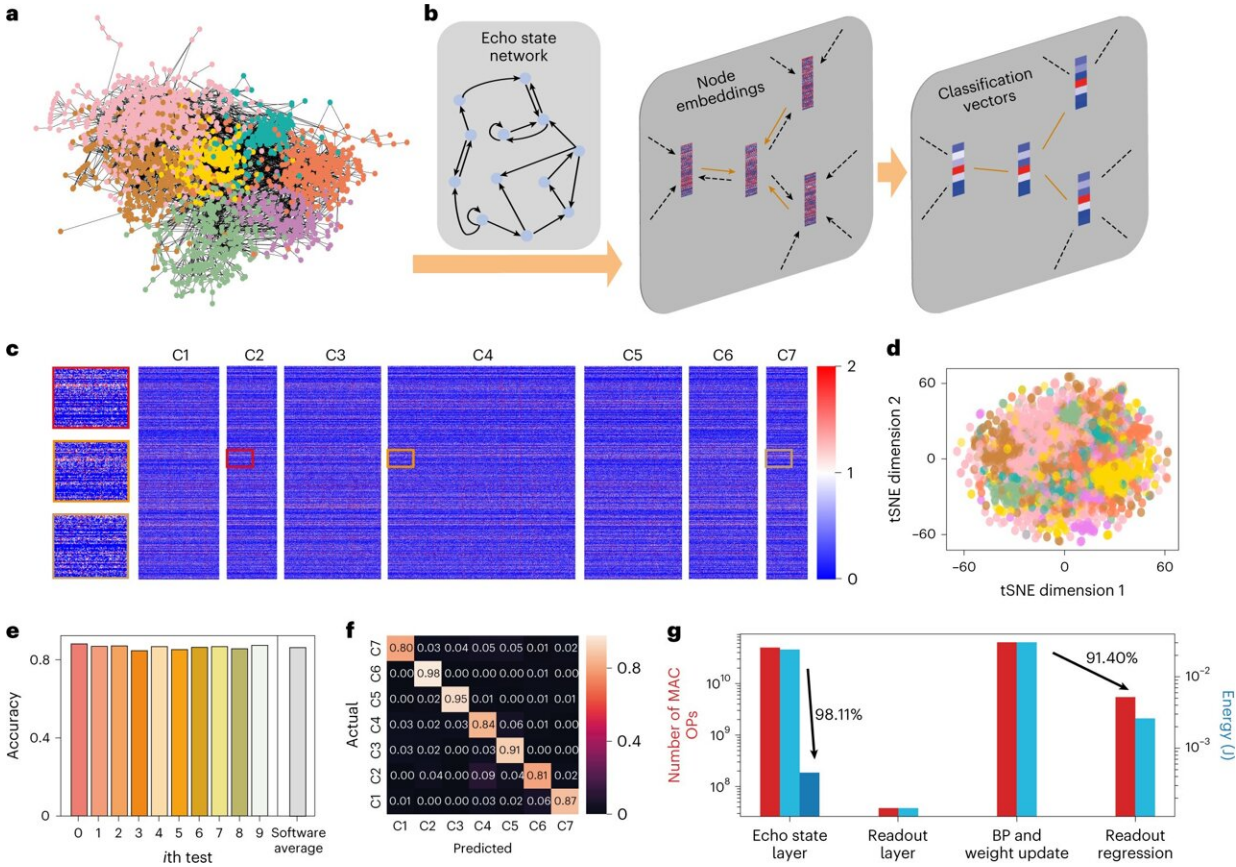


# A system integrating echo state graph neural networks and analogue random resistive memory arrays

March 15 2023, by Ingrid Fadelli



Node classification of a citation network. a, An illustration of the large-scale citation network CORA. Each node in the graph is a scholarly article, while an edge indicates a citation between two papers. There are a total of seven article categories, indicated by node colors, according to their discipline. b, The node classification scheme. The input graph is first embedded using the ESGNN according to the protocol shown in Fig. 1f and Methods, followed by a graph

convolution layer serving as the readout to produce a classification vector for each node. c, An illustration of simulated node embeddings. Colored boxes on the left are the zoom-in of node embedding details. d, A node embedding mapped to a 2D spacing using t-SNE, showing clear clustering of nodes of the same categories. e, The accuracy of ten random tests for node classification and the software baseline. The average accuracy is 87.12%, comparable to state-of-the-art algorithms. f, The normalized confusion matrices of the simulated classification results. g, A breakdown of the estimated OPs (red bars) and the associated energy consumption (light-blue bars for a state-of-the-art GPU; dark-blue bars for a projected random resistive memory-based hybrid analogue–digital system). In a forward (backward) pass, the fully trainable model on a state-of-the-art GPU and ESGNN on a projected random resistive memory-based hybrid analogue–digital system consume approximately 24.20 mJ (approximately 63.18 GOPs) and approximately 599.47  $\mu$ J (approximately 5.43 GOPs), respectively, revealing a >40.37 fold improvement in the inference energy efficiency (and an approximately 91.40% reduction of the backward pass complexity). Credit: *Nature Machine Intelligence* (2023). DOI: 10.1038/s42256-023-00609-5

Graph neural networks (GNNs) are promising machine learning architectures designed to analyze data that can be represented as graphs. These architectures achieved very promising results on a variety of real-world applications, including drug discovery, social network design, and recommender systems.

As graph-structured data can be highly complex, graph-based machine learning architectures should be designed carefully and effectively. In addition, these architectures should ideally be run on efficient hardware that support their computational demands without consuming too much power.

