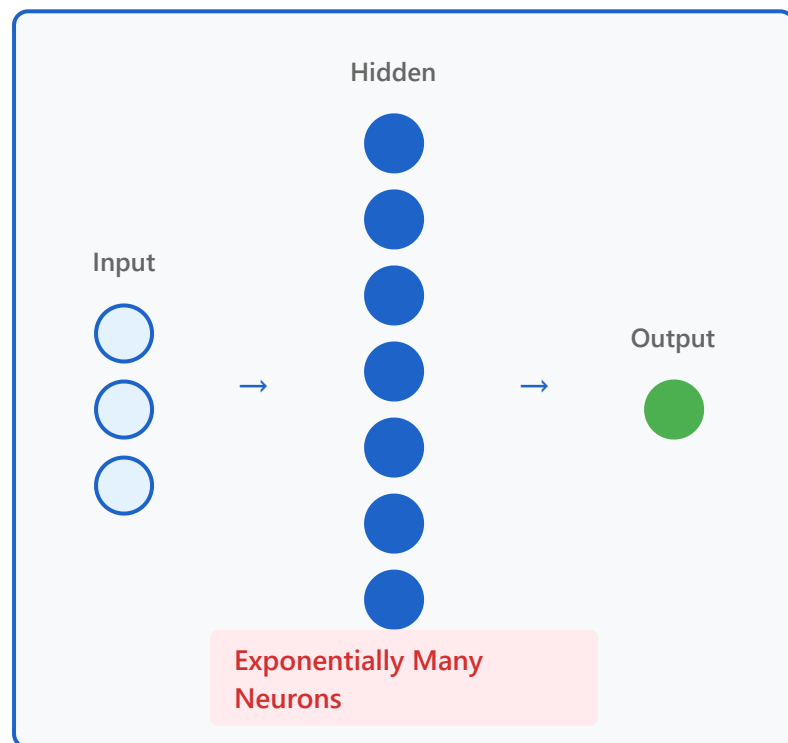


Expressiveness and Depth

Shallow & Wide

Single Hidden Layer

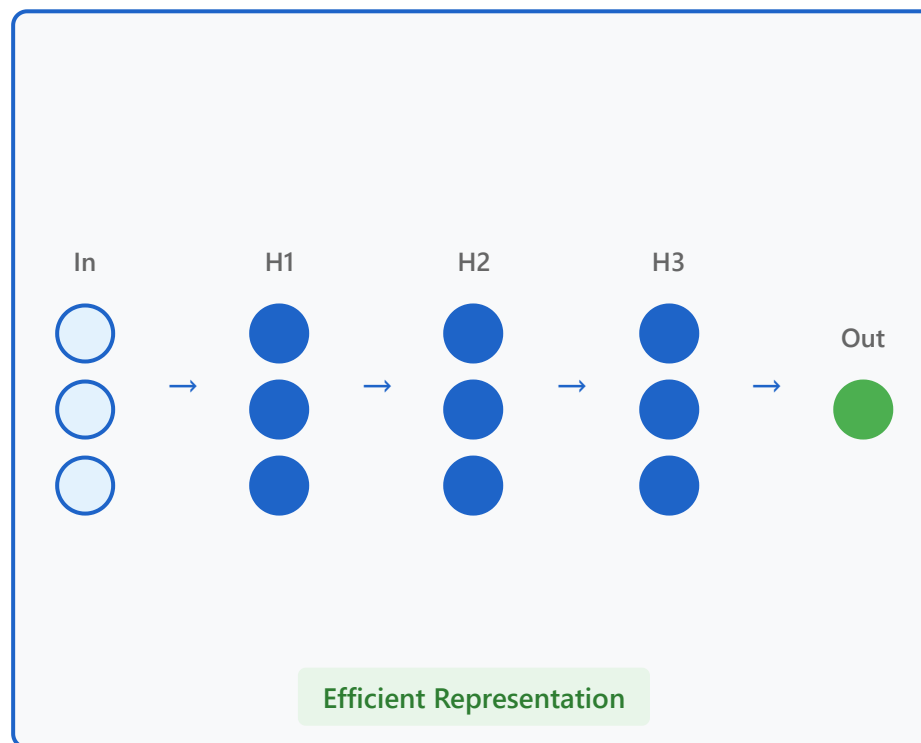


Number of Parameters

$O(2^n)$ vs $O(n^2)$

Deep & Narrow

Multiple Hidden Layers



Feature Learning

Flat vs **Hierarchical**

Generalization

Weaker vs Stronger

Compositionality

Limited vs Natural

Key Insight: Deep networks learn hierarchical abstractions efficiently, achieving exponential expressiveness with polynomial parameters through compositional depth