A framework for the fifth generation

by BRIAN R. GAINES
University of Toronto
Toronto, Ontario, Canada

ABSTRACT

The Japanese initiative in scheduling a development program for a fifth generation of computers has shocked a drowsy West into realizing that computer technology has reached a new maturity. We are ready to take a step forward and integrate into systems the advances in very-large-scale integration, artificial intelligence, database management systems, and the human–computer interface of the last decade. Suddenly, work on the fringes of the computer industry, particularly that in artificial intelligence, is perceived as central and of commercial and military strategic importance. This paper examines the economic, historic, social, and technical logic behind the fifth-generation program from several perspectives. It gives a basis for evaluating the program, responses to it, and its effect on our industry and society.

INTRODUCTION

The Japanese initiative in scheduling a development program for a fifth generation of computers[1,2] has shocked a drowsy West into realizing that computer technology has reached a new maturity. We are ready to take a step forward and integrate into systems the advances in very-large-scale integration (VLSI), artificial intelligence (AI), database management systems (DBMS), and the human–computer interface (HCI) of the last decade. Suddenly, work on the fringes of the computer industry, particularly that in AI, is perceived as central and of commercial and military strategic importance.

This recognition, and the accompanying prestige and funding, is clearly welcome to many, and Western research has been quickly redirected to match the Japanese program.[3,4] Wry remarks may be made that prophets have no honor in their own country. However, it is accepted that the Japanese track record of competition in the auto and semiconductor industries[5,6] does give this nation the right to credibility for its high-technology planning. This credibility is now gratefully accepted as legitimizing much related research in the West.

The activity triggered by the Japanese initiative is not a good environment in which to examine the current status of computing technology and our long-range objectives. The objectives have been set. The paradigm has become accepted, and a new value system has been widely adopted. Such shifts are common in science and technology[7] and the very fact that one has occurred will soon be forgotten.

However, there are many reasons why it is appropriate to look now at some of the deeper factors underlying the fifth-generation programs. AI was oversold in the 1960s and there was a backlash in the 1970s. Is fifth generation being oversold now and, if so, what might be the consequences? The Japanese program looks like system integration of every frontier technology in computer science. What will we have at the end and why should we want it? The program has been justified in terms of economics, market requirements, and leadership in technology. Are there other rationales for it? In the computer industry there has always been an interplay between technology-led and market-led developments. What are the technology and market forces on fifth-generation developments? The pioneers of computing technology raised questions about its long-term impact on the structure of our society. What are the implications of achieving the fifth-generation program goals?

These are not just broad questions of general interest. They also have key relevance to the planning of the fifth generation and responses to it. In establishing such activities we are engaged in both predictive and normative technological forecasting. Past forecasts for the computer industry have proved notoriously wrong. Some promising technologies have not matured; some have advanced much more rapidly than expected. In addition, new and unexpected technologies have emerged. In general, overall trends have occurred well before their predicted arrival dates. Can we learn any lessons from this history in planning current programs? If we cannot predict, how do we manage the change, the unexpected failures, and the unexpected opportunities? Will spinoffs outside the program be more important than the goals achieved within it?

This paper provides a framework for discussing these questions by examining fifth-generation computing from several perspectives: historical, social, economic, and technical. It suggests that we are in a time of fundamental change in the effect of computing on society, and society on computing, and that unusual perspectives from outside the industry can help us to focus on the key issues within it.

SURPLUS CHIPS AND THE PULL OF THE CONSUMER MARKET

Much of the publicity attending the Japanese proposals has emphasized the massive computing power envisioned for next-generation machines. It is true that MIPS abound and that MLIPS have been introduced[1] (one LIP is one logical inference per second), but this does not mean that the overall focus is on giant machines. The dominant objective of the fifth-generation proposals is *ease of use of computers by people,* and it is possible to see this as having come out of the inexorable logic of VLSI chip manufacturing. We are moving toward a surplus of chips that can be absorbed only by the worldwide consumer market.

