A Continuous-Time Ising Machine Using Coupled Inverter Chains Featuring Fully-Parallel One-Shot Spin Updates

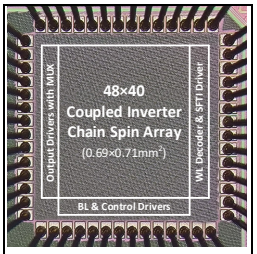
Chengshuo Yu[1,2], Junjie Mu[1], Kevin Chai[2], Tony Kim[1], Bongjin Kim[3]

[1]Nanyang Technological University, [2] Institute of Microelectronics, A*STAR, Singapore, [3]University of California, Santa Barbara, CA

Unconventional computers based on superconducting qubits [1] and analog/digital hardware accelerators [2-5] have been developed and solved hard combinatorial optimization problems (COPs) with higher efficiency than classical computers by integrating the Ising models in hardware and observing their natural convergence behavior towards the lower Ising Hamiltonian (or energy). Quantum annealer based on superconducting qubits [1] has demonstrated its capability of finding solutions to COPs using the quantum tunneling effect. However, the quantum computer requires operating at ultra-low temperatures (15-20mK), consumes excessive power (>10kW), and is not scalable yet to solve real-world problems. Recently, ASIC hardware accelerators [2-5] using low-cost CMOS technologies have implemented the Ising model by integrating many physical spins (i.e., computing units of the Ising model) to accelerate the finding of solutions to hard COPs. The hardware accelerators, so-called Ising machines, solved hard COPs much faster than classical computers and consumed several orders of magnitude lower energy. Nevertheless, they still have to overcome the fundamental scalability issues for solving large-scale COPs in the real world. The performance of existing discrete-time Ising machines has been limited by the exponentially increasing number of operation cycles when the problem size increases for the large-scale COPs. In addition, prior works have not been able to integrate massive random number generators (RNG), the essential building blocks of the Ising machines for finding better solutions, that could occupy a significant hardware footprint. Prior works have implemented them off-chip, or a few have been integrated on-chip and shared by many spins. This paper proposes a novel continuous-time analog Ising machine based on coupled inverter chain circuits to address the scalability issues of the existing discrete-time Ising machines. Besides the benefits of the continuous-time operation, which significantly lowers the computing energy and latency, we introduce an equalization method for finding better solutions without RNGs.

Fig. 1 (left) compares the workflows of the prior discrete-time and the proposed continuous-time Ising machines. A COP is first mapped to the Ising model, a network of spins and interactions between them. Then, the Ising model is mapped to the hardware Ising machine by assigning spins and programming their interaction coefficients. The prior discrete-time Ising machine iterates the sequential operation of spin operation, random number generation, and the readout of spins until the Ising Hamiltonian (i.e., a key performance indicator, where the lower Hamiltonian indicates the better solution) converges to the lowest possible value (i.e., the optimal solution). In contrast, this work adopts a continuous-time Ising machine based on physical couplings between inverter chains that converges to the lowest Hamiltonian in one shot operation. Therefore, we can achieve much faster TTS than prior works based on iterative discrete-time operations. Fig. 1 (right) summarizes the limitations of prior works and the key features of the proposed Ising machine. The proposed coupled inverter chain circuit minimizes the overhead of hardware spin implementations. Besides, the continuous-time operation enables all spins to operate in parallel, while prior discrete-time machines cannot update connected spins in parallel to prevent possible oscillation in the spin states. Instead of adding random fluctuations for annealing operations using RNGs, we equalize spin circuits to find better solutions at the early stage of the Ising operation while the analog spins resolve their spin states.

