. In a manner similar to that of the clamp electrode device, this system was reduced to a two-dimensional model. The inner diameter of the artery was taken as 2.5 mm based on an average diameter of rabbit iliac arteries. Since this models an intravascular procedure, the artery was assumed to be embedded in a large block of tissue, as depicted in Fig. 3. This two-dimensional model assumes that the electrodes are infinite in the axial direction, providing an overestimate on the resulting tissue temperature since, in reality, the electrodes are insulated on their ends and only contact the artery over 2 cm of their length.

The electrical pulse is modeled in a manner similar to that of the clamp electrode device using the Laplace equation, as given in Eqs. (1) and (2). The electrodes utilize a biopolar design with two electrodes having a positive potential and two electrodes having a potential of zero. All boundaries of the system not in contact with the electrodes were assumed to have a zero electric flux boundary condition:

$$\frac{\partial\phi}{\partial n} = 0 \tag{6}$$

Since the artery is assumed to be embedded within the tissue, the Pennes bioheat equation was used to determine the temperature distribution:

$$\nabla(k_t\nabla T) + \omega_b C_b(T_a - T) + q_{JH} + q + \sigma|\nabla\phi|^2 = \delta\rho_t C_t\frac{\partial T}{\partial t} \tag{7}$$

where $k_t$ is the thermal conductivity of the tissue, $T$ is the temperature, $\omega_b$ is the blood perfusion rate, $C_b$ is the heat capacity of the blood, $T_a$ is the arterial tissue temperature, $q_{JH}$ is heat generation obtained from the Joule heating (Eq. (2)), $q$ is the basal metabolic heat generation, $\rho_t$ is the tissue density, and $C_t$ is the tissue heat capacity. We assumed that the metabolic heat source was insignificant [26]. The initial temperature of the entire domain was set at the physiologic arterial tissue temperature ($T_a$). The boundaries along the inner surface of the artery were taken to be adiabatic in order to predict maximal temperature rise along the arterial wall. The outer boundary of the large block of tissue was held at constant physiological temperature ($T_a$) throughout the simulation. The thermal and biological properties used in this analysis are given in Table 2.

*2.2.3 Determining Thermal Damage*. The full procedure utilized $N$ number of square dc pulses of length $t_1$ and a pulse frequency rate of $f$. The temperature increases during each pulse due to resistive heating. Heat is dissipated due to conduction to the electrodes and surrounding air (for the clamp electrode case) and due to conduction to the surrounding tissue for the catheter electrode design. By incorporating intervals between pulses, where there is no resistive heating, the local rise in tissue temperature is kept to a minimum. In order to solve for the temperature distribution over the course of the procedure for a multiple pulse protocol, MATLAB 2008Rb (version 7.7) was used to run COMSOL Multiphysics 3.5a. A finite element mesh was incorporated that utilized triangular elements and the mesh size was varied in order to validate the accuracy of the solution. The coupled electric field and heat transfer equations were solved at each time step after each pulse and after each resting interval and the transient solution obtained at the end of each time step was used as the initial condition for the next time interval. The maximum arterial tissue temperature

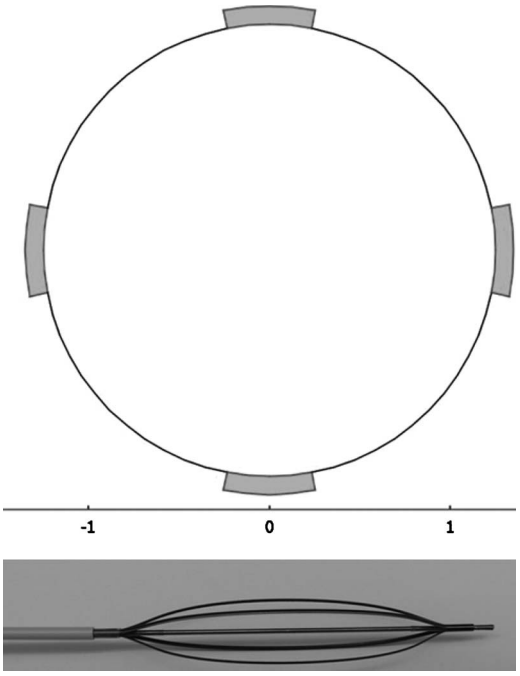



**Fig. 3 Two-dimensional geometry for the endovascular device. The four electrode nickel titanium wire electrodes (0.5 × 0.4 mm² in cross-section) run parallel to the longitudinal axis of the artery and lay pressed against the inner artery wall (2.5 mm in diameter). The electrodes are insulated from the arterial lumen space and the whole construct is modeled as being embedded in a very large block of tissue (not shown in full). Dimensions shown here are in millimeters.**

**Table 2 Thermal and biological constants used in the simulation**

| | | | Tissue | Blood | Catheter electrode |
|---|---|---|---|---|---|
| Electrical conductivity | $\sigma$ | S/m | 0.6 [25] | – | $4.032\times10^6$ |
| Heat capacity | $C$ | J/kg K | 3750 [26] | 3640 [27] | 100 |
| Density | $\rho$ | kg/m$^3$ | 1000 [26] | 1000 [28] | – |
| Thermal conductivity | $k$ | W/m K | 0.5 [26] | – | – |
| Perfusion rate | $\omega$ | 1/s | – | 0.0005 [29] | – |
| Tissue temperature | $T_a$ | °C | 37 | | |

was stored directly after the completion of each pulse as well as once every second for three minutes after the last pulse in order to account for the entire thermal damage due to Joule heating effects [9]. The maximum tissue temperature was used in order to ensure that a conservative estimate of thermal damage would be obtained.

Since the thermal damage to biological tissue is dependent on both temperature and time, the Arrhenius equation is often used to quantify these effects [30–36]. This model uses Maxwell–Boltzmann statistics to describe how biological molecules at a temperature $T$ are converted from a viable state to a thermally damaged state at a rate $K$ [31]. This reaction can be described by a first-order chemical rate process [37]:

$$K = A\exp\left(\frac{-\Delta E}{RT}\right) = \frac{d\Omega}{dt} \tag{8}$$

where $R$ is the ideal gas constant, $A$ is the measurement of molecular collision frequency, $\Delta E$ is the activation energy needed for the molecules to denature, $t$ is the time, and $\Omega$ is the accumulated damage. The damage parameter $\Omega$ can be expressed as the logarithm of the relative concentration of the undamaged molecules at time zero and time $\tau$:

$$\Omega(\tau) = \ln\left[\frac{C(0)}{C(\tau)}\right] \tag{9}$$

where $C(0)$ and $C(\tau)$ are the amount of undamaged molecules at time zero and time $\tau$, respectively. From Eq. (9), it can be seen that an example of $\Omega=1$ corresponds to 63.2% of the arterial tissue molecules having reached a thermally damaged state [38]. The Arrhenius equation given in Eq. (8) can be used to calculate the Henriques and Moritz thermal damage integral:

$$\Omega(t) = \int A\exp\left(\frac{-\Delta E}{RT}\right)dt \tag{10}$$

The values of $A$ and $\Delta E$ are based on experimental data and depend on the type of tissue under consideration [38]. For this analysis, these parameters are taken from a previous study [39] where $A$ and $\Delta E$ were determined for arterial tissue and the values are listed in Table 3. Equation (10) was applied to the entire procedure. By utilizing the maximum tissue temperature at each time step, an upper bound on the potential thermal damage to the tissue was obtained.

For the clamp electrode, the electric parameters that were analyzed by this model were determined from previous experiments [8] to produce NTIRE in the arterial tissue. These parameters consisted of 90 pulses of 70 V (corresponding to a 1750 V/cm electric field). Each pulse was 100 $\mu$s in length and the pulse frequency was 4 Hz. The electric parameters used for the endovascular electrodes also consisted of 90 pulses of 100 $\mu$s in length and a pulse frequency of 4 Hz. A voltage of 600 V was used, corresponding to an electrical field of 1000 V/cm or higher. These parameters are summarized in Table 4.

**Table 3 Constants used in the Arrhenius equation for arterial tissue**

| | | | | |
|---|---|---|---|---|
| Frequency factor | $A$ | 1/s | $5.6\times10^{63}$ | [33] |
| Activation energy | $\Delta E$ | J/mol | 430,000 | [33] |
| Ideal gas constant | $R$ | J/mol K | 8.314 | [33] |

**2.3 Results.** The solution to the Laplace equation for the electric potential distribution is static and independent of time. For each applied pulse, the electric field is nontransient. The electric field obtained from the clamp electrode is constant over the entire artery at 1750 V/cm due to the simple geometry. The electric field distribution for the catheter electrode design is shown in Fig. 4. The maximum temperature obtained for each model was recorded after each pulse and during the simulated cool down period following the applied pulse procedure. These results are shown in Fig. 5. The overall maximum temperature for the clamp electrode device obtained from the simulation was 45.25°C. The electric parameters applied to the endovascular electrode device induced a maximum temperature of 66.8°C. The maximum tissue temperature was obtained immediately after the 90th pulse for both electrode designs.

The Arrhenius integral (Eq. (10)) was evaluated to quantify the thermal damage obtained over the entire heating and cooling phases. This gave a value of $\Omega=7.163\times10^{-6}$ for the clamp electrode design, corresponding to negligible damage to the molecules due to Joule heating effects. The endovascular design resulted in $\Omega=0.0159$, indicating that approximately 1.6% of the molecules in the areas of maximal temperature became thermally damaged. Since the accumulated damage is very low for both designs and since these damage values were calculated based on the maximum tissue temperature rather than an average temperature value, giving an overestimate of any damage that might occur, it was determined that both NTIRE protocols could be safely used as a cell ablation method without risking thermal damage to the ECM.

## 3 Experiment

The electric field obtained using the clamp electrode device is perpendicular to the artery's vascualr smooth muscle cells (VSMC). The endovascular device, however, results in an electric field that is circular in orientation about the center axis of the artery. Thus, experimental results are necessary to compare the effectiveness of both these techniques in efficiently ablating the arterial cells without causing damage to the ECM.

**3.1 Methods.** The experimental procedures were described in earlier publications [8,18,40]. Specifically, the clamp electrode

**Table 4 Electric parameters analyzed**

| Parameter | | | Clamp electrode | Endovascular electrode |
|---|---|---|---|---|
| Applied voltage | $V_o$ | V | 70 | 600 |
| Electric field | – | V/cm | 1750 | >1000 |
| Pulse length | $t_1$ | $\mu$s | 100 | 100 |
| No. of pulses | $N$ | | 90 | 90 |
| Frequency | $f$ | Hz | 4 | 4 |

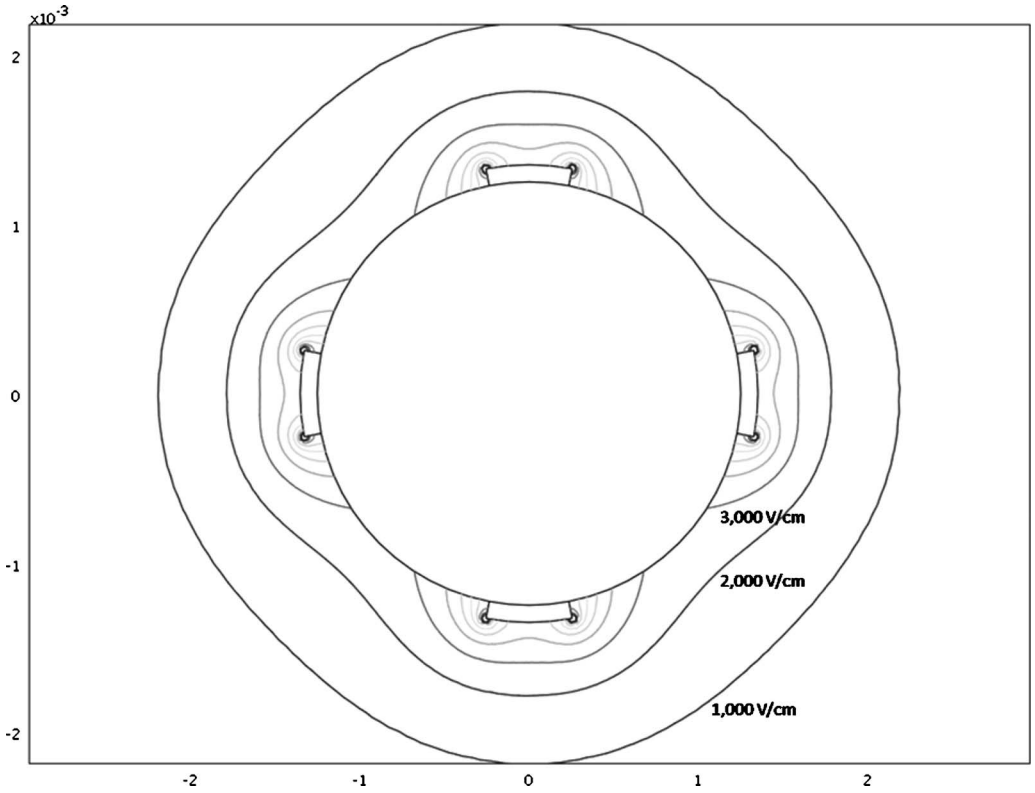



**Fig. 4 Two-dimensional electric field distribution. The resulting electric field is shown for the catheter electrode device. The outermost contour corresponds to 1000 V/cm and the electric field increases by 1000 V/cm for each contour moving in toward the electrodes. A spike in the electric field is seen at the corner of the electrodes due to edge effects. The model dimensions are shown in meters.**

procedure applied an electrical pulse to the rat carotid artery using Sprague-Dawley rats, in compliance with the Principals of Laboratory Animal Care as well as the Guide for the Care and Use of Laboratory Animals (NIH publication No. 85-23, revised 1985). Briefly, the animals were anesthetized with an intramuscular injection of ketamin and xylazine (90 mg/kg and 10 mg/kg, respectively) and vaporized isoflurane was used as anesthesia throughout the procedure. Using sterile techniques, the left common carotid artery was exposed and an electrode clamp, such as the one shown in Fig. 1, was used to apply an electric pulse close to the carotid artery's bifurcation. When gently pressing the carotid artery, the measured distance between the electrodes was approximately 0.4 mm. The electrodes were used to apply a sequence of 90 dc pulses

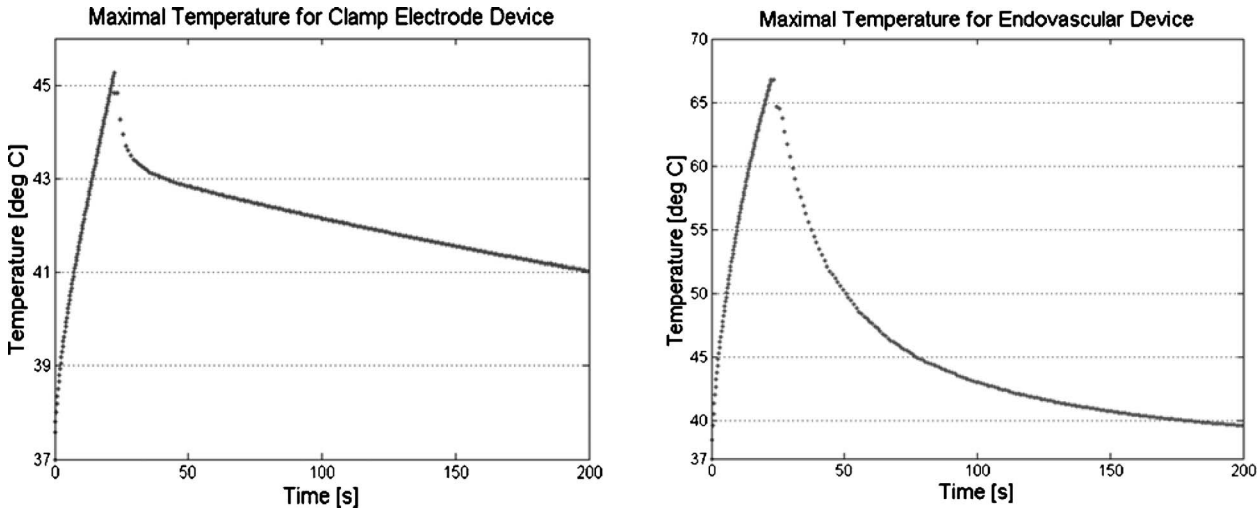



**Fig. 5 Transient solution of the maximum tissue temperature. The maximum temperature obtained for each time step over the course of the simulation is plotted for the clamp electrode design (left), indicating that the overall peak temperature is reached immediately after the final electrical pulse as expected. The maximum temperature obtained for the first 200 $\mu$m of the biological tissue domain are shown for the endovascular device (right).**

of 70 V, 100 $\mu$s length, and a frequency of 4 Hz using a high voltage pulse generator designed for electroporation procedures (ECM 80, Harvard Apparatus, Holliston, MA). The procedure was repeated in two successive locations along the carotid artery, resulting in the treatment of approximately 1.5 cm in length. The right common carotid artery was left alone and served as a control. All animals were kept alive for 7 days prior to being euthanized.

In order to experimentally test the endovascular electrode device, the iliac artery of New Zealand white rabbits was chosen as the model for this study since its dimensions are similar to that of the human coronary artery. The use of these animals was approved by the Institutional Animal Care and the Use Committee of ISIS services facility in Berkeley. Anesthesia was induced by ketamin (35 mg/kg) and xylazine (5 mg/kg) and this was followed by endotracheal intubation and isoflurane for anesthesia maintenance. Sterile techniques were used throughout the procedure. A 4F introducer was placed in the right femoral artery, the endovascular device was inserted in a retrograde manner, and angiography guidance was used to advance the catheter to the aortic bifurcation. The endovascular NTIRE device was inflated along the first two centimeters of the right iliac artery and an electrical sequence of 90 pulses of 600 V, 100 $\mu$s length, and a 4 Hz frequency was applied using a high voltage pulse generator (ECM 830, Harvard Apparatus, Holliston, MA). The endovascular NTIRE device was then removed and control angiography was performed to confirm patency of the vessel. The iliac artery was ligated and the surgical would was sutured closed. Animals recovered and were housed in the animal facility for 7 days prior to being euthanized.

For both experimental groups, animals were anesthetized with ketamin and xylazine and prior to being euthanized. Three centimeter segments of both iliac arteries from the rabbit groups and 1.5 cm of both carotid arteries from the rat groups were harvested, fixed in formalin, and submitted to independent pathology labs (Charles River Laboratories Pathology Associates, Fredrick, MD and Pathology Associates, Inc., Berkeley, CA). The samples were embedded in paraffin and cut into 5$\mu$. Each sample was stained with hematoxylin and eosin (H&E), and select samples were stained with Masson's trichrome and elastic Van Gieson (EVG) in order to determine the ability of NTIRE to ablate the vascular cells and the effect of the ablation method on the ECM, particularly the collagen and elastin fibers.

**3.2 Results.** Histological analysis of the rat carotid arteries treated by the clamp electrode device and the rabbit iliac arteries treated with the endovascular device indicate that both methods can result in an artery that is greatly decellularized, especially throughout the medial layer, as shown in Fig. 6. As can be seen in the figure, the endothelial layer has begun to regenerate by 7 days after treatment and this is especially evident in the carotid artery treated with the clamp electrode device. Further analysis with Masson's trichrome stain was used to demonstrate the loss of cell nuclei and vascular smooth muscle fibers in the treated arteries as well as the presence of collagen fibers after treatment (Fig. 7). EVG staining of both arteries showed intact elastin fibers as well as preservation of the vessel wall after NTIRE-treatment, as seen in Fig. 8.

## 4 Discussion

We have demonstrated the potential use of NTIRE and the systemic host response (potentially the immune system) to derive a functional decellularized tissue scaffold. Two different methods for applying NTIRE to the artery were analyzed and experimentally tested, illustrating the versatility of NTIRE as a tool for tissue engineering. We show that a decellularized artery can be developed in vivo, both by using an electrode clamp on the outside of the artery and by minimally invasive techniques, applying the electrical pulse from the inside of the artery using an endovascular electrode device. Such decellularized arteries could prove beneficial as potential grafts for revascularization surgeries.

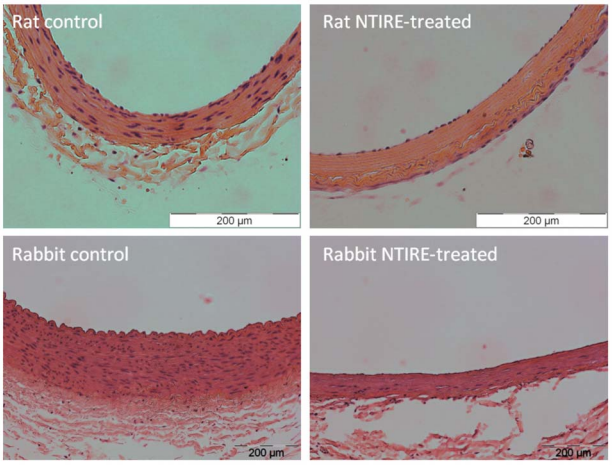



**Fig. 6 H&E staining results for both clamp electrodes and endovascular device experiments. H&E staining shows that the NTIRE-treatment of the rat carotid artery using the clamp electrode device resulted in an artery that was almost completely decellularized (top right) when compared with the control (top left). Treatment of the rabbit iliac artery using the endovascular electrode device resulted in the complete absence of VSMC at one week after treatment (bottom right) as compared with the control (bottom left).**

Many previously developed tissue decellularization methods require chemicals, enzymes, or lengthy procedures and although many of these methods have shown promise, some may cause damage to the ECM and lead to graft failure as well as slow down the graft development process [17]. The experimental results presented here indicate that within 7 days, the NTIRE and the host response treatment procedure results in a decellularized artery and that both treatment method designs were able to produce these

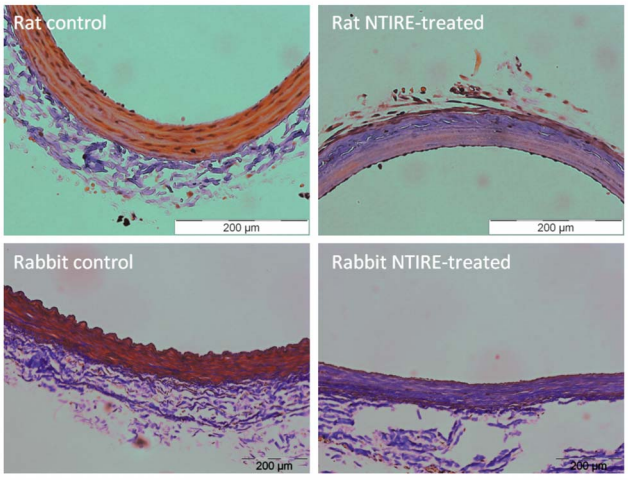



**Fig. 7 Masson's trichrome stain for both clamp electrodes and endovascular device experiments. Masson's trichrome stain demonstrates absence of cell nuclei (stained dark brown) and VSMC fibers (red) 7 days after NTIRE-treatment using both the clamp electrode device (upper right) and the endovascular device (lower right) when compared with their respective controls (upper left and lower left). These images also indicate that an abundance of collagen fibers remains after both treatment methods (stained blue).**

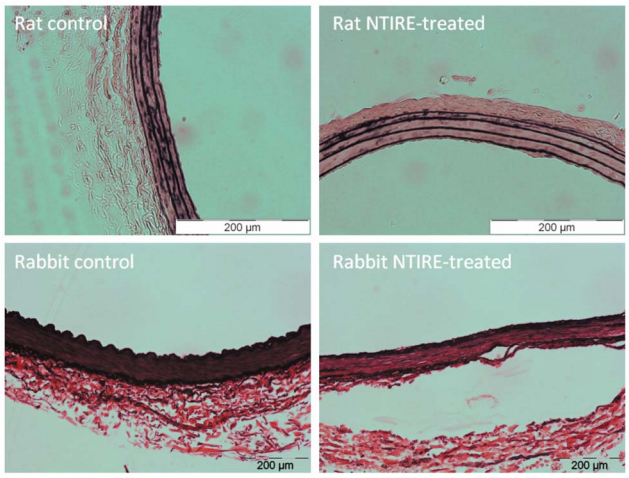



**Fig. 8 EVG staining for both clamp electrodes and endovascular device experiments. EVG staining demonstrates that the elastin fibers of the arteries treated by both the clamp electrode device (top right) and the endovascular device (bottom right) are undamaged 7 days after the NTIRE-treatment when compared with their respective controls (top left and bottom left).**

results. This is illustrated in Fig. 6 using H&E staining and is seen in the lack of cell nuclei and vascular muscle fibers when stained with Masson's trichrome (Fig. 7).

Further analysis of the resulting ECM supports the initial claim that NTIRE, by its very nature, does not harm the natural scaffold structure. After both treatment methods, there remains an abundance of collagen fibers (Fig. 7) and EVG staining indicates that the elastin fibers are undamaged 7 days after NTIRE-treatment (Fig. 8). These results are expected since IRE selectively disrupts the cell membrane's lipid bilayer. Thermal damage is eliminated by controlling the electrical parameters and minimizing Joule heating. Mathematical modeling of the effect of these electrical parameters using the Arrhenius equation gave thermal damage values of $\Omega = 7.163 \times 10^{-6}$ and $\Omega = 0.0159$ for the clamp electrode device and the endovascular device, respectively. This represents only 1.6% damage of the tissue molecules for the larger thermal damage case. This estimate gave an upper bound on potential tissue damage due to Joule heating since it utilized the maximum temperature seen throughout the tissue over the entire course of the simulation. Further efforts were made so that this model would overpredict the amount of thermal damage. For example, both electrode devices were modeled as being two-dimensional, resulting in an assumption that an electrode pulse was being applied along an infinite length of the artery when in reality, the electrical pulse was only applied along 0.5–2 cm of the artery's axis. Also, when modeling the electrode clamp device, only conduction between the artery, electrode clamp, and air were considered as a cooling method, ignoring any heat loss due to natural convection. The electrical and thermal models of the endovascular device did not incorporate heat convection due to the adjacent vein and the tissue conductivity used in both models was taken as 0.6 S/m. This is a significantly low value, probably lower than the true conductivity of the vessel wall [25], which is composed primarily of smooth muscle cells and elastic fibers. This lower conductivity results in a higher resistive heating to the tissue, further increasing the thermal damage value and allowing an increased factor of safety in this thermal damage estimate.

Lack of damage to the ECM is further illustrated by its ability to support additional cell growth after the NTIRE-treatment process. As can be seen in Fig. 6, the medial and adventitia layers are almost completely devoid of cells although a sparse layer of cells is evident, lining the artery lumen. These results are very important and indicate that the ECM is able to retain its function throughout the decellularization procedure and that potential issues with thrombogenicity that have been seen with many other tissue engineered grafts may be avoided. As illustrated in the histological analysis, both the clamp electrode device and the endovascular device result in an ECM that is able to support such cell growth within a week after treatment.

NTIRE is a very simple and controllable cell ablation technology. This has been shown from previous studies for the treatment of cancer [4,5]. Not only can electrical parameters be chosen such that thermal damage is avoided, but the electrical parameters and electrodes can also be designed in such a way as to control the electric field and thus, the extent of cell ablation. The clamp electrode device, as modeled here, results in an electrical field of 1750 V/cm between the two electrodes. Cells ablated using this device must be contained in the tissue placed between the two electrodes. This electric field can be controlled further using more complex electrode geometry, as demonstrated for the endovascular electrode device. Previous studies have shown that when using 90 electric pulses, an electric field of 1750 V/cm is required for successful cell ablation [8]. As shown here, the resulting electrical field can easily be modeled even when, as with the endovascular electrode device, the electric field varies with spatially. As illustrated in Fig. 4, the electric field and hence, the extent of cell ablation, can easily be visualized.

The clamp electrode and the endovascular electrode device apply NTIRE to the artery utilizing different methods, resulting in advantages and disadvantages to both techniques. The clamp electrode utilizes a uniform electric field between the electrodes. As a result, all tissue within the ablation region experiences the same electric parameters. The endovascular device, on the other hand, results in an electric field profile that decreases with distance from the electrodes and that spikes around the corners of the electrodes, as seen in Fig. 4. Finite element modeling also indicates that using the clamp electrode results in much less thermal damage than seen with the endovascular device. Nonetheless, although the endovascular device may result in greater thermal damage and a varying electric field, it still performs well. The electric field may vary due to the endovascular device's more complicated geometry but the device is still very simple to model and results in NTIRE that is very controllable. From the finite element modeling, it was determined that the endovascular device only results in 1.6% molecule denaturation at the points of maximum temperature at the corners of the electrodes. All other areas of the tissue would experience much less thermal damage. Also, although the electric field varies with the distance from the electrodes, NTIRE is unique in that it either results in cell death or leaves the cells undamaged. Thus, by knowing the electric field necessary to induce electroporation, the area of cell ablation can be easily predicted. The endovascular technique described in this paper allows for tissue ablation utilizing minimally invasive methods, reducing the risk of pain, infection, and other complications that can be experienced with the open surgery needed to apply NTIRE using the electrode clamp device. As can be seen, although both methods have their own advantages, the endovascular technique may become the preferred option due to its minimally invasive characteristics.

This study demonstrates the ability to obtain a decellularized artery for use as a tissue graft using two different application techniques: the clamp electrode applied to the outside of the artery and the endovascular electrode device applied minimally invasively. It is shown that within one week after treatment, the artery becomes decellularized. New endothelial cell growth is seen along the lumen layer, demonstrating the ECM can still support cell growth. This study illustrates that with the support of mathematical modeling, NTIRE can be used to decellularize arteries and perhaps other tissue types that differ in location, size, and species. NTIRE is a simple, controllable, and versatile cell ablation method that shows great promise in obtaining decellularized tissue constructs for use as tissue grafts.

## Nomenclature

$\phi$ = electric potential
$\sigma$ = electrical conductivity
$q_{JH}$ = heat generation per unit volume
$V_o$ = potential difference
$\rho$ = material density
$C$ = heat capacity
$k$ = thermal conductivity
$T_o$ = initial body temperature
$k_t$ = tissue thermal conductivity
$T$ = temperature
$\omega_b$ = blood perfusion rate
$C_b$ = blood heat capacity
$T_a$ = arterial tissue temperature
$q$ = basal metabolic heat generation
$\rho_t$ = tissue density
$C_t$ = tissue heat capacity
$K$ = rate molecules convert from viable to thermally damaged states
$R$ = ideal gas constant
$A$ = molecular collision frequency
$\Delta E$ = activation energy
$t$ = time
$\Omega$ = thermal damage parameter
$C(0)$ = amount of undamaged molecules at time zero
$C(\tau)$ = amount of undamaged molecules at time $\tau$
$t_1$ = pulse length
$N$ = number of pulses
$f$ = pulse frequency

## References


[1] Rubinsky, B., 2007, "Irreversible Electroporation in Medicine," Technol. Cancer Res. Treat., **6**(4), pp. 255–259.
[2] Lavee, J., Onik, G., Mikus, P., and Rubinsky, B., 2007, "A Novel Nonthermal Energy Source for Surgical Epicardial Artrial Ablation: Irreversible Electroporation," Heart Surg. Forum, **10**(2), pp. E162–E167.
[3] Rubinsky, B., Onik, B., and Mikus, P., 2007, "Irreversible Electroporation: A New Ablation Modality—Clinical Implications," Technol. Cancer Res. Treat., **6**(1), pp. 37–48.
[4] Miller, L., Leor, J., and Rubinsky, B., 2005, "Cancer Cell Ablation With Irreversible Electroporation," Technol. Cancer Res. Treat., **4**(6), pp. 699–705.
[5] Onik, G., Mikus, P., and Rubinsky, B., 2007, "Irreversible Electroporation: Implications for Prostrate Ablation," Technol. Cancer Res. Treat., **6**(4), pp. 295–300.
[6] Onik, G., and Rubinsky, B., 2010, "Irreversible Electroporation: First Patient Experience Focal Therapy of Prostate Cancer," *Irreversible Electroporation*, Springer, Berlin, pp. 235–247.
[7] Thomson, K., 2010, "Human Experience With Irreversible Electroporation," *Irreversible Electroporation*, Springer, Berlin, pp. 249–254.
[8] Maor, E., Ivorra, A., and Rubinsky, B., 2009, "Non Thermal Irreversible Electroporation: Novel Technology for Vascular Smooth Muscle Cell Ablation," PLoS ONE, **4**(3), pp. e4757, 1–9.
[9] Maor, E., and Rubinsky, B., 2010, "Endovascular Non-Thermal Irreversible Electroporation: A Finite Element Analysis," ASME J. Biomech. Eng., **132**, p. 031008.
[10] Maor, E., Ivorra, A., Leor, J., and Rubinsky, B., 2007, "The Effect of Irreversible Electroporation on Blood Vessels," Technol. Cancer Res. Treat., **6**(4), pp. 307–312.
[11] Huynh, T., Abraham, G., Murray, J., Brockbank, K., Hagen, P.-O., and Sullivan, S., 1999, "Remodeling of an Acellular Collagen Graft Into a Physiologically Responsive Neovessel," Nat. Biotechnol., **17**, pp. 1083–1086.
[12] Clarke, D., Lust, R., Sun, Y., Black, K., and Ollerenshaw, J., 2001, "Transformation of Nonvascular Acellular Tissue Matrices Into Durable Vascular Conduits," Ann. Thorac. Surg., **71**, pp. S433–S436.
[13] Conconi, M. T., De Coppi, P., Di Liddo, R., Vigolo, S., Zanon, G. F., Parnigotto, P. P., and Nussdorfer, G. G., 2005, "Tracheal Matrices, Obtained by a Detergent-Enzymatic Method, Support *In Vitro* the Adhesion of Chondrocytes and Tracheal Epithelial Cells," Transpl Int., **18**, pp. 727–734.
[14] Flynn, L., Semple, J., and Woodhouse, K., 2006, "Decellularized Placental Matrices for Adipose Tissue Engineering," J. Biomed. Mater. Res. Part A, **79**(2), pp. 359–369.
[15] Ott, H., Matthiesen, T., Go, S.-K., Black, L., Kren, S., Netoff, T., and Taylor, D., 2008, "Perfusion-Decellularized Matrix: Using Nature's Platform to Engineer a Bioartificial Heart," Nat. Med., **14**, pp. 213–221.
[16] Campbell, B., and Campbell, J., 2007, "Development of Tissue Engineered Vascular Grafts," Curr. Pharm. Biotechnol., **8**, pp. 43–50.
[17] Gilbert, T., Sellaro, T., and Badylak, S., 2006, "Decellularization of Tissues and Organs," Biomaterials, **27**, pp. 3675–3683.
[18] Maor, E., Ivorra, A., Mitchell, J., and Rubinsky, B., "Vascular Smooth Muscle Cell Ablation With Endovascular Non Thermal Irreversible Electroporation," J. Vasc. Interv. Radiol., accepted.
[19] Zhang, A., Mi, X., Yang, G., and Xu, L., 2009, "Numerical Study of Thermally Targeted Liposomal Drug Delivery in Tumor," J. Heat Transfer, **131**(4), p. 043209.
[20] Becker, S., and Kuznetsov, A., 2007, "Local Temperature Rises Influence In Vivo Electroporation Pore Development: A Numerical Stratum Corneum Lipid Phase Transition Model," ASME J. Biomech. Eng., **129**, pp. 712–721.
[21] Becker, S., and Kuznetsov, A., 2008, "Thermal In Vivo Skin Electroporation Pore Development and Charged Macromolecule Transdermal Delivery: A Numerical Study of the Influence of Chemically Enhanced Lower Lipid Phase Transition Temperatures," Int. J. Heat Mass Transfer, **51**(7–8), pp. 2060–2074.
[22] Becker, S., and Kuznetsov, A., 2007, "Numerical Assessment of Thermal Response Associated With In Vivo Skin Electroporation: The Importance of the Composite Skin Model," ASME J. Biomech. Eng., **129**(3), pp. 330–340.
[23] Becker, S., and Kuznetsov, A., 2007, "Thermal Damage Reduction Associated With In Vivo Skin Electroporation: A Numerical Investigation Justifying Aggressive Pre-Cooling," Int. J. Heat Mass Transfer, **50**(1–2), pp. 105–116.
[24] Becker, S. M., and Kuznetsov, A. V., 2006, "Numerical Modeling of In Vivo Plate Electroporation Thermal Dose Assessment," ASME J. Biomech. Eng., **128**(1), pp. 76–84.
[25] Gabriel, S., Lau, R., and Gabriel, C., 1996, "The Dielectric Properties of Biological Tissues: II. Measurements in the Frequency Range 10 Hz to 20 GHz," Phys. Med. Biol., **41**(11), pp. 2251–2269.
[26] Davalos, R., Rubinsky, B., and Mir, L., 2003, "Theoretical Analysis of the Thermal Effects During In Vivo Tissue Electroporation," Bioelectrochemistry, **61**, pp. 99–107.
[27] Lee, R., and Despa, F., 2005, "Distinguishing Electroporation From Thermal Injuries in Electrical Shock by MR Imaging," Conference Proceedings−IEEE Engineering in Medicine and Biology Society, Vol. 6, pp. 6544–6546.
[28] Liu, J., and Xu, L. X., 1999, "Estimation of Blood Perfusion Using Phase Shift in Temperature Response to Sinusoidal Heating at the Skin Surface," IEEE Trans. Biomed. Eng., **46**(9), pp. 1037–1043.
[29] Wissler, E. H., 1998, "Pennes' 1948 Paper Revisited," J. Appl. Physiol., **85**(1), pp. 35–41.
[30] Tropea, B., and Lee, R., 1992, "Thermal Injury Kinetics in Electrical Trauma," J. Biomech. Eng., **114**, pp. 241–250.
[31] Lee, R., 1991, "Physical Mechanics of Tissue Injury in Electrical Trauma," IEEE Trans. Educ., **34**(3), pp. 223–230.
[32] Chang, I., and Nguyen, U., 2004, "Thermal Modeling of Lesion Growth With Radiofrequency Ablation Devices," Biomed. Eng. Online, **3**(27), pp. 1–19.
[33] Agah, R., Pearce, J., Welch, A., and Motamedi, M., 1994, "Rate Process Model for Arterial Tissue Damage: Implications on Vessel Photocoagulation," Lasers Surg. Med., **15**, pp. 176–184.
[34] Orgill, D., Solari, M., Barlow, M., and O'Connor, N., 1998, "A Finite-Element Model Predicts Thermal Damage in Cutaneous Contact Burns," J. Burn Care Rehabil., **19**(3), pp. 203–209.
[35] Lee, R., and Astumian, R., 1996, "The Physicochemical Basis for Thermal and Non-Thermal 'Burn' Injuries," Burns, **22**(7), pp. 509–519.
[36] Wright, N., 2003, "On a Relationship Between the Arrhenius Parameters From Thermal Damage Studies," J. Biomech. Eng., **125**(2), pp. 300–304.
[37] Maor, E., Ivorra, A., Leor, J., and Rubinsky, B., 2008, "Irreversible Electroporation Attenuates Neointimal Formation After Angioplasty," IEEE Trans. Biomed. Eng., **55**(9), pp. 2268–2274.
[38] Diller, K., and Pearce, J., 1999, "Issues in Modeling Thermal Alterations in Tissues," Ann. N.Y. Acad. Sci., **888**, pp. 153–164.
[39] Pearce, J., and Thomsen, S., 1992, "Kinetic Models of Tissue Fusion Processes," Proc. SPIE, **1643**, pp. 251–260.
[40] Maor, E., 2009, "Fundamental Study on the Effects of Irreversible Electroporation Pulses on Blood Vessels With Application to Medical Treatment," Ph.D. thesis, University of California Berkeley, Berkeley, CA.

# Cryosurgery: Analysis and Experimentation of Cryoprobes in Phase Changing Media

**Avraham Shitzer**
Department of Mechanical Engineering,
Technion, Israel Institute of Technology,
Haifa 32000, Israel
e-mail: mersasa@tx.technion.ac.il

*This article presents a retrospective of work performed at the Technion, Israel Institute of Technology, over the last 3-odd decades. Results of analytical and numerical studies are presented briefly as well as in vitro and in vivo experimental data and their comparison to the derived results. Studies include the analysis of both the direct (Stefan) and the inverse-Stefan phase-change heat transfer problems in a tissue-simulating medium (gel) by the application of both surface and insertion cryoprobes. The effects of blood perfusion and metabolic heat generation rates on the advancement of the freezing front are discussed. The simultaneous operation of needle cryoprobes in a number of different configurations and the effects of a thermally significant blood vessel in the vicinity of the cryoprobe are also presented. Typical results demonstrate that metabolic rate in the yet nonfrozen tissue, will have only minor effects on the advancement of the frozen front. Capillary blood perfusion, on the other hand, does affect the course of change of the temperature distribution, hindering, as it is increased, the advancement of the frozen front. The volumes enclosed by the "lethal" isotherm (assumed as −40°C), achieve most of their final size in the first few minutes of operation, thus obviating the need for prolonged applications. Volumes occupied by this lethal isotherm were shown to be rather small. Thus, after 10 min of operation, these volumes will occupy only about 6% (single probe), 6–11% (two probes, varying distances apart), and 6–15% (three probes, different placement configurations), relative to the total frozen volume. For cryosurgery to become the treatment-of-choice, much more work will be required to cover the following issues: (1) A clear cut understanding and definition of the tissue-specific thermal conditions that are required to ensure the complete destruction of a tissue undergoing a controlled cryosurgical process. (2) Comprehensive analyses of the complete freeze/thaw cycle(s) and it effects on the final outcome. (3) Improved technical means to control the temperature variations of the cryoprobe to achieve the desired thermal conditions required for tissue destruction. (4) Improvement in the pretreatment design process to include optimal placement schemes of multiprobes and their separate and specific operation. (5) Understanding the effects of thermally significant blood vessels, and other related thermal perturbations, which are situated adjacent to, or even within, the tissue volume to be treated.* [DOI: 10.1115/1.4002302]

*Keywords: inverse problem, lethal temperature, cooling rate, frozen volume*

## 1 Introduction

Cryosurgery is a medical technique involving the application of extremely low (cryogenic) temperatures with an aim to destroy abnormal or diseased tissues by freezing [1]. Applications include the treatment of a variety of benign and malignant skin conditions and diseases of internal organs (e.g., prostate). Cryosurgery was in use as early as in the middle 19th century when James Arnott, an English physician, used salt-ice mixtures to treat malignant tumors [2]. The so-called "modern era" of cryosurgery was evidently initiated in 1961, when Irving Cooper, an American neurosurgeon, developed a liquid nitrogen-based cryosurgical system [3]. In this system the cryofluid was used in a heat conducting metal tube, the closed end of its tip was brought into contact with the treated tissue. This design still forms the basis for a variety of instruments used in this field to date.

The destruction of biological tissues by the application of freezing-thawing cycles may be achieved by either, or a combination of the following processes: immediate and delayed. The immediate process involves direct destruction of cells and is dominated by the various modes of extra- and intracellular ice formation during the freezing stage. The delayed, post-application destruction mechanisms are due to damages resulting from the destruction of blood vessels and/or an invoked immune system response.

The key advantages of the cryosurgical procedure are as follows: (a) it is minimally invasive, (b) the application is localized (in situ), (c) it causes minimal trauma to the patient, (d) it involves minimal loss of blood, (e) it possesses anesthetic capabilities due to the subfreezing temperatures involved, (f) its application duration is relatively short, (g) it is cycleable and repeatable (h), it requires minimal hospitalization, or it might even be applied ambulatorily, (i) it incurs relatively low costs, and (j) in certain cases, it may invoke a delayed immune system response to remove the undesired, recently frozen tissue.

In spite of these apparent advantages, the cryosurgical technique is still finding only limited applications, mainly in urology [4] and in dermatology [5]. The reason for this lack of wider applicability stems mainly from the following reasons: (a) the uncertainty in the final outcome of the procedure since the identical physical mechanism, namely, *in vivo* tissue freezing, is also used to achieve quite the opposite outcome—tissue preservation, compare Ref. [6], (b) the inability to ensure complete freezing of the entire target volume, in a short application due usually to its irregular shape, (c) possible undesired excessive damage to sur-

Manuscript received January 7, 2010; final manuscript received February 3, 2010; published online September 27, 2010. Assoc. Editor: Andrey Kuznetsov.

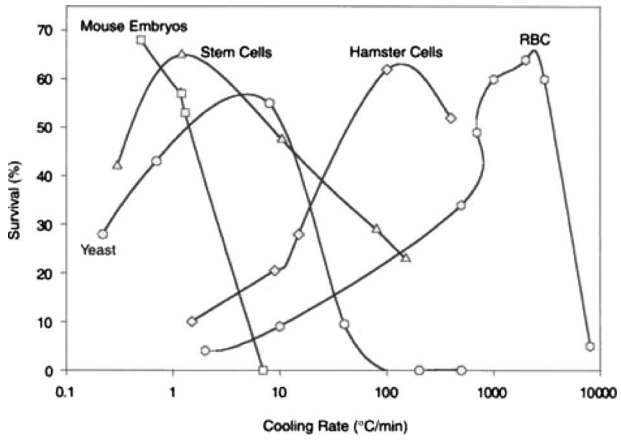



**Fig. 1 Cell survival versus cooling rate applied [11]. (This figure originally appeared in an article by J.K. Critser and L.E. Mobraaten in ILAR Journal 41(1). It is reprinted with permission from the ILAR Journal, Institute for Laboratory Animal Research, The National Academies, Washington DC (www.nationalacademies.org/ilar)).**

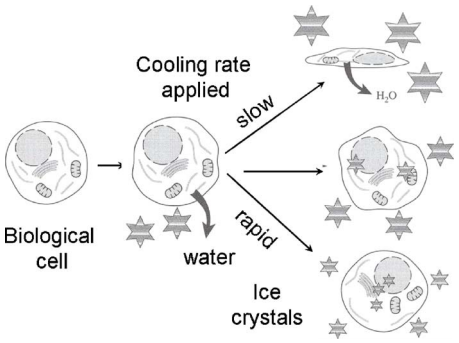



**Fig. 2 Effects of cooling rates applied to biological cells on intra- and extracellular water freezing**

rounding healthy tissues and adjacent major blood vessels, (d) difficulties in controlling the extent of the frozen volumes, (e) difficulties in monitoring the precise temperature distributions in the tissue during each freeze-thaw cycle, and (f) susceptibility to thermal perturbations by adjacent thermally significant blood vessel(s).

The simultaneous application of multiprobes, coupled with the use of modern imaging techniques, e.g., ultrasound, computerized tomography (CT) and magnetic resonance imaging (MRI), have alleviated some of these hindrances to a certain extent and have extended the use of cryosurgery in the treatment of prostate and liver cancer [1,4]. These imaging techniques, however, are still incapable of acquiring detailed temperature distribution data inside the frozen volume. Ultrasound, which is the most commonly applied imaging technique in cryosurgery, has its own limitations. Among these is the opaqueness of the frozen phase to sound waves due to ice formation. Thus, obtaining detailed, temporal, three-dimensional temperature distribution data inside the freezing tissue behind the frozen front by ultrasound is not possible. Currently, even the more expensive and complex CT or MRI imaging techniques, are incapable of providing temperature distribution details inside the frozen volume.

An effective cryosurgical treatment, particularly of malignant tissues, requires that optimal tissue destruction conditions be achieved throughout the entire volume of the tumor, including a predetermined margin for certainty. It is well known that different types of tissue cells exhibit different sensitivities to freezing but, as a general rule, the lower the temperature achieved, the higher the probability of destruction. In the treatment of cancer, a sufficiently low temperature, termed "lethal temperature," should be ensured to achieve effective destruction of the target volume. Mazur [7] defined the lethal temperatures for cell destruction within the range of $-5\,^{\circ}C$ to $-50\,^{\circ}C$. Intracellular ice forms in prostate cells at temperatures below $-40\,^{\circ}C$ [8], which, therefore, serves as a target temperature in this application. Other investigators suggest that even higher temperatures, e.g., $-20\,^{\circ}C$, may be lethal to the cells [9].

The cooling rate maintained at the freezing front, has also been implicated as a key factor, which determines the probability of survival of the frozen tissues [10]. Accordingly, for each cell type, there exist certain ranges of cooling rates, which, when applied at the phase-transition region behind the freezing front, would increase the probability of either cell survival or cell destruction. Typical plots of measured cell survival versus cooling rates shown in Fig. 1 [11] resemble bell-shaped curves. This implies that optimal conditions for cell destruction may be achieved only at either low or high cooling rates. At low cooling rates, of a few degrees Centigrade per minute, extracellular ice crystals are formed first. This disrupts the osmotic balance between the intra- and the extracellular fluids causing osmotic drying of the cell and its eventual shrinkage. This process, termed the "solution effect," may render the cell nonviable leading to its death upon thawing. Sharp intracellular ice crystals are formed at high cooling rates of hundreds degrees Centigrade per minute. This process, termed the "mechanical effect," may cause cell membrane leakage and, consequently, compromise its integrity leading to cell death upon thawing. At intermediate cooling rates, however, the experimental data suggest high probabilities of cell survival, rather than cell destruction. These processes were reviewed by Gage and Baust [12] and more recently by Hoffman and Bischof [13]. Figure 2 demonstrates schematically the effects of various cooling rates on biological cell behavior during phase-change.

It follows from the above discussion that precise and detailed information of the temperature field, which develops during the tissue freezing process, is essential to the surgeon, both at the pretreatment design stage as well as during the application stage. Such information may be obtained by solving the dynamic heat transfer problem involved. The problem to be solved, referred to as the "Stefan" problem, is nonlinear mainly due to the removal of the heat of fusion, which is liberated at the moving front separating the frozen and nonfrozen phases. In biological tissues, unlike pure substances, e.g., pure water, phase-change transition occurs over a temperature range rather than at a single temperature, a factor which further complicates the analysis. An additionally, rather complex associated problem, is the "inverse-Stefan" problem in which a certain predetermined cooling rate is to be imposed and maintained at the freezing front.

Analytical solutions to the Stefan problem are few [14–16] as are those involving phase-change in biological tissue [17–19]. Consequently, investigators revert to employing numerical solution techniques, e.g., front tracking method [20] or the enthalpy method [21,22]. The application of multiprobes in tissuelike substances were analyzed by Keanini and Rubinsky [23] who presented a general technique for optimizing cryosurgical procedures. Rabin and Stahovich [24] and Rabin et al. [25] introduced cryoheaters as a means of controlling the extent of the multiprobes' frozen region in order to protect certain tissue regions from being affected during the cryosurgical process. No experimental data were presented in any of these latter studies.

Rewcastle et al. [26] analyzed the ice ball formation around a single cryoprobe using an axisymmetric, finite difference model. Model predictions were compared with measured data and were shown to conform to within $\pm 5\,^{\circ}C$. Jankun et al. [27] developed an interactive software simulation package (CRYOSIM) for cryoablation of the prostate by liquid nitrogen-operated cryoprobes. The model is based on a finite difference numerical technique. Acquired ultrasound data were used to perform on-line adjust-

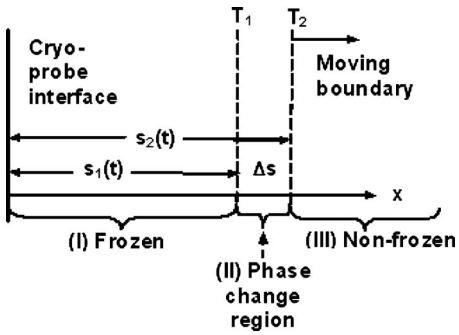



**Fig. 3 Schematic representation of phase change in biological materials**

ments of model parameters and in the prediction of the temporal variations of the isothermal surfaces in the prostate. The package facilitates the simulation and tracking of the "therapeutic temperature" in the prostate. Baissalov et al. discussed a semi-empirical treatment planning model for optimization of multiprobes [28]. A finite elements procedure was developed to analyze the 3D heat transfer problem in a model of the prostate. Predicted locations of the frozen front were compared with experimental X-ray readings to within $\pm 2$ mm. The optimization procedure was demonstrated by simulating the placement of six cryoprobes at equal radial distances in a medium around a central urethral warmer. Rewcastle et al. [29] presented a 3D finite difference analysis of ice ball formation for one, three, and five cryoprobes. Results of the 3.4 mm OD diameter, high pressure Argon-operated probes with 30 mm long active segments, yielded good conformity to experimental data in gelatin. The authors defined an "ablation ratio" as the percentage volume of a certain temperature, which is considered ablative or lethal to the tissue, relative to the total frozen volume. The value of this ratio for their three-probe configuration was calculated at 0.21. Wan et al. [30] presented a finite element model of multiprobe cryosurgery of the prostate, which was based on a variational principle. Model results were verified by comparison to an analytical solution of an idealized problem and to experimental data obtained for a single probe. The case of six-probe symmetric positioning in the prostate was simulated and 2D quadrant results were demonstrated. The authors define a "freezing exposure index," which relates the combined effect of freezing temperature and the duration the tissue is held at this temperature, as an index of damage caused to the tissue.

This article presents a retrospective work performed at the Technion, Israel Institute of Technology, over the last 3-odd decades. Results of analytical and numerical solutions are presented briefly as well as experimental data and their comparison to the derived results. Studies include the analysis of both the direct (Stefan) and the inverse-Stefan phase-change heat transfer problems in a tissue-simulating medium (gel) by the application of both surface and insertion cryoprobes. The effects of blood perfusion and metabolic heat generation rates on the advancement of the freezing front are presented. The simultaneous operation of needle cryoprobes in a number of different configurations and the effects of a thermally significant blood vessel in the vicinity of the cryoprobe are also presented and discussed.

## 2 Formulation of the Phase-Change Problem

With reference to Fig. 3, and without loss of generality, the phase-change problem is presented as one-dimensional in Cartesian coordinates. The phase changing medium (PCM) is assumed to behave such as a "real" substance in which phase transition occurs over a range of temperatures, unlike "pure" substances, which are characterized by a single phase-transition temperature. Three stages are identified in the evolution of the freezing front in the PCM.

Stage 1: The PCM is initially at a certain spatially distributed temperature and remains completely nonfrozen (liquid stage) through the end of this stage. At the beginning of the cryosurgical procedure and thereafter, the temperature of the cryoprobe (heat sink) is being lowered according to a certain preplanned time-dependent trajectory, removing sensible heat from the medium. This stage ends when the cryoprobe-medium interface reaches $T_2$, which defines the upper temperature boundary of the phase-change region.

Stage 2: During this stage an intermediate phase (liquid + solid) begins to form and expands adjacent to the cryoprobe. Solid ice crystals co-exist with the yet nonfrozen surrounding medium. Sensible heat with gradually increasing portions of latent heat are removed from the PCM in this region. Only sensible heat is pumped out of the rest of the medium, which remains nonfrozen. This stage ends when the cryoprobe-medium interface reaches $T_1$, which defines the lower temperature boundary of the phase-change region below which the entire PCM is in the frozen state (solid).

Stage 3: In this stage three regions exist in the medium: (a) a completely frozen region (solid) adjacent to the cryoprobe, $0 > x > s_1(t)$, the upper temperature boundary of which is $T_1$, (b) an intermediate region $s_1(t) > x > s_2(t)$ in which phase transition occurs, bounded by $T_1$ and $T_2$, and (c) a yet nonfrozen region (liquid), the lower boundary of which is located at $x = s_2(t)$ and is bounded by $T_2$ and $T_\infty$.

Similar to other problems involving heat transfer in biological entities, the heat balance of the phase-change problem is assumed to be governed by the bioheat equation [31], Eq. (1).

$$\rho c_p \frac{\partial T}{\partial t} = k \frac{\partial^2 T}{\partial x^2} + w_b c_b (T_a - T) + q_m \tag{1}$$

where $\rho$ and $\rho_b$ are the tissue and blood densities, $kg/m^3$, respectively, $c_p$ and $c_b$ are the tissue and blood specific heat capacities, $J/kg\,K$, respectively, $T(x,t)$ is the tissue temperature, °C, x is the length coordinate, m, t is the time, s, k is the tissue thermal conductivity, $W/m\,K$, $w_b$ is the capillary blood perfusion rate, $kg_b/m^3\,s$, $T_a$ is the blood temperature, °C, and $q_m$ is the metabolic heat generation rate, $W/m^3$.

Equation (1) may be assumed to apply in tissue regions, wherein blood flow and metabolic heat generation persist. In the present case this equation applies in region III. In region I, upon completion of freezing, both capillary blood perfusion ($w_b = 0$) and metabolic heat generation ($q_m = 0$) cease and Eq. (1) reduces to the simple transient heat equation. Blood continues to flow in major blood vessels, at least for awhile, even under these conditions, and would require the formulation of an additional, coupled heat balance equation, as shown subsequently. In the phase-transition region II a gradual cessation of both capillary blood perfusion ($w_b \to 0$) and metabolic heat generation ($q_m \to 0$) are assumed.

The following general boundary and continuity conditions may be specified for this problem:

$$\text{at } x = 0 \text{ (cryoprobe-medium interface):} \quad T_I(x,0) = T_{pr}(t) \tag{2}$$

$$\text{at } x = s_1(t) \text{ (beginning of phase-change region):} \quad T_I = T_{II} = T_1 \tag{3a}$$

$$\text{and} \quad k_I \frac{\partial T_I}{\partial x} = k_{II} \frac{\partial T_{II}}{\partial x} \tag{3b}$$

$$\text{at } x = s_2(t) \text{ (end of phase-change region)} \quad T_{II} = T_{III} = T_2 \tag{4a}$$

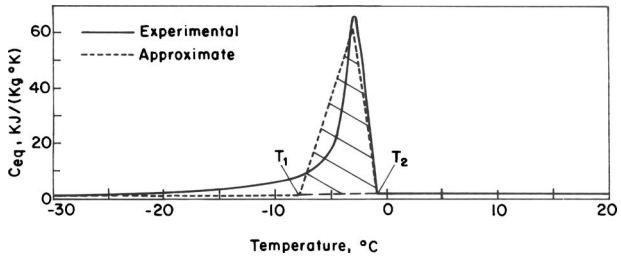



**Fig. 4 Equivalent specific heat for "Tylose" (24% methylcellulose/76% water by mass) as a function of temperature in the phase-change region [32]. (Copyright 1974 by Elsevier. Reproduced with permission of Elsevier via Copyright Clearance Center.)**

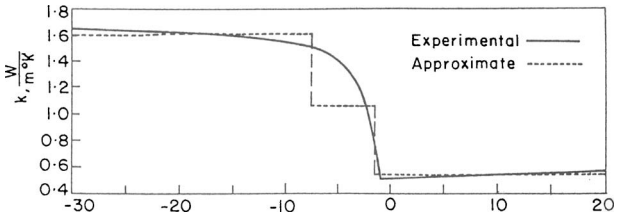



**Fig. 5 Thermal conductivity of "Tylose" as a function of temperature in the phase-change region [32]. (Copyright 1974 by Elsevier. Reproduced with permission of Elsevier via Copyright Clearance Center.)**

$$\text{and} \quad k_{II}\frac{\partial T_{II}}{\partial x} = k_{III}\frac{\partial T_{III}}{\partial x} \tag{4b}$$

In the intermediate region II, wherein the liquid and solid phases co-exist, accounting for the latent heat of freezing yields the following heat balance equation:

$$\rho c_{II}\frac{\partial T_{II}}{\partial t} = k_{II}\frac{\partial^2 T_{II}}{\partial x^2} + \rho L\frac{df_s}{dt} \tag{5}$$

where L is latent heat of freezing, J/kg, and $f_s$ is the solid fraction in this region defined by Eq. (6).

$$f_s = \frac{1}{L}\int_{T_1}^{T_{II}} (c - c_{eq})dT \tag{6}$$

where $c_{eq}$ is equivalent specific heat capacity, e.g., Fig. 4, which approximates the absorbed latent heat in this region, and

$$f_s = 0 \quad \text{for} \quad T_{II} = T_2 \tag{7a}$$

$$f_s = 1 \quad \text{for} \quad T_{II} = T_1 \tag{7b}$$

$$\text{at } x \to \infty \text{ (constant temperature):} \quad T(\infty, t) = T_\infty \tag{8}$$

At t=0 the initial temperature distribution of the tissue is given by

$$T(x,0) = T_{Init}(x) \tag{9}$$

## 3 Tissue Thermophysical and Physiological Properties

Tissue thermophysical properties assume different values in the nonfrozen and frozen regions. Table 1 lists plausible values for various tissue properties as applied in our studies. In the intermediate phase-transition region II, these properties vary appreciably with the temperature. Data in the literature present experimentally measured variations for latent heat and thermal conductivity of high water content materials, compare. Figs. 4 and 5 [32]. This behavior may be assumed as a first approximation of tissue behavior under similar conditions. Accounting for these changes in the analysis may be done by either following the measured changes precisely or approximately, as shown by the broken lines in Figs. 4 and 5.

Capillary blood perfusion and metabolic heat generation decrease from their normothermia values as the tissue temperature is lowered. As the upper temperature bound of the phase-transition range $T_1$ is reached, both these variables, may be assumed to decay to zero. Certain courses of change of these variables may be assumed, e.g., Ref. [33].

**3.1 Solutions of the Phase-Change Problem.** The phase-change problem may be characterized either as a regular Stefan or an inverse-Stefan problem. In the former the course of change of the temperature at the interface with the cryoprobe is assumed to be known. Accordingly, the temperature distribution in the freezing medium, the time-dependent location of the freezing front and the temperature rate of change at this front are to be determined. In the inverse-Stefan problem, on the other hand, a desired temperature rate of change is to be imposed at the freezing front. The purpose of the analysis is to determine the temperature changes (forcing function) that should be applied at the tissue-cryoprobe interface in order to satisfy this condition.

Due to the inherent nonlinearity of this class of problems, the analytical solutions are difficult to derive and are therefore rather limited. Numerical solution schemes have greatly expanded the computational capabilities and dominate this field of research. Below are presented a few of our analytical derivations and additional numerical treatments.

*3.1.1 Analytical Solutions of the Stefan Problem in Biological Media*. Rubinsky and Shitzer [17] presented an analytical solution in Cartesian coordinates of a Stefan-like problem in a biological tissue in contact with a cryoprobe. Equation (1) was solved subject to the assumption $T_2 \to T_1$, i.e., a single phase-transition temperature similar to pure substances. The effects of blood perfusion and of metabolic heat generation rates, on medium temperature variations, heat flux and frozen front location are shown in Figs. 6 and 7, respectively. Table 2 lists the parameters used in the computational cases. According to these figures, the heat flux in the medium increase as both $w_b$ and $q_m$ decrease. In general, varying

**Table 1 Ranges of thermophysical and physiological property values**

| Property | Units | Range of values | | References |
|---|---|---|---|---|
| | | Nonfrozen | Frozen | |
| Specific heat capacity of tissue | kJ/kg K | 3.52–4.1 | 1.84–1.9 | 17, 34, and 36 |
| Specific heat capacity of blood | kJ/kg K | 3.64 | (-) | [15] |
| Thermal conductivity | W/m K | 0.385–0.63 | 1.3–2.25 | 17, 34, 36, 38, and 39 |
| Density | $kg/m^3$ | 1000 | 1000 | [34] |
| Latent heat of freezing | kJ/kg | 233–330 | | 17, 34, and 38 |
| Phase change temperature | °C | −1 to −2 | −7 to −8 | 17, 34, 37, and 39 |
| Capillary blood perfusion rate | $kg/m^3$ s | 0–10 | (-) | [17] |
| Metabolic heat generation rate | $kW/m^3$ | 0–251 | (-) | [17] |

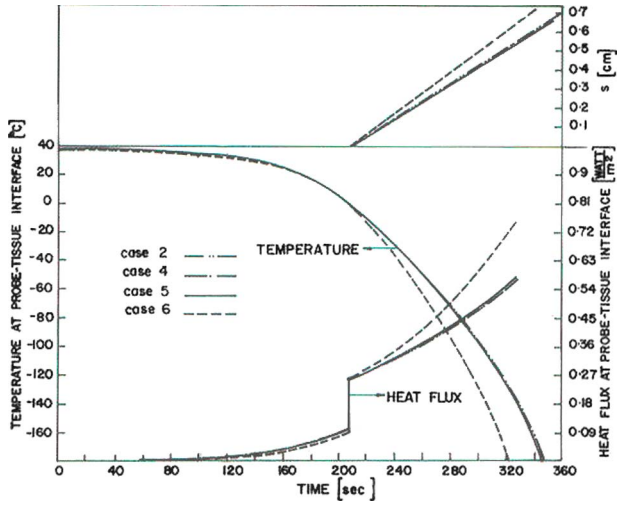



**Fig. 6 Cryoprobe temperature and heat flux variations and frozen front location versus blood perfusion rates [17]**

the metabolic rate in the yet nonfrozen tissue will have minor effects on all the plotted variables, except when it reduces to zero.

**Table 2 Combinations of physiological parameters used in the computations [17]**

| Case | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $q_m$, kW/m$^3$ | 251 | 251 | 251 | 167 | 84 | 0 |
| $w_b$, kg/m$^3$ s | 10 | 7.0 | 4.0 | 7.0 | 7.0 | 0 |
| $T_{arterial}$, °C | 37.7 | 38 | 38.7 | 37.6 | 37.3 | 37 |

Blood perfusion, on the other hand, does affect the course of change of the plotted variables as it changes, hindering, as is to be expected, the advancement of the frozen front.

