NEURAL INFORMATION PROCESSING SYSTEMS

Memory-Integrated Reconfigurable Adapters: A Unified Framework for Settings with Multiple Tasks

Susmit Agrawal* (U. Tübingen), Krishn Vishwas Kher* (IIT-H), Saksham Mittal (IIT-H), Swarnim Maheshwari (IIT-H), Vineeth N. Balasubramanian (IIT-H, MSR India)

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

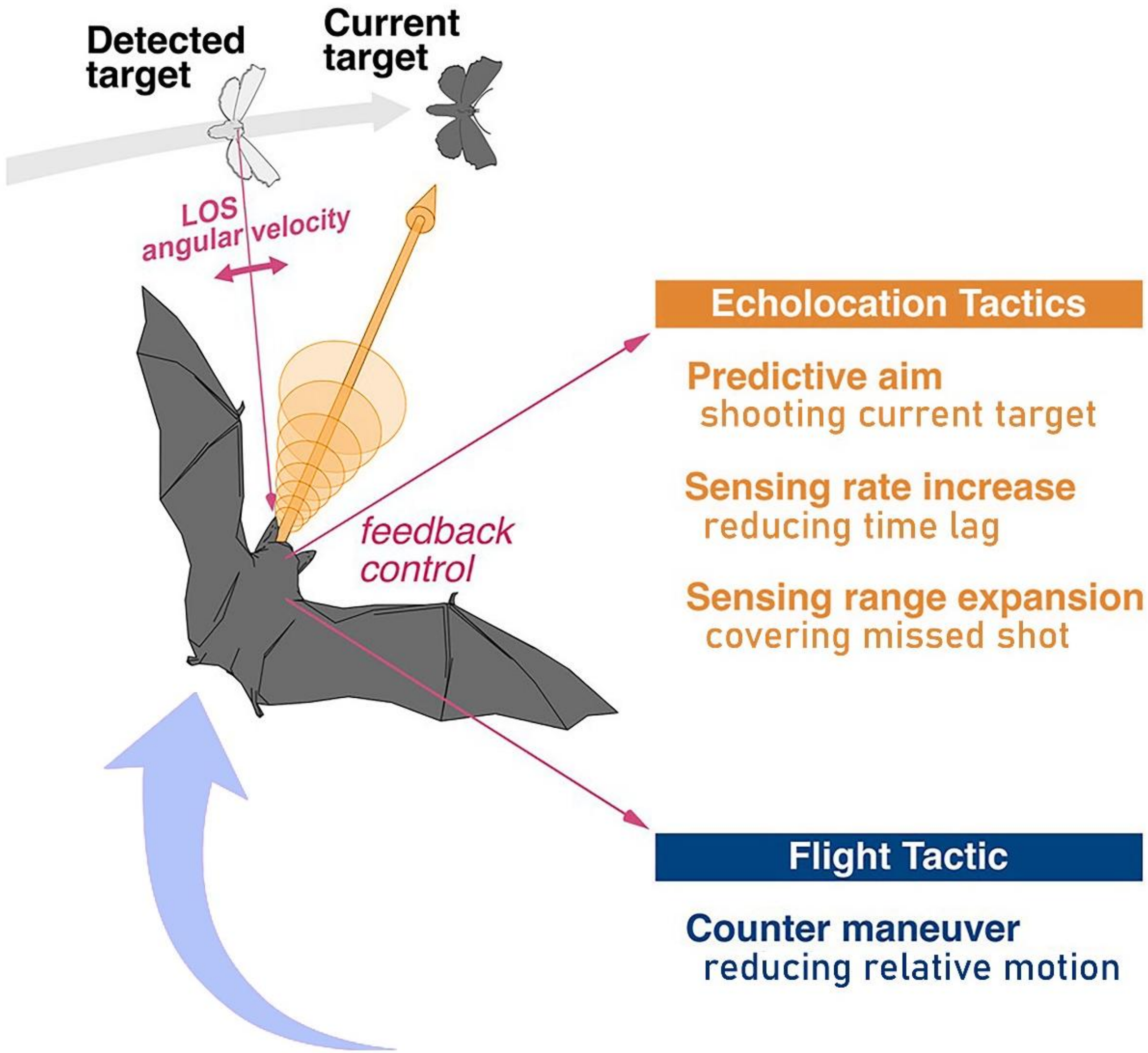
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

imprs-is

Microsoft Research

Task Switching in the Natural World

- Organisms often constantly switch across disparate tasks within fractions of seconds.
- This behaviour is not confined to fixed environments.
- They can often infer/learn a new task in a new environment on-the-fly.
- Certain neuroscience studies attribute this phenomenon to the brain's ability to repurpose the same circuitry for varied tasks without dismantling the core wiring.



