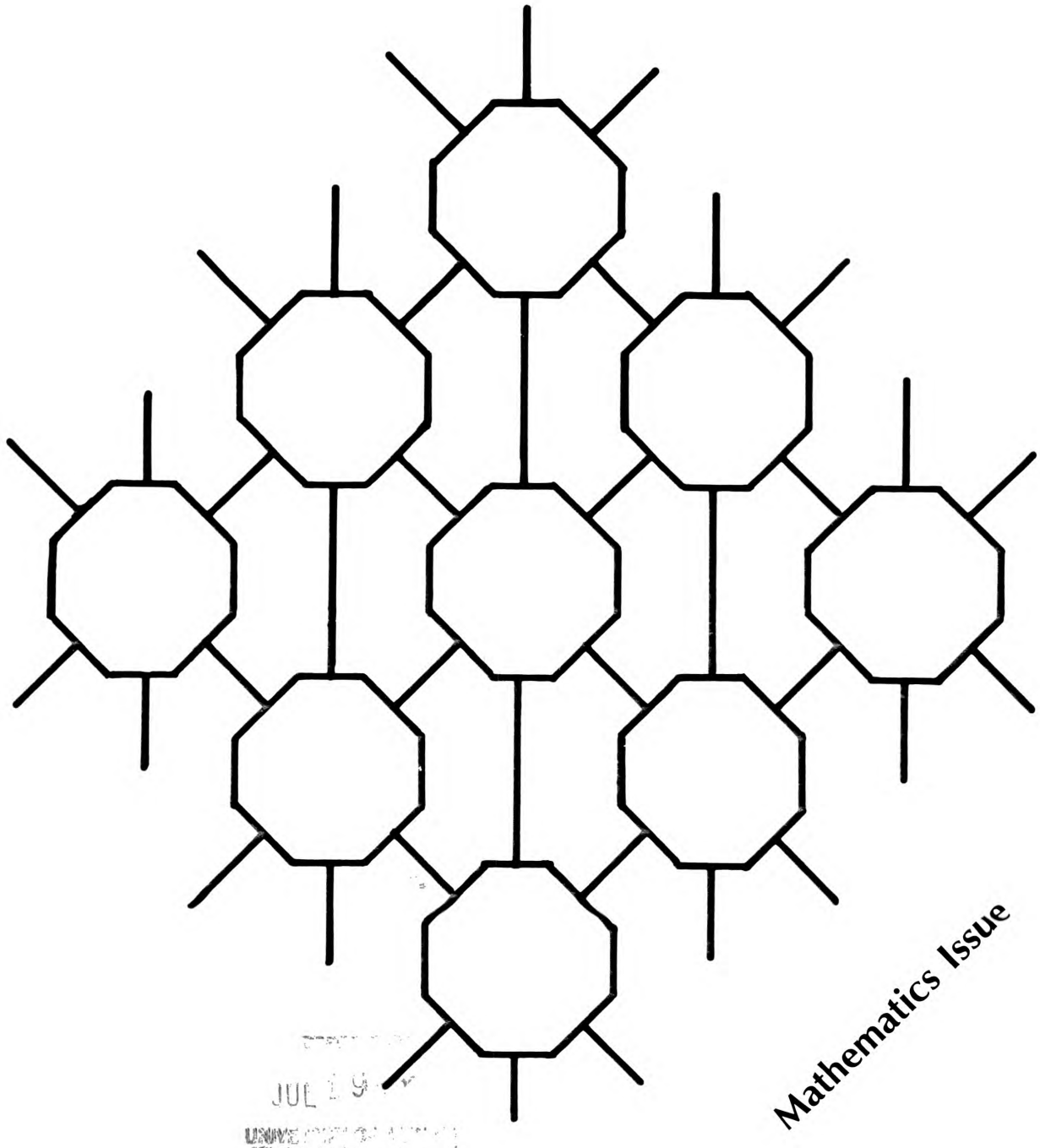
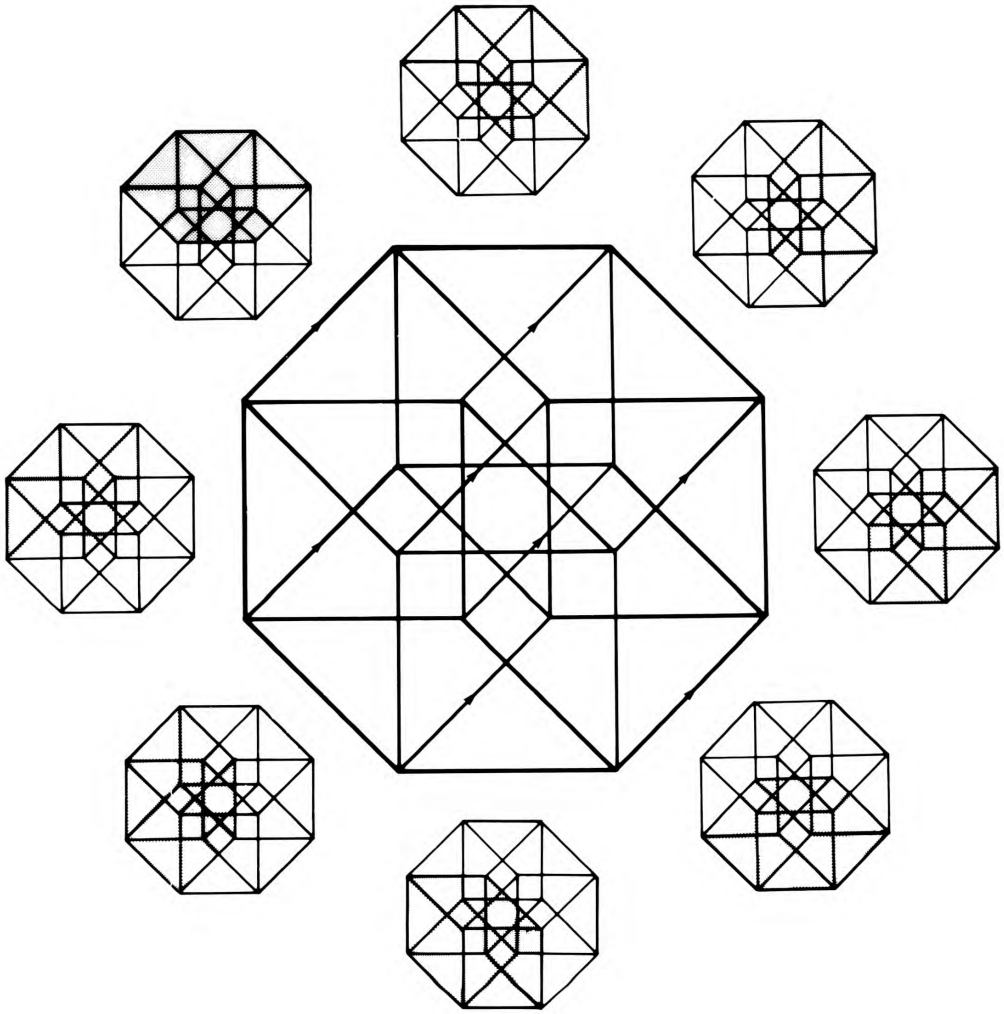
Naval Research Reviews

Office of Naval Research
Two/1983
Vol XXXV



Mathematics Issue



The 4th Dimension

The 4th and higher dimensions have been well understood by mathematicians from an analytic point of view. Few investigators think of a 4th spatial dimension as real in the same sense that we think of our 3 dimensional world. One exception to this point of view is Professor Thomas Banchoff of Brown University. Banchoff has made his life-work the study of the topological and geometric properties of 4 and higher dimensional shapes and surfaces. He has used computer graphics to generate 3-dimensional slices of higher dimensional forms. By studying these 3-d projections, Banchoff gains insight into the higher dimensional shape and has actually discovered a large number of new topological and geometric properties by seeing them.

Although apparently unrelated to statistical methods, Banchoff's methods, in fact, are closely related. With the advent of computerized instrumentation, the collection of multi-attribute (multi-dimensional) data has become commonplace. The analysis of multi-attribute data is difficult not only because it is difficult to use graphical methods for data that are more than 3-dimensional, but also because the standard generalizations of 1-dimensional formulae often prove to be ineffective as analysis tools in higher dimensions. Banchoff's current project for ONR involves close work with statisticians to develop his methods to gain insight into high-dimensional statistical data sets. The ORION-1 system discussed elsewhere in this issue is an example of a computer graphics tool designed to gain this insight.

One particularly stunning example of Banchoff's work is a film entitled, "The Hypercube: Projection and Slices," which has won prizes at two international scientific film festivals. In this film, Banchoff and his colleague, Charles Strauss, take the viewer through a series of mental steps leading to the 4th dimension. Think of a line, they say—a one-dimensional object with two ends of vertices. Then move the line along an axis perpendicular to it to form a square (two dimensions, four vertices). Move the square in a third perpendicular direction to form a cube (three dimensions, eight vertices). Then think of a moving the cube in a fourth dimension perpendicular to the other three to form a hypercube, a four-dimension perpendicular to the other eight cubical faces. The information you get depends on the angle you view it from. As it rotates, the inner cube appears to move up through the top of the object, flatten, and turn inside out.

Naval Research Reviews

Office of Naval Research
Two/1983
Vol XXXV

ARTICLES

DEPARTMENTS

About Our Cover

A very large scale integrated (VLSI) circuit technology will require a very high density per chip. Because it will be impossible to design each circuit element individually for such high density chips, one design strategy is to replicate one relatively simple processor in a space-filling array. The mathematical challenge is to exploit this array of relatively simple processor to carry out a complex task quickly. The array on the front cover is a prototypical hex-connected array which may be used to do numerical linear algebra. *(See page 9).*

Naval Research Reviews publishes articles about research conducted by the laboratories and contractors of the Office of Naval Research and describes important naval expermental activities. Manuscripts submitted for publication, correspondence concerning propective articles, and changes ot address, should be directed to Code 732, Office of Navay Research, Arlington, Va 22217. Requests for subscriptions should be directed to the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402. *Naval Research Reviews* is published from appropriated funds by authority of the Office of Naval Research in accordance with Navy Publications and Printing Regulation NAVSO P-35.

Introductory Note

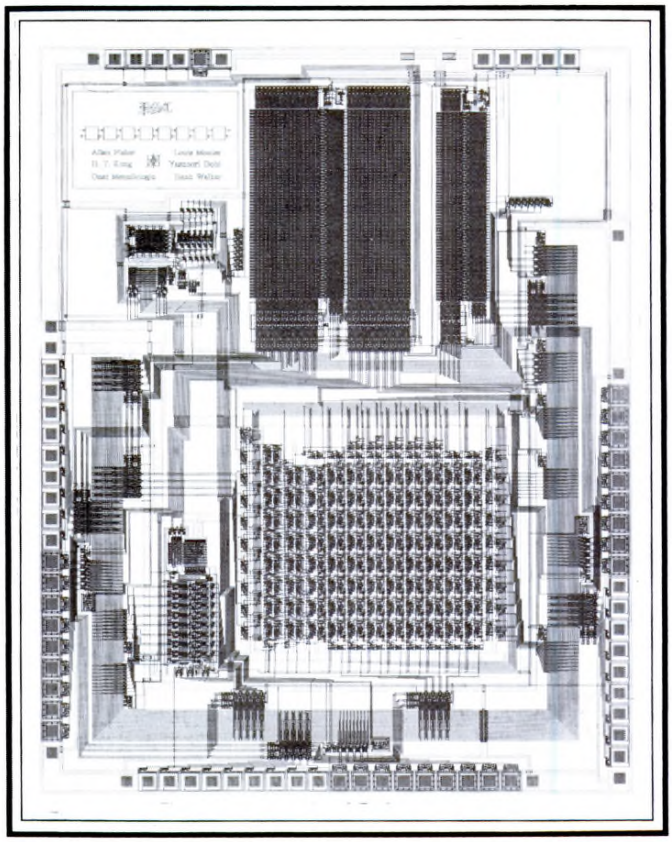
Since its creation in 1946, the Office of Naval Research (ONR) has recognized the importance of investment in research in the mathematical sciences (mathematics, statistics, computer science, operations research). This recognition has enabled ONR to be a pioneering force in both the development and applications of new mathematical concepts. Many of the advances in computers and computing, signal processing, scheduling and logistics can be traced directly to research by eminent scientists supported by ONR over the last 35 years. Research advances in other fields, ranging from management science to radar to laser physics, depend partially upon improved analysis capabilities, techniques, and algorithms developed under ONR support in the mathematical sciences.

The continuing computer revolution promises to increase the already large role that the mathematical sciences play in the modern Navy and in our technological society. First, increased computational speed and reduced costs permit mathematics to be utilized, via computation, to an ever widening class of applications. Second, increased computational capabililty has given scientists the opportunity to address more complex problems using more sophisticated mathematical science techniques. Thus, the need for mathematical science research has grown in parallel with our increased computational capability.

Given this ubiquitious nature of the mathematical sciences, it is clearly impossible to treat prospects for mathematical science research fully in this single issue. Some of the more complex and difficult areas of mathematical research were not appropriate to this issue. We have chosen, therefore, four articles which illustrate some of the new and challenging areas and opportunities which are faced by mathematicians today in carrying out ONR's mission of providing technological superiority to the U.S. Navy.

Credit for organizing this issue of *Naval Research Reviews* devoted to mathematical sciences sould be given to Dr. J.C.T. Pool, former Division Director of the Mathematical Sciences Division.

Charles J. Holland
Office of Naval Research



VLSI chip designed at Carnegie—Mellon University with Office of Naval Research funding. This chip is an example of the special purpose processor designed to implement software in silicon.

An Introduction to Systolic Algorithms and Architectures

by

Jon Louis Bentley
H.T. Kung
Carnegie-Mellon University

Introduction

The design tradeoffs inherent in good engineering often imply that there is no single best design for solving a given family of tasks. Consider, for instance, the task of designing a general-purpose vehicle. An F-14 augmented with Phoenix missles is an excellent vehicle for the task of establishing air superiority, and a ten-speed bicycle augmented with a backpack is an excellent vechicle for the task of going to a neighborhood grocery to purchase a carton of milk. It is hard to imagine a single general-purpose vehicle that would do as well at both these tasks as each special-purpose vehicle does individually.

The vast majority of computer users, however, do in fact use devices labelled "general-purpose" computers. Futhermore, in most task domains, these general-purpose devices have been remarkably successful. Most computer users do well to work on a general-purpose computing device: it offers advantages of ease of use and generality that usually far outweigh the disadvantages of increased cost and decreased performance (as compared to special-purpose machines). In some applications, though, the cost and performance requirements are so demanding that no general-purpose machine can meet them, so the computer engineer has no choice but to turn to a special-purpose computer to solve his problem.

In this paper we will survey the class of *systolic* special-purpose computer architectures and algorithms, which are particularly well-suited for implementation in very large scale integrated circuitry (VLSI). The principles of systolic algorithms were first enunciated by H.T. Kung and Charles Leiserson [1] in work supported by the Office of Naval Reserach (ONR) at Carnegie-Mellon University in 1978. Since that time they have been studied extensively by Kung's research group at Carnegie-Mellon (the basic research has been supported by ONR; the development activity surrounding the design of prototype chips have been funded by the Defense Advanced Research Projects Agency (DARPA) and by hundreds of other researchers throughout the world (the several dozen papers cited at the end of this article represent only a small fraction of the technical literature dealing with various aspects of systolic arrays).

This paper is a brief introduction to systolic arrays for a reader with a broad technical background and some experience in using a computer, but who is not neccessarily a computer scientist. Kung [2] provides more technical introduction to the subject. In the next section we will briefly survey the technological advances in VLSI that led to the development of systolic algorithms and architectures; this section establishes the ground rules underlying the rest of the paper. An overview of systolic architectures is given in the section entitled "Systolic Arrays as Special-Purpose VLSI Devices." Then in the section, "An Example," we will study in detail one special-purpose systolic architecture for efficiently solving a particular problem. In the section, "An Overview of Systolic Designs," we will turn to an overview of systolic algorithms. The last section discusses future directions.

The term systolic refers to the regular flow of data through the machine, which is similar to the rhythmic flow of blood through the body caused by the systolic contractions of the heart; this definition will be claborated later in the paper.

VLSI: The Opportunity and the Problem

The integrated circuitry that revolutionized electronics in the 1960's in now referred to historically as SSI: small scale integration. It was followed by MSI (medium-scale), LSI (large-scale) and VLSI (very-large scale integration). The history and the future of the miniaturization underlying the string of names has been graphically described in an analogy due to Charles Seitz (quoted here from "About the Cover", *Lambda* Magazine, First Quarter, 1980):

> In the mid-1960's the complexity of a chip was comparable to the street network of a small town. Most people can navigate such a network by memory without difficulty. Today's microprocessor, using a five-micron technology, is comparable to the entire San Francisco Bay Area. By the time a one-micron technology is solidly in place, designing a chip will be comparable to planning a street network covering all of California and Nevada at urban densities. The ultimate one-quarter micron technology will likely be capable of producing chips with the intricacy of an urban grid covering the entire North American continent.

In absolute terms, the largest current chips contain almost a million transistors, yet their cost and energy consumption is not considerably greater than their counterparts of less than two decades ago.

The technology alone is only part of the good news of VLSI; its inherent potential is likely to come within the grasp of a new generation of circuit designers. At the birth of integrated circuitry, design in the medium was a skill limited to a few experts in this important (but narrow) art. This was changed dramatically by the efforts of Carver Mead and Lynn Conway [3], who reduced the practice of VLSI design to a small set of engineering principles that is now routinely taught in one-semester university courses. After such a course, senior computer science undergraduates (with little, if any, background in electrical engineering) are typically able to design a small circuit in VLSI technology.

The promise of VLSI is not without its challenges, though. From a viewpoint of the technology alone, the medium presents several new technical challenges. Primary among those is a high "communication cost". Previous generations of circuit designers have had the view that computation is expensive while communicating the results of the computation is almost free. In VLSI, though, the communication costs are high in execution time, power dissipation, and chip area [4]. Another technological observation is that the switching speed of VLSI circuits is not that much greater than the SSI speed of the 1960's. The most important challenge of VLSI, though, goes beyond the technology itself to its use: there is an incredibly high cost in designing VLSI circuitry. It is not atypical for some chips to have a design cost as high as one million dollars and a design time of several years.