Budman et al. provided a combined analytical and integral solution around an embedded general purpose cryoprobe [34]. The liquid nitrogen-operated probe was tested experimentally in pure water with a 2% gelatin agar. Results were compared with the analytical predictions and are shown in Fig. 8.

*3.1.2 Analytical Solutions of the Inverse-Stefan Problem in Biological Media*. Rabin and Shitzer [19] extended the analysis of the inverse-Stefan problem to a freezing biological medium. The tissue model, assumed initially at a uniform temperature, was depicted by a semi-infinite, thermally homogeneous, and isotropic medium. The problem was analyzed by the "enthalpy method" [35] with the latent heat of freezing expressed by the variable

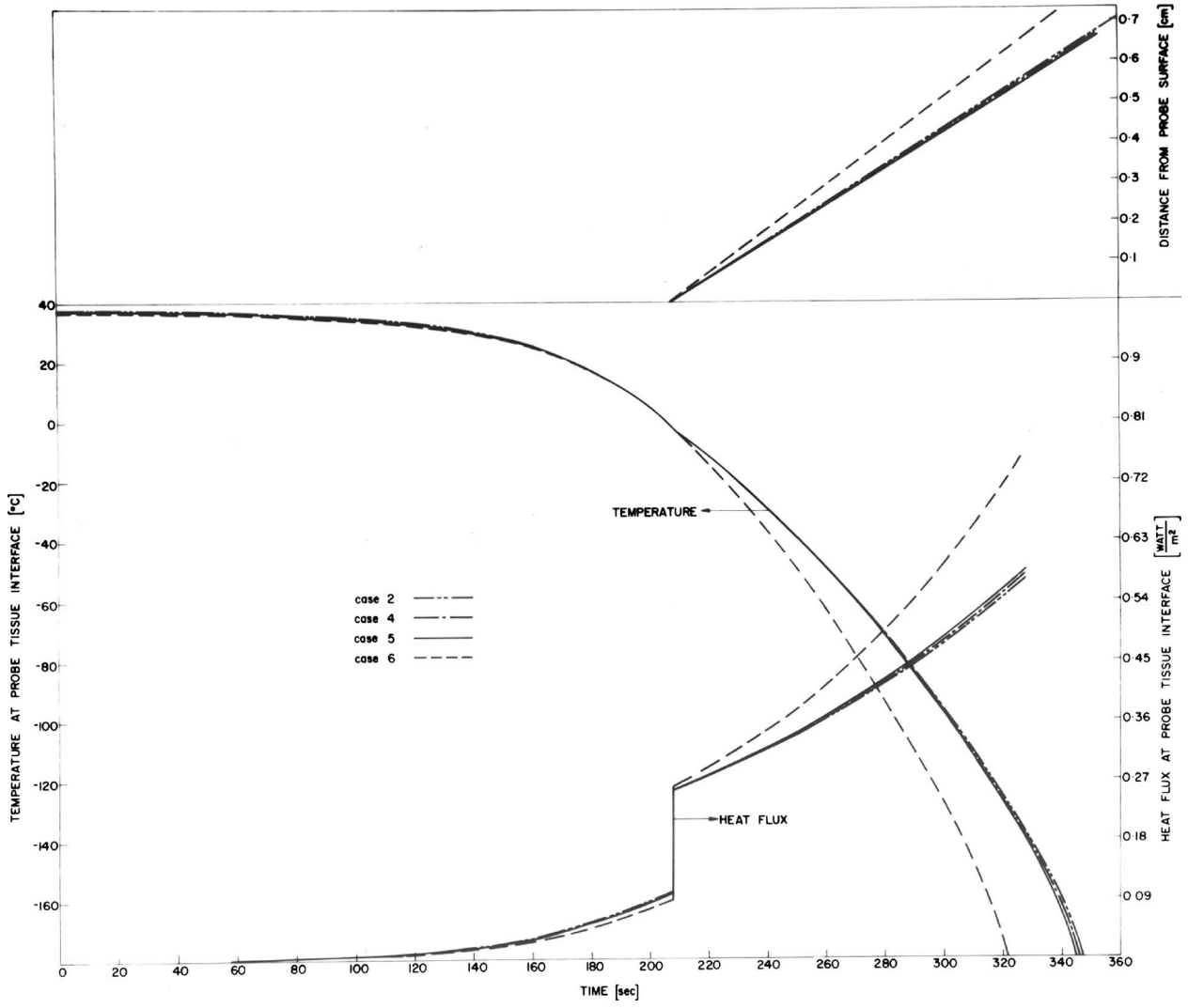



**Fig. 7 Cryoprobe temperature and heat flux variations and frozen front location versus metabolic heat rates [17]**

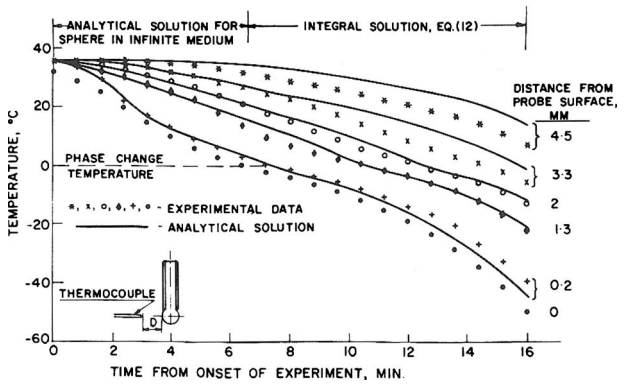



**Fig. 8 Comparison of experimental and analytical temperature distributions around a general purpose cryoprobe [34]**

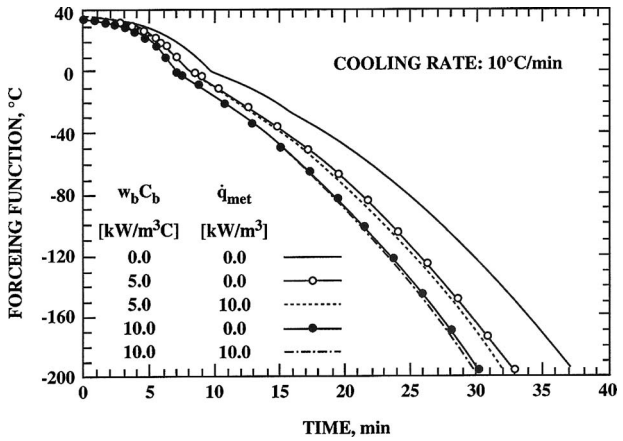



**Fig. 9 Forcing functions of the inverse-Stefan problem for various combinations of heat sources and a cooling rate of 10°C/min at the frozen front [19]**

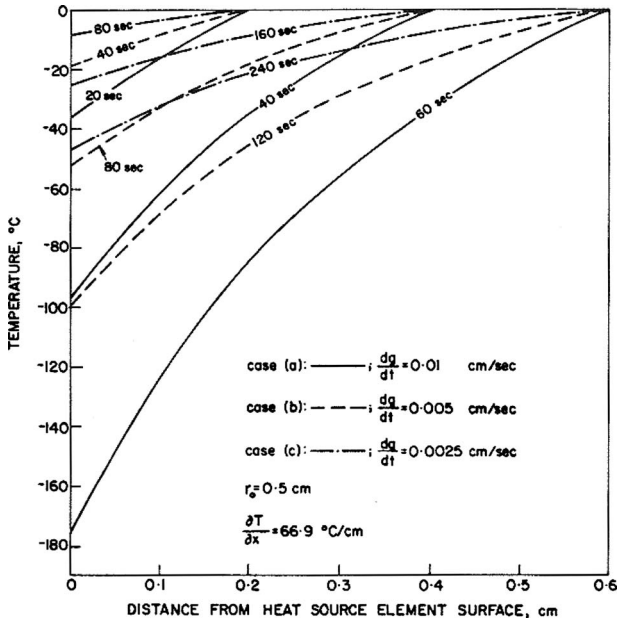



**Fig. 10 Temperature distributions in a PCM for a constant heat flux and various velocities of the moving boundary. Spherical geometry [36].**

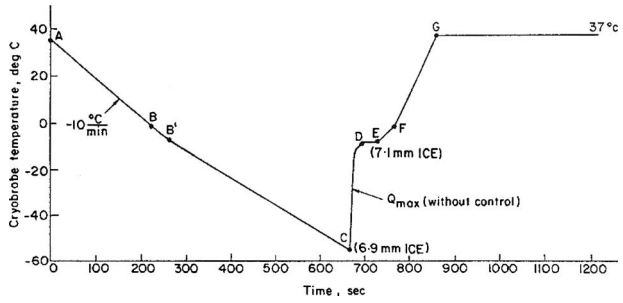



**Fig. 11 Calculated forcing function for the controlled freezing/thawing processes for a maximal cooling/warming rate of 10°C/min [37]**

volumetric specific heat. Results compared very well to the analytic solution presented by Rubinsky and Shitzer [36] by setting the phase-transition temperature range to $T_1-T_2=0.1°C$, closely approximating an ideal PCM. Assuming a cooling rate of 10°C/min, the calculated forcing function at the cryoprobe-medium interface is plotted in Fig. 9 as a function of capillary blood perfusion and metabolic heat generation rates. It is observed that blood perfusion has a much more profound effect than does metabolic heat generation. When comparing the behavior of the forcing function to the case of a nonbiological medium ($w_b \rightarrow 0$ and $q_m \rightarrow 0$), maximally assumed blood perfusion and metabolic heat generation rates reduce the duration of the cryotreatment by about 20%, the depth of freezing by about 15% and the phase-transition width by about 37%.

*3.1.3 Analytical Solutions of the Inverse-Stefan Problem in Nonbiological Media.* The inverse-Stefan problem in a nonbiological medium ($w_b \rightarrow 0$ and $q_m \rightarrow 0$) was solved analytically by Rubinsky and Shitzer [36] in Cartesian and spherical coordinates. Assuming a single phase-change temperature and a known temperature gradient at the moving freezing front, a series solution was derived for the temperature distribution in the medium. This solution facilitates the calculation of the temperature variations required at the cryoprobe-medium interface (forcing function) so as to achieve the desired thermal conditions at the moving front. Figure 10 presents temperature distributions in a PCM, depicted by spherical geometry, for a constant heat flux at and various velocities of the moving boundary. It is seen that the higher the velocity of the freezing front, the steeper are the temperature gradients in the medium. This demands lower temperatures to be

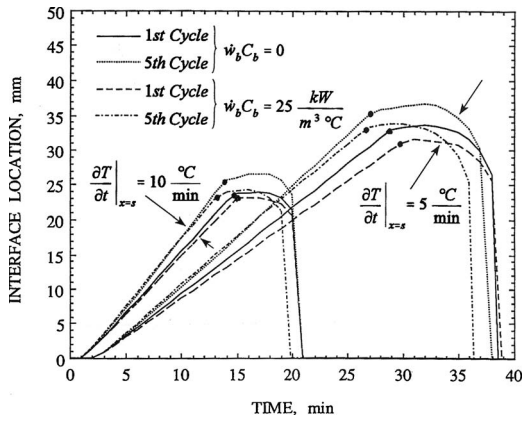



**Fig. 12 Freezing front locations, for the first and fifth repeated freezing/thawing cycles, for various blood perfusion rates and cooling/warming rates. The dots indicate end of cooling and initiation of warming in each cycle [40].**

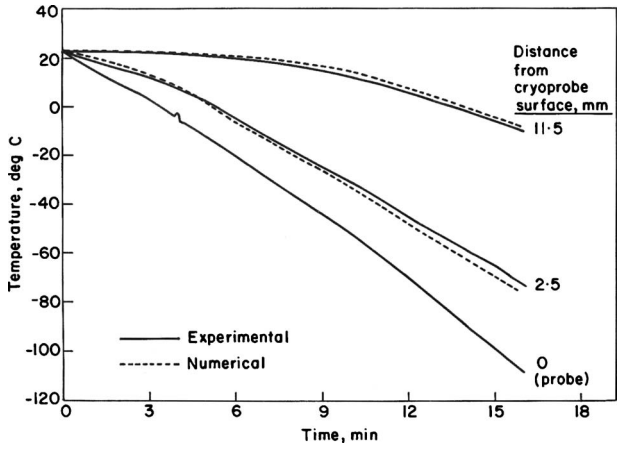



**Fig. 13 Comparison of experimental and computational finite element results [39]**

imposed at the cryoprobe-medium interface and may, therefore, limit the extent of freezing, should the required front velocity be excessively high.

An integral solution of a one-dimensional inverse-Stefan problem was presented by Budman et al. [37]. They assumed phase change to occur over a range of temperatures and analyzed both the controlled freezing and the subsequent thawing processes of the PCM. A constant cooling rate, $\partial T/\partial t = H$, was imposed at the lower phase-transition temperature $T_1$. Medium thermophysical properties were those shown in Table 1 and Figs. 4 and 5. Figure 11 shows the variations of the calculated forcing function during a single freeze-thaw cycle for uniform cooling/heating rates of $\pm 10°C/min$. Control characteristics of the cryosurgical process in a nonideal medium simulating a biological tissue were determined by Budman et al. [38]. Using a proportional-integral controller, the stability of the process was analyzed. It was shown that the desired cooling rate at the phase-change front deviated from the desired value only by about 1%.

Rabin and Shitzer [33] solved the inverse-Stefan problem of a biological tissue undergoing repeated freezing/thawing cycles. The locations of the freezing front following the first and fifth cycles are plotted in Fig. 12. Data are shown for two cooling/warming rates of $5°C/min$ and $10°C/min$ and for two cases of capillary blood perfusion: $w_bC_b = 25\ kW/m^3°C$ and for no blood perfusion. It is seen that the freezing front velocity is almost linearly dependent on time during the freezing stage, for all cases shown. The location of the freezing front seems to be only weakly dependent on blood perfusion during the first cycle. It does penetrate deeper during the initial portion of the thawing stage, in the case of no blood perfusion due to thermal inertia. In all cases studied, the maximal depth of freezing during the fifth cycle was about 7% larger than in the first cycle.

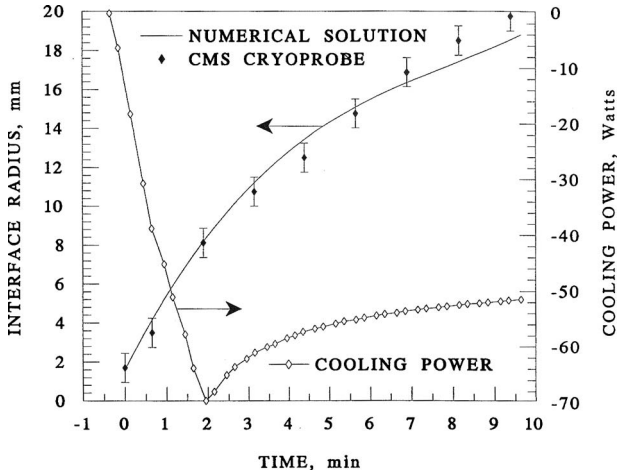



**Fig. 14 Comparison of measured [41] and calculated freezing front location around a 3.4 mm cylindrical Accuprobe. Required cooling power was estimated by the numerical code [40].**

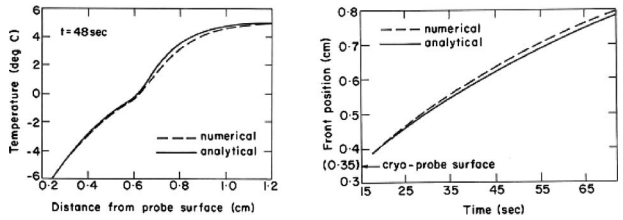



**Fig. 15 Comparison of numerical and analytical temperature distributions (left) and freezing front location (right) for a single, infinitely long cryoprobe embedded in a PCM [21]**

*3.1.4 Numerical Solutions of the Inverse-Stefan Problem in Biological Media*. Budman et al. [39] performed a numerical analysis of the freezing process in a nonideal medium. A new cryosurgical device was developed, which facilitated the achievement of a specified cooling rate at the phase-change front by accurately controlling probe temperature variations. An electrical heating element was wrapped around the shaft of the cryoprobe as a means to control its temperature by off-setting the excessive heat extraction of the cryogen. The system was tested in an aqueous-mashed potatoes solution. A closed loop control system was used with the cryoprobe surface temperature as the feedback variable. The energy balance equations were solved by a finite elements scheme and showed good conformity to measured results, Fig. 13. The tracking accuracy of the desired temperature trajectory of the cryoprobe was to within $\pm 4°C$. Supercooling of the PCM and liquid nitrogen boiling instabilities were identified as the main reasons for deviations from the set-point.

An axisymmetric finite difference solution of the inverse-Stefan problem in a PCM simulating a biological tissue was presented by Rabin and Shitzer [40]. The cylindrical probe, with a finite active length, was assumed to be embedded in an infinite PCM. The solution was verified both against an analytical derivation of a simpler case and experimental data [41], showing good conformity. A parametric study included the cooling power of the cryoprobe and the dimensions of the frozen region, Fig. 14. For a 3.4 mm o.d. probe and 10 min of operation, the required heat extraction power increased gradually for the first 2 min of operation, reaching a maximum of about 70 W. Thereafter, as the heat load on the probe decreased with the thickening of the frozen phase, which acted as a thermal insulator, the cooling power decreased slowly to about 50 W.

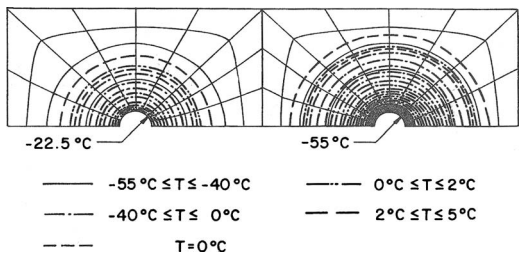



**Fig. 16 Temperature distributions after 60 s for nonsymmetrical activation of two adjacent cryoprobes at −55°C/1°C/s (right) and −22.5°C/0.5°C/s (left) [21]**

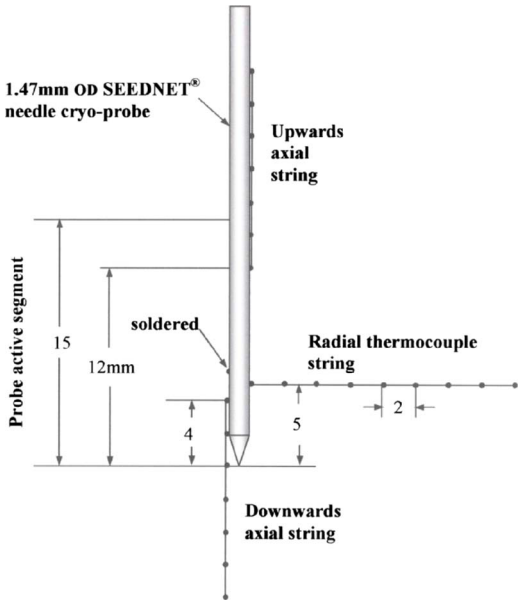



**Fig. 17 Placement of the thermocouple junctions adjacent to the insertion cryoprobe. Dimensions in millimeter [44].**

*3.1.5 Numerical Analyses of the Stefan Problem for Multiprobe Insertion in a PCM*. Finite element analysis of the temperature field around two adjacent cryoprobes was presented by Weill et al. [21]. The infinitely long cylindrical probes were assumed to be embedded in an infinite medium, initially at a constant temperature (5°C). For simplicity the medium was assumed to be homogeneous and isotropic and the effects of both metabolic heat generation and capillary blood perfusion were neglected. In order to test the solution of the developed numerical scheme, the calculated results of a single embedded probe were compared with available analytical results, Fig. 15. The performance of two embedded cryoprobes was calculated for both symmetrical and nonsymmetrical operations. In the symmetrical mode, the surface temperatures of both probes was dropped abruptly and maintained thereafter at this level. Calculated isotherms around each probe were circular during the first minute of operation and became skewed afterwards, "bulging" at the sections away from the centerline separating the probes [21]. The nonsymmetrical operation

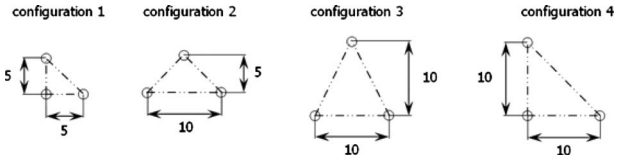



**Fig. 19 Insertion configurations for three cryoneedles [43]. Dimensions in mm. (Copyright 2007 by Elsevier. Reproduced with permission of Elsevier via Copyright Clearance Center.)**

involved simultaneously dropping of one probe to −55°C while the other was dropped to only −22.5°C. In addition, cooling rates applied were 1°C/min and 0.5°C/min, respectively. Temperature distributions around both probes are shown in Fig. 16. As is to be expected, the temperature field generated by the higher temperature probe lags considerably behind that produced by the lower temperature probe.

Magalov et al. [42–44] applied ANSYS 7.0 to analyze the thermal behavior of one, two, and three insertion needle cryoprobes operated by high pressure Argon gas. A single and two SEEDNET® cryoneedles, placed at varying distances apart, were tested in Agar gel the temperatures of which were monitored radially and axially, as shown in Fig. 17 [44]. Figure 18 compares measured and computed isothermal contours of 0°C, −20°C, and −40°C for a single probe after 6 min and 10 min, and 0°C and −40°C isotherms for two probes, 10 mm apart, after 10 min of operation. The synergistic effects of three cryoneedles were studied numerically. All probes were operated uniformly and were placed at the apexes of four triangles, as determined by the 5 × 5 mm placement template used in this industry [45], Fig. 19. Figure 20 shows the locations of the 0°C, −20°C, and −40°C isotherms for all placement configurations after 1 min and 10 min of operation. For all cases studied the volumes enclosed by the lethal isotherm (assumed here as −40°C), achieve most of their size in the first minutes of operation thus obviating the need for prolonged applications [43]. Furthermore, it was demonstrated that percentage volumes occupied by this lethal isotherm, relative to the total frozen volume, after 10 min of operation, will be only about 6% (single probe), 6–11% (two probes, varying distances apart) and 6–15% (three probes, different placement configurations) [43].

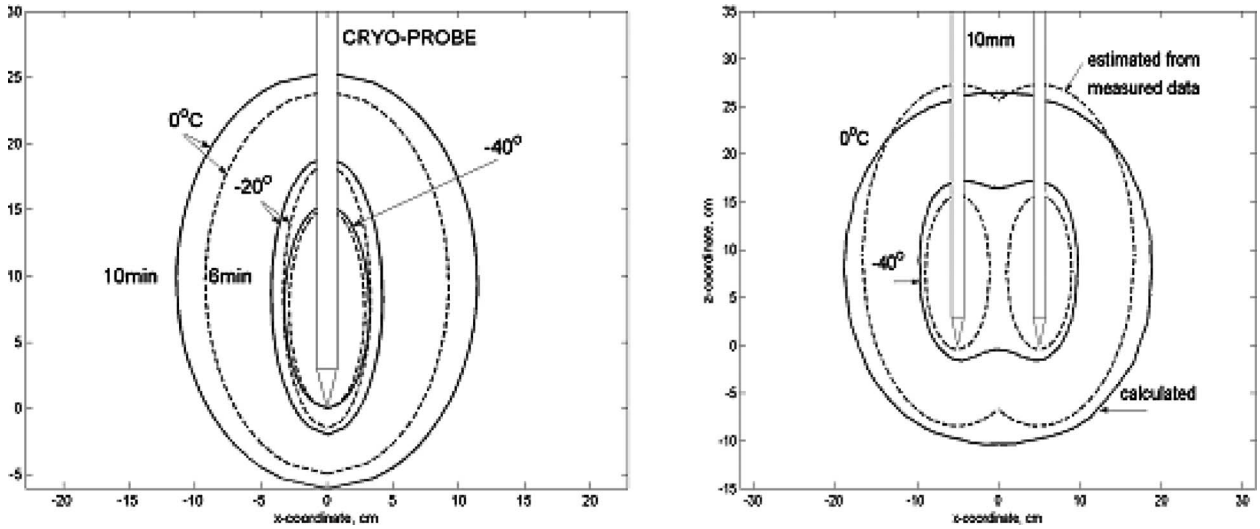



**Fig. 18 Measured and estimated isothermal contours for one (left) and two 10 mm apart (right) cryoprobes after 10 min of operation [44]**

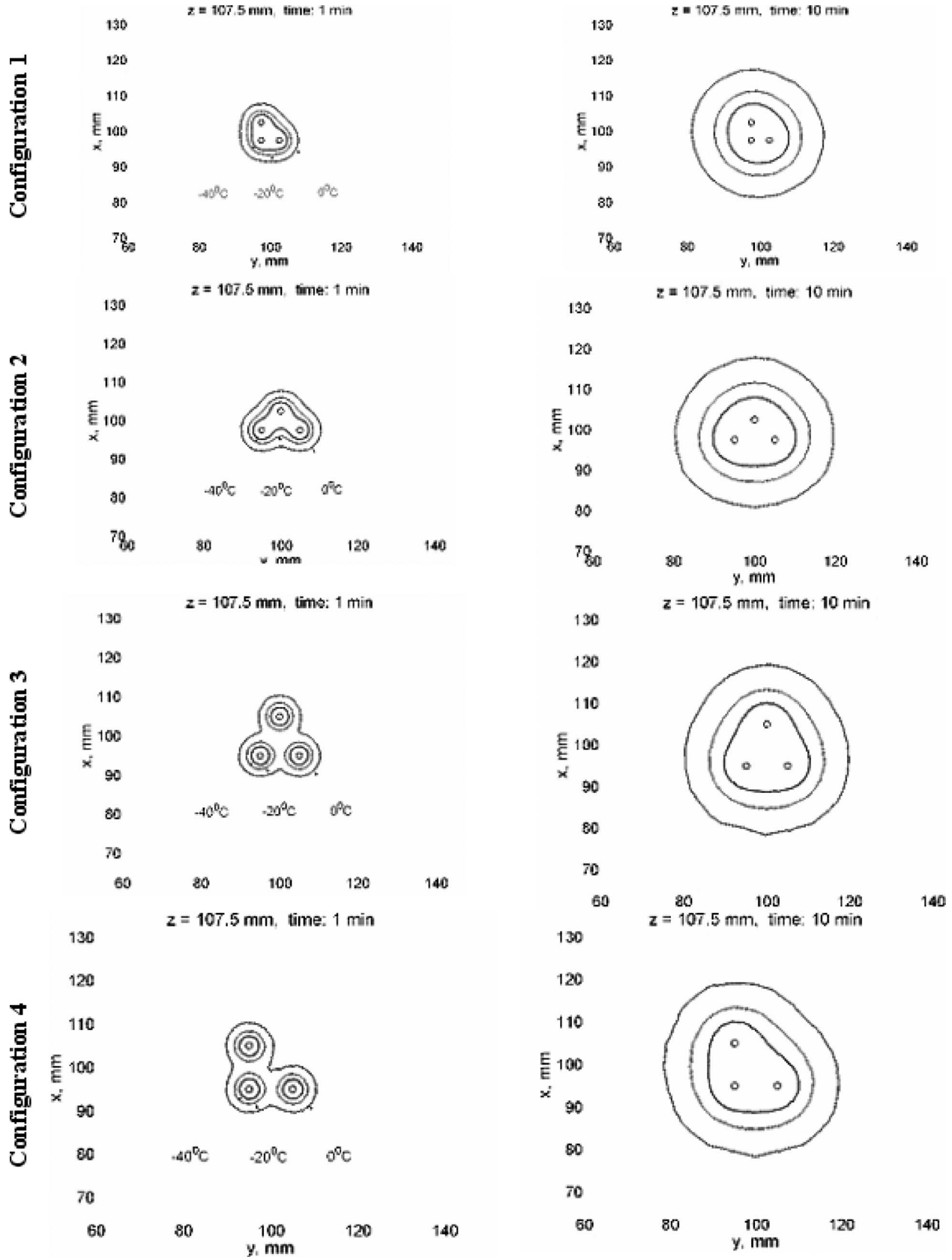



**Fig. 20 Top views of 0°C, −20°C, and −40°C isothermal contours for different placement configurations after 1 min and 10 min of operation of three cryoneedles [43]. (Copyright 2007 by Elsevier. Reproduced with permission of Elsevier via Copyright Clearance Center.)**

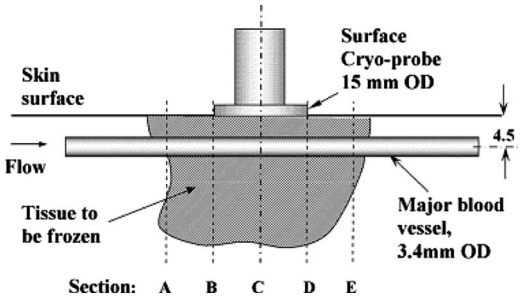



**Fig. 21 Schematic view of the tumor, embedded blood vessel, and surface cryoprobe [48]**

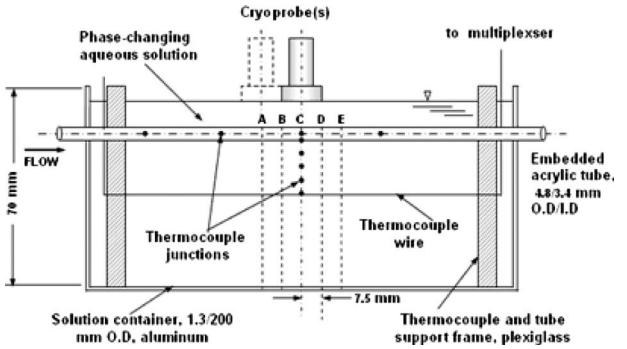



**Fig. 22 Cross-sectional view of the test section holding the embedded tube and thermocouples in the PCM [47]**

*3.1.6 Effects of an Embedded Thermally Significant Blood Vessel*. The effects of an embedded, thermally significant, blood vessel on the temperature field in a tissue-simulating PCM subjected to freezing by a surface cryoprobe were studied experimentally [46–48] and numerically [49]. The surface cryoprobe was operated by liquid nitrogen and was placed flush with the surface of the PCM, Fig. 21 [48]. A tube simulating a blood vessel was embedded in the PCM parallel and close to the surface, coincident with the cryoprobe centerline, Fig. 22 [47]. The tube was perfused by warm water at a constant temperature. Computer connected thermocouples were arranged on a single plane inside the PCM and were spaced sparingly, utilizing the symmetry of the setup to minimize thermal perturbations by the thermocouple wires. Each experiment involved five repetitions with the probe placed alternately at each one of the sections A–E, Fig. 22.

Variations of the isothermal surfaces in the PCM were interpolated from the recorded temperatures. A numerical solution of the phase-change problem under these conditions was developed using ANSYS 7.0. The numerical code was adapted to include the thermal interaction between the PCM and the water perfused em-

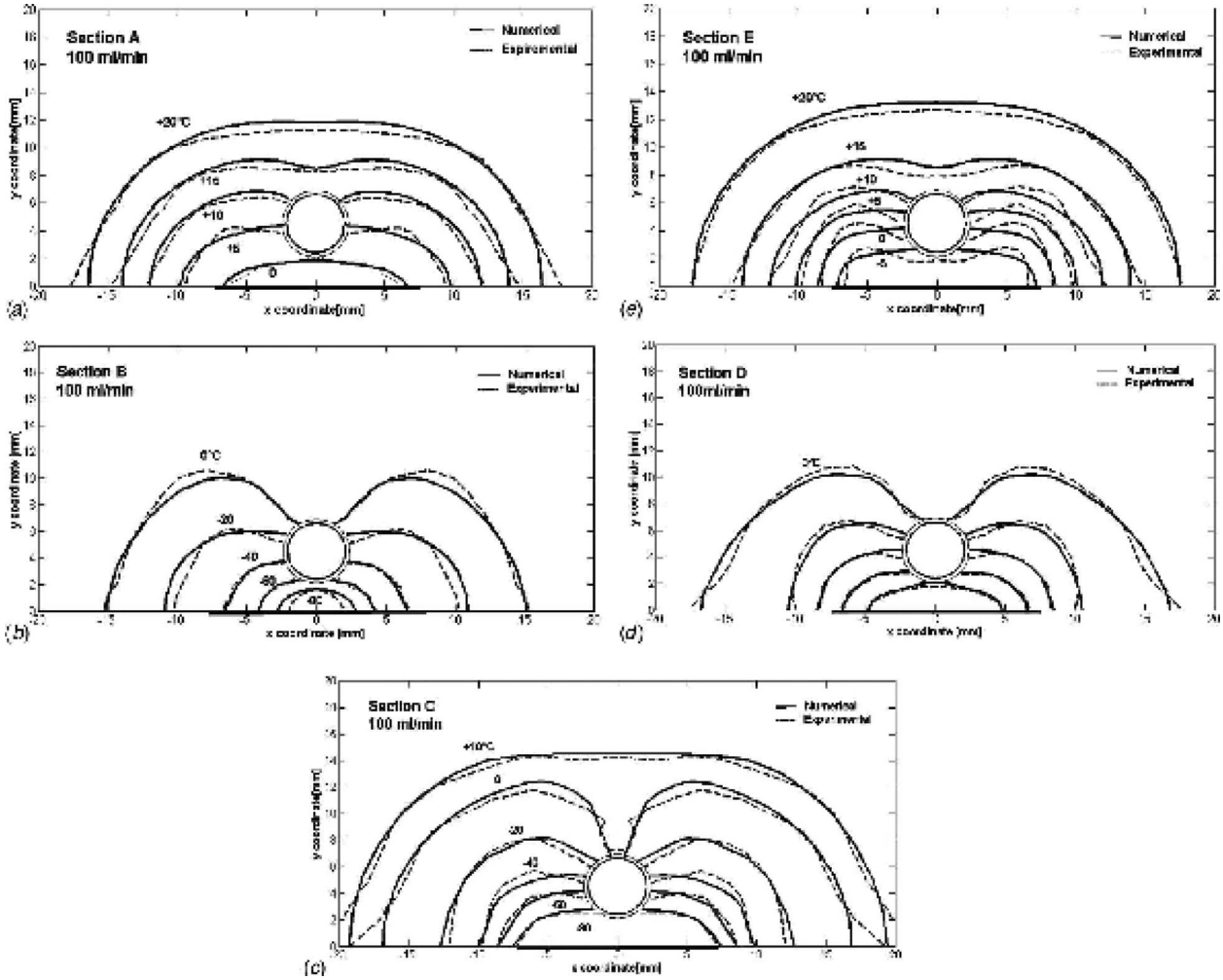



**Fig. 23 Measured and computed isotherms in sections A–E in the PCM after 20 min of cryoprobe operation at −8°C/min. Embedded tube flow rate: 100 ml/min [49].**

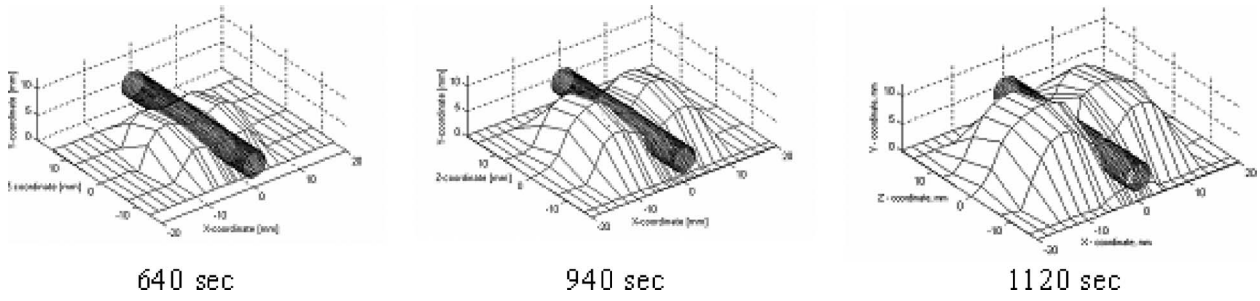



**Fig. 24 Three-dimensional snap shots of the progression of the 0°C isothermal surface in the vicinity of the embedded tube [47]. Cryoprobe operates from underneath.**

bedded tube [49]. Figure 23 compares the measured and computed isothermal contours in the PCM after 20 min of operation of the cryoprobe and 100 ml/min flow rate in the tube. Figure 24 depicts 3D approximations of the frozen front in the PCM at three instances, clearly showing the gradual engulfing of the embedded tube by the PCM [47].

**3.2 In Vivo Experiments.** Based on the previous work of Budman et al. [39], Rabin and Shitzer [50], and Rabin et al. [51] developed a new cryosurgical device, which was tested in vivo on the skeletal muscle of a rabbit's hind leg. Cryoprobe-medium interface temperature was controlled to follow the desired trajectory by activating an electrical heater wrapped around the cryoprobe shaft. Figure 25 shows the histological appearance of muscle cells 7 days post-cryotreatment. A gradual interface, of about 0.5 mm thickness, is observed between the cryo-injured tissue and the adjacent noninjured cells. The difference between the two regions is quite pronounced manifesting the potency of the cryotreatment.

## 4 Conclusion

The studies cited in this article summarize work done in our laboratory of Heat Transfer in Biological Systems for the last 3-odd decades. These studies address a variety of issues relating to the analysis of heat exchange driven by cryoprobes during the phase-change process in biological and nonbiological media alike. The studies included the development of both analytical solutions, wherever applicable, and numerical solution schemes, which, recently, rely for the most part on available commercial codes. In certain cases experimental studies, conducted both in vivo and in vitro, complemented the analyses.

For cryosurgery to become the treatment-of-choice, much more work will be required, among other things to cover the following issues:

1. a clear cut understanding and definition of the tissue-specific conditions that are required to ensure the complete destruction of a tissue undergoing a controlled cryosurgical process. The current conflict and resulting ambiguity between specifying either a tissue-specific lethal temperature or, conversely, a desired cooling rate at the freezing front, hinder progress in this inherently complex subject.
2. more comprehensive analyses, which address, in addition to the freezing stage, the more involved controlled thawing stage and its consequences on the overall outcome of the complete freeze/thaw cycle(s).
3. improved technical means to control the temperature variations of the cryoprobe to achieve the desired thermal conditions required for tissue destruction.
4. improvement in the pretreatment design process to include optimal placement schemes of multiprobes and their separate and specific operation.
5. understanding the effects of thermally significant blood vessels, and other related thermal perturbations, which are situated adjacent to, or even within, the tissue to be destroyed.

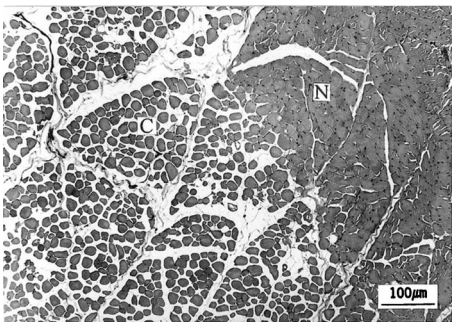



**Fig. 25 Low power magnification of the interface between normal (N) and cryodamaged fibers (C) 7 days post-cryotreatment [51]. Copyright 1996 by Elsevier. Reproduced with permission of Elsevier via Copyright Clearance Center.**

## Acknowledgment

The studies presented in this retrospective article are the results of concerted efforts by many individuals, most of whom were graduate students under the author's guidance at the Technion. They are listed here in chronological order: Boris Rubinsky, Hector Budman, Anne Weill, Yoed Rabin, Yehoshua Chayut, Loay Massalha, Zaur Magalov, Genady Beckerman, and Noga Rybko. Colleagues of mine have also contributed in their specialized fields: Stefano Del Giudice, Yehoshua Dayan, Pinhas Bar-Yoseph, Raymond Coleman, Rosalie Ber, Daniel Mordohovich, and David Degani. Dr. Nir Berzak conducted experiments with the SEEDNET® cryoprobe. The cryosurgical equipment in our laboratory was purchased by a generous grant from the Caesaria Foundation, Israel. The James H. Belfer Chair in Mechanical Engineering provided financial support throughout the years.

I am forever indebted and grateful to all those individuals and organizations for their participation, support, and contributions.

## References

[1] Onik, G., 1996, "Cryosurgery," Crit. Rev. Oncol. Hematol., **23**, pp. 1–24.
[2] Rubinsky, B., and Onik, G., 1991, "Cryosurgery: Advances in the Application of Low Temperatures to Medicine," Int. J. Refrig., **14**, pp. 190–199.
[3] Shepherd, J., and Dawber, R. P. R., 1982, "The Historical and Scientific Basis of Cryosurgery," Clin. Exp. Dermatol., **7**, pp. 321–328.
[4] Gage, A. A., and Huben, R. P., 2000, "Cryosurgical Ablation of the Prostate," Semin. Urol. Oncol., **5**, pp. 11–19.
[5] Zouboulis, C. C., 1998, "Cryosurgery in Dermatology," Eur. J. Dermatol., **8**, pp. 466–474.
[6] http://en.wikipedia.org/wiki/Cryopreservation
[7] Mazur, P., 1968, "Physical-Chemical Factors Underlying Cell Injury in Cryosurgical Freezing," *Cryosurgery*, R. Rand, A. Rinfret, and H. von Leden, eds., Thomas, Springfield, IL, pp. 32–51.
[8] Saliken, J. C., Donnelly, B. J., and Rewcastle, J. C., 2002, "The Evolution and State of Modern Technology for Prostate Cryosurgery," Urology, **60**, pp. 26–33.
[9] Zisman, A., Pantuck, A. J., Cohen, J. K., and Belldegrun, A. S., 2001, "Prostate Cryoablation Using Direct Transperineal Placement of Ultrathin Probes Through a 17-Gauge Brachytherapy Template—Technique and Preliminary

Results," Urology, **58**, pp. 988–993.
[10] Orpwood, R. D., 1981, "Biophysical and Engineering Aspects of Cryosurgery," Phys. Med. Biol., **26**, pp. 555–575.
[11] Crister, J. K., and Morbaaten, L. E., 2000, "Cryopreservation of Murine Spermatozoa," Institute for Laboratory Animal Research, J. Cryobiology of Embryos, Germ Cells & Ovaries, **41**(1), pp. 197–206.
[12] Gage, A., and Baust, J., 1998, "Mechanisms of Tissue Injury in Cryosurgery," Cryobiology, **37**, pp. 171–186.
[13] Hoffmann, N. E., and Bischof, J. C., 2002, "The Cryobiology of Cryosurgical Injury," Urology, **60**, pp. 40–49.
[14] Stefan, J., 1891, "Uber die Theorie des Eisbildung, insbesondere uber die Eisbildung im Polarmeere," Ann. Phys. Chem., **42**(2), pp. 269–286.
[15] Neumann, F., 1912, "Lectures Given in the 1860's," *Die Partielen Differentialglechungen der Mathematischen Physik*, 5th ed., B Riemann and H. Weber, eds., Vieweg und Sohn, Braunschweig, Vol. 2, pp. 117–121.
[16] Ozisik, M. N., and Uzzell, J. C., 1979, "Exact Solution for Freezing in Cylindrical Symmetry With Extended Freezing Temperature Range," ASME J. Heat Transfer, **101**, pp. 331–334.
[17] Rubinsky, B., and Shitzer, A., 1976, "Analysis of a Stefan-Like Problem in a Biological Tissue Around a Cryosurgical Probe," ASME J. Heat Transfer, **38**, pp. 514–519.
[18] Trezek, G. J., 1985, "Thermal Analysis for Cryosurgery," *Heat Transfer in Medicine and Biology: Analysis and Applications*, A. Shitzer and R. C. Eberhart, eds., Plenum, New York, pp. 239–259.
[19] Rabin, Y., and Shitzer, A., 1995, "Exact Solution for the Inverse Stefan Problem in Non-Ideal Biological Tissues," ASME J. Heat Transfer, **117**, pp. 425–431.
[20] Shamsundar, N., and Sparrow, E. M., 1975, "Analysis of Multidimensional Conduction Phase Change Problem via the Enthalpy Method," ASME J. Heat Transfer, **97**, pp. 333–340.
[21] Weill, A., Shitzer, A., and Bar-Yoseph, P., 1993, "Finite Elements Analysis of the Temperature Field Around Two Adjacent Cryo-Probes," ASME J. Biomech. Eng., **115**, pp. 374–379.
[22] Dalhuijsen, A. J., and Segal, A., 1986, "Comparison of Finite Element Techniques for Solidification Problems," Int. J. Numer. Methods Eng., **23**, pp. 1807–1829.
[23] Keanini, R. G., and Rubinsky, B., 1992, "Optimization of Multiprobe Cryosurgery," ASME J. Heat Transfer, **114**, pp. 796–801.
[24] Rabin, Y., and Stahovich, T. F., 2003, "Cryoheater as a Means of Cryosurgery Control," Phys. Med. Biol., **48**, pp. 619–632.
[25] Rabin, Y., Lung, D. C., and Stahovich, T. F., 2004, "Computerized Planning of Cryosurgery Using Cryoprobes and Cryocatheters," Technol. Cancer Res. Treat., **3**(3), pp. 229–243.
[26] Rewcastle, J. C., Sandison, G. A., Hahn, L. J., Saliken, J. C., McKinnon, J. G., and Donnelley, B. J., 1998, "A Model for the Time-Dependent Thermal Distribution Within an Iceball Surrounding a Cryoprobe," Phys. Med. Biol., **43**, pp. 3519–3534.
[27] Jankun, M., Kelly, T. J., Zaim, A., Young, K., Keck, R. W., Selman, S. H., and Jankun, J., 1999, "Computer Model for Cryosurgery of the Prostate," Comput. Aided Surg., **4**, pp. 193–199.
[28] Baissalov, R., Sandison, G. A., Donnelley, B. J., Saliken, J. C., McKinnon, J. G., Muldrew, K., and Rewcastle, J. C., 2000, "A Semi-Empirical Planning Model for Optimization of Multiprobe Cryosurgery," Phys. Med. Biol., **45**, pp. 1085–1098.
[29] Rewcastle, J. C., Sandison, G. A., Muldrew, K., Saliken, J. C., and Donnelley, B. J., 2001, "A Model for the Time Dependent Three-Dimensional Thermal Distribution Within Iceballs Surrounding Multiple Cryoprobes," Med. Phys., **28**(6), pp. 1125–1137.
[30] Wan, R., Liu, Z., Muldrew, K., and Rewcastle, J. C., 2003, "A Finite Element Model for Ice Ball Evolution in a Multi-Probe Cryosurgery," Comput. Methods Biomech. Biomed. Eng., **6**(3), pp. 197–208.
[31] Pennes, H. H., 1948, "Analysis of Tissue and Arterial Blood Temperature in the Resting Human Forearm," J. Appl. Physiol., **1**, pp. 93–122.
[32] Bonacina, C., Comini, G., Fasano, A., and Primicerio, M., 1974, "On the Estimation of Thermophysical Properties in Nonlinear Heat Conduction Problems," Int. J. Heat Mass Transfer, **17**, pp. 861–867.
[33] Rabin, Y., and Shitzer, A., 1997, "Combined Solution of the Inverse Stefan Problem for Successive Freezing/Thawing in Nonideal Biological Tissues," ASME J. Biomech. Eng., **119**(2), pp. 146–152.
[34] Budman, H., Shitzer, A., and Del Giudice, S., 1986, "Investigation of Temperature Fields Around Embedded Cryoprobes," ASME J. Biomech. Eng., **108**(1), pp. 42–48.
[35] Goodman, T. R., 1958, "The Heat Balance Integral and Its Application to Problems Involving a Change of Phase," Trans. ASME, **80**, pp. 335–342.
[36] Rubinsky, B., and Shitzer, A., 1978, "Analytic Solutions to the Heat Equation Involving a Moving Boundary With Application to the Change of Phase Problem (The Inverse Stefan Problem)," ASME J. Heat Transfer, **100**, pp. 300–303.
[37] Budman, H., Shitzer, A., and Dayan, Y., 1995, "Analysis of the Inverse-Stefan Problem of Freezing and Thawing of a Binary Solution During Cryosurgical Processes," ASME J. Biomech. Eng., **117**(2), pp. 193–202.
[38] Budman, H., Dayan, J., and Shitzer, A., 1991, "Control of the Cryosurgical Process in Non-Ideal Materials," IEEE Trans. Biomed. Eng., **38**(11), pp. 1141–1153.
[39] Budman, H., Dayan, J., and Shitzer, A., 1991, "Controlled Freezing of Non-Ideal Solutions With Applications to Cryosurgical Processes," ASME J. Biomech. Eng., **113**(4), pp. 430–437.
[40] Rabin, Y., and Shitzer, A., 1998, "Numerical Solution of the Multidimensional Freezing Problem During Cryosurgery," ASME J. Biomech. Eng., **120**, pp. 32–37.
[41] Chang, Z., Finkelstein, J. J., and Baust, J., 1994, "Development of a High-Performance Multiprobe Cryosurgical Device," Biomed. Instrum. Technol., **42**, pp. 383–390.
[42] Magalov, Z., Shitzer, A., and Degani, D., 2006, "Simulation of Cryo-Ablation of the Prostate by 1, 2 and 3 Embedded Cryo-Surgical Probes," Proceedings of the 13th International Heat Transfer Conference, Vol. BHT-04, pp. 1–12.
[43] Magalov, Z., Shitzer, A., and Degani, D., 2007, "Isothermal Volume Contours Generated in a Freezing Gel by Embedded Cryo-Needles With Applications to Cryo-Surgery," Cryobiology, **55**(2), pp. 127–137.
[44] Magalov, Z., Shitzer, A. and Degani, D., 2008, "Experimental and Numerical Study of 1, 2 and 3 Embedded Needle Cryo-Probes Simultaneously Operated by High Pressure Argon Gas," ASME J. Heat Transfer, **130**(3), p. 032301.
[45] Galil-Medical, Ltd., Yokneam, Israel, 2003, Experimental Data, private communication.
[46] Chayut, Y., and Shitzer, A., 1996, "Simulating the Effects of a Large Blood Vessel on the Temperature Field around a Surface Cryoprobe," ASME Advances in Heat and Mass Transfer in Biotechnology, L. J. Hayes and S. Clegg, eds., ASME, New York, pp. 21–22.
[47] Massalha, L., and Shitzer, A., 2004, "Freezing by a Flat, Circular Surface Cryoprobe of a Tissue Phantom With an Embedded Cylindrical Heat Source Simulating a Blood Vessel," ASME J. Biomech. Eng., **126**(6), pp. 736–744.
[48] Rybko, N., Shitzer, A., and Degani, D., 2009, "Experimental simulation of a Thermally Significant Blood Vessel in a Tissue Subjected to Cryo-Surgery," Proceedings of the Seventh World Conference on Experimental Heat Transfer, Fluid Flow and Thermodynamics, Krakow, Poland, pp. 193–200.
[49] Beckerman, G., Shitzer, A., and Degani, D., 2009, "Numerical Model of the Effects of a Thermally-Significant Blood Vessel on Solidification by a Circular Surface Cryosurgical Probe Compared to Experimental Data," ASME J. Heat Transfer, **131**(5), p. 051101.
[50] Rabin, Y., and Shitzer, A., 1996, "A New Cryosurgical Device for Controlled Freezing. Part 1: Setup and Validation Tests," Cryobiology, **33**, pp. 82–92.
[51] Rabin, Y., Coleman, R., Mordohovich, D., Ber, R., and Shitzer, A., 1996, "A New Cryosurgical Device for Controlled Freezing. Part 2: In Vivo Experiments on Skeletal Muscle of Rabbit Hindlimbs," Cryobiology, **33**, pp. 93–105.

# Analysis of Heat Transfer in Consecutive Variable Cross-Sectional Domains: Applications in Biological Media and Thermal Management

**Shadi Mahjoob**

**Kambiz Vafai**[1]
e-mail: vafai@engr.ucr.edu

Department of Mechanical Engineering,
University of California,
Riverside, CA 92521

*Temperature prescription and control is important within biological media and in bioheat transport applications such as in hyperthermia cancer treatment in which the unhealthy tissue/organ is subject to an imposed heat flux. Thermal transport investigation and optimization is also important in designing heat management devices and small-scale porous-filled-channels utilized in electronic and biomedical applications. In this work, biological media or the stated heat management devices with a nonuniform geometry are modeled analytically as a combination of convergent, uniform and/or divergent configurations. The biological media is represented as blood saturated porous tissue matrix while incorporating cells and interstices. Two primary models, namely, adiabatic and constant temperature boundary conditions, are employed and the local thermal nonequilibrium and an imposed heat flux are fully accounted for in the presented analytical expressions. Fluid and solid temperature distributions and Nusselt number correlations are derived analytically for variable cross-sectional domain represented by convergent, divergent, and uniform or any combination thereof of these geometries while also incorporating internal heat generation in fluid and/or solid. Our results indicate that the geometrical variations have a substantial impact on the temperature field within the domain and on the surface with an imposed heat flux. It is illustrated that, the temperature distribution within a region of interest can be controlled by a proper design of the multisectional domain as well as proper selection of the porous matrix. These comprehensive analytical solutions are presented for the first time, to the best of the authors' knowledge in literature.*
[DOI: 10.1115/1.4002303]

*Keywords: bioheat, nonuniform geometry, electronic and biomedical applications, variable area domain, porous media*

## 1 Introduction

Analyses of temperature distribution and heat transfer through small-scale channels subject to an imposed heat flux are key issues in a variety of applications such as biomedical devices [1–5], cooling of electronic devices and heat pipe technology [6–13]. The channels can be filled with porous inserts, which have been shown to be highly effective in heat transfer enhancement and thermal management [14–19] by providing an extensive surface area between solid and fluid phases.

Temperature control and prescription is crucial in bioheat transport applications such as in hyperthermia cancer treatment, where the unhealthy tissue/organ is subject to an imposed heat flux during the course of the treatment. The biological media can be modeled as blood saturated porous tissue matrix consisting of interstices and cells. Mahjoob and Vafai [20,21] investigated transport through biological media for uniform single and multilayer tissue matrix structures. They had investigated several important parameters affecting transport through the biological media such as volume fraction of the vascular space, organ/tissue depth, imposed hyperthermia heat flux, metabolic heat generation, and body core temperature. Aspects related to modeling in porous media incorporated in this work are given in Refs. [22–30]. Nield and Kuznetsov [31] performed an investigation related to forced convection in a channel filled with a porous medium with counterflow.

The geometrical configuration of the channels is also of importance when designing these heat management devices. The temperature distribution on the surface of the devices connected to the channels can be controlled and optimized utilizing proper uniform and nonuniform (convergent or divergent) geometries or a combination thereof. Mahjoob and Vafai [8] developed an analytical solution for a convergent single channel while incorporating the local thermal nonequilibrium condition. The effects of several pertinent parameters on the temperature distribution and heat transfer coefficient such as inclination angle, interfacial fluid-solid heat exchange, ratio of fluid to solid effective thermal conductivities, and imposed heat flux were investigated [8].

In the present work, the effect of a geometrically nonuniform domain is studied by modeling it analytically as a sequential series of convergent, divergent and/or uniform configurations. For the first time, to the best of the authors' knowledge,fluid and solid temperature distributions and the heat transfer coefficient are derived for variable cross-sectional media such as convergent, divergent, and uniform channels filled with a porous medium or any combination thereof of these geometries while incorporating the possibility of existence of internal heat generation in fluid or solid or both phases as well as the local thermal nonequilibrium condition.

[1]Corresponding author.

Manuscript received May 29, 2010; final manuscript received July 31, 2010; published online September 27, 2010. Assoc. Editor: Peter Vadasz.

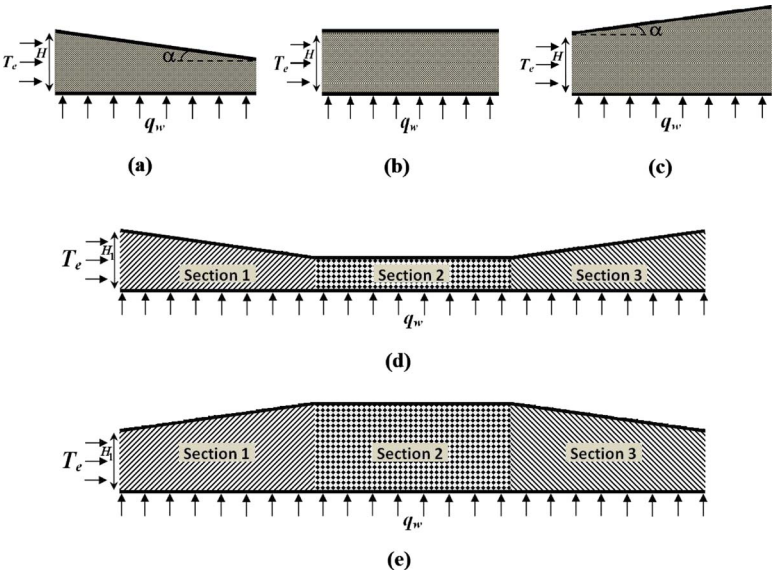



**Fig. 1 Schematic diagram of a channel filled with a porous medium subject to a constant heat flux on one side and either an adiabatic or a constant temperature wall on the other side (at the upper wall): (*a*) convergent channel, (*b*) uniform channel, (*c*) divergent channel, (*d*) variable cross-sectional domain made of convergent-uniform-divergent sections, and (e) variable cross-sectional domain made of divergent-uniform-convergent sections**

## 2 Modeling and Formulation

**2.1 Problem Description.** In this work, variable cross-sectional media are represented by channels with nonuniform (convergent or divergent) and uniform geometries filled with a porous medium subject to forced convection and a uniform heat flux. These domains can represent either a biological media or small-scale heat management devices, where the temperature distribution and heat transfer aspects are analyzed within them. Special consideration is given at the intersection of two consecutive sections to ensure proper accounting of the interface boundary conditions. The schematic diagrams of the type of geometries that were analyzed in this work are presented in Figs. 1(*a*)–1(*c*). The analytical solutions that are obtained in this work can be used for any combination of these geometries such as those shown in Figs. 1(*d*) and 1(*e*). In principal, various variable area media such as biological tissue can be modeled with convergent and/or divergent modular sections such as those shown in Figs. 1(*d*) and 1(*e*).

One side of each modular component is subject to a constant heat flux and the other side is subject to either an adiabatic or a constant temperature condition while accounting for internal heat generation from the solid or fluid phases (or both) within the variable cross-sectional domain. It should be noted that the results presented for the adiabatic boundary are also applicable for a symmetric domain in which the heat flux is imposed from both sides of the module. In Fig. 1, $H$ refers to the thickness at the entrance of each section of the module while that of the first section of a multicomponent module is referred to by $H_1$. The angle between the inclined wall and the longitudinal direction is $\alpha$. Parameter $x_0$ is the longitudinal coordinate of the starting point of each section in a multicomponent channel or that of a single-section channel. Flow is considered to be thermally and hydraulically fully developed within an isotropic and homogeneous porous medium. Radiation and natural convection are neglected while assuming constant properties.

**2.2 Governing Equations.** The governing energy equations for fluid and solid phases, incorporating local thermal nonequilibrium condition and fluid and solid internal heat generations are as follows [20–30].

Fluid phase

$$k_{f,\text{eff}}\nabla_y^2\langle T_f\rangle^f + h_{sf}a_{sf}(\langle T_s\rangle^s - \langle T_f\rangle^f) + \varepsilon\dot{q}_f = \varepsilon\rho c_p\langle u\rangle^f\frac{\partial\langle T_f\rangle^f}{\partial x} \quad (1)$$

Solid phase

$$k_{s,\text{eff}}\nabla_y^2\langle T_s\rangle^s - h_{sf}a_{sf}(\langle T_s\rangle^s - \langle T_f\rangle^f) + (1-\varepsilon)\dot{q}_s = 0 \quad (2)$$

where

$$k_{f,\text{eff}} = \varepsilon k_f + k_{f,\text{dis}} \quad (3)$$

$$k_{s,\text{eff}} = (1-\varepsilon)k_s \quad (4)$$

and parameters $k_{f,\text{eff}}$, $k_{s,\text{eff}}$, $k_f$, $k_s$, $k_{f,\text{dis}}$, $\varepsilon$, $\rho$, and $c_p$ are the fluid and solid effective and regular thermal conductivities, fluid dispersion thermal conductivity, porosity, fluid density, and specific heat capacity, respectively. $\langle T_f\rangle^f$, $\langle T_s\rangle^s$, $\langle u\rangle^f$, $\dot{q}_f$, and $\dot{q}_s$ represent the intrinsic phase average fluid and solid temperatures, intrinsic fluid phase average velocity, and internal heat generation within the fluid and solid phases, respectively. The fluid-solid interfacial heat transfer coefficient is represented by $h_{sf}$ and the specific surface area by $a_{sf}$.

**2.3 Normalization.** The governing Eqs. (1) and (2) are normalized by using the following nondimensional variables.

$$\eta = \frac{y}{H \pm (x-x_0)\tan\alpha} \quad \theta = \frac{k_{s,\text{eff}}(\langle T\rangle - T_w)}{q_w H} \quad \kappa = \frac{k_{f,\text{eff}}}{k_{s,\text{eff}}}$$

$$\psi = \frac{(H \pm (x-x_0)\tan\alpha)^2}{H^2} \quad \Phi = \frac{(1-\varepsilon)H\dot{q}}{q_w} \quad \text{Bi} = \frac{h_{sf}a_{sf}H^2}{k_{s,\text{eff}}} \quad (5)$$

Note that the sign $\pm$ in this work refers to whether the modular section is a divergent channel (+) or a convergent one (−). For a multisectional domain, $H$ is the thickness of each section's entrance. In Eq. (5), parameters $\eta$ and $\theta$ represent the nondimensional transverse coordinate and nondimensional temperature, re-

spectively. Parameter $\kappa$ represents the ratio of fluid to solid effective thermal conductivities and the parameter $\psi$ is the upper wall shape factor. The parameter $\Phi$ is the nondimensional internal heat generation. The parameter Bi is an equivalent Biot number indicating the ratio of the conduction resistance within the solid matrix to the thermal resistance corresponding to the internal convective heat exchange between the solid matrix and the fluid phase [29]. For brevity, the intrinsic volume averaging sign ($\langle\ \rangle$) is dropped in the following sections.

**2.4 Normalized Governing Equations and Boundary Conditions.** Utilizing Eq. (5) and after some modifications, the governing Eqs. (1) and (2) are presented as

$$\kappa\frac{\partial^4\theta_f}{\partial\eta^4}-(1+\kappa)\mathrm{Bi}\psi\left(\frac{\partial^2\theta_f}{\partial\eta^2}\right)=-(1+\Gamma)\mathrm{Bi}\psi^{3/2} \tag{6}$$

$$\kappa\frac{\partial^4\theta_s}{\partial\eta^4}-(1+\kappa)\mathrm{Bi}\psi\left(\frac{\partial^2\theta_s}{\partial\eta^2}\right)=-(1+\Gamma)\mathrm{Bi}\psi^{3/2} \tag{7}$$

where $\Gamma$ is zero for the adiabatic boundary condition (model I). For the constant temperature boundary condition (model II), $\Gamma$ is represented by

$$\Gamma=\frac{1}{q_w}\left(k_{f,\mathrm{eff}}\frac{\partial T_f}{\partial y}+k_{\mathrm{s,eff}}\frac{\partial T_s}{\partial y}\right)\Bigg|_{y=H\pm(x-x_0)\tan(\alpha)} \tag{8}$$

The imposed constant heat flux ($q_w$) is distributed between the fluid and solid phases based on the physical values of their effective thermal conductivities and temperature gradients [8,20,21,28–30]. Boundary conditions are normalized using Eq. (5) and additional boundary conditions are obtained by evaluating the second or third order derivatives of $\theta_f$ and $\theta_s$ at the boundaries. This results in

$$\theta_f|_{\eta=0}=\theta_s|_{\eta=0}=0 \tag{9}$$

$$\left.\frac{\partial^2\theta_f}{\partial\eta^2}\right|_{\eta=0}=\frac{(1+\Gamma)\psi^{1/2}+\psi\Phi_s}{\kappa} \tag{10}$$

$$\left.\frac{\partial^2\theta_s}{\partial\eta^2}\right|_{\eta=0}=-\psi\Phi_s \tag{11}$$

*2.4.1 Model I: Adiabatic Boundary Condition.*

$$\left.\frac{\partial\theta_f}{\partial\eta}\right|_{\eta=1}=\left.\frac{\partial\theta_s}{\partial\eta}\right|_{\eta=1}=0 \tag{12}$$

$$\left.\frac{\partial^3\theta_f}{\partial\eta^3}\right|_{\eta=1}=\left.\frac{\partial^3\theta_s}{\partial\eta^3}\right|_{\eta=1}=0 \tag{13}$$

*2.4.2 Model II: Constant Temperature Boundary Condition.*

$$\theta_f|_{\eta=1}=\theta_s|_{\eta=1}=\theta_c \tag{14}$$

$$\left.\frac{\partial^2\theta_f}{\partial\eta^2}\right|_{\eta=1}=\frac{(1+\Gamma)\psi^{1/2}+\psi\Phi_s}{\kappa} \tag{15}$$

$$\left.\frac{\partial^2\theta_s}{\partial\eta^2}\right|_{\eta=1}=-\psi\Phi_s \tag{16}$$

where

$$\theta_c=\frac{k_{\mathrm{s,eff}}(T_c-T_w)}{q_wH} \tag{17}$$

**2.5 Fluid, Solid, and Wall Temperature Distributions.** The fluid and solid phase temperature distributions are derived by solving the presented governing equations and utilizing the given Neumann and Dirichlet boundary conditions. After a lengthy analysis, the temperature distributions for the fluid and the solid phases are derived for the adiabatic (model I) and constant temperature (model II) boundary conditions, which are applied at the upper wall of the channel.

