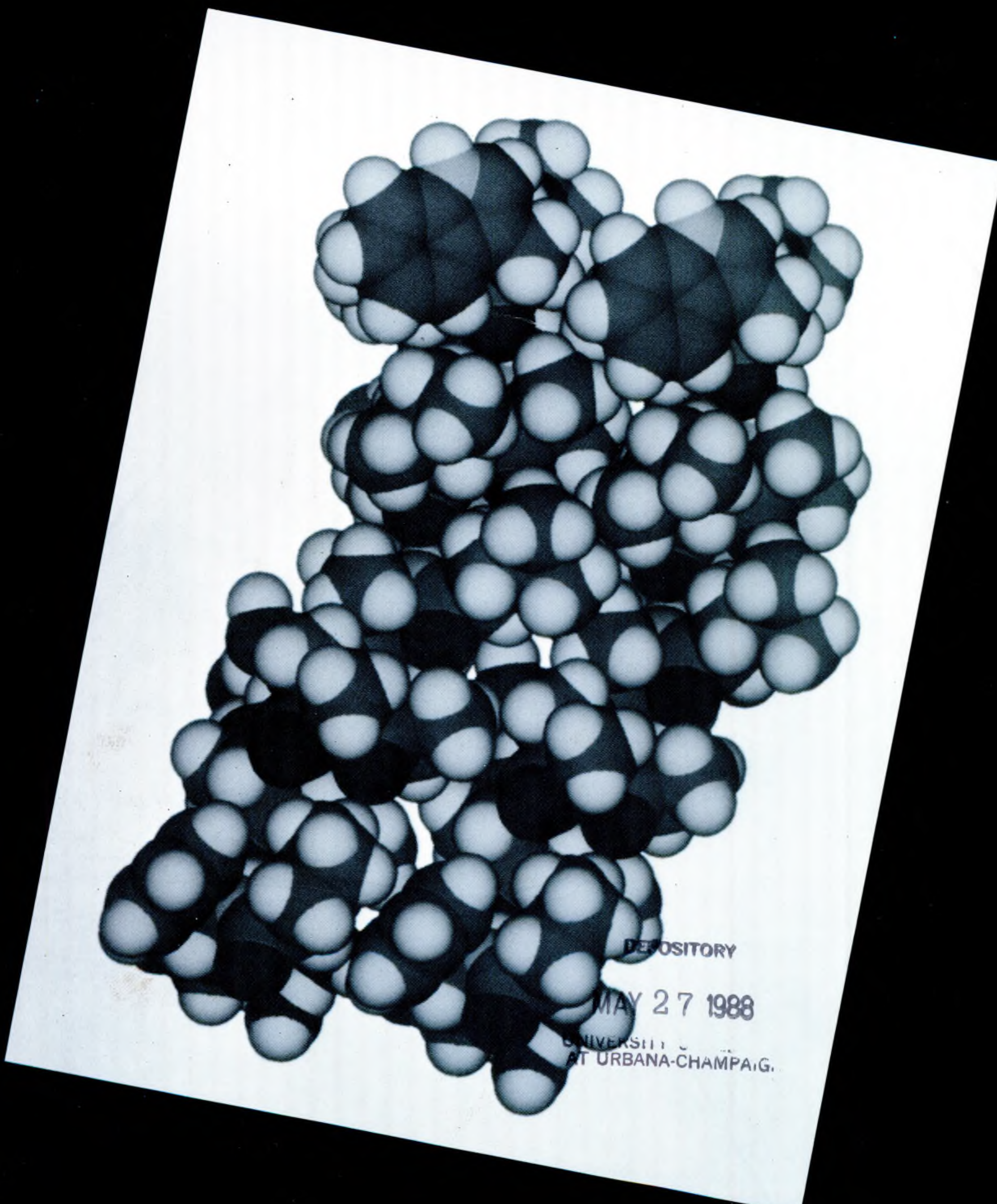
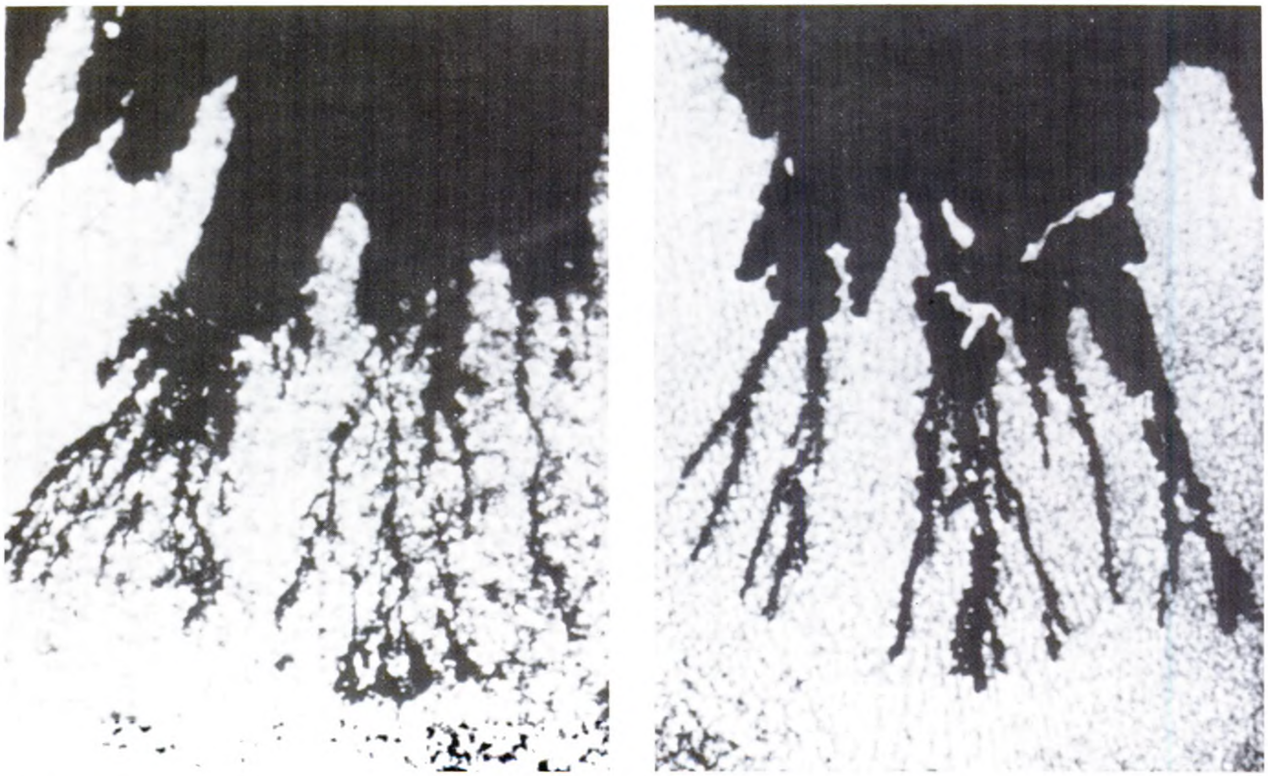
Naval Research Reviews

Office of Naval Research
Four / 1987
Vol XXXIX



Stress Corrosion Cracking



Photomicrographs (magnification about 500x) of a base metal and a weld metal illustrate the typical branching characteristics of stress corrosion cracking in high-strength steel. It occurs by the simultaneous action of stress and the operating environment. This failure mode is of particular concern with respect to new, high-strength structural materials targeted for use in future Navy platforms.

ONR supported research over the last two decades has identified the mechanism of environmentally induced cracking in high strength steels as being caused by the presence of hydrogen, has developed test techniques for its detection, and has provided the fundamental basis for engineering evaluations. These evaluations have led to the issuance of specifications and standards by NASA, U.S. Air Force, and the National Association of Corrosion Engineers, based on restricting the strength level of materials. To date, these new specifications have been effective in preventing envioronmental cracking failures in service.

Naval Research Reviews

Office of Naval Research
Four/1987
Vol XXXIX

Articles

Departments

About the Cover

Three dimensional model of a 16-residue peptide, related to a naturally occurring antibiotic, whose structure was determined by X-ray diffraction analysis. Two identical molecules are displayed, nestled together as they occur in the crystal. Each molecule, shown vertically, is in the form of a curved helix. See article on page 2.

Naval Research Reviews publishes articles about research conducted by the laboratories and contractors of the Office of Naval Research and describes important naval experimental activities. Manuscripts submitted for publication, correspondence concerning prospective articles, and changes of address, should be directed to Code 732, Office of Naval Research, Arlington, Va. 22217-5000. 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 Regulations NAVSO P-35.

FROM X-RAY REFLECTIONS

by Jerome Karle and Isabella L. Karle
Naval Research Laboratory

Dr. Jerome Karle of the Naval Research Laboratory (NRL), Washington, D.C., was awarded the 1985 Nobel Prize in Chemistry for his work in developing direct methods for the determination of crystal structures by X-ray diffraction analysis. Dr. Karle shared the award with Dr. Herbert A. Hauptman, a former scientist at NRL working with Dr. Karle and now with the Medical Foundation of Buffalo, New York. In 1950, Karle and Hauptman published a fundamental research paper entitled, "The Phases and Magnitudes of the Structure Factors." The paper formed the basis for further research in structure determination by the "direct method" in X-ray crystallography.

Karle's research has been concerned with diffraction theory and its applications to the determination of atomic arrangements in various states of aggregation, gases, liquids, amorphous solids, fibers and crystalline molecules and macromolecules. This research has resulted in new techniques for structure determination with a broad variety of application. The subject of the paper presented here is the three-dimensional structures (atomic arrangements) of molecules, as obtained from X-ray diffraction applied to the crystalline state.

With the advent of modern technology such as computers, electronics, instrumentation and sophisticated programming, it has become possible to mount a small single crystal of an unknown substance on an automatic X-ray diffractometer and, with minimal human intervention, see the complete geometric arrangement of atoms in the molecule displayed on a video screen. Depending upon the size of the molecule the process may take as short a time as one day. Automatic structure determination is feasible today for very many small and medium sized structures, particularly if the crystal contains a center of symmetry. Larger and more complex structures require the special expertise of professional crystallographers.

The present ease of structure determination and even the possibility of deriving a structure from X-ray data, stems from the fundamental theoretical and practical advances made in solving the phase problem. When a collimated beam of X-rays of a particular wave-length impinges upon a single crystal, and that crystal is rotated about various axes, different sets of scattering planes acting like a ruled grating come into position for the X-ray beam to be diffracted. The intensity and angular orientation of hundreds or thousands of reflections are measured and recorded by an automated diffractometer (or, formerly, on photographs). The phases associated with the diffracted beams, however, with rare exception, cannot be measured experimentally.

TO A COMPLETE MOLECULAR STRUCTURE

Values for both the intensity and phase for each reflection are needed to calculate the three-dimensional electron density of a unit cell of a crystal. Ideally the electron density is zero everywhere except at the positions where atoms occur, thus giving a direct measure of the x, y and z coordinates of each atom. Furthermore, the density representing atoms is proportional to their atomic weights, permitting an identification of the atomic species. The electron density function ρ is represented by

$$\varrho(x,y,z) = \frac{1}{V} \sum_{\substack{h,k,\ell \\ -\infty}}^{\infty} \left| F_{hk\ell} \right| \exp(i\phi_{hk\ell}) \exp[-2\pi i(hx+ky+\ell z)] \tag{1}$$

where $|F_{hk\ell}|$ is proportional to the square root of the measured intensity, the h,k,ℓ values characterize the planes in the crystal from which reflections occur and their values are obtained directly from experiment, and $\phi_{hk\ell}$ is the phase of each reflection which is not available by direct measurement. The lack of knowledge of phase values, which range from 0 to 2π, had made structure analysis extremely difficult in the past. Trial and error methods were used for simple structures and vector search methods [1] combined with the tactic of introducing a heavy atom chemically into a light atom structure [2] proved to be quite successful. However, structure analyses of very many substances, particularly organic and biological substances containing only light atoms, were essentially impossible to carry out.

There had been an urgent need to develop a procedure for phase determination that was not dependent on the presence of a heavy atom in a crystal. Such a procedure, now commonly called the "direct method" of phase determination, has been developed. It was recognized by H. Hauptman and J. Karle [3] that the number of independent reflections measured in an X-ray pattern is many times greater than the number of unknown quantities in a crystal, the unknown quantities being the three coordinates for each atom in an asymmetric unit and, consequently, that there exists a large overdeterminacy in a mathematical sense. For practical purposes, it has been more expedient to determine the phases associated with the measured intensities as an intermediate step and then to use them in Eq. (1) to derive the atomic positions. On the basis that the electron density function [Eq. (1)] must be a non-negative function, Karle and Hauptman [4] presented a complete set of inequalities relating the structure factors $F_{hk\ell}$ in any crystal. Relatively simple relationships among the phases, $\phi_{hk\ell}$, dependent on the experimentally measured magnitudes of $|F_{hk\ell}|$, were readily extracted from the inequalities. The following are examples of the most useful relationships*:

$$\phi_{h_{1+2}k_{1+2}\ell_{1+2}} \approx \phi_{h_1k_1\ell_1} + \phi_{h_2k_2\ell_2} \tag{2}$$

(valid for reflections with large $|E_{hk\ell}|$, where $|E_{hk\ell}|$ are the normalized values for $|F_{hk\ell}|$, and

$$\tan \phi_{h_{1+2}k_{1+2}\ell_{1+2}} = \frac{\sum \left| E_{h_1k_1\ell_1} E_{h_2k_2\ell_2} \right| \sin(\phi_{h_1k_1\ell_1} + \phi_{h_2k_2\ell_2})}{\sum \left| E_{h_1k_1\ell_1} E_{h_2k_2\ell_2} \right| \cos(\phi_{h_1k_1\ell_1} + \phi_{h_2k_2\ell_2})} \tag{3}$$

A *modus operandi* for applying the above relationships, along with some auxiliary ones, to experimental data in order to derive phases, and, in turn, the structures, has been developed by Karle and Karle.[5] The phase derivation is initiated with very few phases, up to three known phases from assigning an origin for the unit cell and several unknown phases represented by algebraic symbols. The process cascades quickly with the use of Eq. (2). In the cascading process, an unknown phase for a reflection, with a particular h,k,ℓ value is determined by a number of different combinations of known phases (or phases determined in terms of algebraic symbols). Since there are multiple indications for a particular phase, the algebraic symbols can

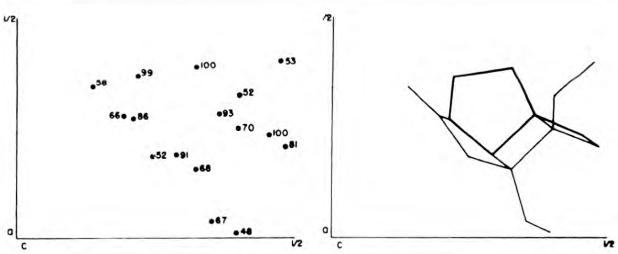
*Note that the $h_{1+2}=h_1+h_2$, $k_{1+2}=k_1+k_2$ and $\ell_{1+2}=\ell_1+\ell_2$.

be evaluated. The phase values are refined by Eq. (3) and phase determination is extended to reflections with smaller $|E_{hk\ell}|$ values, also by Eq. (3), known as the tangent formula. Finally, an electron density function is calculated with Eq. (1) that reveals the positions of the atoms. Figure 1 is an example of the maxima read from a Fourier series map calculated from the experimental X-ray data and the phases derived by the foregoing procedure for an unknown photorearrangement product.[6]

The direct method for phase determination has been formulated subsequently into a number of commonly used computer programs, among them MULTAN [7], SHELX [8], and XRAY 76 [9], by means of which it is often possible to obtain the solution of a crystal structure quite automatically. Thousands of structures of crystals containing only light atoms, as well as crystals containing heavier atoms, have been solved by direct method of phase determination.

