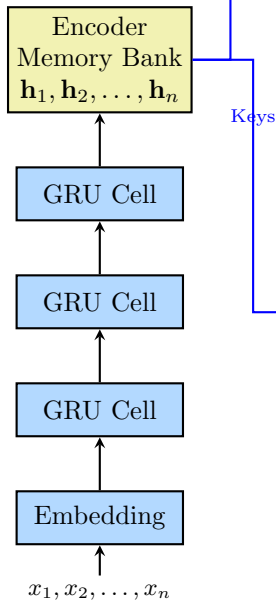
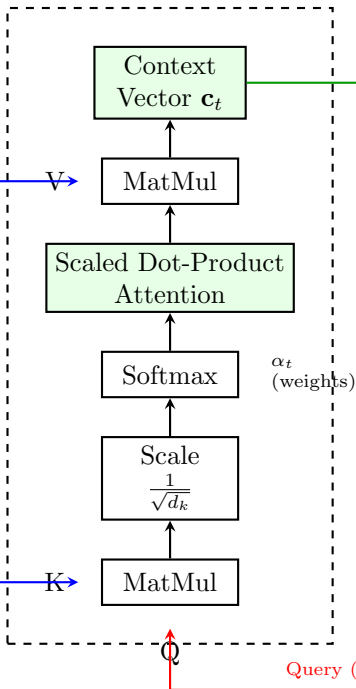


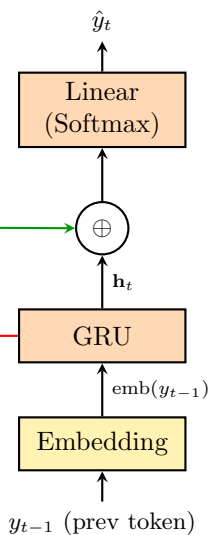
ENCODER



Cross-Attention



DECODER



Luong (post-RNN) Attention:

0. $y_0 = \text{< bos >}$
1. $\text{emb}(y_{t-1}) \rightarrow \text{GRU} \rightarrow \mathbf{h}_t$
2. \mathbf{h}_t queries encoder $\rightarrow \mathbf{c}_t$
3. $[\mathbf{h}_t, \mathbf{c}_t] \rightarrow \text{output}$

Decoding Step t