

An Ultra-Low Power and Fast Ising Machine using Voltage-Controlled Magnetoresistive Random Access Memory

Authors: Yihao Zhang^{1,†}, Sai Li^{1,2,†,*}, Albert Lee³, Zheng Zhu³, Lang Zeng^{1,2}, Peng Wang^{1,2}, Lei Gao⁴, Di Wu³, Weisheng Zhao^{1,2,*}

¹School of Integrated Circuit Science and Engineering, Beihang University, Beijing, 100191, China.

²National Key Lab of Spintronics, Institute of International Innovation, Beihang University, Hangzhou, 311115, China.

³InstonTech, Suzhou, 215000, China.

⁴Empyrean Technology Co., Ltd., Beijing, 100102, China.

[†]These authors contributed equally: Yihao Zhang, Sai Li

*Corresponding author. Email: saili@buaa.edu.cn, weisheng.zhao@buaa.edu.cn

Abstract

Physics-inspired computing paradigms, such as Ising machines, are emerging as promising hardware alternatives to traditional von Neumann architectures for tackling computationally intensive combinatorial optimization problems (COPs). While quantum, optical, and electronic devices have garnered significant attention for their potential in realizing Ising machines, their translation into practical systems for industry-relevant applications remains challenging, with each approach facing specific limitations in power consumption and speed. To address this challenge, we report the first chip-level spintronic Ising machine using voltage-controlled magnetoresistive random access memory. The core of our design leverages magnetic tunnel junctions (MTJs) driven by the voltage-controlled magnetic anisotropy effect to realize the probabilistic update of Ising spins through a new mechanism. It enables a latency below 1 ns and an energy consumption under 40 fJ per spin update, achieving a 1000× improvement over previous current-driven MTJ-based implementations. We map two real-world COPs in electronic design automation—global routing and layer assignment—onto the Ising model and demonstrate high-quality results with an energy efficiency of 2.5×10^4 solutions per second per watt. This outperforms state-of-the-art quantum and graphics processing units by six and seven orders of magnitude, respectively. These results establish voltage-controlled spintronics as a compelling route towards next-generation physics-inspired machine intelligence, offering a paradigm for ultra-low-power, high-speed, and scalable computation.

Introduction

Von Neumann architectures falter with energy-constrained, compute-intensive nondeterministic polynomial time (NP)-hard COPs, where solution search time and resource demands scale exponentially with problem size¹, impacting diverse scientific and industrial domains, such as materials science², finance³, transportation⁴, communications⁵, and integrated circuit design⁶.

Physics-inspired hardware solvers implementing energy-minimization principles, exemplified by Ising machines, have emerged as a promising alternative for tackling computationally hard COPs over the past decade⁷. By mapping these problems onto the Ising model, such machines leverage Ising lattice dynamics to efficiently explore high-dimensional parameter spaces and converge toward ground states, demonstrating remarkable efficiency for finding high-quality solutions⁸. However, realizing large-scale, practical Ising machines remains challenging across various implementation paradigms. Quantum processing units (QUPUs)^{9,10}, while powerful in principle¹¹, contend with cryogenic operating temperatures, sparse topological constraints and decoherence issues¹². Optical approaches, dependent on high-precision components, struggle with high energy consumption and stability issue^{13–16}. Conventional complementary metal-oxide-semiconductor (CMOS) approaches, spanning graphics processing units (GPUs)¹⁷, field-programmable gate arrays (FPGAs)¹⁸, and application-specific integrated circuits^{19–24}, although widely explored but often suffer from significant digital overheads^{25,26}, and computational latency²⁷. Moreover, CMOS ring oscillators, are typically limited by the insufficient reconfigurability of their coupling, thus restricting their application to basic problems such as Max-cut^{28–30}.

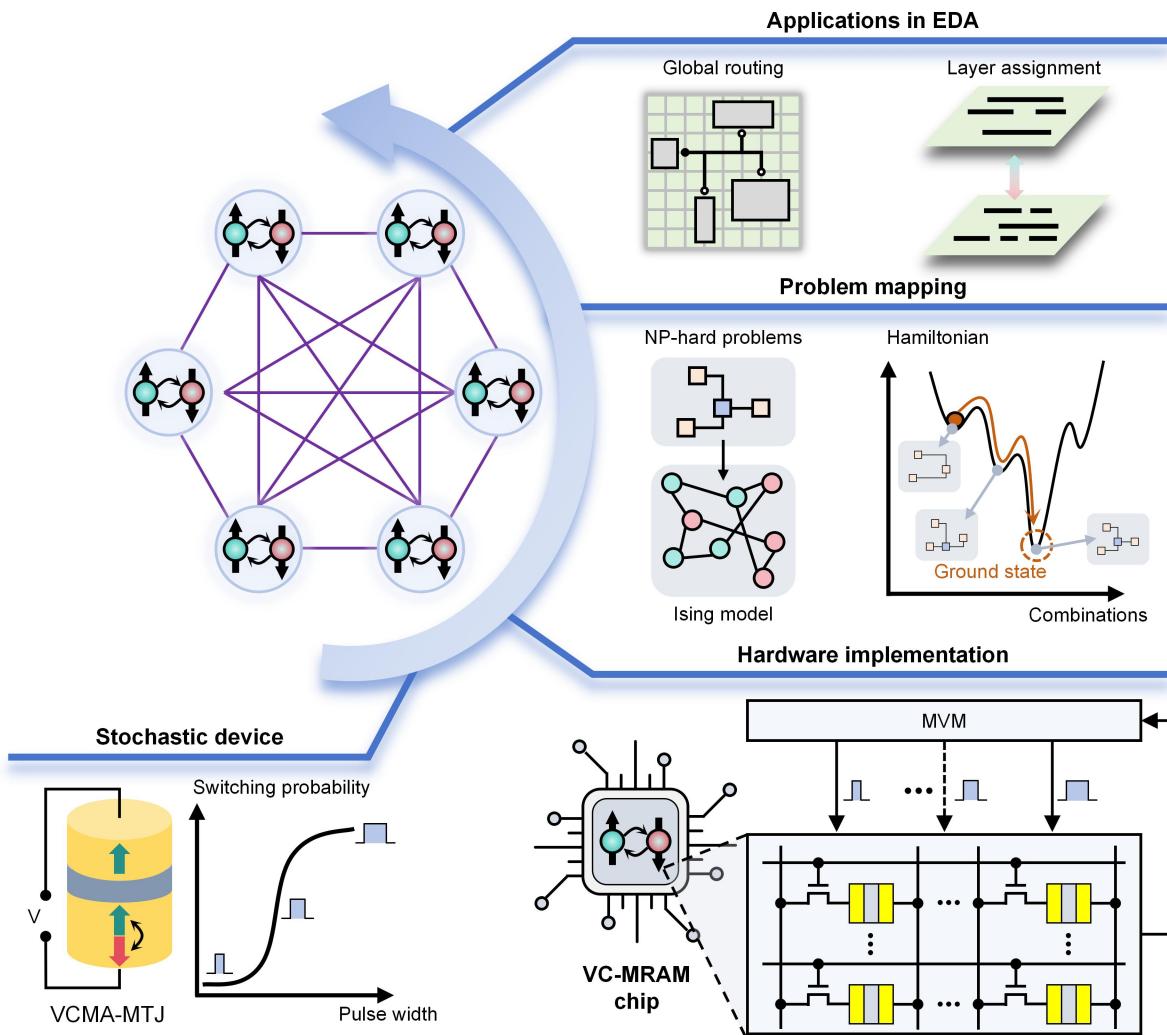


Fig. 1 Overview of our voltage-controlled spintronic Ising machine. (i) Stochastic device: harnessing intrinsic probabilistic switching in VCMA-MTJ devices for in-situ Ising spin update. (ii) Hardware implementation: CMOS-compatible fabrication of VC-MRAM enables chip-level Ising machine. (iii) Problem mapping: Framework converts real-world NP-hard COPs into Ising formulations. (iv) EDA applications: Successful solution of global routing and multilayer assignment challenges in EDA.

To overcome these limitations, recent exploration has focused on nanodevice-based Ising machines, which process efficient computation directly using unique physical properties in an impact CMOS-compatible systems. Manifold approaches have emerged, employing mechanisms such as stochastic switching (e.g., in MTJs^{31–36}), bistable phase-transition dynamics (with nano-oscillators^{37,38}), or in-memory matrix operations acceleration (using resistive random access memory^{39–41} and ferroelectric field effect transistor⁴²). Within this landscape of nanodevices, spintronics, specifically stochastic MTJs, is inherently compatible with Ising formulations due to its intrinsic probabilistic state transitions. Existing spintronic Ising machines predominantly utilize two types of current-driven devices including superparamagnetic tunnel junctions (SMTJ)^{31–34} and spin-orbit torque (SOT) MTJs^{35,36}. However, these implementations have typically been limited to single devices or small-scale systems, where two main challenges still need to be addressed to enlarge the advantage in large-scale spintronic Ising machines. First, current-driven switching is limited by inefficiencies in energy consumption, speed, which obstruct performance scalability. Second, the absence of chip-level validation hinders the development of a flexible platform for mapping to practical computing applications, significantly affecting the performance of Ising machines and their ability to deliver high-quality solutions.

In this work, we introduce MTJs driven by voltage-controlled magnetic anisotropy (VCMA) effect as Ising spins (Fig. 1) to address these challenges. The intrinsic stochasticity of VCMA-MTJs offers precise tuning of switching probability as a function of pulse width, enabling in-situ, significantly improved speed (<1 ns) and energy efficiency (<40 fJ) of spin update, achieving both a 1000× improvement over conventional current-driven MTJ schemes. We fabricate 96-kb voltage-controlled magnetoresistive random access memory (VC-MRAM) chips using 40-nm CMOS and 70-nm MTJ technologies, demonstrating the first chip-level spintronic Ising machine, hereafter referred to as the voltage-controlled spintronic Ising machine (VSIM). Two representative NP-hard COPs in very large-scale integration (VLSI) electronic design automation (EDA)—global routing and layer assignment—are mapped onto the Ising model and

solved experimentally. It achieves 2.5×10^4 solutions per second per watt on 100-node instances, surpassing state-of-the-art Ising machines by three to seven orders of magnitude. These results establish voltage-controlled spintronics as an energy-efficient and high-speed platform for physics-inspired computing, with CMOS-compatible architecture paving the way for Ising machines capable of addressing industry-scale optimization challenges.

