Challenges of Massive Parallelism

Hiroaki Kitano

Center for Machine Translation
Carnegie Mellon University
5000 Forbes
Pittsburgh, PA 15213 U.S.A.
hiroaki@cs.cmu.edu

Software Engineering Laboratory
NEC Corporation
2-11-5 Shibaura, Minato
Tokyo 108 Japan
kitano@ccs.mt.nec.co.jp

Abstract

Artificial Intelligence has been the field of study for exploring the principles underlying *thought,* and utilizing their discovery to develop useful *computers.* Traditional AI models have been, consciously or subconsciously, optimized for available computing resources which has led AI in certain directions. The emergence of massively parallel computers liberates the way intelligence may be modeled. Although the AI community has yet to make a quantum leap, there are attempts to make use of the opportunities offered by massively parallel computers, such as memory-based reasoning, genetic algorithms, and other novel models. Even within the traditional AI approach, researchers have begun to realize that the needs for high performance computing and very large knowledge bases to develop intelligent systems requires massively parallel AI techniques. In this Computers and Thought Award lecture, I will argue that massively parallel artificial intelligence will add new dimensions to the ways that the AI goals are pursued, and demonstrate that massively parallel artificial intelligence is where AI meets the real world.

1. Introduction

Massively Parallel Artificial Intelligence is a field of study exploring methodologies for building intelligent systems and investigating computational models of thought and behavior that use massively parallel computing models [Waltz, 1990, Kitano et al., 1991a]. Traditionally, artificial intelligence research has two goals: as an engineering discipline, artificial intelligence pursues methods to build useful and intelligent systems. As a scientific discipline, artificial intelligence aims at understanding the computational mechanisms of thought and behavior. I believe that massively parallel artificial intelligence will be the central pillar in the ultimate success of artificial intelligence in both engineering and scientific goals.

Massively parallel artificial intelligence does not merely speed up traditional AI models. In fact, many traditional models are not appropriate for massively parallel implementation. Vast majorities of the AI models proposed so far are strongly influenced by the performance of existing von Neumann architectures. In a narrow sense, a von Neumann architecture can be represented by a single processor architecture with a memory to store instructions and data. Even in the early 80s, available computing power for most AI researchers was far less than that for personal computers in 1993, and hopelessly slower than advanced RISC-based workstations. Additionally, memory space has increased drastically in 10 years.

Within the hardware constraints to date, sequential rule application, for example, has been an optimal implementation strategy. While I agree that this idea is not merely an implementation strategy (in fact there are numbers of cognitive bases for sequential rule application), hardware constraints have prevented AI researchers from seriously investigating and experimenting on other, more computationally demanding approaches. For example, the memory-based reasoning model could not have been experimentally supported, using only the computing power and memory space available in the early 1980s. Genetic algorithms and neural networks could not be seriously investigated without a major leap in computing power.

One might argue that if a program were run for, say, a month, experimental results could have been obtained even in the early 1980s. However, in practice, dominating computational resources for such a substantial period of time would be impossible for the vast majority of researchers, and painfully long turn around time for experiments can be a major discouraging factor in promoting research activities based on computationally demanding paradigms.

This hardware factor inevitably guided, consciously or subconsciously, AI in a particular direction. Although the issues and approach may vary, several researchers have been expressing concern over the influence of existing computer architectures on our models of thought. David Waltz stated:

> the methods and perspective of AI have been dramatically skewed by the existence of the common digital computer, sometime called the

von Neumann machine, and ultimately, AI will have to be based on ideas and hardware quite different from what is currently central to it.[Waltz, 1988]

and proposed memory-based reasoning. Rodney Brooks noted in the his 1991 Computers and Thought lecture

the state of computer architecture has been a strong influence on our models of thought. The Von Neumann model of computation has lead Artificial Intelligence in particular directionspBrooks, 199l]

and proposed the subsumption architecture.

Therefore, it is critically important to understand the nature of the computational resources which are available today and will be available in the near future. I argue that massively parallel artificial intelligence will add a new dimension to our models of thought and to the approaches used in building intelligent systems. In the next section, the state-of-the-art in computing technology will be reviewed and the inevitability of massive parallelism will be discussed.

2. Massively Parallel Computers

Advancements in hardware technologies have always been a powerful driving force in promoting new challenges. At the same time, available hardware performance and architectures have been constraints for models of intelligence. This section reviews progress in device technologies and architecture, in order to give an idea of the computing power and memory space which could be available to AI researchers today.

2.1. Devices

Device technology progress is fast and steady. Memory chips, DRAM, capacity increases at a 40% a year rate. 64M DRAM will be available shortly, and successful experimental results have been reported for a $0.1\mu m$ design rule. Figure 1 shows how this trend would continue into the future.

Microprocessor performance increases at the rate of 20 to 30% a year. In 1990, engineers at Intel corporation predicted that processors in the year 2,000 (Micro2000) would contain over 50 million transistors per square inch and run with a 250MHz clock[Gelsinger et al., 1989]. The Micro2000 would contain four 750 MIPS CPUs run in parallel and achieves 2000 MIPS. In 1992, they have commented that their prediction was too modest! This implies the possibility of achieving low cost, high performance computing devices. Thus, some of the research on current massively parallel computers may be conducted on workstations in the future.

On the other hand, uniprocessor supercomputer performance is improving at a substantially lower rate

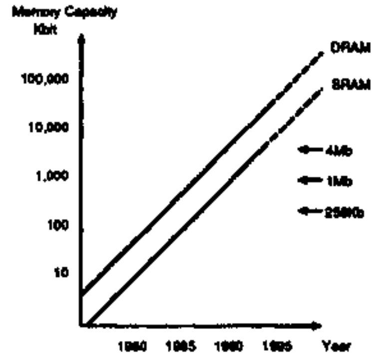


Figure 1: Memory Capacity for Mass Production Memory Chips

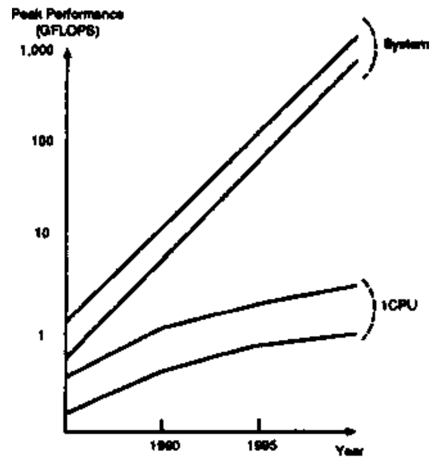


Figure 2: Peak Performance for Supercomputers

than the microprocessor improvement rate. All newly announced vector supercomputers use multiprocessing technologies, so that the slow processor performance improvement rate can be compensated for (Figure 2). In addition, massively parallel machines are already competitive with current supercomputers in some areas [Myczkowski and Steele, 1991, Sabot et al., 1991, Hord, 1990]. It is expected that, in the near future, the majority of high-end supercomputers will be massively parallel machines.

This implies that the progress in microprocessors will face a slow down stage in the future, and that an architecture for a higher level of parallelism will be introduced. This increasing availability of high performance computers, from personal computers to massively parallel supercomputers, will allow experiments on simulated massive parallelism to be made at numerous places. Thus, computational constraints should no longer impose restrictions on how thought is modeled.

Special purpose devices provide dramatic speed up for certain types of processing. For example, associative memory is a powerful device for many AI applications. It provides a compact and highly parallel pattern matcher, which can be used for various applications such as memory-based reasoning, knowledge base search, and genetic algorithms [Stormon et al., 1992, Twardowski, 1990, Kitano et al., 1991c]. The benefit of associative memory is that it can attain very high parallelism in a single chip. Although only a few simple operations, such as bit pattern matching, can be accomplished, these are central operations for AI.

There are also efforts to build high performance and large scale neural network device chips (Graf et al., 1988, Ae and Aibara, 1989, Mead, 1989]. Although current devices have not attained implementation of sufficiently large networks, which can cope with many real applications, there are some promising research results. For example, the use of wafer scale integration for neural networks may provide for scaling up to a reasonable size [Yasunaga et al., 1991].

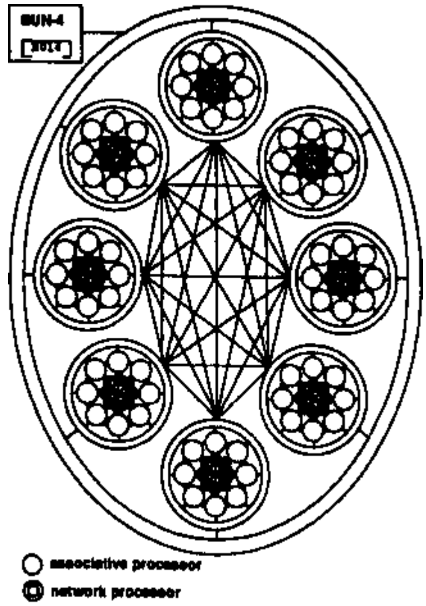
2.2. Architecture

The history of massively parallel computers can be traced back to Illiac-IV [Bouknight et al., 1972], DAP [Bowler and Pawley, 1984], and MPP [Batcher, 1980]. Illiac-IV was the first SIMD supercomputer with 64 processors operating at a rate of 300 MIPS. MPP consists of 16,384 1-bit processors. DAP ranges from 1,024 to 4,096 processors. These early machines were built particularly for scientific supercomputing.

However, the real turning point came with the development of the connection machine [Hillis, 1985]. The connection machine was motivated by NETL [Fahlman, 1979] and other work on semantic networks, and was originally designed with AI applications in mind. Hillis noted:

> This particular application, retrieving commonsense knowledge from a semantic network, was one of the primary motivations for the design of the Connection Machine[Hillis, 1985].

The CM-2 Connection Machine, which is the second commercial version of the Connection Machine, consists of large numbers of 1-bit processors and floating point processing units [TMC, 1989]. Various systems has been implemented, such as rule-based inference systems [Blelloch, 1986, Blelloch, 1986], a massively parallel assumption-based truth maintenance system [Dixon and de Kleer, 1988], a text retrieval system [Stanfillet al., 1989], stereo vision programs [Drumheller, 1986], a frame-based knowledge representation system [Evett et al.. 1990a], heuristic search programs [Evett et al., 1990b], parallel search techniques [Geller, 1991], a classifier system [Robertson, 1987], case-base retrievers [Kolodner, 1988, Cook, 1991, Kettler et al., 1993], a motion control system [Atkeson and Reinkensymeyer, 1990], and others (See [Kitano et al., 1991a] for a partial list of systems developed). The newest version, CM-5 [TMC, 1991], uses 32-bit SPARC chips interconnected via a tree structured network. Companies, such as MasPar Computer [Blank, 1990, Nickolls, 1990], Intel, Cray, IBM, NCR, Fujitsu, NEC and others, have started development projects for massively parallel computers, and some of these are already on the market.



The block diagram of IXM2 associative memory processor

Figure 3: IXM-2

Although most massively parallel machines developed so far employ distributed memory architectures, there have been some attempts to build a shared memory machine. KSR-1 [KSR, 1992] by Kendall Square Research is an attempt to overcome the difficulties in software development which often arise on distributed memory machines. KSR-1 employs the ALLCACHE architecture, which is a virtual shared memory (VSM) model. In a VSM model, a physical architecture resembles that of a distributed memory architecture, but local memories are virtually shared by other processors, so that programmers do not have to worry about the physical location of data.

Independently from these commercial machines, several research machines have been developed at universities and research institutions. These projects include the J-Machine [Dally et al., 1989] at MIT, SNAP [Moldovan et al., 1990] at the University of Southern California (USC), and IXM-2 [Hieuchi et al., 1991] at ElectroTechnical Laboratory (ETL). I have been involved in some of these projects, as will be described below.

The IXM-2 massively parallel associative memory processor was developed at ETL by Higuchi and his colleagues [Higuchi et al, 1991], DCM-2 consists of 64 associative processors and 9 communication processors (figure 3). The associative processor is organized using a T800 transputer [inmos, 1987], 4K words associative memory, SRAM, and other circuitry. Due to an intensive use of associative memory, IXM-2 exhibits 256K parallelism for bit pattern matching. IXM-2 is now being used for various research projects at ETL, Carnegie Mellon University, and the ATR Interpreting Telephony Research Laboratory. The application domains range through genetic algorithms, knowledge-base search, and example-based machine translation.

