Electronics Letters

LETTER OPEN ACCESS

Low Power CMOS Stochastic Bit Based Ising Machine and Its Application to Graph Coloring Problem

Honggu Kim | Dongjun Son | Yerim An | Yong Shim

Department of Intelligent Semiconductor Engineering, Chung-Ang University, 84, Heukseok-ro, Seoul, Dongjak-gu, Republic of Korea

Correspondence: Yong Shim (yongshim@cau.ac.kr)

Received: 18 October 2024 | **Revised:** 5 March 2025 | **Accepted:** 17 March 2025

Funding: This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (No. 2021R1C1C100875214). This research was supported by the National Research Council of Science & Technology (NST) Grant by the Korea government (MSIT) (No. GTL24041-000). The EDA tool was supported by the IC Design Education Center (IDEC), Korea.

ABSTRACT

The Ising spin model is an efficient method for solving combinatorial optimization problems (COPs) but faces challenges in conventional Von-Neumann architectures due to high computational costs, especially with the growing data volume in the IoT era. To address this problem, we proposed low power CMOS stochastic bit based Ising machine to efficiently compute COPs. By adopting compute-in-memory (CIM) approach for parallel spin computation, we achieved energy efficient spin computing. Furthermore, we harnessed the inherent randomness of CMOS stochastic bit to prevent Ising computing process from being stuck into local minima, effectively mitigating the power penalty associated with the random number generators (RNGs) in the conventional CMOS based Ising machines. We demonstrated the feasibility of our design by solving NP-complete graph coloring problem with four vertices and three colors using TSMC 65 nm GP process. Moreover, the proposed CMOS stochastic bit based spin unit consumes the lowest power/spin among the state-of-the-art Ising machine researches, with power/spin of 1.07 μW and energy/spin of 107 fJ.

1 | Introduction

Combinatorial optimization problems (COPs) come into great use in a wide range of applications such as supply chain optimization, VLSI physical design and transportation. Most of COPs are non-deterministic polynomial-time hard or complete, demanding exponential computational resources with respect to its computing dimension [1]. With the advent of a flood of information, the computing energy required to solve COPs within the conventional Von-Neumann architecture has skyrocketed, necessitating a shift to new computing paradigm to enable its energy efficient application [2].

In this context, several Ising machines have been developed to accelerate Ising computation. Quantum annealing processor utilizing quantum tunneling effect of superconductor for spin updates has been demonstrated in [3]. However, the practical application of quantum annealing faces significant challenges, such as extremely low temperature operating condition and excessive power consumption. For these reasons, low-power CMOS based Ising machines have been researched to achieve practical and energy efficient spin computing acceleration [4–6].

CMOS based Ising machines generally employ compute-in-memory (CIM) scheme to enhance the energy efficiency that was hindered by the memory bandwidth bottleneck, and facilitate massive parallel computing of spin interactions for fast convergence to global minimum energy. Furthermore, to realize the escape from convergence to local minimum, they mostly adopted off-chip random number generators (RNGs) [5, 6]. However,

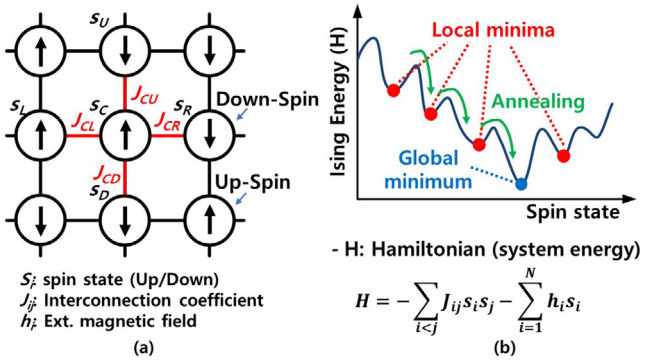

FIGURE 1 | (a) Basic topology of Ising model and (b) system energy versus spin state.