Fig. 2 (top) describes the concept of the proposed Ising machine with an example of a pair of coupled inverter chain spins (σ_1 and σ_2). The interaction coefficient (i.e., $J_{1,2}$ = +1, 0, or -1) is programmed using three pairs of switches and two inverters in the coupling module. The spin state is represented by an internal circuit node voltage (i.e., the spin state is +1 or -1 when the voltage of the corner nodes is H or L, where H/L represents the supply or the ground level). Note that the intermediate voltage level represents a superposed spin state when the spins interact or at the equalization mode. Fig. 2 (top-right) shows a detailed schematic diagram of the proposed spin circuit. A compact spin integrates a core inverter chain with eight equalization switches and two coupling modules. The inverter chain spin interacts with four neighboring spins, updates, and stores the spin state as internal node voltages at the corner of the inverter chain. Note that all eight internal node voltages are equalized to a half VDD (core supply) so that it has a 50% probability to converge to a spin-up or spin-down state before spins interact. A 6-bit SRAM embedded per spin circuit stores two 3bit interaction coefficients to program couplings between spins. A register is also embedded per spin to read out spin states by forming a shift register when it combines with other registers from the spins in the same row of the two-dimensional spin array (Fig. 3). Fig. 2 (bottom) illustrates the equalization (EQ) and interaction (ITR) modes of the proposed Ising machine. Four spins in a lattice graph topology are shown with their interaction coefficients for simplicity. Initially, the disconnected spin states are random based on process variation. Then, we equalize spin states by connecting the inputs and outputs of the inverters by enabling eight EQ switches in the inverter chain. As a result, the spin states are pushed to the superposed state (half-VDD) between spin-up (VDD or σ=+1) and spin-down (GND or σ=-1). After that, we disable the EQ switches and connect the spins based on the programmed coefficients by enabling the ITR signal for spins to start their interactions. Then, the spins update their states, lowering the overall Ising Hamiltonian.



65nm die micrograph

Fig. 3 shows the overall architecture of the proposed 65nm test chip with 1,920 (=48×40) integrated spins. Besides the core spin array, the test chip includes a wordline decoder and a bitline driver for 288× wordlines and 40× bitlines to program 11.25Kb coefficient SRAM. In addition, a control signal driver sends EQ and ITR signals to all spins, and a readout circuit with 4:1 multiplexers reads out spin states.

Fig. 4 (top) shows the measured Hamiltonians and their distributions over 500 iterations with/without equalization. Here, we map a hard Max-Cut problem to the spin array of the fabricated test chip with random interaction coefficients. The distributions of the measured Hamiltonians and the simulated Ising model baseline are shown in Fig. 4 (top-left). The measured result with equalization shows lower Ising Hamiltonians on average with less variation than those without equalization, showing the effectiveness of the proposed equalization for better solutions. Fig. 4 (top-right) shows the measured transient Hamiltonian values. The simulated Hamiltonian values with transient spin maps are shown in Fig. 4 (bottom-left). The spin interaction coefficients are obtained based on the pre-defined ground state (with a spin map showing a pixel image '2023'), and we extracted transient spin maps (from 1ns to 15ns) and their Ising Hamiltonians. Note that all the spin states are equalized before the interaction begins. When the interaction begins, the extracted spin maps show equalized spin states (gray colors). Then, the spin states quickly converge to either a spin-up (white pixel) or a spin-down (black pixel), and finally, a clear image is shown at 15ns. Meanwhile, we can observe the transient Hamiltonian values converging to the ground state (i.e., H=-1984). Fig. 4 (bottom-right) plots the simulated TTS and its comparison with the baseline (i.e., sequential processing of the Ising model using the discrete-time computer operating at 200MHz). The result shows the TTS gains ranging from 12.6× for the smallest (2×2 spins) to 1478× for the largest spin array (32×32 spins) compared to the baseline.

Fig. 5 (top) shows the measured Ising Hamiltonians with core supply voltage swept from 0.6V to 1.2V for the hard Max-Cut problem with random interaction coefficients. We also measured five different test chips, and the results are shown in Fig. 5 (bottom). Fig. 6 compares the state-of-the-art Ising machines. The fabricated 65nm continuous-time Ising machine test chip with 1,920 spins consumes 15mW at 0.8V core supply and occupies 0.49mm^2.

Acknowledgement:

This work was supported by A*STAR under Grant A19E8b0102.

References:

[1] M. Johnson et al., *Nature*, vol. 473, pp. 194-198, 2011.
[2] T. Takemoto et al., *IEEE ISSCC*, pp. 52-53, 2019.
[3] Y. Su et al., *IEEE ISSCC*, pp. 480-481, 2020.
[4] T. Takemoto et al., *IEEE ISSCC*, pp. 64-66, 2021.
[5] I. Ahmed et al., *IEEE JSSC*, vol. 56, pp. 2870-2880, Sep. 2021.

