A Novel Structure to Realize Memory Cell Fluctuation Under Voltage Control in Ising Chip

Zhi WANG[1,2], Shu-ming CHEN[1,2,*] and Jian ZHANG[1,2]

[1]College of Computer, National University of Defense Technology,

[2]State Key Laboratory of High Performance Computing, National University

of Defense Technology

[*]smchen_cs@163.com

Keywords: Ising chip, Indeterminate computing, Memory cell fluctuation, Voltage control.

Abstract. New computing architecture utilizing the noise of integrated circuit to realize indeterminate computing has been proposed previously to solve combinatorial optimization problems. But the solution accuracy of realizing indeterminate computing with circuit's noise performed significantly worse than that with external random number. Because large scale of time is spent to introduce randomness in reading memory cells by row, which reduces the effective computing time. This paper presents a novel structure to introduce probabilistic behavior with just one read operation. It improves the rate of effective computing time and can get close solution accuracy to the scheme using external random number but with less cost.

Introduction

Most combinatorial optimization problems belong to NP-hard class problems [1]. Generally, finding the global optimum solution to them in conventional computer needs exponential time. CMOS implementation of Ising computing has been proposed previously to solve this problem with its own convergence property [2]. It maps the problem to the coefficients of Ising model firstly and then activates the ground-state search by interaction between spins. Finally the solution is acquired through observing the ground state [3]. The purpose of Ising computer is to find a spin configuration that minimizes the energy of the system [4], which corresponds to the ground state [5]. But the energy profile of the Ising model has peaks and troughs as shown in Fig. 1. The energy goes down along the energy profile by interaction between spins but will fall into a local optimum and be stuck at these points in many cases [6]. Then the system needs to change to other states randomly to prevent being stuck at local optimum.
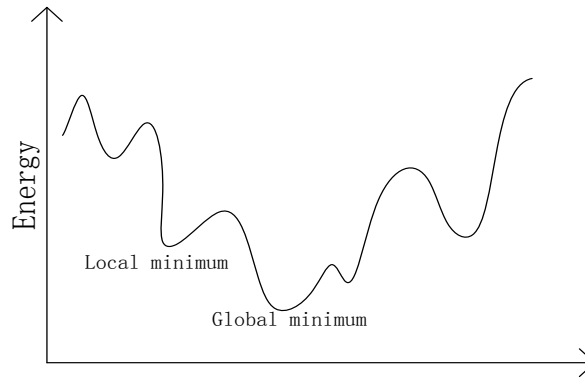
Figure 1. The energy profile of Ising model

A lot of randomization methods have been proposed to help the system to escape from the local optimum. Such as using highly amplified Gaussian thermal noise [7], using additional linear feedback shift register to generate a pseudorandom bit stream [8], using optoelectronic CMOS circuit [9], using external random number [2]. All the methods above need extra hardware cost to intentionally induce randomness, which will add the cost of VLSI implementation. A method utilizing the intrinsic noise of integrated circuit is proposed by Hitachi [2] to induce randomness without any extra hardware. It destroys the memory cells randomly by reading them under low supply voltage. But the solution accuracy of this method performed significantly worse than that using external random number. The reason is that common bit line is shared by a column and common word line is shared by a row in SRAM array, which makes it difficult to read all memory cells storing spin configuration by one time but a row at most. As a result, the rate of effective computing time is reduced.

This paper presents a novel structure to realize memory cell fluctuation under voltage control, which can intentionally induce memory error to all spin configurations with just one dummy read operation. In the proposed structure, all spin configurations are stored by new 8T-SRAM but the interaction coefficients are still stored by 6T-SRAM. All 8T-SRAMs share the common word line but have their own bit lines. Finally, it improves effective computing time and can get close solution accuracy to the scheme using external random number but with less cost.

Memory Cell Fluctuation under Voltage Control

The robustness of a SRAM cell can be expressed by immunity to noise, which is defined as the static noise margin (SNM) [10]. SNM quantifies the voltage noise necessary to flip the cell's state and reduces with the supply voltage [10,11]. It can be estimated by drawing and mirroring the inverter characteristics and finding the maximum possible square between them [12].