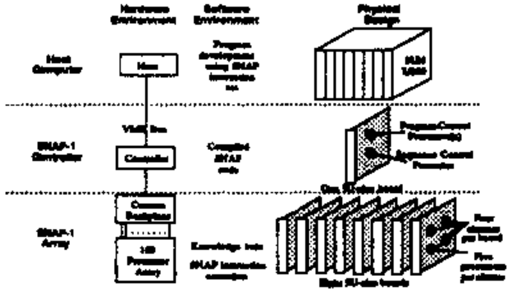


Figure 4: SNAP-1

The Semantic Network Array Processor (SNAP) is now being developed by Professor Dan Moldovan's team at the University of Southern California. A SNAP-1 prototype has been developed and has been operational since the summer of 1991. SNAP-1 consists of 32 clusters interconnected via a modified hypercube (Figure 4). Five TMS320C30 processors form a cluster, which store up to 1,024 semantic network nodes. Processors within a cluster communicate through a multi-port memory. Several systems have been implemented on SNAP, such as the DmSNAP machine translation system [Kitano et al., 1991b], the PASS parallel speech processing system [Chung et al., 1992], and some classification algorithms [Kim and Moldovan, 1990].

These trends in hardware and architectures will enable massive computing power and memory space to be available for AI research.

2.3. Wafer-Scale Integration for Memory-Based Reasoning

WSI-MBR is a wafer-scale integration (WSI) project designed for memory-based reasoning. WSI is the state-of-the-art in VLSI fabrication technology, and has been applied to various domains such as neural networks [Yasunaga et al., 1991]. WSI fabricates one large VLSI-based system on a wafer as opposed to conventional VLSI production which fabricates over 100 chips from one wafer. The advantage of WSI is in its size (high integration level), performance, cost, and reliability:

Size: WSI is compact because nearly all circuits necessary for the system are fabricated on a single wafer.

Performance: WSI has a substantial performance advantage because it minimizes wiring length.

Cost: WSI is cost effective because it minimizes the requirement for using expensive assembly lines.

Reliability: WSI is reliable because it eliminates the bonding process which is the major cause of circuit malfunctions.

However, there is one big problem in WSI fabrication: defects. In conventional VLSI fabrication, one wafer consists of over 100 chips. Typically, there are certain percentages of defective chips. Traditionally, chips with defects have been simply discarded and the chips without defects have been used. To estimate the faults in a WSI chip, we can use the Seeds model[Seeds, 1967]: $Y = e^{-\sqrt{DA}}$, where Y is a yield of the wafer, D is the fault density which is, in the fabrication process being used, about 1 fault per cm^2, and A is the chip area. This is a reasonable rate for the current fabrication process. However, even this level of fault would cause fatal problems for an attempt to build, for example, an entire IBM 370 on one wafer. Unless sophisticated defect-control mechanisms and robust circuits are used, a single defect could collapse an entire operation. But, redundant circuits diminish the benefits of the WSI. This trade-off has not been solved.

MBR is ideal for WSI, and avoids the defect problem because it does not rely upon any single data unit. WSI-MBR is a digital/analog hybrid WSI specialized for memory-based reasoning. The digital/analog hybrid approach has been used in order to increase parallelism and improve performance. In the digital computing circuit, a floating point processor part takes up the most of chip area. On the other hand, the analog circuit requires only a fraction of the area for implementation of equivalent floating point operation circuits and drastic speed up can be attained. For detailed arguments on the relative advantages of analog and digital circuits, see [Kitano and Yasunaga, 1992]. Use of the less area-demanding analog approach provides two major advantages over the digital approach: (1) increased parallelism, and (2) speed up due to relaxed wiring constraints (critical paths and wire width). Expected performance with the current design is 70Tflops on a single 8 inch wafer. Using wafer stack and micro-pin technologies, peta-flops on a few hundred million record system would be technically feasible!

3. Grand Challenge AI Applications

From the engineering point of view, the ultimate success in artificial intelligence can be measured by the economic and social impact of the systems deployed in the real-world. Applications with significant economic and social impacts provide us with some grand challenge AI applications.

Grand challenge AI applications are applications with significant social, economic, engineering, and scientific impacts. The term *Grand Challenge* was defined as

> a fundamental problem in science or engineering, with broad economic and scientific impact, whose solution will require the application of high-performance computing resources.

in the U.S. High Performance Computing Act of 1991. Thus grand challenge AI applications are *AI* applications with significant social, scientific, and engineering impacts. Typical examples are Human Genome project [NRC, 1988] and speech-to-speech translation system projects. Two workshops have been organized to discuss grand challenge AI applications.

In February 1992, NSF sponsored *The Workshop on High Performance Computing and Communications for*

Grand Challenge Applications: Computer Natural Language and Speech Processing, and Artificial Intelligence. In October 1992, I organized *The Workshop on Grand Challenge AI Applications* in Tokyo — participated in by a group of leading Japanese researchers[Kitano, 1992b]. Details on these workshops and some of the on-going grand challenge AI applications can be found in [Kitano et al., 1993].

Typical applications discussed include: a speech-to-speech translation system, human genome analysis, Global Architecture for Information Access (GAIA) — a highly intelligent information access system, Shogi and Go systems which beat Meijin (Grand Master), intelligent robots and programs which undertake daily tasks, and intelligent vehicles. There have also been discussions as to wehat technologies are needed to support grand challenge AI applications. The conclusions from both workshops are surprisingly similar. The common thread was the need for massive computing power, massive memory storage, massive data resource, and ultra high band-width communication networks.

For a grand challenge to succeed, our computer systems must be able to be scaled up. For example, natural language systems must be scaled up to the point where a dictionary of the system contains all common words and many domain-specific terms, and to where grammar rules cover most syntactic structures. For any serious machine translation system, this means that the dictionary contains from half a million to a few million word entries, and the grammar rules amount to over 10,000.

In addition, for systems to be scaled up, and to be deployed in the real world, they must be able to cope with a noisy environment and incomplete data resources.

4. The Bottlenecks

This section discusses limitations of "the traditional AI approach." The traditional AI approach can be characterized by several salient features:

Formal Representations: Most traditional models use rigid and formal knowledge representation schemes. Thus, all knowledge must be explicitly represented in order for the system to use that knowledge. There is no implicit knowledge in the system.

Rule driven inferencing: Reasoning is generally driven by rules or principles, which are abstract and generalize knowledge on how to manipulate specific knowledge.

Strong Methods: Since the system depends on explicit knowledge and rules, domain theory must be understood in order to build any system based on the traditional approach.

Hand-Crafted Knowledge Bases: Knowledge and rules have been hand-coded at extensive labor costs. In many cases, coding must be carried out by experts in the field.

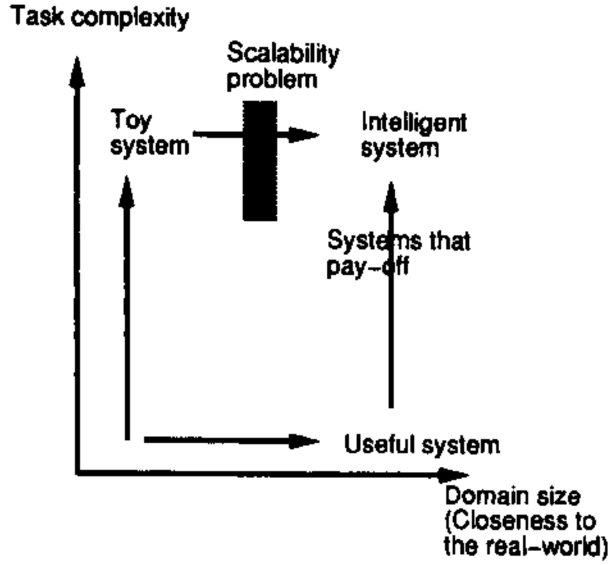


Figure 5: Approaches for Building Intelligent Systems

In essence, these are derived from the physical symbol system hypothesis and the heuristic search. As Dave Waltz stated:

> For thirty years, virtually all AI paradigms were based on variants of what Herbert Simon and Allen Newell have presented as physical symbol systems and heuristic search hypotheses.

The fundamental assumption in the traditional approach is that experts know the necessary knowledge regarding the problem domain, and that expert knowledge can be explicitly written using formal representations. Toy problems, such as the blocks world and the Tower of Hanoi, meet this condition. And thus, historically, many AI research efforts have been carried out on domains such as the blocks world and the Tower of Hanoi. The intention of such work is that by proving the effectiveness of the method using small and tractable tasks the method can be applied to real-world problems. To rephrase, most research has been carried out in such a way that researchers develop a highly intelligent system in a very restricted domain. Researchers hoped that such systems could be scaled up with larger funding and increased effort (Figure 5: Independently, Brooks and Kanade uses similar figures).

However, experiences in expert systems, machine translation systems, and other knowledge-based systems indicate that scaling up is extremely difficult for many of the prototypes. For example, it is relatively easy to build a machine translation system, which translates a few very complex sentences. But, it would be far more difficult to build a machine translation system which correctly translates thousands of sentences of medium complexity. Japanese companies have invested a decade of time and a serious amount of labor inputs to developing commercial machine translation systems, based on traditional models of language and intelligence. So far, a report on high quality machine translation systems has not been published.

There are several reasons why the traditional AI approach fails in the real-world. The basic assumption is that knowledge exists somewhere in the mind of experts, so that, if it can be written down in operational form — the expertise can be transferred to the system. However, this assumption does not stand in the real-world. The following three factors prevent the traditional AI approach from real-world deployment.

Incompleteness: It is almost impossible to obtain a complete set of knowledge for a given problem domain. Experts themselves may not have the knowledge, or the knowledge may be tacit so that it can not be expressed in a formal manner. Thus, a certain portion of the knowledge is always absent.

Incorrectness: There is no guarantee that expert knowledge is always correct and that encoding is perfect. A large knowledge base, which contains over 10,000 frames, must include a few errors. If there were a 0.5% error rate, there would be over 5,000 incorrect frames in a million frame knowledge-base!

Inconsistenet: The set of knowledge represented may be inconsistent, because (1) contextual factors to maintain consistency were ignored, or (2) expert knowledge is inconsistent.

These realities in regard to real world data are devastating to the traditional AI approach. Moreover, in addition to these problems, there are other problems, such as:

Human Bias: The way knowledge is expressed is inevitably influenced by available domain theory. When the domain theory is false, the whole effort can fail.

Tractability: The system will be increasingly intractable, as it scales up, due to complex interaction between piece wise rules.

Economics: When rules are extracted from experts, system development is a labor intensive task. For example, even if MCC's CYC [Lenat and Guha, 1989] system provides a certain level of robustness for knowledge-based systems, proliferation of such systems would be limited due to high development cost.

Empirically, efforts to eradicate these problems have not been successful. In essence, AI theories for the real-world must assume data resources to be inconsistent, incomplete, and inaccurate.

Lenat and Feigenbaum pointed out the scaling up problem for expert systems, and proposed the Breadth Hypothesis (BH) [Lenat and Feigenbaum, 1991] and the CYC project [Lenat and Guha, 1989]. While there is some truth in the BH, whether or not the robustness can be attained by a pure symbolic approach is open to a question. A series of very large knowledge based systems projects, such as CYC [Lenat and Guha, 1989], knowledge-based machine translation (KBMT) [Goodman and Nirenberg, 199l], corporate-wide CBR system [Kitano et al., 1992], and large-scale CBR systems [Kettler et al., 1993] will be important test-beds for this approach.

For these systems to succeed, I believe that incorporation of mechanisms to handle messy real world data resources will be necessary.

5. Computing, Memory, and Modeling

One promising approach for building real world AI applications is to exploit the maximum use of massively parallel computing power and data resources. In essence, I argue that an approach emphasizing massive computing power, massive data resources, and sophisticated modeling will play a central role in building grand challenge AI applications.

Computing: The importance of computing power can be represented by the Deep-Thought Chess machine[Hsu, 1991, Hsu et al., 1990, Hsu, 1990]. Deep-Thought demonstrates the power of computing for playing chess. It was once believed that a strong heuristic approach was the way to build the grand master level chess machine. However, the history of chess machines indicates that computing power and chess machine strength have almost direct co-relation (Figure 6: reproduced based on [Hsu et al., 1990]). Deep-Thought-II consists of 1,000 processors computing over a billion moves per second. It is expected to beat a grand master. Deep-Thought exemplifies the significance of the massive computing. Similarly, real-time processing, using a very large knowledge source requires massive computing power [Evett et al., 1990a, Evett et al., 1990b, Geller, 1991].

