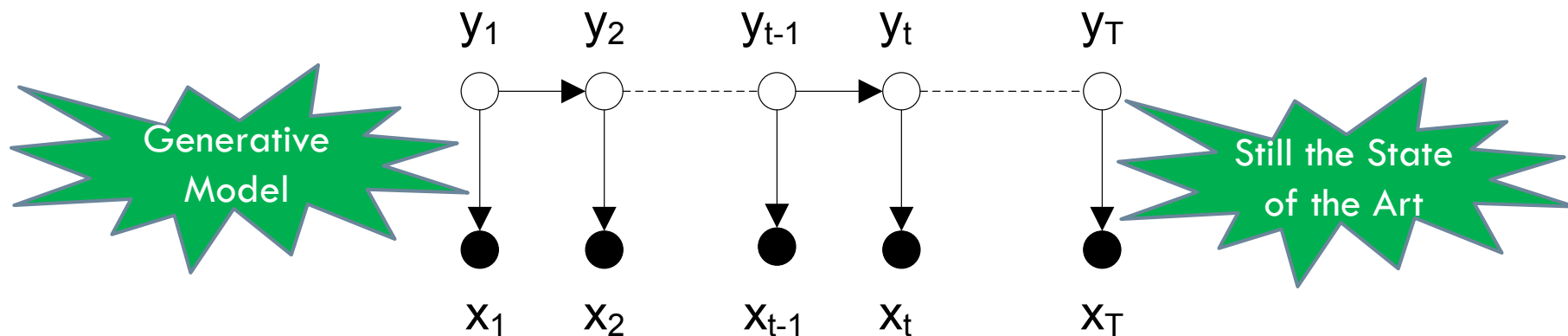


DEEP STRUCTURED DIRECT MODEL IN SPEECH PROCESSING

Dong Yu
Microsoft Research

HMM based Speech Recognition

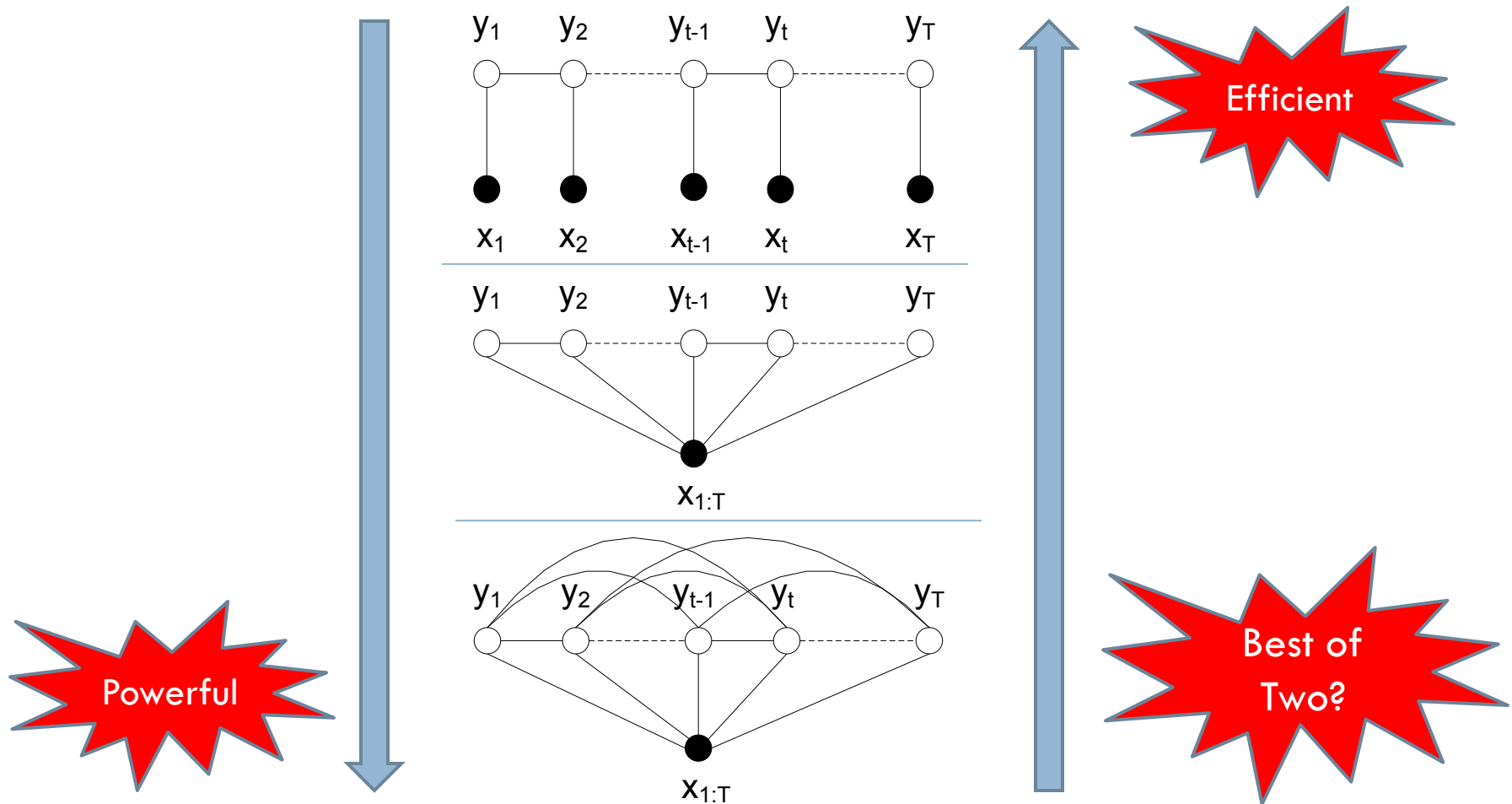
2



- Models joint distribution
- Observations are assumed to be independent given the state
- Short-to-mid-range dependency is modeled using context-dependent units: tri-phone or quin-phone
- Difficult to model long-range dependency: accents, noise, pronunciation, speaking rate

