## UC SANTA BARBARA



May 14, 2025 <u>Sonia Fernandez</u>

## Energy and memory: A new neural network paradigm

Listen to the first notes of an old, beloved song. Can you name that tune? If you can, congratulations — it's a triumph of your associative memory, in which one piece of information (the first few notes) triggers the memory of the entire pattern (the song), without you actually having to hear the rest of the song again. We use this handy neural mechanism to learn, remember, solve problems and generally navigate our reality.

"It's a network effect," said UC Santa Barbara mechanical engineering professor <u>Francesco Bullo</u>, explaining that associative memories aren't stored in single brain cells. "Memory storage and memory retrieval are dynamic processes that occur over entire networks of neurons."

In 1982 physicist John Hopfield translated this theoretical neuroscience concept into the artificial intelligence realm, with the formulation of the Hopfield network. In doing so, not only did he provide a mathematical framework for understanding memory storage and retrieval in the human brain, he also developed one of the first recurrent artificial neural networks, known for its ability to retrieve complete patterns from noisy or incomplete inputs. Hopfield won the Nobel Prize for his work in 2024.

However, according to Bullo and collaborators Simone Betteti, Giacomo Baggio and Sandro Zampieri at the University of Padua in Italy, the traditional Hopfield network model is powerful, but it doesn't tell the full story of how new information guides memory retrieval. "Notably," they say in a <u>paper published in the journal Science</u> <u>Advances</u>, "the role of external inputs has largely been unexplored, from their effects on neural dynamics to how they facilitate effective memory retrieval." The researchers suggest a model of memory retrieval they say is more descriptive of how we experience memory.

"The modern version of machine learning systems, these large language models they don't really model memories," Bullo explained. "You put in a prompt and you get an output. But it's not the same way in which we understand and handle memories in the animal world." While LLMs can return responses that can sound convincingly intelligent, drawing upon the patterns of the language they are fed, they still lack the underlying reasoning and experience of the physical real world that animals have.

"The way in which we experience the world is something that is more continuous and less start-and-reset," said Betteti, lead author of the paper. Most of the treatments on the Hopfield model tended to treat the brain as if it was a computer, he added, with a very mechanistic perspective. "Instead, since we are working on a memory model, we want to start with a human perspective."

The main question inspiring the theorists was: As we experience the world that surrounds us, how do the signals we receive enable us to retrieve memories?

As Hopfield envisioned, it helps to conceptualize memory retrieval in terms of an energy landscape, in which the valleys are energy minima that represent memories. Memory retrieval is like exploring this landscape; recognition is when you fall into one of the valleys. Your starting position in the landscape is your initial condition.

"Imagine you see a cat's tail," Bullo said. "Not the entire cat, but just the tail. An associative memory system should be able to recover the memory of the entire cat." According to the traditional Hopfield model, the cat's tail (stimulus) is enough to put you closest to the valley labeled "cat," he explained, treating the stimulus as an initial condition. But how did you get to that spot in the first place?

"The classic Hopfield model does not carefully explain how seeing the tail of the cat puts you in the right place to fall down the hill and reach the energy minimum," Bullo said. "How do you move around in the space of neural activity where you are storing these memories? It's a little bit unclear."

The researchers' Input-Driven Plasticity (IDP) model aims to address this lack of clarity with a mechanism that gradually integrates past and new information, guiding the memory retrieval process to the correct memory. Instead of applying the two-step algorithmic memory retrieval on the rather static energy landscape of the original Hopfield network model, the researchers describe a dynamic, input-driven mechanism.

"We advocate for the idea that as the stimulus from the external world is received (e.g., the image of the cat tail), it changes the energy landscape at the same time," Bullo said. "The stimulus simplifies the energy landscape so that no matter what your initial position, you will roll down to the correct memory of the cat." Additionally, the researchers say, the IDP model is robust to noise — situations where the input is vague, ambiguous, or partially obscured — and in fact uses the noise as a means to filter out less stable memories (the shallower valleys of this energy landscape) in favor of the more stable ones.

"We start with the fact that when you're gazing at a scene your gaze shifts in between the different components of the scene," Betteti said. "So at every instant in time you choose what you want to focus on but you have a lot of noise around." Once you lock into the input to focus on, the network adjusts itself to prioritize it, he explained.

Choosing what stimulus to focus on, a.k.a. attention, is also the main mechanism behind another neural network architecture, the transformer, which has become the heart of large language models like ChatGPT. While the IDP model the researchers propose "starts from a very different initial point with a different aim," Bullo said, there's a lot of potential for the model to be helpful in designing future machine learning systems.

"We see a connection between the two, and the paper describes it," Bullo said. "It is not the main focus of the paper, but there is this wonderful hope that these associative memory systems and large language models may be reconciled."

Media Contact **Sonia Fernandez** Senior Science Writer

## About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.