*2.5.1 Model I: Adiabatic Boundary Condition.*

$$\theta_f=\frac{\psi^{1/2}}{1+\kappa}\left(\eta\left(\frac{\eta}{2}-1\right)-\frac{1+(1+\kappa)\psi^{1/2}\Phi_s}{(1+\kappa)\mathrm{Bi}\psi}\left\{1-\frac{e^{\lambda\eta}+e^{\lambda(2-\eta)}}{1+e^{2\lambda}}\right\}\right) \tag{18}$$

$$\theta_s=\frac{\psi^{1/2}}{1+\kappa}\left(\eta\left(\frac{\eta}{2}-1\right)+\frac{\kappa(1+(1+\kappa)\psi^{1/2}\Phi_s)}{(1+\kappa)\mathrm{Bi}\psi}\left\{1-\frac{e^{\lambda\eta}+e^{\lambda(2-\eta)}}{1+e^{2\lambda}}\right\}\right) \tag{19}$$

where

$$\lambda=\sqrt{\mathrm{Bi}\psi(1+\kappa)/\kappa} \tag{20}$$

As such, the temperature difference between the solid and the fluid phases and the wall surface temperature, which is subject to an imposed heat flux, can be written as

$$\Delta\theta=\theta_s-\theta_f=\frac{1+(1+\kappa)\psi^{1/2}\Phi_s}{(1+\kappa)\psi^{1/2}\mathrm{Bi}}\left(1-\frac{e^{\lambda\eta}+e^{\lambda(2-\eta)}}{1+e^{2\lambda}}\right) \tag{21}$$

$$\begin{aligned}T_w=&\frac{q_w}{k_{f,\mathrm{eff}}+k_{\mathrm{s,eff}}}\Bigg(\frac{H\pm(x-x_0)\tan(\alpha)}{3}\\&+\frac{k_{\mathrm{s,eff}}q_w+(1-\varepsilon)\dot{q}_s(k_{f,\mathrm{eff}}+k_{\mathrm{s,eff}})[H\pm(x-x_0)\tan(\alpha)]}{h_{sf}a_{sf}(1+\kappa)[H\pm(x-x_0)\tan(\alpha)]q_w}\\&\times\left\{1-\frac{1}{\lambda}\frac{e^{2\lambda}-1}{e^{2\lambda}+1}\right\}\Bigg)\\&+\frac{q_w+(\varepsilon\dot{q}_f+(1-\varepsilon)\dot{q}_s)[H\pm0.5(x-x_0)\tan(\alpha)]}{\rho c_pu_iH}(x-x_0)+T_i\end{aligned} \tag{22}$$

*2.5.2 Model II: Constant Temperature Boundary Condition.* The governing equations (6) and (7) are solved while utilizing the boundary conditions for a constant temperature upper wall. This results in the fluid and solid temperature profiles, which can be rewritten using Eq. (8) so that the presented solution will be independent of the term $\Gamma$. As such the term $\Gamma$ and the fluid and solid temperature distributions are derived as

$$\Gamma=\frac{2(1+\kappa)\theta_c}{\psi^{1/2}}+1 \tag{23}$$

$$\begin{aligned}\theta_f=\frac{1}{1+\kappa}\Bigg(&\eta[(\psi^{1/2}+(1+\kappa)\theta_c)\eta-\psi^{1/2}]\\&-\frac{[2\psi^{1/2}+(1+\kappa)(2\theta_c+\psi\Phi_s)]}{(1+\kappa)\mathrm{Bi}\psi}\left\{1-\frac{e^{\lambda\eta}+e^{\lambda(1-\eta)}}{1+e^{\lambda}}\right\}\Bigg)\end{aligned} \tag{24}$$

$$\begin{aligned}\theta_s=\frac{1}{1+\kappa}\Bigg(&\eta[(\psi^{1/2}+(1+\kappa)\theta_c)\eta-\psi^{1/2}]\\&+\frac{\kappa[2\psi^{1/2}+(1+\kappa)(2\theta_c+\psi\Phi_s)]}{(1+\kappa)\mathrm{Bi}\psi}\left\{1-\frac{e^{\lambda\eta}+e^{\lambda(1-\eta)}}{1+e^{\lambda}}\right\}\Bigg)\end{aligned} \tag{25}$$

$$\Delta\theta = \theta_s - \theta_f = \frac{2\psi^{1/2} + (1+\kappa)(2\theta_c + \psi\Phi_s)}{(1+\kappa)\mathrm{Bi}\psi}\left\{1 - \frac{e^{\lambda\eta} + e^{\lambda(1-\eta)}}{1+e^{\lambda}}\right\} \tag{26}$$

where $\lambda$ is defined in Eq. (20). The wall temperature, which is subject to an imposed heat flux, is derived to be

$$T_w = \Omega + (T_{w,x=x_0} - \Omega)\exp\left(\frac{-(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})(x - x_0)}{\rho c_p u_i H^2\left\{\frac{1}{\mathrm{Bi}(1+\kappa)}\left(1 + \frac{2(1-e^{\lambda_0})}{\lambda_0(1+e^{\lambda_0})}\right) + \frac{1}{3}\right\}}\right) \quad \text{for} \quad \alpha = 0 \tag{27}$$

$$T_w = \exp\left(-\int \Omega_1 d(x - x_0)\right)\left\{\int\left[\exp\left(\int \Omega_1 d(x - x_0)\right)\right] \times \Omega_2 d(x - x_0) + \Omega_3\right\} \quad \text{for} \quad \alpha \neq 0 \tag{28}$$

where

$$\Omega = \frac{H}{k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}}}\{q_w + 0.5H[\varepsilon\dot{q}_f + (1-\varepsilon)\dot{q}_s]\} + T_c \tag{29}$$

$$\Omega_1 = \frac{\frac{(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})}{\rho c_p u_i H[H \pm (x - x_0)\tan(\alpha)]} - \frac{\pm 2\tan(\alpha)[H \pm (x - x_0)\tan(\alpha)]}{\mathrm{Bi}\psi^2(1+\kappa)H^2}\left(1 + \frac{2(1-e^{\lambda})}{\lambda(1+e^{\lambda})}\right) - \frac{\pm 2\lambda_0\tan(\alpha)}{\kappa H\lambda^4(1+e^{\lambda})^2}(2\lambda e^{\lambda} - e^{2\lambda} + 1)}{\frac{1}{\mathrm{Bi}\psi(1+\kappa)}\left(1 + \frac{2(1-e^{\lambda})}{\lambda(1+e^{\lambda})}\right) + \frac{1}{3}} \tag{30}$$

$$\Omega_2 = \Omega_1 T_c + \frac{1}{\frac{1}{\mathrm{Bi}\psi(1+\kappa)}\left(1 + \frac{2(1-e^{\lambda})}{\lambda(1+e^{\lambda})}\right) + \frac{1}{3}}\left\{\begin{array}{l} \frac{q_w + 0.5[\varepsilon\dot{q}_f + (1-\varepsilon)\dot{q}_s] \times [H \pm (x - x_0)\tan(\alpha)]}{\rho c_p u_i H} - \\ \frac{\pm\tan(\alpha) q_w H^2}{k_{s,\mathrm{eff}}(1+\kappa)^2\mathrm{Bi}[H \pm (x - x_0)\tan(\alpha)]^2}\left(1 + \frac{2(1-e^{\lambda})}{\lambda(1+e^{\lambda})}\right) - \\ \frac{\pm q_w\lambda_0\tan(\alpha)}{(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})\mathrm{Bi}\lambda^2(1+e^{\lambda})^2} \times \left(\frac{2H}{(1+\kappa)[H \pm (x - x_0)\tan(\alpha)]} + \Phi_s\right) \times \\ (2\lambda e^{\lambda} - e^{2\lambda} + 1) + \frac{\pm q_w\tan(\alpha)}{12(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})} \end{array}\right\} \tag{31}$$

$\Omega_3$ can be evaluated utilizing the following boundary condition:

$$T_w|_{x=x_0} = T_{w,x=x_0} \tag{32}$$

For a single module domain or for the first component of a multicomponent domain, $T_{w,x=x_0}$ can be evaluated from the following equation. This value also indicates the wall temperature at the channel's entrance ($T_{w,e}$).

$$T_{w,x=x_0} = T_{w,e} = \frac{\frac{\frac{2k_{s,\mathrm{eff}}}{q_w H}T_c + \frac{2}{1+\kappa} + \Phi_s}{\mathrm{Bi}(1+\kappa)}\left(1 + \frac{2(1-e^{\lambda_0})}{\lambda_0(1+e^{\lambda_0})}\right) + \frac{k_{s,\mathrm{eff}}}{q_w H}\left(T_i - \frac{T_c}{3}\right) + \frac{1}{6(1+\kappa)}}{\frac{2k_{s,\mathrm{eff}}}{3q_w H}\left\{\frac{3}{\mathrm{Bi}(1+\kappa)}\left(1 + \frac{2(1-e^{\lambda_0})}{\lambda_0(1+e^{\lambda_0})}\right) + 1\right\}} \tag{33}$$

and

$$\lambda_0 = \sqrt{\mathrm{Bi}(1+\kappa)/\kappa} \tag{34}$$

For other sections of a multicomponent domain, $T_{w,x=x_0}$ is extracted from the immediate last component of the module under consideration. This process is described in more detail later on.

**2.6 Heat Transfer Correlations.** The wall heat transfer coefficient is obtained from

$$h_w = \frac{q_w}{T_w - T_{f,m}} \tag{35}$$

The Nusselt number at the channel wall subject to a constant heat flux can be represented as

$$\mathrm{Nu}_w = \frac{h_w D_h}{k_{f,\mathrm{eff}}} = \frac{-2}{\kappa\theta_{f,m}} \tag{36}$$

*2.6.1 Model I: Adiabatic Boundary Condition.*

$$\mathrm{Nu}_w = \frac{2(1+\kappa)}{\kappa\psi^{1/2}\left(\frac{1}{3} + \frac{1 + (1+\kappa)\psi^{1/2}\Phi_s}{(1+\kappa)\mathrm{Bi}\psi}\left\{1 - \frac{1}{\lambda}\frac{e^{2\lambda} - 1}{e^{2\lambda} + 1}\right\}\right)} \tag{37}$$

*2.6.2 Model II: Constant Temperature Boundary Condition.*

$$\mathrm{Nu}_w = \frac{2}{\kappa\left(\frac{[2\psi^{1/2} + (1+\kappa)(2\theta_c + \psi\Phi_s)]}{(1+\kappa)^2 \mathrm{Bi}\psi}\left\{1 + \frac{2(1-e^{\lambda})}{\lambda(1+e^{\lambda})}\right\} + \frac{\psi^{1/2}}{6(1+\kappa)} - \frac{\theta_c}{3}\right)} \tag{38}$$

**2.7 Simplified Solution Based on Local Thermal Equilibrium Assumption.** The energy equation for the one equation model (utilizing the assumption of local thermal equilibrium between fluid and solid phases) can be obtained by adding Eqs. (1) and (2) with the following boundary conditions:

$$\theta|_{\eta=0} = 0 \tag{39}$$

and

$$\left.\frac{\partial\theta}{\partial\eta}\right|_{\eta=1} = 0 \quad \text{for model I: adiabatic boundary condition} \tag{40}$$

$$\theta|_{\eta=1} = \theta_c$$

for model II: constant temperature boundary condition (41)

where $\theta_c$ is defined in Eq. (17).

Based on the simplified governing equations and boundary conditions, the following relationships are obtained for the wall temperature, the temperature distribution and the Nusselt number for models I and II.

*2.7.1 Model I: Adiabatic Boundary Condition.*

$$\theta = \frac{\psi^{1/2}}{1+\kappa}\left(\frac{\eta}{2} - 1\right)\eta \tag{42}$$

$$T_w = \frac{q_w[H \pm (x - x_0)\tan(\alpha)]}{3(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})} + \frac{q_w + (\varepsilon\dot{q}_f + (1-\varepsilon)\dot{q}_s)[H \pm 0.5(x - x_0)\tan(\alpha)]}{\rho c_p u_i H}(x - x_0) + T_i \tag{43}$$

$$\mathrm{Nu}_{w,TE} = \frac{6(1+\kappa)}{\kappa\sqrt{\psi}} \tag{44}$$

*2.7.2 Model II: Constant Wall Temperature Boundary Condition.*

$$\theta = \frac{\eta}{1+\kappa}[(\psi^{1/2} + (1+\kappa)\theta_c)\eta - \psi^{1/2}] \tag{45}$$

$$T_w = \Omega + (T_{w,x=x_0} - \Omega)\exp\left(\frac{-3(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})}{\rho c_p u_i H^2}(x - x_0)\right) \quad \text{for} \quad \alpha = 0 \tag{46}$$

$$T_w = T_c + \left(\frac{1}{4(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})} \pm \frac{3}{\rho c_p u_i H \tan(\alpha)}\right)\frac{q_w[H \pm (x - x_0)\tan(\alpha)]}{\left(1 \pm \frac{3(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})}{\rho c_p u_i H \tan(\alpha)}\right)} + \frac{3[\varepsilon\dot{q}_f + (1-\varepsilon)\dot{q}_s] \times (H \pm (x - x_0)\tan(\alpha))^2}{2\rho c_p u_i H \tan(\alpha)\left(\frac{3(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})}{\rho c_p u_i H \tan(\alpha)} \pm 2\right)} + \Omega'\left(\frac{H \pm (x - x_0)\tan(\alpha)}{H}\right)^{\pm(-3(k_{f,\mathrm{eff}}+k_{s,\mathrm{eff}})/(\rho c_p u_i H \tan(\alpha)))}$$

for $\alpha \neq 0$ (47)

$$\mathrm{Nu}_{w,TE} = \frac{-12(1+\kappa)}{\kappa(2(1+\kappa)\theta_c - \psi^{1/2})} \tag{48}$$

where $\Omega$ is defined in Eq. (29) and

$$\Omega' = \frac{3}{2}\left\{\pm\frac{-3q_w H}{2[\rho c_p u_i H \tan(\alpha) \pm 3(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})]} \pm \frac{-H^2[\varepsilon\dot{q}_f + (1-\varepsilon)\dot{q}_s]}{2\rho c_p u_i H \tan(\alpha) \pm 3(k_{f,\mathrm{eff}} + k_{s,\mathrm{eff}})} + T_i + T_c\right\} \tag{49}$$

For a single-section module or for the first component of a multicomponent domain, $T_{w,x=x_0}$ can be evaluated from the following equation:

$$T_{w,x=x_0} = T_{w,e} = \frac{1}{2}\left(3T_i - T_c + \frac{q_w H}{2k_{s,\mathrm{eff}}(1+\kappa)}\right) \tag{50}$$

For the subsequent sections of a multicomponent domain, $T_{w,x=x_0}$ is extracted from the immediate last component of the domain under consideration. This process is described in more detail below.

**2.8 Special Considerations for Multicomponent Domains.** The presented equations for the temperature distributions and the Nusselt number are valid for every section of a multicomponent domain. However, special attention should be given in designating the boundary condition at the entrance of each section after the very first module. The boundary condition, at the interface between two sections should link the data between the two consecutive modules. For instance, the entrance velocity should be adjusted based on the continuity equation. As such, it can be evaluated based on the entrance heights of the consecutive sections as well as the flow velocity at the entrance of the module under consideration. Note that interface height is an important parameter for the value of Biot number for the module under consideration.

Other important parameters are the wall temperature and mean flow temperature at the entrance of a given section. The wall temperature at the entrance ($T_{w,x=x_0}$) can be taken to be the same as the wall temperature at the end of the section immediately preceding it. Then, the mean flow temperature at the entrance of the section can be evaluated from the expressions given below for models I and II.

*2.8.1 Model I: Adiabatic Boundary Condition.*

$$T_i = \frac{q_w H}{k_{s,\mathrm{eff}}}\theta_{f,m}|_{x=x_0} + T_{w,x=x_0} \tag{51}$$

For local thermal nonequilibrium condition (general form)

$$T_i = \frac{-q_w H}{k_{s,\text{eff}}(1+\kappa)}\left(\frac{1}{3} + \frac{1+(1+\kappa)\Phi_s}{(1+\kappa)\text{Bi}}\left\{1 - \frac{1}{\lambda_0}\frac{e^{2\lambda_0}-1}{e^{2\lambda_0}+1}\right\}\right) + T_{w,x=x_0} \tag{52}$$

and for local thermal equilibrium assumption $T_i$ can be obtained from

$$T_i = \frac{-q_w H}{3k_{s,\text{eff}}(1+\kappa)} + T_{w,x=x_0} \tag{53}$$

*2.8.2 Model II: Constant Temperature Boundary Condition*. For local thermal nonequilibrium assumption (general form)

$$T_i = \frac{2}{3}\left\{\frac{3}{\text{Bi}(1+\kappa)}\left(1 + \frac{2(1-e^{\lambda_0})}{\lambda_0(1+e^{\lambda_0})}\right) + 1\right\}T_{w,x=x_0} - \frac{q_w H}{6k_{s,\text{eff}}(1+\kappa)} + \frac{T_c}{3} - \frac{2k_{s,\text{eff}}(1+\kappa)T_c + q_w H[2+(1+\kappa)\Phi_s]}{\text{Bi}k_{s,\text{eff}}(1+\kappa)^2}\left(1 + \frac{2(1-e^{\lambda_0})}{\lambda_0(1+e^{\lambda_0})}\right) \tag{54}$$

For local thermal equilibrium assumption $T_i$ can be obtained from

$$T_i = \frac{1}{3}\left(2T_{w,x=x_0} + T_c - \frac{q_w H}{2k_{s,\text{eff}}(1+\kappa)}\right) \tag{55}$$

In Eqs. (51)–(55), $T_i$ should be evaluated based on the values from the downstream component and $T_{w,x=x_0}$ is evaluated based on the wall temperature value at the exit of the upstream component of a multicomponent domain. The channel may compose of several components with different attributes (such as porosity, porous matrix thermal conductivity, specific surface area, and fluid-solid interstitial heat transfer coefficient). However, in the case of a large difference in the physical properties of the components (not the geometry of components), some modifications may be required to achieve an exact matching at the interface due to the discontinuity of the properties.

## 3 Result and Discussions

The derived analytical solutions are first compared against pertinent available analytical solutions as well as some numerical simulations. Lee and Vafai [29] and Marafie and Vafai [30] investigated forced convection through a channel with an imposed constant heat flux boundary condition. As such, fluid and solid temperature profiles obtained from the present work at zero inclination angle with no internal heat generation are compared with the analytical solutions given by Lee and Vafai [29] and the analytical and numerical results from Marafie and Vafai [30]. These comparisons, which are shown in Fig. 2, display an excellent agreement between the results. The small deviation in numerical results from the analytical ones is due to the usage of a slightly different Darcy number in the numerical investigations [30].

The comprehensive nature of the derived analytical solutions enables one to investigate various multicomponent consecutive, convergent, divergent, or uniform channels. This requires proper accounting of the interface boundaries between the adjacent components of a multicomponent channel. The fluid and solid temperature distributions at different axial locations of some typical multicomponent domains are presented in Figs. 3–5. The inclined angles, entrance thickness of the channel, imposed heat flux, porous matrix properties, flow rate, and fluid properties are similar in these figures. The lower wall of the multicomponent channel is subject to a constant heat flux and the upper wall is subjected to either an adiabatic boundary condition (Figs. 3 and 4) or a constant temperature boundary (Fig. 5). The temperature quantities in these figures, are normalized by the mean flow temperature at the entrance of the channel ($T_e$) and the entrance thickness of the channel ($H_1$).

In Fig. 3, a consecutive multicomponent domain is investigated consisting of a convergent (with 5 deg inclination angle), uniform and divergent (with 5 deg inclination angle) sections, respectively. As can be seen in Fig. 3, there is an excellent agreement at the interfaces of the components ($X=5$, $X=10$), between the results obtained from the analytical solution for the upstream and downstream solutions. Comparing Figs. 3(*a*) and 3(*b*) indicate that an increase in specific surface area ($a_{sf}$) or fluid-solid interstitial heat transfer coefficient ($h_{sf}$) decreases the temperature difference between solid and fluid phases.

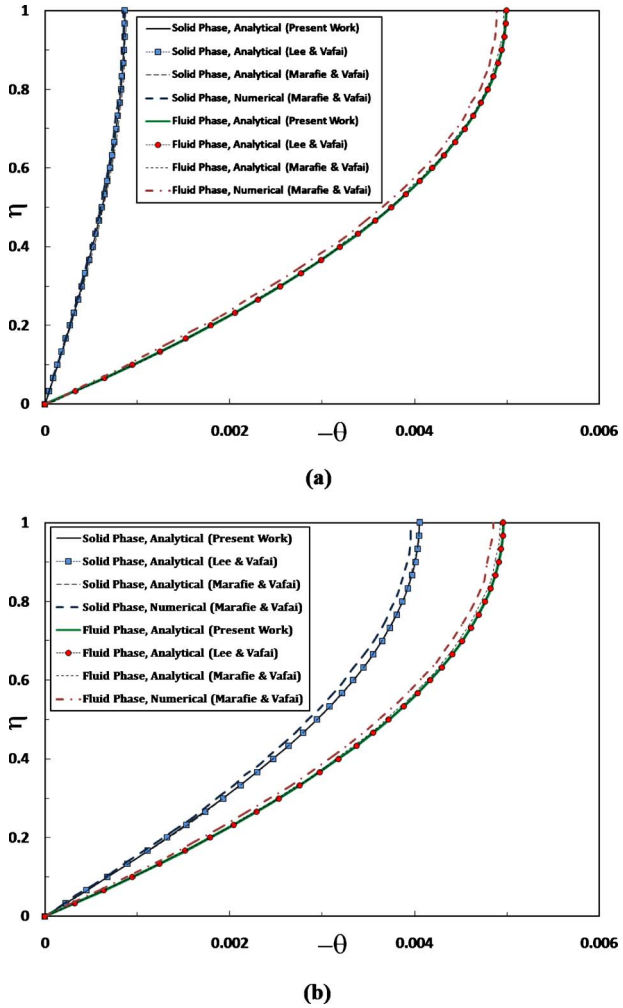
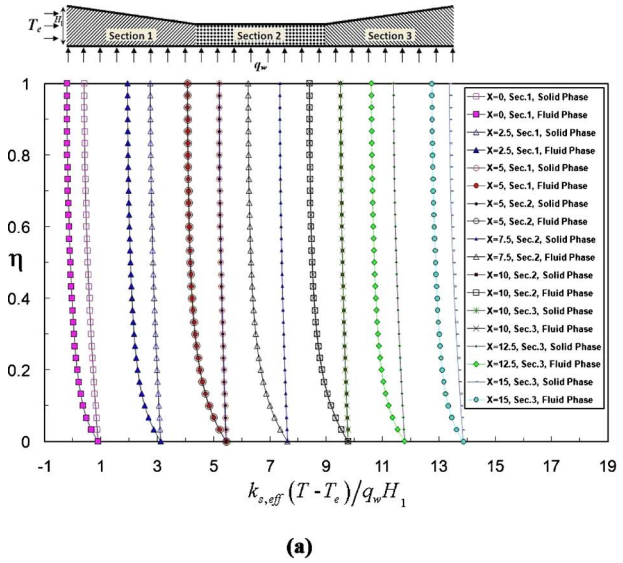


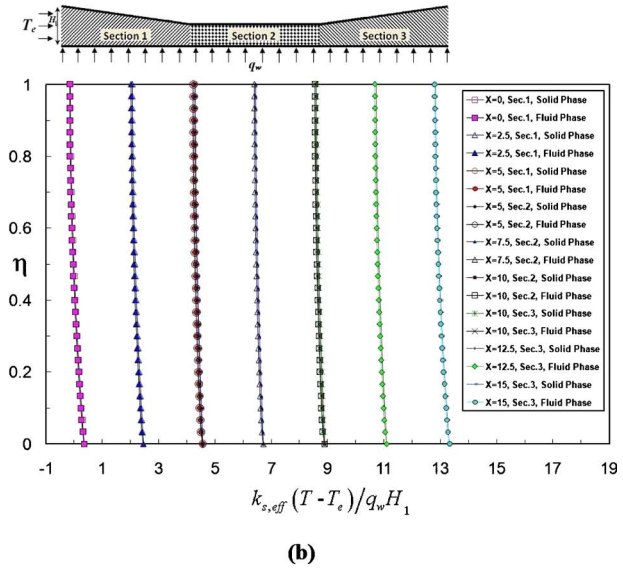

**Fig. 2 Comparison of the present analytical fluid and solid temperature distributions at zero inclination angle with the analytical results of Lee and Vafai [29] and analytical-numerical results of Marafie and Vafai [30] for $\kappa=100$, $\dot{q}_f=\dot{q}_s=0$: (*a*) Bi =0.5 and (*b*) Bi=10**

Figure 4 displays the fluid and solid temperature distributions at different locations for a divergent-uniform-convergent multisectional channel. The results once again confirm the excellent matching of fluid and solid temperature distributions at the interface between consecutive components ($X=5$, $X=10$). Comparing Figs. 3 and 4, one can see the substantial effect of the variable cross-sectional configurations on the fluid and solid temperature distributions and the wall temperature distribution. These results point out the possibility of controlling the wall temperature and temperature distribution inside each module by a proper design of multicomponent domain type of porous insert and the inlet flow attributes.

Figure 5 displays the temperature distributions at different axial locations of a consecutive multicomponent domain consisting of a convergent (with 5 deg inclination angle), uniform and divergent (with 5 deg inclination angle) modules (Fig. 5(*a*)) and also one

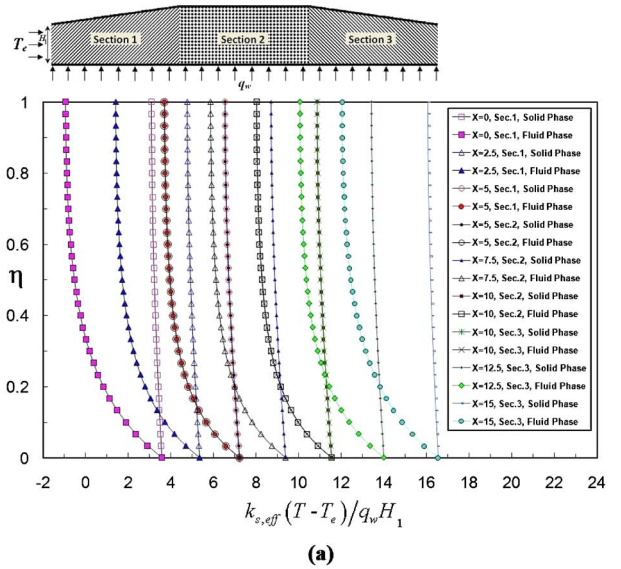



**(a)**

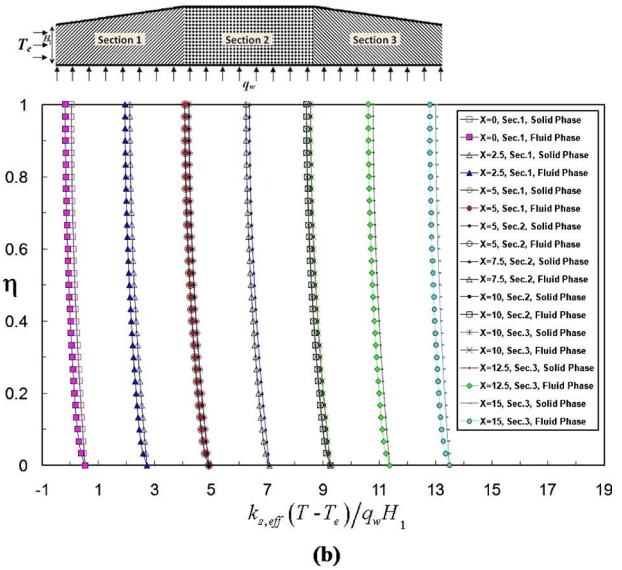



**(b)**

**Fig. 3 Fluid and solid temperature distributions at different axial locations of a variable cross-sectional domain made of convergent ($\alpha$=5 deg)-uniform-divergent ($\alpha$=5 deg) sections, subject to an adiabatic boundary at the upper wall for $\kappa$=0.01, $\dot{q}_f$=$\dot{q}_s$=0: (*a*) Bi=0.5 and (*b*) Bi=10**




**(a)**




**(b)**

**Fig. 4 Fluid and solid temperature distributions at different axial locations of a variable cross-sectional domain made of divergent ($\alpha$=5 deg)-uniform-convergent ($\alpha$=5 deg) sections, subject to an adiabatic boundary at the upper wall for $\kappa$=0.01, $\dot{q}_f$=$\dot{q}_s$=0: (*a*) Bi=0.5 and (*b*) Bi=10**

composed of divergent (with 5 deg inclination angle), uniform and convergent (with 5 deg inclination angle) sections (Fig. 5(*b*)). The lower wall of the multicomponent channel is subject to a constant heat flux and the upper wall has a constant temperature condition and the results are based on utilizing local thermal equilibrium assumption. Once again the results shown in this figure display an excellent matching at the interfaces between different neighboring sections ($X$=5 and $X$=10) of the variable cross-sectional domain. The presented analytical results in this work can be valid for a wide range of module thicknesses. It should be noted that the upper wall temperature has an important role on the temperature distribution within each module. For Fig. 5, the upper wall temperature is considered to be the same as that of the mean temperature at the channel's entrance so as to concentrate on the effects of geometry itself rather than the influence of upper wall temperature on the domain under consideration. Since the convergent or divergent inclination angle has a substantial effect on the temperature distribution, a multicomponent channel can be designed to provide the required temperature profile or a uniform temperature on a surface subject to high heat flux such as in electronic cooling applications. The convergent section can cool the domain and the surface efficiently while the temperature can be managed by adding uniform or divergent sections afterwards (Fig. 5(*a*)). As can be seen in Fig. 5(*b*), utilizing a channel with a divergent starting section can considerably reduce the cooling effects of the working flow and the upper wall temperature while a uniform channel can maintain the desired temperature on the surface.

## 4 Conclusions

A comprehensive analytical investigation of forced convection through a variable cross-sectional domain is carried out. Results obtained form this analysis are pertinent in bioheat transport through variable cross-sectional organ/tissue or in designing thermal management devices as well as in understanding porous medium based heat exchangers. Heat generation within fluid and solid phases is incorporated in the analysis to represent a more adoptive solution. The multisectional domain may compose of

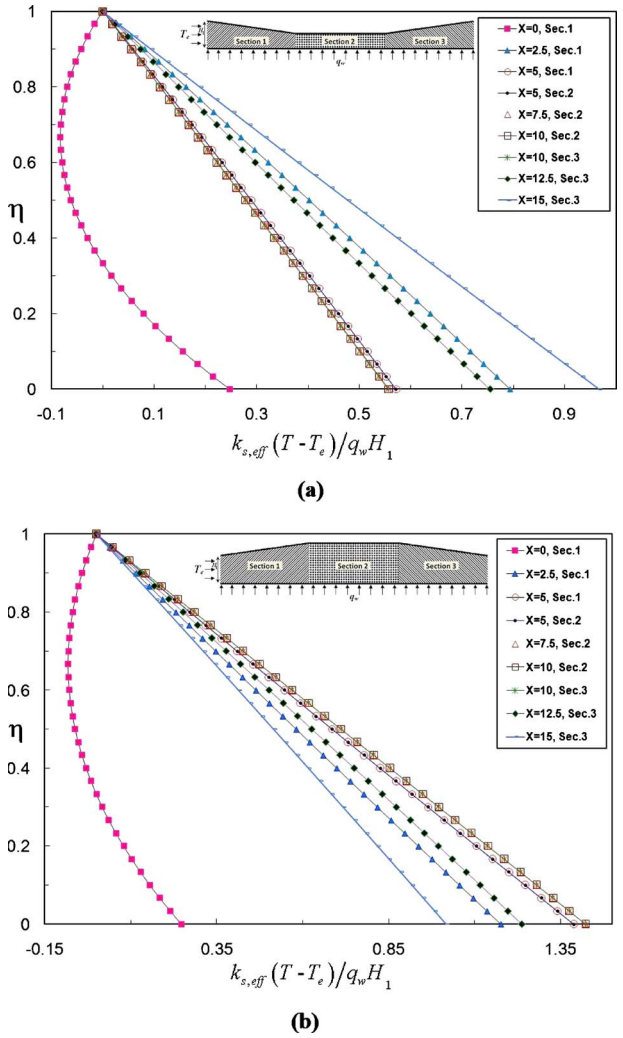



**Fig. 5 Temperature distributions at different axial locations of a variable cross-sectional domain made of (*a*) convergent ($\alpha$ =5 deg)-uniform-divergent ($\alpha$=5 deg) sections; (*b*) divergent ($\alpha$=5 deg)-uniform-convergent ($\alpha$=5 deg) sections, subject to a constant temperature at the upper wall, for $\kappa$=0.01 and $\dot{q}_f$ =$\dot{q}_s$=0.**

variable configurations (convergent, uniform or divergent) with different physical properties such as porosity, specific surface area, fluid-solid interstitial heat transfer coefficient, and different porous matrices. To cover more adoptive and practical conditions, each domain is subject to an imposed heat flux on one side and either a thermally insulated or a constant temperature boundary condition on the other side. The results obtained from the present derived analytical solutions were compared with the available analytical and numerical results in literature and were found to be in a very good agreement. These analytical solutions are presented for the first time, to the best of the authors' knowledge in literature. The temperature profiles were found to match very well at the interface between consecutive sections of a multicomponent domain. The results show that the geometrical variations have a substantial impact on the temperature field within the domain and on the surface with an imposed heat flux. As such, the temperature distributions within the variable cross-sectional domain or at the surface with an imposed heat flux can be controlled by a proper design of the system as well as proper selection of the porous matrix.

## Nomenclature

$a_{sf}$ = specific surface area ($m^{-1}$)
Bi = modified Biot number, $h_{sf}a_{sf}H^2/k_{s,\text{eff}}$
$c_p$ = fluid specific heat capacity (J $kg^{-1}$ $K^{-1}$)
$D_h$ = hydraulic diameter of the channel, $2H$ (m)
$H$ = entrance thickness of channel or each section of a multisection channel (m)
$H_1$ = entrance thickness of the first section of a multisection channel (m)
$h_{sf}$ = fluid-solid interstitial heat transfer coefficient (W $m^{-2}$ $K^{-1}$)
$h_w$ = wall heat transfer coefficient, $q_w/(T_w-T_{f,m})$ (W $m^{-2}$ $K^{-1}$)
$k_f$ = fluid thermal conductivity (W $m^{-1}$ $K^{-1}$)
$k_{f,\text{dis}}$ = fluid dispersion thermal conductivity (W $m^{-1}$ $K^{-1}$)
$k_{f,\text{eff}}$ = effective thermal conductivity of the fluid phase (W $m^{-1}$ $K^{-1}$)
$k_s$ = solid thermal conductivity (W $m^{-1}$ $K^{-1}$)
$k_{s,\text{eff}}$ = effective thermal conductivity of the solid phase (W $m^{-1}$ $K^{-1}$)
$\text{Nu}_w$ = Nusselt number at the wall
$\text{Nu}_{w,TE}$ = Nusselt number at the wall for local thermal equilibrium model
$q_w$ = imposed heat flux at the wall (W $m^{-2}$)
$\dot{q}$ = internal heat generation (W $m^{-3}$)
$T$ = temperature (K)
$T_e$ = mean flow temperature at the channel's entrance (K)
$T_{f,m}$ = mean flow temperature (K)
$T_c$ = upper wall constant temperature (K)
$T_i$ = mean flow temperature at the entrance of a single-section channel or at each section of a multicomponent channel (K)
$T_w$ = temperature of the wall subject to an imposed heat flux (K)
$T_{w,e}$ = wall temperature at the entrance of a channel (K)
$T_{w,x=x_0}$ = wall temperature at the entrance of a single-section channel or at each section of a multicomponent channel (K)
$u$ = fluid velocity (m $s^{-1}$)
$u_i$ = fluid velocity at the entrance of a channel or each section of a multicomponent channel (m $s^{-1}$)
$x$ = longitudinal coordinate (m)
$x_0$ = longitudinal coordinate at the entrance of a channel or each section of a multicomponent channel (m)
$X$ = nondimensional longitudinal coordinate, $x/H_1$
$y$ = transverse coordinate (m)

## Greek Symbols

$\alpha$ = inclination angle
$\varepsilon$ = porosity
$\eta$ = nondimensional transverse coordinate, $y/(H\pm(x-x_0)\tan\alpha)$
$\kappa$ = ratio of the effective fluid thermal conductivity to that of the solid, $k_{f,\text{eff}}/k_{s,\text{eff}}$
$\lambda$ = parameter, $\sqrt{\text{Bi}\psi/(1+\kappa)/\kappa}$
$\lambda_0$ = parameter, $\sqrt{\text{Bi}(1+\kappa)/\kappa}$
$\rho$ = fluid density (kg $m^{-3}$)
$\Gamma$ = parameter used in model II: constant temperature boundary condition
$\Delta\theta$ = nondimensional temperature difference between solid and fluid phases

$\theta$ = nondimensional temperature, $k_{s,\text{eff}}(\langle T\rangle - T_w)/(q_w H)$
$\theta_c$ = nondimensional upper wall constant temperature, $k_{s,\text{eff}}(T_c - T_w)/(q_w H)$
$\theta_{f,m}$ = nondimensional fluid mean temperature
$\Phi$ = nondimensional internal heat generation, $(1-\varepsilon)H\dot{q}/q_w$
$\psi$ = upper wall shape factor, $(H \pm (x-x_0)\tan\alpha)^2/H^2$
$\Omega$ = a dimensional parameter (K)
$\Omega_1$ = a dimensional parameter ($m^{-1}$)
$\Omega_2$ = a dimensional parameter ($m^{-1}K$)
$\Omega_3$ = a dimensional parameter (K)
$\Omega'$ = a dimensional parameter (K)

**Subscripts/Superscripts**

$f$ = fluid phase
$f,m$ = fluid mean
$c$ = constant temperature wall
eff = effective property
$s$ = solid phase
$w$ = wall subject to an imposed heat flux

**Symbols**

$+$ = applied in equations for a divergent channel wherever $\pm$ sign is utilized
$-$ = applied in equations for a convergent channel wherever $\pm$ sign is utilized
$\langle\ \rangle$ = intrinsic volume average of a quantity

## References


[1] Yoon, D. S., Lee, Y. S., Lee, Y., Cho, H. J., Sung, S. W., Oh, K. W., Cha, J., and Lim, G., 2002, "Precise Temperature Control and Rapid Thermal Cycling in a Micromachined DNA Polymerase Chain Reaction Chip," J. Micromech. Microeng., **12**, pp. 813–823.
[2] Lee, D. S., Tsai, C. Y., Yuan, W. H., Chen, P. J., and Chen, P. H., 2004, "A New Thermal Cycling Mechanism for Effective Polymerase Chain Reaction in Microliter Volumes," Microsyst. Technol., **10**, pp. 579–584.
[3] Mahjoob, S., Vafai, K., and Beer, N. R., 2008, "Rapid Microfluidic Thermal Cycler for Polymerase Chain Reaction Nucleic Acid Amplification," Int. J. Heat Mass Transfer, **51**(9–10), pp. 2109–2122.
[4] Hühmer, A. F. R., and Landers, J. P., 2000, "Noncontact Infrared-Mediated Thermocycling for Effective Polymerase Chain Reaction Amplification of DNA in Nanoliter Volumes," Anal. Chem., **72**(21), pp. 5507–5512.
[5] Swerdlow, H., Jones, B. J., and Wittwer, C. T., 1997, "Fully Automated DNA Reaction and Analysis in a Fluidic Capillary Instrument," Anal. Chem., **69**(5), pp. 848–855.
[6] Zhao, C. Y., and Lu, T. J., 2002, "Analysis of Microchannel Heat Sinks for Electronics Cooling," Int. J. Heat Mass Transfer, **45**(24), pp. 4857–4869.
[7] Fedorov, A. G., and Viskanta, R., 2000, "Three-Dimensional Conjugate Heat Transfer in the Microchannel Heat Sink for Electronic Packaging," Int. J. Heat Mass Transfer, **43**(3), pp. 399–415.
[8] Mahjoob, S., and Vafai, K., 2009, "Analytical Characterization and Production of an Isothermal Surface for Biological and Electronics Applications," ASME J. Heat Transfer, **131**(5), p. 052604.
[9] Kim, S. J., 2004, "Methods for Thermal Optimization of Microchannel Heat Sinks," Heat Transfer Eng., **25**(1), pp. 37–49.
[10] Vasiliev, L. L., 2008, "Micro and Miniature Heat Pipes—Electronic Component Coolers," Appl. Therm. Eng., **28**(4), pp. 266–273.
[11] Udell, K. S., 1985, "Heat Transfer in Porous Media Considering Phase Change and Capillarity—The Heat Pipe Effect," Int. J. Heat Mass Transfer, **28**(2), pp. 485–495.
[12] Vafai, K., and Wang, W., 1992, "Analysis of Flow and Heat Transfer Characteristics of an Asymmetrical Flat Plate Heat Pipe," Int. J. Heat Mass Transfer, **35**, pp. 2087–2099.
[13] Vafai, K., Zhu, N., and Wang, W., 1995, "Analysis of Asymmetric Disk-Shaped and Flat Plate Heat Pipes," ASME J. Heat Transfer, **117**, pp. 209–218.
[14] Mahjoob, S., and Vafai, K., 2008, "A Synthesis of Fluid and Thermal Transport Models for Metal Foam Heat Exchangers," Int. J. Heat Mass Transfer, **51**(15–16), pp. 3701–3711.
[15] Hunt, M. L., and Tien, C. L., 1988, "Effects of Thermal Dispersion on Forced Convection in Fibrous Media," Int. J. Heat Mass Transfer, **31**(2), pp. 301–309.
[16] Zhao, C. Y., Kim, T., Lu, T. J., and Hodson, H. P., 2004, "Thermal Transport in High Porosity Cellular Metal Foams," J. Thermophys. Heat Transfer, **18**(3), pp. 309–317.
[17] Boomsma, K., Poulikakos, D., and Zwick, F., 2003, "Metal Foams as Compact High Performance Heat Exchangers," J. Mech. Mater., **35**, pp. 1161–1176.
[18] Ebara, S., Toda, S., and Hashizume, H., 2000, "Application of Porous Matrix to High Heat Load Removal System," Heat Mass Transfer, **36**, pp. 273–276.
[19] Rachedi, R., and Chikh, S., 2001, "Enhancement of Electronic Cooling by Insertion of Foam Materials," Heat Mass Transfer, **37**, pp. 371–378.
[20] Mahjoob, S., and Vafai, K., 2009, "Analytical Characterization of Heat Transfer Through Biological Media Incorporating Hyperthermia Treatment," Int. J. Heat Mass Transfer, **52**(5–6), pp. 1608–1618.
[21] Mahjoob, S., and Vafai, K., 2010, "Analysis of Bioheat Transport Through a Dual Layer Biological Media," ASME J. Heat Transfer, **132**, p. 031101.
[22] Vafai, K., and Tien, C. L., 1981, "Boundary and Inertia Effects on Flow and Heat Transfer in Porous Media," Int. J. Heat Mass Transfer, **24**, pp. 195–203.
[23] Quintard, M., and Whitaker, S., 2000, "Theoretical Analysis of Transport in Porous Media," *Handbook of Porous Media*, K. Vafai, ed., Marcel Dekker, New York, pp. 1–52.
[24] Khaled, A.-R. A., and Vafai, K., 2003, "The Role of Porous Media in Modeling Flow and Heat Transfer in Biological Tissues," Int. J. Heat Mass Transfer, **46**, pp. 4989–5003.
[25] Vafai, K., and Tien, H. C., 1989, "A Numerical Investigation of Phase Change Effects in Porous Materials," Int. J. Heat Mass Transfer, **32**, pp. 1261–1277.
[26] Amiri, A., and Vafai, K., 1994, "Analysis of Dispersion Effects and Non-Thermal Equilibrium, Non-Darcian, Variable Porosity Incompressible Flow through Porous Medium," Int. J. Heat Mass Transfer, **37**, pp. 939–954.
[27] Nield, D. A., and Bejan, A., 2006, *Convection in Porous Media*, Springer, New York.
[28] Alazmi, B., and Vafai, K., 2002, "Constant Wall Heat Flux Boundary Conditions in Porous Media Under Local Thermal Non-Equilibrium Conditions," Int. J. Heat Mass Transfer, **45**, pp. 3071–3087.
[29] Lee, D. Y., and Vafai, K., 1999, "Analytical Characterization and Conceptual Assessment of Solid and Fluid Temperature Differentials in Porous Media," Int. J. Heat Mass Transfer, **42**, pp. 423–435.
[30] Marafie, A., and Vafai, K., 2001, "Analysis of Non-Darcian Effects on Temperature Differentials in Porous Media," Int. J. Heat Mass Transfer, **44**, pp. 4401–4411.
[31] Nield, D. A., and Kuznetsov, A. V., 2008, "A Bioheat Transfer Model: Forced Convection in a Channel Occupied by a Porous Medium With Counterflow," Int. J. Heat Mass Transfer, **51**, pp. 5534–5541.

# Modeling Bidirectional Transport of New and Used Organelles in Fast Axonal Transport in Neurons

**A. V. Kuznetsov**
Department of Mechanical and Aerospace Engineering,
North Carolina State University,
Campus Box 7910,
Raleigh, NC 27695-7910
e-mail: avkuznet@eos.ncsu.edu

*This paper develops a model for simulating transport of newly synthesized material from the neuron body toward the synapse of the axon as well as transport of misfolded and aggregated proteins back to the neuron body for recycling. The model demonstrates that motor-assisted transport, much similar to diffusion, can occur due to a simple concentration difference between the cell body and the synapse; organelles heading to the synapse do not need to attach preferably to plus-end-directed molecular motors, same as organelles heading to the neuron body for recycling do not need to attach preferably to minus-end-directed molecular motors. The underlying mechanics of molecular-motor-assisted transport is such that organelles would be transported to the right place even if new and used organelles had the same probability of attachment to plus-end-directed (and minus-end-directed) motors. It is also demonstrated that the axon with organelle traps and a region with a reversed microtubule polarity would support much smaller organelle fluxes of both new and used organelles than a healthy axon. The flux of organelles is shown to decrease as the width of organelle traps increases.*
[DOI: 10.1115/1.4002304]

*Keywords: molecular motors, fast axonal transport, neurons, axons and dendrites, intracellular organelles, vesicle traps, traffic jams*

## 1 Introduction

Neurons are highly polarized cells with long cellular extensions, axons, and dendrites; axons transmit electrical signals while dendrites receive electrical signals; human axons can reach up to one meter in length. Since diffusion is not a sufficiently fast process to transport large intracellular organelles over such distances, axons are strongly dependent on active transport powered by molecular motors. Active transport toward microtubule (MT) plus-ends is powered by kinesin-family molecular motors while that toward MT minus-ends is accomplished by dynein-family molecular motors [1,2].

Axons support little biosynthesis of proteins or membrane, therefore, newly synthesized organelles must be constantly imported from the synthetically active cytoplasm of the cell body and transported to the distal axon and its synapse [3]. This process is called anterograde axonal transport and since the MT polarity orientation in healthy axons is uniform (the plus end of each MT is directed away from the neuron soma toward the axon terminal, synapse), anterograde motor-driven transport is powered by kinesin-family molecular motors. Misfolded and aggregated proteins return back to the neuron body for efficient degradation by the process called retrograde axonal transport, which is powered by dynein-family molecular motors. Retrograde axonal transport is thus also required for cellular maintenance [2].

Many neurodegenerative diseases, including Alzheimer disease and the Down syndrome, are linked to swellings occurring in long arms of neurons [4]. Electron micrographs of cross sections through axonal swellings [3] show that these swellings accumulate large amounts of mitochondria, large multivesicular bodies, and other types of intracellular organelles. It is believed that these swellings result from traffic jams caused by the failure of intracellular machinery responsible for fast axonal transport [5,6]; such a traffic jam causes distal axon starvation because of the lack of organelle delivery to that region and also mechanical damages due to organelle accumulation in the traffic jam region.

Continuum-based models of organelle transport have been investigated in Refs. [7–12]; the models utilized in these references have been based on modified Smith–Simmons equations [13]. This paper develops a model for simulating transport of two species of organelles, newly synthesized organelles that are being produced at neuron body and transported to the synapse and used organelles that are returned from the synapse to the cell body for recycling. The question investigated in this paper is whether the transport of these two species of organelles can occur simply due to the difference in concentrations between the cell body and the synapse or some kind of bias in the attachment to a certain type of molecular motors is needed. In other words, the question investigated here is whether it is necessary for the new organelles to preferably attach to plus-end-directed molecular motors and for the used organelles to preferably attach to minus-end-directed molecular motors for the transport to occur in the right direction. (The question of whether such bias actually exists in axonal transport is not discussed here since there is not enough conclusive experimental data to answer it at this point.)

The model developed in this paper relies on recent experimental research by Shemesh et al. [14] who investigated tau-induced traffic jams in axons. Using live confocal microscope imaging, Shemesh et al. [14] demonstrated that (a) tau overexpression in cultured Aplysia neurons alters the polar orientation of the axonal MTs to different degrees and (b) demonstrated that the polar mismatching of the MTs leads to organelle traffic jams. Shemesh et al. [14] also found that individual MTs within the swollen axonal segment still support organelle transport. This indicates that the reason for the failure of the intracellular transport machinery is not the disruption of the motor binding sites on MTs. According to mechanistic explanation of the formation of traffic jams proposed in Ref. [14], overexpression of tau protein results in local fragmentation of MTs; this disrupts the parallel alignment of MTs and MT fragments lose their directional orientation. This research also relies on experimental findings reported in Erez and Spira [15]

Manuscript received November 11, 2009; final manuscript received February 8, 2010; published online September 27, 2010. Assoc. Editor: Peter Vadasz.

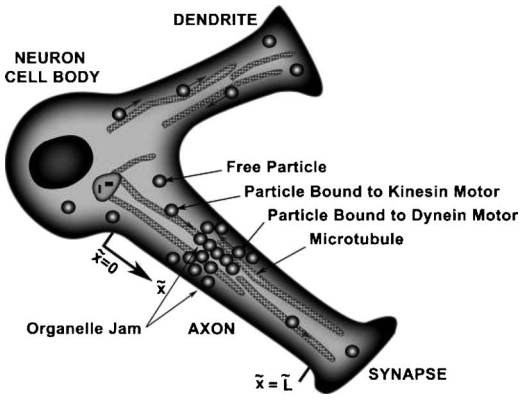



**Fig. 1 Schematic diagram of a neuron cell with a dendrite and axon; also, a traffic jam in the axon resulting from crowding of organelles at a certain location in the axon**

and Erez et al. [16] who demonstrated and explained the phenomenon of organelle trap formation by MTs with opposing polarities.

## 2 Governing Equations

**2.1 Axon With a Reversed MT Polarity in Region 3 and Two MT Vesicle Traps, Fig. 2(*b*).** Figure 1 shows a schematic diagram of a neuron [17]. The model is based on modified Smith–Simmons equations governing molecular-motor-assisted transport in neurons [13]. The kinetic diagram showing various organelle populations and kinetic processes between them is displayed in Fig. 2(*a*) (this diagram is guided by Fig. 4 in Jung and Brown [18]). Schematic diagram of the coordinate system and MT orientations for an axon with two vesicle traps is shown in Fig. 2(*b*).

*2.1.1 Region 1:* $0 \le x \le L_1$ . The equations governing transport of organelles newly synthesized in the cytoplasm of the cell body

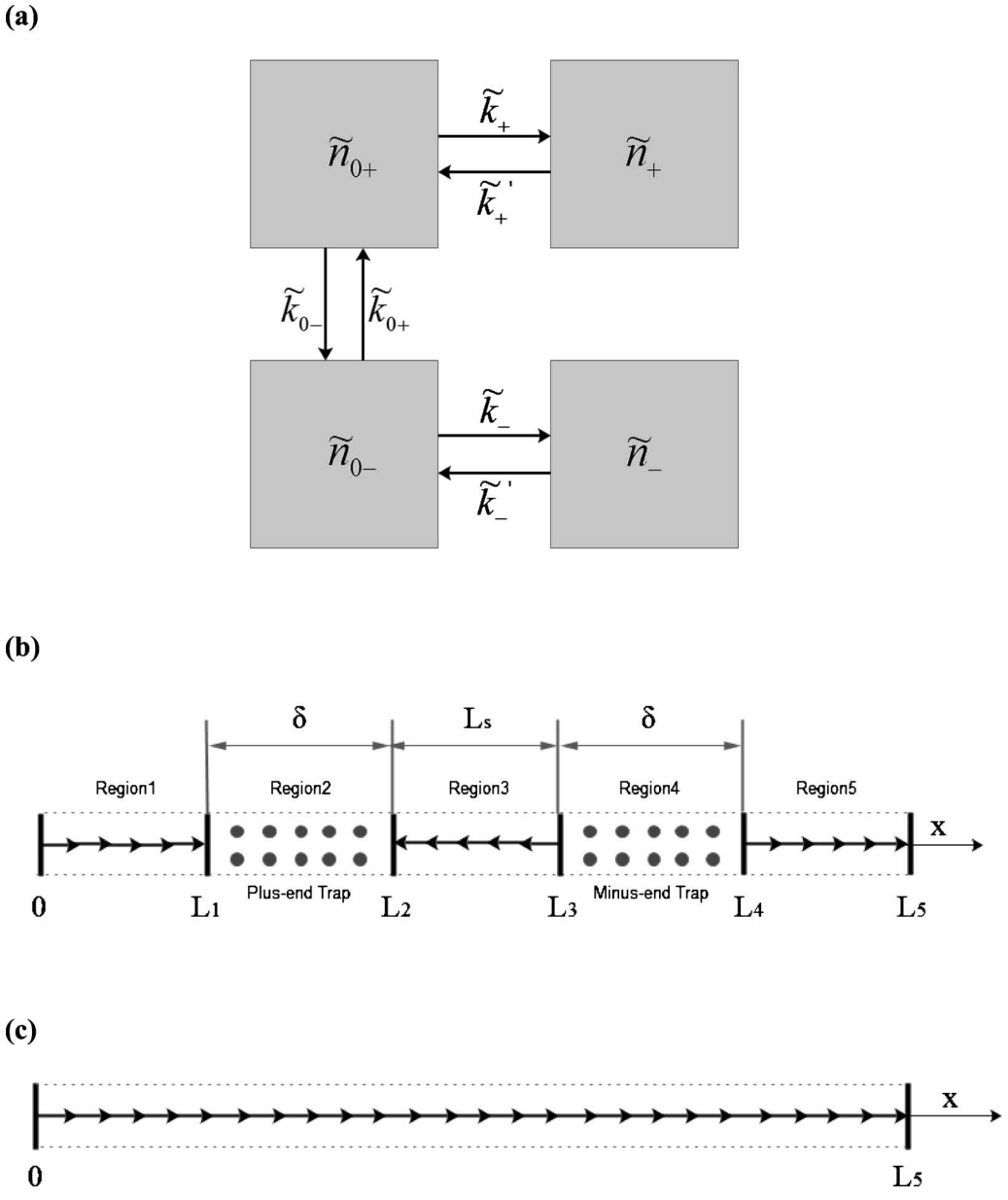



**Fig. 2 (*a*) Kinetic diagram showing various organelle populations and kinetic processes between them, (*b*) coordinate system for the axon with two vesicle traps, and (*c*) coordinate system for the healthy axon**

that are being transported toward the axon synapse (concentrations of these organelles are denoted by superscript "new") are as follows:

$$\frac{\partial n_{0+}^{\text{new}}}{\partial t} = D_{0+}\frac{\partial^2 n_{0+}^{\text{new}}}{\partial x^2} - (1+k_{0-})n_{0+}^{\text{new}} + k_{0+}n_{0-}^{\text{new}} + k'_{+}n_{+}^{\text{new}} \quad (1a)$$

$$\frac{\partial n_{0-}^{\text{new}}}{\partial t} = D_{0-}\frac{\partial^2 n_{0-}^{\text{new}}}{\partial x^2} - (k_{-}+k_{0+})n_{0-}^{\text{new}} + k_{0-}n_{0+}^{\text{new}} + k'_{-}n_{-}^{\text{new}} \quad (2a)$$

$$\frac{\partial n_{+}^{\text{new}}}{\partial t} = n_{0+}^{\text{new}} - k'_{+}n_{+}^{\text{new}} - \frac{\partial n_{+}^{\text{new}}}{\partial x} \quad (3a)$$

$$\frac{\partial n_{-}^{\text{new}}}{\partial t} = k_{-}n_{0-}^{\text{new}} - k'_{-}n_{-}^{\text{new}} - v_{-}\frac{\partial n_{-}^{\text{new}}}{\partial x} \quad (4a)$$

The equations governing transport of old, damaged components, and endocytic material that are returned to cell body for recycling (concentrations of these organelles are denoted by superscript "used") are as follows:

$$\frac{\partial n_{0+}^{\text{used}}}{\partial t} = D_{0+}\frac{\partial^2 n_{0+}^{\text{used}}}{\partial x^2} - (1+k_{0-})n_{0+}^{\text{used}} + k_{0+}n_{0-}^{\text{used}} + k'_{+}n_{+}^{\text{used}} \quad (1b)$$

$$\frac{\partial n_{0-}^{\text{used}}}{\partial t} = D_{0-}\frac{\partial^2 n_{0-}^{\text{used}}}{\partial x^2} - (k_{-}+k_{0+})n_{0-}^{\text{used}} + k_{0-}n_{0+}^{\text{used}} + k'_{-}n_{-}^{\text{used}} \quad (2b)$$

$$\frac{\partial n_{+}^{\text{used}}}{\partial t} = n_{0+}^{\text{used}} - k'_{+}n_{+}^{\text{used}} - \frac{\partial n_{+}^{\text{used}}}{\partial x} \quad (3b)$$

$$\frac{\partial n_{-}^{\text{used}}}{\partial t} = k_{-}n_{0-}^{\text{used}} - k'_{-}n_{-}^{\text{used}} - v_{-}\frac{\partial n_{-}^{\text{used}}}{\partial x} \quad (4b)$$

New and used components thus satisfy the same governing equations, the difference is only in the boundary conditions. Since the new components are synthesized in the cell body and transported toward the axon synapse, their concentration is high at $x=0$ and low at $x=L_5$. On the contrary, for the used components, the concentration is high at the synapse and low in the cell body. For all other regions, only differential equations for the new components are given since equations for the used components are identical. For simplicity, all transport coefficients for the new and used components are assumed to be the same, although they can be easily made different in the model.

Dimensionless variables in Eqs. (1*a*), (2*a*), (3*a*), and (4*a*) and Eqs. (1*b*), (2*b*), (3*b*), and (4*b*) are introduced as follows:

$$D_{0\pm} = \frac{\tilde{D}_{0\pm}\tilde{k}_{+}}{\tilde{v}_{+}^2}, \quad k_{-} = \frac{\tilde{k}_{-}}{\tilde{k}_{+}}, \quad k'_{\pm} = \frac{\tilde{k}'_{\pm}}{\tilde{k}_{+}}, \quad k_{0\pm} = \frac{\tilde{k}_{0\pm}}{\tilde{k}_{+}}, \quad L_i = \frac{\tilde{L}_i\tilde{k}_{+}}{\tilde{v}_{+}} \quad (i=1,\ldots,5) \quad (5)$$

$$v_{-} = \frac{\tilde{v}_{-}}{\tilde{v}_{+}}, \quad x = \frac{\tilde{x}\tilde{k}_{+}}{\tilde{v}_{+}}, \quad t = \tilde{t}\tilde{k}_{+}, \quad n_{0\pm}^{\text{new}} = \tilde{n}_{0\pm}^{\text{new}}\frac{\tilde{v}_{+}^3}{\tilde{k}_{+}^3}, \quad n_{\pm}^{\text{new}} = \tilde{n}_{\pm}^{\text{new}}\frac{\tilde{v}_{+}^3}{\tilde{k}_{+}^3} \quad (6)$$

$$n_{0\pm}^{\text{used}} = \tilde{n}_{0\pm}^{\text{used}}\frac{\tilde{v}_{+}^3}{\tilde{k}_{+}^3}, \quad n_{\pm}^{\text{used}} = \tilde{n}_{\pm}^{\text{used}}\frac{\tilde{v}_{+}^3}{\tilde{k}_{+}^3} \quad (7)$$

where $\tilde{D}_{0+}$ is the diffusivity of free (off-track) organelles with plus-end-directed motors attached to them (motors transporting organelles toward the MT plus-ends); $\tilde{D}_{0-}$ is the diffusivity of free (off-track) organelles with minus-end-directed motors attached to them (motors transporting organelles toward the MT minus-ends); $\tilde{t}$ is the time; $\tilde{n}_{0+}$ is the number density of free organelles with plus-end-directed molecular motors attached to them; $\tilde{n}_{0-}$ is the number density of free organelles with minus-end-directed motors attached to them; $\tilde{n}_{+}$ is the number density of organelles transported on MTs by plus-end-directed motors; $\tilde{n}_{-}$ is the number density of organelles transported on MTs by minus-end-directed molecular motors; $\tilde{x}$ is the linear coordinate along the MT tracks; $\tilde{L}_i$ $(i=1,\ldots,5)$ are the positions of the right-hand side boundaries of five axon regions, see Fig. 2(*b*); $\tilde{v}_{-}$ is the average velocity of organelles moving toward the MT minus ends (in an axon this is the motor velocity generated by dynein-family molecular motors), $\tilde{v}_{-}$ is negative; $\tilde{v}_{+}$ is the average velocity of organelles moving toward the MT plus ends (in an axon this is the motor velocity generated by kinesin-family molecular motors), $\tilde{v}_{+}$ is positive; $\tilde{k}_{+}$ is the first order rate constant for binding to MTs for plus-end-oriented organelles (organelles with plus-end-directed molecular motors attached to them); $\tilde{k}_{-}$ is the first order rate constant for binding to MTs for minus-end-oriented organelles (organelles with minus-end-directed molecular motors attached to them); $\tilde{k}'_{+}$ and $\tilde{k}'_{-}$ are the first order rate constants for detachment from MTs for plus-end-oriented and minus-end-oriented organelles, respectively; $\tilde{k}_{0+}$ is the first order rate constant describing the probability for a free minus-end-oriented organelle to become a free plus-end oriented organelle (to detach from a minus-end-directed molecular motor and attach to a plus-end-directed molecular motor); and $\tilde{k}_{0-}$ is the first order rate constant describing the probability for a free plus-end-oriented organelle to become a free minus-end oriented organelle (to detach from a plus-end-directed molecular motor and attach to a minus-end-directed molecular motor).

*2.1.2 Region 2: $L_1 \le x \le L_2$, the Plus-End Vesicle Trap.* In region 2, transport of new components in governed by the following equations:

$$\frac{\partial n_{0+}^{\text{new}}}{\partial t} = D_{0+}\frac{\partial^2 n_{0+}^{\text{new}}}{\partial x^2} - k_{0-}n_{0+}^{\text{new}} + k_{0+}n_{0-}^{\text{new}} \quad (8)$$

$$\frac{\partial n_{0-}^{\text{new}}}{\partial t} = D_{0-}\frac{\partial^2 n_{0-}^{\text{new}}}{\partial x^2} - k_{0+}n_{0-}^{\text{new}} + k_{0-}n_{0+}^{\text{new}} \quad (9)$$

Note that although there are no MT tracks in region 2, organelles can still switch the type of molecular motor they are attached to; this is the reason for the kinetic terms on the right-hand side of Eqs. (8) and (9).

The used components in region 2 satisfy the same Eqs. (8) and (9) with the superscript new replaced by the superscript used.

*2.1.3 Region 3: $L_2 \le x \le L_3$, Region With a Reversed MT Polarity (Also Referred to as the MT Swirl Region).* To model the reversal of the MT polarity in region 3, the signs before the last terms on the right-hand sides of Eqs. (12) and (13) are reversed. The new components are governed by the following equations:

$$\frac{\partial n_{0+}^{\text{new}}}{\partial t} = D_{0+}\frac{\partial^2 n_{0+}^{\text{new}}}{\partial x^2} - (1+k_{0-})n_{0+}^{\text{new}} + k_{0+}n_{0-}^{\text{new}} + k'_{+}n_{+}^{\text{new}} \quad (10)$$

$$\frac{\partial n_{0-}^{\text{new}}}{\partial t} = D_{0-}\frac{\partial^2 n_{0-}^{\text{new}}}{\partial x^2} - (k_{-}+k_{0+})n_{0-}^{\text{new}} + k_{0-}n_{0+}^{\text{new}} + k'_{-}n_{-}^{\text{new}} \quad (11)$$

$$\frac{\partial n_{+}^{\text{new}}}{\partial t} = n_{0+}^{\text{new}} - k'_{+}n_{+}^{\text{new}} + \frac{\partial n_{+}^{\text{new}}}{\partial x} \quad (12)$$

$$\frac{\partial n_{-}^{\text{new}}}{\partial t} = k_{-}n_{0-}^{\text{new}} - k'_{-}n_{-}^{\text{new}} + v_{-}\frac{\partial n_{-}^{\text{new}}}{\partial x} \quad (13)$$

The used components in region 3 satisfy the same Eqs. (10)–(13) with the superscript new replaced by the superscript used.