Figure 1

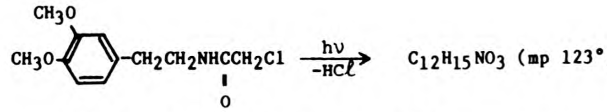
(Left) Positions of the peaks in a three-dimensional election density map for an unknown photocyclization production. The numbers next to the dots represent the height in the b direction. (Right) Peaks separated by distances of 1.35-1.60 A connected to reveal molecular formula and relative stereoconfiguration[6].



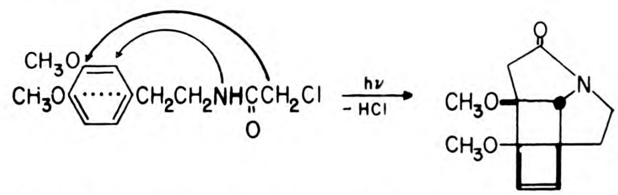
Molecular Formula Elucidation

Structure analysis is fundamental and indispensable to many scientific disciplines, such as mineralogy, geology, chemistry, molecular biology and physics, for which information about the geometric arrangement of atoms in a molecule or polyion, stereochemistry in molecules having chiral atoms, preferred conformations in flexible molecules, precise bond lengths and angles, coordination polyhedra, and aggregation of molecules into crystals by hydrogen bonding or electrostatic attractions, e.g., is essential. The following examples will illustrate some of the uses of structure analysis.

One of the basic requirements in organic chemistry is the elucidation of the molecular formula and stereoconfiguration of a new natural product, a synthetic product or intermediate, or a photorearrangement product. In the photocyclization of pharmacodynamic amines the following reaction occurred,[10]



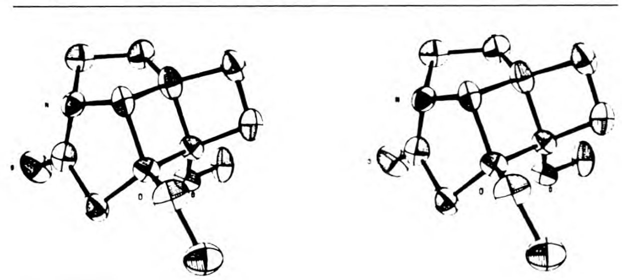
where only the empirical formula of the unknown product was known. Structure analysis by X-ray diffraction of a single crystal of the product [6] produced an electron density map, from which the peaks have been plotted in Figure 1. A calculation of the distances between the peaks (representing the positions of the atoms) immediately indicated which atoms were bonded chemically. When the appropriate bonds were drawn, the unexpected molecular formula and relative stereoconfiguration were revealed. Figure 2 shows a drawing of the molecule made by computer using the experimentally determined coordinates. A schematic indication of the photocyclization reaction, based on the known formula of the product, can be displayed as the following:



The O and N atoms were initially identified by the heavier weights of the peaks in the density map and this assignment was corroborated later by a least-squares refinement and the good fit to all the experimentally measured intensities. The positions of all the hydrogen atoms were found in a subsequent "difference" map that was calculated by Eq. (1) in which the $|F_{hk\ell}|$ values were replaced by the difference between each observed $|F_{hk\ell}|$ value and the $|F_{hk\ell}|$ value calculated on the basis of the C, N, and O positions.

Figure 2

A stereodiagram of the photoproduct from Figure 1 drawn by computer using the experimentally determined coordinates[6]. The ellipsoids represent thermal motion of the atoms at a 50% probability level.



In another typical example, the study of the chemical reactions of L-ascorbic acid, it was observed that acrolein reacts with ascorbic acid to give a crystalline product that shows no olefinic protons and carbons, respectively in the ^{1}H and ^{13}C NMR spectrum. Elemental analysis confirmed that the product is a 1:1 adduct. X-ray analysis elucidated the structural formula of the addition product to be that shown in Figure 3.[11] After the structure of the tricyclic lactone was unequivocably determined, it was possible to confirm and complete the interpretation of its spectra. Scheme 1 outlines a possible mechanism [11] for the Michael reaction in which L-ascorbic acid is a donor to acrolein to give the novel tricyclic hemiacetal lactone *3* (shown in Figure 3).

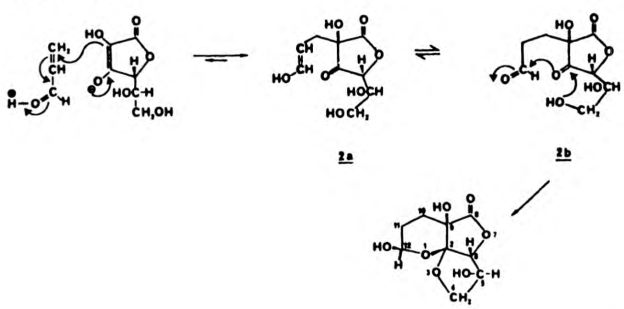
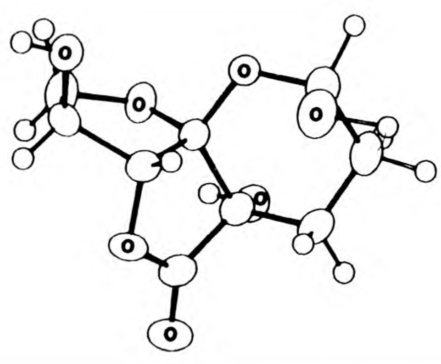


Figure 3

The molecular formula of the 1:1 adduct of L-ascorbic acid and acrolein as determined by X-ray diffraction analysis[11].

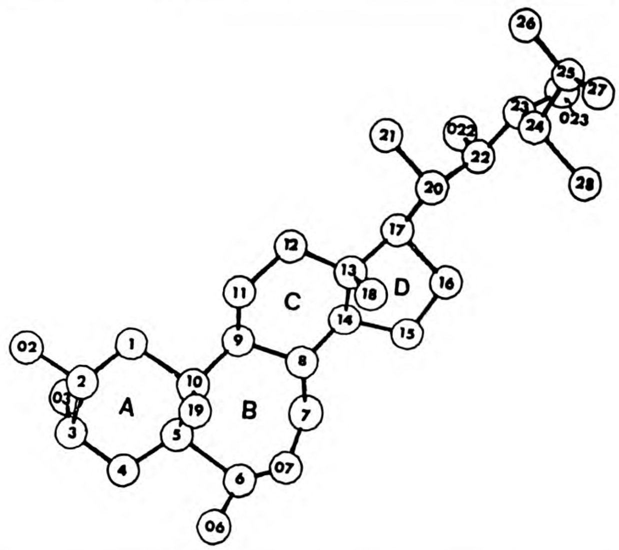


The establishing of the structural formula of a new natural product, particularly if it is available only in extremely minute quantities, is always a challenge. If a single crystal can be grown with a cross-section of 0.05 to 0.5 mm, then X-ray diffraction analysis is the method of choice. Our next example concerning agricultural chemistry demonstrates the benefits accrued from the cooperation of scientists from different disciplines, namely crystallography, organic chemistry, plant physiology and insect behavior.

As the demand for food and biomass continues to grow, plant hormones that stimulate faster growth and increased production are eagerly sought. Brassinolide is one such extremely effective hormone that is active in nanogram quantities. It doubles plant growth both by cell elongation and cell division. Since brassinolide occurs naturally in extremely minute quantities in the pollen of the rape plant, *Brassica napus L.* (15 mg isolated from *500 pounds* of pollen that was brushed from the legs of bees), chemical synthesis was imperative. However, the lack of a chemical formula was an insurmountable obstacle until a tiny single crystal was grown and X-ray diffraction analysis of the structure by Judith Flippen-Anderson of our laboratory established that the steroid nucleus of the molecule had an unprecedented seven-atom B-ring lactone, Figure 4, a molecular feature that appears to be responsible for promoting plant growth.[12–16] Brassinolide and several analogs have now been synthesized and are being evaluated in the greenhouse and in field production of potatoes, soybeans and other major vegetables.

Figure 4

Molecular structure and stereoconfiguration of brassinolide as determined by X-ray analysis[12].

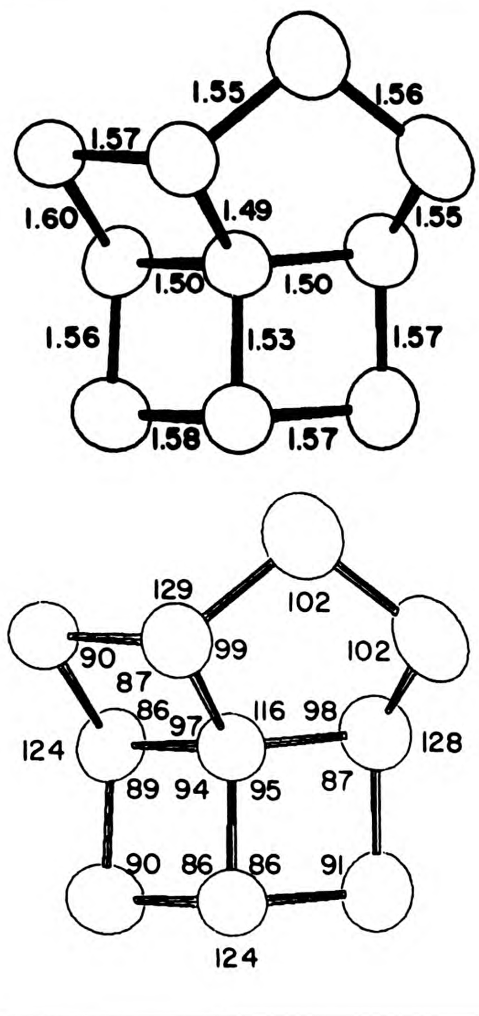


Precise Bond Lengths and Angles

Values for bond lengths and bond angles in normal organic compounds have been measured quite precisely in thousands of crystal structures and found to occur within narrow ranges for particular kinds of bonds, i.e., single, double, triple or aromatic, between specific elements. Single C–C bonds usually fall within 1.52–1.54 and an sp^3 carbon normally has bond angles near the tetrahedral value, i.e. $\sim 110°$. Organic molecules with small rings, however, must have bond angles with values far from the ideal tetrahedral. The strain in such molecules is also exhibited by bond lengths outside the normal range. Prof. William Agosta of Rockefeller University has recently synthesized homo-fenestrane consisting of three 4-membered rings and one 5-membered ring all fused together, in which all the C–C bonds are single and all the carbon atoms are "tetrahedral". The emphasis in this structural problem was not the elucidation of the molecular formula but the determination of the effect of the strain in the molecule on the values of the bond lengths and angles.

Figure 5

Bond lengths and bond angles in the homofenestrane moiety of a crystalline derivative[17].



The results from a crystal structure determination by Dr. Clifford George of our laboratory are displayed in Figure 5. The four C–C bonds involving the central C atom have been compressed to values of as low as 1.48 Å, whereas, all the peripheral C–C bonds have been stretched, especially those in the four-membered rings where the lengths range from 1.57–1.60 Å. Bond angles about the central C atom deviate from 110° by as much as 19°. Within the four-membered rings bond angles range from 86° to 97°. On the periphery, C–C–C bond angles are large, several of them having values of 129°. Despite the obvious deviations from normal values, the derivatized molecule is stable under ordinary conditions and exposure to X-rays.[17]

Conformation

In some classes of substances the molecular formulas sre known and the bond lengths and angles are expected to have normal values. Molecules like peptides, however, are very flexible with rotations possible about every N—C^α and C^α—C′ bond in the backbone and about many bonds in the side chains. It is of great importance to discover the manner of folding of the backbones and the preferred conformation of the side chains in order to correlate the composition of the peptide sequence and the structure with the varied biological functions of peptides.

Endogenous enkephalin, a linear pentapeptide functioning as a natural analgesic with opiate-like activity similar to morphine, occurs naturally in the brain, the gut and the spinal cord.[18] The extreme flexibility of enkephalin, as compared to the fairly rigid molecule of morphine, necessitated the delineation of probable conformations for the peptide. In the present study of the crystal structure of [Leu^5]enkephalin,[19] the serendipitous occurrence of four molecules of the peptide, each having a different conformation, and a large amount of solvent surrounding the peptide molecules may give good indications for the preferred conformations in solution.

The four different conformations of [Leu^5]enkephalin are displayed in Figure 6. The individual molecules have been aligned to illustrate the similarities in the extended backbone conformations and the differences in the side-chain conformations. The four independent molecules crystallize side-by-side to form an infinite antiparallel pleated beta-sheet in the crystal, Figure 7. Twenty independent hydrogen bonds between NH and C=O moieties stabilize the beta-sheet. Side chains of enkephalin molecules project above and below the rather rigid sheets into the space between the sheets. Since the separation between adjacent sheets is ~ 12 Å, there is a large volume that is filled with mobile solvent that surrounds the side chains.

This particular structure consists of more than 210 independent C, N and O atoms, a similar number of H atoms, and no heavier atoms. It illustrates the complexity of structures that can be solved by direct phase determination. Only organic structures have been used as examples in this paper. Structure analysis by direct phase determination for X-ray intensity data has also been used extensively for metal clusters, polyions, metalloorganic complexes, inorganic substances, minerals and alloys.

Figure 6

Four different conformations assumed by enkephalin molecules crystallizing in the same cell of a crystall[19].