Fig. 2(a) shows the circuit diagram of a 6T SRAM. Assume that 0 is stored at Q and both bit lines are precharged to V_{DD} before reading. Upon the rise of WL, the BL line discharges through M6-M3 and the intermediate node Q is pulled up toward the precharged value of BL. This voltage rise of Q must stay low enough, which in the worst case could flip the cell [13,14]. Fig. 2(b) shows the comparison between stand-by SNM and read SNM, where read SNM is much smaller than stand-by SNM. So the bit error rate (BER) would increase when memory read operation is executed under low supply voltage than that without read operation.
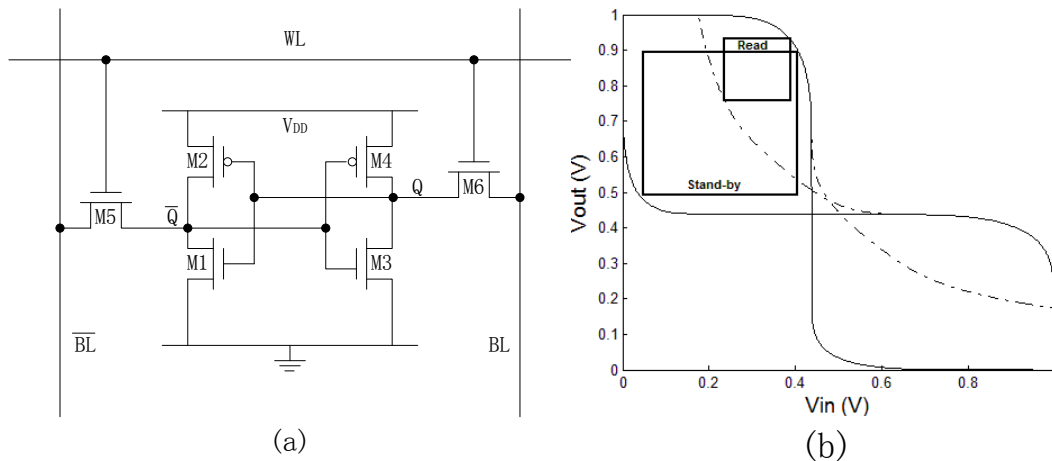
Figure 2. 6-T SRAM and graphical representation of SNM

Bit error rate of memory cells when read operation is executed has been measured in [6]. The flipping rate of the spin needs to reduce along with the computing time in the process of ground-state search [15], so we can make the Ising chip work between local search and randomly flipping the spin to satisfy the demand of ground-state search by controlling the supply voltage.

Proposed Design

The Ising chip is extended by the spin unit, which is connected to 5 nearest-neighbor spin units and forms a 2-layer 2-dimensional lattice [2]. A spin unit has a memory cell array to represent a spin and coefficients, and the decision circuit that determines the next state of the spin. We can access the memory cells via word lines, bit lines and IO circuits [2]. Common bit line is shared by a column and common word line is shared by a row in SRAM array, so a row of memory cells can be dummy read at most by one time under lower supply voltage. Then the effective computing time becomes less in the same solution time, which could be increased if the time to randomly flip all spin configurations can be reduced.

We have found that it is the memory cells that store spin configuration should be changed when the system falls into a local optimum but the coefficients can't be changed. So we can use a new 8-T SRAM shown in Fig. 3 to store the spin configuration for being controlled in an expedient way and the coefficients are still stored in 6-T SRAM. The increased two NMOS transistors are connected to global word line PWL and have their own bit lines. Each 8-T SRAM will not be influenced by other 8-T SRAMs for independent bit lines and can be controlled by the global word line PWL. However, the original word lines and bit lines are still used to read and write the spin configuration in normal situation.
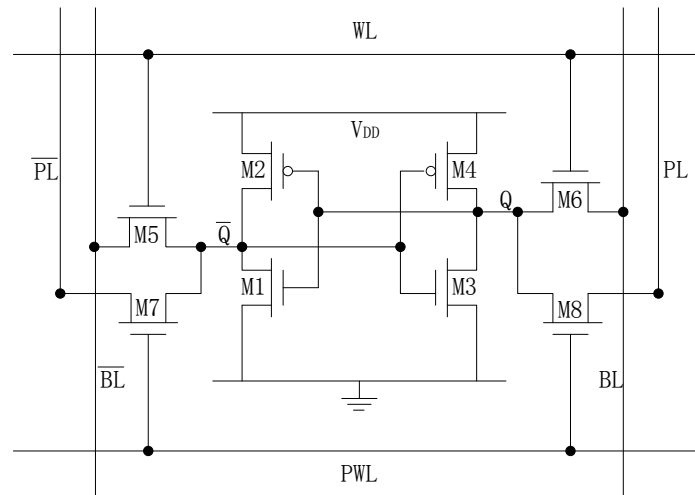
Figure 3. The proposed 8-T SRAM