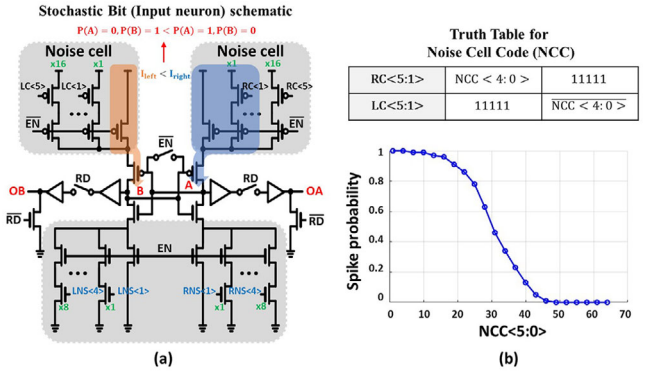

FIGURE 2 | (a) Stochastic bit and (b) output pulse probability of stochastic bit depending on the input PMOS noise cell codes (LCs and RCs).

when these RNGs are integrated on-chip, they introduce substantial power overhead, posing a significant challenge in low-power CMOS circuit design.

In this work, we proposed a low-power CMOS stochastic bit based architecture for an energy-efficient Ising machine, optimized for Ising spin computing. The stochastic bit used in this work serves a dual purpose: (1) as a memory element, and (2) as a stochastic computing element, enabling the efficient implementation of compute-in-memory (CIM) schemes. Furthermore, the stochastic bit leverages an energy-efficient latch operation to effectively eliminate additional power demands that may be introduced by the random number generators (RNGs) in conventional CMOS-based Ising machines.

2 | Ising Spin Model and Annealing Process

Figure 1a shows the basic topology of Ising model, which performs spin operations based on magnetic spin physics to lower the system energy and find the optimal solution for NP problems. s_i denotes the spin state at location i, J_{ij} represents the interaction coefficients of adjacent spins s_i and s_j, and h_i signifies the external magnetic field which works as a local bias of a spin at location i. Over time, each spin continuously interacts and updates its state based on the current spin configuration and interaction coefficients, aiming to minimize the system energy H depicted in Figure 1b. The first term of the system energy H in Figure 1b represents the total interaction energy between spins, while the second term accounts for the sum of local energies due to external fields trying to align them in a specific direction. As the system energy converges towards its ground state, the model reaches a global optimal solution. The operational dynamics of the Ising spin model are as follows:

- Adjacent Ising spins perform majority vote to determine the state of a single Ising unit.
- If $N_s > N_o$, the spin state is set to +1; if $N_s < N_o$, the spin state is set to −1; if $N_s = N_o$, the spin state is set to either +1 or +1 with the equal probability.

where N_s represents the number of neighboring spin pairs in the same orientation (i.e., having the same sign), and N_o denotes the number of neighboring pairs with opposite orientations (i.e., opposite signs).

However, as shown in Figure 1b, the overall system is susceptible to becoming trapped in local minima. To mitigate this problem, random fluctuations are introduced into the Ising spin computation process, allowing the system to escape local minima and continue converging toward the global optimum solution, reducing the likelihood of the premature system convergence.

3 | Stochastic Bit for Random Error Bits Insertion

To address the challenge of escaping local minimum energy states during the Ising computation process, recent studies have commonly employed off-chip RNGs to introduce random fluctuations. However, when these RNGs are integrated on the same CMOS wafer, it may contribute to high power penalty of the overall system. To mitigate this issue, we utilized the low-power CMOS-based stochastic bits, developed from prior research [7], as a source of randomness.

Figure 2a illustrates the schematic of stochastic bit. This circuit primarily utilizes two sources of randomness: (1) the thermal noise inherent in CMOS devices, and (2) the meta-stability of cross coupled inverters when their input and output are shorted. Moreover, the output pulse probability of this Stochastic bit changes with the asymmetric intensity of PMOS Noise cell codes, consisting of Left Codes (LCs) and Right Codes (RCs) applied to the left and right PMOS noise cells, respectively.

