
Part I Research Trends in the ICT Inspired by Life

“Where the world ceases to be the scene of our personal hopes and wishes, where we face it as free beings admiring, asking and observing, there we enter the realm of Art and Science.”

—Albert Einstein—

1 The ICT Inspired by Life

— The Design of ICT System Inspired by Evolution of Life and Brain Functions —

SAWAI Hidefumi

The significance of “biologically inspired paradigm” as the creation of a new paradigm is described showing the scientific position in the course of biological evolution and the hierarchy of nature by referring to the Darwinian theory of evolution as a change of paradigm in the scientific revolutions. The structure and functions of brain as a by-product of biological evolution, the various information processing models with time and space structures by deducing from the brain functions, the genetic and evolutionary algorithms as the models of evolution itself, and the algorithm based on sexual selection, are described. Furthermore, as the results of trend survey toward the recently developing complex network science research, the useful design guidelines for constructing the future information and communications society can be obtained.

Keywords

Life, Evolution, Brain, Information, Network

1 Why Draw Inspiration from Life?

The reasons for drawing inspiration from life are self-evident. Take, for example, the brain, the seat of intelligence and awareness. The product of four billion years of sustained evolution, this unparalleled object is most definitely not an overnight creation.

1.1 Proximate and Ultimate Factors

To be “inspired by life” and to build intelligent information and communication systems, we must not only investigate the func-

tions of the organisms currently found on Earth, but penetrate deep to discover how the diverse and ingenious functions observed today were acquired over the long course of evolution. In other words, we need to look back at the developmental stages in the evolution of living organisms — an ultimate factor — in addition to proximate factors, proximate factors being the adaptation of various organisms to their immediate environments. The relationship between proximate factors and ultimate factors is briefly discussed below. No understanding of the historical background

and significance of modern life science or cognitive science is complete without an understanding of this relationship.

1.2 Nature's Hierarchy

"The most beautiful thing we can experience is the mysterious. It is the source of all true art and science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed."

—Albert Einstein—

It is believed that the universe was created about 15 billion years ago and the Solar System and the Earth approximately 4.6 billion years ago. Primitive life forms are believed to have appeared around 3.8 – 4 billion years ago. Chemical evolution during the first 600 million years after the formation of Earth eventually led to the origin of life.

Table 1 presents the hierarchy of nature and the disciplines and research topics associated with each level of this hierarchy. At the smallest scales are elementary particles,

atoms, and molecules. The topics discussed in this special issue that correspond to this level of the hierarchy are artificial chemistry in Chapter 2 and molecular communications in Part II. The next level of the hierarchy is genetic. Section 3 of Chapter 1 introduces genetic algorithms and algorithms based on genetic duplication. Beyond this is the amino-acid level of this hierarchy; topics associated with this level include chemical genetic algorithms (CGAs) and chemical genetic programming (CGP), presented in Section 4. Part II discusses molecular communication and motor proteins as topics at the protein level of the hierarchy.

As issues at the cellular level, Section 2 of Chapter 1 discusses neural network modeling based on the model of neurons and synapses. Note that Ca ion diffusion, a topic discussed in Part II, is also associated with this hierarchical level. At the level of tissues and organs, researchers are currently pursuing studies of brain machine interfaces (BMIs) to extract and apply neuronal activity outside the brain as part of efforts intended to investigate con-

Table 1 Nature's hierarchy

Hierarchical Levels	Fields & Subjects of Research
Elementary particles, atoms, and molecules	Nano-bio science, artificial chemistry, molecular communication
Genes	Molecular biology, neutral theory of evolution, virus evolution
Amino acids	Chemical genetic algorithm/programming (CGA/CGP)
Proteins	Protein engineering, molecular communication (motor proteins)
Cells	Neurons, molecular communication (Ca ion diffusion)
Tissues and organs	Tissue engineering, brain and mind, consciousness
Organism	Origin of species, sexual selection, game theory
Population	Immune system, ESS*, evolution of altruistic behavior, multi-agent system
Species	Darwinian theory of evolution, Neo-Darwinism, co-evolution
Ecosystem	Theory of habitat segregation, mimicry, migration strategy of population
Earth	The Gaia hypothesis, environmental problems
The Solar system, the Galaxy, and the universe	Origin of life

*ESS: Evolutionary Stable Strategy

consciousness, the state emerging from the brain functions. *The Origin of Species* by Darwin is associated with the organism level of the hierarchy. Section 3.4 of Chapter 1 discusses algorithms inspired by the theory of sexual selection. At the level of populations and species, perhaps the most characteristic topic is co-evolution involving multiple agents; the acquisition of behavioral strategy by multiple agents in CGP discussed in Section 4.2 is indeed based on the mechanism of co-evolution. At the ecosystem level, the topic presented is parallel distributed processing for parameter-free genetic algorithms (presented in Section 3.2 of Chapter 1), a model inspired by the migrational strategies of various populations. At the level of the Earth, environmental problems constitute the most pressing topics. At the level of the universe, when we consider that the origin of life is a lifeless molecule, we understand that the vast scale of the universe is closely linked to small-scale structures at the level of elementary particles, atoms, and molecules. We also consider how what we observe of the universe (nature's hierarchy) reflects a mode of observation unique to humans.

A satisfactory Theory of Everything, or TOE, that provides a unified explanation for all levels of the hierarchy presented in Table 1 has yet to be established. Micro-scale theories such as elementary particle theory and quantum mechanics and macro-scale theories such as electrodynamics, Newtonian dynamics, and Einstein's general theory of relativity have resolved many of nature's mysteries. But the handling of meso-scale problems remain a challenge, and the respective theories remain works in progress, despite the emergence of studies on complexity (complex science) through chaos theory. As is well-known, the discovery of DNA's double helix structure by Watson and Crick in 1953 triggered rapid progress in molecular biology and has pushed the discipline to its current prominence. In the information sciences, research has declined in the area of artificial intelligence, applications of which include expert systems that seek to

embody human expertise. However, studies continue on artificial neural networks, artificial life, and artificial chemistry, which have come to be regarded as fundamental areas of study. Chapters 1 and 2 of this issue introduces recent research achievements in this area. While Descartes's formulation of the mind-body duality has sequestered physical and spiritual or mental issues as problems in totally distinct dimensions, recent progress in cognitive science and the development of technologies for non-invasive measurement of brain functions such as f-MRI, MEG, NIRS, and EEG have transformed the study of human consciousness and subjective perception into subjects of natural science (cognitive science). This is a potentially epochal event in the history of scientific and technological development.

1.3 Scientific Revolutions as Paradigm Shifts

“Possibly we shall know a little more than we do now. But the real nature of things, that we shall never know, never.”

—Albert Einstein—

Let's examine how scientific revolutions can be viewed as paradigm shifts, drawing on Thomas Kuhn's *The Structure of Scientific Revolutions*[1]. According to Kuhn's definition, “paradigms” are “universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners.” Kuhn states that “when the profession can no longer evade anomalies that subvert the existing tradition of scientific practice—then begin the extraordinary investigations that lead the profession at last to a new set of commitments, a new basis for the practice of science. The extraordinary episodes in which that shift of professional commitments occurs are the ones known in this essay as scientific revolutions. They are the tradition-shattering complements to the tradition-bound activity of normal science.” As a prime example of a “scientific revolution” he gives the discovery of the general theory of

relativity by Albert Einstein to explain the nature of space and time—a theory that completely overturned the concept of absolute space and time presented by Newton[2]. Newtonian dynamics failed to explain Mercury’s perihelion movement (precession), the gravitational redshift of the light spectrum, or the gravitational diffraction of light, all phenomena successfully and quantitatively explained for the first time by expressing spaces in which gravitational fields are found as distortions of space by metric tensor $\{g_{ij}\}$ in Riemannian space and deriving the equation of the gravitational field based on Einstein’s general theory of relativity.

Sir Arthur S. Eddington, a British astronomer, observed that the amount of diffracted light arriving from the opposite side of the Sun during a solar eclipse curved by the gravitational field of the Sun exactly matched the values predicted by Einstein’s field equations, demonstrating the validity of Einstein’s theory. This example illustrates how scientific revolutions involve a shift from an old paradigm to a new one. But what does the creation of a new “inspired by life” paradigm mean? To understand this, we must first learn from the theory of Darwin, the father of evolution.

1.4 Darwin’s Theory of Evolution

“The most incomprehensible thing about the world is that it is at all comprehensible.”

—Albert Einstein—

In his book, *The Origin of Species* published in 1859, Darwin writes the following[3]:

It is interesting to contemplate a tangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex a manner, have all been produced by laws acting around us. These laws, taken in the largest sense, being Growth with Reproduction; Inheritance which is almost implied by reproduc-

tion; Variability from the indirect and direct action of the conditions of life, and from use and disuse: a Ratio of Increase so high as to lead to a Struggle for Life, and as a consequence to Natural Selection, entailing Divergence of Character and the Extinction of less-improved forms. Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows. There is grandeur in this view of life, with its several powers, having been originally breathed by the Creator into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and most wonderful have been, and are being evolved.

Richard Leakey, an evolutionary biologist, justifies Darwin’s status as the father of the modern theory of evolution as follows[4]. First, Darwin meticulously and systematically categorized each and every piece of evidence associated with the problem of evolution. In his youth, Darwin boarded the surveying vessel HMS Beagle as a naturalist and spent five productive years aboard the vessel along its various journeys (1831–1836). Introducing him to diverse geological and biological phenomena and giving him time to accumulate knowledge, and formulate ideas, this long voyage was crucial to his future emergence as an eminent naturalist. As early as 1837, he became convinced that species were neither permanent nor fixed. During the period between 1837 and 1859, he read widely, meditated deeply, and carried out carefully designed experiments. This extended period of preparation may explain why *The Origin of Species* succeeds in covering such a wide range of topics and at such depth. Second, claims Leakey, Darwin was able to present a mechanism that convincingly explained how changes occur in species—namely, Natural Selection. Darwin first conceived the idea of natural selection in 1838, inspired by *An Essay on the Principle of Population* by Thomas Robert Malthus, a parson and socio-economist of the early 19th century. While Malthus was

primarily interested in population, his essay points out that the number of offspring produced by an organism is greater than the number expected to live to reproductive age, identifying this as a general principle in nature.

Based on Darwin's theory of evolution, John Holland in the 1960s advanced the idea of genetic algorithms (GA) as a computational model[6]. Genetic algorithms and their evolutionary forms (e.g., genetic programming and evolutionary computation) are discussed in greater detail in a later chapter.

1.5 The Phylogenetic Tree

"The distinction between past, present, and future is only a stubbornly persistent illusion."

—Albert Einstein—

In 1837, Darwin realized that the evolutionary genealogy between organisms could be more clearly represented using a tree diagram. (Figure 1 is a schematic drawing of a phylogenetic tree.)

Over 10 million rich and diverse species populate the globe today, all descended from a single primordial cell existing some four billion years ago. Advances in fossil survey and DNA analysis techniques have made it possible to put the time at which humans diverged from orangutans at some 13 million years ago; from gorillas at some 6.5 million years ago; and from chimpanzees at some five million years ago. Gradual accumulations of small dif-

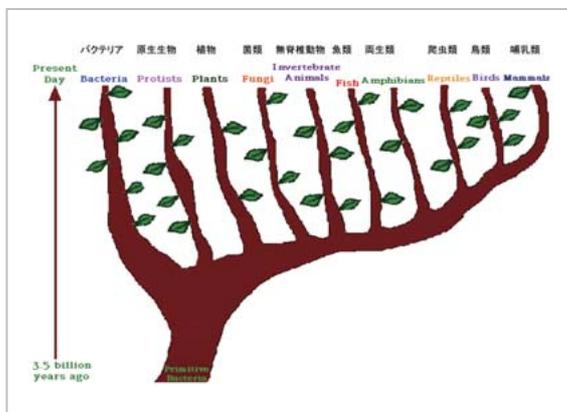


Fig.1 Schematic diagram of a phylogenetic tree

ferences in genotype created by mutations eventually result in significant differences in phenotype, in due course leading to brains and intelligent organisms capable of adapting to their environment. The term “intelligence” as used here will refer both to the adaptive functions of an organism that affect its survival and the exquisite functions emergent from the brain.

