FEATURE ARTICLES

How Changes in Computer Technology Are Revolutionizing the Practice of Chemistry†

GEORGE C. LEVY

NIH Resource, Chemistry Department, CASE Center and Northeast Parallel Architectures Center, Syracuse University, Syracuse, New York 13244-1200

Received July 1, 1988

During the early 1980s, two major developments in computer technology changed the way chemists approached their science. The advent of the microprocessor and then the PC changed experimental chemistry, while the availability of two classes of computer, the superminicomputer and supercomputer, greatly influenced computational chemistry. In the past two years, graphics workstation computers have begun to affect the practice of chemistry by combining fast, high-resolution, multiwindow graphics with superminicomputer power. In 1988, the advent of a new class of computer—the graphics supercomputer—offers extraordinary promise to both theoretician and experimentalist. In these systems, near-Cray compute power is combined with ultrahigh-speed 3-dimensional graphics for unparalleled visualization of molecular processes and other complex events. This is made practical not just by computing and graphics power but by use of ultrahigh internal bandwidths inside the graphics supercomputers. Another major development in scientific computing is the evolving concept of the laboratory computer network. Current network designs include hierarchical configurations incorporating various levels of computers—through supercomputers—either locally or via national or regional networks. New software methods are also having impact on chemical research, allowing, for example, the scientist to better abstract information from noisy or incomplete experimental data. Use of parallelism (multiple CPUs) in new design workstation computers will extend their power, by the early 1990s, past that of current supercomputer mainframes. Within five years the chemist will have $10 million of 1985 computer power on his desk, for considerably less than $100 000, along with visualization tools and software only dreamed of in 1985. This next revolution in computing technology, incorporating advanced artificial intelligence, will probably begin providing "nonprogrammed" problem-solving for the chemist. Computers will design pathways to requested goals, without the necessity of algorithm or code development. Arthur C. Clark's HAL 9000 computer will become fact, and it may sit (entirely) on a chemist's desk.

It is not difficult to recognize today the profound influence of computers on the practice of chemistry. However, many chemists today see the impact of computer technology developed three or more years ago, and they are unaware of extraordinary changes under way at this time. Computing technology develops at a rate far exceeding that of laboratory scientific instrumentation, and innovation is accelerating! New generations of computers appear every 15–18 months. It is interesting to note that the physical characteristics of computers over the past 30 years (Figure 1) reflect only an aspect of this revolution in scientific research. Along with an extraordinary increase in computing power and logarithmically decreasing costs (per unit of computer power), there are other developments in computing architectures, software, and methodology that are changing the way we approach chemical research—both experimental and theoretical aspects.

The use of the computer in experimental science includes data acquisition, experiment control, primary data reduction, postprocessing, and computational support of experimental results. For theoretical scientists, the computer *is* the data acquisition and processing vehicle. Today, the theorist uses the computer to develop fundamental new ideas. An historical view of the impact of the computer in the laboratory over the past 30 years is given in Figure 2, which also attempts to predict trends through the year 2000.

This paper summarizes current trends, discusses their importance to the practice of chemistry, and extrapolates the technology and its applications forward to foresee the state of chemical research in coming years.

RECENT TRENDS

During the early and mid-1980s, two major developments in computer technology changed the way chemists approached their science. The advent of the microprocessor and then the PC changed forever experimental chemistry, while the general availability of two classes of computer, the supercomputer and the minisupercomputer, greatly influenced computational chemistry.

The latest major trend, however, is the development of the workstation computer, first introduced in the early 1980s but just now affecting the practice of chemistry by combining fast, high-resolution, multiwindow graphics and built-in computer networking with near-departmental compute power. In 1988, the advent of a new class of workstation computer—the graphics supercomputer (or *personal* graphics supercomputer)—offers extraordinary promise to both theoretician and experimentalist. In these systems, near-Cray compute power

† This paper is based on invited presentations at the North American Chemical Congress and the Conference on Industrial Applications of Computational Chemistry at the Theory Center, Cornell University, June 1988.

0095-2338/88/1628-0167$01.50/0

Table I. New Graphics Supercomputers

graphics supercomputer[a] and price range	no. of processors and design[b]	benchmarked LINPACKS (64 bit) performance	bandwidth, internal bus (MB/s)
Apollo DN-10000 $80–150K	1–4 RISC + FPUs	5 (1 processor)	150
Ardent Titan $80–225K	1–4 MIPS RISC, 1–4 FPUs	6 (1 processor) ~15 (4 processors)	250
Stellar GS 1000 $100–175K	1 four-stream RISC, 1 FPU (WEITEK 2167), 1 Intel 386 diagnostic processor	8.6	320–1300

[a] All three systems have UNIX operating systems and optimizing C and FORTRAN compilers; each suppotrs X.11 windows and PHIGS+, NFS, TCP/IP, and other industry standards. [b] All processors use reduced instruction set computers (RISC).