*2.1.4 Region 4: $L_3 \le x \le L_4$, the Minus-End Vesicle Trap*. In region 4, the new components are governed by the following equations:

$$\frac{\partial n_{0+}^{\text{new}}}{\partial t} = D_{0+}\frac{\partial^2 n_{0+}^{\text{new}}}{\partial x^2} - k_{0-}n_{0+}^{\text{new}} + k_{0+}n_{0-}^{\text{new}} \tag{14}$$

$$\frac{\partial n_{0-}^{\text{new}}}{\partial t} = D_{0-}\frac{\partial^2 n_{0-}^{\text{new}}}{\partial x^2} - k_{0+}n_{0-}^{\text{new}} + k_{0-}n_{0+}^{\text{new}} \tag{15}$$

As in region 2, kinetic terms in Eqs. (14) and (15) model organelles switching the type of molecular motor they are attached to. The used components in region 4 satisfy the same Eqs. (14) and (15) with the superscript new replaced by the superscript used.

*2.1.5 Region 5: $L_4 \le x \le L_5$* . In region 5, the new components are governed by the following equations:

$$\frac{\partial n_{0+}^{\text{new}}}{\partial t} = D_{0+}\frac{\partial^2 n_{0+}^{\text{new}}}{\partial x^2} - (1 + k_{0-})n_{0+}^{\text{new}} + k_{0+}n_{0-}^{\text{new}} + k'_{+}n_{+}^{\text{new}} \tag{16}$$

$$\frac{\partial n_{0-}^{\text{new}}}{\partial t} = D_{0-}\frac{\partial^2 n_{0-}^{\text{new}}}{\partial x^2} - (k_{-} + k_{0+})n_{0-}^{\text{new}} + k_{0-}n_{0+}^{\text{new}} + k'_{-}n_{-}^{\text{new}} \tag{17}$$

$$\frac{\partial n_{+}^{\text{new}}}{\partial t} = n_{0+}^{\text{new}} - k'_{+}n_{+}^{\text{new}} - \frac{\partial n_{+}^{\text{new}}}{\partial x} \tag{18}$$

$$\frac{\partial n_{-}^{\text{new}}}{\partial t} = k_{-}n_{0-}^{\text{new}} - k'_{-}n_{-}^{\text{new}} - v_{-}\frac{\partial n_{-}^{\text{new}}}{\partial x} \tag{19}$$

Traffic jams in the intracellular flow of organelles are steady-state objects because of a large timescale involved in their formation. Therefore, transient terms in Eqs. (1*a*), (2*a*), (3*a*), and (4*a*), Eqs. (1*b*), (2*b*), (3*b*), and (4*b*), and Eqs. (8)–(19) for the new components as well as in equations analogous to Eqs. (8)–(19) for the used components are neglected and the resulting steady-state equations are solved subject to the following boundary conditions.

At the left-hand side boundary of region 1, $x=0$, the boundary conditions for the new components are

$$n_{0+}^{\text{new}} = N_{0+}^{\text{new}}, \quad n_{0-}^{\text{new}} = N_{0-}^{\text{new}}, \quad n_{+}^{\text{new}} = \sigma_0 N_{0+}^{\text{new}} \tag{20a}$$

while the boundary conditions for the used components are

$$n_{0+}^{\text{used}} = N_{0+}^{\text{used}}, \quad n_{0-}^{\text{used}} = N_{0-}^{\text{used}}, \quad n_{+}^{\text{used}} = \sigma_0 N_{0+}^{\text{used}} \tag{20b}$$

In Eqs. (20*a*) and (20*b*), the dimensionless parameters are defined as follows:

$$N_{0+}^{\text{new}} = \tilde{N}_{0+}^{\text{new}}\frac{\tilde{v}_{+}^3}{\tilde{k}_{+}^3}, \quad N_{0-}^{\text{new}} = \tilde{N}_{0-}^{\text{new}}\frac{\tilde{v}_{+}^3}{\tilde{k}_{+}^3},$$

$$N_{0+}^{\text{used}} = \tilde{N}_{0+}^{\text{used}}\frac{\tilde{v}_{+}^3}{\tilde{k}_{+}^3}, \quad N_{0-}^{\text{used}} = \tilde{N}_{0-}^{\text{used}}\frac{\tilde{v}_{+}^3}{\tilde{k}_{+}^3} \tag{21}$$

where $\tilde{N}_{0+}$ and $\tilde{N}_{0-}$ are constant number densities of free particles with kinesin/dynein motors attached to them, respectively, maintained at $x=0$ and $\sigma_0$ is the degree of loading of plus-end-oriented organelles at $x=0$.

In simulations, $N_{0+}^{\text{new}}$ and $N_{0-}^{\text{new}}$ are assumed high (both equal to 0.1) since new components are being synthesized at the cell body while $N_{0+}^{\text{used}}$ and $N_{0-}^{\text{used}}$ are assumed low (both equal to 0.005) since used components are being recycled at the cell body.

At the boundary between regions 1 and 2, $x=L_1$, the matching conditions for the new components are

$$n_{0+}^{\text{new}}|_{x=L_1-} = n_{0+}^{\text{new}}|_{x=L_1+}, \quad n_{0-}^{\text{new}}|_{x=L_1-} = n_{0-}^{\text{new}}|_{x=L_1+}, \quad n_{-}^{\text{new}}|_{x=L_1-} = \sigma_{L_1} n_{0-}^{\text{new}}|_{x=L_1+} \tag{22}$$

$$\left(-D_{0+}\frac{dn_{0+}^{\text{new}}}{dx} + n_{+}^{\text{new}}\right)\Bigg|_{x=L_1-} = -D_{0+}\frac{dn_{0+}^{\text{new}}}{dx}\Bigg|_{x=L_1+} \tag{23}$$

$$\left(-D_{0-}\frac{dn_{0-}^{\text{new}}}{dx} + v_{-}n_{-}^{\text{new}}\right)\Bigg|_{x=L_1-} = -D_{0-}\frac{dn_{0-}^{\text{new}}}{dx}\Bigg|_{x=L_1+} \tag{24}$$

where $\sigma_{L_1}$ is the degree of loading of minus-end-oriented organelles at $x=L_1$. In Eqs. (22)–(24), $x=L_1-$ means on the left-hand side of the boundary $x=L_1$ while $x=L_1+$ means on the right-hand side of the boundary $x=L_1$. The flux of plus-end oriented organelles (see Eq. (23)) and the flux of minus-end oriented organelles (see Eq. (24)) are matched separately because switching molecular motors requires an elaborate molecular mechanism and cannot happen instantaneously as organelles cross the boundary $x=L_1$. Thus, Eqs. (23) and (24) assume that organelles retain whatever molecular motor they are attached to as they cross the boundary between the two regions.

The matching conditions at $x=L_1$ for the used components are identical to those for the new components and thus are also given by Eqs. (22)–(24) (with superscript new replaced by superscript used). The only difference between the new and used components are the boundary conditions at the neuron body, $x=0$, and at the synapse, $x=L_5$.

At the boundary between regions 2 and 3, $x=L_2$, the matching conditions for the new components are

$$n_{0+}^{\text{new}}|_{x=L_2-} = n_{0+}^{\text{new}}|_{x=L_2+}, \quad n_{0-}^{\text{new}}|_{x=L_2-} = n_{0-}^{\text{new}}|_{x=L_2+}, \quad n_{-}^{\text{new}}|_{x=L_2+} = \sigma_{L_2} n_{0-}^{\text{new}}|_{x=L_2-} \tag{25}$$

$$\left(-D_{0+}\frac{dn_{0+}^{\text{new}}}{dx} - n_{+}^{\text{new}}\right)\Bigg|_{x=L_2+} = -D_{0+}\frac{dn_{0+}^{\text{new}}}{dx}\Bigg|_{x=L_2-} \tag{26}$$

$$\left(-D_{0-}\frac{dn_{0-}^{\text{new}}}{dx} - v_{-}n_{-}^{\text{new}}\right)\Bigg|_{x=L_2+} = -D_{0-}\frac{dn_{0-}^{\text{new}}}{dx}\Bigg|_{x=L_2-} \tag{27}$$

where $\sigma_{L_2}$ is the degree of loading of minus-end-oriented organelles at $x=L_2$. Note that on the left-hand side of Eqs. (26) and (27), the signs before the terms expressing the flux of organelles by motor-driven transport, $n_{+}^{\text{new}}$ and $v_{-}n_{-}^{\text{new}}$, have been changed compared with signs before the similar terms in Eqs. (23) and (24); this is because the polarity of MTs in region 3 has been reversed. For the same reason, the boundary condition given by the third equation in (25) imposes an end-condition on $n_{-}^{\text{new}}$ (rather than on $n_{+}^{\text{new}}$).

The boundary conditions for the used components at $x=L_2$ are identical to those for the new components and thus are also given by Eqs. (25)–(27) with superscript new replaced by superscript used.

At the boundary between regions 3 and 4, $x=L_3$, the matching conditions for the new components are

$$n_{0+}^{\text{new}}|_{x=L_3-} = n_{0+}^{\text{new}}|_{x=L_3+}, \quad n_{0-}^{\text{new}}|_{x=L_3-} = n_{0-}^{\text{new}}|_{x=L_3+}, \quad n_{+}^{\text{new}}|_{x=L_3-} = \sigma_{L_3} n_{0+}^{\text{new}}|_{x=L_3+} \tag{28}$$

$$\left(-D_{0+}\frac{dn_{0+}^{\text{new}}}{dx} - n_{+}^{\text{new}}\right)\Bigg|_{x=L_3-} = -D_{0+}\frac{dn_{0+}^{\text{new}}}{dx}\Bigg|_{x=L_3+} \tag{29}$$

$$\left(-D_{0-}\frac{dn_{0-}^{\text{new}}}{dx} - v_{-}n_{-}^{\text{new}}\right)\Bigg|_{x=L_3-} = -D_{0-}\frac{dn_{0-}^{\text{new}}}{dx}\Bigg|_{x=L_3+} \tag{30}$$

where $\sigma_{L_3}$ is the degree of loading of plus-end-oriented organelles at $x=L_3$. Similar to the third equation in (25), the third equation in (28) imposes an end-condition on $n_{+}^{\text{new}}$ because the polarity of

MTs in region 3 has been reversed.

The boundary conditions for the used components at $x=L_3$ are identical to those for the new components and thus are also given by Eqs. (28)–(30) with superscript new replaced by superscript used.

At the boundary between regions 4 and 5, $x=L_4$, the matching conditions for the new components are

$$n_{0+}^{\text{new}}|_{x=L_4^-} = n_{0+}^{\text{new}}|_{x=L_4^+}, \quad n_{0-}^{\text{new}}|_{x=L_4^-} = n_{0-}^{\text{new}}|_{x=L_4^+}, \quad n_{+}^{\text{new}}|_{x=L_4^+} = \sigma_{L_4} n_{0+}^{\text{new}}|_{x=L_4^-} \tag{31}$$

$$\left(-D_{0+}\frac{dn_{0+}^{\text{new}}}{dx} + n_{+}^{\text{new}}\right)\Bigg|_{x=L_4^+} = -D_{0+}\frac{dn_{0+}^{\text{new}}}{dx}\Bigg|_{x=L_4^-} \tag{32}$$

$$\left(-D_{0-}\frac{dn_{0-}^{\text{new}}}{dx} + v_- n_{-}^{\text{new}}\right)\Bigg|_{x=L_4^+} = -D_{0-}\frac{dn_{0-}^{\text{new}}}{dx}\Bigg|_{x=L_4^-} \tag{33}$$

where $\sigma_{L_4}$ is the degree of loading of plus-end-oriented organelles at $x=L_4$. The boundary conditions for the used components at $x=L_4$ are identical to those for the new components and thus are also given by Eqs. (31)–(33) with superscript new replaced by superscript used.

At the right-hand side boundary of region 5, $x=L_5$, the boundary conditions for the new components are

$$n_{0+}^{\text{new}} = N_{L_5^+}^{\text{new}}, \quad n_{0-}^{\text{new}} = N_{L_5^-}^{\text{new}}, \quad n_{-}^{\text{new}} = \sigma_{L_5} N_{L_5^-}^{\text{new}} \tag{34a}$$

while the boundary conditions for the used components are

$$n_{0+}^{\text{used}} = N_{L_5^+}^{\text{used}}, \quad n_{0-}^{\text{used}} = N_{L_5^-}^{\text{used}}, \quad n_{-}^{\text{used}} = \sigma_{L_5} N_{L_5^-}^{\text{used}} \tag{34b}$$

In Eqs. (34*a*) and (34*b*), the dimensionless parameters are defined as follows:

$$N_{L_5^+}^{\text{new}} = \tilde{N}_{L_5^+}^{\text{new}}\frac{\tilde{v}_+^3}{\tilde{k}_+^3}, \quad N_{L_5^-}^{\text{new}} = \tilde{N}_{L_5^-}^{\text{new}}\frac{\tilde{v}_+^3}{\tilde{k}_+^3},$$

$$N_{L_5^+}^{\text{used}} = \tilde{N}_{L_5^+}^{\text{used}}\frac{\tilde{v}_+^3}{\tilde{k}_+^3}, \quad N_{L_5^-}^{\text{used}} = \tilde{N}_{L_5^-}^{\text{used}}\frac{\tilde{v}_+^3}{\tilde{k}_+^3} \tag{35}$$

where $\tilde{N}_{L_5^+}$ and $\tilde{N}_{L_5^-}$ are constant number densities of free particles with kinesin/dynein motors attached to them, respectively, maintained at $x=L_5$ and $\sigma_{L_5}$ is the degree of loading of minus-end-oriented organelles at $x=L_5$.

In simulations, $N_{L_5^+}^{\text{new}}$ and $N_{L_5^-}^{\text{new}}$ are assumed low (both equal to 0.01) since new components are being utilized at the synapse while $N_{L_5^+}^{\text{used}}$ and $N_{L_5^-}^{\text{used}}$ are assumed high (both equal to 0.05) since used components are being produced at the synapse as organelles become old and damaged.

The total dimensionless flux of new organelles is the sum of fluxes of new organelles due to diffusion and motor-driven transport

$$j^{\text{new}} = j_{\text{diff}}^{\text{new}} + j_{\text{motor}}^{\text{new}} \tag{36}$$

where

$$j_{\text{diff}}^{\text{new}} = -D_{0+}\frac{dn_{0+}^{\text{new}}}{dx} - D_{0-}\frac{dn_{0-}^{\text{new}}}{dx} \tag{37}$$

The flux of new organelles by motor-driven transport $j_{\text{motor}}^{\text{new}}$ in regions 1 ($0\le x\le L_1$) and 5 ($L_4\le x\le L_5$) is calculated as

$$j_{\text{motor}}^{\text{new}} = n_{+}^{\text{new}} + v_- n_{-}^{\text{new}} \tag{38}$$

while in region 3 (the region where MT polarity is reversed, ($L_2\le x\le L_3$) it is calculated as

$$j_{\text{motor}}^{\text{new}} = -n_{+}^{\text{new}} - v_- n_{-}^{\text{new}} \tag{39}$$

where $j_{\text{motor}}^{\text{new}}=0$ in regions 2 ($L_1\le x\le L_2$) and 4 ($L_3\le x\le L_4$) because there are no MTs in these regions.

The total, diffusion, and motor-driven fluxes of used organelles are given by equations identical to Eqs. (36)–(39) with superscript new replaced by superscript used.

**2.2 Healthy Axon (No Vesicle Traps), Fig. 2(*c*).** Schematic diagram of the coordinate system and MT orientation for a healthy axon is displayed in Fig. 2(*c*). In this case the problem is governed by Eqs. (1*a*), (2*a*), (3*a*), and (4*a*) and Eqs. (1*b*), (2*b*), (3*b*), and (4*b*) that are valid over the whole axon length, $0\le x\le L_5$, and must be solved subject to boundary conditions given by Eqs. (20*a*), (20*b*), (34*a*) and (34*b*).

Equation (36) for $j^{\text{new}}$ and $j^{\text{used}}$ still holds; Eq. (37) for $j_{\text{diff}}^{\text{new}}$ and $j_{\text{diff}}^{\text{used}}$ also holds while $j_{\text{motor}}^{\text{new}}$ and $j_{\text{motor}}^{\text{used}}$ must be now calculated by Eq. (38) everywhere in the axon.

## 3 Results and Discussion

Computations are performed for the following values of dimensionless parameters: $D_{0+}=D_{0-}=0.6$, $k_{0+}=k_{0-}=0.05$, $k_-=1$, $k'_{\pm}=0.5$, $L_1=3$, $L_2=6$, $L_3=9$, $L_4=12$, $L_5=15$, $N_{0+}^{\text{new}}=N_{0-}^{\text{new}}=0.1$, $N_{0+}^{\text{used}}=N_{0-}^{\text{used}}=0.005$, $N_{L_5^+}^{\text{new}}=N_{L_5^-}^{\text{new}}=0.01$, $N_{L_5^+}^{\text{used}}=N_{L_5^-}^{\text{used}}=0.05$, $v_-=-1$, $\sigma_0=0.1$, and $\sigma_{L_i}=0.1$ ($i=1,\ldots5$). Figure 3(*a*) displays distributions of the number density of new and used free organelles with plus-end-directed motors attached to them. Simulation results are consistent with experimental observations reported in Erez and Spira [15]; indeed, for the axons with traps, the free plus-end oriented organelles accumulate in region 2 (the plus-end trap region). The peak concentration of new organelles occurs at the left-hand side boundary of region 2 while that of used organelles occurs at the right-hand side boundary of region 2. Figure 3(*b*) displays distributions of the number densities of new and used free organelles with minus-end-directed motors attached to them. Again, in agreement with Ref. [15], for the axon with traps the free minus-end oriented organelles accumulate in region 4 (the minus-end trap region). The peak concentration of new organelles occurs at the left-hand side boundary of region 4 while the peak concentration of used organelles occurs at the right-hand side boundary of region 4. The concentration of free organelles in an axon with traps is markedly larger than in a healthy axon.

Figure 4(*a*) displays distributions of the number densities of new and used organelles transported on MTs by plus-end-directed motors. In the trap regions (regions 2 and 4), concentrations of organelles transported on MTs are zero because all organelles are assumed severed over the activity of MT-severing proteins due to the loss of tau. The detached ends of MTs require organelle dissociation and re-association to continue past the trap regions. The resistance to organelle flux due to the need for dissociation and re-association at the boundaries of the trap regions is the reason why the concentrations of MT-associated organelles peak at these boundaries. For the axon with traps, the concentration of new organelles transported on MTs takes on its maximum value at the left-hand side boundary of region 2 (the plus-end trap) while the concentration of used organelles takes on its maximum value at the right-hand side boundary of region 2. Figure 4(*b*) displays distributions of the number densities of organelles transported on MTs by minus-end-directed molecular motors. For the axon with traps, the concentration of new minus-end oriented MT-associated organelles takes on its maximum value at the left-hand side boundary of region 4 (the minus-end trap) while the concentration of used organelles takes on its maximum value at the right-hand side boundary of region 4. The concentrations of MT-associated organelles are much lower for the healthy axon than for the axon with traps.

Figure 5(*a*) displays distributions of the flux of new and used organelles due to diffusion while Fig. 5(*b*) displays distributions of the flux of new and used organelles due to motor-driven transport. For the healthy axon, the diffusion flux of new organelles is on average positive while the diffusion flux of used organelles is on average negative; this is because new organelles have high

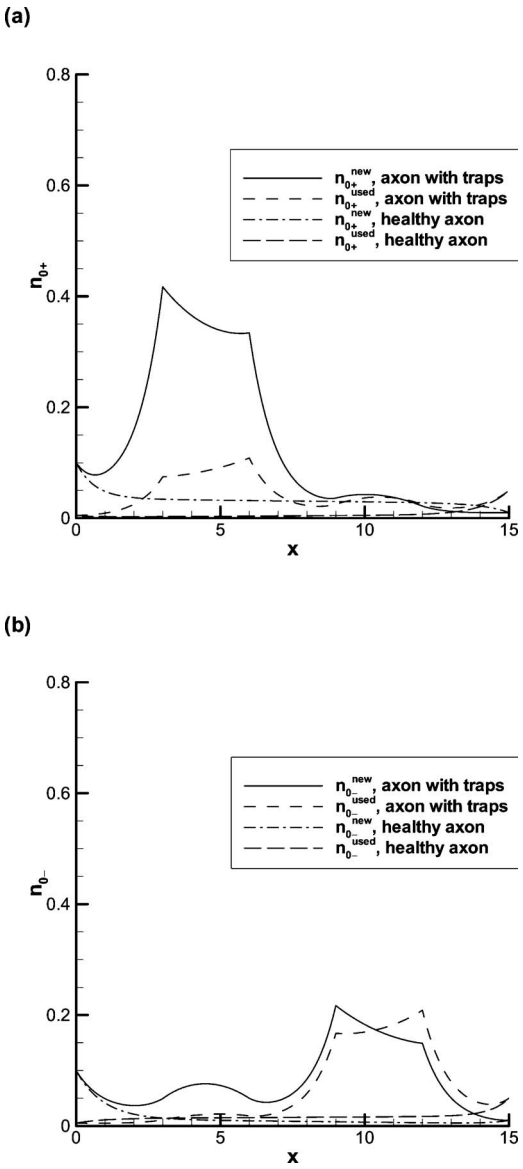



**Fig. 3** (*a*) **Distributions of the number densities of new and used free organelles with plus-end-directed motors attached to them, $n_{0+}^{new}(x)$ and $n_{0+}^{used}(x)$, and (*b*) distributions of the number densities of new and used free organelles with minus-end-directed motors attached to them, $n_{0-}^{new}(x)$ and $n_{0-}^{used}(x)$, for a healthy axon and an axon with traps**

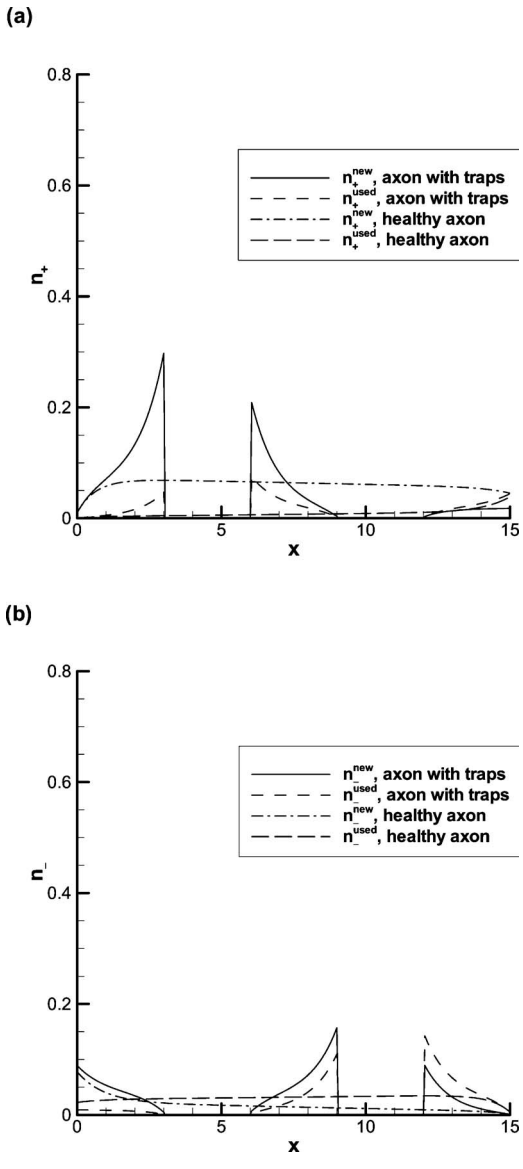



**Fig. 4** (*a*) **Distributions of the number densities of new and used organelles transported on MTs by plus-end-directed motors, $n_{+}^{new}(x)$ and $n_{+}^{used}(x)$, and (*b*) distributions of the number densities of organelles transported on MTs by minus-end-directed molecular motors, $n_{-}^{new}(x)$ and $n_{-}^{used}(x)$, for a healthy axon and an axon with traps**

concentration at the cell body (at $x=0$) and low concentration at the axon synapse (at $x=L_5$). Used organelles, on the contrary, have high concentration at the synapse and low concentration at the cell body. The same positive/negative bias applies to the average motor-driven flux of new/used organelles in a healthy axon. It should be noted that in a healthy axon the motor-driven fluxes of new/used organelles are markedly larger than the diffusion fluxes; this is because motor-driven transport is a much more efficient transport mechanism than diffusion. The distribution of diffusion and motor-driven fluxes in an axon with traps is much more complicated. One obvious feature is whenever the diffusion flux takes a sharp minimum, the motor-driven flux takes a sharp maximum (this happens, for example, at $x=L_1$), and vice versa. This is because the total flux, which according to Eq. (36) is the sum of the diffusion and motor-driven fluxes, in a steady-state problem must be constant independent of $x$; therefore, whenever the diffusion flux decreases, the motor-driven flux must increase to keep the total flux independent of $x$. Over the trap regions (regions 2 and 4), the motor-driven flux is zero and all transport occurs due to the diffusion flux, which is positive for new organelles and negative for used organelles.

Figure 6(*a*) displays distributions of the total number densities of new and used organelles, $n_t=n_{0+}+n_{0-}+n_{+}+n_{-}$, which is the quantity accessible to experiments. For the healthy axon, the total concentration of new organelles gradually decreases from $x=0$ to $x=L_5$ while for used organelles the total concentration gradually increases from $x=0$ to $x=L_5$. For the axon with traps, the total concentrations of new and used organelles take maximum values at the boundaries of the trap regions, not inside of them. This is because $n_t$ has four components, $n_{0+}$, $n_{0-}$, $n_{+}$, and $n_{-}$, two of which, $n_{+}$, and $n_{-}$, are equal to zero over regions 2 and 4. This is consistent with the mechanistic explanation of traffic jam formation put forward by Shemesh et al. [14] who found that organelles accumulate at points of MT polar mismatching and suggested that this accumulation leads to build up of the local mechanical pres-

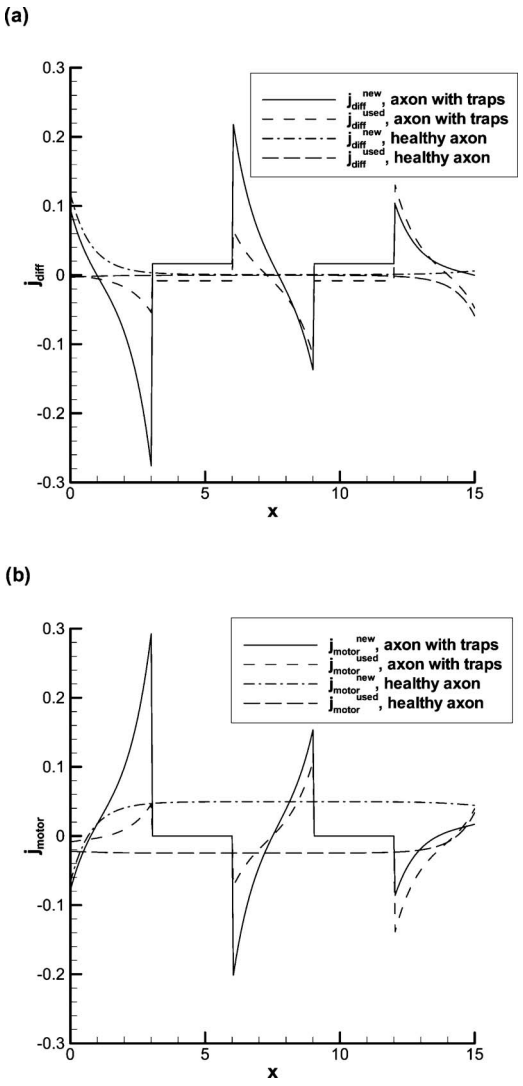



**Fig. 5** (*a*) **Distributions of the flux of new and used organelles due to diffusion, $j_{diff}^{new}(x)$ and $j_{diff}^{used}(x)$, and (*b*) distributions of the flux of new and used organelles due to motor-driven transport, $j_{motor}^{new}(x)$ and $j_{motor}^{used}(x)$, for a healthy axon and an axon with traps**

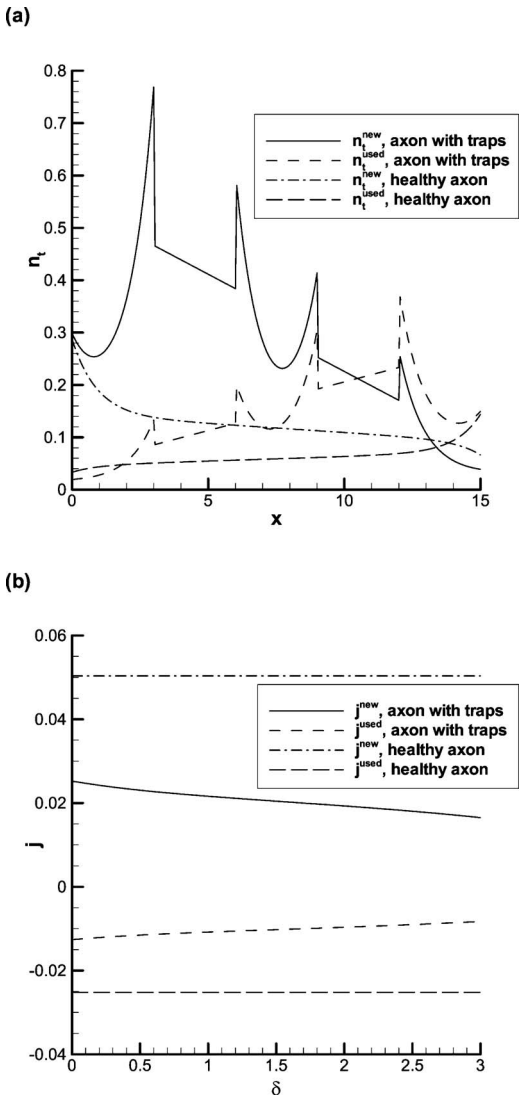



**Fig. 6** (*a*) **Distributions of the total number densities of new and used organelles, $n_t^{new}(x)$ and $n_t^{used}(x)$, and (*b*) total flux of organelles, $j^{new}(\delta)$ and $j^{used}(\delta)$, versus the width of the trap for a healthy axon and an axon with traps**

sure that augments the damages to the axon. It is interesting that $n_t(x)$ for both new and used organelles are linear over regions 2 and 4, although, $n_{0+}(x)$ and $n_{0-}(x)$ are not linear in these regions (see Figs. 3(*a*) and 3(*b*)). The linearity of $n_t(x)$ is explained by the fact that $n_t(x)=n_{0+}(x)+n_{0-}(x)$ over regions 2 and 4 and also kinetic terms cancel out when one adds Eqs. (8) and (9) for region 2 and Eqs. (14) and (15) for region 4, respectively, leaving the diffusion terms alone.

As mentioned before, for a steady-state problem for an axon with a given geometry, the total flux of organelles is a constant independent of $x$. Figure 6(*b*) shows how the total fluxes of new and used organelles depend on the widths of the trap regions, $\delta$ (see Fig. 2(*b*)). The thickness of the swirl region (region 3) and the length of the axon are kept constant, therefore, a variation of the width of regions 2 and 4 (the traps) results in a variation of the width of regions 1 and 5 to keep the length of the axon constant. For the healthy axon, there are no traps (Fig. 2(*c*)); therefore, the fluxes of new and used organelles are independent of $\delta$ with the total flux of new organelles being positive and that of used organelles being negative. This is a significant result because no bias in motor attachment for the new and used organelles is assumed; a new organelle has the same probability of attachment to a plus-end-directed molecular motor as a used one. The only difference between the new and used organelles are their concentrations at $x=0$ and $x=L_5$. The model predicts that the motor-assisted transport would work without any bias in motor attachment; all is needed to transport a certain type of organelles from point A to point B is to keep the concentration of these organelles high at point A and low at point B. The organelles do not need to be selective with respect to attachment to either plus-end-directed or minus-end-directed molecular motors; the same kinetic constants can be applied to all types of organelles, new or used, across the board. For the axon with traps, the total fluxes of organelles are significantly smaller than for the healthy axon; the fluxes decrease as the width of the trap regions increases. It is interesting that the total fluxes of organelles are about twice smaller for the axon with traps even if $\delta=0$; this is because of region 3 where the MT polarity is reversed and also because of the resistance to transport of organelles that is induced at the boundaries of region 3 by the processes of organelle dissociation and re-association to MTs. These processes are accounted for in the model through the degrees of loading at the boundaries of region 3, $\sigma_{L_2}$ and $\sigma_{L_3}$.

## 4 Conclusions

The most important finding of this paper is that no bias with respect to organelles' attachment to molecular motors is needed in order to transport organelles from one location to another location in the axon. New organelles have a higher concentration at the cell body and a lower concentration at the axon synapse; therefore, they are transported toward the synapse. Used organelles have a higher concentration at the synapse and a lower concentration at the cell body; therefore, they are transported to the neuron body. Importantly, the probability of the new organelles to attach to kinesin motors is assumed to be the same in the model as that for the used organelles, that is, new organelles do not need to attach preferably to plus-end-directed motors and used organelles do not need to attach preferably to minus-end-directed motors to accomplish their transport. This suggests that in order to regulate transport of organelles, the cell just needs to regulate organelle concentrations, which can be done naturally by synthesizing organelles at one location and degrading them at another location.

The total concentration of organelles in an axon with traps is much larger than in a healthy axon; this is consistent with damages observed in diseased axons due to the increased local mechanical pressure caused by organelle overcrowding.

The model demonstrates that the total fluxes (by diffusion and motor-driven transport) get substantially reduced by the presence of the MT swirl and two MT traps. This can be related to axon starvation that is observed in neurons affected by a range of neurodegenerative disorders. The model also demonstrates that the increase of the width of regions with severed MTs (organelle traps) further decreases the fluxes of both new and used organelles. Interestingly, even if the trap width is zero, the organelle flux in a diseased axon is still markedly smaller than in a healthy axon, which is explained by the fact that organelles still need to go through the region with a revised MT polarity (the MT swirl region) as well as overcome the resistance due to the need of dissociation and re-association to MTs at the points of MT mismatching (at the boundaries of the swirl region).

### Nomenclature

- $D_{0+}$ = dimensionless diffusivity of free organelles with plus-end-directed motors attached to them, $\tilde{D}_{0+}\tilde{k}_+/\tilde{v}_+^2$
- $\tilde{D}_{0+}$ = diffusivity of free (off-track) organelles with plus-end-directed motors attached to them ($\mu m^2/s$)
- $D_{0-}$ = dimensionless diffusivity of free organelles with minus-end-directed motors attached to them, $\tilde{D}_{0-}\tilde{k}_+/\tilde{v}_+^2$
- $\tilde{D}_{0-}$ = diffusivity of free (off-track) organelles with minus-end-directed motors attached to them ($\mu m^2/s$)
- $j$ = total dimensionless flux of organelles (by diffusion and motor-driven transport)
- $j_{\text{diff}}$ = dimensionless flux of organelles due to diffusion
- $j_{\text{motor}}$ = dimensionless flux of organelles due to motor-driven transport
- $\tilde{k}_+$ = first order rate constant for binding to MTs for plus-end-oriented organelles (1/s)
- $k_-$ = dimensionless first order rate constant for binding to MTs for minus-end-oriented organelles, $\tilde{k}_-/\tilde{k}_+$
- $\tilde{k}_-$ = first order rate constant for binding to MTs for minus-end-oriented organelles (1/s)
- $k'_+$ = dimensionless first order rate constant for detachment from MTs for plus-end-oriented organelles, $\tilde{k}'_+/\tilde{k}_+$
- $\tilde{k}'_+$ = first order rate constant for detachment from MTs for plus-end-oriented organelles (1/s)
- $k'_-$ = dimensionless first order rate constant for detachment from MTs for minus-end-oriented organelles, $\tilde{k}'_-/\tilde{k}_+$
- $\tilde{k}'_-$ = first order rate constant for detachment from MTs for minus-end-oriented organelles (1/s)
- $k_{0+}$ = dimensionless first order rate constant describing the probability for a free minus-end-oriented organelle to become a free plus-end oriented organelle, $\tilde{k}_{0+}/\tilde{k}_+$
- $\tilde{k}_{0+}$ = first order rate constant describing the probability for a free minus-end-oriented organelle to become a free plus-end oriented organelle (1/s)
- $k_{0-}$ = dimensionless first order rate constant describing the probability for a free plus-end-oriented organelle to become a free minus-end oriented organelle, $\tilde{k}_{0-}/\tilde{k}_+$
- $\tilde{k}_{0-}$ = first order rate constant describing the probability for a free plus-end-oriented organelle to become a free minus-end oriented organelle (1/s)
- $L_i$ = dimensionless positions of the right-hand side boundaries of five axon regions, $\tilde{L}_i\tilde{k}_+/\tilde{v}_+$ ($i=1,\ldots,5$)
- $\tilde{L}_i$ = positions of the right-hand side boundaries of five axon regions ($i=1,\ldots,5$), see Fig. 2(*b*) ($\mu$m)
- $n_{0+}$ = dimensionless number density of free organelles with plus-end-directed motors attached to them, $\tilde{n}_{0+}(\tilde{v}_+^3/\tilde{k}_+^3)$
- $\tilde{n}_{0+}$ = number density of free organelles with plus-end-directed motors attached to them ($1/\mu m^3$)
- $n_{0-}$ = dimensionless number density of free organelles with minus-end-directed motors attached to them, $\tilde{n}_{0-}(\tilde{v}_+^3/\tilde{k}_+^3)$
- $\tilde{n}_{0-}$ = number density of free organelles with minus-end-directed motors attached to them ($1/\mu m^3$)
- $n_+$ = dimensionless number density of organelles transported on MTs by plus-end-directed motors, $\tilde{n}_+(\tilde{v}_+^3/\tilde{k}_+^3)$
- $\tilde{n}_+$ = number density of organelles transported on MTs by plus-end-directed motors ($1/\mu m^3$)
- $n_-$ = dimensionless number density of organelles transported on MTs by minus-end-directed molecular motors, $\tilde{n}_-(\tilde{v}_+^3/\tilde{k}_+^3)$
- $\tilde{n}_-$ = number density of organelles transported on MTs by minus-end-directed molecular motors ($1/\mu m^3$)
- $n_t$ = dimensionless total number density of organelles, $n_{0+}+n_{0-}+n_+ +n_-$
- $N_{0+}$ = dimensionless number density of free particles with plus-end-directed motors attached to them maintained at $x=0$, $\tilde{N}_{0+}(\tilde{v}_+^3/\tilde{k}_+^3)$
- $\tilde{N}_{0+}$ = constant number density of free particles with plus-end-directed motors attached to them maintained at $x=0$ ($1/\mu m^3$)
- $N_{0-}$ = dimensionless number density of free particles with minus-end-directed motors attached to them maintained at $x=0$, $\tilde{N}_{0-}(\tilde{v}_+^3/\tilde{k}_+^3)$

$\tilde{N}_{0^-}$ = constant number density of free particles with minus-end-directed motors attached to them maintained at $x=0$ $(1/\mu\text{m}^3)$

$N_{L_{5^+}}$ = dimensionless number density of free particles with plus-end-directed motors attached to them maintained at $x=L_5$, $\tilde{N}_{L_{5^+}}(\tilde{v}_+^3/\tilde{k}_+^3)$

$\tilde{N}_{L_{5^+}}$ = constant number density of free particles with plus-end-directed motors attached to them maintained at $x=L_5$ $(1/\mu\text{m}^3)$

$N_{L_{5^-}}$ = dimensionless number density of free particles with minus-end-directed motors attached to them maintained at $x=L_5$, $\tilde{N}_{L_{5^-}}(\tilde{v}_+^3/\tilde{k}_+^3)$

$\tilde{N}_{L_{5^-}}$ = constant number density of free particles with minus-end-directed motors attached to them maintained at $x=L_5$ $(1/\mu\text{m}^3)$

$t$ = dimensionless time, $\tilde{t}\tilde{k}_+$

$\tilde{t}$ = time (s)

$\tilde{v}_+$ = average velocity of organelles moving toward the MT plus end $(\mu\text{m/s})$

$v_-$ = dimensionless average velocity of organelles moving toward the MT minus end, $\tilde{v}_-/\tilde{v}_+$

$\tilde{v}_-$ = average velocity of organelles moving toward the MT minus end $(\mu\text{m/s})$

$x$ = dimensionless linear coordinate along the MT tracks, $\tilde{x}\tilde{k}_+/\tilde{v}_+$

$\tilde{x}$ = linear coordinate along the MT tracks (see Fig. 1) $(\mu\text{m})$

**Greek Symbols**

$\delta$ = dimensionless width of regions with severed MTs (MT traps)

$\sigma$ = degree of loading

**Superscripts**

new = organelles newly synthesized in the cytoplasm of the cell body that are being transported toward the axon synapse

used = old damaged components, and endocytic materials that are returned to the cell body for recycling

## References


[1] Vallee, R. B., and Bloom, G. S., 1991, "Mechanisms of Fast and Slow Axonal Transport," Annu. Rev. Neurosci., **14**, pp. 59–92.

[2] Holzbaur, E. L. F., 2009, "Axonal Transport and Neurodegenerative Disease," *Intracellular Traffic and Neurodegenerative Disorders*, P. St. George-Hyslop et al., eds., Springer, Berlin, pp. 27–29.

[3] Hurd, D. D., and Saxton, W. M., 1996, "Kinesin Mutations Cause Motor Neuron Disease Phenotypes by Disrupting Fast Axonal Transport in Drosophila," Genetics, **144**, pp. 1075–1085.

[4] Stokin, G. B., Lillo, C., Falzone, T. L., Brusch, R. G., Rockenstein, E., Mount, S. L., Raman, R., Davies, P., Masliah, E., Williams, D. S., and Goldstein, L. S. B., 2005, "Axonopathy and Transport Deficits Early in the Pathogenesis of Alzheimer's Disease," Science, **307**, pp. 1282–1288.

[5] Goldstein, L. S. B., 2001, "Kinesin Molecular Motors: Transport Pathways, Receptors, and Human Disease," Proc. Natl. Acad. Sci. U.S.A., **98**, pp. 6999–7003.

[6] Martin, M. A., Iyadurai, S. J., Gassman, A., Gindhard, J. G., Jr., Hays, T. S., and Saxton, W. M., 1999, "Cytoplasmic Dynein, the Dynactin Complex, and Kinesin are Independent and Essential for Fast Axonal Transport," Mol. Biol. Cell, **10**, pp. 3717–3728.

[7] Kuznetsov, A. V., and Hooman, K., 2008, "Modeling Traffic Jams in Intracellular Transport in Axons," Int. J. Heat Mass Transfer, **51**, pp. 5695–5699.

[8] Kuznetsov, A. V., and Avramenko, A. A., 2009, "A Minimal Hydrodynamic Model for a Traffic Jam in an Axon," Int. Commun. Heat Mass Transfer, **36**, pp. 1–5.

[9] Kuznetsov, A. V., and Avramenko, A. A., 2009, "A Macroscopic Model of Traffic Jams in Axons," Math. Biosci., **218**, pp. 142–152.

[10] Kuznetsov, A. V., 2010, "Effect of the Degree of Polar Mismatching on Traffic Jam Formation in Fast Axonal Transport," Comput. Methods Biomech. Biomed. Eng., in press.

[11] Kuznetsov, A. V., 2010, "Effect of Vesicle Traps on Traffic Jam Formation in Fast Axonal Transport," Math. Biosci., **226**, pp. 147–155.

[12] Kuznetsov, A. V., 2010, "Effect of the Width of Regions With Severed Microtubules on Transport of Organelles Down the Axon," Int. Commun. Heat Mass Transfer, **37**, 443–448.

[13] Smith, D. A., and Simmons, R. M., 2001, "Models of Motor-Assisted Transport of Intracellular Particles," Biophys. J., **80**, pp. 45–68.

[14] Shemesh, O. A., Erez, H., Ginzburg, I., and Spira, M. E., 2008, "Tau-Induced Traffic Jams Reflect Organelles Accumulation at Points of Microtubule Polar Mismatching," Traffic (Oxford, U. K.), **9**, pp. 458–471.

[15] Erez, H., and Spira, M. E., 2008, "Local Self-Assembly Mechanisms Underlie the Differential Transformation of the Proximal and Distal Cut Axonal Ends Into Functional and Aberrant Growth Cones," J. Comp. Neurol., **507**, pp. 1019–1030.

[16] Erez, H., Malkinson, G., Prager-Khoutorsky, M., De Zeeuw, C. I., Hoogenraad, C. C., and Spira, M. E., 2007, "Formation of Microtubule-Based Traps Controls the Sorting and Concentration of Vesicles to Restricted Sites of Regenerating Neurons After Axotomy," J. Cell Biol., **176**, pp. 497–507.

[17] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P., 2008, *Molecular Biology of the Cell*, 5th ed., Garland Science, New York.

[18] Jung, P., and Brown, A., 2009, "Modeling the Slowing of Neurofilament Transport Along the Mouse Sciatic Nerve," Phys. Biol., **6**, p. 046002.

**Peter Vadasz**[1]
Professor
Fellow ASME
e-mail: peter.vadasz@nau.edu

**Alisa S. Vadasz**
Adjunct Assistant Professor

College of Engineering, Forestry and Natural Sciences,
Northern Arizona University,
P.O. Box 15600,
Flagstaff, AZ 86011-5600;
and Faculty of Engineering,
University of KwaZulu-Natal,
Durban 4041, South Africa

# Metabolic Mass Transfer in Monotonic Growth of Microorganisms

*Microorganism growth and reproduction have been traditionally modeled independently of the direct effect of the metabolic process. The latter caused inconsistencies between the modeling results and experimental data. A major inconsistency was linked to the experimentally observed lag phase in the growth process. Attempts to associate the lag phase to delay processes have been recently proven incorrect. The only other alternative is the existence of unstable stationary states resulting from the explicit inclusion of the metabolic mass transfer process via the resource consumption and utilization. The proposed theory that accounts for the latter is presented, analyzed, and compared with experimental data both qualitatively as well as quantitatively.* [DOI: 10.1115/1.4002416]

*Keywords: metabolic mass transfer, monotonic growth, microorganisms, lag*

## 1 Introduction

The diversity of models being used for predictive growth of microorganisms does not provide an acceptable general theory that supports all experimental results. Historically, the first attempt to propose a predictive model for population dynamics was due to Malthus [1], followed by Verhulst [2], who introduced the logistic growth model (LGM), and Pearl [3], who showed that the LGM can reproduce accurately experimental growth data in some circumstances. The latter were expanded, leading to a diversity and wide variety of models. We counted at least 20 different models that are being used in "predictive microbiology" or in the field of "ecology" to describe population growth, 14 of them being consistently used in the past decade (for a list, see Vadasz and Vadasz [4]). It is widely agreed (Baty and Delignette-Muller [5], Augustin and Carlier [6], and Baranyi [7]) that the lag phase, a typical phase in microorganism growth, cannot be estimated accurately by most existing models. Additionally, most of the existing models cannot capture typical phases or growth regimes such as growth followed by decay (decline phase) or oscillations. The existence of a decay phase where the cell count declines is a typical phase in batch growth, which has been well documented experimentally and is included in most textbooks as a classical growth phase. Reports on oscillations in microbial growth are less common, but Vadasz et al. [8–10] presented both experimental evidence as well as theoretical arguments demonstrating the existence of damped oscillatory growth even for a single species, i.e., a particular strain of yeast.

A qualitative description of a typical microbial growth curve is presented in Fig. 1, where five distinct phases are present (and possibly six if a decay/decline phase is also present). The classical way of describing a typical microbial growth curve is via four phases (and possibly five if a decay/decline phase is also present), i.e., the lag phase, which may or may not be present, an exponential growth phase, an inhibition phase, and a stationary phase. We introduced a further subdivision of the exponential growth phase as it becomes convenient for the purpose of identifying different stages of growth that result in our study. As such, we subdivided the exponential phase into logarithmic exponential (LogEx) and regular exponential (REx) growth phases. We will define each one of these phases in what follows. The first is a lag phase identified by an approximately constant value of population number at the initial growth stage on a logarithm of the cell concentration versus time graph ($\ln x$ versus $t$). Therefore, the lag phase is characterized by the condition $\mathrm{d}\ln x(t)/\mathrm{d}t \cong 0$ for times between $0 \le t \le \lambda$, where $\lambda \equiv \lambda_{cl}$ or $\lambda \equiv \lambda_{rd}$ is the lag duration that will be defined and discussed later.

The lag phase is followed by a LogEx phase identified by an exponential growth on a logarithmic versus time graph ($\ln x$ versus $t$). We introduce the terminology of LogEx phase in order to distinguish it from the REx growth phase that follows, the latter being identified by an approximately straight line on the $\ln x$ versus $t$ graph. The inhibition (Inhib) phase is next, which expresses itself as an exponential consolidation toward the stationary constant value on the $\ln x$ versus $t$ graph. The stationary (Stat) phase is usually the last phase (if decay/decline is absent) and can be identified by an approximately constant value of the population number at the final growth stage. The growth curve, which starts with an almost constant value during the lag phase, becomes concave-up during the LogEx and at the beginning of the REx phase and converts to a concave-down curve at the end of the REx phase and during the inhibition and stationary phases. The point where the curve shifts from concave-up to concave-down is the logarithmic inflection point (LIP), which is substantially distinct from the regular inflection point (RIP) that appears on typical sigmoid curves produced by the LGM and presented on an $x$ versus $t$ graph. The salient features of microbial growth that include the lag, LogEx, and the beginning of the REx phase (prior to the LIP point) can only be observed on an $\ln x$ versus $t$ graph. The logistic model, for example, cannot reproduce a LIP, but it may reproduce a RIP, depending on the initial conditions.

The classical definition of the lag duration $\lambda_{cl}$ is due to Pirt [11] as the time value where the tangent line to the growth curve at the LIP, where the specific growth rate is maximum, i.e., at $y_{LIP} = \ln(x_{LIP})$, intersects the line parallel to the $t$ axis at the value of the initial condition $y = y_o = \ln(x_o)$ (the black dot in Fig. 1). The classical definition of the lag has a tendency to overestimate the lag duration, as observed in Fig. 1, where $\lambda_{cl}$ is clearly positioned within the LogEx phase, while in reality it should be the border point between the lag and LogEx phases. A redefined lag duration introduced by Vadasz and Vadasz [12,13] is presented in Fig. 1, the latter being linked with different models such as Baranyi and Roberts [14] and Vadasz and Vadasz [15,16].

[1]Corresponding author.

Contributed by the Heat Transfer Division of ASME for publication in the JOURNAL OF HEAT TRANSFER. Manuscript received November 11, 2009; final manuscript received August 10, 2010; published online September 30, 2010. Assoc. Editor: Andrey Kuznetsov.

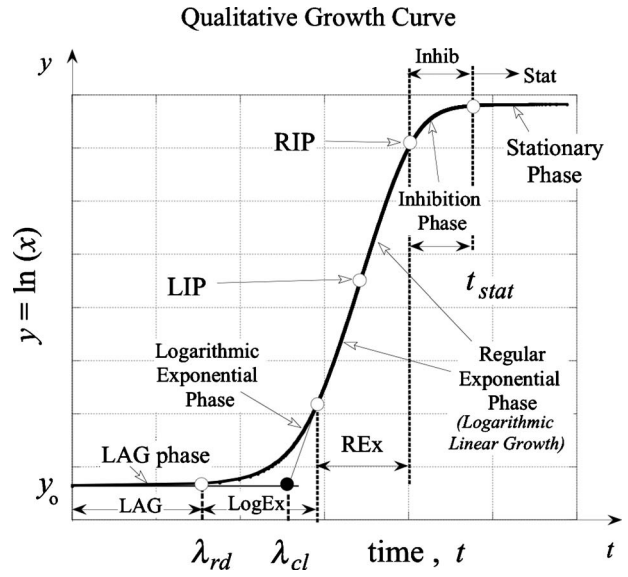



**Fig. 1 A typical microbial growth curve (reproduced from Vadasz and Vadasz [22])**

The curve presented in Fig. 1 represents monotonic growth. A monotonic growth curve is defined as any growth curve that does not reach a maximum, nor a minimum, except at $t=0$ or asymptotically as $t\to\infty$. Growth followed by decay is a typical example of nonmonotonic growth; oscillation is another. All the qualitative features (phases) that are listed and described in Fig. 1 can, in principle, be captured by one first-order autonomous differential equation, except for the lag phase (phase 1) or the decay phase (phase 6). On the other hand, all presently available models, except for the Baranyi and Roberts [14] model, are first-order autonomous systems. As a result, they have no ability to capture the lag nor the decay phase. Baranyi and Roberts [14] were the first to realize that it is impossible to describe microbial growth by using a first-order autonomous system. Baranyi and Roberts [14] introduced a second-order system and obtained a nonautonomous first-order system via an explicitly time dependent adjustment function that can be evaluated by following the Michaelis–Menten kinetics. Vadasz and Vadasz [13] showed how the Baranyi and Roberts model [14] can be transformed into an equivalent autonomous system and how its properties can be uncovered by this transformation.

One additional important biological feature is related to experimental evidence related to reinoculation conditions. Maier [17] introduced this point by stating the conditions that are necessary to encounter a lag phase in microbial growth as follows: (i) dependence on the type of medium as well as on the initial inoculum size and (ii) dependence on the initial growth rate. For example, if an inoculum is taken from an exponential phase culture (high initial growth rate) in trypticase soy broth (TSB) and is placed into a fresh TSB medium at a concentration of $10^6$ cells/ml (large initial inoculum size) under the same growth conditions (temperature and shaking speed) (similar type of medium), there will be no noticeable lag phase. However, if the inoculum is taken from a stationary phase culture (very small initial growth rate), there will be a lag phase. Similarly, if the inoculum is placed into a medium other than TSB, for example, a mineral salt medium with glucose as the sole carbon source (a more stressed type of medium), a lag phase will be observed. Finally, if the inoculum size is small, for example, $10^4$ cells/ml, a lag phase will be observed until the population reaches approximately $10^6$ cells/ml. A similar description was presented by Baranyi and Roberts [14]. Because of the lack of clarity (Swinnen et al. [18]) regarding the accurate definition of the "physiological state of the cells at inoculation," Baranyi and Roberts [14] linked the latter to "the previous history of the cell," a true statement that might have misleading connotations to delay-type models (May [19–21]), which cannot capture correctly the lag phase as demonstrated by Vadasz and Vadasz [22]. Vadasz and Vadasz [13] showed the accurate definition of Baranyi and Roberts's [14] parameter representing the physiological state of the cells at inoculation by deriving the latter link to the initial growth rate at inoculation.

An autonomous neoclassical model proposed by Vadasz and Vadasz [12] based on Vadasz and Vadasz [15,16] was shown to capture all the qualitative features that appear in experiments for monotonic growth of microorganisms, such as lag, LIP, convex and concave curves on the phase diagram, and the LGM as a special case. Vadasz and Vadasz [12] showed that their proposed model fits well the experimental results for five distinct sets of data. In addition, the model that Vadasz and Vadasz [15,16] derived from first principles was shown to capture additional qualitative features that appear in nonmonotonic regimes of growth, such as growth followed by decay (decline) as well as oscillations.

To summarize this introduction, one may conclude that there appear to be only two models that capture all qualitative features of monotonic growth of microorganisms, and they are the Baranyi and Roberts [14] model and the neoclassical model proposed by Vadasz and Vadasz [15,16,12]. Both models incorporate the mechanism of metabolic mass transfer explicitly. All other models cannot capture the lag phase naturally.

## 2 Model Formulation and Metabolic Mass Transfer

Vadasz and Vadasz [12] used a constitutive relationship between the total viable biomass and the cell concentration, i.e., $M(x)$. This constitutive relationship represents the result of the *actual* nutrient (or other resources) consumption and utilization rather than the *potential* nutrient (or other resources) availability and quality, as described via the environmental conditions, such as pH and temperature. To describe the former, it is essential to distinguish between the deterministic description of mechanisms within a single cell and the statistically averaged behavior of a cohort of cells in terms of their interaction with their environment and the corresponding interaction between the cells themselves. Let $M$ represent the total viable cell biomass in the solution or per unit volume of the solution. Then, we may define the average viable cell biomass as $m=M/x$. The expression $M=mx$ is therefore an accurate representation (by definition) of the relationship between the *total* viable cell biomass $M$ and the *average* viable cell biomass $m$. By using only this simple representation relating the total viable cell biomass to a statistically averaged value of a "typical cell viable biomass" (the latter reflecting a representative average cell), simple but intriguing effects are revealed when following simple derivations. For example, evaluating the *rate of total viable biomass production* during the cell growth process, i.e., $\mathrm{d}M/\mathrm{d}t$ leads to

$$\frac{\mathrm{d}M}{\mathrm{d}t}=\frac{\mathrm{d}(mx)}{\mathrm{d}t}=m\frac{\mathrm{d}x}{\mathrm{d}t}+x\frac{\mathrm{d}m}{\mathrm{d}t} \tag{1}$$

which can be presented in the following alternative form explicitly for the *rate of average viable biomass production*:

$$\frac{\mathrm{d}m}{\mathrm{d}t}=\frac{1}{x}\frac{\mathrm{d}M}{\mathrm{d}t}-m\left(\frac{1}{x}\frac{\mathrm{d}x}{\mathrm{d}t}\right) \tag{2}$$

It is evident from Eq. (2) that the rate of average viable biomass production is equal to the rate of total viable biomass production per unit cell minus an additional quantity reflecting the effect of the cells specific growth rate $(\mathrm{d}x/\mathrm{d}t)/x$ on the rate of average viable biomass production. When the growth rate is positive, reflecting a positive growth in the population number, the rate of average viable biomass production is reduced when compared with the viable biomass production per cell. The opposite occurs when the cell concentration declines, representing a negative

growth. Only in the stationary phase do the two rates of viable biomass production become equal, i.e., when $\dot{x}=0$ (stationary phase conditions) $\dot{m}=\dot{M}/x$ (where we introduced the Newtonian time derivative notation $\dot{x}=dx/dt$, $\dot{m}=dm/dt$, and $\dot{M}=dM/dt$). There is therefore a fundamental distinction between the rate of average viable biomass production represented by $\dot{m}=dm/dt$ and the *average rate of viable biomass production* represented by $\dot{M}/x$. They become identical only in the stationary phase. This reduction in the average viable biomass is a result of the division process that increases the number of cells without affecting the total viable biomass. This effect was demonstrated generally and in more detail by Vadasz and Vadasz [15,16].

In addition to introducing in the model the latter effect by relating the growth of the cell concentration to the total viable cell biomass, this model includes an additional effect that is missing in all other existing models. Most of the existing population models include two distinct mechanisms in the form of growth inertia, as described by Ginzburg [23] and Vadasz and Vadasz [22,15,16] and a resistance to growth that produces the inhibition phase on the growth curve. Vadasz and Vadasz [12] proposed to include a third mechanism that is not included in any existing model. This third mechanism reflects storage effects on microbial growth. Cells have the ability to store nutrient as well as "potential life" as they grow in size toward division.

The proposed neoclassical model was derived from first biological as well as physical principles by Vadasz and Vadasz [15,16] and can be presented in the form of an autonomous system of two first-order differential equations:

$$\dot{x}=\mu_{\max}\left[1-\frac{x}{\delta}+\frac{(z-\mu_{\max})}{\mu_{\max}(1+r_m x)}\right]x \tag{3a}$$

$$\dot{z}=k_o f(x,\dot{x}) \tag{3b}$$

where $x$ is the cell concentration, $\dot{x}=dx/dt$ is the growth rate, $z$ is related (but not identical) to the nutrient (or resource) consumption/utilization rate, and $\dot{z}=dz/dt$ is related (but not identical) to the nutrient consumption acceleration. The denominator in the third term of Eq. (3*a*) is related to the net average biomass; hence, the third term is a representation of the excess nutrient consumption/utilization per unit of viable biomass in the population (see Vadasz and Vadasz [15,16] for details). The other symbols are as follows: $\mu_{\max}$ is the maximum specific growth rate, $\delta$ is the carrying capacity of the environment, and $r_m=m_1/m_o$ is the ratio between the viable biomass constant coefficients $m_1$ and $m_o$ corresponding to a linear constitutive relationship between the total viable biomass and the cell concentration in the form $M(x)=m_o+m_1x$. The accurate form of the function $f(x,\dot{x})$ affects only nonmonotonic growth regimes such as growth followed by decay (decline), as presented by phase 6 (not shown in Fig. 1), or oscillatory growth. Vadasz and Vadasz [15,16] used a linear approximation for this function to show very good agreement of the model with experimental results of nonmonotonic growth.

For monotonic growth, it is sufficient to mention that when the value of the coefficient $k_o$ is identically zero, Eq. (3*b*) produces the solution $z=z_o=\text{const}$, where $z_o$ represents the initial condition of $z$ at $t=0$ defined by using Eq. (3*a*) at $t=0$ in the form

$$z_o=\mu_{\max}+(1+r_m x_o)\left[\left(\frac{\dot{x}}{x}\right)_{t=o}+\mu_{\max}\left(\frac{x_o}{\delta}-1\right)\right] \tag{4}$$

The latter produces monotonic growth, leading to a growth curve that does not exhibit a maximum nor a minimum except at the boundaries at $t=0$ or $t\to\infty$. Monotonic growth is regularly obtained in laboratory experiments, whether in batch or continuous environments, and is also observed in nature. In the event that the value of $k_o$ is not identically zero but rather very small, one can anticipate an original monotonic growth curve over the short to medium time scales, followed at much longer (slower) time scales by a possibly nonmonotonic behavior. Examples of nonmonotonic features obtained by using Eq. (3) were presented by Vadasz and Vadasz [15,16] and were shown to compare well with experimental data. However, for monotonic growth, the system of Eqs. (3*a*) and (3*b*) degenerates to one first-order differential equation in the form

$$\dot{x}=\mu_{\max}\left[1-\frac{x}{\delta}+\frac{(z_o-\mu_{\max})}{\mu_{\max}(1+r_m x)}\right]x \tag{5}$$

which was obtained from the second-order system (3) as a mathematically degenerated case. The only memory, which this equation has that it originated from a second-order system, is the term $z_o$, which depends on two initial conditions, i.e., on the initial cell concentration $x_o$ as well as on the initial growth rate $\dot{x}_o$. It is easy to observe that when $z_o=\mu_{\max}$, corresponding to a particular set of initial conditions, Eq. (5) produces the particular case of Pearl's [3] LGM. The system represented by Eq. (5) is an autonomous system.

## 3 Transition Points on the Phase Diagram

The analysis of the monotonic growth following Eq. (5) is undertaken first by plotting the solution on a phase diagram in terms of the specific growth rate (per capita growth rate) $(\dot{x}/x)$ versus the cell concentration $x$. There are a variety of regimes, linked to the behavior of the solution, that depend on different ranges of the parameters $\mu_{\max}$, $r_m$, and $\delta$ and on the initial conditions expressed by $x_o$ and $\dot{x}_o$ (or, alternatively, $z_o$). For plotting the solution on a phase diagram, one presents Eq. (5) in the following form:

$$\frac{\dot{x}}{x}=\mu_{\max}\left[1-\frac{x}{\delta}+\frac{(z_o-\mu_{\max})}{\mu_{\max}(1+r_m x)}\right] \tag{6}$$

which can be expanded by using a common denominator on its right hand side. One may separate the parameter domain into two distinct regimes, namely, $r_m<1/\delta$ and $r_m>1/\delta$. The solution for $r_m<1/\delta$ and $r_m>1/\delta$ produces a variety of curves on the phase diagram. The regime corresponding to $r_m<1/\delta$ was presented in a dimensionless form by Vadasz and Vadasz [15,16] showing that the curves corresponding to $z_o>\mu_{\max}$ are concave, while the curves corresponding to $z_o<\mu_{\max}$ are convex; however, no LIP and no lag are possible in this parameter regime.

