

# Non-equilibrium Monte Carlo

Masoud Mohseni,<sup>1, 2,\*</sup> Navid Anjum Aadit,<sup>2, 3, †</sup> and Aaron Lott<sup>2, ‡</sup>

<sup>1</sup>*Large Scale Integrated Photonics (LSIP), Hewlett Packard Labs, Santa Barbara, CA, USA*

<sup>2</sup>*USRA Research Institute for Advanced Computer Science (RIACS)*

<sup>3</sup>*Department of Electrical and Computer Engineering,  
University of California, Santa Barbara, CA, USA*

(Dated: October 10, 2023)

In this note, we provide a simple but general documentation for an important class of probabilistic solvers known as Non-equilibrium Monte Carlo (NMC).

## I. INTRODUCTION

Optimization, the quest to find the best solution among a set of feasible solutions, lies at the heart of many scientific, engineering, and industrial applications. The complexity of many real-world problems, especially those characterized by vast and rugged solution spaces, makes the optimization task a formidable one. For problems described by non-convex energy landscapes, the existence of numerous local minima poses a particular challenge. An algorithm aiming to find the global optimum might often get ensnared by these local minima, thereby failing to reach the most optimal solution.

Traditional optimization techniques, like simulated annealing (SA), have enjoyed considerable success in a variety of domains. SA, inspired by the annealing process in metallurgy, probabilistically accepts solutions that are worse than the current one to escape local minima. However, as problems scale up in complexity and the energy landscapes become more non-convex, the chances of SA (and similar algorithms) getting trapped increases. The main shortcoming of such techniques is their equilibrium nature, wherein they rely on a balance between exploration and exploitation, defined by a temperature parameter. As this balance is difficult to maintain across varying landscape complexities, there's an inherent limitation to how well these methods can adapt.

Monte Carlo (MC) methods, grounded in statistical mechanics, have provided a foundational framework for many optimization techniques, including SA. Yet, standard MC approaches, too, suffer from similar equilibrium constraints, making them less agile in navigating intricate energy landscapes.

In this paper, we explore the Non-Equilibrium Monte Carlo (NMC) approach, building on prior work [1]. While the previous research implemented NMC in multi-replica systems, we present both single replica and multi-replica variations. By sidestepping the equilibrium constraints and introducing mechanisms that better adapt to the underlying problem structure, NMC promises a more robust approach to complex optimization challenges. This

adaptability becomes especially evident when we integrate NMC with other techniques. Our results underscore the enhanced efficacy of NMC, particularly when faced with hard non-convex optimization challenges.

## II. NON-EQUILIBRIUM MONTE CARLO (NMC) ALGORITHM

**Motivation:** Traditional Monte Carlo methods, rooted in equilibrium principles, often falter in local minima, particularly within non-convex energy landscapes. The NMC method presents a dynamic countermeasure. By leveraging the Loopy Belief Propagation (LBP) upon pinpointing a potential local minimum, it offers a nuanced exploration of the minima. This approach emphasizes the computation of local energy landscape properties, enabling the formulation of nonlocal moves in configuration space.

### A. Understanding the Local Geometry

In the realm of discrete spaces, deriving the local geometry isn't as straightforward as merely computing derivatives. Instead, one must estimate specific measures, most notably the first, second, or even higher-order marginals, which manifest as magnetizations and correlations. Here, a surrogate variable  $r$  comes into play. Initialized to the detected state  $s^*$ , this surrogate variable serves as a diagnostic tool to probe the nuances of the local energy landscape.

### B. Constructing the Surrogate Hamiltonian

The surrogate Hamiltonian, represented as  $H_\epsilon(r)$ , embodies the mechanism that influences the evolution of the variable  $r$ . Formally, it's articulated as:

$$H_\epsilon(r) = H(r) - \lambda \sum_i \epsilon_i s_i^* r_i \quad (1)$$

Where  $\epsilon_i$  is defined as  $|h_i| + \sum_j |J_{ij}|$  and  $\lambda$  acts as a scaling parameter. The architecture of this Hamiltonian ensures the locality of sampling over crucial vari-

\* masoud.mohseni@hpe.com

† maadit@ucsb.edu

‡ plott@usra.edu

ables, particularly in problems characterized by highly heterogenous graph topologies.

