A Flip-count-based Dynamic Temperature Control Method for Constrained Combinatorial Optimization by Parallel Annealing Algorithms

Genta INOUE\(^{\dagger}\), Member, Daiki OKONOGI\(^{\dagger}\), Satoru JIMBO\(^{\dagger}\), Nonmembers, Thiem Van CHU\(^{\dagger}\), Masato MOTOMURA\(^{\dagger}\), and Kazushi KAWAMURA\(^{\dagger}\), Members

SUMMARY  
Annealing machines use an Ising model to represent combinatorial optimization problems (COPs) and minimize the energy of the model with spin-flip sequences. Pseudo temperature is a key hyperparameter to control the search performance of annealing machines. In general, the temperature is statically scheduled such that it is gradually decreased from a sufficiently high to a sufficiently low values. However, the search process during high and low temperatures in solving constrained COPs does not improve the solution quality as expected, which requires repeated preliminary annealing for pre-tuning. This paper proposes a flip-count-based dynamic temperature control (FDTC) method to make the preliminary annealing unnecessary. FDTC checks whether the current temperature is effective by evaluating the average number of flipped spins in a series of steps. The simulation results for traveling salesman problems and quadratic assignment problems demonstrate that FDTC can obtain comparable or higher solution quality than the static temperature scheduling pre-tuned for every COP.

key words: combinatorial optimization, Ising model, annealing machines, hyperparameter control, parallel annealing algorithms

1. Introduction

Annealing machines have attracted attention as domain-specific computers for solving real-world combinatorial optimization problems (COPs) [1–7]. They utilize an Ising model to represent COPs in a unified way [8]. The Ising model has binary variables called spins, and its Hamiltonian (i.e., energy function) is defined as a function of the spin state. For a given Ising model, annealing machines search for low-energy states and obtain optimized solutions. In the search process, the spin state is repeatedly updated according to a specific annealing algorithm. In order to accelerate this search process, previous studies developed various parallel annealing algorithms that can update multiple spins simultaneously [9–13]. Especially, ratio-controlled parallel annealing (RPA) [12] is effective for a wide range of COPs since it can change the spin-update parallelism depending on the target problem.

Most real-world COPs are classified into constrained COPs that aim to minimize (or maximize) one or more objective functions under several constraints. Constrained COPs are converted to the Ising model such that feasible solutions (i.e., solutions satisfying the constraints) have lower energies than infeasible solutions; otherwise, lowering the Hamiltonian does not lead to feasible solutions. As a result of this conversion, a large energy gap is generated between feasible and infeasible solutions.

Parallel annealing algorithms update the spin state by stochastically judging whether each spin should be flipped or not. The spin-flip probability is determined based on the energy decrease caused by the spin flip and a hyperparameter called pseudo temperature. The temperature is initially set to be high, which aggressively accepts spin flips that increase the energy. Then, it is monotonically decreased for finally fixing the spin state. However, it is not so easy to specify a temperature scheduling (i.e., a set of the initial temperature and the decreasing rate) which is fit for the target problem.

Furthermore, when the target problem is a constrained COP like a traveling salesman problem (TSP) or a quadratic assignment problem (QAP), to specify an appropriate temperature scheduling becomes more difficult. Firstly, the spin state cannot be a feasible solution during high temperatures; this is because high-energy states are filled with infeasible solutions due to the Ising-model formulation of constrained COPs. Thus, the search process during high temperatures may not be effective to minimize the objective functions. Secondly, the spin state is hard to escape from local minima during low temperatures; this is because state transitions rarely occur due to the energy gap between feasible and infeasible solutions. Thus, the search process during low temperatures may not make sense. In summary, the effective temperature range is fairly narrow when constrained COPs are solved by annealing machines.

For this reason, pre-tuning the temperature range is an essential process to maximize the annealing performance. However, such a pre-tuning approach contains several challenges. Firstly, the effective temperature range is different depending on the target problem, which requires repeated preliminary annealing. Secondly, the temperature scheduling pre-tuned for a COP is not always effective for the COP since the search progress also has an effect on the effective range. Thirdly, the conventionally-used static temperature scheduling has no mechanisms to escape from local minima and eventually suffers from state stagnation.

Dynamic temperature control is a promising approach.
to overcome these challenges. Improved parallel annealing (IPA) [14] was proposed by integrating a per-step dynamic temperature control method into a parallel annealing algorithm. This method detects an empty step (i.e., a step where no spins flip), and increases the temperature at the following step. IPA can keep the temperature from being too low depending on the search progress for a COP, and hence mitigate state stagnation. However, IPA does not have no mechanisms to detect whether the current temperature is effective or not. As a result, preliminary annealing is still required to specify the baseline temperature scheduling. Moreover, the per-step control may cause drastic temperature fluctuation, preventing the stable convergence.

This paper proposes a flip-count-based dynamic temperature control (FDTC) method to make the preliminary annealing unnecessary. By evaluating the average number \( c_{\text{avg}} \) of flipped spins in a series of steps, this method enables checking whether the current temperature is effective or not. More specifically, the temperature is drastically decreased in the case that \( c_{\text{avg}} \) is too large, whereas increased when \( c_{\text{avg}} \) is too small. The simulation results for TSPs and QAPs demonstrate that the solution quality by FDTC is comparable to or higher than the static temperature scheduling pre-tuned for every COP. Furthermore, it is also confirmed that FDTC works well when the initial temperature is extremely high or extremely low.

