Qiankun WANG, Xingchen LI, Bingzhe WU, Ke YANG, Wei HU, Guangyu SUN, Yuchao YANG, 2023. COPPER: a combinatorial optimization problem solver with processing-in-memory architecture. *Frontiers of Information Technology & Electronic Engineering*, 24(5):731-741. https://doi.org/10.1631/FITEE.2200463

COPPER: a combinatorial optimization problem solver with processing-in-memory architecture

Key words: Combinatorial optimization; Chaotic simulated annealing; Processing-in-memory

Corresponding author: Guangyu SUN
E-mail: gsun@pku.edu.cn
ORCID: https://orcid.org/0000-0002-7315-6589

Motivation

1. Combinatorial optimization problems (COPs) aim to find the optimal solution under specific conditions in a discrete space. They are of great importance in various fields.
2. Many COPs are NP-complete, and as the problem size increases, the cost of solving them escalates rapidly.
3. Chaotic simulated annealing (CSA) is an effective method for solving COPs, but the traditional computational hardware incurs significant time and energy costs in executing the CSA algorithm.

Method

Based on the Hopfield neural network, CSA can solve COPs well.

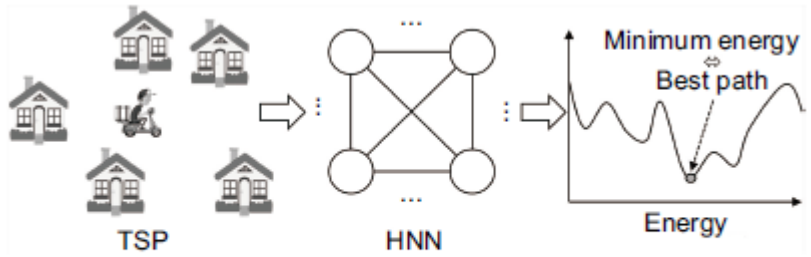


Fig. 1 Solving the traveling salesman problem (TSP) with a Hopfield neural network (HNN)

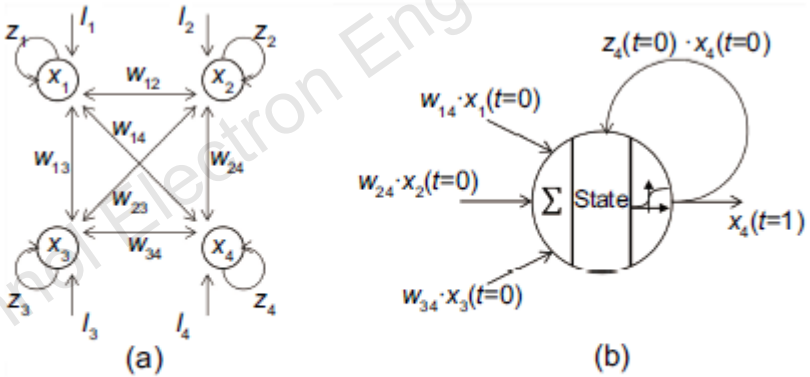


Fig. 2 Chaotic simulated annealing (CSA) (a) and its neuron (b)

Table 2 Performance of chaotic simulated annealing (CSA) and simulated annealing (SA)

City number	Real best	Iteration number	P_{feasible} (%)	Best		Average	
				CSA	SA	CSA	SA
10	1.768	906	96	**1.768**	**1.768**	1.776	**1.773**
20	4.172	1187	91	**4.172**	**4.172**	**4.316**	4.647
30	4.293	2962	96	**4.565**	4.575	**4.838**	4.980
40	5.451	4855	99	**5.502**	5.726	**5.705**	6.364
50	5.939	6903	99	**6.086**	6.444	**6.379**	7.180

The column "iteration number" is the average number of iterations needed by CSA to reach stability, and P_{feasible} means the percentage of obtaining a feasible solution using CSA. The bold values represent the better solutions from CSA and SA

