History of Key Technologies AIAA 82-4075

Space Guidance Evolution—A Personal Narrative

50th Anniversary 1981

Richard H. Battin
Charles Stark Draper Laboratory, Inc., Cambridge, Mass.

Introduction

THE prospect of preparing a comprehensive history of space guidance and navigation was, initially, a delight to contemplate. But, as the unproductive weeks went by, the original euphoria was gradually replaced by a sense of pragmatism. I reasoned that the historical papers which had the greatest appeal were written by "old timers" telling of their personal experiences. Since I had lived through the entire space age, and had the good fortune of being involved in many of the nation's important aerospace programs, I decided to narrow the scope to encompass only that of which I had personal knowledge. (It is, however, a sobering thought that you might qualify as an "old timer.")

The story begins in the early 1950's when the MIT Instrumentation Laboratory (later to become the Charles Stark Draper Laboratory, Inc.) was chosen by the Air Force Western Development Division to provide a self-contained guidance system backup to Convair in San Diego for the new Atlas intercontinental ballistic missile. The work was contracted through the Ramo-Wooldridge Corporation, and the technical monitor for the MIT task was a young engineer named Jim Fletcher who later served as the NASA Administrator.

The Atlas guidance system was to be a combination of an on-board autonomous system, and a ground-based tracking and command system. This was the beginning of a philosophic controversy, which, in some areas, remains unresolved. The self-contained system finally prevailed in ballistic missile applications for obvious reasons. In space exploration, a mixture of the two remains.

The electronic digital computer industry was in its infancy then, so that an on-board guidance system could be mechanized only with analog components. Likewise, the design and analysis tools were highly primitive by today's standards. It is difficult to appreciate the development problems without considering the available computational aids.

Computing in the Fifties

When I joined the MIT Instrumentation Lab in 1951, digital computation was performed with electrically driven mechanical desk calculators by a battery of young female operators. For analog computation, an electronic analog computer marketed by the Reeves Instrument Company, called the REAC, was used. The big innovation, which signalled the demise of the desk computers, was the IBM Card Programmed Calculator (CPC) acquired in 1952. Floating point calculations could now be made at the fantastic rate of one hundred per minute. But read-write memory was at a premium, and consisted of 27 mechanical counters each holding a ten decimal digit number with sign and housed in bulky units known as "ice boxes."

Development of the all-electronic digital computer was well underway at MIT in the early 1950's. Project Whirlwind produced an enormous machine, completely filling a large building off-campus, which boasted 1024 sixteen bit words electrostatically stored on cathode-ray tubes. We were fortunate to have access (albeit somewhat limited) to this marvel of the electronic age. (Today, of course, that same capability can be had on a single silicon chip.)

Richard H. Battin received an S.B. degree in Electrical Engineering in 1945 and a Ph.D. in Applied Mathematics in 1951—both from the Massachusetts Institute of Technology. Currently, he is Associate Department Head of the Charles Stark Draper Laboratory, Inc. and Adjunct Professor of Aeronautics and Astronautics at MIT. In 1956, in collaboration with Dr. J. Halcombe Laning Jr., he co-authored *Random Processes in Automatic Control*—a book which has appeared in Russian, French, and Chinese editions. His 1964 book *Astronautical Guidance* was also published in a Russian edition. Dr. Battin is a Fellow of the AIAA, and a member of the International Academy of Astronautics and the National Academy of Engineering. In 1972, he and David G. Hoag were presented by the AIAA with the Louis W. Hill Space Transportation Award (now called the Goddard Astronautics Award) "for leadership in the hardware and software design of the Apollo spacecraft primary control, guidance and navigation system which first demonstrated the feasibility of on-board space navigation during the historic flight of Apollo 8." He received the AIAA Mechanics and Control of Flight Award for 1978 and the Institute of Navigation Superior Achievement Award for 1980. "In recognition of outstanding teaching" the students of the MIT Department of Aeronautics and Astronautics honored him in 1981 with their first Teaching Award. Some of his past AIAA activities include Technical Committee memberships on Guidance and Control (1964-66), Computer Systems (1968-70), and Astrodynamics (1974-76, Chairman 1978-79). Currently, he serves the AIAA on the Board of Directors as a Director—Technical and as a member of the Honors and Awards Committee, the Technical Activities Committee, and the Institute Development Committee.

Received Aug. 7, 1981.

EDITOR'S NOTE: This manuscript was invited as a History of Key Technologies paper as part of AIAA's 50th Anniversary celebration. It is not meant to be a comprehensive survey of the field. It represents solely the author's recollection of events at the time, and is based upon his own experiences.

In the summer of 1952, following about six months experience as a user of Whirlwind, my boss, Dr. J. Halcombe Laning Jr., became enamored of the idea that computers should be capable of accepting conventional mathematical language directly, without the time consuming intermediate step of recasting engineering problems in an awkward, and all too error-prone, logic that was far removed from the engineer's daily experiences. Over the next few months he personally brought this idea to fruition with the successful development of the first algebraic compiler called, affectionately, "George" (from the old saw "Let George do it").

Of some interest are the first compiler statements successfully executed by "George":

```
x = 1,
Print x.
```

Unfortunately, this is not as well-known as

"Watson! Come here. I need you."

since few programmers are aware of this bit of folklore.

The first nontrivial program performed by George was a set of six nonlinear differential equations describing the lead-pursuit dynamics of an air-to-air fire control problem. The power of this grandfather of all compilers was aptly demonstrated—the equations were programmed in less than one hour, and successfully executed on the very first trial.

When "peripherals" were added to the Whirlwind computer, Hal Laning encouraged Neal Zierler to collaborate in extending, perfecting, and documenting[1] George. In June of 1954, almost two years after Hal had begun his work, John Backus and a team of programming researchers from IBM came to MIT for a demonstration of George. They were beginning work on a programming system for IBM's newly announced 704 calculator. As a result of this visit, algebraic expressions found their way into the Fortran language.[2]

For historical interest, a program I wrote in March 1954 using the George compiler to compute the Atlas missile trajectory is reproduced in Fig. 1. The notation was constrained by the symbol availability on a Flexowriter, a specially designed typewriter that produced a coded pattern of holes in a paper tape. Since only superscripts were available, subscripts were indicated with a vertical slash prefix. The upper case letter D in the program denotes d/dt. The symbols F^2 and F^3 designate the sine and cosine functions.

The use of and interest in George began to wane when our laboratory acquired its own stored program digital computer—an IBM type 650 Magnetic Drum Data Processing Machine—in the fall of 1954. But three years later, when tapes were available, Hal, with the help of Phil Hankins and Charlie Werner, initiated work on MAC—an algebraic programming language for the IBM 650, which was completed by early spring of 1958. Over the years MAC became the work-horse of the laboratory, and many versions were written to be hosted on the IBM 650, 704, 7090, and 360, as well as the Honeywell H800, H1800, and the CDC 3600.

MAC is an extremely readable language having a three-line format, vector-matrix notations and mnemonic and indexed subscripts.[3] (I had left the laboratory for "greener pastures" during the period of MAC's creation, and will always regret not participating in its development. But I take some solace in having originated the three-line format, which permits exponents and subscripts to assume their proper position in an equation. The idea was offered to IBM to use in Fortran but was dismissed as being "too hard to keypunch.") Unfortunately, after all these years of yeoman service, MAC seems destined to share the fate of Sanskrit, Babylonic cuneiform and other ancient but dead languages.

The high-order language called HAL, developed by Intermetrics, Inc. and used to program the NASA space shuttle

```
a=0.0065646,
e=0.472,
h=0.1,
p|^1=0,
p|^2=0.7853982,
n=0,
2  n=n+1,
q|^o=p| n+0.68766002,
q|^1=0,
q|^2=0,
r|^1=0,
r|^2=0,
t=0,
PRINT p| n.
CR5,
CR8,
CR9,
CR6,
CR7,
1  Dr|^1=r|^2,
5  r|^o=1/(1-eF^3(q|^o-p| n)),
Dq|^o=-1/r|^o r|^o,
Dr|^2=r|^o q|^2 q|^2-2q|^2/r|^o
+a(1-3(F^3(q|^o))^2+3q|^1 F^2(2q|^o))/(r|^o)^4,
Dq|^1=q|^2,
Dq|^2=-2eF^2(q|^o)(q|^2/r|^o+r|^1/(r|^o)^4)+2r|^2/(r|^o)^3
-2r|^2 q|^2/r|^o-a(F^2(2q|^o)+2q|^1 F^3(2q|^o))/(r|^o)^5,
8  t|^1=419.703t,
9  r=13571000(r|^o+r|^1),
6 PRINT t|^1,r,q|^o.
7  PRINT t,r|^o,r|^1,r|^2,q|^1,q|^2.
j=p| n-0.6981317-q|^o,
CP1,
k=n-1.5,
CP2,
STOP
```

Fig. 1 Atlas trajectory program illustrating the "George" compiler.

avionics computers, is a direct offshoot of MAC. Since the principal architect of HAL was Jim Miller, who co-authored with Hal Laning a report[3] on the MAC system, it is a reasonable speculation that the space shuttle language is named for Jim's old mentor, and not, as some have suggested, for the electronic superstar of the Arthur Clarke movie "2001-A Space Odyssey."

Since MAC was not then available on our IBM 650, some of the early analysis of the Atlas guidance system was made using a program, which Bob O'Keefe, Mary Petrick, and I developed, known as the MIT Instrumentation Laboratory Automatic Coding 650 Program or, simply, MITILAC.[4] We modeled the coding format to resemble that used for the CPC to minimize the transitional shock to those laboratory engineers who, though still uncomfortable with the digital computer, were beginning to wean themselves away from their more familiar analog devices.

MITILAC was soon superceded by BALITAC,[5] a mnemonic for Basic Literal Automatic Coding, because MITILAC programs were inefficient consumers of machine time. Besides, laboratory problems like Atlas guidance generally involved three-dimensional dynamics so that direct codes were provided (for the first time ever) to perform vector and matrix operations. (The coding format was alphanumeric, which was no easy trick to implement without an "alphabetic device"—obtainable from IBM for an additional monthly rental of $350 but too expensive for our budget.)

Delta Guidance

Initially, Hal Laning and I were the only ones at the laboratory involved in the analytical work for the Atlas guidance system. With no vast literature to search on "standard" methods of guiding ballistic missiles, we "invented" one.