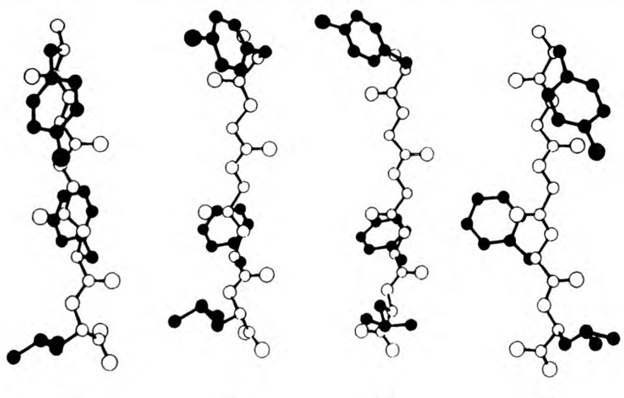
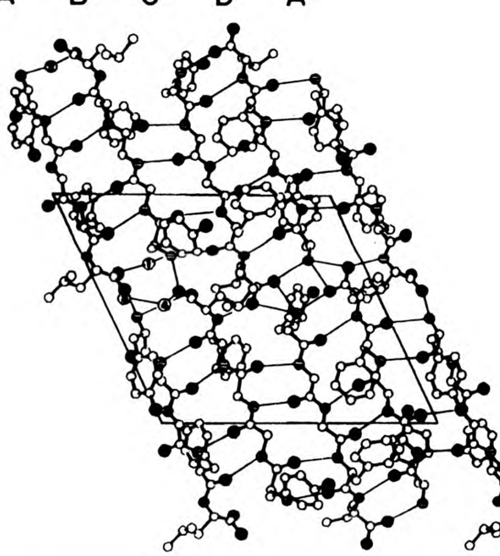


Figure 7

An infinite antiparallel pleated beta-sheet formed by the four conformers of enkephalin[19] shown in Figure 6.



Biography

Dr. Jerome Karle joined the Naval Research Laboratory (NRL) in 1944. Since 1968, when a Chair of Science was established in his name, Dr. Karle has been Chief Scientist of the Laboratory for the Structure of Matter at NRL. His work in crystal structure analysis was recognized by the 1985 Nobel Prize in Chemistry.

Dr. Isabella Karle, the coauthor, is married to Jerome Karle and is also a renowned chemist at NRL, who at times collaborates with her husband on research projects. She is noted for her pioneering work in bridging the gap between the theoretical formulas for crystal structure determination and the practical methology required to achieve useful applications.

References

1. A. L. Patterson, "A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals," *Phys. Rev.* **46,** 372–376 (1935); "A Direct Method for the Detemination of the Components of Interatomic Distances in Crystals," *Z. Kristallogr.* **90,** 517–542 (1935).

2. G. H. Stout and L. H. Jensen, *X-ray Structure Determination,* MacMillan, New York (1968).

3. H. Hauptman and J. Karle, "The Structure of Atoms from Diffraction Studies," *Phys. Rev.* **77,** 491–499 (1950).

4. J. Karle and H. Hauptman, "The Phases and Magnitudes of the Structure Factors," *Acta Crystallogr.* **3,** 181–187 (1950).

5. J. Karle and I. L. Karle, "The Symbolic Addition Procedure for Phase Determination for Centrosymmetric and Noncentrosymmetric Crystals. *Acta Crystallogr,*" **21,** 849–859 (1966).

6. I. L. Karle, J. W. Gibson and J. Karle, "The Identification and Crystal Structure of 1,2,5a, 7b-Tetrahydro-5a,5b-dimethoxy-5bH-cyclobuta[1,4] cyclobuta[1,2,3gh]-pyrrolizin-4(5H)-one, a Photolysis Product from N-Chloroacetyl-3,4-dimethoxyphenethylamine," *Acta Crystallogr.* **B25,** 2034–2039 (1969).

7. G. Germain, P. Main and M. M. Woolfson, "The Application of Phase Relationship to Complex Structures III. The Optimum Use of Phase Relationships," *Acta Crystallogr.* A*27*, 368–376 (1971).

8. G. M. Sheldrick, *SHELX 76, Program for Crystal Structure Determination,* University of Cambridge, England (1976).

9. J. M. Stewart, "The XRAY 76 System, Tech. Rep. TR-446," Computer Science Center, University of Maryland, College Park, Md. (1976).

10. O. Yonemitsu, Y. Okuno, Y. Kanaoka, I. L. Karle and B. Witkop, "Photocyclizations of Pharmacodynamic Amines. II. X-Ray Analysis of a Noncentrosymmetric Tetracyclic Indole," J. *Amer. Chem. Soc. 90*, 6522–6523 (1968).

11. G. Fodor, R. Arnold, T. Mohacsi, I. L. karle and J. Flippen-Anderson, "A New Role for L-Ascorbic Acid: Michael Donor to α,β—Unsaturated Carbonyl Compounds," *Tetrahendron 39*, 2137–2145 (1983).

12. M. D. Grove, G. F. Spencer, W. K. Rohwedder, N. Madava, J. F. Worley, J. D. Warthen, Jr., G. L. Steffens, J. L. Flippen-Anderson and J. C. Cook, Jr." "Brassinolide, a Plant Growth-Promoting Steroid Isolated from *Brassica napus* Pollen," Nature *281*, 216–217 (Sept. 20, 1979).

13. M. J. Thompson, N. Madava, J. L. Flippen-Anderson, J. F. Worley, S. R. Dutky, W. E. Robbins and W. Lusby, "Synthesis of Brassino Steroids," *J. Org. Chem. 44*, 5002–5004 (1979).

14. "New Chemicals promise Larger Crops," reported by Thomas H. Maugh II, *Science 212*, 33–34 (April 3, 1981).

15. "Steroid Plant Growth promoter Identified," C & E News, p. 20 (Nov. 5, 1979).

16. Malcolm J. Thompson, Nagabhushanam Mandava, Anita S. Worley, Werner J. Meudt, Samson R. Dutky, Williams E. Robbins, Judith L. Flippen-Anderson, Patent #182,210 (Plant Growth Promoting Brassinosteroids) August 28, 1980.

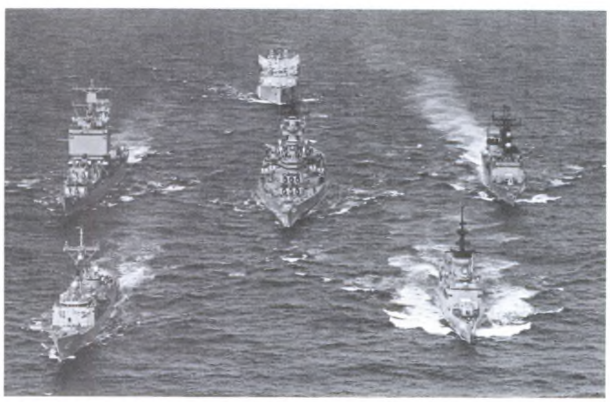
17. V. Bhasker Rao, C. F. George, S. Wolff and W. C. Agosta, "Synthetic and Structural Studies in the [4.4.4.5]Fenestrane Series," *J. Amer. Chem. Soc. 107*, 5732–5739 (1985).

18. J. Hughes, T. W. Smith, H. W. Kosterlitz, L. A. Fothergill, B. A. Morgan and H. R. Morris, "Identification of Two Related Pentapeptides from the Brain with Potent Opiate Agonist Activity," *Nature* (London) 258, 577–579 (1975).

19. I. L. Karle, J. Karle, D. Mastropaolo, A. Camerman and N. Camerman, "[Leu^5]Enkephalin: Four Cocrystallizing Conformers with Extended Backbones that Form an Antiparallel β-Sheet;" *Acta Crystallogr.* B*39*, 625–637 (1983).

Operations Research in Optimization at the Naval Postgraduate School

by Gerald G. Brown
Naval Postgraduate School



This is a brief description of a long-term research program in optimization theory and practice conducted within the Operations Research Department of the Naval Postgraduate School (NPS). Our program is unique in that our theoretical research is predominantly motivated by pressing real-world problems for which we are allowed very little time to organize and produce viable solutions. The principal sources of these problems are our research sponsors and graduate students with inflexible due dates. We show how beneficial the interaction of research and instruction at NPS can be by use of a typical thesis project. Over the years, our optimization research program has embraced many fields of application and theoretical development out of immediate necessity to support our demanding clients. We have also served as an important conduit between our civilian and military collaborators at other institutions who have assisted us in our research. Projects in the program have been supported by the Office of Naval Research, the National Science Foundation, Department of Energy, Defense Logistics Agency and Joint Chiefs of Staff.

Background

The *Operations Research* (OR) Program at NPS is unique. Operations Research, defined by Webster as the application of scientific and especially mathematical methods to the study and analysis of complex overall problems, was born of military necessity during World War II. Thus, it is not surprising that the OR program at NPS is the oldest and one of the largest such graduate engineering programs in the U.S. Students are usually commissioned officers from uniformed services of the U.S., and from Australia, Canada, Egypt, Germany, Greece, Israel, Korea, Mexico, Turkey and many other allied countries. Typical students have an undergraduate engineering degree, five to ten years of successful professional experience, and the aptitude and diligence to make them the envy of any university. Most of our students pursue the Master's Degree in a program lasting eight to nine quarters. Their first year is occupied by core coursework. The second year begins with a half-quarter experience tour, during which each student is transferred to an operations analysis activity somewhere (e.g., Washington, D.C., Hawaii, Iceland, etc.) and given actual work to do in his or her primary professional specialty. Returning from this taste of reality, advanced coursework, electives, and a thesis complete the program. Although a handful of doctoral students contribute to our research programs, it is principally the Master's thesis requirement which occupies us.

Theses frequently focus on real problems which the student will face immediately following graduation. A thesis topic may be theoretical, or quite practical, but it must be accepted by an advisor and a second reader; these faculty members often act as collaborators with the thesis student, providing an enriched environment for all.

According to Webster, *optimize* is making as perfect, effective, or functional as possible. Optimization in Operations Research evokes a scientific and especially a mathematical connotation for this definition. Computation provides the experimental evidence with which mathematical theory is scientifically tested. Optimization is one of the fundamental areas of study in Operations Research at NPS.

A Case Study

To illustrate the singular nature and pace of our research program in optimization, consider a typical thesis project. Navy LT Clarke Goodman completed his core studies and spent his experience tour in Norfolk, VA. While working on his assigned project, LT Goodman became fascinated with an unrelated activity occupying a signficant number of people and directly affecting the entire U.S. Atlantic Fleet. LT Goodman witnessed the preparation of an *Annual Employment Schedule.* The employment schedule directs the activities of all air, marine, submarine, and surface untis for the following quarter, and informs such units of planned activities for a year in advance. Having been the object of such employment schedules for years, LT Goodman was naturally interested in the reasoning underlying their formulation and in possibly improving their effectiveness and fairness. His questions were unofficially entertained by members of the planning staff.

At NPS, LT Goodman convinced Professor Kevin Wood that the employment schedule problem was a worthy topic, and enlisted my support as a second reader. LT Goodman and Wood worked together to formalize the reasoning for and render a tractable model of the employment scheduling process. They decided to restrict their attention to major *surface combatants* (e.g., frigates, destroyers, cruisers, carriers, etc.) and to use an actual unclassified annual schedule as a mechanism to test their model. LT Goodman returned to Norfolk to formally extract additional data and to receive criticism and additional guidance on his proposed approach.

LT Goodman then devised:

1. A data base for all surface combatants in the U.S. Atlantic Fleet (over 100 ships), including the weapons systems and capabilities of each ship as well as restrictions for fixed commitments (e.g., yard periods, required exercises, or even administrative fiats).
2. A data base for all *major events* in the following year (e.g., fleet exercises, deployment, etc.), including the complementary requirements of each activity (e.g., numbers and types of ships, weapons systems, substitutability of units, mission effectiveness, etc.)
3. A schedule generator which efficiently produces large numbers of alternate employment schedules for each ship, recognizing the unique attributes of each ship and honoring its fixed commitments. A crucial feature of this generator is a mechanism to estimate the contribution of each ship to the mission at hand, given its other commitments as well as its crew and weapons systems endowment.
4. An optimization model to select from the large number of candidate ship schedules a particular set of schedules which satisfies event requirements with maximal effectiveness.
5. A report writer to produce an annual employment schedule in its published format, as well as analyses of the criteria by which such schedules are evaluated (e.g., at-sea time, time between deployments, etc.)

The optimization model can be illustrated by the simplified example in Figure 1.

Figure 1

Set Partition Coefficient Matrix

	Col: Row	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Row Sum
Ship	1	1	1	1	1	1											= 1
	2						1	1	1	1							= 1
	3										1	1	1	1	1	1	= 1
Req't/ Event	4								1						1	1	= 1
	5		1		1			1				1			1		= 1
	6			1						1				1			= 1
	7		1			1			1				1		1		= 1

The object is to select from this matrix a set of columns which collectively has exactly one 1 in each row. This deceptively simple model is known as a *set partition problem.* LT Goodman used each column to represent an alternate annual ship schedule, and each row to represent mutual exclusion among all schedules for each ship (ship rows), or appropriate assignment of proper ships and weapons to each event (requirement/event rows). Furthermore, each of his columns has a value (not shown in Figure 1) which contributes additively to the value of the fleet employment schedule and this total value must be maximized.

As the number of rows and columns in the matrix grows, the resulting set partition optimization problem becomes an integer linear program of great combinatoric difficulty. To appreciate the task at hand, consider an artificial case in which only one hundred randomly generated columns need be evaluated. There are 2 to the power 100 (as a decimal number, this is more that 1 followed by 31 zeroes) different candidate solutions to try. Because LT Goodman expected to generate set partitions with tens of thousands of rows plus columns, we contacted Prof. Glenn Graves at UCLA whom we knew to be solving similar problems in crew scheduling for United Airlines. Graves agreed to help us cope with whatever Goodman produced.

Over the next *few weeks,* the data structures and logic of the schedule generator were refined technically to efficiently produce columns restricted to those most likely to be used in an optimal set partition (fleet employment schedule). Considerable modelling effort was devoted to the evaluation of each candidate ship schedule so that the multiple criteria of combat effectiveness, crew fatigue, substitutability of assets, equitable distribution of workload among ships, etc., could all be effectively addressed. In concert, we developed a set–partition solver based on existing general-purpose experimental optimization software (called the X-system by Brown and Graves) and some special-purpose procedures contributed by a prior Master's thesis by Marine CAPT Dan Bausch. The enriched set of (about 20 thousand) columns yielded by the generator was attacked in many experiments aimed at discovering and exploiting special structure in the set–partition matrices and thus reducing solution efforts.