The rest of this paper is organized as follows. Sect. 2 introduces the Ising model and RPA. Then, in Sect. 3, the effective temperature control is discussed. After that, Sect. 4 proposes FDTC, and Sect. 5 evaluates its performance. Finally, the paper is wrapped up in Sect. 6.

2. Preliminaries

2.1 Ising Model Formulation for Constrained COPs

The Ising model can represent various COPs in a unified way. It consists of \( N \) spins \( \sigma = \{\sigma_i|1 \leq i \leq N, \sigma_i \in \{-1, 1\}\}, \) and its Hamiltonian \( H \) is defined as a function of the spin state \( \sigma \) as follows:

\[
H(\sigma) = -\sum_{(i,j)} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i,
\]

where \( J_{ij} \in \mathbb{R} \) and \( h_i \in \mathbb{R} \) are the interaction coefficient between the \( i \)-th and \( j \)-th spins and the external magnetic field on the \( i \)-th spin, respectively. When a COP is converted to the Ising model, it is required to determine \( J \in \mathbb{R}^{N \times N} \) and \( h \in \mathbb{R}^N \) such that the energy \( H(\sigma) \) in Eq. (1) becomes lower for better spin states (i.e., better solutions). Efficient Ising-model formulations of COPs have been discussed in [8, 15].

The Hamiltonian for a constrained COP is composed of the linear combination of the objective term \( H_{\text{obj}} \) and the constraint term \( H_{\text{con}} \) as follows:

\[
H(\sigma) = H_{\text{obj}}(\sigma) + \alpha H_{\text{con}}(\sigma),
\]

where \( \alpha > 0 \) is a constraint coefficient. In Eq. (2), \( H_{\text{con}}(\sigma) = \)

0 if the spin state \( \sigma \) is a feasible solution satisfying all the constraints; otherwise, \( H_{\text{con}}(\sigma) > 0 \). Therefore, solving a constrained COP corresponds to searching for a spin state \( \sigma \) that lowers \( H_{\text{obj}}(\sigma) \) while satisfying \( H_{\text{con}}(\sigma) = 0 \). Note that a sufficiently large value is used as \( \alpha \) to ensure that feasible solutions have lower energies than infeasible solutions.

2.2 RPA

RPA [12] is one of the parallel annealing methods, which can control the spin-update parallelism by a hyperparameter \( \epsilon \). The spin-update ratio \( \epsilon \) (0 < \( \epsilon \) < 1) prevents the spin state from changing too vastly.

The algorithm of RPA is shown in Algorithm 1. RPA searches for an optimized (i.e., low-energy) state by repeatedly updating multiple spins in parallel. In a step \( s \), spin \( i \) is stochastically flipped based on the flip probability \( \sigma_i \) calculated from the previous spin state \( \sigma_{(s-1)} \) and the temperature \( T \). Note that \( 2 \ln \sigma_{(s-1)} \) shows the energy decrease caused by a single flip of spin \( i \). In other words, spin \( i \) is likely flipped when this value is large.

The temperature \( T \) also affects the flip probabilities. It is initially set to be high for aggressively accepting spin flips that increase the energy. Then, it is monotonically decreased for finally fixing the spin state. In Algorithm 1, it is updated per frame (i.e., some grouped steps). Conventionally, the temperature scheduling is given by:

\[
T_f = T_{\text{init}} \times r^{f-1},
\]

where \( T_f \) shows the temperature after updating \( f \) times. As in Eq. (3), a temperature scheduling is specified by the initial
3. Discussions on Effective Temperature Control

3.1 Temperature Scheduling for Constrained COPs

This section demonstrates how the annealing performance varies depending on the temperature scheduling given by Eq. (3), and then identifies what factor is important for efficiently solving constrained COPs. We here conduct a preliminary experiment to examine the impact of the initial temperature under a fixed decreasing rate. In this experiment, we use a 14-city TSP as the target problem. Then, we prepare the baseline temperature scheduling such that the initial temperature $T_{\text{init}}$ is sufficiently high and the decreasing rate $r$ ensures the convergence in 100 frames. When the decreasing rate is fixed, the lower the initial temperature is, the smaller number of frames is required for the convergence. Thus, we set the initial temperature as $T_{\text{init}} = \frac{1}{9} (10^n)$ and the number of frames as $100 - 10n (0 \leq n \leq 9)$, and solve the TSP instance.

Figure 1 shows the experimental results. In $n \leq 3$, the solution quality is almost the same despite the number of annealing steps is smaller for larger $n$. This means that the search process at high temperatures is ineffective to improve the solution quality in solving constrained COPs. Also, the solution quality clearly gets worse in $n \geq 5$. Since the spin state is easily trapped in local minima due to the energy gap between feasible and infeasible solutions, the search process at low temperatures is also less effective. From these results, we can confirm that the search process during temperatures between $T_{\text{init}} \times r^{10}$ and $T_{\text{init}} \times r^{30}$ is the most effective. Thus, to continue the annealing at such temperatures is considered to be important for efficiently solving constrained COPs.

