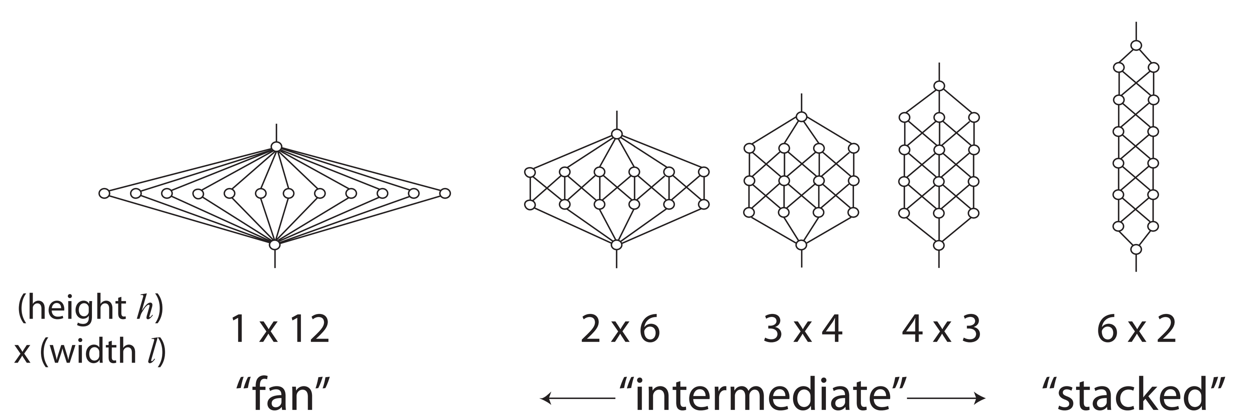
Learning and Memory Tradeoffs in Neural Networks

Ann M. Hermundstad, Kevin S. Brown, Danielle S. Bassett, and Jean M. Carlson
Department of Physics, University of California, Santa Barbara

ABSTRACT

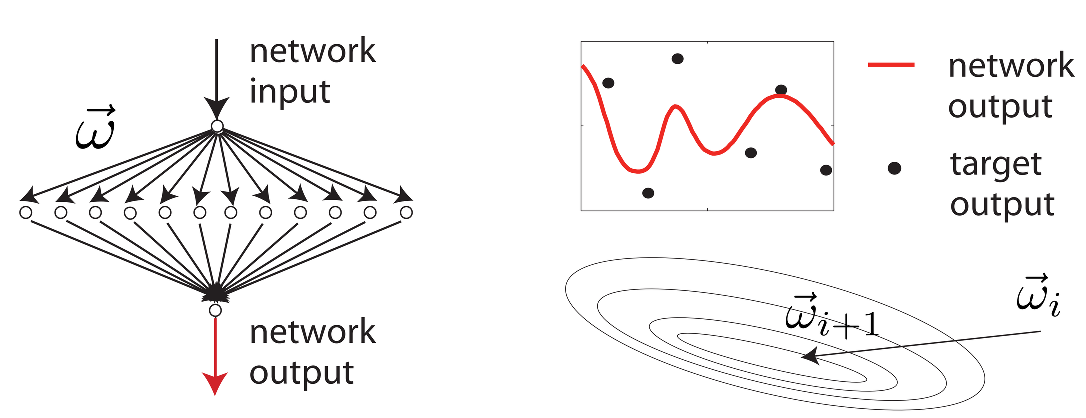
We compare the performance of parallel and layered network architectures during sequential tasks that require both acquisition and retention of information, thereby identifying tradeoffs between learning and memory processes. Performance is evaluated by statistically analyzing the error in representations of external information while varying the initial network state and the structure of the external information. We link performance to complexity in network architecture by characterizing local error landscape curvature. We find that variations in error landscape structure give rise to tradeoffs in performance; these include the ability of the network to maximize accuracy versus minimize inaccuracy and produce specific versus generalizable representations of information.