Many of those attempting to buy certain processor and memory chips today will find the concept of a surplus ludicrous. However, scarcity of the latest devices is brought about by the current extreme price competition in small computer systems. Simplistically, all micros have much the same hardware and are differentiated primarily by cost. Sixty-four kilobyte rams and processors with inbuilt dma and communications make a big difference in board size and manufacturing cost. The newest devices will always be in great demand and short supply. For the previous generation of devices history suggests there will always be a price war between suppliers because of oversupply.

This oversupply problem will worsen with the increasing power of the coming generation of chips. The increasing number of gates in modern VLSI is achieved through decreasing element size that increases speed. Increases in complexity and in speed are positively correlated rather than subject to tradeoff. Hence the potential power of modern ic's has already

become extremely high. The problem is to know what to do with this power. Processor, memory, and support chips account for the bulk of high-volume demand, with communications and graphics forming a second tier.[8] The $12 billion turnover of the semiconductor industry corresponds to some 25 chips a year for every person in the world—about 1 million transistor equivalents a person a year!

We are manufacturing more and more of a product whose power is already great and increasing. The professional markets absorb much of this capacity now, but only the consumer markets can support it in the long term. However, computer technology plays as yet only a major role in these markets—one limited by the technical skills required of computer users. The key thrust of the fifth-generation program is to overcome this limitation and make computers accessible to all by moving from information processing to knowledge processing, from a machine-like interface with the user to a human-like interface through speech, writing, and vision.

Figure 1 shows the economic logic behind the fifth generation and shows the way this leads to the specified technical program. The chip surplus problem is resolved by aiming for consumer markets, and this requires that anyone be able to use a computer and have the motivation to do so in a domestic context. This in turn requires improvements to the human–computer interface, realistic simulation of real and fantasy worlds, and the capability to encode expertise for use by others. These requirements then lead to projects for speech recognition, high-resolution graphics, languages for knowledge-processing, and so on. The customizing of computers is an important capability that must not be lost and yet the volumes involved are too large for professional programming. Hence automatic programming in some form or another is also necessary. Thus the logic of the marketplace leads to the projects that form the fifth-generation program.