Conventionally, we need repeated preliminary annealing to obtain this effective temperature range. It is a time-consuming task that is required for every COP. Moreover, the pre-tuned temperature scheduling may not be appropriate depending on the search progress. In addition, the static temperature scheduling may suffer from state stagnation since it has no mechanisms to escape from local minima.

3.2 Per-step Dynamic Temperature Control

IPA [14] is a parallel annealing, which dynamically controls the temperature based on spin flip. The scheduling basically follows preset monotonically cooling per step yet temporarily increases the temperature just after an empty step. In IPA, the temperature where an empty step occurs is considered to be too low, sustaining the effective solution search by temporarily heating.

However, the per-step temperature control in IPA may cause drastic temperature fluctuation. Table 1 shows the spin-flip frequency for 1,000 consecutive annealing steps when a COP is solved by RPA. Although more than half of the steps are empty, two or more spins are flipped at 20% of the steps. This means that many spins are still flippable during the current and slightly lower temperatures. Since the low-energy state is obtained during low temperatures where empty steps occur enough, occasional empty steps should be tolerated without heating. Nevertheless, the temperature will likely be increased at the next update in IPA, and the heating can continue even in several following steps. As a result, the temperature may not reach such a low throughout IPA execution. Flipping many spins by heating too much, the previous search progress is not reflected and is wasteful.

Moreover, IPA has no mechanisms to detect too-high temperatures and to avoid them. Like too-low temperatures are avoided by temporarility heating, too-high temperatures are expected to be avoided; otherwise, the temperature may take much time to reach the comfortable range. However, too-high temperatures cannot be detected only by monitoring if the step is empty because empty steps cannot occur there.

4. FDTC: Flip-count-based Dynamic Temperature Control

4.1 Motivation

Securing effective search duration longer affects the anneal-
ing performance and depends on how long the search is conducted at comfortable temperatures. Nevertheless, a preset temperature schedule without pre-tuning usually has a range to spare to cover the comfortable range. Therefore, in order to improve the performance, it is necessary to eliminate the uncomfortable temperatures by dynamically controlling the temperature during the annealing execution. That includes avoiding also too-high temperatures, not only too-low ones. In addition, the temperature evaluation and update should be stable; otherwise, they can make obtaining a decent solution harder.

Firstly, in order to detect both too-high and too-low temperatures, we focus on the spin-flip counts, instead of the absence of spin-flip, as a criterion for temperature evaluation. Like IPA, monitoring if the step is empty can be an excellent criterion when aiming only to detect too-low temperatures. However, considering also too-high temperatures, it is required to evaluate the temperature by comparing the flip count with each threshold for high and low.

Secondly, we guess that quenching (i.e., more rapid cooling than usual) is suitable for quickly reaching the comfortable range from higher temperatures. Though quick cooling can be realized even under classically decreasing schedules by making the decreasing rate small, the constantly rapid transition can also shorten the comfortable solution search. On the other hand, by temporarily quenching only when the temperature is regarded as too high, the temperature transition is kept gentle after quickly reaching the comfortable range. Moreover, the temperature transition lower than the preset range is possible by keeping quenching. Therefore, a comfortable solution search can be conducted even if the preset range is a little higher than the comfortable.

Thirdly, we guess that the temperature evaluation and update get more stable by conducting them per frame. Spins are flipped stochastically, and empty steps can occur even at temperatures where spins can be flipped enough. After some empty steps, spin-flip often occurs and finds better solutions in the neighborhood. Thus, in order to search closely in the neighborhood around a local optimum, it is favorable to stay at the temperature when empty steps occasionally occur for a while. However, dynamically updating the temperature per step, heating occurs soon after an empty step, and then the spin state can disappear from the previous neighborhood before a careful local search. For this issue, per-frame temperature evaluation tolerates occasional empty steps by evaluating the number of flipped spins together in a frame, and per-frame temperature update forces constant temperature search as long as it is in the frame even if many empty steps occur there.

4.2 FDTC Algorithm

Based on the motivation, we propose a flip-count-based dynamic temperature control (FDTC) method. Algorithm 2 shows the algorithm of RPA with the proposed FDTC method. This method newly introduces two thresholds for quenching ($\theta_q$) and heating ($\theta_h$). The output is the optimized spin state and its energy, corresponding to the minimum energy in the whole search process. This is because the dynamic temperature control has no guarantee that the spin state finally converges. Under FDTC, the temperature is updated every frame, basically following the baseline temperature scheduling with the decreasing rate $R$. However, it is quenched with $R^2$ if the average number $c_{avg}$ of flipped spins in a frame is larger than $\theta_q$, whereas is heated with $R^{-1}$ if $c_{avg}$ is less than $\theta_h$.

