Optimizing Working Memory Sequence Manipulation through Permutation Group Decomposition

Kevin Bien[1], Junfeng Zuo[2,3], Wen-Hao Zhang[3]

[1]Texas Christian University, [2]Peking University, [3]UT Southwestern Medical Center

NEURAL INFORMATION PROCESSING SYSTEMS

Abstract

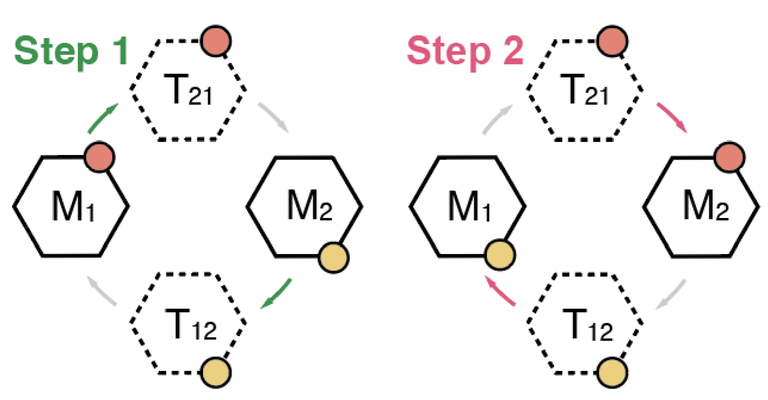
- Working memory (WM) maintenance and manipulation is essential for higher cognitive reasoning
- We design a biologically plausible neural circuit model for WM sequence manipulation by decomposing complex permutations into swapping sequences
- One-to-one correspondence between circuit model and Cayley graphs of the permutation group enables analysis of the trade-offs between circuit efficiency, complexity, and robustness

Swapping WM Items

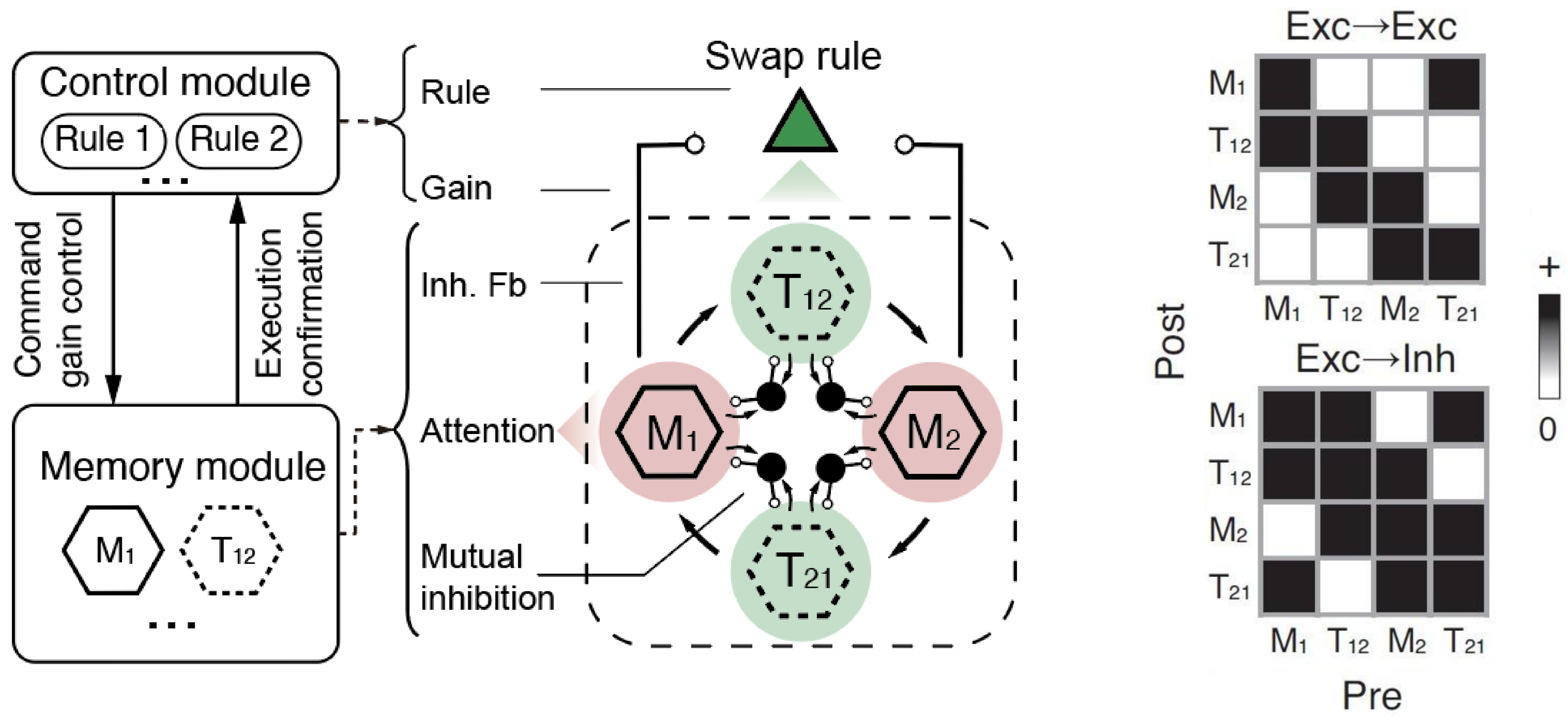
- WM items swapped through **parallel swapping** (1)