### C. LBP Execution for Magnetizations & Correlations

Leveraging the Loopy Belief Propagation (LBP), we can infer values of magnetizations and correlations. For the surrogate Hamiltonian, this is crucial to ensure efficient sampling. We begin by enforcing a stringent locality on the surrogate Hamiltonian. Every Hamiltonian is tied to a basin of attraction characterized by  $s^*$ , scaled by an inhomogeneous vector  $\epsilon$ , with a suitably large magnitude  $\lambda$ . Each  $\epsilon_i$  for every site is defined as  $\epsilon_i = |h_i| + \sum_j |J_{ij}|$ . This ensures that our initial epsilon dominates the energy scale of the site, providing a locality to our surrogate Hamiltonian on vital variables.

For the initial LBP messages, we utilize the definitions  $h_{i \rightarrow j}$  and  $u_{i \rightarrow j}$ , iteratively updating them as  $\lambda$  is adjusted. The LBP equations that drive this process are:

$$h_{i \rightarrow j} = h_i + \lambda \epsilon_i s_i^* + \sum_{k \in \partial i \setminus j} u_{k \rightarrow i} \quad (2)$$

$$u_{i \rightarrow j} = \beta^{-1} \operatorname{arctanh} [\tanh(\beta J_{ij}) \tanh(\beta h_{i \rightarrow j})] \quad (3)$$

From which magnetizations and correlations can be deduced:

$$\langle r_i \rangle = \tanh \left[ \beta \left( h_i + \lambda \epsilon_i s_i^* + \sum_{j \in \partial i} u_{j \rightarrow i} \right) \right] \quad (4)$$

$$\langle r_i r_j \rangle = \frac{\tanh(\beta J_{ij}) + \tanh(\beta h_{i \rightarrow j}) \tanh(\beta h_{j \rightarrow i})}{1 + \tanh(\beta J_{ij}) \tanh(\beta h_{i \rightarrow j}) \tanh(\beta h_{j \rightarrow i})} \quad (5)$$

### D. Effective Interactions and Rigidity Rankings

Post magnetization and correlation estimation, effective interactions amongst variables are gauged, allowing the subsequent ranking of rigid variables. Using this hierarchical ranking, we can then define a correlation threshold cutoff, vital for determining surrogate backbones. These backbones, in turn, influence the iterative two-temperature Markov chain process, optimizing our traversal through the energy landscape.

To understand the core nature of rigid variables and their interactions, our methodology uses LBP sampling. With these messages in place, our algorithm embarks on growing clusters of connected variables or spins. These clusters, envisioned as droplet-like excitations in the context of the original spin-glass problem, are developed around “seed” variables. These seeds are variables with effective couplings exceeding a specified seed correlation threshold. From each seed, a cluster evolves, including neighboring spins with effective couplings that are above

a predetermined correlation threshold cutoff. The process is iterative, consuming spins with strong correlation until no further additions are feasible. The process then pivots to the next available seed.

It’s crucial to highlight the significance of the correlation threshold cutoff. Acting as a linchpin, it determines the surrogate backbones’ dimension and form, subsequently influencing the efficiency of cluster updates. Empirical investigations underscore the following insights about this threshold:

- There exists a viable correlation threshold value ensuring surrogate Hamiltonian backbones have substantial sizes without causing percolation.
- Within its acceptable range, the value of the correlation threshold remains robust.
- Optimal correlation thresholds find correct surrogate backbones that enable nonlocal cluster moves, augmenting the sampling efficacy for a single MCMC replica.

### E. Iterative Two-Temperature Exploration

The quintessence of our approach in the NMC algorithm lies in the non-equilibrium inhomogeneous sampling over specific subgraphs, facilitating a rapid convergence towards the relevant low-energy states in the configuration space. Within a single replica, operating at a predetermined temperature  $\beta$ , the strategy is meticulously orchestrated.

