provision for instrumentation, but provision of instrumentation is in competition with many other objectives of URI. It seems to me that the ability to acquire significant instrumentation resources from the DoD is now substantially diminished and likely not to return unless political pressure is used.

In summary, my points are:

- You don't use what you don't own.
- Standardization aids communication and algorithm exchange and thus is highly desirable.
- Movement from minis to workstations seems prevalent and also desirable.
- The time is now to explore supercomputing.
- Provisions for maintenance and support personnel are key elements of planning.
- Acquisition of equipment (other than supercomputer access) from federal sources is likely to be somewhat more difficult in the future.

I mention in closing that Wegman (1986) contains some personal perspectives on how computing relates to statistics.

ADDITIONAL REFERENCE

WEGMAN, E. J. (1986). Midcourse musings. *IMS Bull.* **15** 238–241.

Rejoinder

William F. Eddy

I would like to thank all of the discussants for their uniformly positive comments; I wish they had been members of the Workshop. I would also like to take this opportunity to publicly thank all of the members of the Workshop for their hard work that led to the report. I am sorry that the publication schedule prevented them from having the opportunity to join me in this response.

The activities begun in the Workshop are continuing; we are organizing a session at Computer Science and Statistics: 19th Symposium on the Interface to be held at Temple University on March 8–11, 1987. This session will provide what I hope becomes a continuing public forum for discussion of both technical issues and some of the broader matters raised by this report.

SUPERCOMPUTERS

Prem Goel, David Scott, and Ed Wegman all mentioned the National Science Foundation (NSF) (and other) supercomputer centers. I agree completely with Wegman's principle: *you don't use what you don't own.* I also agree with Goel's recommendation to let others provide access to these centers. And I agree with Scott that supercomputers are not a panacea. Personally, I am not sure that supercomputers will have much positive impact on statistical research. Supercomputers are very good at doing linear algebraic calculations but are distinctly not cost effective for many other kinds of calculations.

An hour on my local Cray X-MP/48 is currently valued at $1000. An hour on a Cray X-MP is roughly equal to 30 hours on a VAX 11/780 if the particular problem is not amenable to vectorization; if the problem is totally vectorizable a Cray hour is roughly equal to 300 VAX hours (although, see Dongarra and Hewitt, 1986, for a report on a particular calculation where a Cray hour is roughly equal to 7000 VAX hours). I am able to buy a VAX 11/780 equivalent for $6000. I would much rather have 10 such machines than 60 hours on a Cray because they will support a much wider range of computing activities (and they will last longer). The most common statistical use of supercomputers is for large simulation experiments. It is interesting to note that if such an experiment is not vectorizable but is decomposable into several parallel computing tasks (such as one for each independent sample), the ten VAX equivalents together operate at roughly one-third the speed of the Cray.

The major negative impact of supercomputer centers on statistics comes from the developing sense within the funding agencies that general computational needs are being satisfied by the national centers together with a few dollars in individual research grants to buy inexpensive workstations. While workstations and supercomputers can satisfy a large fraction of the needs, there will continue to be highly diverse and specialized needs for other sorts of computing equipment. Section 5.2 of the report tried to point out that some aspects of research in computational statistics are more like computer science than like applied mathematics; I believe that graphics and parallel computation both provide fertile ground for statistical research and both require specialized equipment that is expensive and difficult to acquire without substantial external support.

STANDARDIZATION

Doug Bates, Andreas Buja, Ed Fowlkes, Jon Kettenring, David Scott, and Edward Wegman have all referred to our recommendation concerning

standardization. Lest there be any misunderstanding, I would like to give some examples to clarify what I believe was the intent of the recommendation.

