Development of a Combinatorial Optimization Calculation Engine

Takayuki Shibasaki
Research Manager
Digital Annealer Project, Technical Development Group
Fujitsu Laboratories LTD.

1. Introduction

It is said that improving performance in semiconductor integrated circuits through scaling is slowing down and will essentially come to an end in several years[1]. One means of improving performance without depending on scaling is to employ specialized domain-specific hardware, but this creates another problem in that increasing the degree of specialization narrows the application area making it difficult to recoup development costs. It is therefore important to focus on domains that can be applied as much as possible to a broad range of fields.

The need for selecting an optimal combination from among many combinations is present in a variety of fields. In general, the time needed to solve a combinatorial optimization problem increases dramatically as the scale of the problem grows, and as a result, conventional computers are limited in the scale that they can handle. Against this background, we have developed Digital Annealer, an optimization calculation engine for solving combinatorial optimization problems. This paper describes Digital Annealer architecture and introduces acceleration and scaling technologies.

2. Digital Annealer architecture

2.1 Search technique using the Ising energy function

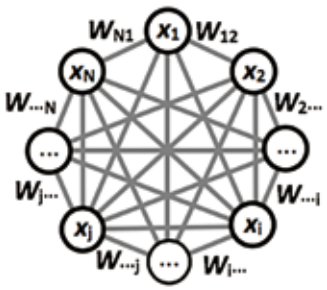
Digital Annealer performs statistical searching in parallel based on a Markov Chain Monte Carlo method (MCMC), which entails minimizing the Ising energy function shown by Eq. (1) at high speed.

$$E(X) = -(1/2)\sum_{i,j} W_{ij}\, x_i\, x_j - \sum_i b_i\, x_i \qquad (1)$$

$$x_i \in \{0,1\}\ (i = 1, 2, \cdots, N),\ W_{ii} = 0,\ W_{ij} = W_{ji}$$

Here, X is a set of bits with $X = (x_1, x_2, \ldots, x_N)$. N bit values x_i, $(i = 1, 2, \cdots, N)$ expresses a combination. W_{ij} denotes the combination coefficient between bit i and bit j and b_i is a bias term with respect to each bit. Digital Annealer features a full-connection architecture that can express mutual interaction among all bits (Figure 1).

Figure 1: Schematic of full-connection structure



2.2 Acceleration by parallel trials

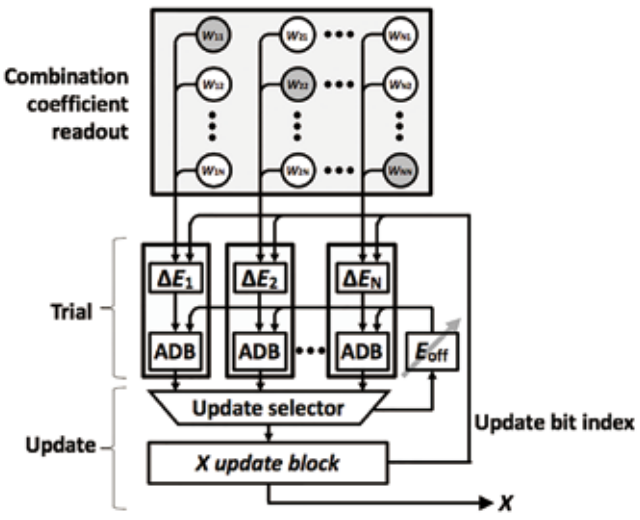
The operating cycle of the Digital Annealer is divided into two phases: the trial phase that selects bit inversion so as to satisfy an acceptance criterion and an update phase that inverts the selected bit (Figure 2).

The trial phase takes the current bit set $X = (x_1, x_2, \ldots x_N)$ and inverts a single bit value x_i to 1- x_i to obtain N-element neighboring state $X^{(i)}$ to be examined. Letting ΔE_i denote the increase in energy $E(X)$ obtained by making a transition from current state X to neighboring state $X^{(i)}$, the Metropolis-Hastings criterion given by Eq. (2) is used to decide whether to accept the bit inversion.

$$A(\Delta E_i) = \min[1, exp(-\beta \Delta E_i)] \qquad (2)$$

Here, $A(\Delta E_i)$ denotes the inversion-acceptance probability for energy change ΔE_i when inverting bit x_i to 1-x_j and β (=1/T) is the reciprocal of temperature T used in the simulated annealing method. Now, an Acceptance Decision Block (ADB) established for each bit compares the value of ΔE_i for each state change with an appropriate random number (noise) and outputs a binary flag that takes on the value of 1 with the acceptance probability of Eq. (2). A value of 1 for this binary flag means that inverting the corresponding bit is good. Finally, the update selector selects a

Figure 2: Digital Annealer configuration



single bit to be inverted based on those bits whose binary flags took on the value of 1. If there is no bit candidate for updating, the update selector outputs a flag with the value 0.