NEURAL NETWORK MODEL

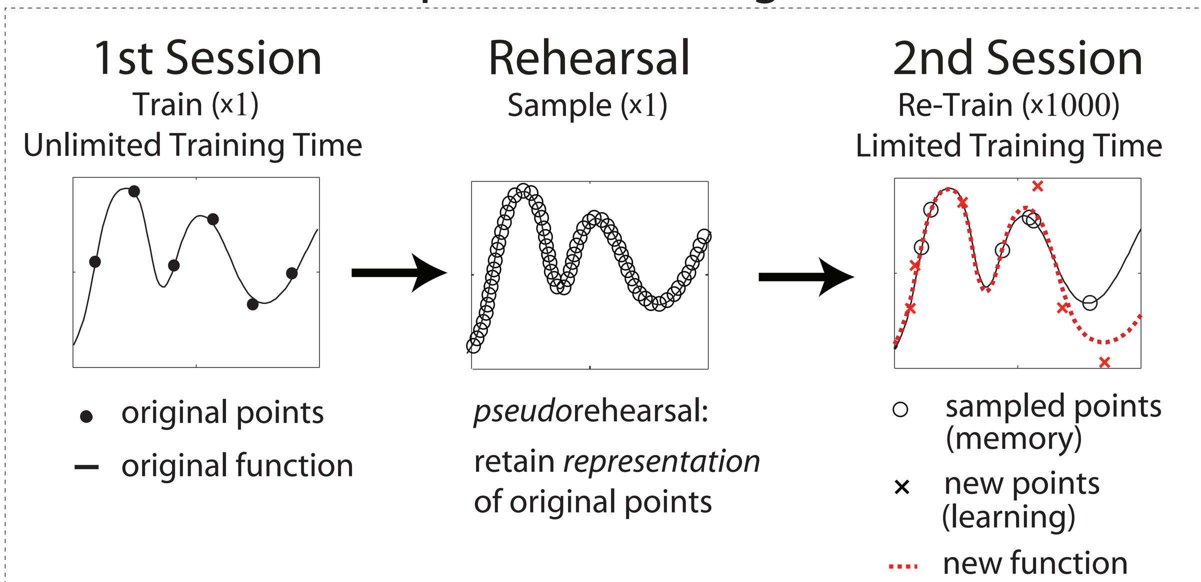


Function Approximation by Gradient Descent

- feedforward connections
- sigmoidal transfer functions
- variable connection weights $\vec{\omega}$
- adjust $\vec{\omega}$ to minimize error $E(\vec{\omega})$ between network and target output
- search error landscape for minima