Memory: The need for memory can be argued from a series of successes achieved in memory-based reasoning. Starting from the initial success of MBRtalk[Stanfill and Waltz, 1988], memory-based reasoning has been applied to various domains, such as protein structure prediction [Zhang et al., 1988], machine translation [Sato and Nagao, 1990, Sumita and Iida, 1991, Furuse and Iida, 1992, Kitano and Higuchi, 1991a, Kitano. 1991], census classificationfCreecy et al., 1992], parsing[Kitano and Higuchi, 1991b, Kitano et al., 1991b], and weather forecasting. PACE, a census classification system, attained 57% classification accuracy for occupation codes and 63% classification accuracy for industry codes [Creecy et al, 1992]. The AICOS expert system attained only 37% for the occupation codes and 57% for industry codes. These successes can be attributed to the superiority of the approach which places memory as the basis for intelligence, rather than fragile hand-crafted rules. In a memory-based reasoning system, the quality of the solution depends upon the amount of data collected. Figure 7 shows the general relationship between the

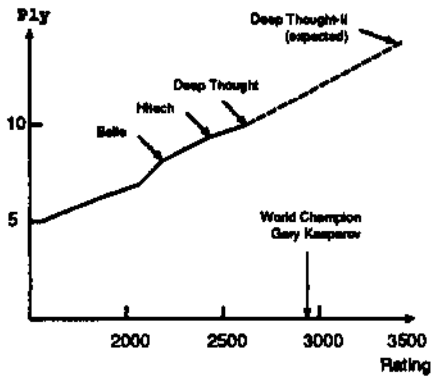


Figure 6: Progress of Chess Machines

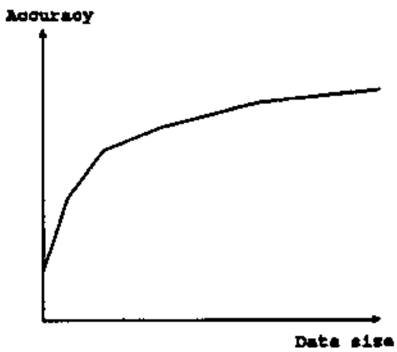


Figure 7: Memory and Accuracy

amount of data and solution quality. The success of memory-based reasoning demonstrates the significance of a massive memory or data-stream.

Modeling: The importance of modeling can be discussed using the SPHINX speech recognition system[Lee, 1988]. Using massive computing and a massive data-stream is not sufficient to build such artificial intelligence systems. The critical issue is how the application domain is modeled. Figure 8 shows the improvement in recognition rate, with modeling sophistication. Even if massively parallel machines and large data resources are used, if the modeling is not appropriate, only a poor result can be expected. SPHINX exemplifies the significance of modeling.

There are several reasons that these factors are important, the following analysis may help in understanding the effectiveness of the approach.

Distribution Many of the phenomena that AI tries to deal with are artifacts determined by human beings, such as natural languages, society, and engineering systems. However, it is acknowledged that even these phenomena follow basic statistical principles, as applied in nature. Normal distribution (also called Gaussian distribution) and Poisson distribution are important distributions. In physics, quantum statistics (such as Bose-Einstein statistics and Fermi-Dirac statistics) and a classical statistics (Maxwell-Boltzmann statistics) are used. AI, however, is

Figure 8: Modeling and Accuracy

not mature enough to establish statistics systems for various phenomena. However, using statistical ideas, even in a primitive manner, can greatly help to understand the nature of many phenomena.

Zipf's law, for example, describes distributions of types of events in various activities. In Zipf's law, the multiplication of the rank of an event (r) and the frequency of the event (f) is kept constant *($C = rf$).* For example, when $C = 0.1$, an event of rank 1 should have a 10.0% share of all events, a rank 2 event should occupy 5.0%, and so on (Figure 9). The sum of the top 10 events occupies only 29.3%. Despite the fact that the occurance probability for an individual event, greater than the 11-th rank, has only a fraction of a percent (e.g. 0.5% for the 11-th rank), the sum of these events amounts to 70%. This law can be observed in various aspects of the world. AI models using heuristics rules will be able to cover events of high frequency relatively easily, but extending the coverage to capture irregular and less frequent events will be major disturbing factors. However, the sum of these less frequent events will occupy a significant proportion of the time, so that the system must handle these events. In fact, about 60% to 70% of grammar rules in typical commercial machine translation systems are written for specific linguistic phenomena, whose frequencies are very low. This empirical data confirms Zipfs law in the natural language. Memory-based reasoning is a straightforward method to cover both high frequency and low frequency events. In fact, the relationship between data size and accuracy of a MBR system and the accumulated frequency in Zipf's law is surprisingly similar (Figure 10).

The Central Limit Theorem Inaccuracy, or noise, inherent in large data resources can be mitigated using the very nature of these large data resources. Assuming that inaccuracy on data is distributed following a certain distribution (not necessary Gaussian), the central limit theorem indicates that the distribution will be Gaussian over large data sets. A simple analysis illustrates the central limit theorem. Assuming the linear model, $y = \mu + \epsilon$ where y is an observed value, μ is a true value, and ϵ is noise, the central limit theorem indicates that the ϵ will follow a normal distribution, regardless of the original

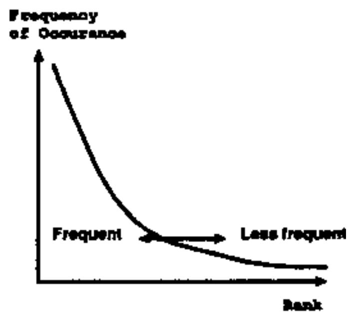


Figure 9: Zipf's Law

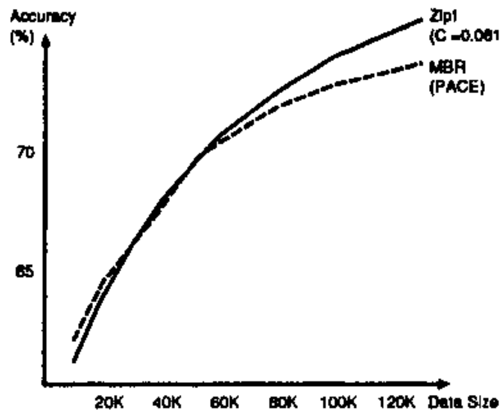


Figure 10: Zipf's Law and MBR

distribution which produces ϵ. When the expected value of ϵ is 0, the observer will get the true value.

In fact, adding various types of noise to memory-based reasoning systems does not cause major accuracy degradation. Figure 11 shows accuracy degradation for MBRtalk (a MBR version of NETtalk [Sejnowski and Rosenberg, 1987]), when noise is added to the weights. The noise follows a uniform distribution. N% noise means that a weight can be randomly changed in the $W \pm \frac{N \times W}{100}$ range, where W is a weight value. The degradation rate is much smaller, when larger data sets are used.

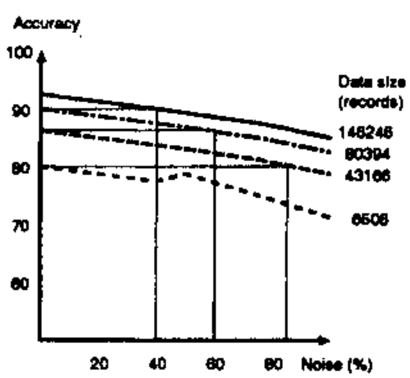


Figure 11: Accracy Degradation by Noisy Data

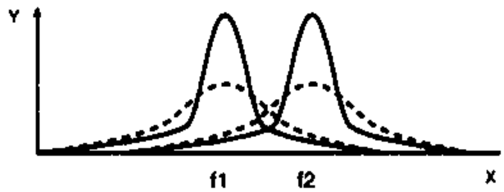


Figure 12: Overlapping Noisy Data

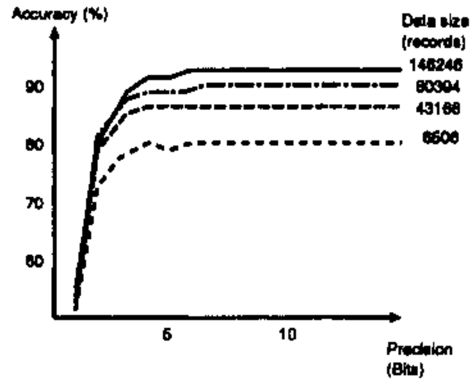


Figure 13: Accracy Degradation and Precision

The Law of Large Numbers According to the law of large numbers, the peak in a data distribution will be narrower with a large number of data sets. Assume that there are two close true values. Data distributions may overlap due to noise. However, the law of large numbers indicates that, by collecting larger data sets, overlap can be mitigated (Figure 12). (I have not yet confirmed whether these effects can be observed in memory-based reasoning. However, the relationship between accuracy and data size implies that the law is in effect (Figure 13).)

Nonlinear Boundary A vast majority of real world phenomena are non-linear. Any attempt to approximate non-linear boundaries, using Unear or discrete methods, will create certain levels of inaccuracy. A simple illustration is shown in Figure 14. The real boundary is shown by the (non-linear) solid line. Available data points are shown by O and X. Assuming that a categorization has been made on both the Y and X axis as A1, A2, A3 and B1, B2, B3, the region can be expressed as (A2 and (B2 or B3)). However, there are areas which do not fit the rectangular area described by this symbolic representation. The proponents of the symbolic approach may argue that the problem can be circumvented by using fine-grained symbol systems. However, there is always an area where symbol systems and the real boundary do not fit. Using an infinite number of symbols would eliminate the error — but this would no longer be a symbol system! Therefore, for the non-linear boundary problems, symbol systems are inherently incomplete.

The use of a large number of data points and statistical computing methods, however, can better mitigate the error with significantly lower human effort. This observation has two major implications: First, it implies

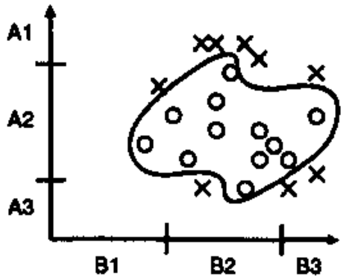


Figure 14: Nonlinear boundary of the problems

that a massively parallel memory-based approach may be able to attain a certain level of competence. This will be discussed in the next section. Second, it has major implications as to how to build very large knowledge-based systems. My recommendation is to use redundancy and population encoding even when the system is based on the traditional AI approach.

6. Speech-to-Speech Translation

The idea discussed above is also effective in speech-to-speech translation systems. Speech-to-speech translation is one of the major topics of the grand challenge AI applications, and its success will have major social and scientific impacts. Speech-to-speech translation systems are generally composed of a speech recognition module, a translation module, and a speech synthesis module. The first group of speech-to-speech translation systems appeared in the late 1980s. These include Speech Trans [Saito and Tomita, 1988] and Dm-Dialog [Kitano, 1990a] developed at Carnegie Mellon University, and SL-Trans [Morimoto et al., 1989] developed at the ATR Interpreting Telephony Research Laboratories. These systems were immediately followed by a second group of systems, which included JANUS [Waibel et al., 1991] at CMU, and ASURA [Kikui et al., 1993] at ATR. This section discusses limitations with the traditional AI approach for natural language processing and how massively parallel artificial intelligence can mitigate these problems.

6.1. Traditional View of Natural Language

The traditional approach to natural language processing has been to rely on extensive rule application. The basic direction for the approach is to build up an internal representation, such as a parse-tree or case-frame, using a set of rules and principles. In essence, it follows the traditional approach to artificial intelligence.

In the early 1970s, there were efforts to develop natural language systems for small and closed domains. Woods developed LUNAR, which could answer questions about moon rocks[Woods et al., 1972]. In the LUNAR system, an ATN represented possible sequences of syntactic categories in the form of a transition network[Woods, 1970]. Winograd developed the famous SHRDLU program, which involved simple questions about the blocks world [Winograd, 1972]. These efforts heavily involved manipulation of the world model and a procedural model for sentence analysis. The world models, grammar, and inputs were assumed to be complete. These systems are typical examples of the traditional AI approach.

In the mid 70s, Schank proposed Conceptual Dependency theory [Schank, 1975J and the conceptual information processing paradigm. His claim was that there are sets of primitives (such as ATRANS, PTRANS, and PROPEL) which represent individual actions, and that language understanding is a heavily semantic-driven process, where syntax plays a very limited role. Schank called his approach a *cognitive simulation.* From the late 1970s to the early 1980s, there was a burst of ideas regarding conceptual information processing, mostly advocated by Schank's research group at Yale. These ideas are well documented in *Inside Computer UnderstandinglSchank,* 1981] and in *Dynamic Memory* [Schank, 1982]. Although the emphasis changed from syntax to semantics, the basic framework followed the traditional AI approach — the systems represented knowledge using primitives and used heuristics to derive a CD representation from an input sentence. Again, this requires that the world model and the knowledge must be complete and hand-crafted.