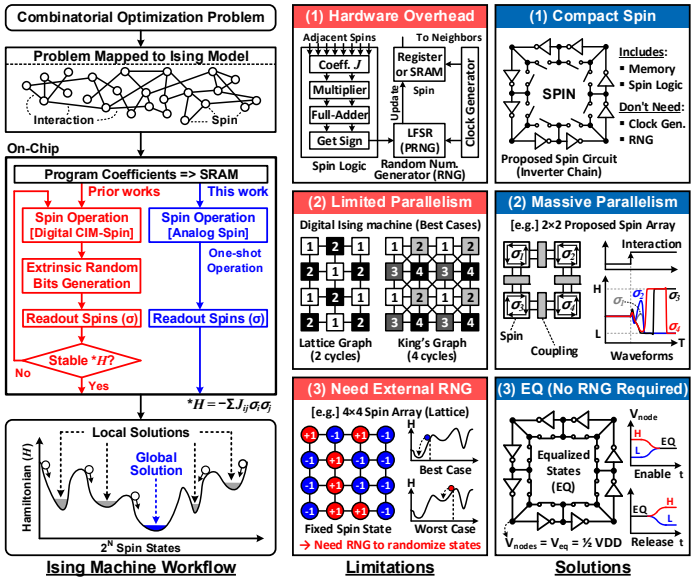


Fig. 1. Ising machine workflow (left) and design limitations of prior works and potential solutions (right).

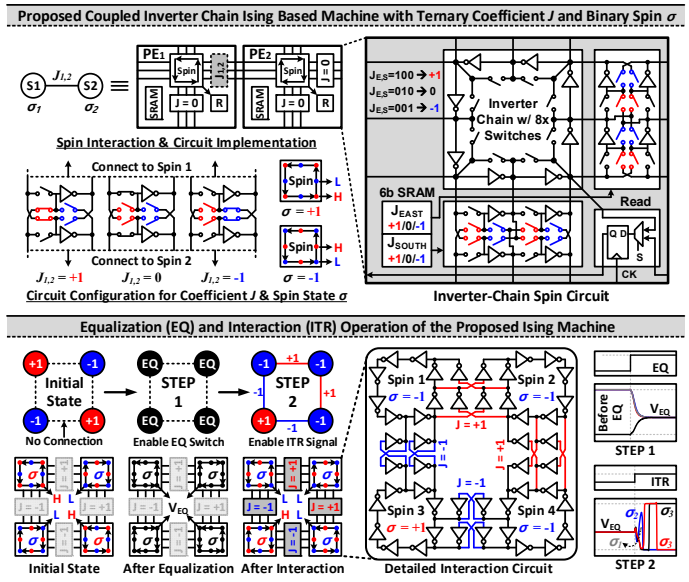


Fig. 2. Concept of the proposed Ising machine using inverter chain (top) and equalization and interaction operation (bottom).

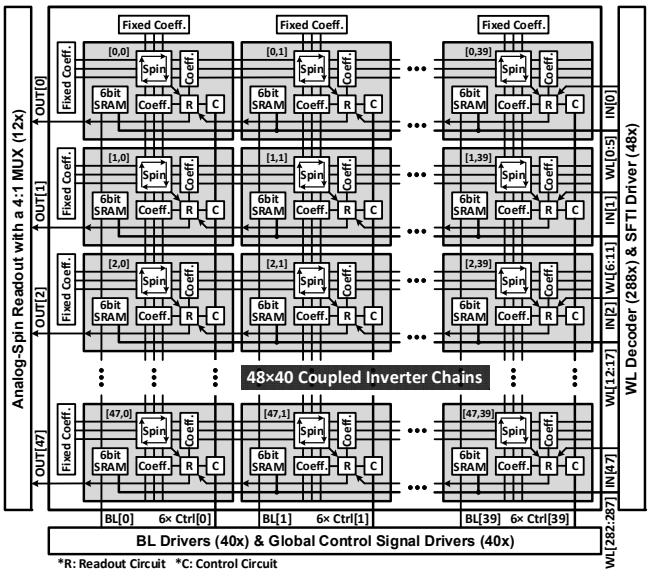


Fig. 3. Overall architecture of the proposed continuous-time Ising machine with 48x40 coupled-inverter chains.