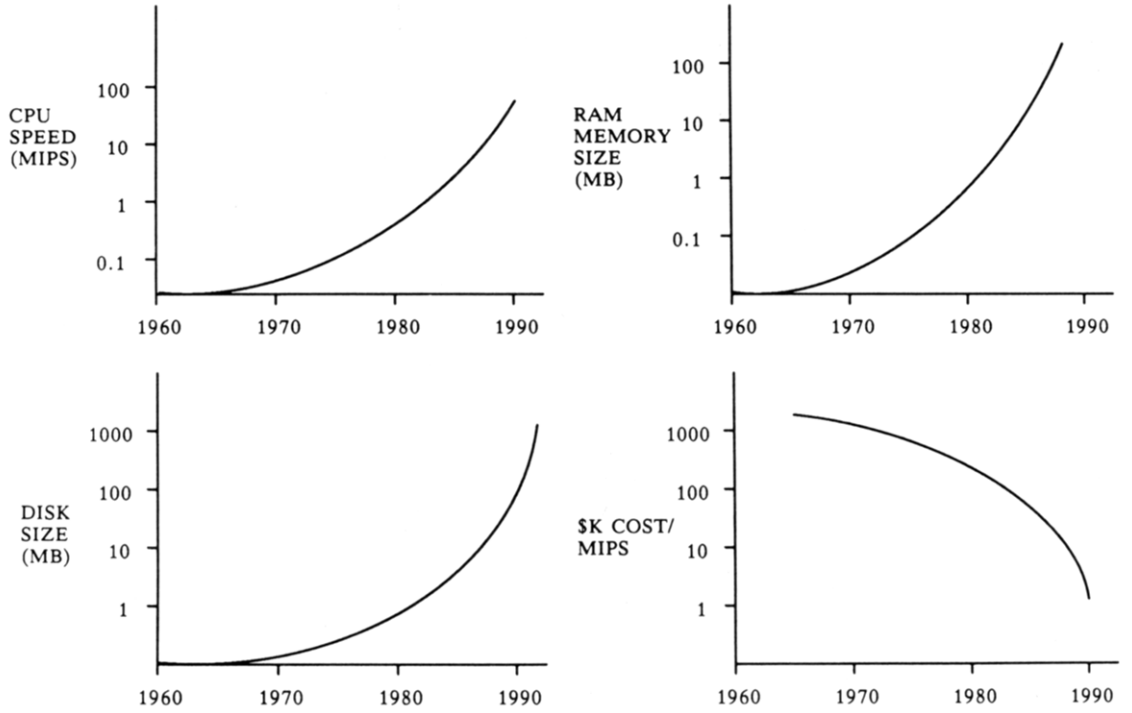

Figure 1. Trends in laboratory computer technology: A surprising logarithmic trend.

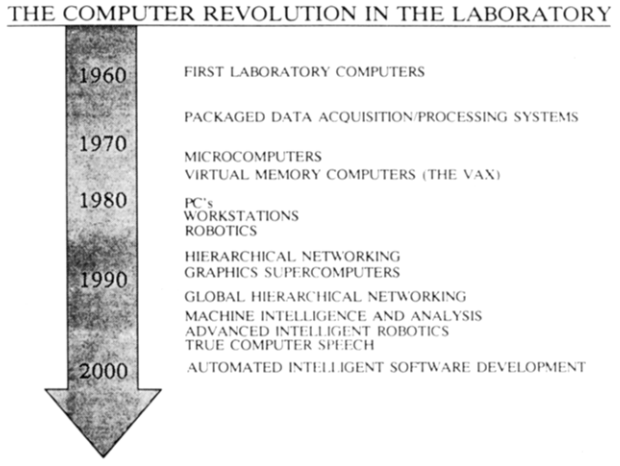

Figure 2. Reviewing and forecasting major developments in laboratory computing.

is combined with ultrahigh-speed 3-dimensional graphics for unparalleled visualization of molecular processes and other complex events. This is made practical not just by computing and graphics power but by use of ultrahigh internal bandwidths inside these machines. This avoids the relatively slow network or other connections between more usual central processing units and graphics systems, which can severely limit interactive modeling and other graphics. For example, the Stellar graphics supercomputer, announced in March 1988, has an aggregate internal bandwidth exceeding 1 GB/s. Other hardware enhancements combine in the Stellar to support fast, smooth 3-dimensional rotations of high molecular weight molecules—even with space-filling shaded-sphere representations. In the Ardent Titan system, up to 4 CPUs and 4 multipipelined floating point processors can simultaneously act to reach compute power close to that of a Cray 1.

The Stellar, Ardent Titan, and Apollo DN-10000 graphics supercomputers (also announced in 1988) will usher in a new age of scientific investigation. *And this revolution is just beginning.* Table I lists the three new systems with some details. While there are obvious similarities among the computers listed in Table I (for example, they all use RISC technology), there are significant differences as well. Each system supports parallelism at several levels, although in functionally different ways. The Stellar uses new-design high-density application-specific integrated circuits (ASICs) for its CPU, while the Ardent relies on advanced implementation of RISC chips supplied by MIPS Computer, as well as multiple pipelined floating point processors. (The deskside Stellar, in fact, is said to have more logic gates than in a room-sized IBM-3090.) The new Apollo RISC CPU architecture, called PRISM (parallel reduced instruction set multiprocessor) is a major depature for the original workstation company.

The real revolution represented by the graphics supercomputer arises from its ability to provide to the scientist a tightly integrated computing and visualization environment. Software now being ported to these computers will provide unprecedented molecular modeling capability. Within one to two years, chemists will be able to see intermolecular contacts, solvation, stereochemical interactions—on the largest molecules

Table II. Computer Speed Limits: PC to Supercomputer

category and examples	minimum system cost[a] ($K) (1988)	64-bit LINPACK benchmark (MFlops)[b] 1988	1991 (est)
PC PC/AT	5	0.012	1
"original" superminicomputer VAX 11/780		0.14	
desktop workstation Sun 4/110, Apollo DN3500, VAXstation 3200	30–40	0.5	3
workstation Sun 4/2608 HP 835SRX, Apollo 4500, Silicon Graphics 4D/80GXT, Masscomp 6400	60–90	1–2	5–10
graphics supercomputer Ardent, Stellar, Apollo	80–150	5–15	25–50
minisupercomputer Alliant, Convex	500	10–20	50–100
supercomputer ETA-10, Cray XMP/48	5000	20–100	1000

[a] Complete system, including disks, tape, color graphics, etc. [b] For 100 × 100 matrices using Fortran coding (J. Dongarra, Argonne National Laboratory, Technical Memorandum No. 23, February 2, 1988); much larger numbers result for 1000 × 1000 matrices, where vector processors become more efficient. For general-purpose computers, advertised MFlops are typically 1 order of magnitude higher than LINPACK numbers.

and with 3D-stereo vision. Two or more orders of magnitude increase in speed for calculation and drawing of smooth-shaded solid atom spheres makes this possible.

The graphics supercomputer is not alone in bringing revolution to the chemist. As indicated in Table II, throughout the spectrum of scientific computing, changes are accelerating. Calculations not practical today will certainly be performed in the early 1990s, and many theoretical approximations will be challenged and tested.