To summarize the above, we should keep in mind the following two facts.

- **The Potential.** VLSI currently offers a huge number of transistors on an integrated circuit; that number promises to increase substantially in the next decade. The large number of transistors offers increased functionality in digital designs, increased memory, and a potential increase in computing speed if the logical devices are run in parallel.
- **The Challenge.** To exploit the promise of VLSI, a design must minimize communication and design costs. Communication costs are minimized by regular structures that communicate over short wires; design costs are minimized by using regular structures rather than designing new components from scratch.

Systolic Arrays as Special-Purpose VLSI Devices

In this section we will consider various approaches to exploiting the opportunity of VLSI while respecting its challenges, and then focus on the single approach of systolic arrays. The following two approaches represent two ends of the VLSI spectrum.

- **Further Down Old Paths.** We can try to use VLSI to make more efficient implementations of the traditional von Neumann computer architectures. This approach has been quite successful for systems implemented in non-VLSI technology, where communication is not a serious issue. It has become apparent in the past several years, though, that this approach will never be able to exploit fully the potential of VLSI due to the "von Neumann bottleneck": no matter how powerful we make a processor or how large and fast our memory, they are still doomed

to communicate through a very small, low-bandwidth (relative) channel of the memory access mechanism. Specifically, if we have a powerful processor connected to a hundred million words of memory, at every particular clock cycle only a few of the words in memory can be communicated to the processor; the rest of the data will be unused.

- **Radically New General-Purpose Architectures.** Several ways around the von Nuemann bottleneck have been proposed recently by various researchers, as surveyed by Treleaven [5]. Although these methods do appear promising for the distant future, most of them will entail a fundamental change in our view of computer hardware and software. It is not clear that any of these methods will have a substantial impact on the practice of software engineering in this century.

Note that these approaches suffer from different flaws: the former is too concervative, while the latter is too risky.

Throughout the rest of this paper, we will focus on an approach that is midway between the previous two: we will keep the standard von Neumann architecture for most of the system, but we offload the computation-intensive tasks onto special-purpose devices that involve a different computational structure. For instance, if we find that a certain set of matrix computations is in the time bottleneck of our system, we can offload those operations onto a special matrix processor, while performing the rest of the system on the original general-purpose machine.

There are many approaches to building special-purpose computer architectures; the systolic approach is characterized by Kung [2] as follows.

> In a systolic system, data flows from the computer memory in a rhythmic fashion, passing though many processing elements before it returns to memory, much as blood circulates to and from the heart. The system works like an automobile assembly line where different people work on the same car at different times and many cars are assembled simultaneously. An assembly line is always linear, however, and systolic systems are sometimes two-dimensional. They can be rectangular, triangular, or hexagonal to make use of high degrees of parallelism. Moreover, to implement a variety of computations, data flow in a systolic system may be at multiple speeds in multiple directions — both inputs and (partial) results flow, whereas only results flow in classical pipelined systems. Generally speaking, a systolic system is easy to implement because of its regularity and easy to reconfigure (to meet various outside constraints) because of its modularity.

At this point we can explain the name systolic arrays: the regular flow of the data through the processor is similar to the rhythmic flow of blood through the body, generated by the systolic contractions of the heart.

To make these abstract notions more concrete, we will consider an algorithm from everyday life that illustrates many of the features of systolic arrays: how do we ensure that each of the 200 guests at a wedding shakes hands with each of the 15 members of the wedding party? A simple approach has a single Master of Ceremonies orchestrate each of the $200 \times 15 = 3000$ handshakes, and each is performed at a separate time. This approach suffers from two fatal flaws. First, the job of the Master of Ceremonies involves handling massive amounts of data and is hopelessly confusing. Second, even the rapid rate of ten seconds per handshake would yield a total time for the affair of over eight hours. The typical (systolic) organization for this task is less difficult to manage and more efficient in time; by arranging a receiving line of the 15 members of the wedding party, and then having the 200 guests file past the party, all hands can be shaken in just a little under forty minutes.

The pipelining of the receiving line is obviously very efficient, and provides a handy analogy for one kind of systolic algorithm. The 15 members of the wedding party correspond to 15 processors arranged in a linear array, and the 200 guests can be viewed as data, passing regularly from the input of the receiving line to its output (the first and last members of the line). The analogy weakens as we consider the computation performed in this task; the goal of each guest is to collect the 15 handshakes, and each member of the party performs the function of providing one of the handshakes. This analogy should become clearer as we consider in the next section a problem whose systolic solution uses the same structure as the receiving line.

An Example

In this section we will study the following problem, which arises in a number of pattern recognition applications.

The Nearest Neighbors Problem. Given a set of N training points and a set of M sample points, for each sample point determine its nearest neighbor among the training points. We assume that each point is described by k coordinates (that is, the points are in k-dimensional space) and that the distance between points is the standard Euclidean distance.

For definiteness, we will assume that we are solving the particular problem that $N = 1000$, $M = 20000$, and $k = 3$ (that is, we are searching for the nearest neighbors of 20000 points among a set of 1000 points in Euclidean 3-space). This problem can be solved easily for any particular input; for each of the M sample points, we compute its distance to all N training points, and return the point with the minimum distance. The only drawback of this method is its excessive run time. Assuming that distances can be calculated in just ten microseconds (an optimistic assumption for many of today's high-performance general-purpose computers), the $M \times N$ distances computed require over three minutes of computer time.

We will now investigate a systolic algorithm that can solve the nearest neighbors problem in less than a quarter of a second. The algorithm has the same structure as the receiving line we studied earlier. In this case, though, the receiving line contains the 1000 training points. As each of the 20000 sample points files past the training points, it computes the distance to the particular point, and "remembers" the minimum distance seen so far and the point realizing that distance. The resulting systolic array is shown in Figure 1; each box in that figure is a small processor capable of storing two points, computing the distance between the points, performing distance comparison, and communicating with its two neighbors.

The systolic algorithm proceeds in two phases: the first sets up the "receiving line", and the second passes the "guests" through the line. In the first phase, each of the 1000 training points is loaded into a separate processor in the systolic array. The second phase is responsible for computing nearest neighbors; it does this by writing the 20000 sample points into the input of the systolic array and then reading them from the output. During this process, each processor is controlled by a global clock, and performs the following actions at each clock cycle:

- Read from the left neighbor a sample point, the name of its nearest neighbor seen so far, and the distance to the nearest neighbor.
- Compute the distance from the sample point to the training point contained in this cell.
- Compare the distance just computed with the distance to the nearest neighbor seen so far (the minimum distance is then used by the next step).
- Make available to the cell to the right the following information: the sample point, the minimum of the closest distance so far and the distance to this training point, and the name of the point realizing the minimum distance. This information will be read by the cell to the right at the next clock cycle.

Notice that the rightmost cell in the systolic array outputs the sample points, their distances to their nearest neighbors, and the names of the nearest neighbors. We give input data to the array at each clock cycle in packets containing a sample point and a fictitious nearest neighbor so far that is "infinitely" far away (that is, we give it the greatest possible distance).

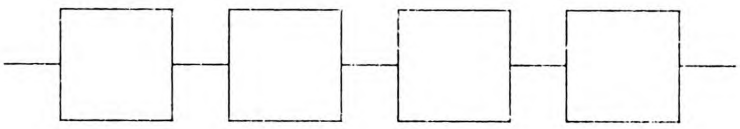


Figure 1. A linear array for the nearest neighbors problem.

The total time required by this method is 1000 units to load the array and 20000 units to pass the entire set of sample points through the array, for a total of 21000 time units. If we asume, as before, a ten-microsecond clock, then the entire problem can be solved in about 0.2 seconds (a factor of 1000 faster than the straightforward solution). In general, computing the nearest neighbors of M sample points among N training points can be accomplished in $M+N$ time units. (Notice that use of N processors yields a speedup factor of N over the straightforward single-processor algorithm.) If the performance of a solution to the nearest neighbors problem is in the time bottleneck of a system, then this approach could clearly have a substantial impact. This systolic approach is particularly well-suited to implementation in VLSI; it uses only local communication, and it utilizes only one basic cell that can be designed once and then copied many times. Note, though, that the cost of designing and implementing a systolic solution to this problem will be significantly greater than merely writing a (short) program to solve the task on a general-purpose computer: this is a tradeoff of design time against processing time.

There are a number of variations possible on the basic theme of using a linear array to solve the nearest neighbors problem. In the previous solution, the training points remained in the same processor while the sample points marched past them, carrying the answers along with them. We could have organized the system so that the M sample points each stayed in a processor while the N training points moved past them; this would still required approximately $M+N$ time units. There are two possible motivations for this alternative approach:

- If there are fewer sample points than training points (that is, if $M<N$), then this requires fewer processors.
- Because the answers remain in the same processor as they are computed, this approach might involve less data movement.

Kung [2] investigates a problem from a signal-processing application with a similar structure, and shows how a half-dozen linear-array systolic algorithms are each appropriate under various conditions.

We will now briefly consider two problems that can lead to other variations.

- *What happens if there is an overflow?* Suppose we have a systolic device of 1000 processors, but we have to solve a problem with $M=20000$ and $N=2000$. We can solve this problem by decomposing the training point into two sets of 1000 points each (so that they will fit in the array), and then passing the 20000 sample points through the array twice (once for each of the set). This solution will not be as efficient as the case when all points fit in the array, but the array costs only half as much.
- *What happens if points are not atomic elements?* Throughout this discussion we have assumed that points in Euclidean k-space were given as basic units; in fact, on almost all computers (systolic and general-purpose alike), such items will have to be implemented in terms of other primitives. Figure (2) shows one possible solution for the nearest neighbor problem if we must deal with one-dimensional scalar values as primitives. This particular array is specialized to hold four points in three-space, named a, b, c and d; the coordinates of each point are stored in the columns of the array. The sample points then flow from left to right, slanted at a 45° angle. The cumulative sum of the squares of the differences of the coordinates flows from the top to the bottom of each column; when it reaches the circular nodes at the bottom it is compared to the minimal distance seen so far, and the minimum of the two (along with the point realizing that distance) flows to the right. Finally, the output of the nearest neighbor to each of the sample point flows out the rightmost circle. This discussion is only a sketch of the problem; filling in the details of the data flow is a fine exercise for the industrious reader.

Figure 3 shows a systolic array for the nearest neighbors problem based on the non-linear structure of a binary tree. The training points are stored in the square nodes; the sample points flow in the topmost circular node and are broadcast by the circles to the squares, the distance are computed in the squares, and the minimum of the distances is computed by the triangles and returned out the bottom triangle. This structure yields a solution that has the same throughput for the problem as the linear array, but has much faster response time. The details of this scheme are given by Bentley and Kung [6].

This simple example illustrates many principles

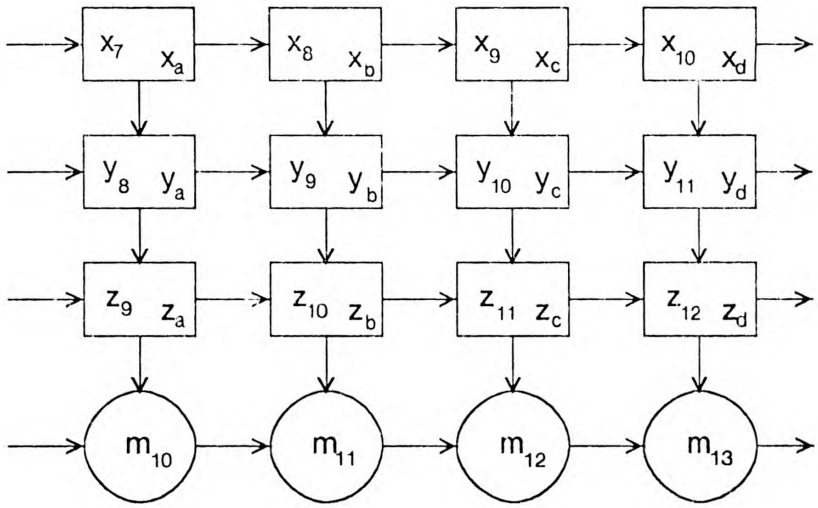


Figure 2. A Two-dimensional array for the nearest neighbors problem.

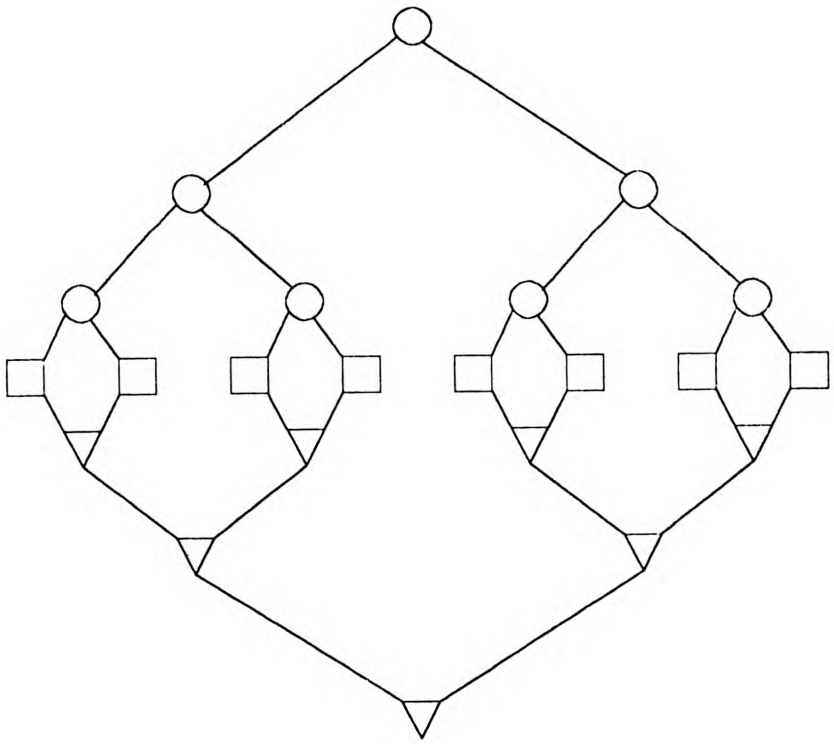


Figure 3. A tree-structured machine for the nearest neighbors problem.

of systolic algorithms for VLSI implementation. The choice of problem is crucial in any system; we solved the particular problem of nearest neighbor searching, but our work is justified only if that problem is of great importance in some application. Another approach would be to build a special-purpose processor to solve the more general problem of searching; Bentley and Kung [6] describe one approach to this problem. Our experience in this section has touched on many aspects of data flow in systolic systems: we saw two distinct ways in which the data can flow in simple linear arrays. We also saw three distinct geometric structures for solving the problem: a one-dimensional linear array, a two-dimensional array, and a tree. In the next section we will see that these structures are just a few scattered points in much larger design space.

An Overview of Systolic Designs

In the previous section we studied one systolic algorithm in detail; in this section we shall take a broader (but shallower) view and survey the field of systolic designs. In the following subsection: "Design Techniques," we will study a set of techniques used in the design of systolic systems; "A Catalog of Results," we will survey some of the theoretical results those techniques have achieved; and "Reduction to Practice," we will describe how the theoretical ideas can be, and have been, implemented in current technology.

Design Techniques

The example above showed that there are many possible systolic solutions to a given problem, and many variants of a basic solution. During the early development of systolic algorithms, each new problem seemingly required a fundamentally different solution. As time passed, though, and a larger set of problems and solutions was assembled, it became clear that there is a fundamental set of design techniques underlying most systolic algorithms. In this section we will briefly survey some of the more important of those techniques.

One of the most obvious properties of a systolic design is its basic geometric structure. In the previous section we saw several possible geometries; the following list categorizes those and several other popular layouts for systolic algorithms.

- *Linear Array.* Figure 1 shows a simple structure ideal for solving problems with an inherently linear structure.
- *Rectangular Array.* Figure 2 is an instance of a rectangular structure suitable for certain matrix applications and for database applications in which each record contains several fields. Variations of the basic rectangular array include traingular and trapezoidal arrays.
- *Tree Structures.* Figure 3 shows one variant of a tree structure—several others are suitable for various problems.
- *Hexagonal Array.* Figure 4 from Reference 1 is a structure well-suited to matrix operations such as multiplication of band matrices.

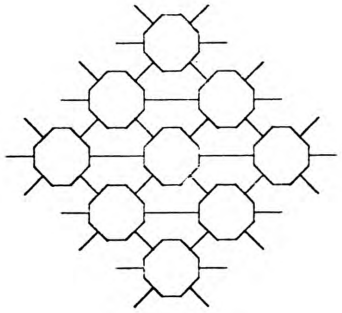


Figure 4. A hexagonal array.

Once the geometric structure is chosen, the designer must still decide how data will flow through it. In the previous section we saw how one kind of data can stay stationary while the other kind flows past it. There are many more complicated kinds of data flow; the following set of flows just scrapes the surface of data flow for the simple structure of a linear array.

- One kind of data flows left while the other flows right; the answers stay in the same place. Only half the processors are used on any clock cycle; the unused half can be used to solve an entirely distinct problem.
- One kind of data flows left, and the other kind stay in the same place; the answers flow left at twice the speed.
- There is only one kind of data, but it flows left or right as a function of the data contained in its neighbors.

Kung [2] shows that there are at least eight different data flows through a linear array which solve a single problem. Data flow through a more complex geometric structure can often be even more subtle.

Two theoretical fields are developing the potential to aid the designer of systolic systems. The field of lower bounds seeks to show that a certain set of resources is necessary to solve a problem systolically. Thompson [7] developed an appropriate mathematical model of VLSI and derived fundamental results using general techniques; since then the techniques have been extended and applied to many other problems (see, for instance, Brent and Kung [8]). Hong and Kung [9] have described a general model of special-purpose computing devices that allows one to derive bounds on the efficiency of off-loading computation. A second theoretical field is developing a set of transformations that allow systolic designers to assume that they have certain freedom not actually present in most VLSI technology; transformations are then used to convert the resulting designs into systolic designs appropriate for the technology at hand (Leiserson and Sax [10] describe such transformations).

