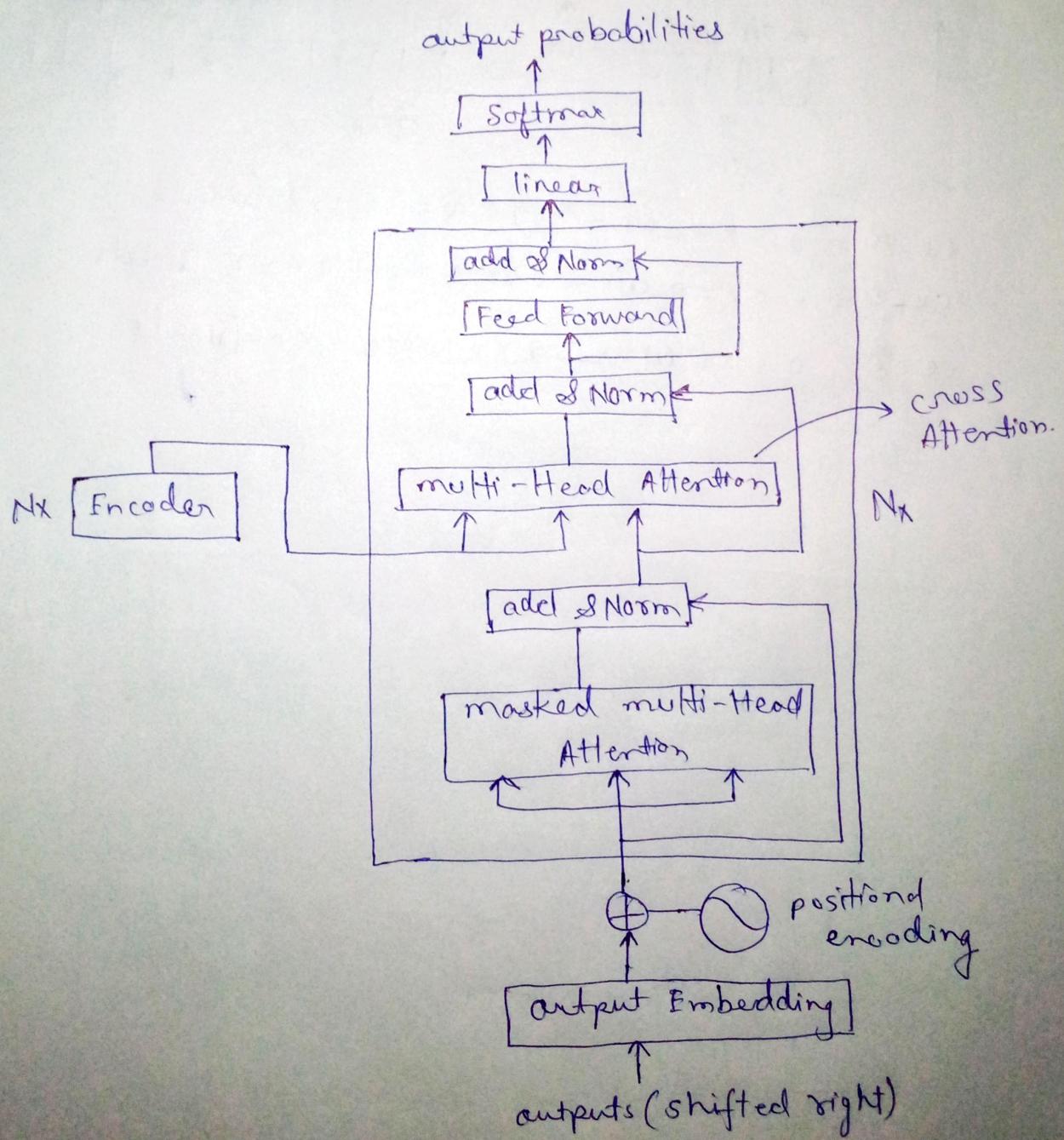


20 Mar 2025

TRANSFORMER'S DECODER ARCHITECTURE

Here the explanation of transformer's decoder architecture in training perspective.