Researchers at University of Hong Kong, the Chinese Academy of Sciences, InnoHK Centers and other institutes worldwide recently

developed a software-hardware system that combines a GNN architecture with a resistive memory, a [memory solution](#) that stores data in the form of a resistive state. Their paper, published in *Nature Machine Intelligence*, demonstrates the potential of new hardware solutions based on resistive memories for efficiently running graph machine learning techniques.

"The efficiency of digital computers is limited by the von-Neumann bottleneck and slowdown of Moore's law," Shaocong Wang, one of the researchers who carried out the study, told Tech Xplore. "The former is a result of the physically separated memory and processing units that incurs large energy and time overheads due to frequent and massive data shuttling between these units when running graph learning. The latter is because transistor scaling is approaching its physical limit in the era of 3nm technology node."

Resistive memories are essentially tunable resistors, which are devices that resist the passage of electrical current. These resistor-based memory solutions have proved to be very promising for running artificial [neural networks](#) (ANNs). This is because individual resistive memory cells can both store data and perform computations, addressing the limitations of the so-called Naumann bottleneck.

"Resistive memories are also highly scalable, retaining Moore's law," Wang said. "But ordinary resistive memories are still not good enough for graph learning, because graph learning frequently changes the resistance of resistive memory, which leads to a large amount of energy consumption compared to the conventional digital computer using SRAM and DRAM. What's more, the resistance change is inaccurate, which hinders precise gradient updating and weight writing. These shortcomings may defeat the advantages of resistive memory for efficient graph learning."

The key objective of the recent work by Wang and his colleagues was to overcome the limitations of conventional resistive memory solutions. To do this, they designed a resistive memory-based graph learning accelerator that eliminates the need for resistive memory programming, while retaining a high efficiency.

They specifically used echo state networks, a reservoir computing architecture based on recurrent neural network with a sparsely connected hidden layer. Most of these networks' parameters (i.e., weights) can be fixed random values. This means that they can allow resistive memory to be immediately applicable, without the need for programming.

"In our study, we experimentally verified this concept for graph learning, which is very important, and in fact, quite general," Wang said.

"Actually, images and sequential data, such as audios and texts, can also be represented as graphs. Even transformers, the most state-of-the-art and dominant deep learning models, can be represented as graph neural networks."

The echo state graph neural networks developed by Wang and his colleagues are comprised of two distinct components, known as the echo state and readout layer. The weights of the echo state layer are fixed and random, thus they do need to be repeatedly trained or updated over time.

"The echo state layer functions as a graph convolutional layer that updates the hidden state of all nodes in the graph recursively," Wang said. "Each node's hidden state is updated based on its own feature and the hidden states of its neighboring nodes in the previous time step, both extracted with the echo state weights. This process is repeated four times, and the hidden states of all nodes are then summed into a vector to represent the entire graph, which is classified using the readout layer. This process is repeated for four times, and then the hidden states of all nodes are summed together into a vector, as the representation of the

entire graph, which is then classified by the readout layer."

The software-hardware design proposed by Wang and his colleagues has two notable advantages. Firstly, the echo state neural network it is based on requires significantly less training. Secondly, this neural network is efficiently implemented on a random and fixed resistive memory that does not need to be programmed.

"Our study's most notable achievement is the integration of random resistive memory and echo state graph neural networks (ESGNN), which retain the energy-area efficiency boost of in-memory computing while also utilizing the intrinsic stochasticity of dielectric breakdown to provide low-cost and nanoscale hardware randomization of ESGNN," Wang said. "Specifically, we propose a hardware-software co-optimization scheme for graph learning. Such a codesign may inspire other downstream computing applications of resistive memory."

In terms of software, Wang and his colleagues introduced a ESGNN comprised of a large number of neurons with random and recurrent interconnections. This neural network employs iterative random projections to embed nodes and graph-based data. These projections generate trajectories at the edge of chaos, enabling efficient feature extraction while eliminating the arduous training associated with the development of conventional graph neural networks.

"On the hardware side, we leverage the intrinsic stochasticity of dielectric breakdown in resistive switching to physically implement the random projections in ESGNN," Wang said. "By biasing all the resistive cells to the median of their breakdown voltages, some cells will experience dielectric breakdown if their breakdown voltages are lower than the applied voltage, forming random resistor arrays to represent the input and recursive matrix of the ESGNN. Compared with pseudo-random number generation using [digital systems](#), the source of

randomness here is the stochastic redox reactions and ion migrations that arise from the compositional inhomogeneity of resistive memory cells, offering low-cost and highly scalable random resistor arrays for in-memory computing."

In initial evaluations, the system created by Wang and his colleagues achieved promising results, running ESGNNs more efficiently than both digital and conventional resistive memory solutions. In the future, it could be implemented to various real-world problems that require the analysis of data that can be represented as graphs.

Wang and his colleagues think that their software-hardware system could be applied to a wide range of machine learning problems, thus they now plan to continue exploring its potential. For instance, they wish to assess its performance in sequence analysis tasks, where their echo state network implemented on memristive arrays could remove the need for programming, while ensuring low power consumption and high accuracy.

"The prototype demonstrated in this work was tested on relatively small datasets, and we aim to push its limits with more complex tasks," Wang added. "For instance, the ESN can serve as a universal graph encoder for feature extraction, augmented with memory to perform few-shot learning, making it useful for edge applications. We look forward to exploring these possibilities and expanding the capabilities of the ESN and memristive arrays in the future."

**More information:** Shaocong Wang et al, Echo state graph neural networks with analogue random resistive memory arrays, *Nature Machine Intelligence* (2023). [DOI: 10.1038/s42256-023-00609-5](https://doi.org/10.1038/s42256-023-00609-5)

Citation: A system integrating echo state graph neural networks and analogue random resistive memory arrays (2023, March 15) retrieved 26 April 2024 from <https://techxplore.com/news/2023-03-echo-state-graph-neural-networks-1.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.