A department, which is new to the computing business, would be ill-advised to purchase one VAX/VMS monochrome workstation, one IBM AT personal computer, and one SUN 3 color workstation. Although each has some highly desirable features (and certain features of each might even be deemed essential for support of some particular research activity within the department) the total incompatibility among the three hardware/software systems prevents (1) the development of a common environment among all researchers in the department and, hence, the development of a useful "institutional" memory; (2) the intercommunication of data, programs, or text among graduate students, faculty, and staff; (3) the sharing of on-line disk storage and off-line backup storage; (4) the use of an alternate machine in case of equipment failure; (5) economies of scale in the acquisition of hardware upgrades, software, or maintenance; etc.

As another example, the Statistics Department at Carnegie-Mellon University (CMU) chose to use the VMS operating system when it acquired its first VAX, primarily because VMS supported the DECnet network protocol which was in wide use on campus and which provided access to the central Xerox 9700 laser printer. It was natural then to continue using VMS as we started acquiring VAX workstations. We have recently acquired a number of IBM RT workstations and AT&T 3B2/400 supermicrocomputers; these run, respectively, 4.2bsd Unix and System V Release 3 Unix operating systems and, respectively, TCP/IP and 3BNET network protocols. Thus, we now have many kinds of hardware, several different operating systems, and at least three different network protocols. While our operation is not as efficient as if everything were identical, we now have a broad enough base of experience that we do not have major difficulties using this diverse environment except for intercommunication among the different systems. In the meantime the University has decided to make TCP/IP its standard network, in part because NSF uses this protocol for its supercomputer network and in part because the Department of Defense uses this protocol for ARPAnet. By the time this appears in print we expect to have made TCP/IP our standard network protocol.

As a third example, a department which decides to use 4.2bsd Unix together with TCP/IP and the SUN Network File System has adopted a system software standard which actually provides a great deal of flexibility in the acquisition of hardware. While this environment has a number of undesirable features, many major vendors provide hardware and software which is compatible with this environment. Adopting this standard provides a very large number of options in the selection of equipment. On the other hand if all other systems at the local university are IBM equipment, then the decision to adopt such a system software standard would be foolhardy.

UNIVERSITY COMPUTER CENTERS

Unlike Wegman, I suspect that despite the proliferation of workstations, university computer centers will not become obsolete. At CMU the development of a campus wide computer network appears to have imparted new life to the central computer center as it attempts to transform itself from a provider of CPU cycles into a provider of communications services (a local "phone" company), hardware and software maintenance services, hardware sales, etc. The installation and maintenance of the university network (with something approaching 15,000 connection ports) is a *very* big job. And just keeping abreast of the changing technology is a major task. When the CMU Statistics Department moved into new quarters in January 1984 we had provided for our future communication needs by installing large conduits in every interior wall and installing one telephone outlet and two standard terminal outlets (RS-232-C) in every office. In the intervening two and one-half years we have rewired each office three times: first, to provide Ethernet service for the workstations we have acquired for our research; second, to provide IBM token ring service for the campus wide educational workstation network; and third, to provide digital telephone service. I am very thankful the conduits in the walls are large and would emphasize the need for careful planning for future communication needs.

A STATISTICS NETWORK

Bates has pointed to a significant area where statisticans lag behind other disciplines in their use of department and individual computer facilities: the use of national networks for the semiautomatic distribution of data and software. Logically, this use should be preceded by the development of centralized mail facilities of the type described by Buja, Fowlkes, and Kettenring. As they note, sending and receiving mail between two individuals on different networks can be closely akin to a research problem; I find the send and receive paths are often totally different for one individual correspondent. Such a centralized service can provide not only automatic mail forwarding but also an on-line directory, bulletin boards for the community at large and for specialized subgroups, and data and software distribution services. Starting up such a service will cost something in the neighborhood of $75,000–$100,000. But, once established, continuing service of this type, piggy-backed upon existing (inter)national networks, can be provided for much less than $20,000 per year. I would rank this as the single

most important computer-related activity that statisticians are neglecting now.