AI Can't Task-Switch (As Well)

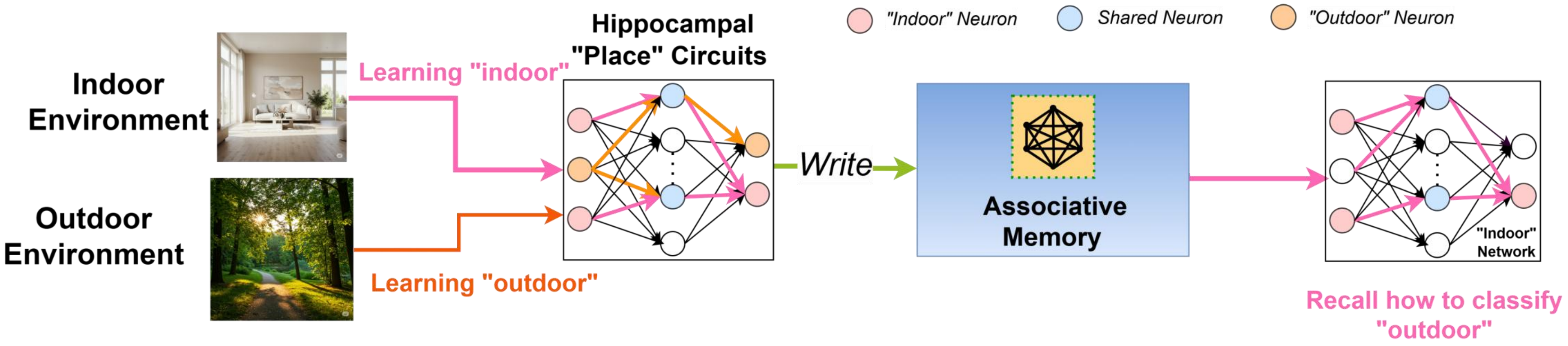
- AI has looked at this idea for long, but approaches to tackle this have often remained siloed.
- Domain Generalization mimics changes in environments.
- Incremental Learning emulates adapting to new tasks.
- Lack of an overarching framework that tackles both problems simultaneously.

Data Availability	**DG**	**DIL**	**CIL**
Tasks arrive sequentially?	✗	✓	✓
Same label sets across tasks?	✓	✓	✗
Task identifier available at inference?	✗	✗	✗
Test distribution seen during training?	✗	✓	✓

What’s Missing?

Deep Learning has predominantly overlooked explicit memory-based mechanisms that biology suggests are fundamental to rapid and efficient adaptation. Integrating biologically plausible associative memory models into deep learning frameworks should allow for rapid task switching.