2 Information Processing Based on Brain Function Modeling

2.1 Brain Structures and Their Functions

This section presents a summary of the structure and functions of the brain. As shown in Fig. 2, the brain is divided into left and right hemispheres. The brain can also be divided into the following units: the cerebrum, consisting of frontal, parietal, temporal, and occipital lobes; the cerebellum; the brain stem; and the spinal cord. Each part has a modular structure consistent with its function. The brain constitutes a system of immense complexity, composed of some 100 billion neurons (counting both the cerebrum and cerebellum), with each neuron connected to other neurons by several thousand to tens of thousands of synapses. This complex organ is the product of evolution. Figures 3 and 4 present the structures of a neuron and a neural network, respectively. As shown in Fig. 5, a synapse in essence is the space between two neurons

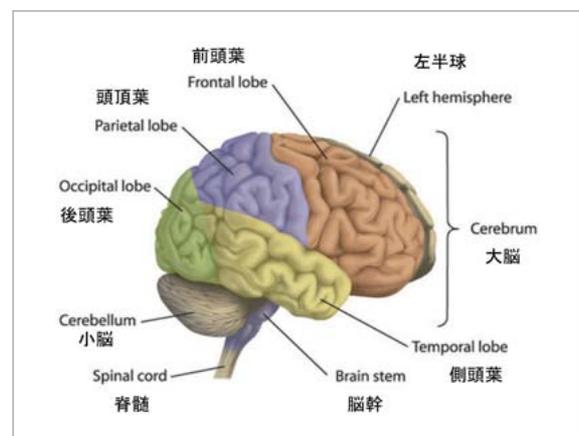


Fig.2 Brain structure

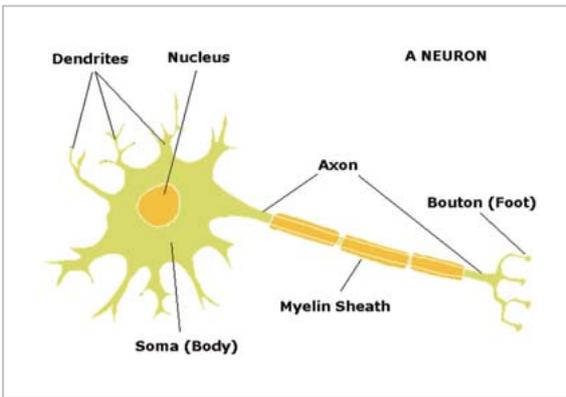


Fig.3 Neuron (Nerve cell)

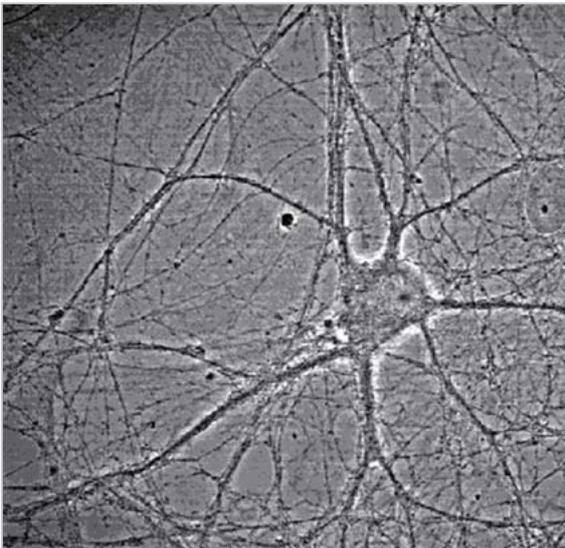


Fig.4 Neural network

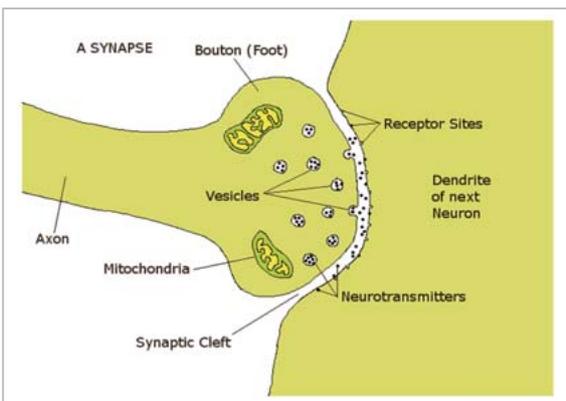


Fig.5 Synapse

across which electric pulses traveling along an axon are relayed to an adjacent dendrite via chemical transmitters.

A neural network is an example of an information processing model based on brain

function. Presented below are modeling and design methods for neural network architectures suitable for speech or image pattern recognition.

Special focus will be given to the feature extraction technique tailored to the temporal structure of speech and the spatial structure (information structure) of images.

2.2 Basic Architecture of Neural Networks

Figure 6 is a model of the neuron proposed by McCulloch & Pitts in 1943. The input signal x_i ($i = 1, 2, \dots, n$) is input to the neuron, with the links representing synapses weighted by w_i . The resulting internal potential u is the sum of these products, or $\sum w_i x_i$. The sum is input to the threshold function $y = f(u)$, which is $y = 0$ when the value of u is below a certain threshold value of θ and $y = 1$ when u is greater than the threshold. Thus, this neuron model adopts a weighted majority logic based on threshold value. The threshold function $y = f(u)$ is generalized into a differentiable sigmoid function $y = 1 / \{1 + \exp(-x + \theta)\}$ and applied to the learning rule by error back-propagation for the artificial neural network (ANN) shown in Fig. 7.

2.3 Architecture of Time-Delay Neural Networks

Figure 8 shows a unit used in a time-delay neural network (TDNN)[7]. The architecture shown here was proposed to process time-series data—for example, speech. To the left is the input unit, which is linked to the host output unit via synapse W , which can be a non-time-delay concatenation or a time-delay concatenation with a delay of D_1, D_2, \dots, D_n . This architecture is suitable for processing speech signal patterns that have temporal structures. Weighted summation (Σ) is performed on the input signals, and the resulting value is sent to the sigmoid function $F(x) = 1 / \{1 + \exp(-x)\}$ and the result output.

Figure 9 shows the architecture of a TDNN designed to distinguish the voiced plosives /b, d, g/ in the Japanese language[7].

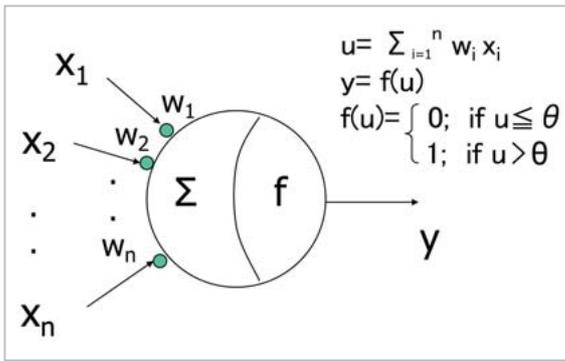


Fig.6 Formal neuron in McCulloch & Pitts

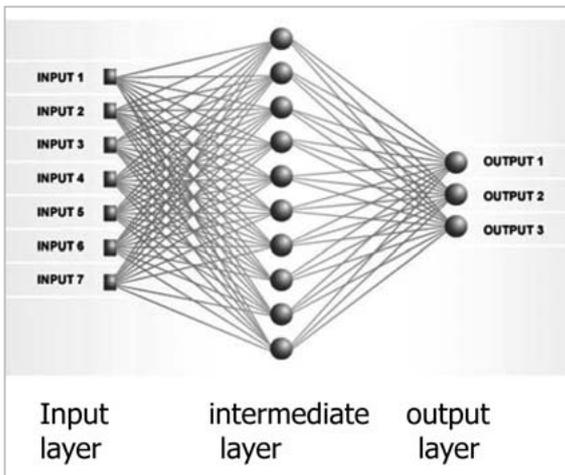


Fig.7 Artificial neural network (ANN)

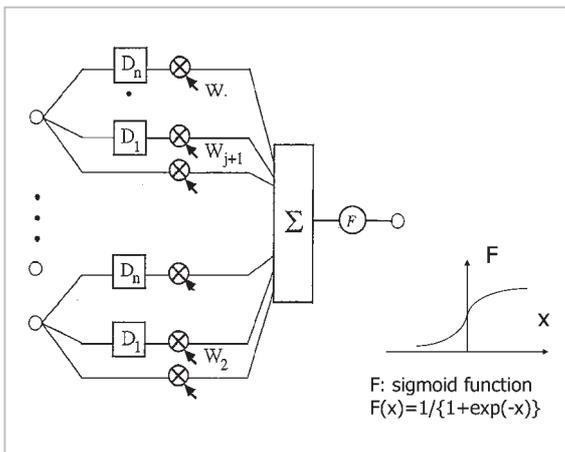


Fig.8 Unit of the time-delay neural network (TDNN)

From bottom to top are the input layer, hidden layer 1, hidden layer 2, and output layer. The input layer consists of a total of 240 units, or 15 and 16 units in each of the x- and y-axis directions.

The x- and y-axis correspond to the time-

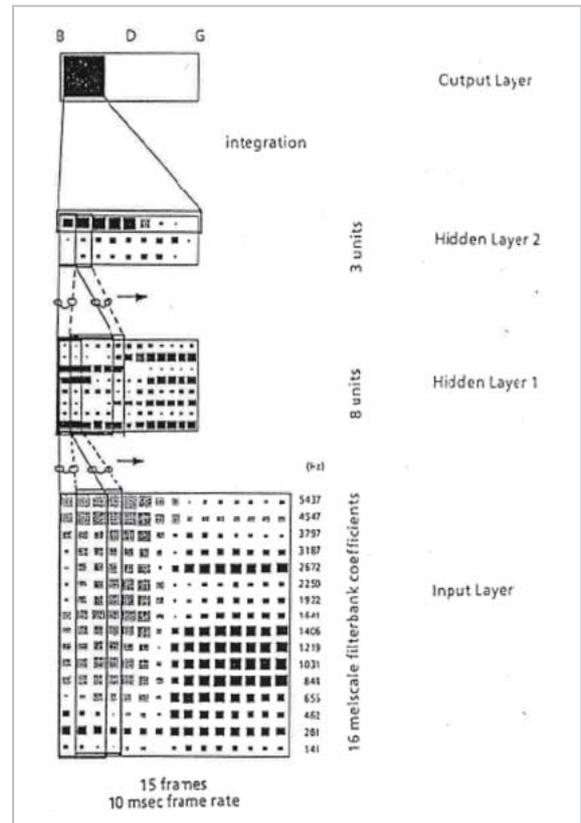


Fig.9 Time-delay neural network (TDNN) architecture

and frequency-axis, respectively. The time-frequency spectrum (sound spectrum) produced by frequency analysis performed every 10 ms is input to the input layer. This time-frequency spectrum is shifted by a time window of 30 ms (3 units) and after being multiplied by the weighting factors of synapses having time-delay, it is linked to the host unit in hidden layer 1. As in the input layer, the total of 40 units in hidden layer 1, five units in the x-axis (time-axis) direction, and the eight units in the y-axis (frequency-axis) direction are linked to hidden layer 2 with the time-delay. In hidden layer 2, each of the nine units in the x-axis direction is assigned to one of three categories /b, d, g/, and concatenated to the output unit with the time-delay. This TDNN is trained by error-propagation. The results of discrimination testing for input phonemes not used for training indicate that the TDNN achieves a speaker-dependent recognition rate of approximately 98–99 %, representing a reduction in misrecognition rate to approxi-

mately 1/4 that of the conventional HMM (Hidden Markov Model) widely used for speech recognition applications. (HMM is associated with a recognition rate of 91–97 %.) By expanding the phonemic category in a similar manner, we should be able to build a TDNN for all Japanese phoneme groups.

Figure 10 shows the modular architecture of a TDNN capable of discriminating the 18 consonants of the Japanese language[8]. As the figure shows, the 18 consonants are divided into six groups—the voiced plosives /b, d, g/, unvoiced plosives /p, t, k/, nasals /m, n, N/,

fricatives /s, sh, h, z/, affricates /ch, ts/, and liquids and semi-vowels /r, w, y/. A TDNN capable of distinguishing between the six phoneme groups is designed so that the results of discrimination processing within each phoneme group and those from group discrimination processing can be linked in the output layer.

Figure 11 shows a TDNN system with a TDNN capable of distinguishing between the Japanese language vowels /a, i, u, e, o/ and a TDNN capable of distinguishing between the six consonant groups and the vowel group added to the TDNN in Fig. 10, in addition to a speaker-dependent recognition TDNN expanded to operate as a speaker-independent recognition network[9]. As the figure shows, this TDNN is a large-scale network with a 3D structure. Its scale enables speech recognition of all Japanese consonants and vowels in speaker-independent mode. Adding a discrimination unit (Q) for the presence/lack of voice (voiced/unvoiced) makes it possible to automatically recognize Japanese phonemes (called phoneme spotting) simply by scanning vocalized speech along the temporal direction. Readers are referred to reference[9] for more information on the recognition performance of these networks.