LT Goodman and the professors enlisted to support him assembled in less than two months a prototypic system which accepts a user-friendly input script with ship and event descriptions, and manual preassignments of ships to events, and automatically produces a complete annual employment schedule in less than a minute of time-shared computer time. Objective comparison of the quality of the schedules produced automatically and those produced manually and subsequently published for actual use reveal significant improvements in all criteria. The sheer speed of response of the system suggests that schedulers could employ it as an interactive tool to incorporate non-quantifiable criteria or frequent changes in actual schedules. Technical accomplishments of this work have been presented in international research society meetings and will appear in a research paper in the open literature.

Optimization Research Program

Our objective for the overall optimization research program at NPS is to qualitatively improve the efficiency with which we can solve real-life problems using optimization. Our methods include theoretical as well as extensive empirical investigation of problems presented to us, and of problems posed as wholly artificial test cases.

Many of the classes of problems with which we must deal (e.g., integer programs) are believed to be intractable in the sense that there is strong theoretical evidence (but as yet no proof) that there can never be an algorithm to efficiently compute optimal solutions for all problem instances. Despite the daunting theoretical computational complexity of these problems, the gloomy prospects for success are rarely borne out. Real-life problems seem to exhibit special properties which render them solvable by an approach which identifies and exploits the specializations. This has led us to a guiding ***Principle of Optimization:*** **careful modelling of real-life systems that are shaped by strong economic, physical, social, and/or rational forces yields problems that are dramatically easier to solve than others.** Following this principle, we seek means to effectively identify, extract, express, and exploit in our models and solution methods problem features reflecting these underlying forces.

Solution methods can be characterized as *direct,* in which the problem is solved outright, or *indirect,* in which related sub-problems are solved and their solutions used to assemble a solution to the original problem.

Direct methods benefit from *factorization,* in which some subset of the problem's rows or columns are identified as exhibiting special structure and treated separately by the solution method. Networks are an excellent example of such structure, and we have found that networks are contained within most real-life models.

Indirect methods break up the problem into more manageable pieces, solve the pieces and assess the solutions achieved to see if they solve the original problem. If not, the pieces can be altered, and resolved, perhaps many times, to obtain solutions closer to that required by the original problem. Frequently, solutions achieved by indirect means are approximate and our goal is to achieve minimal error at reasonable cost. The classical indirect approach is *decomposition,* or its currently fashionable heuristic simplification called *Lagrangean relaxation.* We have used both approaches, especially when the *manageable pieces* thus produced are networks. In addition to repeatedly solving the manageable pieces, decomposition usually requires repeated solution of a restricted master problem to which constraints are successively added to ensure convergence. Lagrangean relaxation can be arranged to omit this difficult step, but our experience is that convergence is not so easy to prove theoretically or achieve in practice.

Both direct and indirect methods are improved, especially with integer models, if we modify them to allow violation of each constraint at some cost. Thus, the constraints become aspirations, or goals, and the choice of whether, or by how much to violate constrains is made by the model. We refer to models with linear violation penalties as *elastic* models, and in our experience there is much to recommend them. At first glance, elastic models appear to cheat by violating constraints. However, solutions which offer enormous improvement via some slight constraint violation can be very attractive in real-world problems. Our research has produced new solution methods for network models, such as:

- ☐ Assignment (e.g., men to jobs, weapons to targets, etc.)
- ☐ Transportation (e.g., rail shipment from depots to units)
- ☐ Capacitated Transshipment (e.g., capacitated physical distribution systems)
- ☐ Generalized Networks (e.g., amplification and/or attenuation in transit)
- ☐ Integer Generalized Networks (e.g., flows restricted to one of two values)

as well as for other models, such as:

- ☐ Linear Programming
- ☐ Mixed Integer Linear Programming
- ☐ Nonlinear Programming
- ☐ Quadratic Assignment (e.g., travelling salesman problems)

In cases for which our computer programs do not require extensive field support (which we are not able to provide), we have made the programs available to other academic researchers. Thus, problems formerly at the threshold of the state-of-the-art have been reduced to routine features within even more ambitious models.

The optimization group at NPS has also agreed to apply, when possible, our more complex experimental computer systems to problems supplied to us with a publication release for our research findings.

A partial topical list of applications of optimization which we have pursued includes:

- ☐ Physical distribution system design
- ☐ Energy production, distribution and consumption
- ☐ Ship scheduling and fleet employment
- ☐ Munitions procurement, storage and shipment
- ☐ Production scheduling
- ☐ Search, detection and surveillance
- ☐ Vehicle dispatching and routing
- ☐ Target assignment
- ☐ Mobilization
- ☐ Manpower
- ☐ Network reliability
- ☐ Engineering design
- ☐ Real-time process optimization
- ☐ Capital budgeting

An important role for NPS and its research programs is to serve as a medium for communication among academic, military, and civilian researchers. The list above gives ample evidence of the general usefulness of optimization. We have scrupulously shown how to schedule supertankers with the same techniques used for Naval surface combatants, how to bake crackers and cookies or produce ammunition, how to dispatch trucks or allocate sorties to military aircraft, how to build capital portfolios or plan flight tests of new fighter aircraft, and so forth for many application areas.

We are concerned that we can often solve models much more efficiently than we can implement, validate, support or understand them. Accordingly, in concert with our modelling research, we have developed several prototypic user-friendly model-building and solution-interpretation systems. (Necessity leads to invention: a misspelling-tolerant system for naming model entities was developed following a comical late-night terminal session in which students futilely sought transportation data for *Albuquerque.*)

Even more ambitiously, the entire process of formulation, model statement, data specification, and model validation and interpretation has been formally addressed by Professor Gordon Bradley, Ph.D. student Army MAJ Bob Clemence, and Professor Art Geoffrion (UCLA). Geoffrion has specified a formal approach called *structured modelling,* and Bradley and MAJ Clemence have extended the concept and implemented a system called LEXICON to illustrate the efficacy of the approach. They have extended this work by developing the concept of typing and type validation that allows automatic verification that the model correctly expresses the modeler's intentions.

All these research activities have contributed to a rather large optimization research software suite which enables quick response to new modelling challenges. Yet, there are always demands for solutions to larger, more difficult models. But such is the nature of research, each breakthrough leads to more challenges for our program to pursue.

Biography

Dr. Gerald G. Brown is Professor of Operations Research at the Naval Postgraduate School. He chairs the Doctoral Program in Operations Research, and serves as the National Research Council Postdoctoral Research Advisor in Optimization. His contributions to computational techniques in mathematical modelling have received wide recognition. He has had extensive consulting experience with large U.S. corporations and the federal government in the application of state-of-the-art mathematical programming techniques to a wide variety of problems.

References

1. Bausch, D.O., "Computational Advances in the Solution of Large-Scale Set Covering and Set Partitioning Problems," *MS Thesis,* Naval Postgraduate School (1982) 77 pages.

2. Bradley, G.H., "Survey of Deterministic Networks," *AIIE Transactions, 7* (1) (1975), pp. 222–234.

3. Bradley, G.H., G.G. Brown, and G.W. Graves, "Design and Implementation of Large Scale Primal Transshipment Algorithms," *Management Science, 24* (1) (1977), pp. 1–34.

4. Bradley, G.H., "Optimization Models to Construct Capital Portfolios," *Proceedings of the National Communications Forum 86,* held Sept. 29–Oct. 1 at Chicago, Illinois (1986).

5. Bradley, G.H., and R.D. Clemence, "A Type Calculus for Execitable Modeling Languages," *IMA Journal of Mathematics in Management* (to appear).

6. Brown, G.G., A.M. Geoffrion, and G.H. Bradley, "Production and Sales Planning with Limited Shared Tooling at the Key Operation," *Management Science, 27* (3) (1981) pp. 247–259.

7. Brown, G.G., C.E. Goodman, and R.K. Wood, "Annual Scheduling of Atlantic Fleet Naval Combatants," (in review) 28 pages.

8. Brown, G.G., and G.W. Graves, "Elastic Programming: A New Approach to Large-Scale Mixed-Integer, Nonlinear Optimization," presented at ORSA/TIMS National Meeting held November 17–19 at Las Vegas, Nevada (1975).

9. Brown, G.G., and G.W. Graves, "Real-Time Dispatch of Petroleum Tank Trucks," *Management Science, 27* (1) (1981) pp. 19–32.

10. Brown, G.G., G.W. Graves, and M.D. Honczarenko, "Design and Operation of a Multicommodity Production/Distribution System Using Primal Goal Decomposition," *Management Science, 33* (11) (1987) pp.1469–1480.

11. Brown, G.G., G.W. Graves, and D. Ronen, "Scheduling Ocean Transportation of Crude Oil," *Management Science, 33* (3) (1987) pp. 335–346.

12. Brown, G.G., and R.D. McBride, "Solving Generalized Networks," *Management Science, 30* (12) (1984) pp. 1497–1523.

13. Brown, G.G., R.D. McBride, and R.K. Wood, "Extracting Embedded Generalized Networks from Linear Programming Problems," *Mathematical Programming, 32* (1985) pp. 11–31.

14. Clemence, R.D., "LEXICON: A Structured Modelling System for Optimization," *MS Thesis,* Naval Postgraduate School (1984) 97 pages.

15. Goodman, C.E., "Annual Scheduling of Atlantic Fleet Naval Combatants," *MS Thesis,* Naval Postgraduate School (1985) 103 pages.

16. Rosenthal, R.E., "A Nonlinier Network Flow Algorithm for Maximization of Benefits in a Hydroelectric Power System," *Operation Research, 29* (1981) pp. 763–786.

17. Harrison, T.P., and R.E. Rosenthal, "An Implicit/Explicit Approach to Multiobjective Optimization with an Application to Forest Management Planning," *Decision Success* (1988) to appear.

18. Rosenthal R.E., "Principles of Multiobjective Optimization," *Decision Sciences, 16* (1985) pp. 133–152.

19. Rosenthal, R.E., P. Choypeng, and P. Puakpong, "Optimal Ship Routing and Personnel Assignment for Naval Recruitment in Thailand," *Interfaces, 16* (4) (1986) pp. 47–52.

20. Staniec, C.J., "Solving the Multicommodity Transshipment Problem," *PhD Dissertation,* Naval Postgraduate School (1987).

Profiles in Science

Robert J. Markunas

Mr. Robert J. Markunas is Associate Director of the Center for Semiconductor Research at the Research Triangle Institute (RTI), North Carolina, and manager of the Department for Semiconductor Device Research. Since 1983, Mr. Markunas has been a principal investigator for the Office of Naval Research.

His early research at RTI centered on liquid phase eiptaxial growth of III-V compounds for cascade solar cells, on Impact Avalance Transit Time (IMPATT) diodes, and Metal Insulator Semiconductor Field Effect Transistor (MISFET) structures. Together with Professor G. Lucovscky of North Carolina State University, Mr. Markunas also explored plasma deposition of dielectrics on Indium Gallium Arsenide (InGaAs). This work culminated in demonstrations of a hightransconductance, relatively stable InGaAs MISFET technology. Later, he directed research in plasma processing of semiconductor materials, and he was principal investigator for the further development of a Germanium/Gallium Arsenide, High Electron Mobility Transistor/High Hole Mobility Transistor (HEMT/HHMT) technology and for the development of an InGaAs MISFET technology. The special focus of his work has been on the development of deposition technologies for low temperature processing and the growth of alternative materials such as semiconducting diamond. This work has involved the development of a variety of in vacuo processes and in situ characterization techniques for process control, culminating in an integrated Ultra High Vacuum processing environment for materials and device research.

Mr. Markunas' recent research activity on a number of ONR and joint ONR/Strategic Defense Inititive Organization-Innervative Science and Technology programs has resulted in low temperature ($T < 300°$ C) gate insulator processing for silicon, ultralow temperature epitaxial growth ($T < 300°$ C) of silicon, germanium, and alloys, and the nucleation and epitaxial growth of diamond on single crystal nickel.

ARCTIC LOWS*

by Paul F. Twitchell
Office of Naval Research

Introduction

Polar lows, arctic lows, arctic cyclones, polar vortices and other descriptive terms such as "comma clouds" are used to describe high latitude mesoscale storm systems. As noted at a 1985 Navy sponsored workshop on this topic there has been in the past considerable confusion on what to call these storms. The report on that workshop opened with the following:[1]

"The history of meteorology is replete with instances of some phenomenon in the atmosphere that defies an adequate description. We know that something exists, sometimes with disastrous consequences to people and their possessions, but its origins and evolution, and characteristics are only vaguely understood. Furthermore, it may even be hard for meteorologists to decide on what to call it."

At that May 1985 workshop, an arctic low was defined to be a shallow vortex usually formed near the edge of an ice pack and developing gale-to hurricane-force winds in a few hours. Figure 1 is an enhanced National Oceanic and Atmospheric Administration (NOAA) meteorological satellite infrared image of that type of storm. For reference, the "eye" of the storm is about 70 km. across; the entire width of the picture is about 300 km.

At an earlier mini-symposium on polar lows held in August 1984 in Copenhagen, the "comma clouds" were discussed by I. Orlanski.[2] They were defined after a published definition by Richard Reed to be a storm that forms in polar air streams in the proximity of fronts and jet stream.[3] For example, Mullen described explosive cyclogensis associated with a "comma cloud" of 6–8 January 1981 over the eastern Pacific.[4] Figure 2 is a typical "comma cloud" which is generally found at lower latitudes and the size makes it relatively easy for detection by conventional satellites. Also, at mid-latitudes are extra tropical cyclone scale storms associated with fronts such as the explosive North Atlantic storms described by Frederick Sanders at the 1977 Workshop on Maritime Meteorology.[5] Figure 3 is a Sanders' type storm. These storms, which are referred to as "explosive" or "bombs", develop from first detection to maturity is often in a full day or more. With this background, let's focus on the topic of polar lows, which for this article includes storms also defined as arctic lows.

*Paper presented at the Department of Defense Symposium and workshop on Arctic and Arctic-Related Environmental Sciences, John Hopkins University, 1987.

Figure 1

NOAA-7 meteorolgical satelitte infrated image of a storm near 68)N, 3)W on 27 February 1984. Bright white are cold middle clouds (3-7 km.) banded cyclonically around an inner darker warm eye. Scale of this pciture is total width less than 300 km. (Picture courtesy M. Shapiro and L. Fedor, NOAA Wave Propagation Laboratory.)