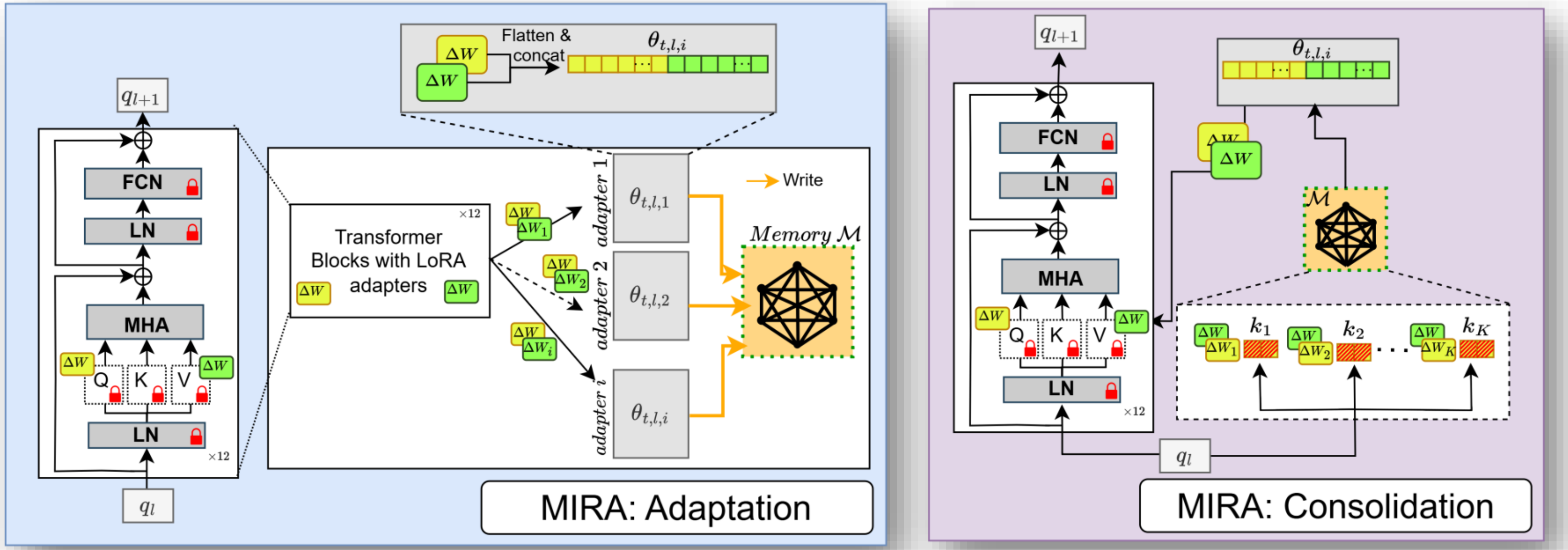
Hypothesis Realization

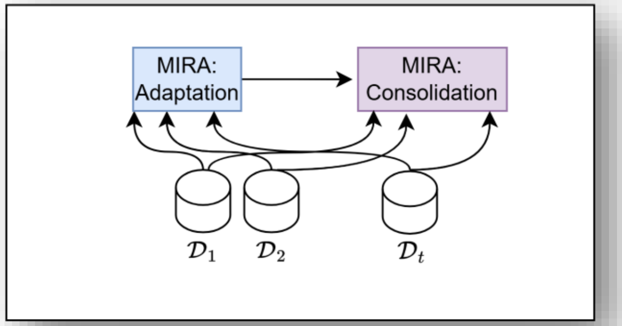


Associative Memory (AM) Primer

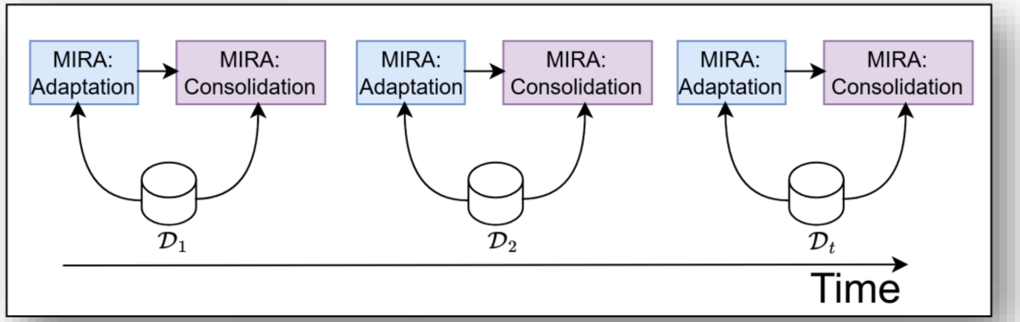
- Biologically plausible content-addressable memory system.
- Stores patterns, retrieves via associations based on query vector and stored patterns.
- Typically consists of three components:
  - **A memory vault**: a substrate on which it is instantiated.
  - **An energy function**: designed so that patterns lie very close to local minima of the function.
  - **Set of optimization programs**: for each operation such as *read*, *write*, *forget*, *update*, etc.
- Examples: Universal Hopfield Networks, Predictive Coding Networks, etc.