Utilizing insights from Loopy Belief Propagation (LBP), we identify clusters or backbones of rigid variables tethered within a particular basin of attraction. Once discerned, these backbones undergo a thermal cycle, with their temperature typically elevated by a factor ranging between 2 to 10 compared to the ambient system (keeping non-backbone variables fixed). This selective thermal perturbation is designed to liberate the system from potential local minima, enhancing the exploration of novel configurations.

Subsequent to this heating cycle, the exterior variables—those outside the backbone—are updated, maintaining the backbone in its perturbed state. This ensures that the broader system configuration commences alignment with the prevalent energy landscape. Finally, a holistic update is executed, encompassing both the backbone and its exterior, culminating in a coherent progression of the system’s state.

Reiterative cycling of this adaptive sampling strategy ensures comprehensive exploration of the configuration space. While each iteration may not strictly adhere to global detailed balance conditions, the cumulative effect through iterations ensures robust sampling of the relevant low-energy states. Thus, our methodology efficaciously navigates the energy landscape, achieving an approximation of the system’s partition function without necessitating strict global detailed balance conditions.

## F. Family of Algorithms and Variants of NMC

The Non-equilibrium Monte Carlo (NMC) methodology exhibits immense versatility and adaptability. The foundational principles of NMC can be integrated seamlessly into an array of optimization algorithms, thereby enhancing their efficacy. Below, we elucidate NMC’s potential integration with adaptive parallel tempering (APT):

### 1. NMC in Adaptive Parallel Tempering (NPT)

Parallel Tempering (PT) [2–4] is a powerful sampling technique that facilitates exploration over a broad energy landscape by simulating replicas of the system at different temperatures. The higher temperature replicas can easily surmount barriers between local minima, while the colder ones give accurate statistics around the global minimum. A temperature schedule that maintains the fixed swap probability independent of the replica’s temperature is efficient [5]. A number of methods have been developed to construct such a schedule adaptively [6–9]. Adaptive Parallel Tempering (APT) [1, 10] fine-tunes this approach by adjusting the temperatures dynamically based on the performance of the replicas, often making the search more efficient.

However, the colder replicas in APT often get trapped in local minima due to the low temperatures, making their contributions less significant over time. This is where the Non-Equilibrium Monte Carlo (NMC) method can make a substantial impact.

By incorporating NMC into the APT framework, we aim to rejuvenate the colder replicas, thus ensuring they remain active and productive. The integration is achieved through the following steps:

1. **Monitoring Stage:** Track the progress of each replica. If a colder replica remains stagnant (i.e., trapped in a local minimum) for a predetermined number of iterations, flag it for an NMC intervention.
2. **Invocation of NMC:** Once flagged, pause the standard APT updates for the specific replica and initiate the NMC cycles.
  - (a) **Exploration Phase:** Apply the first cycle of NMC, which involves identifying the backbone structures using techniques like LBP and then heating up only the backbone, keeping the rest fixed. This allows for localized, high-temperature movements without disturbing the entire structure.
  - (b) **Exploitation Phase:** Here, the outside of the backbone is updated while the backbone remains frozen. This helps in refining the solution and moving towards more energetically favorable configurations.

(c) **Unlearning Phase:** The entire structure is then allowed to update freely, ensuring any artificially induced biases from the first two phases are minimized.

3. **Frequency of Invocation:** To maintain a balance between exploration and exploitation, it’s crucial to control how frequently NMC is invoked on the colder replicas. Too frequent invocations might overheat the system, leading to loss of detailed balance, while too rare invocations won’t be beneficial in terms of escaping local minima. Empirical tuning might be necessary based on the specific problem at hand.
4. **Reintegration:** After the NMC cycles are completed, reintegrate the replica into the standard APT framework and continue the iterative process.

In essence, by leveraging the local exploration capabilities of NMC within the global structure of APT, we can ensure that even the coldest replicas remain diverse and active, thereby enhancing the overall efficiency and effectiveness of the sampling process.