In the mid 80s, the syntactic approach gained renewed interest. It was motivated by unification-based grammar formalisms, such as Lexical-Functional Grammar (LFG:[Kaplan and Bresnan, 1982], Generalized Phrase Structure Grammar (GPSG:[Gazdar et al., 1985]), and Head-driven Phrase Structure Grammar (HPSG:[Pollard and Sag, 1987]). Introduction of a powerful information manipulation operation, such as unification and well-formalized theories, resulted in some success in developing experimental systems. Once again, however, these approaches were within the traditional AI approach. The frustrating fact is that the major emphasis of these theories was to determine whether or not a given sentence was grammatical. This is a bit of exaggerating, but the point is that modern linguistic theories ignore many phenomena important for building natural language systems. For example, linguistic theories typically do not explain how people understand ungrammatical sentences. Most linguistic theories merely put * in front of the ungrammatical sentence. Many linguistically interesting sentences, such as extremely deep center-embedded sentences, almost never appear in reality. In addition, these theories do not entail a theory on how to disambiguate interpretations, which is a major problem in natural language processing.

Therefore, although there have been progress and changes in emphasis, these approaches are all within the traditional approach towards artificial intelligence. As I have argued in the previous section, systems based on the traditional approach inevitably face a number of problems. For example, developers of commercial machine translation consider that the following problems are inherent in their approach, which is grounded in traditional AI and NLP:

Performance: Performance of most existing machine translation systems is not sufficient for real-time tasks such as speech-to-speech translation. It takes a few seconds to a few minutes to translate one sentence.

Scalability: Current machine translation systems are difficult to scale up because their processing complexity makes the systems' behavior almost intractable.

Quality: Intractability of a system's behavior combined with other factors lowers the quality of translations.

Grammar Writing: By the same token, grammar writing is very difficult since a complex sentence has to be described by piecewise rules. It is a hard and time consuming task, partly due to the intractability of the system's behavior, when these rules are added into the whole system.

Notice that the capability to cope with ungrammatical sentences is not even listed, because such a feature is not listed in the initial specification of the system. Obviously, I do not intend to claim that massively parallel artificial intelligence will immediately open an avenue for high performance and robust natural language systems. The accomplishment of such specifications seems to be far in the future. However, I do argue that reliance on models which assume complete knowledge will never accomplish the goal!

6.2. Memory-Based Model of Natural Language

The alternative to the traditional approach to natural language processing is a memory-based model. The memory-based approach to intelligence has been explicitly discussed since the early 80s. In 1981, Nagao proposed a model of translation based on analogy at a NATO conference (later published in [Nagao, 1984]). This model is the precursor for recent research on memory-based and example-based models of translation. Nagao argued that humans translate sentences, by using similar past examples of translation. In 1985, Stanfill and Waltz proposed a memory-based reasoning paradigm[Stanfill and Waltz, 1988, Stanfill and Waltz, 1986]. The basic idea of memory-based reasoning places memory at the foundation of intelligence. It assumes that large numbers of specific events are stored in memory, and response to new events is handled by first recalling past events which are similar to the new input, and invoking actions associated with these retrieved events to handle the new input. In the same year, Riesbech and Martin proposed the direct memory access parsing model[Rie8beck and Martin, 1986], They argued that the *Build-and-Store* approach toward parsing is incorrect, and proposed the *Recognize-and-Record* approach. The main thrust of the DMAP was to view language understanding as retrieval of episodic memory. DMAP can be viewed as an application of case-based reasoning to parsing.

Despite certain differences among these models, the common thread is to view memory as the foundation of intelligence. This idea runs counter to most AI approaches which place rules or heuristics as the central thrust of reasoning. At the same time, differences between models later became very important. For example, DMAP, in essence, uses complex indexing and heuristics which follows the tradition of AI. Thus, the weaknesses of traditional AI were also exposed when scaling up the DMAP system.

Massively parallel memory-based natural language processing directly inherits these ideas. For example, in the memory-based parsing model, parsing is viewed as a memory-search process which locates the past occurrence of similar sentences. The interpretation is built by activating past occurrences. Rationales for memory-based natural language processing include:

Very Large Finite Space: The Very Large Finite Space (VLFS) concept is critical to the memory-based approach. The most frequently raised concern with memory-based approaches to natural language processing is how these models account for the productivity of language — the ability to generate an infinite set of sentences from a finite grammar. Opponents to the memory-based approach would claim that due to the productivity of language, this approach cannot cover the space of all possible sentences. However, the productivity of language is incorrect. The productivity of language ignores resource boundedness. It should be noted that only a finite number of sentences can be produced when the following conditions stand:

Condition 1: Finite Vocabulary

Condition 2: Finite Grammar Rules

Condition 3: Finite Sentence Length

Conditions 1 and 2 stand, since people only have finite vocabulary and grammar at a given time. Obviously, Condition 3 stands, as there is no infinite length sentence. For example, the longest sentence in one recent CNN Prime News report was 48 words long. 99.7% of sentences were within 25 words of length. Logically, natural language is a set of sentences within a very large finite space (VLFS).

Similarity: In memory-based natural language processing, input sentences are matched against all relevant data to find similar sentences in the data-base. The assumption is that *similar problems have a similar solution.* In fact, a series of experimental data on example-based machine translation indicate reasonable accuracy can be maintained for a relatively broad space. Figure 15 illustrates the relationship between accuracy and normalized distance in EBMT. (This figure is reproduced based on [Sumita and Iida, 1992].) Therefore, a memory-based approach may be able to cover a real solution space with implementable numbers of examples.

Collective Decision: The memory-based approach can take advantage of the redundancy of information implicit in a large data resource. An advantages of this collective decision making is the increased robustness against the inaccuracy of individual data. As has been argued previously, the central limit theorem and the law of large numbers ensure the robustness of the system.

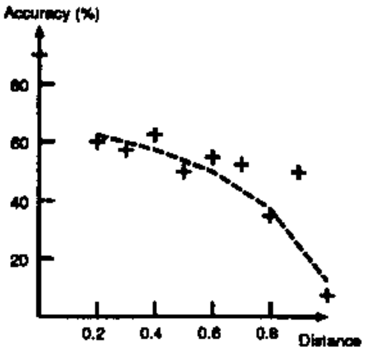


Figure 15: Normalized Distance and Accuracy

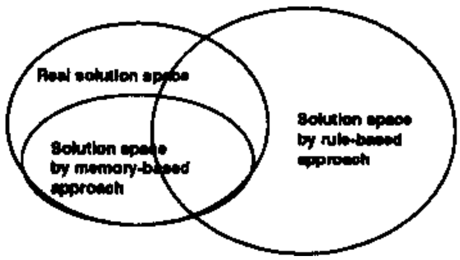


Figure 16: Solution Spaces

Real Space Covering: Assuming that there is a set of grammar rules which covers a certain portion of natural language sentences, the grammar not only covers sentences which actually appear in the real world, but it also covers sentences which are grammatical, but also never produced in reality. An empirical study revealed that only 0.53% of possible sentences considered to be grammatical are actually produced. Figure 16 shows how solution spaces overlap. The memory-based approach never overgenerates because the memory contains only examples of actual sentences used in the past.

In addition to these rationales, it should be noted that the memory-based approach keeps examples, whereas neural networks and statistical approaches do not. This difference is important because the memory-based approach will be able to capture singular events and higher order nonlinearities, while neural networks and statistical approaches often fail to capture these. For neural networks and statistical methods, the ability to capture singular and nonlinear curvature is determined by their network structure or the model. In memory-based reasoning, there are chances that dense lower order interpolation may approximate higher order nonlinear curvature.

6.3. Systems

Since 1988, I have been building several massively parallel natural language systems. ΦDMDIALOG, or DM-DIALOG for short, is a representative system resulting from the early work. ΦDMDIALOG is a speech-to-speech dialog translation system between Japanese and English which accepts speaker-independent, continuous speech. It has been publicly demonstrated since March 1989. The first version of ΦDMDIALOG had a vocabulary of only 70 words. It was extended to 300 words in the second version. Major characteristics of ΦDMDIALOG are:

Massively Parallel Computing Model: Knowledge representation and algorithms are designed to exploit the maximum level of parallelism.

Memory-Based Approach: The knowledge-base consists of a large collection of translation examples and templates. Parsing is viewed as a recall of past sentences in the memory network.

Parallel Constraint Marker-Passing: The basic computational mechanism is marker-passing in which markers carry various information among nodes in the memory network [Charniak, 1983, Charniak, 1986]. This is a useful mechanism which has been studied in various applications [Hendler, 1989a, Hendler, 1989b, Hendler, 1988, Hirst, 1986, Norvig, 1989]. Markers are not bit markers, but are structured markers which carry data structures. In addition, propagation paths for each type of marker are restricted by the types and orders of links to be traversed.

Integration of Speech and NLP: The architecture enables interaction between speech processing and natural language processing. The bottom up process provides the likelihood that a certain interpretation could be correct. The top-down processing imposes constraints and *a priori* probability of phoneme and word hypotheses.

The first version of ΦDMDIALOG was strongly influenced by the idea of DMAP, but substantial extensions has been made. It was a memory-based approach to natural language processing since the parsing process is to recall past sentences in the memory network. However, the focus of the system was on accomplishing speech-to-speech translation. Thus, the system design has been rather conservative by today's standards in the memory-based translation community. In fact, ΦDMDIALOG used a complicated indexing mechanism for the memory network, and markers carried feature structures when markers were propagating to a part of the network used for abstract memory. It used three abstraction levels — specific cases, generalized cases, and unification grammar (Figure 17). Therefore, ΦDMDIALOG can be considered as a mixture of the traditional approach and a memory-based approach.

Despite the fact that the system was not a full implementation of memory-based natural language processing paradigm, ΦDMDIALOG demonstrated various interesting and promising characteristics such as a high level of parallelism, simultaneous interpretation capability, and robustness against inconsistent knowledge. One of the salient features of the model is the integration of speech processing and linguistic processing. In ΦDMDIALOG, activation of the network starts bottom up triggered by external inputs. The activation propagates upward in the network to the word, sentence, and discourse levels. At each level, predictions have been made which are rep-

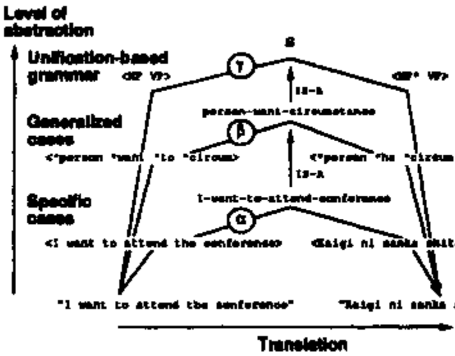


Figure 17: Translation Paths

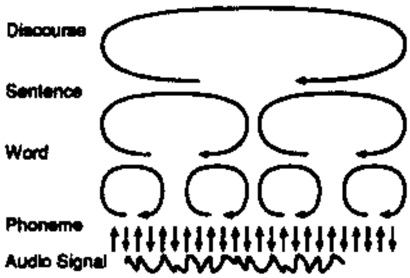


Figure 18: Bottom-Up Activation and Top-Down Prediction

resented by downward propagation of prediction markers with *a priori* probability measures. There are multiple levels of loops in the system (Figure 18).

However, the performance of the system on a serial machine was far from satisfactory. The parallelism was simulated on software in the first version. In addition, increasing emphasis on robustness against ill-formed inputs and the inconsistency of the knowledge-base further degraded the total performance. Fortunately, the high level of parallelism in □DMDIALOG enabled re-implementing the system on various massively parallel computers.

Massively parallel implementations of versions of □DMDIALOG started in the fall of 1989. Two different approaches have been examined. One approach was to emphasize the DMAP approach using complex indexing. DMSNAP was implemented on the Semantic Network Array Processor (SNAP) to examine this approach. One other approach was to focus on memory-based parsing using simpler indexing and a large number of examples. ASTRAL was implemented on IXM-2 based on this approach.

Implementations of $DMDIAL0G on SNAP began in October 1989 when I started a joint research project with Dan Moldovan's group at the University of Southern California. At that time, Moldovan's group had been designing SNAP-1. I worked together with Moldovan's to implement DMSNAP, a SNAP version of □DMDIALOG. The first version was completed by the end of 1990. DMSNAP emphasized complex indexing and dynamic memory network modification to create discourse entities. This follows an approach taken in the case-based reasoning community [Hammond, 1986, Kolodner, 1984, Riesbeck and Schank, 1989]. In a sense, DMSNAP is much closer to the original DMAP idea, rather than to the memory-based reasoning approach.