Specifically, an increase in the magnitude of RCs with all LCs deactivated results in an elevated output spike generation rate. Conversely, an increase in the magnitude of LCs with all RCs deactivated, leads to a reduction in the output spike generation rate. The modulation of the spike generation rate is determined in a sigmoid manner with respect to the code sets (LCs and RCs), as illustrated in the graph in Figure 2b. The truth table of the code set (LC and RC) is shown in the upper side of Figure 2b. By appropriately selecting the PMOS noise cell code sets, it is possible to introduce a desired level of random fluctuations into the Ising computation process.

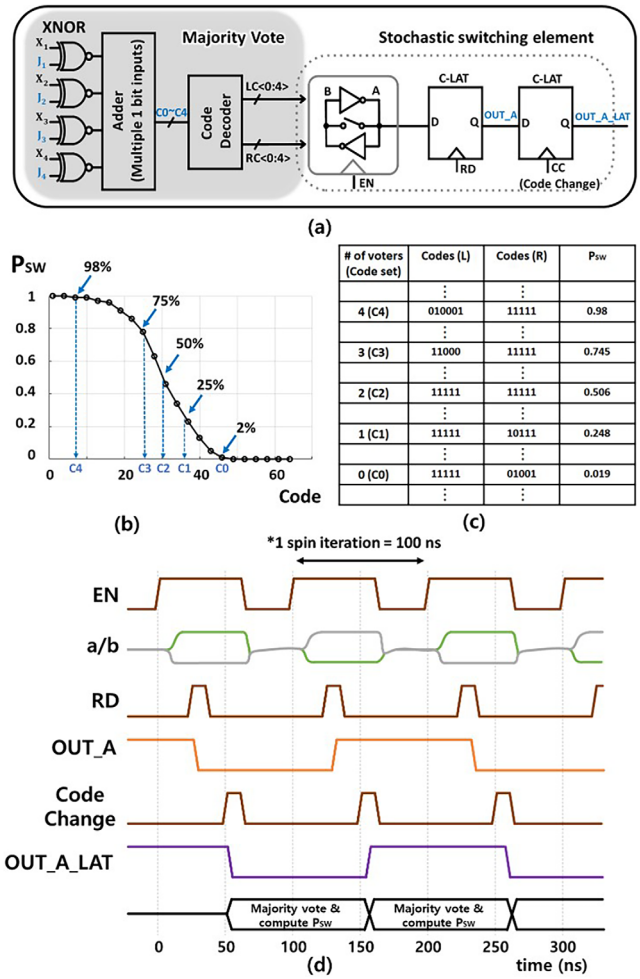

FIGURE 3 | (a) Single Ising unit with XNOR based majority vote logic and code decoder, (b) selected code from majority vote based on adjacent spin interactions, (c) entire truth table of switching probability of stochastic bit and (d) operational timing diagram of single Ising unit.

4 | Proposed Stochastic Bit Based Ising Spin Unit

Figure 3a depicts the structure of a single Ising unit employing a stochastic bit for both purposes: (1) storing spin state and (2) spin computing with the incorporation of random fluctuations. The spin state (A) is stored within the cross-coupled inverter at each spin iterations. Additionally, C latches are utilized to deliver the current spin state to adjacent spins to compute the next spin state. The majority vote circuit consisting of XNOR gate, adder, and code decoders determines the next spin state based on the current spin states of the adjacent spins.

Each XNOR gate performs the bit-wise multiplication of adjacent spin states (X_i) with their corresponding interaction coefficients (J_i), and the multiplied results are summed through the adder logic, resulting in C_i. When C_i is high (e.g., 4), then certain code sets (LC and RC) is selected via the code decoder logic to configure the stochastic bit for generating a high-probability output pulse. Conversely, when the C_i is low (e.g., 1), the code decoder logic selects the code set (LC and RC) that decreases the probability of the stochastic bit generating an output pulse.