As shown in Fig. 4, the word line PWL of all 8-T SRAMs are connected, but the increased bit lines are only connected to an 8-T SRAM. When flipping the spin configuration randomly is needed, the supply voltage is lowered to the required value at first. Then all of the increased bit lines are precharged to the supply voltage. Finally the global word line PWL are asserted to dummy read all spin configurations. As a result, we can destroy all spin configurations randomly by one dummy read operation to get more time for effective computing. When the problem size is N*M*2, the time to dummy read all spin configurations by the proposed structure is 1/N than that by original structure, which helps to increase the time for effective computing and then improves the final solution accuracy.
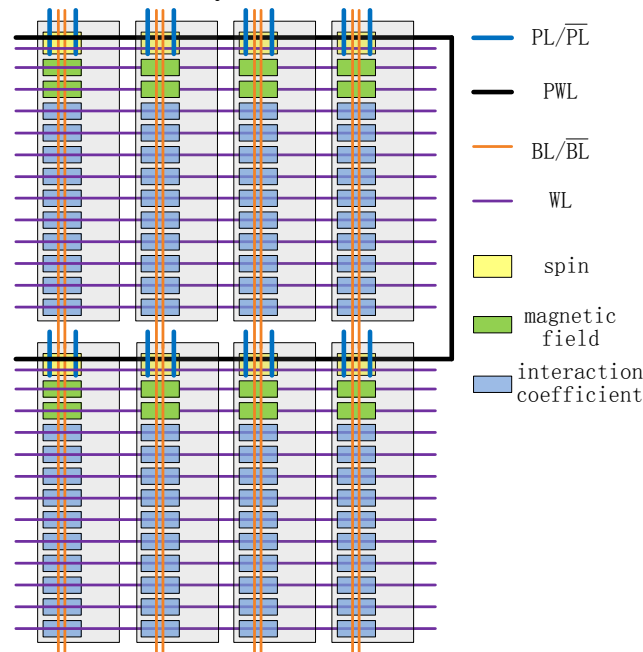

Figure 4. Connection of Ising chip

Results and Discussion

Fig. 5(a) shows the energy history of ground-state search in original structure, final energy stays high for limited effective computing time. When external random number is used to randomly flip the spin configuration, final energy becomes less for more

