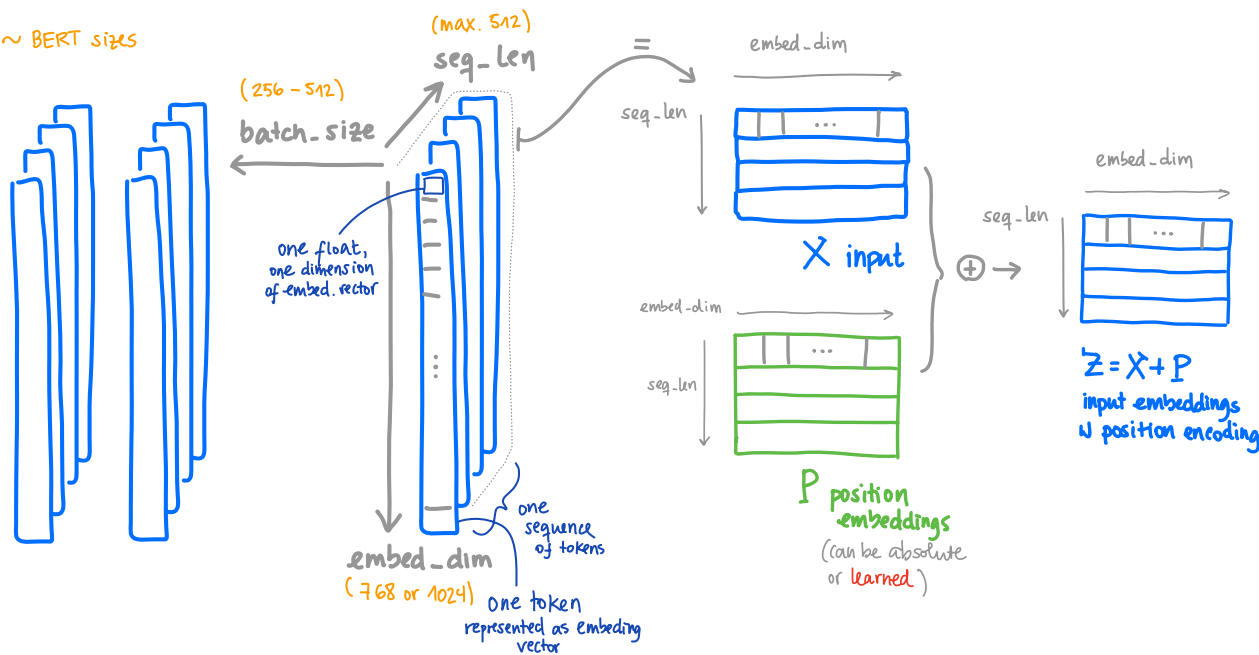
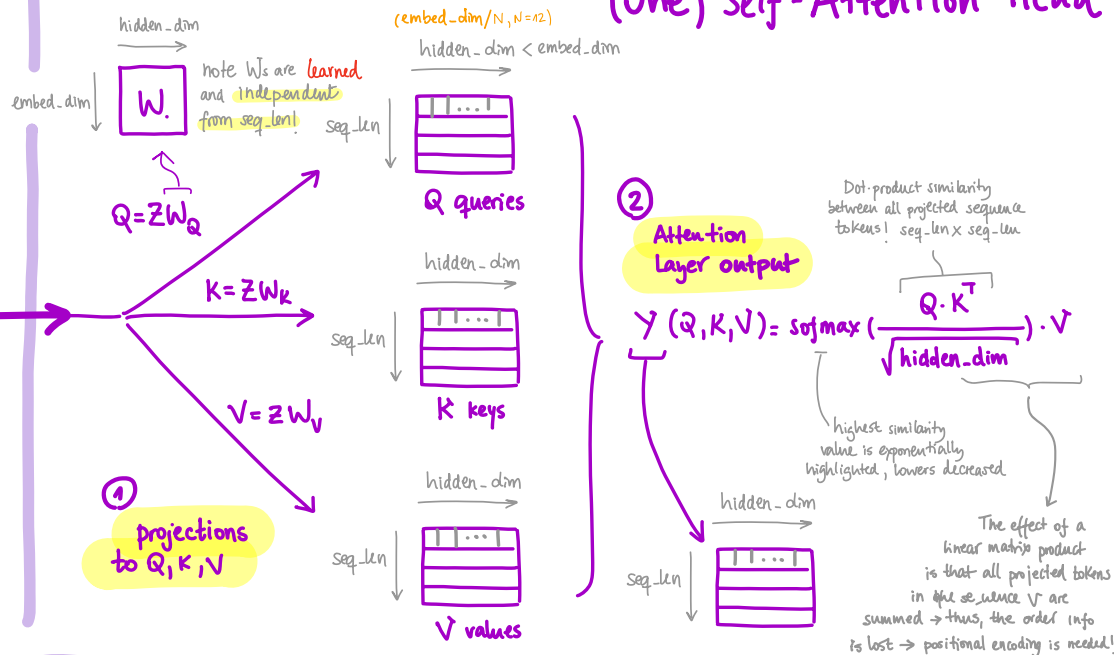


SIMPLIFIED TRANSFORMER-ENCODER ARCHITECTURE

~ BERT sizes



(One) Self-Attention Head



Multi-head Attention

(N attention layers)
(12 in BERT)