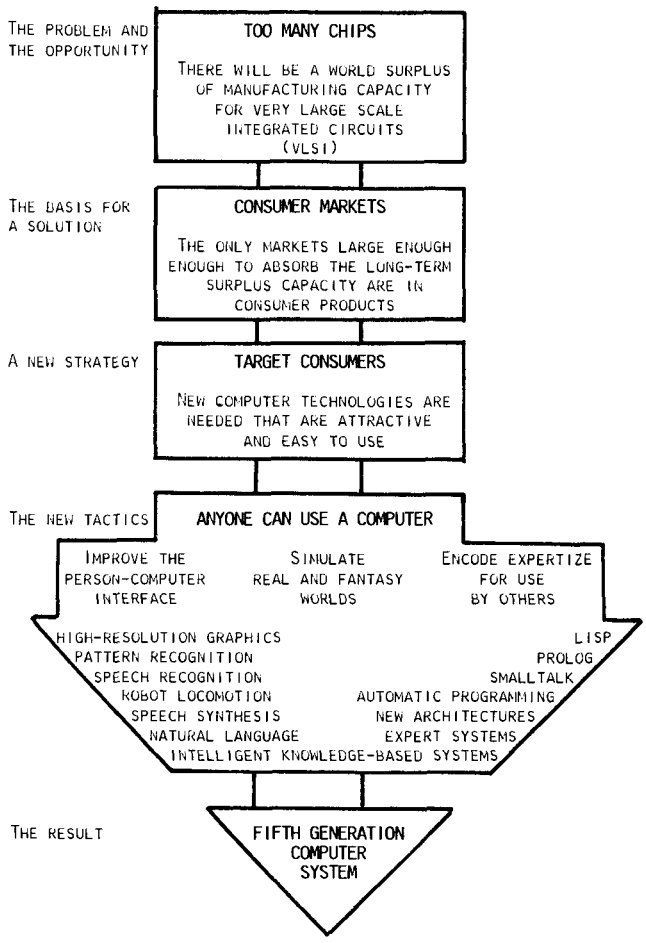

Figure 1—The logic behind the development of fifth-generation computer systems

This is not to say that a new generation of giant machines will not also come out of the program. They are probably necessary in any event as design tools for the advanced technologies underlying the consumer market machines. Yesterday we needed the large machines to help us in designing chips. Tomorrow we will need them to help us in designing knowledge structures. However, the dramatic swing of computer technology into the consumer market through the advent of the personal computer[10] will go much further as the fifth-generation objectives are achieved.

THE CHANGING GOALS OF AI RESEARCH

Perhaps the greatest element of surprise in Japanese program was its dependence on advances in AI research. This has always been an important part of frontier research in computer science but has not had major commercial significance. However, the argument of the previous section leads to a basic requirement for computing technologies that are part of AI, notably knowledge processing, natural language communication, and speech recognition. The commercial requirement that AI be routinely available marks the beginning of a new era for this field of research. The cycle of AI research shown in Figure 2, from overoptimism in the 1960s through disenchantment in the 1970s to commercial success in the 1980s, is a significant historic framework for fifth-generation developments.

The accepted starting period for Era 1 of research on AI is the late 1950s, with work by Newell and Simons on the General Problem Solver; McCarthy on programs with common sense; Selfridge on Pandemonium; Rosenblatt on the Perceptron; Widrow on Adalines; Solomonoff on mechanized induction; and McCulloch, Farley, von Foerster, Greene, and others on neural nets. Minsky's 1961 survey[11] (citing all these researchers) gives a fairly comprehensive feeling for this era worldwide. The logic behind much of the work is that new computer forms, organized as aggregates of simple, self-organizing elements, could be induced to perform a task by means of learning mechanisms, such as mimicking, reward, and punishment, similar to those of animal learning.

These objectives and this type of work characterized the worldwide goals and approaches at that time. They should be placed in the context of the early development of digital computers, where slow, expensive, unreliable machines with mercury-delay line memories were still in use, programmed in various forms of assembler or autocode. The JOSS, Culler Fried, PLATO, and Project MAC experiments on time-

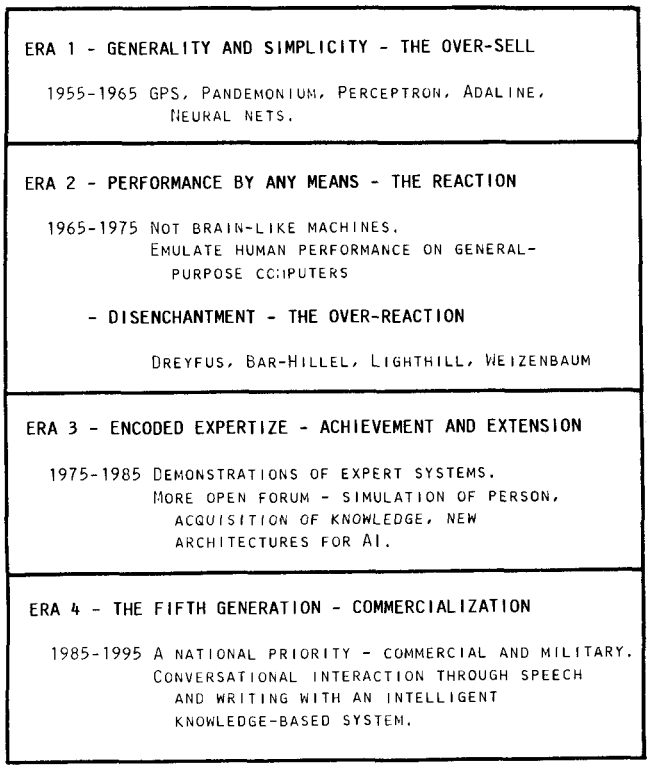

Figure 2—Changing goals in the different eras of AI research

sharing were part of the same era.[12] The atmosphere was one of immense excitement over the potential for computer systems to lead to the "augmentation of human reasoning" and "man-machine symbiosis," to use the terms of books and papers of that era.[13]