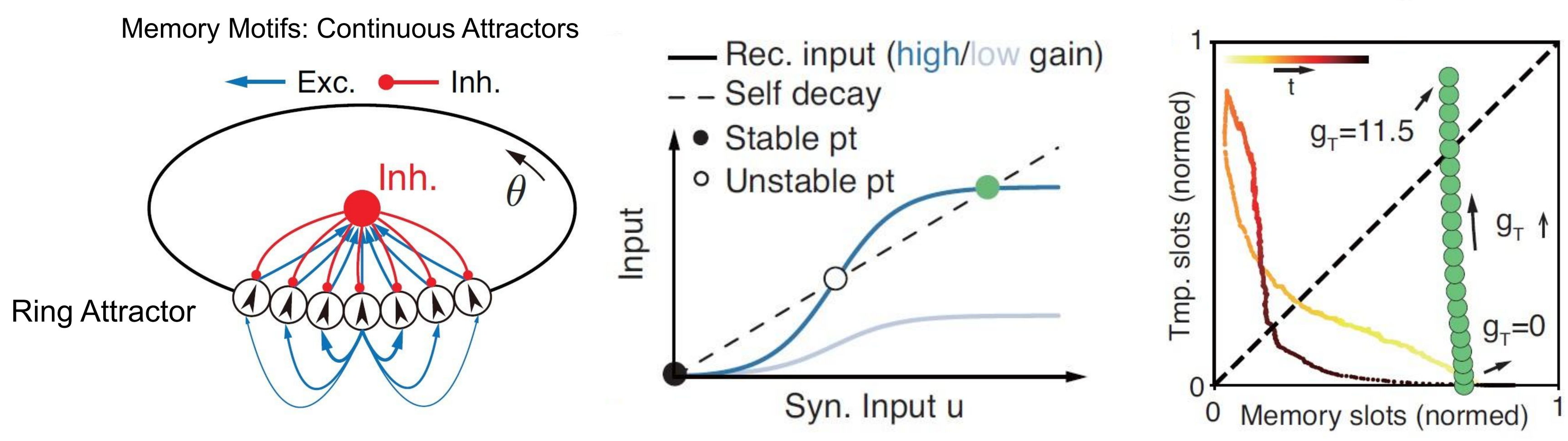
Parallel Swap(M_1,M_2)

Step 1
Processor 1: $T_{21} \leftarrow M_1$
Processor 1: $T_{12} \leftarrow M_2$
Step 2
Processor 1: $M_1 \leftarrow T_{12}$
Processor 2: $M_2 \leftarrow T_{21}$



- A neural circuit to implement parallel swapping (2):
  - WM items stored in memory motifs, which form the **memory module**
  - Memory motifs subject to gain modulation from neurons in the **control module**
  - **Mutual inhibition** between memory/temporary motifs