An independent research program to implement a version of $DMDIAL0G on the IXM2 massively parallel associative memory processor began in March 1990 with Tetsuya Higuchi at ElectroTechnical Laboratory (ETL). IXM2 is an associative memory processor designed and developed by Higuchi and his colleagues at ETL. It is a faithful hardware implementation of NETL [Fahlman, 1979], but using associative memory chips. ASTRAL, the IXM2 version of □DMDIALOG was completed in the summer of 1990. In ASTRAL, complex indexing was eliminated, and a large set of sentence templates were used as a central source of knowledge[Kitano and Higuchi, 1991b].

These differences in emphasis have been employed to make maximum use of architectural strengths for each kind of massively parallel computer. Comparisons between different implementations of □DMDIALOG revealed contributing factors and bottlenecks for robust speech-to-speech translation systems. DMSNAP faced a scaling problem due to reliance on complex indexing, and a performance bottleneck due to node instantiation which inevitably involves an array controller — a serial process. In addition, it is difficult to maintain multiple contexts before an interpretation is uniquely determined. On the other hand, ASTRAL exhibited desirable scaling properties and single millisecond order parsing performance. The DMSNAP project has been re-directed to place more emphasis on the memory-based approach without complex indexing and node instantiation.

While I have been working on massively parallel implementations and performance aspects of memory-based approach to natural language, a series of reports has been made on the quality of the translation attained by memory-based models. Sato-san developed MBT-I[Sato, 1992] for word selection in Japanese-English translation, and attained an accuracy of 85.9%. He extended the idea to the transfer phase of translation in his MBT-II [Sato and Nagao, 1990]. A group at the ATR Interpreting Telephony Research Laboratory developed Example-Based Machine Translation (EBMT:[Sumita and Iida, 1991]) and Transfer-Driven Machine Translation (TDMT:[Furuse and Iida, 1992]). EBMT translates Japanese phrase of type ''A no B" with an accuracy of 89%. TDMT applied the memory-based translation model to translate whole sentence, and attained an accuracy of 88.9%. Since the architecture of TDMT is almost identical to the baseline architecture of □DMDIALOG, the high translation accuracy of TDMT and the high performance of massively parallel versions of $DMDIAL0G indicates a promising future for the approach.

Since 1992, a new joint research project has begun to implement EBMT and TDMT on various massively parallel computers. EBMT was already implemented on the IXM2, CM-2, and SNAP-1 machines [Sumita et al.,

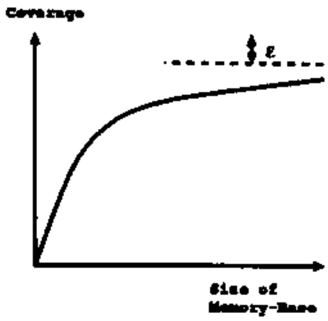


Figure 19: Asymptotic Convergence

1993]. Early results shows that the EBMT approach fits well with massively parallel computers and scales well. Implementation of TDMT is in progress. Independently, Sato has implemented MBT-III [Sato, 1993] on CM-5 and nCUBE machines.

6.4. Limitations and Solutions

Although I strongly believe that the memory-based approach will be a powerful approach in many application domains, there are limitations to this approach.

First, there is the problem of data collection. Here, Zipf s law which was a strong supporting rationale plays a killer role. Although most problems can be defined as having VLFS as their solution space, it is economically and practically infeasible to collect all solutions in a memory-base. Thus, a certain part of the solution space must be left out. If solutions, which are not covered, can be derived from similar examples in the memory-base, the entire solution space can be covered. However, it is most likely that rare examples are truly irregular so that any similar examples cannot derive correct solution. For this type of problems, the only solution at this moment is to keep adding rare events to the memory-base. It should be noted, however, that this problem is not unique to a memory-based approach. In rule-based systems, rules which are specific to rare events must keep being added to a rule-base.

Second, the memory-based approach is not free from the human bias problem. Just like the traditional AI approach, the representations of example and modeling are created by a system designer. For example, current memory-based translation systems use a hand-crafted thesaurus to determine a distance between words. A series of experiments indicated that inaccuracy of thesaurus-based distance calculations is the major source of translation errors. Methods to refine domain theories need to be incorporated [Shavlik and Towell, 1989, Towell et al., 1990]. If inappropriate representation and modeling have been adopted, the system will exhibit poor performance. Even if an appropriate representation and modeling is used, it is implausible that the representation and model would be complete so that 100% accuracy could be attained. Some bias in representation and modeling would be inevitable. When there is some bias, in Figure 19 will not be zero. This means that even if the memory-base contains data to cover the entire solution space, certain errors (ε) still remain. A hybrid approach has been proposed and tested in [Zhang et al., 1992] to overcome the problem of biased internal representations. Their experiment shows certain improvements can be gained by using multiple strategies.

Third, pure memory-based approach models form only a part of the cognitive process. Although the memory-based process may plays an important roles in intelligent activities, use of rules in the human cognitive process cannot be ignored. I believe that there is a place for rules. However, it is likely to be a very different role from what has been proposed in the past. A rule should be created as the result of autonomous generalization over a large set of examples - it is not given a priori. Resource bounds are the major reason for using rules in this context. There must be a cost for Btoring in memory. Thus, a trade-off exists between memory and rules. Pinker [Pinker, 1991] made an interesting observation on the relationship between memory and rules for English verbs. In English, the 13 most frequent verbs are irregular verbs: *be, have, do, say, make, go, take, come, see, get, know, give, find*. Most other verbs are regular. Low frequency irregular verbs are reshaped, and become regular over the centuries. Pinker proposed the rule-associative-memory theory, which considers that irregular verbs are stored in the memory and that regular verbs are handled by rules. In this model, generalization takes place over a set of regular verbs.

Finally, from the engineering viewpoint, explicit rules given *a priori* work as a monitor. For example, human translators use rules which define how specific terms should be translated. Even for human experts, such rules are given in an explicit form. A memory-based model should be able to incorporate such reality. Thus, I recently proposed a model of memory-based machine translation which combines a memory-based approach and a rule-based approach in a novel fashion [Kitano, 1993].

7. Grand Breakthrough

Massively parallel artificial intelligence can be an ideal platform for the scientific and engineering research for next generation computing systems and models of thought. Although discussions have been focused on the memory-based approach, even this approach involves several trails in traditional AI, which are potentially undesirable as a model of thought. Massively parallel memory-based reasoning and massively parallel VLKB search use explicit symbols and do not involve strong learning and adaptation mechanisms by themselves. Thus, they are inherently biased by the features and representation schemes defined by the system designer. Although the problems of representation, domain knowledge, and knowledge-base building can be greatly mitigated by memory-based paradigm, they are not totally free from the problem. Therefore, while there are many places where the present form of massively parallel AI can be useful, we must go beyond this for a next generation paradigm to come into play. At least, it will be a significant scientific challenge. I argue that massively parallel artificial intelligence itself will provide an ideal experimental basis for the genesis of new generation technologies. In this section, therefore, I focus on scientific aspects, rather than engineering. I believe that the following properties will characterizes the realistic theory for intelligence:

Emergence: Intelligence is an emergent property. It emerges from as a result of interactions between subcomponents. Emergence can take place in various levels such as emergence of intelligence, emergence of structure, and emergence of rules for structure development. In this section, the term *emergence* is used to indicate emergence of intelligence.

Evolution: Intelligence is one of the by-products of evolution. Intelligence is the memory of matters since the birth of this universe. Thus, a true theory of intelligence must be justified from an evolutional context.

Symbiosis: Intelligence emerges as a result of symbiotic computing. The essence of intelligence is diversity and heterogeneity. Symbiosis of diverse and heterogeneous components is the key to the intelligence.

Diversity: Diversity is the essence of intelligence. No meaningful artifacts can be created without substantial diversity, the reality of our brain, body, and the eco-system demonstrates the significance of the diversity.

Motivation: Intelligence is driven by motivation. Obviously, learning is a key factor in intelligence. For learning theories to be legitimate from evolutional and psychological contexts, they must involve theories on the motivation for learning. Essentially, learning and other intelligent behavior are driven to the direction which maximizes survivability of the gene.

Physics: Intelligence is governed by physics. Ultimately, intelligent system must be grounded on physics. Nano-technology and device technology provide direct grounding, and digital physics may be able to ground the intelligence on hypothetical physics.

7.1. Emergence

Intelligence is an emergent property. For example, natural language emerged from ethological and biological interaction under a given environment. Our linguistic capability is an emergent property based on our brain structure, and evolved under selectional pressure where individuals with better communication capacity survived better.

Brooks proposed the subsumption architecture as an alternative to a Sense-Model-Plan-Action (SMPA) architecture [Brooks, 1986]. The subsumption architecture is a layered architecture in which each layer is defined in a behavior-based manner. The network consists of simple Augmented Finite State Machines (AFSM), and has no central control mechanism. A series of mobile robots based on the subsumption architecture demonstrated that a certain level of behaviors which would appear to be intelligence can emerge.

Currently, the structure of the system is given *a priori* in the subsumption architecture. Human bias and the problems of the physical symbol systems take place again because definitions for the units of behavior and internal circuitry for each layer must be defined by a designer. In addition, manual designing of the system which contains large numbers of layers for higher levels of intelligence is be extremely difficult.

The research front now has to go into the next level of emergence — the emergence of structure. Structure could emerge using the principles of self-organization and evolution [Jantsch, 1980, Nicolis and Prigogine, 1977]. Self-organization is a vague notion which can be interpreted in many ways. However, since the *a priori* definition for a structure is infeasible, self-organization of a system's structure through dynamic interaction with an environment needs to take place. DARWIN-m developed by Edelman's group [Edelman, 1987, Edelman, 1989] shows some interesting properties for a self-organizing network without explicit teaching signals. However, DARWIN-III is limited to synaptic weight modifications, and has no structure emergence property.

Independently, recent studies by Hasida and his colleagues propose the emergence of information flow as a result of dynamic interaction between the environment and constraints imposed orthogonally [Hasida et al., 1993]. Although this work is still at a preliminary stage, it has an interesting idea that the system has no defined functional unit. Functionality emerges through dynamic interaction with the environment.

The concept of a *Holonic computer* was proposed by Shimizu [Shimizu, et al., 1988]. A holonic computer consists of devices which have non-linear oscillation capability (van del Pol oscillator), which are claimed to be self-organizing. Shimizu emphasizes an importance of *Holonic Loop,* which is a dynamic interaction of elements and top-down constraints emerged from bottom up interactions. Figure 20 is a conceptual image of next generation emergent computers (Needs for the limbic system and the endogenous system will be discussed later).

Although these models have not captured the emergence of structure by themselves, combining these ideas with evolutional computing may enable the emergence of structures and the emergence of structure generating mechanisms.

7.2. Evolution

Intelligence is a by-product of evolution. Species with various forms of intelligence find their niches for survival. Different species could have different forms of intelligence as they have been evolved under different environment, and hence different selectional pressures. If the environment for human beings had been different from what has

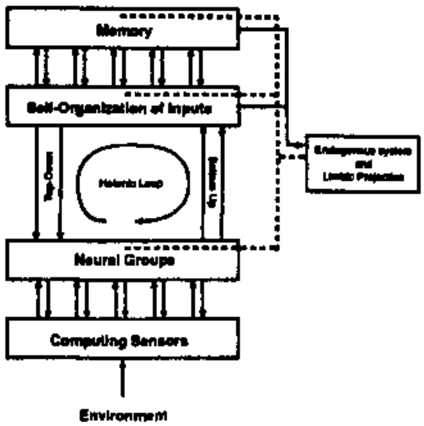


Figure 20: An Architecture for Next Generation Computers?

been imposed as the selectional pressure for mankind, the form of human intelligence could have been very different. Dolphins, for example, are considered to have a comparable numbers of neurons, however, human beings and dolphins have evolved to have different kinds of brains. Physiological differences of the brain inevitably affects the form of intelligence.

The structure of our brain is also affected by our evolutional path. Obviously, brain is not a *tabula rasa,* not only a global structure such as the existence of a brain stem, neo-cortex, and other brain components, but also with numbers of local circuits which are genetically defined such as the hypercolumn, Papez circuits, and hippocampal loops. Investigation of these existing structures of brain and how the brain has evolved may provide substantial insights.