The value of $\theta_q$ is likely given based on how many simultaneously flipped spins are enough (e.g., the number of flipped spins enough for the state transition between feasible solutions), and giving a large enough value to $\theta_q$ does not hinder the annealing compared with the conventional. In this paper, flipping spins at more than half the intra-frame steps on average is regarded as enough for state transition. When the average flip count $c_{avg}$ is less than 1.0 in a frame, that guarantees that at least $(1 - c_{avg}) \times I$ steps are empty. When $\theta_h < 1.0$, heating does not occur unless more than $(1 - \theta_h) \times I$ steps are empty and sometimes does not occur.

### Algorithm 2 RPA with the FDTC method

**Input:**
- # of spins: $N$, initial spin state: $\sigma^{(0)}$
- interaction coefficients: $J_i$, external magnetic fields: $h$
- # of annealing steps: $S$, # of frames: $F$
- # of intra-frame annealing steps: $I = S/F$
- initial temperature: $T_{init}$, basic decreasing rate: $R$
- thresholds for quenching and heating: $\theta_q$ and $\theta_h$
- spin-update ratio: $\epsilon$

**Output:**
- optimized spin state: $\sigma^{opt} = \arg\min_i \{H(\sigma_i^{(s)})\}$
- optimized energy: $H(\sigma_{opt})$

1. for $s = 1$ to $S$ do
2.   if $s \mod I = 1$ then
3.     if $s = 1$ then
4.       $T \leftarrow T_{init}$
5.     else
6.       $c_{avg} \leftarrow c_{total}/I$
7.       $r \leftarrow \text{rate\_calculation}(R, c_{avg}, \theta_q, \theta_h)$
8.       $T \leftarrow T \times r$
9.   end
10. end
11. for $i = 1$ to $N$ do
12.   Local field calculation: $h_i = \sum_j J_{ij} \sigma_j^{(s-1)} + h_i$
13.   Flip probability calculation: $p_i \leftarrow \text{sigmoid} \left(\frac{-2h_i \sigma_i^{(s-1)}}{T}\right)$
14.   Random number generation: $\text{rand} \in [0, 1)$
15.   if $p_i \times \epsilon > \text{rand}$ then
16.     $c_{total} \leftarrow c_{total} + 1$
17.     $\sigma_i^{(s)} \leftarrow -1 \times \sigma_i^{(s-1)}$
18.     $\sigma_i^{(s)} \leftarrow \sigma_i^{(s-1)}$
19. end
20. function rate\_calculation($R$, $c_{avg}$, $\theta_q$, $\theta_h$)
21.   if $c_{avg} > \theta_q$ then
22.     $r \leftarrow R^2$
23.   else if $c_{avg} < \theta_h$ then
24.     $r \leftarrow R^{-1}$
25.   else
26.     $r \leftarrow R$
27. return $r$
5. Simulation Results

5.1 Problems

We solved ten constrained COPs for the simulation: half are traveling salesman problems (TSPs) from TSPLIB [16], and the rest are quadratic assignment problems (QAPs) from QAPLIB [17]. TSPs aim to minimize the path when visiting each listed city once and returning to the initial one. When formulating one with an Ising model, the inter-city distances are given. An Ising model of a TSP has two one-hot constraints: forces each city to be visited exactly once, and the other forces the traveler to visit exactly one city simultaneously. QAPs aim to minimize the total cost when assigning each listed facility to a location. When formulating one with an Ising model, the inter-location distance and the inter-facility supply weights are given. Like TSPs, an Ising model of a QAP has two one-hot constraints: one forces each facility to be assigned exactly once, and the other forces each location to have exactly one facility.

In this paper, the shared constraint coefficient is given to the two constraints. The number of spins and constraint coefficient $\alpha$ for each problem are listed in Table 2. The coefficient of each TSP is equal to its maximum inter-city distance. That of each QAP is based on the doubled product of the maximum inter-location distance and the maximum inter-facility supply weights: the exact values are given to chr12a, scr15, and chr22a, and the tripled one is given to kria30a.

5.2 Settings

The common parameter settings are listed in Table 3. The threshold value for quenching $\theta_q$ is set to 4.0 because the minimum hamming distance of Ising models between two feasible solutions of TSPs and QAPs is four.

Figure 1 in Sect. 3.1 shows that the energies get worse when skipping more than 40 initial frames in this experiment. Figure 2 shows a transition of the average flip counts around the 40th frame when annealing is conducted under the whole baseline temperature scheduling in the same experiment. Since the average numbers around the 40th frame are approximately 0.5, and the later numbers are less than that, the threshold value for heating $\theta_h$ is empirically set to 0.5. That indicates that heating does not occur unless more than half the intra-frame steps are empty. In all cases, initial spin states are random.

As mentioned in Sect. 3.1, pseudo temperature $T$ is typically controlled from a sufficiently large value to a sufficiently small value for a given Ising model. Thus, we used the following setting ($T_{\text{init}}, R$) as a basic pseudo temperature controller:

$$T_{\text{init}} = 0.01 \times N \times \max |J_{ij}|, \quad (4)$$

$$R = \left( \frac{10 \times \min J_{ij} \max |J_{ij}|}{T_{\text{init}}} \right)^{\frac{1}{1.75}}. \quad (5)$$

This setting means that, when the conventional temperature scheduling in Eq. (3) is used, $T = 0.01 \times N \times \max |J_{ij}|$ in the first frame, whereas $T = 10 \times \min J_{ij} \max |J_{ij}|$ in the last frame. When constrained COPs in Table 2 are solved by Algorithm 1 under this setting, it is expected that many spins are flipped in the first frame and no spins are flipped in the last frame.

