ADAPTIVE ASSOCIATIVE MEMORY WITH DIFFERENTIABLE CONTENT-ADDRESSABLE MEMORIES FOR ONLINE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Associative memory is a unifying abstraction underlying attention mechanisms, energy-based models, and adaptive inference systems. Analog content-addressable memories (CAMs) enable graded similarity search directly in hardware, supporting soft retrieval and online adaptation within a single substrate. We study differentiable CAM (diff-CAM) as a general associative memory primitive. Through controlled simulations, we characterize its retrieval robustness, learning dynamics, and stability–plasticity tradeoffs under distribution drift and representation shift. Compared to static CAMs, adaptive diff-CAM exhibits rapid rebinding and improved associative inference under non-stationary inputs. These results position diff-CAM as a hardware-aligned substrate for efficient and adaptive memory-augmented systems.

1 INTRODUCTION

Associative memory systems retrieve stored patterns based on partial or noisy cues and form a foundational component of biological and artificial intelligence. Classical associative memories, such as Hopfield networks and content-addressable memories (CAMs), implement retrieval through similarity-based access to stored representations. Recent advances in analog and differentiable CAM hardware enable continuous-valued storage and retrieval with low latency and energy consumption Li et al. (2020); Mao et al. (2022); Pedretti et al. (2022). However, most prior algorithmic evaluations of CAMs focus on narrow tasks such as nearest-neighbor retrieval and clustering Ameli et al. (2025); Saha et al. (2023), leaving open whether CAMs can function as general-purpose associative memory substrates supporting soft inference and continual adaptation.

Approaches such as Elastic Weight Consolidation Kirkpatrick et al. (2017) and Test-Time Training Sun et al. (2020) address distribution shift via parameter regularization or self-supervision but lack content-addressable memory and soft associative inference. Modern Hopfield networks Krotov & Hopfield (2016) implement continuous attractor dynamics but lack local online plasticity. In contrast, adaptive diff-CAM integrates soft retrieval with similarity-gated local updates, enabling unsupervised online adaptation and exposing a stability–plasticity tradeoff Grossberg (1987); Ramsauer et al. (2020); Zenke et al. (2017).

We study differentiable CAM (diff-CAM) as a trainable associative memory primitive. Our focus is on two core capabilities: (i) *soft associative inference*, in which a query retrieves a graded mixture of stored memories rather than a single nearest neighbor, and (ii) *continual adaptation*, in which memory contents are updated online under non-stationary input distributions.

We characterize the retrieval robustness, learning dynamics, and stability–plasticity tradeoffs of diff-CAM in controlled settings, including abrupt distribution shifts and representation drift. Compared to static CAMs, adaptive diff-CAM exhibits rapid reassociation and improved associative inference under non-stationary inputs, positioning diff-CAM as a hardware-aligned substrate for adaptive memory-augmented systems.

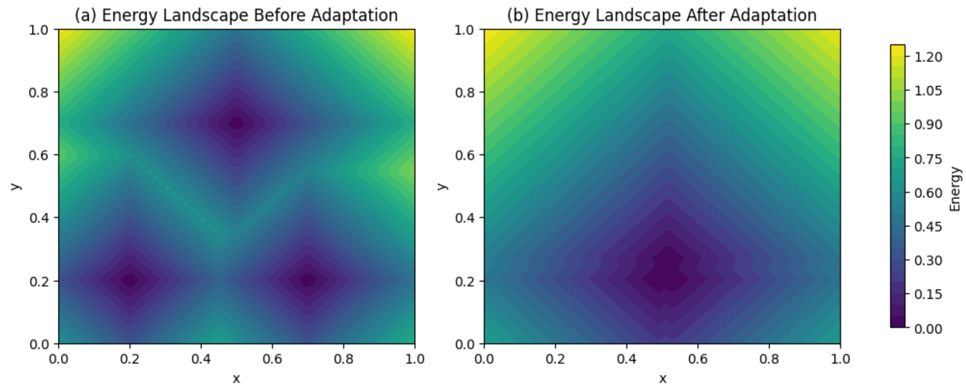

Figure 1: Reshaping of the associative energy landscape through plasticity. Energy landscapes computed from diff-CAM memories before (left) and after (right) continual adaptation. Online memory updates shift low-energy regions toward frequently observed inputs, illustrating dynamic reconfiguration of associative structure.