Creating structural systems using evolutional computing, such as genetic algorithms, generally requires use of a developmental phase. While the biological theory for development has not been established, there are some useful mathematical tools to describe development — the Lyndenmayer system [Lindenmayer, 1968, Lindenmayer, 1971] and the Cellular automaton. For example, the L-system can be augmented to handle graph rewriting so that a wide range of development processes can be described (Figure 21). Wilson proposed using the L-system in the context of artificial life [Wilson, 1987], However, I believe I was the first to actually implement and experiment with this evolution of development [Kitano, 1990b]. Since the L-system is a descriptive model, and not a model which describes mechanisms for development, new and more biologically grounded models should be introduced. Nevertheless, the encouraging results achieved in a series of experiments [Kitano, 1990b, Kitano, 1992a, Gruau, 1992] demonstrate that the evolution of development is an important issue.

In addition, the evolutional perspective leads us to investigate the mechanism of immune systems [Tonegawa, 1983], projections from the limbic system, and the endogenous system [McGaugh, 1989]. The central nervous system, immune system, and endogenous system

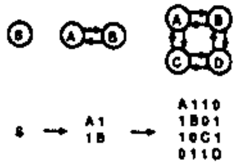


Figure 21: Development using L-system

are known to have the same origin, and to have differentiated in the course of evolution. Projections from the limbic system, which is an old part of the brain, controls awake/sleep rhythms and other basic moods of the brain. These systems play a central role in building a motivation for the system.

7.3. Symbiosis

Symbiotic computing is an idea whereby we view an intelligent system as a symbiosis between various different kinds of subcomponents, each of which is differentiated through evolution. Symbiosis has three major types: host-parasite, fusion, and network. The idea of symbiosis itself is originated in the field of biology. One of the strong proponents of symbiosis is Lynn Margulis. Margulis considered symbiosis as an essential principle for the creation of eukaryotic cells. She claims

> some parts of eukaryotic cells resulted directly from the formation of permanent associations between organisms of different species, the evolution of symbiosis [Margulis, 1981].

While the serial endosymbiosis theory (SET) [Taylor, 1974] is a symbiosis theory for eukaryotic cells, which is a fusion, there are symbioses at various levels.

I claim that symbiosis is one of the central principles for models of thought and life. Particularly, symbiosis is the key for evolution of intelligence. Although genetic algorithms and other genetically-inspired models provide powerful adaptation capabilities, it would be extremely difficult and time-consuming to create highly complex structures from scratch. It may be difficult for genetic algorithms alone to create complex structures with high multimodality. Symbiosis can take place at various level from cells to global multi-agent systems. At the neural level, *neural symbiosis* may be a useful idea to consider.

In *The Society of Mind,* Marvin Minsky postulates intelligence to emerge from the society of large numbers of agents [Minsky, 1986]. The society of mind idea is one instance of symbiosis. It is a weak form of symbiosis, which can be categorized as a network, because individual agents retain their independency. Symbiotic computing offers a more dynamic picture. In symbiotic computing, subcomponents can be merged to form a tight symbiosis or loosely coupled to form a weak symbiosis. A component can be differentiated to form several kinds of components, but could be merged again later. In the light of symbiotic computing, the society of minds and the subsumption architecture are instances, or snapshots, of symbiotic computing.

7.4. Diversity

Diversity is an essential factor for intelligence. For example, in order for the symbiosis to be effective, or even for symbiosis to take place, components involved in symbiosis must have different functions. Thus, the assumption of symbiosis is diversity. Evolution often drives species into niches so that differentiation between species takes place.

Intelligence is a manifestation of a highly complex and dynamic system whose components maintain a high level of diversity. It is analogous to big artifacts such as big buildings and bridges. In physics, particle physicists try to discover the unified theory. Numbers of theories such as the superstring theory and the supersymmetry theor have been proposed. Although these theories comes very close to the unified theory, or even if one of them is the unified theory, they never explain how a building can be built.

Buildings are built using various different components interrelated to each other just like symbiosis. It is impossible to build any meaningful architecture using a single type of component. Theories which claim that one particular mechanism can create intelligence fall into this fallacy. Thus, it is wrong to claim that rules alone can be the mechanism behind intelligence. It is wrong to claim that memory alone can be the mechanism behind intelligence.

Many models of neural networks make the same mistake. Most of models are too simplified as they usually assume a single neuron type. The real brain consists of numbers of different types of neurons interconnected in a specific and perhaps genetically specified manner. In addition, most neural networks assumes electric impulses to be the sole source of inter-neuron communication. Recent biological studies revealed that hormonal effects also affect the state of neurons.

Even if mechanisms at microscopic levels can be described by relatively small numbers of principles, it is analogous to a theory of particle physics. In order for us to formulate the model of thought, we must find how these components create diverse substructures and how they interact.

7.5. Motivation — Genetic Supervision

No learning takes place without motivation. This is the factor which is critically lacking in much current AI research. The term *motivation* is used as a *drive for learning.* Although, both conscious and subconscious drives are important, the following discussion focuses on a subconscious drive.

One example would illustrate the point of my discussion. Infants learn to avoid dangerous things without any teaching signal from the environment. When an infant touches a very hot object, the infant reflexively removes its hand and screams. After several such experiences, the infant would learns to avoid the hot object. The question is why this infant learns to avoid hot objects rather than learning to prefer them. There must be some innate drive, which guides learning in a particular direction. Learning is genetically supervised. We learn to avoid dangerous things because individuals whose learning is driven to that direction have survived better than a group of individuals whose learning is driven to prefer dangerous things. Therefore, learning is genetically supervised in a particular direction which improves the reproduction rate. In addition, studies on child language acquisition seems to support the existence of an innate mechanism for selective learning. Not just direction and focus, but also the learning mechanisms are determined genetically.

Some species of bird demonstrate a form of learning called *imprinting.* For these birds, the first moving object which they see after hatching will be imprinted on their brain as being their mother. Imprinting is an innate learning which allows birds to quickly recognize their mother, so that they can be fed and protected well. While there are certain risks that an individual might imprint something different, evolutional choice has been made to use the particular mechanism under the given environment.

How knowledge should be organized and represented largely depends upon the purpose of learning. If the learning is to avoid dangerous things, features of objects such as temperature, speed, distance, and other factors could be important. In addition, reasoning mechanisms to provide quick real-time response will be used rather than accurate but slow mechanisms. The subsumption architecture captured a mechanism which is largely primitive but important for survival. Thus, these mechanisms must be composed of relatively simple circuitry and provide real-time and autonomous reactions without central control.

The idea of reinforcement learning [Watkins, 1989] provides a more realistic framework of learning than most traditional models of learning which assume explicit teaching signals. Reinforcement learning assumes that the action module received a scalar value feedback called *reward,* according to their action under a certain environment. I feel this is closer to reality. The shortcoming of reinforcement learning is that the evaluation function to provide the reward signal to the action module, is given *a priori.* Ackely and Littman proposed Evolutional Reinforcement Learning (ERL:[Ackley and Littman, 1992]) to remedy this problem. In ERL, the evaluation function co-evolves with the action module so that the individual witha better evaluation function learns more effectively. Independently, Edelman uses *the value system* to drive learning of DARWIN-III to a desired direction. I call this class of models *Genetic Supervision.* A general architecture for genetic supervision should look like Figure 22.

For genetic supervision to be grounded in the biological and evolutional context, several new factors must be considered such as the influence of hormones over mem-

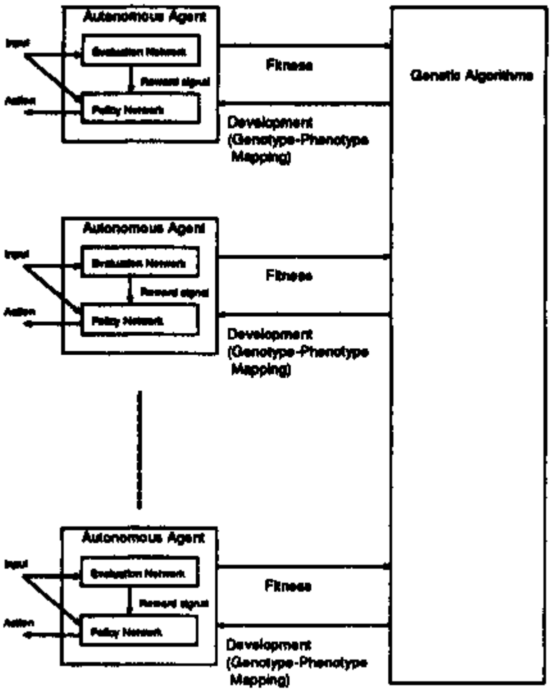


Figure 22: Learning and Evolution

ory modulation [McGaugh, 1989] and A_6 and A_{10} neural systems. These systems have an older origin than neocortex, and control deep emotional and motivational aspects of our behavior. Due to the fact these systems were established at an early stage of evolution, their signal processing-transmission capabilities are limited. Generally, signals from these systems act as an analog global modulator. Nevertheless, I believe that these factors play critically important roles in the next generation of intelligence theories.

7.6. Physics

Since intelligence is a result of evolution and emergence, it is governed by physics. As long as models of intelligence deal with stationary models, the significance of physics does not come into play. However, once the research is directed to more basic properties such as emergence, development, and evolution, physics cannot be ignored. For example, morphogenesis involves cell movements and cell divisions which are largely influenced by physics. Morphogenesis is not necessary, or it can be significantly simplified, if there is no physics involved. However, at the same time, theories without physics inevitably assume *a priori* external constraints. Basic configurations of gene and reproduction mechanisms could have been different if a different physics has been applied. Recent discoveries from the U.S. Space Shuttle missions, particularly by Japanese experiments in Endeavor, uncovered the importance of gravitational influence in cell division and the gene repair. The semantics of the world could have been very different if a spontaneous breakdown of symmetry took place in a very different way. This argument largely assumes artificial life research to be the basis for artificial intelligence. I agree with Belew that artificial life is "a constructive lower bound for artificial intelligence" [Belew, 1991]. Brooks made a good argument for the *embodiment* of robots in the real world at a macroscopic level. The argument in this section is at a more microscopic level. My argument is that physics will be necessary to make emergence, development, and evolution to take place. It should be noted that this is one extreme, and much AI research will be carried out without physics. But, as research delves into basic mechanisms, physics will have to be considered. There are two promising directions, both approaches whcih should be pursued for physical grounding of AI and Alife — Nanotechnology and digital physics.

Nano-technology can be a driving force grounding AI and ALife research to actual physics. The same physics which has been applied to ourselves for billions of years can be applied to AI. This approach will result in enormous impact to society because the very definition of natural life and intelligence will be contested in the light of the state-of-the-art technologies.

To the contrary, digital physics will be able to create a new physics. Massively parallel machines in the near future may be able to produce physics in themselves. This was one of the motivations for building the connection machine [Hillis, 1985]. Just as Langton termed Artificial Life *life-as-it-could-be* [Langton, 1989], digital physics may allow us to pursue the *universe-as-it-could-be.*

8. Conclusion

The arguments developed in this paper can be expressed in a single statement: *Massively parallel artificial intelligence is where AI meets the real world.* The phrase *AI meets the real world* has been used in various contexts, such as *Robotics is where AI meets the real world.* Nevertheless, the phrase has its own significance and that is why it has been repeatedly used.

Artificial intelligence is the field of studying models of intelligence and engineering methods for building intelligent systems. The model of intelligence must face the biological reality of the brain, and intelligent systems must be deployed in the real world. The real world is where we live, and embodies vastness, complexity, noise, and other factors which are not easy to overcome. In order for the AI to succeed, powerful tools to tackle complexity and vastness of the real world phenomena must be prepared. In physics, there are powerful tools such as mathematics, particle accelerators, the hubble telescope, and other experimental and conceptual devices. Biology made a leap when DNA sequencing and gene recombination methods were discovered.

Artificial intelligence, however, has been constrained by available computing resources for the last thirty years. The conceptual devices have been influenced by hardware architectures to date. This is represented by the traditional AI approach. Fortunately, however, the progress of VLSI technology liberates us from the computational constraints.

Computationally demanding and data-driven approaches, such as memory-based reasoning and genetic algorithms emerge as realistic alternatives to the traditional approach. This does not mean that the traditional approach will be eliminated. There are suitable applications for traditional models, and there are places where massively parallel approach will be suitable. Thus, there will be co-habitation of different approaches. However, the emergent new approach will be a powerful method to challenge the goals of AI, which has not been accomplished so far.

Massive computing power and memory space, along with new ideas for AI, will allow us to challenge the realities in our engineering and scientific discipline. For AI to be a hard-core science and engineering discipline, real world AI must be pursued and our arsenals for attacking the problems must be enriched. I hope that discussions developed in this paper contribute to the community making a quantum leap into the future.

