

# シミュレーテッド・アニーリングに向けた ログ・エンコーディングによる ReRAM Computation-in-Memory

三澤奈央子<sup>1</sup> 松井千尋<sup>1</sup> 竹内健<sup>1</sup>

**概要:** 組合せ最適化問題を効率的に解くために, ReRAM Computation-in-Memory (CiM) のアレイ面積を削減し, ReRAM デバイスのエラーを考慮したログ・エンコーディングによる ReRAM CiM を提案する. 組合せ最適化問題の1つであるナップサック問題をシミュレーテッド・アニーリング (SA) を用いて ReRAM CiM で解く場合, 従来の ReRAM CiM へのマッピング方法では, ナップサック容量に応じて, アレイ面積が線形的に増加していた. そこで, ログ・エンコーディングによるマッピング方法を提案し, アレイ面積を 97.6%削減する. また, SA に向けて ReRAM デバイスの非線形なエラー特性を考慮することにより, 許容できる BER が 10 倍, 許容できるビット精度が 5 ビット精度まで向上する.

**キーワード:** シミュレーテッド・アニーリング, ReRAM, Computation-in-Memory

## 1. はじめに

組合せ最適化問題の1つであるナップサック問題をシミュレーテッド・アニーリング (SA) を用いて ReRAM Computation-in-Memory (CiM) で解く場合, 従来の ReRAM CiM へのマッピング方法では, ナップサック容量に応じてアレイ面積が線形的に増加する[1]. そこで, 本論文では, ログ・エンコーディングによりハミルトニアンを定式化し, CiM にマッピングする方法を提案する[2]. また, ReRAM デバイスの書き換え回数の増加に伴い起こる非線形なビット・エラーを, SA へのエラー注入により再現し, デバイスのエラーを最適化した ReRAM CiM を提案する[2].

## 2. ログ・エンコーディングを用いた SA による ReRAM CiM のアレイ面積の削減

ナップサック問題は, 選択したアイテムの総重量がナップサック容量以内でなければならないという制約があるため, ペナルティ項を含めハミルトニアンを定式化する必要がある. 提案のハミルトニアンは, ペナルティ項をログ・エンコーディングより定式化し (図 1(a)) [3], そのハミルトニアンを二次形式に変換して ReRAM CiM にマッピングする (図 1(b)). 二次形式のハミルトニアンのスピン  $q^T$ , QUBO 行列  $Q$ , スピン  $q$  が, それぞれ ReRAM CiM の Bit-line (BL), ReRAM のコンダクタンス, Word-line (WL) にマッピングされ, 積和演算を用いてハミルトニアンを計算する[4, 5].

図 2 に, アイテム数が 10, ナップサック容量が 100 のナップサック問題を例に, 提案のログ・エンコーディングと従来のリニア・エンコーディングによる ReRAM CiM のアレイ面積の比較を示す. 従来の ReRAM CiM が 110 WLs × 110 BLs であるのに対し, 提案の ReRAM CiM は 17 WLs × 17 BLs となり, 従来の ReRAM CiM よりも 97.6%, アレイ面積を削減できる.

**Knapsack problem** (e.g. items: 10, Knapsack capacity: 100)

Item No.	1	2	...	10	$W$ : Knapsack capacity
Value	$v_1$	$v_2$	...	$v_N$	$v_\alpha$ : Value of each item
Weight	$w_1$	$w_2$	...	$w_N$	$w_\alpha$ : Weight of each item

(a) Hamiltonian  $H$

$$H = -\underbrace{\sigma \sum_{\alpha=1}^N v_\alpha x_\alpha}_{\text{Cost function}} + \underbrace{\mu \left( W - \sum_{i=1}^{\lfloor \log_2 W - 1 \rfloor + 1} 2^i y_i - \sum_{\alpha=1}^N w_\alpha x_\alpha \right)^2}_{\text{Penalty function}} \quad \text{Eq. (1)}$$