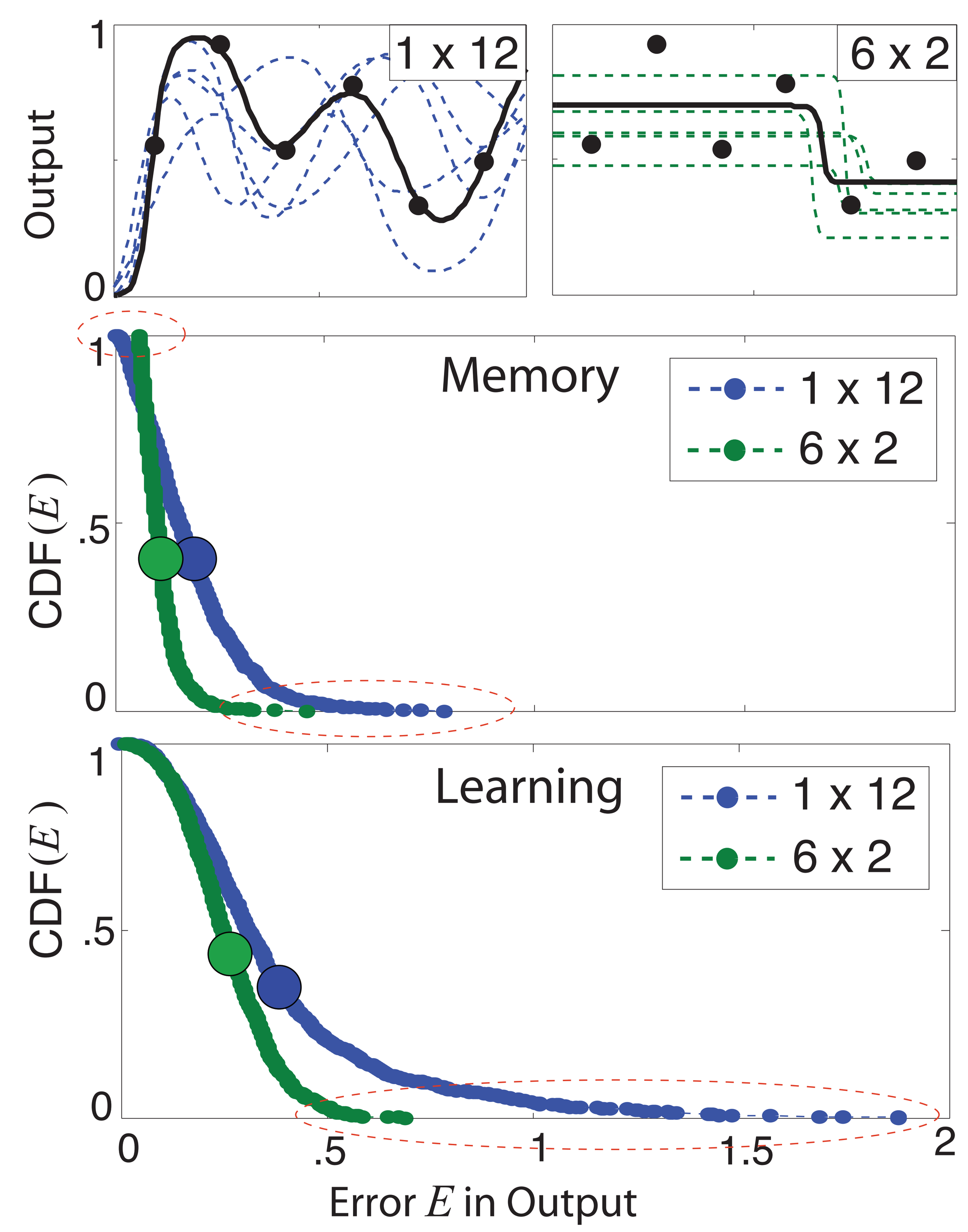
Sequential Training Method



TRADEOFFS IN LEARNING AND MEMORY

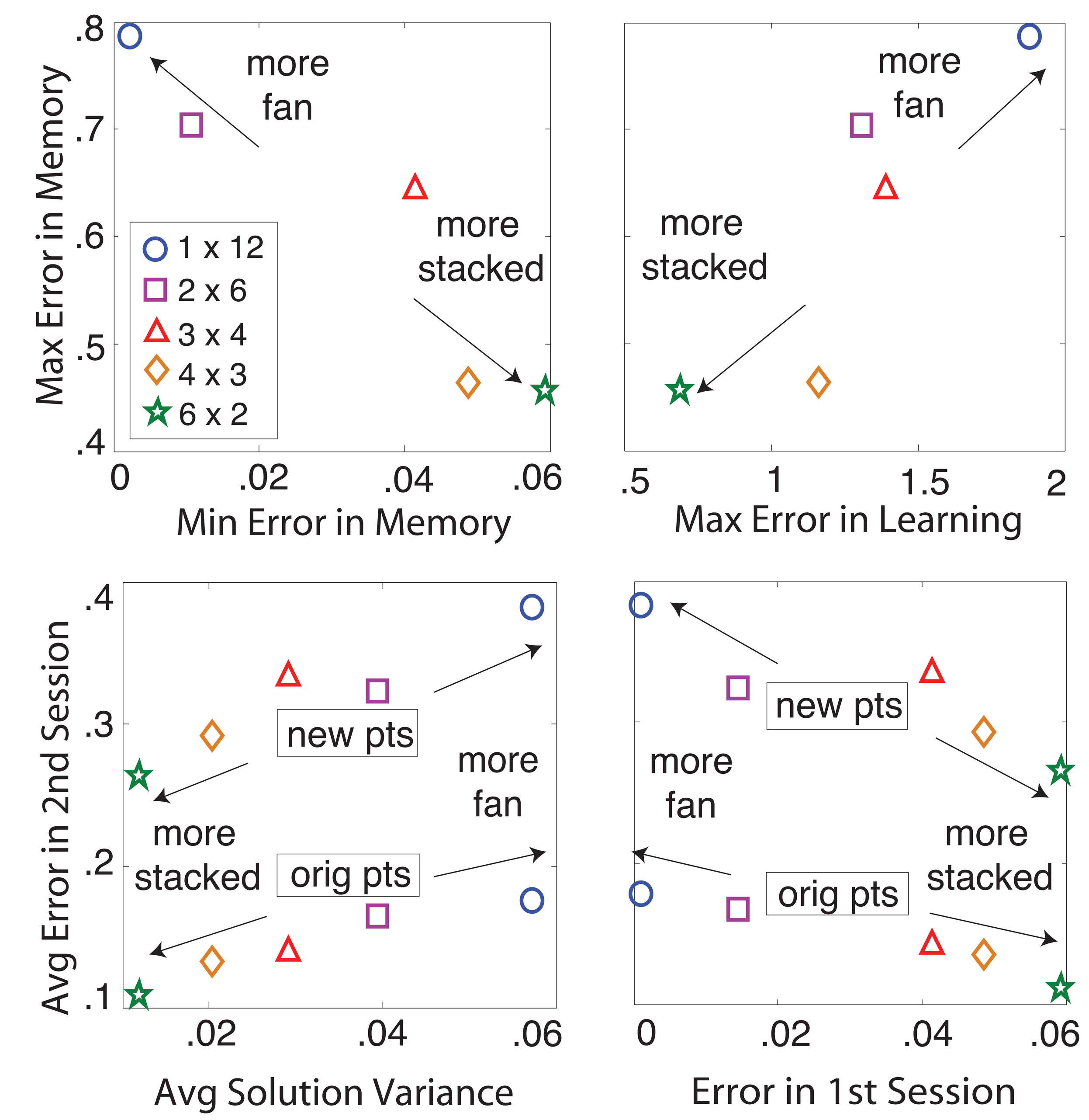
Effects of Structural Layering

- minimizes failure in both learning and memory
- decreases solution variance and average error