Acknowledgements

There are just too many people to thank: Makoto Nagao, Jaime Carbonell, Masaru Tomita, Jay McClleland, David Waltz, Terry Sejnowski, Alex Waibel, Kai-Pu Lee, Jim Gellar, David Goldberg, Dan Moldovan, Tetsuya Higuchi, Satoshi Sato, Hideto Tomabechi, Moritoshi Yasunaga, Ron DeMara, Toyoaki Nishida, Hitoshi Iida, Akira Kurematsu, Hiroshi Okuno, Koiti Hasida, Katashi Nagao, Toshio Yokoi, Mario Tokoro, Lori Levin, Rodney Brooks, Chris Langton, Katsu Shimohara, Scott Fahlman, David Touretzky, Hideo Shimazu, Akihiro Shibata, Keiji Kojima, Hitoshi Matsubara, Eiichi Osawa, Eiichirou Sumita, Kozo Oi... I cannot list them all. Sorry if you are not in the list. Jim Hendler gave me various advices for the final manuscript of this paper. I owe him a lot.

This research is supported, in part, by National Science Foundation grant MIP-9009111 and by Pittsburgh Supercomputing Center IRI-910002P.

References

[Ackley and Littman, 1992] Ackley, D. and Littman, M., "Interactions Between Learning and Evolution," *Artificial Life II,* Addison Wesley, 1992.

[Ae and Aibara, 1989] Ae, T. and Aibara, R., "A neural network for 3-D VLSI accelerator," *VLSI for Artificial Intelligence,* Kluwer Academic Pub., 1989.

[Atkeson and Reinkensymeyer, 1990] Atkeson, C. and Rienkensymeyer, D., "Using Associative Content-Addressable Memories to Control Robots," *Artificial Intelligence at MIT,* 1990.

[Batcher, 1980] Batcher, K., "Design of a Massively Parallel Processor," *IEEE Transactions on Computers* 29, no.9, September 1980.

[Belew, 1991] Belew, R., "Artificial Life: A Constructive Lower Bound for Artificial Intelligence," *IEEE Expert,* Vol. 6, No. 1, 8-15, 1991.

[Blank, 1990] Blank, T., "The MasPar MP-1 Architecture," *Proc. of COMPCON Spring 90,* 1990.

[Blelloch, 1986] Blelloch, G. E., *AFS-1: A programming language for massively concurrent computers,* Tech Report 918, MIT AI Lab., 1986.

[BleUoch, 1986] Blelloch, G. E., "CIS: A Massively Concurrent Rule-Based System," *Proceedings of the National Conference on Artificial Intelligence (AAAI-86),* 1986.

[Bouknight et al., 1972] Bouknight, W. J., et al., "The ILLIACIV System," *Proceedings of the IEEE,* 369-388, April, 1972.

[Bowler and Pawley, 1984] Bowler, K. C. and Pawley, G. S., "Molecular Dynamics and Monte Carlo Simulation in Solid-State and Elementary Particle Physics," *Proceedings of the IEEE,* 74, January 1984.

[Brooks, 1991] Brooks, R., "Intelligence Without Reason," *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-91),* Sydney, 1991.

[Brooks, 1986] Brooks, R., "A robust layered control system for a mobile robot," *IEEE J. Robotics and Automation,* RA-2, 14-23, April, 1986.

[Charniak, 1983] Charniak, E., "Passing Markers: A Theory of Contextual Influence in Language Comprehension," *Cognitive Science,* 7, 171-190, 1983.

[Charniak, 1986] Charniak, E., "A Neat Theory of Marker Passing," *Proc. of AAAI-86,* 1986.

[Chung et al., 1992] Chung, S., Moldovan, D., and DeMara, R., *A Parallel Computational Model for Integrated Speech and Natural Language Understanding,* University of Southern California, TR PKPL-92-2, also as IMPACT-92-008-USC, 1992.

[Cook, 1991] Cook, J., "The Base Selection Task in Analogical Planning," *Proc. of IJCAI-91,* 1991.

[Creecy et al., 1992] Creecy, R., Masand, B., Smith, S., and Waltz, D., "Trading MIPS and Memory for Knowledge Engineering: Automatic Classification of Census Returns on a Massively Parallel Supercomputer," *Communications of the ACM,* 1992.

[Dally et al., 1989] Dally, W. J., et. al. "The J-Machine: a fine-grain concurrent computer," Information Processing 89, Elsevier North Holland, 1989.

[Dixon and de Kleer, 1988] Dixon, M. and de Kleer, J., "Massively Parallel Assumption-based Truth Maintenance," *Proceedings of the the National Conference on Artificial Intelligence (AAAI-88),* 1988.

[Drumheller, 1986] Drumheller, M., "Connection Machine stereomatching," *Proceedings of the National Conference on Artificial Intelligence (AAAI-86),* 1986.

[EDR, 1988] Japan Electric Dictionary Research Institute, "EDR Electric Dictionaries," Technical Report, Japan Electric Dictionary Research Institute, 1988.

[Edelman, 1989] Edelman, G., *The Remembered Present: A Biological Theory of Consciousness,* Basic Books, 1989.

[Edelman, 1987] Edelman, G., *Neural Darwinism: The Theory of Neuronal Group Selection,* Basic Books, 1987.

[Evett et al., 1990a] Evett, M., Hendler, J., and Spector, L., PARKA: Parallel knowledge representation on the Connection Machine, Tech Report CS-TR-2409, University of Maryland, 1990.

[Evett et al., 1990b] Evett, M., Hendler, J., Mahanti, A., and Nau, D., "PRA* A memory-limited heuristic search procedure for the connection machine," Proceedings - Frontiers of Massively Parallel Computing, 1990.

[Fahlman, 1979] Fahlman, S. E., NETL: A System for Representing and Using Real World Knowledge, MIT Press, 1979.

[Furuse and Iida, 1992] Furuse, O. and Iida, H., "Cooperation between Transfer and Analysis in Example-Based Framework," Proc. of COLING-92, Nantes, 1992.

[Gazdar et al., 1985] Gazdar, G., Klein, G., Pullum, G., and Sag, I., Generalized Phrase Structure Grammar, Harvard University Press, 1985.

[Geller, 199l] Geller, J., Advanced Update Operations in Massively Parallel Knowledge Representation, New Jersey Institute of Technology, CIS-91-28, 1991.

[Gelsinger et al., 1989] Gelsinger, P., Gargini, P., Parker, G., and Yu, A., "Microprocessors circa 2000," IEEE Spectrum, Oct. 1989.

[Goodman and Nirenberg, 1991] Goodman, K. and Nirenberg, S. (Eds.), The KBMT Project: A case study in Knowledge-Based Machine Translation, Morgan Kaufmann, 1991.

[Graf et al., 1988] Graf, H., et. al., "VLSI implementation of a neural network model," IEEE Computer, Vol. 21, No. 3, 1988.

[Gruau, 1992] Gruau, F., Celluar Encoding of Genetic Neural Netowrks, Ecole Normale Suerieure de Lyon, Lyon, 1992.

[Hammond, 1986] Hammond, K., Case-Based Planning: An Integrated Theory of Planning, Learning, and Memory, Ph. D. Thesis, Yale University, 1986.

[Hasida et al., 1993] Hasida, K., Nagao, K., and Miyata, T., "Joint Utterance: Intrasentential Speaker/Hearer Switch as an Emergent Phenomenon," Proc. of IJCAI-93, 1993.

[Hendler, 1989a] Hendler, J., "Marker-passing over Microfeatures: Towards a Hybrid Symbolic/Connectionist Model," Cognitive Science, 13, 79-106, 1989.

[Hendler, 1989b] Hendler, J., "The Design and Implementation of Marker-Passing Systems," Connection Science, Vol. 1, No. 1, 1989.

[Hendler, 1988] Hendler, J., Integrating Marker-Passing and Problem-Solving, Lawrence Erlbaum Associates, 1988.

[Higuchi et al., 199l] Higuchi, T., Kitano, H., Furuya, T., Kusumoto, H., Handa, K., and Kokubu, A., "IXM2: A Parallel Associate Processor for Knowledge Processing," Proceedings of the National Conference on Artificial Intelligence, 1991.

[Hillis, 1985] Hillis, D., The Connection Machine, The M.I.T. Press, 1985.

[Hirst, 1986] Hirst, G., Semantic Interpretation and the Resolution of Ambiguity, Cambridge University Press, 1986.

[Holland, 1975] Holland, J., Adaptation in Natural and Artificial Systems, University of Michigan, 1975.

[Hord, 1990] Hord, M., Parallel Supercomputing in SIMD Architectures, CRC Press, 1990.

[Hsu, 1991] Hsu, F. H., "The Role of Chess in AI Research," Proc. of IJCAI-91, 1991.

[Hsu, 1990] Hsu, F. H., Large Scale Parallelization of Alpha-Beta Search: An Algorithmic and Architectural Study with Computer Chess, CMU-CS-90-108, Ph.D. Thesis, 1990.

[Hsu et al, 1990] Hsu, F. H., Anantharaman, T., • Campbell, M. and Nowatzyk, A., "A Grandmaster Chess Machine," Scientific American, Vol. 263, No. 4, Oct., 1990.

[IEDM, 1991] IEEE, Proc. of the International Electron Device Meeting (IEDM-91), 1991.

[Inmos, 1987] Inmos, IMS T800 Transputer, 1987.

[Jantsch, 1980] Jantsch, E., The Self-Organizing Universe, Pergamon Press, 1980.

[Kaplan and Bresnan, 1982] Kaplan, R. and Bresnan, J., "Lexical-Functional Grammar: A Formal System for Grammatical Representation," The Mental Representation of Grammatical Relations, The MIT Press, 1982.

[KSR, 1992] Kendall Square Research, KSR-1 Technical Summary, Kendall Square Research, 1992.

[Kettler et al, 1993] Kettler, B., Andersen, W., Hendler, J. and Evett, M. "Massively Parallel Support for a Case-based Planning System", Proceedings of the Ninth IEEE Conference on AI Applications, Orlando, Florida, March 1993.

[Kikui et al., 1993] Kikui, K., Seligman, M., Takezawa, T., "Spoken Language Translation System: ASURA," Video Proceedings of IJCAI-93, 1993.

[Kitano, 1993] Kitano, H., "A Comprehensive and Practical Model of Memory-Based Machine Translation," Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-93), 1993.

[Kitano et al., 1993] Kitano, H., Wah, B., Hunter, L., Oka, R., Yokoi, T., and Hahn, W., "Grand Challenge AI Applications," Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-93), 1993.

[Kitano, 1992a] Kitano, H., Neurogenetic Learning Carnegie Mellon University, CMT Technical Report, 1992.

[Kitano, 1992b] Kitano, H. (Ed.), Proceedings of the Workshop on Grand Challenge AI Applications, Tokyo, 1992. (in Japanese)

[Kitano and Yasunaga, 1992] Kitano, H. and Yasunaga, M., "Wafer Scale Integration for Massively Parallel Memory-Based Reasoning," Proceedings of the National Conference on Artificial Intelligence (AAAI-92), 1992.

[Kitano et al., 1992] Kitano, H., Shibata, A., Shimazu, H., Kajihara, J., and Sato, A., "Building a Large-Scale and Corporate-Wide Case-Based System," Proc. of AAAI-92, 1992.

[Kitano et al., 1991a] Kitano, H., Hendler, J., Higuchi, T., Moldovan, D. and Waltz, D., "Massively Parallel Artificial Intelligence," Proceeding of the International Joint Conference on Artificial Intelligence (IJCAI-91), 1991.

[Kitano et al, 1991b] Kitano, H, Moldovan, D., Cha, S., "High Performance Natural Language Processing on Semantic Network Array Processor," Proceeding of the International Joint Conference on Artificial Intelligence (IJCAI-91), 1991.

[Kitano et al., 1991c] Kitano, H., Smith, S. F., and Higuchi, T., "GA-1: A Parallel Associative Memory Processor for Rule Learning with Genetic Algorithms," Proceedings of the International Conference on Genetic Algorithms (ICGA-91), 1991.

[Kitano and Higuchi, 1991a] Kitano, H. and Higuchi, T., "High Performance Memory-Based Translation on IXM2 Massively Parallel Associative Memory Processor," Proceeding of the National Conference on Artificial Intelligence (AAAI-91), 1991.

[Kitano and Higuchi, 1991b] Kitano, H. and Higuchi, T., "Massively Parallel Memory-Based Parsing," Proceeding of the International Joint Conference on Artificial Intelligence (IJCAI-91), 1991.

[Kitano and Moldovan, 1991] Kitano, H. and Moldovan, D., "Semantic Network Array Processor as a Massively Parallel Platform for Real-time and Large-Scale Natural Language Processing," Proc of COLING-92, 1992.