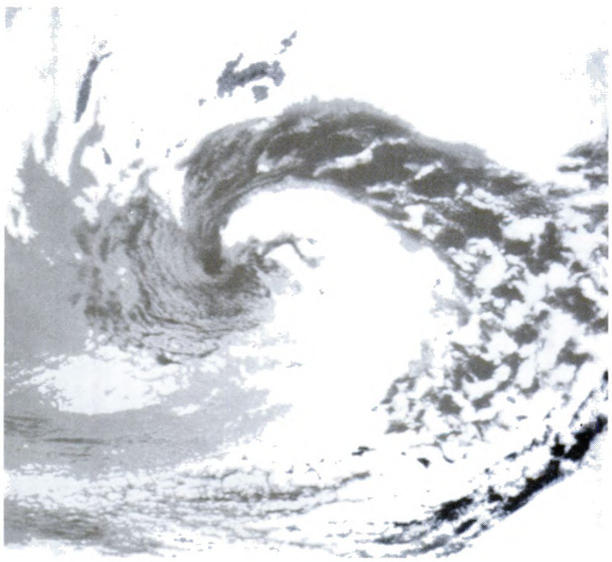


Figure 2

Example of comma cloud off U.S. East Coast. Length of comma cloud about 900 km.

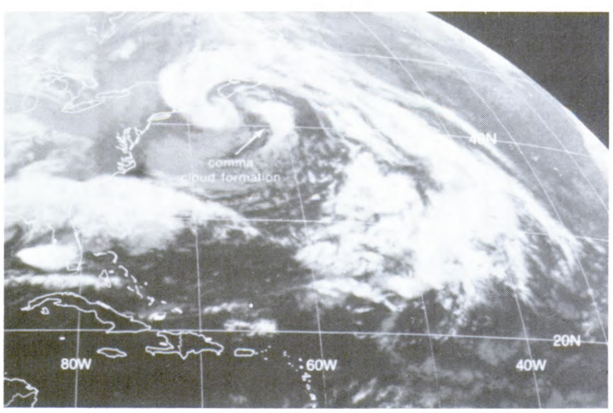
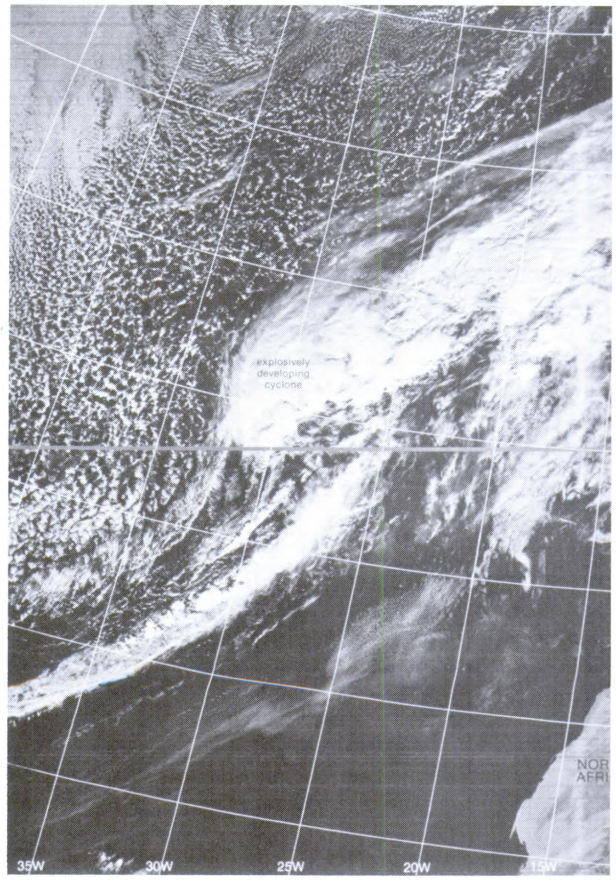


Figure 3

Example of North Atlantic explosively developing storm on polar front, note size. Storm clouds cover a region of approximately 2.5 x 10^4 km^2.



At the International Conference on Polar Lows, held near Oslo, 20–23 May 1986, the Director of the Norwegian Polar Low Project, Magne Lystad, defined a polar low to be, ". . . small often intense low pressure system that occurs in Norwegian Arctic waters." He continued in his opening address to describe an intense polar low to be: "The wind speed increases to storm force in a short time (½-2 hours) and the wind direction is changing. Heavy snowfall takes place, and visibility is poor. Sometimes high waves accompany the polar low, and they may occur simultaneously with the onset of strong winds." (Storm force winds are greater than 25 m/sec or 50 knots.) Only a few hours after the storm shown in Figure 1 was believed to have started, a scientist on a NOAA aircraft reported 300m. flight level wind speeds exceeding 75 knots. The chaotic seas were estimated from the aircraft to have wave heights up to 10 meters. Magne Lystad's definition is geographically biased by his project sponsors, primarily the Norwegian oil industry. Polar lows also occur over the Bering, Beaufort, Chukchi Seas and northern Gulf of Alaska.

High latitude storms have always been a threat to life and property; but in recent years, the increased activity (commercial and military) at high latitude and the need for protecting these investments is rapidly growing. For example, in September 1986, a battle group operating in the Norwegian Sea suddenly encountered winds of 45 knots and high seas. The following are direct quotes from the USS NIMITZ post deployment report.

"A polar low passed east of USS NIMITZ on 14 September in the southern Norwegian Sea. Winds and seas doubled within minutes and seas built rapidly to 18 feet."

"No operational model exists that can accurately forecast the development and intensity of polar lows. A numerical weather prediction model that can forecast these extremely intense systems is needed."

Another example of the problem was illustrated on 13–15 October 1985 when the United States Coast Guard icebreaker POLAR SEA encountered an unpredicted arctic low near Barrow Alaska. The ship had just completed helicopter recovery operations that require good weather conditions. Less than 5 hours later, the ship measured winds at 75 knots, seas estimated fifteen feet, icing of superstructure, low visibility in snow and the ship sustained damage from this unpredicted storm. Excerpts from the POLAR SEA situation report message from 13 October 1985 read as follows:

P 14024 OCT 85
FM USCGC POLAR SEA TO AIG EIGHT NINE FIVE SIX
INFO REQIFUA/COGDTHIRTEEN SEATTLE WA//O//

SUBJ: AWS-85 DAILY SITREP: 13 OCT 85
1. 140200Z OCT 85 POSIT: 70–53N 160–42W
2. WX: TEMP: 26F SST: 32F
BARO: 29.24R SIG: OVC/SNOW
WINDS: 240/67KTS VIS: .3NM
ICE: OPEN WATER SEAS: 240/7FT, CONFUSED, INCREASING.
3. SUMMARY OF ACTIVITIES:
A. ARRIVED VICINITY OF PT. BARROW LATE P.M. 12 OCT. MAJORITY OF TRANSIT TO PT. BARROW WAS THROUGH FY THIN AND YOUNG ICE WHICH DISALLOWED FOR ANY FURTHER GLOBAL LOAD DATE COLLECTING.
B. 130900U: CONDUCTED FLT QTRS FOR PAX/CARGO TRANSFER ———
C. 131227U: RECOVERED HELOS, ———
D. 131700U: WINDS SUDDENLY AND UNEXPECTANTLY INCREASED TO 60KTS AND CONTINUED TO INCREASE TO 70KTS WITH GUSTS TO 75KTS; WINDS OF THIS MAGNITUDE NOT FORECASTED. OMEGA ANTENNA BLOWN DOWN. BOTH ANTENNAS HAVE BEEN RECOVERED. ADVISED BARROW RADIO VIA H.F. OF PRESENT WINDS/WX. BARO.
4. INTENTIONS:
A. PROCEED TO KODIAK. SOA MAY HAVE TO BE REDUCED IF CONDITIONS CONTINUE DETERIORATE.

The 14 October supplementary situation report message from the POLAR SEA summarizes the ships encounter with an unpredicted small arctic lows.

P 141800Z OCT 85
FM USCGC POLAR SEA TO AIG EIGHT NINE FIVE SIX
INFO RHWIFUA/COGDTHIRTEEN SEATTLE WA//O//
SUBJ: AWS-85 DAILY SITREP: 14 OCT 85 SUPPLEMENTARY
1. 141606Z OCT 85 POSIT: 70–17N 164–44W
2. WX: TEMP: 26F SST: 32F
BARO: 29.60R SIG: OVC/SNOW/FOG
WINDS: 270/30KTS DECREASING
VIS: 5NM
SEAS: 240/3 SWELL: 280/10
3. SUMMARY OF ACTIVITIES:
A. WINDS IN EXCESS OF 50KTS CONTINUED THROUGHOUT THE EVENING; MAXIMUM SEAS APPROX 15 FT WITH FREEZING SPRAY. SOA REDUCED TO 6KTS; PRESENTLY 8 HRS BEHIND SCHEDULE. MODERATE ICING HAS OCCURRED ON SUPERSTRUCTURE. AS A RESULT OF STORM. THE FOLLOWING EQUIP HAS BEEN DAMAGED/LOST:
(1) BOTH FORECASTLE LIFE RINGS/STROBE LIGHTS APPARENTLY WASHED OVERBOARD.
(2) OMEGA ANTENNA BLOWN DOWN
(3) 3 ICE ACCUMULATOR UNITS DAMAGED AND INOP. PRESENT WX HAS PRECLUDED A SAFE WORKING PLATFORM TO INVESTIGATE ALL UNITS. MORE MAY BE INOP.
(4) NAVSAT UNIT PRESENTLY INOP. ———
(5) UPPER RADAR ANTENNA INOP; REPAIRS TO UNIT NOT POSSIBLE AT THIS TIME DUE TO PRESENT WX/ICE.

B. PRESENTLY PROCEEDING APPROX 220T AT 8-10KTS. WINDS/SEAS ARE SLOWLY ABATING. LAST NIGHTS WEATHER CONDITIONS CAUSED BY A WIND GRADIENT WHICH WAS FORMED AS A RESULT OF A QUICK FORMING LOW TO THE WEST MEETING A HIGH TO THE EAST.
4. INTENTIONS:
A. CONTINUE TO PROCEED SOUGH AT BEST SPEED FOR PRESENT WX CONDITIONS.
B. PLAN TO ASSESS DAMAGE/REPAIR ——— AT FIRST OPPORTUNITY WHICH WILL ALLOW A SAFE WORKING ———

A less robust ship than the POLAR SEA may not have survived such a surprise storm. Whatever these storms are called, polar lows, arctic lows, etc., they are a threat to life, property, and ability of the Navy to operate effectively at high latitudes.

In 1983, a Navy research program to address high latitude weather was formulated; and a small funded effort on arctic lows commenced in October 1985 as part of the Office of Naval Research's, Accelerated Research Initiative known as Heavy Weather at Sea. The arctic low program was coordinated with NOAA and Norwegian programs. The progress made in understanding the genesis of the arctic or polar low in the past few years will now be discussed.

Present State of Knowledge

Climatology

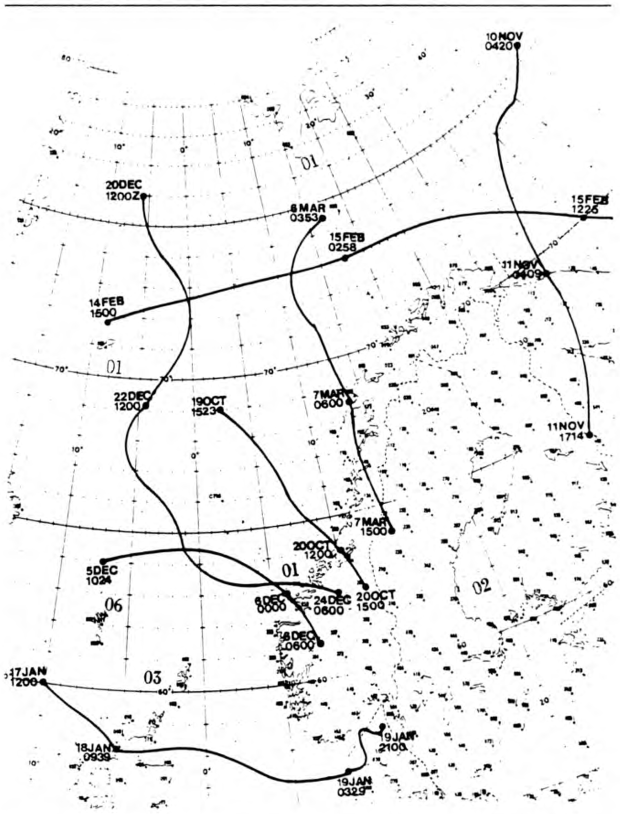
To provide a basis for planning research on polar lows, it is useful to examine historical data and document fundamental information on frequency of storm occurrence, where they most often form, and their tracks after they are detected. The last word "detected" has been a traditional problem; however, with the advent of modern environmental satellites there is a better chance of early detection by research scientists which will be discussed later. Detection of these storms with remote sensing techniques is in itself a research topic.

From the climatological studies, there are clues to the research scientists for processes contributing to storm generation. For example, one class or type of storm could be isolated for study since it generally forms in a certain area relative to high coastal terrain, another type may be associated with oceanic water mass boundaries, or the ice edge. For planning costly at-sea experiments or military exercises, the environmental knowledge of where and when to optimize an experiment or exercise is required. Results from such studies have been reported at the specialized meetings on the topic of arctic or polar lows and have or soon will appear in the published literature. For example, Steven Businger has published an article on the climatology of polar lows.[6] Kari Wilhelmsen, Dag Roger Kristoffersen and others working under the Norwegian polar low project have compiled statistics based on satellite and conventional data for storms over the Norwegian and Barents Seas. The Norwegian reports reveal that many of these storm's life cycle is only about one day, typically (according to Kristoffersen) only fifteen hours from detection to land fall on the Norwegian coast with gale winds and high waves. On Figure 4 are reported representative tracks of polar lows during the winter of 1983/1984.[7] Wilhelmsen documented 83 storms that affected the Norwegian coast during the winters of 1982/83, 1983/84 and 1984/85.[8] The maximum frequency of occurrence (one or more storms per week) was reported during the months of January and March. Not fully explained is the drop each winter of storm occurrence in February, fewer than April for the three winters studied.

Although limited (geographically and years studied), these climatogical investigations have been used in planning experiments, developing numerical simulations, and gaining insight on how to use satellite data for early detection.