Figure 12 outlines a different approach

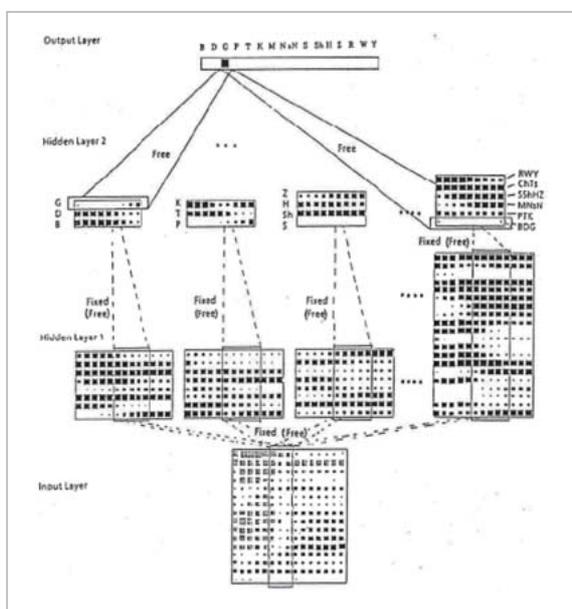


Fig. 10 Modular architecture of all consonant network

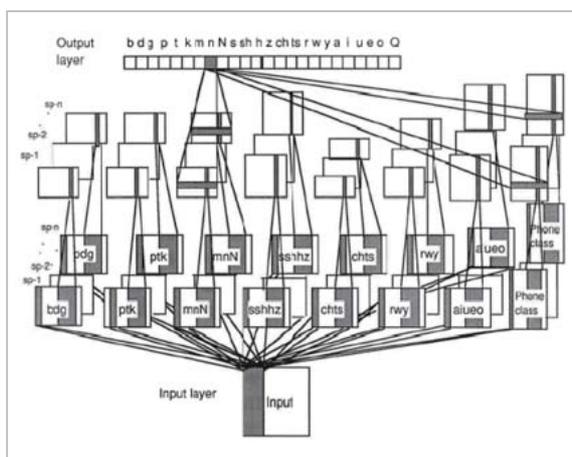


Fig. 11 Large-scale TDNN architecture for speaker-independent recognition

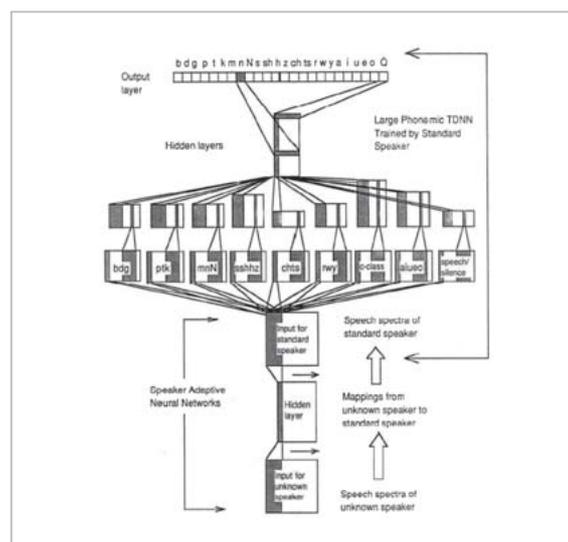


Fig. 12 Large-scale TDNN architecture with speaker-adaptive neural network

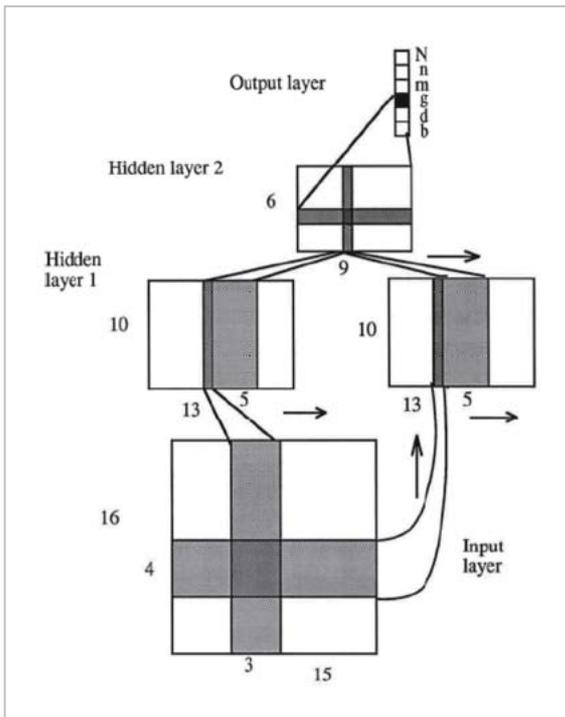


Fig. 13 Frequency-time-shift-invariant TDNN architecture

that also results in a speaker-adaptive neural network[10]. The three lowermost layers are neural networks that perform speech spectra mapping to adapt the speech of an unknown speaker to that of a standard speaker used in training. This approach—providing the mapping to the TDNN in advance—may represent a highly effective strategy for applying a TDNN trained to a standard speaker to recognize the speech of unknown speakers.

2.4 Expansion of Time-Delay Neural Network[11]

Figure 13 shows an expanded architecture model of a new TDNN designed to absorb temporal and frequency fluctuations in speech. Time and frequency windows are installed in the input layer, and extractions of feature quantities (the feature quantity associated with temporal fluctuations and those associated with frequency fluctuations) are integrated in hidden layer 1. The signal is ready for output by the output layer after integration in hidden layer 2. The figure is an expanded TDNN designed to make precise distinctions within a category consisting of six phonemes associat-

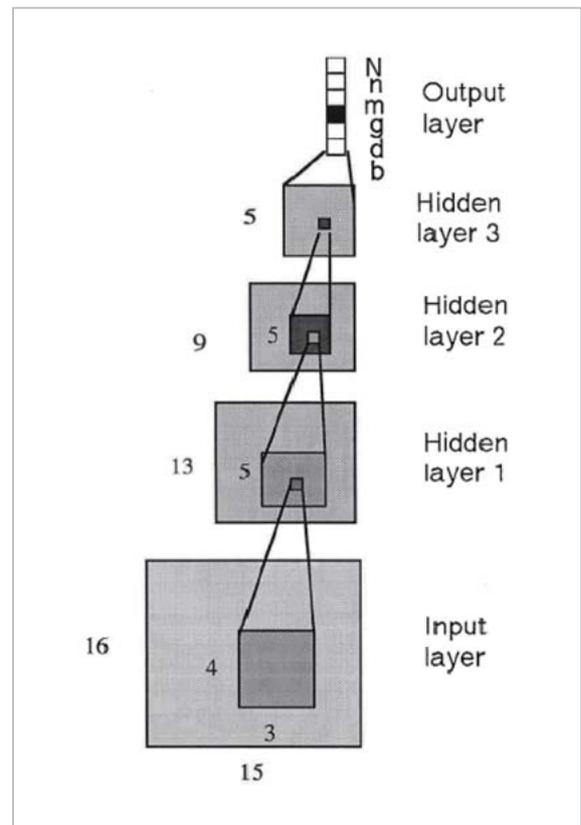


Fig. 14 Block-windowed neural network (BWNN) architecture

ed with high misrecognition rates: the voiced plosives /b, d, g/ and nasals /m, n, N/. The recognition performance of this TDNN is discussed in reference[11].

Figure 14 is a neural network with block windows inspired by the neocognitron, extensively investigated for applications in handwritten letter recognition. As with handwritten letter recognition, it is necessary to absorb fluctuations in the absorption along the time (x-axis) and frequency (y-axis) directions in speech pattern recognition. Thus, we can build an architecture that promotes such absorption by installing block-shaped windows in the lower layers so that the feature quantities can be integrated in succession as connections are made upwards.

2.5 Architecture for Rotation-Invariant Pattern Recognition[12]

The architecture in Fig. 15 is an expanded NN having axial symmetry created by expanding the translation invariance of TDNN to

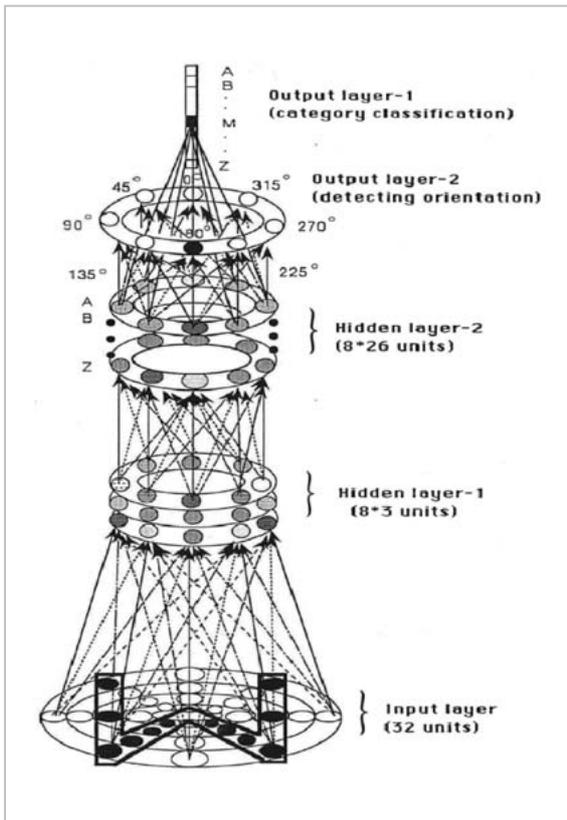


Fig. 15 Axially symmetric neural network architecture

rotational invariance.

The synapse weighting factors having parallel assignments from the bottom to upper layers retain the original values acquired through training. This architecture means that if the synapse weighting factors can be learned using error back-propagation for every pattern of the letter pattern class categories (A–Z) input at a given rotational position (0 deg., for example), the neural network can correctly recognize the class category and rotation angle when an arbitrary class category is input at any angle. The figure shows a neural network architecture capable of recognizing the 26 letters of the alphabet at rotation angle intervals of 45 degrees. From bottom to top are the input layer, hidden layer 1, hidden layer 2, output layer 2 (for rotation angle recognition), and output layer 1 (for class category recognition). This neural network has an axially symmetric structure. Proper execution of the recognition function requires careful alignment of the center of the pattern of the image

fed to the input layer.

3 Information Processing Inspired by Evolutionary Computation

An information processing model inspired by biological evolution is a computational algorithm based on genetic algorithms and evolutionary computation. Table 2 is a compilation of the evolutionary computation algorithms introduced in this and subsequent sections.

In 1859, Charles Darwin released his book, *The Origin of Species*, which dealt with the genetics and evolution of organisms. Prompted by the book, John Holland advanced the idea of genetic algorithms (GA) in the 1960s. Genetic algorithms have been applied to problems in numerous fields, including functional optimization, combinatorial optimization, and parameter optimization in machine design. In recent years, genetic algorithms have been integrated with other techniques such as Evolutionary Strategies (ES) by Rechenberg and Evolutionary Programming (EP) by L. Fogel to form a research discipline known as Evolutionary Computation (EC).

3.1 Parameter-free Genetic Algorithms Inspired by Disparity Theory of Evolution^[13]

This section discusses a new algorithm, the parameter-free genetic algorithm (PfGA), which requires no initial setting of genetic parameters such as initial population size, crossover rate, or mutation rate.

The PfGA was inspired by and builds on the disparity theory of evolution proposed by Furusawa et al.[20], which itself is based on mutations in the double strands of DNA (Fig. 16). According to disparity theory, when the double strands of DNA unwind and a copy of each is created, a difference emerges in the rate of mutation between leading and lagging strands. This is because the direction of replication in the former is the same as the direction of unwinding, while the direction of repli-

Table 2 Some examples of evolutionary computation algorithms

Life phenomenon, theory of evolution (from micro to macro)	Evolutionary computation algorithm
Disparity evolution (Furusawa) [20]	Parameter-free GA [13]
Theory of gene-duplication (Ohno) [16]	Gene duplicating GA [15]
Codon to amino acid translation	Chemical GA [21], GP [22]
Sexual selection (Darwin) [4]	Evolutionary computation based on sexual selection [17]
Ecosystem	Hierarchical parallel-distributed GA [14]

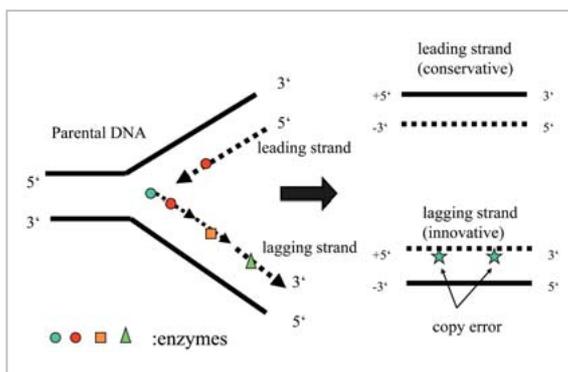


Fig. 16 Hypothesis based on the disparity theory of evolution

cation in the latter is in the opposite direction. While mutations are generally rare in the leading strand (conservative), the lagging strand displays comparatively high mutation rates (innovative). Disparities in replication errors accumulate through crossovers and mutations over generations, creating DNA diversity within a single population via DNA strands experiencing little or no mutation and strands with accumulated mutations. The former type promotes the stability of the population, while the latter promotes flexibility. In the case of PfGA, the former corresponds to the optimized individual at a specific point in time, while the latter corresponds to the offspring produced by crossover and mutation. Thus, under the disparity theory of evolution, the mechanism by which diversity is retained while maintaining a balance between exploitation (localized search) and exploration (global search) can be understood as a balance between genetic stability and flexibility.

In PfGA, the population is regarded as a set of all possible solutions in the whole

search space. In this whole search space S , a local sub-population S' is set. Two individuals are selected from this sub-population S' as parents. The parents are subjected to crossover and mutation to generate a family (S'') of four, with two offspring. The fitness of the four individuals within this family is then evaluated to select or eliminate individuals to evolve local population S' and to execute a search for the solution.

Below are the steps in the basic algorithm for the PfGA (See Fig. 17).

1. An individual is randomly extracted from S and is regarded to be initial local population S' .
2. An individual is randomly extracted from S and added to the local population S' .
3. Two individuals are randomly extracted from local population S' for use as parent 1 (P_1) and parent 2 (P_2) in the multi-point crossover.
4. Of the two individuals generated by the crossover, one is randomly selected and inverse mutation is applied at a random number of points at random positions.
5. Selection and elimination is performed for a total of four individuals (referred to as a family) consisting of the two generated offspring (C_1 and C_2) and the two parents (P_1 and P_2), by selecting either one or three members of the family to be returned to local population S' , based on the calculated fitness.
6. If the local population size $|S'| \geq 2$, then return to Step 3; if $|S'| = 1$, then return to Step 2 and repeat the cycle.

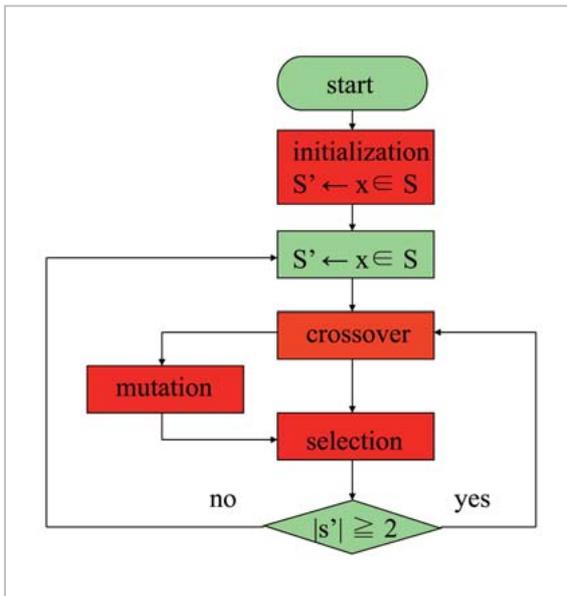


Fig. 17 Flowchart of parameter-free GA (PfGA)

Multi-point crossover is used as the crossover mode in PfGA. In multi-point crossover, both the number and positions of crossover points are determined randomly, and crossover is executed between chromosomes of two different individuals. Mutations use the erroneous copy of chromosomes generated during the crossover. Thus, of the two offspring produced, one is randomly selected and a partial inversion of the gene sequence carried out to create mutations at a random number of points at random positions. Here, of the two offspring produced by the crossover, one is left untouched by mutation to allow one offspring to retain at least a portion of the parents' traits. In this manner, the PfGA is implemented to execute genetic manipulation based on random numbers and to minimize *ad hoc* choices.