computing time shown in Fig. 5(b). The supply voltage for the proposed structure in ground-state search is shown in Fig. 5(c). Local search is executed when V_{DD} is 1V and dummy read is executed when it is less than 1V. Fig. 5(d) is the corresponding energy history. The final energy is close to that got from the scheme using external random number.
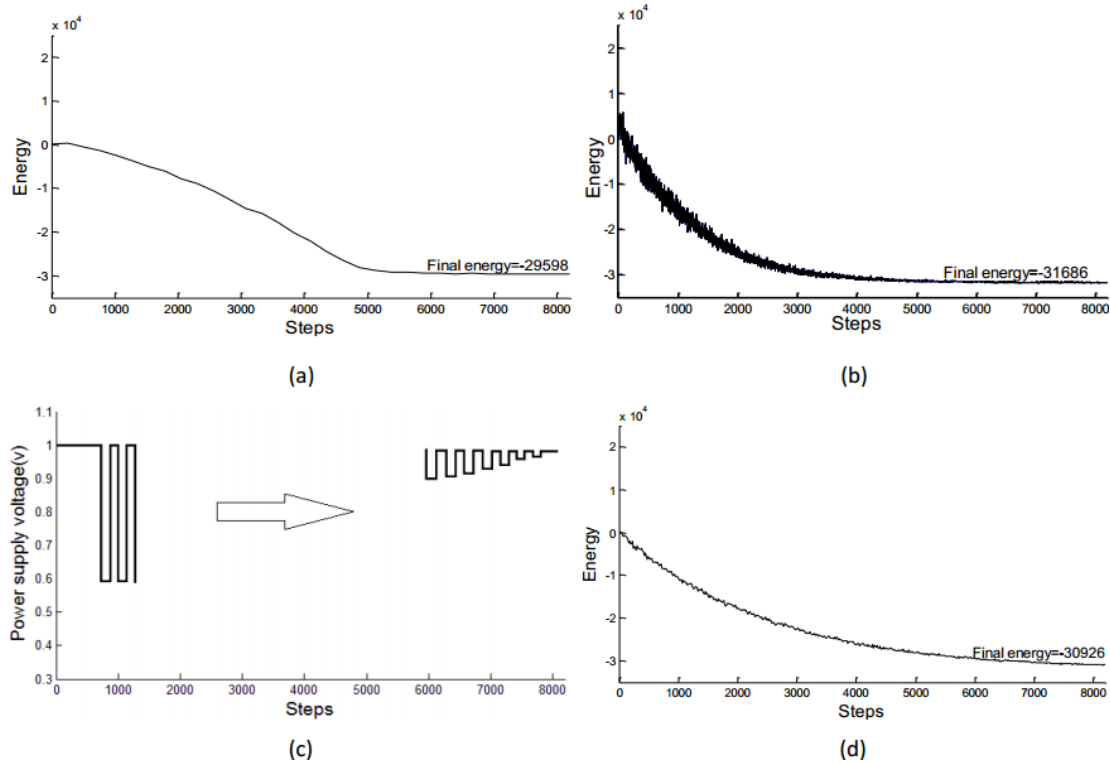

Figure 5. Energy and voltage history of ground-state search

The solution accuracy is reflected by the value of system energy. Lower energy stands for higher solution accuracy when the problem is the same size. But it is difficult to compare it when the size is different. Therefore, the relative energy defined in Eq. 1 is introduced as a metric for comparison.

$$R(s) = \frac{E(s)}{E(S_{Non-randomness})}. \tag{1}$$

Where E(s) is the energy got from the compared solution, E($s_{Non\text{-}randomness}$) is for local search. Larger R(s) shows higher solution accuracy.

Fig. 6(a) shows the solution accuracy comparison for different methods in different problem size when the ratio r of the coefficients +1 is 0.5. Fig. 6(b) shows the solution accuracy differs with r when the problem size is 20k. The proposed structure can improve the solution accuracy compared to the original one, which is close to that got from the scheme using external random number but with less hardware cost.
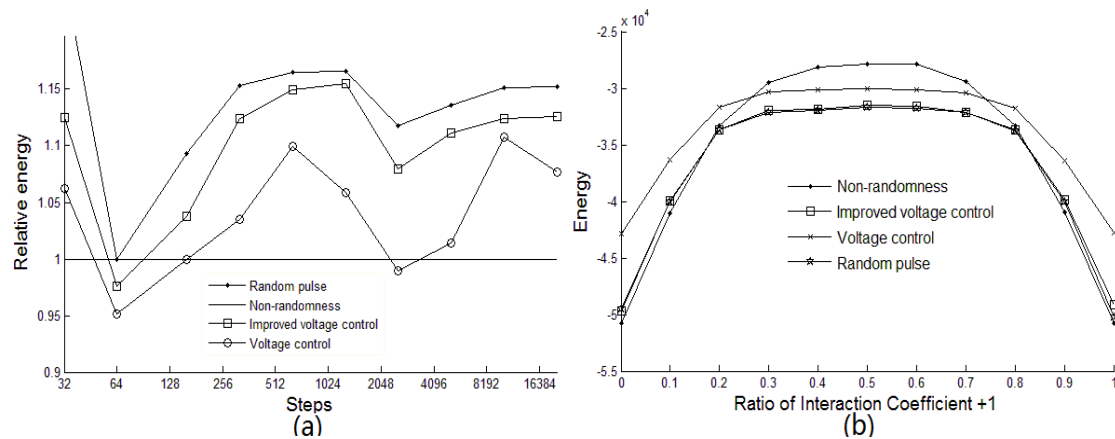
Figure 6. Comparison of solution accuracy