Suppose r and v are the position and velocity vectors of a vehicle, and r_T is the position vector of the target. Then along any free-fall, target-intersecting trajectory there is a functional relation among the vectors r, v, and r_T. Call it

$$F(r,v,r_T)=0 \tag{1}$$

At the end of the powered or thrusting portion of the flight, this relationship must be satisfied if the missile is to hit the target.

A reference powered flight trajectory is chosen for which the "cut-off criterion" of Eq. (1) will be satisfied. Let the function F be expanded in a Taylor series about the reference cut-off values r_0, v_0. Thus

$$F(r_0+\Delta r,\ v_0+\Delta v,\ r_T) = F(r_0,\ v_0,\ r_T) + \left.\frac{\partial f}{\partial r}\right|_0 \Delta r + \left.\frac{\partial F}{\partial v}\right|_0 \Delta v + \ldots \tag{2}$$

where the function and its derivatives on the right-hand side are all evaluated on the reference path. (For each value of r along the reference trajectory there is a value of v for which F will vanish. Thus, each point is a potential cut-off point.)

In essence, then, the zeroth order term on the right of Eq. (2) is zero by definition of the reference path, and the linear terms are driven to zero by an autopilot. Thus, the function F, with off-nominal arguments, will eventually vanish (assuming second and higher order terms are negligible). There are, of course, complications of detail, which shall be ignored in this discussion.

The particular function F chosen for this purpose was

$$F(r,v,r_T) = (v\times r)\cdot[v\times(r_T-r)] + \mu r_T\cdot\left(\frac{r_T}{r_T}-\frac{r}{r}\right) \tag{3}$$

where μ is the Earth's gravitation constant. It is not a difficult exercise to show that $F=0$ is necessary and sufficient for a target intercept. However, I am unable to recall from whence the expression came. Since at that time neither Hal nor I were celestial mechanists (nor acquainted with any), the mystery is all the more puzzling.

Though simple in concept, the Delta guidance method (as it came to be called) is not easy to mechanize especially with analog hardware. First, considerable reference data must be stored; second, a complete navigation system is required; and third, time of flight errors are uncompensated, which will most certainly compromise system accuracy unless separately handled (with additional hardware, of course). Nevertheless, this is the system we were determined to make work, despite all of its deficiencies, until I made my first trip to Convair San Diego in the summer of 1955.

The Convair Legacy

The key figures at Convair were Charlie Bossart, the Chief Engineer, and Walter Schweidetzky, head of the guidance group. Walter had worked with Wernher von Braun at Peenemuende during World War II, and had a most delightful Spanish wife who served as our interpreter during the inevitable evening adventure in Tijuana.

I returned to Cambridge spouting a new vocabulary: "correlated flight path" and "correlated velocity" —"velocity-to-be-gained" and "distance-to-be-gained." The correlated flight path was a predetermined, free-fall reference trajectory designed to intercept the target. The nominal missile flight path would intersect the correlated path at the nominal cut-off point. To each point in time on the missile trajectory corresponded a point on the reference trajectory so that the missile velocity vector v_m was related in a one-to-one manner to a corresponding reference velocity v_c—the correlated velocity. The velocity-to-be-gained vector was the difference $v_g = v_c - v_m$; distance-to-be-gained was the time integral of v_g. A page from my old notebook illustrating these concepts is reproduced as Fig. 2.

As nearly as I can recall, the Convair mechanization proposal went something like this: If r is the position vector of the missile, and $r+\Delta r$ is the position of the correlated point on the reference path, then the correlated velocity would be obtained by a polynomial approximation utilizing a family of constant momentum (or constant energy) trajectories all passing through the target. In addition, a functional relationship between v_g and Δr could be determined since v_g was well approximated by the integral of the thrust acceleration a_T. (Near the cut-off point, the difference between the gravity terms along the actual and reference paths rapidly approach zero.) In short, Δr could be represented as

$$\Delta r = (s_g/v_g)\,v_g \tag{4}$$

where $s_g = \int v_g \mathrm{d}t$. An iteration loop was implied since v_c is computed from a polynomial function of $r+\Delta r$ while v_g is determined as $v_c - v_m$.

The Convair engineers recognized that the velocity-to-be-gained vector eventually remains essentially parallel to a fixed direction in inertial space, and proposed a number of control schemes to drive the velocity-to-be-gained to zero.

The immediate outcome of my trip to San Diego, and subsequent debriefing by Hal Laning, was his total abandonment of the Delta system. From that moment, the Delta guidance development was my millstone to bear. Hal no longer appeared interested in guiding the Atlas missile, or in anything else for that matter. But after what seemed like an eternity (several weeks at least), Hal reappeared with an idea and needed a sympathetic ear. It had to do with a redefinition of the concept of correlated velocity, and a simple differential equation for velocity-to-be-gained. In a few days, Delta guidance would be an orphan.

The Q-System

If r is the radius vector representing the position of the missile at an arbitrary time t after launch, the correlated velocity vector v_c was now to be defined as the velocity required by the missile at the position $r(t)$, in order that it

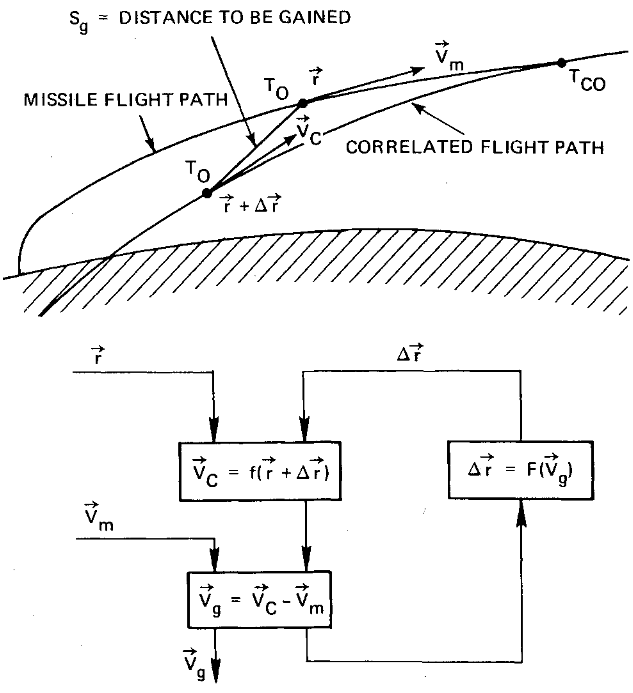

Fig. 2 Early concept of the correlated trajectory.

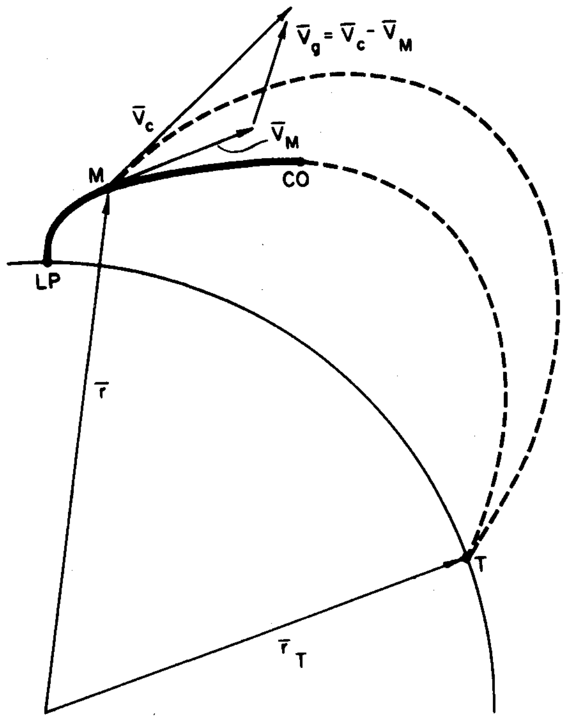

Fig. 3 Correlated trajectory and velocity-to-be-gained (from Ref. 6).

might travel thereafter by free-fall in vacuum into a desired terminal condition (here considered to be coincidence of the missile and a target on the Earth's surface although other applications to be discussed later are possible). For the definition of v_c to be unique, a further condition must be stipulated, such as the time at which the missile and target shall coincide. (This requirement would alleviate one of the deficiencies of Delta guidance.)

The point M in Fig. 3 represents the missile position at time t; the heavy line through M is the powered flight path terminating at the cut-off point (CO) in the elliptical free-fall trajectory shown as a dashed line to the target position T. Tangent to the correlated velocity vector v_c is a second ellipse, which would be followed by the missile in free-fall if it, indeed, possessed the velocity v_c at the point M.

Suppose, now, that when the missile is at the point M at time t, a "correlated missile" is simultaneously located at the same position. The correlated missile is assumed to experience only gravity acceleration g and moves with velocity v_c. The actual missile has velocity v_m and is affected by both gravity g and engine thrust acceleration a_T. During a time interval Δt, the two missiles are allowed to move "naturally" with the result that they will diverge in position by the amount

$$\Delta r = (v_m - v_c)\,\Delta t \tag{5}$$

Each experiences a velocity change given by

$$\Delta v_c = g\Delta t \qquad \Delta v_m = (g + a_T)\,\Delta t \tag{6}$$

At the end of this time interval, imagine that the correlated missile is brought back into coincidence with the actual missile. This change in position must be accompanied by a corresponding change in velocity if the terminal conditions imposed on the correlated missile are to remain satisfied. The appropriate change may be expressed as

$$\Delta v_c = Q\Delta r \tag{7}$$

where the elements of the matrix Q are the partial derivatives of the components of the velocity v_c with respect to the components of the position vector r. It is understood, in carrying out the differentiations, that the target location r_T and the time of free flight t_{ff} remaining (as well as t itself) are held fixed in the process. Thus, we have

$$Q = \frac{\partial v_c}{\partial r}\bigg|_{r_T, t_{ff}} \tag{8}$$

The total change in v_c as a result of these two steps is then

$$\begin{aligned} \Delta v_c &= g\Delta t + Q(v_m - v_c)\,\Delta t \\ &= g\Delta t - Qv_g\Delta t \end{aligned} \tag{9}$$

Finally, the change in velocity-to-be-gained Δv_g is simply the difference between Δv_c and Δv_m so that

$$\Delta v_g = -a_T\Delta t - Qv_g\Delta t \tag{10}$$

Division by Δt, and letting Δt approach zero, produces the fundamental differential equation for velocity-to-be-gained

$$\frac{\mathrm{d}v_g}{\mathrm{d}t} = -a_T - Qv_g \tag{11}$$

Behold the absence of the gravity vector! The necessity to compute Earth's gravity, an implied feature of Delta guidance, had vanished. In effect, almost all of the difficulties of the guidance problem were now bound up in the matrix Q. (Hal had a marvelous blackboard derivation of the fundamental equation utilizing block diagrams and an eraser. The audience never failed to be impressed when the block labeled g magically disappeared.)