There are other trends that are affecting the practice of chemistry today. Laboratory data acquisition computers can simultaneously capture experimental data, sampling in a routine way, at up to 10 MHz with 12–15-bit accuracy, while the scientist simultaneously does significant UNIX-based postprocessing or theoretical computations. UNIX is ordinarily not oriented toward real-time data acquisition, but a few companies and particularly Masscomp have built hardware and software extensions for UNIX workstations that support simultaneous high-speed data acquisition and general purpose and graphics computing.

Current development in scientific computing rests on several advances in computing technology and applications. Central to these was the advent of computer networking standards and the emergence of standard operating systems.

COMPUTER NETWORKING

Until the mid-1980s, most scientific computing was performed on isolated computers. Over the past three years, extraordinary progress has been made in networking, due at one level to the emergence of the UNIX workstation computer with integrated Ethernet hardware and software and, at another level, due to the emergence of the personal computer. While the personal computer revolution has had the largest impact to date, in terms of the future, the most important of these two trends is the linkage between UNIX and high-speed computer networking. While UNIX is by no means the ideal operating system for scientists, it is rapidly becoming the de facto standard for virtually all scientific computing platforms.

Ethernet computer networks have been available since the early 1980s; however, the existence of several alternative and incompatible Ethernet network protocols (DECnet, XNS, TCP/IP) prevented universal implementation. In the past three years the TCP/IP protocol has emerged as dominant. This nonproprietary protocol is now universally supported on all UNIX workstations and on most mainframes, supercomputers, minisupers, and minicomputers. Even Digital Equipment VMS computers can support TCP/IP through third-party vendor software. (In fact, many UNIX computers now additionally support the DECnet protocol, again via third-party software.)

Ethernet networking is efficient and inexpensive. It is a broadcast topology network (one "linear" cable connects all nodes equally and all nodes can receive any "broadcast") and can cover a multibuilding site with appropriate hardware. Logical extension of an Ethernet-based network beyond one site (a wide-area network) is also possible by use of high-speed optical fibers and hardware/software products that act as bridges between local Ethernets. Simultaneous use of more than one networking protocol (e.g., TCP/IP and DECnet) on a single Ethernet system poses no problem whatever.

While the speed of Ethernet is high by historical standards (10 MHz, with sustained net throughput near 300 kB/s, depending on task and protocol), significant increases in networking speed will be needed in the future. Within two to three years a 100-MHz scheme (FDDI) will become available, and much higher (>100 MB/s) inter-CPU links are now in design and testing. These latter systems will initially be used to link supercomputers and special processors such as advanced graphics systems.

In the U.S., the nearly universal use of TCP/IP for wide-area networking is now allowing a variety of resources to assist the scientist. NSFNet, ARPAnet, and regional networks such as NYSERnet (N.Y. State Education & Research Network) are now bringing national supercomputer resources and other specialized computing to individual scientists via T1 (1.5 MHz) links running TCP/IP. These links can be, in practice, only a factor of 2–3 slower than a local Ethernet, and thus a real revolution is in the offing. In fact, using the X-graphics windows protocol (which is emerging as *the* universal standard), it is now practical for scientists to see on a local graphics workstation high-speed color graphics computing that is actually resident on a machine hundreds or thousands of miles away. Further, virtually all brands of computers and workstations are adapting this fully compatible standard.

The X-windows standard, developed at Project Athena (Massachusetts Institute of Technology, under funding from IBM, Digital Equipment, and others), is designed for network-transparent graphics. Any graphics station or terminal supporting X-server code can display graphics generated by applications running on that machine or on any networked computer that supports X-windows at the client level. So long as the X-client implementation supports the specific connectivity being used (e.g., TCP/IP, DECnet, RS-232), *it does not matter how the connection is made.* Virtually all computers supporting X-client code support TCP/IP-based networks. Table III gives a partial list of computers supporting X by the end of 1988.

X-Windows is very efficient on Ethernet-based networks, and its speed across the network is not greatly slower than local graphics on a typical workstation. With slower connections, X-graphics across the net is of course slower but performance is not degraded significantly across T1 (1.5 MHz) links running TCP/IP, and across 56 kbaud modem connections X still operates quite usefully.

LABORATORY COMPUTER NETWORKS

The nature of modern laboratory computer networks[1] still depends on whether experiments are PC-based (MS-DOS) or