The excitement and the funding died down in the 1960s as a result of two main factors. First, the work did not fulfill the promises made on its behalf: neural nets did not self-organize into brains; learning machines did not learn; perceptron-like elements did not recognize patterns. Second, the conventional digital computer became more reliable, smaller, faster, and cheaper, and the advantages of its sheer generality became widely realized. Effort switched from designing brain-like machines to emulating human-like activities on general-purpose computers. The focus of attention also became human *performance,* not human learning or human simulation. The initial rationale for this shift was that if we could not program a computer to perform a task then it was unlikely that we could program it to learn to perform that task; and if we could not program it somehow then we certainly could not in a way that simulated a person. This became the aim of legitimate artificial intelligence research in the late 1960s and early 1970s: the performance of tasks that required intelligence when performed by people.

While the AI community regrouped around the revised goals in Era 2, many outside took the opportunity to express disenchantment with the whole endeavor. Dreyfus in his 1972 book,[14] *What Computers Can't Do,* detailed many of the over-optimistic claims for AI research and the ensuing under-achievement. He pointed to weaknesses in the philosophical and methodological foundations of work in Era 1. His 1979 revised edition reported some of the intervening debate resulting from his book and the changing priorities in AI research. It was a well-balanced report that can be criticized primarily because it was out of date by the time it was published. Those in the AI community felt they had already learned the lessons of Era 1 and changed their strategy.

Lighthill's negative report[15] in 1973 on behalf of the SRC on the status of AI research and the appropriate level of its funding in the U.K. reflected the disenchantment at the end of Era 1 with its oversell and lack of achievement. However, there was another motive behind the commissioning of that report, which was the fear of some of those responsible for developing computer science departments in the U.K. that AI research would be funded most strongly outside those departments. At a mundane level the misgivings came from the realization that the expensive and powerful computers then necessary for AI research might not be used to improve the research facilities of recently formed computer science departments. More fundamentally it was sensed that developments in AI might prove a major part of the future foundations of computer science. Whatever the motivation, the outcome of the report was very negative for AI research in the U.K.[16] during Era 2.

Weizenbaum's 1976 book,[17] *Computer Power and Human Reason,* was a far more personal statement than Dreyfus's by someone who had been responsible for one of the early achievements of AI research. His ELIZA program was widely acclaimed in the late 1960s as the first successful attempt at passing the Turing test for AI. It could carry out "cocktail party" conversation with a person at a terminal that was remarkably human-like. However, the acclaim became embarrassment when it was realized the simple mechanisms of ELIZA illustrated the weakness of the Turing test rather than a major advance in AI. People were all too ready to discern intelligence in machines and men, and commonsense human judgment in the matter was not an adequate criterion. The AI community was forced to reconsider what was meant by "intelligence."

AI work in Era 2 shifted from requirements for power and generality, consideration of computer architecture, and the simulation of human operation. Instead, it emphasized requirements to encode human expert knowledge and performance, by whatever means, for emulation by the computer. These targets resulted in practical achievements in the development of systems that could perform diagnostic inference tasks just as well as human experts, and the late 1970s and early 1980s became the era of *expert system* research.[18]

The strength of this paradigm shift cannot be overemphasized. It defined the boundaries of an AI community that established parameters for funding and publication. Focusing efforts on performance led to achievements recognized outside this community and hence established the legitimacy of AI research. In recent years the community has become strong enough to begin to reabsorb some of the earlier objectives and concepts. Human simulation rather than just emulation is an objective of cognitive science. Certain aspects of knowledge acquisition processes subsumed under learning are