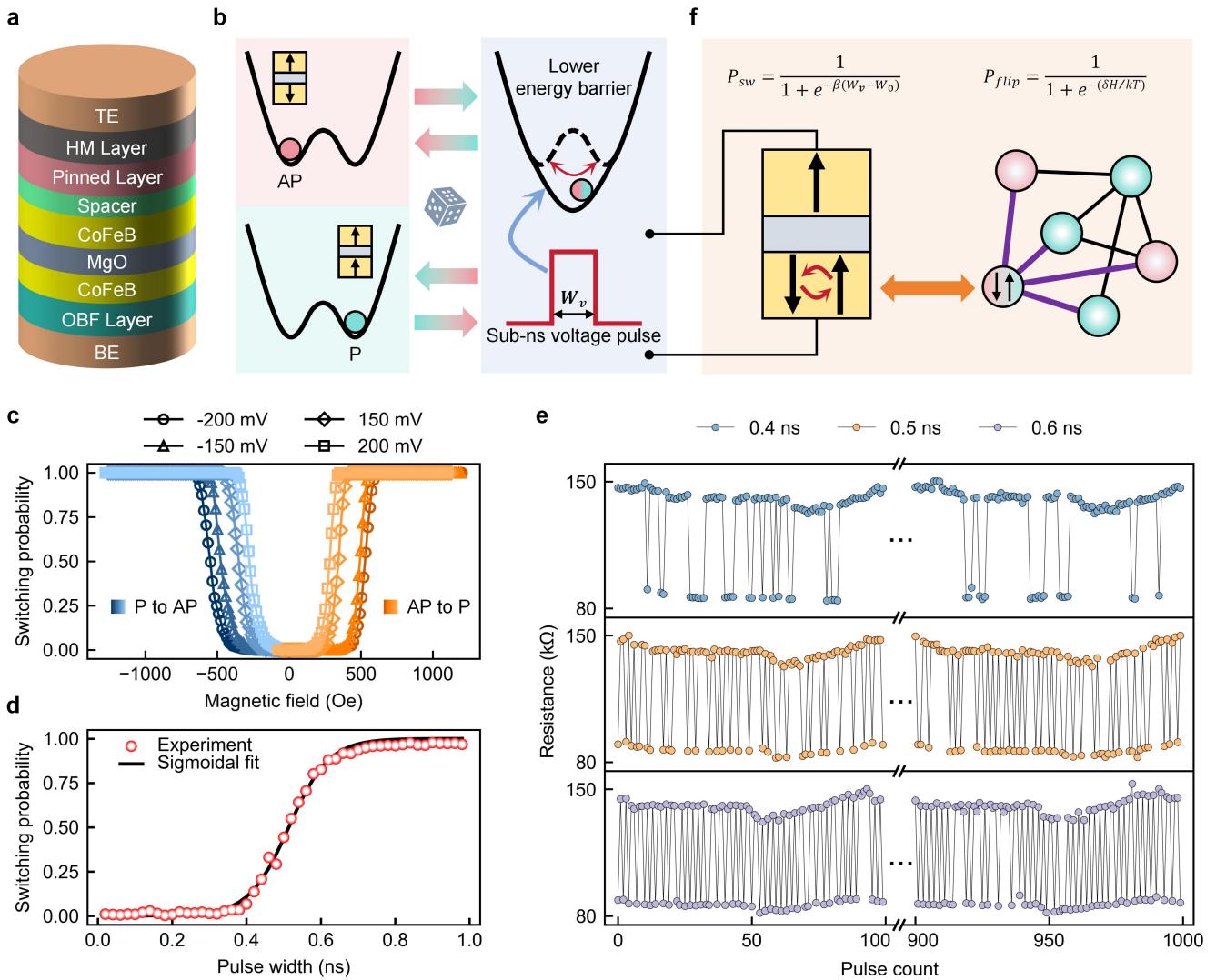


Fig. 2 Ising spins using VCMA-MTJs. **a**, VCMA-MTJ multilayer stack architecture. **b**, Voltage-controlled magnetic anisotropy reduces the energy barrier, enabling stochastic switching between P and AP states. **c**, P_{sw} versus magnetic field under different voltage biases. It indicates the symmetry of the P-to-AP and AP-to-P switching. **d**, Sigmoidal P_{sw} versus W_v , enabling the in-situ ultra-fast (sub-ns) Ising computing under a write voltage of 2.2 V. **e**, Resistance measurements after each of 1000 consecutive 0.4 ns, 0.5 ns, and 0.6 ns write pulses, corresponding to P_{sw} of ~10%, ~50% and ~90%.

VCMA-MTJs for Ising spins

The Hamiltonian of an all-to-all Ising model can be expressed as

$$H = - \sum_{i < j} J_{ij} s_i s_j - \sum_i h_i s_i \quad (1)$$

where H represents the total energy of the system, while s_i denotes the i -th spin that can adopt either of two states $\{-1, +1\}$, corresponding to spin-down or spin-up configurations. The coupling strength between spins is characterized by the matrix J , and h represents the external magnetic field. The spin dynamics can be simulated through Monte Carlo sampling, specifically using the well-established Glauber approach⁴³:

$$P_{flip} = \frac{1}{1 + e^{-(\delta H/kT)}} \quad (2)$$

where P_{flip} denotes the probability of spin flipping, δH represents the energy reduction in system Hamiltonian upon flipping, k is the Boltzmann constant and T is the temperature. To emulate the dynamical evolution of the Ising model, an annealing process is employed, during which k is fixed at a constant value (set to 1 in this work), and T gradually decreased in an iterative manner. The probabilistic flipping of Ising spins, representing thermal fluctuations in physical systems, is crucial for escaping local minima, as it allows flips even when δH is negative (indicating an energy increase), provided T is sufficiently high.

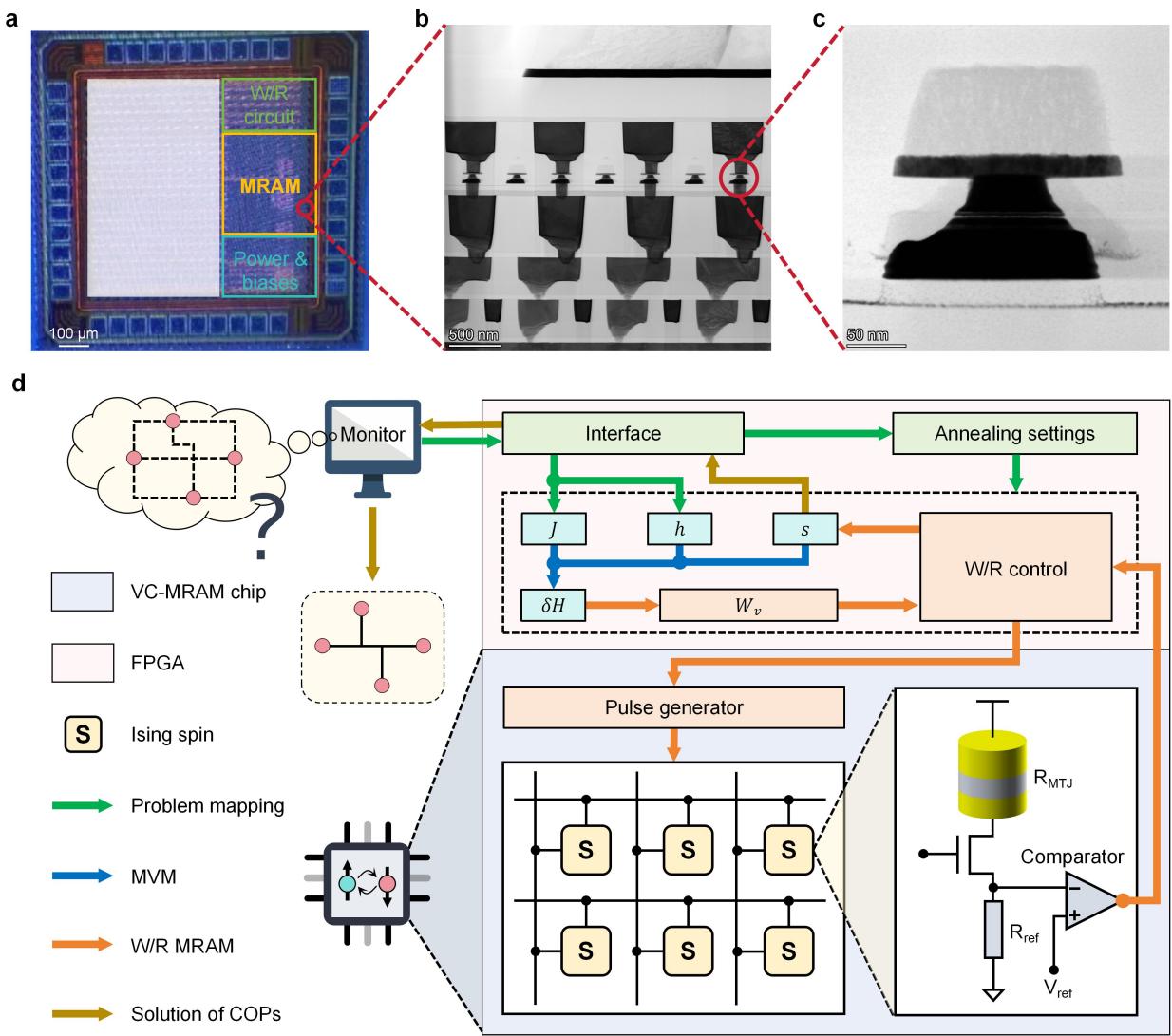


Fig. 3 Implementation of VSIM with the VC-MRAM chip. **a**, Die micrograph of the fabricated VC-MRAM chip. **b**, Cross-sectional TEM image illustrating the vertical integration of MTJs with CMOS circuitry. **c**, TEM image of a single MTJ with a nominal diameter of 70 nm. **d**, VSIM architecture: (i) The host monitor interfaces with FPGA to configure Ising parameters J and h ; (ii) FPGA modulates in-MRAM pulse generator to drive MTJ for Ising computing; (iii) Spin states s are sampled for subsequent computation.