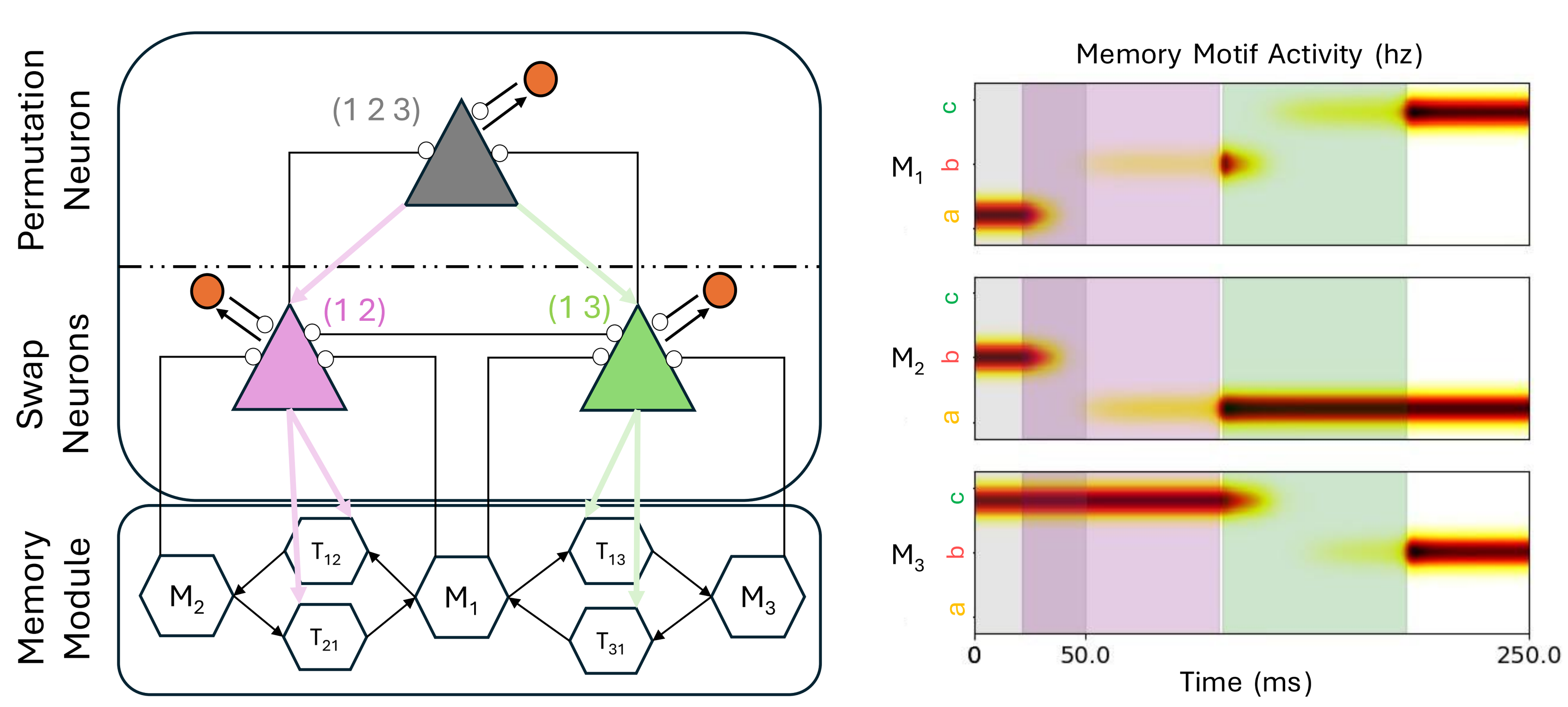


From Swaps to Sequence Manipulation

- Any n-item permutation can be decomposed into sequences of swaps

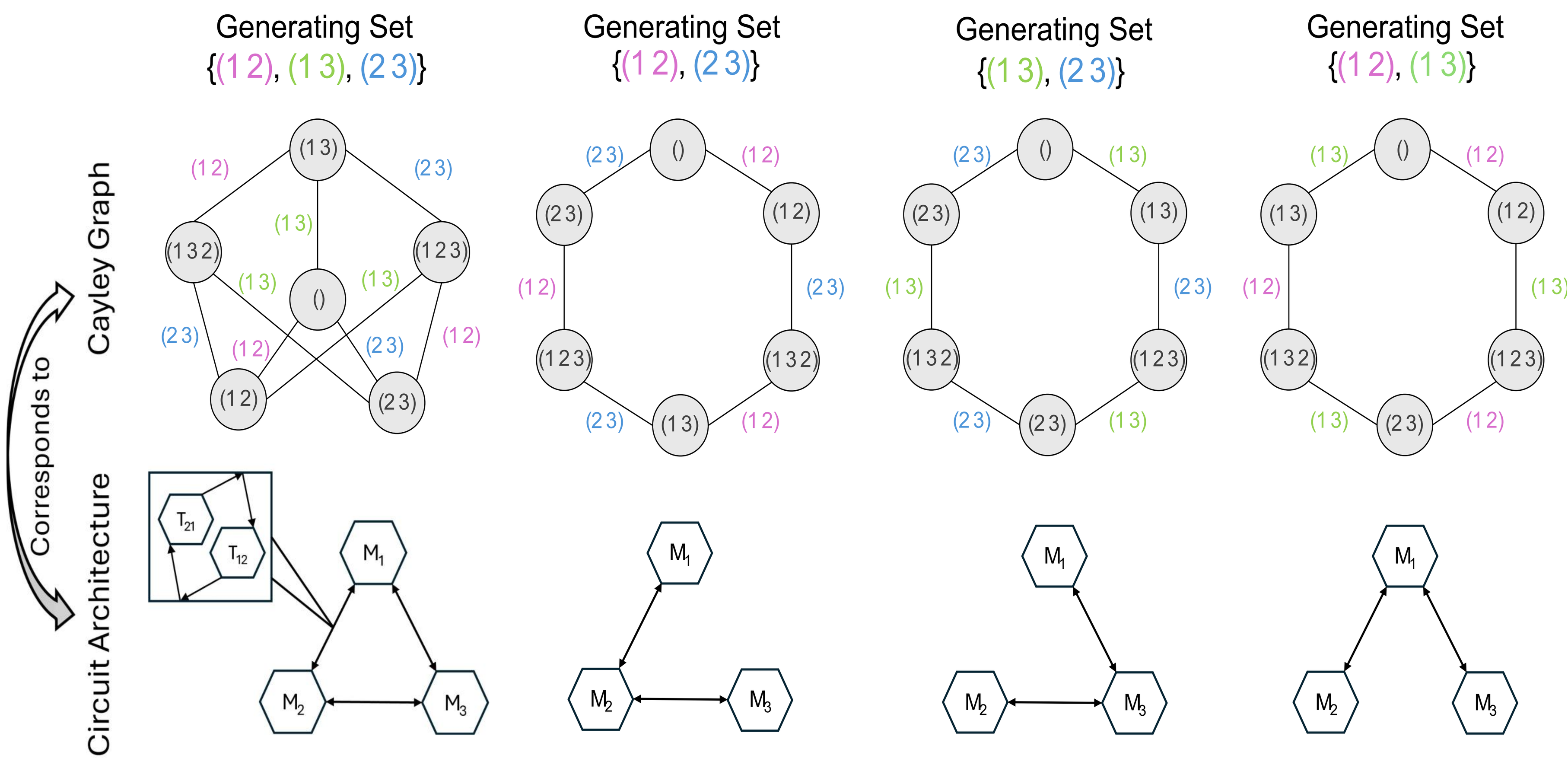
$$(1\ 2\ 3) : abc \xrightarrow{(1\ 2\ 3)} cab$$
$$= (1\ 3)(1\ 2) : abc \xrightarrow{(1\ 2)} bac \xrightarrow{(1\ 3)} cab$$

- **Generating set**: set of swaps that can be used to produce any permutation
- Swapping circuit can be extended for sequence manipulation:
  - **Control tree** modulates gain to swap neurons
  - Mutual inhibition between swap neurons ensures ordered execution
  - Connections between memory motifs determines generating set used