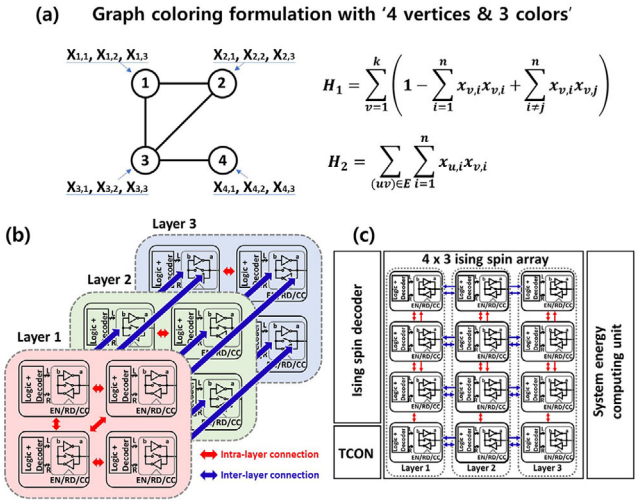

FIGURE 4 | (a) Graph coloring problem formulation with four vertices and three different colors and its system energy, (b) 3-dimensional integrated 2-dimensional Ising network construction based on the stochastic bit Ising unit and (c) realization of 3-dimensional Ising network within CMOS layout via 2D array flattening.

As illustrated in Figure 3b, a predefined set of biased switching probabilities based on each code set (LC and RC) are established, where the switching probabilities increase in proportion to the majority vote result. Figure 3c presents the corresponding truth table for the switching probabilities of the stochastic bit, based on the summed majority votes (C_i). For a high majority vote (e.g., C3), the selected code set configures a higher switching probability (e.g., $P_{SW} = 0.745$), whereas a low majority vote (e.g., C1) results in a lower switching probability (e.g., $P_{SW} = 0.248$).

These code sets (LCs and RCs) are stored in code decoder logic using D flip-flops, and one code bundle among them is uploaded to PMOS noise cell of stochastic bit based on the result of majority vote from adjacent spins. Finally, next spin computation is proceeded based on this uploaded code set. Figure 3d shows the operational timing diagram of a single Ising unit. The Ising unit computes spin state (A) based on the majority vote result at EN signal, reads out and stores this spin state in the first C latch (OUT_A) at the RD signal, and subsequently fetches this spin state in the second C latch (OUT_A_LAT) at Code Change (CC) signal to use it for interactions with adjacent spins at the next spin operation step. The synchronous three-phase operations (EN, RD, and CC) are critical in maintaining consistent operational speed for the Ising spin machine, even as the array size scales to accommodate larger NP problem instances.

5 | Application of Stochastic Bit Based Ising Machine to Graph Coloring Problem

Figure 4a shows an example of Ising model application to NP complete graph coloring problem, with four vertices and three colors (R,G and B) and the corresponding system energy formula. The spin states for the three colors R, G, and B at vertex location i are denoted as $X_{i,1}$, $X_{i,2}$, and $X_{i,3}$, respectively. The first term H_1 imposes a constraint ensuring that each vertex possesses only one color at a given time, while the second term H_2 enforces

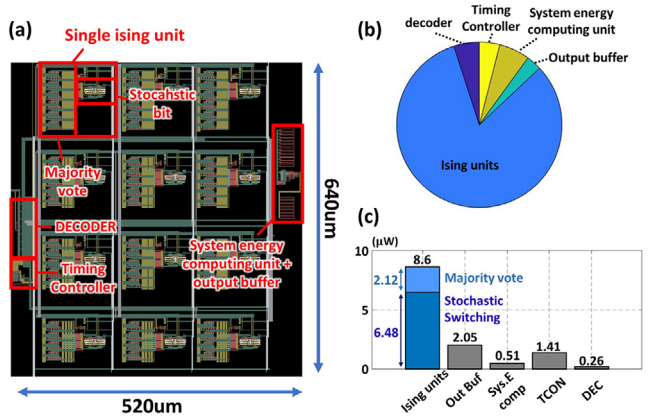

FIGURE 5 | (a) Overall layout of proposed CMOS stochastic bit based insing machine, (b) area breakdown and (c) power breakdown