For selection and elimination, to retain the diversity of the local population while maintaining a balance between global search (exploration) and local search (exploitation), the fitness of the four members of the family (S'') are compared and selections made in the manner presented in the following four cases as shown in the left plot of Fig. 18. Local population S' is evolved while dynamically maintaining a balance between global and local

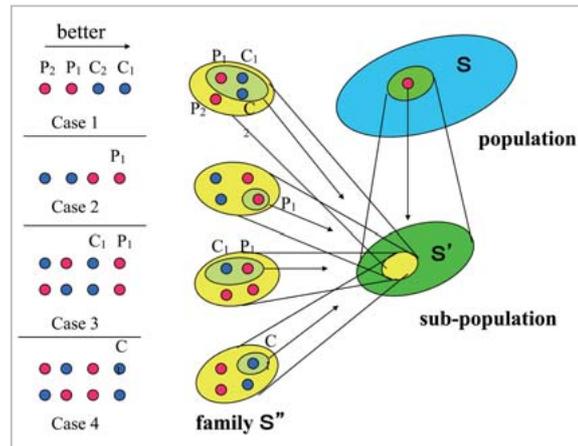


Fig. 18 Population and selection rule in PfGA

search and concurrently changing the size of the local population S' based on an implicit set of rules for switching between Cases 1–4, depending on the relative superiority of the fitness of family members. This feature improved the search efficiency of PfGA over other GAs of fixed population size, since it eliminates the need to perform unproductive searches. In addition, the best individual among the four family members is always returned to the local population S' . Thus, the family may be said to be adopting an elite-preserving strategy: The algorithm guarantees the retention of the best individual at a given point in time while simultaneously performing an active search over a very wide (neighborhood) space. If a better individual than the currently best individual is generated, the center of the search transfers to that offspring; if not, the current best individual is retained. This avoids fitness degradations during the course of evolution.

3.2 Parallel Distributed Processing Techniques for Parameter-free Genetic Algorithms^[14]

This section describes techniques for parallel distributed processing related to the parameter-free genetic algorithm (PfGA) inspired by ecosystems. In general, the main goal of parallel processing in any processing, including GAs, is to increase processing speed. However, we can dramatically enhance the

efficiency of search problem processing based on a GA by introducing interactions between individuals by migration, rather than simply dividing up the task. In the case of a coarse-grained parallel GA, the local population is treated as the processing unit, and individuals are migrated between local populations at appropriate frequencies. In a fine-grained GA, the neighborhood of a given individual is treated as the processing unit, and overlaps are set among neighbors. The former is frequently referred to as the island model, wherein a single local population constitutes the deme of a single species. This paper uses this as the model.

Two types of parallel processing architectures are used: the uniformly-distributed type and the master-slave type. The uniformly-distributed type corresponds to a situation in which all local populations have the same role and local population monitoring functions are absent.

On the other hand, in the master-slave type, a master local population is equipped with a function for monitoring the processing of other local populations, called slaves. Fig. 19 shows the uniformly-distributed-type PfGA architecture. From the whole search space S , an N number of local populations S'_i ($i = 1, \dots, N$) is derived; in each local population S'_i , there exists a family (S''_i) that performs PfGA crossovers, mutations, and selection. Migration of individuals may occur between any of the local populations.

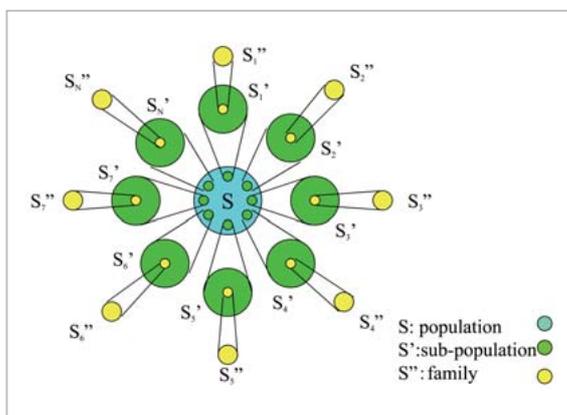


Fig.19 Uniformly-distributed type architecture

Figure 20 shows the master-slave architecture. S'_0 is the master local population, while S'_i ($i = 1, \dots, N$) are the slave local populations. The master S'_0 consistently (or at constant intervals) seeks to identify the best individual in all slave populations.

Several migration strategies may come to mind, but here we adopt the following two. In the first, an individual in a given local population is copied and distributed to other local populations only when a good individual emerges. This is called the direct migration type. The disadvantage of this method is that the same individual is retained by other local populations after migration, threatening system diversity. We adopt the second strategy, in which good individuals are gathered from multiple local populations and two individuals are arbitrarily selected to be the new parents. They bear two offspring (by crossover and mutation), and one to three members of the family are distributed according to the selection rules in PfGA to arbitrarily selected local populations. Since this migration method implements meta-level PfGA operations from the perspective of the local population, it is called the hierarchical migration type (Fig. 21).

We performed parallel processing for the four different combinations of parallel architecture (uniformly-distributed/master-slave) and migration type (direct/hierarchical) to investigate the effects of migration.

An evaluation of search performance

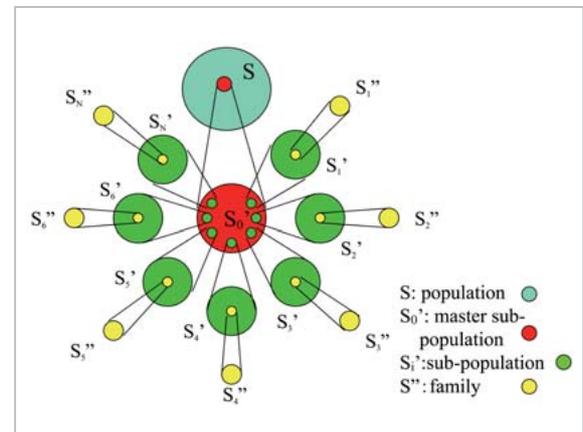


Fig.20 Master-slave type architecture

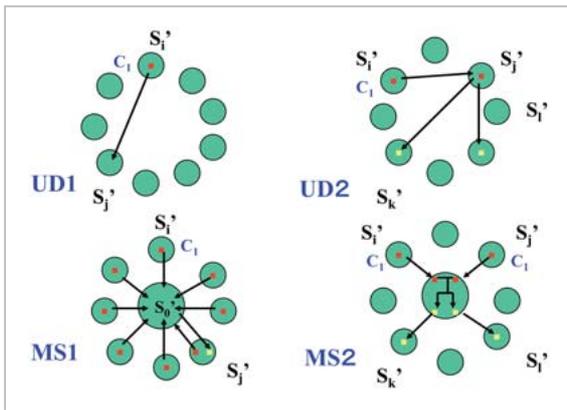


Fig.21 Migration strategy selection method: UD1 (top left), UD2 (top right), MS1 (bottom left), and MS2 (bottom right)

showed that increasing the number of local populations reduced the number of evaluations required before success by a ratio of $1/N$. Among the four types of architecture/migration method examined, search performance, from high to low, fell into the following sequence: UD1, MS1, MS2, UD2. We confirmed that increasing the number of local populations increases the chance of success through the effects of migration relative to serial processing.

3.3 Evolutionary Computation Based on Gene Duplication^[15]

This section discusses the gene-duplicating GA (GDGA) inspired by the theory of gene-duplication proposed by Susumu Ohno in the 1970s. The theory of gene-duplication claims that the replication and reuse of gene fragments in the evolution of all organisms, from viruses and plants to animals, fuels a drive toward life forms of ever-growing sophistication.

Ohno distills this phenomenon into a succinct formula: “a single creation and one hundred plagiarisms.”^[16] Gene duplication is assumed to occur by unequal crossover between chromatids on a single chromosome, unequal crossover between homologous chromosomes during the meiotic process, and partial repetitive duplication of DNA. Inspired by this gene duplication mechanism, we have proposed four gene duplication models: gene

concatenation, gene-prolonging, gene coupling, and extended gene coupling.

This computational method is based on a divide-and-conquer GA in which a given problem is broken down into sub-problems, which are then combined to obtain the solution for the original problem. Each individual concatenates the partial solution each had accumulated up to that point in time, then the individual migrates between the local populations. This strategy makes it possible to obtain the solution more efficiently and quickly.

Gene duplication, a powerful tool in solving multi-dimensional functional optimization problems, is a genetic operator applicable to individuals of different gene lengths. It is applied by first coding variables to genes for each subdimension, then setting the fitness function for each subspace and running the GA to obtain the (quasi) optimum solution. By concatenating individuals owning the gene corresponding to this (quasi) optimum solution, we can solve an optimization problem in higher dimensions. This algorithm is implemented by having individuals with differing gene lengths migrate between local populations. Overall, the algorithm performs crossover, mutation, and selection within a local population; between separate local populations, it performs duplication and migration, in that sequence.

In a simulation evaluating the search performance of the four types, we used a functional optimization benchmark problem to compare success rates, probabilities, and convergence performance in obtaining the optimum solution. We found that increasing the number of migrating individuals increases population diversity, thereby confirming improvements in convergence performance and the effectiveness of this computation method.

3.4 Evolutionary Computation Inspired by Sexual Selection^[17]

The theory of sexual selection seeks in part to explain the extensive differences between the phenotype and behavior of the two sexes in certain sexually-reproducing

organisms. Certain well-known examples include competition between males (generating antlers) and female preferences (generating peacock plumage).^{[18][19]} Numerous hypotheses based on female preference seek to explain the evolution of traits that appear disadvantageous from the perspective of natural selection.

Two famous hypotheses are the runaway hypothesis and the excellent gene hypothesis. The former assumes that female preferences within a population are always biased and that male traits are always variable due to random mutations. In such cases, males with the preferred traits are more likely to father large numbers of offspring, regardless of fitness in terms of natural selection, thereby conferring an indirect advantage to the gene manifesting that trait and resulting in the rise of the male trait preferred by females. The latter hypothesis claims that the male shows off the quality of his genes even if doing so comes at some cost, with the result that the trait and the female preference for that trait grow more common. However, the actual processes at work in sexual selection remain unclear.

The genetic assignment of sex to individuals has led to a state in which an individual of one of the sexes effectively observes and chooses the phenotype of the other, and we will focus on the effect of such asymmetric roles of males and females on the process of evolution.

We will also examine the role of mutation (the simplest transition rule) as a genetic operator that drives organisms toward the direction of evolution based on the fitness landscape. A model is assumed in which a mutation rate is coded in the gene as a parameter for mutation and in which fitness varies.

We will focus in particular on the interaction between sexual selection and mutation.

Working from these perspectives, we introduce an evolutionary computational model in which mutation rates are encoded in the gene and also account for sex and sexual selection. Based on this model, we investigate how mutation rates become self-adaptive

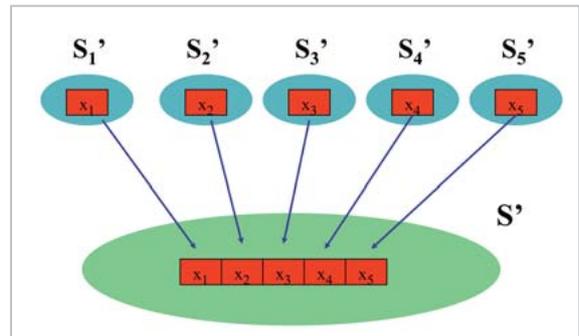


Fig.22 Gene duplication in a gene-concatenating model

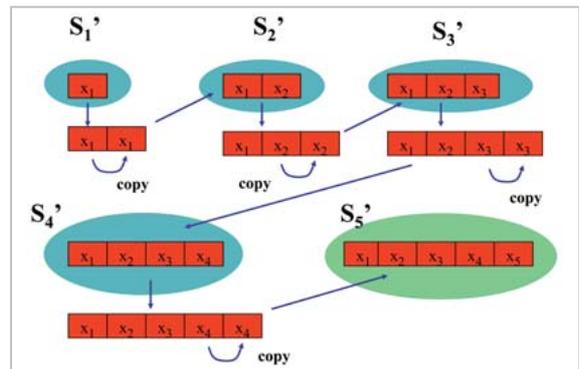


Fig.23 Gene duplication in a gene-prolonging model

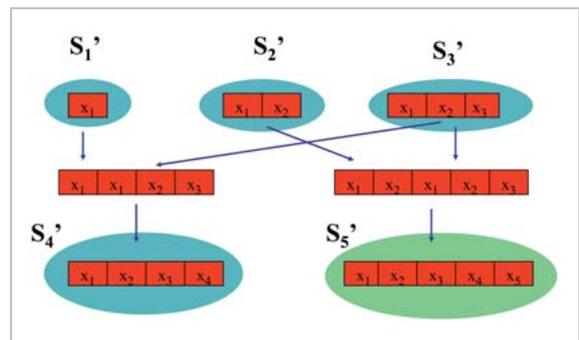


Fig.24 Gene duplication in a gene-coupling model

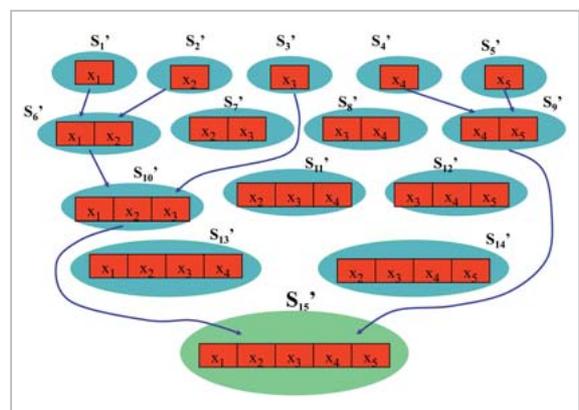


Fig.25 Gene duplication in an extended gene-coupling model

within a population and how the direction of evolution is determined in phenotypic space.
[Constructing a computational model]

How does an asymmetric relationship between the sexes whereby one sex observes and selects the phenotype of the other affect mutation rates? We propose a model based on sexual reproduction with its own mutation rate encoded into the genes. For the sake of convenience, the observing and observed sex, respectively, are regarded as female and male.