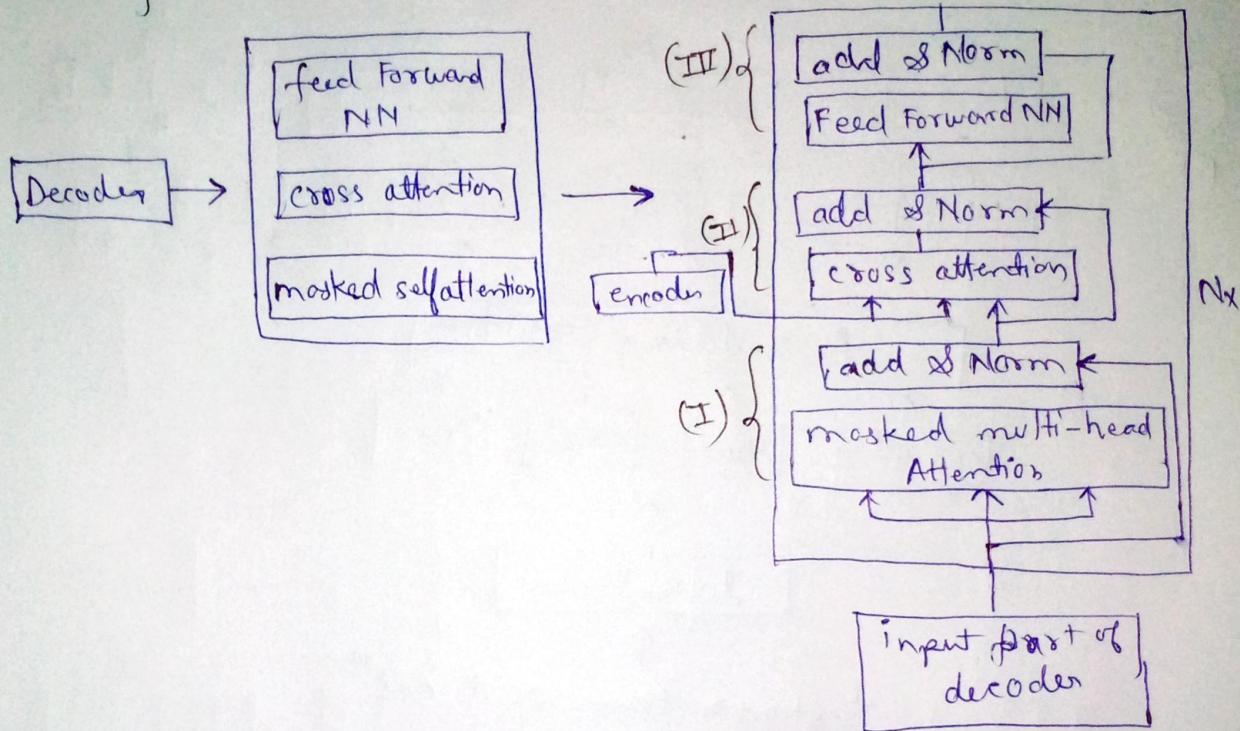
Decoder Architecture:



- same as the encoder. in original Research paper there is 6 encoders and 6 decoders is used.