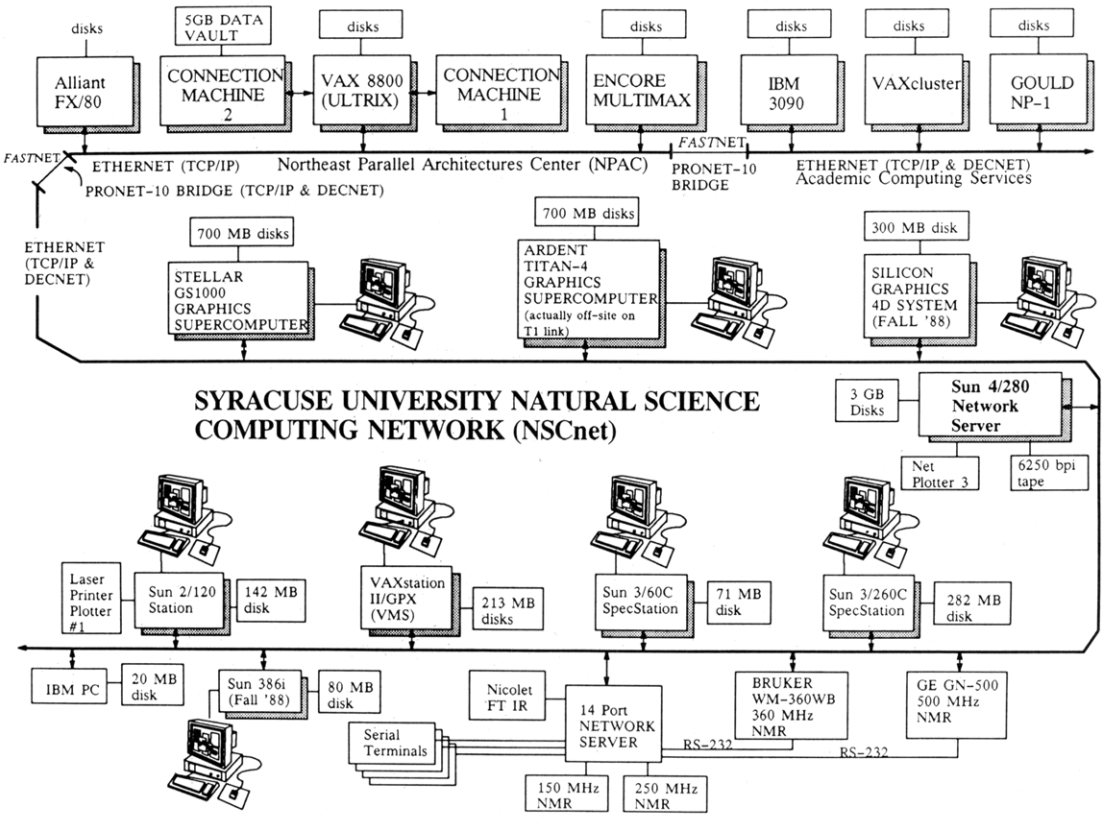

Figure 3. The hierarchical[2] computer network at Syracuse University.

Table III. Representative Computers Supporting the X-Windows Standard by 1988 Year End

(A) server level (includes client level)	
(1) workstations and graphics supercomputers	
Apollo	Masscomp
Ardent	Silicon Graphics
Digital Equipment	Sony NEWS
Hewlett-Packard	Stellar
IBM PC/RT (AIX)	Sun Microsystems
(2) PCs	
AT&T	IBM PS/2 (Unix)
IBM PC (3rd parties)	Macintosh II (Unix)
(B) client level	
(1) minisupercomputers and supercomputers	
Alliant	Thinking Machines Connection Machine
Convex	(through VAX front end)
Cray	ETA-10
(2) superminicomputers	
Digital Equipment VAX (VMS and Ultrix)	IBM 9370
Gould	IBM 43X1
Hewlett-Packard	etc.
(3) mainframes	
IBM 3090, etc.	

workstation/minicomputer-based (UNIX or VMS, RSX/11, etc.). However, this distinction is expected to disappear over the next few years as UNIX comes to new 32-bit PC systems (which actually will become PC-workstation computers). One model for a large scientific computing local area network (LAN), the Hierarchial computer network,[2] has been designed and implemented at Syracuse University. This network, NSCNet, in its current configuration, is shown in Figure 3. The hierarchical computer network has computing nodes with varied capabilities, in a configuration where the applications can benefit from and utilize appropriate nodes. Network graphics is supported throughout, thereby ensuring that a broad range of applications software is widely available, with users able to visualize results from any node.

In the Syracuse University network, NSCNet nodes range in power from IBM-PC to Stellar and Ardent graphics supercomputers. Furthermore, the powerful massively parallel 32 768 processor Connection Machines (with a VAX 8800 front end computer) and Alliant FX/80 systems of the Northeast Parallel Architectures Center at Syracuse University are available for demanding applications that can benefit from parallelization and vectorization. Large tasks can also benefit from the large CPU memories and high disk I/O bandwidths of these systems.

NEW COMPUTER ARCHITECTURES

In the past several years, major developments have occurred to make computers faster and more suited for large computation. This paper will not summarize all of these technical details, except to indicate that some of these trends can affect the way we perform scientific calculations. In particular, most new computer designs incorporate extensive pipelining and concurrency/parallelism. These can enhance scientific vector processing, but optimal execution even on modestly parallel computer architectures requires significant recoding, even in cases where compilers claim automated parallelization.

There are a variety of parallel computer architectures today, ranging from loosely coupled large processors with a large "shared" memory (coarse-grained parallelism) to fine-grained parallel *n*-dimensional arrays of "smaller" processors with dedicated (and sometimes shared) memories that in aggregate can have huge compute power. Examples of the former include the Cray XMP/48 supercomputer, IBM 3090 Model 600 mainframe with vector facilities, Alliant FX/series minisupercomputers, and the VAX 6000 series superminicomputers; examples of fine-grained parallel systems include the Hypercube from Intel, Butterfly and Monarch from BBN, and Connection Machines 1 and 2 from Thinking Machines, Inc.

Some new designs have individual processors with multiple streams of instructions/processing, such as in the Stellar GS1000. Most new designs for scientific computation in-

corporate separate (parallel to the CPU) floating point or vector processing units, often based on Weitek or similar digital signal processing (DSP) chips. The latter, until recently, were limited to 32-bit precision, but current designs incorporate single and dual 64-bit chip sets with speeds only attainable in the early 1980s on that time's supercomputers.

A number of special-purpose processors are being offered today. One type that may prove useful to the future chemist is the database machine. This type of computer is designed to efficiently manage and retrieve information from very large databases. Next generation laboratory information management systems (LIMS) in larger company networks may use this technology, for example.

Another specialized machine is designed for artificial intelligence (AI) languages such as LISP. This type of machine can significantly accelerate applications that are AI intensive. Future laboratory expert systems might utilize this type of technology, incorporated as chip or board-level special processors in general purpose machines.

Looking further ahead, we can anticipate that laboratory computers will acquire observational (sensory) input, particularly vision and tactile feedback for robotic experimentation. These developments, however, will require breakthrough advances in AI software, pattern recognition, and other areas.

SOFTWARE DEVELOPMENTS

Five areas of software development are important to the scientist: (1) computer operating system environments; (2) language compilers and software development tools; (3) discovery of new algorithms and methodology; (4) nonalgorithmic programming and artificial intelligence; (5) realization of new application programs.

(1) Operating Systems. During the mid-1980s, scientific computing was dominated by MS-DOS at the low end and on larger systems by VMS on ubiquitous VAX computers. In 1986–87, while VMS still reigned as the most common environment for chemical computations, UNIX operating systems began to support increased chemical computation. Further, in 1987 the two significant versions of UNIX, System V from AT&T and Berkeley UNIX 4.2–4.3 (commercialized largely by Sun Microsystems), began to merge into a unified UNIX standard, confirmed in early 1988 by a joint announcement from AT&T and Sun. This is an outstanding opportunity for scientific computing, since large applications can be developed to be system independent.