GRADUATE EDUCATION

Lynne Billard and Jessica Utts have both raised the issue of education of graduate students. There is no substitute for experience as a teacher; thus, in departments such as my own, students quickly become moderately good at dealing with the idiosyncracies of the local system and getting basic tasks completed. We separate training in the use of the facilities from education about computation more generally. We begin each year with several (noncredit) lectures introducing the facilities and their use. In a forthcoming issue of *The American Statistician*, some of my colleagues and I have described what we believe to be the heart of graduate education in computational statistics (Eddy et al., 1987).

PUBLICATIONS

Buja, Fowlkes, and Kettenring have made a recommendation to the profession which I would like to underline here: the adoption of a very small number of typesetting systems as standards. It is interesting to note that the original version of this report was prepared on a 300 dot per inch laser printer in the Universe type font and then later reformatted for a phototypesetter in the Times Roman font for printing and distribution. The originals of four of the Comments were prepared on similar laser printers in Times Roman and only three were traditional typewriter copy. Unfortunately, all of them had to be retypeset for publication here, with the consequent reproofreading and undiscovered new errors.

The major document production systems used by statisticians, in decreasing order of use, are *troff*, TEX, and Scribe. TEX is probably the best of these for mathematical formulae (although *eqn* afficionados might argue) and Scribe is probably the best of these for general document production (although LATEX fans may argue). This could be taken as an argument for making *troff* the standard. I hope that the major statistical associations can reach a general agreement on this matter because it is already important and can only increase in importance. I believe that both the AMS and SIAM have already addressed this problem and have "adopted" standard systems; TEX by the AMS and *troff* by SIAM. Standardization across societies would be highly desirable.

INTERDEPARTMENTAL COOPERATION

I congratulate Prem Goel on his cooperative arrangement with the Mathematics Department. I doubt if mathematics is the right department with which most other university statistics departments should join forces. Some statistics departments have opted for isolation of their computing facilities (an unfortunate choice); more commonly, statistics departments have affiliated with their local computer science department. We in Statistics at CMU cooperate most closely with our Psychology(!!) Department. In any case, I strongly endorse cooperation with other groups but would caution that too close an affiliation can defeat the purpose of owning and operating one's own facilities.

Utts has made a number of recommendations which relate to interdepartmental cooperation. I can only applaud these ideas and add that some universities provide a good environment for statistics and some do not. Collaborative research should be strongly encouraged; application of statistics (which might lead to new theory) should be more highly valued than the theory of statistics (which lead to new applications). In this vein, Jerry Sacks and Ingram Olkin are co-chairmen of a committee which is currently preparing a report on the opportunities for cross-disciplinary research in statistics. I hope their report will stimulate more interdepartmental collaboration of statisticians and suggest that computation can provide a sharp focus for a great deal of cross-disciplinary research.

L'ENVOI

I feel the need to repeat here the point made by Lynne Billard concerning the rapid change in technology. Computation is a continuing expenditure; it is not sufficient to purchase a superminicomputer and then sit back. In Statistics at CMU the capital funds committed to computing have averaged nearly $200,000 per year over the last 5 years; roughly one-third has come from external funding agencies, one-third from hardware vendors, and one-third from internal sources. We are now throwing away (literally!) state of the art microcomputers purchased 5 years ago.

As with anything revolutionary, computing is a sword (one edge is enough for statistics). There is a serious danger that this sword will cleave the statistical research community into the *haves* and the *have nots*. Those with good facilities are going to find their research and educational programs supported and broadened by computation; those with poor or no facilities are going to continue in the old ways.

ADDITIONAL REFERENCES

DONGARRA, J. J. and HEWITT, T. (1986). Implementing dense linear algebra algorithms using multitasking on the Cray X-MP-4 (or approaching the Gigaflop). *SIAM J. Sci. Statist. Comput.* **7** 347–350.

EDDY, W. F., JONES, A. C., KASS, R. E. and SCHERVISH, M. J. (1987). Graduate education in computational statistics. To appear in *Amer. Statist.* **41**.