We compare the solution quality when four different pseudo temperature control policies (Baseline, Pre-tuned, IPA [14], and FDTC) are used in solving each COP. We used Algorithm 1 for Baseline and Pre-tuned, and Algorithm 2 for FDTC. In Baseline and FDTC, the temperature setting ($T_{\text{init}}, R$) is uniformly given by Eqs. (4) and (5). In Pre-tuned, it is prepared for each COP instance after the Baseline is executed once. In this evaluation, we obtained the temperature

<table>
<thead>
<tr>
<th>Type</th>
<th>Problem</th>
<th># of spins $N$</th>
<th>Coeff. $\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSP</td>
<td>burma14</td>
<td>196</td>
<td>1,261</td>
</tr>
<tr>
<td></td>
<td>ulysses16</td>
<td>256</td>
<td>2,789</td>
</tr>
<tr>
<td></td>
<td>ulysses22</td>
<td>484</td>
<td>2,789</td>
</tr>
<tr>
<td></td>
<td>bayg29</td>
<td>841</td>
<td>586</td>
</tr>
<tr>
<td></td>
<td>el106</td>
<td>2,001</td>
<td>86</td>
</tr>
<tr>
<td>QAP</td>
<td>chr12a</td>
<td>144</td>
<td>18,624</td>
</tr>
<tr>
<td></td>
<td>scr15</td>
<td>225</td>
<td>29,340</td>
</tr>
<tr>
<td></td>
<td>chr22a</td>
<td>484</td>
<td>9,100</td>
</tr>
<tr>
<td></td>
<td>kra30</td>
<td>900</td>
<td>9,960</td>
</tr>
<tr>
<td></td>
<td>wil50</td>
<td>2,500</td>
<td>1,872</td>
</tr>
</tbody>
</table>

Table 2: List of the Ising models used in this paper.

Table 3: Common parameter settings.

<table>
<thead>
<tr>
<th>Param.</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td># of trials</td>
<td>100</td>
</tr>
<tr>
<td># of frames $F$</td>
<td>100</td>
</tr>
<tr>
<td># of intra-frame steps $I$</td>
<td>1,000</td>
</tr>
<tr>
<td>$\epsilon$ of IPA</td>
<td>0.2</td>
</tr>
<tr>
<td>Threshold of quenching $\theta_q$</td>
<td>4.0</td>
</tr>
<tr>
<td>Threshold of heating $\theta_h$</td>
<td>0.5</td>
</tr>
</tbody>
</table>
setting \((T_{\text{init}}, R)\) tuned for a COP instance according to the following procedures:

1. Record the per-frame trace data of pseudo temperature \(T_f\) and the average number \(c_f\) of flipped spins, which are obtained in the process of Baseline execution.
2. Identify the first frame \(f_1\) that satisfies \(c_{f_1} \geq 4\). Then, set \(T_{\text{ini}} = T_{f_1}\).
3. Identify the last frame \(f_2\) that satisfies \(c_{f_2} > 0\). Then, set \(R = (T_{\text{ini}}/T_{f_1})^{\frac{1}{T_{f_2}}}, \) where \(T_{\text{fin}} = T_{f_2+1}\).

Under this setting, it is expected that approximately four spins (i.e., not many spins) are flipped in the first frame and the spin state converges near the last frame. On the other hand, IPA based on RPA updates the temperature per step, and its initial temperature, decreasing rate, and unique parameter are set to \(T_{\text{ini}} = 1 \times 10^3, r = 0.97\) and \(x = 90\), respectively [14].

5.3 Results

Table 4 shows the simulation results, which summarizes the minimum, maximum, average of the energies. The values in each row can be obtained after a COP instance is solved 100 times by a pseudo temperature control policy. Also, the last column shows how many feasible solutions are included in the 100 solutions. As shown in Table 4, FDTC averagely outperforms Baseline in all cases and Pre-tuned in most cases. The solution qualities on eil51, chr22a, and kra30a under FDTC are inferior to those under Pre-tuned but close. Those results indicate that FDTC performs comparably to Pre-tuned even without pre-tuning temperature. However, Table 4 shows that FDTC outperforms IPA on larger problems yet underperforms on smaller ones.

Figure 3(a)-(e) show the transitions of the temperatures and energies in the four cases. Comparing Figure 3(a) and (b), it can be reconfirmed that the duration of effective search
get longer by pre-tuning. Figure 3(c) shows that some temperature steps at the initial stage are skipped by quenching in FDTC, and the solution search is sustained by heating. Quenching stops just around the temperature where Tuned-Exp. starts, and heating occurs around the one where the minimum energy is obtained under Tuned-Exp. Considering that the minimum energy is obtained sometime after the first heating, heating contributes to searching for better solutions, and so does quenching by bringing the first heating forward. Figure 3(d) indicates that global and local searches in IPA were repeated. IPA is good at finding a number of local optimums by the frequent switch of those search styles, but the momentary local search hardly refines the solutions anymore. On the other hand, FDTC is suitable for finding not so many but better local optimums by steady local search throughout a frame. Therefore, FDTC is expected to overwhelm IPA when solving huge problems.