2 diff-CAM as an Associative Memory

We model a diff-CAM memory as a set of stored analog keys, where similarity between an input query and each memory entry is computed via an analog match-line current. This current defines a differentiable similarity score. Memory adaptation is implemented through updates to the stored parameters, corresponding to changes in the conductance values of the underlying devices.

2.1 Continuous learning in diff-CAMs

For retrieval, we adopt a softmax weighting over distance, consistent with modern Hopfield networks and attention-based associative memory formulations Ramsauer et al. (2020); Krotov & Hopfield (2016). Given query q, a single diff-CAM cell provides a distance, with the prepheral circuit, the association weights can be computed as:

$$w_i(t) = \frac{\exp(-\alpha \parallel q_t - c_i(t) \parallel_1)}{\sum_j \exp(-\alpha \parallel q_t - c_j(t) \parallel_1)}. \tag{1}$$

Here α is an inverse temperature controlling the sharpness of associative retrieval. Retrieval is implemented via softmax-weighted similarity over memory entries. An adaptive CAM augments retrieval with online memory updates:

$$c_i(t+1) = c_i(t) + \eta \Delta_i(q_t, c_i(t)), \tag{2}$$

trading stability for responsiveness under non-stationary inputs. Static CAM corresponding to the special case $c_i(t+1) = c_i(t)$. Here, $\eta > 0$ is a learning rate, and Δ_i is similarity dependent update rule. Note that this update rule is fully differentiable. Here the updates are local, depending only on q_t and c_i. Learning is softly gated by retrieval strength, and no explicit supervision is required. This is analogous to Hebbian-style associative learning under soft assignment. This transforms CAM from a static lookup structure into a plastic associative memory, trading stability for responsiveness under non-stationary inputs. The L1 distance reflects the Manhattan metric implemented by diff-CAM match-line currents, aligning the algorithmic model with the underlying hardware substrate. Similar lifelong adaptation of external memories has been demonstrated in memristive few-shot learning via online prototype augmentationZhou et al. (2023).

2.2 Adaptive associative Memory under Distribution Shift

We consider an online associative inference problem under distribution shift, where a stream of queries evolve over time and the underlying input distribution is non-stationary. The objective is

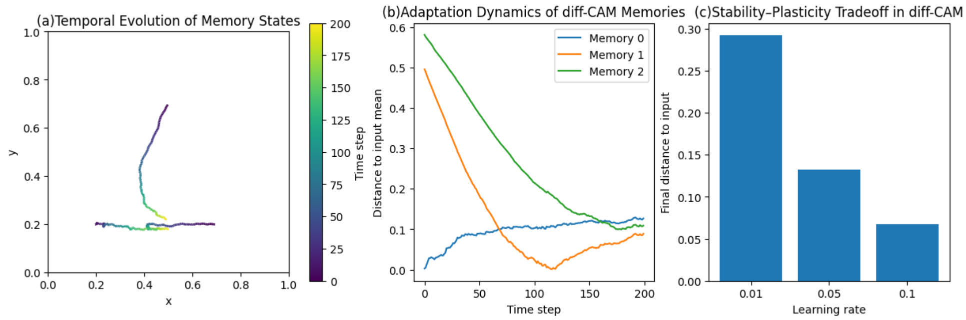

Figure 2: (a) Continual adaptation of stored memories. Trajectories of diff-CAM memory patterns under an online rewrite rule driven by a drifting input distribution. Memory states adapt smoothly over time, demonstrating plasticity without retraining or backpropagation. (b) Adaptation dynamics of memory entries, measured as distance to the current input mean over time. Adaptive diff-CAM tracks the drifting distribution, while frozen memory (static CAM) would remain misaligned. (c) Stability–plasticity tradeoff as a function of learning rate. Larger learning rates enable faster adaptation to distribution shift but lead to less stable memory states, revealing a continuous tradeoff between responsiveness and associative sharpness.

to evaluate how well an associative memory maintains meaningful retrieval as the data distribution drifts. Specifically, we model queries as Let queries arrive sequentially as

$$q_t \sim \mathcal{N}(\mu_t, \sigma^2 I), \qquad \mu_t = \mu_0 + vt \tag{3}$$

where the drift velocity v controls the rate of distribution shift. This captures gradual structured changes commonly observed in continual learning and streaming settings. Static attention corresponds to a frozen associative memory that cannot adapt to this drift, while adaptive diff-CAM updates its memory contents online to track the evolving input distribution.