Architecture Comparison using Cayley Graphs

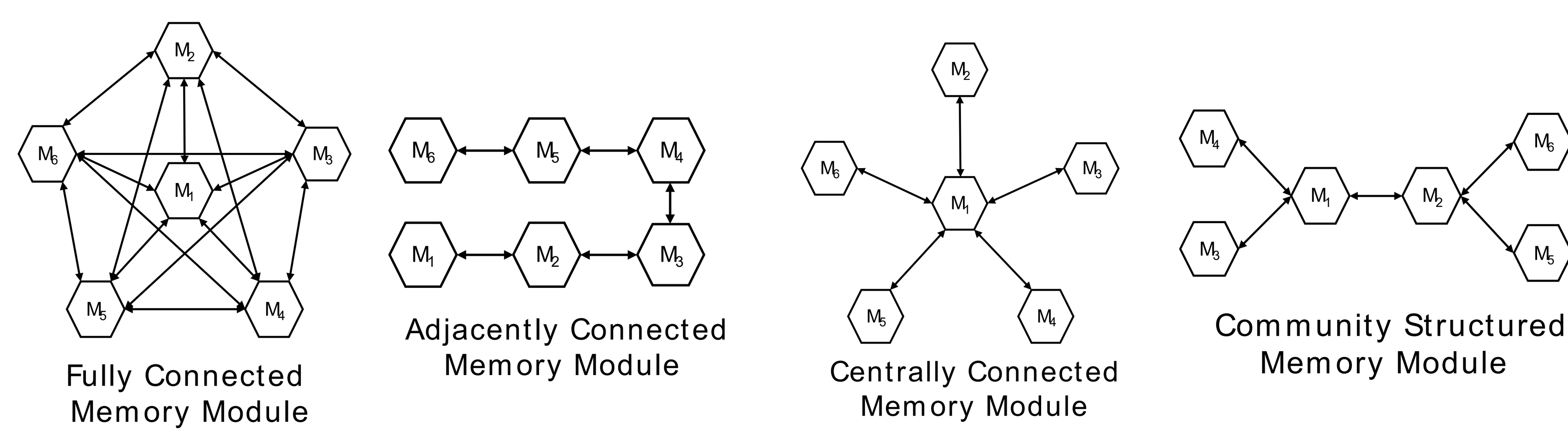
- **Cayley graphs** used to analyze how different generating sets/circuit architectures represent the permutation group



- More complex circuit ⇒ more highly connected Cayley graph ⇒ decomposition using fewer swaps

Balancing Efficiency, Complexity, Robustness

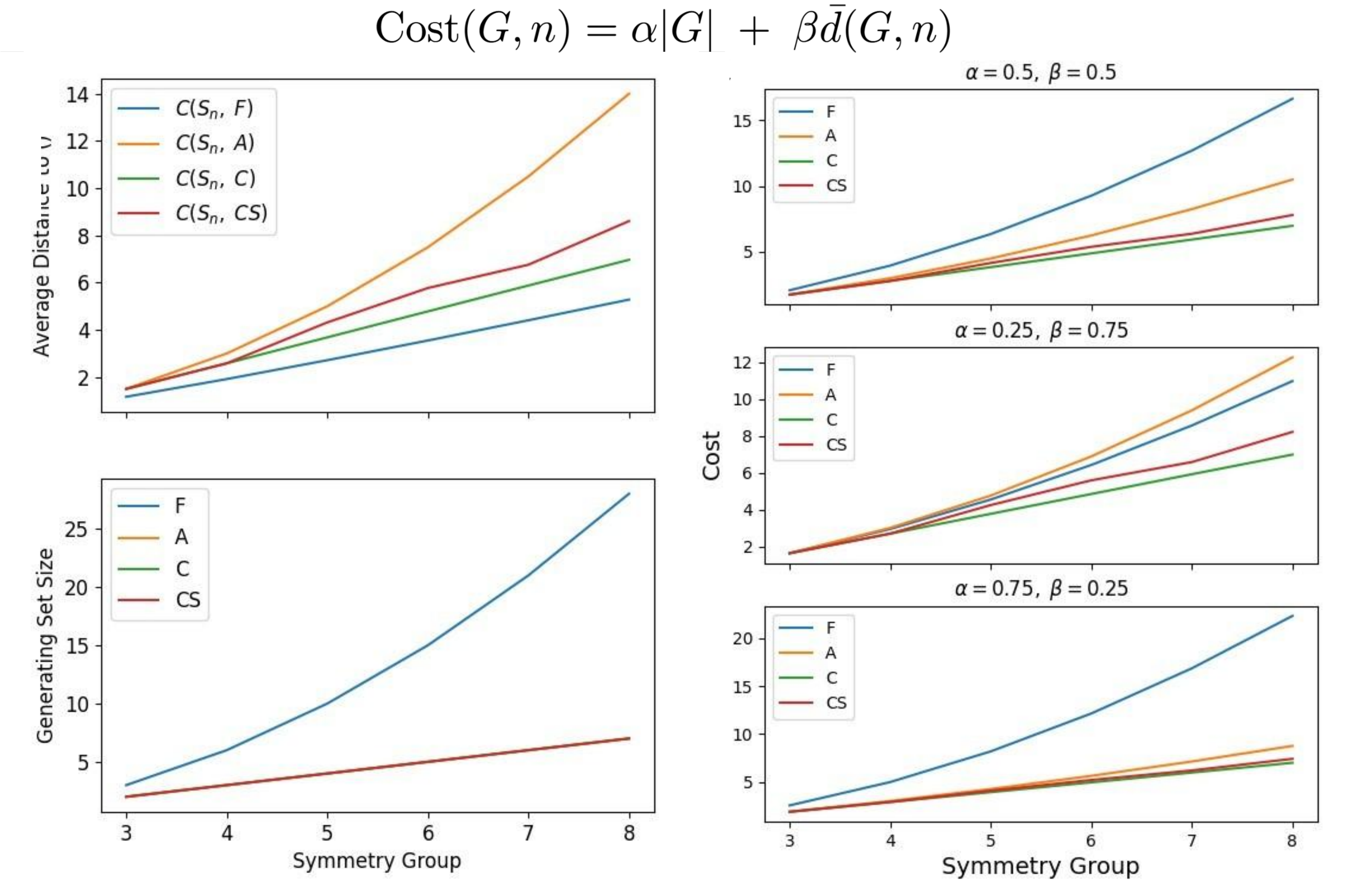
- Four candidate memory module architectures compared to investigate efficiency/complexity trade-off