To say that calculating the elements of the Q matrix was not a simple exercise would be a gross understatement. In our final report[6] on the Q-system it took fourteen pages of an appendix just to describe the necessary equations. Of what possible use could the v_g differential equation be if the coefficient matrix was that complex? (Had Delta guidance been abandoned too cavalierly?) We were encouraged to proceed because the Q matrix was so simple in the hypothetical case of a flat Earth with constant gravity.

With the vector g a constant, it is not difficult to show that

$$r_T = r + v_c t_{ff} + \tfrac{1}{2} g + t_{ff}^2 \tag{12}$$

is the appropriate relation for the problem variables. Therefore, it follows at once that

$$Q = -(1/t_{ff})I \tag{13}$$

where I is the identity matrix, and the v_g equation is simply

$$\frac{\mathrm{d}v_g}{\mathrm{d}t} = \frac{1}{t_{ff}} v_g - a_T \tag{14}$$

(This differential equation is technically unstable, so that errors in v_g will increase with time. But the "time constant" associated with this instability is the missile time of free flight. Since no more than one fourth of the flight time is spent in the powered mode, the magnification of any errors will not exceed 4/3 in any case.)

The general nature of the Q's for ICBM applications is illustrated in Fig. 4. They correspond to a range of 5500 n.mi. and a coordinate system for which the x,z plane is approximately directed toward the target with the x-axis elevated 20 deg above the local horizontal at the launch point. (The matrix Q is symmetrical—more about this later.) It is seen from the figure that the Q's are slowly varying functions of time suggesting that they may be adequately approximated by

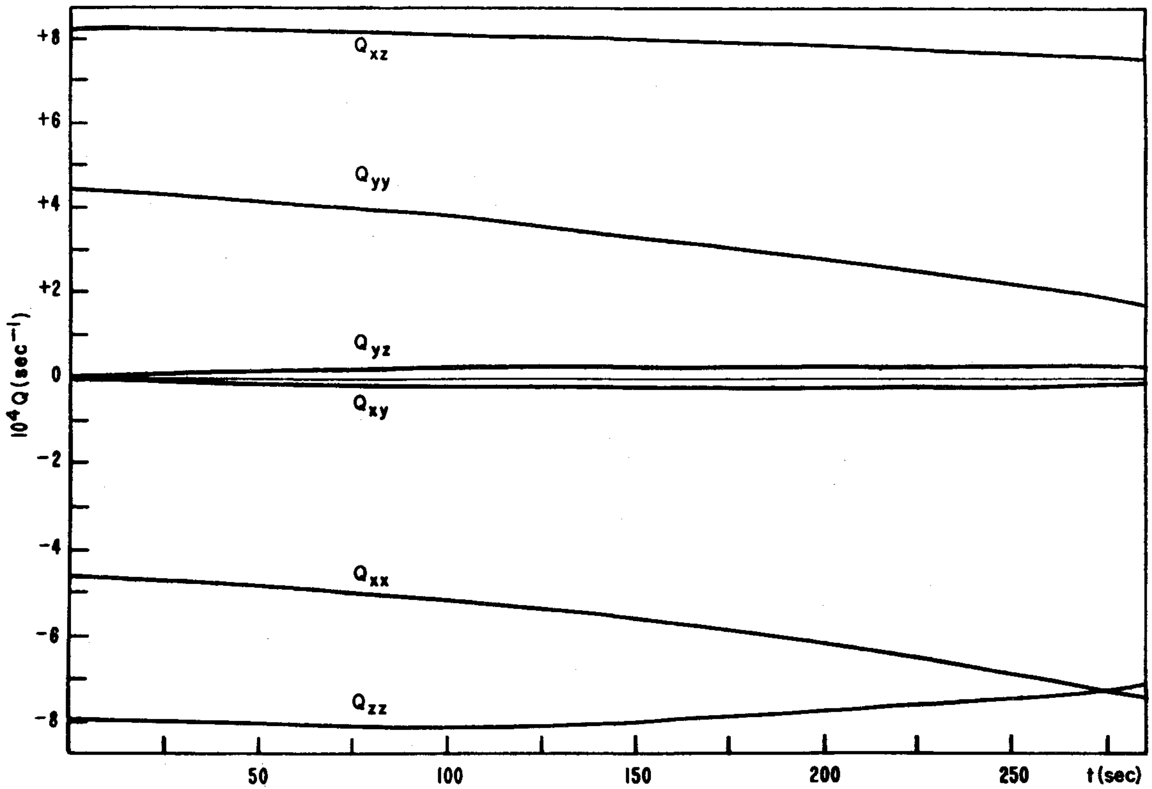

Fig. 4 Q coefficients for 5500 mile ICBM trajectory (from Ref. 6).

simple polynomials.* Indeed, for IRBM (intermediate range ballistic missile) applications, for which the range is 1500 miles or less, the Q's could be taken as constants with acceptable accuracy (less than a nautical mile).

The computation of the velocity-to-be-gained vector is only one element of the Q-guidance scheme. Of equal importance is a method to control the missile in pitch and yaw, in order that the thrust acceleration will cause all three components of v_g to vanish simultaneously.

The elegant solution to this control problem came as a brilliant burst of insight. It was all so simple and obvious! If you want to drive a vector to zero, it is sufficient to align the time rate of change of the vector with the vector itself. Therefore, the components of the vector cross product $v_g \times dv_g/dt$ could be used as the basic autopilot rate signals—a technique that became known as "cross-product steering."

With this control method, it was clear that the v_g vector would eventually vanish. However, the effect on fuel economy was not so obvious. Therefore, an optimization program was constructed utilizing the calculus of variations to study optimum fuel trajectories[7] (one of the earliest such applications made on a digital computer). The upshot was a confirmation of our suspicion that a good approximation to optimum fuel usage was, indeed, provided by cross-product steering.

Almost ten years later, Fred Martin reconsidered this problem in his MIT doctoral thesis.[8] One interesting little tidbit bears repeating here. Fred was able to show, using elementary methods only, that cross-product steering is optimum for the flat Earth hypothesis. His argument went as follows:

Form the scalar product of Eq. (14) and the vector v_g to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(v_g \cdot v_g) = \frac{\mathrm{d}}{\mathrm{d}t} v_g^2 = \frac{2}{T-t} v_g^2 - 2a_T \cdot v_g$$

where T is the time of impact at the target $(t_{ff} = T-t)$. Then integrate from present time t to the time of engine cut-off t_{co}. After integrating by parts we have

$$\int_t^{t_{co}} [2(T-t)a_T \cdot v_g - v_g^2]\mathrm{d}t = (T-t)v_g^2 \qquad (15)$$

Now, for any particular time t, the right-hand side of Eq. (15) is determined. Therefore, to minimize the integration interval $t_{co}-t$, we should maximize $a_T \cdot v_g$—i.e., align the thrust direction with the v_g vector. In the special case of a flat Earth, [check Eq. (14)] this requirement is equivalent to cross-product steering.

The vector block diagram of Fig. 5 shows the basic simplicity of an analog mechanization of the Q-system for an IRBM application. By use of the Qv_g signals as a torque feedback to the pendulous integrating gyro (PIG) units, the output of the latter can be made available as shaft rotations proportional to the components of v_g. Voltage signals can, therefore, be obtained, which are proportional to the v_g components by exciting potentiometers on the v_g shafts. These signals can, in turn, be fed into constant gains at the torquing amplifier inputs to provide the necessary multiplications and summations that constitute the matrix vector product Qv_g. The thrust acceleration sensed by the PIG's varies from approximately 50 to 200 ft/s^2, while the product Qv_g is of the general order of magnitude of 20 ft/s^2 at launch, and decreases to zero at cut-off. Thus, the Qv_g product is of the nature of a correction term, which, although far from negligible, does not have to be instrumented with the same precision as the integral of the thrust acceleration itself.

*For a single reference trajectory, the Q's may be regarded as functions only of time. However, for an actual missile trajectory with missile parameters different from nominal values, the mathematically correct Q's depend also on missile position. It is only an engineering approximation to regard the Q's as time-programmable.

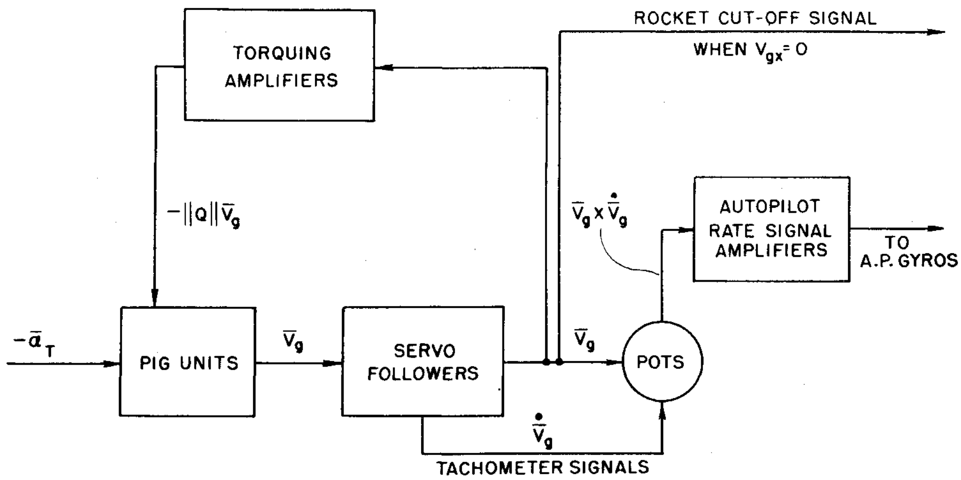

Fig. 5 Vector schematic of IRBM guidance computer (from Ref. 6).

As a result, the accuracy requirement on each component in the computation of Qv_g is about one-quarter of one percent for a one mile miss at the target—well within the range of analog technology available at the time.

Control signals for pitch and yaw are obtained simply using tachometers mounted on the follow-up servos, which produce signals proportional to the derivative of v_g. These are used as excitation for potentiometers mounted on the v_g shafts. The resulting signals are combined to give the appropriate components of the vector cross product, which are then transmitted to the autopilot as appropriate command rates in pitch and yaw.