A Catalog of Results

In previous sections we have seen one problem in detail and several different systolic arrarys for solving it under various constraints; we have also surveyed a number of design techniques for building systolic systems. This section catalogs a number of other problems that have been solved by systolic systems. The purpose of this catalog is two-fold: to provide references for readers interested in exploring a systolic solution to a particular computational problem they might face, and to show the breadth of the approach.

Most compute-bound problems are inherently regular; they are definable in terms of simple recurrences or inner loops, such as the nearest neighbors problem considered earlier. By exploiting that regularity, systolic algorithms for these problems can usually be found without too much difficulty. Systolic designs are now available for the following problems.

- Signal and image processing:
  - FIR and IIR filtering and 1-D convolution [11,12,13] 2-D convolution and correlation [14,15,16,17,18];
  - discrete Fourier transform [12,13];
  - interpolation [15];
  - 1-D and 2-D median filtering [19]; and
  - geometric warping [15].
- Matrix arithmetic:
  - matrix-vector multiplication [1];
  - matrix-matrix multiplication [1,20];
  - matrix triangularization (solution of linear systems, matrix inversion) [21,1] QR-decomposition (eigenvalue, least squares computations) [22,21]; and
  - solution of triangular linear systems [1].
- Non-numeric applications:
  - data structures — stack and queue [23], searching [6,24,25], priority queue [26], and sorting [26,25];
  - graph algorithms — transitive closure [27], minimum spanning trees [28], and connected com ponents [29];
  - geometric algorithms—convex hull generation [30];
  - language recognition—string matching [31] and regular expres sion [32];
  - dynamic programming [28];
  - polynomial algorithms —polynomial multiplication and division [33], and polynomial greatest common divisors [34];
  - relational data-base operations [35,36]; and
  - Monte Carlo Simulation [37].

Reduction to Practice

The design of systolic algorithms, on which we have been concentrating so far in this paper, constitutes just one phase of the total effort of constructing a systolic device for use in a computational system. Following this phase are specification of the system architecture, system design, and integration to the host system. For well-understood, highly important inner loops a systolic processor could be cost-effectively built for each single application. For example, a large quanity of low-cost, high-performance sonar bouys with some filtering and

pattern matching capabilities could conceivably be built based on the systolic approach. For applications where more flexibility is desired, however, multi-purpose systolic arrays could be more attractive. An example is the ESL systolic processor, a linearly connected array of 28 cells which can be microcoded to perform 2-D convolution, matrix multiplication, and other similar operations [13,17]. Figure 5 dipicts the ESL systolic processor, which has been running since January 1982, and can be called as a subroutine by FORTRAN programs running on a host VAX-11/780.

The highest degree of flexibility can be obtained by allowing programmability in cells as well as reconfigurability of cell interconnections. The *programmable systolic chip* (PSC) being developed at Carnegie-Mellon (CMU) is intended to explore the design space of systolic processors where this highest degree of flexibility is achieved. The chip will form the basic cells for a wide variety of systolic arrays (see Figure 6). Note that conventional, commercially available microprocessor components are not adequate for forming cells in systolic arrays, because they do not have enough I/O bandwidth to pass data from cell to cell with a speed sufficient to balance the computation speed. Moreover, unlike the PSC, conventional microprocessors do not have fast, on-chip multiplier circuits which are crucial for high-speed signal and image processing. By using arrays consisting of tens or hundreds of copies of the PSC chip, we expect to get orders of magnitude improvements in speed for application including:

- Convolution-like operations in signal and image processing.
- Reed-Solomon encoding and decoding for error corrections.
- Encryption and finger-print calculations.
- Number theoretic transforms.
- Disk sorting.

Consider the application to Reed-Solomon error-correcting codes as an example. Suppose that each codeword consists of 224 information symbols

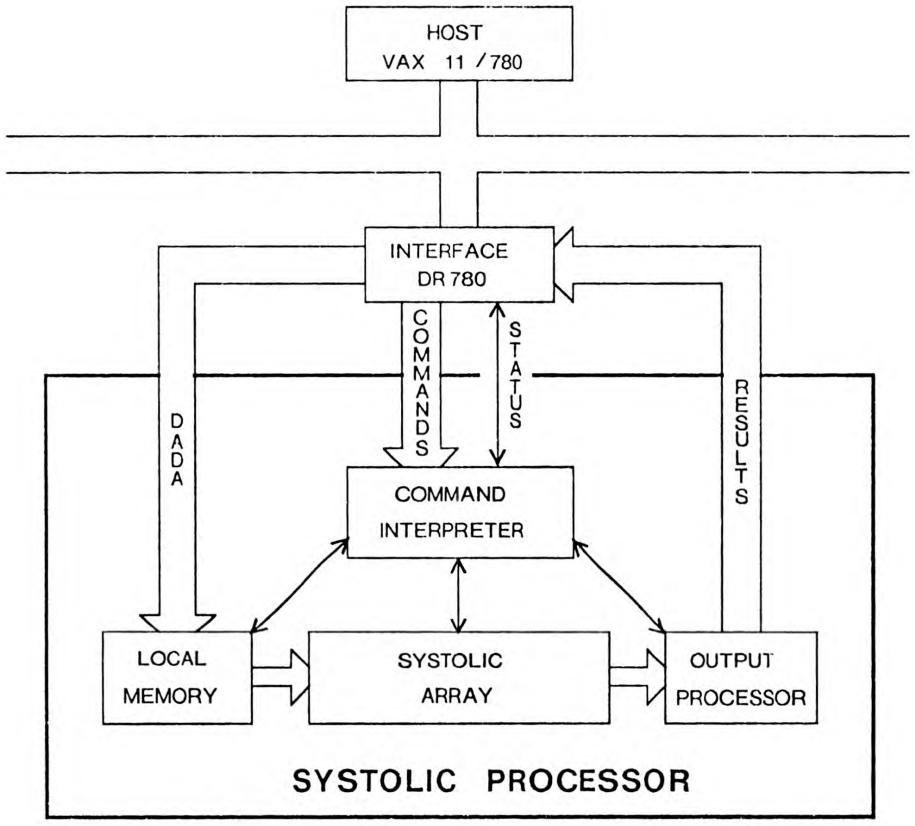


Figure 5. The ESL systolic processor.

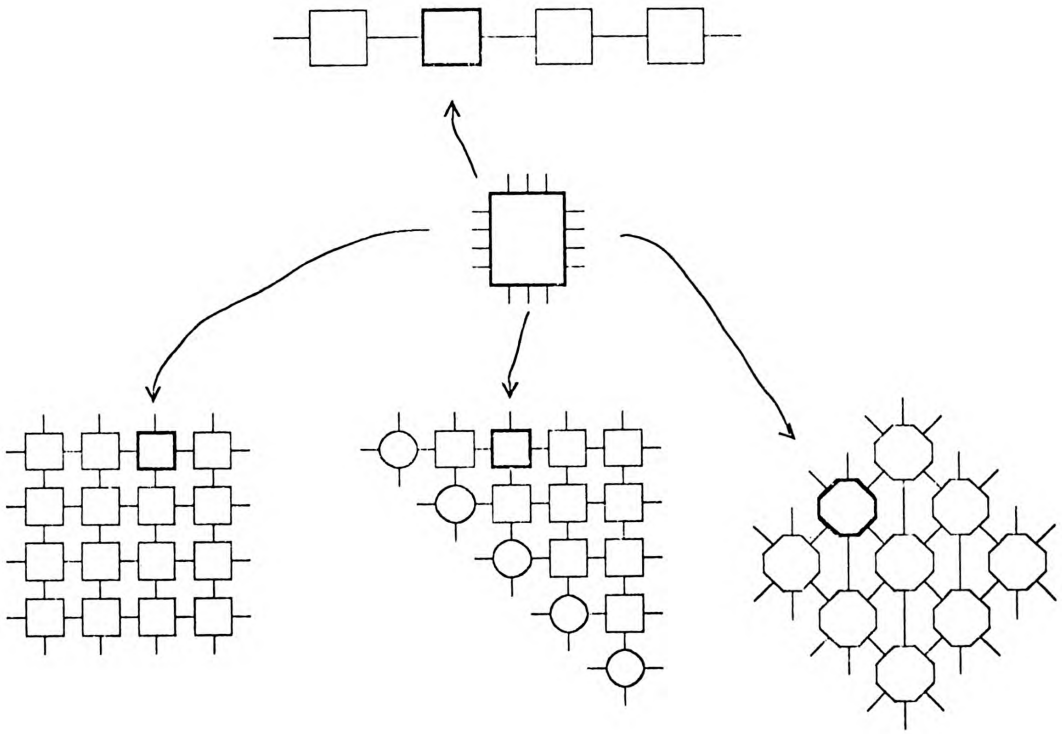


Figure 6. A building-block chip for a variety of systolic arrays.

followed by 32 check symbols and each symbol is an 8-bit interger; using the Reed-Solomon scheme, errors involving no more than 16 symbols can be corrected. We estimate that by using a linear array of 112 PSC chips the Reed-Solomon decoding can be performed in a throughput of 10 million bits per second. Encoding is much easier; it requires only about 30 PSC chips to achieve the same throughput. As far as we know, the fastest existing Reed-Solomon decoder with the same characteristics uses about 500 chips but achieves a throughput of only 1 million bits per second. The performance of the PSC-based system is largely due to the use of systolic array; for example, a highly effective systolic array has been developed for implementing the extended Euclidean computation, the most diffcult task in the decoder implementation [35]. Five people are currently working on the PSC project (July 1982). Microprograms for most of the applications mentioned above have been written and tested, using a register level simulator generated from a hardware description of the chip. Layouts for the major components, the multiplier-accumulator, the 4K (64 x 64-bits) memory for the control store, and the ALU, are complete. The architecture of the chip has been finalized, and the chip is to be completely laid out and simulated by September 1982.

We see that there is a spectrum of possibilities for systolic processor designs, ranging from single-purpose processors to general-purpose, programmable processors such as the PSC chip, as depicted conceptually in Figure 7. The choice of a particular architecture in the design space depends on each individual application problem at hand.

It is worthwhile to note that a systolic processor has recently been implemented by the Naval Ocean Systems Center in San Diego as a testbed for a variety of systolic array configurations, and to understand the architectural features needed by various systolic algorithms [38]. This processor has an 8x8 array of cells, each containing a Intel Arithmetic Processing Unit and microprocessor for achieving the fexibility needed in the testbed.

Future Directions

With the development of systolic architectures, more and more special-purpose systems may become feasible—especially systems that implement fixed, well-understood computation routines. The ultimate goal is effective use of systolic processors in general computing environments to off-load regular, compute-

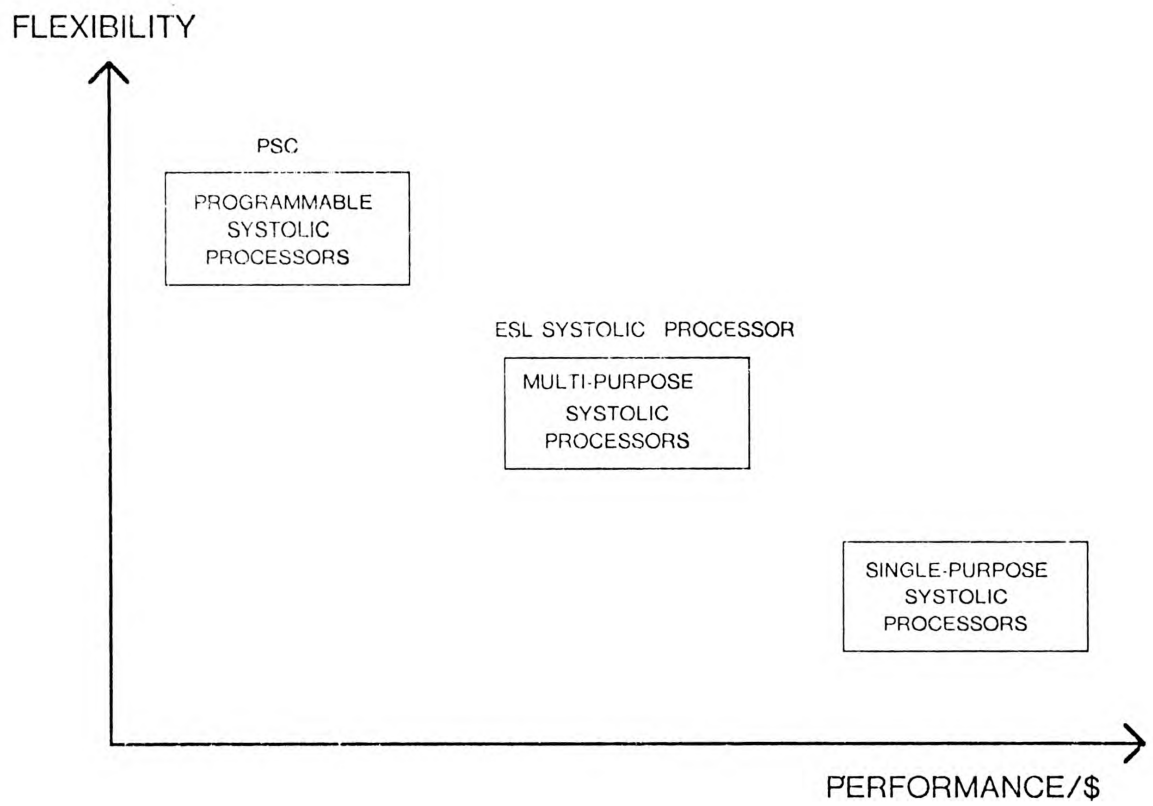


Figure. 7. Design space for systolic processors.

bound computations. To achieve this goal further research is needed in two areas. The first concerns the system integration: we must provide a convenient means for incorporating high-performance systolic processors into a complete system and for understanding their effective utilization from a system point of view. The *Universal Host* project at CMU, sponsored by ONR, directly addresses this system integration problem. We are in the process of constructing a compact host system that can be plugged in with interchangeable, high-performance, special-purpose modules to fit various application requirements. The second research area is to specify more building-blocks such as the PSC chip described earlier for a variety of systolic processors so that, once built, these building-blocks can be programmed to form basic cells for a number of systolic systems. The building-block approach seems inherently suitable to systolic architectures since they tend to use only a few types of simple cells. By combining these building-blocks regularly, systolic systems geared to different applications can be obtained with little effort.

We believe that because of the simplicity and regularity inherent in systolic structures, it will eventually be possible to have an *integrated* treatment for all the design efforts needed to the construction of a systolic system, starting from application to the final system integration. A consequence of this integration is that application people will have direct influence on the specification and design of custom VLSI systems to meet their needs, while enjoying fast turn-around time to build the systems. ■

References

1. Kung, H.T. and C.E. Leiserson, "Systolic Arrays (for VLSI)," I.S. Duff, and G.W. Stewart, (editors), *Sparse Matrix Proceedings 1978,* Society for Industrial and Applied Mathematics, (1979) 256-282. A slightly different version appears in *Introduction to VLSI Systems* by C.A. Mead and L.A. Conway, Addison-Wesley, (1980) Section 8.3.

2. Kung, H.T., "Why Systolic Architectures?" *Computer Magazine* 15(1) (January 1982) 37-46.

3. Mead, C.A. and L.A. Conway, *Introduction to VLSI Systems.* Addison-Wesley, Reading, Massachusetts, (1980).

4. Sutherland, I.E. and C.A. Mead, "Microelectronics and Computer Science," *Scientific American* 237(3) (September 1977) 210-228.

5. Treleaven, P.C., "VLSI Processor Architectures," *Computer Magazine* 15(6) (June 1982) 33-45.

6. Bentley, J.L. and H.T. Kung, "A Tree Machine for Searching Problems," *Proceedings of 1979 International Conference on Parallel Processing,* IEEE, (August 1979), 257-266.

7. Thompson, C.D., *A Complexity Theory for VLSI,* Phd thesis, Carnegie-Mellon University, Computer Science Department (1980).

8. Brent R.P. and H.T. Kung, "The Area-Time Complexity of Binary Multiplication," *Journal of the ACM* 28(3), (July 1981) 521-534.

9. Hong, J.W. and H.T. Kung, "I/O Complexity: The Red-Blue Pebble Game," *Proceedings of the Thirteenth Annual ACM Symposium on Theory of Computing,* ACM SIGACT, (May 1981) 326-333.

10. Leiserson, C.E. and J.B. Saxe, "Optimizing Synchronous Systems," *Proceeding of the 22nd Annual Symposium on Foundations of Computer Science,* IEEE Computer Society, (October 1981) 23-36.

11. Cappello, P.R. and K. Steiglitz, "Digital Signal Processing Applications of Systolic Algorithms," H.T. Kung, R.F. Sproull, and G.L. Steele, Jr. (editors), *VLSI Systems and Computations,* Computer Science Department, Carnegie-Mellon University, Computer Science Press, Inc., (October 1981) 245-254.

12. Kung, H.T., "Let's Design Algorithms for VLSI Systems," *Proceedings of Conference on Very Large Scale Integration: Architecture, Design, Fabrication,* California Institute of Technology, (January 1979) 65-90. Also available as a CMU Computer Science Department technical report, (September 1979).

13. Kung, H.T., "Special-Purpose Devices for Signal and Image Processing: An Opportunity in VLSI," *Proceedings of the SPIE, Vol. 241, Real-Time Signal Processing III,* The Society of Photo-Optical Instrumentation Engineers, (July 1980) 76-84.

14. Blackmer, J., P. Kuckes and G. Frank, "A 200 MOPS Systolic Processor, *Proceedings of SPIE Symposium," Vol. 298, Real-Time Signal Processing IV.* The Society of Photo-Optical Instrumentation Engineers, (August 1981).