### 2. Comparisons with Isocluster Move (ICM)

The Isocluster Move (ICM) technique offers an interesting contrast to the NMC strategy. At the heart of ICM lies the identification of disagreeing clusters — groups of variables that diverge in their configurations across different replicas. Once identified, these clusters are strategically swapped among replicas, promoting better exploration and potentially facilitating the system to escape local minima. While the underlying philosophy of enhancing exploration is shared with NMC, the methodologies diverge in their approach. While NMC focuses on perturbing the system using LBP to escape local configurations, ICM seeks to capitalize on the diversity across replicas by judiciously swapping clusters. Benchmarking NMC against ICM not only provides insights into the relative strengths and efficiencies of both methodologies but also opens the door to potential hybrid techniques that might amalgamate the best of both worlds.

In summary, the NMC’s inherent flexibility allows it to be seamlessly integrated into a myriad of optimization and sampling techniques. Each integration, while preserving the core principles of NMC, offers a tailored approach to tackle specific challenges posed by the parent algorithm, making the NMC a versatile tool in the optimization toolbox.

Our empirical assessment drew a comparative analysis across four computational techniques: Simulated Annealing (SA), Non-Equilibrium Monte Carlo (NMC), Adaptive Parallel Tempering (APT), and fixed- $\beta$  MCMC. The derived approximation ratio is computed from the best energy found with the particular algorithm with respect to the putative ground state.

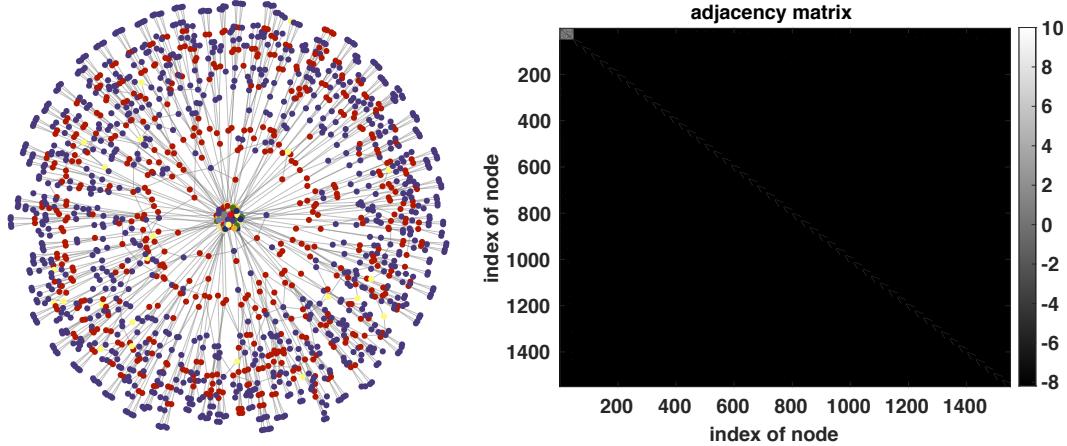


Figure 1. The graph and adjacency matrix representation of the 1550-node crafted backbone problem.

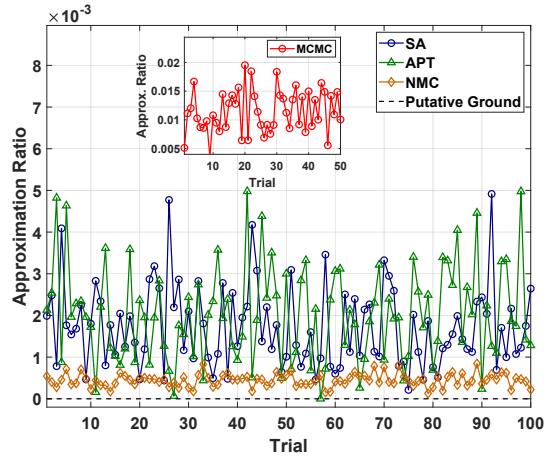


Figure 2. Performance comparison of SA, NMC, APT, and fixed- $\beta$  MCMC on the crafted backbone problem. The inset highlights the suboptimal performance of MCMC in terms of energy approximation ratio.

### III. RESULTS

#### A. Benchmarking on a Crafted Backbone Problem

We investigated an illustrative example consisting of 1550 nodes, structured to manually feature a backbone using a Wishart [11] instance (Figure 1). The core of this structure, formed by the initial 50 nodes, is all-to-all. Connections within this backbone were intensively weighted, being ten times stronger (10X) than those of the external tree structure, which were given 1X weights. To add more intricacies to the topology, random tree leaves were interconnected using these standard 1X weights.