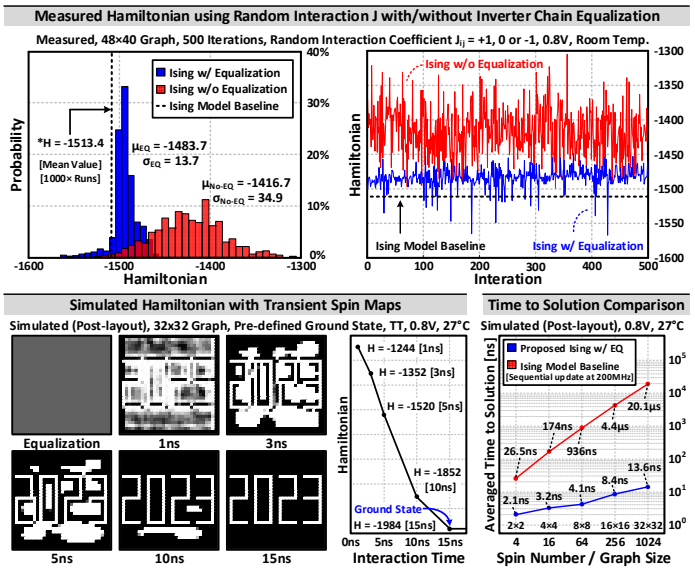


Fig. 4. Measured Hamiltonian with/without equalization (top) and simulated Hamiltonian with transient spin maps (bottom).

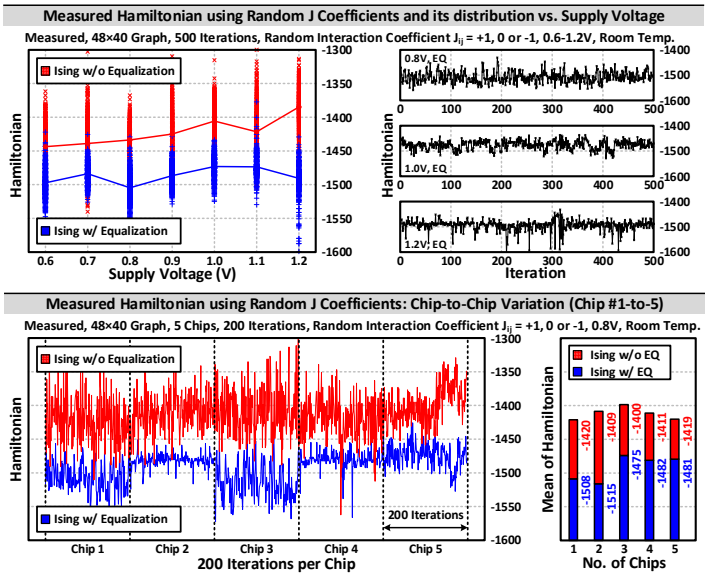


Fig. 5. Measured Hamiltonian using random J coefficients versus supply sweep (top) and chip-to-chip variation (bottom).

	D-Wave [1]	ISSCC'19 [2]	ISSCC'20 [3]	ISSCC'21 [4]	JSSC'21 [5]	This work
Technology (Circuit Type)	Super Conductor	40nm CMOS (Digital)	65nm CMOS (Digital)	40nm CMOS (Digital)	65nm CMOS (Mixed-Signal)	65nm CMOS (Analog)
Operating Temp.	15mK	Room Temp.	Room Temp.	Room Temp.	Room Temp.	Room Temp.
Spin Type	Qubit	SRAM	Register	Flip-Flop	Ring-Oscillator	Inverter-Chain
Random State	O	X	X	X	X	O
Spin Update Method	Qubit Interaction	Majority Voting	Digital MAC	Digital MAC	Oscillator Coupling	Inverter-Chain Coupling
Computing Type	Quantum Computing	Discrete-Time Digital	Discrete-Time Digital	Discrete-Time Digital	Continuous-Time Analog	Continuous-Time Analog
Ising Hamiltonian Lowering Method	Quantum Annealing	Simulated Annealing	Simulated Annealing	Simulated Annealing	No	Inverter-Chain Equalization
Natural Coupling	O	X	X	X	O	O
# of Spins	128-2K	30K	480	16K	560	1920
Core Area	N/A	10.42mm^2	0.403mm^2	10.812mm^2	0.53mm^2	0.49mm^2
Spin Area	13,888μm^2	339μm^2	832μm^2	552μm^2	950μm^2	224μm^2
Power Supply	N/A	1.1V	0.5-1.2V	1.1V	1V	0.6-1.2V
Power@Supply (Frequency)	25kW	N/A	86.4μW@0.6V (64MHz)	N/A	23mW@1V	15mW@0.8V
Power/Spin	12.2W	N/A	0.18μW	N/A	41μW	7.8μW
Time to Solution (Problem Size)	N/A	22μs (30K Spins)	30 Cycles (480 Spins)	4ms (9×16K Spins)	1-10μs (560 Spins)	<30ns (1920 Spins)

Fig. 6. Comparison with state-of-the-art Ising machines.