A real-valued genetic algorithm (real-valued GA) is used as the evolutionary computational model. The methods of genetic recombination are chromosomal exchange between individuals and isotropic mutation of each gene.

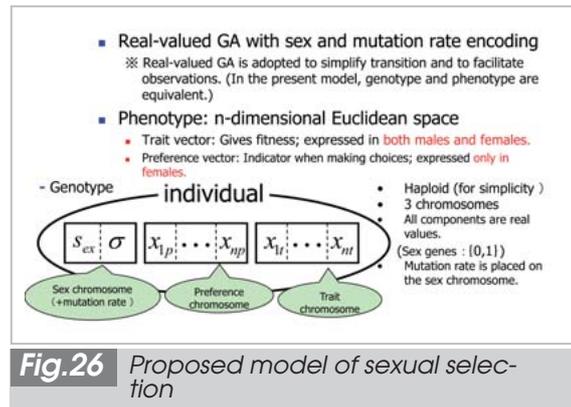
Sexual selection will focus on relative phenotypic value. (Example: a strong preference for taller individuals or individuals of a certain stronger coloring [e.g., bluer].) In such modeling schemes, the direction of the transition of next-generation males in phenotypic space is determined by the direction of female preferences.

[Individual phenotype]

Each individual is assigned a sex and with two broad phenotypic categories, trait and preference. Traits are expressed in both sexes and determine the individual's fitness. In contrast, preferences are expressed only by females and act as a mechanism by which males are assessed and selected. Preferences do not affect natural selection. Traits and preferences are represented in phenotypic space by trait vector $t = (x_t^1, x_t^2, \dots, x_t^n)$ and preference vector $p = (x_p^1, x_p^2, \dots, x_p^n)$ in n-dimensional Euclidean space.

Sexually-reproducing organisms are diploid in nature, but here, given the emphasis on interactions due to preference rather than the mode of reproduction itself, the model assumes that the genotype of each individual is haploid for the sake of simplicity, and an individual will have three kinds of chromosomes: sex chromosome, preference chromosome, and a trait chromosome (Fig. 26).

The sex gene is coded in a single bit and



the other genes are encoded as real values. However, the preference gene in the male is considered not expressed, creating a buffer for male preference and sustaining diversity in preference. Furthermore, we place the gene encoding for the mutation rate (σ) for the trait and preference genes on the sex chromosome, making the expressed mutation rate dependent on sex.

[Natural selection]

A known hummingbird species displays sexual dimorphism in beak morphology, with the males and females of the species feeding on flowers of different shapes. This constitutes habitat segregation in the form of resource partitioning. The male and female members of this species can be considered to have followed different paths in natural selection.

Natural selection is posited to operate separately on the sexes. This renders a constant sex ratio while allowing the sexes to generate different traits (sex difference).

[Sexual selection]

Under sexual selection, a female chooses a male to form a pair, and the offspring produced are one male and one female to maintain the constant sex ratio. All females will always be part of a pair at least once, while males are allowed to pair only when selected by a female as a preferred male. In short, this population is polygamous.

- (1) The female i observes M numbers of males at random as potential mates. The average trait vector t_i^0 of the observed male population is calculated, and the relative trait vector $t_{ij} = t_j - t_i^0$ of male j ($j = 1, 2,$

..., M) is used as the selection target.

- (2) The difference in direction θ_{ij} between t_{ij} and the female preference vector p_i is calculated. The strength of preference is defined as given by $\cos(\theta_{ij})$. The male having a trait vector in the direction closer to the direction of the female preference vector will be strongly favored.
- (3) The most preferred male is selected deterministically.

By using the relative trait vector as the selection target, we can search for the direction toward which the male population should shift in phenotypic space—or the direction of evolution—based on the direction of the female preference vector.

In sexual selection, males preferred by more females (i.e., attractive males) gain the advantage, and selection works directly on the male, manifesting as the number of offspring in the next generation. While females in this model are exempt from the direct operation of sexual selection, through genetic exchange with preferred males, the offspring of females having a genetic preference for males with advantages in terms of natural and sexual selection is likely to have the advantage in the next generation. This means sexual selection works indirectly to the advantage of females having genes for such preferences.

[Genetic Manipulation]

Two types of chromosomal exchange and mutation are used as methods of genetic manipulation. With respect to interactions between mutation rate and sexual selection, parent chromosomes are exchanged at a constant probability when parent genes are copied to produce offspring.

Mutations are accomplished by imparting perturbations that follow the normal distribution function $N(0, \sigma)$ on the offspring trait and preference genes x^i, x^p ($i = 1, 2, \dots, n$), respectively.

Here, the standard deviation σ corresponds to the mutation rate, which is varied adaptively by gene encoding.

$$\sigma = \sigma + \delta \quad \delta \sim N(0, \sigma_0)$$

Note here that the standard deviation σ_0 of the normal distribution function of the perturbation imparted to σ is constant.

[Steps in simulation]

- (1) Population is initialized at a sex ratio of 1:1.
- (2) The following procedure is repeated until the termination condition is satisfied:
 - (a) Natural selection
 - (b) Sexual selection
 - (c) Genetic manipulation

[Problem]

Traits and preferences, respectively, are represented using two-dimensional vectors $t = (x_t, y_t)$ and $p = (x_p, y_p)$ in examining the following functional maximization problem.

$$Max f(t) = x_t - 2 \sin(\pi x_t) - 0.001 y_t^2 \exp(x_t)$$

The initial population is positioned near the point of origin, while each individual's fitness is determined by trait alone. In the above equation, local optimum solutions may be found on the line $y_t = 0$ with period 2. The problem, then, is to find a way to balance the search for the local optimum solution and the escape from the local optimum solution.

The effects of the 3rd term become large with increasing x_t as the search progresses, and a slight transition of y_t from 0 results in a precipitous decline in fitness, rendering the search more difficult. Although there is no upper limit to the function, GA restricted to isotropic mutation alone places a practical

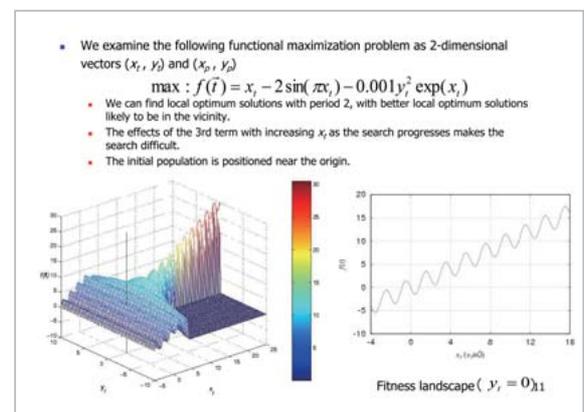


Fig.27 Validation of the proposed model with an illustrative problem

limit on the search. This extremely simple model is an example of how adaptive acquisition of a direction advantageous to evolution (positive x_i direction) concurrent with a search for the optimum solution makes it possible to search as efficiently as possible.

Natural selection is executed by performing a roulette-wheel selection based on g , where $g = \exp(\alpha f(t))$ (α : scaling rate) when $f(t)$ is fitness. This is selected to make the search progress more smoothly when the exponential term in the above equation starts to have a strong effect on the latter part of the search and when the average fitness of the population has decreased exponentially.

We use preference vector $p = (x_p, y_p)$ as the unit vector to normalize the results after each mutation. This means that the preference selection of a male by a female is based solely on the direction of the trait.

[Results of experiment]

As we can see from Fig. 28, the search in the proposed method progressed the most among the three types of strategies (random mating, SGA, proposed method), indicating the effects of sexual selection on the search. We see no apparent differences in mutation rates between the sexes for the results of random mating. Under the proposed method, as the search proceeded, the male mutation rate surpassed the mutation rate for females, confirming that female preference triggers unequal mutation rates.

[Search process]

Under the present method, when the search process becomes trapped in a local optimum solution, a runaway situation between the male trait and female preference results that intermittently drives explosive evolution out of the equilibrium state. As the search progresses, differences appear in average mutation rates between male and female populations and a division of roles arises. The sex exercising choice (female) carries out a conservative search with low mutation rates and the chosen sex (male) carries out an innovative search. This results in only the male population performing the search when escap-

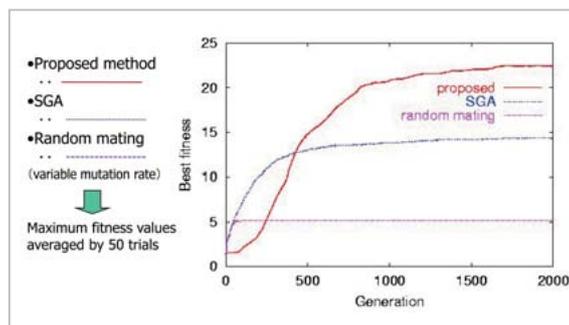


Fig.28 Experimental results

ing from the local optimum solution, while females dedicate themselves to maintaining present conditions and standing by to move on to better solutions through chromosomal exchange only after they have been found by the male population. In many sexually-reproducing organisms, the production processes differ for reproductive cells between the sexes, and sperm cells have higher mutation rates than ova. The similarity between the characteristics of organisms and the proposed method is quite interesting. The advantages of balancing an innovative and conservative search during the search process through the acquisition of various mutation rates have been discussed in relation to the Neo-Darwinian algorithm by Wada et al.[20]. Since the error frequency varies between the two strands in DNA, they proposed a model in which the mutation rate varies within a single individual, thereby permitting a wide range of searches. This realizes a broad range of mutation rates within the population, allowing them to perform an extensive search in problems involving a high level of risk while maintaining present conditions.

4 Information Processing Based on the Modeling of Cells in Early Stages of Evolution

4.1 Chemical Genetic Algorithm (CGA)[21]

The mechanism of cell metabolism emerged over an astoundingly lengthy evolutionary process. Modeling this process should make it possible to search efficiently for an

optimum solution by techniques totally different from conventional methods. This section discusses the chemical genetic algorithm (CGA), used for solving difficult problems by dynamically converting them into simpler problems—in short, by dynamically changing mapping from the genotype to the phenotype, as inspired by the mechanism of cell metabolism in the early stages of the evolutionary process. The section will also discuss chemical genetic programming (CGP), in which this solution method is applied to the evolution of programming, and introduce the problem of symbolic regression in artificial intelligence and describe the results of its application to the acquisition of multi-agent behavioral strategy.

[Mechanisms of cell metabolism generated in the early stages of the evolutionary process]

In the early stages of cell evolution, cells are believed to have acquired their present metabolic processes by dynamically changing the mapping behavior from genotype to phenotype mapping (Fig. 29). Fig. 30 presents a model based on this mechanism.

[CGA generation cycle] (See Fig. 31)

The steps in the CGA generation cycle are given below:

1. Initialization: First, we prepare a number, N , of cells having the structure presented in Fig. 30. In the initial state, no cell possesses aminoacyl tRNA (aa-tRNA), tRNA, or outputs amino acids. However, they do have random DNA strands and amino values.
2. Chemical reaction: The following 4-step reaction takes place in all cells: transcription, tRNA-amino acid reaction, translation into internal amino acid, and translation into output amino acid. In the several generations of the early stage, this reaction produces new tRNA and aa-tRNA, and their sizes grow. Within the next few generations, we exceed the size of the molecular pool size.
3. Selection: The fitness of the cell is calculated based on the output amino acid,

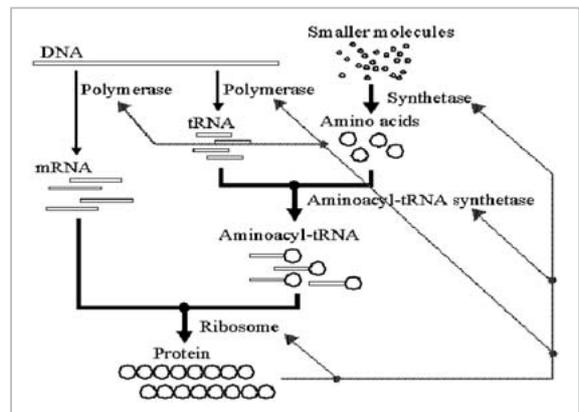


Fig.29 Biochemical reactions for translating genetic information in a cell

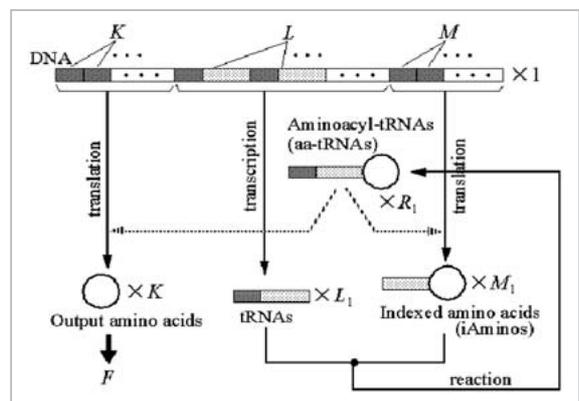


Fig.30 Cell structure used in CGA

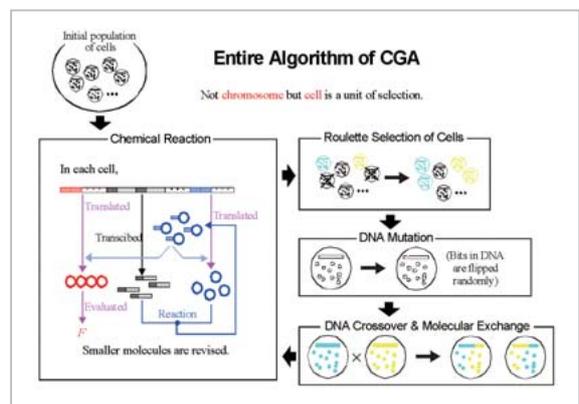


Fig.31 Entire algorithm of CGA

and cells marking high fitness are selected by roulette-wheel selection. The selected cells are regenerated, and the complete internal information (DNA, 3 molecular pools) of each cell is copied to the daughter cell.