			HARDWARE	SOFTWARE	EFFECTS
I	1953 -58	GEE WHIZ !	VACUUM TUBES & DELAY LINES	NONE	TECHNICIANS & FEARS OF AUTOMATION
II	1958 -66	PAPER PUSHERS	TRANSISTORS & MAGNETIC CORES	COMPILERS & I/O CONTROL	PROLIFERATION OF EDP GROUPS & RIGIDITY
III	1966 -74	COMMUN-ICATORS	LARGE-SCALE INTEGRATED CIRCUITS & INTERACTIVE TERMINALS	OPERATING SYSTEMS & COMMUN-ICATIONS	CENTRALIZATION OF EDP AS NEW FUNCTION
IV	1974 -82	INFORM-ATION CUSTOD-IANS	VERY LARGE FILE STORES & SATELLITE COMPUTERS	VIRTUAL MACHINES	REDISTRIBUTION OF MANAGEMENT FUNCTIONS
V	1982 -	ACTION AIDS	MAGNETIC BUBBLE, & LASER HOLO-GRAPHIC & DISTRIBUTED	INTERACTIVE LANGUAGES & CONVENIENT SIMULATION	SEMI-AUTOMATIC OPERATING DECISIONS

Figure 3—Characteristics of the five generations as seen in 1974

proving significant. The first steps are being taken toward new computer architectures aimed at AI requirements.

It would be foolish to pretend that the academic debates and funding struggles of the 1970s culminated in a clear realization of the significance of the three eras described. The oversell of Era 1 remained in many people's minds. Regrouping under siege and bitter struggle was the perception of Era 2 for many others. A nation of the East drew the attention of the West to the achievements of its AI research in Era 3 and the significance of this for the world economy.

We are now in Era 4, with the understanding and development of fifth-generation computers being treated as a national priority by many countries. The defense and commercial implications of achieving even some of the Japanese objectives are a more compelling argument than any the AI community have been able to muster. The sociocultural implications might also be overwhelming. The AI community of Eras 2 and 3 no longer exists. The financial pull of industry has fragmented effort, and it will be some time before new patterns of activity are clarified.

THE PREVIOUS GENERATIONS

In analyzing and forecasting the effects of fifth-generation developments it is useful to look at past definitions and projections. Some 10 years ago Withington[19] analyzed the 30 years of computer history to 1974 in terms of three generations of machines and projected these forward to two further generations. Figure 3 presents his main arguments in tabular form.

Withington's retrospective of the first three generations shows that they may be distinguished by hardware, software, or use, since all three are correlated. His description of the fourth generation then coming into use emphasizes large file stores and satellite computers, which came about as database and distributed systems technology. Virtual machine software also came about but did not have as much effect as expected. Other articles of the mid-1970s emphasize "universal emulators" and the need for each generation of hardware to be able to run the software of all previous generations. Many manufacturers experimented with such emulation of competitors' machines but it has not become widely adopted, possibly because of software copyright and licensing problems. This was also the era of hardware/software unbundling.

Withington's projections of the fifth generation, like all at that time, fell far short of target. Magnetic bubble technology lost out to advances in semiconductors. Interactive languages and convenient simulation were a major part of the fourth generation thrust. We can now see that VLSI goes into the hardware column and AI into the software column for the fifth generation. What we cannot see is what the sixth-generation extension of this table should be! It seems likely, however, that the often-quoted results will be continued, that each generation of about six years leads to a 10 times increase in speed, 20 times increase in memory, 10 times in reliability, and 1/10 in component cost.[20]

Withington's titles for each generation are evocative. The 1950s were an era of "gee whiz computers can do anything." This was necessary to support the belief of their users that one day they might become cost-effective. It took at least one more generation for this to happen. His name for the fifth generation is surprisingly apt; expert systems are certainly "action aids." However, it is a matter of faith today that they, and other developments coming out of AI, will become cost-effective in a wide range of applications. At least one more generation will be required, but the Japanese have already triggered this faith. We are at the beginning of a new cycle of events.