15. Kung, H.T. and R.L. Picard, "Hardware Pipelines for Multi-Dimensional Convolution and Resampling," *Proceedings of the 1981 IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management,* IEEE Computer Society Press, (November 1981) 273-278.

16. Kung, H.T., L.M. Ruane, D.W.L. Yen, "A Two-Level Pipelined Systolic Array for Convolutions," H.T. Kung, R.F. Sproull, and G.L. Steele, Jr. (editors), *VLSI Systems and Computations,* Computer Science Department, Carnegie-Mellon University, Computer Science Press, Inc., (October 1981) 255-264.

17. Kung, H.T. and S.W. Song, "A Systolic 2-D Convolution Chip," K. Preston, Jr. and L. Uhr (editors), *Multicomputers and Image cessing: Algorithms and Programs* (1982) 373-384. An extended abstract appears in *Proceedings of 1981 IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management* (November 11-13, 1981) 159-160.

18. Yen, D.W.L. and A.V. Kulkarni, "The ESL Systolic Processor for Signal and Image Processing," *Proceedings of the 1981 IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management,* (November 1981) 265-272.

19. Fisher, A., "Systolic Algorithms for Running Order Statistics in Signal and Image Processing," H.T. Kung, R.F. Sproull, and G.L. Steele, Jr. (editors), "VLSI Systems and Computations,". Computer Science Department,

Carnegie-Mellon University, Computer Science Press, Inc., (October 1981) 265-272.

20. Weiser, U. and A. Davis, "A Wavefront Notation Tool for VLSI Array Design," H.T. Kung, R.F. Sproull, and G.L. Steele, Jr. (editors), *VLSI Systems and Computations.* Computer Science Department, Carnegie-Mellon University, Computer Science Press, Inc. (October 1981) 226-234.

21. Gentleman, W.M. and H.T. Kung, "Matrix Triangularization by Systolic Arrays," *Proceedings of SPIE Symposium, Vol. 298, Real-Time Signal Processing IV.* The Society of Photo-Optical Intrumentation Engineers, (August 1981).

22. Bojanczyk, A., R.P. Brent, and H.T. Kung, *Numerically Stable Solution of Dense System of Linear Equations Using Mesh-Connected Processors,* Technical Report, Carnegie-Mellon University, Computer Science Department (May 1981). The final version of the paper is to appear in *SIAM Journal on Scientific and Statistical Computing.*

23. Guibas, L.J. and F.M. Liang, "Systolic Stacks, Queues, and Counters," *Proceedings of the Conference on Advanced Research in VLSI,* Cambridge, Massachusetts (January 1982).

24. Ottmann, T., A.L. Rosenberg, and L.J. Stockmeyer, *A Dictionary Machine for VLSI.* Technical Report RC 9060 (#39615), IBM Thomas J. Watson Research Center, Yorktown Heights, New York (1981).

25. Song, S.W., *On a High-Performance VLSI Solution to Database Problems.* PhD thesis, Carnegie-Mellon University, Computer Science Department (July 1981). Also available as a CMU Computer Science Department technical report (August 1981).

26. Leiserson, C.E., "Systolic Priority Queues," *Proceedings of Conference on Very Large Scale Integration: Architecture, Design, Fabrication,* California Institute of Technology (January 1979) 199-214. Also available as a CMU Computer Science Department technical report (April 1979).

27. Guibas, L.J., H.T. Kung, and C.D. Thompson, "Direct VLSI Implementation of Combinatorial Algorithms," *Preceedings of Conference on Very Large Scale Integration: Architecture, Design, Fabrication,* California Institute of Technology (January 1979) 509-525.

28. Bentley, J.L., "A Parallel Algorithm for Constructing Minimum Spanning Trees," *Journal of Algorithms* (1980) 1:51-59.

29. Savage, C., "A Systolic Data Structure Chip for Connectivity Problems," H.T. Kung, R.F. Sproull, and G.L. Steele, Jr. (editors), *VLSI Systems and Computations,* Computer Science Department, Carnegie-Mellon University, Computer Science Press, Inc. (October 1981) 296-300.

30. Chazelle, Bernard, "Computational Geometry on a Systolic Chip, " Technical Report, Carnegie-Mellon University, Computer Science Department (May 1982).

31. Foster, M.J. and H.T. Kung, "The Design of Special-Purpose VLSI Chips," *Computer 13(1)* (January 1980) 26-40. Reprint of the paper appears in *Digital MOS Integrated Circuits,* edited by M.I. Elmasry, IEEE Press Slected Reprint Series, (1981) 204-217. A preliminary version of the paper, entitled "Design of Special-Purpose VLSI Chips: Example and Opinions," also appears in *Proceedings of the 7th International Symposium on Computer Architecture,* La Baule, France, (May 1980) 300-307.

32. Foster, M.J. and H.T. Kung, "Rec ,nize Regular Languages With Prograr nable Building-Blocks," J.P. Gray (editor) *VLSI 81.* Academic Press, (August 1981) 75-84. The final version is to appear in *Journal of Digital Systems.*

33. Kung, H.T., "Use of VLSI in Algebraic Computation: Some Suggestions," P.S. Wang (editor), *Proceedings of the 1981 ACM Symposium on Symbolic and Algebraic Computation,* ACM SIGSAM, (August 1981) 218-222.

34. Brent, R.P. and H.T. Kung, "*Systolic VLSI Arrays for Polynomial GCD Computation,*"

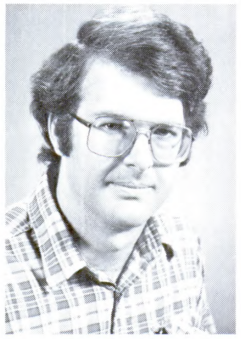
Technical Report, Carnegie-Mellon University, Computer Science Department, (May 1982).

35. Kung, H.T. and P.L. Lehman, "Systolic (VLSI) Arrays for Relational Database Operations," *Proceedings of ACM-SIGMOD 1980 International Conference on Management of Data,* ACM, (May 1980) 105-116. Also available as a CMU Computer Science Department technical report, (August 1979).

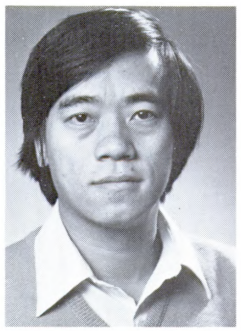
36. Lehman, P.L., "A Systolic (VLSI) Array for Processing Simple Relational Queries," H.T. Kung, R.F. Sproull, and G.L. Steele, Jr. (editors), *VLSI Systems and Computations,* Computer Science Department, Carnegie-Mellon University, Computer Science Press, Inc., (October 1981) 285-295.

37. Whiteside, R.A., P.G. Hibbard, and N.S. Ostlund, *Systolic Algorithms for Monte Carlo Simulations,* Draft, CMU Computer Science Department.

38. Bromley, K., J.J. Symanski, J.M. Speiser, and H.J. Whitehouse, "Systolic Array Processor Developments," H.T. Kung, R.F. Sproull, and G.L. Steele, Jr. (editors), *VLSI Systems and Computations,* Computer Science Department, Carnegie-Mellon University, Computer Science Press, Inc., (October 1981) 273-284.



Dr. Jon Louis Bentley is an Associate Professor of Computer Science and Mathematics at Carnegie-Mellon University. His research interests include the design and analysis of computer algorithms and their application in computer systems; other areas in which he has worked include software engineering and novel computer architectures. He has consulted for Bell Laboratories and IBM Corporation.



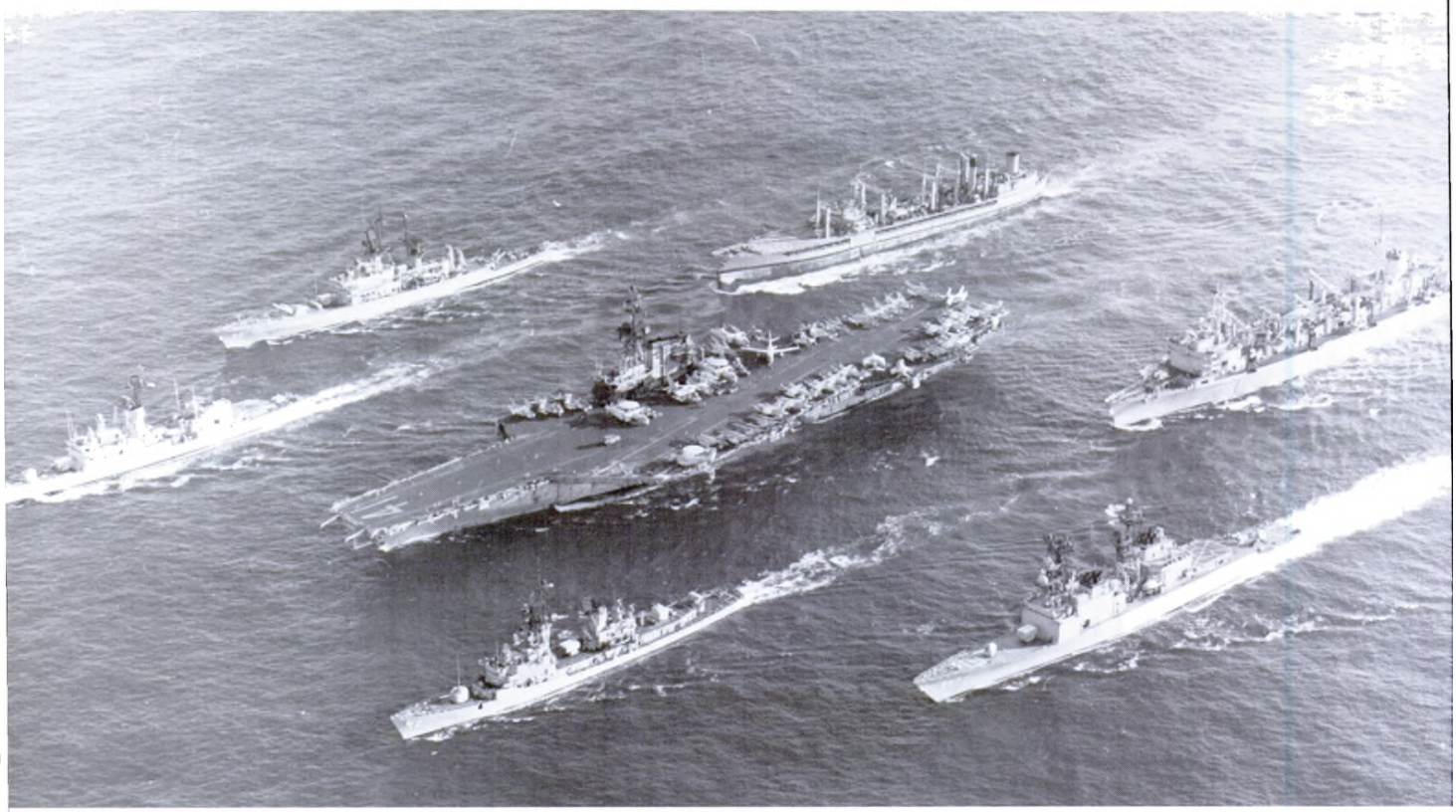
Dr. H. T. Kung is Professor of Computer Science at Carnegie-Mellon University, where he leads a research group in the design and implementation of very large scale integrated (VLSI) systems. He is interested in paradigms of mapping algorithms and systems directly onto silicon chips, and pioneered the concept of systolic processing. For five years he has been a principal investigator for the Office of Naval Research.

Profiles in science

Walter L. Smith is a Professor and Chairman of Statistics at the University of North Carolina at Chapel Hill and has been for many years a principal investigator for the Office of Naval Research. Professor Smith formulated a number of concepts pertaining to applied probability theory which are now important features in that field. He was the first to realize the usefulness of the Key Renewal Theorem, which has found a considerable range of applications. He introduced successively the theories of Regenerative Stochastic Processes, Cumulative Processes, Quasi-Poisson Processes, and simultaneously with (but independently of) Paul Levy, Semi-Markov Processes. Recently he pioneered the theory of Branching Processes in Random Environments. He is currently interested in improvements and refinements of formulae used to study multivariate versions of the Cumulative Processes to enhance their practical applications.

He was awarded the Adams Prize by the University of Cambridge in 1961. He is a Fellow of the Institute of Mathematical Statistics, the American Statistical Association, the Royal Statistical Society, and the Cambridge Philosophical Society. He is also an Elected Member of the International Statistical Institute and the Operations Research Society.

System Theoretic Challenges in Military C^3 Systems



by

Michael Athans
Massachusetts Institute of Technology

Introduction

The purpose of this paper is to examine certain generic problems of military Command, Control and Communications (C^3) systems from the viewpoint of modern systems engineering and to suggest research avenues that appear to be promising in advancing the state-of-the-art for C^3 processes. The research described here is funded in part by the Office of the Naval Research and the Air Force Office of Scientific Research.

The field of Command and Control has received a lot of attention in the past few years (see reference 1 for a recent collection of views on the subject by several high level Department of Defense (DOD) and industrial contributors). The reason of this flurry of interest appears to be due to several factors:

(1) There seems to be some dissatisfaction with C^3 related procurements. To a certain degree this dissatisfaction may stem from unrealistic expectations on the part of the military user community on potential performance. There have been several comments voiced that the hardware specifications and system-wide performance were decided by engineers that are not aware of the "true" needs of operational commanders.

(2) There is an increasing awareness that it is very difficult to arrive at rational design specifications for the performance of a C^3 system under a variety of peacetime, crisis-management, and hot war conditions. The integration of specifications for individual hardware components into a system-wide peformance index is a very complex task, and there does not exist a systematic methodology to aid the systems engineer to carry out the necessary cost-performance-reliability tradeoffs.

(3) The increasing range, speed, and accuracy of sophisticated weapons systems enlarge the volume of responsibility of commanders, thus necessitating an increased geographical dispersion of the commanders' surveillance and warfare assets.

(4) The increased speed of weapons systems have reduced the time available for making tactical decisions by human decision makers. This requires an ever increasing use of automation at all levels.

(5) Centralized command posts are vulnerable to enemy attack. This necessitates the geographical distribution of the command hierarchy and requires increased tactical communication among command centers for force coordination.

(6) The necessary secure and reliable communications links represent a very vulnerable component of the C^3 system. Modern electronic warfare (EW) disrupts communications through jamming during the time that tactical communications are needed most for force coordination. Similiar communications constraints can occur due to environmental conditions (terrain, weather, etc.). High power communications are also undesirable because the emitted electromagnetic energy can be detected by enemy sensors, such as high-frequency direction finders (HFDF).

(7) Active, high signal-to-noise ratio sensors, such as radar, although excellent in the surveillance problem, suffer from similar electromagnetic detection problems. Passive sensors are less vulnerable, but they do not provide accurate information unless they are internetted with a high-bandwidth communications network.

The above remarks indicate some of the complex systems engineering issues that have to be addressed in C^3 processes. It should be self-evident that one deals with very complex large-scale distributed estimation and decision processes with a high degree of uncertainty due to environmental variables and enemy actions. The entire command and control process is a cybernetic process with several feedback loops. Clearly, if a relevant set of theories and tools are developed for military C^3 systems, then these results would be applicable to similar problems arising in the civilian sector (power systems, transportation systems, etc.) which operate in a less stressful and less hostile environment.

Some Definitions

The words "command-and-control (C^2)," "command, control and communications (C^3)," "com-

mand, control, communications and intelligence (C^3I)" have different meanings to different people, and their indiscriminate use can create a great degree of confusion. For this reason, in this paper some distinction between the C^2 and C^3 terms will be made.

C^2 Process: The means by which a team of military warfare commanders (WC) make decisions that relate to the deployment and motion of the resources and assets assigned to them to carry-out a military mission specified by higher authority.

C^2 Authority and Responsibility: This refers to the fact that a particular commander can instruct, direct, position, move, and, in general, control the human and physical assets assigned to him by his superior commander. The responsibility is to accomplish the specific objectives spelled out to him by his superior commander.

C^2 Organization: The semihierarchical way and organizational rules by which the human commanders organize themselves in terms of C^2 authority and responsibility by warfare area and/or geographical sector. Rules of engagement and tactical doctrine are an integral part of the C^2 organization.

We remark that in this paper we shall use the term C^2 to primarily relate to human decision making. Obviously commanders need real-time information to make decisions and weapons systems that eventually implement a subset of their decisions. This leads us to the following definitions.

C^3 System Elements: The physical and technological hardware and software that generate, manipulate, communicate, and display information and the weapon systems. Thus typical C^3 System Elements are:

(a) Sensors (Fixed or Moving);
(b) Communication Links (mostly radio for tactical C^2 and related devices;
(c) Computers and Displays (hardware, software, firmware, decision aids) viewed as systems;
(d) Weapon platforms and weapons systems.

C^3 Systems: The physical system and its architecture that defines the interconnection of the C^3 elements. Thus it is the C^3 system that:

(a) generates data and information for different WC's in the C^2 organization;
(b) allows for coordination among several WC's;
(c) provides means for implementing the decisions generated by the C^2 process.