The more interesting growth regime corresponding to $r_m>1/\delta$ is presented on the phase diagram in Fig. 2 showing the solutions in terms of the specific growth rate $(\dot{x}/x)$ versus the cell concentration $x$. A straight line corresponding to the LGM solution, which occurs when $z_o=\mu_{\max}$, divides the phase plane into two regions, namely, $z_o>\mu_{\max}$ and $z_o<\mu_{\max}$. In the region where $z_o>\mu_{\max}$, the curves are concave, while in the region corresponding to $z_o<\mu_{\max}$, the curves are convex. The positive $x$ axis consists of a continuous distribution of stationary points where $(\dot{x}/x)=0$. The stationary points to the right of the point $x_A$ represented by the continuous thick line are stable, while the stationary points to the left of $x_A$ represented by the dotted line are unstable, as observed in Fig. 2 by following the direction of the arrows representing the solution moving in the positive time direction. Any point on the phase plane represents a possible initial condition. Once such a point is set (i.e., an initial condition for both $x_o$ and $\dot{x}_o$), the solution follows the corresponding curve that passes through that point in the direction of the arrows toward a stationary point. From Eq. (4), it is easy to observe that the value of $z_o$ is identical to the value of the specific growth rate at $x_o=0$, i.e., $z_o=(\dot{x}/x)_{x=0}$. The value of $z_o$ can therefore be established by the point where the curves cross the $(\dot{x}/x)$ axis. In Fig. 2, it is evident that the region corresponding to $z_o>\mu_{\max}$ (on the right side and above the straight line representing the LGM) is qualitatively similar to that obtained previously when $r_m<1/\delta$. However, the region corresponding to $z_o<\mu_{\max}$ can be further divided into four important regions, as follows:

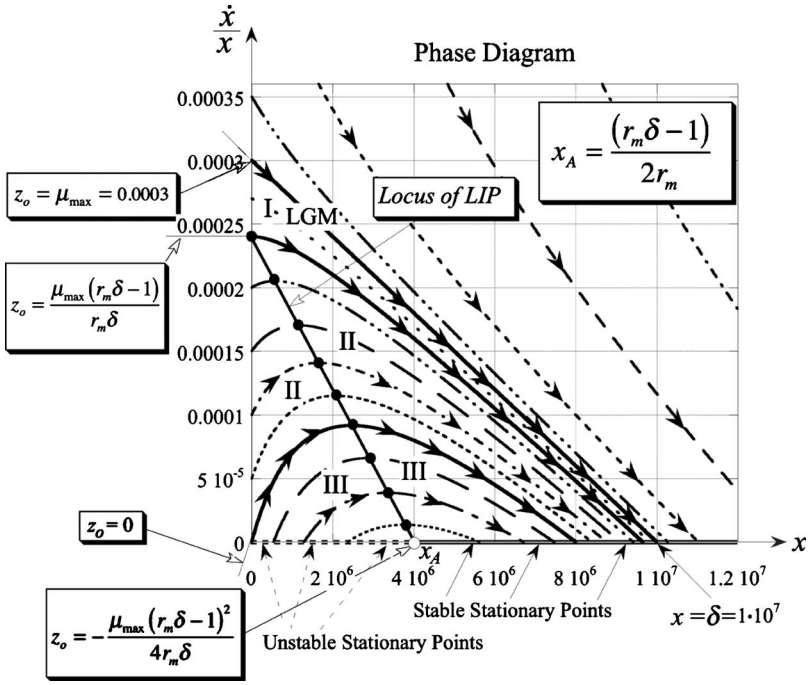



**Fig. 2 Phase diagram for the solution of monotonic growth corresponding to $r_m\delta>1$ in terms of the specific growth rate $\dot{x}/x$ versus the cell concentration $x$ for $r_m=5\times10^{-7}$ (cell/ml)$^{-1}$, $\delta=10^7$ cells/ml, and $\mu_{max}=3\times10^{-4}$ s$^{-1}$ (reproduced from Vadasz and Vadasz [12])**

(i) Region I: $\mu_{max}(r_m\delta-1)/r_m\delta<z_o<\mu_{max}$, convex curves but no LIP and no lag.
(ii) Region II: $0<z_o<\mu_{max}(r_m\delta-1)/r_m\delta$, LIP exists but no lag.
(iii) Region III: $-\mu_{max}(r_m\delta-1)^2/4r_m\delta<z_o<0$, both LIP and lag are possible.
(iv) Region IV: $z_o<-\mu_{max}(r_m\delta-1)^2/4r_m\delta$, the solution leads to extinction. (below the $x$ axis and therefore not shown in Fig. 2).

The following statements regarding the properties of the solution in each one of these regions were derived in detail by Vadasz and Vadasz [12]. In Region I, the curves are convex, but there is no possibility of a logarithmic inflection point, as the curves have no maxima in the non-negative $x$ domain. In Region II, a logarithmic inflection point exists (the maximum on the phase diagram), but no lag is possible. In Region III, both a LIP and a lag become possible. In Region IV (not shown in the figure), the solution leads always to extinction; i.e., the trivial stationary point $x_{1s}=0$ becomes the only globally stable solution. The reason for the possibility of a lag in Region III is the existence of additional positive but unstable stationary points in this region to the left of point $x_A$ represented by the dotted line on the $x$ axis in Fig. 2. When the initial conditions are sufficiently close to one of these unstable stationary points, the solution spends a relatively long time to escape from its neighborhood. This implies that if $0<\dot{x}_o \ll \mu_{max}\delta$, i.e., $\dot{x}_o$ is positive but sufficiently small and $x_o<x_A$, then a lag exists. It is shown in the next subsection that the value of $x_A$ is given by $x_A=(r_m\delta-1)/2r_m$.

The condition for existence of a LIP is that the second time derivative of the logarithm of the cell concentration should vanish at the location of the LIP, implying that $[d(\dot{x}/x)/dt]_{LIP}=0$ at the logarithmic inflection point. The latter condition is further developed by using the chain rule, leading to $[d(\dot{x}/x)/dx]_{LIP}=0$, i.e., when $(\dot{x}/x)$ has a maximum on the $(\dot{x}/x)$ versus $x$ phase diagram. By applying the latter condition to Eq. (6), i.e., undertaking the derivative with respect to $x$ of the right hand side of Eq. (6), and equating it to zero produces a quadratic algebraic equation for $x_{LIP}$ in the form

$$x_{LIP}^2 + \frac{2}{r_m}x_{LIP} + \left[\frac{1}{r_m^2} + \frac{\delta(z_o-\mu_{max})}{\mu_{max}r_m}\right] = 0 \tag{7}$$

Solving Eq. (7) for $x_{LIP}$ and accounting only for the non-negative solution (only non-negative values of $x$ are biologically meaningful) reveals the location of the LIP in the form

$$x_{LIP} = -\frac{1}{r_m} + \sqrt{\frac{(\mu_{max}-z_o)\delta}{\mu_{max}r_m}} \tag{8}$$

Further derivations (see Ref. [12] for details) reveal the locus of all logarithmic inflection points on the $(\dot{x}/x)$ versus $x$ phase diagram that lie on the straight line,

$$\mu_{LIP} = \left(\frac{\dot{x}}{x}\right)_{LIP} = \frac{\mu_{max}}{\delta}\left[\frac{(r_m\delta-1)}{r_m} - 2x_{LIP}\right] = \mu_{max}\left[1 - \frac{1}{r_m\delta} - 2\frac{x_{LIP}}{\delta}\right] \tag{9}$$

In Eq. (9), it can be observed that for $r_m\delta\gg1$ and $(x_{LIP}/\delta)\ll1$ the maximum specific growth rate $\mu_{max}$ and the specific growth rate at the logarithmic inflection point $\mu_{LIP}$ become equal, i.e., $\mu_{max}=\mu_{LIP}\equiv(\dot{x}/x)_{LIP}$. The location of the LIP as evaluated above represents the maxima of the curves on the $(\dot{x}/x)$ versus $x$ phase diagram, and therefore all logarithmic inflection points in Regions II and III lie on the straight line expressed by Eq. (9), as indicated in Fig. 2.

The first appearance of a LIP occurs when $x_{LIP}=0$ (Fig. 2), which upon substitution into Eq. (9) yields

$$\left(\frac{\dot{x}}{x}\right)_{\substack{\text{first}\\ \text{LIP}}} = (z_o)^{\text{first}}_{\text{LIP}} = \frac{\mu_{\max}(r_m\delta - 1)}{r_m\delta} \tag{10}$$

The last appearance of LIP in the first quadrant of the $(\dot{x}/x)$ versus $x$ phase diagram (i.e., were $\dot{x} \geq 0$) occurs at $x_A$ where $(\dot{x}/x)_{\text{LIP}} = 0$. By substituting the latter condition into Eq. (9) yields

$$x_A = \frac{(r_m\delta - 1)}{2r_m} \tag{11}$$

The corresponding value of $z_{o,A}$ is evaluated by substituting $x_A$ from Eq. (11) and $(\dot{x}/x)_A = 0$ in Eq. (4) to produce (see Fig. 2)

$$z_{o,A} = -\frac{\mu_{\max}(r_m\delta - 1)^2}{4r_m\delta} \tag{12}$$

The special curve representing Pearl's [3] LGM is obtained as a particular case corresponding to $z_o = \mu_{\max}$. Substituting the latter into Eq. (6) yields $\dot{x}/x = \mu_{\max}[1 - x/\delta]$, which describes a straight line on the $(\dot{x}/x)$ versus $x$ phase diagram, as presented in Fig. 2. The location of this straight line corresponding to the LGM is invariant to any changes in the value of $r_m$. Its location on the phase plane depends only on the values of $\mu_{\max}$ and $\delta$. This is an important property of the LGM, indicating that its location on the phase diagram depends on $\mu_{\max}$ and $\delta$ only.

While different lag definitions were proposed in the professional literature, e.g., Pirt [11] and Wangersky and Cunningham [24], the definition that is used here is the one that is consistent with Vadasz and Vadasz [12] findings that the lag is essentially related to the existence of unstable stationary points, i.e., the continuous collection of points represented by the dotted line on the $x$ axis to the left of $x_A$ in Fig. 2. One therefore defines the lag duration $\lambda_{\text{rd}}$ as the amount of time that elapses until the solution reaches a value, which is by a certain percentage above the corresponding unstable stationary point $x_{3s}$. This implies that $x_\lambda = bx_{3s}$, where $b > 1$ is a constant that specifies how far $x_\lambda$ is from $x_{3s}$. The lag duration, $\lambda_{\text{rd}}$, is therefore defined as the time needed for the solution to reach the value $x_\lambda$ for any predetermined value of $b$. This definition is very similar to the way one defines the time needed for a monotonic solution to reach a steady state. In most of the computations, a value of 3.5% above $x_{3s}$ was used as the lag threshold, i.e., $b = 1.035$. The location of $x_\lambda$ on the phase diagram can therefore be established for any value of $x_{3s}$ and a corresponding choice for $b > 1$.

## 4 Closed Form Analytical Solution

Equation (5) has a closed form solution that can be obtained by direct integration in terms of its corresponding stationary points. The first stationary point is the trivial one $x_{1s} = 0$. The nontrivial stationary points are obtained by solving Eq. (6) for steady state, i.e., $(\dot{x}/x) = 0$, leading to the following equation for these stationary points, $x_{2s}$ and $x_{3s}$:

$$x_s^2 - \frac{(r_m\delta - 1)}{r_m}x_s - \frac{z_o\delta}{\mu_{\max}r_m} = 0 \tag{13}$$

The roots of Eq. (13) are

$$x_{2,3s} = \frac{(r_m\delta - 1)}{2r_m}\left[1 \pm \sqrt{1 + \frac{4z_o r_m\delta}{\mu_{\max}(r_m\delta - 1)^2}}\right] \tag{14}$$

where $x_{2s} \geq x_{3s}$. While the negative stationary points are of no interest to cell growth (there is no meaning to the negative cell number), their importance here is in the sense that they affect mathematically the transient solution. The closed form solution to Eq. (5) can be presented by defining the following parameter:

$$\alpha = \frac{2r_m\delta z_o - \mu_{\max}(r_m\delta - 1)}{\mu_{\max}(r_m\delta - 1)\sqrt{1 + \dfrac{4z_o r_m\delta}{\mu_{\max}(r_m\delta - 1)^2}}} \tag{15}$$

Then, the solution can be expressed in terms of $x_{2s}, x_{3s}$ and $\alpha$ in the form

$$\ln\left[\frac{x|x - x_{3s}|^{(\alpha-1)/2}}{|x - x_{2s}|^{(\alpha+1)/2}}\right] = z_o t + c_1 \tag{16}$$

where $c_1$ is an integration constant that is related to the initial conditions by the following relationship $c_1 = \ln[x_o|x_o - x_{3s}|^{(\alpha-1)/2}/|x_o - x_{2s}|^{(\alpha+1)/2}]$. Equation (16) represents the closed form solution to Eq. (5) for monotonic growth. It is presented in an implicit form, and there is no explicit analytical expression for $x$ as a function of $t$. Nevertheless, one can use the property of the solution being monotonic and evaluate $t$ as a function of $x$ by using Eq. (16) for values of $x_o \leq x \leq x_{2s}$ when $x_{2s}$ is the stable positive stationary point or for values of $0 \leq x \leq x_o$ when $x_{1s} = 0$ is the stable stationary point. Note that when $x_{3s}$ is non-negative (the points on the $x$ axis represented as a dotted line to the left of $x_A$ in Fig. 2) it is always globally unstable as can be observed from Fig. 2. The monotonic behavior of the solution guaranties that for each value of $x$ there is one and only one value of $t$. By using this procedure, one can vary the values of $x$ within the range indicated above and obtain the corresponding values of $t$, producing therefore the numerical values needed for plotting the resulting solution of $x$ as a function of $t$.

The occurrence of the LIP in time can be evaluated by substituting Eq. (8) for $x_{\text{LIP}}$ into the closed form solution Eq. (16) in the form

$$t_{\text{LIP}} = \frac{1}{z_o}\ln\left[\left(\frac{x_{\text{LIP}}}{x_o}\right)\left(\frac{|x_{\text{LIP}} - x_{3s}|}{|x_o - x_{3s}|}\right)^{(\alpha-1)/2}\left(\frac{|x_o - x_{2s}|}{|x_{\text{LIP}} - x_{2s}|}\right)^{(\alpha+1)/2}\right] \tag{17}$$

where $x_{\text{LIP}}$ is evaluated by using Eq. (8). Since the lag duration, $\lambda$, was defined as the time needed for the solution to reach the value $x_\lambda = bx_{3s}$ for any predetermined value of $b > 1$, one can use this definition into the closed form solution Eq. (16) to evaluate the lag duration $\lambda_{\text{rd}}$ in the form

$$\lambda_{\text{rd}} = \frac{1}{z_o}\ln\left[\left(\frac{bx_{3s}}{x_o}\right)\left(\frac{(b-1)|x_{3s}|}{|x_o - x_{3s}|}\right)^{(\alpha-1)/2}\left(\frac{|x_o - x_{2s}|}{|bx_{3s} - x_{2s}|}\right)^{(\alpha+1)/2}\right] \tag{18}$$

## 5 Results and Discussion

The analytical closed form solution linked with the computational procedure described above was used by Vadasz and Vadasz [12] to produce results that demonstrate the natural occurrence of a LIP and a lag phase. The impact of the parameters $\mu_{\max}$ and $r_m$ and of the initial conditions $\dot{x}_o$ and $x_o$ on the lag duration $\lambda$ and on the LIP location in time, $t_{\text{LIP}}$, was also investigated from these solutions.

The results of the solution in the time domain showing the effect of the maximum specific growth rate on the LIP as well as on the lag, obtained by using Eq. (16) for different values of $\mu_{\max}$ ranging from $\mu_{\max} = 0.3 \times 10^{-3}$ s$^{-1}$ up to $\mu_{\max} = 2 \times 10^{-3}$ s$^{-1}$, are presented in Fig. 3 corresponding to parameter values of $r_m = 10^{-5}$ (cells/ml)$^{-1}$ and $\delta = 1.5 \times 10^8$ cells/ml. The accurate location of the lag and LIP is presented in the figure by the indicated points. As indeed expected, the smaller the maximum specific growth rate $\mu_{\max}$, the longer the lag duration $\lambda_{\text{rd}}$ and the larger the value of $t_{\text{LIP}}$. Deriving a relationship showing how the LIP and the lag depend on the maximum specific growth rate $\mu_{\max}$ becomes appealing. Baranyi and Roberts [14] indicated that the lag duration is inversely proportional to the maximum specific growth rate $\mu_{\max}$. This implies that $\lambda = c_o/\mu_{\max}$, where $c_o$ is a constant, and

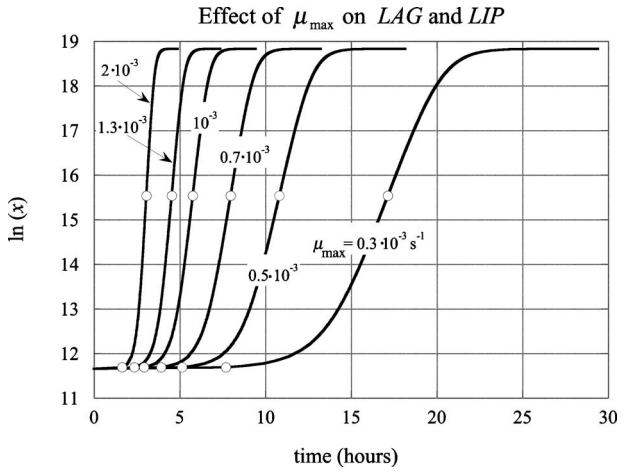
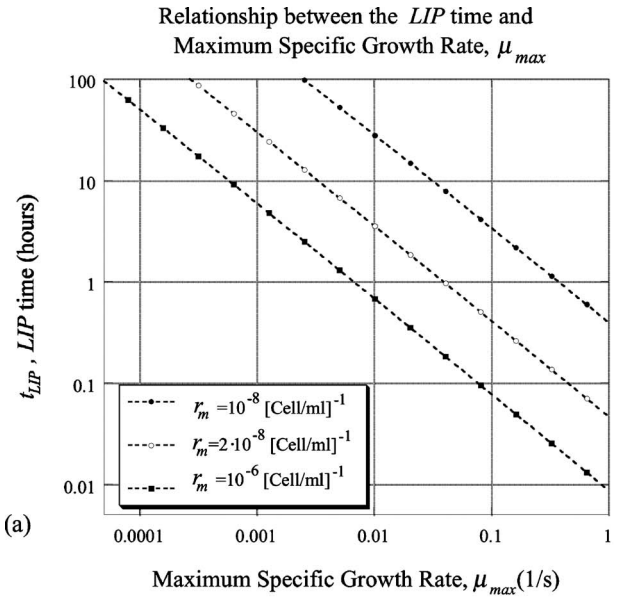



**Fig. 3 The effect of the maximum specific growth rate, $\mu_{max}$, on LIP and lag. Analytical results in the time domain for $r_m = 10^{-5}$ (cell/ml)$^{-1}$ and $\delta = 1.5\times10^8$ cells/ml, subject to initial conditions of $x_o = 1.1458\times10^5$ cells/ml, $\dot{x}_o = 7.5\times10^{-3}$ cells/(ml s), and different values of $\mu_{max}$ ranging from $\mu_{max} = 0.3\times10^{-3}$ s$^{-1}$ and up to $\mu_{max} = 2\times10^{-3}$ s$^{-1}$. (Reproduced from Vadasz and Vadasz [12].)**

after applying the log function on both sides of this equation it can be presented as $\log\lambda = \log c_o - \log\mu_{max}$, showing that such an inversely proportional relationship should appear as a straight line on a log-log scale. Vadasz and Vadasz [12] tested the neoclassical model solution to reveal whether it captures this fact, as well as the relationship between the LIP and the maximum specific growth rate. They used a variation in values of $\mu_{max}$ over three orders of magnitude and evaluated the LIP time $t_{LIP}$, producing the results presented on a log-log scale in Fig. 4(*a*) for three different values of $r_m$. The results indicate that the LIP time $t_{LIP}$ is accurately related as inversely proportional to the maximum specific growth $\mu_{max}$. The corresponding results for the lag duration using the present lag definition, presented in Fig. 4(*b*) for three different values of $r_m$, show also that the lag duration is inversely proportional to the maximum specific growth rate $\mu_{max}$. The latter is an approximation as observed by the very slight deviations of the dotted line (representing the model's results) from the straight line (continuous) in Fig. 4(*b*). The impact of the initial growth rate at inoculation on this inversely proportional relationship between the LIP time $t_{LIP}$, lag duration $\lambda_{rd}$, and maximum specific growth rate $\mu_{max}$ is presented in Fig. 5, where it is evident that this relationship holds when $z_o \ll \mu_{max}$, but when $z_o$ is closer to $\mu_{max}$ it breaks down. The effect of the initial growth rate, $\dot{x}_o$, on the solution is presented in Fig. 6, corresponding to an initial cell concentration of $x_o = 1.1458\times10^5$ cells/ml and to the following parameter values $r_m = 10^{-5}$ (cells/ml)$^{-1}$, $\mu_{max} = 3\times10^{-4}$ s$^{-1}$, and $\delta = 1.5\times10^8$ cells/ml. The initial growth rate used varied from $\dot{x}_o = 7.5\times10^{-4}$ cells/(ml s) to $\dot{x}_o = 50$ cells/(ml s). The anticipation that the lag duration is reduced as the initial growth rate increases is confirmed by these results. Furthermore, only five of the eight curves produce a lag. The curves corresponding to initial growth rates larger than or equal to 1 cell/(ml s), i.e., $\dot{x}_o > 1$ cell/(ml s), have no lag, while the last curve corresponding to $\dot{x}_o = 50$ cells/(ml s) shows no LIP as well. The effect of the initial cell concentration, $x_o$, on the solution as evaluated by using the closed form analytical solution (16) is presented on a phase diagram in terms of $(\dot{x}/x)$ versus $x$ in Fig. 7, corresponding to an initial growth rate of $\dot{x}_o = 1$ cell/(ml s) and to the following parameter values: $r_m = 5\times10^{-4}$ (cells/ml)$^{-1}$, $\mu_{max} = 3\times10^{-4}$ s$^{-1}$, and $\delta = 1.5\times10^8$ cells/ml. The initial cell concentration used varied from $x_o = 1\times10^7$ cells/ml to $x_o = 6\times10^7$ cells/ml. The lag



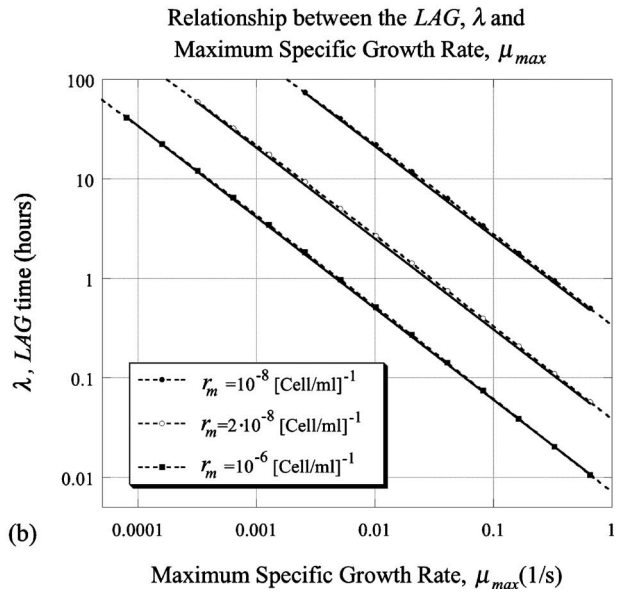
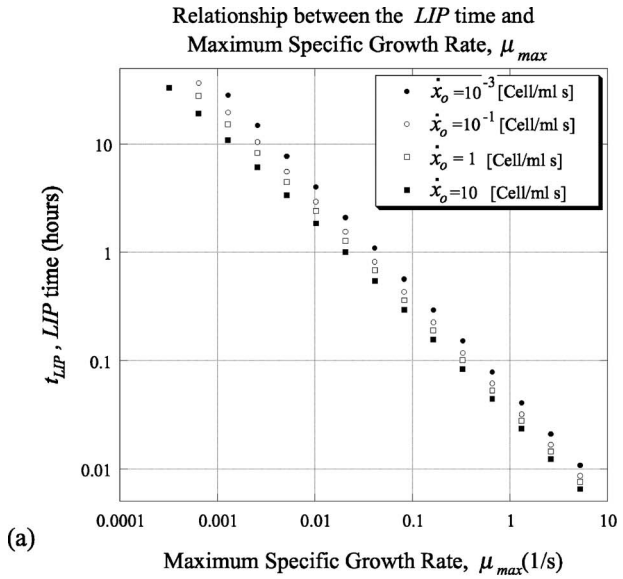




**Fig. 4 (*a*) The relationship between the maximum specific growth rate $\mu_{max}$ and LIP for three different values of $r_m$. (*b*) The relationship between the maximum specific growth rate $\mu_{max}$ and lag for three different values of $r_m$. (Reproduced from Vadasz and Vadasz [12].)**

and LIP location on this phase diagram, as evaluated by the solution using Eqs. (17) and (18), is revealed by the black point markers. The neoclassical definition of the lag duration is different from the classical definition primarily because the former is a direct consequence of the properties of the neoclassical model, while the classical definition is an arbitrary geometric definition. The objective of this section is to show that our definition of lag is perfectly consistent with the classical one. The classical definition of lag is based upon the geometry of the growth curve presented in Fig. 1. It is essentially the value of time at the intersection between the tangent line to the growth curve at $(t_{LIP}, y_{LIP})$ with the line $y = y_o$, where by definition $y = \ln(x)$. The equation of the tangent line to the growth curve at $(t_{LIP}, y_{LIP})$ is obtained in the form $y = y_o + \dot{y}_{LIP}(t-\lambda) = y_o + \mu_{LIP}(t-\lambda)$, or alternatively in terms of $y_{LIP}$



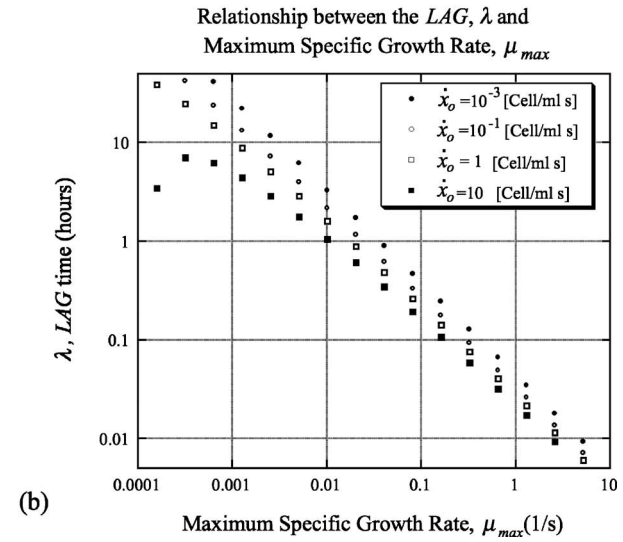




**Fig. 5 (*a*) The relationship between the maximum specific growth rate $\mu_{max}$ and LIP for four different values of the initial growth rate $\dot{x}_o$. (*b*) The relationship between the maximum specific growth rate $\mu_{max}$ and lag for four different values of the initial growth rate $\dot{x}_o$. (Reproduced from Vadasz and Vadasz [12].)**

and $t_{LIP}$ in the form $y=y_{LIP}+\dot{y}_{LIP}(t-t_{LIP})=y_{LIP}+\mu_{LIP}(t-t_{LIP})$. From the first of these equations, one can obtain the classical definition of lag in the form

$$\lambda_{cl}=t-\frac{(y-y_o)}{\dot{y}_{LIP}}=t-\frac{1}{\mu_{LIP}}\ln(x/x_o) \tag{19}$$

By using now the definition of $\mu_{LIP}\equiv\dot{y}_{LIP}\equiv(\dot{x}/x)_{LIP}$ corresponding to the neoclassical model from Eq. (9), we obtain the classical lag definition in the form

$$\lambda_{cl}=t_{LIP}-\frac{1}{\mu_{max}\left[1+\frac{1}{r_m\delta}-2\sqrt{\frac{(1-z_o/\mu_{max})}{r_m\delta}}\right]}\ln\left(\frac{x_{LIP}}{x_o}\right) \tag{20}$$

In order to check that the neoclassical definition of lag is consistent with the classical one, we evaluated accurately the neoclassical lag duration according to the accurate Eq. (18) and the accurate stationary point solution (14). The resulting values were compared with the classical definition of the lag evaluated by using Eq. (19) and are presented in Fig. 8. The wide range of variation in the lag was accomplished by varying the value of $\mu_{max}$ over six orders of magnitude from $\mu_{max}=2\times10^{-4}$ s$^{-1}$ up to $\mu_{max}=1\times10^{2}$ s$^{-1}$. The lag threshold used in Eq. (18) for its neoclassical definition was taken at about 1% above $y_{3s}=\ln(x_{3s})$, which is equivalent to 17% above $x_{3s}$, or $b=1.17$. Figure 8 shows that the neoclassical definition of lag overlaps with its classical

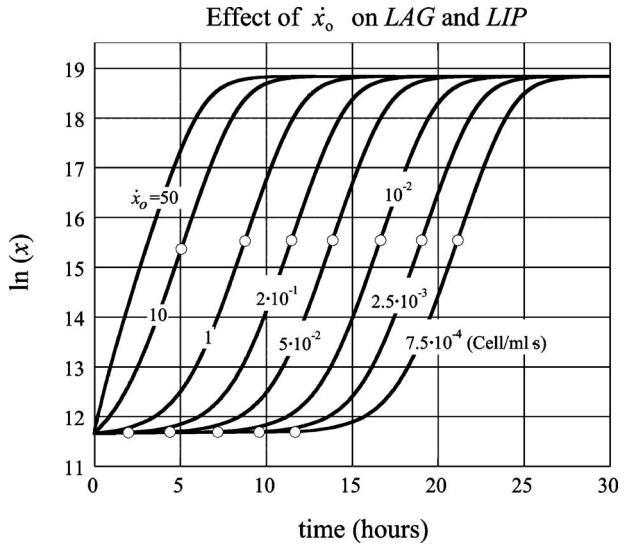



**Fig. 6 The effect of the initial growth rate $\dot{x}_o$ on LIP and lag. Analytical results for $r_m=10^{-5}$ (cell/ml)$^{-1}$, $\mu_{max}=3\times10^{-4}$ s$^{-1}$, and $\delta=1.5\times10^{8}$ cell/ml, subject to the initial condition of $x_o=1.1458\times10^{5}$ cell/ml, and for different values of $\dot{x}_o$ ranging from $\dot{x}_o=7.5\times10^{-4}$ cell/(ml s) up to $\dot{x}_o=50$ cells/(ml s). (Reproduced from Vadasz and Vadasz [12].)**

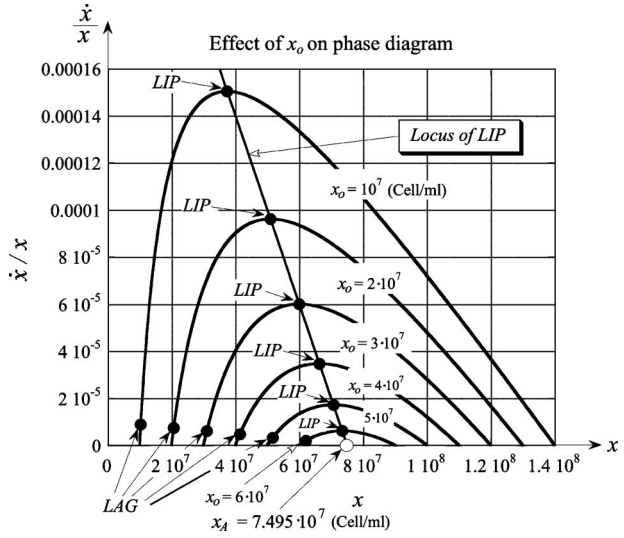



**Fig. 7 The effect of the initial cell concentration $x_o$ on LIP and lag on a phase diagram in terms of the specific growth rate $\dot{x}/x$ versus the cell concentration $x$. Analytical results for $r_m=5\times10^{-4}$ (cell/ml)$^{-1}$, $\mu_{max}=3\times10^{-4}$ s$^{-1}$, and $\delta=1.5\times10^{8}$ cells/ml, subject to the initial growth rate condition of $\dot{x}_o=1$ cell/(ml s), and for different values of $x_o$ ranging from $x_o=1\times10^{7}$ cells/ml up to $x_o=6\times10^{7}$ cells/ml. (Reproduced from Vadasz and Vadasz [12].)**

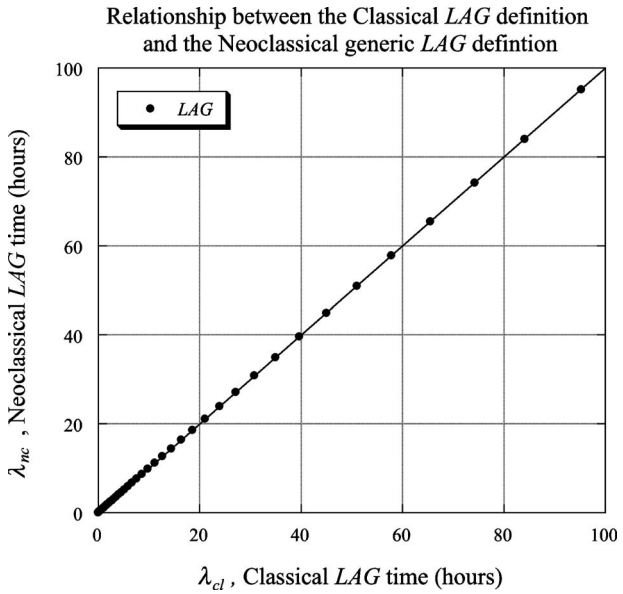



**Fig. 8 Comparison of the neoclassical values of the lag duration with the classical lag duration values obtained by a variation in $\mu_{max}$ over six orders of magnitude from $\mu_{max}=2\times10^{-4}$ $s^{-1}$ up to $\mu_{max}=1\times10^{2}$ $s^{-1}$. The value of $b$ for evaluating the neoclassical lag was $b=1.17$, i.e., about 1% above $y_{3s}=\ln(x_{3s})$, or 17% above the unstable stationary points $x_{3s}$. The markers represent the evaluated results. (Reproduced from Vadasz and Vadasz [12].)**

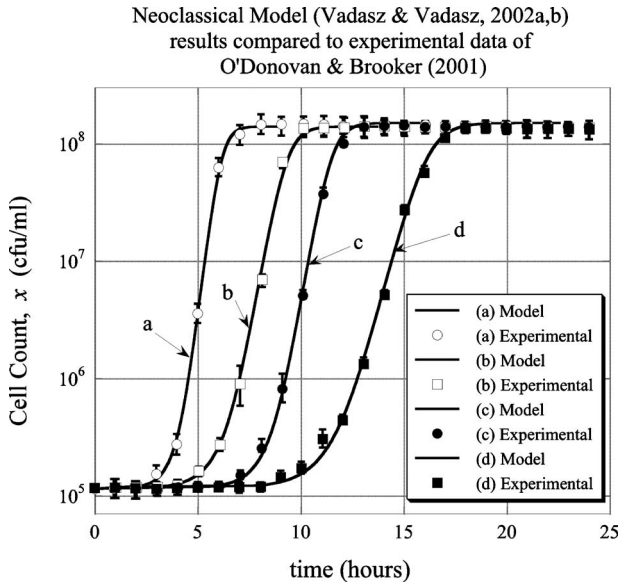



**Fig. 9 Comparison of the neoclassical model analytical solution for monotonic growth based on Vadasz and Vadasz [15,16] with the experimental data based on O'Donovan and Brooker [25] (here redrawn from published data). The initial cell concentration is $x_o=1.1458\times10^{5}$ (cfu/ml), and the estimated parameter values are as presented by Vadasz and Vadasz [12]. (Reproduced from Vadasz and Vadasz [12].)**

counterpart as their results lie on a straight line inclined at a 45 deg angle, identifying a complete overlap.

While the previous sections introduced results that show evidently that the neoclassical model captures all qualitative features that have been revealed experimentally, the most important test of this model is the direct comparison of its results to experimental data. The data set used was from O'Donovan and Brooker [25], showing the growth of *Streptococcus gallolyticus* subject to different concentrations of acacia condensed tannins (ACT), which affect substantially the lag phase. The latter represents microbial growth curves subject to different levels of stressed environmental conditions. The experimental data presented in Fig. 2 of O'Donovan and Brooker [25] were digitized and used to test the model results. A trial and error approach for parameter estimation was applied to each data set separately and yielded the four different curves for each of the data sets presented by O'Donovan and Brooker [25] in their Fig. 2. Each curve is linked to three different parameter values and initial growth rates listed in Ref. [12]. The model results are presented in Fig. 9 in comparison with the digitized data from O'Donovan and Brooker [25] for each of the four different curves. It is evident from the results that a very good fit was obtained by using only three parameter values that control the monotonic version of the neoclassical model (Vadasz and Vadasz [15,16]) and without using a systematic solution to the inverse problem for parameter estimation.

It is important to emphasize that all four data sets presented by O'Donovan and Brooker [25] describe convex curves when plotted on the phase diagram in terms of $\dot{x}/x$ versus $x$, and therefore models such as Gompertz [26] cannot capture them correctly because the Gompertz [26] model produces concave curves on the phase diagram.

In addition, the neoclassical model was shown (see Fig. 4(*b*)) to capture the inversely proportional relationship between the lag duration and the maximum specific growth rate as part of the solution as well as producing a complete fit with the classical definition of the lag. The Baranyi and Roberts [14] model exhibits very similar but not identical features as the neoclassical model presented here, as shown in Ref. [13].

## 6 Conclusions

The variety of models that are currently being used in predictive microbiology or "microbial ecology" aiming at reproducing the growth curve of microorganisms cannot capture a lag phase naturally. The only two models that capture qualitative features, which appear in experiments and naturally capture the lag phase, are the Baranyi and Roberts [14] model and the neoclassical model (Vadasz and Vadasz [12] and Vadasz and Vadasz [15,16]). In both models, the lag is associated with unstable stationary points and linked to the general behavior of solutions having initial conditions in the neighborhood of unstable stationary points. This general behavior consists of solutions departing very slowly away from this neighborhood. Another important feature common to these two models is that they include explicitly the metabolic mass transfer process in the model formulation.


## Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. CTS-0500466. The authors wish to thank the NSF for the funding support.


## Nomenclature

**Latin Symbols**

$M$ = total viable biomass
$m$ = average viable biomass
$\dot{M}$ = rate of total viable biomass production
$\dot{m}$ = rate of average viable biomass production
$r_m$ = ratio between the viable biomass constant coefficients
$t$ = time
$t_{LIP}$ = time associated with the LIP
$x$ = population number/cell concentration

$\dot{x}$ = absolute growth rate
$\dot{x}/x$ = specific (per capita) growth rate
$y$ = logarithm of the cell concentration, equals $\ln x$
$z$ = metabolic variable related to the nutrient (resource) consumption/utilization rate

**Greek Symbols**

$\delta$ = carrying capacity of the environment
$\mu_{max}$ = maximum specific growth rate
$\lambda$ = lag duration

**Subscripts**

cl = classical
rd = redefined
$o$ = related to initial conditions

## References


[1] Malthus, T. R., 1798, *An Essay on the Principle of Population*, Penguin, Harmondsworth, UK.
[2] Verhulst, P. F., 1838, "Notice sur la loi que la population suit dans son accroissement," Corr. Math. et Phys. Publ. par A. Quetelet. T., **10**, pp. 113–121.
[3] Pearl, R., 1927, "The Growth of Populations," Q. Rev. Biol., **2**(4), pp. 532–548.
[4] Vadasz, P., and Vadasz, A. S., 2008, "Microbial Models," *Ecological Models* (Encyclopedia of Ecology, Vol. 3), S. E. Jørgensen and B. D. Fath, eds., Elsevier, Oxford, pp. 2369–2389.
[5] Baty, F., and Delignette-Muller, M. L., 2004, "Estimating the Bacterial Lag Time: Which Model, Which Precision," Int. J. Food Microbiol., **91**, pp. 261–277.
[6] Augustin, J. C., and Carlier, V., 2000, "Mathematical Modelling of the Growth Rate and Lag Time for *Listeria monocytogenes*," Int. J. Food Microbiol., **56**, pp. 29–51.
[7] Baranyi, J., 2002, "Stochastic Modelling of Bacterial Lag Phase," Int. J. Food Microbiol., **73**, pp. 203–206.
[8] Vadasz, A. S., Vadasz, P., Abashar, M. E., and Gupthar, A. S., 2001, "Recovery of an Oscillatory Mode of Batch Yeast Growth in Water for a Pure Culture," Int. J. Food Microbiol., **71**(2–3), pp. 219–234.
[9] Vadasz, A. S., Vadasz, P., Abashar, M. E., and Gupthar, A. S., 2002, "Theoretical and Experimental Recovery of Oscillations During Batch Growth of a Mixed Culture of Yeast in Water," World J. Microbiol. Biotechnol., **18**(3), pp. 239–246.
[10] Vadasz, A. S., Vadasz, P., Gupthar, A. S., and Abashar, M. E., 2002, "Theoretical and Experimental Recovery of Oscillations During Batch Yeast Growth in a Pure Culture Subject to Nutritional Stress," J. Mech. Med. Biol., **2**(2), pp. 147–163.
[11] Pirt, S. J., 1975, "Growth Lag," *Principles of Microbe and Cell Cultivation*, Blackwell, London.
[12] Vadasz, P., and Vadasz, A. S., 2005, "Predictive Modeling of Microorganisms: LAG and LIP in Monotonic Growth," Int. J. Food Microbiol., **102**, pp. 257–275.
[13] Vadasz, P., and Vadasz, A. S., 2007, "Biological Implications From an Autonomous Version of Baranyi and Roberts Growth Model," Int. J. Food Microbiol., **114**, pp. 357–365.
[14] Baranyi, J., and Roberts, T. A., 1994, "A Dynamic Approach to Predicting Bacterial Growth in Food," Int. J. Food Microbiol., **23**, pp. 277–294.
[15] Vadasz, P., and Vadasz, A. S., 2002, "The Neoclassical Theory of Population Dynamics in Spatially Homogeneous Environments—Part I: Derivation of Universal Laws and Monotonic Growth," Physica A, **309**(3–4), pp. 329–359.
[16] Vadasz, P., and Vadasz, A. S., 2002, "The Neoclassical Theory of Population Dynamics in Spatially Homogeneous Environments—Part II: Non-Monotonic Dynamics, Overshooting and Oscillations," Physica A, **309**(3–4), pp. 360–380.
[17] Maier, R. M., 2000, "Bacterial Growth," *Environmental Microbiology*, R. M. Maier, I. L. Pepper, and C. P. Gerba, eds., Academic, New York, pp. 43–59.
[18] Swinnen, I. A. M., Bernaerts, K., Dens, E. J. J., Geeraerd, A. H., and Van Impe, J. F., 2004, "Predictive Modeling of the Microbial Lag Phase: A Review," Int. J. Food Microbiol., **94**, pp. 137–159.
[19] May, M. R., 1973, "Time-Delay Versus Stability in Population Models With Two and Three Trophic Levels," Ecology, **54**, pp. 315–325.
[20] May, M. R., 1978, "Mathematical Aspects of the Dynamics of Animal Populations," S. A. Levin, ed., *Studies in Mathematical Biology—Part II: Populations and Communities* (Studies in Mathematics Vol. 16), The Mathematical Association of America, Washington, D.C., pp. 317–366.
[21] May, M. R., 1981, "Models for Single Populations," *Theoretical Ecology*, Blackwell Scientific, Oxford, pp. 5–29.
[22] Vadasz, P., and Vadasz, A. S., 2010, "On the Distinction Between Lag and Delay in Population Growth," Microb. Ecol., **59**(2), pp. 233–245.
[23] Ginzburg, L. R., 1986, "The Theory of Population Dynamics: I. Back to First Principles," J. Theor. Biol., **122**, pp. 385–399.
[24] Wangersky, P. J., and Cunningham, W. J., 1957, "Time Lag in Population Models," Cold Spring Harbor Symp. Quant. Biol., **22**, pp. 329–338.
[25] O'Donovan, L., and Brooker, J. D., 2001, "Effect of Hydrolysable and Condensed Tannins on Growth, Morphology and Metabolism of *Streptococcus gallolyticus* (*S. caprinus*) and *Streptococcus bovis*," Microbiology, **147**, pp. 1025–1033.
[26] Gompertz, B., 1825, "On the Nature of the Function Expressive of the Law of Human Mortality, and a New Mode of Determining the Value of Life Contingencies," Philos. Trans. R. Soc. London, **115**, pp. 513–583.

# Human Eye Response to Thermal Disturbances

**Maryam Shafahi**

**Kambiz Vafai**[1]
e-mail: vafai@engr.ucr.edu

Department of Mechanical Engineering,
University of California Riverside,
Riverside, CA 92521

*Human eye is one of the most sensitive parts of the body when exposed to a thermal heat flux. Since there is no barrier (such as skin) to protect the eye against the absorption of an external thermal wave, the external flux can readily interact with cornea. The modeling of heat transport through the human eye has been the subject of interest for years, but the application of a porous media model in this field is new. In this study, a comprehensive thermal analysis has been performed on the eye. The iris/sclera section of the eye is modeled as a porous medium. The primary sections of the eye, i.e., cornea, anterior chamber, posterior chamber, iris/sclera, lens, and vitreous are considered in our analysis utilizing a two-dimensional finite element simulation. Four different models are utilized to evaluate the eye thermal response to external and internal disturbances. Results are shown in terms of temperature profiles along the pupillary axis. Effects of extreme ambient conditions, blood temperature, blood convection coefficient, ambient temperature, sclera porosity, and perfusion rate on different regions of the eye are investigated. Furthermore, the role of primary thermal transport mechanisms on the eye subject to different conditions is analyzed.* [DOI: 10.1115/1.4002360]

*Keywords: eye thermal modeling, eye thermal disturbance, bioheat, biological tissue, porous media, aqueous humor*

## 1 Introduction

Thermal disturbances are more pronounced in the eye due to an insufficient blood flow circulation and lack of skin as a protecting layer. Lack of blood flow in the interior part of the eye makes it more vulnerable compared with other organs even in the case of weak thermal interactions.

Although recent noninvasive methods for temperature measurement of the eye, such as infrared (IR), are more convenient in comparison to the contact thermometry; only corneal temperature can be obtained using these methods. In biological systems, which are not amenable to direct investigation, such as human eye, computational modeling is the preferred tool to represent the transport phenomena [1–4]. Thermal modeling of the eye is important as it can provide one with a tool to investigate the effect of external heat sources as well as in predicting the abnormalities within the eye. Hot or cold weather can create a thermal load on the eye. Moreover, in order to optimize laser therapy or surgery in ophthalmology, it is essential to have a better understanding of the thermal response in different sections of the eye for an imposed heat flux.

In previous thermal models of the eye, heat transfer in different parts was primarily modeled using conduction and in a few cases by natural convection [2–12]. Ooi and Ng [2] studied the effect of aqueous humor (AH) hydrodynamics on the heat transfer within the eye. They also investigated the effect of different variables on the thermal modeling of the eye [3]. Chua et al. [5] modeled the impact of the aging of the human eye, thermal conductivity of the lens and a radiant flux on the temperature profile.

Lagendijk [6] performed experiments on the normal and heated rabbit eye and developed a conduction model to achieve the temperature distribution in human and rabbit eyes. Scott [8] utilized a 2D finite element method to obtain the temperature profile based on heat conduction in different sections of the eye. Effect of lens thermal conductivity, evaporation from cornea surface, blood flow in choroid, ambient temperature, ambient convection coefficient and blood temperature on the eye temperature distribution was studied based on a conduction model. Using a 2D model, Scott [9] also calculated the temperature change in intra-ocular media subject to an infrared radiation. Kumar et al. [10] developed a model to study the buoyancy effect on the convective motion for different geometrical configurations of a rabbit eye. Flyckt et al. [11] performed a comprehensive study on the cooling process of the eye by blood flow circulation. Amara [12] presented a conduction model to study the interaction between a laser flux and ocular media numerically. Jacobs [13] reviewed the multiscale hierarchical system of the eye considering the appropriate approaches to be taken to understand the complex properties and functions of the ocular tissues. Mapstone [14,15] recorded the cornea temperature subject to different ambient temperatures.

Although porous media models have been used in biomedical research [16–25], the previous investigations on the human eye did not utilize porous media modeling in representing the thermal transport. The first such attempt was done by the current authors at an ASME Heat Transfer Conference [1]. Porous media models have been utilized to study the characteristics of the bifurcation airflow and mass transfer within a lung [17], countercurrent bioheat transfer between terminal arteries and veins in the circulatory system [18], bioheat transport in biological media considering local thermal nonequilibruim [19,20], and arterial transport incorporating multi porous media layers [21,22].

The present work utilizes four pertinent thermal models to analyze heat transfer in six primary sections of the human eye. In model I, thermal transport occurs by conduction in all the domains. Model II accounts for natural convection in the anterior chamber; in model III iris/sclera is considered as a porous medium and model IV accounts for the existence of a porous medium and natural convection in anterior chamber. To the best of authors' knowledge, this is the first use of porous media modeling in analyzing the eye thermal characteristics and also the first comprehensive comparison among the pertinent thermal transport mechanisms within different regions of the eye.

## 2 Modeling and Formulation

**2.1 Anatomy of the Eye.** The eye is assumed to be a sphere with a given diameter. Its geometrical configuration such as thickness of the cornea, anterior chamber, lens, vitreous and sclera are

[1]Corresponding author.

Manuscript received March 25, 2010; final manuscript received April 1, 2010; published online September 30, 2010. Assoc. Editor: Andrey Kuznetsov.

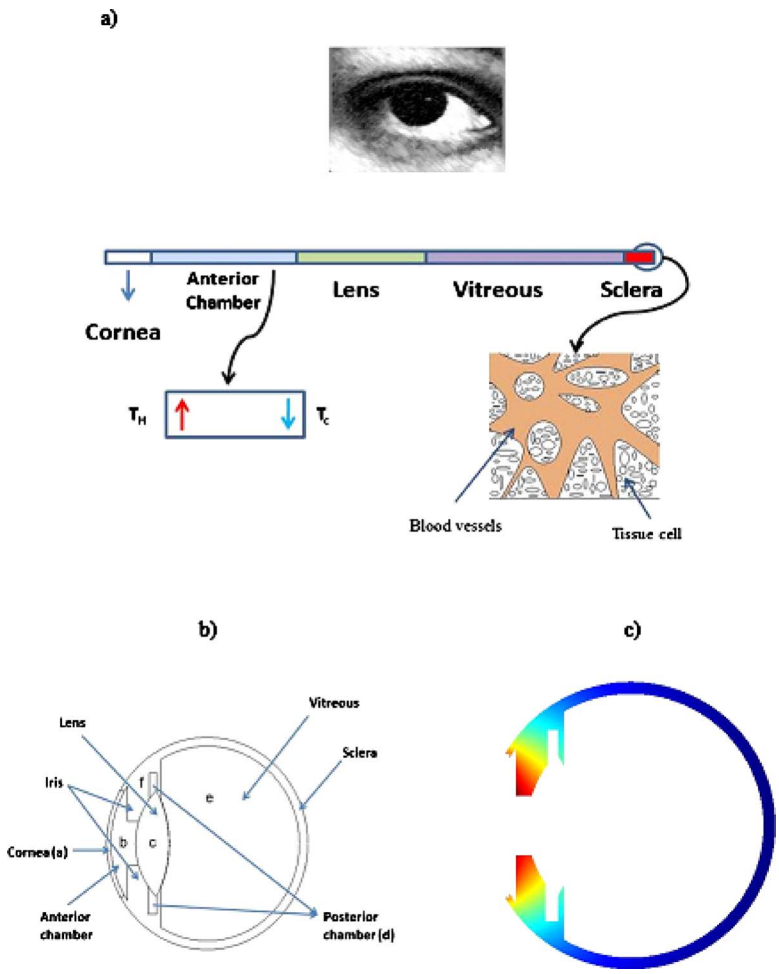



**Fig. 1 Schematic of different primary sections of the eye along the pupillary axis: (*a*) different regions of the eye along the pupillary axis with a display of special attributes, (*b*) display of the primary regions within the eyeball, and (*c*) display of the blood/tissue interaction within Iris/sclera domain**

based on the physiological data. There is blood flow in the iris/sclera part, which plays a crucial rule to adjust the eye temperature with the rest of the body. Different regions of the eye along the papillary axis with a display of special attributes as well as the two-dimensional eye configuration and display of the blood/tissue interaction within Iris/sclera domain are shown in Fig. 1.

**2.2 Model I.** In this model, conduction is assumed to be the dominant heat transfer mechanism in all of the subdomains of the eye. Each subdomain has its own thermal properties. The governing equation is as follows:

$$\rho_i c_i \frac{\partial T_i}{\partial t} = \nabla \cdot (k_i \nabla T_i) \quad i = a,b,c,d,e,f \tag{1}$$

where $i$ denotes each subdomain in the eyeball. As shown in Fig. 1, $a$, $b$, $c$, $d$, $e$, and $f$ refer to cornea, anterior chamber, lens, posterior chamber, vitreous and iris/sclera, respectively.

**2.3 Model II.** Model II accounts for the existence of AH in the anterior chamber. The energy transfer through this part occurs by both conduction and natural convection. The governing equations for five domains: cornea, posterior chamber, lens and vitreous regions are the same as that given in Eq. (1). The anterior chamber heat transfer is modeled as follows:

Continuity

$$\nabla \cdot \mathrm{v}_i = 0; \quad i = b \tag{2}$$

Momentum

$$\rho_i \frac{\partial \mathrm{v}_i}{\partial t} + \rho_i \mathrm{v}_i \nabla \cdot \mathrm{v}_i = -\nabla p_i + \rho_i g \beta_i (T_i - T_{\mathrm{ref}}); \quad i = b \tag{3}$$

The effect of buoyancy term is taken into account using the Boussinesq approximation, where $\beta$ is the volume expansion coefficient and $T_{\mathrm{ref}}$ is the reference temperature.

The energy equation is as follows:

$$\rho_i c_i \frac{\partial T_i}{\partial t} - \nabla \cdot (k_i \nabla T_i) = -\rho c_i \mathrm{v}_i \,.\, \nabla T_i; \quad i = b \tag{4}$$

where the viscous dissipation effect is neglected.

**2.4 Model III.** Blood flow in the sclera/iris keeps the eye temperature close to the other body organs. This part is modeled as a porous medium while incorporating blood circulation through the tissue. Assuming a local thermal equilibrium between the blood and the tissue and accounting for the blood perfusion term, a modified Pennes equation is used in this work [16,21].

$$(1-\varphi_i)\rho_i c_i \frac{\partial T_i}{\partial t} = \nabla \cdot ((1-\varphi) k_i \nabla T_i) + \rho_b c_b \omega (T_{bl} - T_i); \quad i = f \tag{5}$$

**Table 1 Summary of the pertinent equations utilized in the four primary eye thermal models**

| | Governing equations |
|---|---|
| Model I (conduction) | $\rho_i c_i \frac{\partial T_i}{\partial t} = \nabla \cdot (k_i \nabla T_i); \quad i = a,b,c,d,e,f$ |
| Model II (conduction+natural convection) | $\rho_i c_i \frac{\partial T_i}{\partial t} = \nabla \cdot (k_i \nabla T_i); \quad i = a,c,d,e,f$ |
| | $\nabla \cdot \mathrm{v}_i = 0; \quad i = b$ |
| | $\rho_i \frac{\partial \mathrm{v}_i}{\partial t} + \rho_i \mathrm{v}_i \nabla \cdot \mathrm{v}_i = -\nabla p_i + \rho_i g \beta_i (T_i - T_{ref}); \quad i = b$ |
| | $\rho_i c_i \frac{\partial T_i}{\partial t} - \nabla \cdot (k_i \nabla T_i) = -\rho c_i \mathrm{v}_i . \nabla T_i; \quad i = b$ |
| Model III (conduction+porous media model) | $\rho_i c_i \frac{\partial T_i}{\partial t} = \nabla \cdot (k_i \nabla T_i); \quad i = a,b,c,d,e$ |
| | $(1-\varphi_i)\rho_i c_i \frac{\partial T_i}{\partial t} = \nabla \cdot ((1-\varphi) k_i \nabla T_i) + \rho_{bl} c_{bl} \omega (T_{bl} - T_i); \quad i = f$ |
| Model IV (conduction+natural convection+porous media model) | $\rho_i c_i \frac{\partial T_i}{\partial t} = \nabla \cdot (k_i \nabla T_i); \quad i = a,c,d,e$ |
| | $\nabla \cdot \mathrm{v}_i = 0; \quad i = b$ |
| | $\rho_i \frac{\partial \mathrm{v}_i}{\partial t} + \rho_i \mathrm{v}_i \nabla \cdot \mathrm{v}_i = -\nabla p_i + \rho_i g \beta_i (T_i - T_{ref}); \quad i = b$ |
| | $\rho_i c_i \frac{\partial T_i}{\partial t} - \nabla \cdot (k_i \nabla T_i) = -\rho c_i \mathrm{v}_i . \nabla T_i; \quad i = b$ |
| | $(1-\varphi_i)\rho_i c_i \frac{\partial T_i}{\partial t} = \nabla \cdot ((1-\varphi) k_i \nabla T_i) + \rho_{bl} c_{bl} \omega (T_{bl} - T_i); \quad i = f$ |

**2.5 Model IV.** In this model, the natural convection in anterior chamber and blood circulation in the iris/sclera as a porous medium are both accounted for. The governing equations are the same as model III in all the eye subdomains except anterior chamber, which follow the natural convection equations (Eqs. (2)–(4)).

**2.6 Boundary Conditions.** Radiation, evaporation and convection are accounted for at the cornea surface and the boundary condition for all the models on the cornea surface can be presented as follows:

$$n . (-k_i \nabla T_i) = E + h_{\mathrm{am}}(T_i - T_{\mathrm{am}}) + \sigma\varepsilon(T_i^4 - T_{\mathrm{am}}^4) \quad \text{on} \quad \Omega_i \quad i = a \tag{6}$$

where $\Omega_i$ is the external surface area corresponding to section $i$, $E$ is the tear evaporation heat loss, and $T_{\mathrm{am}}$ and $h_{\mathrm{am}}$ are ambient temperature and convection coefficient, respectively.

At the sclera surface the thermal exchange between the eye ball and blood flow occurs through convection. The boundary condition at this interface can be presented as

$$n \cdot (k_i \nabla T_i) = h_{bl}(T_{bl} - T_i) \quad \text{on} \quad \Omega_i \quad i = f \tag{7}$$

Subdomains properties used in our simulations are based on the physiological data for the eye. A summary of the models utilized in the current work are given in Table 1. Different domain sizes and physical properties are taken from [2].

**2.7 Numerical Solution.** The domain is discretized using triangular elements with the Lagrange quadratic shape functions. The set of partial differential equations along with their related boundary conditions are coupled and are solved numerically by the finite element method using the commercial code COMSOL multiphysics program. The system of algebraic equations is solved with the Unsymmetric Multifrontal Method (UMFPACK) solver. The mesh independency has been verified by progressively increasing the number of elements insuring that the results are invariant. For example, increasing the number of elements from 5012 to 20,048 produced less than 1% change in the results.

## 3 Results and Discussion

The thermal analysis of the eye was carried out by utilizing four primary models. The eye response to various internal and external disturbances, such as fluctuations in blood temperature, blood convection coefficient, ambient temperature, ambient convection coefficient, extreme ambient conditions, iris/sclera porosity, and perfusion rate was investigated.

**3.1 Comparison With Experimental Data.** Models I–IV results are compared with the experimental data from Mapstone [14,15] and Lagrndijk [6] in Figs. 2 and 3. Lagendijk had assumed that the rabbit body temperature was 38.8°C and obtained the temperature of cornea, behind the lens, retina and rectal (within the sclera region) for a normal and heated rabbit eye. As Fig. 2

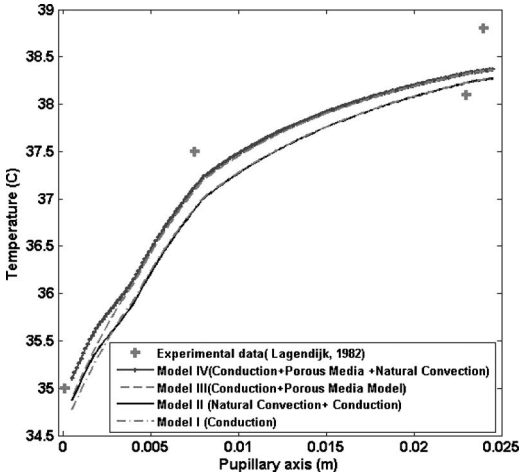



**Fig. 2 Comparison of the analyzed models with Lagendijk's experimental data [6]; $h_{\mathrm{am}}=20$ W/m$^2$ K and $T_{\mathrm{am}}=25$°C**

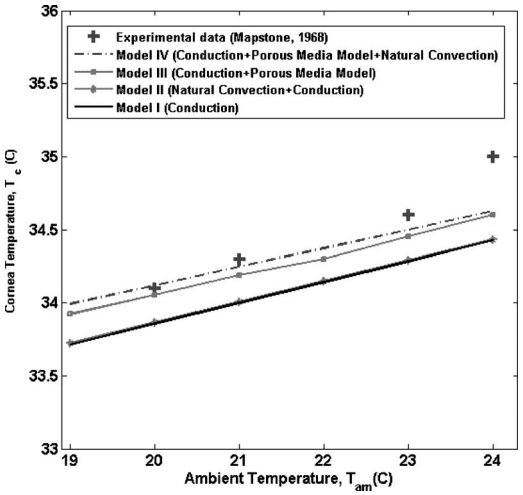



**Fig. 3 Comparison of the analyzed models with Mapstone's experimental data [14]: $h_{am}$=10 W/m$^2$ K and E=40 W/m$^2$**

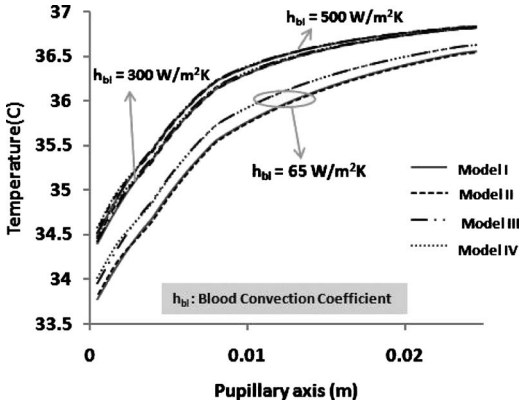



**Fig. 5 Effect of blood convection coefficient on the pupillary axis temperature profile for the analyzed models; $h_{am}$ =10 W/m$^2$ K and $T_{am}$=25°C**

shows for a normal eye recorded temperatures are better estimated by porous media models. Overall, the two profiles provided by porous media models are in better agreement with the experimental data for the rabbit eye. Figure 3 shows the cornea temperature corresponding to different ambient temperatures. In this figure, model IV provides a very good match with the experimental observations. It should be noted that the temperature profiles from models I and II are approximately the same.

**3.2 Effect of the Blood Temperature and Post Sclera Conditions.** Blood circulation plays a crucial role in adjusting the eye temperature subject to a thermal load. In models I and II, blood circulation is taken into account only through the iris/sclera boundary conditions. In models III and IV, in addition to this boundary condition, the blood circulation is also directly taken into account through the perfusion term within the governing equations. This more closely mimics the physiological conditions. As can be seen in Fig. 4, models I and II predictions are almost identical when moving from anterior chamber toward sclera. Models III and IV show a smaller temperature gradient across the pupillary axis with an average temperature which is closer to the blood temperature. This is due to the presence of blood flow in the iris/sclera domain, which covers a substantial surface area of the eye ball as seen in Fig. 1(*c*). Figure 5 shows that for normal values of blood convection coefficient, there is a noticeable difference between the porous media models and models I and II. However, for very large blood convection coefficients, this difference diminishes significantly. This observation illustrates that the convective boundary condition on the iris/sclera surface has a substantial impact in all the models.