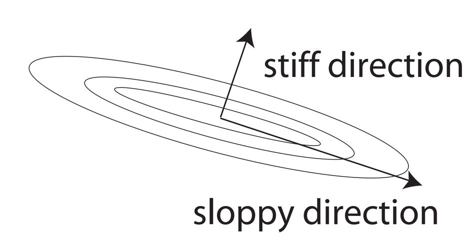
Tradeoffs Across Architectures

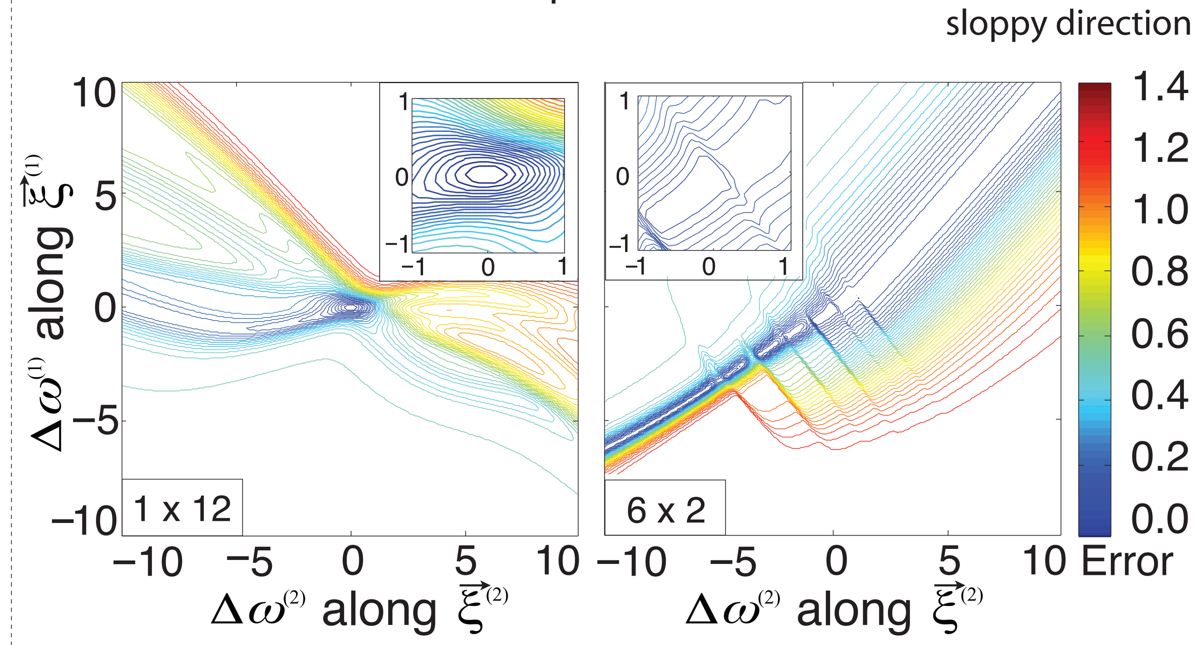
- maximizing success versus minimizing failure in memory
- average error in first versus second training session



ERROR LANDSCAPES

eigenvectors/values $(\vec{\xi}, \lambda)$ of error Hessian give directions/degrees of stiffness in error landscape