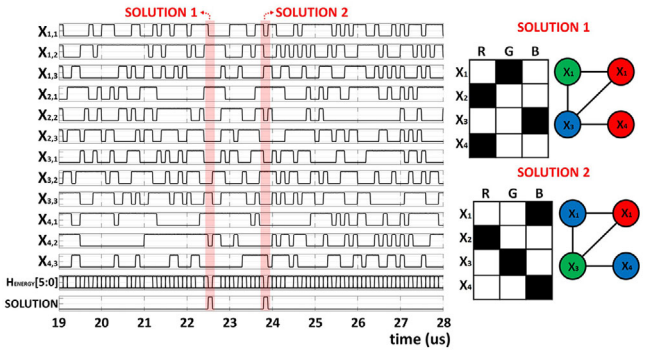

FIGURE 6 | Post-layout simulated result of the proposed architecture, applied to a graph coloring problem with four vertices and three colors.

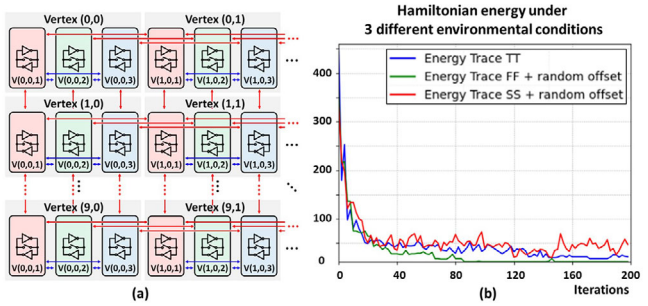

FIGURE 7 | (a) Scaling methodology of the proposed stochastic bit based Ising machine to a 10 × 10 × 3 spin network and (b) software-emulated Hamiltonian energy traces, accounting for process variations and random device mismatch.

the constraint that every adjacent vertices must have distinct colors, with n denoting the total number of colors. Violations of these constraints incur energy penalties, making adjustments to the spin states to align with the specified constraints. Through the multiple operation cycles, the system minimizes its energy, ultimately achieving the ground state where the system energy is zero. This ground state indicates the valid solution for the graph coloring problem, where all constraints are satisfied.

To map the four vertices and three color variables to the Ising model, we constructed a three-dimensional Ising network, as depicted in Figure 4b. In one layer, the four vertices are interconnected according to the graph connectivity in Figure 1a, with each layer dedicated to one of the three colors: R, G, and B. This network architecture incorporates spin connections for the spin interactions and enforce energy penalties based on the energy formula in Figure 4a. To implement this 3D Ising network on CMOS hardware, we flattened the 3D structure into a 2D layout, as shown in Figure 4c. Spin decoder logic is utilized to select and upload code sets (LCs and RCs) for each Ising spin unit. These units are interconnected within a 4 x 3 Ising spin array, maintaining the topology depicted in Figure 4b. Additionally, a system energy computing unit is employed to calculate the overall system energy. Note that the interconnection coefficients are determined analytically using the Hamiltonian energy formulations (H_1 and H_2) [1], ensuring precise alignment with the energy minimization principles of the Ising model.

6 | Performance Analysis

The entire system was designed using the TSMC 65 nm GP process, operating with a 0.65 V supply voltage (V_{DD}) and a main clock frequency (f_{CK}) of 50 MHz. As depicted in Figure 5a, the system comprises 12 stochastic bit based Ising units, decoder logic, timing controller, system energy computing unit and output buffer. Figures 5b and 5c provide breakdowns of the area and power consumption of the implemented system, respectively. The Ising units occupy the largest portion of both total area and power consumption, primarily due to their deployment as a uniform array. Furthermore, we designed a single Ising unit to occupy redundant silicon space to mitigate parasitic coupling effects among individual Ising units and thereby reduce degradation in computing accuracy. Ongoing research aims to optimize the dimensions of the single Ising unit further, thereby minimizing the overall area consumed by the Ising spin array while maintaining performance integrity.