**3.3 Effect Of Ambient Conditions.** Figure 6 shows the effect of ambient convection coefficient on the pupillary axis temperature profile subject to a normal outside temperature. The eye thermal response to hot or cold ambient conditions is illustrated in Fig. 7. As can be seen, the porous media models provide a buffer against the outside thermal variations and result in an eye temperature distribution, which is closer to the temperature of the body. This is what is expected to occur based on biological information. It can be seen in these figures that the temperature profiles predicted by models, which account for the natural convection display a change in the curvature at the interface of the anterior chamber and the lens. For example, under cold ambient conditions given in Fig. 7(*b*), models II and IV result in cornea and anterior chamber temperatures, which are higher than those given by models I and III. This is due to the natural convection and formation of a circulatory cell within the anterior chamber. The dominant mechanism in the lens is conduction and for the cold ambient conditions, the higher temperature fluid in the anterior chamber transfers the thermal energy form the lens/AH interface toward the cornea/AH interface. This results in a temperature drop after the anterior chamber and the lens interface. Models I and III, which are only based on conduction in the cornea and the anterior cham-

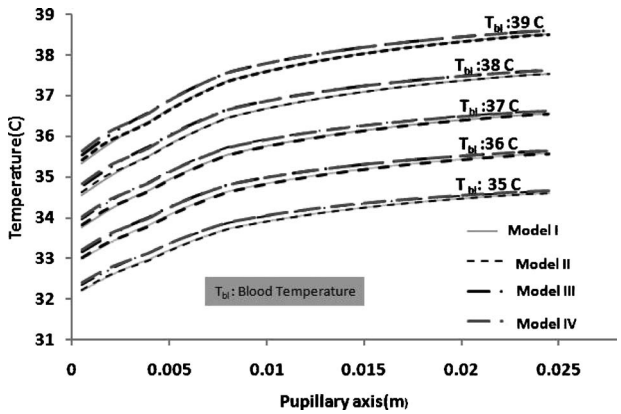



**Fig. 4 Effect of the blood temperature on the pupillary axis temperature profile for the analyzed models; $h_{am}$=10 W/m$^2$ K and $T_{am}$=25°C**

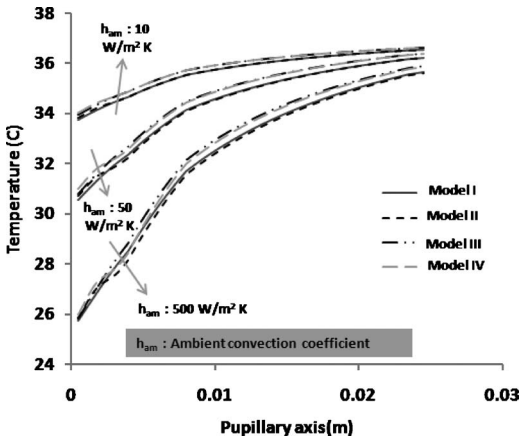



**Fig. 6 Effect of the ambient convection coefficient $h_{am}$ on the pupillary axis temperature profile for the analyzed models; $T_{am}$:25°C**

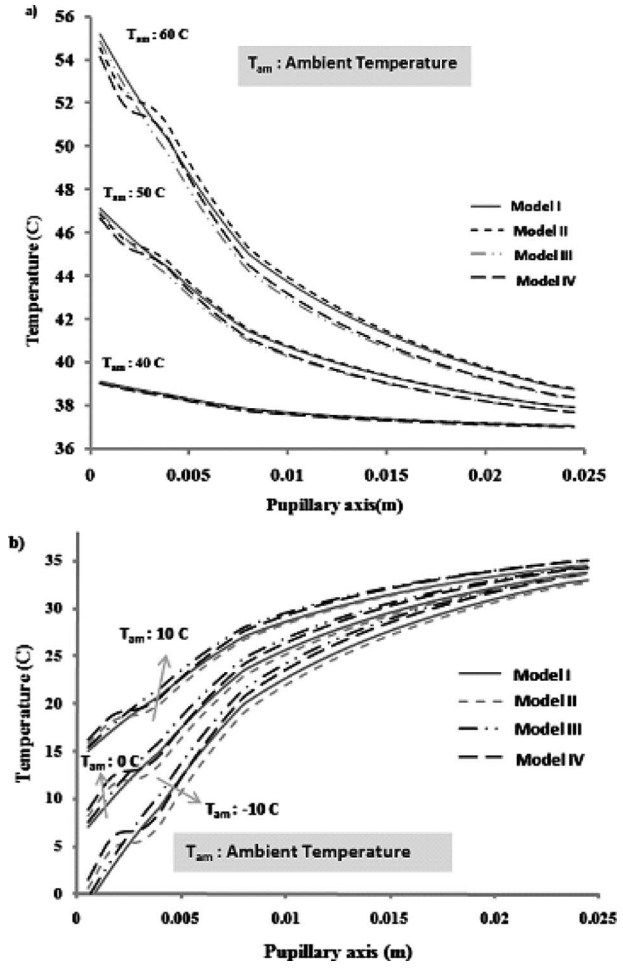



**Fig. 7 Effect of the extreme ambient conditions on the eye thermal response along the pupillary axis: (*a*) hot ambient conditions $T_{am}$=40°C, 50°C, and 60°C; $h_{am}$=200 W/m$^2$ K and (*b*) cold ambient condition, $T_{am}$=−10°C, 0°C, 10°C; $h_{am}$ =200 W/m$^2$ K**

ber, produce a lower temperature in this region compared with those from models II and IV due to lack of this circulatory cell within the anterior chamber. The existence of the buoyancy cell for the cold ambient conditions can be seen in Fig. 8.

For the hot ambient conditions, the reverse of the described phenomena occurs as seen in Fig. 7(*a*). As it can be seen, the porous media models provide the proper buffer and result in the closest average temperature to the blood temperature for the eyeball for both extreme cold and hot ambient conditions within the cornea and the anterior chamber. This is expected to occur based on biological information. This observation shows that under extreme thermal disturbance, AH natural convection plays its own role to establish the eye ball temperature along with a substantial influence of blood circulation in the iris/sclera.

**3.4 Effect of Aqueous Humor Flow Field on the Thermal Response of the Eye.** Natural convection within the anterior chamber can play an important role in establishing the thermal response of the eye. In Fig. 9 the influence of ambient conditions on the velocity field within the anterior chamber for models II and IV are shown. It can be seen that when ambient convection coefficient changes from 50 W/m$^2$ K to 500 W/m$^2$ K, the maximum velocity within the anterior chamber increases by 50%. The presence of porous media in domain *f* affects the velocity field as it augments the movement within the natural convection cell, as can

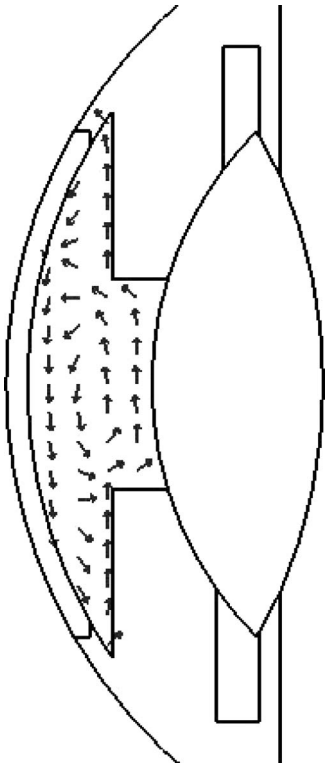


**Fig. 8 Display of velocity distribution inside the anterior chamber when exposed to a cold ambient condition**

be seen in Figs. 9(*d*)–9(*f*).

Natural convection in the anterior chamber maintains the temperature of the eye close to the other body organs when subjected to a thermal disturbance. When the thermal disturbance is more pronounced the differences among the four models becomes more noticeable.

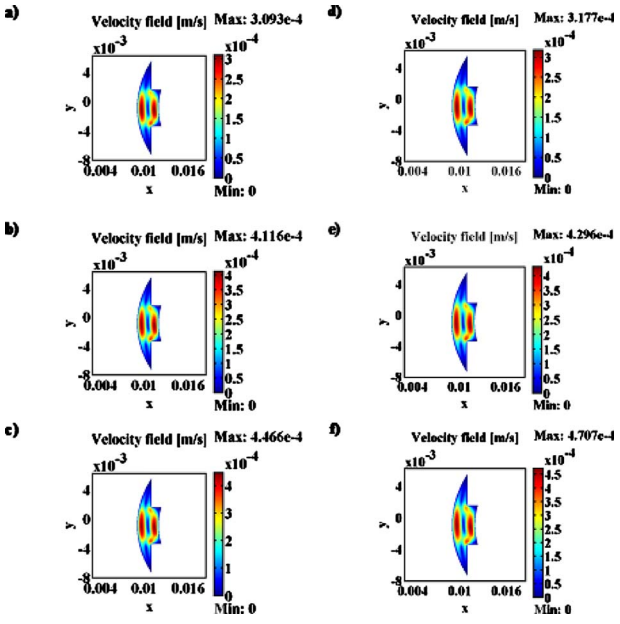



**Fig. 9 Effect of the ambient convection coefficient on the velocity distribution within the anterior chamber for models II and IV; $T_{am}$=25°C. Model II: (*a*) $h_{am}$=50 w/m$^2$ K, (*b*) $h_{am}$ =200 w/m$^2$ K, (*c*) $h_{am}$=500 w/m$^2$ K; model IV: (*d*) $h_{am}$ =50 w/m$^2$ K, (*e*) $h_{am}$=200 w/m$^2$ K, and (*f*) $h_{am}$=500 w/m$^2$ K.**

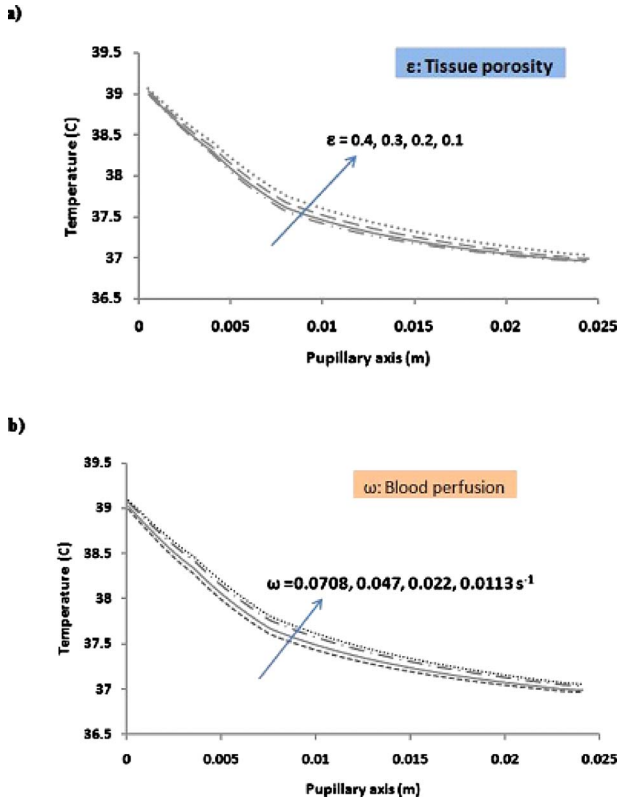



**Fig. 10 Effect of variations in the iris/sclera characteristics on the eye thermal response along the pupillary axis for Model IV: (*a*) Effect of variations in tissue porosities and (*b*) effect of variations in blood perfusion rates; $h_{am}=200$ W/m$^2$ K and $T_{am}=40$°C**

**3.5 Effect of Iris/Sclera Porosity and Blood Perfusion Rate.** As was mentioned earlier, Iris/sclera domain is considered as a porous medium with a given porosity and blood perfusion rate. As such the sclera/iris tissue characteristics has a significant influence on the thermal response of the eye. The effect of variations in the sclera domain porosity or perfusion rate on the temperature profile of the eye is shown in Figs. 10(*a*) and 10(*b*), respectively, for an ambient temperature of 40°C. It can be seen that an increase in either blood perfusion or porosity of the sclera enables the eye to better adapt to the internal body temperature.

## 4 Conclusions

The thermal response of a human eye to the internal and external disturbances was analyzed by investigating four primary thermal models. The effect of fluctuations in blood temperature, blood convection coefficient, ambient temperature, ambient convection coefficient, iris/sclera porosity, and blood perfusion rate on the temperature response for the eye was analyzed. The models, which were investigated in this study were compared with the available experimental data and the porous media models, were found to provide the best agreement. It was established that model IV, which includes the presence of blood circulation in the iris/sclera and hydrodynamics of AH region, maintains the closest temperature difference between the eyeball and the body organs when exposed to extreme thermal variations. The blood flow and AH circulation both have an important effect in adjusting the eyeball temperature. This effect becomes more pronounced for larger thermal disturbances, which is the case for cryosurgery and laser irradiation.

## Nomenclature

- c = specific heat, J/kg K
- E = evaporation rate of tear, W/m$^2$
- H = convection coefficient, W/m$^2$ K
- k = thermal conductivity, W/m K
- $n$ = normal vector
- $p$ = pressure, Pa
- $T$ = temperature, °C
- v = velocity, m/s

### Greek Symbols

- $\rho$ = density, kg/m$^3$
- $\mu$ = viscosity, Ns/m$^2$
- $\beta$ = volume expansion coefficient, 1/K
- $\varphi$ = porosity
- $\omega$ = perfusion rate, 1/s

### Subscripts

- $a$ = cornea
- am = ambient
- $b$ = anterior chamber
- bl = blood
- $c$ = lens
- $d$ = posterior chamber
- $e$ = vitreous
- $f$ = iris/sclera
- $i$ = subdomain index

## References


[1] Shafahi, M., and Vafai, K., 2009, "Thermal Modeling of the Human Eye as a Porous Structure," ASME Paper No. HT2009-88138.
[2] Ooi, E., and Ng, E. Y. K., 2008, "Simulation of Aqueous Humor Hydrodynamics in Human Eye Heat Transfer," Comput. Biol. Med., **38**, pp. 252–262.
[3] Ng, E. Y. K., and Ooi, E. H., 2006, "FEM Simulation of the Eye Structure With Bioheat Analysis," Comput. Methods Programs Biomed., **82**, pp. 268–276.
[4] Narasimhan, A., Kumar Jha, K., and Gopal, L., 2010, "Transient Simulations of Heat Transfer in Human Eye Undergoing Laser Surgery," Int. J. Heat Mass Transfer, **53**, pp. 482–490.
[5] Chua, K. J., Ho, J. C., Chou, S. K., and Islam, M. R., 2005, "On the Study of the Temperature Distribution Within a Human Eye Subjected to a Laser Source," Int. Commun. Heat Mass Transfer, **32**, pp. 1057–1065.
[6] Lagendijk, J. J. W., 1982, "A Mathematical Model to Calculate Temperature Distribution in Human and Rabbit Eye During Hyperthermic Treatment," Phys. Med. Biol., **27**, pp. 1301–1311.
[7] Emery, A. F., Kramar, P., Guy, A. W., and Lin, J. C., 1975, "Microwave Induced Temperature Rises in Rabbit Eyes in Cataract Research," ASME J. Heat Transfer, **97**, pp. 123–128.
[8] Scott, J., 1988, "A Finite Element Model of Heat Transport in the Human Eye," Phys. Med. Biol., **33**, pp. 227–241.
[9] Scott, J., 1988, "The Computation of Temperature Rises in the Human Eye Induced by Infrared Radiation," Phys. Med. Biol., **33**, pp. 243–257.
[10] Kumar, S., Acharya, S., Beuerman, R., and Palkama, A., 2006, "Numerical Solution of Ocular Fluid Dynamics in a Rabbit Eye: Parametric Effects," Ann. Biomed. Eng., **34**, pp. 530–544.
[11] Flyckt, V. M. M., Raaymakers, B. W., and Lagendijk, J. J. W., 2006, "Modeling the Impact of Blood Flow on the Temperature Distribution in the Human Eye and the Orbit: Fixed Heat Transfer Coefficients Versus the Pennes Bioheat Model Versus Discrete Blood Vessels," Phys. Med. Biol., **51**, pp. 5007–5021.
[12] Amara, E. H., 1995, "Numerical Investigations on Thermal Effects of Laser–Ocular Media Interaction," Int. J. Heat Mass Transfer, **38**, pp. 2479–2488.
[13] Jacobs, M. D., 2009, "Multiscale Systems Integration in the Eye," WIREs Systems Biology and Medicine, **1**, pp. 15–27.
[14] Mapstone, R., 1968, "Determinants of Corneal Temperature," Br. J. Ophthalmol., **52**, pp. 729–741.
[15] Mapstone, R., 1968, "Measurement of Corneal Temperature," Exp. Eye Res., **7**, pp. 237–242.
[16] Nakayama, A., and Kuwahara, F., 2008, "A General Bioheat Transfer Model Based on the Theory of Porous Media," Int. J. Heat Mass Transfer, **51**, pp. 3190–3199.
[17] Kuwahara, F., Yoshihik, S., Jianjun, L., and Nakayama, A., 2009, "A Porous Media Approach for Bifurcating Flow and Mass Transfer in a Human Lung," ASME J. Heat Transfer, **131**, p. 101013.
[18] Nakayama, A., Kuwahara, F., and Liu, W., 2009, "A Macroscopic Model for Countercurrent Bioheat Transfer in a Circulatory System," J. Porous Media, **12**, pp. 289–300.
[19] Mahjoob, S., and Vafai, K., 2009, "Analytical Characterization of Heat Transfer Through Biological Media Incorporating Hyperthermia Treatment," Int. J. Heat Mass Transfer, **52**, pp. 1608–1618.

[20] Mahjoob, S., and Vafai, K., 2010, "Analysis of Bioheat Transport Through a Dual Layer Biological Media," ASME J. Heat Transfer, **132**, p. 031101.
[21] Khanafer, K., and Vafai, K., 2009, "Synthesis of Mathematical Models Representing Bioheat Transport," *Advances in Numerical Heat Transfer*, Vol. 3, W. J. Minkowycz and E. M. Sparrow, eds., Taylor & Francis, London, pp. 1–28.
[22] Khakpour, M., and Vafai, K., 2008, "A Critical Assessment of Arterial Transport Models," Int. J. Heat Mass Transfer, **51**, pp. 807–822.
[23] Khakpour, M., and Vafai, K., 2008, "A Comprehensive Analytical Solution of Macromolecular Transport Within an Artery," Int. J. Heat Mass Transfer, **51**, pp. 2905–2913.
[24] Khaled, A.-R. A., and Vafai, K., 2003, "The Role of Porous Media in Modeling Flow and Heat Transfer in Biological Tissues," Int. J. Heat Mass Transfer, **46**, pp. 4989–5003.
[25] Khanafer, K., and Vafai, K., 2006, "The Role of Porous Media in Biomedical Engineering as Related to Magnetic Resonance Imaging and Drug Delivery," Heat Mass Transfer, **42**, pp. 939–953.

# Modeling Bioheat Transport at Macroscale

**Liqiu Wang**[1]
e-mail: lqwang@hku.hk

**Jing Fan**

Department of Mechanical Engineering,
University of Hong Kong,
Pokfulam Road, Hong Kong

*Macroscale thermal models have been developed for biological tissues either by the mixture theory of continuum mechanics or by the porous-media theory. The former uses scaling-down from the global scale; the latter applies scaling-up from the microscale by the volume averaging. The used constitutive relations for heat flux density vector include the Fourier law, the Cattaneo–Vernotte (Cattaneo, C., 1958, "A Form of Heat Conduction Equation Which Eliminates the Paradox of Instantaneous Propagation," Compt. Rend., **247**, pp. 431–433; Vernotte, P., 1958, "Les Paradoxes de la Théorie Continue de I'equation de la Chaleur," Compt. Rend., **246**, pp. 3154–3155) theory, and the dual-phase-lagging theory. The developed models contain, for example, the Pennes (1948, "Analysis of Tissue and Arterial Blood Temperature in the Resting Human Forearm," J. Appl. Physiol., **1**, pp. 93–122), Wulff (1974, "The Energy Conservation Equation for Living Tissues," IEEE Trans. Biomed. Eng., **BME-21**, pp. 494–495), Klinger (1974, "Heat Transfer in Perfused Tissue I: General Theory," Bull. Math. Biol., **36**, pp. 403–415), and Chen and Holmes (1980, "Microvascular Contributions in Tissue Heat Transfer," Ann. N.Y. Acad. Sci., **335**, pp. 137–150), thermal wave bioheat, dual-phase-lagging (DPL) bioheat, two-energy-equations, blood DPL bioheat, and tissue DPL bioheat models. We analyze the methodologies involved in these two approaches, the used constitutive theories for heat flux density vector and the developed models. The analysis shows the simplicity of the mixture theory approach and the powerful capacity of the porous-media approach for effectively developing accurate macroscale thermal models for biological tissues. Future research is in great demand to materialize the promising potential of the porous-media approach by developing a rigorous closure theory. The heterogeneous and nonisotropic nature of biological tissue yields normally a strong noninstantaneous response between heat flux and temperature gradient in nonequilibrium heat transport. Both blood and tissue macroscale temperatures satisfy the DPL-type energy equations with the same values of the phase lags of heat flux and temperature gradient that can be computed in terms of blood and tissue properties, blood-tissue interfacial convective heat transfer coefficient, and blood perfusion rate. The blood-tissue interaction leads to very sophisticated effect of the interfacial convective heat transfer, the blood velocity, the perfusion, and the metabolic reaction on blood and tissue macroscale temperature fields such as the spreading of tissue metabolic heating effect into the blood DPL bioheat equation and the appearance of the convection term in the tissue DPL bioheat equation due to the blood velocity.* [DOI: 10.1115/1.4002361]

*Keywords: bioheat transport, mixture theory, porous-media theory, dual-phase-lagging, blood-tissue interaction, macroscale, modeling*

## 1 Introduction

The accurate description of heat transport in biological tissues is essential not only for fundamental understanding of biological processes/functions but also for many medical operations of thermal therapy, cryopreservation, and biopreservation [1–5]. Biological tissues are composed of dispersed cells separated by voids. Blood flows into these tissues through arteries and perfuses to the cells via blood capillaries. Returned blood from the capillaries is collected in veins and then pumped back to the heart. Heat transport in biological tissues is thus enriched by heat conduction in tissue and vascular system, blood-tissue convection, and perfusion through the capillaries within the tissue and also metabolic heat generation.

Heat transport in biological tissues may be studied from a molecular point of view, from a microscopic point of view, or from a macroscopic point of view [6]. The macroscale is a phenomenological scale that is much larger than the microscale of cells and voids and much smaller than the system length scale. The interest in the macroscale rather than the molecular scale and the microscale comes from the fact that a prediction at the molecular scale or the microscale is complicated because of either the huge numbers of particles at molecular scale or the complex microscale anatomical structure of biological media, and that we are usually more interested in large scales of heat transport for practical applications. Existence of such a macroscale description equivalent to the microscale behavior requires a good separation of length scale and has been well discussed in Ref. [7].

The macroscopic thermal models for blood-perfused tissues have been developed by two approaches: (1) scaling-down from the global scale based on the mixture theory of continuum mechanics and (2) scaling-up from the microscale based on porous-media theory. We present a concise synthesis of these two approaches and the models developed. Instead of a comprehensive review, our emphasis is limited to identify the essence of the two approaches and to discuss the fundamental heat-conduction theories they involved.

## 2 Scaling-Down by Mixture Theory

To develop a macroscale model of heat transport in living biological tissues, the mixture theory of continuum mechanics views

[1]Corresponding author.
Manuscript received March 26, 2010; final manuscript received March 31, 2010; published online September 30, 2010. Assoc. Editor: Peter Vadasz.

blood and tissues as a mixture of continuum deformable media. In this approach, no microscale presentation of the system is provided and microscale quantities are not introduced. Phase properties are defined at the macroscale. The global balance equations are formed in terms of macroscale properties and with additional terms accounting for the interaction between blood and tissue. These global equations can then be localized to obtain the macroscale point equations. Required constitutive equations are supplied by direct postulation of desirable relations at macroscale.

There are two drawbacks associated with this approach. The first is the lack of connection between microscale and macroscale properties. The second is the difficulty in extension to multiphase systems with distinct properties of interfaces and common curves.

In this approach, a global balance is written with the integrand properties at the macroscale. For a property $\psi$ of the $\alpha$-phase, the conservation equation reads [8]

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V (\rho\varepsilon\psi)^{\alpha}_{\mathrm{mac}}\mathrm{d}V + \int_S \mathbf{n}\cdot[(\rho\varepsilon)^{\alpha}_{\mathrm{mac}}(\mathbf{v}^{\alpha}_{\mathrm{mac}}-\mathbf{w})\psi^{\alpha}_{\mathrm{mac}} - (\varepsilon\mathbf{i})^{\alpha}_{\mathrm{mac}}]\mathrm{d}S - \int_V (\rho\varepsilon G)^{\alpha}_{\mathrm{mac}}\mathrm{d}V - \int_V (\rho\varepsilon f)^{\alpha}_{\mathrm{mac}}\mathrm{d}V = 0 \tag{1}$$

where $t$ is time, the superscript $\alpha$ is used to indicate the $\alpha$-phase properties, and the subscript "mac" is used to indicate the macroscale properties. $\rho$, $\varepsilon$, and $\mathbf{v}$ are the density, the volume fraction, and the velocity, respectively. $V$ is the global volume at a given time instant $t$. The boundary of $V$, denoted by $S$, may have a velocity $\mathbf{w}$ in general. $\mathbf{i}$ is the diffusive flux of $\psi$ across the boundary, $\mathbf{n}$ is the unit vector normal to $S$ and pointing outward from $V$, $G$ is the term accounting for production of $\psi$ within the volume, and $f$ is the external supply of $\psi$. This equation is a mathematical statement of the physical principle that the rate of change of some property in a volume is equal to the net flux of that property across the boundary of the volume plus production of the property and the external supply.

To transform Eq. (1) to a differential form at macroscale, we must use the two multiscale theorems (Eqs. (311) and (320) in Ref. [6]) to bring the time derivative inside the integral and to convert the boundary integral to a volume integral so that

$$\int_V \left[\frac{\partial(\rho\varepsilon\psi)^{\alpha}_{\mathrm{mac}}}{\partial t} + \nabla\cdot(\rho\varepsilon\mathbf{v}\psi)^{\alpha}_{\mathrm{mac}} - \nabla\cdot(\varepsilon\mathbf{i})^{\alpha}_{\mathrm{mac}} - (\rho\varepsilon G)^{\alpha}_{\mathrm{mac}} - (\rho\varepsilon f)^{\alpha}_{\mathrm{mac}}\right]\mathrm{d}V = 0 \tag{2}$$

Because the size of the volume is arbitrary, by the localization theorem [6], the integrand in Eq. (2) must be zero as long as Axiom 1 in Ref. [6] is satisfied so that the macroscale point equation can be obtained as

$$\frac{\partial(\rho\varepsilon\psi)^{\alpha}_{\mathrm{mac}}}{\partial t} + \nabla\cdot(\rho\varepsilon\mathbf{v}\psi)^{\alpha}_{\mathrm{mac}} - \nabla\cdot(\varepsilon\mathbf{i})^{\alpha}_{\mathrm{mac}} - (\rho\varepsilon G)^{\alpha}_{\mathrm{mac}} - (\rho\varepsilon f)^{\alpha}_{\mathrm{mac}} = 0 \tag{3}$$

Consider $\psi$ being the internal energy of biological tissue. Equation (3) reduces to

$$\frac{\partial(\rho c\varepsilon T)^{t}_{\mathrm{mac}}}{\partial t} = -\nabla\cdot(\varepsilon\mathbf{q})^{t}_{\mathrm{mac}} + (\rho\varepsilon q_m)^{t}_{\mathrm{mac}} + (\rho\varepsilon q_c)^{t}_{\mathrm{mac}} + (\rho\varepsilon q_p)^{t}_{\mathrm{mac}} + (\rho\varepsilon q_e)^{t}_{\mathrm{mac}} \tag{4}$$

where the superscript $t$ is used to indicate the tissue properties, $c$ is the specific heat, And $\mathbf{q}$ is the heat flux density vector. $q_m$, $q_c$, and $q_p$ are the volumetric rate of heat generation by the metabolic heating, the blood interfacial convective heat transfer, and the blood perfusion, respectively. $q_e$ is the volumetric rate of external heat supply like the one used in hyperthermia therapy. There exists some confusion in the literature between $q_c$ and $q_p$. The former comes actually from the interfacial convective heat transfer between the blood and the blood vessel, driving by the temperature difference between them. The latter, on the other hand, stems from the energy exchange due to the blood perfusion, the mass-transfer process of nutritive delivery of arterial blood to capillaries in the biological tissue.

The existing thermal models for biological tissues developed by this approach differ from each other mainly on how to model the heat flux density vector $\mathbf{q}$. Three constitutive relations for $\mathbf{q}$ have been used: the Fourier law, the Cattaneo–Vernotte (CV) relation, and the dual-phase-lagging (DPL) relation.

**2.1 The Fourier Law.** The Fourier law was the first constitutive relation of heat flux density and was proposed by the French mathematical physicist Joseph Fourier in 1807 based on experimentation and investigation [9]. For heat conduction in a homogeneous and isotropic medium, the Fourier law of heat conduction reads

$$\mathbf{q}(\boldsymbol{r},t) = -k\,\nabla\,T(\boldsymbol{r},t) \tag{5}$$

where $\boldsymbol{r}$ stands for the material point, $t$ stands for the time, $T$ stands for the temperature, and $\nabla$ stands for the gradient operator. $k$ is the thermal conductivity of the material, which is a thermodynamic property. By the state theorem of thermodynamics, $k$ should be a function of two independent, intensive properties (normally pressure and temperature) [10]. The second law of thermodynamics requires that $k$ is positive definite [9,11,12]. In engineering applications, we often take $k$ as a material constant because variations in pressure and temperature are normally sufficiently small. The value of $k$ is material dependent. If the material is not homogeneous or isotropic, $k$ becomes a second-order tensor [9,11–13]. Along with the first law of thermodynamics, this equation leads to the classical *parabolic* heat-conduction equation

$$\frac{\partial T}{\partial t} = \alpha\Delta T + \frac{\alpha}{k}F \tag{6}$$

Here $\alpha$ is the thermal diffusivity of the material, $F$ is the rate of internal energy generation per unit volume, and $\Delta$ is the Laplacian.

By using the Fourier law, Eq. (4) yields a group of thermal models for biological tissues

$$\frac{\partial(\rho\varepsilon cT)^{t}_{\mathrm{mac}}}{\partial t} = \nabla\cdot(\varepsilon k\,\nabla\,T)^{t}_{\mathrm{mac}} + (\rho\varepsilon q_m)^{t}_{\mathrm{mac}} + (\rho\varepsilon q_c)^{t}_{\mathrm{mac}} + (\rho\varepsilon q_p)^{t}_{\mathrm{mac}} + (\rho\varepsilon q_e)^{t}_{\mathrm{mac}} \tag{7}$$

The thermal models in this group includes the classical Pennes model [14], the Wulff model [15], the Klinger model [16], and the Chen and Holmes model [17]. Table 1 lists $(\rho c)^{t}_{\mathrm{mac}}$, $\varepsilon^{t}_{\mathrm{mac}}$, $k^{t}_{\mathrm{mac}}$, $(\rho\varepsilon q_m)^{t}_{\mathrm{mac}}$, $(\rho\varepsilon q)^{t}_{\mathrm{mac}}$, $(\rho\varepsilon q_p)^{t}_{\mathrm{mac}}$, and $(\rho\varepsilon q_e)^{t}_{\mathrm{mac}}$ in these models.

The Fourier law of heat conduction is an early empirical law. It assumes that $\boldsymbol{q}$ and $\nabla T$ appear at the same time instant $t$ and consequently implies that thermal signals propagate with an infinite speed. If the material is subjected to a thermal disturbance, the effects of the disturbance will be felt instantaneously at distances infinitely far from its source. Although this result is physically unrealistic, it has been confirmed by many experiments that the Fourier law of heat conduction holds for many media in the usual range of heat flux $\mathbf{q}$ and temperature gradient $\nabla T$ [9].

**2.2 The CV Constitutive Relation.** With the development of science and technology such as the application of ultrafast pulse-laser heating on metal films, heat conduction appears in a range of high heat flux and high unsteadiness. The drawback of infinite heat propagation speed in the Fourier law becomes unacceptable. This has inspired the work of searching for new constitutive rela-

**Table 1 Pennes, Wulff, Klinger, and Chen and Holmes models with the confusion between $q_c$ and $q_p$ corrected (superscript $b$ indicating blood properties, $v_h$ indicating local mean blood velocity, $\mathbf{v}_p$ indicating mean perfusion velocity, $\Delta h$ indicating enthalpy of formation in metabolic reaction, $\phi$ indicating extent of reaction, $\omega^b$ indicating blood perfusion rate, and $\omega^*$ indicating total perfusion bleed-off to the tissue only from the microvessels)**

| Model | $(\rho c)^t_{\text{mac}}$ | $\varepsilon^t_{\text{mac}}$ | $k^t_{\text{mac}}$ | $(\rho\varepsilon q_m)^t_{\text{mac}}$ | $(\rho\varepsilon q_c)^t_{\text{mac}}$ | $(\rho\varepsilon q_p)^t_{\text{mac}}$ | $(\rho\varepsilon q_e)^t_{\text{mac}}$ |
|---|---|---|---|---|---|---|---|
| Pennes | Constant $(\rho c)^t_{\text{mac}}$ | 1 | Constant $k^t_{\text{mac}}$ | $(\rho\varepsilon q_m)^t_{\text{mac}}$ | 0 | $(\rho c)^b_{\text{mac}}\omega^b(T^b_{\text{mac}}-T^t_{\text{mac}})$ | 0 |
| Wulff | Constant $(\rho c)^t_{\text{mac}}$ | 1 | Constant $k^t_{\text{mac}}$ | $\rho^b_{\text{mac}}v_h\Delta h\nabla\phi$ | $-(\rho c)^b_{\text{mac}}v_h\nabla T^t_{\text{mac}}$ | 0 | 0 |
| Klinger | Constant $(\rho c)^t_{\text{mac}}$ | 1 | Constant $k^t_{\text{mac}}$ | $(\rho\varepsilon q_m)^t_{\text{mac}}$ | $-(\rho c)^b_{\text{mac}}\mathbf{v}^b_{\text{mac}}\nabla T^t_{\text{mac}}$ | 0 | 0 |
| Chen and Holmes | Constant $(\rho c)^t_{\text{mac}}$ | 1 | $k_{\text{eff}}=(1-\varepsilon^t_{\text{mac}})k^b_{\text{mac}}+(\varepsilon k)^t_{\text{mac}}\approx k^t_{\text{mac}}$ | $(\rho\varepsilon q_m)^t_{\text{mac}}$ | $-(\rho c)^b_{\text{mac}}\mathbf{v}_p\cdot\nabla T^t_{\text{mac}}$ | $(\rho c)^b_{\text{mac}}\omega^*(T^b_{\text{mac}}-T^t_{\text{mac}})$ | 0 |

tions. Among many proposed relations [9], the constitutive relation proposed by Cattaneo [18] and Vernotte [19,20],

$$\mathbf{q}(\mathbf{r},t)+\tau_q\frac{\partial\mathbf{q}(\mathbf{r},t)}{\partial t}=-k\nabla T(\mathbf{r},t) \tag{8}$$

is the most widely accepted. This relation is named the CV constitutive relation after the names of the proposers. Here $\tau_q>0$ is a material property and is called the *relaxation time*. The corresponding heat-conduction equation is thus

$$\frac{\partial T}{\partial t}+\tau_q\frac{\partial^2 T}{\partial t^2}=\alpha\Delta T+\frac{\alpha}{k}\left(F+\tau_q\frac{\partial F}{\partial t}\right) \tag{9}$$

Unlike its classical counterpart equation (6), this equation is of *hyperbolic* type, characterizes the combined diffusion and wavelike behavior of heat conduction, and predicts a *finite speed*,

$$V_{\text{CV}}=\sqrt{\frac{k}{\rho c\tau_q}} \tag{10}$$

for heat propagation [21].

Note that the CV constitutive relation is actually a first-order approximation of a more general constitutive relation (single-phase-lagging model [22]),

$$\mathbf{q}(\mathbf{r},t+\tau_q)=-k\nabla T(\mathbf{r},t) \tag{11}$$

according to which the temperature gradient established at a point $\mathbf{r}$ at time $t$ gives rise to a heat flux vector at $\mathbf{r}$ at a *later* time $t+\tau_q$. There is a finite built-up time $\tau_q$ for the onset of heat flux at $\mathbf{r}$ after a temperature gradient is imposed there. Thus $\tau_q$ represents the time lag needed to establish the heat flux (the result) when a temperature gradient (the cause) is suddenly imposed. The higher $\partial\mathbf{q}/\partial t$ corresponds to a larger derivation of the CV constitutive relation from the classical Fourier law.

The value of $\tau_q$ is material dependent [23–25]. For most solid materials, $\tau_q$ varies from $10^{-10}$ s to $10^{-14}$ s. For gases, $\tau_q$ is normally in the range of $10^{-8}$–$10^{-10}$ s. The value of $\tau_q$ for some biological materials and materials with nonhomogeneous inner structures can be up to $10^2$ s [26–31]. The long time delay in heterogeneous materials comes from the fact that the structural heat interaction occurs at multiscales [32]. Therefore, the thermal relaxation effects can be of relevance even in common engineering applications where the time scales of interest are of the order of a fraction of 1 minute.

Three factors contribute to the significance of the second term in the hyperbolic heat-conduction Eq. (9): the $\tau_q$ value, the rate of change of temperature, and the time scale involved. The wave nature of thermal signals will be over the diffusive behavior through this term when [22]

$$\frac{\partial T}{\partial t}\gg\frac{T_r}{2\tau_q}\exp(t/\tau_q) \tag{12}$$

where $T_r$ is a reference temperature. Therefore, the wavelike features will become significant when (1) $\tau_q$ is large, (2) $\partial T/\partial t$ is high, or (3) $t$ is small. Some typical situations where hyperbolic heat conduction differs from classical parabolic heat conduction include those concerned with a localized moving heat source with a high intensity, a rapidly propagating crack tip, shock wave propagation, thermal resonance, interfacial effects between dissimilar materials, laser material processing, transport in biological systems, and laser surgery [22–25,33–36].

When $\tau_q\to\infty$ but $k_e=k/\tau_q$ is finite, the CV constitutive relation (8) and the hyperbolic heat-conduction equation (9) become [34]

$$\frac{\partial\mathbf{q}(\mathbf{r},t)}{\partial t}=-k_e\nabla T(\mathbf{r},t) \tag{13}$$

and

$$\frac{\partial^2 T}{\partial t^2}=\alpha_e\Delta T+\frac{\alpha_e}{k_e}\frac{\partial F}{\partial t} \tag{14}$$

where $\alpha_e=k_e/(\rho c)$, $\rho$ and $c$ are the density and the specific heat of the material, respectively. Therefore, when $\tau_q$ is very large, a temperature gradient established at a point of the material results in an *instantaneous heat flux rate* at that point, and vice versa. Equation (14) is a classical wave equation that predicts thermal wave propagation with speed $V_{\text{CV}}$, like Eq. (9). A major difference exists, however, between Eqs. (9) and (14): The former allows damping of thermal waves, the latter does not [21].

Using the CV relation (Eq. (8)) as the constitutive relation for $\mathbf{q}^t_{\text{mac}}$, Eq. (4) yields

$$\frac{\partial(\rho\varepsilon cT)^t_{\text{mac}}}{\partial t}+\tau_q\frac{\partial^2(\rho\varepsilon cT)^t_{\text{mac}}}{\partial t^2}=\nabla\cdot(\varepsilon k\nabla T)^t_{\text{mac}}+\left(1+\tau_q\frac{\partial}{\partial t}\right)(\rho\varepsilon q_m+\rho\varepsilon q_c+\rho\varepsilon q_p+\rho\varepsilon q_e)^t_{\text{mac}} \tag{15}$$

This is known as a hyperbolic bioheat equation. It can well predict the experimental results of some biological materials [28] and has been used for the blood perfusion rate measurement [37–39], for the explanation of temperature oscillations in biological systems [40], for the prediction of temperature and thermal stress during skin cryopreservation [41], and for the prediction of temperature and thermal dose distributions in living tissue during thermal therapies [42]. The hyperbolic model also predicts considerably different temperature and thermal damage in skin tissues under different heating from the Pennes model [43,44] and provides more realistic predictions both for the case of heating with a high flux under an extremely short duration [43–45] and for the thermal behavior in surgical techniques with laser and radiofrequency heating [46].

**2.3 The Dual-Phase-Lagging Constitutive Relation.** It has been confirmed by many experiments that the CV constitutive relation generates a more accurate prediction than the classical Fourier law. However, some of its predictions do not agree with experimental results either [9,25,33]. A thorough study shows that the CV constitutive relation has only taken account of the fast-transient effects but not the microstructural interactions. These two effects can be reasonably represented by the dual-phase-lag between $\mathbf{q}$ and $\nabla T$, a further modification of Eq. (5) [25,33],

$$\mathbf{q}(\mathbf{r},t+\tau_q) = -k\nabla T(\mathbf{r},t+\tau_T) \tag{16}$$

According to this relation, the temperature gradient at a point $\mathbf{r}$ of the material at time $t+\tau_T$ corresponds to the heat flux density vector at $\mathbf{r}$ at time $t+\tau_q$. The delay time $\tau_T$ is interpreted as being caused by the microstructural interactions (small-scale heat transport mechanisms occurring at the microscale or small-scale effects of heat transport in space) such as phonon-electron interaction or phonon scattering and is called the *phase-lag of the temperature gradient* [25,33]. The other delay time $\tau_q$ is interpreted as the relaxation time due to the fast-transient effects of thermal inertia (or small-scale effects of heat transport in time) and is called the *phase-lag of the heat flux*. Both of the phase-lags are treated as intrinsic thermal or structural properties of the material. The corresponding heat-conduction equation reads [47]

$$\frac{1}{\alpha}\frac{\partial T(\mathbf{r},t')}{\partial t} = \Delta T(\mathbf{r},t'-\tau) + \frac{1}{k}F(\mathbf{r},t'), \quad t' = t+\tau_q, \quad \tau = \tau_q - \tau_T,$$
$$\text{for} \quad \tau_q - \tau_T > 0 \quad \text{and} \quad t' > \tau_q \tag{17}$$

or

$$\frac{1}{\alpha}\frac{\partial T(\mathbf{r},t'-\tau)}{\partial t} = \Delta T(\mathbf{r},t') + \frac{1}{k}F(\mathbf{r},t'-\tau), t' = t+\tau_T, \tau = \tau_T - \tau_q,$$
$$\text{for} \quad \tau_q - \tau_T < 0 \quad \text{and} \quad t' > \tau_T \tag{18}$$

Unlike the relation (11) according to which the heat flux is the result of a temperature gradient in a transient process, the relation (16) allows either the temperature gradient or the heat flux to become the effect and the remaining one to be the cause. For materials with $\tau_q > \tau_T$, the heat flux density vector is the result of a temperature gradient. It is the other way around for materials with $\tau_T > \tau_q$. The relation (11) corresponds to the particular case where $\tau_q > 0$ and $\tau_T = 0$. If $\tau_q = \tau_T$ (not necessarily equal to zero), the response between the temperature gradient and the heat flux is instantaneous; in this case, the relation (16) is identical to the classical Fourier law (5). It may also be noted that while the classical Fourier law (5) is macroscopic in both space and time and the relation (11) is macroscopic in space but microscopic in time, the relation (16) is microscopic in both space and time. Also note that Eqs. (17) and (18) are of the delay and advance types, respectively. While the former has a wavelike solution and possibly resonance, the latter does not [47]. Both single-phase-lagging and dual-phase-lagging heat conduction have been shown to be admissible by the second law of extended irreversible thermodynamics [25], by the Boltzmann transport equation [47,48], and by the Galilean principle of relativity [49].

Expanding both sides of Eq. (16) using the Taylor series and retaining only the first-order terms of $\tau_q$ and $\tau_T$, we obtain the following constitutive relation that is valid at point $\mathbf{r}$ and time $t$,

$$\mathbf{q}(\mathbf{r},t) + \tau_q\frac{\partial \mathbf{q}(\mathbf{r},t)}{\partial t} = -k\left\{\nabla T(\mathbf{r},t) + \tau_T\frac{\partial}{\partial t}[\nabla T(\mathbf{r},t)]\right\} \tag{19}$$

which is known as the Jeffreys-type constitutive equation of heat flux [34]. In literature this relation is also called the *dual-phase-lagging constitutive relation*. When $\tau_q = \tau_T$, this relation reduces to the classical Fourier law (5), and it reduces to the CV constitutive relation (8) when $\tau_T = 0$.

Eliminating $\mathbf{q}$ from Eq. (19) and the classical energy equation leads to the dual-phase-lagging heat-conduction equation that reads, if all thermophysical material properties are assumed to be constant

$$\frac{\partial T}{\partial t} + \tau_q\frac{\partial^2 T}{\partial t^2} = \alpha\Delta T + \alpha\tau_T\frac{\partial}{\partial t}(\Delta T) + \frac{\alpha}{k}\left(F + \tau_q\frac{\partial F}{\partial t}\right) \tag{20}$$

This equation is *parabolic* when $\tau_q < \tau_T$ [21]. Although a wave term $\tau_q\partial^2 T/\partial t^2$ exists in the equation, the mixed derivative $\alpha\tau_T\partial(\Delta T)/\partial t$ completely destroys the wave structure. The equation, in this case, therefore predicts a nonwavelike heat conduction that differs from the usual diffusion predicted by the classical parabolic heat conduction (6). When $\tau_q > \tau_T$, however, Eq. (20) can be approximated by Eq. (9) and then predominantly predicts wavelike thermal signals.

The dual-phase-lagging heat-conduction equation (20) forms a generalized, unified equation that reduces to the classical parabolic heat-conduction equation when $\tau_T = \tau_q$, the hyperbolic heat-conduction equation when $\tau_T = 0$ and $\tau_q > 0$, the energy equation in the phonon scattering model when $\alpha = \tau_R c^2/3$, $\tau_T = (9/5)\tau_N$, and $\tau_q = \tau_R$ [34,50], and the energy equation in the phonon-electron interaction model when $\alpha = k/(c_e + c_l)$, $\tau_T = c_l/G$, and $\tau_q = 1/G((1/c_e)+(1/c_l))^{-1}$ [51–53]. In the phonon scattering model, $c$ is the average speed of phonons (sound speed), $\tau_R$ is the relaxation time for the Umklapp process in which momentum is lost from the phonon system, and $\tau_N$ is the relaxation time for normal processes in which momentum is conserved in the phonon system. In the phonon-electron interaction model, $k$ is the thermal conductivity of the electron gas, $G$ is the phonon-electron coupling factor, and $c_e$ and $c_l$ are the heat capacity of the electron gas and the metal lattice, respectively. This, together with its success in describing and predicting phenomena such as ultrafast pulse-laser heating, propagation of temperature pulses in superfluid liquid helium, nonhomogeneous lagging response in porous media, thermal lagging in amorphous materials, and effects of material defects and thermomechanical coupling, heat conduction in nanofluids, bicomposite media, and two-phase systems [21,25,54–63], has given rise to the research effort on various aspects of dual-phase-lagging heat conduction [21,25].

The dual-phase-lagging heat-conduction equation (Eq. (20)) has been shown to be well-posed in a finite region of $n$-dimensions ($n \geq 1$) under any linear boundary conditions including Dirichlet, Neumann, and Robin types [63]. Solutions of one-dimensional (1D) heat conduction have been obtained for some specific initial and boundary conditions in [25,33,54,64–70]. Analytical solutions have also been obtained in Ref. [21] for regular 1D, 2D, and 3D heat-conduction domains under essentially arbitrary initial and boundary conditions. The solution structure theorems were also developed for both mixed and Cauchy problems of dual-phase-lagging heat-conduction equations in Refs. [21,71] by extending those theorems for hyperbolic heat conduction [36]. These theorems build relationships among the contributions (to the temperature field) by the initial temperature distribution, the source term, and the initial time-rate of the temperature change, uncovering the structure of the temperature field and considerably simplifying the development of solutions. Xu and Wang [72] addressed thermal features of dual-phase-lagging heat conduction (particularly conditions and features of thermal oscillation and resonance and their contrast with those of classical and hyperbolic heat conduction). The issues associated with the Galilean principle of relativity have also been discussed in Ref. [49] for both single- and dual-phase-lagging heat-conduction models in moving media.

An experimental procedure for determining the value of $\tau_q$ has been proposed in Ref. [73]. The general problem of measuring short-time thermal transport effects has been discussed in Ref. [74]. Three methods have been developed in Ref. [21] for measuring $\tau_q$. The equivalence has also been built-up in Refs. [25,55–60] between the Fourier heat conduction in porous media and the dual-phase-lagging heat conduction.

Tzou [25,69] also generalized Eq. (19), for $\tau_q \gg \tau_T$, by retaining terms up to the second order in $\tau_q$ but only the term of the first order in $\tau_T$ in the Taylor expansions of Eq. (16) to obtain a $\tau_q$-second-order dual-phase-lagging model

$$\mathbf{q} + \tau_q\frac{\partial \mathbf{q}}{\partial t} + \frac{1}{2}\tau_q^2\frac{\partial^2 \mathbf{q}}{\partial t^2} = -k\left[\nabla T + \tau_T\frac{\partial}{\partial t}(\nabla T)\right] \tag{21}$$

For this case, the dual-phase-lagging heat-conduction equation (20) is generalized into

$$\frac{\partial T}{\partial t}+\tau_q\frac{\partial^2 T}{\partial t^2}+\frac{\tau_q^2}{2}\frac{\partial^3 T}{\partial t^3}=\alpha\Delta T+\alpha\tau_T\frac{\partial}{\partial t}(\Delta T)+\frac{\alpha}{k}\left(F+\tau_q\frac{\partial F}{\partial t}+\frac{\tau_q^2}{2}\frac{\partial^2 F}{\partial t^2}\right) \tag{22}$$

which is of hyperbolic type and thus predicts thermal wave propagation with a finite speed [25,69]

$$V_T=\frac{1}{\tau_q}\sqrt{\frac{2k\tau_T}{\rho c}} \tag{23}$$

The thermal wave from Eq. (9) is obviously different from that in Eq. (22). While the former is caused only by the fast-transient effects of thermal inertia, the latter comes from these effects, as well as the delayed response due to the microstructural interaction. Tzou [25] refers to the former wave as the CV-wave and the latter wave as the T-wave. By Eqs. (10) and (23), we have

$$V_T=\sqrt{\frac{2\tau_T}{\tau_q}}V_{\mathrm{CV}} \tag{24}$$

Therefore, the T-wave is always slower than the CV-wave because Eqs. (21) and (22) are valid only for $\tau_q \gg \tau_T$. This has been shown by the heat propagation in superfluid helium at extremely low temperatures [25]. It is interesting to note that Eq. (21) is the simplest constitutive relation that accounts for the dual-phase-lagging effects and yields a heat-conduction equation of hyperbolic type. If the second-order term in $\tau_T$ is also retained, the resulting heat-conduction equation will no longer be hyperbolic [25]. It is also of interest to note that Eq. (22) closely resembles the energy equation describing the ballistic behavior of heat transport in an electron gas [25,53].

Using the DPL relation [Eq. (19)] as the constitutive relation for $\mathbf{q}^t_{\mathrm{mac}}$, Eq. (4) yields

$$\frac{\partial(\rho\varepsilon cT)^t_{\mathrm{mac}}}{\partial t}+\tau_q\frac{\partial^2(\rho\varepsilon cT)^t_{\mathrm{mac}}}{\partial t^2}=\nabla\cdot[(\varepsilon k)^t_{\mathrm{mac}}\nabla T^t_{\mathrm{mac}}]+\tau_T\frac{\partial}{\partial t}\{\nabla\cdot[(\varepsilon k)^t_{\mathrm{mac}}\nabla T^t_{\mathrm{mac}}]\}+\left(1+\tau_q\frac{\partial}{\partial t}\right)(\rho\varepsilon q_m+\rho\varepsilon q_c+\rho\varepsilon q_p+\rho\varepsilon q_e)^t_{\mathrm{mac}} \tag{25}$$

This is known as the DPL model of bioheat transfer. By fitting the experiments data of meat samples in Ref. [28], $\tau_T$ and $\tau_q$ were found to be around 0.05 s and 15 s respectively. The readers are also referred to Refs. [2,75] for the other types of DPL models of bioheat transfer that developed by using Eq. (4) and the other order Taylor expansions of Eq. (16).

The thermal relaxation time is normally large for biological tissue so that the DPL model of bioheat transfer has received increasingly attention. A recent study of bioheat transfer in skin tissue shows that both $\tau_T$ and $\tau_q$ play a significant role for temperature, thermal stress, and thermal damage of skin tissue [76]. The study on the temperature rise behaviors in biological tissues during hyperthermia treatment also reveals their importance at the early stage of heating [77]. The DPL model also predicts significantly different temperature and thermal damage in laser-irradiated biological tissues from both the hyperbolic thermal wave and the Fourier-type Pennes models [78,79].

## 3 Scaling-Up by Porous-Media Theory

To develop a macroscale model of heat transport in living biological tissues, the method of volume averaging starts with a microscale description. Both conservation and constitutive equations are introduced at the microscale. The resulting microscale field equations are then averaged over a representative elementary volume (REV), the smallest differential volume resulting in statistically meaningful local averaging properties, to obtain the macroscale field equations. In the process of averaging, the *multiscale theorems* are used to convert integrals of gradient, divergence, curl, and partial time derivatives of a function into some combination of gradient, divergence, curl, and partial time derivatives of integrals of the function and integrals over the boundary of the REV [6]. Often in order to make averaging procedure tractable and to obtain desirable results, some assumptions are made before, during, and after averaging. These assumptions typically relate to the spatial and/or temporal distribution of properties, expected order of magnitude of various terms, and existence of certain relations among various properties, all based on intuitive and somewhat heuristic arguments.

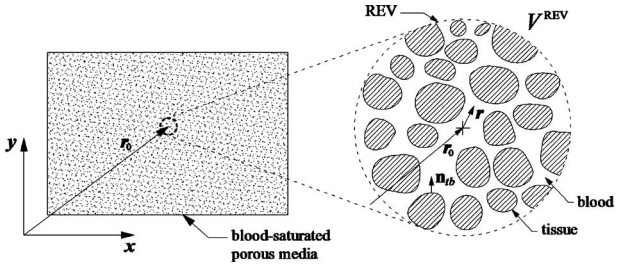



**Fig. 1 Blood-saturated porous media and REV**

The macroscale field equations obtained are not a closed system for determination of velocity, pressure, and temperature because of unclosed terms reflecting the microscale effect. To form a closed system, the approach used for Reynolds-stress closures in turbulence is usually employed to develop governing differential equations and boundary conditions for spatial deviations of pressure, velocity and temperature, the difference between microscale and macroscale values [80–85]. Resulting closure model is a set of differential equations defined on the microscale, which is difficult to solve due to complex microscale geometry. The closure problem is usually solved over a unit cell for a spatially periodic model of a porous medium. This leads to a local closure problem in terms of closure variables and a method of predicting the macroscale parameters in the macroscale models. The readers are referred to Refs. [6,86,87] for the details of the method of volume averaging and to Ref. [6] for a summary of the other methods of obtaining macroscale models.

**3.1 Microscale Model.** For developing a set of the volume-averaged (macroscale) governing equations for the blood flow and bioheat transfer, biological tissue is simplified to be a blood-saturated porous matrix including cells and interstices, the so-called tissue that is considered as a solid matrix [4]. Neglect the gravitational effect and assume that blood is incompressible and Newtonian. By the conservation of mass, momentum, and energy and the Fourier law of heat conduction, we have the microscale model for blood flow and heat conduction in biological tissues (Fig. 1) [4]

In the blood phase,

$$\nabla\cdot\mathbf{v}^b_{\mathrm{mic}}=0 \tag{26}$$

$$\rho^b_{\mathrm{mic}}\frac{\partial\mathbf{v}^b_{\mathrm{mic}}}{\partial t}+\rho^b_{\mathrm{mic}}\mathbf{v}^b_{\mathrm{mic}}\cdot\nabla\mathbf{v}^b_{\mathrm{mic}}=-\nabla p^b_{\mathrm{mic}}+\mu^b_{\mathrm{mic}}\nabla^2\mathbf{v}^b_{\mathrm{mic}} \tag{27}$$

$$(\rho c)^b_{\mathrm{mic}}\frac{\partial \mathrm{T}^b_{\mathrm{mic}}}{\partial t}+(\rho c)^b_{\mathrm{mic}}\mathbf{v}^b_{\mathrm{mic}}\cdot\nabla \mathrm{T}^b_{\mathrm{mic}}=\nabla\cdot(k^b_{\mathrm{mic}}\nabla \mathrm{T}^b_{\mathrm{mic}}) \tag{28}$$

In the tissue phase,

$$(\rho c)^t_{\mathrm{mic}}\frac{\partial \mathrm{T}^t_{\mathrm{mic}}}{\partial t}=\nabla\cdot(k^t_{\mathrm{mic}}\nabla \mathrm{T}^t_{\mathrm{mic}})+(q_m)^t_{\mathrm{mic}} \tag{29}$$

Boundary conditions,

$$\text{B.C.1}\quad \mathbf{v}^b_{\mathrm{mic}}=\mathbf{v}^b_{\mathrm{mic}}|_{A_{bt}}\quad\text{at}\quad A_{bt} \tag{30}$$

$$\text{B.C.2} \quad T^b_{\text{mic}} = T^t_{\text{mic}} \quad \text{at} \quad A_{bt} \tag{31}$$

$$\text{B.C.3} \quad \mathbf{n}_{bt} \cdot k^b_{\text{mic}} \nabla T^b_{\text{mic}} = \mathbf{n}_{bt} \cdot k^t_{\text{mic}} \nabla T^t_{\text{mic}} \quad \text{at} \quad A_{bt} \tag{32}$$

Here the subscript "mic" is used to indicate the microscale properties. Superscripts $b$ and $t$ refer to the blood and tissue phases, respectively. $\mathbf{v}$, $p$, and $T$ are the velocity, the pressure, and the temperature, respectively. $\rho$, $\mu$, $c$, and $k$ are the density, the viscosity, the specific heat, and the thermal conductivity, respectively. $q_m$ is the volumetric rate of heat generation by the metabolic reaction. $A_{bt}$ represents the area of the blood-tissue interface contained in the REV; $\mathbf{n}_{bt}$ is the outward-directed surface normal from the $b$-phase toward the $t$-phase, and $\mathbf{n}_{bt}=-\mathbf{n}_{tb}$ (Fig. 1).

**3.2 Scaling-Up by Volume Averaging.** By applying the superficial averaging process to Eqs. (26), (28), and (29), we have

$$\frac{1}{V^{\text{REV}}}\int_{V^b} \nabla \cdot \mathbf{v}^b_{\text{mic}} dV = 0 \tag{33}$$

$$\frac{1}{V^{\text{REV}}}\int_{V^b} (\rho c)^b_{\text{mic}} \frac{\partial T^b_{\text{mic}}}{\partial t} dV + \frac{1}{V^{\text{REV}}}\int_{V^b} (\rho c)^b_{mic} \mathbf{v}^b_{\text{mic}} \cdot \nabla T^b_{\text{mic}} dV = \frac{1}{V^{\text{REV}}}\int_{V^b} \nabla \cdot (k^b_{\text{mic}} \nabla T^b_{\text{mic}}) dV \tag{34}$$

and

$$\frac{1}{V^{\text{REV}}}\int_{V^t} (\rho c)^t_{\text{mic}} \frac{\partial T^t_{\text{mic}}}{\partial t} dV = \frac{1}{V^{\text{REV}}}\int_{V^t} \nabla \cdot (k^t_{\text{mic}} \nabla T^t_{\text{mic}}) dV + \frac{1}{V^{\text{REV}}}\int_{V^t} (q_m)^t_{\text{mic}} dV \tag{35}$$

where $V^{\text{REV}}$, $V^b$, and $V^t$ are the volumes of the REV, blood in REV, and tissue in REV, respectively. We should note that the superficial temperature is evaluated at the centroid of the REV, whereas the phase temperature is evaluated throughout the REV. Neglecting variations of $\rho c$ within the REV and considering the system to be rigid so that $V^b$ and $V^t$ are time independent, the volume-averaged form of Eqs. (26), (28), and (29) are

$$\langle \nabla \cdot \mathbf{v}^b_{\text{mic}} \rangle = 0 \tag{36}$$

$$(\rho c)^b_{\text{mac}} \frac{\partial \langle T^b_{\text{mic}} \rangle}{\partial t} + (\rho c)^b_{\text{mac}} \langle \mathbf{v}^b_{\text{mic}} \cdot \nabla T^b_{\text{mic}} \rangle = \langle \nabla \cdot (k^b_{\text{mic}} \nabla T^b_{\text{mic}}) \rangle \tag{37}$$

and

$$(\rho c)^t_{\text{mac}} \frac{\partial \langle T^t_{\text{mic}} \rangle}{\partial t} = \langle \nabla \cdot (k^t_{\text{mic}} \nabla T^t_{\text{mic}}) \rangle + \langle (q_m)^t_{\text{mic}} \rangle \tag{38}$$

where the angular brackets indicate superficial quantities such as

$$\langle T^b_{\text{mic}} \rangle = \frac{1}{V^{\text{REV}}}\int_{V^b} T^b_{\text{mic}} dV \tag{39}$$

and

$$\langle T^t_{\text{mic}} \rangle = \frac{1}{V^{\text{REV}}}\int_{V^t} T^t_{\text{mic}} dV \tag{40}$$

Applying the spatial averaging theorem (Theorem 40 in Ref. [6]) to Eq. (36) leads to

$$\langle \nabla \cdot \mathbf{v}^b_{\text{mic}} \rangle = \nabla \cdot \langle \mathbf{v}^b_{\text{mic}} \rangle + \frac{1}{V^{\text{REV}}}\int_{A_{bt}} \mathbf{n}_{bt} \cdot \mathbf{v}^b_{\text{mic}} dA = 0 \tag{41}$$

The surface integral represents the volumetric rate of blood bleeding off to the solid matrix through the interfacial vascular wall. Most microcirculatory systems are with a capillary blood filtration larger than reabsorption so that there is a net filtration of blood from the intravascular to the extravascular regions. However, this surface integral is negligibly small because the lymphatic system brings the excess blood from the interstitium to the intravascular compartment [4]. Therefore, Eq. (41) reduces into

$$\nabla \cdot \langle \mathbf{v}^b_{\text{mic}} \rangle = 0 \tag{42}$$

Note that the superficial average defined in Eqs. (39) and (40) is an unsuitable macroscale variable because it can yield erroneous results. For example, if the temperature of the blood were constant, the superficial average would differ from it [88]. On the other hand, intrinsic phase averages do not have this shortcoming. These averages are defined by

$$\langle T^b_{\text{mic}} \rangle^b = \frac{1}{V^b}\int_{V^b} T^b_{\text{mic}} dV \tag{43}$$

and

$$\langle T^t_{\text{mic}} \rangle^t = \frac{1}{V^t}\int_{V^t} T^t_{\text{mic}} dV \tag{44}$$

Also, intrinsic averages are related to superficial averages by

$$\langle T^b_{\text{mic}} \rangle = \varepsilon^b \langle T^b_{\text{mic}} \rangle^b \tag{45}$$

and

$$\langle T^t_{\text{mic}} \rangle = \varepsilon^t \langle T^t_{\text{mic}} \rangle^t \tag{46}$$

where $\varepsilon^b$ and $\varepsilon^t$ are the volume fractions of the blood and tissue with $\varepsilon^t = 1-\varepsilon^b$.

By applying Eqs. (42), (45), and (46) and the spatial averaging theorem (Theorem 40 in Ref. [6]) and neglecting variations of physical properties within the REV, Eqs. (37) and (38) become

$$\underbrace{(\rho c \varepsilon)^b_{\text{mac}} \frac{\partial \langle T^b_{mic} \rangle^b}{\partial t}}_{\text{accumulation}} + \underbrace{(\rho c \varepsilon)^b_{\text{mac}} \langle \mathbf{v}^b_{\text{mic}} \rangle^b \cdot \nabla \langle T^b_{\text{mic}} \rangle^b}_{\text{convection}} = \underbrace{\nabla \cdot \left\{ k^b_{\text{mac}} \left[ \varepsilon^b_{\text{mac}} \nabla \langle T^b_{\text{mic}} \rangle^b + \frac{1}{V^{\text{REV}}}\int_{A_{bt}} \mathbf{n}_{bt} \tilde{T}^b_{\text{mic}} dA \right] \right\}}_{\text{conduction}}$$

$$\underbrace{-(\rho c)^b_{\text{mac}} \nabla \cdot \langle \tilde{\mathbf{v}}^b_{\text{mic}} \tilde{T}^b_{\text{mic}} \rangle}_{\text{dispersion}} + \underbrace{\frac{1}{V^{\text{REV}}}\int_{A_{bt}} \mathbf{n}_{bt} \cdot k^b_{\text{mic}} \nabla T^b_{\text{mic}} dA}_{\text{interfacial flux}} - \underbrace{(\rho c)^b_{\text{mac}} \frac{1}{V^{\text{REV}}}\int_{A_{bt}} \mathbf{n}_{bt} \cdot \mathbf{v}^b_{\text{mic}} T^b_{\text{mic}} dA}_{\text{perfusion}} \tag{47}$$

and

$$\underbrace{(\rho c\varepsilon)^t_{\mathrm{mac}}\frac{\partial\langle T^t_{\mathrm{mic}}\rangle^t}{\partial t}}_{\text{accumulation}} = \underbrace{\nabla\cdot\left\{k^t_{\mathrm{mac}}\left[\varepsilon^t_{\mathrm{mac}}\nabla\langle T^t_{\mathrm{mic}}\rangle^t + \frac{1}{V^{\mathrm{REV}}}\int_{A_{bt}}\mathbf{n}_{tb}\widetilde{T}^t_{\mathrm{mic}}dA\right]\right\}}_{\text{conduction}} + \underbrace{\frac{1}{V^{\mathrm{REV}}}\int_{A_{bt}}\mathbf{n}_{tb}\cdot k^t_{\mathrm{mic}}\nabla T^t_{\mathrm{mic}}dA}_{\text{interfacial flux}} + \underbrace{\varepsilon^t_{\mathrm{mac}}\langle(q_m)^t_{\mathrm{mic}}\rangle^t}_{\substack{\text{metabolic}\\ \text{thermal source}}}$$
$$\underbrace{-(\rho c)^b_{\mathrm{mac}}\frac{1}{V^{\mathrm{REV}}}\int_{A_{bt}}\mathbf{n}_{tb}\cdot\mathbf{v}^b_{\mathrm{mic}}\mathrm{T}^b_{\mathrm{mic}}dA}_{\text{perfusion}} \tag{48}$$

Here $\tilde{\mathbf{v}}^b_{\mathrm{mic}}=\mathbf{v}^b_{\mathrm{mic}}-\langle\mathbf{v}^b_{\mathrm{mic}}\rangle^b$, $\widetilde{T}^b_{\mathrm{mic}}=T^b_{\mathrm{mic}}-\langle T^b_{\mathrm{mic}}\rangle^b$, and $\widetilde{T}^t_{\mathrm{mic}}=T^t_{\mathrm{mic}}-\langle T^t_{\mathrm{mic}}\rangle^t$. A rigorous closure is not available at present for these spatial deviation velocity and temperature in the context of bioheat transfer. Nakayama and Kuwahara [4] recently modeled the blood interfacial convective heat transfer by using the Newton law of cooling and approximated the effect of the blood perfusion by using the perfusion rate and macroscale temperature difference between blood and tissues. This leads to a simplified two-equation macroscale model

$$(\rho c\varepsilon)^b_{\mathrm{mac}}\left(\frac{\partial T^b_{\mathrm{mac}}}{\partial t}+\mathbf{v}^b_{\mathrm{mac}}\cdot\nabla T^b_{\mathrm{mac}}\right)=\nabla\cdot(\mathbf{K}^b_{\mathrm{mac}}\cdot\nabla T^b_{\mathrm{mac}})+(\rho\varepsilon q_c)^b_{\mathrm{mac}}+(\rho\varepsilon q_p)^b_{\mathrm{mac}} \tag{49}$$

and

$$(\rho c\varepsilon)^t_{\mathrm{mac}}\frac{\partial T^t_{\mathrm{mac}}}{\partial t}=\nabla\cdot(\mathbf{K}^t_{\mathrm{mac}}\cdot\nabla T^t_{\mathrm{mac}})+(\rho\varepsilon q_c)^t_{\mathrm{mac}}+(\rho\varepsilon q_p)^t_{\mathrm{mac}}+(\varepsilon q_m)^t_{\mathrm{mac}} \tag{50}$$

where

$$T^b_{\mathrm{mac}}=\langle T^b_{\mathrm{mic}}\rangle^b \tag{51}$$

$$T^t_{\mathrm{mac}}=\langle T^t_{\mathrm{mic}}\rangle^t \tag{52}$$

$$\mathbf{K}^b_{\mathrm{mac}}=(\varepsilon\mathbf{k})^b_{\mathrm{mac}}+(\mathbf{k}_{\mathrm{dis}})^{bt}_{\mathrm{mac}} \tag{53}$$

$$\mathbf{K}^t_{\mathrm{mac}}=(\varepsilon\mathbf{k})^t_{\mathrm{mac}} \tag{54}$$

$$(\rho\varepsilon q_c)^b_{\mathrm{mac}}=-(\rho\varepsilon q_c)^b_{\mathrm{mac}}=-ha(T^b_{\mathrm{mac}}-T^t_{\mathrm{mac}}) \tag{55}$$

$$(\rho\varepsilon q_p)^b_{\mathrm{mac}}=-(\rho\varepsilon q_p)^b_{\mathrm{mac}}=-(\rho c\omega)^b_{\mathrm{mac}}(T^b_{\mathrm{mac}}-T^t_{\mathrm{mac}}) \tag{56}$$

Here $\mathbf{K}^b_{\mathrm{mac}}$ and $\mathbf{K}^t_{\mathrm{mac}}$ are the effective thermal conductivity tensor of blood and tissue, respectively, $\mathbf{k}$ is the thermal conductivity tensor, $\mathbf{k}_{\mathrm{dis}}$ is the thermal dispersion conductivity tensor, $h$ is the convective heat transfer coefficient ($\mathrm{W/m^2\,K}$), $a$ is the volumetric contact area between tissue and blood ($\mathrm{m^2/m^3}$), and $\omega$ is the perfusion rate ($\mathrm{kg/m^3\,s}$). Therefore, the microscale effects are represented by $\mathbf{k}_{\mathrm{dis}}$, $ha$, and $\omega$ in Eqs. (49) and (50). A rigorous theory regarding $\mathbf{k}_{\mathrm{dis}}$, $ha$, and $\omega$ is, however, not available at present.