Our experimental evaluations incorporated Simulated Annealing (SA), Non-Equilibrium Monte Carlo (NMC), Adaptive Parallel Tempering (APT), and fixed- $\beta$  MCMC, as

well as two hybrid techniques: APT combined with NMC (NPT) and APT combined with ICM (APT + ICM).

Figure 2 unveils that the NMC method consistently outstrips the other techniques. This can be attributed to its innovative feature that facilitates effective navigation through the energy landscape, particularly adept at evading local minima. Simulated Annealing (SA), given its inherent adaptability in traversing intricate energy landscapes, manifests itself as a robust second-best. In contrast, the APT's performance seems hampered, potentially due to the fragmented nature of its approach, dividing its computational strength across various replicas, thereby diluting the individual impact of each. The results of the fixed-temperature MCMC are delineated in the inset of the figure due to their egregiously poor performance, driving them out of the standard approximation ratio range. Meanwhile, noteworthy is the performance of NMC, which, with only two temperatures and a single replica, significantly outperforms the other techniques, highlighting its efficiency and effectiveness in navigating the energy landscape.

#### B. Results on Deceptive Cluster Loop (DCL) Problem

For a more sophisticated benchmark, we considered the Deceptive Cluster Loop (DCL) problem [11], using planted Ising instances mapped onto a Chimera graph of 1865 nodes. The nature of the DCL problem inherently embeds a non-trivial backbone, which was dynamically discerned using the LBP technique.

From Figure 4, it becomes evident that in this instance, SA is the most effective at navigating the problem's deceptive landscape. NMC, using its adaptability to the dynamically detected backbone, delivers commendable performances in generating high-quality solutions. APT, consistent with our earlier benchmarks, is somewhat limited in its capabilities, mainly due to its computational

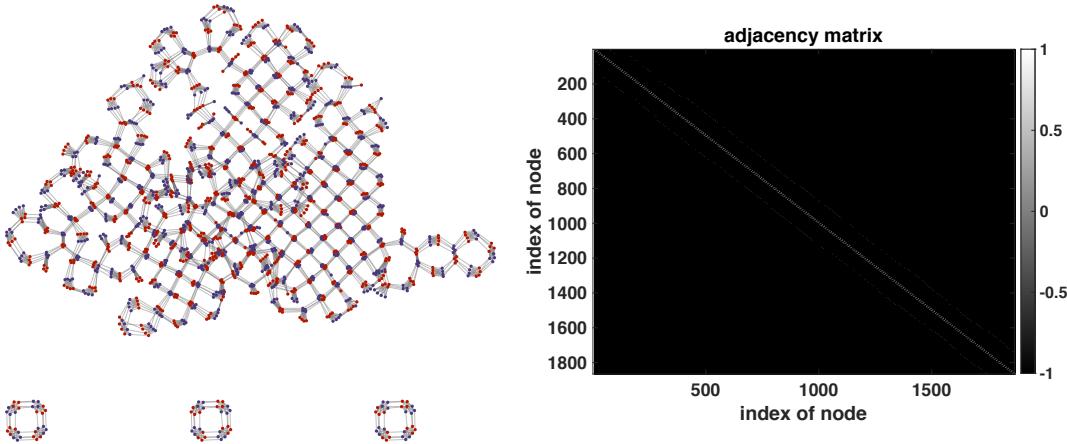


Figure 3. Graph and adjacency matrix representation of the 1865-node Chimera graph for the DCL problem.