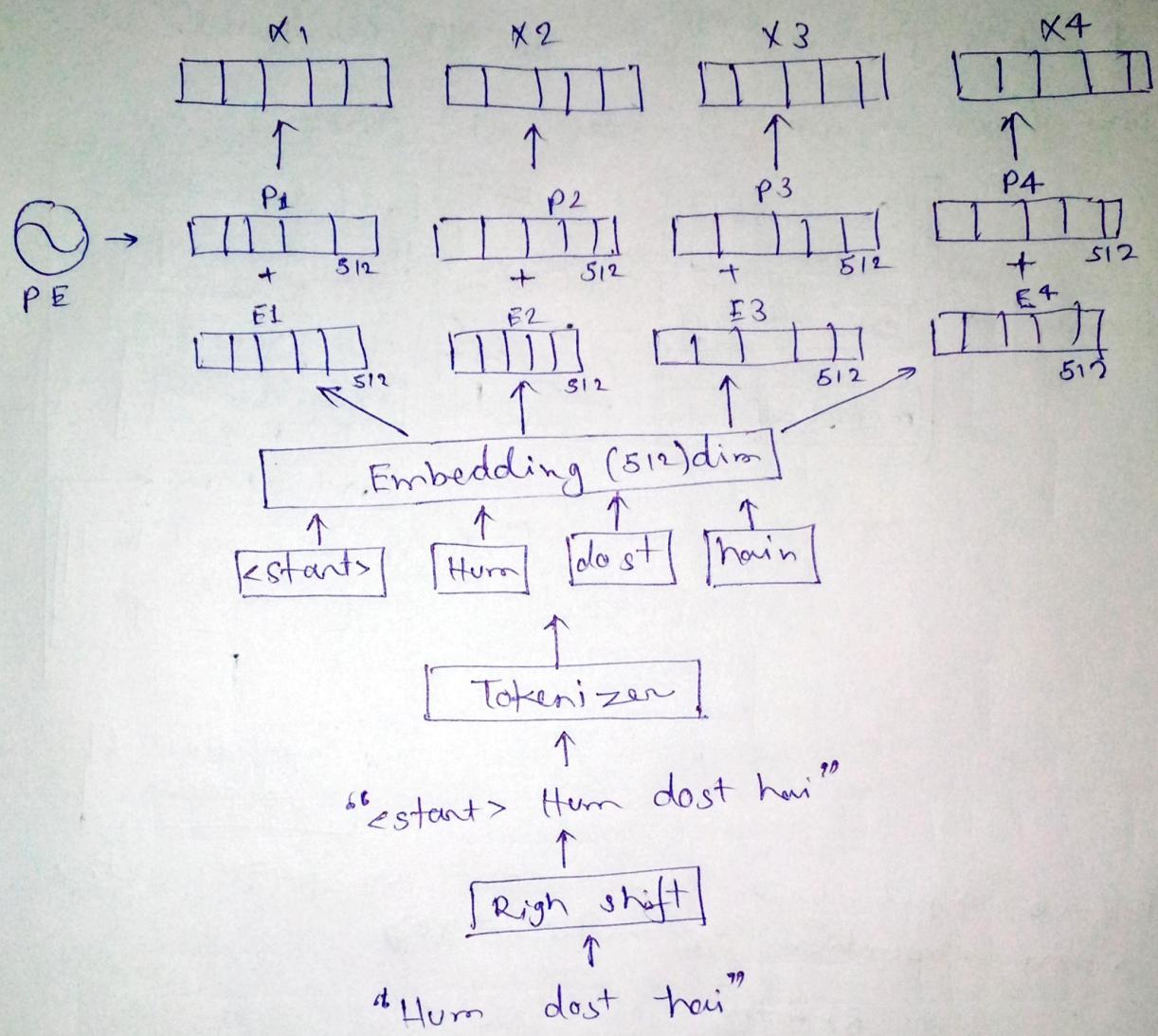
- Research Paper: "Attention Is All You Need".