It should be noted that the physical elements of the C^3 systems, as well as the warfare commanders (WC) are naturally and/or intentionally geographically distributed; geographic distribution reduces the vulnerability of the C^2 organization and of the C^3

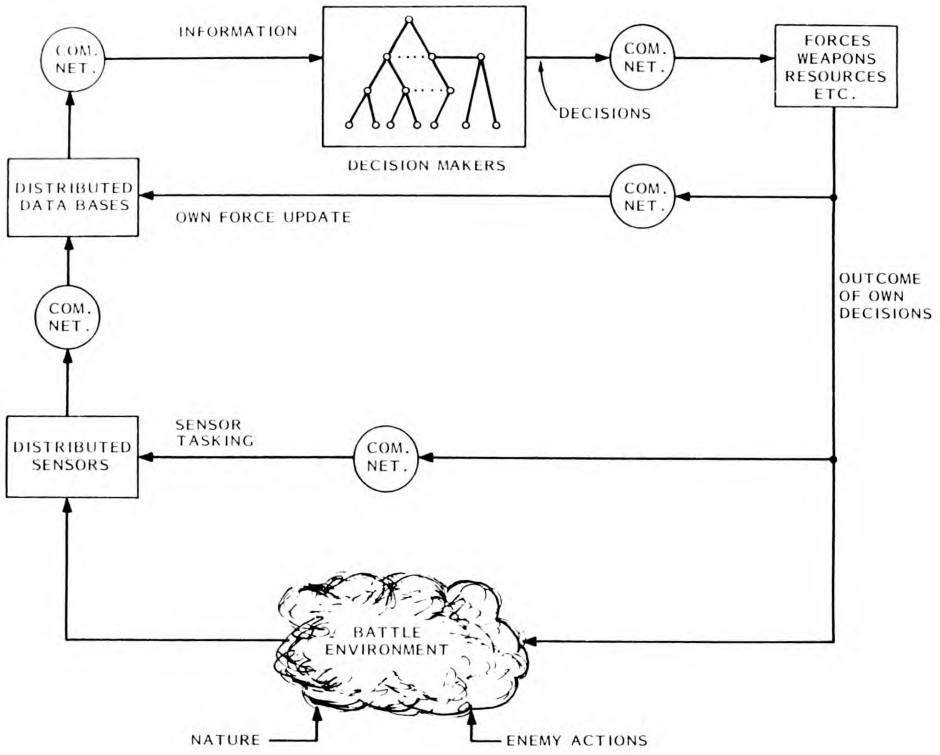


Figure 1. The C^3 Process and the C^3 System.

system to enemy attack but strains the radio communication requirements* necessary for coordination among WC's and to carry out the C^2 process.

Figure 1 represents a highly simplified schematic of the C^2 process interacting with the C^3 system. It represents a highly multivariable dynamic and stochastic feedback control process or a cybernetic process. The emphasis of Figure 1 is upon the functions of the organic C^3 system as providing information (not data!) to the C^2 organization and acting as an actuator or effector that suitably transforms the collective decisions of the WC's into physical events.

Parenthetically we remark that the inclusion of weapons as part of the C^3 system is not a widely accepted convention. It is the author's opinion that they must be included since at a minimum they will generate tracks at a subset of sensors and they must be sorted out from the enemy and neutral objects and targets. Also, when we consider coordination problems among WC's with respect to proper utilization of multiwarfare-capable assets, it becomes very difficult to define multiwarfare-capability at a systems level without explicit consideration of the weapons systems that are carried in that asset. Finally, the whole issue of fratricide cannot be addressed in the absence of weapons.

The function of the C^3 system is to provide relevant, accurate, and timely information to the commanders so that they can make correct decisions regarding the deployment and movement of their forces and resources, carry out the necessary resource allocations, and achieve the objectives assigned to them. To accomplish their missions the commanders must establish a command and control (C^2) organizational structure that can deal effectively with rapidly changing tactical situations. The necessary C^2 organizational structure is very much dependent upon sensors, communications, and weapons technology. This is a key point to keep in mind. For example, in the absence of tactical communications, the organizational structure must be such that each commander can operate in an open-loop manner (or according to doctrine in military language). Clearly, a different organizational structure will be more effective if communications are available, so that force and resource coordination can be carried out in real time (closed loop command and control).

Radio communications bandwith is the most precious and vulnerable element in the C^3 system because of its susceptibility to jamming. Spread spectrum techniques can reduce the jamming vulnerability, but there is just so much bandwith in the universe. UHF communications are less susceptible to enemy intercepts and localization, but constrain the system to line-of-sight communications.

The distributed nature of the sensors and the distributed nature of human decision makers obviously interact with the architecture of the C^3 system. It is unrealistic to expect that all relevant information is transmitted in a timely manner to all decision makers; this would consume valuable (and vulnerable) communications resources and require a large time-delay. Even if such "global" information could be transmitted to all human decision makers; they would be "swamped" with information, and they would be unable to use it effectively under the stringent time constraints within which commanders must operate in a stressful tactical engagement.

The above remarks illustrate the fascinating system-theoretic issues that arise in military command and control. These problems suggest new and unexplored avenues for theoretical research in modern systems, estimation, and control theory. The problems are clearly of the large-scale system variety, but quite different from those considered in the system-theoretic literature.

Human Decision Makers

In this paper, the tactical aspects of command and control will be discussed leaving aside the strategic and planning issues; these are extremely important, and quite often the success of a mission will crucially depend upon careful planning, resource allocation, and logistic support. These set up some of the goals and constraints that a commander must operate under in any given tactical situation.

The purpose of a C^3 system is to provide the necessary information to a *team* of human decision makers so that they can carry out the mission assigned to them.

It is very important to stress that tactical decisions in a modern warfare environment are carried out by a team of commanders, rather than by a single commander. The reason for this is obvious. Weapons technology has progressed to the point that no single human decision maker can be an expert on all offensive and defensive options; even if he were, there would be no time to absorb all information and execute all decisions. To be sure, there is and there always will be a traditional military command hierarchy; however, from a functional point of view, modern warfare requires a much more fluid structure as far as tactical decision making is concerned.

To illustrate this concept, let us consider a naval battle group or task force consisting of at least one aircraft carrier together with several other platforms

(cruisers, destroyers, etc.) and support vessels. Such a task force is capable of conducting *simultaneously* anti-submarine warfare (ASW), anti-air warfare (AAW), and anti-surface warfare (ASUW) in the sense that it possesses both the sensor and weapons resources, distributed among the platforms, to engage simultaneously submarines, aircraft, ships and missiles (which can be launched from submarines, airplanes, and ships as well as land bases). In order for the task force to be successful in such a complex engagement requires the coordinated decision making of several commanders, who may not necessarily be colocated. The U.S. Fleet has adopted the doctrine of Composite Warefare Commander (CWC) who oversees the actions of an ASW Commander (ASWC), and AAW Commander (AAWC), and an ASUW Commander (ASUWC). Each of the three subordinate commanders has a functional responsibility for defensive and offensive decision and actions in his area of "expertise." Interestingly enough the CWS "governs by negation;" the CWC monitors the decisions of the ASWC, AAWC, and ASUWC and he only intervenes when either an inappropriate (in his mind) decision is contemplated or when there is obvious need to resolve conflicts related to resource allocation.

The three main commanders (ASWC, AAWC, ASUWC) are those that issue C^2 orders. However, their actions are influenced by inputs of coordinators (such as the screen coordinator, electronic warfare coordinator, helicopter coordinator, etc.) who have responsibility for a certain function associated with resources controlled by the commanders (ASWC, AAWC, and ASUWC). These coordinators monitor the status of the specific problem assigned to them and provide recommendation to the commanders on what to do. Interestingly enough one can abstract the objectives of the coordinators in terms of specific detection, survival, deception, and kill probabilities in their specific area of responsibility. Also note that the coordinators do not serve necessarily as staff to one of the commanders; rather they assess situations and generate recommendations that may involve the assets of more than one commander.

The above discussion illustrates that although there is a clear cut command hierarchy (e.g., admiral, captain, etc.), the team-decision mechanism in a tactical situation is *not* hierarchical is nature (where the word hierarchical is used in the strict system theoretic context).

Such team coordinator issues are not unique to the Navy. Even more complex tactical coordination issues arise when one considers a land warfare scenario which involves the coordinated actions of both Air Force and Army assets. In this case the problems of tactical coordination are further complicated by the existence of traditional chains of command between two services as well as necessary coordination with Allied ground and air assets.[2,3]

Thus, if one examines tactical warfare one must immediately come to grips with problems of distributed decision making by teams of expert commanders. For maximum effectiveness the resources available must be coordinated in time and in space. The complexity of the problem requires a division of responsibility and a certain autonomy in the decision making process of an individual commander. On the other hand, the commanders must communicate selectively for best resource utilization; communications may be subject to constraints especially when the commanders are geographically distributed to decrease vulnerability. It is precisely the function of the C^3 system to provide to each commander the right information at the right time so that he can accomplish the mission assigned to him, and to allow for the necessary coordination of the team resources and decisions.

The most information that a commander needs is the location, velocity, and identity of his own, the enemy's and neutral objects. Past motion history is often important in deducing the enemy intentions. Fuel and weapons availability related to one's own assets is obviously important also. Such information must be gathered by the C^3 surveillance system. It is important to stress that in a tactical situation, surveillance and tracking information can be collected by a variety of sensors. Often the sensor measurements must be "fused," i.e. correlated with each other so as to provide an accurate situation estimate (some sensors such as radar provide very good position information but not identity; others, such as ESM can provide some identity information but poor position accuracy). Because the sensors are geographically distributed, there is a significant time delay before an accurate situation plot can be generated and transmitted to the commanders.

In a tactical situation each commander is presented with a view of the "state of the world" which he knows can be inaccurate and not necessarily very timely. Nonetheless, on the basis of this incomplete information, each comander must make decisions consistent with the constraints imposed by the preplanned actions. Typical decisions involve:

(a) Control of surveillance resources (e.g. turn on a radar, launch a reconnaisance aircraft, etc.) to gather more information or clarify ambiguous information;

(b) Control of electromagnetic radiation (e.g. communicate or not, jamming strategies, etc.);

(c) Control of resources (e.g. relative position of ships, aircraft, tanks, troops, etc. and control of their movements);
(d) Assignment of weapons to targets (e.g. sortie planning by deciding what aircraft, from what bases should be armed, with what weapons to attack what targets or other objects of military value);
(e) Weapons control.

These decisions are always made by human commanders. Some of the decisions are strategic in nature, i.e., they are the outcome of extensive preparation and planning. In a system-theoretic context, the strategic decisions and planning are roughly equivalent to the establishment of desired open-loop controls and trajectories; and one can argue that such strategic or command decisions are the outcome of a static or dynamic deterministic optimization problem. In this phase, intelligence information is crucial. In the planning phase, many details are not taken into account. Uncertainty is usually handled by planning in detail alternate options; specific and unambiguous objectives and directives are command for execution and implementation by the appropriate commanders. Generally[5], subordinate commanders are told who, what, where, when, and why. In general, subordinate commanders are not told "how"; it is up to the subordinate commander to develop a detailed tactical plan of action to accomplish the objective assigned to him.

It is useful, in the opinion of the author, to think of the "command" function in C^2 systems as being completely analogous to the open-loop control concept of modern control theory. The command function effectively specifies the reference trajectories in time and space (and alternate ones) for the mission to be performed. Although the tactical actions will be executed by a geographically distributed set of subordinate commanders according to their best decisions (remember the "how" is not specified), the availability of a global open-loop plan provides valuable open-loop *coordination* among distinct units. For example, neighboring platoons in a land warfare scenario will know the planned location and plans of adjacent units, as well as the time and place of close air-support operations. This helps a platoon commander plan his own course of action by correctly allocating his own defensive and offensive resources.

The word "control" in C^2 systems refers to the function that indeed the preplanned courses of action are more or less being accomplished in a *tactical* situation. In the author's opinion, the "control" function is completely analogous to "feedback." Real-time surveillance provides the commander with an estimate of the "state of the world"; he compares this with the desired state associated with the command process, and real-time decisions are made to control in real time the available resources to correct for undesirable deviations.[4]

The nature of real time information by the tactical surveillance system can impact the overall decision process in different ways. A crucial piece of "global" information is one that violates a key assumption* under which the operational battle plans were made. This may necessitate the implementation of alternative operational plans at the highest level. If such a contingency were taken into account in the planning process, the alternate plans can be sent easily and rapidly to the subordinate commanders. If, on the other hand, this contingency was not anticipated, and there is no time for drawing another set of plans (or if the communications environment does not permit transmission of the commands), the tactical situation will become chaotic at least in the short run.

The feedback "control" function requires the real-time reallocation by an individual warfare commander of the resources to meet his objectives. Sometimes he can accomplish this with the resources alloted to him; in certain cases, the situation necessitates the real-time coordination of the resources of two district commanders, and this obviously requires tactical communications.

The above discussion illustrates the nature of the decisions that have to be carried out by individual commanders, at different levels of the command hierarchy. The quality of the decisions made by an individual commander depend on the following key factors:

(a) The nature, quality, and especially timeliness of the available information (this can be greatly influenced by a superior C^3 system);
(b) The rules of the engagement (these act as constraints upon the commander's decision process);
(c) The goals and objectives assigned to him by superior commanders at the strategic planning phase;

**A good example is the naval battle at Midway Island during World War II. The Japanese plans to invade Midway were based upon the intelligence assumption that the U.S. carriers were at Pearl Harbor. By breaking the Japanese code, the U.S. Naval forces knew of the intended invasion of Midway, and Admiral Nimitz dispatched a three carrier force near Midway. The presence, location, and size of the U.S. task force did not become apparent to the Japanese until the battle started. This may have been a deciding factor in the battle outcome, in which the Japanese lost four of their large aircraft carriers, and abandoned the invasion plan.*

(d) The commander's available resources (these again act as constraints to his decisions);
(e) The planning horizon time;
(f) The complexity of the tactical situation vs. the time available to arrive at a satisfactory decision (computer decision aide for the commander can help him to either arrive at better decisions within a given time limit and/or complete his decisions sooner).

The overall global outcome of any engagement will clearly depend upon the quality of the decisions of the distributed commanders.

System - Theoretic Research Issues

In the above section we discussed generic issues associated with the Command and Control process. In this section we outline some relevant research directions, from a system-theoretic viewpoint, that are necessary to develop methodologies and theories for C^3 systems.

Clearly military C^3 systems fall in the category of large-scale distributed decision systems. In spite of the many advances in large-scale estimation and control systems[8], we do not have a unified theoretical methodology for such systems.

Although C^3 systems fall in the category of large scale systems, nonetheless they are characterized by certain key properties and attributes that must be taken in account. These are:

(1) They are event-driven,
(2) The dynamics* tend to be trivial, and there is not a significant dynamic coupling among the system elements,
(3) The coupling occurs primarily at the resource allocation level and the coordination of the resources in time and in space.

Military C^3 systems are characterized by geographical dispersion and phenomena that involve multiple time scales (for example, ASW operations are slow as compared to AAW operations). This spatial and temporal decomposition naturally leads to command distribution without the need for extensive coordination in many cases.

From a technical point of view, the distributed detection and estimation problems that arise in the surveillance area and in the sensor tasking area appear to be those that are most amenable to analytical treatment. Unified theoretical and algorithmic approaches are needed in: the generic problem of multiple geographically distributed sensors tracking multiple geographically distributed objects, the networking of these sensors, the necessary distributed data base management issues, and the communications requirements. A recent paper[6] discusses some of the issues in the C^3 surveillance functions so no more comments on that topic will be given here.

The greatest challenge by far is to understand the interactions of a distributed team of human decision makers with the mechanistic and electronic components of the C^3 system. Fundamental understanding is required of the proper functional organizational structure of human commanders with the organization and architecture of the underlying C^3 system. It should be stressed that technological advances in the mechanistic and electronic components of the C^3 systems (weapons, sensors, computers, communications, etc.) will have a definite impact upon the organizational structure of the commanders, including their geographical distribution to reduce the vulnerability of centralized command and control centers. Tactical coordination will increase as multipurpose platforms (ships, aircraft, etc.), capable of performing simultaneously many functions (e.g., ASW, AAW, ASUW, EW), become increasingly available.

One cannot avoid facing squarely the issue that we must develop *normative* mathematical models for individual decision making. One approach that is under consideration[9] is to attempt to model the decisions of a *well-trained* expert human commander as the output of a constrained optimization (static or dynamic) problem. In this formulation, the constraints on the optimization problem are related to the information available to the commander, the rules of engagement, the assigned resources, the desired mission, the planning horizon, and time-deadlines. The greatest problem area is to quantify the objective function used by the human consistent with his limitations in problem solving. A key constraint that must be included in the problem formulation is the limitations of human short-term memory (STM)[10]. Information that is being processed by the central nervous system has to be held in STM, which represents a memory of notoriously short term capacity; human performance on cognitive tasks is dramatically sensitive to the limits of STM. On the other hand, an experienced well-trained expert has stored a tremendous amount of information in long term memory (LTM). As Simon points out[10], "the accumulation of experience may

*By this we mean the trajectories of aircraft, ships, missiles, tanks, etc. The motion of these resources does not create a dynamic coupling as in the case with large scale power systems.

allow humans to behave in ways that are very nearly optimal in situations to which their experience is pertinent, but will be of little help when genuinely novel situations are presented." It may very well be possible to abstract the human's objectives by suitable combinations of conditional detection probabilities, deception probabilities, survival probabilities, and kill probabilities to properly take into account the risk-taking and risk-aversion characteristics of human decision making.[3, 10, 11, 12]

Normative models for humans are not new in control theory. There have been several successful and validated models of a human acting as a control.[7, 13] Recently, these normative models have been extended to the case of a single decision maker having to accomplish a multiplicity of dynamic tasks under time deadline constraints[7]. These results are very encourgaging, not only because there is good agreement between the theoretical predictions of the normative optimal decision model for the human and the experimental data, but also because they demonstrate that for rapid tactical-like problems with significant uncertainty the planning horizon of the human tends to be short. This is consistent with the limitations of short-term memory discussed above which precludes the human from solving in his head stochastic dynamic programming problems with long planning horizons.

It is important to stress that such normative models of warfare commanders cannot be developed in a vacuum; they must be developed in the context of the physical C^3 system. An expert naval anti-air warfare (AAW) commander would not do a very good job if all of a sudden 100 Backfires appeared 150 miles from the carrier. This would represent a "novel" situation. However, such an event is extremely unlikely; it is precisely the intelligence and surveillance function of the C^3 system to prevent having the AAW commander having to face such a novel situation.

We shall refer to such normative models of an individual warfare commander as his *Principal Expert Model* (PEM). The PEM reflects the superior ability (expertise) of a well-trained commander to make tactical decisions in the warfare area/sector of his C^2 authority and responsibility.

Figure 2 illustrates the key elements of the PEM. The "blocks" within the PEM are those that must be reflected in the human-optimization model.