The key to implementing the Ising model lies in employing an efficient building block to realize Ising spins with a tunable P_{flip} . To this end, we adopt perpendicular VCMA-MTJ devices (Fig. 2a) as the Ising spin, where the bistable magnetization switching probability (P_{sw}) is purely controlled by an electric field and depends on the duration of the writing pulse. The MTJ is designed with a high resistance-area product (RA) of $300 \Omega \cdot \mu\text{m}^2$, where a thick MgO layer is served to eliminate spin-transfer torque current contribution and reduce the switching energy consumption significantly. As shown in Fig. 2b, a single voltage pulse lowers the energy barrier between the anti-parallel (AP) and parallel (P) states (representing the +1 and -1 states of Ising spins, respectively) with a certain polarity and regulate the stochastic magnetization precession⁴⁴. It is worth mentioning that the MTJ achieves a symmetry of the switching process as demonstrated in Fig. 2c, where the P_{sw} for AP-to-P and P-to-AP transitions are nearly identical under equivalent voltage conditions. As a result, the probabilistic flipping required for an Ising spin can be induced in VCMA-MTJs by the same voltage pulse irrespective of the initial P or AP state, obviating reset operations of the device. To precisely determine the P_{sw} , we measure the MTJ state 1000 times after applying the identical voltage pulse for each measurement. Fig. 2d shows that P_{sw} as a function of voltage pulse width (W_v) follows a sigmoid relationship:

$$P_{sw} = \frac{1}{1+e^{-\beta(W_v-W_0)}} \quad (3)$$

where β and W_0 represent fitting coefficients. Fig. 2e illustrates the resistance measurements obtained following each pulse in a series of 1000 consecutive write operations, with pulse durations of 0.4 ns, 0.5 ns, and 0.6 ns, corresponding to P_{sw} of approximately 10%, 50%, and 90%. The results indicate that the MTJ devices can achieve the full P_{sw} range, from 0% to 100%, within a large voltage window of 1.9 V to 2.2 V (Supplementary Note 1). Voltage-controlled probabilistic switching enables in-situ annealing, where the applied

pulse width is determined as

$$W_v = \frac{\delta H}{\beta kT} + W_0 \quad (4)$$

to ensure that the P_{sw} of MTJs matches the P_{flip} of Ising spins. This sigmoidal switching characteristic, controlled by sub-nanosecond voltage pulses, renders the VCMA-MTJ a natural embodiment of Ising spins (Fig. 2f).

Hardware implementation of VSIM

We fabricated 96-kb VC-MRAM chips (Fig. 3a) using a 40-nm CMOS process, incorporating MTJs with diameters of approximately 70 nm (Fig. 3b, c). Little die-to-die resistance variations (see Supplementary Fig. 1e) provide high-quality of the P_{sw} under the applied voltage. An integrated pulse generator within the VC-MRAM (see Supplementary Note 3) applies precisely timed voltage pulses to the MTJs, which function as voltage-controlled Ising spins. We implemented VSIM on our MRAM chip, assisted by an FPGA for control and interfacing. The system design diagram is illustrated in Fig. 3d, and the experimental prototype is shown in Extended Data Fig. 1. NP-hard COPs are mapped onto the Ising model by determining the necessary Ising parameters J and h (the specific mapping process will be discussed later). These Ising parameters and the VSIM annealing settings are configured via the monitor. During the annealing process, δH of flipping s_i is calculated as

$$\delta H = -2s_i(\sum_j J_{i,j}s_j + h_i) \quad (5)$$

which is subsequently used to determine the corresponding W_v according to Eq. (4). The write/read (W/R) control module drives the pulse generator in the MRAM based on the W_v . The Ising spins are updated accordingly, stimulated by voltage pulses of target width. After a series of annealing iterations, the final states of the Ising spins, corresponding to the solution of the COP, are transferred back to the monitor.

VSIM capitalizes on the intrinsic stochasticity of VCMA-MTJs to emulate Ising spins, significantly reducing area overhead compared to CMOS schemes reliant on pseudo-random number generators (PRNGs), while achieving faster and more energy-efficient Ising computations than previous current-driven MTJs. It is also worth noting that our experimental results confirm exceptional device endurance up to 10^{13} cycles (see Supplementary Excel data), with ongoing testing expected to further validate its exceptional longevity. Compared to other current-driven MTJ-based Ising spin implementations, the VCMA-MTJ's reduced reliance on thermal effects ensures enhanced device tolerance. And the VC-MRAM chip offers exceptional scalability and higher area density by eliminating the need for large write currents, enabling the use of minimum-sized access transistors in advanced technology nodes. Moreover, the FPGA provides the flexibility to tackle various NP-hard COPs by simply adjusting the Ising parameters J and h for different problems. This capability establishes VSIM as a highly efficient, universal Ising machine. We next demonstrate VSIM's effectiveness by solving two key NP-hard COPs in EDA.

VSIM for global routing

VLSI EDA is a powerful tool for the design of increasingly complex integrated circuits in the modern semiconductor industry, a field where many core VLSI design problems are NP-hard COPs⁴⁵. The VLSI design process is typically segmented into three critical phases: component placement, physical routing determination, and layer assignment⁴⁶. Here, we start from mapping the global routing problem onto the Ising model as it plays a crucial role in minimizing routing distance across the chip, which is efficiently implemented on-chip using the VSIM system.

The global routing problem can be effectively modeled as a rectilinear Steiner minimum tree (RSMT) problem. To achieve this mapping, consider a directed weighted graph $G = (V, E)$, where the weight of each edge $(u, v) \in E$ is denoted as $c_{u,v}$. Given a set $U \subseteq V$ of terminal vertices and a root vertex $v_0 \in U$, the Steiner minimum tree is a spanning tree $T = (V_T, E_T)$ that satisfies $U \subseteq V_T \subseteq V$ and $E_T \subseteq E$, while minimizing the total cost. Vertices in the set $W = V \setminus U$ are referred to as Steiner vertices⁴⁷. The RSMT problem is a specific variant of the Steiner minimum tree problem in routing scenarios⁴⁸ (Fig. 4a). Directly mapping the RSMT problem to the Ising model is not sufficiently concise; therefore, we first employ a quadratic unconstrained binary optimization formulation as proposed in reference⁴⁹. The variables include edge variables x^e and order variables x^o , both of which take values in {0, 1}. Specifically, when $x_{u,v}^e = 1$, it indicates that edge (u, v) is included in the Steiner tree while $x_{u,v}^e = 0$ indicates it is not included. If $x_{u,v}^o = 1$, it signifies that u is closer to the root vertex than v , and 0 otherwise (Fig. 4b). The total Hamiltonian for the RSMT problem is stated as:

$$H_{RSMT}(x) = H_o + \lambda_1 H_{c1} + \lambda_2 H_{c2} + \lambda_3 H_{c3} + \lambda_4 H_{c4} \quad (6)$$

Here, H_o represents the optimization objective, specifically the total cost of all edges in the Steiner tree:

$$H_o = \sum_{\substack{(u,v) \in E \\ v \neq v_0}} c_{u,v} x_{u,v}^e \quad (7)$$