Conclusions

In this work, the robustness of 6T-SRAM is analyzed firstly. And it is found that the BER would increase when memory read is executed than that without it. Then we proposed to store the spin configuration with 8T-SRAM and set global word line and local bit lines to it for reducing the time to dummy read all spin configurations one time, which will increase effective computing time and improve solution accuracy. Final solution accuracy of the proposed new structure is compared to other methods as the problem varies. Results show that the proposed structure can improve solution accuracy than the original one, which is close to the scheme using external random number but with less hardware cost. In the future work, we should control the BER of the spin configuration to better satisfy the process of CMOS annealing, which will help to find better states in ground-state searches.

References

[1] J. K. Lenstra, AHGR. Kan, Computational complexity of discrete optimization problems, Annals of Discrete Mathematics. 4 (1979) 121-140.

[2] M. Yamaoka, C. Yoshimura, M. Hayashi, T. Okuyama, H. Aoki, H. Mizuno, 20k-spin Ising chip for combinational optimization problem with CMOS annealing, ISSCC. 24.3 (2015) 1-3.

[3] M. Yamaoka, C. Yoshimura, M. Hayashi, T. Okuyama, H. Aoki, H. Mizuno, A 20k-Spin Ising Chip to Solve Combinatorial Optimization Problems With CMOS Annealing, IEEE J. Solid-State Circuits. 50.1 (2016) 303-309.

[4] C. Yoshimura, M. Yamaoka, H. Aoki, H. Mizuno, Spatial computing architecture using randomness of memory cell stability under voltage control, ECCTD. (2013) 1-4.

[5] S. G. Brush, History of the Lenz-Ising model, Reviews of modern physics. 39.4 (1967) 883.

[6] C. Yoshimura, M. Yamaoka, M. Hayashi, T. Okuyama, H. Aoki, K. Kawarabayashi, H. Mizuno, Uncertain behaviours of integrated circuits improve computational performance, Scientific reports. 5 (2015).

[7] J. Alspector, B. Gupta, RB. Allen, Performance of a stochastic learning

Microchip, NIPS. (1989) 748-760.

[8] MB. Parker, R. Chu, A VLSI-efficient technique for generating multiple uncorrelated noise sources and its application to stochastic neural networks, IEEE Transactions on circuits and systems. 38.1 (1991) 109.

[9] A. Dupret, E. Belhaire, JC. Rodier, P. Lalanne, D. Prevost, P. Garda, P. Chavel, An optoelectronic CMOS circuit implementing a simulated annealing algorithm, IEEE Journal of Solid-State Circuits. 31.7 (1996) 1046-1050.

[10] E. I. Vatajelu, J. Figueras, Supply voltage reduction in SRAMs: Impact on static noise margins, AQTR. 1 (2008) 73-78.

[11] Da-Wei L, Jun-Rui Q, Shu-Ming C, The supply voltage scaled dependency of the recovery of single event upset in advanced complementary metal-oxide-semiconductor static random-access memory cells, Chinese Physics B. 22.2 (2013) 029402.

[12] E. Seevinck, F. J. List, J. Lohstroh, Static-noise margin analysis of MOS SRAM cells, IEEE Journal of solid-state circuits. 22.5 (1987) 748-754.

[13] J. Singh, D. K. Pradhan, S. Hollis, S. P. Mohanty, A single ended 6T SRAM cell design for ultra-low-voltage applications, IEICE Electronics Express. 5.18 (2008) 750-755.

[14] Peng L, Min-Xuan Z, Zhen-Yu Z, Quan D, Design of Analysis Platform Used for Studying Soft Error Characteristic of 3D SRAM. (2015).

[15] S. Kirpatrick, C. D. Gelett, M. P. Vecchi, Opimization by simulated annealing, Science. 220 (1983) 671.