Method

Quantize weights and outputs of CSA:
Modify CSA to make it hardware-friendly

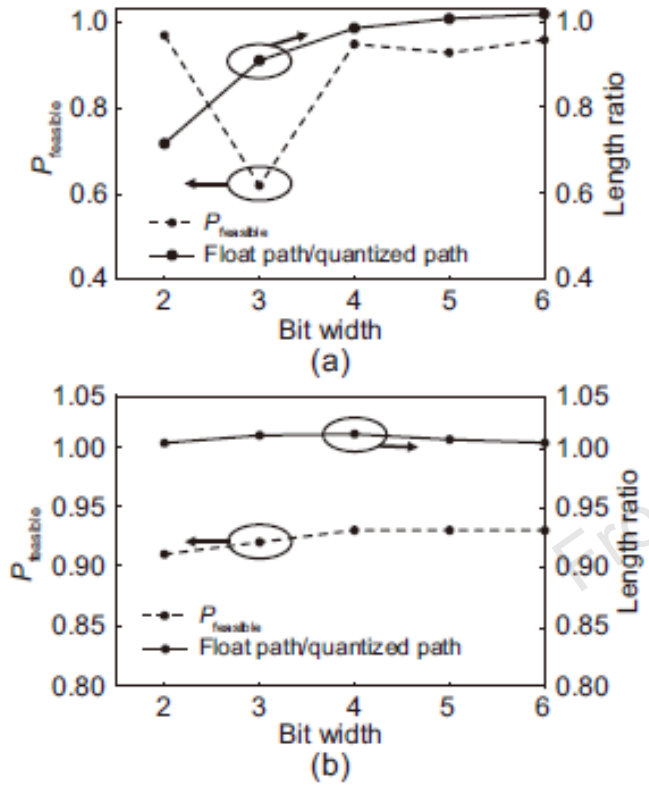


Fig. 4 Quantization influence: (a) weight quantization; (b) output quantization

$$x_i(t) = \frac{1}{1 + \mathrm{e}^{-y_i(t)/\epsilon}},$$

$$y_i(t+1) = ky_i(t) + \alpha \left(\sum_{j=1, j\neq i}^{n} w_{ij} x_j(t) + I_i \right) - z_i(t)(x_i(t) - I_0),$$

$$z_i(t+1) = (1-\beta) z_i(t),$$

$$a_i = \mathrm{Truncate}(y_i(t) << p),\ x_i(t) = \mathrm{LUT}(a_i).$$

$$z_i(t+1) = z_i(t) - \delta.$$

Table 3 Influence of the adaptation

Scheme	P_{feasible} (%)	Iteration number	Length
Original	96	2962	4.838
Linear decay	93	3081	4.875
Output stability	92	2093	4.806