Unfortunately, a group of major vendors including IBM, Digital Equipment, Hewlett-Packard, and Apollo have set up a consortium group to develop an independent (and potentially rival) UNIX standard. As scientists, we can only hope that an irreparable schism does not prevent the further growth of UNIX as a nearly universal computing environment for the next decade. (Note that the author does not promote UNIX as an ideal operating system choice; it is, however, the multivendor standard environment for the next decade.) (It is interesting to note that both of the new generation PC's—IBM PS/2 and Macintosh II—will offer UNIX operating systems.)

(2) Languages. Although FORTRAN 77 is still with us, and in fact is still dominating individual scientific code development, the language C is becoming the standard in commerical scientific programming. C is closely linked to UNIX and is always included in implementations of UNIX (most UNIX operating systems are written in C). It is interesting that Pascal, strongly popular in the early 1980s, is rapidly fading from the scene.

Most companies manufacturing computers oriented for the scientific community are offering vectorizing (and, where appropriate, parallelizing) C and FORTRAN compilers. Not all of these compilers, however, are equally successful at automatically finding opportunities for acceleration of program execution. There are advantages in designing FORTRAN compilers for concurrent acceleration of programs and, to date, less success has been achieved with parallelizing C compilers. The new FORTRAN 8X standard (which may well become FORTRAN 9X) will contain enhanced capabilities for vector and parallel code execution.

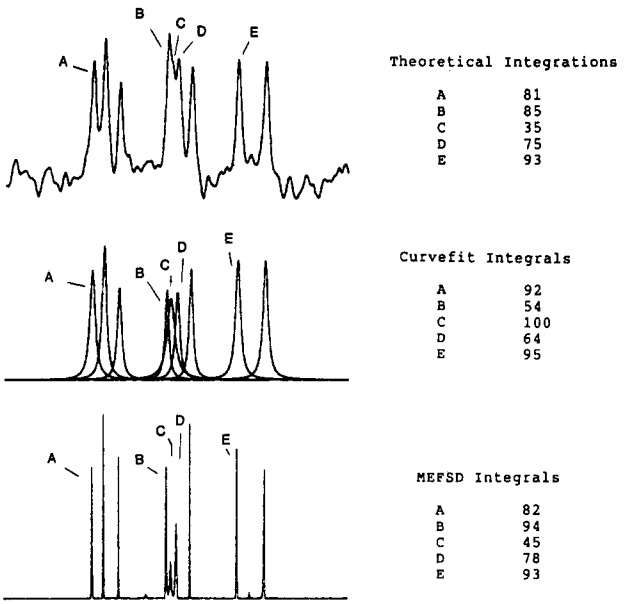

Figure 4. Comparison of maximum entropy spectral deconvolution and Lorentzian curve fitting. For this synthetic data set, the MEM approach gives significantly better quantification, especially for the heavily overlapped C peak.

(3) Discovery of New Algorithms. In many cases, the practicality of large scientific computation rests on development of a new method for a desired calculation. One well-known example is the Cooley–Tukey fast Fourier transform (FFT) algorithm, developed approximately 25 years ago.[3] The advent of new computer architectures, particularly those incorporating massive parallelism,[4] brings a need for a new way of solving problems on the computer. We are accustomed to thinking sequentially, for the most part, and not to working a problem through several hundreds or thousands of channels.

(4) Artificial Intelligence. In 1988, chemists think of artificial intelligence (AI) in terms of expert systems being designed to elucidate molecular structures from spectroscopic or other data. This is, of course, one significant goal for AI, but others are expected to have major impact on future chemical research. Some of these have examples today, such as automated diagnostics or calibrations for complex instrumentation, or the beginnings of intelligent robotics. Another area of current interest is intelligent statistical analysis of laboratory experiment design and data results, using pattern recognition methods, in part derived from chemometrics methodology.

While current and near-term AI applications largely fall under the expert systems category, it is expected that the next decade will see the beginning of automated algorithm development and automated programming (see Figure 2).

(5) Realization of New Application Programs. The practicing chemist, of course, cares most for the scientific results that can be achieved from experimentation or theoretical work. This rests on applications software developed by instrument manufacturers, independent software vendors, academic scientists, and others. The variety of today's chemical applications software is astounding and growing in type and quantity. Much software is available through inexpensive distribution sources such as the Quantum Chemistry Program Exchange (QCPE),[19] which has a broad library of chemical applications, including some very large programs. QCPE and similar

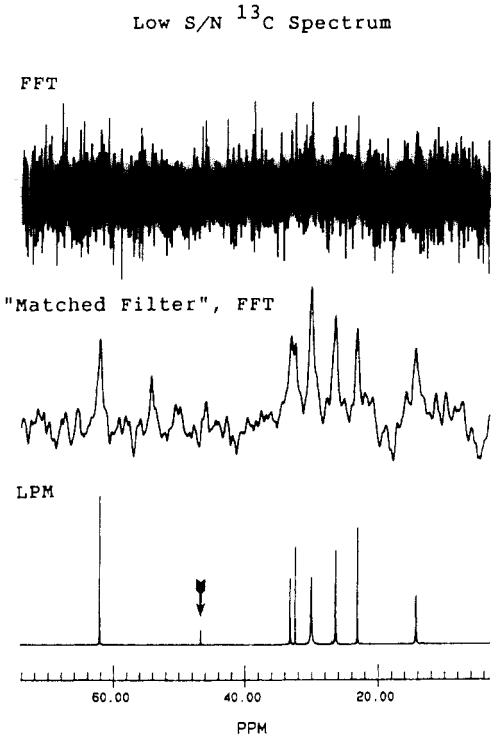

Figure 5. ^{13}C Spectrum of 1-octanol obtained at conditions of low S:N and poor resolution. At top is the spectrum using FFT alone. Application of a "matched filter" exponential apodization (middle) increases S:N but degrades resolution. Forward/backward linear prediction with 5-Hz-line-width reduction gives a reconstruction with enhanced resolution and virtually no noise. The arrow indicates a remaining noise peak that LP was unable to discriminate due to the low signal level. Also, the two resonances near 30 ppm were too close to be resolved by the experimental conditions and thus are represented by one peak in both conventional and LP methods.

sources, however, cannot provide significant support to users of their libraries.