2.3 Few-shot Rebinding under Representation Shift

Beyond gradual distribution drift, intelligent memory systems must rapidly rebind stored content when representations change, for example due to sensor recalibration, feature drift, or embedding updates in a perception pipeline. In such cases, retrieval fails not because memory is noisy, but because the coordinate system of stored patterns becomes misaligned with incoming queries. We formalize this as a representation shift problem. Let $\phi_0(x)$ denote the original embedding of inputs and $\phi_1(x)$ a shifted embedding obtained after a system update (e.g., a rotated or perturbed feature space). Memory is initialized with prototypes $\{c_i\}$ stored in the original representation space. At time t_0, the representation changes abruptly and queries are drawn from $\phi_1(x)$.

$$\phi_1(x) = T\big(\phi_0(x)\big), \qquad T \in \mathcal{F}, \tag{4}$$

where $\phi : \mathcal{X} \to \mathbb{R}^d$ denotes the embedding function mapping inputs to a feature space, and T models a representation shift such as a rotation, affine transform, or feature reparameterization induced by a system update. Memory is initialized as $c_i(t_0^-) = \mathbb{E}_{x\sim\mathcal{D}_i}[\phi_0(x)]$. After the shift, queries are drawn from $\phi_1(x)$ and memory is updated online:

$$c_i(t+1) = c_i(t) + \eta w_i(t)\big(\phi_1(x_t) - c_i(t)\big). \tag{5}$$

Retrieval accuracy is evaluated on held-out post-shift queries as a function of the number of support samples per class.

We compare frozen and adaptive CAM, evaluating post-shift retrieval accuracy as a function of support samples per class.

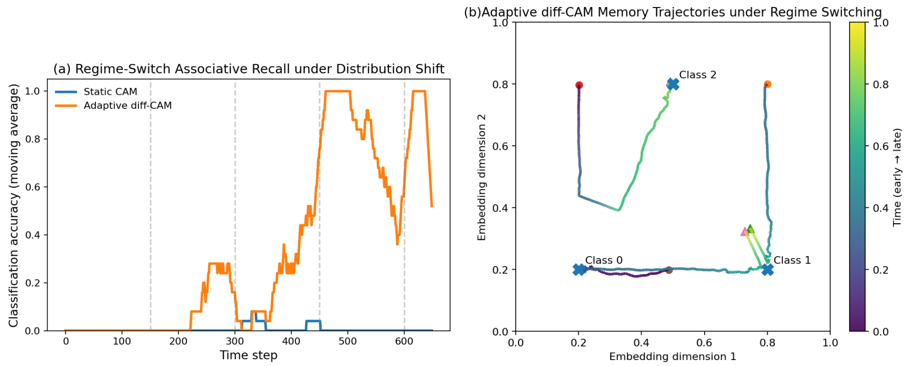

Figure 3: Few-shot rebinding under regime-switching inputs. (a) Retrieval accuracy over time as the active class switches. Adaptive diff-CAM rapidly recovers after each switch, while static CAM fails. (b) Memory trajectories in embedding space, showing online rebinding toward the active class.

3 Results

We evaluate diff-CAM in a controlled associative memory setting. Queries are drawn from a drifting input distribution , simulating non-stationary environments common in continual learning and agentic systems. We compare two configurations: (I) Static memory: retrieval only, no updates. (II) Adaptive memory: retrieval with online memory updates.

Figure 1 provides a physical energy interpretation of diff-CAM retrieval: plastic updates reshape the associative energy landscape, shifting basins of attraction toward frequently observed inputs. Regions of low energy correspond to memory states that are preferentially retrieved by nearby queries. As the input distribution drifts, online updates continuously deform this landscape, deepening basins around recently observed patterns while flattening regions associated with rarely used memories. diff-CAM tracks drifting inputs via plasticity (Fig. 2(a)), and reveals a stability–plasticity tradeoff controlled by the learning rate (Fig. 2(c)).

As shown in Fig. 3, adaptive diff-CAM rapidly rebinds memory entries after regime switches, enabling few-shot recovery, while static CAM fails under abrupt distribution shifts.