Figure 4

Examples of polar low tracks. (reference 5)



Theory and Numerical Simulation

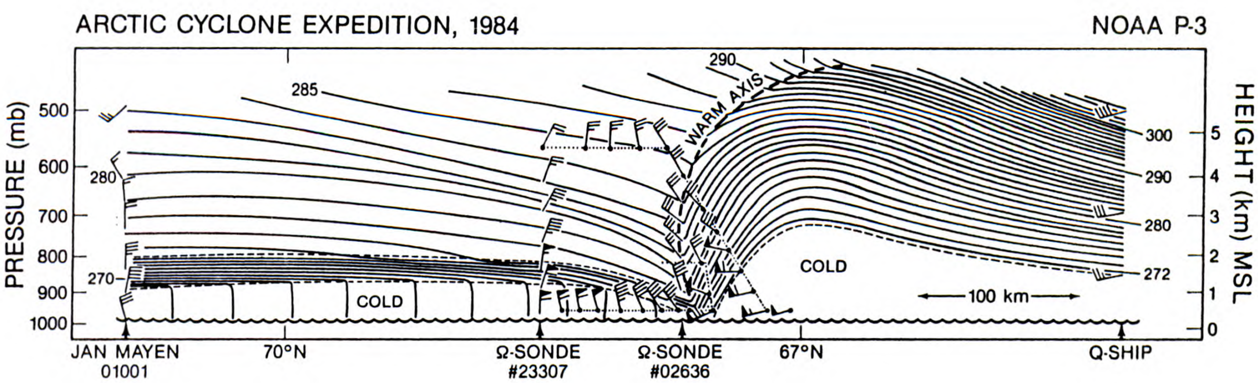
Numercial models have been formulated and numerical simulations carried out in an attempt to understand the physical processes that create the polar or arctic lows. With each generation of numerical models, there is increasing complexity due largely to new, more detailed observations. Early numerical simulations of Atlantic polar lows by investigators in the United Kingdom indicated the genesis of these storms was entirely due to baroclinic proesses [9], and later Duncan noted the baroclinicity is related to a reverse shear flow in the atmospheric boundary layer. [10] Other mechanisms for storm genesis were evolving from theoretical and numerical simulation studies; for example, Erik Rasmussen [11] suggested convective precipitation and related atmospheric instability known as Conditional Instability of the Second Kind (CISK) is a major process in the formation of polar lows. In the United States, a group at Pennsylvania State University (supported in part by the U.S. Navy through the former Naval Air Systems Command field facility Naval Environmental Prediction Research Facility) developed fine mesh (50 km grid spacing) models and applied these models to Atlantic and Pacific polar lows. The Pennsylvania State group concluded baroclinicity and CISK were both important in development of these lows. [12]

At the Norwegian Meteorological Institute, there was a focused effort from 1982 and 1985 to develop a numerical model capable of predicting polar lows that impinge upon the Norwegian coast or otherwise impact off-shore commerce. Sensitivity experiments were conducted using a 25 km and 50 km numerical grid model. [13] These experiments indicated baroclinic instability was present in all cases; but in two cases the baroclinicity was shallow, that is below approximately 3000 meters. In all cases, a double front (air mass discontinuity) was created by the model. The model does not predict condensation and convection very well for the arctic type storms, but the indication is CISK plays a role in the rapid development of these storms. To understand the inner structure of polar lows and learn how to parameterize convection, it was recommended by Magne Lystad that further studies be conducted using aircraft data such as obtained during the 1984 NOAA Arctic Cyclone Experiment. [14] Figure 5 is an example of data which can be obtained by aircraft on the internal structure of these storms. [15] The Office of Naval Research did plan research flights for the winter 1985/86 but the aircraft developed problems and the expedition was non-productive. A limited effort for the Bering Sea and vicinity was completed in February 1987, using the NOAA WP-3 research aircraft.

From ongoing theoretical work and analysis of the recent numerical simulations, new knowledge is unfolding on the genesis and prediction of polar or arctic lows. For example, those investigators studying CISK type tropical hurricane or tyhoon models have provided a better understanding of polar or arctic low dynamics. From early CISK tropical or hurricane models which did not consider sea surface temperature but did include moisture in the boundary layer (latent heat release), the CISK process has been determined to be important in the rapid intensification phase of the polar or arctic low life cycle. The CISK models do simulate the sometimes observed warm core or eye of the polar low, but most polar or arctic lows do not have the axial symmetric vortex usually found in tropical systems.

Figure 5

Cross section of polar low indicating shallow depth and limited extent of high winds. (reference 16)



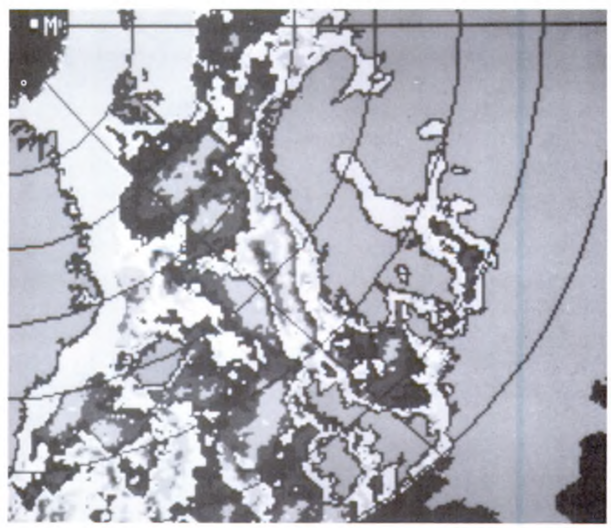
The middle latitude findings on "explosive bombs" which relate upper level, (500 mb or 18,000 feet) larger scale vorticity cascading down to the smaller surface eddies in the atmospheric flow, has applications to the polar or arctic low genesis. The difference is primarily spatial, and temporal. At high latitudes, it is rare that a small upper level (probably at 700 mb or 9,000 feet) transient disturbance, that inputs energy to a surface system in an hour or two, can be detected by the conventional sparse upper air observations obtained at twelve-hour intervals.

Using models (e.g., Anthes and Warner) the high latitude ocean cyclogenesis problem can be studied.[16] In general, these models do not develop the polar or arctic storm fast enough, even when boundary layer moisture and horizontal advection of moisture is introduced, as can be done over the central United States or just off shore at middle latitudes. Even when forced by having the model assimilate a special observation, the development is slow relative to the high latitude storms.

Observational and theoretical investigations at NOAA, University of Wisconsin-Madison, University of Copenhagen, NASA-Goddard Space Flight Center, are in progress to use satellite data for ingestion into prediction models. The planned work at Wisconsin was described, by Thomas Achtor at the 1985 Boulder workshop.[1] He noted how High Resolution Infrared Sounder (HRIS) and the Scanning Multichannel Microwave Radiometer (SMMR) can provide vertical temperature profiles approximately every 75 km along the Polar orbiting satellite track. The objective of the Wisconsin effort is to incorporate satellite derived data into numerical prediction models. Similarly the goal of the investigations at Copenhagen is to determine whether data derived from satellite data can be used to detect and analyze polar lows.[17] They used data from TIROS-N satellite instruments known as the High-resolution Infrared Radiometric Sounder (HIRS), the Microwave Sounding Unit (MSU) and the Stratospheric Sounding Unit (SSU). These instruments collectively are referred to as TOVS for TIROS-N Operational Vertical Sounder. From preliminary results, the Copenhagen group concludes that polar lows can be detected, and useful information or the vertical structure of the atmosphere can be obtained for numerical models. The future of some of these satellite systems now in orbit is questionable; however, another recently launched system is now in orbit, called Special Sensor Micro-wave Imager (SSM/I). One of the recently launched Defense Meteorological Satellite Program (DMSP) operational satellites has an SSM/I instrument and subsequent DMSP vehicles are scheduled to carry the SSM/I. For that reason, the Office of Naval Research is supporting work at NASA-GSFC Laboratory for Oceans to determine if the SSM/I would be useful for detecting polar or arctic lows. Using existing SMMR data to simulate the SSM/I signals, Per Gloersen has demonstrated that indeed the SSM/I system can be used to identify polar lows over the water from the sea surface signal, indicating the presence of high winds. Figure 6 is an example of the type imagery that the SSM/I system will likely produce. This SMMR derived image is for the same storm shown in Figure 1 which was the storm the NOAA P-3 penetrated on 27 February 1984. The algorithm developed by NASA indicates agreement with the NOAA scientists on the aircraft estimated surface wind at about the same time and place. The preliminary results for applying operational satellite system derived data to the detection of polar or arctic lows is encouraging.

Figure 6

Example of composite surface wind and vertical water content. The black and white reproduction of the false color display can crudely be interpreted that dark areas are indicative of high winds. (Courtesy of Per Gloersen.)



Summary and Plans

In very short periods of time, a great deal of information has been consolidated that relates to polar lows, and new knowledge on the physical processes that generate these storms has been acquired. However, if an operational forecast procedure is to emerge from this research, there remains a critical need for more aircraft flights in, near, and if possible, in the region of a storm development before it actually forms. Using present knowledge, it should be possible to conduct a field experiment and vector an aircraft from real-time satellite data. With on-board satellite data reception capability, the scientist on the aircraft could "finetune" their mission in flight to optimize data acquisition.

At high latitudes, a ship can frequently (about every hour above 72°) receive operational environmental satellite data. If the prediction model can be simplified and tailored for satellite data assimilation, it is possible that a small computer could run the model. The technology (computers/satellites) is or soon will be available; however, to achieve the goal of a system like the Navy Tactical Environmental Support System, having an arctic low module, requires a better understanding of the physics that creates and drives these dangerous storms. Furthermore, it appears that even with the best physics in the numerical models, the small size of the polar or arctic low remains a problem for a reasonable numerical model to predict. (Reasonable here means model grid size and related computer power.) For that reason, the "priming" of the model with satellite derived data on storm location, possibly strength (central pressure), nearby winds, selected vertical temperature and moisture structure, sea surface temperatures, or possibly other parameters could be assimilated into prediction models in essentially real time.

The numerical prediction models developed to the present need to be improved if they are to be useful, for example, the convective parameterization borrowed from middle latitude models requires modification. The Norwegian model development is promising and the Office of Naval Research's plan is to arrange in early 1987 for one of the Norwegian numerical modelers to work with atmospheric numerical modeling community in Boulder, CO. If successful in developing a reliable model using research quality initialization data and state-of-art computers in Boulder, then the next step will be to incorporate satellite data. The premise is that in an operational scenario the data used for research model development will not exist but relatively crude satellite data should be available.

Biography

Dr. Paul F. Twitchell manages the Ocean/Atmosphere/Polar Program of the Applied Research and Technology Directorate at the Office of Naval Research. Previously, Dr. Twitchell managed the Environmental Program at the Naval Air Systems Command and the Naval Environmental Prediction Research Facility in Monterey, California. For many years he was a meteorologist and physical science administrator at the Office of Naval Research Detachment, Boston and was a visiting professor in the Oceanography Department at the United States Naval Academy.

References

1. Kellogg, W.W. and P.F. Twitchell 1986: Summary of the Workshop on Arctic Lows 9–10 May 1985, Boulder, CO. *Bulletin of American Meteorological Society,* 67, 186–193.

2. Orlanski, I., 1984: Explosive cyclogensis. *Proceedings,* Mini-Symposium on Polar Lows 30–31 August. Geophysical Inst., Univ. of Copenhagan. (In Proceedings of the Symposium published by Geophysical Institute, University of Copenhagan.)

3. Reed, R.J., 1979: Cyclogensis in polar air streams. *Monthly Weather Reviews,* 107, 38–52.

4. Mullen, S.L., 1983: Explosive cyclogensis associated with cylones in polar air streams. *Monthly Weather Reviews,* 111, 1537–1553.

5. Twitchell, P.F., 1978: Meeting Review-Workshop on maritime meteorology, 8–9 September 1977. *Bulletin of American Meteorological Society,* 59, 172–174.

6. Businger, S., 1985: The synoptic climatology of polar low outbreaks. *Tellus,* 37A, 419–432.

7. Kristoffersen, D.R., 1985: Polar lows observed the winter 1983–84 in the adjacent seas of Norway. Norwegian Polar Low Project, Technical Report No. 9. (See Gronas, et al for availability.)

8. Wilhelmsen, K., 1985: Climotological study of polar lows near Norway. Norwegian Polar Low Project, Technical Report No. 4. (See Gronas, et al for availability.)

9. Harold, T.W., and K.A. Browning, 1969: The polar low as a baroclinic disturbance, *Quarterly Journal of Royal Meteorological Society,* 95, 710–723.

10. Duncan, C.N., 1978: Baroclinic instability in a reversed shear flow. *Meterological Magazine,* 107, 17–23.

11. Rasmussen, E., 1979: The polar low as an extra tropical CISK disturbance. *Quarterly Journal of Royal Meteorological Society,* 105, 531–549.

12. Sardie, J.M., and T.T. Warner, 1984: The genesis of polar lows. *Proceedings Mini-Symposium on Polar Lows,* Geophysical Institute, University Copenhagan. (See Orlanski for availability.)

13. Gronas, S., A. Foss, and M. Lystad, 1986: Numerical simulation of polar lows. Norwegian Polar Low Project Tech. Rpt. No. 18. (Available from Oceanor A/S, Klaebuveien 153, P. Box 2905—Tempe, 7001 Tyondheim, Norway.)

14. Lystad, M., 1986: Polar lows in the Norwegian, Greenland, and Barents Seas, Summary Presentation at Polar Low Conference, Oslo, Norway, 20–23 May 1986. (In Proceedings of the International Conference on Polar Lows, available from the Meteorological Institute, Oslo, Norway.)

15. Shapiro, M.A., L.S. Fedor, and T. Hampel, 1986: Arctic Cyclone Expedition, 1984: Research aircraft observations of fronts and polar lows over the Norwegian and Barents Sea. Norwegian Polar Low Project Tech. Rpt. No. 26. (See Gronas, et al for availability.)

16. Anthes, R.A., and T.T. Warner, 1978: Development of hydrodynamic models suitable for air pollution and other meso meteorological studies. *Monthly Weather Review,* 106, 1045–1078.

17. Steffensen, M., and E. Rasmussen, 1986: An investigation in the use of TOVS-data in polar low research. *Norwegian Polar Low Project Technical Report No. 25.* (See Gronas, et al for availability.)

THE EVOLUTION OF COGNITIVE SCIENCE*

by Dr. Herbert A. Simon
Carnegie-Mellon University

Cognitive science is the study of intelligent systems, whether these systems be human beings or machines. In recent years, the study of intelligence has been guided by the Physical Symbol Hypothesis. This hypothesis, which is to be tested empirically and not accepted on faith, asserts that a system is capable of behaving intelligently if and only if it is a physical symbol system—that is, if and only if it has the following characteristics:

☐ It can process symbols, that is, patterns of some kind that can denote ("stand for") other patterns or things in the outside world.

☐ It can store these symbols in one or more memories.

☐ In its memories, it can organize symbols into relational structures, and can reorganize and revise these structures.

☐ It can input symbols and symbol structures, and output them.