Here, we explain why the parallel trials scheme results in high-speed processing compared with conventional MCMC. Given the system in state X, let P_{single} denote the probability of making a transition to some new state for the case of a single trial in ordinary MCMC. Denoting the increase in energy when making a transition to neighboring state $X^{(i)}$ as ΔE_i, P_{single} is given by Eq. (3) using the transition-acceptance probability of Eq. (2).

$$P_{single} = (1/N) \sum_{i=1}^{N} A(\Delta E_i) \quad (3)$$

The probability of selecting a specific bit i is uniformly 1/N and the probability that an inversion can occur when that bit is selected is $A(\Delta E_i)$. Consequently, since any bit is fine, the probability that an inversion can occur is given by Eq. (3).

We now determine probability $P_{parallel}$ of making a transition to a new state under the parallel trials scheme of Digital Annealer. Here, we assume the selection rule that one bit out of N bits is selected with the probability given by Eq. (3). Now, considering that the probability of discovering a state as a transition destination is small when approaching an optimal solution so that $A(\Delta E_i) \ll 1$ (i=1, 2, ⋯, N), we get:

$$P_{parallel} = 1 - \prod_{i=1}^{N}(1 - A(\Delta E_i)) \cong \sum_{i=1}^{N} A(\Delta E_i) = N \cdot P_{single} \quad (4)$$

Equation (4) shows that the probability of making a transition to a new state in the case of parallel trials is N times that of the single search scheme. This technique is also advantageous in that no convergence problem occurs, which differs from a parallel trials scheme that simply updates multiple bits in parallel.

2.3 Acceleration by offset addition

If the state of a system falls into a local minimum of the Ising energy function, the probability of exiting that state into another one may be quite low even if using parallel trials. In such a case, the system will stay at the same local minimum for some number of cycles adding to the time taken until convergence. To therefore decrease the time stuck in a local minimum, we implemented a means of subtracting a positive offset E_{off} from the energy increment. This is practically equivalent to multiplying the common factor $\exp(\beta \cdot E_{off}) > 1$ by the acceptance probability of state inversion. To execute this, we add a fixed incremental value to the offset using an offset generator whenever a new candidate for bit inversion cannot be found and continue increasing the offset until the next state is found. This method dynamically controls the offset value so that the probability of finding the next destination for state inversion is 1 thereby shortening the time stuck in a local minimum.

2.4 Exchange Monte Carlo Method

There are a variety of acceleration techniques in stochastic searching using replicas. The simplest of these techniques is simple parallel operation, which gives the same problem to multiple replicas and performs statistically independent searches. Given M replicas, let the state vector of those replicas with the lowest objective function (energy) be the solution. Now, denoting the accuracy rate of individual replicas when annealing each replica for a certain number of cycles as P_0, accuracy rate P_{total} (M) for all M replicas can be given by Eq. (5).

$$P_{total}(M) = 1 - (1-P_0)^M \quad (5)$$

Thus, to obtain a total accuracy rate P_{total} (k) of 99%, the accuracy rate P_0 of individual replicas must be P_0=0.99 for M=1, P_0=0.37 for M=10, and P_0=0.045 for M=100. The time required (number of cycles) to obtain a correct answer becomes shorter by the amount deemed acceptable in making the target accuracy rate of individual anneal blocks smaller.

A serious problem in the simulated annealing method is that the state vector can become stuck near a local minimum as the temperature drops thereby slowing down the process of arriving at a true solution. This is known as the "hardly relaxing" problem in simulated annealing[2]. While Digital Annealer achieves high-speed processing by shortening the average trial time, it does not in essence solve the "hardly relaxing" problem. In addition, simple parallel operation as well cannot solve this problem.

