Sequence Encoding
Self-Attention

國立臺灣大學 資訊工程學系
陳縕儂 助理教授
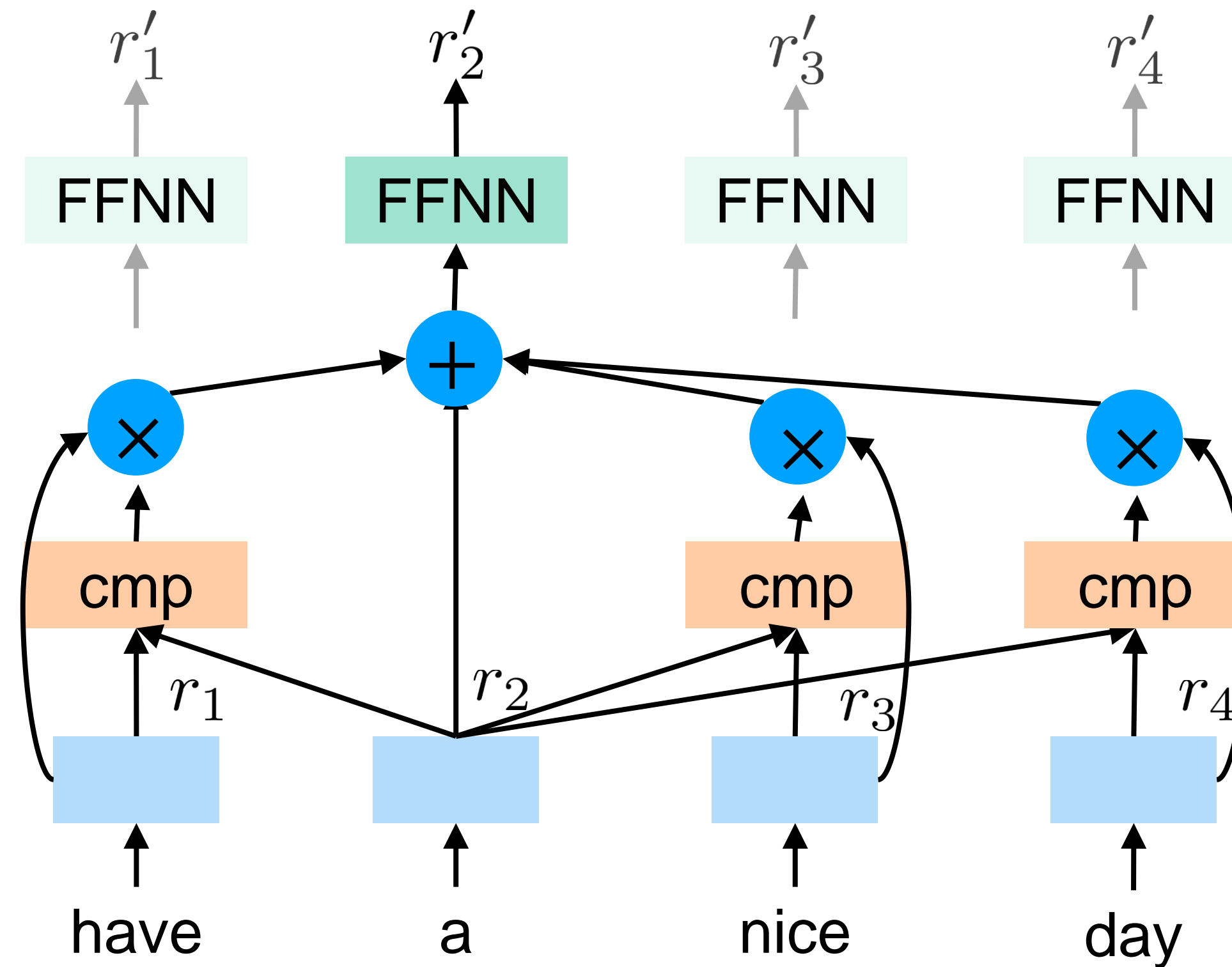http://vivianchen.idv.tw

Attention

- Encoder-decoder model is important in NMT
- RNNs need **attention mechanism** to handle long dependencies
- Attention allows us to access any state