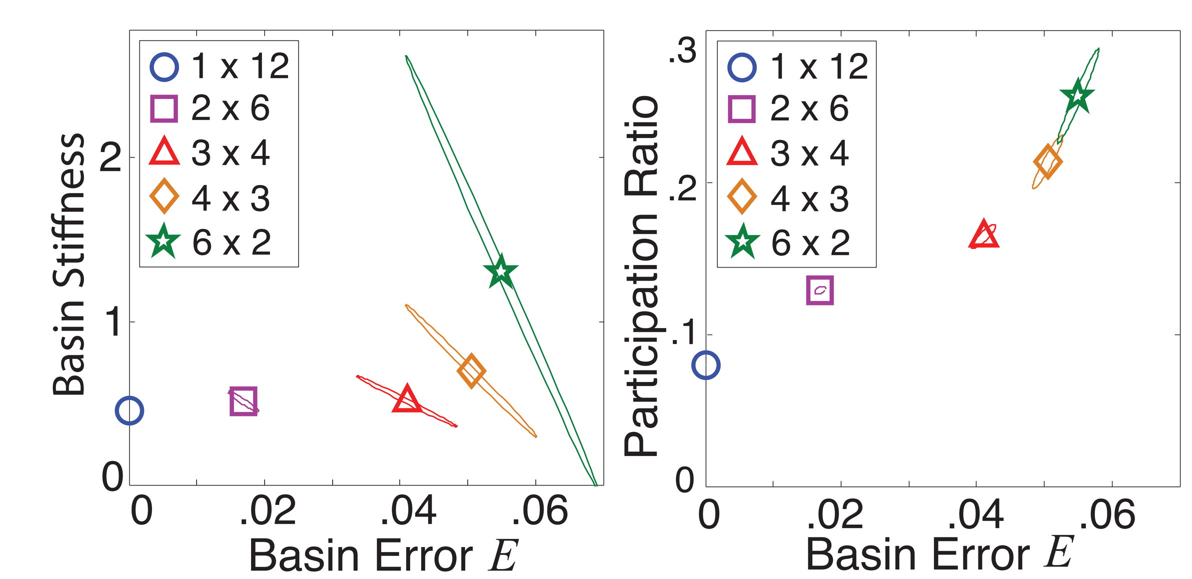


- fan landscape: single deep, narrow minimum, smooth landscape features
- stacked landscape: several putative minima separated by small barriers, rough landscape features

LANDSCAPE STRUCTURE

participation ratio ρ: eigenvector localization

$$\rho = \sum^{N} (\xi_n)^4 \quad \begin{cases} 1 & \text{single weight controls motion} \\ 1/\sqrt{N} & \text{all weights contribute equally} \end{cases}$$

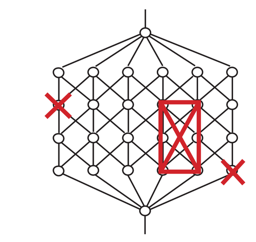


- layering increases variability in basin error E, basin stiffness λ, and participation of network connections ρ
- high error basins are shallower (lower degree of curvature) and require the adjustment of fewer weights (higher participation ratio)

FUTURE DIRECTIONS

Lesion Formation

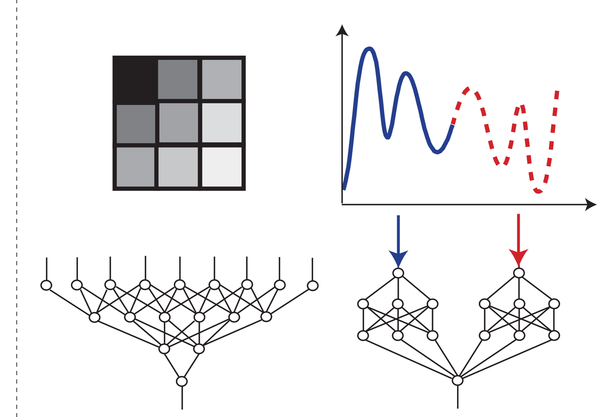
degradation of nodes and connections



- Are some architectures more stable to degradation?
- How do networks maintain functionality in the presence of structural changes?

Interaction with Environment

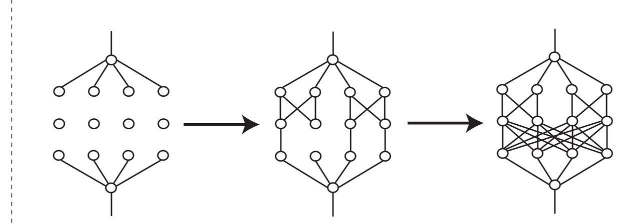
modularity and hierarchy in structure and function



- How does network structure (topological or functional) take advantage of structure in the environment?
- How does a network balance integration and segregation of information?

Network Evolution

adaptive architectures and learning algorithms



- How do structure and function coevolve?
- How do learning algorithms take advantage of dynamic structural changes?

CONCLUSIONS

Parallel Architectures: accurate but highly specific

- low error, stiff basins with low participation ratios
- must adjust large fractions of weights to navigate narrow basins
  - enables accurate solution given unlimited training time
  - prone to catastrophic failure when time is limited

Layered Architectures: coarse but generalizable

- basins with varying depth, stiffness, and participation ratio
- can adjust small fractions of weights to navigate shallow basins
  - enables coarses solutions in limited training time
  - prevents consistent accuracy when time is unlimited

Structural layering increases variability in error landscape features, which in turn gives rise to differences in performance and accounts for functional tradeoffs between learning and memory processes