Now focus on one Decoder Architecture:

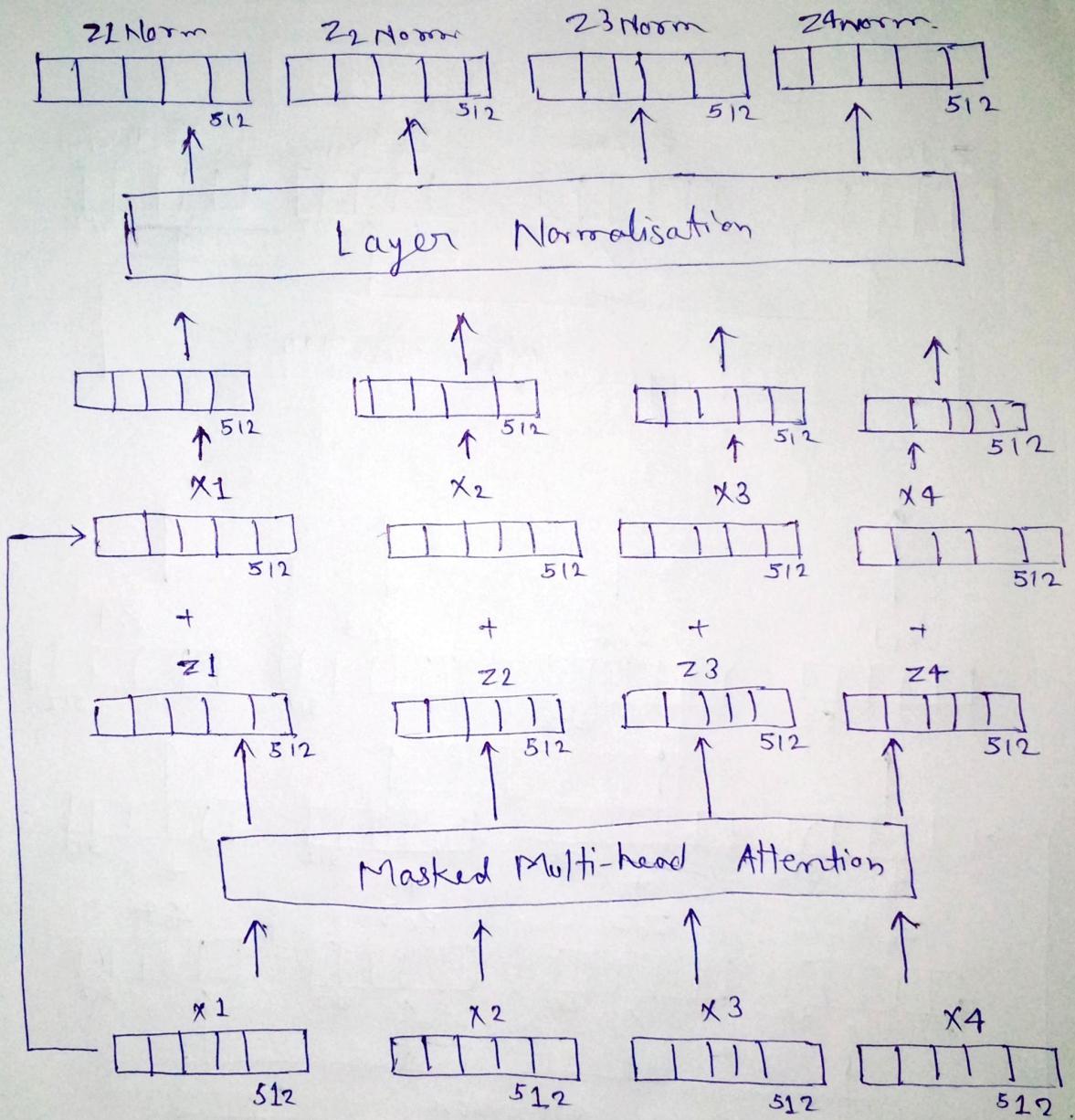


Before input goes into decoder we perform the some operation or architecture