4 Conclusion

We presented a unified view of differentiable analog content-addressable memory as an associative memory primitive. Through systematic experiments, we showed that diff-CAM supports attention-like retrieval, continual adaptation, and controllable stability–plasticity tradeoffs. By framing diff-CAM beyond specific applications, our work highlights its potential role in future memory-augmented and agentic AI systems, particularly where energy efficiency and online adaptation are critical.

The soft associative retrieval implemented by diff-CAM is closely related to modern Hopfield networks, which perform retrieval via softmax-weighted similarity over stored patterns Ramsauer et al. (2020). In contrast to digital implementations, diff-CAM computes similarity through analog energy evaluation, with match-line currents corresponding to L1 distances between queries and stored keys. This yields an attention-like weighting without explicit dot products or digital normalization, effectively implementing a physical attention kernel. Memory adaptation in diff-CAM further enables test-time reassociation, positioning diff-CAM as a hardware instantiation of continuous associative memory with plasticity. Unlike digital Hopfield networks, diff-CAM realizes these dynamics through physical energy evaluation in memory, rather than explicit optimization.

These results position diff-CAM as more than a task-specific accelerator: its analog similarity implements soft associative inference, while local plasticity enables continual test-time adaptation without backpropagation. Together, these properties suggest diff-CAM as a hardware substrate in which storage, retrieval, and learning are tightly integrated.

References

Sara Ameli, Jiaao Yu, and John Paul Strachan. Associative clustering with analog content-addressable memory. In *Proceedings of the International Symposium on Memory Systems*, pp. 84–93, 2025.

Stephen Grossberg. Competitive learning: From interactive activation to adaptive resonance. *Cognitive science*, 11(1):23–63, 1987.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

Dmitry Krotov and John J Hopfield. Dense associative memory for pattern recognition. *Advances in neural information processing systems*, 29, 2016.

Can Li, Catherine E Graves, Xia Sheng, Darrin Miller, Martin Foltin, Giacomo Pedretti, and John Paul Strachan. Analog content-addressable memories with memristors. *Nature communications*, 11(1):1638, 2020.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.

Ruibin Mao, Bo Wen, Arman Kazemi, Yahui Zhao, Ann Franchesca Laguna, Rui Lin, Ngai Wong, Michael Niemier, X Sharon Hu, Xia Sheng, et al. Experimentally validated memristive memory augmented neural network with efficient hashing and similarity search. *Nature communications*, 13(1):6284, 2022.

Giacomo Pedretti, Catherine E Graves, Thomas Van Vaerenbergh, Sergey Serebryakov, Martin Foltin, Xia Sheng, Ruibin Mao, Can Li, and John Paul Strachan. Differentiable content addressable memory with memristors. *Advanced electronic materials*, 8(8):2101198, 2022.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Thomas Adler, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, et al. Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*, 2020.

Bishwajit Saha, Dmitry Krotov, Mohammed J Zaki, and Parikshit Ram. End-to-end differentiable clustering with associative memories. In *International Conference on Machine Learning*, pp. 29649–29670. PMLR, 2023.

Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pp. 9229–9248. PMLR, 2020.

Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *International conference on machine learning*, pp. 3987–3995. PMLR, 2017.

Houji Zhou, Ling Yang, Han Bao, Jiancong Li, Yi Li, and Xiangshui Miao. Memristive cosine-similarity-based few-shot learning with lifelong memory adaptation. *Advanced Intelligent Systems*, 5(2):2200173, 2023.

Table 1: Comparison of online adaptation and memory mechanisms.

Method	Online adaptation	Label-free update	Memory structure	Note
EWC Kirkpatrick et al. (2017)	×	×	Parameter regularization	Protects weights, but no content-addressable structure
GEM Lopez-Paz & Ranzato (2017)	✓	×	Episodic Memory	Replay with constraints to avoid forgetting
Test-Time Training Sun et al. (2020)	✓	✓	Feature adaptation	Self-supervised updates at test time
Hopfield Dense Memory Krotov & Hopfield (2016)	✓	×	Attractor dynamics	Static patterns, not plastic
Adaptive diff-CAM (ours)	✓	✓	Plastic CAM entries	Local associative update, no labels

Table 2: Comparison of online adaptation and memory mechanisms.