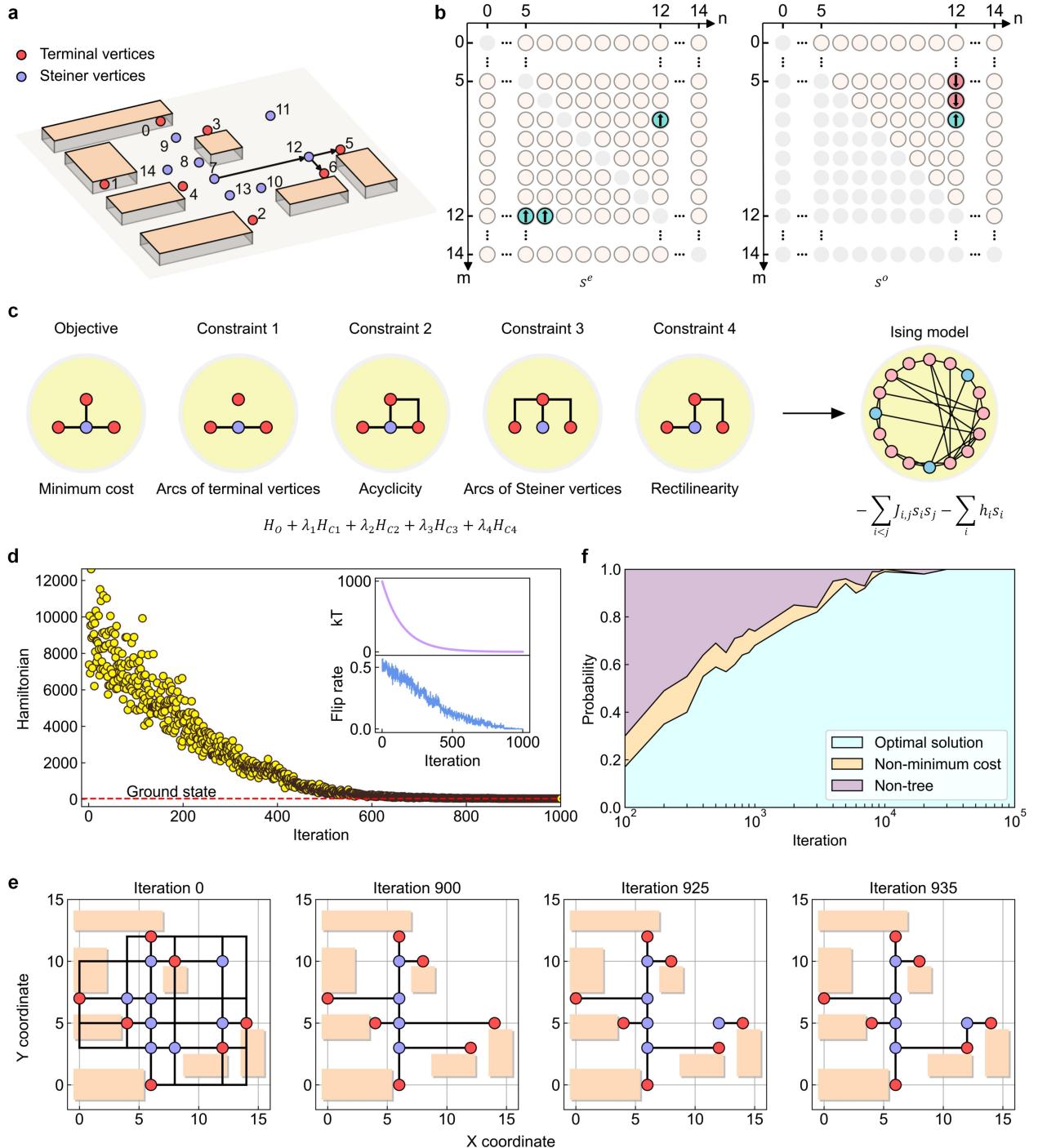


Fig. 4 Global routing demonstration on VSIM. **a**, Global routing problem formulated as a RSMT problem, where terminal vertices correspond to circuit module pins and Steiner vertices represent auxiliary routing nodes for wirelength optimization. **b**, Variable definitions: $s_{i,j}^e$ encodes edge connectivity between vertices i and j while $s_{i,j}^o$ specifies topological ordering relative to the routing root. **c**, Ising model mapping framework of the RSMT problem with dual components: (i) Objective term minimizes total wirelength cost; (ii) Constraint terms enforce appropriate arcs of terminal vertices and Steiner vertices, acyclicity and rectilinearity. **d**, Hamiltonian evolution trajectory of the problem in **a** across 1000 iterations. Dashed line represents the ground state. Insets: kT profile and flip rate of Ising spins during the annealing process. **e**, Progressive routing solutions extracted at four representative annealing stages in **d**. **f**, Solution validity statistics from 100 independent trials per parameter set, classifying results as: optimal solutions, suboptimal trees and invalid solutions.

$H_{c1} \sim H_{c4}$ denote four essential constraints for Steiner tree formation, with corresponding penalty coefficients $\lambda_1 \sim \lambda_4$ that must be carefully selected for obtaining the optimal performance. These constraints serve distinct purposes (Fig. 4c). H_{c1} ensures that each terminal vertex (except the root) has precisely one incoming arc:

$$H_{c1} = \sum_{v \in U \setminus \{v_0\}} (1 - \sum_{(u,v) \in E} x_{u,v}^e)^2 \quad (8)$$

H_{c2} enforces acyclicity, a fundamental requirement for tree structures. H_{c3} penalizes configurations where Steiner vertices either have multiple incoming arcs or possess outgoing arcs without any incoming ones. H_{c4} addresses the rectilinearity constraint specific to

routing scenarios. When two vertices u and v cannot be connected by a straight segment, a positive penalty is imposed. H_{c2} is expressed as a function of both x^e and x^o , while other constraints are formulated solely in terms of x^e (detailed mathematical formulations for H_{c2} , H_{c3} , and H_{c4} are provided in Supplementary Note 4). The spin variables s in the Ising model is derived through a composite transformation of x^e and x^o :

$$s = \begin{pmatrix} s^e \\ s^o \end{pmatrix} = 2 * \begin{pmatrix} x^e \\ x^o \end{pmatrix} - 1 \quad (9)$$

where s^e and s^o are corresponding Ising spins for edge and order variables, respectively. Therefore, we successfully construct the mapping from the RSMT problem to the Ising model, yielding J and h by comparing Eq. (1) and Eq. (6). The complete derivation process is also detailed in Supplementary Note 4.

After J and h are mapped and configured in VSIM, the MRAM chip initiates in-situ annealing process, leveraging the intrinsic stochasticity of VCMA-MTJs. As shown in Fig. 4d, while the total Hamiltonian exhibits occasional increases, it follows a distinct downward trend overall, converging to the ground state rapidly within 1000 iterations. Throughout this process, the temperature T decreases exponentially to achieve fast convergence. Simultaneously, Ising spin flips are governed by both the system's energy change (δH) and thermal fluctuations related to T , as described in Eq. (3). The flip rate, defined as the proportion of spins that flip in the entire system per iteration during the annealing process, is a statistically measured value. At high T , the flip approaches 0.5, reflecting a randomized state, while at low T , it tends towards 0 as the system stabilizes. Fig. 4e shows the routing results at four representative iterations, illustrating the evolution from a random initial state through suboptimal solutions to the optimal solution. This progression underscores MRAM's ability to provide sufficient stochasticity, enabling the system to escape local minima and converge toward the ground state.

We categorize solutions into three types: optimal solutions (successful routing with minimized wire length), suboptimal trees (successful routing without minimized wire length), and invalid solutions (unsuccessful routing, i.e., containing unconnected terminal vertices). For penalty coefficients set to $\lambda_1 = \lambda_2 = 10$ and $\lambda_3 = \lambda_4 = 6$, the probability distribution of these categories over iterations is shown in Fig. 4f. The probability of achieving the optimal solution—termed the success probability—exceeds 95% beyond 10^4 iterations. The calibration of penalty coefficients is crucial for maintaining high success probability, requiring a careful balance between optimization objectives and constraints. Insufficient coefficient values hinder proper Steiner tree construction (resulting in invalid solutions), while excessive values compromise minimal cost achievement (yielding suboptimal trees). A detailed analysis of success probability variations with different coefficient settings is presented in Supplementary Note 5. Notably, the aforementioned Ising mapping method, along with the VSIM-based solution framework, can be applied to a wide range of routing problems.

VSIM for layer assignment

Layer assignment involves allocating wire segments across multiple metal layers, a task that is considerably more complex than global routing. Following standard practice, we assign horizontal and vertical segments to separate layers to minimize interference. Intersecting horizontal and vertical segments require vias for electrical connectivity (Fig. 5a). This layer assignment strategy addresses two critical issues: managing local wire density⁵⁰ and minimizing via usage⁴⁶. Wire density minimization reduces to the Max-cut problem⁵¹ and incorporating via minimization transforms this into a multi-objective optimization challenge. We formulate it as a density-driven via-aware layer assignment (DVLA) problem, mapping it to the Ising model for efficient solution implementation on VSIM.

In our four-layer model (L_1 - L_4), horizontal and vertical segments occupy alternating odd- and even-numbered layers respectively (horizontal in L_1/L_3 , vertical in L_2/L_4). Let Ising spin variables $s = [s^h \ s^v]$ directly encode the layer assignment for each segment. If $s_i^h = -1$ ($s_i^v = -1$), it indicates that the i -th horizontal (vertical) segment is assigned to L_1 (L_2); while if $s_i^h = +1$ ($s_i^v = +1$), it is assigned to L_3 (L_4). The total Hamiltonian for the DVLA problem can be expressed as $H_{DVLA} = \lambda_D H_D + \lambda_V H_V$, where H_D and H_V represent the density-driven and via-aware contributions to the overall objective, respectively. The penalty coefficients λ_D and λ_V determine the relative weight of each component in the optimization process. Denote the local density matrices of horizontal (vertical) segments as w^h (w^v). The local density between two horizontal (vertical) segments is calculated as l/d , where l is the overlap length in the horizontal (vertical) direction, and d is the distance in the vertical (horizontal) direction (Fig. 4b). The density-driven term H_D can then be expressed as:

$$H_D = \frac{1}{2} \sum_{0 \leq i < j < N_h} w_{i,j}^h (s_i^h s_j^h - 1) + \frac{1}{2} \sum_{0 \leq m < n < N_v} w_{m,n}^v (s_m^v s_n^v - 1) \quad (10)$$

where N_h and N_v denote the numbers of horizontal and vertical segments, respectively. The via-aware term H_V is formulated in Supplementary Note 6, along with a clear derivation of the Ising mapping results. Fig. 5c visualizes the corresponding Ising model of the problem presented in Fig. 5a.