1. Shifting
2. Tokenization
3. Embedding
4. Positional Encoding

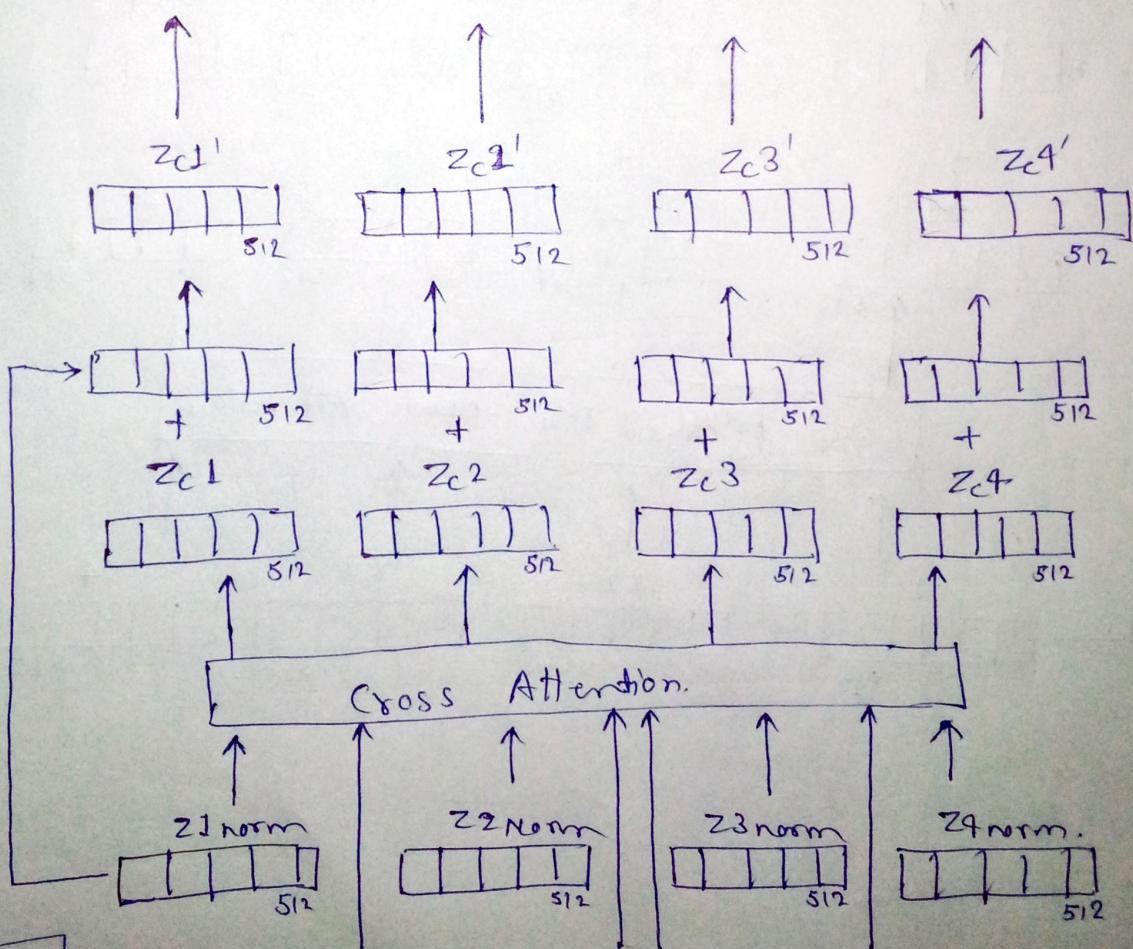
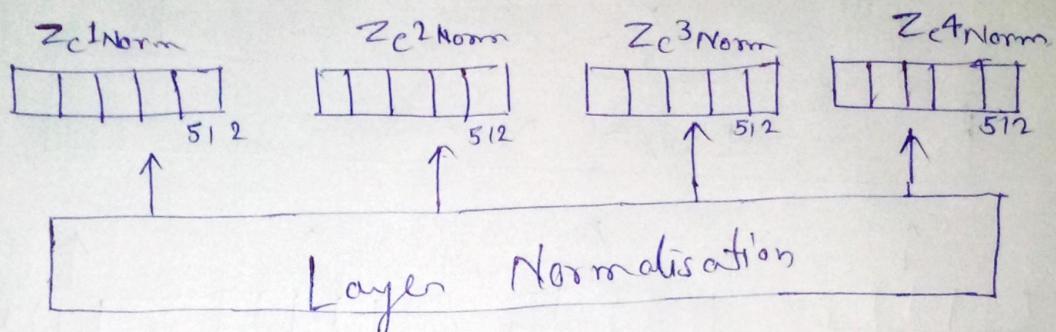