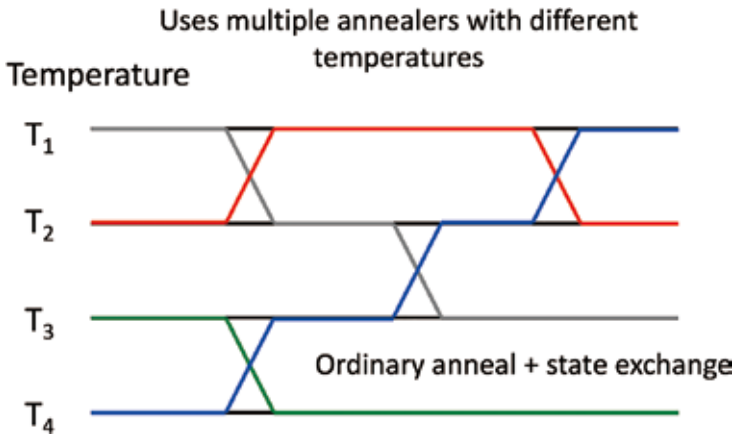
In the field of statistical physics, methods for solving the "hardly relaxing" problem were first devised in the mid-1990s. These methods conduct a stochastic search using multiple statistical ensembles (replicas) having different parameters (such as temperature). The exchange Monte Carlo method is one such method that uses multiple replicas[3]. This method prepares M replicas with different temperatures ($T_1 > T_2 > \cdots > T_{Ne}$) and conducts a stochastic search on each. It also exchanges state vectors between replicas with adjacent temperatures under certain conditions (using the Metropolis flow rule). Exchanging states in this manner can create a path between low-temperature replicas (that easily fall into "hardly relaxing") and high-temperature replicas (that have no "hardly relaxing" problem) and thereby solve the "hardly relaxing" problem overall (Figure 3). Digital Annealer supports both simple parallel operation and the exchange Monte Carlo method.

3. Scaling technologies

3.1 Approaches to scaling up

Problems of increasingly larger scale must be supported to expand the application domain of Digital Annealer. We have therefore been engaged in the ongoing development of scaling

Figure 3: Stochastic searching by the exchange Monte Carlo method



technologies from both hardware and software perspectives. At present, the number of bits *N* that can be handled by first-generation Digital Annealer hardware is 1,024 bits. To begin with, we will expand *N* in second-generation Digital Annealer hardware to 8,192 bits. Then, by combining with technology for decomposing and solving the problem by software means, we will enable Digital Annealer to deal with even larger problems than that capable by only hardware scaling thereby expanding its application domain even further.

3.2 Problem decomposition technology

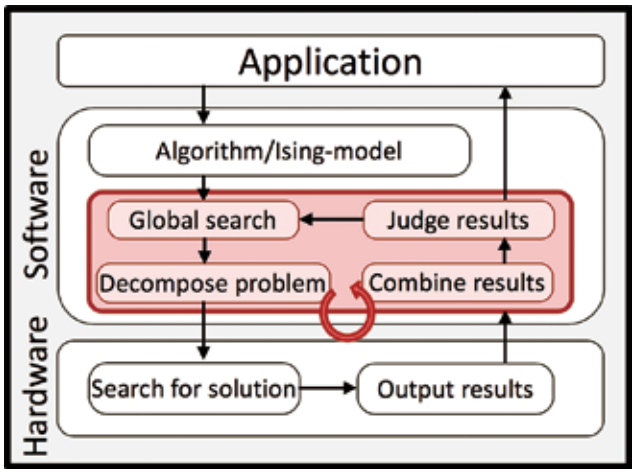
Simply decomposing a large-scale problem into portions that can be input into the system's hardware and optimizing each of these portions does not make for total optimization. This is because mutual relationships exist among these portions such that once one is optimized another is affected. Here, extracting a portion of the problem, fixing the states of the locations not extracted, and optimizing only the extracted location can achieve a partial optimization, but it cannot obtain a sufficient effect if an appropriate location has not been selected.

Against the above background, we developed technology that can process a problem on a scale larger than that possible on hardware alone. Applying this technology to the second-generation Digital Annealer will enable application to problems on a scale of 100,000 bits.

This technology, called the problem decomposition method, begins by determining an initial solution to the entire problem by performing a short-time total search. Next, it extracts a portion of the problem on a scale that can be input to the hardware and searches for a solution to that part using Digital Annealer. It then repeats a flow that returns that result to the total search any number of times while changing the location to be extracted. Finally, it derives a solution for the large-scale problem overall (Figure 4).

An important factor in improving the efficiency of optimizing the entire problem is determining what portions to extract taking the characteristics of the problem into account. For this