A similar averaging has also been applied to blood momentum equation (Eq. (27)). The result is available, for example, in Refs. [84–86,88]. To examine the countercurrent heat exchange between the arterial and venous blood vessels in the circulatory systems, a three-energy equation model has also been developed in Ref. [4] by following a similar approach.

When the system is isotropic and the physical properties of blood and tissue are constant, Eqs. (49) and (50) reduce to

$$\gamma^b_{\mathrm{mac}}\left(\frac{\partial T^b_{\mathrm{mac}}}{\partial t}+\mathbf{v}^b_{\mathrm{mac}}\cdot\nabla T^b_{\mathrm{mac}}\right)=(k_{\mathrm{eff}})^b_{\mathrm{mac}}\Delta T^b_{\mathrm{mac}}-ha(T^b_{\mathrm{mac}}-T^t_{\mathrm{mac}})-(\rho c\omega)^b_{\mathrm{mac}}(T^b_{\mathrm{mac}}-T^t_{\mathrm{mac}}) \tag{57}$$

and

$$\gamma^t_{\mathrm{mac}}\frac{\partial T^t_{\mathrm{mac}}}{\partial t}=(k_{\mathrm{eff}})^t_{\mathrm{mac}}\Delta T^t_{\mathrm{mac}}+ha(T^b_{\mathrm{mac}}-T^t_{\mathrm{mac}})+(\rho c\omega)^b_{\mathrm{mac}}(T^b_{\mathrm{mac}}-T^t_{\mathrm{mac}})+(\varepsilon q_m)^t_{\mathrm{mac}} \tag{58}$$

where the effective thermal conductivities $(k_{\mathrm{eff}})^b_{\mathrm{mac}}$ and $(k_{\mathrm{eff}})^t_{\mathrm{mac}}$ are

$$(k_{\mathrm{eff}})^b_{\mathrm{mac}}=(\varepsilon k)^b_{\mathrm{mac}}+(k_{\mathrm{dis}})^{bt}_{\mathrm{mac}} \tag{59}$$

$$(k_{\mathrm{eff}})^t_{\mathrm{mac}}=(\varepsilon k)^t_{\mathrm{mac}} \tag{60}$$

The effective thermal capacities $\gamma^b_{\mathrm{mac}}$ and $\gamma^t_{\mathrm{mac}}$ are

$$\gamma^b_{\mathrm{mac}}=(\rho c\varepsilon)^b_{\mathrm{mac}} \tag{61}$$

$$\gamma^t_{\mathrm{mac}}=(\rho c\varepsilon)^t_{\mathrm{mac}} \tag{62}$$

**3.3 Analysis.** Rewrite Eqs. (57) and (58) in their operator form

$$\begin{bmatrix}\gamma^b_{\mathrm{mac}}\frac{\partial}{\partial t}+\gamma^b_{\mathrm{mac}}\mathbf{v}^b_{\mathrm{mac}}\cdot\nabla-(k_{\mathrm{eff}})^b_{\mathrm{mac}}\Delta+G & -G\\ -G & \gamma^t_{\mathrm{mac}}\frac{\partial}{\partial t}-(k_{\mathrm{eff}})^t_{\mathrm{mac}}\Delta+G\end{bmatrix}\times\begin{bmatrix}T^b_{\mathrm{mac}}\\ T^t_{\mathrm{mac}}\end{bmatrix}=\begin{bmatrix}0\\ (\varepsilon q_m)^t_{\mathrm{mac}}\end{bmatrix} \tag{63}$$

where the lumped convection-perfusion coefficient $G$ is

$$G=ha+(\rho c\omega)^b_{\mathrm{mac}} \tag{64}$$

We then obtain an uncoupled form by evaluating the operator determinant and dividing by $G(\gamma^b_{\mathrm{mac}}+\gamma^t_{\mathrm{mac}})$

$$\frac{\partial T^b_{\mathrm{mac}}}{\partial t}+\tau_q\frac{\partial^2 T^b_{\mathrm{mac}}}{\partial t^2}+\frac{\gamma^b_{\mathrm{mac}}}{\gamma^b_{\mathrm{mac}}+\gamma^t_{\mathrm{mac}}}\mathbf{v}^b_{\mathrm{mac}}\cdot\nabla T^b_{\mathrm{mac}}=\alpha\Delta T^b_{\mathrm{mac}}+\alpha\tau_T\frac{\partial}{\partial t}(\Delta T^b_{\mathrm{mac}})+\frac{\alpha}{k}\left[F(\mathbf{r},t)+\tau_q\frac{\partial F(\mathbf{r},t)}{\partial t}\right]^b_{\mathrm{mac}} \tag{65}$$

$$\frac{\partial T^t_{\mathrm{mac}}}{\partial t}+\tau_q\frac{\partial^2 T^t_{\mathrm{mac}}}{\partial t^2}+\frac{\gamma^b_{\mathrm{mac}}}{\gamma^b_{\mathrm{mac}}+\gamma^t_{\mathrm{mac}}}\mathbf{v}^b_{\mathrm{mac}}\cdot\nabla T^t_{\mathrm{mac}}=\alpha\Delta T^t_{\mathrm{mac}}+\alpha\tau_T\frac{\partial}{\partial t}(\Delta T^t_{\mathrm{mac}})+\frac{\alpha}{k}\left[F(\mathbf{r},t)+\tau_q\frac{\partial F(\mathbf{r},t)}{\partial t}\right]^t_{\mathrm{mac}} \tag{66}$$

where

$$\tau_q = \frac{\gamma^b_{\text{mac}}\gamma^t_{\text{mac}}}{G(\gamma^b_{\text{mac}} + \gamma^t_{\text{mac}})} \tag{67}$$

$$\tau_T = \frac{\gamma^b_{\text{mac}}(k_{\text{eff}})^t_{\text{mac}} + \gamma^t_{\text{mac}}(k_{\text{eff}})^b_{\text{mac}}}{G[(k_{\text{eff}})^b_{\text{mac}} + (k_{\text{eff}})^t_{\text{mac}}]} \tag{68}$$

$$k = (k_{\text{eff}})^b_{\text{mac}} + (k_{\text{eff}})^t_{\text{mac}} \tag{69}$$

$$\rho c = \gamma^b_{\text{mac}} + \gamma^t_{\text{mac}} \tag{70}$$

$$\alpha = \frac{k}{\rho c} = \frac{(k_{\text{eff}})^b_{\text{mac}} + (k_{\text{eff}})^t_{\text{mac}}}{\gamma^b_{\text{mac}} + \gamma^t_{\text{mac}}} \tag{71}$$

$$\left[F(\mathbf{r},t) + \tau_q \frac{\partial F(\mathbf{r},t)}{\partial t}\right]^b_{\text{mac}} = (\varepsilon q_m)^t_{\text{mac}} - \frac{(k_{\text{eff}})^b_{\text{mac}}(k_{\text{eff}})^t_{\text{mac}}}{G}\Delta^2 T^b_{\text{mac}} - \frac{\gamma^b_{\text{mac}}}{G}\mathbf{v}^b_{\text{mac}} \cdot \left[\gamma^t_{\text{mac}}\frac{\partial}{\partial t}\nabla T^b_{\text{mac}} - (k_{\text{eff}})^t_{\text{mac}}\nabla \Delta T^b_{\text{mac}}\right] \tag{72}$$

$$\left[F(\mathbf{r},t) + \tau_q \frac{\partial F(\mathbf{r},t)}{\partial t}\right]^t_{\text{mac}} = -\frac{(k_{\text{eff}})^b_{\text{mac}}(k_{\text{eff}})^t_{\text{mac}}}{G}\Delta^2 T^t_{\text{mac}} - \frac{\gamma^b_{\text{mac}}}{G}\mathbf{v}^b_{\text{mac}} \cdot \left[\gamma^t_{\text{mac}}\frac{\partial}{\partial t}\nabla T^t_{\text{mac}} - (k_{\text{eff}})^t_{\text{mac}}\nabla \Delta T^t_{\text{mac}}\right] + \frac{1}{G}\left[\gamma^b_{\text{mac}}\frac{\partial}{\partial t} + \gamma^b_{\text{mac}}\mathbf{v}^b_{\text{mac}} \cdot \nabla - (k_{\text{eff}})^b_{\text{mac}}\Delta + G\right](\varepsilon q_m)^t_{\text{mac}} \tag{73}$$

Therefore, both $T^b_{\text{mac}}$ and $T^t_{\text{mac}}$ are governed by dual-phase-lagging heat-conduction equations (Eqs. (65) and (66)) with $\tau_q$ and $\tau_T$ as the phase lags of the heat flux and the temperature gradient, respectively [21,25,89]. Here, $F(\mathbf{r},t)$ is the volumetric heat source. $k$, $\rho c$, and $\alpha$ are the effective thermal conductivity, capacity, and diffusivity, respectively. While the heat conduction in blood and tissue is assumed to be Fourier-type at microscale (Eqs. (28) and (29)), it is DPL-type at macroscale. It is thus more proper to use the DPL constitutive relation for the heat flux density vector in developing macroscale bioheat equations via scaling-down from the global scale based on the mixture theory of continuum mechanics.

Although both Eqs. (25) and (66) are of DPL-type for macroscale tissue temperature, the distinct difference exists between them. The explicit relation between $T^t_{\text{mic}}$ and $T^t_{\text{mac}}$ and the way for computing $\tau_q$ and $\tau_T$ are not available in the former; they are available in the latter (Eqs. (52), (67), and (68)). The former involves both $T^b_{\text{mac}}$ and $T^t_{\text{mac}}$; the latter contains only the tissue temperature $T^t_{\text{mac}}$. The effect of blood conduction and convection is not considered in the former; it is incorporated in the latter.

There is a metabolic heat term $(\varepsilon q_m)^t_{\text{mac}}$ in the blood energy equation (Eqs. (65) and (72)) and a convective term $\mathbf{v}^b_{\text{mac}} \cdot \nabla T^t_{\text{mac}}$ in the tissue energy equation (Eq. (66)). Therefore, both microscale metabolic heat generation in tissue and microscale convection in blood are with their macroscale manifestation in both blood and tissue. The interaction between blood and tissue also yields a very rich way that the interfacial convective heat transfer, the blood velocity, the perfusion, and the metabolic reaction affect $T^b_{\text{mac}}$ and $T^t_{\text{mac}}$ (Eqs. (65), (66), (72), and (73)). It would be very difficult to model this rich interaction by the mixture theory of continuum mechanics.

Consider

$$\frac{\tau_T}{\tau_q} = 1 + \frac{(\gamma^2)^b_{\text{mac}}(k_{\text{eff}})^t_{\text{mac}} + (\gamma^2)^t_{\text{mac}}(k_{\text{eff}})^b_{\text{mac}}}{\gamma^b_{\text{mac}}\gamma^t_{\text{mac}}[(k_{\text{eff}})^b_{\text{mac}} + (k_{\text{eff}})^t_{\text{mac}}]} \geq 1 \tag{74}$$

By the condition for the existence of thermal waves that requires $\tau_T/\tau_q < 1$ [21,72], therefore, there is no thermal wave in bioheat transfer based on the model in Eqs. (65) and (66). It is also interesting to note that although each $\tau_q$ and $\tau_T$ is $G$ dependent, the ratio $\tau_T/\tau_q$ is not. Therefore the evaluation of $\tau_T/\tau_q$ will be much simpler than $\tau_q$ or $\tau_T$. Based on some simplified versions of Eq. (66), the lagging behavior has been recently examined in Refs. [75,90].

## 4 Concluding Remarks

Macroscale thermal models have been developed for biological tissues either by the mixture theory or by the porous-media theory. The former considers blood and tissues as a mixture of continuum deformable media and develops the macroscale point equations via scaling-down the global balance equations. In this approach, neither microscale presentation of the system nor microscale quantities are introduced. Phase properties are defined at the macroscale. The global balance equations are formed in terms of macroscale properties and with additional terms accounting for the interaction between blood and tissue. Required constitutive equations for the heat flux vector are supplied by the Fourier law, the CV relation, or the DPL relation. The thermal models developed in this approach include the classical Pennes model, the Wulff model, the Klinger model, and the Chen and Holmes model, the thermal wave bioheat model, and the DPL bioheat model. The heterogeneous and nonisotropic feature of biological tissue yields normally a strong noninstantaneous response between heat flux and temperature gradient in nonequilibrium heat transport. The DPL bioheat model is thus recommended in order to catch such a lagging behavior.

The porous-media theory considers biological tissue as a blood-saturated porous matrix, including cells and interstices, and develops the macroscale point equations via scaling-up the microscale model. In this approach, both conservation and constitutive equations are introduced at the microscale. The resulting microscale field equations are then averaged over a REV to obtain the macroscale field equations. In the process of averaging, the multiscale theorems are used to convert integrals of gradient, divergence, curl, and partial time derivatives of a function into some combination of gradient, divergence, curl, and partial time derivatives of integrals of the function and integrals over the boundary of the REV. The closure model must be provided for the unclosed terms in macroscale field equations that represent the microscale effect in order to form a closed system. The macroscale model developed by this approach shows the DPL-type bioheat transfer at macroscale for both blood and tissue phases and the sophisticated effect of the interfacial convective heat transfer, the blood velocity, the perfusion, and the metabolic reaction on macroscale temperature fields in blood and tissue.

Therefore, the mixture theory and porous-media approaches are top-down and bottom-up approaches, respectively, in nature. Simplicity is the main advantage of the former. However, it offers no connection between microscale and macroscale properties and is not capable to accurately describe the rich blood-tissue interaction. The porous-media approach successfully overcomes these drawbacks, thereby offering an effective way for developing accurate macroscale thermal models for biological tissues. The requirement of a rigorous closure model presents, however, unique challenges for materializing its promising potential. Future research is in great demand to define the potential of this promising approach by developing a rigorous closure theory.

## Acknowledgment

The financial support from the Research Grants Council of Hong Kong (Grant GRF718009) is gratefully acknowledged.

## References


[1] Cho, Y. I., ed., 1992, *Bioengineering Heat Transfer* (Advances in Heat Transfer Series No. 22), Academic, London.

[2] Xu, F., Lu, T. J., Seffen, K. A., and Ng, E. Y. K., 2009, "Mathematical Modeling of Skin Bioheat Transfer," Appl. Mech. Rev., **62**, p. 050801.

[3] Vadasz, P., ed., 2008, *Emerging Topics in Heat and Mass Transfer in Porous Media: From Bioengineering and Microelectronics to Nanotechnology* (Theory and Applications of Transport in Porous Media Series No. 22), Springer-Verlag, Berlin.

[4] Nakayama, A., and Kuwahara, F., 2008, "A General Bioheat Transfer Model Based on the Theory of Porous Media," Int. J. Heat Mass Transfer, **51**, pp. 3190–3199.

[5] Khaled, A. R. A., and Vafai, K., 2003, "The Role of Porous Media in Modeling Flow and Heat Transfer in Biological Tissues," Int. J. Heat Mass Transfer, **46**, pp. 4989–5003.

[6] Wang, L. Q., Xu, M. T., and Wei, X. H., 2008, "Multiscale Theorems," Adv. Chem. Eng., **34**, pp. 175–468.

[7] Auriault, J. L., 1991, "Heterogeneous Medium: Is an Equivalent Macroscopic Description Possible?," Int. J. Eng. Sci., **29**, pp. 785–795.

[8] Eringen, A. C., and Ingram, J. D., 1965, "A Continuum Theory of Chemically Reacting Media," Int. J. Eng. Sci., **3**, pp. 197–212.

[9] Wang, L. Q., 1994, "Generalized Fourier Law," Int. J. Heat Mass Transfer, **37**, pp. 2627–2634.

[10] Cengel, Y. A., and Boles, M. A., 2006, *Thermodynamics: An Engineering Approach*, 5th ed., McGraw-Hill, Boston.

[11] Wang, L. Q., 1995, "Properties of Heat Flux Functions and a Linear Theory of Heat Flux," Int. J. Mod. Phys. B, **9**, pp. 1113–1122.

[12] Wang, L. Q., 2001, "Further Contributions on the Generalized Fourier Law," Int. J. Transp. Phenom., **2**, pp. 299–305.

[13] Wang, L. Q., 1996, "A Decomposition Theorem of Motion," Int. J. Eng. Sci., **34**, pp. 417–423.

[14] Pennes, H. H., 1948, "Analysis of Tissue and Arterial Blood Temperature in the Resting Human Forearm," J. Appl. Physiol., **1**, pp. 93–122.

[15] Wulff, W., 1974, "The Energy Conservation Equation for Living Tissues," IEEE Trans. Biomed. Eng., **BME-21**, pp. 494–495.

[16] Klinger, H. G., 1974, "Heat Transfer in Perfused Tissue I: General Theory," Bull. Math. Biol., **36**, pp. 403–415.

[17] Chen, M. M., and Holmes, K. R., 1980, "Microvascular Contributions in Tissue Heat Transfer," Ann. N.Y. Acad. Sci., **335**, pp. 137–150.

[18] Cattaneo, C., 1958, "A Form of Heat Conduction Equation Which Eliminates the Paradox of Instantaneous Propagation," Compt. Rend., **247**, pp. 431–433.

[19] Vernotte, P., 1958, "Les Paradoxes de la Théorie Continue de l'equation de la Chaleur," Compt. Rend., **246**, pp. 3154–3155.

[20] Vernotte, P., 1961, "Some Possible Complications in the Phenomena of Thermal Conduction," Compt. Rend., **252**, pp. 2190–2191.

[21] Wang, L. Q., Zhou, X. S., and Wei, X. H., 2008, *Heat Conduction: Mathematical Models and Analytical Solutions*, Springer-Verlag, Heidelberg.

[22] Tzou, D. Y., 1992, "Thermal Shock Phenomena Under High-Rate Response in Solids," Annu. Rev. Heat Transfer, **4**, pp. 111–185.

[23] Chandrasekharaiah, D. S., 1986, "Thermoelasticity With Second Sound: A Review," Appl. Mech. Rev., **39**, pp. 355–376.

[24] Chandrasekharaiah, D. S., 1998, "Hyperbolic Thermoelasticity: A Review of Recent Literature," Appl. Mech. Rev., **51**, pp. 705–729.

[25] Tzou, D. Y., 1997, *Macro-to Microscale Heat Transfer: The Lagging Behavior*, Taylor & Francis, Washington, DC.

[26] Herwig, H., and Beckert, K., 2000, "Experimental Evidence About the Controversy Concerning Fourier or Non-Fourier Heat Conduction in Materials With a Nonhomogeneous Inner Structure," Heat Mass Transfer, **36**, pp. 387–392.

[27] Kaminski, W., 1990, "Hyperbolic Heat Conduction Equation for Materials With a Nonhomogeneous Inner Structure," ASME J. Heat Transfer, **112**, pp. 555–560.

[28] Mitra, K., Kumar, S., Vedavarz, A., and Moallemi, M. K., 1995, "Experimental Evidence of Hyperbolic Heat Conduction in Processed Meat," ASME J. Heat Transfer, **117**, pp. 568–573.

[29] Graßmann, A., and Peters, F., 1999, "Experimental Investigation of Heat Conduction in Wet Sand," Heat Mass Transfer, **35**, pp. 289–294.

[30] Roetzel, W., Putra, N., and Das, S. K., 2003, "Experiment and Analysis for Non-Fourier Conduction in Materials With Non-Homogeneous Inner Structure," Int. J. Therm. Sci., **42**, pp. 541–552.

[31] Vedavarz, A., Mitra, K., Kumar, S., and Moallemi, M. K., 1992, "Effect of Hyperbolic Heat Conduction on Temperature Distribution in Laser Irradiated Tissue With Blood Perfusion," Adv. Bio. Heat Mass Transfer, ASME HTD, **231**, pp. 7–16.

[32] Luikov, A. V., 1966, "Application of Irreversible Thermodynamics Methods to Investigation of Heat and Mass Transfer," Int. J. Heat Mass Transfer, **9**, pp. 139–152.

[33] Tzou, D. Y., 1995, "A Unified Field Approach for Heat Conduction From Micro- to Macro-Scales," ASME J. Heat Transfer, **117**, pp. 8–16.

[34] Joseph, D. D., and Preziosi, L., 1989, "Heat Waves," Rev. Mod. Phys., **61**, pp. 41–73.

[35] Joseph, D. D., and Preziosi, L., 1990, "Addendum to the Paper Heat Waves," Rev. Mod. Phys., **62**, pp. 375–391.

[36] Wang, L. Q., 2000, "Solution Structure of Hyperbolic Heat-Conduction Equation," Int. J. Heat Mass Transfer, **43**, pp. 365–373.

[37] Liu, J., Zhang, X. X., Wang, C. C., and Liu, W. Q., 1997, "Engineering Investigation on Medical Application Approaches for the Thermal Wave Effects in Living Tissue," Space Med. Med. Eng. (Beijing), **10**, pp. 135–139.

[38] Liu, J., Zhang, X. X., and Liu, W. Q., 1999, "The Thermal Pulse Decay Method for Invasive Measurement of Blood Perfusion of Tissue in Vivo," Prog. Nat. Sci., **9**, pp. 179–184.

[39] Zhu, T. C., and Feng, X. Z., 2001, "Numerical Analysis of the Relationship Between Blood Flow Coefficient and Living Tissue Thermal Behavior," Chin. J. Hemort., **11**, pp. 182–183.

[40] Chato, J. C., and Lee, R. C., 1998, "The Future of Biothermal Emerging," Ann. N.Y. Acad. Sci., **858**, pp. 1–20.

[41] Deng, Z. S., and Liu, J., 2003, "Non-Fourier Hear Conduction Effect on Prediction of Temperature Transients and Thermal Stress in Skin Cryopreservation," J. Therm. Stresses, **26**, pp. 779–798.

[42] Shih, T. C., Kou, H. S., Liauh, C. T., and Lin, W. L., 2005, "The Impact of Thermal Wave Characteristics on Thermal Dose Distribution During Thermal Therapy: A Numerical Study," Med. Phys., **32**, pp. 3029–3036.

[43] Liu, J., 2000, "Preliminary Survey on the Mechanisms of the Wave-Like Behaviors of Heat Transfer in Living Tissues," Forsch. Ingenieurwes., **66**, pp. 1–10.

[44] Liu, J., Chen, X., and Xu, L. X., 1999, "New Thermal Wave Aspects on Burn Evaluation of Skin Subjected to Instantaneous Heating," IEEE Trans. Biomed. Eng., **46**, pp. 420–428.

[45] Ma, N., Jiang, S., Li, H., and Zhang, X., 2003, "Analysis of Non-Fourier Effect and Laser-Induced Thermal Damage of Laser-Irradiated Layered Human Skin Tissue," Space Med. Med. Eng. (Beijing), **16**, pp. 133–137.

[46] Tung, M. M., Trujillo, M., Lopez-Molina, J. A., Rivera, M. J., and Berjano, E. J., 2009, "Modeling the Heating of Biological Tissue Based on the Hyperbolic Heat Transfer Equation," Math. Comput. Modell., **50**, pp. 665–672.

[47] Xu, M. T., and Wang, L. Q., 2005, "Dual-Phase-Lagging Heat Conduction Based on Boltzmann Transport Equation," Int. J. Heat Mass Transfer, **48**, pp. 5616–5624.

[48] Cheng, L., Xu, M. T., and Wang, L. Q., 2008, "From Boltzmann Transport Equation to Single-Phase-Lagging Heat Conduction," Int. J. Heat Mass Transfer, **51**, pp. 6018–6023.

[49] Cheng, L., Xu, M. T., and Wang, L. Q., 2008, "Single- and Dual-Phase-Lagging Heat Conduction Models in Moving Media," ASME J. Heat Transfer, **130**, pp. 121302.

[50] Guyer, R. A., and Krumhansi, J. A., 1966, "Solution of the Linearized Boltzmann Equation," Phys. Rev., **148**, pp. 766–778.

[51] Anisimòv, S. I., Kapeliovich, B. L., and Perelman, T. L., 1974, "Electron Emission From Metal Surfaces Exposed to Ultra-Short Laser Pulses," Sov. Phys. JETP, **39**, pp. 375–377.

[52] Kaganov, M. I., Lifshitz, I. M., and Tanatarov, M. V., 1957, "Relaxation Between Electrons and Crystalline Lattices," Sov. Phys. JETP, **4**, pp. 173–178.

[53] Qiu, T. Q., and Tien, C. L., 1993, "Heat Transfer Mechanisms During Short-Pulse Laser Heating of Metals," ASME J. Heat Transfer, **115**, pp. 835–841.

[54] Tzou, D. Y., and Zhang, Y. S., 1995, "An Analytical Study on the Fast-Transient Process in Small Scales," Int. J. Eng. Sci., **33**, pp. 1449–1463.

[55] Vadasz, P., 2005, "Absence of Oscillations and Resonance in Porous Media Dual-Phase-Lagging Fourier Heat Conduction," ASME J. Heat Transfer, **127**, pp. 307–314.

[56] Vadasz, P., 2005, "Explicit Conditions for Local Thermal Equilibrium in Porous Media Heat Conduction," Transp. Porous Media, **59**, pp. 341–355.

[57] Vadasz, P., 2005, "Lack of Oscillations in Dual-Phase-Lagging Heat Conduction for a Porous Slab Subject to Imposed Heat Flux and Temperature," Int. J. Heat Mass Transfer, **48**, pp. 2822–2828.

[58] Vadasz, P., 2006, "Exclusion of Oscillations in Heterogeneous and Bi-Composite Media Thermal Conduction," Int. J. Heat Mass Transfer, **49**, pp. 4886–4892.

[59] Vadasz, P., 2006, "Heat Conduction in Nanofluid Suspensions," ASME J. Heat Transfer, **128**, pp. 465–477.

[60] Wang, L. Q., and Wei, X. H., 2008, "Equivalence Between Dual-Phase-Lagging and Two-Phase-System Heat Conduction Processes," Int. J. Heat Mass Transfer, **51**, pp. 1751–1756.

[61] Wang, L. Q., and Wei, X. H., 2009, "Nanofluids: Synthesis, Heat Conduction and Extension," ASME J. Heat Transfer, **131**, pp. 033102.

[62] Wang, L. Q., and Wei, X. H., 2009, "Heat Conduction in Nanofluids," Chaos, Solitons Fractals, **39**, pp. 2211–2215.

[63] Wang, L. Q., and Xu, M. T., 2002, "Well-Posedness of Dual-Phase-Lagging Heat Conduction Equation: Higher Dimensions," Int. J. Heat Mass Transfer, **45**, pp. 1165–1171.

[64] Dai, W. Z., and Nassar, R., 2002, "An Approximate Analytical Method for Solving 1D Dual-Phase-Lagging Heat Transfer Equations," Int. J. Heat Mass Transfer, **45**, pp. 1585–1593.

[65] Antaki, P. J., 1998, "Solution for Non-Fourier Dual Phase Lag Heat Conduction in a Semi-Infinite Slab With Surface Heat Flux," Int. J. Heat Mass Transfer, **41**, pp. 2253–2258.

[66] Dai, W. Z., and Nassar, R., 1999, "A Finite Difference Scheme for Solving the Heat Transport Equation at the Microscale," Numer. Methods Partial Differ. Equ., **15**, pp. 697–708.

[67] Lin, C. K., Hwang, C. C., and Chang, Y. P., 1997, "The Unsteady Solutions of a Unified Heat Conduction Equation," Int. J. Heat Mass Transfer, **40**, pp. 1716–1719.

[68] Tang, D. W., and Araki, N., 1999, "Wavy, Wavelike, Diffusive Thermal Responses of Finite Rigid Slabs to High-Speed Heating of Laser-Pulses," Int. J. Heat Mass Transfer, **42**, pp. 855–860.

[69] Tzou, D. Y., 1995, "The Generalized Lagging Response in Small-Scale and

High-Rate Heating," Int. J. Heat Mass Transfer, **38**, pp. 3231–3240.
[70] Tzou, D. Y., and Chiu, K. S., 2001, "Temperature-Dependent Thermal Lagging in Ultrafast Laser Heating," Int. J. Heat Mass Transfer, **44**, pp. 1725–1734.
[71] Wang, L. Q., Xu, M. T., and Zhou, X. S., 2001, "Well-Posedness and Solution Structure of Dual-Phase-Lagging Heat Conduction," Int. J. Heat Mass Transfer, **44**, pp. 1659–1669.
[72] Xu, M. T., and Wang, L. Q., 2002, "Thermal Oscillation and Resonance in Dual-Phase-Lagging Heat Conduction," Int. J. Heat Mass Transfer, **45**, pp. 1055–1061.
[73] Mengi, Y., and Turhan, D., 1978, "The Influence of Retardation Time of the Heat Flux on Pulse Propagation," ASME J. Appl. Mech., **45**, pp. 433–435.
[74] Chester, M., 1966, "High Frequency Thermometry," Phys. Rev., **145**, pp. 76–80.
[75] Tzou, D. Y., "Nonequilibrium Transport: The Lagging Behavior," Adv. Transport Phenom., in press.
[76] Xu, F., Seffen, K. A., and Lu, T. J., 2008, "Non-Fourier Analysis of Skin Biothermomechanics," Int. J. Heat Mass Transfer, **51**, pp. 2237–2259.
[77] Liu, K. C., and Chen, H. T., 2009, "Analysis for the Dual-Phase-Lag Bio-Heat Transfer During Magnetic Hyperthermia Treatment," Int. J. Heat Mass Transfer, **52**, 1185–1192.
[78] Zhou, J., Chen, J. K., and Zhang, Y. W., 2009, "Dual-Phase Lag Effects on Thermal Damage to Biological Tissues Caused by Laser Irradiation," Comput. Biol. Med., **39**, pp. 286–293.
[79] Zhou, J., Zhang, Y. W., and Chen, J. K., 2009, "An Axisymmetric Dual-Phase-Lag Bioheat Model for Laser Heating of Living Tissues," Int. J. Therm. Sci., **48**, pp. 1477–1485.
[80] Goyeau, B., Benihaddadene, T., Gobin, D., and Quintard, M., 1997, "Averaged Momentum Equation for Flow Through a Nonhomogenenous Porous Structure," Transp. Porous Media, **28**, pp. 19–50.
[81] Haro, M. L., Rio, J. A., and Whitaker, S., 1996, "Flow of Maxwell Fluids in Porous Media," Transp. Porous Media, **25**, pp. 167–192.
[82] Quintard, M., and Whitaker, S., 1994, "Transport in Ordered and Disordered Porous Media III: Closure and Comparison Between Theory and Experiment," Transp. Porous Media, **15**, pp. 31–49.
[83] Wang, L. Q., 1997, "Frame-Indifferent and Positive-Definite Reynolds Stress-Strain Relation," J. Fluid Mech., **352**, pp. 341–358.
[84] Whitaker, S., 1986, "Flow in Porous Media I: A Theoretical Derivation of Darcy's Law," Transp. Porous Media, **1**, pp. 3–25.
[85] Whitaker, S., 1996, "The Forchheimer Equation: A Theoretical Development," Transp. Porous Media, **25**, pp. 27–61.
[86] Whitaker, S., 1999, *The Method of Volume Averaging*, Kluwer Academic, Dordrecht.
[87] Wang, L. Q., 2000, "Flows through Porous Media: A Theoretical Development at Macroscale," Transp. Porous Media, **39**, pp. 1–24.
[88] Quintard, M., and Whitaker, S., 1993, "One- and Two-Equation Models for Transient Diffusion Processes in Two-Phase Systems," Adv. Heat Transfer, **23**, 369–464.
[89] Fan, J., and Wang, L. Q., 2010, "Is Classical Energy Equation Adequate for Convective Heat Transfer in Nanofluids?," Advances in Mechanical Engineering, **2010**, 719406.
[90] Zhang, Y. W., 2009, "Generalized Dual-Phase Lag Bioheat Equations Based on Nonequilibrium Heat Transfer in Living Biological Tissues," Int. J. Heat Mass Transfer, **52**, pp. 4829–4834.

# Skin Electroporation With Passive Transdermal Transport Theory: A Review and a Suggestion for Future Numerical Model Development

**S. M. Becker**
University of Canterbury,
Department of Mechanical Engineering,
Private Bag 4800,
Christchurch 8140, New Zealand
e-mail: sid.becker@canterbury.ac.nz

*Skin electroporation is an approach used to enhance the transdermal transport of large molecules in which the skin is exposed to a series of electric pulses, resulting in the structural alteration of the stratum corneum. This article suggests the use of passive transdermal transport models in combination with models depicting the electrically induced structural alterations in order to advance the modeling development of transport associated with skin electroporation. A review of the major physical phenomena observed in skin electroporation transport experiments is provided. A compendium of representative models is made available through a review of the current understanding of the two fields: (1) porous media descriptions of nondestructive transdermal transport and (2) modeling electroporation related structural changes within the skin. To show the applicability and potential of merging transdermal transport modeling with skin electroporation modeling, an example model is developed that combines a brick and mortar style skin representation with a thermodynamic based model of skin electroporation.*
[DOI: 10.1115/1.4002362]

*Keywords: transdermal transport, skin, stratum corneum, computational, porous media, skin electroporation*

## 1 Introduction

The motivation behind this work is to create an awareness of the need for computational model development focusing on transport associated with skin electroporation. The intent of this article is to provide the necessary background information for researchers to combine passive transdermal transport models with existing models of electroporation. The following introduction to skin electroporation is a descriptive narrative on the results of experimental observations that provide the primary physical phenomena that the successful skin electroporation transport model must reflect. The second section of the article provides a review of models used to depict nondestructive transdermal transport. This is followed by a section that reviews the methods that have been used to describe permeability increases associated with skin electroporation. In Sec. 4, concepts from these two fields (porous media transdermal transport modeling and electroporation modeling) are combined in order to develop a model that captures mass transport associated with skin electroporation.

**1.1 Skin Electroporation.** The thin (10–50 $\mu$m) outermost layer of the epidermis is called the *stratum corneum* (*SC*). This is the most transport resistive component of the skin. This high resistance to permeability presents a major obstacle for successful transdermal drug delivery. The *SC* is composed of 15–20 layers of corneocytes (flat dead cell shells), which are interconnected by a lipid lamellar bilayer structure in a crystalline-gel phase. Diffusion transport through the *SC* is primarily associated to occur within the lamellar lipid structure of the *SC* [1,2].

Skin electroporation has been shown to greatly increase the success of transdermal delivery through this resistive layer [3,4]. During skin electroporation, the skin is exposed to a series of intense electric pulses that change the structure of the *SC* by creating microscopic aqueous pores within the lipid filled spaces [5,6]. This alteration of the *SC* structure allows for increased molecular transport and radically reduces the electrical resistance of the skin [5,7,8]. Typically skin electroporation is conducted on a skin fold in which a section of skin coated with an applicator gel is pinched between two electrodes that deliver dc electric pulses to the skin (see Fig. 1(*a*)).

Electroporation pulses can be classified into two regimes: short and long pulses. Because the physical effects of the two regimes differ greatly, the choice of which regime to use is dependent largely on the drug size (with molecules of lower molecular weight requiring pulses of shorter duration) [3].

*1.1.1 Short Pulse (Nonthermal).* Electroporation pulse times classified as short duration are typically less than 100 $\mu$s. When the drop in electric potential across the skin reaches some critical value (30–100 V) [9], skin resistance (both ionic and molecular) drops several orders of magnitude [3]. In the short pulse regime, it is believed that nanometer sized pores develop within the individual lipids. Although this phenomenon has been documented in experimental studies (both in vitro and in vivo) [3], the exact mechanism behind the sudden development of these pores is not fully understood. Nevertheless, a theoretical description of this process has been developed for single lipid bilayer electroporation in which a mechanistic model connects the transbilayer electric potential to the resulting pore size and distribution within the single bilayer [10–12]. Single lipid bilayer electroporation should be contrasted with skin electroporation. The *SC* cannot be represented by a single bilayer because the *SC* is made up of a lipid-corneocyte matrix in which the lipids are arranged in a lamellar network of about 100 lipid bilayers. In the short pulse regime,

Manuscript received January 31, 2010; final manuscript received March 16, 2010; published online September 30, 2010. Assoc. Editor: Andrey Kuznetsov.

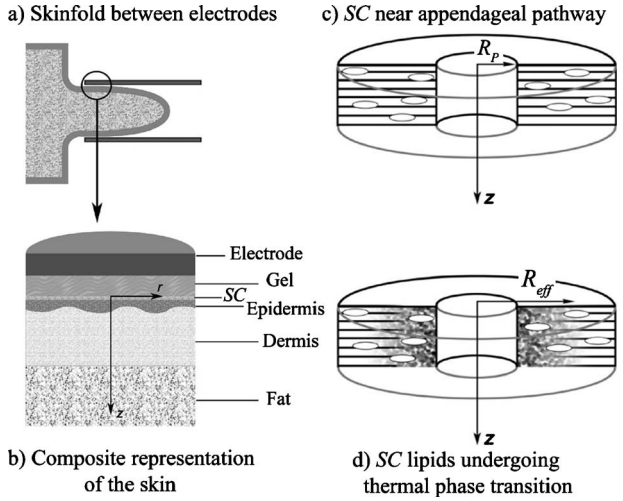



**Fig. 1 ((*a*) and (*b*)) Electroporation skin fold overview; ((c) and (d)) close-up of thermal transition near a pre-existing appendageal pore**

drug molecules passing through the *SC* must navigate through the electropores in each of the 100 bilayers. Only small drugs are capable of being transdermally transmitted through the small, tortuous pathways created during short pulse electroporation.

*1.1.2 Long Pulse and Thermal Effects*. Longer duration pulses (up to several hundred milliseconds) make possible two important secondary effects: electrophoresis and Joule heating. Even with increased *SC* permeability, larger molecules experience great viscous resistance to transport through the tortuous *SC*. The long pulse electric field provides electrophoretic forces that drive the large charged molecule through the layers of the *SC*. Long pulse skin electroporation studies have shown that electro-osmosis plays a negligible role, while electrophoresis is the primary contributor in transdermal delivery of charged molecules [3,4,13].

The second important effect of long duration skin electroporation pulses is Joule heating, which results in localized temperature rises that contribute to the increased permeability of the *SC* by lipid chain melting [3,14–16]. Experimental observations show that at temperatures around 70 °C, the barrier function of the *SC* is dramatically reduced [17,18] because the *SC* lipid lamellar structure experiences a fluidizing phase transition [17,19,20]. It is generally accepted that there exist three main endothermic transitions in the *SC* within the temperature range 60–130 °C, which have been independently confirmed by differential scanning calorimetry [19–21]. Figure 2 shows a representative heat versus temperature curve in which the *SC* endothermic transitions are evident as

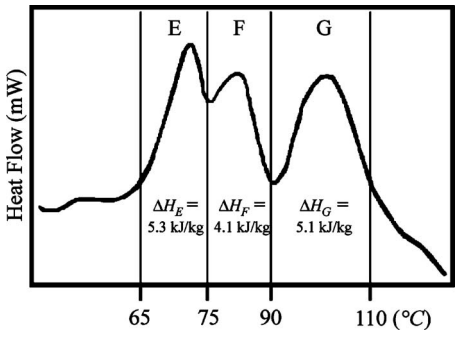



**Fig. 2 A representative heat versus temperature curve of *SC* lipids in thermal phase transition. Latent heat and phase transition temperature range values taken from data provided in Ref. [20].**

the peaks. The two high temperature phase transitions, F and G, are of less direct importance to skin permeability and are associated with secondary melting of lipids covalently bonded to corneocytes and protein denaturation, respectively [20–22]. Phase change E, however, has been attributed to increased permeability by the disordering of the *SC* lipid lamellar barrier structure [20] and has been documented in numerous differential scanning calorimetry studies at endothermic peak temperatures of 65–72 °C [23], X-ray diffraction microscopy [24], polarized light thermal microscopy [22], and experimental studies linking *SC* permeability increases to temperature [17,18].

Early researchers of skin electroporation used the phrase local transport region (LTR) to describe a phenomenon in which the applied pulses result in "large" micrometer sized regions of increased *SC* permeability [6,14]. The results of in vitro studies show that within the LTR the electrical and mass permeabilities may be many orders of magnitude greater than outside the LTR [16,25] and that the development of LTRs is always associated with thermal effects [26]. Experimentally, it has been shown that in some cases high temperature contours associated with Joule heating may originate near skin appendages (sweat gland) [5]. Direct evidence of the localized moving heat front and LTRs is found in in vitro studies in which human *SC* is removed and electroporated under observation [5,7–9,15,16,25]. Depending on pulse intensity, the scale in size of these high temperature fronts is on the order of 100 $\mu$m occurring on a time scale of 10–100 ms [16]. In Figs. 1(*c*) and 1(*d*), a representation of the *SC* prior to electroporation with a pre-existing pore of radius, $R_P$, shows that the lipid structure connecting the corneocytes of the *SC* is lamellar and uniform. Figure 1(*d*) shows the lamellar structure in phase transition *E* and that the effective pore radius, $R_{\text{eff}}$, extends to the region at which *SC* lipids are unaffected by the temperature rises.

The distinction should be made between the nonthermal primary nanometer sized pores that occur within a single bilayer and the Joule heating associated micrometer to millimeter sized LTRs that pass through the entire *SC*. Also, the *SC* thermal LTR formations should not be confused with actual macroscopic voids within the skin.

**1.2 Electroporation Transport Modeling Considerations.** A successful electroporation model must capture two transient processes: (1) transport through the *SC* microstructure and (2) the decrease in the *SC*'s resistance to transport that occurs as a result of the applied electric pulse. While models of these two processes do exist separately, current electroporation models either entirely neglect transport through the skin or do not detail the effect of the *SC*'s microstructure. The following two sections review models of nondestructive transport (transport without structural alterations to the *SC*) and then current models depicting the structural changes in the skin's architecture resulting from electroporation.

## 2 Nondestructive Transport Modeling: The *SC* as a Porous Medium

Nondestructive transdermal drug delivery methods that rely on pure diffusion or very low voltage, electrically enhanced diffusion (iontophoresis) are associated with solutes of low molecular weight. The transient equations governing the transport of solute through the skin are for pure diffusion,

$$\frac{\partial C^*}{\partial t} = \nabla \cdot (D_{\text{eff}} \nabla C^*) \tag{1}$$

and for nondestructive electrically enhanced diffusion,

$$\frac{\partial C^*}{\partial t} = \nabla \cdot (D_{\text{eff}} \nabla C^*) + \nabla \cdot (m_{\text{eff}} C^* \nabla \phi) - U_{\text{eff}} \cdot \nabla C^* \tag{2}$$

where $t$ is the time, $C^*$ is the local solute concentration, $D_{\text{eff}}$ is the effective diffusion coefficient, $\phi$ is the electric potential resulting from the applied electric field, $m_{\text{eff}}$ is the effective electrophoretic mobility coefficient, which describes the ability of the electric

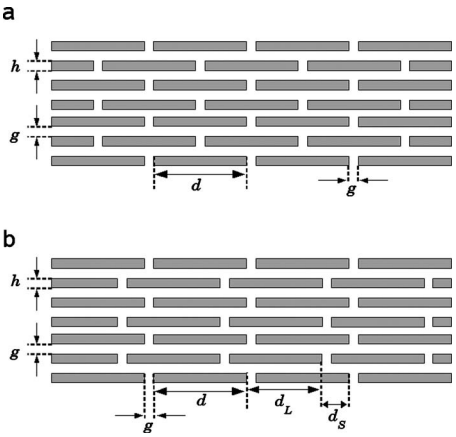



**Fig. 3 *SC* brick and mortar representations: (*a*) symmetric and (*b*) asymmetric**

field to move the solute, and $U_{\text{eff}}$ is the effective electro-osmotic flow.

Theoretical models are validated by a comparison with experimental diffusion studies in which a skin sample is placed between donor and receiver chambers of a Franz diffusion cell [27,28]. In these studies, the term "permeability" is often used to represent the steady state dimensionless flux across the barrier membrane of thickness $L$. If the solute concentration in the donor cell is $C_O$ and the receiver cell concentration, $C_R$, is negligible, then the permeability may be represented by the relation

$$P = D_{\text{eff}} \frac{(C_O - C_R)}{LC_O} = D_{\text{eff}} \frac{1}{L} \tag{3}$$

The challenge to researchers has been how to define the effective transport coefficients (diffusion, electrophoretic mobility, and electro-osmotic flow) based on the architecture and the physics of the *SC*. Porous media representations of the *SC* in which the barrier function is represented using tortuosity, $\tau$, and porosity, $\varepsilon$, have been used to help depict these transport coefficients. The two primary porous media approaches to describe the permeability of hydrophilic solutes within the *SC* are based either on the overall structure of the *SC* lipid-corneocyte architecture or on the nanoscale structure and molecular behavior of the *SC* lipids themselves.

**2.1 Brick and Mortar Models.** Brick and mortar depictions of the SC are based on the lipid-corneocyte matrix in which the mortar represents the permeable lipid filled spaces within the *SC*, while the corneocytes, which are either completely impermeable or highly impermeable, are represented by the bricks. The representation of the *SC* architecture consisting of a symmetric brick and mortar model was introduced in transdermal diffusion studies [29] (see Fig. 3(*a*)). A nonsymmetric brick and mortar concept of the *SC* (see Fig. 3(*b*)) has been introduced, which more accurately describes the lateral diffusion path by using a parameter based on both long and short lateral diffusion path lengths [30]. Extensions to this geometric representation include a trapezoidal representation of the corneocyte [31] and an irregular description of the *SC* architecture, which is based on actual micrographs of mouse *SC* [32,33].

A comparison of the models based on the geometry of the brick and mortar model representation (Figs. 3(*a*) and 3(*b*)) is provided in the study [34] that derives a two-tortuosity model accounting for both the total lipid filled space as well as the lateral diffusion pathway and compares the model results with the more traditional model descriptions of tortuosity, $\tau$, and porosity, $\varepsilon$ (see Table 1). In the descriptions of Table 1, $L$ is the overall *SC* thickness, N refers to the number of layers within the *SC*, and the other geometric parameters of the *SC* representation are shown in Figs. 3(*a*) and 3(*b*). The corneocyte offset, $\omega$, is defined as the ratio of long to short lateral paths $\omega = d_L/d_S$. In that study, comparisons are made based on the permeability, which is defined as

$$P = \frac{\varepsilon}{\tau} \frac{K_b D_b}{L} \tag{4}$$

The study compares the predictions of $D_b$, the solute-*SC* diffusion coefficient, and $K_b$, the solute-lipid bilayer partition coefficient (a measure of the solute's solubility within the lipid bilayers). The study finds that the two-tortuosity model most closely predicts $K_b$ and $D_b$ of hydrophobic permeants.

**2.2 Lipid Structure Based Models: Small Hydrophilic Solutes.** While models depicting the transport of hydrophobic solutes have been extensively studied [30,35,36] in what are called free volume diffusion models, in this article the discussion is restricted to hydrophilic solutes. This is primarily because the nature of electroporation creates water filled disruptions within the SC lipid structure. For hydrophilic solutes, an aqueous pore model has been proposed in which the solute travels through nanometer sized water filled voids within the *SC* that exist as a result of defects within the lipid bilayers, which "manifest themselves as separation of grain boundaries, lattice vacancies, multi-molecular voids due to missing lipids, or defects caused by steric constraints placed by keratinocytes on intercellular lipid bilayers" [37]. The

**Table 1 Description of *SC* tortuosity and porosity based on brick and mortar structures**

| Type | Porosity $\varepsilon$ | Tortuosity $\tau$ |
|---|---|---|
| Symmetric [29] | $\frac{2g}{d}$ | $1 + \left(\frac{d}{h} + \frac{g}{h}\right) / 2\left(1 + \frac{g}{h}\right)$ |
| Symmetric [50] | $\frac{g}{g+d}$ | $\left(Nh + (N-1)\left(\frac{d+g}{4}\right)\right)/L$ |
| Symmetric [51] | $\left(\left(\frac{g+d}{2}\right)^2 - \left(\frac{d}{2}\right)^2\right) / \left(\frac{d}{2}\right)^2$ | $\left(L + (N-1)\left(\frac{d+g}{2}\right)\right)/L$ |
| Asymmetric [30] | $\frac{g}{(g+d)}$ | $\left(Nh + (N-1)\left(\frac{\omega}{(1+\omega)^2}\right)d\right)/L$ |
| Asymmetric [34] (Two-Tortuosity) | $\frac{g}{(g+d)}$ | $\tau_{\text{flux}} = \frac{Nh + (N-1)g + (N-1)\frac{\omega}{(1+\omega)^2}d}{Nh + (N-1)g}$<br>$\tau_{\text{volume}} = \frac{Nh + (N-1)g + (N-1)d}{Nh + (N-1)g}$ |

aqueous pore membrane model represents these defects with a system of cylindrical tortuous pores that traverse the *SC* and are described in terms of radial size and tortuosity.

Although indirect physical experimental evidence supports the presence of these pathways, it should be recognized when implementing this method that there is not explicit evidence that this route actually exists [38,39]. The concept has been used to describe transdermal transport for both passive diffusion [37–39] as well as nondestructive electrically enhanced diffusion [28].

The study [27] combines the porous media approach with a hindrance factor, which is based on a spherical solute transported through a fluid filled cylindrical pore. The flux, $J$, can be represented by modifying the Nernst–Plank equation to include porosity, tortuosity, and hindrance to the transport:

$$J = \frac{\varepsilon}{\tau}\left\{ - H_F' D\left( \frac{dC^*}{dx} - \frac{C^* zF}{RT}\frac{d\phi}{dx} \right) + W_F' U C^* \right\} \tag{5}$$

The electrophoretic component of Eq. (5) makes use of the Einstein–Smoluchowski relation, which relates the electrophoretic mobility of Eq. (2) to the diffusion coefficient,

$$m = D\frac{zF}{RT} \tag{6}$$

where $z$ is the charge of the ionic permeant (solute), $F$ is Faraday's number, $R$ is the gas constant, $T$ is the absolute temperature, and $\phi$ is the electric potential. The term $H_F'$ is the diffusion hindrance factor, which is associated with the Brownian motion and electrophoretic migration. The term $W_F'$ is a convective flow hindrance factor associated with the bulk convective flow (contributions from electro-osmosis and pressure) at an average flow velocity of $U$. It may be assumed that the electrokinetic transport hindrance factors $H_F'$ and $W_F'$ are approximately equal to their counterparts for passive diffusion ($W_F$) and pressure induced convective flow ($W_F$) [40].

The hindrance factors are characterized by the ratio of the ionic solute radius, $r_S$, to the defect pore radius, $r_P$:

$$\lambda = \frac{r_S}{r_P} \tag{7}$$

It may be assumed that the lipid bilayer membranes of the *SC* have small pores of radii in the range $10 \leq r_P \leq 20$ Å [27]. For medium and small molecule transports ($\lambda < 0.4$), hindrance factors are represented by [37,41]

$$H_F(\lambda) = (1-\lambda)^2(1 - 2.104\lambda + 2.09\lambda^3 - 0.948\lambda^5) \tag{8}$$

$$W_F(\lambda) = (1-\lambda)^2(2-(1-\lambda)^2)(1 - 0.667\lambda^2 - 0.163\lambda^3) \tag{9}$$

## 3 Skin Electroporation Models

The previous models focusing on nondestructive transport are limited to very small molecular transport. Electroporation, which facilitates the transport of larger molecular transport by altering the *SC*, is represented by models that may be characterized as either empirical or mechanistic.

**3.1 Empirical Skin Electroporation Models.** Empirical models typically use electrical conductivity as a physical parameter to quantify the electroporation of the skin. This is because (1) the electroporation related changes in the *SC* microstructure are accompanied by the order of magnitude increases in *SC* electrical conductivity, $\sigma_{SC}$ [9] and (2) skin electrical characteristics (conductivity, current, and voltage drop) can be directly measured experimentally. One major benefit of the empirical approach is that it can describe electroporation in the short pulse regime without having to keep track of the permeability of each of the 100 *SC* bilayers.

For a low frequency or dc pulse, the electric potential distribution within the *SC* is usually solved from the Laplace equation,

$$\nabla \cdot (\sigma_{SC} \nabla \phi) = 0 \tag{10}$$

where $\phi$ is the electric potential. Because the capacitive charging time associated with non-Ohmic behavior of the *SC* is very short (negligible at pulsing times greater than 1 ms [42]), this behavior is typically neglected in studies that are less concerned with individual nanoscale electropore creations in a single lipid bilayer and focus instead on describing the behavior of the *SC* as a whole.

Recent novel empirical approaches directly relate the degree of permeability (represented by increases in the electrical conductivity) to the local electric field [43,44]. These studies use data from skin electroporation experiments, which monitor the relationship between electrical conductivity changes and the electric field magnitude. The electric field is determined from the gradient in the potential:

$$E = \nabla \phi \tag{11}$$

In the parametric study [44], a finite element analysis is conducted of a skin fold undergoing electroporation by using experimentally determined relations between electrical conductivity and electric field magnitude for three tissue types: subcutaneous, *SC*, and the combined dermis and epidermis (excluding *SC*). The drop in conductivity is modeled as a step function of the magnitude of the electric field consisting of four different steps. The *SC* conductivity is represented as

$$\sigma_{SC}\ (\mathrm{S/m}) = \begin{cases} E\ (\mathrm{V/m}) < 600: & \sigma_{SC} = 0.0005 \\ 600 < E < 800: & \sigma_{SC} = 0.0165 \\ 800 < E < 1000: & \sigma_{SC} = 0.06 \\ 1000 < E < 1200: & \sigma_{SC} = 0.178 \\ 1200 < E: & \sigma_{SC} = 0.5 \end{cases} \tag{12}$$

The study allows for a macroscopic representation of the current-applied voltage relation that closely resembles that of experimental data and compares well with experimental findings relating the electrical behavior of the skin fold at various applied voltages.

The concept of using the electric field magnitude to define the degree of permeability of the skin and underlying tissue is used in a detailed model [43] that accounts for tissue damage in which a single value of post-electroporated tissue electrical conductivity increase is used. The basis of this study is that, experimentally, it has been shown that electroporation begins at a specific transmembrane voltage $E_O$, but that when a certain voltage is exceeded, $E_1$, irreversible cellular damage begins. The study relies on experimentally derived values of $E_O$ and $E_1$ to represent the threshold values of various tissues undergoing electroporation. The electric field dependent electrical conductivity, $\sigma(E)$, is related to the degree of electroporation by some relationship between electrical conductivity before permeabilization, $\sigma_O$, and a maximum value of electrical conductivity due to electroporation, $\sigma_1$. Another novel concept of this study is the variety of functional dependencies that are investigated to depict the relation between conductivity and potential drop. These include a simple step function, a linear dependence, exponential dependence, and a sigmoid based dependence. The study concludes that the following exponential function best approximates the experimental data of skin electrical conductivity changes associated with electroporation,

$$\sigma(E) = (\sigma_O - \sigma_1)\frac{\exp\left[\dfrac{E-E_1}{B}\right] - 1}{\exp\left[\dfrac{E-E_O}{B}\right] - 1} + \sigma_1 \tag{13}$$

where $B$ is a shape parameter, and the parameter values corresponding to skin are $B = 30$, $E_O = 400$ V/cm, $E_1 = 900$ V/cm, $\sigma_O = 0.002$ S/m, and $\sigma_1 = 0.16$ S/m. The exponential function is used to model the electrical conductivity changes of the skin as a whole (*SC*, epidermis, and dermis) and not separately. The study finds that the empirical model allows for computational results

that agree very well with experiments also conducted in that study.

**3.2 Thermodynamic Skin Electroporation Model.** Recent studies have proposed a thermodynamically based model to describe the structural changes associated with skin electroporation [45,46] by taking into account that for longer duration electroporation pulses, the increase in permeability associated with the formation of the LTR has been linked to sudden local rises in temperature. Experimental findings show that the electroporated *SC* experiences localized regions of higher permeability, that the skin experiences local heat fronts that originate in skin appendageal ducts and grow radially outward, forming regions of high permeability, and that the temperatures associated with these heated regions have been observed to reach temperatures above lipid thermal phase transition temperatures.

To describe the degree of lipid disorder, these studies borrow from methods that have been traditionally designed to model melting and solidification processes occurring over a temperature range [47], the analogy being that the thermally influenced transition of the *SC* lipid microstructure from structured to disordered may be represented by the same function that models the solid structure thermal transition from solid to liquid.

The function referred to as the lipid melt fraction is used to describe the degree of lipid disorder, and it is defined as

$$\varphi = \frac{(H - c_{SC}T)}{\Delta H_E} \tag{14}$$

where $c_{SC}$ is the *SC* specific heat capacity and $\Delta H_E$ is the latent heat associated with phase transition *E*. The total enthalpy, *H*, is defined as

$$H = \int_{T_{E1}}^{T} c_{SC,\text{app}} d\tau \quad [T_{E1} < T \leq T_{E2}] \tag{15}$$

where $T_{E1}$ and $T_{E2}$ are the temperatures over which transition *E* takes place. To simplify the description considerably, a rectangular shaped specific heat versus temperature curve is used to model phase transitions so that the apparent *SC* specific heat, $c_{SC,\text{app}}$, is represented by

$$c_{SC,\text{app}} = c_{SC} + c_{SC,L} \tag{16}$$

where $c_{SC,L}$ is the latent specific heat,

$$c_{SC,L} = \frac{\Delta H_{PC}}{T_{PC2} - T_{PC1}} \tag{17}$$

where $\Delta H$ is the latent heat, $T_2$ and $T_1$ are the representative phase change end and beginning temperatures, and the subscript *PC* refers to one of the three phase transitions: E, F, or G. Figure 2 lists the values used in computations for the three *SC* phase change transitions.

The lipid melt fraction of Eq. (14) is a thermodynamically based function that represents the degree of lipid disorder in the range $0 \leq \varphi \leq 1$, where $\varphi = 0$ corresponds to the insulative unaltered lamellar lipid structure and $\varphi = 1$ corresponds to permeable, fully disrupted *SC* lipid architecture.

The electroporation altered lipid structure is related to increases in ionic and mass transport coefficients through the lipid melt fraction by the relations

$$\sigma^*_{SC} = \sigma_{SC} + \varphi(\sigma_{SC,\text{melt}} - \sigma_{SC}) \tag{18}$$

$$m^*_{SC} = m_{SC} + \varphi(m_{SC,\text{melt}} - m_{SC}) \tag{19}$$

$$D^*_{SC} = D_{SC} + \varphi(D_{SC,\text{melt}} - D_{SC}) \tag{20}$$

where $\sigma_{SC}$ is the normal electrical conductivity, $\sigma_{SC,\text{melt}}$ is the electrical conductivity associated with the *SC* after full lipid melting, $m_{SC}$ is the unperturbed mobility of the solute in the *SC*, $m_{SC,\text{melt}}$ is the mobility associated with the *SC* after full lipid melting, $D_{SC}$ corresponds to the initial diffusion coefficient of the solute in the *SC*, and $D_{SC,\text{melt}}$ is the diffusion coefficient associated with the *SC* after full lipid melting [45,46].

To provide closure for the thermodynamic model's enthalpic relations, Eqs. (14) and (15), the Joule heating and associated temperature distribution within the electroporated *SC* must be accurately defined. This is done in the parametric study [45] in which a physical model of a skin fold is developed, which includes the electrode, applicator gel, and skin composite layers. Figure 1(*b*) shows the composite representation of this configuration consisting of electrodes, gel, and skin. The skin model consists of four composite layers: *SC*, epidermis, dermis, and subcutaneous fat.

Pennes' bioheat equation [48] is used along with an additional source term to describe thermal energy,

$$\rho_i c_i \frac{\partial T}{\partial t} = \nabla \cdot (k_i \nabla T) - \omega_{m,i} c_b (T - T_a) + q'''_i + Q_J \tag{21}$$

where $\rho$ is the density, $c$ is the specific heat, $k$ is the thermal conductivity, $T$ is the temperature, $T_a$ is the arterial blood temperature, $q'''$ is the metabolic volumetric heat generation, $\omega_m$ is the nondirectional blood flow associated with perfusion, and $t$ is the time. The parameter $c_b$ in the perfusion term is the specific heat of blood, which is assigned a value of $c_b = 3800$ J/kg K. The subscript $i$ refers to the gel, electrode, or one of the four composite skin layers.

The second and third terms on the right hand side of Eq. (21) denote perfusion and metabolic heat generation and are not present in the nonliving epidermis, electrode, and gel layers. The term, $Q_J$, in Eq. (21) is the Joule heat generated from the induced electric field. Joule heating occurs during the applied electric pulse in the gel and tissue composite layers, and it is defined as

$$Q_J = \sigma_i |\nabla \phi|^2 \tag{22}$$

where $\sigma_i$ is the composite electrical conductivity and $\phi$ is the electric potential of the applied electric field, which is solved from the Laplace equation:

$$\nabla \cdot (\sigma_i \nabla \phi) = 0 \tag{23}$$

The thermodynamic approach requires considerably more theoretical analysis than the empirical model because the solutions of the electric field, the thermal energy, and the variable transport coefficients are coupled implicitly. However, such an approach allows researchers to predict and optimize solute transport and lipid behavior based on experimental parameter variables. Both studies [45,46] provide parametric investigations on the effects of pulse spacing, pulse duration, and *SC* lipid thermal dependency on LTR size, LTR growth rate, and solute deposition.

## 4 Electroporation Transport Modeling: Combining Porous Media With Electroporation

In the previous section's review of the field, models have been presented, which depict different approaches to modeling passive transdermal transport. Concepts in modeling the structural changes associated with skin electroporation have also been reviewed, listing the representative models of multiple studies. The final section of this article proposes that in the future, researchers attempting to model transdermal transport associated with skin electroporation consider using well understood porous media descriptions of the *SC* in combination with one of the models describing skin electroporation (either empirical or thermodynamic). To show an example of one of the possible combinations of passive transport models with electroporation models, the following section combines a brick and mortar description of the *SC* together with the mechanistic thermodynamic model of *SC* structural alterations associated with electroporation.