**Cost function**      **Penalty function**

$x_\alpha \in \{0, 1\}$  Item is 0: unselected, 1: selected

$y_i \in \{0, 1\}$  Extra variable for total weight

$\sigma, \mu$ : hyper parameter

(b) Quadratic form  $H$

$$H = q^T Q q \quad \text{Eq. (2)}$$

Spin  $q$

$$q^T = [x_1, \dots, x_N, y_1, \dots, y_{\lfloor \log_2 W - 1 \rfloor}] \quad \text{Eq. (3)}$$

QUBO matrix  $Q$

$$Q = \begin{bmatrix} \mu w_1^2 - 2\mu W w_1 - \sigma v_1 & \dots & 2^{1+\lfloor \log_2 W - 1 \rfloor + 1} \mu w_1 \\ 0 & \dots & \vdots \\ \vdots & \dots & \vdots \\ 0 & \dots & 2^{2+\lfloor \log_2 W - 1 \rfloor + 1} \mu \end{bmatrix} \quad \text{Eq. (4)}$$

図 1 ログ・エンコーディングを用いたハミルトニアンの ReRAM CiM へのマッピング方法. (a) ナップサック問題を定式化したハミルトニアン. (b) 二次形式のハミルトニアン.

**Proposed Log-encoding**

$$q^T = [x_1, \dots, x_{10}, y_1, \dots, y_3] \quad \text{Eq. (5)}$$

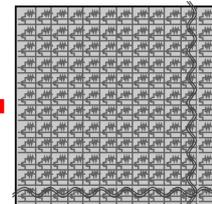
**Conventional linear-encoding**

$$q^T = [x_1, \dots, x_{10}, y_1, \dots, y_{100}] \quad \text{Eq. (6)}$$

ReRAM CiM



**97.6% Area Reduction**



110 BLs

図 2 提案のログ・エンコーディングと従来のリニア・エンコーディングによる ReRAM CiM のアレイ面積の比較

<sup>1</sup> 東京大学大学院  
 Graduate School of Engineering, The University of Tokyo

### 3. SA への ReRAM デバイスのエラー注入

図 3(a)に、ReRAM デバイスの測定にて得られた、書き換え回数における Bit-error rate (BER) を示す。高抵抗状態 (HRS) では、書き換え回数にかかわらずエラーが起こらない一方で、低抵抗状態 (LRS) ではエラーが起こる。図 3(b)に ReRAM にマッピングする QUBO 行列を示す。QUBO 行列の重み  $a$  を “0” もしくは “1” のバイナリ形式で表し、ReRAM の HRS もしくは LRS に記憶する。QUBO 行列は上三角行列のため、“0” が大半を占める。

ReRAM CiM を用いて SA を行う際に、Case1 として “0” を LRS, “1” を HRS, Case2 として “0” を HRS, “1” を LRS に記憶する場合を考える。言い換えれば、Case1 では、信頼性が低い LRS に記憶されている “0” にはエラーが発生するが、信頼性が高い HRS に記憶されている “1” はエラーは発生しない。Case2 では、その逆に “0” にはエラーが発生しないが、“1” にはエラーが発生する。図 3(c)に Case1, 図 3(d)に Case2 の BER とビット精度における SA の成功確率をそれぞれ示す。成功確率は、1 つの問題を 1,000 回試行し、最適解に到達した割合と定義した。Case1 では、BER が高くなると成功確率が低くなる。特にビット精度が低い場合、BER の増加に伴い、顕著に成功確率が低くなる。一方で、Case2 は、Case1 よりも全体的に成功確率が高く、BER の影響を受けにくい。

図 4 に、Case1 と Case2 の SA の成功確率を示す。90%以上の成功確率を基準として比較した。図 4(a)に 10 ビット精度のときの BER による比較を示す。許容できる BER は、Case1 では 1%だが、Case2 では 10%である。図 4(b)に BER が 10%のときのビット精度による比較を示す。Case1 では、10 ビット精度でも基準を満たすことができないが、Case2 では、5 ビット精度まで低くできる。QUBO 行列で大半を占める “0” を信頼性の高い HRS に記憶することで、許容できる BER が 10 倍向上し、ビット精度を 5 ビット精度まで許容できる。