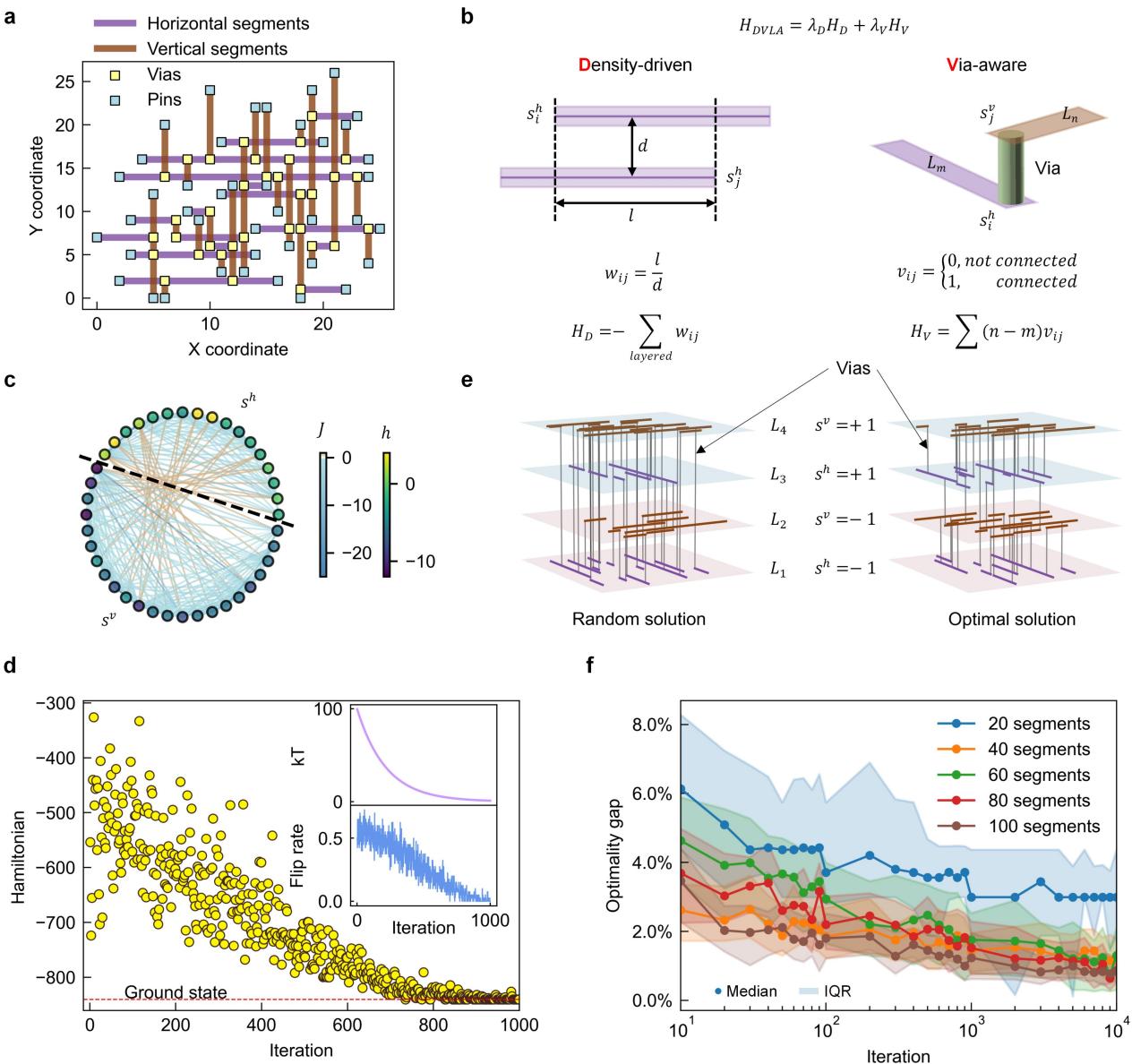


Fig. 5 Demonstration of layer assignment optimization using VSIM. **a**, Layer assignment scenario illustrating horizontal and vertical wire segments, with inter-layer connections facilitated by vias. **b**, The DVLA problem formulation, which simultaneously optimizes wire density distribution and minimizes via count. **c**, Visualization of the Ising model representation corresponding to the problem in **a**. Circles represent Ising spins, with colors indicating external magnetic field strength. Connection colors between spins denote coupling strength. **d**, Dynamic evolution during problem solving, showing Hamiltonian trajectory, kT profile, and spin flip rate throughout the annealing process. **e**, Solution visualization demonstrating the dual optimization objectives: homogenized wire density and minimized via count. **f**, Optimality gap versus iteration count for problems of various sizes. Data points represent median values while shaded regions indicate interquartile ranges (IQR).

Fig. 5d shows the Hamiltonian rapid decrease, reaching the ground state over 1000 iterations with coefficients $\lambda_D = 10$ and $\lambda_V = 4$. Due to the max-cut nature of H_D , the total Hamiltonian becomes negative when λ_D is significantly larger than λ_V . This does not affect the annealing process, as the relative variation of the Hamiltonian, rather than its absolute value, is of primary concern. Fig. 5e illustrates the initial and final solutions to the DVLA problem, demonstrating both homogenized local density and minimized via count. Converging to the optimal solution for the DVLA problem is particularly challenging due to its two competing objectives: reducing wire density by assigning segments to different layers, which inherently increases via requirements, versus minimizing the overall via count. Fig. 5f shows the optimality gap (defined as the relative error between convergence value and minimum value) versus iterations for problems of varying sizes. As iterations exceed 10^4 , the optimality gaps fall below 2%, which is sufficient for real-world EDA designs. Given the capability in exploring complex parameter spaces, VSIM is thus well-suited to efficiently tackle such optimization problems characterized by competing objectives and challenging solution landscapes.

		GPU ¹⁷	QPU ^{10,11}	CIM ^{13,14}	PTNO ³⁷	SOT-MTJ ³⁵	SMTJ ^{31,33}	VSIM
Ising spin	Form	PRNG	Qubit	DOPO ^Ψ	PTNO	SOT-MTJ	SMTJ	VCMA-MTJ
	Number	N/A [†]	5627	100k	8	1	80	96k
	Speed	3.2 μs	7 ns	5 μs	1.85 μs	4 μs	100 μs	< 1 ns
	Energy consumption	38 μJ	- [‡]	-	47 pJ	810 pJ	50 pJ	< 40 fJ
System implementation	Connectivity	All-to-all	Sparse	All-to-all	All-to-all	All-to-all	All-to-all	All-to-all
	Technology	5 nm CMOS	N/A	N/A	Devices	Devices	Devices	MRAM chip
	Temperature	300 K	12 mK	300 K	300 K	300 K	300 K	300 K
	Power	450 W	25 kW	-	32 mW [‡]	-	329 mW	40 mW
Experimental application and performance	Type of COP	Max-cut	Max-cut	Max-cut	Max-cut	Integer-factorization	TSP	RSMT DVLA
	Difficulty level	★	★	★	★	★★	★★★	★★★★★
	Solutions per second per watt [†]	6.95×10^{-4}	5.71×10^{-3}	-	1.69×10^1	-	3.04×10^{-2}	2.50×10^4

Table 1 Comparison of VSIM with state-of-the-art Ising machines. ^Ψ Degenerate optical parametric oscillator. [†] “ - ” indicates that the data is not reported in the corresponding reference; “N/A” denotes not applicable or not suitable for direct comparison. [‡] An estimated extrapolation to a 100-spin system. [†] Normalized to a reference task involving a 100-spin system executing 10^4 iterations to facilitate a fair comparison.

Discussions

To assess the impact of spin quality, we first evaluate VCMA-MTJs compared with CMOS-based Ising spins using linear feedback shift registers (LFSRs) in solving global routing problem (Extended Data Fig. 3). We found that VCMA-MTJs achieved a success probability comparable to 16-bit LFSR. On this basis, VCMA-MTJs demonstrated significant performance advantages over conventional hardwares (see Supplementary Notes 7-9): 10^4 times faster operation with 10^9 times lower energy reduction than central processing units (CPU) and GPU; $10\times$ higher speed, $150\times$ lower energy consumption, and $3000\times$ fewer transistors than FPGA. Furthermore, VCMA-MTJs exhibit remarkable resilience to device variations, sustaining $> 95\%$ success probability even when a fifth of devices operate with peak switching probability (PSP) as low as 60%. Such robustness underscores the viability of Ising machines built on current non-volatile memory technologies for tackling inter-device variability, a critical hurdle in computing hardware.

Evaluating overall system performance, we benchmarked VSIM against state-of-the-art Ising machines in Table 1, including GPU-based solver¹⁷, QPU^{10,11}, coherent Ising machine (CIM)¹³, phase-transition nano-oscillators (PTNO)³⁷, SOT-MTJ³⁵, and SMTJ³³. Among these, VSIM realizes ultra-fast (< 1 ns) and energy-efficient (< 40 fJ/spin) Ising spins by exploiting the intrinsic voltage-tunable stochasticity of VCMA-MTJs. Benefiting from the VC-MRAM chip, VSIM is a low-power (40 mW), scalable system, with a current spin capacity reaching 96k. To enable fair comparisons across different Ising machines, all experimental results were normalized to a standardized benchmark with 100 Ising spins and 10^4 iterations (see Supplementary Notes 10). Under this configuration, VSIM achieves the highest energy efficiency (2.5×10^4 solutions per second per watt), representing an improvement of 1.5×10^3 - 3.6×10^7 times over competing platforms. Moreover, VSIM demonstrates superior capability in solving sophisticated, real-world problems, such as global routing and layer assignment in EDA, highlighting its potential for industrial deployment.

Fundamentally, QPUs offer rapid spin updates yet become highly time-consuming when tackling dense Ising topologies¹¹. Their reliance on cryogenic cooling leads to extremely high power consumption (25 kW), resulting in energy efficiency seven orders of magnitude lower than that of VSIM. Optical approaches such as CIM support large spin counts but require kilometer-scale fiber cavities^{13,14}, which severely impact overall system compactness and performance. Among electronic platforms, GPUs suffer from inefficient Ising emulation and high power cost. Oscillator-based approaches (e.g., PTNO) are faced with restricted programmability and significant power and area overheads due to their reliance on passive coupling components (e.g., capacitors/resistors)³⁷. And the simple bistable dynamics of such oscillators do not fully capture the stochastic nature of Ising spins. In contrast, spintronic approaches provide highly tunable probabilistic behavior, where VSIM achieves up to three orders of magnitude improvement in both speed and energy

consumption per spin update compared to current-driven SOT-MTJ and SMTJ Ising machines. Our findings underscore that integrating a well-suited physical phenomenon, exemplified by VCMA-MTJ probabilistic switching, with robust chip-level architectures capable of addressing large-scale problems, is key to unlocking substantial gains in both hardware performance and algorithm efficiency.

Conclusions and outlook

We propose and demonstrate the use of VCMA-MTJs to construct an Ising machine, which achieves significant improvement in energy consumption ($< 40 \text{ fJ}$) and speed ($< 1 \text{ ns}$) of spin update. Using 96 kb VC-MRAM, we developed VSIM, the first chip-level spintronic Ising machine, as a scalable, ultra-low-power and ultra-fast platform for tackling NP-hard COPs. Experimental results on complex instances show that VSIM delivers an energy efficiency of 2.5×10^4 solutions per second per watt—three to seven orders of magnitude higher compared to quantum, optical, and other electronic counterparts—paving the way for broader deployment in industry-relevant applications.