4. DNA mutation: As in normal GA, point mutations of genes are performed.

5. DNA crossover and molecular exchange between cells: Gene crossovers occur as in normal GA, and half the molecules are exchanged between two cells.
6. Calculation of fitness of the cell population. If the termination conditions are satisfied, the computation is complete. If not, return to Step 2.

[GA evolvability] (See Fig. 32)

By converting the “ragged fitness landscape” as seen in the genotype space presented on the left side of Fig. 32 (a) into a smoothed landscape on the right side (b), we can improve evolvability.

Three types of deception problems (Types I, II, and III) and 2 benchmark functions were used to validate the search performance of CGA. SGA (simple GA) and PfGA (Parameter-free GA) are used for comparisons. In the simple Type I, $F(x)$ assumes the maximum value (optimum value) of 1 when $x_k = 1$ in all dimensions of k ($k = 1, \dots, n$). Type II is an intermediate complex type, and $F(x)$ assumes the maximum value of 1 only when, in each dimension k , $f_k(x)$ randomly takes a maximum at $x_k = 0$ or $x_k = 1$. In Type III, $F(x)$ assumes the maximum value of 1 only when $f_k(x)$ assumes a maximum of 1 at $x_k = \alpha_k$ (α_k is a uniformly random number between 0 and 1) in each dimension of k . As can be seen from Fig. 33, the ratio of the probability of $f(x)$ assuming a maximum (optimum value) of 1 in each dimension to the probability of taking the localized optimum value of 0.8 is 1:4. Thus, the probability of $f(x)$ taking optimum

values at all dimensions ($k = 1, \dots, n$) is $(1/5)^n$. Types I and II constitute special cases of Type III (Fig. 33).

[Results of analysis]

Figures 34 and 35 present the evolution of the function $F(x)$ of CGA and SGA, respec-

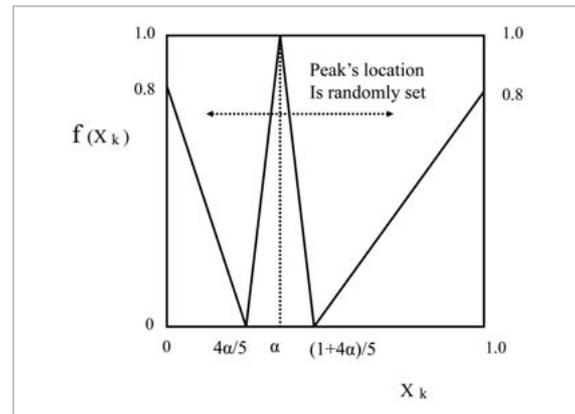


Fig.33 Complex deceptive problem (Type III)

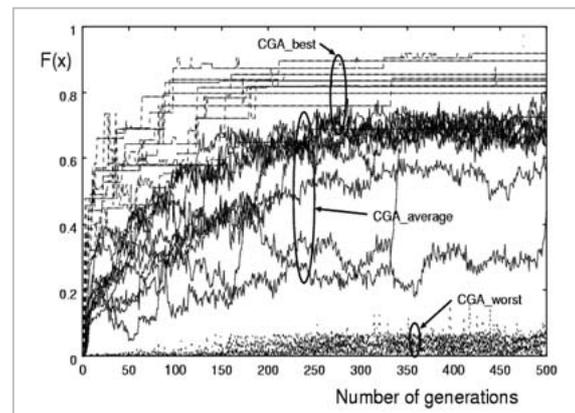


Fig.34 Evolution of CGA for deceptive problem (Type III)

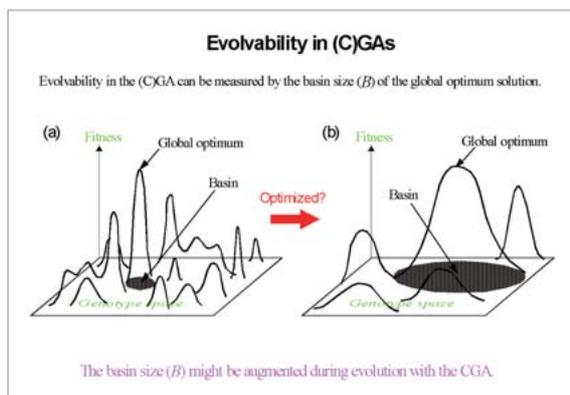


Fig.32 Evolvability in C(GA)

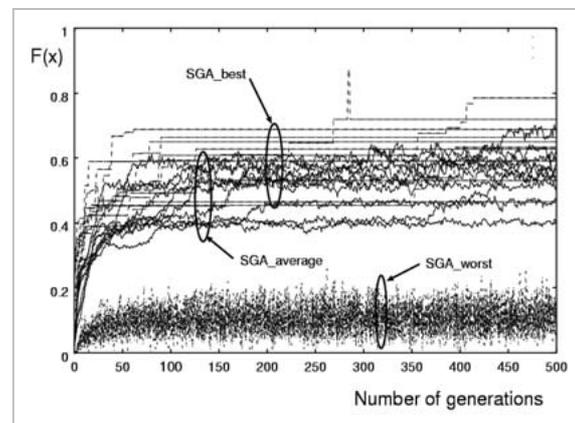


Fig.35 Evolution of SGA for deceptive problem (Type III)

tively. The dispersion of the function values is large for CGA. In contrast, the dispersion is small for SGA. The optimum value of CGA (CGA best) is 0.8 or higher and attains the status of the optimum solution. On the other hand, the optimum value of SGA fails to surpass 0.8 and is trapped in a local optimum solution.

Figures 36 and 37 present the time series of amino value histograms for CGA and SGA, respectively. The results are for a 5-dimensional Type-III deception problem. We see that for all α_k assuming maximum values at each dimension of k , the amino value for CGA exceeds a certain value. In contrast, for SGA, the amino value exceeds a certain value in certain dimensions but not in others. This is a

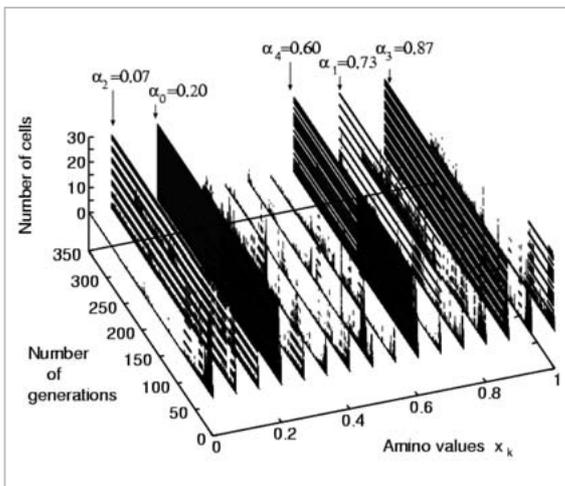


Fig.36 Time series of amino value histogram for CGA

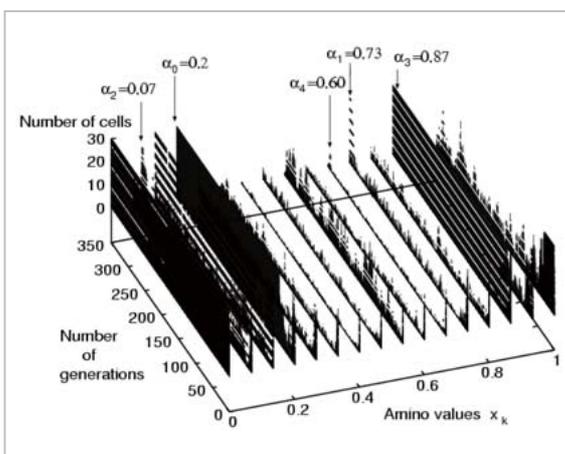


Fig.37 Time series of amino value histogram for SGA

typical example of a search that has fallen into a local optimum solution.

[Performance comparison]

Tables 3 and 4 present comparisons of performance evaluation results for SGA, CGA, and PfGA.

Table 3 gives the results for deception problems. A comparison of the success rates of CGA and SGA shows that the performance of CGA far outpaces SGA. Furthermore, PfGA is 100% successful for all types of deception problems. Table 4 gives the results for benchmark problems (Shekel's foxhole problem, Langerman function), and we see that CGA gives success rates comparable to PfGA.

Figure 38 shows the changes in basin size in the case of CGA. The figure shows a sudden increase in basin size at a certain point in the evolution (100 generations), considered to reflect the achievement of punctuated equilibrium associated with the transition stage in the evolution, mapping from genotype (binary value) to phenotype (function value).

Table 3 Success rate for deceptive problems

GA	SGA	SGA*	CGA	PfGA	SGA	SGA*	CGA	PfGA
Dimension	5	5	5	5	10	10	10	10
Codon length	6	6	6	20	6	6	6	20
scaling	linear	linear	exp.	none	linear	linear	exp.	none
Type I	2%	9%	100%	100%	0%	3%	100%	100%
Type II	12%	20%	100%	100%	0%	5%	100%	100%
Type III	47%	85%	95%	100%	14%	79%	43%	100%

Table 4 Success rate for benchmark problems

GA	SGA	SGA*	CGA (WF)	CGA (NF)	PfGA	SGA	SGA*	CGA (WF)	CGA (NF)	PfGA
Dimension	5	5	5	5	5	10	10	10	10	10
Codon Length	6	6	6	6	20	6	6	6	6	20
scaling	linear	linear	exp.	exp.	none	linear	linear	exp.	exp.	none
Shekel	5%	5%	5%	50%	37%	0%	0%	0%	0%	1.3%
Langerman	41%	47%	13%	35%	83%	0%	0%	0%	3%	1.7%

Figure 39 presents the values of the codon-amino acid translation table. Under the initial conditions of evolution, variations in the values in the table appear to be large changes in amino value per single bit of change in the codon (lower right). But as evolution progresses, the amino value changes only gradually relative to the change in codon value. This corresponds to the evolution (transition) of the landscape represented by the left plot of Fig. 32 into the smooth landscape represented by the right plot. The evolvability of CGA has greatly increased, indicating that the algorithm generates a solution method for difficult problems in an evolutionary manner, while automatically (in evolutionary fashion) converting difficult problems into easier problems. The method for improving mapping techniques from genotype to phenotype through the evolutionary process is a highly generalizable optimization technique. The following

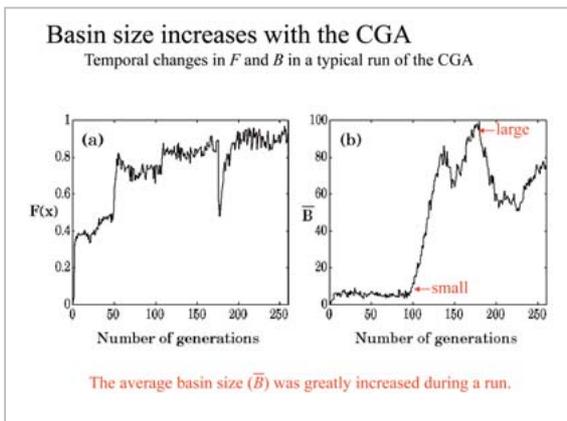


Fig.38 Increase in basin size for CGA

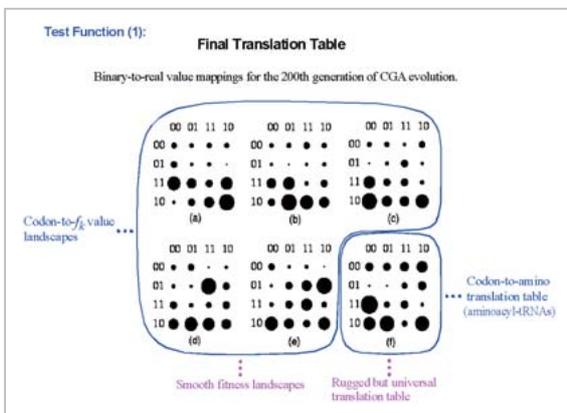


Fig.39 Final translation table in binary-to-real value mapping for CGA

section discusses a method for expanding CGA to genetic programming (CGP).

4.2 Chemical Genetic Programming (CGP)^[22]

Figure 40 shows how CGA is expanded to CGP. A comparison of Fig. 31 and Fig. 40 shows how the genes (DNA sequence) in CGA are converted into a combination of the rewriting rule numbers and the left sides of the rewriting rules. DNA is translated into protein synthesized by concatenating amino acids, after which fitness is calculated. At this point, the other portions of the DNA have been transcribed into tRNA or translated into amino acids. Aminoacyl tRNA, produced by their reaction, acts as a catalyst in this translation process. Modeling these metabolic processes inside the cell results in the evolutionary generation of the rewriting rule itself, and generates rules completely different from those in the initial state to make it possible to acquire rules with higher fitness scores.

[Example of application 1] (Symbolic regression problems)

Figure 41 compares the fitness evolution curve in CGP and conventional GE (grammatical evolution). We see that evolution proceeds faster in CGP than GE, resulting in a good solution after 140 generations. Furthermore, even though the best solution appears faster, the average fitness of the population consistently remains below the best value, indicating that population diversity is sustained.