Brief explanation of (I) Architecture:

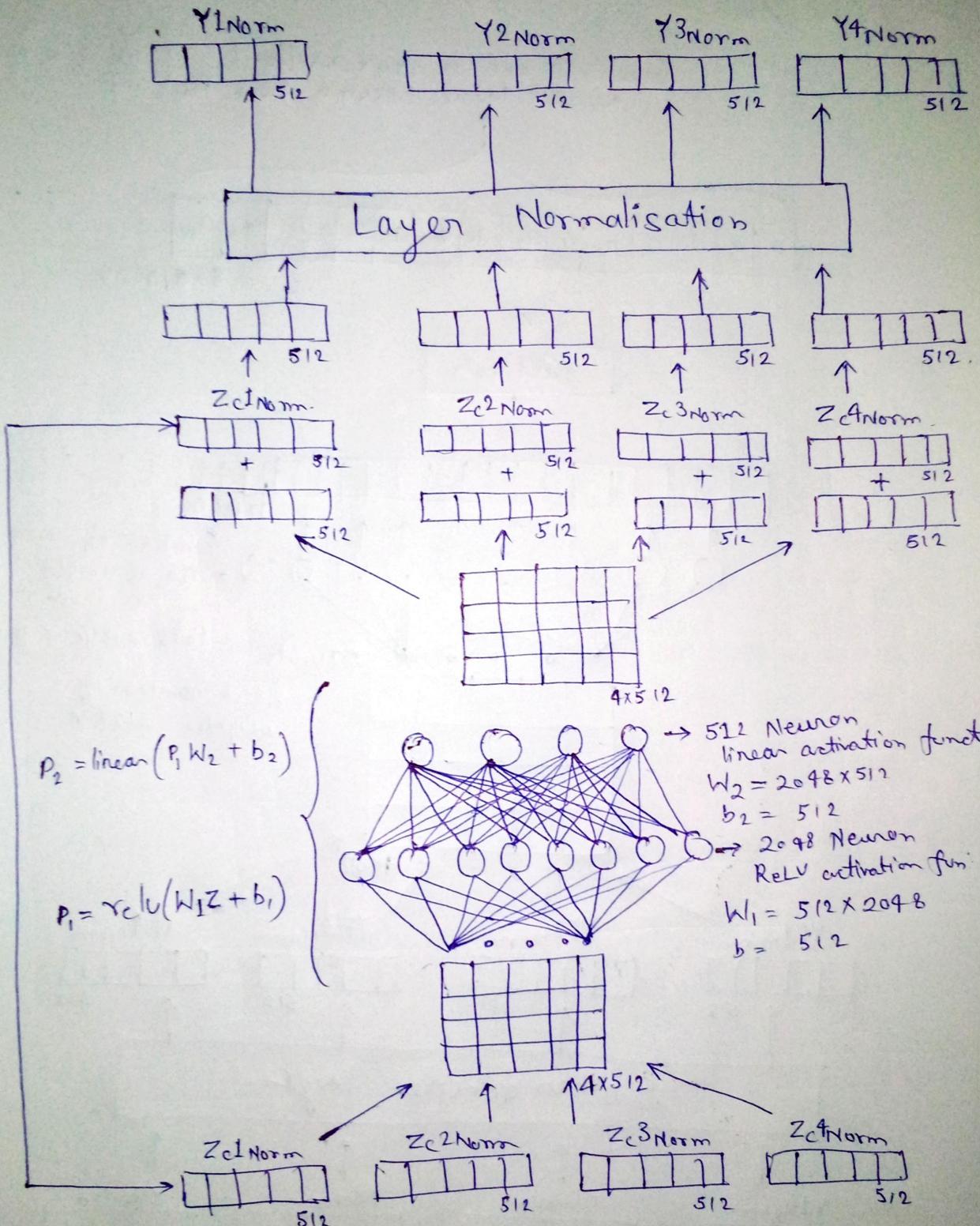


- inputs in cross attention. 2 seq from encoder.
- 1 sequence from decoder.
- 2 seq (key and value) \rightarrow English
- 1 seq (Query) \rightarrow Hindi