5.4 Discussion

In this section, we investigate the effect of hyperparameters in FDTC on the search performance. As shown in Sect. 4.2, FDTC takes thresholds for quenching ($\theta_q$) and heating ($\theta_h$) as input. Moreover, decreasing rates of quenching ($R^2$) and heating ($R^{-1}$) are also considered to be hyperparameters in FDTC. The thresholds and the decreasing rates are discussed in Sect. 5.4.1 and Sect. 5.4.2, respectively.

5.4.1 Thresholds

In the above experiments, the threshold values of quenching ($\theta_q$) and heating ($\theta_h$) are fixed to 4.0 and 0.5, respectively, based on the characteristic of target Ising models. However, it can be considered that there is room to optimize these parameters and enhance the search performance of FDTC. We here used one TSP (bays29) and one QKP (kra30a) to investigate the effect of $\theta_q$ and $\theta_h$ on the search performance. In this experiment, $\theta_q$ was varied within a range larger than 1.0, and $\theta_q$ was varied within a range smaller than 1.0. Note that the settings except the thresholds are as in Sect. 5.2.

Figure 4 shows the experimental results, where each value shows solution quality represented by the rate of energy to the global optimum. Also, the results are illustrated with a heat map where smaller values have lighter colors. From these results, both too large and too small values of $\theta_q$ degrade the solution quality, whereas the best value is within (0.0, 1.0). On the other hand, $\theta_h$ does not affect the solution quality as well as $\theta_q$ does. The aim of quenching is to avoid wasteful global search in the initial stage, and its role ends once the temperature reaches the comfortable
Fig. 4: Comparison of the solution quality for each threshold setting in FDTC. The values are expressed by the rate of energy to the global optimum, i.e., the smaller value is better.

range. Thus, it appropriately works unless $\theta_h$ is too small. Furthermore, the most notable point in these results is that the solution quality is in the same level when $\theta_h$ and $\theta_q$ are set to around their optimal settings. In other words, FDTC can enhance the search performance compared with the classical exponential cooling (Exp.) even when the thresholds are not tuned strictly.

In this paper, we have set the threshold values based on the minimum Hamming distance between feasible solutions. However, the above results indicate that our settings are sufficient but not optimal. In the future, we will examine how to optimize the thresholds without preliminary experiments.

5.4.2 Decreasing Rates

In the above experiments, the decreasing rates of quenching ($r_q$) and heating ($r_h$) are fixed to $R^2$ and $R^{-1}$, respectively. Just one temperature is skipped when the exponent part is $-2$, whereas temperature is returned to the previous one when it is $-1$. By employing these values together with the basic decreasing rate $R$, temperature is expected to be recovered appropriately in just one frame even if quenching and heating overwork. The experimental results in Sect. 5.3 demonstrate that $r_q = R^2$ and $r_h = R^{-1}$ have an effect on the enhancement of the search performance.

However, it is also important to investigate how the decreasing rates affect the search performance to understand the potential of FDTC. We here used one TSP (bayg29) and one QKP (kra30a) to investigate the effect of $r_q$ and $r_h$ on the search performance. In this experiment, the exponent part of $r_q$ was varied within a range larger than 1.0, and that of $r_h$ was varied within a range smaller than 1.0. Note that the other settings are as in Sect. 5.2.

