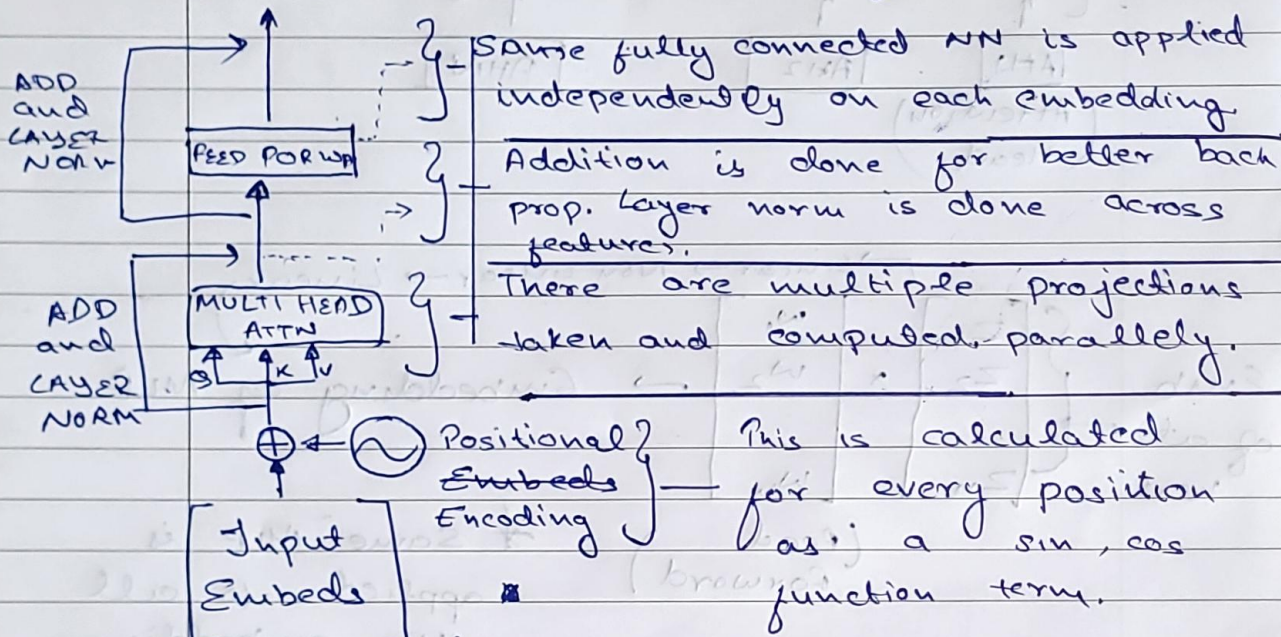


Encoder Decoder and Attention Architectures

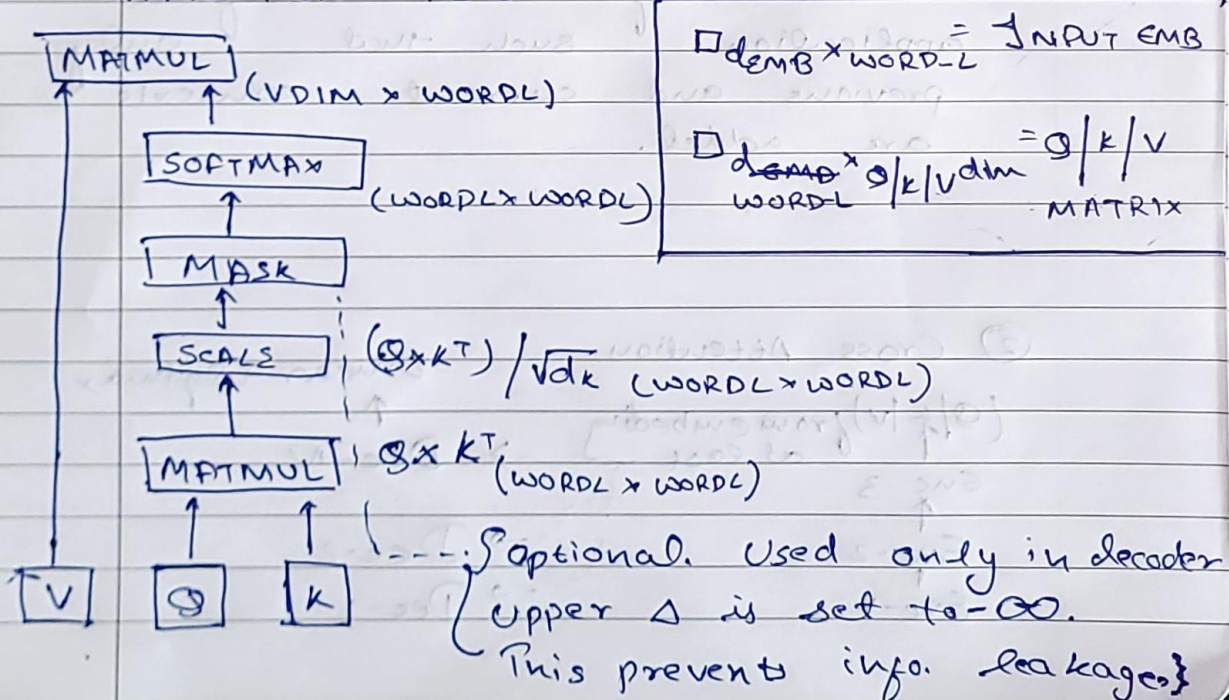
□ (VAL DIMENSION × NUM OF WORDS)