Commercial chemical software ventures provide advanced solutions for a variety of application areas including quantum mechanics, molecular dynamics, spectroscopic and chromatographic data processing, chemical information management, algebraic manipulation, molecular graphics, etc. Over the next five years, many of these areas are expected to undergo revolutionary advances as a result of the technological developments described in this paper. A few of these changes are already visible. For example, several companies are now developing biopolymer molecular modeling software that integrates high-speed conformational calculations with rapid and interactive 3D graphics presentations, using realistic space-filling representations. This was only dreamed of just a few short years ago.

A few examples follow of current advanced chemical applications developed in the author's laboratories. These results center on NMR spectroscopy, but they are representative of other techniques as well and serve as guideposts to the future of computers in chemistry.

SOME APPLICATIONS

In NMR spectroscopy, as in other analytical techniques, the computer has become a central part of the experiment. Unlike several other laboratory methods, however, the experimental requirements of pulse-FT NMR forced an earlier and more demanding role for digital hardware in NMR instrumentation.

During the early and mid-1970s, digital computers in NMR instruments acquired new responsibilities along with data acquisition: experiment control, intermediate storage of data, and some spectral data reduction, such as phasing and simple base-line corrections, plotting. By the end of the 1970s it was clear that the computers within NMR instruments were woefully overworked and, further, that this hardware, which was optimized for NMR data acquisition, was not flexible or powerful enough for significant NMR data processing tasks. The first NMR laboratory computer network with separate data acquisition hardware and a general-purpose multiuser computer was thus born over the period 1978–79.[1b]

Processing of modern NMR experiments requires, at a minimum, the following: (1) pre-Fourier transform operations such as apodization and FID data base-line offset corrections, (2) the Fourier transform itself (which for the largest multidimensional datasets can be a challenge!), (3) postprocessing such as phasing, peak finding, and integration, and (4) output—plotting and peak table listings.

A variety of advanced procedures can also be invoked, such as base-line conditioning and deconvolution and quantitation of overlapping features by curve-fitting[5] or Fourier deconvolutions. Methodologies such as maximum entropy deconvolutions,[6,7] image processing techniques,[8,9] or multivariate analysis[10] are also being proposed as superior to conventional processing for specified applications and tasks. Several examples follow.

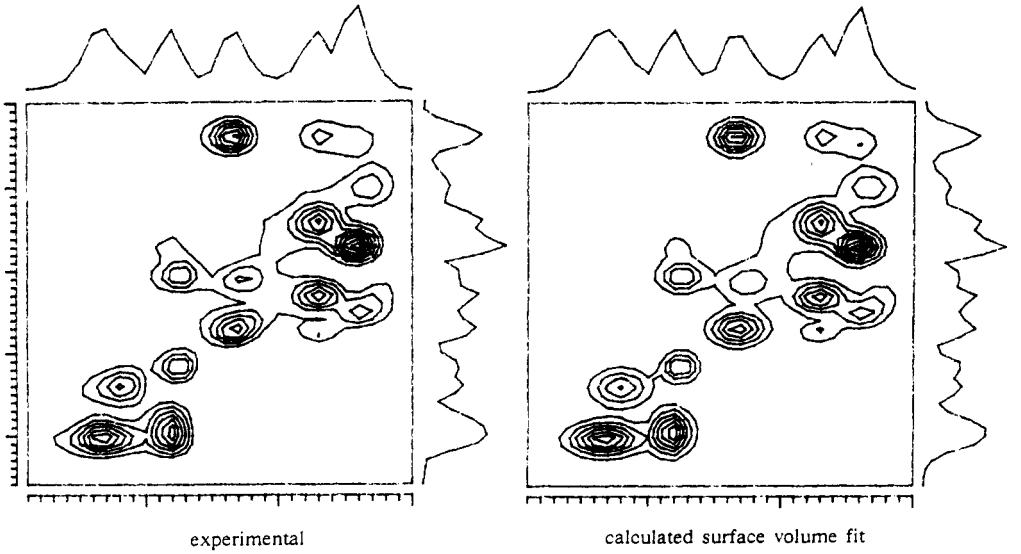

Figure 6. Example of 2-dimensional surface fitting to quantitate overlapping peaks. The software makes initial guesses based on determination aided by cluster analysis.