☐ It can compare symbol structures for identity, and follow different courses of action depending on whether a pair of structures are found to be the same or different.

The claim, then, is that these are precisely the capabilities that intelligent systems employ. If the hypothesis is correct, then two corollaries follow:

1. Human beings are (at least) physical symbol systems, for we are capable of behaving intelligently.

2. Computers can be programmed to exhibit intelligence, for they demonstrably possess the capabilities enumerated above, hence are physical symbol systems.

Cognitive science is concerned with testing and elucidating these two corollaries of the physical system hypothesis in order, first, to gain an understanding of how people think and, second, to learn how computers can be programmed to think. The first of these goals is usually thought to be the province of cognitive psychology, the second, the province or artificial intelligence. Since the two goals are closely connected, I will use the term cognitive science to encompass both of them.

Under this interpretation, cognitive science becomes a meeting place for cognitive psychology, artificial intelligence, linguistics, philosophy (especially epistemology) as well as other disciplines. And it has two intended products: improved human thinking and more powerful machine augmentations of human thinking.

In my account of the historical development of this field, I will not try to extract out the developments that can be attributed exclusively to the Office of Naval Research. ONR has been, and is, a major source of support for research in cognitive science since World War II, and it has played an important role in almost all of the developments to which I shall refer.

* This article is based on a lecture Professor Simon presented at the symposium honoring the 40th anniversary of the Office of Naval Research held in Washington, D.C., August 1986

History Prior to World War II

Prior to the Second World War, research in psychology was largely dominated by the point of view known as "behaviorism." Behaviorism avoids reference to internal events, preferring to look simply for regular patterns connecting stimulus events with the corresponding responses. It has no theory of "mind," as distinguished from behavior.

The principal competing "school" of psychology was Gestalt psychology, which was willing to talk about mind, but which had only vague ways of describing the mental processes it postulated. Some of those associated with the Gestalt and similar viewpoints actually did talk about human problem solving in a way that does not seem alien to our modern information processing point of view, but they lacked a technical vocabulary for making their theories precise. The two psychologists who, by hindsight, look most like ancestors of contemporary cognitive science were Otto Selz and Karl Duncker. Their work was not widely known in the United States before World War II.

In addition to behaviorists and Gestalt psychologists, there were a number of scientists who were seeking to explain human thinking in terms of the physiology of the human brain. But physiological psychology was still in its infancy; its main accomplishments had been to determine the localizations, within the brain, of certain mental functions. Some progress had been made, also, toward understanding the biochemical processes that take place in neurons and at the synapses between neurons. But physiological psychology was (and is) far short of providing a biological base for the symbolic processes that had been observed to underlie thinking. Perhaps the most ambitious attempt to bridge the gap between brain physiology, on the one hand, and experimental evidence about thought processes, on the other, was Donald Hebb's book, *The Organization of Behavior,* published in 1948.

From this description of the point at which psychology had arrived by the middle 1940's it may be seen that a fundamental change was needed in our view of the mind before a rigorous theory could be developed of how human beings processed information and manipulated symbols.

Other developments outside psychology, and specifically in the domain of logic, were meanwhile preparing the way for this new conception of thinking processes. Modern symbolic logic, as exemplified, for example, by A.N. Whitehead and B.N. Russell's notable *Principia Mathematica,*[1] cast an entirely new light on the nature of reasoning. The formalization of the rules of logic showed that symbols could be manipulated in precisely defined ways. In formal logic, symbols became as tangible as so many blocks of wood, to be shaped and altered by rigorous rules of inference.

Of course, logic is not identical with thinking. Logic provides certain normative rules against which the correctness of an inference can be judged, but it says nothing about how inferences are actually achieved, nor does it explain how human beings reach conclusions that cannot be justified rigorously, or that are just plain wrong. Nevertheless, the new developments in symbolic logic generated a completely new way of viewing the nature of thought, and thereby the nature of mind.

The connection became much closer with the invention by Alan Turing in 1936 of the Turing Machine, a conceptual computer that cast a very bright light on the fundamental requirements for a completely general reasoning system. Both the advent of World War II and the fact that actual computers were not yet in existence at the time of Turing's discovery delayed its application to cognition for about two decades, but there is no doubt that Turing himself soon came to realize the implications of his machine, and of computers in general, for cognitive science. Already in 1950, he wrote a remarkable essay on "Minds and Machines" and prepared a program, perhaps the first, for a chess playing computer. (The program was hand simulated, but apparently never actually run on a computer.)

Impact of World War II

The Second World War marks a watershed in the history of cognitive science, among other reasons, because the War saw the birth of the modern computer. But other developments were taking place as well. Just before the outbreak of the War, Claude Shannon (1938) conceived the idea of using Boolean algebra as a formal language to describe switching circuits.[2] Soon Pitts and McCulloch (1943) applied this same formalism to the description of networks.[3]

Information theory, of which Shannon was also one of the originators, was born during this same period, providing additional support for the belief that the work of symbols was measurable. The period of the War saw great progress, too, in the theory of feedback control and the applications of that theory to the design of goal-seeking mechanisms. In 1943, Rosenblueth, Wiener, and Bigelow published their well known paper on "Behavior, Purpose, Teleology,"[4] which developed operational definitions of these concepts and shortly thereafter, a number of investigators realized actual teleological mechanisms in the metal, one of the best-known examples being the nest-seeking "turtles" of Grey Walter.[5]

Within economics and statistics several major contributions to the formal theory of decision making emerged toward the end of the War. The first of these was the theory of maximization of subjective expected utility, a body of theory that continues to provide the foundation for most contemporary mathematical statistics and theoretical economics. A second was the invention of game theory by von Neumann and Morgenstern (1944), with proposals for its application to economic and social phenomena.[6] A third was the development of tools for operations research and management science, among the most notable of which was the invention of the simplex method for linear programming just after the end of the war.

Finally, the end of World War II permitted the wraps to be taken off of the pioneering electronic computers that had been built during the war years. I will have a great deal more to say about computers later, but of course their first impact was as powerful engines for numerical computation.

All of these developments—and especially information theory, the theory of control systems, and computers—were bundled together by Norbert Wiener under the label of "cybernetics," with a major impact upon thinking about information-bearing and goal-seeking systems. But the ideas associated together by cybernetics still did not constitute the cognitive revolution, which only emerged about a decade after the end of World War II.

The Birth of Cognitive Science: 1956

There are several ways of characterizing what was lacking in the cybernetic view of the world in order to permit a full-fledged revolution in conceptions of mind and intelligence. One way of putting it is that the cybernetic concepts were too closely tied to the real numbers. To treat phenomena in cybernetic terms, they had to be arithmetized, and many of the most central phenomena that are related to intelligence—language, for example, human concept formation, problem solving—did not readily lend themselves to numerical treatment. Only in economics, where readily measurable constructs like prices and quantities of goods were at the center of the theory, did numerical translation seem quite natural.

In 1956, four events took place that are often cited as marking the birth of the new cognitive science. One was the publication of *A Study of Thinking* by Bruner, Goodnow, and Austin, which introduced the notion of Strategy (probably derived from the theory of games) into the analysis of human concept formation processes.[7] A second was the publication by Naom Chomsky of one of the first versions of his new theory of transformational grammar, which showed how the syntax of natural languages could be formalized by notations closely resembling computer programming languages.[8, 9, 10] A third was the publication by Newell, Shaw, and Simon of a paper describing the Logic Theorist, a computer program capable of discovering proofs for theorems in symbolic logic—the first example of a program for solving problems by heuristic search. And a fourth was the publication of George A. Miller's celebrated paper about "The Magical Number Seven," an analysis of human short-term memory capacity, which, among other things, demonstrated that the "bit," the fundamental unit of information theory, was not the right unit for characterizing human memory limits.

What these four studies had in common was an approach to mental functions in terms of clearly specified processes that referred to the manipulation of symbols. They showed that rigor could be attained without the use of numbers. And although only one of them—the report on the Logic Theorist—made explicit use of the computer, the relevance to all of them of the concepts of computer programs and programming languages was quite evident. They provided the foundation for viewing the computer as a general symbol processing, and for viewing symbol processing as the basis for intelligence. In them one sees the core of the physical symbol system hypothesis.

The idea that computers are not simply number crunchers, but quite general symbol processing machines, had its earliest manifestations in the architecture of the store-program computer itself. For a computer to use an instruction that has been stored in its memory, the instruction must be interpreted not as a number but as a sentence in the imperative mode: "Do such and such!" Unless the computer understands the instruction as denoting a demand to take a specified action, it cannot execute the instruction. The fact that it was easy to "encode" instructions as strings of binary digits concealed from many that they were not really numbers, to be added to or multiplied with other numbers, but sequences of symbols to be interpreted as sentences. As recently as a decade ago, it was possible to find professional computer scientists who thought that what was "really" stored in computer memories were only binary digits. A major paradigm shift was required to establish computers as general-purpose manipulators of symbols, including symbols that denoted words, pictures, and actions.

The Search Paradigm: 1955-1965

The first decade of research in cognitive science focussed upon problem solving as the quintessential intelligent activity, and upon heuristic search as the central process in problem solving. It had long been noted that when human beings are faced with problems involving large spaces of possible solutions, they do not search these spaces either systematically or randomly, but look for rules of thumb, now usually called "heuristics," that permit the search to be highly selective.

Computer programs embodying the principles of heuristic search soon showed themselves capable of solving a considerable variety of problems that were quite hard for intelligent human beings. One of the central ideas that emerged from this research was to organize heuristic search around means-ends analysis. Means-ends search is achieved by comparing the goal situation for a problem with the situation achieved to date, noting one or more differences between the present and the desired situation, drawing from memory an operator know to be efficacious in removing differences of this kind, and applying the operator to move to a new situation. The new situation will often be closer to the goal than the previous one (the difference that was noted will have been removed), and the same cycle can then be repeated from the new starting point. If we examine the so-called "expert systems" that have proliferated in the past few years for doing a wide variety of artificial intelligence tasks, we observe that means-ends analysis is still the principal inference engine embedded in such programs.

The problem domains on which the AI systems and cognitive simulations of the first decade were tested were mainly "toy" or puzzle domains. The problems were often quite difficult, but they did not require the solver to draw upon any extensive store or real-world knowledge. To transport the missionaries and the cannibals across the river, observing the terms of the traditional problem, one does not have to know anything about real missionaries or real cannibals or, for that matter, about real rivers or boats. All the information that one needs is contained in the brief problem statement of goal and rules.

This selection of problem domains for the early research was in part a response to the usual research strategy of faking up difficulties one at a time. There was enough to be discovered about inference processes and heuristic search without complicating the situation with large quantities or real-world knowledge. Moreover, the memory capacities of early computers were severely limited (it is hard for us now to remember how limited they were), so that it was not practicable to tackle problems that required access to large bodies of information stored in memory. The character of research in cognitive science began to change as the advances in computer hardware, year by year, relaxed the stringency of the limitations on computer memory capacity and speed.

Knowledge-Rich Domains: 1965 and after

The first knowledge-rich domain studied extensively by researchers in AI and cognitive psychology was the game of chess. From one standpoint, the world of chess is not knowledge-rich at all. The game is played on an 8x8 board with 32 pieces of a few different kinds, each with rigorously defined moves. The world of chess is fantastically simpler than the world of real life. Nevertheless, the search space of chess is so enormous that the skilled player must have available in memory a very large body of information about the patterns that will be observed on the chess board in the course of games, and about what actions should be considered when any one of these patterns makes itself evident.

Partly as an historical accident, and partly because of the characteristics of the game—the irregularity of its moves, and the difficulty of characterizing positions or strategies in any simple, computable mathematical form—chess has become the *Drosophila* of artificial intelligence. It provides a standard environment in which research ideas can be tried out and knowledge about heuristic search theory can cumulate. A great deal of the cognitive science, both psychology and AI, has been derived from research on chess along with research on some other standard environments, like the Tower of Hanoi problem and the problem of medical diagnosis.

Studies of chess perception, initiated by the Dutch psychologist Adriaan de Groot, soon revealed that the chess grandmaster can carry out look-ahead search with extreme selectivity because of the knowledge he possesses about where to search. A typical Chess position has a branching factor of about 20 legal moves at each choice point. An exhaustive search even two or three full moves deep would require an enormous number of positions to be examined (a multiplicative factor of 400 per move in depth). The empirical evidence shows that the grandmaster seldom looks at more that 100 of this enormous number of possibilities—but he normally looks at the *right* 100. He sees what is relevant in the position, while the weaker player, searching to about the same extent, sees much that is irrelevant and misses much that is relevant.

The research on human chess players also gradually revealed how the grandmaster achieved this astounding selectivity in search. The grandmaster had stored in memory a large number of patterns (perhaps

50,000) representing configurations of pieces that occur frequently in chess positions during well-played games. When any of these patterns occurs during the course of a game, the chessmaster recognized it at once, just as the reader of these lines recognizes each of the familiar English words on the page—or risks missing the meaning if some are not recognized. Moreover, with each familiar chess pattern, the grandmaster has stored some information about actions that may be favorable whenever that pattern appears. This information provides the raw materials for the expert's reasoning about the position, and his choice of the particular branches of the game tree that will be worth examining.

This insight into chess perception has given us a basic understanding today of the nature of expertness in all domains; and as a result of that understanding, the ability we have acquired to construct expert systems in the form of computer programs. If we examine the structure of such programs—programs for medical diagnosis, for example—we find at their core a large body of expert knowledge that is indexed by recognition cues or patterns. Whenever such a pattern is observed in the problem situation, access is gained to the information relevant to it that is stored in the main memory. The expert is like a large indexed encyclopedia with which is associated some capacity for drawing inferences from the information that is retrieved with the help of its index.

This understanding of the nature of expertise also enables us to see why the expert is so often able to solve problems, or take major steps toward their solution, on the basis of "intuition." Intuition is identified by the ability to reach decisions quickly, and often without being able to explain (or to be aware) how the result was achieved. Experts have intuition because they are able to recognize the key features in situations in the domain of their expertise. And they cannot explain the origins of their intuitions since we have, in general, no conscious access to the grounds for our acts of recognition—no awareness of the specific cues to which we responded, but only knowledge of the result of the whole recognition process. Intuition is simply recognition.

The research on expert performance and efforts to construct automated expert systems have begun to give us some knowledge of the size of the body of knowledge on which the performance depends. I have mentioned 50,000 patterns (or "chunks") as an estimate of the size of the chess grandmaster's memory index. Estimates of the sizes of the vocabularies of college educated professionals are of the same order of magnitude.