Further improvements in scalability and performance could be realized through multi-chip Ising computing architectures^{18,20} and hierarchical co-optimization strategies across global and local levels. These efforts are expected to significantly enhance the hardware efficiency and adaptability of physics-inspired computing systems, expanding their potential across a broad range of application domains in machine intelligence.

Methods

VC-MRAM fabrication

The VCMA-MTJs were fabricated on a 300-mm MRAM pilot line. As illustrated in Fig. 2a, the multilayer stack was deposited in situ at room temperature using physical vapor deposition in a 300-mm cluster tool (Canon-Anelva EC7800), followed by a post-deposition annealing process at 350°C for 1 hour under a 1 T magnetic field. The MgO tunnel barrier exhibited a RA of $300 \Omega \cdot \mu\text{m}^2$. Circular MTJ devices with a nominal diameter of $\sim 70 \text{ nm}$ were patterned using 193-nm immersion lithography and sequentially etched via ion beam etching at both normal and grazing incidence angles. Electrical measurements revealed room-temperature device resistance of $\sim 80 \text{ k}\Omega$ and a magnetoresistance ratio of $\sim 120\%$. Finally, a dual damascene copper top electrode was integrated using standard back-end-of-line processes to establish electrical interconnects.

The VC-MRAM chip integrates readout circuitry and high-speed write circuitry capable of generating pulse widths of sub-nanosecond. A scan-based interface is used to configure both the pulse delay and pulse width for precise timing control during write operations.

Probability Measurement

The P_{sw} versus magnetic field characteristics under varying voltage biases (Fig. 2c) were obtained using a custom-designed ultrafast probe card system. This system enables rapid, high-precision measurements of resistance-magnetic field (R-H) hysteresis loops. For each voltage bias condition, we performed 500 consecutive measurements to ensure statistical robustness. The VC-MRAM test system, incorporating a Xilinx PYNQ-Z2 FPGA development board and a custom-designed PCB hosting the integrated VC-MRAM chip, was used to measure the chip's probabilistic switching characteristics. The FPGA (part of the Zynq-7000 SoC on the PYNQ-Z2 board) orchestrated the measurement sequence: controlling W/R timing for the MRAM chip, setting write voltages via separate on-PCB DACs (TLV5618A), triggering the MRAM's on-chip pulse generator to apply voltage pulses of specific widths to the cells, and reading the resulting cell states through the Xilinx 7 series XADC module located on the PYNQ-Z2 board. Switching probabilities were determined by executing 1000 W/R cycles per cell for every tested pulse width setting.

Implementation of VSIM

Extended Data Fig. 1 illustrates a closed-loop VSIM demonstration platform built with a VC-MRAM chip, a custom-designed PCB, and a Xilinx PYNQ-Z2 FPGA development board. The process begins in the PYNQ-Z2's processing system (PS), where a NP-hard COP is mapped to the Ising model to determine parameters J and h . These parameters are then transferred to the PYNQ-Z2's programmable logic (PL) via the Advanced eXtensible Interface bus. The PL performs MVM operations to calculate the required pulse width and subsequently configures the MRAM's pulse generator accordingly. After triggering a write pulse, the PL reads the resulting spin states from the MRAM. Once the maximum iteration count is reached, the PL sends the final spin states back to the PS, and the corresponding COP solution is displayed on a screen.

Evaluation of device variations

To investigate the impact of device variations, we utilized a full-FPGA implementation of the Ising machine for simulation. In contrast to the VSIM platform which employs VC-MRAM as Ising spins, this FPGA implementation emulates spin behavior using LFSRs, sigmoid LUTs, and digital comparators. Other operations such as MVM on the PL and the communication between the PS and PL were kept identical to VSIM. To simulate devices with varying degrees of sub-100% PSP, the maximum output values of the sigmoid LUTs were configured from 50% to 90% in 10% increments. For each specific configuration of PSP and maximum iteration count, the full-FPGA Ising machine attempted to solve a target COP 100 times to determine the corresponding success probability.

References

1. Korte, B. & Vygen, J. *Combinatorial Optimization: Theory and Algorithms*. vol. 21 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2018).
2. Ichikawa, K., Ohuchi, S., Ueno, K. & Yokoyama, T. Accelerating optimal elemental configuration search in crystal using Ising machine. *Phys. Rev. Res.* **6**, 033321 (2024).
3. Parizy, M., Sadowski, P. & Togawa, N. Cardinality Constrained Portfolio Optimization on an Ising Machine. in *2022 IEEE 35th International System-on-Chip Conference (SOCC)* 1–6 (IEEE, Belfast, United Kingdom, 2022).
4. Bao, S., Tawada, M., Tanaka, S. & Togawa, N. An Ising-Machine-Based Solver of Vehicle Routing Problem With Balanced Pick-Up. *IEEE Trans. Consum. Electron.* **70**, 445–459 (2024).
5. Singh, A. K. *et al.* Uplink MIMO Detection Using Ising Machines: A Multi-Stage Ising Approach. *IEEE Trans. Wirel. Commun.* **23**, 17037–17053 (2024).
6. Xiao, W., Zhang, T., Qian, X., Han, J. & Qian, W. Efficient Approximate Decomposition Solver using Ising Model. in *Proceedings of the 61st ACM/IEEE Design Automation Conference* 1–6 (ACM, San Francisco CA USA, 2024).
7. Cipra, B. A. An Introduction to the Ising Model. *Am. Math. Mon.* **94**, 937–959 (1987).
8. Mohseni, N., McMahon, P. L. & Byrnes, T. Ising machines as hardware solvers of combinatorial optimization problems. *Nat. Rev. Phys.* **4**, 363–379 (2022).
9. Johnson, M. W. *et al.* Quantum annealing with manufactured spins. *Nature* **473**, 194–198 (2011).
10. King, A. D. *et al.* Beyond-classical computation in quantum simulation. *Science* **388**, 199–204 (2025).
11. Hamerly, R. *et al.* Experimental investigation of performance differences between coherent Ising machines and a quantum annealer. *Sci. Adv.* **5**, eaau0823 (2019).
12. King, A. D. *et al.* Coherent quantum annealing in a programmable 2,000 qubit Ising chain. *Nat. Phys.* **18**, 1324–1328 (2022).
13. Inagaki, T. *et al.* A coherent Ising machine for 2000-node optimization problems. *Science* **354**, 603–606 (2016).
14. Honjo, T. *et al.* 100,000-spin coherent Ising machine. *Sci. Adv.* **7**, eab0952 (2021).
15. Roques-Carmes, C. *et al.* Heuristic recurrent algorithms for photonic Ising machines. *Nat. Commun.* **11**, 249 (2020).
16. Hua, S. *et al.* An integrated large-scale photonic accelerator with ultralow latency. *Nature* **640**, 361–367 (2025).
17. Huang, K.-P., Nien, C.-F., Zhang, Y.-T., Lee, C.-K. & Wang, Y.-C. GPU-based Ising Machine for Solving Combinatorial Optimization Problems with Enhanced Parallel Tempering Techniques. in *2024 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)* 636–640 (2024).
18. Tatsumura, K., Yamasaki, M. & Goto, H. Scaling out Ising machines using a multi-chip architecture for simulated bifurcation. *Nat. Electron.* **4**, 208–217 (2021).
19. Yamamoto, K. *et al.* 7.3 STATICA: A 512-Spin 0.25M-Weight Full-Digital Annealing Processor with a Near-Memory All-Spin-Updates-at-Once Architecture for Combinatorial Optimization with Complete Spin-Spin Interactions. in *2020 IEEE International Solid-State Circuits Conference - (ISSCC)* 138–140 (IEEE, San Francisco, CA, USA, 2020).
20. Takemoto, T. *et al.* 4.6 A 144Kb Annealing System Composed of 9×16Kb Annealing Processor Chips with Scalable Chip-to-Chip Connections for Large-Scale Combinatorial Optimization Problems. in *2021 IEEE International Solid-State Circuits Conference (ISSCC)* 64–66 (IEEE, San Francisco, CA, USA, 2021).
21. Su, Y., Kim, T. T.-H. & Kim, B. 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)* 1–3 (IEEE, San Francisco, CA, USA, 2022).
22. Bae, J., Shim, C. & Kim, B. 15.6 e-Chimera: A Scalable SRAM-Based Ising Macro with Enhanced-Chimera Topology for Solving Combinatorial Optimization Problems Within Memory. in *2024 IEEE International Solid-State Circuits Conference (ISSCC)* 286–288 (IEEE, San Francisco, CA, USA, 2024).
23. Chu, Y.-C., Lin, Y.-C., Lo, Y.-C. & Yang, C.-H. 30.4 A Fully Integrated Annealing Processor for Large-Scale Autonomous Navigation Optimization. in *2024 IEEE International Solid-State Circuits Conference (ISSCC)* 488–490 (IEEE, San Francisco, CA, USA, 2024).
24. Wu, Z. *et al.* 37.5 SKADI: A 28nm Complete K-SAT Solver Featuring Dual-Path SRAM-Based Macro and Incremental Update with 100% Solvability. in *2025 IEEE International Solid-State Circuits Conference (ISSCC)* 614–616 (IEEE, San Francisco, CA, USA, 2025).
25. Singh, N. S. *et al.* CMOS plus stochastic nanomagnets enabling heterogeneous computers for probabilistic inference and learning. *Nat. Commun.* **15**, 2685 (2024).
26. Li, M.-C. *et al.* 12.2 p-Circuits: Neither Digital Nor Analog. in *2025 IEEE International Solid-State Circuits Conference (ISSCC)* 1–3 (IEEE, San Francisco, CA, USA, 2025).