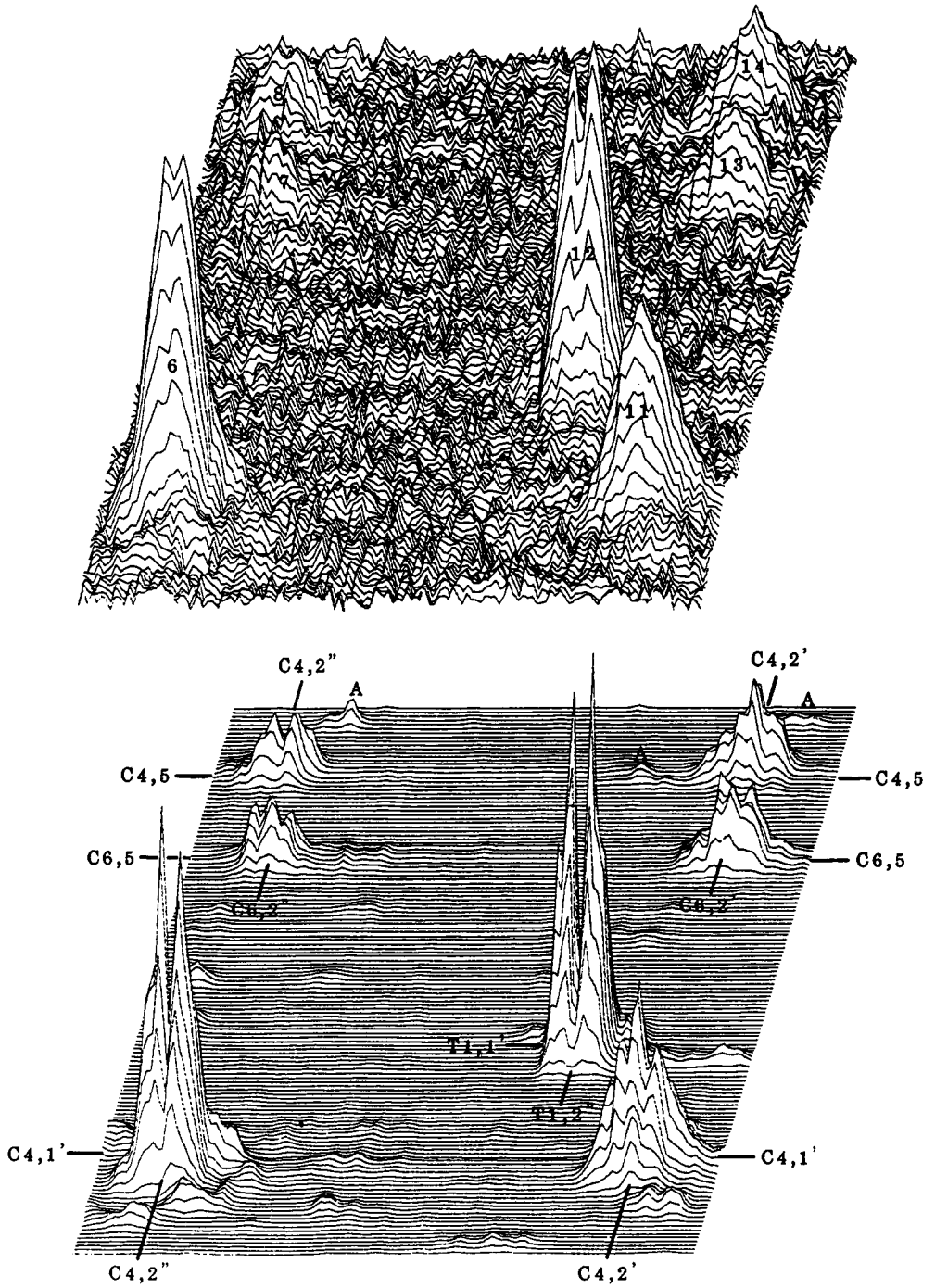

Figure 7. 2-Dimensional spectral deconvolution using the criterion of maximum likelihood (ML). The top spectral segment is a portion of the 2′,2″–1′ and base–5 proton NOESY cross-peak region for the molecule $[d(TAGCGCTA)]_2$. Note the ^1H–^1H spin–spin coupling in the ML reconstruction. Over 1 order of magnitude of signal amplitudes, quantition in the ML reconstruction was seen to be twice as precise as in the conventionally processed spectrum (comparing cross-peak volumes above and below the diagonal).

Maximum Entropy Processing. The maximum entropy method (MEM) can be very useful in signal analysis, where proponents claim significant sensitivity increases versus conventional FFT processing.[6] They also point out that the MEM technique is able to simultaneously "optimize" the spectral signal-to-noise *while* narrowing resonance lines and eliminating peak overlaps. Detractors of the MEM approach correctly point out that without additional input information MEM is not able to detect peaks that are not detectable by careful inspection after FFT processing.[11] Nonetheless, it is clear that MEM dramatically increases the contrast (apparent S:N) between resonance signals and noise while simultaneously affording effective spectral deconvolution.[11,12] This would not be possible by conventional self-deconvolution, even for moderately high S:N spectra (order of 100).

It is also possible that MEM can, in the future, be combined with ancillary data information or that similar but distinct signal analysis methods (such as the maximum likelihood method) will demonstrate superior signal discrimination and spectral quantitation. In tests done thus far[12] on synthetic data and real mixtures of compounds, MEM on extracted spectral segments of low S:N has been shown to give quantitation ranging from equal to that from Lorentzian curve fitting to 2.5 times more precise. Quantitation of small peaks in the MEM approach is limited by coexisting large peaks, but a dynamic range exceeding 10 is achievable with ordinary iterative MEM calculations, and much higher spectral dynamic ranges can be accommodated by inverse transformation of spectral segments.[7,12] Figure 4 shows an example of a maximum entropy quantitation compared with quantitation via

automated curve fitting. In this case the MEM treatment proved quite useful.

Linear Prediction Processing. Another technique that has been proposed for determination of noisy spectra, the linear prediction method (LPM),[13,14] can dramatically enhance the appearance of spectra and also afford complete deconvolution of overlapping lines. An example is given in Figure 5. This method searches for Lorentzian lines (exponential decays) and calculates a best fit to the input data, as a set of theoretical Lorentzians. The method is quite good for contrast enhancement but is not generally reliable as spectral S:N approaches 1.0. Further, calculation times for even moderately complex spectra can be quite excessive.

2D Spectroscopy. NMR signal analysis methods have analogies in multidimensional NMR spectroscopy and imaging. Use of automated 2-dimensional surface fitting on crowded cross-peak regions (Figure 6) can give excellent quantitation for 2D spectra of reasonable S:N. This type of scheme[15] is not simple to implement, however, since errors in determination of both spectral base plane and resonance bandshape can, unless well accounted for, give rise to significant errors. In the example shown, the line shapes are apodized appropriately and corrections are applied for deviations in the local base plane.[15]

As an alternative to surface fitting, maximum entropy and other constrained restoration techniques can effectively deconvolute 2D spectra, reducing spectral overlaps while simultaneously increasing (dramatically) apparent S:N. Figure 7 shows an example: a section of a 2D NOESY spectrum obtained on a DNA duplex oligomer (minihelix) processed conventionally and also by maximum likelihood deconvolution.[16,17] The stacked plots shown in Figure 7 give a good perspective of the utility of these techniques. The deconvolution succeeds in defining ^1H–^1H spin–spin coupling multiplets. Also, in tests of this 2D deconvolution module, it was determined that along with line sharpening and noise suppression spectral quantitation improved considerably, compared with conventional processing. The compute time for these calculations is of the order of 15 min to 1 h on Sun-3 or Sun-4 workstations for a single 256 × 256 data point cross-peak region, but processing of an entire 2K × 2K spectrum can be achieved by using a minisupercomputer or graphics supercomputer in a few hours. The considerable time savings resulting from use of vector processors comes about since such deconvolutions require 250–400 2D Fourier transforms. Other 2D restoration criteria that do not require FT calculations can be selected, and these are generally faster but may not achieve as much enhancement for these types of data.