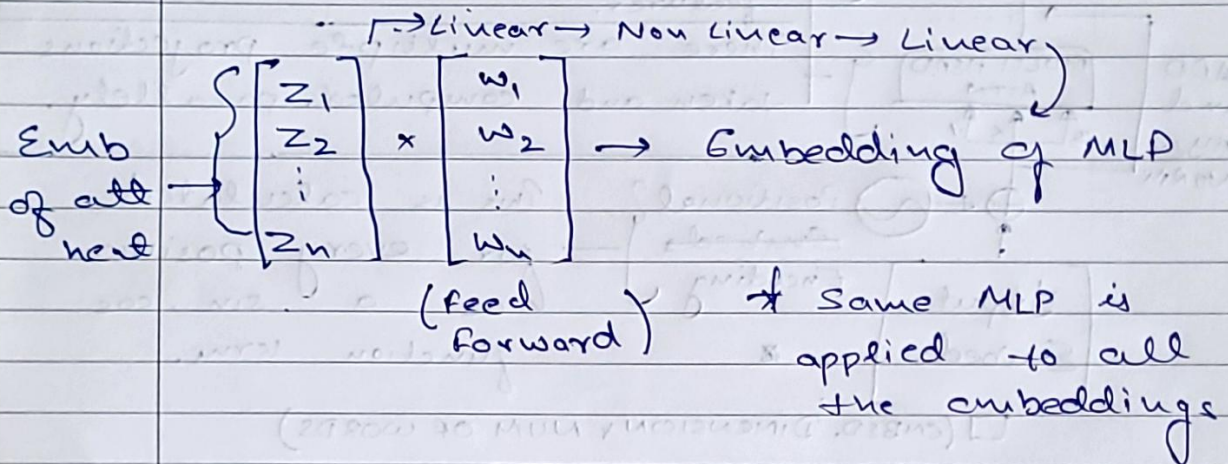
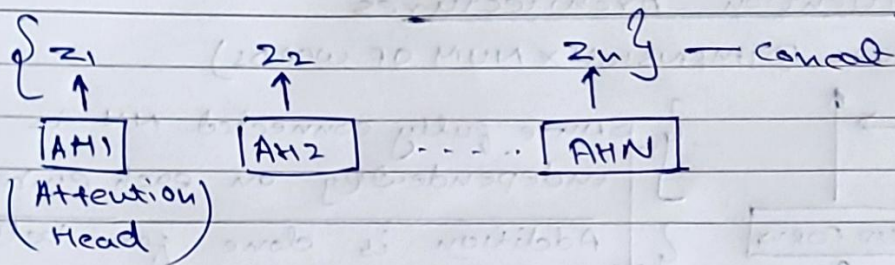


□ (EMBED. DIMENSION × NUM OF WORDS)

ENCODER

ATTENTION HEAD





Decoder:

① Masked Training: A mask of

is added before the softmax application such that only previous and current contents are added.

② Cross Attention