(1) The *Model of the World* in the area of the WC expertise must include *as a minimum* the disposition of his own assets, enemy assets, and capabilities of the sensor and weapon systems under his control. His mental model in his LTM will be updated by tactical information processed by his STM. The tac-

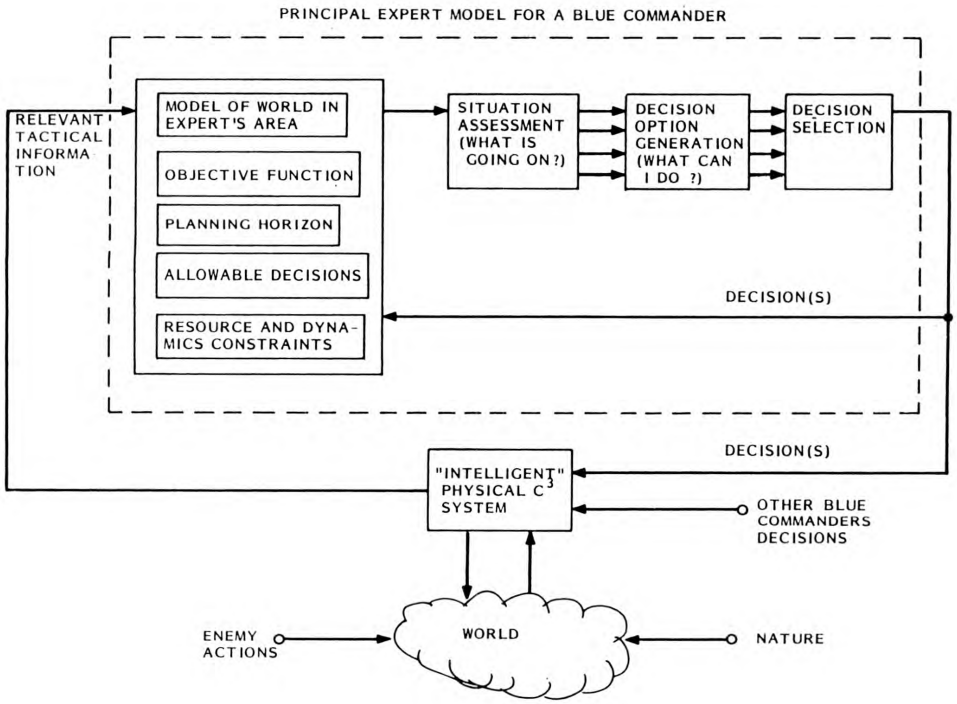


Figure 2. Conceptual Model of the C^2 Decision Process.

tical information and the interactions between STM and LTM leads to a small set of alternate hypotheses which correspond to situation assessment. One can indeed argue that one of the key functions of the physical C^3 systems would be to provide quality and timely information to the WC so as to minimize the situation assessment hypotheses (through reduction in uncertainty) since the generation of multiple "what is going on" hypotheses stress the limitations of the STM.

(2) Based on the situation assessment, a set of decision options must be generated and eventually a particular decision option is selected. This process clearly involves the correlation of tactical training stored in LTM and is influenced by:

   (a) the *objective function* which abstracts the mission responsibility and objectives dictated by his superior command;
   (b) the *planning horizon* that dictates the time-urgency of the WC's tactical decisions;
   (c) any constraints on his *allowable decisions* imposed by the C^2 organization (rules of engagement, EMCON status, use of tactical nuclear weapons etc.);
   (d) available *resources* (platforms, sensors, weapons, etc.) and *dynamic constraints* (speed, maneuvarability).

Obviously the decision option generation is one of the most complex aspects of human decision making. It has been claimed* that experienced commanders tend to generate fewer options as compared to more novice commanders. This can only be explained that superior tactics are better organized as "patterns" in the expert commander's LTM, thus minimizing the "chunks" that are transferred in the STM.

It should be noted that the above qualitative description of the Principal Expert Model (PEM) represents a mild extension of the SHOR (Sense-Hypothesize-Options-Resolution) proposed by Wohl[3]. Once more the quality of the information presented by the physical C^3 system can have a significant impact upon the decision option generation process. For example, gridlock errors among different surveillance platforms can cause a single target to appear as two or more targets; the WC must assess this possibility in his option generation and selection. Similarly any delay in communicating a weapons-release command to the actual weapon release will influence the tactical decision process.

The PEM must be by necessity a detailed model, and even in a single warfare area the amount of tactical information may saturate a single WC; it is for this reason that a major WC is augmented by his staff to avoid information overload. This has led Boettcher and Levis[15] to develop information-theoretic models of human decisions.

The normative models that represent the PEM of a particular WC are not necessarily complex. The reader should keep in mind that the only reason that these models are necessary is to understand the interactions between the physical C^3 system and the commanders within the organization. The decision variables of the commanders are relatively simple at each instant of time since they relate to:

   (a) positions and motions of assets,
   (b) turning on or off sensors and other surveillance assets,
   (c) communication with other commanders,
   (d) timing of above decisions.

The process by which decisions are arrived at are complex; but these are carried out in the mind of the commander who is modeled as an expert in this decision-making task. The only degradation in his decision-making is due to the C^3 system that may give him too much or too little information, and only partial information about the enemy's assets and their motion.

Therefore, the PEM's must be stochastic and dynamic models, and must have a feedback mechanism. Thus, they cannot be purely prescriptive and open-loop (e.g., modification of linear programming algorithms); rather they must represent adequately the cybernetic feedback process of Figure 1*. The successful human control and decision models developed to date [7,13], indeed fall in this category. Thus the PEM's should *not* explain how a WC arrived at his decision; rather it should explain *how* the information generated by the C^3 system pertrubs his inherent tactical decision process.

Based upon private communication with Dr. A. H. Levis, Mr. J. G. Wohl, and Gen. Cushman, U.S. Army (ret.).

This is in philosophical agreement with the central hypothesis proposed by Simon in reference [16], p. 65: "A man, viewed as a behaving system, is quite simple. The apparent complexity of his behavior over time is largely a reflection of the complexity of the environment in which he finds himself." Indeed one could argue that a "good" C^3 system should not increase the complexity of the true tactical scenario through the information that it transmits to the WC's.

Needless to say, the very issue of developing normative models for human decision makers, and especially military commanders, is a very controversial one. Assuming for the time being that such models can be developed and validated, they can be used to represent the most probable decision of an ensemble of well trained commanders in the same area of expertise, together with a measure of variability (e.g., standard deviations). Thus, such normative models cannot be used to model any particular commander. Also, it is unlikely that such normative models can provide us with rules on how "great military geniuses" think.

One can engage in endless discussions on modeling human decision making. It is the opinion of the author that such arguments, although intellectually interesting, have little to do with the problem of properly designing the architecture of a C^3 system that is intended to support the decisions of many commanders. If we do develop adequate normative models, these will be very useful in carrying out engineering and cost/effectiveness tradeoffs on the C^3 system hardware and architectures. They can also be very useful in carrying out computer aided war games, in which the functions of low-level subordinate commanders are replaced by computer algorithms, thus allowing the war game to be played realistically at a global level with many fewer human resources. This would result in significant savings. (A complex naval war game, of the type played at the Naval War College in Newport, R.I., may involve 150 players; the computer is primarily used for bookkeeping purposes.)

Much novel research is needed to combine such normative models of individual commanders into a distributed team decision model. This will allow the study of alternate organization structure in conjunction with the architecture of the C^3 system, as a function of the tactical situation. Very little military doctrine has been developed in defining the command and organizational structure best suited to coordinate forces with a significant content of multipurpose platforms. If a relevant command organization theory could be developed, then it would be very useful in defining suitable adapative changes in command following an initial engagement in which some resources, including those associated with the C^3 system, were destroyed. It would also be useful for counter-C^3 studies, by isolating most vulnerable interfaces between the command organizational structure and the C^3 components. From a system-theoretic viewpoint, very little has been done along these lines. The methodology in reference [14] (which deals with issues of team decision making by a distributed set of "expert" decision makers, each with a limited model of the world) could be a useful first step in this class of complex problems.

Concluding Remarks

In this paper a brief discussion of some generic decision problems in military command and control organizations was presented. Also the need for normative models of both individual and team decision processes was discussed. One should understand how to most effectively structure the human organizational command and control structure in unison with the architecture of the C^3 system so as to best support the decisions of the commanders.

Acknowledgements

The author has greatly benefited by many stimulating discussions with several individuals over the past four years on the subject of C^3 systems. Special thanks are due to Professors R.R. Tenney and W.B. Davenport, Jr., and Dr. A.H. Levis (Massachusetts Institute of Technology), Drs. N.R. Sandell, Jr., and J.G. Wohl (ALPHATECH), Burlington, Massachusetts, Drs. S. Brodsky (Office of Naval Research), Drs. J. Lawson and D. Schutzer (NAVALEX), Professor J.M. Wozencraft Naval Postgraduate School, Dr. D. Brick (U.S. Air Force Electronic Systems Division), Rear Admiral G. Thomas (USN/ret.), and Rear Admiral W. Meyers III (USN/ret.) The views expressed in this paper are solely those of the author. ■

References

1. *Proceedings for Quantitative Assessment of Utility of Command and Control Systems*, Office of the Secretary of Defense with the cooperation of the MITRE Corp., C^3 Division, National Defense University, Ft. Leslie J. McNair, Washington, D.C.; available as report MTR-80W00025, The MITRE Corp., McLean, VA. (January 1980).

2. C^2 Time-Phased Plan (TAFFIS Master Plan), Tactical C^2 Archicture Group, Report MTR-3653 (Revision 1), The MITRE Corp., Bedford, MA (October 1978).

3. J.G. Wohl, "Decision Making for Tactical Air Battle Management," *Proceedings IEEE Conference on Decision and Control*, Albuquerque, N.M. (December 1980).

4. J.S. Lawson, Jr., "Command Control as a Process," *Proceedings IEEE Conference on Decision and Control*, Albuquerque, N.M. (December 1980).

5. "Command and Staff Action," *Publication FMFM 3-1*, Commanding General, Marine Corps Development and Education Command, Quantico, VA. (June 1979).

6. N.R. Sandell, Jr., G.S. Lauer, and L.C. Kramer, "Research Issues in Surveillance for C^3," *Proceedings IEEE Conference on Decision and Control*, Albuquerque, N.M. (December 1980).

7. D.L. Kleiman, P. Krishna-Rao, and A.R. Ephrath, "From OCM to ODM - An Optimal Decision Model of Human Task Sequencing Performance," *Proceedings 1980 International Conference on Cybernetics and Society*, Cambridge, MA (October 1980).

8. N.R. Sandell, Jr., P. Varaiya, M. Athans, and M.G. Safonov, "Survey of Decentralized Control Methods for Large Scale Systems," *IEEE Transactions on Automated Control*, Vol. AC-23, No. 2 (April 1978) 108-128.

9. N.R. Sandell, Jr., "Distributed Decision Making Processes in C^3," *Proceedings 3rd MIT/ONR Workshop on Distributed and Decision Systems Motivated by C^3 Problems*, Silver Springs, MD. (June 1980).

10. H.A. Simon, "On How to Decide What to Do," *The Bell Journal of Economics,* (1978) 494-507.

11. I.L. Janis and L. Mann, *Decision Making*, The Free Press, N.Y. (1977).

12. W.D. Rowe, *An Anatomy of Risk,* J. Wiley N.Y. (1977).

13. W.H. Levinson and S. Baron, "The Optimal Control Model: Status and Future Directions," *Proceedings 1980 International Conference on Cybernetics and Society*, Cambridge, MA. (October 1980).

14. R.R. Tenney, "Distributed Decision Making Using a Distributed Model," Report LIDS-TH-938, Laboratory for Information and Decision Systems, MIT, Cambridge, MA (September 1980).

15. K.L. Boettcher and A.H. Levis, "Modeling the Interacting Decision Maker with Bounded Rationality," Report LIDS-P-1110, MIT Laboratory for Information and Decision Systems, Cambridge, MA (July 1981).

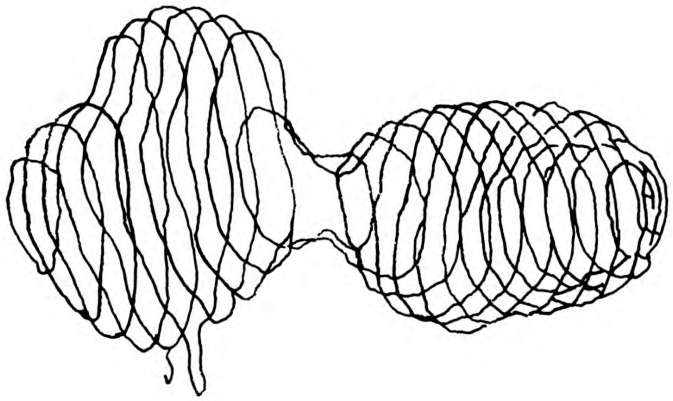
16. H.A. Simon, "The Sciences of the Artificial (Second Edition)," MIT Press, Cambridge, MA (1981).

Dr. Michael Athans is Professor of Systems Science Engineering at the Massachusetts Institute of Technology. From 1974-1981, he was Director of the M.I.T. Laboratory for Information and Decision Systems. He is the co-founder of ALPHATECH, Inc., Burlington, Massachusetts, where he is Chief Consultant and Chairman of the Board. His books and published papers span the areas of system, control, and estimation theory and its applications to the fields of defense, aerospace, and communication (C^3) systems. He is a Fellow of the Institute of Electrical and Electronic Engineering and a Fellow of the American Association for the Advancement of Science.

Orion I: Interactive Computer Graphics in Statistics

by

John Alan McDonald
Stanford University



Representation of the 15% contours of a three dimensional probability density formed by slicing the density by parallel planes.

The Orion I workstation is an experimental computer graphics system designed and built by Jerry Friedman and Werner Stuetzle at the Stanford Linear Accelerator Center (SLIC) in 1980-81[1] with the support of the Office of Naval Research, the U.S. Army Research Office, and the Department of Energy. It is being used to explore applications of computer graphics—especially interactive motion graphics—for statistical data analysis. [2,3]

The methods being developed on Orion I are among the first in the rapidly growing field of computer intensive statistical methods. These new methods, stimulated by the evolution of computing technology, are creating a new paradigm for statistical research; graphical and other computer intensive methods have a strong engineering flavor, in contrast to the more rigorous mathematical nature of traditional statistics.

Traditional Statistical Graphics

Graphical methods are among the oldest in statistics. Examples of graphical representation of quantitative data are known from 3000 B.C.[4] Any elementary text in statistics will discuss in detail the standard statistical graphics, such as the histogram and the scatterplot. However, until recently, the development of new graphical methods has been outside the mainstream of statistical research.

This is, in part, due to limitations of technology. The potential in pencil and paper graphics was thoroughly exploited early in the nineteenth century; most of the standard techniques (histogram, scatterplot, etc.) were in use by the 1830's. Statisitics acquired a real identity separate from mathematics in the twentieth century, especially in the period after World War II. Thus the successful graphical methods were

well entrenched by the time statistics emerged as a field; these established methods left little room for improvement. Only unusually creative researchers were able to suggest useful additions to a statistician's graphical tool kit that could be used with pencil and paper.

Exploratory Data Analysis

One of these unusually creative statisticians was John W. Tukey, who, in the 1960's, developed a family of both graphical and non-graphical statistical methods referred to collectively as exploratory data analysis. [5,6,7] The underlying philosophy of exploratory data analysis is very different from that of classical statistics.

Classical statistics emphasizes mathematical analysis. Problems that arise directly from real data are usually impossible to analyze mathematically, so only simplified, idealized problems are considered. Simplifying assumptions are made so that it is possible to determine rigorously the "best" solution. The implicit hope is that the "best" solution to an idealized problem will be a reasonable solution in more realistic circumstances.

The collection of methods characterized as "exploratory data analysis" approaches statistical problems from the opposite direction. These methods are not optimal in any precise sense. Their use is justified by the fact that they work in practice with real data.

Description and Inference

One reason that graphical methods are so different from classical statistics is that they address different problems.

Using graphical methods in statistics means looking at pictures of data; the purpose of graphical methods is primarily descriptive. Statistical description is concerned with discovering and summarizing features of a particular data set.

Classical statistics concentrates on mathematical analysis of formal methods for inference. An inference is a generalization from a given data set to some larger population (real or hypothetical) from which the data set is presumed to be a sample.

Formal inference is necessarily based on mathemetical models. Description is neccessary, first, to help construct a model, and later, to check to see if a chosen model is really appropriate.

Prim-9

The introduction of modern computer graphics to statistics began at SLAC, in 1972, when Mary Ann Fisherkeller, Jerome H. Friedman, and John W. Tukey developed a system called Prim-9.[8] Prim-9 pioneered the use of real-time motion in statistical graphics, most notably to display three-dimensional scatterplots.

The tradional statistical graphics, the histogram and the scatterplot, are excellent pictures of data in one and two dimensions. However, it is very difficult to see relationships among three variables by looking at histograms and scatterplots.

With Prim-9, it was possible to see the three-dimensional shape of a data set using real-time rotation. If a point cloud is subjected to small rotations, and the rotated point cloud is projected on the display screen quickly enough for the motion to seem smooth; a viewer gets a convincing and accurate perception of the shape of the point cloud as a three-dimensional object.

The three-dimensional scatterplot is a simple idea. It is an obvious and natural extension of the standard, two-dimensional scatterplot. However, this simple first step in new statistical graphics gives statisticians a powerful new tool. It is trivial to see and understand structure in data that is very difficult or impossible to see in any number of two-dimensional pictures.

On Prim-9 in 1972, the display of three-dimensional scatterplots stretched the limits of the available computing power. The computation was done on SLAC's large, central, mainframe computer, and IBM 360/91. The total cost for hardware in Prim-9—the 360/91, the graphics device, and a high speed channel to transfer data from the 360/91 to the graphics device—was millions of dollars.

Prim-9 made a strong positive impression when it first appeared, but it did not have much immediate impact on the practice of statistics. The expense of the hardware made it impractical for such systems to be developed except at special institutions such as SLAC.

Successor to Prim-9 were developed by Peter Huber and his students in Switzerland (Prim-ETH) in 1978 and at Harvard (Prim-H) in 1979-1980. They were (and are —Prim-H is in use and still under development) similar to Prim-9 in basic computing technology and graphical capabilities, though perhaps one tenth as expensive.