### 4. おわりに

表 1 に、本論文のまとめを示す。本論文では、シミュレートッド・アニーリングに向けたログ・エンコーディングによる ReRAM CiM を提案した。提案の ReRAM CiM は、従来の ReRAM CiM に比べて、アレイ面積を 97.6%削減する。また、ReRAM の非線形なエラー特性を考慮し、“0” を信頼性の高い HRS に、“1” を信頼性の低い LRS に記憶することで、許容できる BER が 10 倍、許容できるビット精度が 5 ビットまで向上する。

**謝辞** この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものです。

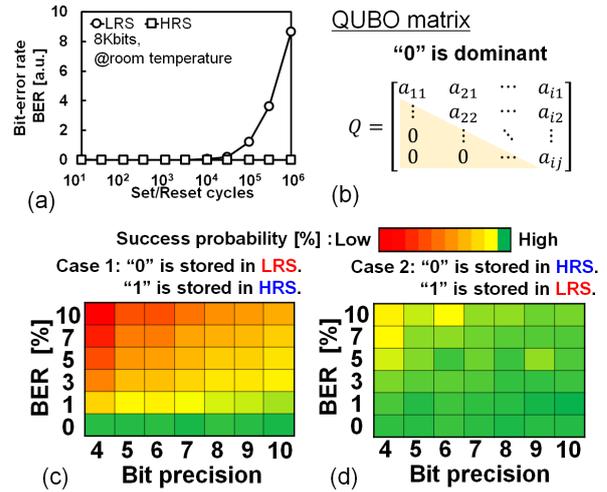


図 3 (a) ReRAM デバイスの測定により得られた Bit-error rate (BER). (b) ReRAM のコンダクタンスにマッピングされる QUBO 行列. (c) Case1 における SA の成功確率. (d) Case2 における SA の成功確率.

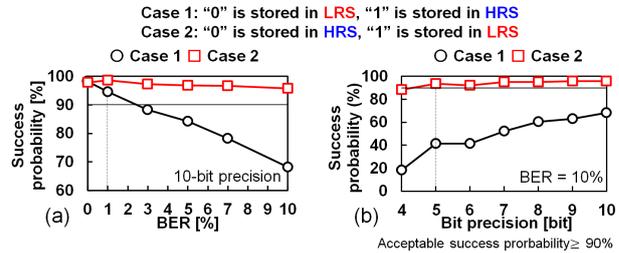


図 4 Case1 と Case2 の SA の成功確率. (a) BER による比較. (b) ビット精度による比較.

表 1 本論文のまとめ. (a) アレイ面積の比較 (図 2). (b) Case1 と Case2 の SA 結果の比較 (図 4).

	Conventional linear-encoding	Proposed log-encoding
(a)	ReRAM CiM 110 WLS × 110 BLs	17 WLS × 17 BLs
		-97.6%
(b)	Case 1	Case 2
Acceptable BER (10-bit precision)	1%	10%
Acceptable bit precision (BER = 10%)	—	≥5-bit

Case 1: “0” is stored in LRS, “1” is stored in HRS  
 Case 2: “0” is stored in HRS, “1” is stored in LRS

### 参考文献

- [1] K. Taoka et al., “Simulated Annealing Algorithm & ReRAM Device Co-optimization for Computation-in-Memory,” *IEEE IMW*, 2021, pp. 1-4.
- [2] N. Misawa et al., “Small Array Area, Memory Error Tolerant ReRAM Computation-in-Memory with Log-encoding Simulated Annealing for Combinational Optimization Problems” in *Ext. Abstr. SSDM*, 2021, pp. 119-120.
- [3] A. Lucas, “Ising formulations of many NP problems,” *Frontiers in Physics*, vol. 2, no.5, pp. 1-15, 2014.
- [4] R. Mochida et al., “A 4M Synapses integrated Analog ReRAM based 66.5 TOPS/W Neural-Network Processor with Cell Current Controlled Writing and Flexible Network Architecture”, in *Symp. VLSI Tech. Dig. Tech. Papers*, 2018, pp. 175-176.
- [5] R. Yasuhara et al., “Reliability Issues in Analog ReRAM Based Neural-Network Processor,” *IEEE IRPS*, 2019, pp. 1-5.