P_{feasible}: the percentage of obtaining a feasible solution

Method

Processing-in-memory hardware architecture based on RRAM and COPPER

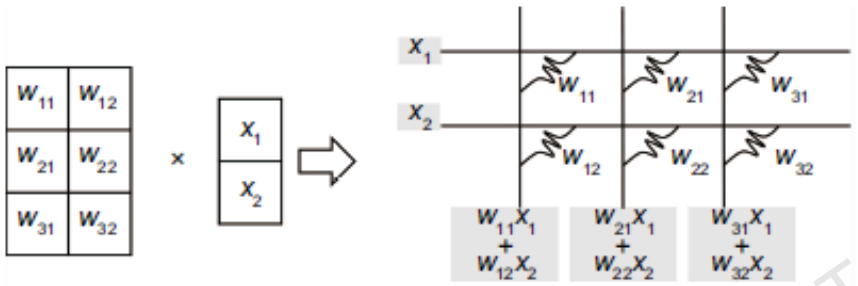


Fig. 3 Memristor crossbar for multiplication

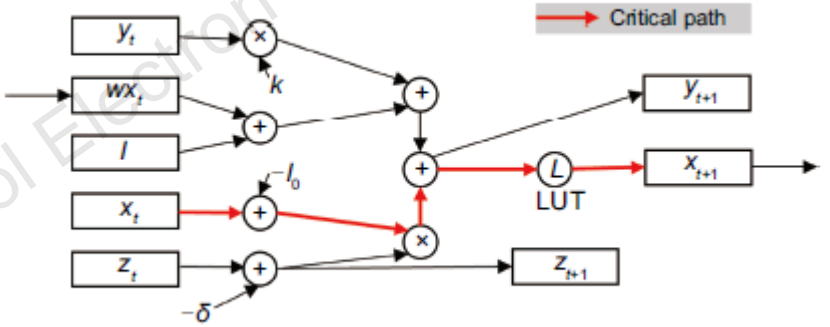


Fig. 7 Peripheral circuit data path

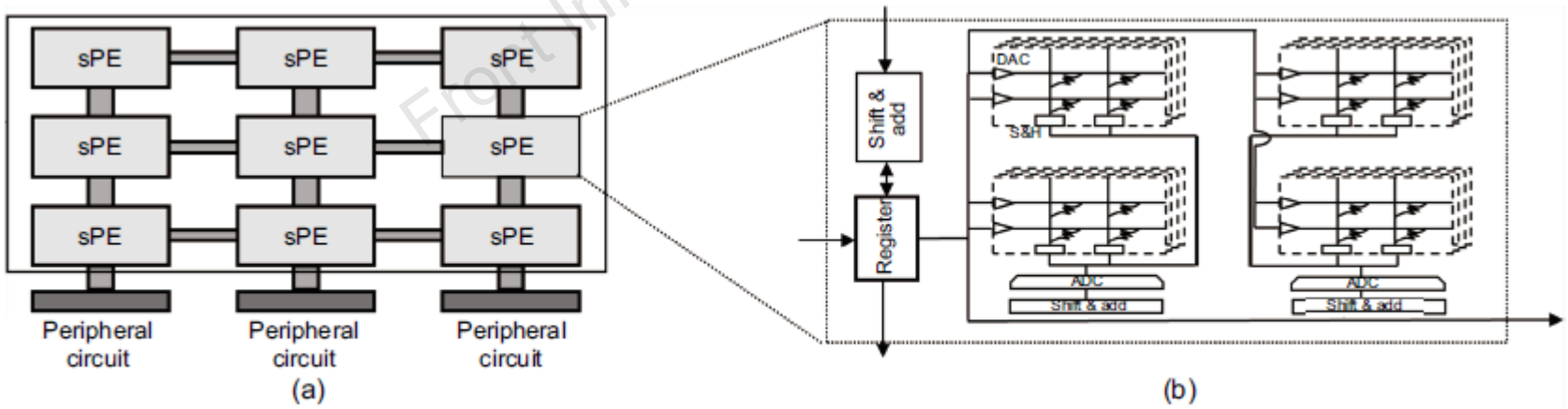
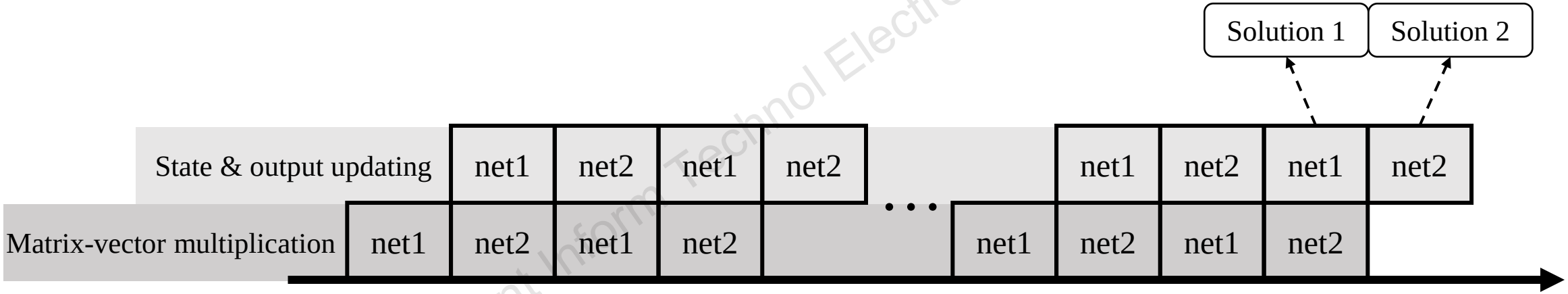


Fig. 6 COPPER overview (a) and subPE (b)

Method

A pipeline is introduced to increase the hardware utilization and save time.



Pipeline of COPPER

Conclusions

1. We explored the influence that quantization brings to CSA, and determined the appropriate bit width.
2. We adapted CSA to hardware without reducing its performance.
3. We designed an efficient COP solver with CSA and PIM, and proposed a pipeline method.

Qiankun WANG received his BS degree in computer science and technology from Peking University, Beijing, China, in 2020. He is currently pursuing his MS degree in software engineering at Peking University. His research interests focus on neuromorphic computing and model compression.

Guangyu SUN is currently an associate professor with the School of Integrated Circuits at Peking University. He received his BS and MS degrees from Tsinghua University, Beijing, in 2003 and 2006, respectively, and his PhD degree from the Pennsylvania State University in 2011. His PhD thesis, Exploring Memory Hierarchy Design with Emerging Memory Technologies, received the 2012 EDAA outstanding dissertation award. His research interests include design and automation for computer architecture, cross-layer co-optimization, emerging memory technologies, etc. He has published more than 150 journal and conference papers on DAC, ISCA, MICRO, HPCA, IEEE TCAD, etc. He won the DAC Under-40 Innovators Award, CCF-IEEE CS Young Computer Scientists Award, Microsoft Research Asia Collaborative Research Award, CCF-Intel Young Faculty Researcher Program Award, and the best paper awards three times. He was the general co-chair of NVMSA2021 and the TPC co-chair of NVMSA2020, RTCSA2019, APPT2017, and NAS2012. He has served as a program committee member and a track chair for over 20 conferences in these areas, including DAC, ICCAD, MICRO, HPCA, etc. He is an associate editor of ACM JETC.