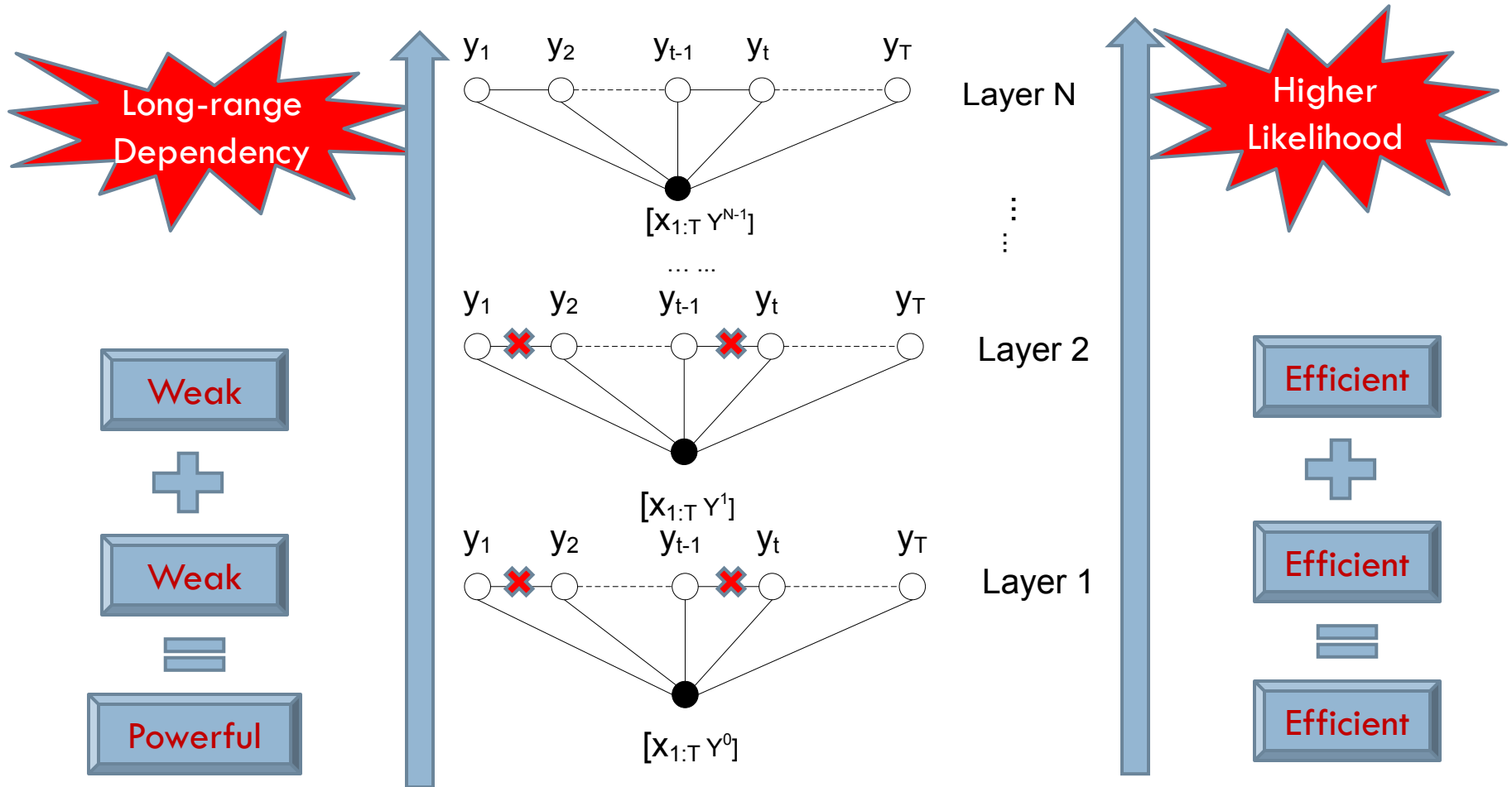
Shallow Conditional Models

3



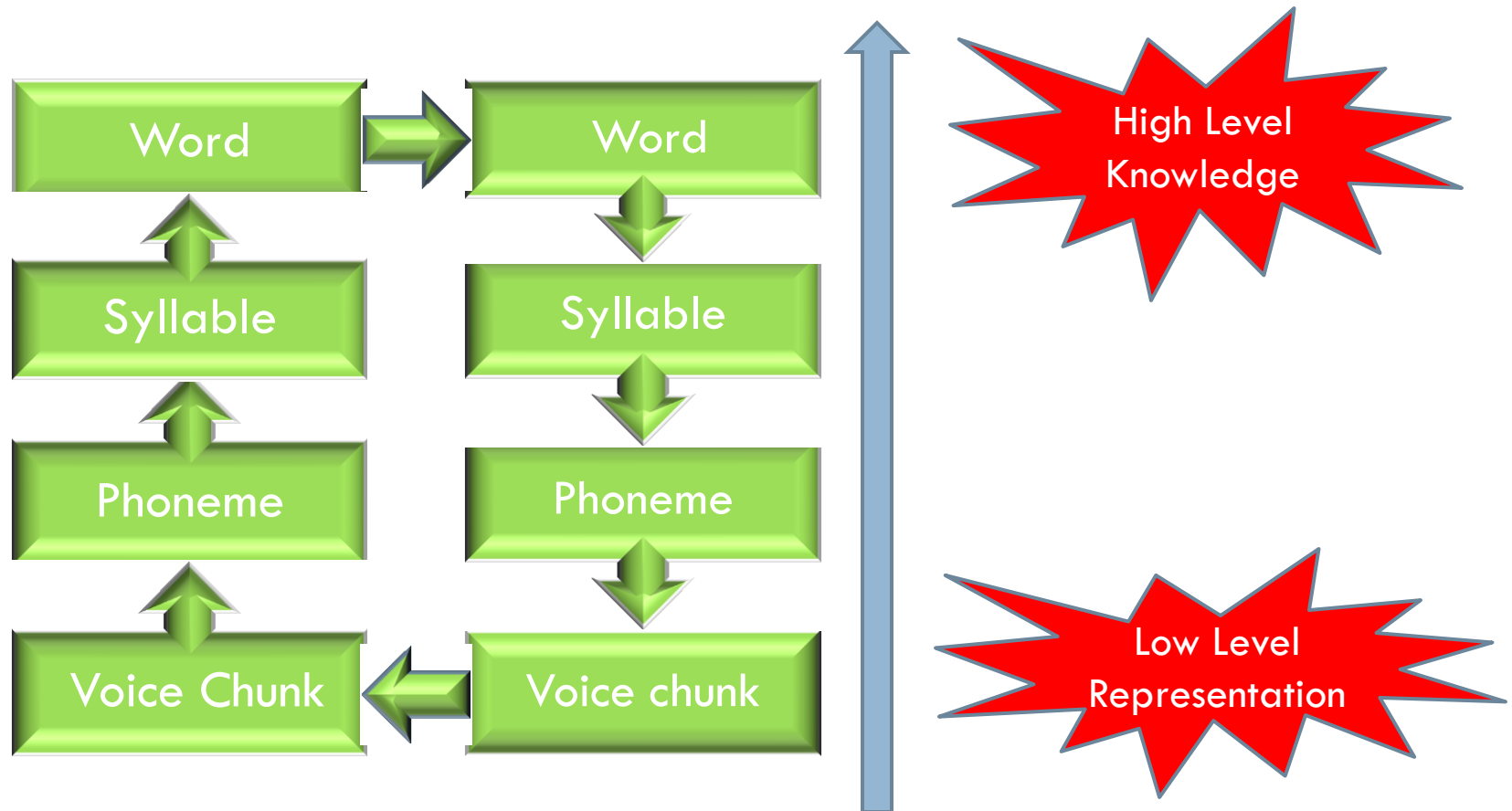