A report[6] on the *Q*-system was presented at the first Technical Symposium on Ballistic Missiles held at the Ramo-Wooldridge Corporation in Los Angeles on June 21 and 22, 1956. In the afternoon of the second day came the only session on Inertial Guidance, and all of the papers except ours dealt with inertial instruments—the *Q*-system had no competition! We could easily have returned to Boston by walking on the clouds.

The *Q*-system was first implemented on the Thor IRBM and then on the Polaris fleet ballistic missile, but not the Atlas for which it had been designed. What system was used for Atlas? Some form of Delta guidance, I've been told.

Symmetry of the Q Matrix

In 1955, the program output from the IBM 650 was a stack of punched cards that had to be printed separately using a Type 418 accounting machine. Hal and I watched the 418 type bars rise and fall, with their characteristic noisy clank, as the first set of Q matrix elements was being tabulated in a neat array format. At the pace of 150 lines per minute, plenty of time was available for a surprising, and totally unexpected, observation. The Q matrix was symmetric! In fact, the off-diagonal elements were asymmetric only in the last decimal place.

Considering the enormous complexity of the program, the phenomenon could not be happenstance. Two conclusions were immediate: 1) the symmetry of the Q matrix *must* be analytically demonstrable, and 2) our computer program *must* be correct. It was only much later that the mathematical proof was supplied. Meanwhile, an instant check was always available on the complicated numerical procedures required to produce the Q matrix.

In an appendix of our report,[6] two different proofs of the symmetry property were given. The first utilized a special coordinate system for which it could be shown that four of the off-diagonal elements of the Q matrix were identically zero. The two remaining corner elements were then shown to be equal by a rather messy, nonintuitive argument requiring five pages of uninspiring and tedious mathematics.

The second proof provided greater physical insight, and involved a hydrodynamical analogy. Correlated velocity was to be visualized as a vector velocity field describing the motion of an inviscid, compressible fluid. The symmetry of the Q matrix was then equivalent to the velocity field having a zero curl ($\nabla \times v_c = 0$). The equation of motion

$$\mathrm{d}v_c/\mathrm{d}t = g$$

is the same as that for an inviscid fluid moving under the action of conservative body forces throughout which the internal pressure gradient is zero. Together with the equation describing the variation in fluid density ρ

$$\mathrm{d}\rho/\mathrm{d}t + \rho \nabla \cdot v_c = 0$$

it follows (with just a little exercise in ingenuity) that

$$(\nabla \times v_c)/\rho = \text{constant}$$

The demonstration concludes with an argument that the fluid is converging on the target point r_T so that the density in the vicinity of r_T is becoming infinite. Hence, the constant is zero, implying that the curl is everywhere zero.

The distribution of the *Q*-system report to those individuals having both a secret clearance and a "need to know" triggered an informal competition. Who would be first with the shortest and most elegant proof of Q matrix symmetry? (Among the contestants, the only one I recall vividly supplied a carefully detailed but erroneous demonstration of non-symmetry.)

In 1960, the original *Q*-system report[6] was reprinted in a shortened form[9] with a new appendix describing my own most recent proof. The key was to establish Q^{-1} as a solution of the matrix Ricatti equation

$$\mathrm{d}Q^{-1}/\mathrm{d}t + Q^{-1}GQ^{-1} = I \tag{16}$$

where G is the gravity gradient matrix

$$G = \partial g/\partial r \tag{17}$$

The symmetry of Q^{-1} follows at once from the differential equation and the terminal condition for Q^{-1}. The matrix G is necessarily symmetrical since g is the gradient of a scalar potential function. Hence, Eq. (16) and its transpose are identical. Also $Q^{-1} = O$ at the terminal point is symmetric. Therefore, Q^{-1} (hence also Q) is everywhere symmetric.

A by-product of this symmetry proof was an alternate computational procedure for determining Q, which is independent of the assumption that the gravity field through which the missile moves is inverse square. Hence, a more precise modeling of Earth gravity could be incorporated when computing the Q matrix.

Five years later, Fred Martin published an explicit expression for the Q matrix as an appendix to his doctoral thesis.[8] The symmetry was now obvious from inspection.

The latest bulletin on the subject is a recent recognition that no elaborate proof is really necessary! The property follows from the inherent nature of symplectic matrices.

It has been known for years that the Q matrix can be formed algebraically from partitions of the state transition matrix for the linearized equations describing a missile in free fall. It has also been known (since 1962 when Jim Potter first called attention to the fact[10]) that the transition matrix is an example of a class of so-called "symplectic" matrices. The virtue of a symplectic matrix is that the inverse is easily obtained by a simple rearrangement of its elements.

One day last year while preparing a lecture for my class, I noticed that the product of the transition matrix, and its inverse, produced a number of symmetric matrices—one of which was Q. The interested reader may wish to verify this for himself.

October 4, 1957 and the Aftermath

Like so many other Americans, the first half of October 1957 found me standing in my yard in the cold but clear early morning hours watching and waiting for the Russian Sputnik to pass overhead. I had been away from MIT for just one year exploiting new and different career opportunities in the alleged greener pastures of industry. A few months later during one of my infrequent telephone conversations with Hal Laning, I learned that he had a simulation of the solar system running on the IBM 650 and was "flying" round trips to Mars.

It didn't take long to wind up my affairs and head back to the Instrumentation Lab. My return practically coincided with the publication of a laboratory report[11] on the technical feasibility of an unmanned photographic reconnaissance flight to the planet Mars. It was asserted by the authors that a research and development program to that end could reasonably be expected to lead to the launching of such a vehicle within the next five to seven years. (It is interesting that the study and report had been sponsored by the Ballistic Missile Division of the U.S. Air Force.)

A small group was forming to flesh out the system proposal for the Mars mission. Hal and I were responsible for the trajectory determination, as well as the mathematical development of a suitable navigation and guidance technique. The project culminated a year or so later in a three volume report,[12] and a full-scale model of the spacecraft.

To my surprise, it quickly became evident that we did not really know how to compute trajectories for the simple two-body, two-point boundary value problem! How could that be possible after all the work on ballistic missile trajectories only a few years earlier? As I reviewed those equations in the Q-system report, the difficulty (but not the solution) was apparent. We had, indeed, developed expressions involving the correlated velocity vector but they were all implicit—v_c never appeared explicitly. These equations were fine for calculating the Q matrix by implicit differentiation but in no way did it seem possible to isolate the velocity vector. (Hal had been calculating round-trip Martian trajectories by "trial and error"—adjusting and readjusting the spacecraft initial conditions and determining the orbit by solving numerically the equations of motion. There had to be a better way!)

I found the clue in the classical treatise on dynamics by Whittaker[13]: "Lambert in 1761 shewed (sic) that in elliptic motion under the Newtonian law, the time occupied in describing any arc depends only on the major axis, the sum of the distances from the center of force to the initial and final points, and the length of the chord joining these points: so that if these three elements are given, the time is determinate, whatever the form of the ellipse." The proof followed, and the section ended with a neat analytical expression for time-of-flight as an explicit function of the problem geometry and the semimajor axis a of the orbit. Given the geometry and the time-of-flight, then a could be determined—not directly but by iteration.

It was the footnote that gave me pause: "It will be noticed that owing to the presence of the radicals, Lambert's theorem is not free from ambiguity of sign. The reader will be able to determine without difficulty the interpretation of sign corresponding to any given position of the initial and final points."

By no means was it obvious to me how to resolve the ambiguity or, more to the point, how to instruct a computer to choose unerringly from among the several alternatives. Whittaker's only reference was to Lagrange (Oeuvres de Lagrange, IV, p. 559) who also failed to address my concerns; but going to the original source did pay dividends. Instead of proceeding immediately to his proof of Lambert's theorem, Lagrange first chatted about the problem from different perspectives†—one of which led me to transform the problem to rectilinear motion. The ambiguity then ceased to exist.

A nontrivial problem remained—to obtain the initial velocity vector in terms of the semimajor axis a. An intense effort produced finally a delightfully elegant expression. We were now able to generate interplanetary trajectories with great aplomb. (My first trajectory program suffered from an annoying deficiency. Time-of-flight is a double-valued function of the semimajor axis a with infinite slope for the minimum energy trajectory—far from ideal for a Newton-Raphson iteration. The difficulty was resolved by a different choice of independent variable against which the time-of-flight is a monotonic function. This small, but necessary, wrinkle was first reported in an appendix to Ref. 9, and practically eliminated the audible vulgarisms that so frequently accompanied the use of the original program.)

With some trepidation, I presented this method[14] of trajectory determination in New York on January 28, 1959 at the annual meeting of the Institute of the Aeronautical Sciences. My scant background in celestial mechanics did little to inspire self-confidence in the novelty of the technique. But, as I later learned, Rollin Gillespie and Stan Ross were in the audience, and had carried a preprint back home to their associate John Breakwell at the Lockheed Missiles and Space Division. They, too, had been grappling with the trajectory problem and (according to Rollin) this was the "breakthrough" they also needed.

The method became the basis of the major orbit determination programs of the Jet Propulsion Laboratory for its series of unmanned interplanetary probes, and of the Navy and Air Force for targeting ballistic missiles. Indeed, in the early sixties, JPL used this technique to generate an enormous set of volumes—similar to the Airline Guide—in which were tabulated daily launch conditions for Venus and Mars missions extending many years into the future.

To support the Mars reconnaissance study project, we confined our attention to trajectories whose flight times were of the order of three years, and which had launch dates in the years 1962-1963. These missions, for which the space vehicle makes two circuits about the sun while the Earth makes three, seemed to provide the greatest flexibility in launch window and passing distance at Mars without placing unreasonable requirements on launch system capabilities. Later we investigated round trip missions to Venus, which could be accomplished with flight times of only a year and a quarter.

One day, when plotting a few of these Venusian reconnaissance trajectories, I was impressed by the proximity of the spacecraft orbit and the Martian orbit resulting from the

†Lagrange's paper would never appear in the *Journal of Guidance, Control, and Dynamics,* or any other modern archival publication, without strong protestations from the editor—"Needs at least a 50% reduction!"