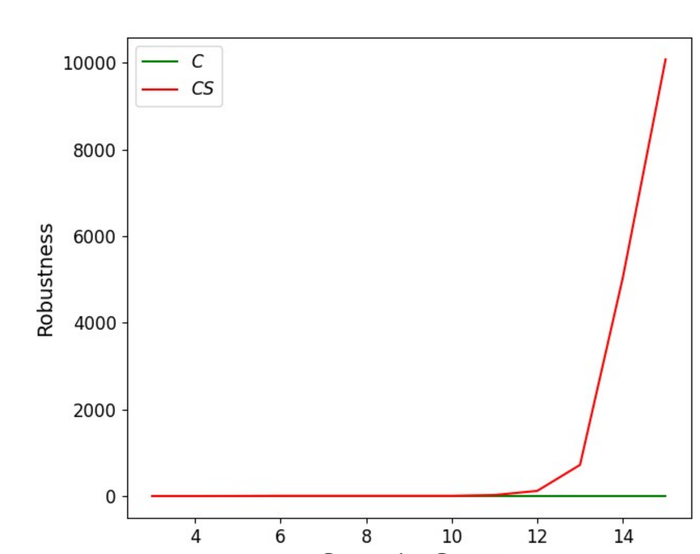
- **Cost function** to measure efficiency/complexity balance:
  - Weighted average of average decomposition length (efficiency) and generating set size (circuit complexity)

$$\bar{d}(G,n) = \frac{1}{n!}\sum_{v} d((),v), \quad v \in C(S_n, G)$$

$$\text{Cost}(G,n) = \alpha|G| + \beta\bar{d}(G,n)$$



- Centrally connected memory module achieves lowest cost but is less robust
  - **Robustness**: minimum number of permutations the circuit can produce after removing one memory motif
- **Community-structured** memory module is more robust and has similar cost to centrally structured architecture ⇒ it achieves optimal scaling for longer sequences



References

- Tian, Z., Chen, J., Zhang, C., Min, B., Xu, B., & Wang, L. (2024). Mental programming of spatial sequences in working memory in the macaque frontal cortex. Science, 385(6716), eadp6091.
- Zuo, Junfeng, Cheng Xue, Si Wu, and Wen-Hao Zhang. "Towards a mental programming neural circuit: Insights from working memory sequence manipulation." *bioRxiv* (2025): 2025-07.