27. Xiao, Z. *et al.* In-Memory Neural Stochastic Differential Equations with Probabilistic Differential Pair Achieved by In-Situ P-Bit Using CMOS Integrated Voltage-Controlled Magnetic Tunnel Junctions. in *2024 IEEE International Electron Devices Meeting (IEDM)* 1–4 (IEEE, San Francisco, CA, USA, 2024).
28. Ahmed, I., Chiu, P.-W., Moy, W. & Kim, C. H. A Probabilistic Compute Fabric Based on Coupled Ring Oscillators for Solving Combinatorial Optimization Problems. *IEEE J. Solid-State Circuits* **56**, 2870–2880 (2021).
29. Moy, W. *et al.* A 1,968-node coupled ring oscillator circuit for combinatorial optimization problem solving. *Nat. Electron.* **5**, 310–317 (2022).
30. Lo, H., Moy, W., Yu, H., Sapatnekar, S. & Kim, C. H. An Ising solver chip based on coupled ring oscillators with a 48-node all-to-all connected array architecture. *Nat. Electron.* **6**, 771–778 (2023).
31. Borders, W. A. *et al.* Integer factorization using stochastic magnetic tunnel junctions. *Nature* **573**, 390–393 (2019).
32. Aadit, N. A. *et al.* Computing with Invertible Logic: Combinatorial Optimization with Probabilistic Bits. in *2021 IEEE International Electron Devices Meeting (IEDM)* 40.3.1-40.3.4 (IEEE, San Francisco, CA, USA, 2021).
33. Si, J. *et al.* Energy-efficient superparamagnetic Ising machine and its application to traveling salesman problems. *Nat. Commun.* **15**, 3457 (2024).
34. Nikhar, S., Kannan, S., Aadit, N. A., Chowdhury, S. & Camsari, K. Y. All-to-all reconfigurability with sparse and higher-order Ising machines. *Nat. Commun.* **15**, 8977 (2024).
35. Yin, J. *et al.* Scalable Ising Computer Based on Ultra-Fast Field-Free Spin Orbit Torque Stochastic Device with Extreme 1-Bit Quantization. in *2022 International Electron Devices Meeting (IEDM)* 36.1.1-36.1.4 (IEEE, San Francisco, CA, USA, 2022).
36. Yang, C.-Y. *et al.* Dual-Function Unipolar Top-pSOT-MRAM for All-Spin Probabilistic Computing with Ultra-Dense Coupling and Adaptive Temporal Coding. in *2024 IEEE International Electron Devices Meeting (IEDM)* 1–4 (IEEE, San Francisco, CA, USA, 2024).
37. Dutta, S. *et al.* An Ising Hamiltonian solver based on coupled stochastic phase-transition nano-oscillators. *Nat. Electron.* **4**, 502–512 (2021).
38. Maher, O. *et al.* A CMOS-compatible oscillation-based VO₂ Ising machine solver. *Nat. Commun.* **15**, 3334 (2024).
39. Cai, F. *et al.* Power-efficient combinatorial optimization using intrinsic noise in memristor Hopfield neural networks. *Nat. Electron.* **3**, 409–418 (2020).
40. Jiang, M., Shan, K., He, C. & Li, C. Efficient combinatorial optimization by quantum-inspired parallel annealing in analogue memristor crossbar. *Nat. Commun.* **14**, 5927 (2023).
41. Yue, W. *et al.* A scalable universal Ising machine based on interaction-centric storage and compute-in-memory. *Nat. Electron.* **7**, 904–913 (2024).
42. Yin, X. *et al.* Ferroelectric compute-in-memory annealer for combinatorial optimization problems. *Nat. Commun.* **15**, 2419 (2024).
43. Mackenzie, N. D. & Young, A. P. Statics and dynamics of the infinite-range Ising spin glass model. *J. Phys. C Solid State Phys.* **16**, 5321–5337 (1983).
44. Shao, Y. & Khalili Amiri, P. Progress and Application Perspectives of Voltage - Controlled Magnetic Tunnel Junctions. *Adv. Mater. Technol.* **8**, 2300676 (2023).
45. Laudis, L. L., Shyam, S., Suresh, V. & Kumar, A. A Study: Various NP-Hard Problems in VLSI and the Need for Biologically Inspired Heuristics. in *Recent Findings in Intelligent Computing Techniques* (eds. Sa, P. K., Bakshi, S., Hatzilygeroudis, I. K. & Sahoo, M. N.) vol. 708 193–204 (Springer Singapore, Singapore, 2018).
46. Fouilhoux, P. & Mahjoub, A. R. Solving VLSI design and DNA sequencing problems using bipartition of graphs. *Comput. Optim. Appl.* **51**, 749–781 (2012).
47. Liu, K. & Dinneen, M. J. Solving the Bounded-Depth Steiner Tree Problem using an Adiabatic Quantum Computer. in *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)* 1–9 (IEEE, Melbourne, Australia, 2019).
48. Lin, S.-E. D. & Kim, D. H. Construction of All Rectilinear Steiner Minimum Trees on the Hanan Grid and Its Applications to VLSI Design. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **39**, 1165–1176 (2020).
49. Fowler, A. Improved QUBO Formulations for D-Wave Quantum Computing. (University of Auckland, 2017).
50. Huang-Yu Chen, Szu-Jui Chou, Sheng-Lung Wang, & Yao-Wen Chang. A Novel Wire-Density-Driven Full-Chip Routing System for CMP Variation Control. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* **28**, 193–206 (2009).
51. Jun-Dong Cho, Raje, S. & Sarrafzadeh, M. Fast approximation algorithms on maxcut, k-coloring, and k-color ordering for VLSI applications. *IEEE Trans. Comput.* **47**, 1253–1266 (1998).

Acknowledgements

We acknowledge financial support from the Beijing Natural Science Foundation (QY24139), National Natural Science Foundation of China (62404015), the Fundamental Research Funds for the Central Universities and the Beijing Outstanding Young Scientist Program.

Authors' contributions

W.Z. and S.L. initialized and supervised the project. Y. Z. and S.L. contributed equally to this work. A.L., Z.Z. and D.W. fabricated the VC MRAM chip. Y. Z. and S.L. performed the measurements and the implementation of VSIM with the help from A.L., L.Z., P.W. and L.G.. Y. Z., S.L., A.L. and W.Z. drafted the manuscript. All authors discussed the results and implications.

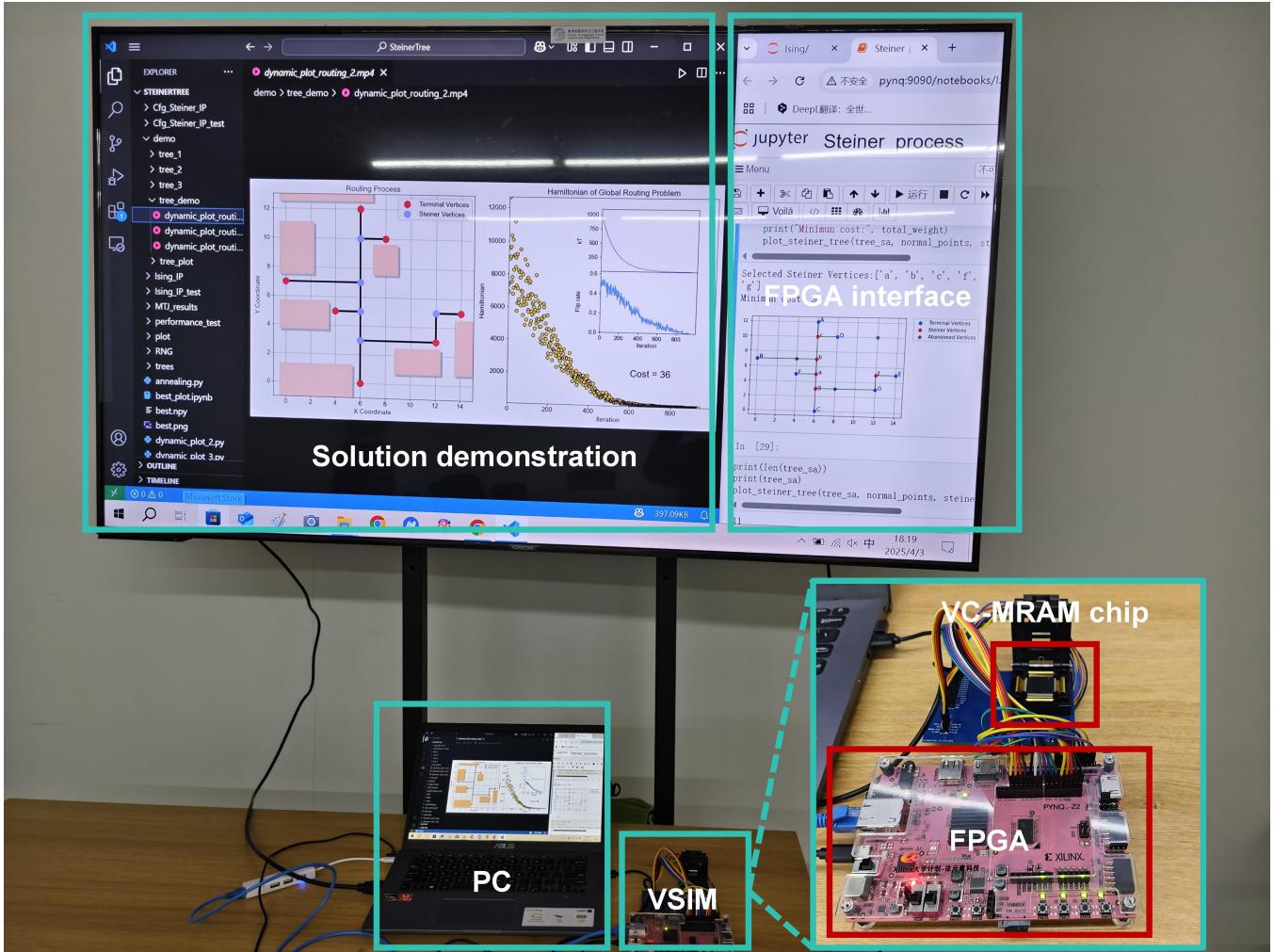
Competing interests

The authors declare no competing interests.