CONCLUSIONS

The Japanese fifth-generation program may have come as a surprise but it can be seen in retrospect as the logical culmination of developments in VLSI and AI technologies. The availability of VLSI in power and in quantity provides a technology that requires the consumer markets to be fully exploited. The availability of AI technology is a prerequisite for the human-like characteristics necessary in systems that can be interacted with easily by users untrained in computing.

That VLSI and AI can be developed over the next six years to provide cost-effective products is a dream. However, it requires no greater faith than that which set off this industry in the 1950s when the first generation of computers promoted the "gee whiz" reaction. During the intervening period the economic basis of our society has changed, and we can now see the knowledge economy[20] forming a new environment for the development of mankind. The fifth generation of computers is a natural product of the ecology[21] of the knowledge environment. By analogy with past predictions of computer

development the fifth generation will be with us sooner than anyone might reasonably expect.

REFERENCES

1. Moto-Oka, T. (ed.) *Fifth Generation Computer Systems.* Amsterdam: North-Holland, 1982.
2. Simons, G. L. *Towards Fifth-Generation Computers.* Manchester: National Computing Centre, 1983.
3. Steier, R. "Cooperation is the Key: An Interview with B. R. Inman." *Communications of the ACM,* 26 (1983), pp. 642–645.
4. *A Programme for Advanced Information Technology: The Report of the Alvey Committee.* London: HMSO, 1982.
5. Kahn, H., and Pepper, T. *The Japanese Challenge.* New York: William Morrow, 1980.
6. Servan-Schreiber, J.-J. *The World Challenge.* New York: Simon & Schuster, 1980.
7. Kuhn, T. S. *The Structure of Scientific Revolutions.* University of Chicago Press, 1962.
8. "Technology update." *Electronics,* 56 (1983), pp. 122–235.
9. Gaines, B. R., and Shaw, M. L. G. *The Art of Computer Conversation.* New Jersey: Prentice Hall, 1984.
10. Weil, U. *Information Systems in the 80's.* New Jersey: Prentice Hall, 1982.
11. Minsky, M. "A Selected Descriptor-Indexed Bibliography to the Literature on Artificial Intelligence." *IRE Transactions on Human Factors in Electronics,* 2 (1961), pp. 39–55.
12. Orr, W. D. (ed.) *Conversational Computers.* New York: Wiley, 1968.
13. Sass, M. A., and Wilkinson, W. D. (eds.) *Computer Augmentation of Human Reasoning.* Washington: Spartan Books, 1965.
14. Dreyfus, H. L. *What Computers Can't Do: The Limits of Artificial Intelligence.* New York: Harper, 1972.
15. Lighthill, J. "Artificial Intelligence: A General Survey." In *Artificial Intelligence: a Paper Symposium.* London: Science Research Council, 1973.
16. Fleck, J. "Development and Establishment in Artificial Intelligence." In Elias, N., Martins, H., and Whitley, R. (eds.) *Scientific Establishments and Hierarchies,* Holland: Reidel, 1982, pp. 169–217.
17. Weizenbaum, J. *Computer Power and Human Reason.* San Francisco: Freeman, 1976.
18. Michie, D. (ed.) *Expert Systems in the Micro Electronic Age.* Edinburgh University Press, 1979.
19. Withington, F. G. "Five Generations of Computers." *Harvard Business Review,* July-August (1974), pp. 99–108.
20. Turn, R. *Computers in the 1980s.* New York: Columbia University Press, 1974.
21. Machlup, F. *Knowledge and Knowledge Production.* Princeton University Press, 1980.
22. Wojciechowski, J. "The Impact of Knowledge on Man: The Ecology of Knowledge." In *Hommage a Francois Meyer.* Marseille: Laffitte, 1983, pp. 161–175.