Architecture Blueprint





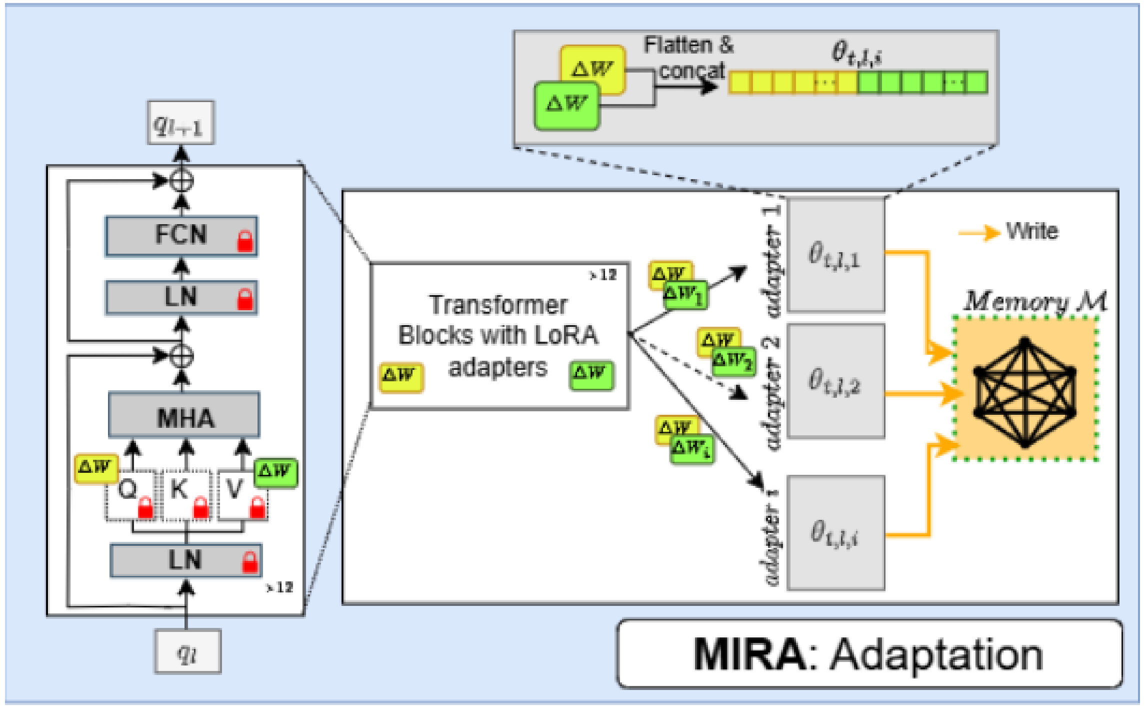
Domain Generalization



Domain/Class Incremental Learning

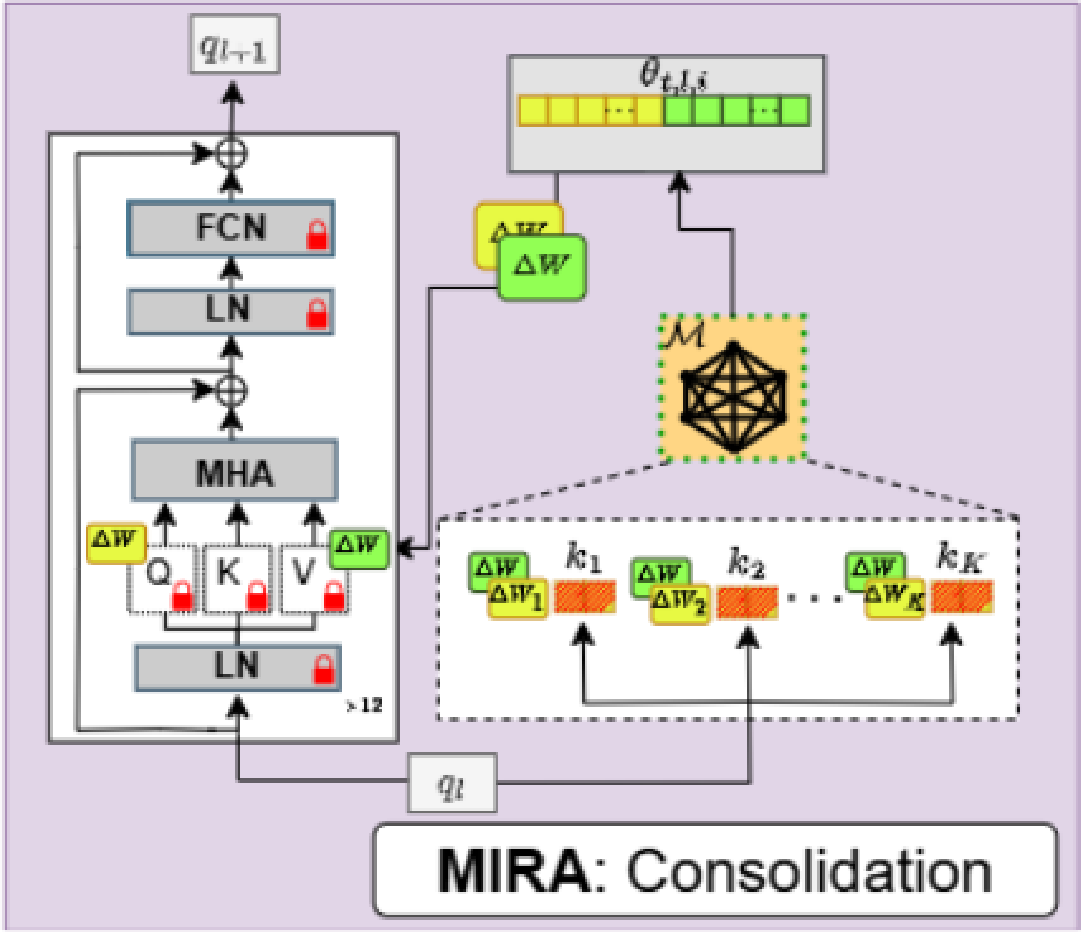
Phase-I: Building An *Expert* Arsenal [Adaptation]