Figure 5 shows the experimental results, where each value shows solution quality represented by the rate of energy to the global optimum. Also, the results are illustrated with a heat map where smaller values have lighter colors. From these results, minor changes in the decreasing rates of quenching and heating do not affect the solution quality so much compared with that of Baseline. Figure 5(a) shows some trends of the solution quality for these decreasing rates (e.g., gradual heating leads to higher solution quality), but...
Table 5: Comparison of the solution quality of RPA under FDTC to that under exponential cooling (Exp.). Up and down arrows in Scd. show if the temperature ranges are higher or lower than the tuned ones in Sect. 5.4.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Scd.</th>
<th>$H_{\text{min}}$</th>
<th>$H_{\text{max}}$</th>
<th>$H_{\text{avg}}$</th>
<th>Std.</th>
<th>CS rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>burma14</td>
<td>Exp. (↑)</td>
<td>5,587</td>
<td>9,723</td>
<td>7,697.59</td>
<td>784.87</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>3,359</td>
<td>4,214</td>
<td>3,948.50</td>
<td>151.17</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>4,695</td>
<td>7,160</td>
<td>5,745.33</td>
<td>537.13</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>3,394</td>
<td>4,205</td>
<td>3,844.40</td>
<td>167.90</td>
<td>1.00</td>
</tr>
<tr>
<td>ulysses16</td>
<td>Exp. (↑)</td>
<td>7,594</td>
<td>9,250</td>
<td>8,483.76</td>
<td>386.71</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>8,174</td>
<td>13,175</td>
<td>10,684.34</td>
<td>1,035.95</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>7,559</td>
<td>8,894</td>
<td>8,220.76</td>
<td>293.13</td>
<td>1.00</td>
</tr>
<tr>
<td>ulysses22</td>
<td>Exp. (↑)</td>
<td>11,681</td>
<td>20,496</td>
<td>16,486.89</td>
<td>1,855.91</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>9,402</td>
<td>11,380</td>
<td>10,495.11</td>
<td>416.38</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>9,631</td>
<td>13,170</td>
<td>11,321.43</td>
<td>744.75</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>8,736</td>
<td>11,138</td>
<td>10,209.27</td>
<td>465.75</td>
<td>1.00</td>
</tr>
<tr>
<td>bayg29</td>
<td>Exp. (↑)</td>
<td>7,056</td>
<td>11,093</td>
<td>9,357.12</td>
<td>695.36</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>2,467</td>
<td>3,009</td>
<td>2,776.61</td>
<td>120.76</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>2,885</td>
<td>4,260</td>
<td>3,514.09</td>
<td>278.32</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>2,336</td>
<td>2,926</td>
<td>2,682.11</td>
<td>106.96</td>
<td>1.00</td>
</tr>
<tr>
<td>eil51</td>
<td>Exp. (↑)</td>
<td>1,083</td>
<td>1,438</td>
<td>1,295.84</td>
<td>70.80</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>933</td>
<td>1,101</td>
<td>1,036.74</td>
<td>36.68</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>1,248</td>
<td>1,717</td>
<td>1,450.88</td>
<td>97.83</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>941</td>
<td>1,087</td>
<td>1,013.85</td>
<td>33.21</td>
<td>0.99</td>
</tr>
<tr>
<td>chr12a</td>
<td>Exp. (↑)</td>
<td>19,408</td>
<td>77,662</td>
<td>56,831.86</td>
<td>11,759.97</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>12,544</td>
<td>20,936</td>
<td>16,428.32</td>
<td>1,933.52</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>13,990</td>
<td>51,724</td>
<td>32,821.30</td>
<td>7,812.16</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>10,214</td>
<td>19,330</td>
<td>14,729.84</td>
<td>1,502.60</td>
<td>1.00</td>
</tr>
<tr>
<td>scr15</td>
<td>Exp. (↑)</td>
<td>81,176</td>
<td>181,086</td>
<td>135,855.02</td>
<td>19,855.38</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>60,770</td>
<td>73,880</td>
<td>67,523.06</td>
<td>2,688.95</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>67,044</td>
<td>106,446</td>
<td>86,538.84</td>
<td>8,269.20</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>58,760</td>
<td>71,428</td>
<td>65,725.66</td>
<td>2,472.58</td>
<td>1.00</td>
</tr>
<tr>
<td>chr22a</td>
<td>Exp. (↑)</td>
<td>9,608</td>
<td>29,954</td>
<td>16,997.72</td>
<td>6,794.03</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>8,180</td>
<td>12,056</td>
<td>10,437.78</td>
<td>695.34</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>9,286</td>
<td>19,908</td>
<td>13,250.18</td>
<td>1,971.10</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>8,508</td>
<td>11,998</td>
<td>9,978.96</td>
<td>622.76</td>
<td>1.00</td>
</tr>
<tr>
<td>kra30a</td>
<td>Exp. (↑)</td>
<td>117,900</td>
<td>137,170</td>
<td>128,945.50</td>
<td>3,791.57</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>106,600</td>
<td>116,490</td>
<td>113,235.70</td>
<td>1,829.64</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>109,290</td>
<td>125,340</td>
<td>116,773.00</td>
<td>2,895.24</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>106,030</td>
<td>114,390</td>
<td>111,040,90</td>
<td>1,580.40</td>
<td>1.00</td>
</tr>
<tr>
<td>wil50</td>
<td>Exp. (↑)</td>
<td>33,328</td>
<td>57,808</td>
<td>53,290.36</td>
<td>679.81</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>FDTC (↑)</td>
<td>52,464</td>
<td>53,600</td>
<td>53,239.52</td>
<td>174.47</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Exp. (↓)</td>
<td>54,356</td>
<td>56,392</td>
<td>55,444.32</td>
<td>527.45</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>FDTC (↓)</td>
<td>52,512</td>
<td>54,000</td>
<td>53,466.08</td>
<td>298.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Figure 5(b) indicates that the trends are fragile and not constant.

In this paper, we have set the exponentially decreasing rates of quenching and heating as if tracing the temperature transition in Baseline. However, the above results indicate that our settings are sufficient but not optimal. In the future, we will examine how to optimize the decreasing rates without preliminary experiments. Additionally, it is necessary to consider the combination of the thresholds and decreasing rates.

5.5 Adaptivity for Uncomfortable Temperature Input

The users often input uncomfortable temperatures. For this issue, FDTC enables practical annealing also with the preset ranges a little off the comfortable range. We evaluated the solution quality under FDTC with two types of temperature ranges.

First, higher-temperature ranges than the tuned ones in Sect. 5.3 are given to the preset. The initial temperature $T_{\text{init}}$ and basic cooling rate $R$ are given as follows:

$$T_{\text{init}} = 100 \times T_{\text{max}},$$

$$R = 100 \times T_{\text{max}},$$

where $T_{\text{max}}$ is the maximum temperature in the tuned range in Sect. 5.3. In the preset range based on these settings, even the final temperature is too high to converge the spin state without quenching.