**4.1 Description of Transport Coefficients.** In order to predict solute behavior during a long skin electroporation pulse, the

**Table 2 Electroporation skin fold thermo-electric and physiological property values [45,46]**

| | | | *SC* | Epidermis | Dermis | Fat | Gel | Electrode |
|---|---|---|---|---|---|---|---|---|
| Thickness | mm | $L$ | 0.0165 | 0.035 | 1.1 | 1.2 | 1 | 1 |
| Thermal conductivity | W/m K | $k$ | 0.2 | 0.209 | 0.293 | 0.23 | 0.6 | 16 |
| Density | kg/m$^3$ | $\rho$ | 1500 | 1110 | 1116 | 850 | 1000 | 7850 |
| Heat capacity (specific) | J/kg K | $c$ | 3600 | 3600 | 3800 | 2300 | 4180 | 450 |
| Perfusion | kg/m$^3$ s | $\omega_m$ | 0 | 0 | 2.33 | 0.545 | – | – |
| Metabolic heat generation | W/m$^3$ | $q'''$ | 0 | 0 | 200 | 5 | – | – |
| Electrical conductivity | S/m | $\sigma$ | $10^{-5}$–$10^{-3}$ | 0.01 | 0.015 | 0.02 | 1.5 | – |
| Electrophoretic mobility | m$^2$/V s | $m$ | $10^{-13}$–$10^{-7}$ | $10^{-10}$ | $2\times10^{-10}$ | $10^{-10}$ | $10^{-8}$ | – |
| Diffusivity | m$^2$/s | $D$ | $10^{-13}$–$10^{-9}$ | $10^{-12}$ | $10^{-12}$ | $10^{-12}$ | $10^{-9}$ | – |

model must describe the transport coefficient related to diffusion and electrophoresis in such a way that considers (1) the transport resistive lipid-corneocyte structure of the *SC* and (2) the drastic changes in the lamellar lipid microstructure that result from lipid melting.

During an electroporation pulse, electrophoretic forces aid the transport of charged solutes. Equation (2), which describes the electrokinetic transport of a charged solute in a homogeneous medium is altered to represent the composite layer nature of the skin fold undergoing skin electroporation (Figs. 1(*a*) and 1(*b*)),

$$\frac{\partial C}{\partial t} = \nabla \cdot (m_i C \nabla \phi) + \nabla \cdot (D_i \nabla C) \tag{24}$$

where the subscript $i$ refers to the gel or one of the four composite skin layers. The term, $C$, is the dimensionless solute concentration in which the concentrations have been normalized by the relation $C = C^*/C_O$, where $C^*$ is the local solute concentration and $C_O$ is the initial concentration within the applicator gel layer.

Equation (24) must be altered to account for transport within the *SC* where the porosity and tortuosity act to inhibit transport. To account for this, the equation governing transport through the *SC* is defined as follows:

$$\frac{\partial C}{\partial t} = \frac{\varepsilon}{\tau}[\nabla \cdot (m_L^* C \nabla \phi) + \nabla \cdot (D_L^* \nabla C)] \tag{25}$$

At this point, a choice one of the *SC* representations of Sec. 2 must be made. In this illustrative example, we choose the brick and mortar description (Figs. 3(*a*) and 3(*b*)). The porosity and tortuosity of Eq. (25) are then defined using any choice of the descriptions provided in Table 2.

The next step is to include a model component that accounts for the structural changes within the *SC* lipids that occur during the electroporation pulse. In this example, the choice is made to use the thermodynamic skin electroporation model, which connects the structural change related increases in mobility and diffusion coefficients to the local thermal energy. It is important to note that in Eq. (25) the mobility $m_L^*$ and diffusivity $D_L^*$ are representative only of the solute within the lipid filled spaces between the corneocytes. These should be contrasted to the thermodynamic model variables of Eqs. (19) and (20) in which the values correspond to those represented by the entire *SC* structure (lipids and corneocytes). The transport coefficients of Eq. (25) may now be defined through the thermodynamic electroporation by the relations

$$m_L^* = m_L + \varphi(m_{L,\text{melt}} - m_L) \tag{26}$$

$$D_L^* = D_L + \varphi(D_{L,\text{melt}} - D_L) \tag{27}$$

where $m_L$ is the mobility in unperturbed lipids, $m_{L,\text{melt}}$ is the mobility within the *SC* lipid structure after full lipid melting, $D_L$ corresponds to the initial diffusion coefficient of the solute in the lipids of the *SC*, and $D_{L,\text{melt}}$ is the diffusion coefficient associated within the fully altered *SC* lipid structure.

In the equation governing transport (Eq. (25)), the brick and mortar representation of the *SC* (in which the tortuosity and porosity are linked to assumed *SC* geometry) has now been combined with the mechanistic thermodynamic skin electroporation model (in which the thermal energy is linked to transport coefficients through Eqs. (14), (26), and (27)).

**4.2 Computational Considerations.** To demonstrate this combined brick and mortar–thermodynamic approach, a computational analysis of a skin fold undergoing electroporation is conducted, reflecting the physical description shown in Fig. 1(*b*). Computations preformed in this proposal rely on symmetry at the axial midplane; therefore, only the half domain is modeled. Figure 1(*b*) shows the composite representation of this configuration consisting of electrode, gel, and skin. The skin model consists of four sections: *SC*, epidermis, dermis, and fat. Computations are made in a cylindrical domain with a 1 mm outer radius. The *SC* is assumed to be transverse by a pre-existing appendageal pore, and this is represented in the radial center by a gel filled 5 $\mu$m cylindrical region. The thermodynamic properties used in Eqs. (14) and (17) are evaluated using data provided in the study [20] shown in Fig. 2. Equations (21)–(23) and (25)–(27) are solved using geometric and thermo-electrical property and values listed in Table 2 (taken from Refs. [45,46]). The electrical potential of Eq. (23) is evaluated assuming no electrical conduction at the outer radial sides (approximating the case of a regular distribution of LTRs) and is prescribed a value of half the applied voltage at the lower computational boundary (corresponding to the midpoint between the electrodes). It is assumed that no solute is transported into the electrode from the gel, and by symmetry no flux in solute occurs at the radial boundaries of the domain. Initially only the dimensionless solute concentration is unity in the gel layer and has a zero value everywhere else. The initial temperature distribution is uniform and equal to the arterial blood temperature 37 °C. The *SC* structural change related electrical behavior of Eq. (18) is found using the following conductivity values: $\sigma_{SC} = 10^{-5}$ S/m and $\sigma_{SC,\text{melt}} = 10^{-3}$ S/m, which are chosen to represent a two order magnitude increase in electrical conductivity with lipid restructuring, as suggested by the results of Ref. [49].

To complete the computation of Eq. (21), it is assumed that the radial boundaries are thermally insulated and a convective boundary condition is imposed on the electrode surface in which the ambient temperature is 20 °C and the convection coefficient has a value of 15 W/m$^2$ K. The transport associated property values that are used to represent solute mobility and diffusion coefficients within the lipids are $m_{SC-L} = 10^{-13}$ m$^2$/V s, $D_{SC-L} = 10^{-13}$ m$^2$/s, $m_{SC-L,\text{melt}} = 10^{-7}$ m$^2$/V s, and $D_{SC-L,\text{melt}} = 10^{-9}$ m$^2$/s. The values used to represent the mass transport coefficients associated with the altered and unaltered *SC* lipids have been chosen to approximate the very resistive conditions a large DNA molecule would experience within the aggregate lipids of the *SC* [40]. To evaluate the porosity and tortuosity, the following input values suggested for normal *SC* [34] were used: $N = 15$, $h = 1$ $\mu$m, $g = 0.1$ $\mu$m, $\omega = 3$, and $d = 40$ $\mu$m for the two brick and mortar style models: symmetric as defined by Ref. [29] and asymmetric as defined by Ref. [30].

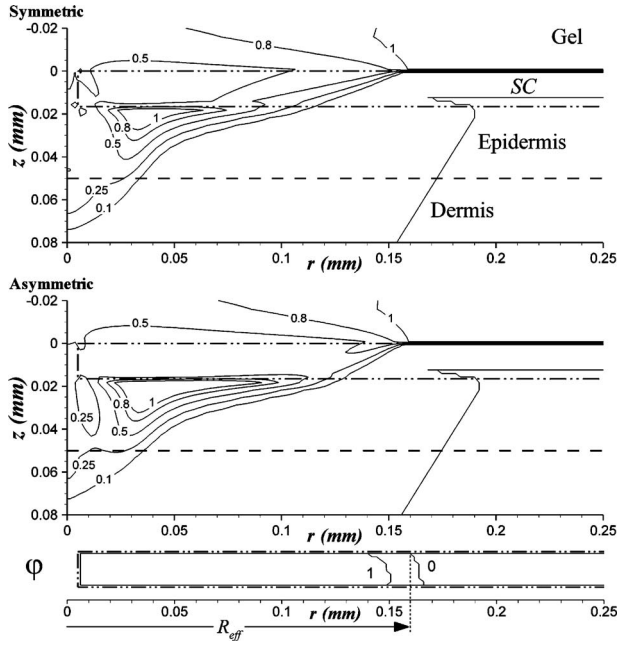



**Fig. 4 Dimensionless solute concentration profile comparisons of symmetric [29] and asymmetric [30] brick and mortar models for a 300 V electroporation pulse at 400 ms. Dashed line indicates epidermis-dermis border; dash-dot-dot represents *SC* boundary.**

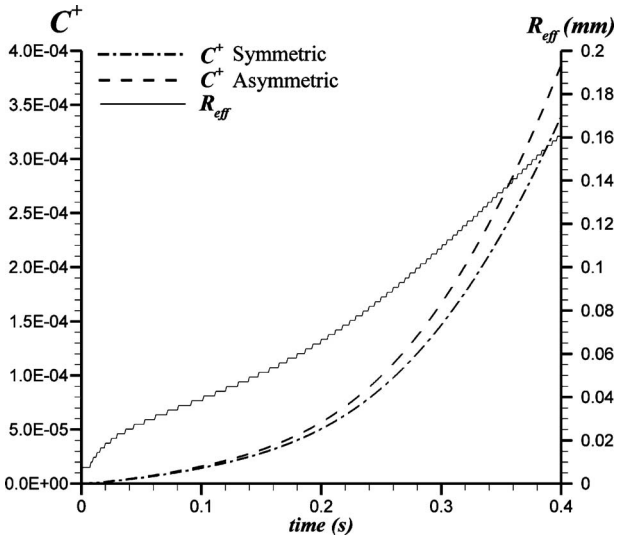



**Fig. 5 A comparison between the dimensionless transient solute concentration transported below the SC surface by a 300 V electroporation pulse for symmetric [29] and asymmetric [30] brick and mortar models**

**4.3 Results.** Figure 4 shows the concentration profiles of the two models at the end of a 400 ms, 300 V pulse. Both cases show depletion of solute in the gel directly above the altered *SC*, indicating large flux rates into the *SC* from the gel. At the end of the pulse, a local accumulation of solute is seen in the higher concentration located at the lower *SC*-epidermis interface. This may be explained by recalling that the contribution by the electrophoretic term of Eqs. (24) and (25) is a function of the gradient in potential. Within the *SC* the apparent electrical conductivity values are much lower than those of the epidermis below (even the conductivity associated with full lipid melting). This results in a larger potential gradient within the *SC* compared with the epidermis below. Effectively, the electrophoretic driving forces that carry the solute through the *SC* experience a sudden drop in magnitude after the *SC*-epidermis junction.

The asymmetric brick and mortar model of Ref. [30] takes into account the asymmetry of the impermeable corneocytes within the *SC* and therefore predicts a lower value in tortuosity than the symmetric *SC* representation. With a less tortuous path, the solute experiences less resistance to transport through the *SC*, which is evident in Fig. 4 by the asymmetric model's slightly larger region of influence in the higher level concentration profiles below the *SC*. To directly quantify this difference, a nondimensional concentration of solute is introduced. The dimensionless solute concentration is defined as the ratio of solute transported beneath the surface of the *SC* to the total solute contained in the applicator gel prior to electroporation:

$$C^+ = \frac{\int_{r=0}^{R_{\max}} \int_{z=0}^{Z_{\max}} C(r,z,t)dzdr}{\int_{r=0}^{R_{\max}} \int_{z=\mathrm{th}_{\mathrm{gel}}}^{0} C(r,z,0)dzdr} \tag{28}$$

By using this concentration description, the value of the relative amount of solute transported below the surface of the *SC* can be shown, and actual value comparisons between the two brick and mortar models at any time during the electroporation pulse can be made. Figure 5 shows the transient dimensionless solute concentration transported into the skin, as described by Eq. (28) for the two models. As implied in Fig. 4, the magnitude of solute transported is greater for the asymmetric case. It is also interesting to note that the values are similar for the first 100 ms of the pulse. This is because at very early stages of the pulse, the LTR has not grown substantially, and the solute is primarily transported through the pre-existing pore, and thus the porous media description of the *SC* does not play a major role at the beginning of the pulse. However, once larger areas of the *SC* lipids become fluidized (as the LTR grows), the primary pathway of the solute is through the *SC*, and the type of porous media description has a greater effect on the amount of solute transported. The dotted line in Fig. 5 represents the effective LTR radius, $R_{\mathrm{eff}}$, which is defined as the minimum radial distance in which the *SC* lipids experience no thermal phase change (i.e., the smallest radial location that the value $\varphi=0$ may be found). At 100 ms, the *SC* area that has experienced the increase in permeability is less than 10% of that at the end of the pulse. Figure 5 shows clearly that for the asymmetric model the values of solute transported are larger and that this continues throughout the remainder of the pulse. This indicates that not only are the values of solute transported greater for an asymmetric model of the *SC*, but so are the rates at which the solute is transported into the skin.

## 5 Conclusions

A review has been presented, offering insight into passive transdermal transport, which treats the SC either as a brick and mortar structure or as a membrane permeated by nanoscale aqueous pores. A review of current models, which capture the structural changes within the *SC* during skin electroporation, has been made, characterizing the models as either empirically or thermodynamically based.

The goal of this article has been to consider the use of known models of nondestructive transdermal transport in conjunction with theoretical descriptions of permeability changes associated with skin electroporation. As a step toward this combined approach, an exploratory model has been presented, which combines the brick and mortar representation of the *SC* with the thermodynamic based model of skin electroporation. The results show that the combined model approach to modeling skin electroporation transport allows parametric influences (such as symmetry in *SC*

microstructure representation) to be easily isolated and examined.

Future attempts at modeling transport during the alteration of the lipid structure of the *SC* are not restricted to the thermal description presented in this study. For instance, during shorter pulse skin electroporation, increases in the *SC* permeability may occur without microscale thermal influences. In this case, it is plausible to approach the electroporation with an empirical model, and the transport properties can be approached using either free volume diffusion or aqueous pathway theory. In such a model, it is possible to envision that the porosity, tortuosity, and existing pathway distributions could all be represented as functions of the local gradient in electric potential.

## Acknowledgment

Special thanks are due to the Alexander von Humboldt Foundation for its generous fellowship award that facilitated the completion of this work. The author also expresses gratitude to Prof. Heinz Herwig for providing facilities at the Institute for Thermo-Fluid Dynamics, Hamburg University of Technology and whose insights and encouragement greatly benefited the quality of this work.

## References

[1] Bouwstra, J. A., Honeywell-Nguyen, P. L., Gooris, G. S., and Ponec, M., 2003, "Structure of the Skin Barrier and Its Modulation by Vesicular Formulations," Prog. Lipid Res., **42**(1), pp. 1–36.
[2] Madison, K. C., 2003, "Barrier Function of the Skin: "La Raison d'Être" of the Epidermis," J. Invest. Dermatol., **121**(2), pp. 231–241.
[3] Denet, A. R., Vanbever, R., and Préat, V., 2004, "Skin Electroporation for Transdermal and Topical Delivery," Adv. Drug Delivery Rev., **56**(5), pp. 659–674.
[4] Regnier, V., De Morre, N., Jadoul, A., and Préat, V., 1999, "Mechanisms of a Phosphorothioate Oligonucleotide Delivery by Skin Electroporation," Int. J. Pharm., **184**(2), pp. 147–156.
[5] Pliquett, U. F., and Gusbeth, C. A., 2000, "Perturbation of Human Skin Due to Application of High Voltage," Bioelectrochemistry, **51**(1), pp. 41–51.
[6] Pliquett, U., Zewert, T. E., Chen, T., Langer, R., and Weaver, J. C., 1996, "Imaging of Fluorescent Molecule and Small Ion Transport Through Human Stratum Corneum During High Voltage Pulsing: Localized Transport Regions Are Involved," Biophys. Chem., **58**(1–2), pp. 185–204.
[7] Prausnitz, M. R., 1996, "Do High Voltage Pulses Cause Changes in Skin Structure?," J. Controlled Release, **40**(3), pp. 321–326.
[8] Vanbever, R., and Preat, V., 1999, "In Vivo Efficacy and Safety of Skin Electroporation," Adv. Drug Delivery Rev., **35**(1), pp. 77–88.
[9] Pliquett, U., Langer, R., and Weaver, J. C., 1995, "Changes in the Passive Electrical Properties of Human Stratum Corneum Due to Electroporation," Biochim. Biophys. Acta, **1239**(2), pp. 111–121.
[10] Krassowska, K., and Filev, P., 2007, "Modeling Electroporation in a Single Cell," Biophys. J., **92**(2), pp. 404–417.
[11] Smith, K. C., Neu, J. C., and Krassowska, W., 2004, "Model of Creation and Evolution of Stable Electropores for DNA Delivery," Biophys. J., **86**(5), pp. 2813–2826.
[12] Stewart, D. A., Gowrishankar, T. R., and Weaver, J. C., 2004, "Transport Lattice Approach to Describing Cell Electroporation: Use of a Local Asymptotic Model," IEEE Trans. Plasma Sci., **32**(4), pp. 1696–1708.
[13] Satkauskas, S., Andre, F., Bureau, M. F., Scherman, D., Miklavcic, D., and Mir, L. M., 2005, "Electrophoretic Component of Electric Pulses Determines the Efficacy of *In Vivo* DNA Electrotransfer," Hum. Gene Ther., **16**(10), pp. 1194–1201.
[14] Pliquett, U. F., Vanbever, R., Preat, V., and Weaver, J. C., 1998, "Local Transport Regions (LTRs) in Human Stratum Corneum Due to Long and Short 'High Voltage' Pulses," Bioelectrochem. Bioenerg., **47**(1), pp. 151–161.
[15] Pliquett, U., Gallo, S., Hui, S. W., Gusbeth, C., and Neumann, E., 2005, "Local and Transient Changes in Stratum Corneum at High Electric Fields: Contribution of Joule Heating," Bioelectrochemistry, **67**(1), pp. 37–46.
[16] Vanbever, R., Pliquett, U. F., Preat, V., and Weaver, J. C., 1999, "Comparison of the Effects of Short, High-Voltage and Long, Medium-Voltage Pulses on Skin Electrical and Transport Properties," J. Controlled Release, **60**(1), pp. 35–47.
[17] Golden, G. M., Guzek, D. B., Kennedy, A. H., Mckie, J. E., and Potts, R. O., 1987, "Stratum Corneum Lipid Phase-Transitions and Water Barrier Properties," Biochemistry, **26**(8), pp. 2382–2388.
[18] Potts, R. O., and Francoeur, M. L., 1990, "Lipid Biophysics of Water Loss Through the Skin," Proc. Natl. Acad. Sci. U.S.A., **87**(10), pp. 3871–3873.
[19] Al-Saidan, S. M., Barry, B. W., and Williams, A. C., 1998, "Differential Scanning Calorimetry of Human and Animal Stratum Corneum Membranes," Int. J. Pharm., **168**(1), pp. 17–22.
[20] Cornwell, P. A., Barry, B. W., Bouwstra, J. A., and Gooris, G. S., 1996, "Modes of Action of Terpene Penetration Enhancers in Human Skin Differential Scanning Calorimetry, Small-Angle X-Ray Diffraction and Enhancer Uptake Studies," Int. J. Pharm., **127**(1), pp. 9–26.
[21] Tanojo, H., Bouwstra, J. A., Junginger, H. E., and Bodde, H. E., 1999, "Thermal Analysis Studies on Human Skin and Skin Barrier Modulation by Fatty Acids and Propylene Glycol," J. Therm Anal. Calorim., **57**(1), pp. 313–322.
[22] Silva, C. L., Nunes, S. C. C., Eusebio, M. E. S., Pais, A. A. C. C., and Sousa, J. J. S., 2006, "Study of Human Stratum Corneum and Extracted Lipids by Thermomicroscopy and DSC," Chem. Phys. Lipids, **140**(1–2), pp. 36–47.
[23] Silva, C. L., Nunes, S. C. C., Eusebio, M. E. S., Sousa, J. J. S., and Pais, A. A. C. C., 2006, "Thermal Behaviour of Human Stratum Corneum—A Differential Scanning Calorimetry Study at High Scanning Rates," Skin Pharmacol. Physiol., **19**(3), pp. 132–139.
[24] White, S. H., Mirejovsky, D., and King, G. I., 1988, "Structure of Lamellar Lipid Domains and Corneocyte Envelopes of Murine Stratum Corneum. An X-Ray-Diffraction Study," Biochemistry, **27**(10), pp. 3725–3732.
[25] Pliquett, U., and Gusbeth, C., 2004, "Surface Area Involved in Transdermal Transport of Charged Species Due to Skin Electroporation," Bioelectrochemistry, **65**(1), pp. 27–32.
[26] Pliquett, U., Gusbeth, C., and Nuccitelli, R., 2008, "A Propagating Heat Wave Model of Skin Electroporation," J. Theor. Biol., **251**(2), pp. 195–201.
[27] Higuchi, W. I., Li, K. S., Abdel-Halim, G., Honggang, Z., and Yang, S., 1999, "Mechanistic Aspects of Iontophoresis in Human Epidermal Membrane," J. Controlled Release, **62**(1–2), pp. 13–23.
[28] Li, S. K., Higuchi, I., Kochambilli, R. P., and Zhu, H., 2004, "Mechanistic Studies of Flux Variability of Neutral and Ionic Permeants During Constant Current dc Iontophoresis With Human Epidermal Membrane," Int. J. Pharm., **273**(1–2), pp. 9–22.
[29] Michaels, A. S., Chandraskeran, S. K., and Shaw, J. E., 1975, "Drug Permeation Through Human Skin: Theory and In Vitro Experimental Measurement," AIChE J., **21**(5), pp. 985–996.
[30] Johnson, M. E., Blankschtein, D., and Langer, R., 1997, "Evaluation of Solute Permeation Through the Stratum Corneum: Lateral Bilayer Diffusion as the Primary Mechanism," J. Pharm. Sci., **86**(10), pp. 1162–1172.
[31] Wang, T., Kasting, G. B., and Nitsche, J. M., 2006, "A Multiphase Microscopic Diffusion Model for Stratum Corneum Permeability. I. Formulation, Solution, and Illustrative Results for Representative Compounds," J. Pharm. Sci., **95**(3), pp. 620–648.
[32] Frasch, H. F., and Barbero, A. M., 2003, "Steady-State Flux and Lag Time in the Stratum Corneum Lipid Pathway: Results From Finite Element Models," J. Pharm. Sci., **92**(11), pp. 2196–2207.
[33] Barbero, A. M., and Frasch, H. F., 2005, "Modeling of Diffusion With Partitioning In Stratum Corneum Using a Finite Element Model," Ann. Biomed. Eng., **33**(9), pp. 1281–1292.
[34] Kushner, J., Deen, W., Blankschtein, D., and Langer, R., 2007, "First-Principles, Structure-Based Transdermal Transport Model to Evaluate Lipid Partition and Diffusion Coefficients of Hydrophobic Permeants Solely From Stratum Corneum Permeation Experiments," J. Pharm. Sci., **96**(12), pp. 3236–3251.
[35] Mitragotri, S., 2002, "A Theoretical Analysis of Permeation of Small Hydrophobic Solutes Across the Stratum Corneum Based on Scaled Particle Theory," J. Pharm. Sci., **91**(3), pp. 744–752.
[36] Lian, G., Chen, L., and Han, L., 2008, "An Evaluation of Mathematical Models for Predicting Skin Permeability," J. Pharm. Sci., **97**(1), pp. 584–598.
[37] Mitragotri, S., 2003, "Modeling Skin Permeability to Hydrophilic and Hydrophobic Solutes Based on Four Permeation Pathways," J. Controlled Release, **86**(1), pp. 69–92.
[38] Kushner, J., Blankschtein, D., and Langer, R., 2007, "Evaluation of the Porosity, the Tortuosity, and the Hindrance Factor for the Transdermal Delivery of Hydrophyllic Permeants in the Context of the Aqueous Pore Pathway Hypothesis Using Dual-Radiolabeled Permeability Experiments," J. Pharm. Sci., **96**(12), pp. 3263–3282.
[39] Tezel, A., Sens, A., and Mitragotri, S., 2003, "Description of Transdermal Transport of Hydrophilic Solutes During Low-Frequency Sonophoresis Based on a Modified Porous Pathway Model," J. Pharm. Sci., **92**(2), pp. 381–393.
[40] Li, S. K., Ghanem, A. H., Teng, C. L., Hardee, G. E., and Higuchi, W. I., 2001, "Iontophoretic Transport of Ogligonucleotides Across Human Epidermal Membrane: A Study of the Nernst-Plank Model," J. Pharm. Sci., **90**(7), pp. 915–931.
[41] Deen, W. M., 1987, "Hindered Transport of Large Molecules in Liquid-Filled Pores," AIChE J., **33**(9), pp. 1409–1425.
[42] Chizmadzhev, Y. A., Indenbom, A. V., Kuzmin, P. I., Galichenko, S. V., Weaver, J. C., and Potts, R. O., 1998, "Electrical Properties of Skin at Moderate Voltages: Contribution of Appendageal Macropores," Biophys. J., **74**(2), pp. 843–856.
[43] Pavselj, N., Bregar, Z., Cukjati, D., Batiuskaite, D., Mir, L. M., and Miklavcic, D., 2005, "The Course of Tissue Permeabilization Studied on a Mathematical Model of a Subcutaneous Tumor in Small Animals," IEEE Trans. Biomed. Eng., **52**(8), pp. 1373–1381.
[44] Pavšelj, N., Préat, V., and Miklavčič, D., 2007, "A Numerical Model of Skin Electropermeabilization Based on *In Vivo* Experiments," Ann. Biomed. Eng., **35**(12), pp. 2138–2144.
[45] Becker, S. M., and Kuznetsov, A. V., 2008, "Thermal *In Vivo* Skin Electroporation Pore Development and Charged Macromolecule Trans-Dermal Delivery: A Numerical Study of the Influence of Chemically Enhanced Lower Lipid Phase Transition Temperatures," Int. J. Heat Mass Transfer, **51**(7–8), pp. 2060–2074.
[46] Becker, S. M., and Kuznetsov, A. V., 2007, "Thermally Induced Pore Growth Associated With In Vivo Skin Electroporation: A Numerical Model of Lipid

Phase Transition Influence," ASME J. Biomech. Eng., **129**(5), pp. 712–721.
[47] Ozisik, M. N., 1993, *Heat Conduction*, Wiley, New York.
[48] Pennes, H. H., 1948, "Analysis of Tissue and Arterial Blood Temperatures in the Resting Forearm," J. Appl. Physiol., **1**, pp. 93–122.
[49] Craane-van Hinsberg, W. H. M., Verhoef, J. C., Junginger, H. E., and Bodde, H. E., 1995, "Thermoelectrical Analysis of the Human Skin Barrier," Thermochim. Acta, **248**, pp. 303–318.
[50] Cussler, E., Hughes, S., Ward, W., and Aris, R., 1988, "Barrier Membranes," J. Membr. Sci., **38**(2), pp. 161–174.
[51] Lange-Lieckfeldt, R., and Lee, G., 1992, "Use of a Model Lipid Matrix to Demonstrate the Dependence of the Stratum-Corneum Barrier Properties on Its Internal Geometry," J. Controlled Release, **20**(3), pp. 183–194.

**Tzuen-Rong J. Tzeng**
e-mail: tzuenrt@clemson.edu

**Yunyan R. Cheng**
e-mail: ycheng@clemson.edu

Department of Biological Sciences,
Clemson University,
132 Long Hall,
Clemson, SC 29634-0314

**Reza Saeidpourazar**
e-mail: rezas@illinois.edu
Department of Materials Science and Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801

**Siddharth S. Aphale**
Department of Mechanical Engineering,
Clemson University,
Clemson, SC 29634

**Nader Jalili**
Piezoactive Systems Laboratory,
Department of Mechanical and Industrial Engineering,
Northeastern University,
Boston, MA 02115

# Adhesin-Specific Nanomechanical Cantilever Biosensors for Detection of Microorganisms

*Lectins (adhesins) on bacterial surfaces play important roles in infection by mediating bacterial adherence to host cell surfaces via their cognate receptors. We have explored the use of α-D-mannose receptors as capturing agents for the detection of Escherichia coli using a microcantilever and have demonstrated that E. coli ORN178, which expresses normal type-1 pili, can interact with microcantilevers functionalized with α-D-mannose and can cause shifts in its resonance frequencies. Although E. coli ORN208, which expresses abnormal pili, binds poorly to α-D-mannose on the nitrocellulose membrane of a FAST slide, it did cause a detectable shift in resonance frequency when interacting with the α-D-mannose functionalized microcantilevers.*
[DOI: 10.1115/1.4002363]

## 1 Introduction

Capability to detect pathogens in small amounts is important for national security and medical diagnostics [1]. The traditional geographical boundaries that once curbed the spread of disease have dissipated due to modern trade and travel [2]. Conventional methods are well established for pathogen detection, but they often require a long analysis time. Antibody-based pathogen sensors offer the potential of a rapid analysis, but the production of specific antibodies is expensive, difficult, and time consuming [3]. From all the incidents over the past decades, such as the anthrax letters, the tomato Salmonella outbreak, and the spinach *Escherichia coli* O157:H7 outbreak, we have learned that platforms that enable accurate and rapid detection of pathogens are especially vital.

Microcantilevers have attracted a lot of attention recently due to their ability to serve as highly sensitive, real-time, multifunctional, and specific biological detectors [1]. First, the ultrasmall size of microcantilevers scales down the amount of the biological sample required for detection [4]. Second, the detection of biological samples with microcantilevers has the potential to be label-free. Therefore, time-consuming and expensive labeling can be eliminated and cross-talk interference between the tags and target samples can be avoided [5]. Third, the conventional micromachining techniques make mass production of microcantilevers easy and inexpensive [4]. Different reagents can be applied during the manufacturing process to arrays of parallel fabricated microcantilevers [5–7], benefitting the area of pathogen detection.

Microcantilevers have a relatively large base with a beam anchored at one end or both ends. The beam resembles a miniature diving board [4]. The detection system involving a microcantilever is called an atomic force microscope (AFM) system [8]. A typical AFM optical detection system consists of a sharp tip mounted to a piezoelectric (PZT) actuator and a laser beam photodetector, which is sensitive to the position of deflection feedback by the end point of the beam in the system [8]. During an AFM test, the PZT scanners maintain the tip at a constant height above the sample surface or at a constant force controlled by computer software feedback mechanisms, while the machine scans the tip over the sample surface [8]. The cantilever surface deflects different light intensities of the laser beam between the upper and lower photodetectors [8].

An AFM microcantilever detection system operates via the adsorption of a sample onto the cantilever surface, and the cantilever responds to stress, damping, and mass changes [1,4]. Microcantilevers bend with the molecular adsorption, which changes the surface stress and is observed as changes in microcantilever deflection [4]. A frequently used parameter is the resonance frequency, which changes sensitively with molecular adsorption [1,4,8]. Microcantilever systems can be divided into three types based on their operation principles: the noncontact mode, the contact mode, and the tapping mode [8,9]. In the noncontact mode, the cantilever is positioned slightly away from the sample surface and is oscillated near its natural resonance frequency [9]. The adsorption of samples onto the cantilever changes its natural resonance frequency, and the system can detect the difference and create topographical information of the samples [9]. In the contact mode, the cantilever tip always contacts the sample surface to test the interaction forces [10]. The cantilever tip is oscillated at or near its natural resonance frequency in the tapping mode, while the tip is allowed to contact the sample surface for a short time [11,12].

Based on the analyte detection method, there are two different

Manuscript received May 16, 2010; final manuscript received July 15, 2010; published online September 30, 2010. Assoc. Editor: Andrey Kuznetsov.

modes in the microcantilever biosensor system [13,14]. The first one is called the static mode, where the adsorption induced deflection is measured [15]. On an asymmetrically functionalized cantilever surface, the adsorption generates a surface stress, which forces the beam to bend and gives a measurable signal [15]. But the static mode is not practical when both sides of the microcantilever are functionalized with the same molecules. The other mode is the dynamic mode, where the shift in resonance frequency is induced by the adsorption [15]. When the target molecules are adsorbed onto the microcantilever surface, its overall mass changes; therefore, the natural frequency is altered by a small but detectable amount. This method is practical for both one-sided and two-sided functionalizations.

Initiation of pathogenesis occurs when bacterial cell surface adhesins bind to host cell surface receptors. The adherence of a particular pathogen is often to specific receptors on the host cell surface. Many pathogens exhibit regional specificity in their pathogenesis, which can be explained through the adherence to unique receptors on specific tissues. Many carbohydrates serve as the recognition molecules on the host cell surface in the adhesion of pathogens to the host [16].

Detection methods based on the binding interactions between bacterial pathogens and the corresponding carbohydrate receptors have been developed [17]. Fluorescence [18] and optically [19] or thermally [20] responsive synthetic glycopolymers have been exploited in such a sensor system. However, the current developing sensors rely on the detection of adhesin molecules [1,21] and are less ideal for real-time detection. The approach demonstrated in this work allows for the detection of label-free whole cells. The objective of this research is to develop and demonstrate that microcantilevers functionalized with $\alpha$-D-mannose, a carbohydrate receptor recognized by bacteria expressing type-1 fimbriae, could be used to detect the presence of *E. coli* ORN178.

## 2 Material and Methods

**2.1 Bacterial Strains.** The *E. coli* strains ORN178 and ORN208 were kindly provided by Chu-Cheng Lin in the Department of Zoology, National Taiwan Normal University, Taiwan. The *E. coli* ORN178 strain expresses type-1 fimbriae, and the *E. coli* ORN208 expresses abnormal type-1 fimbriae. Fluorescent strains for *E. coli* ORN178 and ORN208 were constructed by transforming them with plasmid pGREEN (Carolina Biological, Burlington, NC). The pGREEN plasmid encodes for green fluorescent proteins (GFPs). *E. coli* strains were cultivated on tryptic soy agar plates or in tryptic soy broth at 37°C overnight.

**2.2 Yeast Agglutination Assay.** *E. coli* ORN178-GFP and ORN208-GFP were suspended in phosphate buffer saline (PBS). The bacterial suspensions were mixed with yeast suspension at a ratio of 1 to 1. Wet mount slides were observed under a Motic AE31 inverted phase-contrast/fluorescence microscope with a filter for GFP.

**2.3 Carbohydrate Microarray System.** Whatman MicroCaster™ Manual Microarraying System with an eight-pin hand tool and Whatman FAST slides with 1, 2, 8, or 16 pads were used for the microarray setup. The Whatman FAST slides consist of a nitrocellulose membrane, which binds to proteins and polyacrylamide (PAA) conjugated carbohydrates in an irreversible and noncovalent manner. The size of the printed carbohydrate spots ranges from 500 $\mu$m to 1000 $\mu$m, which is ideal for bacterial hybridization. The volume of the printed carbohydrate solution ranges from 20 nl to 70 nl. With the indexing system, an array of $32\times24$ spots can be precisely printed, a total of 768 spots. PAA conjugated carbohydrate was mixed with protein arraying buffer ($2\times$) from the Whatman FAST PAK 16 protein array kit. The concentrations of carbohydrates printed on the slide were 10 $\mu g/\mu l$, 5 $\mu g/\mu l$, 2.5 $\mu g/\mu l$, 1 $\mu g/\mu l$, and 0.1 $\mu g/\mu l$.

After printing, carbohydrate spots were air-dried at room temperature. Baking the Whatman slide at 80°C for 30 min or leaving it at room temperature overnight irreversibly immobilized the PAA conjugated carbohydrate spots onto the nitrocellulose membrane. Blocking of the printed FAST slides and hybridization of *E. coli* strains to the slides were carried out following the manufacturer's recommended procedures. Briefly, blocking of the printed FAST slides was carried out, with the protein array blocking buffer from the Whatman FAST PAK 16 protein array kit on a microtiter plate shaker at 250 rpm for 30 min at room temperature. After bacterial hybridization, the slides were washed three times, 5 min each with gentle shaking in the protein array wash buffer. A Motic AE31 inverted fluorescence microscope with a filter for GFP was used to observe the results.

**Table 1 Specifications of the one-side gold coated microcantilever from BudgetSensors**

| | Value | Range |
|---|---|---|
| Resonant frequency | 13 kHz | ±4 kHz |
| Force constant | 0.2 N/m | 0.07–0.4 N/m |
| Length | 450 $\mu$m | ±10 $\mu$m |
| Mean width | 50 $\mu$m | ±5 $\mu$m |
| Thickness | 2 $\mu$m | ±1 $\mu$m |
| Tip height | 17 $\mu$m | ±2 $\mu$m |
| Tip set back | 15 $\mu$m | ±5 $\mu$m |
| Tip radius | | <10 nm |
| Reflex coating | 70 nm gold on the detector side | |
| Half cone angle | 20–25 deg along the cantilever axis<br>25–30 deg from the side<br>10 deg at the apex | |

**2.4 Preparation of Nonviable *E. coli* Cells.** *E. coli* ORN178-GFP and ORN208-GFP cells were rendered nonviable by mixing with formaldehyde (25 mg/ml) for 40 min. The bacterial suspensions were then washed three times with distilled water and tested to see if they retain their carbohydrate binding specificity. The viability of the formaldehyde killed bacteria was confirmed by plating aliquots of the bacterial suspension on tryptic soy agar plates.

**2.5 Gold Surface Functionalization.** The immobilization strategy was based on hydrazide-derivatized self-assembled monolayers (SAMs) on the gold surface with an amine group at the free end [22]. The aldehyde group of a reducing sugar then reacted with the amine group by a condensation reaction to form an imine linkage, so that the carbohydrate was immobilized onto the gold surface [22]. The methods explained by Zhi et al. were followed [22]. Gold slides were purchased from Erie Scientific Co. (Portsmouth, NH). Carbohydrate printing and bacteria hybridization were carried out as described in the previous section

**2.6 Microcantilever Biosensor Development.** Gold microcantilevers were purchased from BudgetSensors. This microcantilever product is for contact mode with rotated monolithic silicon on one side and 70 nm gold on the detector side. It has a symmetric shape tip with a resonant frequency of 13 kHz and a force constant of 0.2 N/m. The chip size is $3.4\times1.6\times0.3$ mm$^3$. Detailed information provided by the manufacturer is presented in Table 1.

This one-sided coated gold microcantilever was coated with carbohydrate following the same protocol as previously described in Sec. 2.5. *E. coli* ORN178-GFP and ORN208-GFP were hybridized with the microcantilever using the optimized conditions from the carbohydrate microarray. The microcantilevers were observed under an epifluorescent microscope. Resonant frequencies of the uncoated, carbohydrate coated, and bacteria bound microcantilevers were measured with the Polytec Micro System Analyzer 400 (MSA-400, Fig. 1). MSA-400 has a stroboscopic video microscope, a built-in microscope with a scanning laser Doppler vibro-

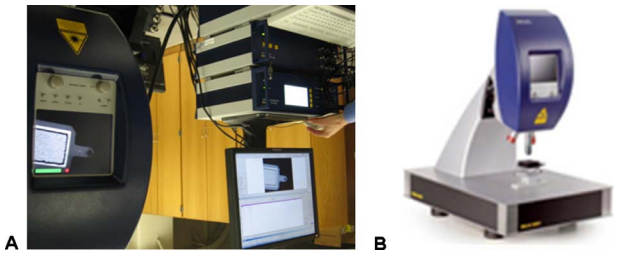


**Fig. 1 Polytec MSA 400 setup: (*a*) utilized at Clemson University Smart Structures and NEMS Laboratory and (*b*) listed on the website of Polytec**

meter, and a white light interferometer [23] was utilized to sense the position feedback of the microcantilever beam. The beam of the microcantilever was actuated by the real-time interface dSPACE 1103. The built-in vibrometer was utilized to analyze the frequency response. We used noncontact mode and dynamic mode so that the actuated microcantilever beam can be analyzed under a continuous domain model.

## 3 Results

**3.1 Yeast Agglutination Assay.** A yeast agglutination assay was carried out to confirm the mannose binding characteristic of type-1 fimbriae expressed on *E. coli* ORN178 (FimH+). Yeast was used to detect the presence of mannose-specific type-1 fimbriae since yeast cells express mannan, a polymer of mannose, on their cell surface [24]. The specific binding interaction involves the FimH adhesin at the end of type-1 fimbriae [25,26] and can be found on many enterobacteria [25]. There are two domains on the FimH adhesin [26]. The lectin domain (Ld) (1–156 aa) binds to mannose with a monomannose binding site at the distal end of the $\beta$-barrel fold [26,27]. The other domain is the pilin domain (Pd) (160–273 aa) playing a role in fimbria incorporation [27].

As depicted in Fig. 2, yeast agglutination was mediated by *E. coli* ORN178 (FimH+) but not by *E. coli* ORN208 (FimH-). This assay confirmed that the adhesin FimH facilitates the agglutination of yeast cells through the interaction of *E. coli* ORN178 pili to mannose polymers on the yeast cell surface.

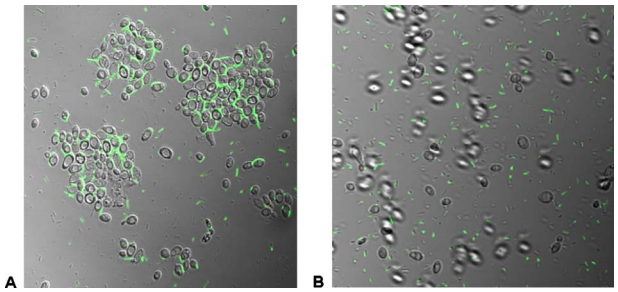


**Fig. 2 Yeast agglutination test. Superimposed images obtained from phase contrast microscope and epifluorescent microscope demonstrating (*a*) the agglutination of yeast cells mediated by *E. coli* ORN178-GFP (FimH+) and (*b*) the absence of yeast cell agglutination with *E. coli* ORN208-GFP (FimH-).**

**3.2 Carbohydrate Microarray Hybridization.** Researchers have shown that glass surfaces can be used to develop a DNA array system [28,29], and nitrocellulose membranes can be used in a protein array system [30]. However, oligosaccharides bind poorly to both surfaces because of their small molecular weight. However, PAA conjugated carbohydrates (GlycoTech, Inc., MD) can be easily immobilized onto a nitrocellulose membrane since polyacrylamide permits the immobilization by its hydrophobic interactions with the nitrocellulose membrane and its high molecular weight. The binding specificity of *E. coli* type-1 fimbriae for $\alpha$-D-mannose-PAA was evaluated on Whatman FAST 16-pad slides.

The binding of *E. coli* ORN178-GFP to $\alpha$-D-mannose-PAA is depicted in Fig. 3. The concentrations of the immobilized carbohydrate dots were 10 $\mu$g/$\mu$l, 5 $\mu$g/$\mu$l, 2.5 $\mu$g/$\mu$l, 1 $\mu$g/$\mu$l, and 0.1 $\mu$g/$\mu$l from top to bottom in each of the six sections. While the carbohydrate concentration decreased, the fluorescence of bound bacterial cells also decreased, indicating that the binding between *E. coli* ORN178 and $\alpha$-D-mannose is carbohydrate concentration dependent (Fig. 3(*a*)). *E. coli* ORN208-GFP (FimH-) with abnormal type-1 fimbriae and *E. coli* O157:H7 not expressing type-1 fimbriae [31] did not bind to $\alpha$-D-mannose-PAA (Figs. 3(*c*) and 3(*f*)). The binding of *E. coli* ORN178-GFP (FimH+) to immobilized $\alpha$-D-mannose-PAA was also inhibited when the bacterial cells were pre-exposed to free $\alpha$-D-mannose (Fig. 3(*e*)). It is demonstrated that the binding of *E. coli* ORN178-GFP (FimH+) to immobilized $\alpha$-D-mannose-PAA is type-1 fimbriae and $\alpha$-D-mannose specific.

**3.3 Binding of Nonviable *E. coli* ORN178 to $\alpha$-D-mannose.** One of the long term goals of this research is to be able to detect pathogens in field samples. Both the target and nontarget microorganisms may have potential virulence. If nonviable cells maintain their binding specificity, we can inactivate field samples prior

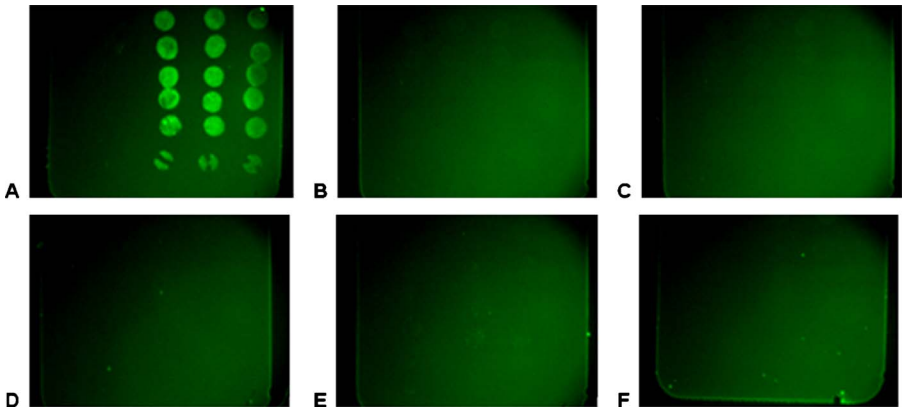


**Fig. 3 Binding of (*a*) *E. coli* ORN178-GFP to immobilized $\alpha$-D-mannose-PAA, (*b*) PBS to immobilized $\alpha$-D-mannose-PAA, (*c*) *E. coli* ORN208-GFP to immobilized $\alpha$-D-mannose-PAA, (*d*) *E. coli* ORN178-GFP to immobilized $\alpha$-D-galactose-PAA, (*e*) *E. coli* ORN178-GFP, pre-exposed to free mannose, to immobilized $\alpha$-D-mannose-PAA, and (*f*) *E. coli* O157:H7-GFP to immobilized $\alpha$-D-mannose-PAA on Whatman FAST 16-pad slides**

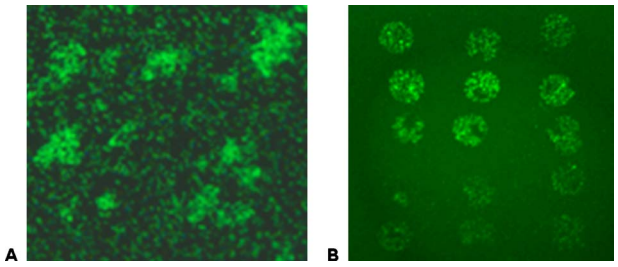


**Fig. 4 Binding of formaldehyde killed *E. coli* ORN178-GFP (FimH+) to (*a*) yeast and to (*b*) $\alpha$-D-mannose-PAA**

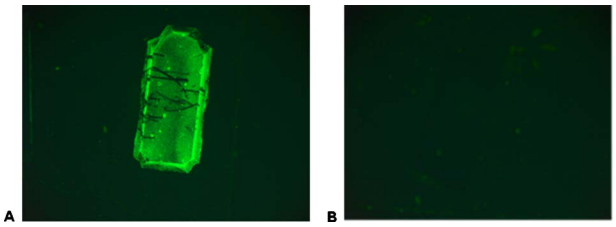


**Fig. 6 Binding results of (*a*) *E. coli* ORN178-GFP (FimH+) and (*b*) *E. coli* ORN208-GFP (FimH-) to $\alpha$-D-mannose-PAA on a gold microcantilever base**

to subjecting them to the detection procedures for safer handling of field samples. To evaluate whether nonviable cells retain the adhesin-receptor specificity, *E. coli* ORN178-GFP (FimH+) was killed using formaldehyde and hybridized with yeast cells (Fig. 4(*a*)) or $\alpha$-D-mannose-PAA, which was immobilized on a Whatman FAST slide (Fig. 4(*b*)). This killed strain retained the ability to agglutinate yeast cells and to bind to $\alpha$-D-mannose. So, the adhesin-receptor specificity was still maintained by the formaldehyde killed *E. coli* ORN178-GFP cells. Formaldehyde preserves or fixes bacterial cells by irreversibly cross-linking primary amino groups in proteins through a–CH2-linkage [32]. Since formaldehyde fixation causes no interference with bacterial adhesion [33], it can serve as a potential reagent in field samples with the ability to kill pathogens, leaving the adhesins intact.

**3.4 Bold Surface Functionalization.** Functionalization of biosensors will be accomplished via attachment of carbohydrates to gold coated microcantilevers. Binding of fluorescent *E. coli* ORN178-GFP to $\alpha$-D-mannose-PAA functionalized gold surface is depicted in Fig. 5. We demonstrated that the carbohydrate can be immobilized onto the gold surface and that *E. coli* ORN178-GFP retains the adhesin-receptor specificity. These results confirm that the protocol can be used with a gold microcantilever.

**3.5 Microcantilever Biosensor Development.** After carbohydrate functionalization and bacteria exposure, the base of the microcantilever was observed under an epifluorescent microscope. As depicted in Fig. 6, we have demonstrated that *E. coli* ORN178-GFP can bind to the gold surface of a microcantilever functionalized with $\alpha$-D-mannose-PAA. Although the beam was not tested, the result gives us the confidence that the *E. coli* ORN178-GFP cells can also bind to the beam since the beam has the same gold coating as the base.

Utilizing MSA-400, we measured the resonance frequencies of uncoated microcantilevers, carbohydrate functionalized microcantilevers, and bacteria bound microcantilevers. The results are depicted in Table 2. The shift in resonance frequency from 10.57 kHz to 10.83 kHz after carbohydrate functionalization demonstrated that the carbohydrate was coated on the surface of the microcantilever beam. The resonance frequency was further shifted to 11.27 kHz when exposed to *E. coli* ORN208-GFP and to 11.37 kHz when exposed to ORN178-GFP.

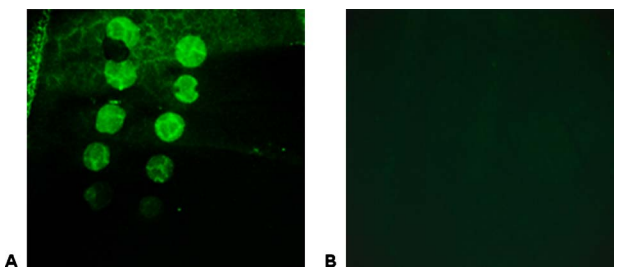


**Fig. 5 Binding results of (*a*) *E. coli* ORN178-GFP (FimH+) and (*b*) *E. coli* ORN208-GFP (FimH-) to $\alpha$-D-mannose-PAA on a gold slide**

## 4 Discussion

Carbohydrate microarray systems provide a platform for presenting many molecules on one chip to identify the bacterial adhesin binding profile to different carbohydrates. A better understanding of bacteria-carbohydrate adhesion interactions would greatly aid in the improvement of diagnosis, detection, and therapeutic tools [34]. The carbohydrate microarray system enables us to identify, verify, and optimize the binding characteristics of *E. coli* ORN178 to its carbohydrate receptors. Here, we have demonstrated the specific binding between *E. coli* ORN178 and $\alpha$-D-mannose-PAA. With the developed carbohydrate microarray system, we will be able to identify the binding specificity of many bacteria to monosaccharides, disaccharides, trisaccharides, oligosaccharides, polysaccharides, glycoproteins, and glycolipids. A carbohydrate library based on the carbohydrate binding specificity can be generated with a single nitrocellulose slide since as many as 768 spots can be printed on it. The information can be analyzed with the help of bioinformatics tools and benefit the understanding of lectin-carbohydrate binding interactions and the adherence patterns to host cell surfaces by pathogens. The information obtained from this carbohydrate microarray can then be used in the biosensor development.

Biosensors utilize interactions between biomolecules as a "sensing" mechanism [4]. Microcantilever sensors have been used in designing many physical [35], chemical [36,37], and biological [38] sensors. They also have the potential as a platform for the study of various molecular interactions, such as cell/protein and drug/protein intermolecular forces [8]. With the information obtained from our carbohydrate microarray system, an array of adhesin-specific microcantilever-based biosensors can be developed to accurately detect the presence of target pathogens in field samples with high sensitivity. However, the present work is performed as a batched process. For a real-time detection of pathogens, additional parameters such as the flow direction, the height of the fluidic cell, the mass transfer within the fluid cell, and the geometric size of the microcantilever's supporting system should be further characterized [39,40].

**Table 2 Frequencies of microcantilevers measured by MSA-400**

| | Resonance frequencies (kHz) |
|---|---|
| Uncoated microcantilever | 10.57 |
| Carbohydrate functionalized microcantilever | 10.83 |
| *E. coli* ORN208-GFP bound microcantilever | 11.27 |
| *E. coli* ORN178-GFP bound microcantilever | 11.37 |

## 5 Conclusions

We have demonstrated through microarray binding that *E. coli* ORN178 expressing type-1 fimbriae binds specifically to $\alpha$-D-mannose, while *E. coli* ORN208 expressing abnormal type-1 fimbriae binds poorly to $\alpha$-D-mannose, and nonviable *E. coli* ORN178 retain this binding characteristic.

We have also demonstrated that *E. coli* ORN178 can interact with microcantilevers functionalized with $\alpha$-D-mannose and cause shifts in their resonance frequencies. Although that *E. coli* ORN208 binds poorly to $\alpha$-D-mannose on the nitrocellulose membrane of a FAST slide, it did cause a detectable shift in resonance frequency when interacting with the $\alpha$-D-mannose functionalized microcantilevers.

To further quantify and elucidate the detection sensitivities of adhesin-specific microcantilever biosensors, a standard curve depicting resonance frequency versus a series of bacterial concentrations needs be established to determine the detection limits of such a system. To further elucidate the detection specificities of adhesin-specific microcantilever biosensors, mixed cultures of target, e.g., *E. coli* ORN178, and nontarget bacteria, e.g., *E. coli* O157:H7, should be prepared, and the shifts in resonance frequencies caused by the mixtures should be compared with those of the pure cultures.

## Acknowledgment

This material is based upon work supported by CSREES/USDA under Project No. SC-1700344 (T.-R. J. T.), as well as the National Science Foundation under MRI Program Grant No. CMMI-0619739 (N. J.).

## References

[1] Weeks, B., Camarero, J., Noy, A., Miller, A., Stanker, L., and De Yoreo, J., 2003, "Development of a Microcantilever-Based Pathogen Detector," Scanning, **25**(6), pp. 297–299.
[2] Uttamchandani, M., Neo, J. L., Ong, B. N. Z., and Moochhala, S., 2009, "Applications of Microarrays in Pathogen Detection and Biodefence," Trends Biotechnol., **27**(1), pp. 53–61.
[3] Skottrup, P. D., Nicolaisen, M., and Justesen, A. F., 2008, "Towards On-Site Pathogen Detection Using Antibody-Based Sensors," Biosens. Bioelectron., **24**, pp. 339–348.
[4] Hansen, K. M., and Thundat, T., 2005, "Microcantilever Biosensors," Methods, **37**(1), pp. 57–64.
[5] McKendry, R., Zhang, J., Arntz, Y., Strunz, T., Hegner, M., Lang, H. P., Baller, M. K., Certa, U., Meyer, E., Güntherodt, H.-J., and Gerber, C., 2002, "Multiple Label-Free Biodetection and Quantitative DNA-Binding Assays on a Nanomechanical Cantilever Array," Proc. Natl. Acad. Sci. U.S.A., **99**(15), pp. 9783–9788.
[6] Lockhart, D. J., and Winzeler, E. A., 2000, "Genomics, Gene Expression and DNA Arrays," Nature (London), **405**, pp. 827–836.
[7] Milburn, C., Zhou, J., Bravo, O., Kumar, C., and Soboyejo, W. O., 2005, "Sensing Interactions Between Vimentin Antibodies and Antigens for Early Cancer Detection," Journal of Biomedical Nanotechnology, **1**, pp. 30–38.
[8] Jalili, N., and Laxminarayana, K., 2004, "A Review of Atomic Force Microscopy Imaging Systems: Application to Molecular Metrology and Biological Sciences," Mechatronics, **14**(8), pp. 907–945.
[9] Basso, M., Giarre, L., Dahleh, M., and Mezic, I., 1998, "Numerical Analysis of Complex Dynamics in Atomic Force Microscopes," *Proceedings of the 1998 IEEE International Conference on Control Applications*, IEEE, New York, Vol. 2, pp. 1026–1030.
[10] Fung, R., and Huang, S., 2001, "Dynamic Modeling and Vibration Analysis of the Atomic Force Microscope," ASME J. Vibr. Acoust., **123**, pp. 502–509.
[11] Salapaka, M. V., Chen, D. J., and Cleveland, J. P., 1998, "Stability and Sensitivity Analysis of Periodic Orbits in Tapping Mode Atomic Force Microscopy," *Proceedings of the 37th IEEE Conference on Decision and Control*, IEEE, New York, Vol. 2, pp. 2047–2052.
[12] Sebastian, A., Salapaka, M., Chen, D., and Cleveland, J., "Harmonic Analysis Based Modeling of Tapping-Mode AFM," Proceedings of American Control Conference, pp. 232–236.
[13] Battiston, F. M., Ramseyer, J. P., Lang, H. P., Baller, M. K., Gerber, C., and Gimzewski, J. K., 2001, "A Chemical Sensor Based on a Microfabricated Cantilever Array With Simultaneous Resonance-Frequency and Bending Readout," Sens. Actuators B, **77**, pp. 122–131.
[14] Braun, T., Huber, F., Ghatkesar, M. K., Backmann, N., Lang, H., and Gerber, C., 2007, "Processing of Kinetic Microarray Signals," Sens. Actuators B, **128**, pp. 75–82.
[15] Braun, T., and Ghatkesar, M. K., 2009, "Measuring the Intrinsic Nanomechanics of Molecular Interactions With Microcantilever Sensors," Eur. J. Nanomedicine, **2**, pp. 13–15.
[16] Sharon, N., 2006, "Carbohydrates as Future Anti-Adhesion Drugs for Infectious Diseases," Biochim. Biophys. Acta, **1760**(4), pp. 527–537.
[17] Mukhopadhyay, B., Martins, M. B., Karamanska, R., Russell, D. A., and Field, R. A., 2009, "Bacterial Detection Using Carbohydrate-Functionalised CdS Quantum Dots: A Model Study Exploiting *E. coli* Recognition of Mannosides," Tetrahedron Lett., **50**(8), pp. 886–889.
[18] Xue, C., Jog, S. P., Murthy, P., and Liu, H., 2006, "Synthesis of Highly Water-Soluble Fluorescent Conjugated Glycopoly(*p* -phenylene)s for Lectin and *Escherichia coli*," Biomacromolecules, **7**(9), pp. 2470–2474.
[19] Disney, M. D., Zheng, J., Swager, T. M., and Seeburger, P. H., 2004, "Detection of Bacteria With Carbohydrate-Functionalized Fluorescent Polymers," J. Am. Chem. Soc., **126**(41), pp. 13343–13346.
[20] Pasparakis, G., Cockayne, A., and Alexander, C., 2007, "Control of Bacterial Aggregation by Thermoresponsive Glycopolymers," J. Am. Chem. Soc., **129**(36), pp. 11014–11015.
[21] Karamanska, R., Clarke, J., Blixt, O., MacRae, J., Zhang, J., Crocker, P., Laurent, N., Wright, A., Flitsch, S., Russell, D., and Field, R., 2008, "Surface Plasmon Resonance Imaging for Real-Time, Label-Free Analysis of Protein Interactions With Carbohydrate Microarrays," Glycoconjugate J., **25**(1), pp. 69–74.
[22] Zhi, Z.-l., Powell, A. K., and Turnbull, J. E., 2006, "Fabrication of Carbohydrate Microarrays on Gold Surfaces: Direct Attachment of Nonderivatized Oligosaccharides to Hydrazide-Modified Self-Assembled Monolayers," Anal. Chem., **78**, pp. 4786–4793.
[23] Saeidpourazar, R., and Jalili, N., 2009, "Towards Microcantilever-Based Force Sensing and Manipulation: Modeling, Control Development and Implementation," Int. J. Robot. Res., **28**, pp. 464–483.
[24] Firon, N., Ashkenazi, S., Mirelman, D., Ofek, I., and Sharon, N., 1987, "Aromatic Alpha-Glycosides of Mannose Are Powerful Inhibitors of the Adherence of Type 1 Fimbriated *Escherichia coli* to Yeast and Intestinal Epithelial Cells," Infect. Immun., **55**(2), pp. 472–476.
[25] Sharon, N., and Ofek, I., 2000, "Safe as Mother's Milk: Carbohydrates as Future Anti-Adhesion Drugs for Bacterial Diseases," Glycoconjugate J., **17**(7–9), pp. 659–664.
[26] Aprikian, P., Tchesnokova, V., Kidd, B., Yakovenko, O., Yarov-Yarovoy, V., Trinchina, E., Vogel, V., Thomas, W., and Sokurenko, E., 2007, "Interdomain Interaction in the FimH Adhesin of *Escherichia coli* Regulates the Affinity to Mannose," J. Biol. Chem., **282**(32), pp. 23437–23446.
[27] Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J., and Knight, S. D., 1999, "X-Ray Structure of the FimC-FimH Chaperone-Adhesin Complex From Uropathogenic *Escherichia coli*," Science, **285**, pp. 1061–1066.
[28] Ono, S. J., Nakamura, T., Ohbayashi, M., Dawson, M., Ikeda, Y., Nugent, A. K., Toda, M., and Jay, G., 2003, "Expression Profiling: Opportunities and Pitfalls and Impact on the Study and Management of Allergic Diseases," J. Allergy Clin. Immunol., **112**(6), pp. 1050–1056.
[29] Jafari, R., Arefi-Khonsari, F., Tatoulian, M., Le Clerre, D., Talini, L., and Richard, F., 2009, "Development of Oligonucleotide Microarray Involving Plasma Polymerized Acrylic Acid," Thin Solid Films, **517**(19), pp. 5763–5768.
[30] Stoevesandt, O., Taussig, M. J., and He, M., 2009, "Protein Microarrays: High-Throughput Tools for Proteomics," Expert Rev. Proteomics, **6**(2), pp. 145–157.
[31] Enami, M., Nakasone, N., Honma, Y., Kakinohana, S., Kudaka, J., and Iwanaga, M., 1999, "Expression of Type I Pili Is Abolished in Verotoxin-Producing *Escherichia Coli* O157," FEMS Microbiol. Lett., **179**(2), pp. 467–472.
[32] Fox, C. H., Johnson, F. B., Whiting, J., and Roller, P. P., 1985, "Formaldehyde Fixation," J. Histochem. Cytochem., **33**(8), pp. 845–853.
[33] Riise, G. C., Larsson, S., and Andersson, B. A., 1994, "Bacterial Adhesion to Oropharyngeal and Bronchial Epithelial Cells in Smokers With Chronic Bronchitis and in Healthy Nonsmokers," Eur. Respir. J., **7**, pp. 1759–1764.
[34] Love, K. R., and Seeberger, P. H., 2002, "Carbohydrate Arrays as Tools for Glycomics," Angew. Chem., Int. Ed., **41**(19), pp. 3583–3586.
[35] Finot, E., Lesniewska, E., Goudonnet, J. P., and Thundat, T., 2001, "Measuring Magnetic Susceptibilities of Nanogram Quantities of Materials Using Microcantilevers," Ultramicroscopy, **86**, pp. 175–180.
[36] Maute, M., Raible, S., Prins, F. E., Kern, D. P., Ulmer, H., Weimar, U., and Göpel, W., 1999, "Detection of Volatile Organic Compounds (VOCs) With Polymer-Coated Cantilevers," Sens. Actuators B, **58**(1–3), pp. 505–511.
[37] Thundat, T., Wachter, E. A., Sharp, S. L., and Warmack, R. J., 1995, "Detection of Mercury Vapor Using Resonating Microcantilevers," Appl. Phys. Lett., **66**(13), pp. 1695–1697.
[38] Meriaudeau, F., Ferrell, T. L., Arakawa, E. T., Wig, A., Passion, A., Thundat, T., Shen, W.-J., Patel, S., and Kraemer, F. B., 2001, "Study of Differential Hormone-Sensitive Lipase Concentrations Using a Surface Plasmon Resonance Sensor," Sens. Actuators B, **73**, pp. 192–198.
[39] Khanafer, K., and Vafai, K., 2005, "Geometrical and Flow Configurations for Enhanced Microcantilever Detection Within a Fluidic Cell," Int. J. Heat Mass Transfer, **48**(14), pp. 2886–2895.
[40] Khanafer, K., Alamiri, A., and Pop, I., 2010, "Fluid-Structure Interaction Analysis of Flow and Heat Transfer Characteristics Around a Flexible Microcantilever in a Fluidic Cell," Int. J. Heat Mass Transfer, **53**(9–10), pp. 1646–1653.