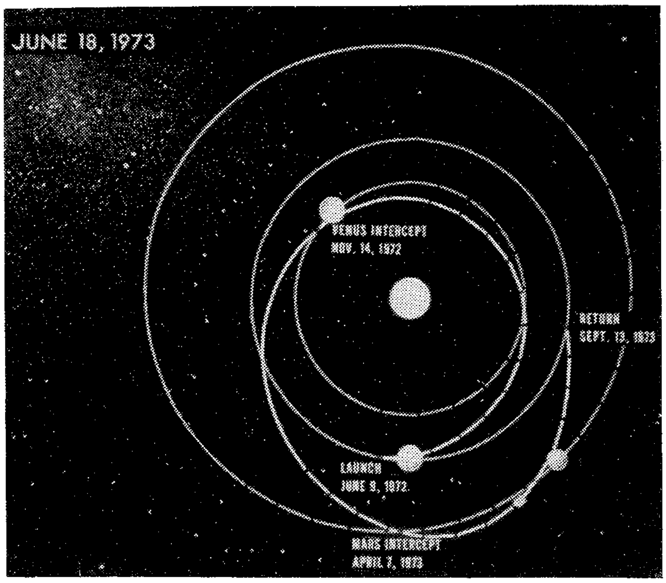

Fig. 6 Double reconnaissance trajectory (from Ref. 10).

increased velocity induced during the Venusian fly-by. The interesting possibility of a dual contact with both planets seemed feasible—a kind of celestial game of billiards. The infrequency of proper planetary configurations would, of course, severely limit the practicality of such a mission if, indeed, one existed at all.

Using trusty "cut and try" methods, I found that ideal circumstances did prevail on June 9, 1972. On that date, a vehicle in a parking orbit launched from Cape Canaveral on a 110 deg launch azimuth course could be injected into just such a trajectory at the geographical location of 5°W and 18°S, and with an injection velocity relative to the Earth of 15,000 ft/s. The first planet encountered would be Venus after 0.4308 year. The vehicle would pass 4426 miles above the surface of the planet and would, thereby, receive from the Venusian gravity field alone a velocity impulse sending it in the direction of Mars. The second leg of the journey would consume 0.3949 year and the spacecraft would then contact Mars, passing 1538 miles above the surface. The trip from Mars back to Earth would last 0.4348 year so that the vehicle would return on September 13, 1973. This truly remarkable orbit is ilustrated in Fig. 6. (At the time, the launch date seemed incredibly far off—twelve whole years! But the day finally came and, sad though it may seem, passed without fanfare or even a comment.)

Although this was the first realistic multiple fly-by mission ever designed, it was not the first ever conceived. That distinction goes to General Gaetano Arturo Crocco who was the Director of Research of the Air Ministry and a Professor of Aeronautics at the University of Rome, Italy. His paper[15] described an Earth to Mars to Venus to Earth mission of one year duration. The orbits were all coplanar; the velocity requirements were enormous; and the reversed itinerary prevented the best utilization of the gravity assist maneuvers. But it was published in 1956—one year before Sputnik. (AIAA members might appreciate knowing that General Crocco was a founding member of the Institute of the Aeronautical Sciences—one of our parent organizations.)

The Mars reconnaissance preliminary design was ready for customer review in the summer of 1959. The Air Force had been our sponsor, and it was there that we expected to turn for authorization to proceed. We were ready to go—"Mars or bust!"—with an enthusiasm that was exceeded only by our naivete. While we had been busy nailing down the myriad of technical problems one by one, the political climate was changing. A new government agency called the "National Aeronautics and Space Administration," not the Air Force, would control the destiny of the Mars probe.

With view-graphs, reports, and the wooden spacecraft model, we headed for Washington instead of Los Angeles, and arrived there on the same day as Chairman Khrushchev. Although our presentation was well received, the high level NASA audience we had expected (including Hugh Dryden, the Deputy Administrator) was attending to the necessary protocol mandated by the Russian visit. We were sent home with a pat on the head and the promise of some future study money. As our dreams of instant glory in interplanetary space began to fade, we secretly took perverted pleasure in having Nikita Khrushchev himself as a ready-made scapegoat. The Russians were formidable opponents indeed!

The NASA study contract allowed our small team to continue the work that had begun under Air Force auspices. For this we were most grateful, but the absence of a specific goal diminished much of the enthusiasm. Now we were simply doing "interplanetary navigation system studies." There certainly was no reason to expect that a new goal lay just over the horizon, which would challenge and excite us beyond our wildest imaginations.

Prelude to Apollo

The general method of navigation that Hal and I had created for the Mars probe mission[16] was based on perturbation theory, so that only deviations in position and velocity from a reference path were used. Data was to be gathered by an optical angle measuring device, and processed by a spacecraft digital computer. Periodic small changes in the spacecraft velocity were to be implemented by a propulsion system as directed by the computer.

The appropriate velocity changes were calculated from a pair of matrices obtained as solutions of the differential equations

$$\mathrm{d}\boldsymbol{R}^*/\mathrm{d}t = \boldsymbol{V}^* \qquad \mathrm{d}\boldsymbol{V}^*/\mathrm{d}t = \boldsymbol{G}\boldsymbol{R}^* \tag{18}$$

where $\boldsymbol{G}$ is the gravity gradient matrix evaluated along the reference path. Boundary conditions were specified at the reference time of arrival t_A at the target as

$$\boldsymbol{R}^*(t_A) = \boldsymbol{O} \qquad \boldsymbol{V}^*(t_A) = \boldsymbol{I} \tag{19}$$

Then if $\delta\boldsymbol{r}(t)$ is the position deviation from the reference path at time t, the required velocity deviation $\delta\boldsymbol{v}(t)$ was found to be

$$\delta\boldsymbol{v}(t) = \boldsymbol{V}^*(t)\boldsymbol{R}^{*-1}(t)\delta\boldsymbol{r}(t) \tag{20}$$

It is a trifle embarrassing to admit that we did not immediately recognize our old friend the Q matrix in Eq. (20). When the dawn came, we were truly nonplussed. Here we were now working in an unclassified area with every intent to freely publish the results—but the Q-system was still classified! That last point had been dramatically emphasized only a year or so earlier. An author, who shall remain nameless, wrote a book on guidance containing a section that made full disclosure of the Q-system. When the U.S. Navy was finally ready to act, the books were on the publisher's loading dock awaiting shipment. All copies—several thousand at least—were seized and burned.

Then and there the matrix product $\boldsymbol{V}^*\boldsymbol{R}^{*-1}$ was christened $\boldsymbol{C}^*$. We reasoned that the Q matrix by itself was just a mathematical collection of partial derivatives. The security classification derived from its use in the velocity-to-be-gained differential equation as applied to ballistic missile guidance. Since the letter "Q" signified absolutely nothing, it would have been pointless to persist in its use in an entirely different context.

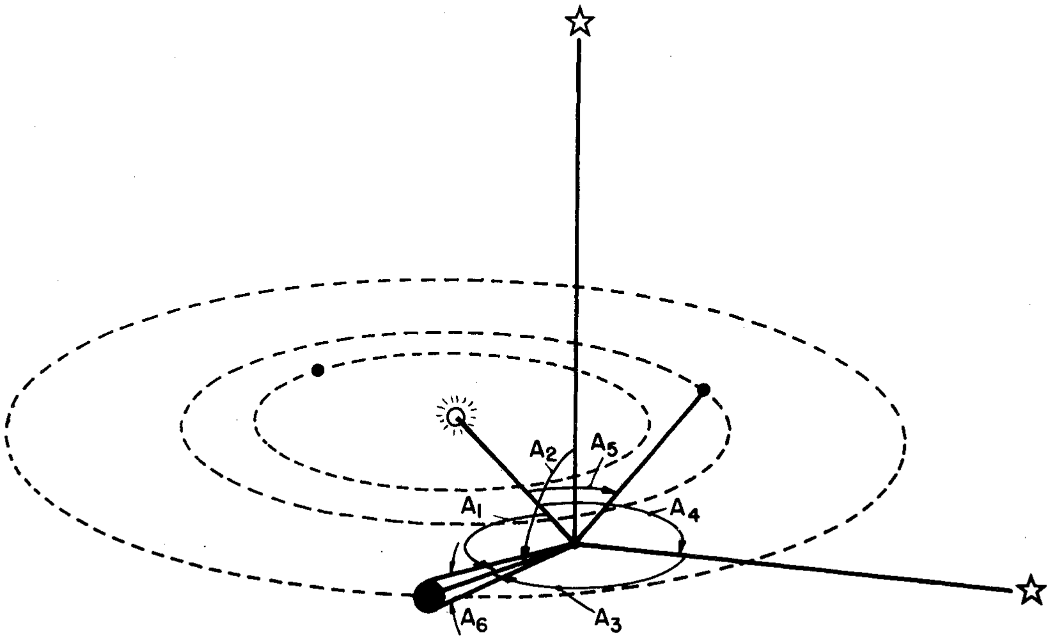

Fig. 7 Mars probe navigation fix (from Ref. 16).

The velocity correction, as calculated from Eq. (20), was perfectly adequate for interplanetary missions, except when the spacecraft was in proximity to the destination planet. With a relatively short time of flight remaining, the constraint imposed on the vehicle by the C^* matrix to reach the target at a predetermined time caused an inordinate expenditure of rocket fuel.

The deficiency was later corrected during our NASA studies with the invention of variable time of arrival guidance.[17] Equation (20) could now be replaced by

$$\delta v(t) = C^*(t)\delta r(t) + R^{*\,T-1}(t)\,v_r(t_A)\,\delta t \qquad (21)$$

where $v_r(t_A)$ is the velocity of the spacecraft relative to the target planet at the nominal time of arrival t_A, and δt is the change in arrival time. The increment δt is chosen to minimize the magnitude of the required velocity correction.

To navigate the Mars probe, a sequence of measurements of angles between selected pairs of celestial bodies, together with the measurement of the angular diameter of a nearby planet, was to be made on board the spacecraft (automatically, of course, under computer control) to obtain a celestial fix. For specificity, the measured angles, illustrated in Fig. 7, were chosen as follows: 1) from the sun to the nearest visible planet P; 2) from Alpha Centauri to P; 3) from that one of Sirius or Arcturus to P such that the plane of measurement is most nearly orthogonal to the plane of the angle measured in 2; 4) from the sun to the same star selected in 3; 5) from the sun to the second closest planet provided that more than one planet is "visible" (at least 15 deg away from the line of sight to the sun); and 6) the angular diameter of P, provided that it exceeds 1 mrad. This strategy for observations ensured that a minimum of four angles would be measured, provided at least one planet was visible. (The three particular stars were selected because they are among the brightest and form roughly an orthogonal triad.) The intended result of these observations was a determination of the coordinates of spaceship position together with a correction to the spaceship clock.

Although the terminology was not yet in vogue, we were in fact dealing with an estimation problem involving a four dimensional "state vector." We linearized the measurements about a reference point, and developed a weighted least-squares procedure to obtain the celestial fix.[16]

So much for the Mars probe, which was now, at best, in a state of limbo. We began working for NASA, and the close technical collaboration that had existed between Hal and myself gradually subsided.