Figure 42 compares the solution generated by CGP and GE. For the target function $2x^6 + 3x^4 + 4x^2 + 100$, CGP gives $2x^6 + 501$,

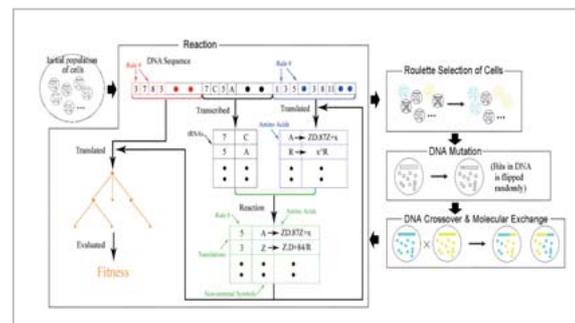


Fig.40 CGP algorithm

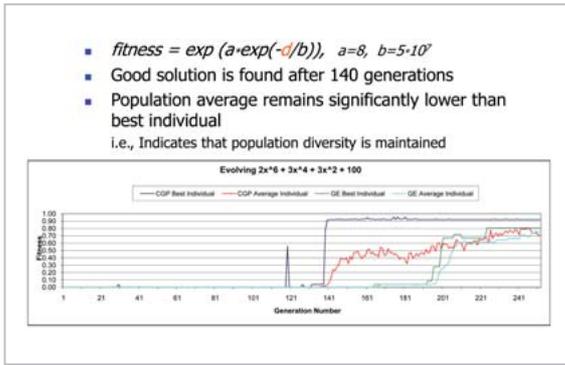


Fig.41 Evolution curve of fitness

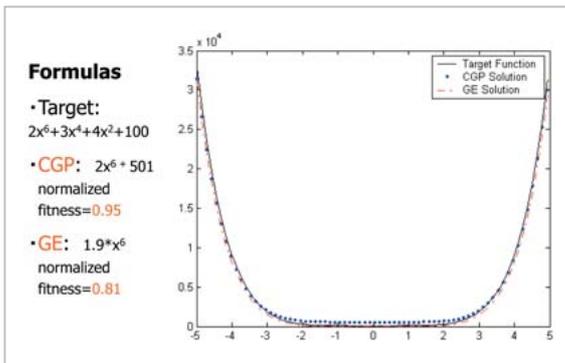


Fig.42 Generated function CGP vs. GE

while GE gives $1.9x^6$. Since the normalized fitness values are 0.95 and 0.81 for CGP and GE, respectively, we may conclude that CGP is superior.

[Example of application 2] (Behavioral strategy of agents)[23]

Figure 43 applies CGP to a multi (2)-agent problem, the game of “tag.” Two agents use CGP to generate in evolutionary fashion a behavioral strategy for catching the other as quickly as possible, or to evade the other for as long as possible.

Table 5 is a list of basic functions used in CGP.

Figure 44 shows the two agents in motion. S denotes the starting point. The objects numbered 1 and 2 in the center are obstacles. We see how the two agents (pursuer and evader) skillfully avoid the obstacle in the chase. Table 6 presents the behavioral strategies generated by the agents. Although the details are not given here, we can appreciate how both agents acquire various behavioral strategies through evolutionary generation.

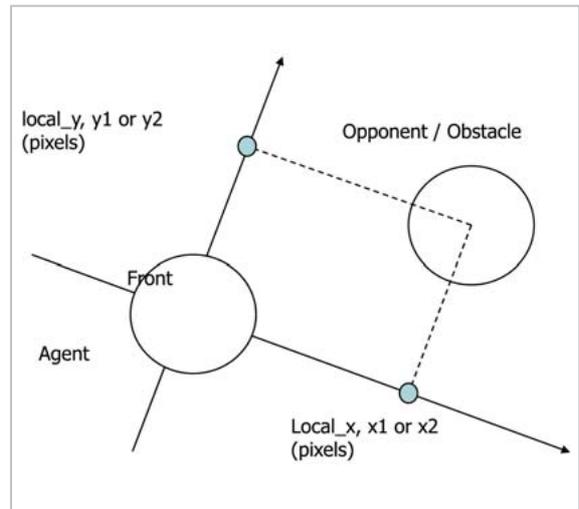


Fig.43 The game of tag

Table 5 List of basic functions

Function	Usage	Description
+	(+ a b)	a plus b
-	(- a b)	a minus b
*	(* a b)	a time b
%	(% a b)	if b = 0 then 1 else a divided by b
min	(min a b)	If a < b then a else b
max	(max a b)	If a > b then a else b
abs	(abs a)	absolute value of a
neg	(neg a)	negative value of a
iflte	(iflte a b c d)	If a <= b then c else d

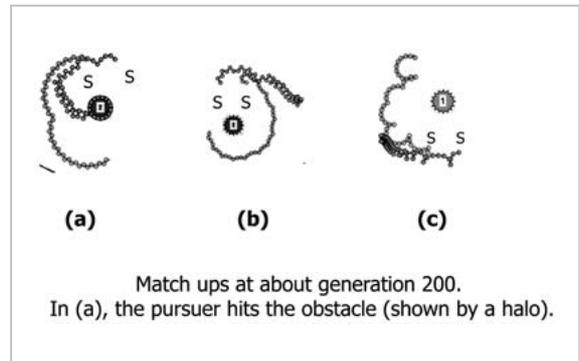


Fig.44 Results of agent behavior

Table 6 Generated strategies

Emergent Strategy	Generation	Player
Tagging	200 – 450	Pursuer
Weaving	200 – 450	Pursuer
Safe Haven	200 – 450	Evader
Obstacle Hugging	200 – 450	Evader
Boundary Diversion	~ 450	Evader
Focused Tagging	~ 450	Pursuer
Point Circling	~ 500	Evader
Arcing	~ 1000	Evader
Shorting	~ 1000	Pursuer

5 Research Trends in Complex Networks

“Man is, at one and the same time, a solitary being and a social being.”

—Albert Einstein—

The information and communication systems permeating today’s society grow in sophistication and complexity seemingly day by day, while assuming greater roles in our lives at the same apparent pace. But the same information and communication systems have been shown to be fragile under conditions such as accidents and disasters. Developing technologies that will overcome such fragility has become essential. To resolve these problems, by focusing on network dynamics, we examined basic technologies for securing self-organized functional network structures that do not rely on an infrastructure and remain operational even in dynamically changing environments. Our goal is to propose a proto-model for a next-generation information and communications system that is highly reliable and has high affinity for human needs and modes of interaction. Here, we examined research trends in complex network sciences and statistical physics, with a special focus on networks with self-restoring functions and requiring only localized information for adaptive evolution, and on the design principles of self-organizing networks that feature node deployment and load distribution functions according to population distribution and economic activity levels and network traffic management functions designed to resolve traffic congestion. In addition, we also examined the spontaneous social network structures created by relationships between real people and considered a network society desirable in the near future.

5.1 Keywords

- Small-world phenomenon (Fig. 45): a network phenomenon emerging from the rewiring of a few edges in a regular network; one in which the average path length (described later) is significantly reduced.

This is known in common parlance as “six-degrees of separation.” Observations of the phenomenon date from experiments performed in the U.S. by the social psychologist Stanley Milgram. From left to right in Fig. 45 are a regular network, a small-world network, and a random network. ρ represents network rewiring probability; higher ρ values indicate higher randomness. The figure shows how randomness increases from regular networks to small-world networks and finally to random networks.

- Scale-free network (Fig. 46): a network whose relationship between the degree (k_i) of a node i in the network and degree distribution $P(k_i)$ follows the power law $P(k_i) \propto k_i^{-\gamma}$. Figure 46 shows the Internet in a scale-free condition, with relatively small numbers of hubs having nodes with large degrees, and with numerous nodes having small degrees. The upper right-hand panel in the figure shows the degree and degree distribution, displaying what is known as the long-tail phenomenon (resembling the long tail of a

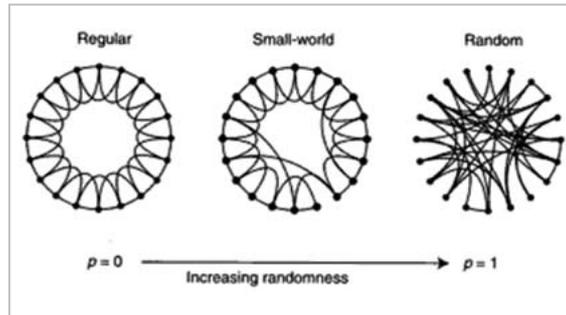


Fig.45 Small-world network [26]

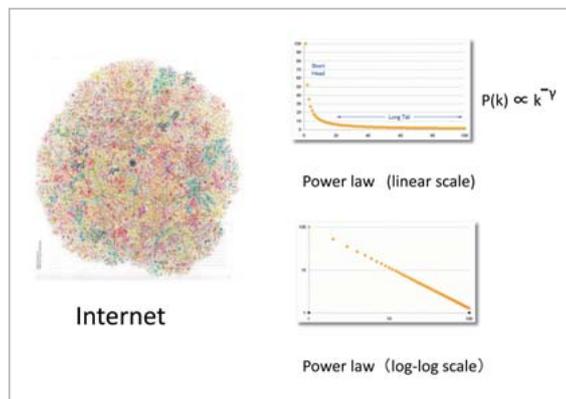


Fig.46 Scale-free network (for Internet)

dinosaur). The lower right-hand panel shows the same data as the upper right-hand panel, but in a log-log plot. The inclination of the linear line here is equal to $-\gamma$. Various scale-free networks in nature and society have their own unique values.

- Pareto's law: A law proposed by the Italian economist Vilfredo Pareto, popularly known as the "80:20 rule." It states that the significant 20% of the element dominates 80% of the whole. It is deeply associated with the relationship between degree k_i and degree distribution $P(k_i)$ in the above scale-free network.
- Robustness: Adaptability and/or fault tolerance of life or systems to environmental change.
- Fragility, vulnerability: The opposite of robustness; the fragility of life or systems (weakness) in the face of environmental change.

In *Persistent Life* by Hiroaki Kitano and Kaoru Takeuchi (Diamond Inc., 2007), the trade-off between robustness and fragility is explained by taking as examples the 2003 New York City blackout, the beef bowl business strategy of Yoshinoya, and diseases such as diabetes and cancer.[40]

5.2 Network Analysis Indicators

- Network G: Set $G = \{V, E\}$ consisting of the set of nodes $V = \{v_1, v_2, \dots, v_n\}$ and set of edges $E = \{e_1, e_2, \dots, e_m\}$.
- Degree: Degree k_i of node v_i refers to the number of edges branching out of node v_i .
- Average path length/average distance between nodes $L(\rho)$: For two arbitrarily chosen points v_i and v_j , the average of the minimum number of edges that must be crossed to reach each other.
- Clustering coefficient $C(\rho)$: An index that takes a value between 0 and 1, which quantifies the degree of concentration of edges constituting network G.
- Network centrality: The network property that plays the central role. When nodes having higher degrees are more central, the network is assumed to have "degree centrality."

The nodes clustered in the highest concentration along the shortest path between two node pairs are described as having "betweenness centrality." The index was proposed by the sociologist Linton Freeman as an index for social network analysis.

- Network density: The higher the number of edges between nodes, the higher the network density.

Figure 47 is a graph showing the relationship between the clustering coefficient $C(\rho)$ and average path length $L(\rho)$ (where ρ is the network rewiring rate). For small values of ρ , both the clustering coefficient $C(\rho)$ and the average path length $L(\rho)$ are large. But as ρ increases, they decrease.

At intermediate values of ρ (around 0.01), $C(\rho)$ is large while $L(\rho)$ becomes small, at which point the small-world phenomenon emerges.

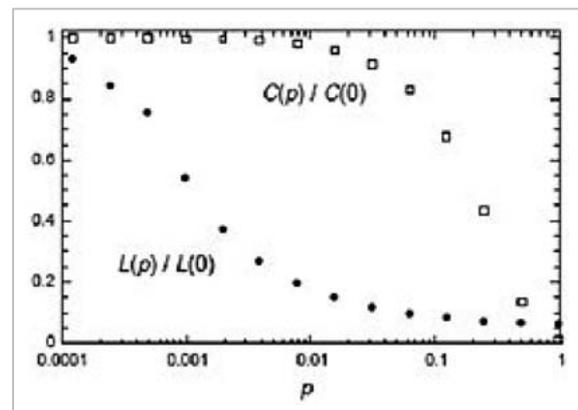


Fig.47 Clustering coefficient $C(\rho)$ and average path length $L(\rho)$

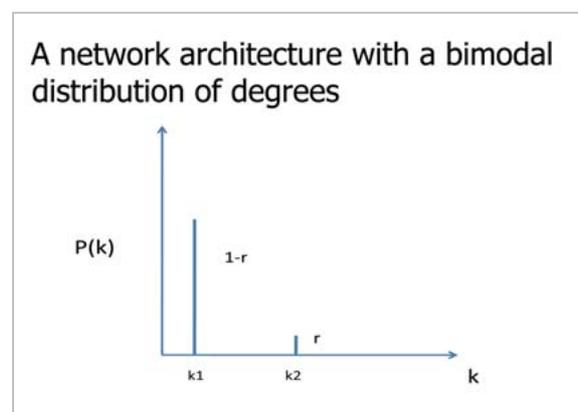


Fig.48 Robust network architecture resistant to both failure and attack

[A network architecture resistant to both random failure and targeted attack]

Although the question, “Do such networks actually exist?” comes to mind, according to “What Kind of Strong Wiring Design Would be Resistant to Both Failure and Attack?” in “Semi-Special Edition: Spread of Complex Network Science” by Toshihiro Tanizawa (pp. 282-289, Information Processing, March 2008 issue, No. 3, Vol. 49), the answer is a bimodal distribution network architecture in which only two types of nodes are present: a small number of large-degree (k_2) hubs and large numbers of small-degree (k_1 , where $k_1 \ll k_2$) nodes.

5.3 General Trends in Research on Network Science

- In *Network Science* by the U.S. National Research Council (<http://www.nap.edu/catalog/11516.html>), the research topics listed in Table 7 are reported to be priority areas.