If acquisition of this amount of knowledge is a prerequisite to expert skill, it is easy to see why no one, at least in the few domains that have been studied carefully, reaches world-class performance with less than a decade of close application to learning the domain. Mozart (starting at the age of 4) was composing for perhaps 14 years before he composed any world-class music. Bobby Fisher took nearly ten years to become a chess grandmaster. Picasso had painted for more that a decade before he began his "Blue Period" paintings, his first notable product.

With these brief samples of the kinds of things we have learned through research in cognitive science, I would like to bring my history up to the present by discussing some of the directions of research that are most active at the present time.

The Frontiers of Cognitive Science Research

Most of you are well aware of the very active research that is going on today in the domain of expert systems. Most of this research lies toward the "applied" rather than the "basic" end of the spectrum. The mechanisms that are being used in most expert systems are the mechanisms that I have described briefly in the previous two sections of this account: especially means-ends analysis and access to large knowledge bases by recognition. There is still much to be learned about how to apply these mechanisms to real world problems, but there are undoubtedly wide areas of application that can be exploited with our present understanding of mechanisms and architecture.

If we search for the most active areas of basic research in cognitive science today, we must look in other directions, three in particular: research on learning, research on problem representations, and research on robotics. I would like to say a few words about each of these topics.

In the literature of psychology, a distinction is often made between "reproductive" thinking and "productive" thinking. Reproductive thinking is the everyday garden variety that uses established methods and, while it may solve problems, does not produce novelty. Productive thinking is the kind that is supposed to lead to new, creative outcomes. There is a good deal of room for skepticism that there really are these two kinds of qualitatively different thinking. It might well be that creative products emerge from thinking when established ways of thought are applied to new situations.

Even if that is true (and I think it is), many people remain unconvinced that computers (or people) are capable of original and creative thinking as long as they are simply proceeding on the basis of procedures that have been programmed into them. For these skeptics, the mere fact that computers can solve difficult problems, even at a professional level, does not prove that they are capable of intelligence, or that computer programs cast much light on the workings of intelligence in the human brain. They will only be convinced that cognitive science is on the right track if it can elucidate the processes of learning, and the processes whereby problem representations are formed in the first place.

Learning

Very early in the history of artificial intelligence, attempts were made to program computers to learn. The most successful of these attempts were the programs that Arthur Samuels wrote in the late 1950's that enabled his computer to learn to play the game of checkers, and to improve its skill to a very high level. Unfortunately, others were not able to extend Samuels' techniques to other task domains, and research on learning languished for a number of years.

Within the past ten years, a number of new ideas have been introduced into learning research. One of these is the notion of the *adaptive production system.* An adaptive production system is a computer program that is capable of creating new productions that can then be inserted into the program (into itself), thereby enabling it to work more effectively in the future. One way of accomplishing this is to give the program one or more worked out examples of a task, showing the solution step by step. The adaptive production system examines each example and uses its capabilities for means-ends analysis to diagnose why each step was taken, what cues in the situation signalled, its appropriateness, and what it accomplished. It uses this information to create new productions capable of performing the task. Systems of this kind have been built, for example, that taught themselves how to solve linear algebraic equations by analysis of worked-out examples.

At the present time, there are a number of exciting new initiatives in learning research that make use of parallel computer architectures. These so-called "connexionist" approaches achieve learning by "tuning" large networks of active parallel computing elements. The analogy to the work of the human retina and associated brain areas in sorting out and identifying visual patterns is an obvious one, but the connexionists are seeking to extend their systems to cognitive as well as perceptual learning tasks. Since these are new developments, it will be some time before we can evaluate them adequately, but there is no doubt that they represent a promising direction of exploration.

These are just samples of the kinds of research on learning that are going on in cognitive science today. Their obvious value lies in the new insights they are providing us about human learning—insights that are just beginning to be translated into classroom and tutorial procedures. Whether, and under what circumstances, it will prove to be efficient to allow artificial intelligence systems to learn, rather than programming them directly for task performance, is a question that must be left for the future to answer.

Representation

It is sometimes asserted that a problem well formulated is half solved. That is often, but not always, true. Some problems are hard even when we have a way of representing them clearly; but others find a speedy solution as soon as we find the right way of looking at them. In any case, finding a good problem formulation may often be an essential creative step towards solution. Understanding why this is so and understanding how effective representations are created are important contemporary goals of cognitive science research.

The "mutilated checkerboard" is a classical problem where discovering the right representation is crucial for solution. Given an 8x8 checkerboard and a set of 32 dominoes, each capable of covering exactly two adjacent squares, it is easy to see how to cover the checkerboard with dominoes, with no square uncovered and no domino left over. Suppose, now, that we cut out the northwest and the southeast corner squares of the checkerboard. Can this mutilated board of 62 squares be covered by 31 dominoes?

This is a very hard problem, which most people are unable to solve even in several hours. It becomes easy as soon as it is observed that each domino covers exactly one red and one black square while the mutilation removes two squares of the same color, so that there are no longer equal numbers of squares of each color to be covered. The re-representation of the problem—replacing it by the abstract problem of simply counting the numbers of squares of each color that are covered—is the key to solution.

Not much progress has yet been made to understand how the new representation for the mutilated checkerboard problem could be discovered. But computer programs have been constructed that can understand written natural language descriptions of problems well enough to create straightforward representations for them. For example, a program called UNDERSTAND will take the instructions for the missionaries and cannibals puzzle and construct from them a symbolic structure of objects and relations among them that represents the missionaries and cannibals situation and allows a problem-solving program to go to work moving the symbolized missionaries and cannibals across the symbolized river. Another program, called ISAAC, is able to read simple physics problems in textbooks, to create representations of the problem situations (which can be drawn as pictures on a cathode ray tube), and then to apply the laws of mechanics to solve the problems.

Human beings frequently find that drawing diagrams helps them to understand and solve problems. Why is a picture or diagram easier to understand than a corresponding description in words? Current research is beginning to explore this question by creating alternative representations of the same problem, one a propositional representation that consists of a rather direct translation of verbal statements, the other a diagrammatic representation that reflects the geometrical relations of the objects described. By comparing the computational efficiencies of these two representations in solving the problem, we begin to understand what work the diagrammatic representation is doing for the problem solver.

It is clear that much remains to be explored in the subject of representation, as in the subject of learning. But the first steps have been taken. We now have computer programs that learn, and we have computer programs that construct their own problem representations. Any notion that computer programs are forever barred from simulating "productive" thinking has been refuted by the reality of these programs.

Robotics

The design and construction of robots has turned out to be a domain of great difficulty in which progress, if steady, has been slow. We might have anticipated this (we didn't). Robotics involves not only thinking in the interior of the brain, but also the acquisition and interpretation of information through the senses, and the implementation of actions through the motor system. The "new" brain, of which we humans are so proud, has been in the process of evolution for one or two million years. The "old" brain of the sensory and motor system, which we share with the other mammals, has been evolving for several hundred millions of years. It should not be surprising that this older system has reached a level of sophistication that is very hard to match or simulate.

Robotics is of great practical significance to both the military and civilian worlds today. It is also of great theoretical significance to cognitive science. A typical expert system is expected to handle intelligently an internal, simplified model of the complex outside world. A robot system, too, has to handle such a model; but it has the additional task of constantly checking the inferences it makes with that model against the realities of the world outside, and adjusting the model to fit the realities. Hence, robotics raises a whole range of fundamental research problems that are absent from most other artificial intelligence research.

Conclusion

Cognitive science attacks one of the great questions that has stirred human minds for centuries: how can matter implement mind? How can a physical system think? The computer has provided us with a new research tool for tackling that fundamental research question.

Cognitive science answers the question of how intelligence is achieved with the physical symbol system hypothesis: a system can be intelligent if and only if it is a physical symbol system. Thirty years of empirical research, AI research, much of it supported by the Office of Naval Resources, have developed and extended that hypothesis and its corollaries, revealing the processes of human thinking and showing how computers can be made to think important thoughts.

Through this empirical research the prospect has been opened of improving and enhancing human thinking and learning, as well as the prospect of assisting and augmenting human thinking with computerized expert systems and robots. It is the task of cognitive science, on both its psychological and artificial intelligence sides, to extend and realize these prospects, and thereby to make major contributions to human productivity and human understanding.

Biography

Herbert A. Simon is a Professor of Computer Science and Psychology at Carnegie-Mellon University, where he has taught since 1949. Educated at the University of Chicago, he is a member of the National Academy of Science, and has received awards for distinguished research from the American Psychological Association, the Association for Computing Machinery, and the American Economic Association. In 1978, he received the Nobel Prize in Economic Sciences.

His books include Administrative Behavior, Human Problem Solving (with Allen Newell), *The New Science of Management Decision* (rev. ed.), *The Sciences of the Artificial,* Models of Thought, and *Scientific Discovery* (with P. Langley, G. Bradshaw, and J. Zypkow).

References

1. Whitehead, A.N. and B.N. Russel, *Principia Mathematics,* University Press, Canbridge MA (1935).

2. Shannon, Claude, "A Symbolic Analysis of Relay and Switching Circuits," *Transactions of the American Institute of Electrical Engineers,* 57: pp. 1–11 (1938).

3. Pitts, W. and W.S. McCulloch, "How We Know Universals," *Bulletin of Mathematical Biophysics,* 9: pp. 127–147 (1947).

4. Rosenblueth, A., N. Wiener, and J. Bigelow, "Behavior, Purpose, and Teleology," *Philosophy of Science,* 10: pp. 18–24 (1943).

5. Walter, W.G., "An Imitation of Life," *Scientific American,* 182: 5 p. 42 (1950).

6. Von Neumann, J. and O. Morgenstern, *Theory of Games and Economic Behavior,* University Press, Princeton (1947).

7. Bruner, J.S., J.J. Goodnow, and G.A. Austin, *A Study of Thinking,* John Wiley and sons, New York, New York (1956).

8. Newell, A., and H. Simon, "The Logic Theory Machine; A Complex Information Processing System," *IRE Transactions on Information Theory,* Vol IT–2, No. 3 (1956)

9. Newell, A., and J.C. Shaw, "Programming the Logic Theory Machine," *Proceedings of the Western Joint Computer Conference,* pp. 230–240 (1957)

10. Newell, A., J.C. Shaw and H.A. Simon, "Emperical Explorations of the Logic Theory Machine; a Case Study in Heuristics," *Proceedings of a Joint Computer Conference,* pp. 218-230 (1957)

Unique Electronics Fire-Protective System

A team of scientists at the Naval Research Laboratory (NRL) has developed and laboratory-tested a new fire protection system capable of protecting electronic equipment aboard ships for as long as 70 minutes during fires up to 2000 degrees Fahrenheit.

NRL's Dr. Joseph T. Leonard and Dr. Shih-yung Hsieh, a former visiting scientist to NRL from Brookhaven National Laboratory and now an NRL contractor, are the principal investigators in this research.

The team reports that the new NRL development not only effectively protects electronic equipment while a fire is being fought, but also allows for uninterrupted operation of electronic systems and a quick cleanup after the fire is extinguished. Such a system could keep electronic equipment operational during a peacetime fire or one caused by hostile fire.

The NRL-developed protection system consists of a waterproof (using O-rings and gaskets) enclosure for any systems, such as a radar or navigation console.

In the proposed system, cool air is continually ducted into the waterproof enclosure. If the electronic equipment is subjected to fire, water spray is automatically activated to cool its exterior surfaces. At the same time, any heat released within the enclosure is effectively carried away by the unit's internal air cooling system. As a result, the electronic equipment within the enclosure remains protected from both the fire and smoke.

Another distinct advantage over conventional systems is the new system's resistance to fire fighting agents, the scientists say. This gives fire fighters the flexibility of using appropriate fire fighting agents to douse the fire without damaging any electronic equipment thereby assuring continued operations of the equipment during and after a fire.

Large-scale validation tests have been conducted on the new concept at NRL and particularly at the Chesapeake Bay Detachment facility of the Navy's Technology Center for Safety and Survivability. In these large-scale tests, temperatures inside prototype protective enclosures did not exceed more than 95 degrees, while outside burning temperatures went as high as 2000 degrees.

As an example of the protection given by this system, a 10-inch color television monitor protected by a new enclosure was tested in a diesel fuel pan fire. During the test, the monitor was subjected to a 20-minute burn, and the monitor, during and after the fire, functioned normally. In another example, a Navy telecommunication system and the same television monitor were exposed to a heptane fuel fire in which both items continued to function during test burns lasting as long as 50 minutes. Also, a Navy Surface Search Radar System (AN/SPS 64 V9) was exposed to heptane fuel fire for 40-minutes. The radar system continued to function normally during and after the fire test.

The NRL scientists note that work is continuing on the research project to even further enhance the new protection method and to develop procurement specifications for these systems.

New Soft X-Ray Lasers

Researchers at the Naval Research Laboratory (NRL) have demonstrated lasing at soft X-ray wavelengths in copper and germanium plasmas that were heated by a high-power infrared laser.

In recent experiments, Tong Nyong Lee, Edgar McLean, and Raymond Elton of the Plasma Physics Division, with the assistance of James Ford and Nicholas Nocerino, reported lasing at various wavelengths in the interval 196–286 Angstroms, using relatively simple, thick slab targets.

The scientists state that although the present output power of these NRL lasers is rather modest (a few kilowatts in a pulse about one-billionth of a second long), their hope is that further improvements will take the power up to megawatts, where many military and civilian applications requiring short laser wavelengths can be realized. These applications include X-ray lithography and microscopy, X-ray holography and flash radiography.

In commenting on the research the team notes that since early 1970 active search has sought to produce extreme ultraviolet and soft X-ray lasers. However, it was not until 1984 that a group at the Lawrence Livermore National Laboratory conclusively demonstrated that they had produced such a laser in a high-temperature plasma created from an ultrathin foil of selenium that was heated by a high-power glass laser.

The NRL researchers say that they built upon this early result, obtaining similar laser gain using a much simpler target (a thick slab instead of an ultrathin foil), and a much lower-power glass laser heating source (approximately one-tenth the power level). In addition, strong lasing lines were seen that had been predicted but not previously observed. The team states that to generate the X-ray lasing light, NRL's Pharos III neodymium glass laser was operated at an infrared wavelength, a pulse energy up to 500 joules, and a pulse duration of two-billionths of a second. This laser had been developed by NRL for use in laser fusion and nuclear weapons effects research. For the X-ray laser experiments, the infrared laser light was focused to a line about two centimeters long and 200 microns wide. The emitted X-ray lasing lines were recorded on a small grazing-incidence spectrograph which had been developed by NRL's Space Science Division for their rocket spectroscopy program.

☆ U.S. GOVERNMENT PRINTING OFFICE: 1988 - 209-320 - 814/80378

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