Orion I

The latest descendant of Prim-9, Orion I, represents a substantial advance over the Prim systems, at least in hardware. Orion I is both more powerful and much less expensive that the Prim systems because it takes advantage of advances in microprocessor and computer graphics technology in the last decade.

Orion I cost approximately $60,000 in 1981. It has much more computing power than the Prim systems. Orion I uses a color graphics device, as opposed to the strictly black and white displays in the Prim systems. Color is essential for several new methods that are used to look at pictures of more than three-dimensional data.

An important implication of Orion I is that such graphics systems will soon be widely available. The price of systems like Orion I is dropping rapidly; a better system could be build today for $25,000, less than half of what it cost eighteen months ago. It is likely that in a few years such graphics systems will be in the price range of personal computers.

Applications of Orion I

The methods we are developing on Orion I emphasize real-time motion and interaction. So it is difficult to do them justice in a verbal description. To compensate for this, we have made two movies, which are available for viewing.[9,10]

Almost any statistical problem will benefit from the addition of interactive graphics. An interactive graphics system makes it possible to bring human intelligence to bear directly on the problem. Humans have powerful abilities for pattern recognition that are so far not equaled by computers. Further, a human can adapt the analysis to unforeseen features of a data set, using judgement and a knowledge of the context of the problem from which the data arose. These points are clearly illustrated in our two films.

Orion I is being used to develop interactive graphical methods for many statistical problems, including: general multivariate analysis, regression, classification, clustering, time series, signal processing, seismic data (multivariate point processes), and analysis of digitized x-rays (interactive image processing).

The first problem, multivariate analysis, refers to the analysis of data sets in which there are many variables. A typical data set will have 100-1000 observations, and for each observation we will have values of perhaps 4-20 variables.

The basic graphical problem posed is this: how can we draw pictures that show us relationships involving 4 or more variables at once? If the data set had only two variables we would just look at an ordinary scatterplot. With three variables we can use a three-dimensional scatterplot. Looking at three-dimensional pictures is a very natural activity for human beings because we live in a three-dimensional world. However, drawing pictures that show four or more variables and are also easily understood is very difficult.

We are pursuing two basic approaches to more than three-dimensional data. The first approach is to find ways of effectively representing as many variables a possible at once in a single picture. For example, we take a three-dimensional scatterplot and color each point according to the value of an additional variable. This gives us a picture that shows some relationships among the four variables. We can add other features to the points in the scatterplot, such as size, shape, or orientation, to code values of more variables. Each additional feature will, unfortunately, add less information to the picture and will make the picture harder to interpret.

Another method puts several different scatterplots on the screen, and uses interactive coloring schemes to show relationships between the contents of the scatterplots.

The second approach to many-dimensional data is called Projection Pursuit.[11,12,13] The basic idea of Projection Pursuit is to compress (or project) the information in the many-dimensional data into a two- or three-dimensional picture, which is easy to look at. We choose the projection to try to capture any interesting structure in the many-dimensional data. ■

References

1. Friedman, J.H. and Stuetzle, W., "Hardware for Kinematic Statistical Graphics," Tech. Rep. Orion 003, Dept. of Statistics, Stanford University, (1982b).

2. Friedman, J.H., McDonald, J.A., and Stuetzle, W., "An Introduction Real Time Graphical Techniques for Analyzing Multivariate Data," *Proceeding 3rd Annual Conference of N.C.G.A.* (1982a).

3. McDonald, J.A., "Interactive Graphics for Data Analysis," Ph.D. thesis, Department of Statistics, Stanford University. Available as Technical Report Orion 011, Department of Statistics, Stanford, (1982).

4. Beniger, James R. and Robyn, Dorothy L., "Quantitative Graphics in Statisitics: A Brief History," *The American Statistician, 32* (1978).

5. Tukey, John W., "The Future of Data Analysis," *Annals of Mathematical Statistics, 32,* (1962).

6. Tukey, J.W., *Exploratory Data Analysis,* Addison-Wesley (1978).

7. Mosteller, F. and Tukey, J.W., *Data Analysis and Regression,* Addison-Wesley (1977).

8. Fisherkeller, M.A., Friedman, J.H., and Tukey, J.W., "Prim-9, An Interactive Multidimensional Data Display and Analysis System," *Proceeding 4th International Congress for Stereology* (1975).

9. Friedman, J.H., McDonald, J.A., and Stuetzle, W., (1982b) "Exploring Data with the Orion I Workstation," a 25 minute 16mm sound film, may be borrowed from: J.H. Friedman, Bin-88, Computation Research Group, Stanford Linear Accelerator Center, P.O. Box 4349, Stanford, CA 94305

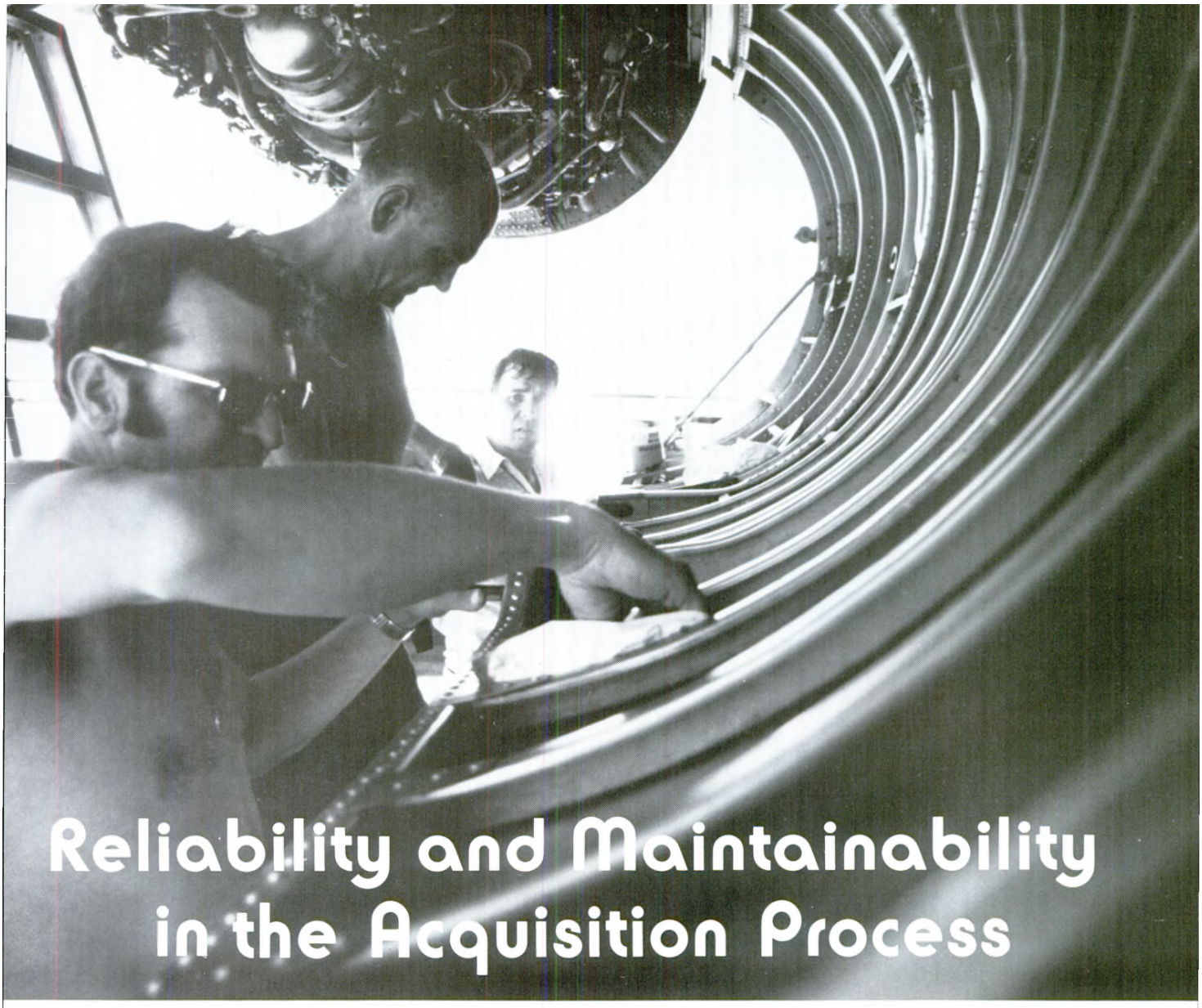
10. Friedman, J.H., McDonald, J.A., and Stuetzle, W., (1982c) "Projection Pursuit Regression with the Orion I Workstation," a 20 minute 16mm sound film, may be borrowed from: J.H. Friedman, Bin-88, Computation Research Group, Stanford Linear Acclerator Center, P.O. Box 4349, Stanford, CA 94305

11. Friedman, J.H. and Stuetzle, W., "Projection Pursuit Regression," *JASA, 76.* (1981).

12. Friedman, J.H. and Stuetzle, W., (1982a) "Projection Pursuit Methods for Data Analysis," in *Modern Data Analysis,* Launer, Robert L. and Siegel, Andrew F., eds., Academic Press, 1982.

13. Friedman, J.H. and Tukey, J.W., "A Projection Pursuit Algorithm for Exploratory Data Analysis," *IEEE Trans. Comput. C-23* (1974).

Dr. John Alan McDonald is the National Science Foundation Mathematical Sciences Postdoctoral Research Fellow at Stanford. He is currently doing research on graphical and computational methods for data analysis. For his Ph.D. thesis, Dr. McDonald developed new methods for data analysis using the Orion I workstation.



Reliability and Maintainability in the Acquisition Process

by

Dr. Thomas C. Varley
Office of Naval Research, Arlington

Introduction

There is an increasing concern as to the high cost of producing and maintaining weapon systems. Current trends indicate that these forces are all moving in the wrong direction. Complexity of systems are increasing, without improved maintainability considerations to lower the operating costs; unit production costs are increasing with ever increasing maintenance personnel skill levels; availability of skilled personnel is decreasing through lower retention of mid-career personnel, while manpower availability is decreasing due to past national demographics. Added to these individual trends is the interaction and relationships they combine to produce. We have lack of in personnel and equipment, and we have increasing operating costs which reduce our capability to create new assets under constrained budgets. We have reduced assets, but they are costing us more. Where does

research in reliability and maintainability (R&M) ir teract with these trends, and what can they do to reverse these trends?

The Trends

The role of R&M in the acquisition process is very broad and can have a significant impact on the long run operational capability of naval forces. One generally thinks of R&M considerations as a method for improving equipment operation, but because of the interactions between all components of the system, improving equipment/system R&M has far greater implications. Consider some of the following statistics: The availability of manpower for the remainder of this century will be 10-25 percent lower than we had in 1976. Figure 1 displays the percentage change over that 20-year time period. It is interesting that the same general picture is true for females, as well as males, in the 17-19 year age range. With fewer personnel available and a greater demand being placed on this population by industry and academic institutions, the Navy must develop ways to decrease its personnel requirements. R&M has potential for doing this.

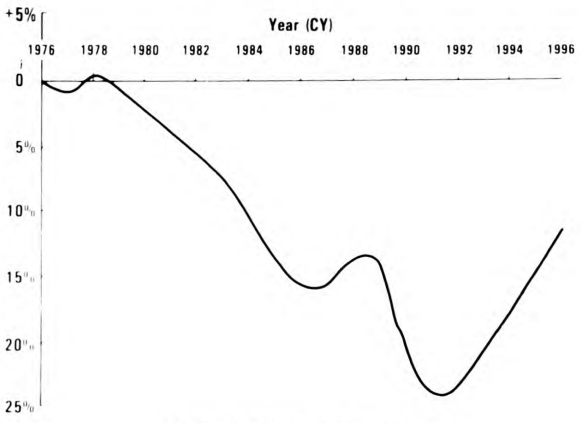
Figure 2 describes a second consideration—the decreasing number of naval aircraft purchases over the last 10 years and their average costs. In 1970, we purchased 134 aircraft for 442 million dollars, while in 1980, we purchased 39 aircraft for 1,365 million dollars—a 400 percent decrease in aircraft; yet, over a 300 percent increase in procurement costs. Associated with this is an annual maintenance cost that averages about 6 percent of its acquisition cost per year per aircraft, a second area where R&M can play a significant role.

The increasing complexity of aircraft is well illustrated by technical manual trends. Figure 3 shows the growth of technical manual size over the last 10 years. This growth is unavoidable in order to provide good documentation, but the control, updating, and training required is a considerable burden on naval resources. This is another area where R&M can play a significant role in alleviating present difficulties.

Weapons Life Cycle

So far, we have been describing some particular areas where improving R&M could be beneficial. How do we take advantage of these potential benefits? First, we have to understand the long life cycle of a major system before we can understand the role R&M plays in the acquisition cycle. Figure 4 defines the various phases of the acquisition process from definition of the need through deployment of the system. To understand the time phase in Figure 4, the period to deployment is between 10-14 years while the initial operating life varies from 15-30 years for aircraft and ships, respectively. Together, the acquisition cycle and the deployment or operating life represent the life cycle of the system. The critical decision point is Milestone III which is the decision to produce the

Percentage Change in 17 to 19-Year-Old U.S. Male Population, 1976-1996



Source: Department of Commerce

Figure 1. *Percentage change in 17 to 19-year-old, U.S. male population, 1976-1996. Courtesy of the Department of Commerce.*

Declining Navy Aircraft Purchases

Fiscal Year	Fighters and Attack Planes Bought	Average Cost of Each Plane, Millions $
1970	134	3.3
1971	110	8.0
1972	114	7.5
1973	147	5.5
1974	149	6.4
1975	92	9.2
1976	77	10.6
Transition Quarter	15	7.3
1977	93	9.7
1978	68	14.1
1979	69	20.2
1980	39	35.0
Total	1107	

Figure 2. *Declining Navy aircraft purchases.*

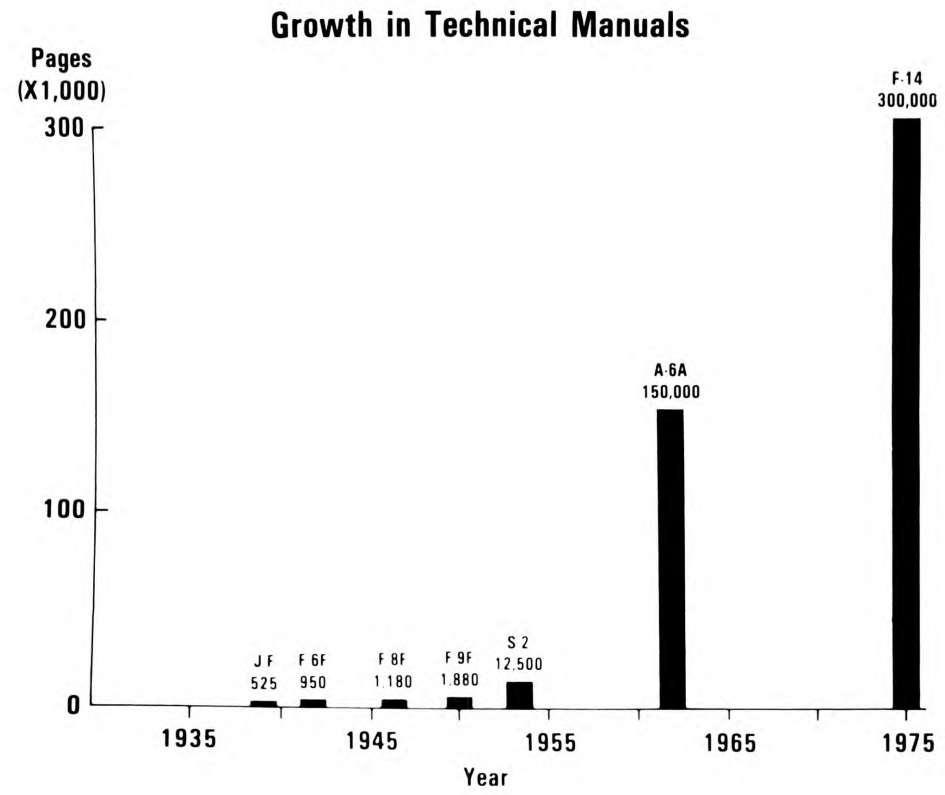


Figure 3. *Growth in technical manuals.*

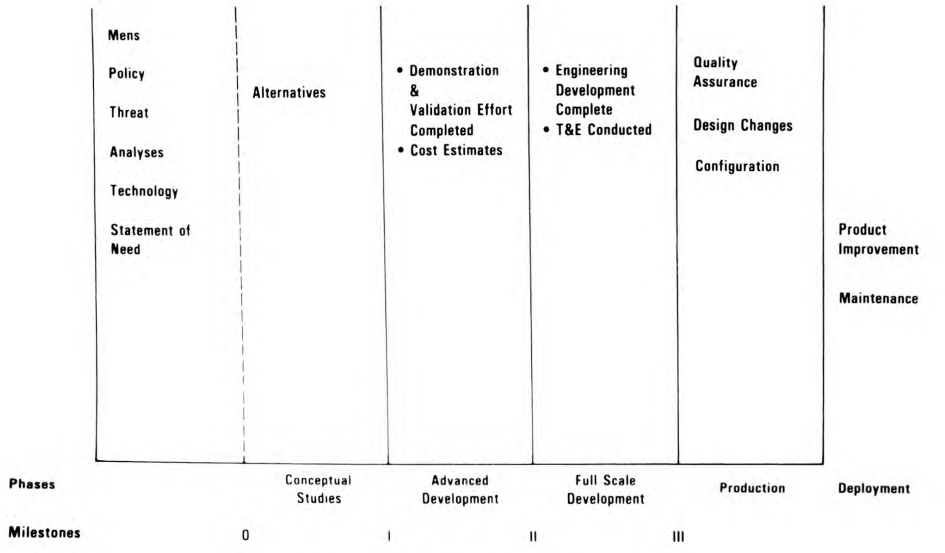


Figure 4. *Phases of the acquisition process.*

system. Once the decision is made for production, future resources need to be allocated for its operation. These resources are of a time-phased nature and can be considered as "flows" through the system over a multi-year period.