Deep Structured Model

4



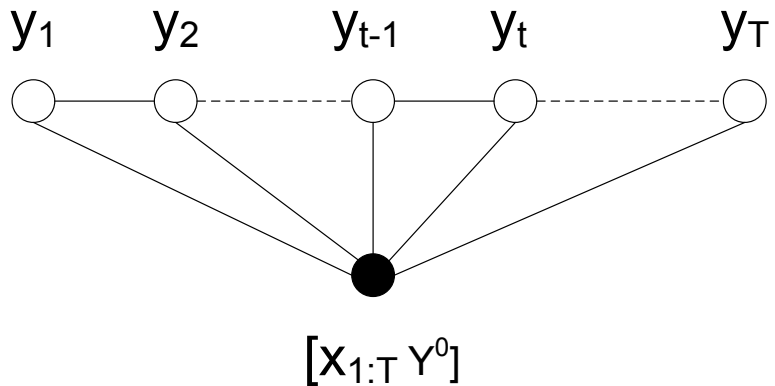
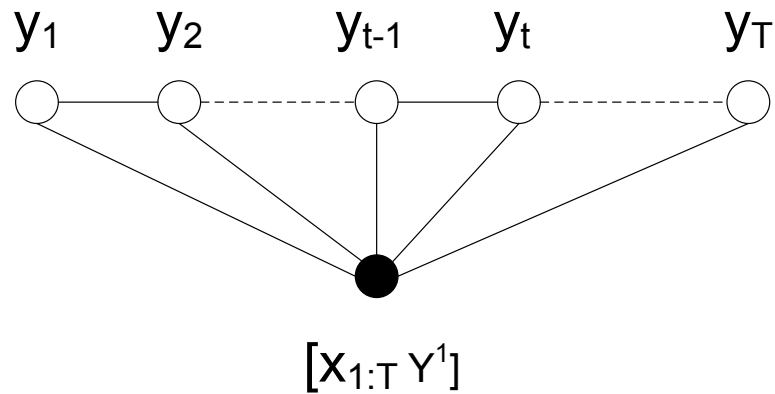
Hierarchical Units

5



Internal Sparse Representation

6



Thanks

7