Figure 4: Problem decomposition method



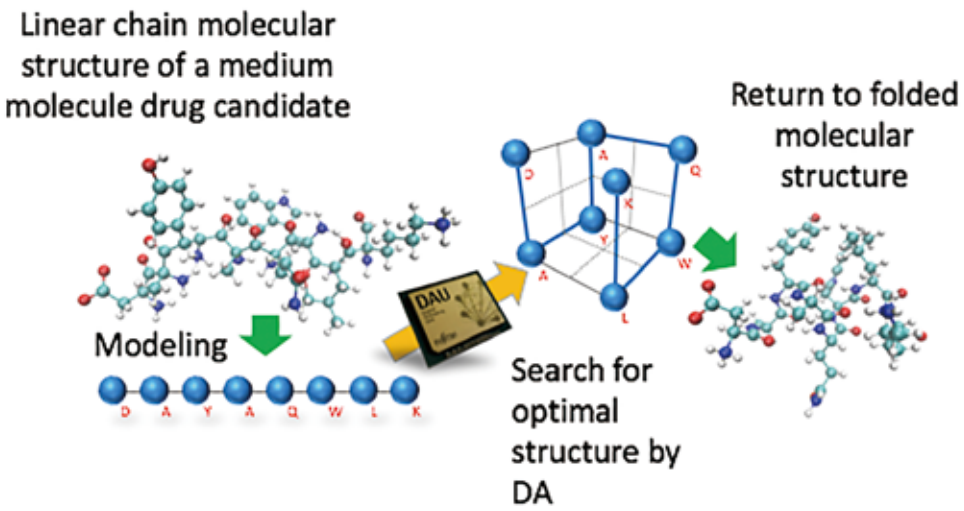
reason, we developed several decomposition methods focusing on relationships within the problem, such as a method that performs extraction centered about elements that easily change within the entire problem and a method that partitions the problem into locations for which inter-element bonding is small. Selecting a decomposition method applicable to the target problem enables solution searching with good efficiency for large-scale problems.

3.3 Application to a stable structure search problem

Focusing on simulation seeking a stable structure of a medium molecule drug candidate, we have confirmed that Digital Annealer using the above problem decomposition technology can be applied to large-scale problems as described below (Figure 5).

In medium molecule drug discovery, a medium molecule drug candidate that connects several to about 50 amino acids in a chain can demonstrate a drug effect by binding strongly with a targeted protein. To this end, the first step is to use Digital Annealer to search for the most stable structure after modeling each amino acid as a point on a lattice based on binding relationships among

Figure 5: Application to a stable structure search problem



those amino acids. The next step is to investigate the binding strength between the amino-acid structure just found and the target protein through docking calculations. Repeating this flow about 1,000 times enables a medium molecule drug candidate with high drug efficacy to be found.

We have demonstrated that applying this problem decomposition technology to the second-generation Digital Annealer can shorten simulation time for a medium molecule drug candidate on a scale of 48 amino acids (30K bits) from several hours by a conventional computer using the same technique for modeling amino acids to several minutes. We expect the application of the developed technology to Digital Annealer to accelerate the development of medium molecule drugs that have been attracting attention as next-generation drugs.

4. Expansion of application fields

To apply Digital Annealer to combinatorial optimization problems in a variety of real-world fields, it will be necessary to develop software for application to real problems in collaboration with specialists in those fields. In this regard, 1QB Information Technologies Inc. (Headquarters: Vancouver, Canada), a company excelling in the development of quantum computing applications, began collaborative work with Fujitsu in 2017 including the construction of an application development platform. Fujitsu Laboratories, meanwhile, entered into a strategic partnership with the University of Toronto also in 2017 establishing a research center in Toronto. It also concluded a comprehensive collaborative activity agreement with Waseda University for joint research on Digital Annealer in 2018 that included the establishment of a joint-research center (Figure 6). Going forward, Fujitsu plans to incorporate the results obtained from such joint research into Digital Annealer business with the aim of promoting solutions to real-world problems and contributing to social development and economic growth.

5. Conclusion

In this paper, we described the architecture of Digital Annealer, an optimization calculation engine for solving combinatorial optimization problems, and introduced acceleration and scaling technologies. Going forward, we plan to incorporate more acceleration and scaling technologies into Digital Annealer as needed. Furthermore, in addition to improving performance, we intend to expand the range of application fields through joint research with diverse research institutions and contribute to business in various fields.

References
[1] R. Colwell, "The Chip Design Game at the End of Moore's Law," Hot Chips 27, 2015.
[2] K. Fukushima, "On The Front Lines of the Monte Carlo Method—Roll the Dice and Integrate Method," Grant-in-Aid for Scientific Research on Priority Areas "Statistical Mechanical Approach to Probabilistic Information Processing (SMAPIP)" (sponsor), Lectures Providing Easy-to-understand Introductions to New Technologies to Young Researcher and Students "Information Processing by Probabilistic Algorithms" 2003. (in Japanese)
[3] K. Hukushima and K. Nemoto, "Exchange Monte Carlo Method and Application to Spin Glass Simulations," J. Phys. Soc. Jpn., 65, pp. 1604-1611 (1996).

Figure 6: Digital Annealer community