Hal Laning renewed his old love affair with the digital computer—however, it was basic computer architecture, and not software, that attracted his interest this time. He joined forces with Ramon Alonzo to develop a design for a small control computer with some unique characteristics for space and airborne applications.[18] Some of those features were variable speed with power consumption proportional to speed, relatively few transistors, parallel word transfer, automatic incrementing of counters and automatic interruption of normal computer processes upon receipt of inputs. The program and constants were stored in a wired-in form of memory called a "core rope," which permitted unusually high bit densities for that time. Such a computer would have been ideally suited for the Mars probe but, in fact, Hal and Ray were unknowingly designing the computer whose technical offspring would take man to the moon.

Meanwhile, I continued alone on the guidance and navigation analysis. There were some annoying problems with our interplanetary navigation algorithm having to do with numerical difficulties encountered in the required matrix inversion associated with the method of least squares.

In the notation used at that time,[19] the least-squares method resulted in the following expression:

$$\delta\tilde{r} = (U_{4m}\Phi^{-1}_{mm}U_{m4})^{-1}\,U_{4m}\Phi^{-1}_{mm}\,\delta\tilde{A}_m \qquad \text{(B-10)‡}$$

Here m is the total number of measurements for any particular fix; U_{m4} is the $m \times 4$ measurement geometry matrix; Φ_{mm} is the diagonal moment matrix of measurement errors; $\delta\tilde{A}_m$ is the vector of measurements; $\delta\tilde{r}_4$ is the least-squares estimate of the four-dimensional state vector. Four measurements are sufficient to determine the spaceship position and the clock correction, so that $r = m - 4$ is the number of redundant measurements.

‡Equations with the prefix B are taken from, and numbered in accordance with, Ref. 19.

To cope with the numerical difficulties of Eq. (B-10), I was determined to obtain, if at all possible, an explicit inverse of the matrix product $(U_{4m}\Phi_{mm}^{-1}U_{m4})^{-1}$. The task became an obsession and the derivation was so involved that only the final expression appeared[19] in an unclassified appendix to an otherwise classified§ report. The result was recorded as follows:

"Now if we define the two square matrices

$$P_{44} = U_{44}^{-1}\Phi_{44}U_{44}^{T-1} \quad \text{(B-12)}$$

$$Q_{rr} = \Phi_{rr} + U_{r4}P_{44}U_{4r} \quad \text{(B-13)}$$

it can be verified directly that

$$(U_{4m}\Phi_{mm}^{-1}U_{m4})^{-1} = P_{44} - P_{44}U_{4r}Q_{rr}^{-1}U_{r4}P_{44} \quad \text{(B-14)}$$

Then by substituting (B-14) into Eq. (B-10), we obtain, after a little manipulation,

$$\delta\tilde{r}_4 = U_{44}^{-1}\{\,\|I_{44}\ \ O_{4r}\| + B_{4r}Q_{rr}\| - A_{r4}\ \ I_{rr}\|\,\}\delta\tilde{A}_m \quad \text{(B-15)}$$

where we have defined the two rectangular matrices

$$A_{r4} = U_{r4}U_{44}^{-1} \quad \text{(B-16)}$$

$$B_{4r} = \Phi_{44}A_{4r} \quad \text{(B-17)}$$

The matrix Q_{rr} may be expressed in terms of A_{r4} and B_{4r} by

$$Q_{rr} = \Phi_{rr} + A_{r4}B_{4r} \quad \text{(B-18)}$$

Equation (B-15) displays explicitly the effect of adding redundant measurements."

Those familiar with the Kalman filter will recognize Eq. (B-14) at once as the covariance matrix update formula. Although the expression (B-15) for the state vector update is not in the customary form, it is evident that the first term on the right is the state estimate using four measurements. The second term may be rewritten as

$$-U_{44}^{-1}B_{4r}Q_{rr}^{-1}A_{r4}\delta\tilde{A}_4 + U_{44}^{-1}B_{4r}Q_{rr}^{-1}\delta\tilde{A}_r$$

with $\delta\tilde{A}_4$ and $\delta\tilde{A}_r$ denoting the two partitions of the vector of measurements $\delta\tilde{A}_m$. Then, substituting from Eqs. (B-12), (B-16), and (B-17), and introducing the matrix

$$W_{4r} = P_{44}U_{4r}Q_{rr}^{-1}$$

called the "weighting matrix" in the current vernacular, the term in question becomes

$$W_{4r}(\delta\tilde{A}_r - U_{r4}U_{44}^{-1}\delta\tilde{A}_4)$$

Since this is precisely the weighted difference between the actual redundant measurements and the predicted values of those measurements, Eq. (B-15) is then exactly equivalent to the now conventional state vector update formula.

Unbeknownst to me at the time, Rudolf Kalman was also addressing the estimation problem, albeit with greater generality and from a more esoteric standpoint. His now classical paper[20] was published almost simultaneously with Ref. 19. About a year later, I learned of Kalman's work from Stan Schmidt at the Ames Research Center.

§The report was classified because it quoted some confidential Centaur missile data.

The Race to the Moon

After the publication of our studies for NASA in the three volume report R-273,[19] a hiatus in further navigation work resulted from an unexpected invitation by the Research and Advanced Development Division of Avco Corporation. They enlisted the support of the Instrumentation Laboratory to design a system for guiding a vehicle propelled electrically by a low-thrust arc jet engine from an Earth-satellite orbit into a lunar orbit. We were so eager to work on a real space program that we fairly leaped at this opportunity. A bright and amiable young engineer, Mike Yarymovych (who will assume the presidency of the AIAA in May), was our principal contact at Avco.

At that time a great deal of work had already been accomplished in optimizing low-thrust escape trajectories utilizing variational techniques, but virtually no attention had been directed to the guidance problem. We succeeded in designing a multiphased guidance scheme—one aspect of which relied heavily on the concepts of velocity-to-be-gained and steering developed for the Q-system. Preliminary results were first published[21] in January of 1961, and Jim Miller carried through the complete development as his doctoral dissertation,[22] which he presented on August 29, 1961 in Los Angeles at the Sixth Symposium on Ballistic Missile and Aerospace Technology.

Meanwhile, in February of that same year, NASA came through with another six-month contract—this time for a preliminary design study of a guidance and navigation system for Apollo to be sponsored by the Space Task Group of NASA. It was time to dust off and re-examine the navigation problem once again.

I learned from Gerald Smith that the Ames Research Center's Dynamic Analysis Branch was working on midcourse navigation and guidance for a circumlunar mission.[23] The Branch Chief, Stanley Schmidt, and his associates were most hospitable during my visit, and gave me a private blackboard lecture describing the filter equations taken from Kalman's year old paper[20] which they were using for their own navigation studies. I also received a copy of Kalman's paper, along with the admonition that it would not be easy reading.

The key idea gleaned from the meeting at Ames was the possibility of eliminating the notion of the navigation fix. I learned that the covariance matrix could be easily extrapolated using the state transition matrix. Navigation measurements could be spaced in time and the update equations could be applied recursively to a full six-dimensional state vector. Indeed, if only one scalar measurement was processed at any one time, the matrix Q_{rr} of Eq. (B-13) would be simply a positive scalar with no matrix inversion required at all!

Kalman's paper was to me so abstruse that it was not clear whether his equations were equivalent to those obtained using the maximum likelihood method. (Indeed, Stan told me during our visit that this question had not yet been resolved to everyone's complete satisfaction.) To settle this in my own mind, I wrote down a linear state vector update equation to process a single measurement and left the weighting vector to be determined so as to minimize the variance of the estimation error. The result agreed with Kalman's.

As a second check, I applied the equations (with the state transition matrix replaced by the identity matrix) to one of the Mars mission position fixes, and processed the measurements one at a time. Again the result was the same.

The recursive navigation algorithm was clearly the best formulation for an on-board computer. But a number of questions still remained. When a single measurement is to be made, which star and planet combination provides the "best" available observation? Does the best observation give a sufficient reduction in the predicted target error to warrant its being made at all? Is the uncertainty in the computer velocity correction a small enough percentage of the correction itself to justify an engine restart and propellant expenditure? Can a statistical simulation of a space flight mission be made without resorting to Monte Carlo techniques? How would cross-correlation effects of random measurement errors affect the estimator? These questions were all addressed in a paper[24] presented in October of 1961, on Friday the thirteenth, at the American Rocket Society's Space Flight Report to the Nation held in the New York Coliseum.

During the preparation of that paper, political events provided a new urgency to our work. On May 25, 1961, President John F. Kennedy in his Special Message to Congress on Urgent National Needs said "I believe that this nation should commit itself to achieve the goal, before this decade is out, of landing a man on the moon and returning him safely to Earth." Less than three months later on August 10, NASA contracted with our laboratory for the development of the Apollo guidance and navigation system—the first major Apollo contract awarded by the space agency.

The history of the Apollo on-board guidance, navigation, and control system was well told[25] by Dave Hoag at the International Space Hall of Fame Dedication Conference in Alamogordo, N.M., during October 1976. With that as background, a description of the Apollo system and its development will not be necessary here. However, a few items require emphasis to provide a proper perspective for the rest of this narrative.

Initially, the specifications were for a completely self-contained system—there would be absolutely no ground communications either verbally or by telemetry with the vehicle. This requirement, which was presumably to prevent an overenthusiastic competitor in the race to the moon from intentionally interfering with an Apollo flight, gradually eroded away—but not before computer algorithms had been designed and implemented which would permit completely autonomous missions.

Several fundamental characteristics of the Apollo guidance computer (AGC) made the implementation of self-contained algorithms a definite challenge: 1) a short word length, 16 bits, necessitating double precision for most calculations; 2) a modest memory size—36,864 read-only, and 2048 read-write registers; and 3) moderate speed—23.4 μs add time. Small as that may seem, it was a major improvement in speed and capacity over that which was available in the fall of 1961. Then the AGC had 4096 words of fixed memory, 256 words of erasable memory, and twice the cycle time. (Over the years, technology advances permitted the expansion in capacity while maintaining the original size of one cubic foot. The physical dimensions could change only at great cost—that was all the space provided for in the spacecraft.)