Second, lower-temperature ranges than the tuned ones in Sect. 5.3 are given to the preset. $T_{\text{init}}$ is given as follows:

$$T_{\text{init}} = T_{\text{min}},$$

where $T_{\text{min}}$ is the minimum temperature in the tuned range in Sect. 5.3. And $R$ is the same as Eq. (7). In the preset range based on these settings, even the initial temperature is too low to sustain the solution search without heating.

The results are shown in Table 5. Up and down arrows
in Schedule (Scd.) mean that the temperature ranges are higher and lower than the tuned ones in Sect. 5.3, respectively. Table 5 indicates that FDTC outperforms Exp. in all cases. Especially in cases with higher temperature ranges, the solutions obtained under FDTC all satisfy the constraints even though the CS rates under Exp. are all low.

Figure 6(a) and (b) show the transitions of the temperatures and energies in the two cases. The figures indicate that the solution searches are actually conducted at the temperatures off the preset ranges by quenching and heating.

6. Conclusion

Annealing machines are a unified way to solve real-world COPs. Most real-world COPs are classified into constrained ones with several constraints and objective functions. However, hyperparameters called temperature must be well-tuned to search for optimal solutions effectively. Conventionally, the temperature is statically scheduled from sufficiently high to sufficiently low values. The proposed FDTC checks whether the current temperature is effective by evaluating the average number of flipped spins in a series of steps, offering appropriate temperature control for constrained COPs. Quenching and heating resolve conditions where the temperature is too high and too low, respectively. The simulation result shows that FDTC adjusts the temperature appropriately and can obtain comparable or higher solution quality than the pre-tuned conventional method for every COPs. Also, the potential of FDTC to overwhelm IPA when solving huge problems is indicated.

Acknowledgment

This work was supported in part by JSPS KAKENHI Grant Number JP23H05489 and JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2112, Japan.

References


Genta Inoue received his B.S. degree from Yamaguchi University in 2022 and M.E. degree in Tokyo Institute of Technology in 2024. He is currently pursuing the Ph.D. degree with Tokyo Institute of Technology. His current research interests include Ising computing.

Daiki Okonogi received his B.S. degree in engineering from NIAD-QE with National Institute of Technology(KOSEN), Gunma College in 2021 and received his M.E. degree in information and communication engineering from the Tokyo Institute of Technology in 2023. He is currently pursuing the Ph.D. degree with the Tokyo Institute of Technology. His current research interests include Ising computing.

Satoru Jimbo received his B.E. degree in mechanics from Tohoku University in 2021 and received his M.E. degree in information and communication engineering from the Tokyo Institute of Technology in 2023. He is currently pursuing the Ph.D. degree with the Tokyo Institute of Technology. His research interests include Ising computing.

Thiem Van Chu completed his Ph.D. at Tokyo Institute of Technology in 2018. Upon graduation, he joined the School of Information Science, Japan Advanced Institute of Science and Technology as an Assistant Professor. In 2020, he moved to Tokyo Institute of Technology. His research interests lie at the intersection of computer architecture, reconfigurable computing, and machine learning.

Masato Motomura received B.S., M.S., and Ph.D. (electrical engineering) from Kyoto University, Kyoto, Japan, in 1985, 1987, and 1996, respectively. In 1987, he joined NEC Central Research Laboratories, Kawasaki, Japan, where he worked on various hardware architectures including string search engines, multi-threaded on-chip parallel processors, embedded DRAM-FPGA hybrid systems, memory-based processors, and reconfigurable systems. From 2001 to 2008, he was with NEC Electronics, Kawasaki, Japan, where he led research and business development of dynamically reconfigurable processor (DRP) that he invented. He was also a visiting researcher at the MIT Laboratory for Computer Science, Cambridge, MA, USA, from 1991 to 1992. He had been a professor at Hokkaido University, Sapporo, Japan, from 2011 to 2019. After that, he has been a professor at Tokyo Institute of Technology, Yokohama, Japan, where he is leading artificially intelligent computing (ArtIC) research unit. He is actively working on reconfigurable and parallel architectures for deep neural networks, machine learning, annealing machines, and intelligent computing in general. He is an IEEE Fellow, and a member of IEICE, IPSJ, JSAT, and EAJ. He was a recipient of the IEEE JSSC Annual Best Paper Award in 1992, the IPSJ Annual Best Paper Award in 1999, the IEICE Achievement Award in 2011, and the ISSCC Silmroad Award as the corresponding author in 2018, respectively.

Kazushi Kawamura received the B. Eng., M. Eng. and Dr. Eng. degrees from Waseda University in 2012, 2013 and 2016, respectively, all in computer science. From 2018 to 2019, he was an Assistant Professor in the Department of Communications and Computer Engineering, Waseda University. He is presently a Specially Appointed Assistant Professor in the Institute of Innovative Research, Tokyo Institute of Technology. His research interests include Ising computing and LSI design methodologies. He is a member of IEEE, IEICE, and IPSJ.