- We instantiate a task expert as a low-rank adapter.
- A standard classification loss is used to train the adapters.
- Adapters are written to an encoder-specific AM at each encoder block and tied to randomly initialized keys.

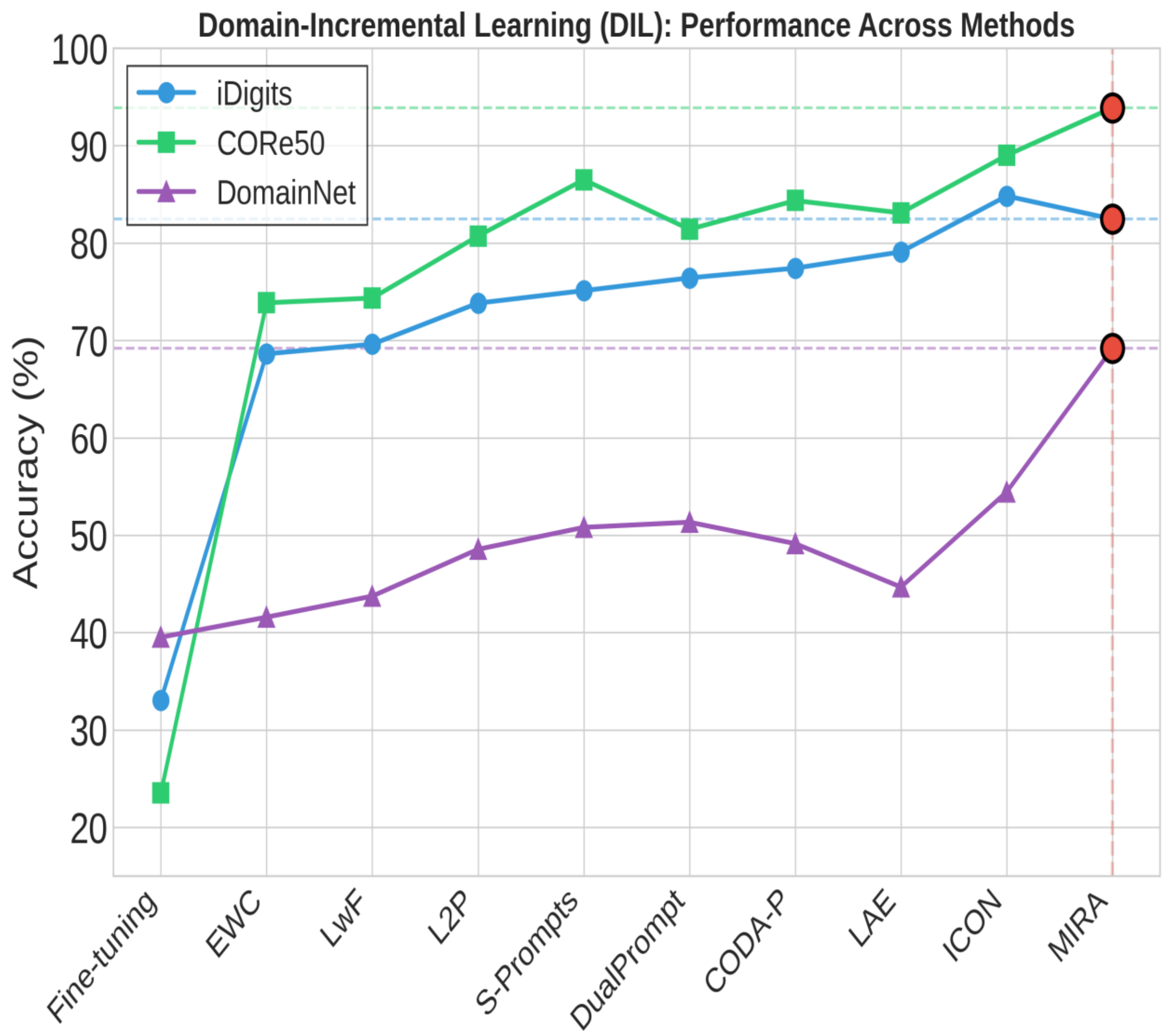


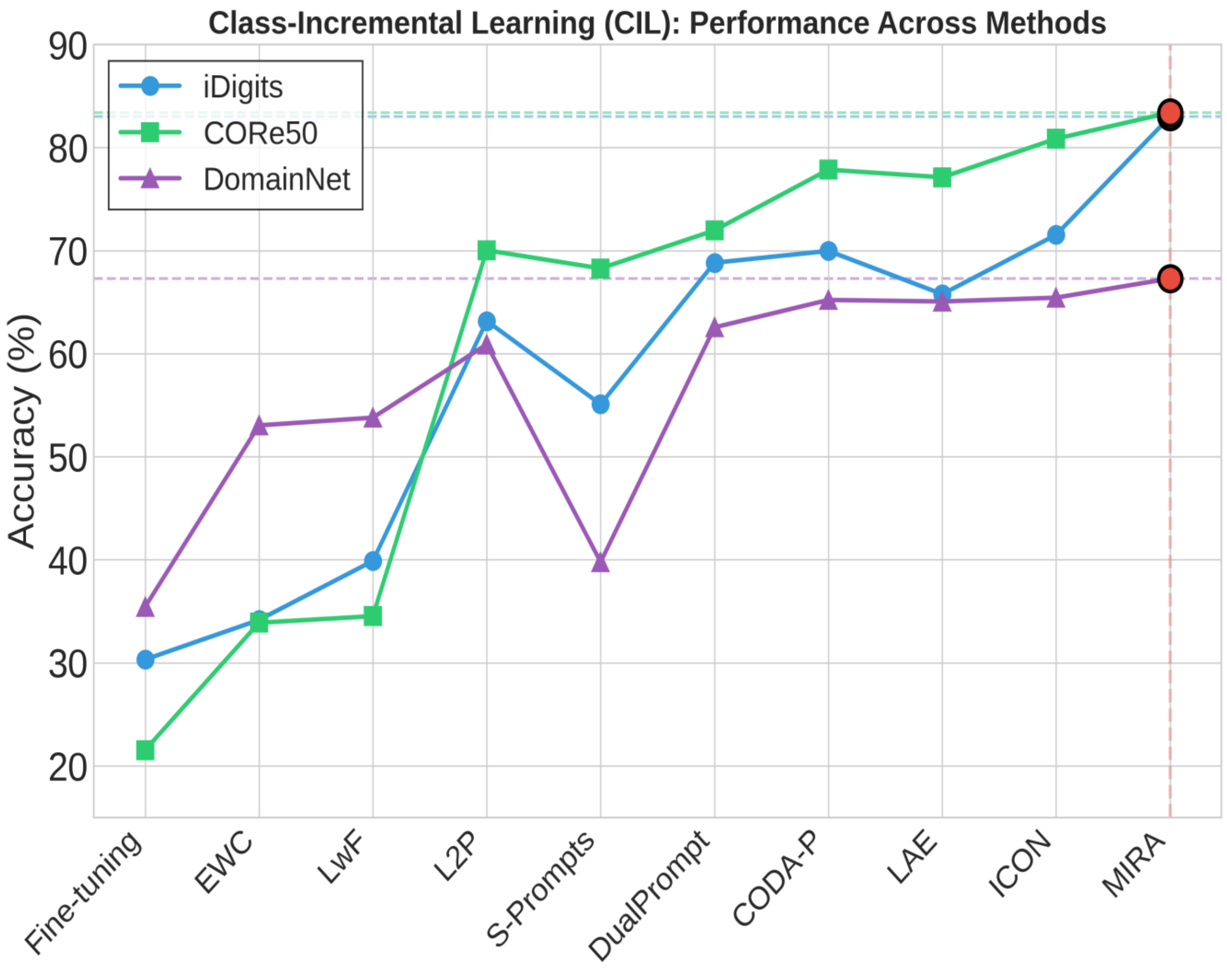
Phase-II: Learning What to Recall, When [Consolidation]