Brief explanation of (II) Architecture:



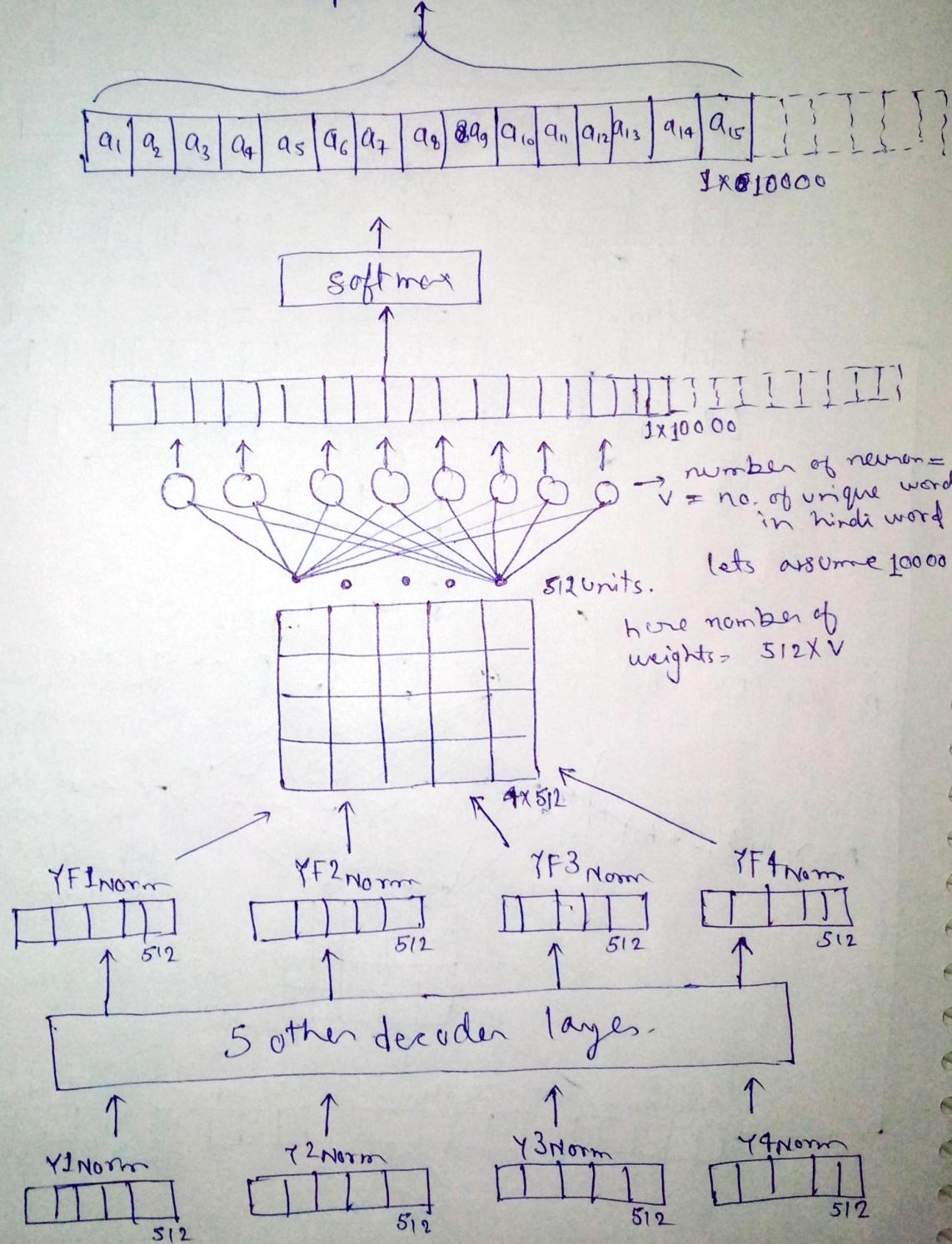
Brief explanation of (III) Architecture:



Brief explanation of output part of the Decoder.