The first mission programming for the AGC was to implement the recursive navigation algorithm. That, at least, we knew how to do! Of course, the program changed many times during the ensuing years but not the concept. A complete description, including all the nitty-gritty, of the final implementation for Apollo is found in Ref. 26. A diagram which we used countless times for customer briefings is reproduced as Fig. 8. Note from the figure that the reference trajectory has been replaced by the integrated vehicle state. The necessity for this important change was obvious when we first addressed the implementation problem. The modification is generally referred to as the "extended" Kalman filter.

Guiding the Apollo vehicle during its many and varied powered maneuvers was another matter. The idea of using the original Q-system for these purposes was soon rejected. Its principal virtue was the ease of mechanization on board the vehicle. But this advantage had to be traded off against the burden placed on ground facilities. (Consider the significant staff and computers of the Dahlgren Naval Weapons Laboratory devoted solely to the task of supplying the necessary targeting data, and the curve-fitted elements of the Q-matrix for the fleet ballistic missiles of the U.S. Navy.) With the AGC we had at our disposal for the first time ever a powerful general purpose digital computer as the key ingredient of a vehicle-borne guidance system. Why not use it?

In fact, the Q matrix could be avoided altogether in the $\boldsymbol{v}_g$ differential equation by simply differentiating the defining equation for velocity-to-be-gained $\boldsymbol{v}_g = \boldsymbol{v}_r - \boldsymbol{v}$ to obtain

$$\frac{d\boldsymbol{v}_g}{dt} = \frac{d\boldsymbol{v}_r}{dt} - \boldsymbol{g} - \boldsymbol{a}_T \tag{22}$$

(The terminology "correlated velocity" $\boldsymbol{v}_c$ was replaced by "required impulsive velocity" $\boldsymbol{v}_r$; missile velocity $\boldsymbol{v}_m$ became vehicle velocity $\boldsymbol{v}$.) If the vector $\boldsymbol{v}_r$ could be expressed in analytical form, then the vector

$$\boldsymbol{p} = d\boldsymbol{v}_r/dt - \boldsymbol{g} \tag{23}$$

could be calculated so that the rate of change of $\boldsymbol{v}_g$ is determined from

$$d\boldsymbol{v}_g/dt = \boldsymbol{p} - \boldsymbol{a}_T \tag{24}$$

(We were no longer concerned about computing gravity—it posed no problem for the AGC.) We had an expression for $\boldsymbol{v}_r$ when the target vector, and the time of flight, are specified. It remained to be seen how many of the major orbital transfer maneuvers could be accomplished conceptually by a single impulsive velocity change, and if simple formulas could be obtained for the corresponding required velocities.

One by one, we accumulated suitable required velocity expressions for a variety of possible Apollo maneuvers. For example, when the Apollo command module returned to Earth, it had to impact the atmosphere at a specified flight path angle—otherwise it might either skip out of the atmosphere or be destroyed by overheating. A simple formula for $\boldsymbol{v}_r$ was obtained (see problem 3.11 in Ref. 10).

Braking into a circular lunar orbit was another mission requirement. We used for $\boldsymbol{v}_r$ the velocity the vehicle should have in order to be in a circular orbit at its present position, and in a specified plane. In this manner, we were able to control the shape and orientation of the final orbit but not its radius. However, it turned out that an empirical relation could be found between the final radius and the pericenter of the approach trajectory so that the desired radius could be established by an appropriate selection of the approach orbit. (This technique was based on an idea developed during our low-thrust guidance work for Avco.)

On the first unmanned guided Apollo flight in August 1966, the required velocity vector was defined so as to achieve an orbit of specified eccentricity and semimajor axis. The list goes on, but does not include, for example, the lunar landing since this maneuver cannot be performed conceptually with a single impulsive burn.

We experimented with a variety of guidance laws to drive the $\boldsymbol{v}_g$ vector to zero: 1) align the thrust acceleration vector $\boldsymbol{a}_T$ with the $\boldsymbol{v}_g$ direction; 2) direct $\boldsymbol{a}_T$ to cause $\boldsymbol{v}_g$ to be aligned with its derivative—cross-product steering; and 3) a combination of both as illustrated in Fig. 9. The scalar mixing parameter γ was chosen empirically to maximize fuel economy. A constant value of γ was usually sufficient for a particular mission phase; however, if required, γ would be allowed to vary as a function of some convenient system variable. Fred Martin found that this third method gave a highly efficient steering law that compared favorably with calculus of variations optimum solutions.[27]

A functional diagram illustrating the computation of the error signal required for control purposes is shown in Fig. 10. Numerical differentiation of the required velocity was simpler than programming the analytically obtained derivative. Near the end of the maneuver, when $\boldsymbol{v}_g$ is small, cross-product steering is terminated, the vehicle holds a constant attitude,

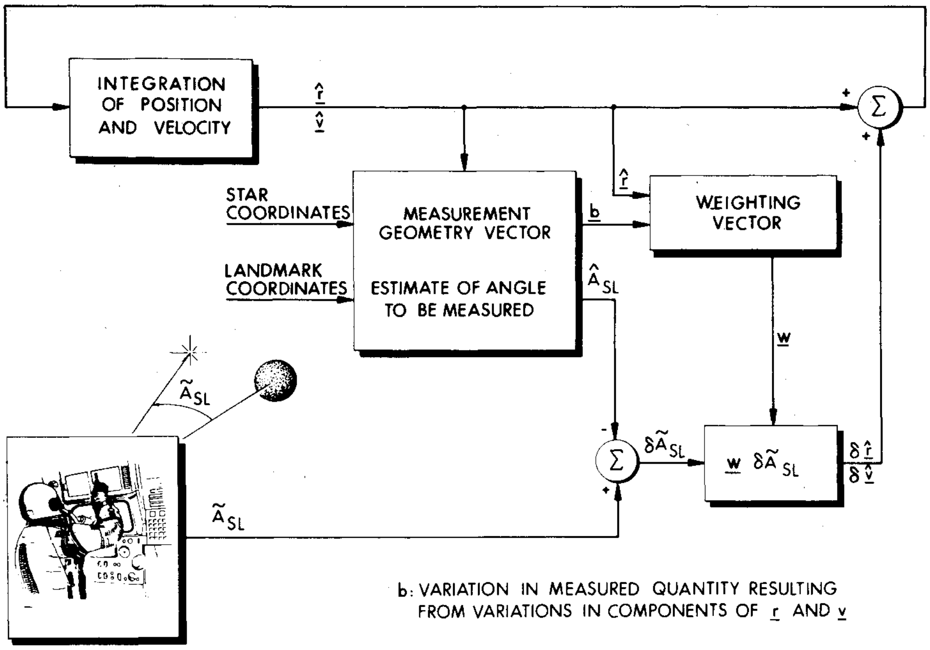

Fig. 8 Apollo coasting flight navigation.

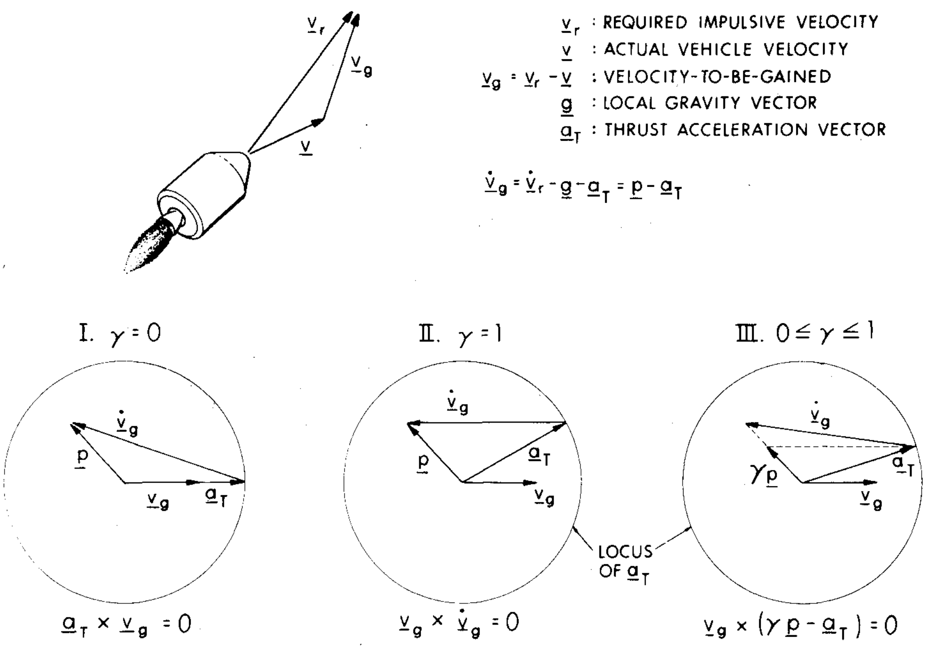

Fig. 9 Velocity-to-be-gained guidance laws (from Ref. 28).

and engine cut-off is made on the basis of the magnitude of the v_g vector. A detailed description of just how this guidance scheme was mechanized in the AGC is provided in Ref. 28.

Steering to intercept a given target at a specified time came to be known as Lambert guidance after Johann Heinrich Lambert, the famous eighteenth century Alsatian scholar who discovered the theorem that bears his name. Since v_r had to be calculated cyclically in real time, Lambert guidance (which also required an iterative solution of Lambert's time equation) placed one of the heaviest burdens on the AGC. The task of completing all the necessary calculations in the time available became a programmer's nightmare. Ever since, the problem has fascinated me, and I am always on the lookout for new and better solutions of Lambert's equation.[29-31]

As the years went by, more and more of the guidance and navigation responsibilities were transferred from the on-board system to the Real Time Control Center (RTCC) in Houston. Much of the capability remained and was used in the AGC as a backup, but the RTCC was primary. Targeting calculations were made in the ground-based computers and Apollo performed most of its maneuvers in the so-called "external Δv" mode.

The Circle Closes

The intense pressure under which we all worked began to ease somewhat after the first landing on the moon. Timothy Brand, who played an active role in mechanizing the powered flight AGC algorithms, had time now to reflect on Lambert guidance performance. Was it possible to avoid the frequent solutions of Lambert's problem, which were necessary to maintain an accurate value of the v_g vector? What could be done about the small yet persistent error in cutoff when estimating the time-to-go until thrust termination? How can a more nearly constant attitude maneuver be attained that would avoid any relatively large turning rates?