- Adapter-modulated network needs to consolidate knowledge learnt in Phase-I for test-time inference.
- Specifically, keys for hetero-associative Hopfield memory indexing into the adapters are *learnt* in this stage.
- Queries are cascaded across encoder blocks of the Transformer.

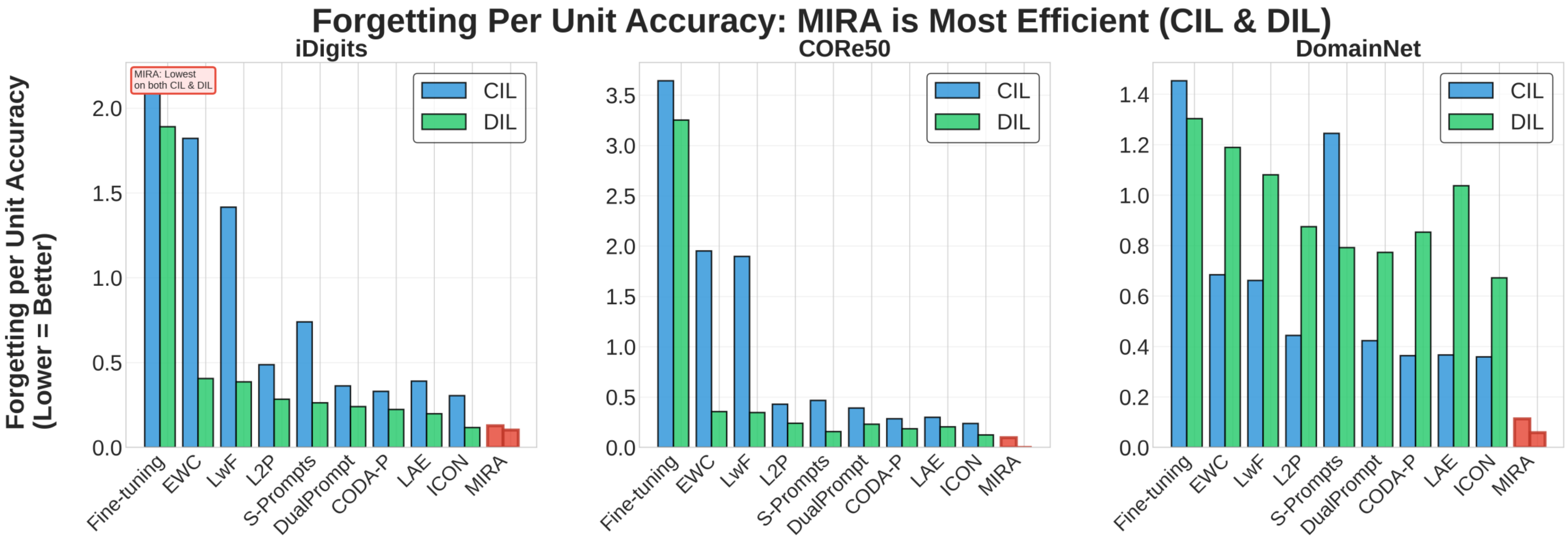


MIRA: SoTA (IL)