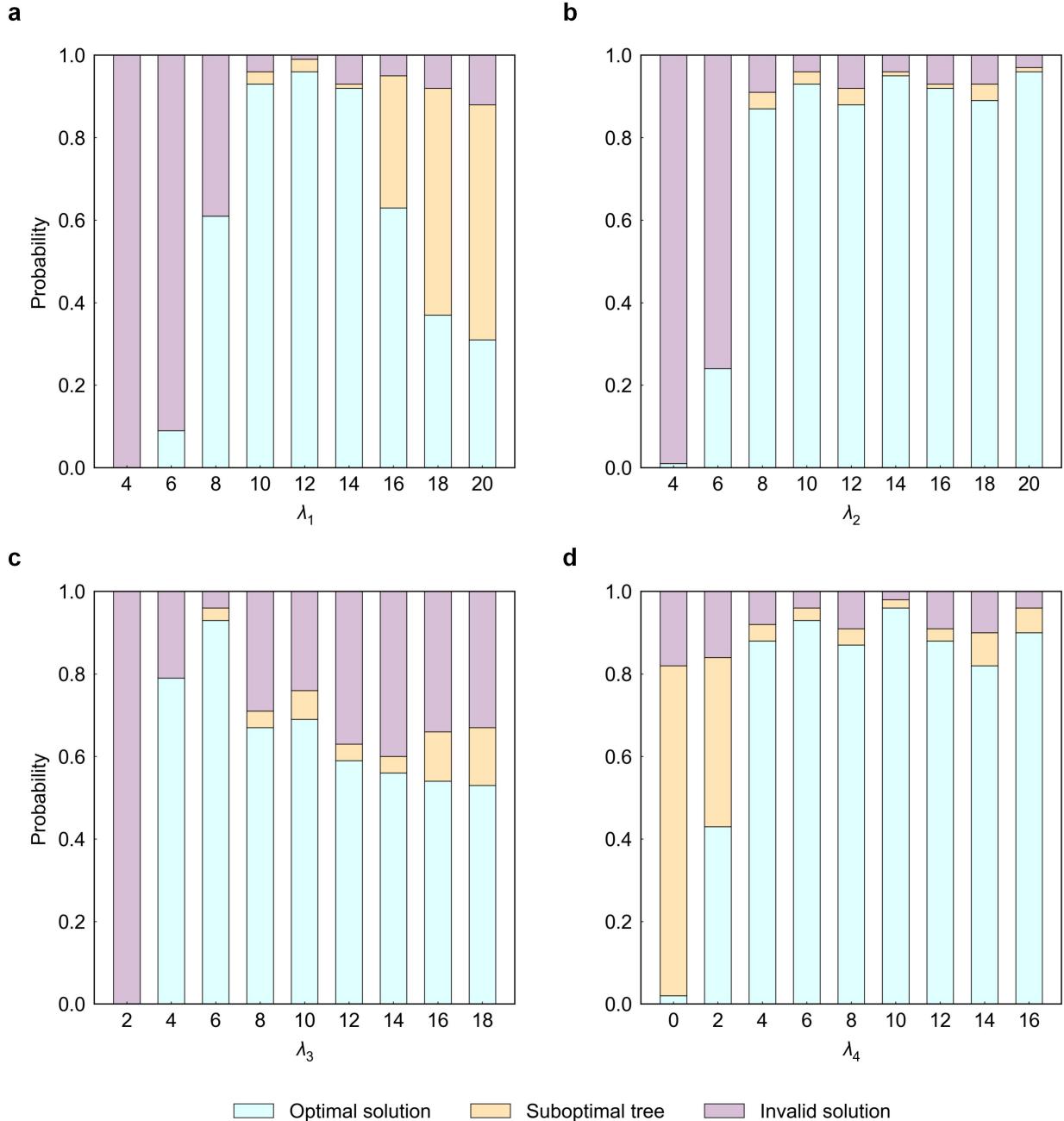
Additional information

1. Supplementary information word
2. Supplementary videos for solving two EDA problems
3. Supplementary Excel data for device endurance test

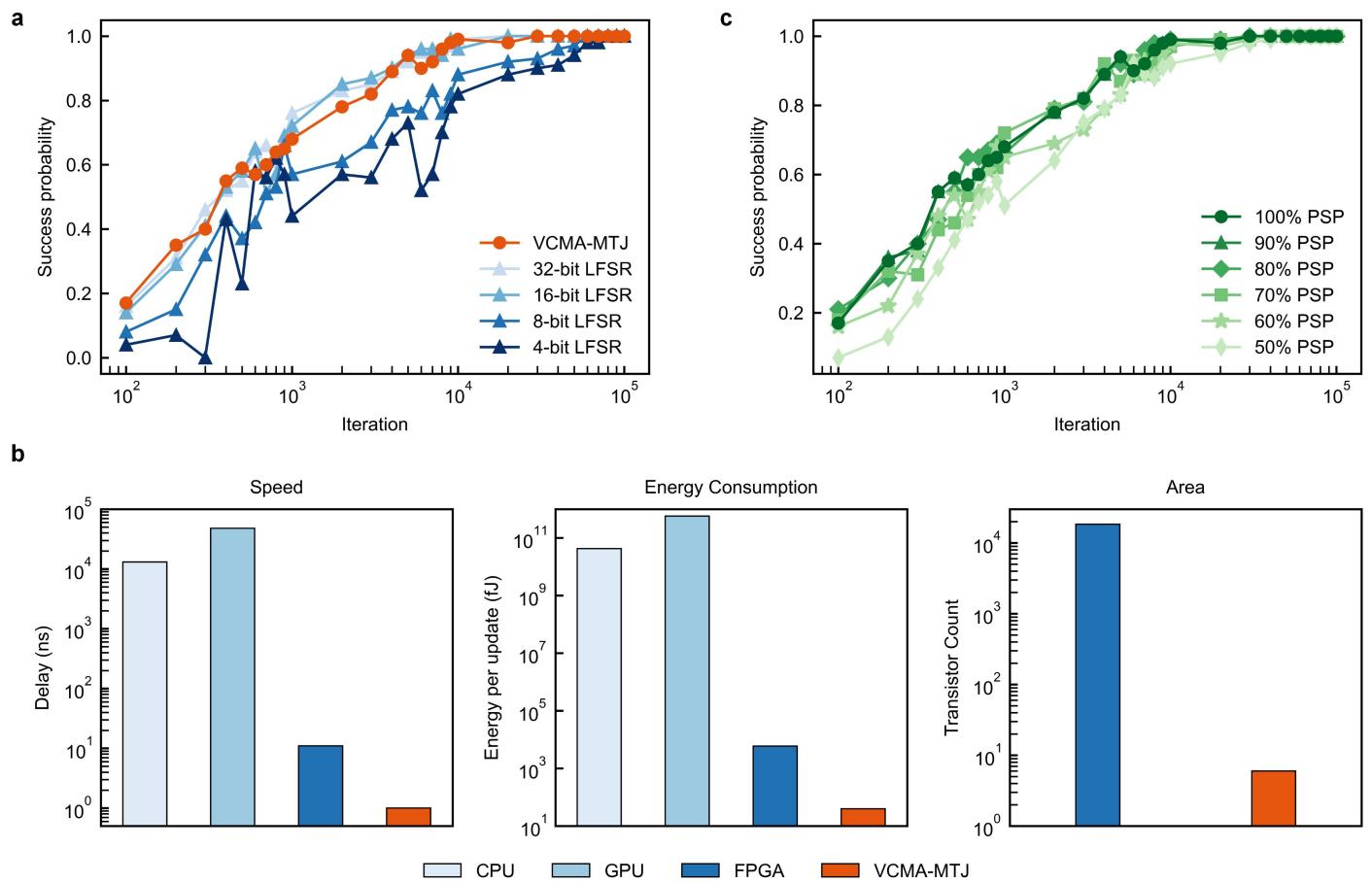
Correspondence and requests for materials should be addressed to Sai Li or Weisheng Zhao



Extended Data Fig. 1 VSIM demonstration platform. This system comprises a VC-MRAM chip mounted on a custom PCB with a Xilinx PYNQ-Z2 FPGA board, all controlled by a host PC. Operation begins on the PC, where a COP is mapped to an Ising model and parameters are configured using a Jupyter Notebook interface. These configurations are transferred to the FPGA, which then drives the VC-MRAM chip through iterative Ising computations. Finally, the resulting spin states, corresponding to the COP solution, are returned to the PC for visualization.



Extended Data Fig. 2 Influence of penalty coefficients on global routing solution probability. Baseline coefficients are set to $\lambda_1 = \lambda_2 = 10$ and $\lambda_3 = \lambda_4 = 6$. Each panel (a-d) shows the probability variation when only one coefficient is changed. Effects of coefficient values: (i) Low λ_1 , λ_2 , and λ_3 lead to invalid solutions. (ii) Low λ_4 yields suboptimal trees, highlighting its importance for finding the ground state. (iii) High λ_1 or λ_3 reduce success probability due to conflicts among constraints or with the optimization objective. (iv) High λ_2 or λ_4 do not reduce success probability, implying no critical conflicts.



Extended Data Fig. 4 Performance evaluation of VCMA-MTJ as Ising spin for global routing applications. **a**, Success probability versus iterations comparing VCMA-MTJ and LFSRs with varying bit-widths. **b**, Hardware performance benchmarking (speed, area, energy consumption) of VCMA-MTJ compared to CPU (Intel Xeon Platinum 8352V), GPU (NVIDIA GeForce RTX 4090) and FPGA (Zynq-7000) implementations. **c**, System robustness against device-to-device variations, demonstrating that success probability remains largely unaffected when PSP exceeds 60%, ensured by the calibration.