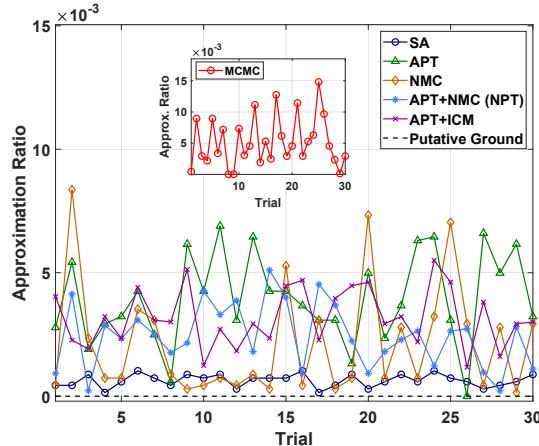


Figure 4. Performance comparison of SA, NMC, APT, fixed- $\beta$  MCMC, APT + NMC (NPT), and APT + ICM on the DCL problem.

resources being spread across multiple replicas. However, enhancing APT with NMC and ICM does show marked improvements in the outcomes. Both NPT and APT + ICM outshine the standalone APT in their solution qual-

ity. Nevertheless, it's critical to highlight that APT + ICM necessitates a more considerable number of replicas compared to NPT, which can affect its scalability and computational efficiency. Moreover, the MCMC method, due to its fixed- $\beta$  nature, falters in this problem too, underlining its difficulties in managing the DCL problem's intricate energy landscape.

#### IV. CONCLUSION

The study showcased the effectiveness of the Non-Equilibrium Monte Carlo (NMC) approach across different problem instances. In the crafted backbone scenario where the backbone was well-defined, NMC showed superior performance. For the Deceptive Cluster Loop (DCL) problem with a non-trivial backbone identified dynamically through LBP, NMC still held its ground as a competitive approach. The integration of Adaptive Parallel Tempering (APT) with NMC also yielded promising results, underscoring the potential of combining different techniques to tackle complex problems effectively. These findings highlight the versatility and robustness of NMC in addressing various computational challenges.

- 
- [1] M. Mohseni, D. Eppens, J. Strumpfer, R. Marino, V. Denchev, A. K. Ho, S. V. Isakov, S. Boixo, F. Ricci-Tersenghi, and H. Neven, [arXiv:2111.13628](https://arxiv.org/abs/2111.13628) (2021).
  - [2] R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
  - [3] C. J. Geyer, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (American Statistical Association, 1991) p. 156–163.
  - [4] K. Hukushima and K. Nemoto, *Journal of the Physical Society of Japan* **65**, 1604 (1996), <https://doi.org/10.1143/JPSJ.65.1604>.
  - [5] A. Kone and D. A. Kofke, *The Journal of Chemical Physics* **122**, 206101 (2005), <https://doi.org/10.1063/1.1917749>.
  - [6] D. A. Kofke, *The Journal of Chemical Physics* **117**, 6911 (2002), <https://doi.org/10.1063/1.1507776>.
  - [7] N. Rathore, M. Chopra, and J. J. de Pablo, *The Journal of Chemical Physics* **122**, 024111 (2005), <https://doi.org/10.1063/1.1831273>.
  - [8] C. Predescu, M. Predescu, and C. V. Ciobanu, *The Journal of Chemical Physics* **120**, 4119 (2004), <https://doi.org/10.1063/1.1644093>.

- [9] C. Predescu, M. Predescu, and C. V. Ciobanu, *The Journal of Physical Chemistry B* **109**, 4189 (2005), pMID: 16851481, <https://doi.org/10.1021/jp045073+>.
- [10] N. A. Aadit, M. Mohseni, and K. Y. Camsari, in *2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)* (IEEE, 2023) pp. 1–2.
- [11] D. Perera, I. Akpabio, F. Hamze, S. Mandra, N. Rose, M. Aramon, and H. G. Katzgraber, *arXiv preprint arXiv:2005.14344* (2020).

## Appendix A: Procedure of NMC Algorithm for Single Replica

The Non-equilibrium Monte Carlo (NMC) method provides a refined and adaptive approach to traditional Monte Carlo simulations, offering an optimized strategy to traverse complex energy landscapes. This is achieved by dynamically identifying and manipulating localized regions of the configuration space while continuing to explore more global structures. The detailed procedure of the NMC strategy is delineated in Algorithm 1.