Acquisition Funds Flow

Figure 4 is one attempt to describe the Navy budget in a very aggregated form with three major categories: procurement funds, research and development (R&D) funds, and something called "cost of operating current assets." These categories represent the flow of funds into the primary functional areas of Navy activity. The flow of funds through the procurement chain result in accumulated assets. These assets are the results of the acquisition process, which has a major decision node on the determination of force characteristics. R&D funds flow two ways; some flow into the force characteristics decision node and are used to develop various alternatives-characteristics, and to produce prototype systems; and some flow into current systems assets through modernization programs inproving their efficiency and effectiveness. The cost of operating current assets flows into net ownership costs which provide the resources to operate the fleet and shore establishment. The level of funds flowing into operations is a major determinant of the level of naval readiness; that is, if fewer funds flow than are required, readiness is lessened.

Readiness resources which are needed can be defined as the amount of funds required to maintain operational commitments including adequate supply of technical spare parts; of personnel skills and grades for manning; of maintenance funds for ship and aircraft overhauls; of training in formal schools, tactics and fleet exercises.

Figure 6 is a more detailed representation of the R&D funds flow interactions in the cost of current assets area. With an increase or decrease in force assets, operational support costs increase or decrease, but not proportionally. The relative costs of maintenance, operating schedules, and manpower affect the degree of change in operation support costs as these new assets are acquired. In addition, R&D funds are expanded for modernization which should reduce the operational costs.

Each of these areas—maintenance, manpower and operations— are candidates for research in R&M. They are dependent on each other. Improvements in one can have either positive or negative effects on the others. Care must be taken to understand what the overall influence is on the entire weapon systems life. Serious investments to reduce the life cycle cost of weapons can have a significant impact on the future composition of naval forces. Given Figures 5 and 6, consider the following scenario. As the Navy receives its budget, a determination is made to fund the current assets account at 100 percent of requirements; the remainder of the budget is apportioned at 20 percent for R&D and the remainder to procurement. If we

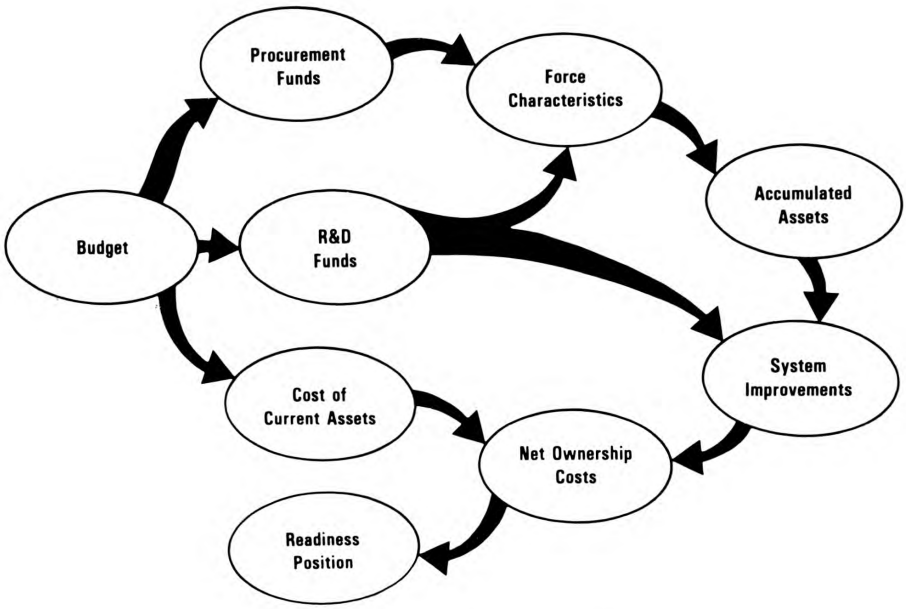


Figure 5. Aggregated budget flow.

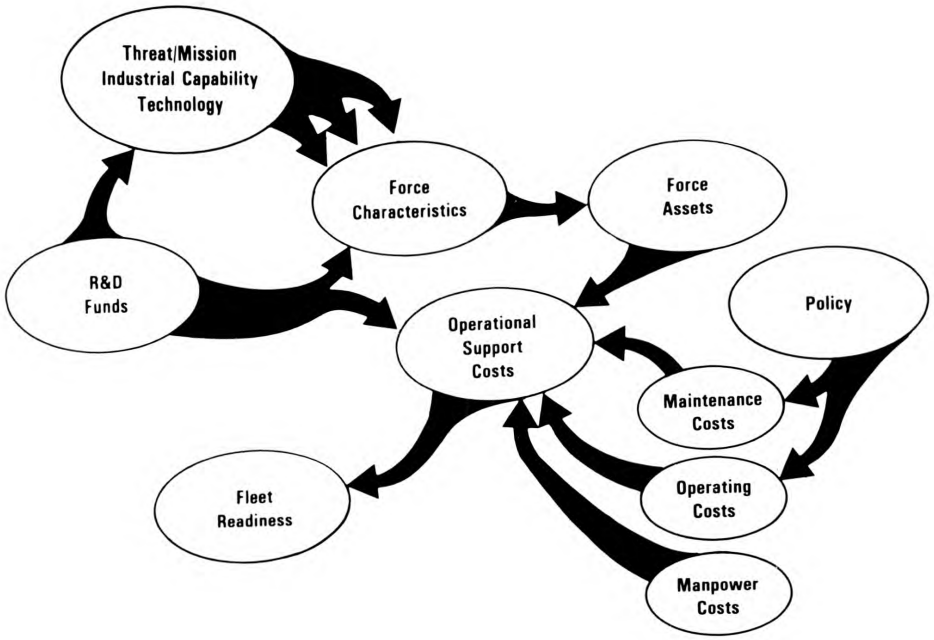


Figure 6. R&D funds flow.

translate this to the FY-81 Navy budget, we would have something on the order of 23 billion on current operations; 4 billion in R&D and 16 billion on procurement out of a 43 billion dollar total Navy budget. A further breakdown of current operations would include 16 billion for operations and maintenance and 7 billion for personnel. If over the next 5 years the Navy budget increases by 7 percent and by improving R&M related activities, the cost of current operations could be held to a 5 percent increase, the distribution for FY-86 would be 29 billion for current operations; 6 billion for R&D, and 23 billion for procurement, out of a 58 billion dollar total Navy budget. That two percent decrease in current operations' requirements costs increases the procurement account by more than 6 percent. Since current operations costs are such a large part of the budget, a small percentage gain can have a large influence on the remaining components of the budget. While the example shown indicated a decrease in current operations, it is also true that an increase in current operations costs has a major effect in reducing the ability to procure new assets. In fact, the latter case is our current difficulty, in that the funds required for current operations has been increasing at a rate that is seriously eroding the asset base.

The Acquisition Cycle

If we want to interact with the acquisition cycle, where are we most effective? Figure 7 describes the relationship between product innovation and process innovation. Most will agree that product innovation is early in the cycle, that technology improvement spawns new products which then are considered as alternatives to system design. As you move towards advanced development where demonstrations, validations and cost estimating are completed, new product alternatives decrease quickly as full-scale development begins and process innovation peaks. There is considerable concern about process innovation or producibility. It seems that actual practice places that process innovation later in the cycle, which tends to lengthen the full-scale development period and causes process innovation to peak at the beginning of production. Another way of stating the concern is that process innovation is of little concern to the manufacturers until after full-scale development has been completed and DSARC III (the decision to move to full production) has been approved.

The ability to influence system characteristics must be accomplished early in the cycle. These new products are the results of both hardware as well as methodology development coming from the technology base. They are evaluated and integrated into the system design during the conceptual effort.

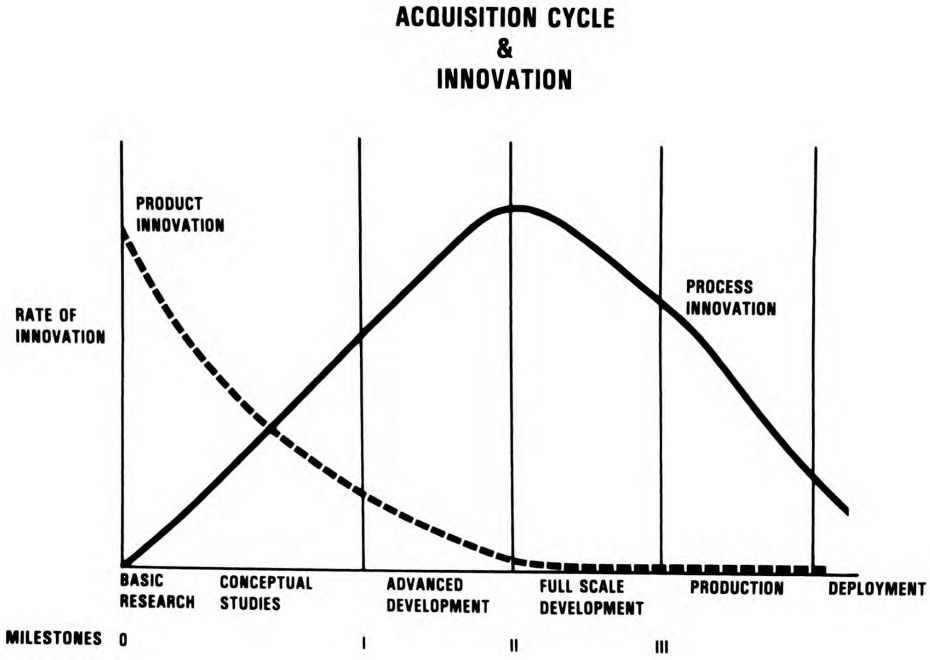


Figure 7. Acquisition cycle and innovation.

While hardware development is important for meeting future threats, methodology development has a significant role in attempting to balance the conflicts and constraints into the best overall design and life cycle operation of the system.

R&M Areas of Need

Figure 8 displays several of these conflicts and constraints over three primary functional areas for system characteristics. The logistics area represents the life of the system and its ownership costs after procurement. Manpower is also a lifetime cost and is separated from the more traditional logistics considerations because of its increasing importance in manning systems and the demographics we discussed earlier. Aquisition represents those actions and activities that are required to bring the system into being. As one reflects on the three functional areas, it is obvious that they are not independent and actions in one area can influence the other areas.

The particular areas of need are an indication of the R&M activity which can influence each of the functional areas, and as a rough scale. A "P" relates to a primary influence while an "S" suggests a secondary influence. Of those needs displayed, availability is the only one considered as having secondary influence for manpower and acquisition, yet availability is not independent of reliability and maintainability. This is especially true when you consider operational availability as opposed to supply availability. Similar cases can be described for each of the other areas, which infer that there are many constraints placed on the "system designer" before the best affordable system can be produced. There are design-to-costs and life-cycle considerations, producibility and risk-cost tradeoffs and manpower constraints. Consider warranties; if a new methodology can be developed to improve our ability to evaluate future cost avoidance due to increased reliability/availability early in the acquisition phase (before production), direct savings could be realized in the logistics and manpower areas during system operating life. Such a methodology should contain concepts for rewarding both the Navy (through reduced life-cycle cost) and the manufacturers (through improved financial incentives during production). This would include a reduction in spare parts support; in organizational, intermediate and depot maintenance; and in training and support personnel at all levels. This is especially important for personnel when you consider that it takes about three inductees to produce one E-6 electronics technician.

Reliability and maintainability considerations are important components in the total acquisition process. They and their related "ilities" that comprise the assurance sciences disciplines, are necessary to provide the capability for developing, testing, evaluating and

FUNCTIONAL AREAS

AREAS OF NEED	LOGISTICS	MANPOWER	ACQUISITION
R&M	P	P	P
Warranties	P	P	P
Quality Control	P	P	P
Availability	P	S	S
T&E	S	S	P
L.C.C.	P	P	P
P3I	P	S	P

Figure 8.

determining the appropriate components and the organization of those components into affordable systems.

The increased complexity of weapon systems with their highly integrated circuits; their design for practical degradation; the need for fault detection and fault isolation schemes and applications for quality control, optimal inspection scheduling and non-destructive testing require increased coordination between the Navy, academic and industry to understand the conflicts and constraints and to develop producible and affordable systems for our National defense missions.

■

Dr. Thomas Varley is Project Director INTRA Type Tactical Development and Evaluation at the Office of Naval Research. His publications and research have been primarily in the field of operation research and statistics.

research notes

SRO Program Initiated in Inverse Problems

One of ONR's special programs is the Selected Research Opportunity (SRO) Program which selects for special funding research areas with great research potential requiring interdisciplinary approaches. There was a SRO program in inverse problems begun in 1983 with awards made to research groups at Cornell University, Iowa State University, and the University of Denver. These three groups were selected from approximately fifty groups that submitted proposals.

Inverse problems have long been of interest to scientists, engineers and to the Navy. However, recent attention to application in nondestructive evaluation, oil exploration, and medical tomography, as well as the more established applications of radar and sonar imaging has led to a greater awareness of and participation in this research field. In addition, recent and anticipated progress in computational capability promises the ability to solve problems of increasing difficulty. ONR decided that now is the appropriate time to expand research in this area to impact on such naval needs as: (1) understanding the ocean bottom structure for acoustic and seismic wave propagation; (2) improving reliability via nondestructive testing procedures; (3) designing optimum waveform and detection algorithms for radar applications; and (4) improving sonar techniques for detection and classification.

Although there is evidence of recent engineering progress in some inverse problem areas, such as computer-aided tomography and non-destructive evaluation, this progress has been in problem areas in which one has a highly controlled environment. This permits the achievement of a superior data set both in terms of quality and quantity. In other areas, however, such as radar classification where the target is uncooperative, data sets are much more limited. It therefore becomes essential to develop insights into the design of systems for collecting and analyzing data. Limitations on uniqueness and accuracy imposed on the solution by the presence of incompleted data must also be investigated.

It is the purpose of this program to develop a better understanding of the fundamentals of inverse problems and limitations imposed by data collection. There are three cases to be considered: (1) too much data (the redundant case) for which it is important to understand how to most effectively use extra data; (2) insufficient data, for which one wants to characterize the class of permissible solutions; and (3) sufficient, but not redundant data, for which one is interested in the stability of the solution obtained. Progress is especially needed to develop approximate and qualitative results which will lay the foundation for solving two- and three- dimensional problems. ■

(Charles J. Holland, ONR)

Seminal Work on Quantization Published After 19 Years

The Institute of Electrical and Electronic Engineers Transactions on Information Theory for March 1982 was issued in two parts. Part 1 was a Special Issue on Quantization. The editorial page for the special issue highlighted Dr. Paul Zador's work as one of the three foci on which the issue is based.

Reference is made to Zador's 1966 Bell Laboratories Memorandum (#18). This is based on his Ph.D dissertation which was published as Stanford Technical Report 92, December 1963, under an Office of Naval Research contract. Zador wrote this under Dr. Herbert Solomon but unfortunately he did not submit it for publication.

After he left Bell Labs, all that existed in addition to the thesis was the unpublished Bell Labs Memorandum of February 1966. When Dr. Neil Sloane of Bell Labs visited Stanford a year or two ago he mentioned Zador's work at Bell Labs in a seminar talk and said that he was trying to trace Zador but learned that the Electrical Engineering Department had no records of him! Dr. Solomon told him that Zador could be found at the Insurance Institute for Highway Safety in Washington, D.C., and Sloane said that he was going to urge Zador to send in a version of his work for the IEEE Special Issue.

Thus, some 19 years had passed before Zador's work made it to print. However, it had become a classic technical report and has served as a basis for many subsequent developments in quantization and coding. The paper, entitled "Asymptotic Quantization Error of Continuous Signals and the Quantization Dimension," was published in *IEEE Transactions on Information Theory,* p. 139 to 149. ■

(Edward Wegman, ONR)

Optimization Applied to Sonar Design

As a result of an Office of Naval Research sponsored workshop on beamforming, a significant application of modern optimization theory to a problem of current naval importance will be attempted. The workshop was held in April 1982. Its purpose was to determine whether the state-of-the-art of nonlinear optimization was sufficiently advanced to have an impact on the very large complex models that arise in sonar design.

One of the mathematicians invited to the workshop, Professor Leon Lasdon, University of Texas at Austin, presented the results of approximately 10 years of ONR funding. Under this funding, Professor Lasdon has developed computer codes using several algorithms, each designed for special instances of nonlinear optimization problems. As a result, ONR's Advanced Conformal Submarines Acoustic Sensors project has a contract with Professor Lasdon to assist in model development and apply these optimization techniques to problems of sonar design.

In addition, ONR has played a small but significant role in the development of simi-infinite programming in this country. This relatively new field analyzes optimization problems where the number of constraints is infinite. With additional theoretical development, there appear to be important applications of semi-infinite programming to sonar design. As a result of the workshop, ONR is examining these potential applications with the intention of including the required basic research in its core program. ■

(Neal Glassman, ONR)

Significant New Book on Game Theory

The Massachusetts Institute of Technology Press recently published *Game Theory in the Social Sciences* by Professor Martin Shubik of Yale University. On the dust jacket, Anatol Rapaport (who is Director of the Institute for Advanced Studies in Vienna) calls it a lucid exposition. . . bound to stimulate. . . creative applications of game theory in exploring . . . reality. Robert Aumann, a leading mathematical game theorist, writes that it is "what the game theory world has been waiting for, for the last 25 years. . . outlines methods and tools of this exciting and dynamic field in a way that is comprehensible. . . to those with. . . little mathematical training. It is bound to become a classic in a short time." Professor Shubik began his long association with the Office of Naval Research in the 1950's when he was a student of the late Professor Oskar Morgenstern, coauthor with John Von Neumann of *Theory of Games and Economic Behavior,* the book that founded game theory. ■

(J.R. Simpson, ONR)

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
ARLINGTON, VA. 22217

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

THIRD-CLASS MAIL
POSTAGE & FEES PAID
USN
PERMIT No. GS-9