Perhaps all of these difficulties had to do with the definition of the v_g vector itself. What if we defined a single coasting trajectory, which coincides with the powered

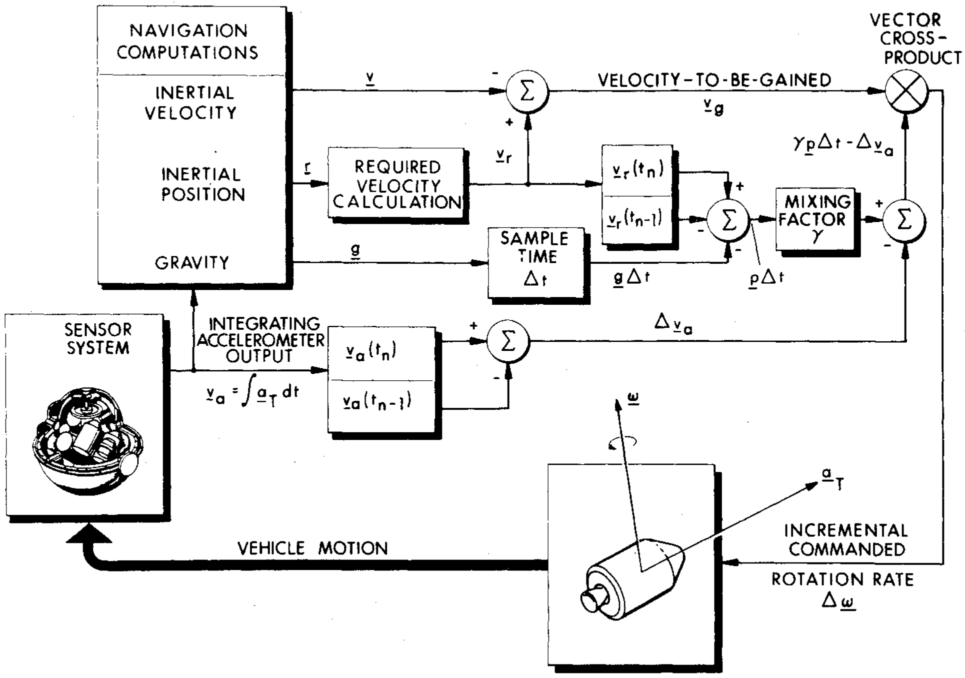

Fig. 10 Apollo cross-product steering (from Ref. 28).

trajectory at thrust termination, and used it as the basis of the velocity-to-be-gained computation? We would have then $v_g = v_r - v$, but now v_r is the velocity along the single coasting path. If r' is the corresponding position vector on the coasting trajectory then

$$dv_r/dt = g(r') \tag{25}$$

would describe the rate of change of v_r. The v_g differential equation would be

$$dv_g/dt = g(r') - g(r) - a_T = \Delta g - a_T \tag{26}$$

with Δg replacing the term $-Qv_g$ in the older version.

The advantages of the new formulation became evident. An easy calculation showed that the contribution of the term Δg is generally much smaller than that of Qv_g. Furthermore, Δg approaches zero at a rate proportional to v_g^2, while the Qv_g term, on the other hand, vanishes like v_g to the first power. Simulations verified that Δg is so small for short maneuvers that a nearly constant attitude can be obtained by merely steering the vehicle so as to align the thrust vector along v_g. Velocity-to-be-gained, under these circumstances is particularly easy to compute—the accelerometer-sensed velocity change is subtracted from the previous value of v_g on each computer guidance cycle.

Tim's technique[32] works well, even for long duration maneuvers, if we periodically create a new coasting flight trajectory. A suitable approximation for $\Delta r = r' - r$ is found to be

$$\Delta r = -(v_g/2a_T)v_g \tag{27}$$

which, when added to current vehicle position, produces the position vector r'. Knowing r' and the target r_T, together with the time of flight, permits a new Lambert solution —hence a new v_r and a new coasting trajectory. Subtracting the current vehicle velocity provides an updated value of v_g with which to begin anew.

If none of these ideas seem familiar, you have forgotten the Convair legacy. What has just been described is essentially what the Convair engineers were advocating those many years ago. I must confess that I did not make the connection between Tim's new technique and the old Convair proposal until I began rummaging through my memorabilia in preparation for this paper. Obviously, Tim knew nothing of this—he was only about ten years old at the time.

Of course, the Tim Brand or the Convair scheme would have been impractical for an on-board implementation to guide the early ballistic missiles. It was feasible only after the small airborne digital computer replaced all those servos, amplifiers, potentiometers, and other analog devices of the good old days.

References

[1] Laning, J.H., Jr. and Zierler, N., "A Program for Translation of Mathematical Equations for Whirlwind I," Engineering Memorandum E-364, MIT Instrumentation Laboratory, Cambridge, Mass., Jan. 1954.

[2] Knuth, D.E. and Pardo, L.T., "The Early Development of Programming Languages," STAN-CS-76-562, Stanford Univ., Stanford, Calif., Aug. 1976, pp. 55-60.

[3] Laning, J.H., Jr. and Miller, J.S., "The MAC Algebraic Language System" Report R-681, MIT Charles Stark Draper Laboratory, Cambridge, Mass., Nov. 1970.

[4] Battin, R.H., O'Keefe, R.J., and Petrick, M.B., "The MIT Instrumentation Laboratory Automatic Coding 650 Program," *IBM Applied Science Division Technical Newsletter*, No. 10, Oct. 1955, pp. 63-79.

[5] Battin, R.H., "Programming Manual for BALITAC 650 Routine," Report R-126, MIT Instrumentation Laboratory, Cambridge, Mass., Aug. 1956.

[6] Laning, J. H., Jr. and Battin, R.H., "Theoretical Principle for a Class of Inertial Guidance Computers for Ballistic Missiles," Report R-125, MIT Instrumentation Laboratory, Cambridge, Mass., June 1956.

[7] Laning, J.H., Jr. and Battin, R.H., "Optimum Fuel Trajectories for a Q-Type Guidance System," Engineering Memorandum E-520, MIT Instrumentation Laboratory, Cambridge, Mass., Feb. 1956.

[8] Martin, F.H., "Closed-Loop Near-Optimum Steering for a Class of Space Missions," Doctor of Science Thesis, Massachusetts Institute of Technology, Cambridge, Mass., May 1965.

[9] Laning, J.H., Jr. and Battin, R.H., "Theoretical Principle for a Class of Inertial Guidance Computers for Ballistic Missiles," Engineering Memorandum E-988, MIT Instrumentation Laboratory, Cambridge, Mass., 1960.

[10] Battin, R.H., *Astronautical Guidance*, McGraw-Hill, New York, 1964.

[11] Laning, J.H. Jr., Frey, E.J., and Trageser, M.B., "Preliminary Considerations on the Instrumentation of a Photographic Reconnaissance of Mars," Report R-174, MIT Instrumentation Laboratory, Cambridge, Mass., April 1959.

[12] "A Recoverable Interplanetary Space Probe," Report R-235, MIT Instrumentation Laboratory, Cambridge, Mass., July 1959.

[13] Whittaker, E.T., *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, Cambridge, England, 1937.

[14] Battin, R.H., "The Determination of Round-Trip Planetary Reconnaissance Trajectories," *Journal of the Aerospace Sciences*, Vol. 26, Sept. 1959, pp. 545-567.

[15] Crocco, G.A., "One Year Exploration Trip Earth-Mars-Venus-Earth," *Proceedings of the Seventh International Astronautical Congress*, Rome, 1956, pp. 227-252.

[16] Battin, R.H. and Laning, J.H., Jr., "A Navigation Theory for Round Trip Reconnaissance Missions to Venus and Mars," *Planetary and Space Science*, Vol. 7, Pergamon Press, London, 1961, pp. 40-56. (Originally published as MIT Instrumentation Laboratory Report R-240, Aug. 1959.)

[17] Battin, R.H., "A Comparison of Fixed and Variable Time of Arrival Navigation for Interplanetary Flight," *Proceedings of the Fifth AFBMD/STL Aerospace Symposium on Ballistic Missile and Space Technology*, Academic Press, New York, 1960, pp. 3-31. (Originally published as MIT Instrumentation Laboratory Report R-283, May 1960.)

[18] Alonso, R. and Laning, J.H., Jr., "Design Principles for a General Control Computer," Report R-276, MIT Instrumentation Laboratory, Cambridge, Mass., April 1960.

[21] Battin, R.H. and Miller, J.S., "Trajectory and Guidance Theory for a Lunar Reconnaissance Vehicle," Report E-993, MIT Instrumentation Laboratory, Cambridge, Mass., Jan. 1961.

[22] Miller, J.S., "Trajectory and Guidance Theory for a Low-Thrust Lunar Reconnaissance Vehicle," Report T-292, MIT Instrumentation Laboratory, Cambridge, Mass., Aug. 1961.

[23] Schmidt, S.F., "The Kalman Filter: Its Recognition and Development for Aerospace Applications," *Journal of Guidance and Control*, Vol. 4, Jan.-Feb. 1981, pp. 4-7.

[24] Battin, R.H., "A Statistical Optimizing Navigation Procedure for Space Flight," *ARS Journal*, Vol. 32, Nov. 1962. pp. 1681-1692.

[25] Hoag, D.G., "The History of Apollo On-Board Guidance, Navigation and Control," *The Eagle Has Returned*, Science and Technology, Vol. 43, American Astronautical Society Publication, 1976.

[26] Battin, R.H. and Levine, G.M., "Application of Kalman Filtering Techniques to the Apollo Program," *Theory and Applications of Kalman Filtering*, edited by C.T. Leondes, AGARDograph 139, 1970, Chap. 14.

[27] Martin, F.H., "Closed-Loop Near-Optimum Steering for a Class of Space Missions," *AIAA Journal*, Vol. 4, Nov. 1966, pp. 1920-1927.

[28] Martin, F.H. and Battin, R.H., "Computer-Controlled Steering of the Apollo Spacecraft," *Journal of Spacecraft and Rockets*, Vol. 5, April 1968, pp. 400-407.

[29] Battin, R.H., "A New Solution for Lambert's Problem," *Proceedings of the XIX International Astronautical Congress*, International Academy of Astronautics, Vol. 2, Astrodynamics and Astrionics, Pergamon Press, New York, 1970, pp. 131-150.

[30] Battin, R.H., "Lambert's Problem Revisited," *AIAA Journal*, Vol. 15, May 1977, pp. 707-713.

[31] Battin, R.H., Fill, T.S., and Shepperd, S.W., "A New Transformation Invariant in the Orbital Boundary-Value Problem," *Journal of Guidance and Control*, Vol. 1, Jan.-Feb. 1978, pp. 50-55.

[32] Brand, T.J., "A New Approach to Lambert Guidance," Report R-694, MIT Charles Stark Draper Laboratory, Cambridge, Mass., June 1971.