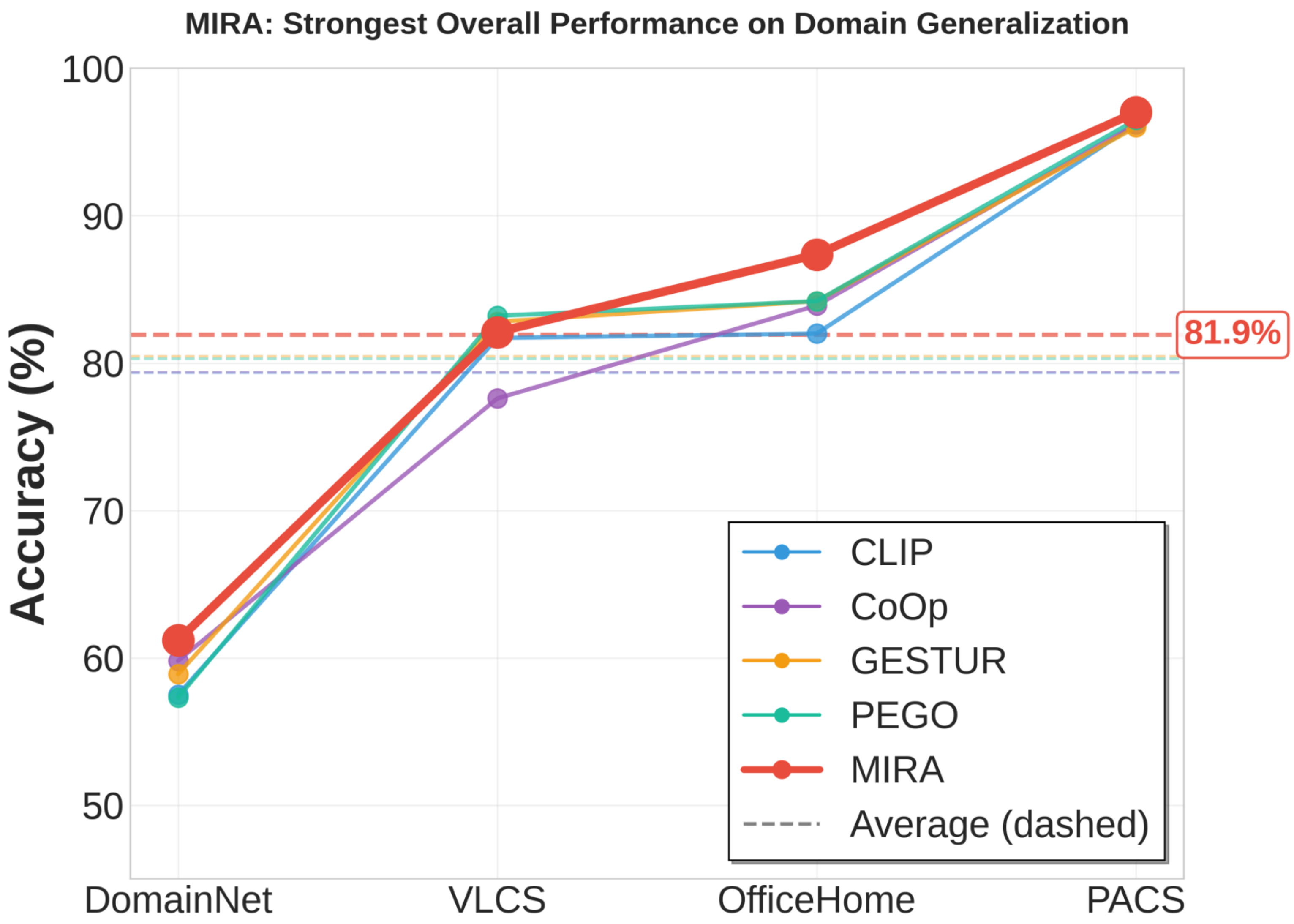

Domain-Incremental Learning (DIL): Performance Across Methods
Accuracy (%)
iDigits
CORe50
DomainNet
Fine-tuning
EWC
LwF
L2P
S-Prompts
DualPrompt
CODA-P
LAE
ICON
MIRA


Class-Incremental Learning (CIL): Performance Across Methods
Accuracy (%)
iDigits
CORe50
DomainNet
Fine-tuning
EWC
LwF
L2P
S-Prompts
DualPrompt
CODA-P
LAE
ICON
MIRA

MIRA: SoTA (IL) [Cont'd]



MIRA: SoTA (DG)



MIRA: Strongest Overall Performance on Domain Generalization
100
90
80
70
60
50
Accuracy (%)
81.9%
DomainNet
VLCS
OfficeHome
PACS
CLIP
CoOp
GESTUR
PEGO
MIRA
Average (dashed)

Summary & Conclusions

- AI systems lack an overarching framework that can rapidly switch across disparate tasks in an environment.
- We posit that memory could play a key role in AI systems to tackle this challenge, inspired from observations in neuroscience.
- We propose a way to integrate a differentiable memory unit (AM) into a Transformer architecture and refine its keys for extrapolation in diverse environments.
- Extensive experiments on IL and DG benchmarks demonstrate the importance of memory for task switching.