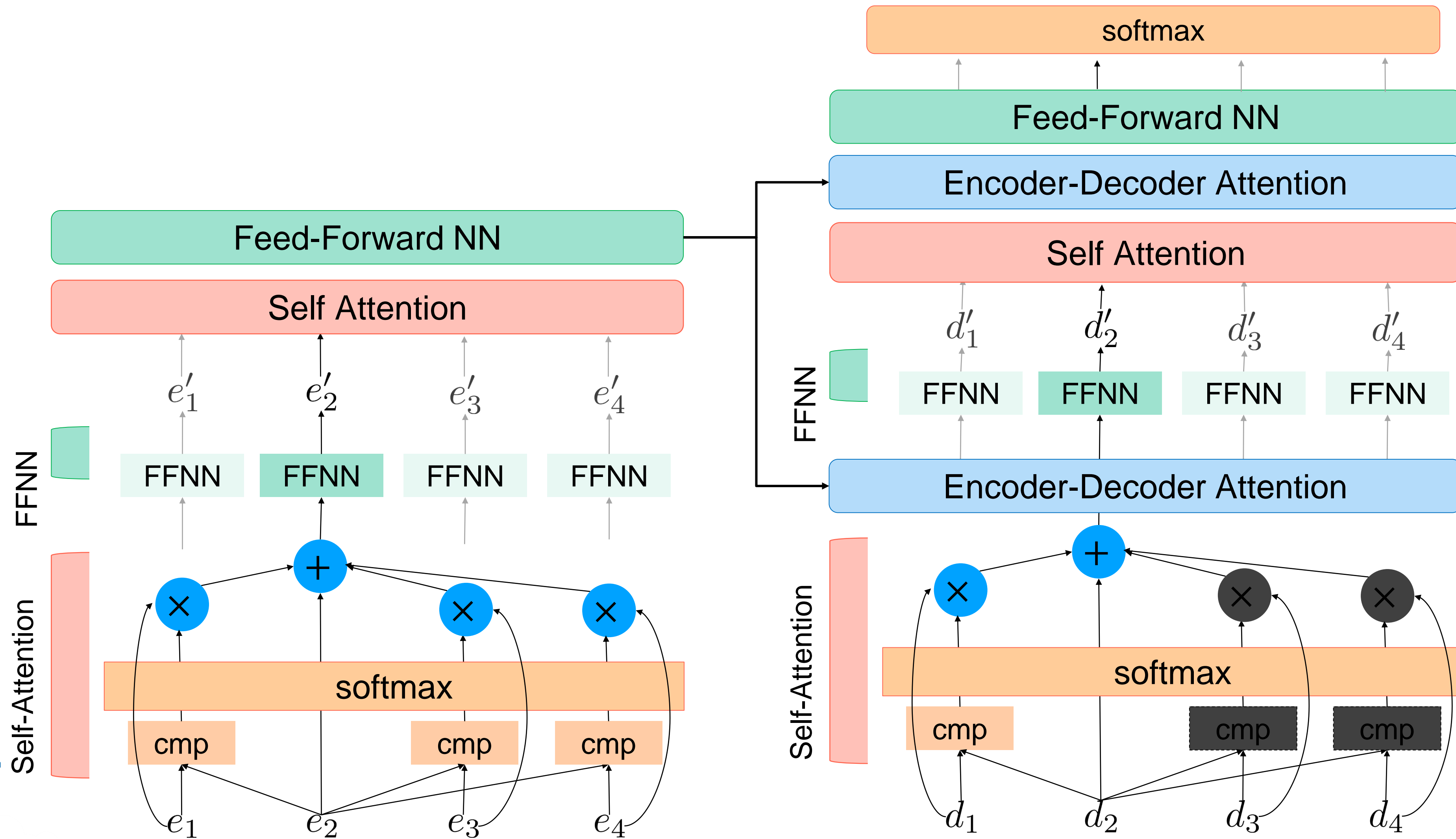Using attention to replace recurrence architectures