[Kitano, 1991] Kitano, H.," DMDULOG: An Experimental Speech-to-Speech Dialogue Translation System," IEEE Computer, June, 1991.

[Kitano, 1990a] Kitano, H.," DMDIALOG: A Speech-to-Speech Dialogue Translation," Machine Translation, 5, 301-338, 1990.

[Kitano, 1990b] Kitano, H., "Designing Neural Networks using Genetic Algorithms with Graph Generation system," Complex Systems, Vol. 4., Num. 4., 1990.

[Kim and Moldovan, 1990] Kim, J. and Moldovan, D., "Parallel Classification for Knowledge Representation on SNAP," Proceedings of the 1990 International Conference on Parallel Processing, 1990.

[Kolodner, 1988] Kolodner, J., "Retrieving Events from a Case Memory: A Parallel Implementation," Proceedings of Case-Based Reasoning Workshop, 1988.

[Kolodner, 1984] Kolodner, J., Retrieval and Organizational Strategies in Conceptual Memory: A Computer Model, Lawrence Erlbaum Assoc, 1984.

[Langton, 1991] Langton, C, Taylor, C, Farmer, J., and Rasmussen, S.., (Eds.) Artificial Life II, Addison Wesley, 1991.

[Langton, 1989] Langton, C, (Ed.) Artificial Life, Addison Wesley, 1989.

[Lee, 1988] Lee, K. F., Large-Vocabulary Speaker-Independent Continuous Speech Recognition: The SPHINX System, CMU-CS-88-148, Ph.D. Thesis, 1988.

[Lenat and Feigenbaum, 1991] Lenat, D. B. and Feigenbaum, E., "On the thresholds of knowledge," Artificial Intelligence, 47, 185-250, 1991.

[Lenat and Guha, 1989] Lenat, D. B. and Guha, R. V., Building Large Knowledge-Based Systems, Addison-Wesley, 1989.

[Lindenmayer, 1971] Lindenmayer, A., "Developmental Systems without Celluar interactions, their Languages and Grammars," J. Theor. Biol, 30, 455-484, 1971.

[Lindenmayer, 1968] Lindenmayer, A., "Mathematical Models for Celluar Interactions in Development," J. Theor. Biol, 18, 280-299, 1968.

[Margulis, 1981] Margulis, L., Symbiosis in Cell Evolution, Freeman, 1981.

[McGaugh, 1989] McGaugh, J., "Involvement of Hormonal and Neuromodulator Systems in the Regulation of Memory Storage," Ann. Rev. Neurosci., 12:255-87, 1989.

[Mead, 1989] Mead, C, Analog VLSI and Neural Systems, Addison Wesley, 1989.

[Minsky, 1986] Minsky, M., The Society of Mind, Simon and Schuster, New York, NY, 1986.

[Moldovan et al., 1990] Moldovan, D., Lee, W., and Lin, C, SNAP: A Marker-Passing Architecture for Knowledge Processing, Technical Report PKPL 90-4, Department of Electrical Engineering Systems, University of Southern California, 1990.

[Morimoto et al., 1989] Morimoto, T., Iida, H., Kurematsu, A., Shikano, K., and Aizawa, T., "Spoken Language Translation," Proc. of InfoJapan-90, Tokyo, 1990.

[Myczkowski and Steele, 199l] Myczkowski, J. and Steele, G., "Seismic Modeling At 14 Gigaflops On The Connection Machine," Supercomputing-91, 1991.

[Nagao, 1984] Nagao, M., "A Framework of a Mechanical Translation between Japanese and English by Analogy Principle," Artificial and Human Intelligence, Elithorn, A., and Banerji, R. (Eds.), Elsevier Science Publishers, B.V. 1984.

[NRC, 1988] National Research Council, Mapping and Sequencing the Human Genome, National Academy Press, Washington, D.C., 1988.

[Nickolls, 1990] Nickolls, J., "The Design of the MasPar MP-1: A Cost Effective Massively Parallel Computer," Proc. of COMPCON Spring 90, 1990.

[Nicolis and Prigogine, 1977] Nicolis, G. and Prigogine, I., Self-Organization in Nonequibillium Systems, Wiley, New York, NY, 1977.

[Norvig, 1989] Norvig, P., "Marker Passing as a Weak Method for Text Inferencing," Cognitive Science, 13, 569-620, 1989.

[Pinker, 1991] Pinker, S., "Rules of Language,"; Science, Vol. 253, 530-535, 1991.

[Pollard and Sag, 1987] Pollard, C. and Sag, L, Information-Based Syntax and Semantics, Volume 1, CSLI Lecture Notes, 13, 1987.

[Quillian, 1968] Quillian, M. R., "Semantic Memory," Semantic Information Processing, Minsky, M. (Ed.), 216-270, The MIT Press, 1968.

[Riesbeck and Martin, 1986] Riesbeck, C. and Martin, C, "Direct Memory Access Parsing," Experience, Memory, and Reasoning, Lawrence Erlbaum Associates, 1986.

[Riesbeck and Schank, 1989] Riesbeck, C. and Schank, R., Inside Case-Baaed Reasoning, Lawrence Erlbaum Associates, 1989.

[Robertson, 1987] Robertson, G., "Parallel Implementation of Genetic Algorithms in a Classifier System," Davis, L., (Ed.) Genetic Algorithms and Simulated Annealing, Morgan Kaufmann Publishers, 1987.

[Rumelhart and McClelland, 1986] Rumelhart, D. and McClelland, J., Parallel Distributed Processing: Explorations in the Microstructure of Cognition, The MIT Press, 1986.

[Sabot et al, 1991] Sabot, G., Tennies, L., Vasilevsky, A. and Shapiro, R., "Compiler Parallelization of an Elliptic Grid Generator for 1990 Gordon Bell Prize," Supercomputing-91, 1991.

[Saito and Tomita, 1988] Saito, H. and Tomita, M., "Parsing Noisy Sentences," Proc. of COLING-88, 1988.

[Sato and Nagao, 1990] Sato, S. and Nagao, M., "Toward Memory-based Translation," Proceedings of COLING-90, 1990.

[Sato, 1992] Sato, S., Example-Based Machine Translation, Ph.D. Thesis, Kyoto University, 1992.

[Sato, 1993] Sato, S., Example-Based Translation of Technical Terms, IS-RR-93-41, Japan Advanced Institute of Science and Technology (JAIST), 1993.

[Schank, 1982] Schank, R., Dynamic Memory, MIT Press, 1982.

[Schank, 1981] Schank, R. and Riesbeck, C, Inside Computer Understanding, Lawrence Erlbaum Associates, 1981.

[Schank, 1975] Schank, R., Conceptual Information Processing, North Holland, 1975.

[Seeds, 1967] Seeds, R., "Yield and Cost Analysis of Bipolar LSI," Proc. of IEEE International Electron Devices Meeting, 1967.

[Sejnowski and Rosenberg, 1987] Sejnowski, T. and Rosenberg, C, "Parallel networks that learn to pronounce English text," Complex Systems, 1: 145-168, 1987.

[Shavlik and Towell, 1989] Shavlik, J. and Towell, G., "An approach to combining explanation-based and neural learning algorithms," Connection Science, 1:233-255, 1989.

[Shimizu, et al, 1988] Shimizu, H., Yamaguchi, Y., and Satoh, K., "Holovision: a Semantic Information Processor for Visual Perception," dynamic Patterns in Complex Systems, (Eds.) Kelso, S., et. al., World Scientific, 42-76, 1988.

[Spiessens and Bernard, 199l] Spiessens, P., and Bernard, M., "A Massively Parallel Genetic Algorithm: Implementation and First Analysis," Proc. of ICGA-91, 1991.

[Stanfill and Waltz, 1988] Stanfill, C. and Waltz, D., "The Memory-Based Reasoning Paradigm," Proceedings of the Case-Based Reasoning Workshop, DARPA, 1988.

[Stanfill and Waltz, 1986] Stanfill, C. and Waltz, D., "Toward Memory-Based Reasoning," Communications of the ACM, 1986.

[Stanfill et al., 1989] Stanfill, C, Thau, R., and Waltz, D., "A Parallel Indexed Algorithm for Information Retrieval," Proceedings of the SIG-IR-89, 1989.

[Stormon et al, 1992] Stormon, C, Troullinos, N., Saleh, E., Chavan, A., Brule, M., and Oldfield, J., "A General-Purpose CMOS Associative Processor IC and System." IEEE Micro, December, 1992.

[Sumita et al., 1993J Sumita, E., Oi, K., Iida, H., Higuchi, T., Kitano, H., "Example-Based Machine Translation on Massively Parallel Processors," Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-93), 1993.

[Sumita and Iida, 1992] Sumita, E., and Iida, H., "Example-Based Transfer of Japanese Adnominal Particles into English," IEICE Transaction of Information and Systems, July, 1992.

[Sumita and Iida, 1991] Sumita, E., and Iida, H., "Experiments and Prospects of Example-Based Machine Translation," Proceedings of the Annual Meeting of the ACL (ACL-91), 1991.

[NOAH, 1992] Systems Research & Development Institute of Japan, A Plan for the Knowledge Archives Project, 1992.

[Towell et al., 1990] Towell, G., Shavlik, J. and Noordewier, M., "Refinement of Approximate Domain Theories by Knowledge-Based Neural Networks," Proc. of AAAI-90, 1990.

[Taylor, 1974] Taylor, F., "Implications and extensions of the serial endosymbiosis theory of the origin of eukaryotes," Taxon, 23:229-258, 1974.

[TMC, 1991] Thinking Machines Corporation, The Connection Machine CM-5 Technical Summary, 1991.

[TMC, 1989] Thinking Machines Corporation, Model CM-2 Technical Summary, Technical Report TR-89-1, 1989.

[Tonegawa, 1983] Tonegawa, S., "Somatic Generation of Antibody Diversity," Nature, Vol. 302, No. 5909, 575-581, April 14, 1983.

[Twardowski, 1990] Twardowski, K., "Implementations of a Genetic Algorithm based Associative Classifier System (ACS)," Proc. of International Conference on Tools for Artificial Intelligence, 1990.

[Waibel et al., 199l] Waibel, A., Jain, A., NcNair, A., Saito, H., Hauptmann, A., and Tebelskis, J., "JANUS: A Speech-to-speech translation system using connectionist and symbolic processing strategies," IEEE Proceedings of the 1991 International Conference on Acoustics, Speech, and Signal Processing, April, 1991.

[Waltz, 1991] Waltz, D., "Eight Principles for Building an Intelligent Robot," Proceedings of the First International Conference on Simulation of Adaptive Behavior: Prom animals to animats, MIT Press, 1991.

[Waltz, 1990] Waltz, D., "Massively Parallel AI," Proceedings of the National Conference on Artificial Intelligence (AAAI-90), 1990.

[Waltz, 1988] Waltz, D., "The Prospects for Building Truly Intelligent Machines," Graubard, S. (Ed.), The Artificial Intelligence Debate, The MIT Press, 1988.

[Waltz and Pollack, 1985] Waltz, D. and Pollack, J., "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation," Cognitive Science, 9(1): 51-74, 1985.

[Watkins, 1989] Watkins, C, Learning with Delayed Rewards, Ph. D. Thesis, Cambridge University, 1989.

[Werner and Dyer, 1991] Werner, G. and Dyer, M., "Evolution of Communication in Artificial Organisms," Artificial Life II, Addison Wesley, 1991.

[Wilson, 1987] Wilson, S., "The genetic algorithm and biological development," Proc. of International Conference on Genetic Algorithms, 1987.

[Winograd, 1972] Winograd, T., Understanding Natural Language, Academic Press, New York, 1972.

[Woods, 1970] Woods, W., "Transition network grammars for natural language analysis," Communications of the ACM, 13:10, 591-606, 1970.

[Woods et al., 1972] Woods, W., Kaplan, R., Nash-Webber, B., The lunar sciences natural language information system: Final report, Bolt, Beranek and Newman Report No. 2378, Cambridge, Mass., 1972.

[Yasunaga et al., 199l] Yasunaga, et al., "A Self-Learning Neural Network Composed of 1152 Digital Neurons in Wafer-Scale LSIs," Proc. of the International Joint Conference on Neural Networks at Singapore (IJCNN-91), 1991.

[Zhang et al., 1988] Zhang, X., Waltz, D., and Mesirov, J., Protein Structure Prediction by Memory-based Reasoning, Thinking Machine Corporation, 1988.

[Zhang et al., 1992] Zhang, X., Mesirov, J., and Waltz, D., "Hybrid System for Protein Secondary Structure Prediction," J. Mol Biol, 225, 1049-1063, 1992.