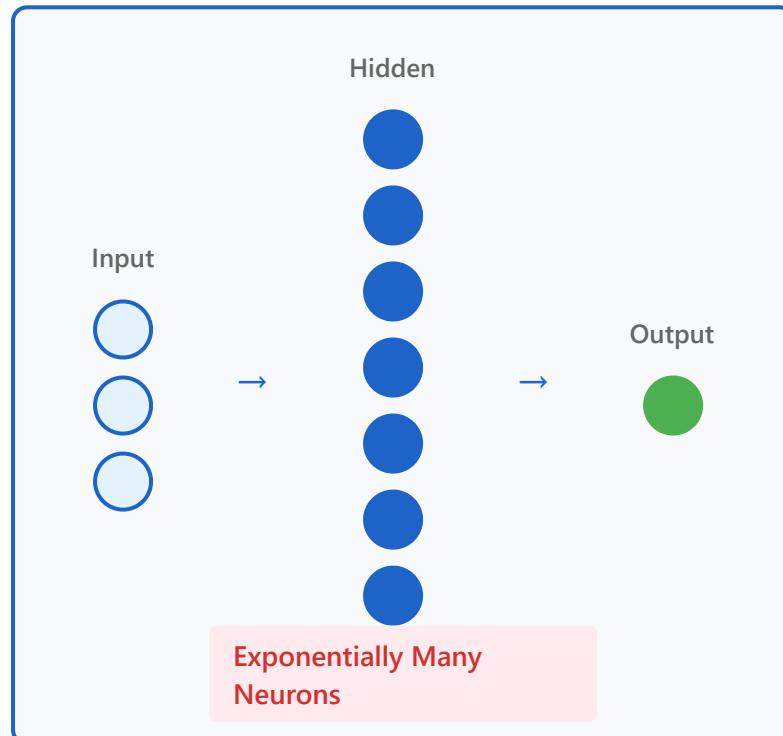


## Expressiveness and Depth

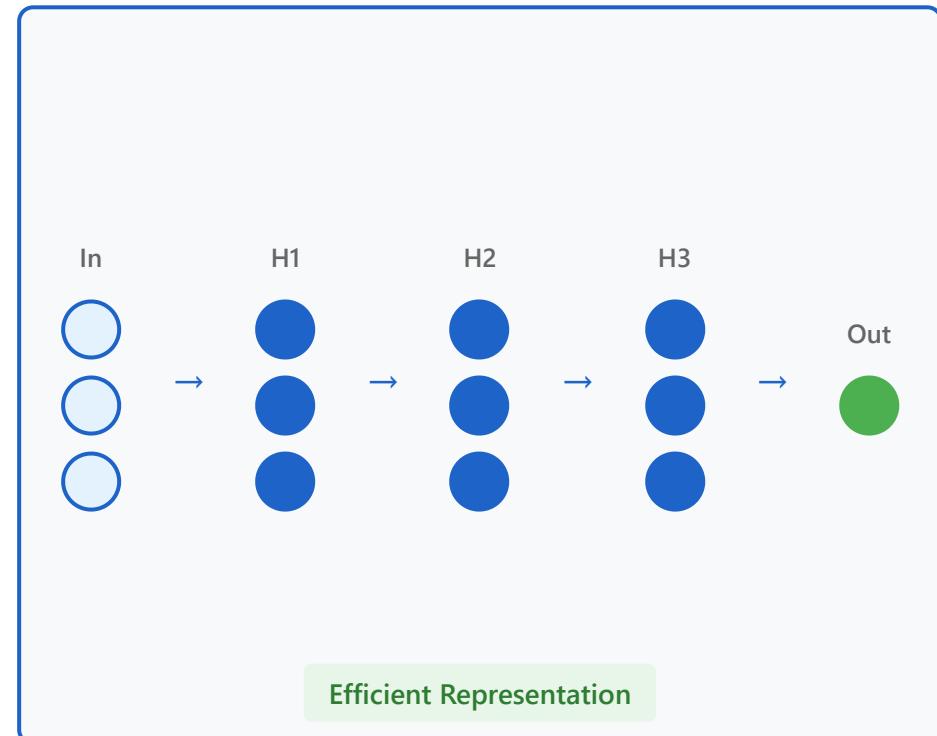
### Shallow & Wide

Single Hidden Layer



### Deep & Narrow

Multiple Hidden Layers



Number of Parameters

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

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