Self-Attention

- Constant "path length" between two positions
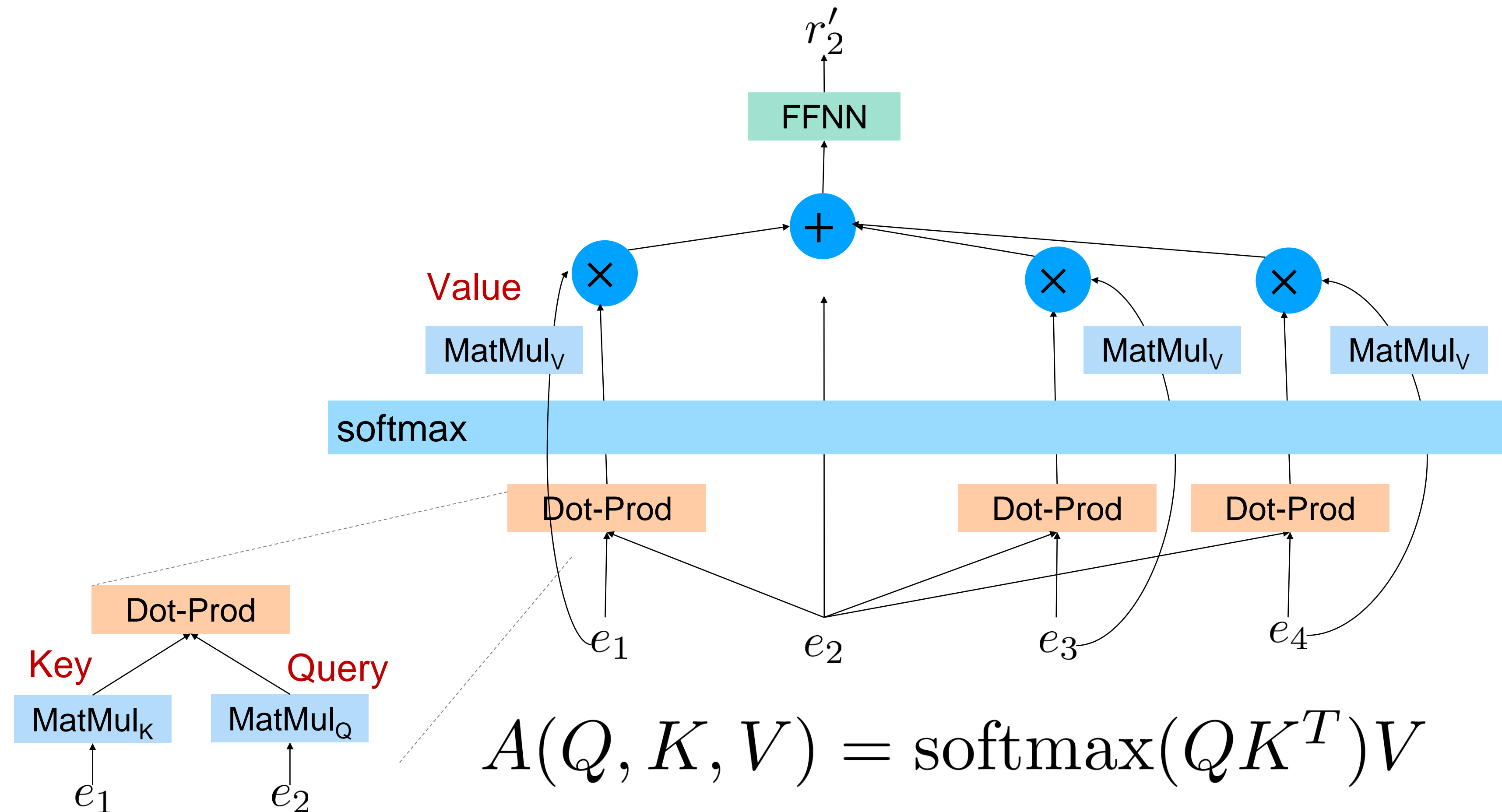- Easy to parallelize

Transformer Idea

Vaswani et al., "Attention Is All You Need", in *NIPS*, 2017.

Encoder Self-Attention (Vaswani+, 2017)

$$A(Q, K, V) = \mathrm{softmax}(QK^T)V$$

Vaswani et al., "Attention Is All You Need", in *NIPS*, 2017.

Decoder Self-Attention (Vaswani+, 2017)

Vaswani et al., "Attention Is All You Need", in *NIPS*, 2017.