### Algorithm 1: NONEQUILIBRIUM MONTE CARLO (NMC) for Single Replica

```

while good solutions not found do
    1. MCMC sampling over the entire problem.
    2. Locate a seed solution:
        • Generate seed: Find a low-energy state as a
          seed solution,  $s^*$ 
        • Construct a localized surrogate Hamiltonian
          around the neighborhood of  $s^*$ 
        • Use LBP to estimate marginals and/or
          correlations over the surrogate problem
        • Threshold the obtained values to identify
          surrogate backbones of rigid variables
    3. Sampling updates:
        while not in a steady state do
            • Non-local updates: Perform
              inhomogeneous MCMC on the backbone at
              elevated temperatures, conditioning on
              non-backbone variables
            • Local updates: Update non-backbone
              variables at original temperature,
              conditioning on the backbone
            • Integration: Perform MCMC sampling on
              the entire problem at original temperature
              to smooth out any inconsistencies
        end
    end

```

The strength of NMC lies in its adaptability and fine-grained exploration, ensuring both global and localized optimizations. By dynamically manipulating subregions of the configuration space based on localized information, the NMC method avoids pitfalls such as getting trapped

in local minima or diffusing too broadly, making it a robust and powerful technique for challenging optimization scenarios.

The NMC (Non-Equilibrium Monte Carlo) algorithm is devised to optimize the state  $s$  at a given  $\beta$  value.

**1. Standard MCMC step:** Perform MCMC with  $10^5 - 10^7$  steps, iterating  $s$  to reach a steady state  $s^*$ .

**2. Polarization via LBP for semi-localized solution:**

- Start with  $\lambda = \infty$  and compute solutions of  $h_{ij}$  and  $u_{ij}$ .
- Initialize  $\lambda$  with  $\lambda_0 = 2 - 10$ .
- Adjust  $\lambda$  incrementally and update the LBP messages. Evaluate stability using a threshold  $x$ .
- Compute polarization  $\langle r_i \rangle$  to determine spins that are involved in the local minimum.

**3. Discovering backbones:**

- Rank polarization by magnitude. Choose a number of seeds  $N_s$  based on the problem and temperature.
- Grow backbones around these seeds, either autonomously based on predefined criteria or using optimized hyperparameters.

**4. Non-equilibrium annealing:**

- Perform iterations that involve:
  - (a) MCMC with heated backbone and fixed non-backbone.
  - (b) MCMC with fixed backbone.
  - (c) Standard MCMC over the entire state.
- If the energy of the resulting state is lower, store it.
- Monitor changes in state using Hamming distance.

## Appendix B: Parameter Summary and Flow for Single Replica NMC

**1. Energy Normalization:** Normalize energy with  $J_{ij} \approx 1$ , thus a characteristic  $\beta$  is 1.

**2. Localization via  $\lambda$ :**

- Input: Seed state  $s^*$  and its corresponding  $\beta$ .
- Output: Magnetization  $\langle r_i \rangle$ .
- Data: Distribution plots of  $\langle r_i \rangle$  and its correlation with  $\beta$ . The plots provide a range of usable  $\beta$  values.

**3. Backbone Identification:**

- Input: Magnetization  $\langle r_i \rangle$ .
- Input: Thresholds  $\alpha_{\text{seed}}$  and  $\alpha_{\text{th}}$  to separate data into two distinct groups. Typically,  $\alpha_{\text{seed}} \approx 30\%$  with a range from 20% to 45%.
- Output: Identified backbone.
- Data: Ratio plots of the backbone size to the total system size, ensuring a broad range of backbones from 1% to as large as 90% of total variables is captured.

#### 4. Inhomogeneous Annealing:

- Input: Identified backbone (and non-backbone) spins.
- Input: High temperature and low temperature values for annealing.
- Strategy:

- Execute MCMC updates with a boosted temperature for the backbone variables while conditioning on non-backbone variables.
- Run another MCMC for the non-backbone variables at the original  $\beta$  of the replica.
- Perform MCMC sampling on the entire problem at the original temperature to reconcile potential discrepancies and ensure a harmonious integration of the backbone and non-backbone regions.
- Output: Updated state  $s^*$  and its energy.
- Data: Efficiency of annealing as a function of high-temperature value. This could provide insights into the energy scale of the problem.