Figure 6 shows the post-layout simulated results of the proposed Ising machine, incorporating stochastic bit based Ising units, applied to a graph coloring problem with four vertices and three colors. Throughout the multiple operation cycles, the spin states undergo continuous flipping, driven by the majority vote from adjacent spin states. Furthermore, despite the initial convergence to a single solution, the system dynamically moves away from this energy minimum state and continues to explore the alternative energy minimum state, leveraging the stochastic characteristics of the stochastic bit.

To evaluate the scalability of the proposed stochastic bit based Ising spin machine for larger NP problem instances, the behavior of the stochastic bit based Ising unit was emulated within a 10 × 10 × 3 Ising spin network at the software level, as depicted in Figure 7a. The stochastic operational characteristics of the 'stochastic bit' were modeled using error CDF fitting, as detailed in prior work [8]. Figure 7b illustrates the Ising spin computing results, highlighting the Hamiltonian energy trajectories converging efficiently toward the ground-state energy. Furthermore, Figure 7b includes results for two additional scenarios—Ising computing under (1) ff and (2) ss process corner variations with

TABLE 1 | Comparison table.

	Naute'11 [3]	CICC'21 [5]	ISSCC'22 [6]	ISSCC'23 [4]	This work	
Technology Node	N/A	65 nm CMOS	65 nm CMOS	65 nm CMOS	65 nm CMOS	
Supply voltage (V)	N/A	0.6	1.2	0.75–1.05	0.65	
Main clock freq. (MHz)	N/A	64	64	N/A	50	
Spin type	Qubit	Register	Register	CMOS latch	CMOS latch	
Source of randomness	N/A	Off-chip RNG	Off-chip RNG	Latch equalization	Thermal noise, latch equalization	
Active spin area (μm^2)	N/A	1671	1320	310	[e]1300	
Spin latency (ns)	N/A	N/A	625	N/A	100	
Ising topology	Chimera	Sparse graph	Lattice graph	Lattice graph	Sparse graph	Lattice graph
# of spins	8	252	256	1,440	12	300
Power/spin	[a]12.2 W	[a,d]1.33 μW	[a,d]1.44 μW	N/A	[b]1.07 μW	[c]1.15 μW
Energy/spin (fJ)	N/A	N/A	[a,d]712.5	[a]226	[b]107	[c]115

[a]Measurement based.
[b]Post layout simulation based.
[c]Software-level emulation based.
[d]Off-chip RNG power excluded.
[e]Assuming Layout optimization, accounting for CMOS device area only.

random device mismatch conditions—both demonstrating consistent convergence to the ground-state energy, thereby validating the robustness of the Ising computing framework across diverse operational conditions.

Table 1 presents a performance comparison with the state-of-the-art research in the field of Ising spin machines. The area of the stochastic bit based Ising unit has been calculated based solely on the CMOS device area, since there are potentials for further area optimization. As a result, stochastic bit based Ising unit demonstrates comparable area efficiency relative to other works. The prior works [5, 6] utilizing off-chip RNGs did not account for their associated power consumption. Nevertheless, our approach not only eliminates the need for RNG integration but also achieves a lowest power consumption of 1.07 μW per spin and an energy efficiency of 107 fJ per spin. Even with a spin count of 300, our software emulation platform maintains outstanding scalability, maintaining power and energy efficiencies of 1.15 μW per spin and 115 fJ per spin, respectively. This demonstrates the energy-efficient nature of CMOS stochastic bit-based Ising spin machine, with minimal power and energy increment despite the increased system complexity.