Recent discussions extending 2D NMR spectroscopy to three and even four dimensions[18] offer exciting possibilities to simplify spectral elucidation of very complex molecules. Just as 2D FT NMR has increased dispersion of signals by placing them in two dimensions, the higher dimensional spectra will have significantly enhanced information content. However, the capabilities of current data processing, with restrictions on our abilities to utilize the new data, will limit these studies for some time. In addition, the experiments themselves require greatly extended periods, except in cases where only selected frequencies are sampled.[15]

CONCLUSIONS

The impact of the computer on modern chemistry is profound, affecting virtually all aspects of experimental and theoretical studies. During the decade of the 1980s, advances in computer hardware technology have accelerated dramatically. The advent of PC's, workstations, and supercomputers have each had their impact. In 1988, the birth of a new class of computers combining aspects of all three of these systems is expected to change forever chemical sciences and their ability to help man to overcome natural limitations in his world.

Changes in computing technology are expected to continue, beyond our ability to forecast the characteristics of computers at the turn of the century. Advances in machine intelligence and automatic programming will lead to revolutionary application of computer technology in the new millenium—which is just a few short years away.

ACKNOWLEDGMENT

Support from the Division of Research Resources, NIH (Grant RR-0317), Syracuse University, and several corporate sponsors (General Electric, Digital Equipment, Star Technologies, Apollo, Sun, Silicon Graphics, Stellar, Ardent, New Methods Research Inc., and others) is gratefully acknowledged, as are technical contributions over the past several years from (S.U.) Pascale Sole, Roy Hoffman, Anil Kumar, Mark Roggenbuck, William Curtiss, Marc Delsuc, Hans Grahn, Phil Borer, Anthony Mazzeo, Steve LaPlante, Karl Bishop, Mats Josefson, (Cornell University) Feng Ni, (New Methods Research, Inc.) Alex Macur, Frank Delaglio, John Begemann, Elliot Dudley, Molly Crowther, John Stanley, David Sternlicht, and others.

REFERENCES AND NOTES

(1) (a) Levy, G. C., Terpstra, D., Eds. *Computer Networks in the Chemical Laboratory*; Wiley-Interscience: New York, 1981. (b) Levy, G. C.; Terpstra, D.; Dumoulin, C. L. In *Computer Networks in the Chemical Laboratory*; Wiley-Interscience: New York, 1981; Chapter 5.
(2) Levy, G. C. *Spectroscopy (Springfield, Oreg.)* **1988**, *3*, 30–36.
(3) Cooley, J. W.; Tukey, J. W. *Math. Comp.* **1965**, *19*, 297.
(4) Potter, J. L., Ed. *The Massively Parallel Processor*; Massachusetts Institute of Technology: Cambridge, MA, 1985.
(5) Kumar, A.; Sotak, C. H.; Dumoulin, G. L.; Levy, G. C. *Comp. Enhanced Spectrosc.* **1983**, *1*, 107–115.
(6) (a) Sibisi, S. *Nature (London)* **1983**, *301*, 134. (b) Sibisi, S.; Skilling, J.; Brereton, R. G.; Laue, E. D.; Brereton, J. *Nature (London)* **1984**, *311*, 446. (c) Laue, E. D.; Skilling, J.; Stauton, J.; Sibisi, S.; Brereton, R. G. *J. Magn. Reson.* **1985**, *62*, 437.
(7) Ni, F.; Levy, G. C.; Scheraga, H. A. *J. Magn. Reson.* **1986**, *66*, 385–390.
(8) Bolton, P. H. "Image Processing of Two-Dimensional NMR Data". *J. Magn. Reson.* **1985**, *64*, 352–355.
(9) (a) Sole, P.; Delaglio, F.; Macur, A.; Levy, G. C. *Am. Lab. (Fairfield, Conn.)* **1988**(Aug). (b) Sole, P.; Delaglio, F.; Levy, G. C. *J. Magn. Reson.* (in press).
(10) (a) Grahn, H.; Delaglio, F.; Roggenbuck, M. W.; Levy, G. C. *Abstracts of Papers*; 29th ENC, Rochester, NY, 1988. (b) Grahn, H.; Delaglio, F.; Roggenbuck, M. W.; Levy, G. C. "Multivariate Techniques for Enhancement of Two-Dimensional NMR Spectra" (submitted for publication). (c) Grahn, H.; Delaglio, F.; Delsuc, M. A.; Levy, G. C. "Multivariate Data Analysis for Pattern Recognition in Two-Dimensional NMR". *J. Magn. Reson.* (in press).
(11) Freeman, R. *A Handbook of Nuclear Magnetic Resonance*; Longman: Birmingham, AL, 1988.
(12) Mazzeo, A. R.; Delsuc, M. A.; Kumar, A.; Levy, G. C. *J. Magn. Reson.* (in press).
(13) Barkhuijsen, H.; De Beer, R.; Bovee, W. M. J.; Van Ormondt, D. *J. Magn. Reson.* **1985**, *61*, 465–481.
(14) Delsuc, M. A.; Ni, F.; Levy, G. C. *J. Magn. Reson.* **1987**, *73*, 548–552.
(15) Delaglio, F.; Grahn, H.; Sole, P.; Kumar, A.; Hoffman, R. E.; Begemann, J.; Macur, A.; Levy, G. C. Unpublished results.
(16) Ni, F. Unpublished results.
(17) Bishop, K. D.; LaPlante, S.; Borer, P.; Levy, G. C. Unpublished results.
(18) (a) Greisinger, C.; Sorenson, O. W.; Ernst, R. R. *J. Magn. Reson.* **1987**, *73*, 574–579. (b) Vuister, G. W.; Boelens, R. *J. Magn. Reson.* **1987**, *73*, 328–333. (c) Hoffman, R. E.; Davies, D. B. *J. Magn. Reson.* (in press).
(19) QCPE, Department of Chemistry, Indiana University, Bloomington, IN 47405 (Richard W. Counts, Ed.).