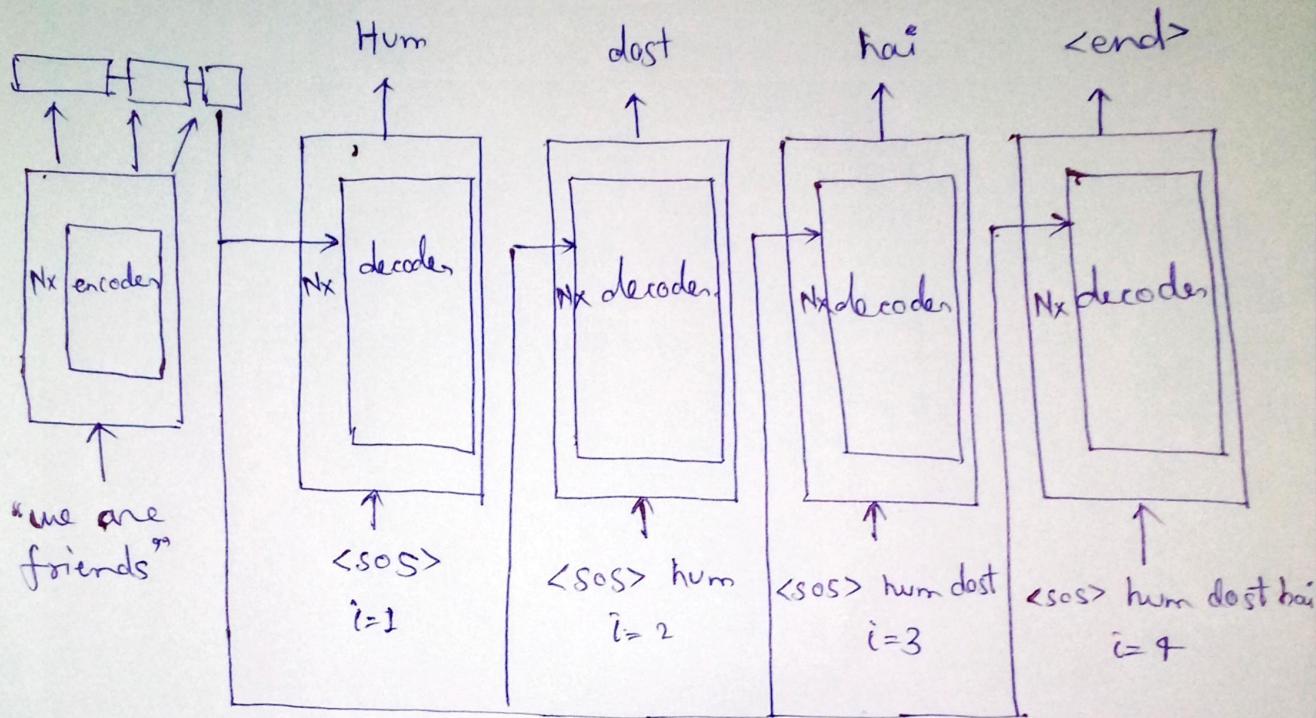
Architecture :-

Greatest one is our resultant output let us assume here $a_4 \left(\sum_{i=1}^{10k} a_i = 1 \right)$



(B)

Here the explanation of transformer's decoder architecture in inference perspective.



Here $i = \text{number of iteration}$

$Nx = 6$ (according to research paper), may be vary