Based on this table, we find that modeling the simulation of very large networks is considered critical in the development of real-world tool sets and that human performance enhancement is the goal of studies of swarming behavior in the development of self-organizing systems and studies on metabolic and gene expression networks. The selection of biological mechanisms and human swarm behavior as research areas for achieving these research goals is noteworthy.

5.4 Research Trends in Social Network-Associated Fields^{[41]-[46]}

[Organization network strategy]

- According to *Global Neighborhoods — Strategies of Successful Organizational Networks* by Toshio Nishiguchi, (NTT Publishing Co., Ltd. 2008), a strong and effective human network may be constructed by combining long-distance relationships (= use of small-world by distant but dear acquaintances) and neighborhood socializing (= rewiring of human networks). Examples given in the book include the robust human network created overseas by former residents

of Wenzhou, China, and the miraculous recovery of Aisin Seiki Co., Ltd., an affiliate subsidiary of Toyota, after a fire (self-organizing human network when faced with emergency), as well as the examples of defense procurement by the British Ministry of Defense and by Japan. The book presents actual examples of several interesting network topologies. The book concludes by summarizing such networks as follows: The secret to successfully managing a complex social network that far surpasses human cognitive limits is to establish a neighborhood social network and to weave into it long-distance relationships by some appropriate rewiring at the edges to transform the entire system into a small-world, and to take the best parts of both. Supporting this network is the social capital that forms the foundation of the relationship of mutual trust and ease with which structural rewiring can be performed.

- *The Science of Building a Network of Human Connections — Probing the Power Hidden in Human-Human Relationships* by Yuki Yasuda (Nikkei Inc., 2004)

“Distant people” are more beneficial. The quality of the links is more important than the number. We turn more often to those we seldom see for support than frequent friends. Using *The Strength of Weak Ties* by Mark Granovetter and the “small world theory” by Duncan J. Watts as the basis of her arguments, she reveals the most efficient method for constructing human connections by network analysis.

- *Network Analysis — What Determines Human Actions* by Yuki Yasuda (Shinyo-sha, 1997)

Covering human relationships from lovers to organization and international networks, Yasuda searches for the common pattern in human-human and human-society relationships, then explains how network analysis techniques can be used to explore factors that determine human actions.

[SNS field]

- The current status of social network services such as Mixi, etc.

Table 7 Areas of research in network science, with categories created by the U.S. National Research Council

Research Area	Key Objective	Time Frame	Commercial Interest
Modeling, simulating, testing, and prototyping very large networks	Practical deployment tool sets	Mid-term	High
Command and control of joint/combined networked forces	Networked properties of connected heterogeneous systems	Mid-term	Medium
Impact of network structure on organizational behavior	Dynamics of networked organizational behavior	Mid-term	Medium
Security and information assurance of networks	Properties of networks that enhance survival	Near-term	High
Relationship of network structure to scalability and reliability	Characteristics of robust or dominant networks	Mid-term	Medium
Managing network complexity	Properties of networks that promote simplicity and connectivity	Near-term	High
Improving shared situational awareness of networked elements	Self-synchronization of networks	Mid-term	Medium
Enhanced network-centric mission effectiveness	Individual and organizational training designs	Far-term	Medium
Advanced network-based sensor fusion	Impact of control systems theory	Mid-term	High
Hunter-prey relationships	Algorithms and models for adversary behaviors	Mid-term	Low
Swarming behavior	Self-organizing UAV/UGV; self-healing	Mid-term	Low
Metabolic and gene expression networks	Soldier performance enhancement	Near-term	Medium

(Cited from: page 5 of <http://www.nap.edu/catalog/11516.html>)

Typical examples of such networks include MySpace in the U.S. and Mixi in Japan. These networks have enjoyed widespread success by providing services that let users create blogs and make connections to friends and other community members.

[Information search]

- Google's business strategy

Based on the hypothesis that "quality sites are sites supported by many other sites"—based on how documents on the WWW are linked together—Google has successfully applied a business strategy that presents the search results for a search term using an algorithm called PageRank™.

[Distribution field]

- Amazon's business strategy

Amazon has applied the long-tail phenom-

enon (power-law = scale-free property) in book marketing, creating a database that lets customers search for and purchase on the Internet, books ranking low on the sales list, and has succeeded in making enormous profits off such books, going against the conventional trend in which best-sellers and long-sellers were the primary source of revenue for book-sellers.

5.5 Fields of Application

One field of application of Complex Network Science is "congestion studies." *Congestion Studies* by Katsuhiro Nishinari (Shincho Sensho, 2006) presents a unified discussion of the phenomenon of congestion in information transferred over the Internet (packets) and of traffic caused by human and automobile trans-

portation from the perspective of networks. He introduces technology for relieving traffic congestion while explaining how congestion can be actively directed to benefit humans (for example, by preventing the spread of wildfires or viral infections).

5.6 Conclusions

“Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world.”

—Albert Einstein—

We have witnessed dramatic progress in research that analyzes various natural and social phenomena from the perspective of networks. Additionally, Internet businesses have emerged that skillfully apply the characteristics of the small world phenomenon and scale-free networks.

We see that universal network principles are in play at various hierarchical levels in both the natural world and human world (society). The field of complex network science has advanced rapidly from the end of the 20th century to the beginning of the 21st century, and there are great expectations for its application to a wide range of academic and business fields. Our hope is that this field will enlighten all those living in the 21st century in the crucial field of knowledge called network litera-

cy. One of the most important themes for the future is to apply this knowledge energetically not only in building future ICT infrastructures, such as new-generation network architectures, but in designing and constructing ideal social systems as well.

6 Future Prospects

Twenty years have passed since studies began in the late 1980s on information processing technologies “inspired by life.” As discussed in this chapter, research “inspired by life” represents a new paradigm, fully worthy of the term “paradigm shift” coined by Thomas Kuhn to describe revolutions in the history of science and technology and marking an epochal transition of the highest significance. If we penetrate into the process of “bio-evolution,” which surpasses human wisdom (once again, let me remind readers that even human wisdom is a product of bio-evolution), we may develop a powerful wellspring of ideas for designing and constructing the future information and communication society.

In this chapter, I have discussed how information processing models “inspired by life” can be applied to solve various real-world problems and provided examples of how such models have actually been applied.

References

- 1 Thomas S. Kuhn, “The Structure of Scientific Revolutions”, The University of Chicago Press, 1962.
- 2 H. Yukawa, “Readings from Einstein”, vol. 1: Special theory of relativity, quantum theory and Brownian motion, vol.2: General theory of relativity and unified theory, vol. 3: Einstein and his thoughts, Kyoritsu Shuppan, 1970-1972.
- 3 Jerry Mayer and John P. Holms, “BITE-SIZE EINSTEIN”, Kotoku-sha, 1997.
- 4 Charles Darwin, “On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life”, London, John Murray, 1859.
- 5 Richard Leakey abridged and introduced. “The illustrated ORIGIN OF SPECIES”, THE ORIGIN OF SPECIES by Charles Darwin, The Rainbird Publishing Group Ltd. 1979.
- 6 J. H. Holland, “Adaptation in Natural and Artificial System”, The University of Michigan Press, 1975.
- 7 A. Waibel, “Phoneme Recognition Using Time-Delay Neural Networks”, Report of Speech Committee, SP87-100, pp.19-24, Dec. 1987.

- 8 A. Waibel, H. Sawai and K. Shikano, "Modularity and Scaling in Large Phonemic Neural Networks", IEEE Trans. on ASSP, Vol.37, No.12, pp1888-1898, Dec. 1989.
- 9 S. Nakamura and H. Sawai, "Performance comparison of neural network architectures for speaker-independent phoneme recognition", Trans. IEICE, D-II, vol. J74-D-II, No.10, pp.1361-1369, Oct. 1991.
- 10 K. Fukuzawa, Y. Komori, H. Sawai and M. Sugiyama, "A Segment-based Speaker Adaptation Neural Network Applied to Continuous Speech Recognition", Proceeding of the ICASSP'92, San Francisco, 55.1, Vol.1, pp.433-436, Mar. 1992.
- 11 H. Sawai, "Frequency-Time-Shift-Invariant Time-Delay Neural Networks for Robust Continuous Speech Recognition", Proceedings of the ICASSP'91, Toronto, S2.1, Vol.1, pp.45-48, May 1991.
- 12 H. Sawai, "Axially Symmetric Neural Network Architecture for Rotation-Invariant Pattern Recognition," IEEE, WCCI(World Congress on Computational Intelligence), Proceedings of the IEEE ICNN'94, Vol.7, pp.4253-4258, Jun-Jul. 1994.
- 13 H. Sawai, S. Kizu, and T. Endo, "Parameter-free Genetic Algorithm (PfGA)", Tech Report, IEICE, Vol.J81-D-II No.2, pp.450-452, Feb. 1998.
- 14 S. Adachi and H. Sawai, "Effects of migration methods in parallel distributed parameter-free genetic algorithm", Trans. IEICE, D-I, Vol.J83-D-I, No.8, pp.834-843, Aug. 2000.
- 15 S. Adachi and H. Sawai, "Evolutionary computation method inspired by gene-duplication: application to functional optimization", J. of IPSJ, Vol.42, No.11, pp.2663-2671, Nov. 2001.
- 16 S. Ohno, "Evolution by Gene Duplication", Springer-Verlag, 1970.
- 17 K. Omori, Y. Fujiwara, S. Maekawa, H. Sawai, and S. Kitamura, "Evolutionary computation based on sexual selections driving asymmetrical mutation", J. of the Institutes of Systems, Control and Information Engineers (ISCIE), Vol.15, No.8, pp.422-429 (2002)
- 18 M. Hasagawa, "Why is the male peacock beautiful?", Kinokuniya Shoten, 1992.
- 19 Y. Iwasa, "Introduction to mathematical biology – investigating the dynamics in biological society", Kyoritsu Shuppan, 1998.
- 20 K. Wada, H. Doi, S. Tanaka, Y. Wada and M. Furusawa, "A neo-Darwinian algorithm: Asymmetrical mutations due to semi-conservative DNA-type replication promote evolution", Proc. Natl. Acad. Sci. USA, Vol.90, pp.11934-11938 (1993)
- 21 H. Suzuki, H. Sawai and Wojciech Piaseczny, "Chemical Genetic Algorithms – Evolutionary Optimization of Binary-to-real-value Translation in GAs", Journal of Artificial Life No.12, pp.1-27, MIT Press (2006).
- 22 W. Piaseczny, H. Suzuki and H. Sawai, "Chemical Genetic Programming - Evolutionary Optimization of the Genotype-to-Phenotype Translation Set", Journal of Artificial Life and Robotics, Vol.9, pp.202-208, Springer (2006)
- 23 C. S. Chan, J. C. Tay and H. Sawai, "Environmental Effects on the Coevolution of Pursuit and Evasion Strategies", Journal of Genetic Programming and Evolvable Machines, Kluwer Academic Publishers, Submitting.
- 24 Albert-Laszlo Barabasi, "Linked: The new science of networks", NHK Shuppan, 2002.
- 25 Mark Buchanan, "NEXUS: Small Worlds and the Groundbreaking Science of Networks", W. W. Norton & Company, Inc., 2002.
- 26 Duncan J. Watts, "Six degrees, The Science of a Connected Age", Hankyu Communications, 2003.
- 27 Duncan J. Watts, "Small world – The dynamics of networks between Orders and Randomness", Tokyo Denki University Press, 1999.
- 28 N. Masuda and N. Konno, "Introduction to complex networks", Sangyo Tosho, 2005.

-
- 29 N. Masuda and N. Konno, "What is complex network?", Kodan-sha, 2006.
 - 30 N. Masuda, "How do we connect each other? – Applying network science", Chuko Shinsho, 2007.
 - 31 Y. Hayashi (Ed.), "How to spread rumor – unraveling the world by network science", DOJIN Sensho009, Kagaku Dojin, 2007.
 - 32 Y. Hayashi (Ed.), "Toolbox of network science – untying the hidden phenomena of connections", Kindai Kagaku-sha, 2007.
 - 33 "Special issue on expansion of complex network science", Vol.49, No.3, Proc. of IPSJ, Mar. 2008.
 - 34 N. Kamibayashi, "From Network to Networking – Social and academic significance of network science", Vol.51, No.8, pp.348-353, System, Control and Information Society, 2007.
 - 35 "The forefront of network science", Mathematical Science, No.520-, Oct. 2006.
 - 36 Steven Strogatz, "SYNC – The Emerging Science of Spontaneous Order", Hayakawa Shobo, 2005.
 - 37 Steven Johnson, "Emergence – The Connected Lives of Ants, Brains, Cities, and Software", SOFT-BANK Publishing, 2004.
 - 38 H. Tanaka, "Life and complex system", Baihu-kan, 2002.
 - 39 H. Tanaka, "Life – Evolutional molecular network – An introduction to system evolutionary biology", Personal Media, 2007.
 - 40 H. Kitano and K. Takeuchi, "Sound life", Diamond-sha, 2007.
 - 41 T. Nishiguchi, "Companionship in the long distance and the neighborhood – Successful strategy of organized network", NTT Shupan, 2008.
 - 42 J. Kanamitsu, "Foundation on the analysis of social network – Towards the social relational capital", Keiso Shobo, 2003.
 - 43 S. Nozawa, "Readings on the theory of network – Family, Community and Social relational capital", Keiso Shobo, 2006.
 - 44 H. Ohe, "Sustainable community network – Regional management strategy in the information society", Nihon Chiiki Shakai Kenkyusho, 2007.
 - 45 Y. Yasuda, "Human network science – Investigation into the power hidden in the relation between humans", Nihon Keizai Shinbun-sha, 2004.
 - 46 Y. Yasuda, "Network analysis – What determine acts ?", Shinyo-sha, 1997.
 - 47 K. Nishinari, "Study on Congestion", Shincho Sensho, 2006.



SAWAI Hidefumi *Dr. Eng.*

*Director, Project Promotion Office,
Kobe Advanced ICT Research Center,
Kobe Research Laboratories
Information Science, Computer Science*