7 | Conclusion

In this work, we proposed low power CMOS stochastic bit based Ising machine to energy efficiently compute combinatorial optimization problems (COPs). To avoid large power consumption of common random number generator (RNG) that is needed to implement escape from the convergence to local minimum state of the system, we leveraged the inherent stochasticity of low power CMOS stochastic bit. We demonstrated the feasibility of the proposed design by applying it to the NP-complete graph coloring problem with four vertices and three color variables. Furthermore, proposed CMOS stochastic bit based spin type exhibits lowest power/spin (1.07 μW) and energy/spin (107 fJ) compared to state-of-the-art Ising spin machine researches.

Author Contributions

Honggu Kim: conceptualization, investigation, methodology, resources, software, validation, visualization, writing – original draft. **Dongjun Son**: resources. **Yerim An**: investigation. **Yong Shim**: formal analysis, funding acquisition, project administration, supervision, writing – original draft.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (No. 2021R1C1C100875214). This research was supported by the National Research Council of Science & Technology (NST) Grant by the Korea government (MSIT) (No. GTL24041-000). The EDA tool was supported by the IC Design Education Center (IDEC), Korea.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. A. Lucas, "Ising Formulations of Many NP Problems," *Frontiers in Physics* 2 (2014): 1–15, https://www.frontiersin.org/articles/10.3389/fphy.2014.00005.

2. M. Horowitz, "Computing's Energy Problem (and What We Can Do About It)," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)* (IEEE, 2014), 10–14, https://doi.org/10.1109/ISSCC.2014.6757323.

3. M. W. Johnson, M. H. Amin, S. Gildert, et al., "Quantum Annealing With Manufactured Spins," *Nature* 473 (2011): 194–198, https://doi.org/10.1038/nature10012. PMID: 21562559.

4. J. Bae, W. Oh, J. Koo, and B. Kim, "CTLE-Ising:A 1440-Spin Continuous-Time Latch-Based Isling Machine With One-Shot Fully-Parallel Spin Updates Featuring Equalization of Spin States," in *2023 IEEE International Solid-State Circuits Conference (ISSCC)* (IEEE, 2023), 142–144, https://doi.org/10.1109/ISSCC42615.2023.10067622.

5. Y. Su, J. Mu, H. Kim, and B. Kim, "A 252 Spins Scalable CMOS Ising Chip Featuring Sparse and Reconfigurable Spin Interconnects for Combinatorial Optimization Problems," in *2021 IEEE Custom Integrated Circuits Conference (CICC)* (IEEE, 2021), 1–2, https://doi.org/10.1109/CICC51472.2021.9431401.

6. Y. Su, T. T. -H. Kim, and B. Kim, "FlexSpin: A Scalable CMOS Ising Machine With 256 Flexible Spin Processing Elements for Solving Complex Combinatorial Optimization Problems," in *2022 IEEE International Solid-State Circuits Conference (ISSCC)* (IEEE, 2022), 1–3, https://doi.org/10.1109/ISSCC42614.2022.9731680.

7. M. Koo, G. Srinivasan, Y. Shim, and K. Roy, "sBSNN: Stochastic-Bits Enabled Binary Spiking Neural Network With On-Chip Learning for Energy Efficient Neuromorphic Computing at the Edge," *IEEE Transactions on Circuits and Systems I: Regular Papers* 67, no. 8 (2020): 2546–2555, https://doi.org/10.1109/TCSI.2020.2979826.

8. H. Kim, Y. An, M. Kim, G. -C. Heo, and Y. Shim, "All Stochastic-Spiking Neural Network (AS-SNN): Noise Induced Spike Pulse Generator for Input and Output Neurons With Resistive Synaptic Array," *IEEE Transactions on Circuits and Systems II: Express Briefs* 72, no. 1 (2025): 78–82, https://doi.org/10.1109/TCSII.2024.3485178.

9. S. Srinivasan, S. Mathew, V. Erraguntla, and R. Krishnamurthy, "A 4Gbps 0.57pJ/bit Process-Voltage-Temperature Variation Tolerant All-Digital True Random Number Generator in 45nm CMOS," in *2009 22nd International Conference on VLSI Design* (IEEE, 2009), 301–306, https://doi.org/10.1109/VLSI.Design.2009.69.