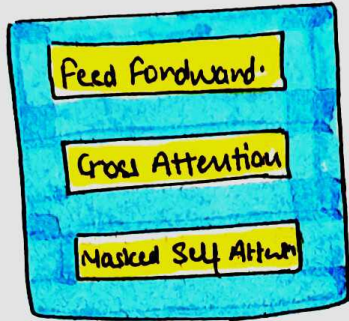
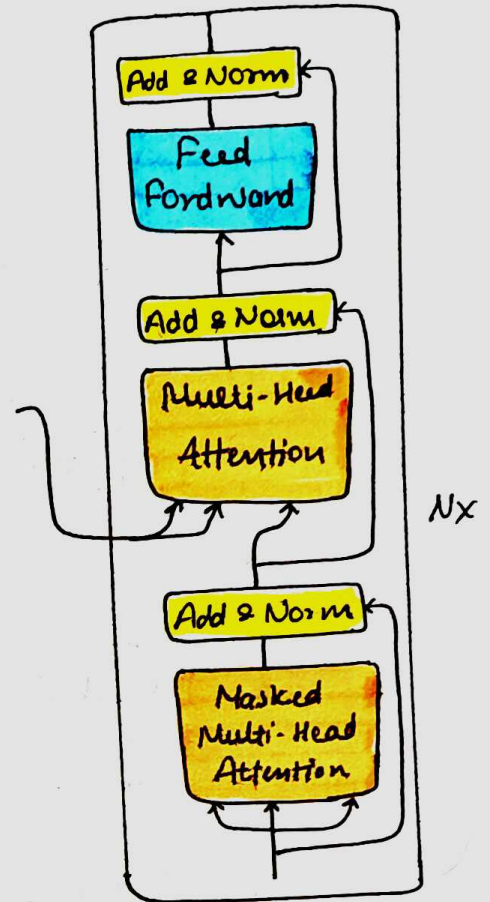
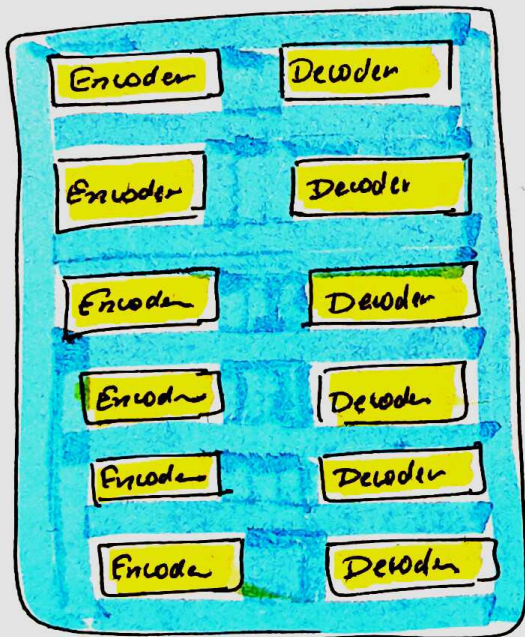


# Transformer Decoder Architecture

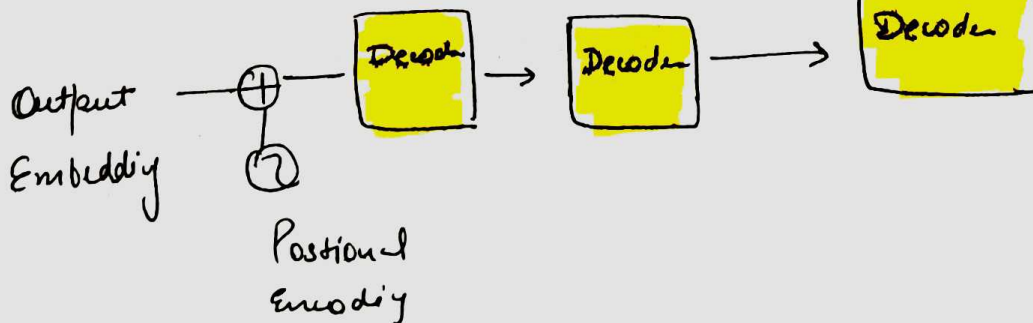
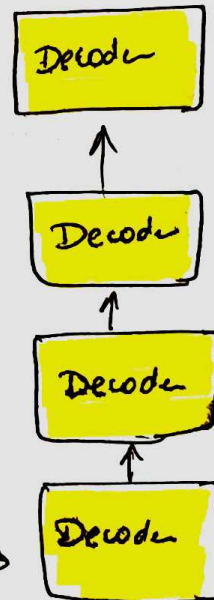
Decoder

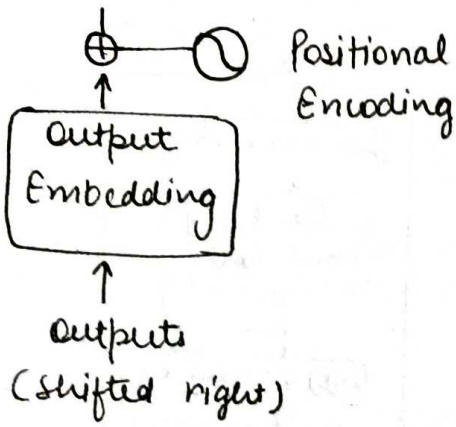


In Transformer Research Paper  
Six decoder use.

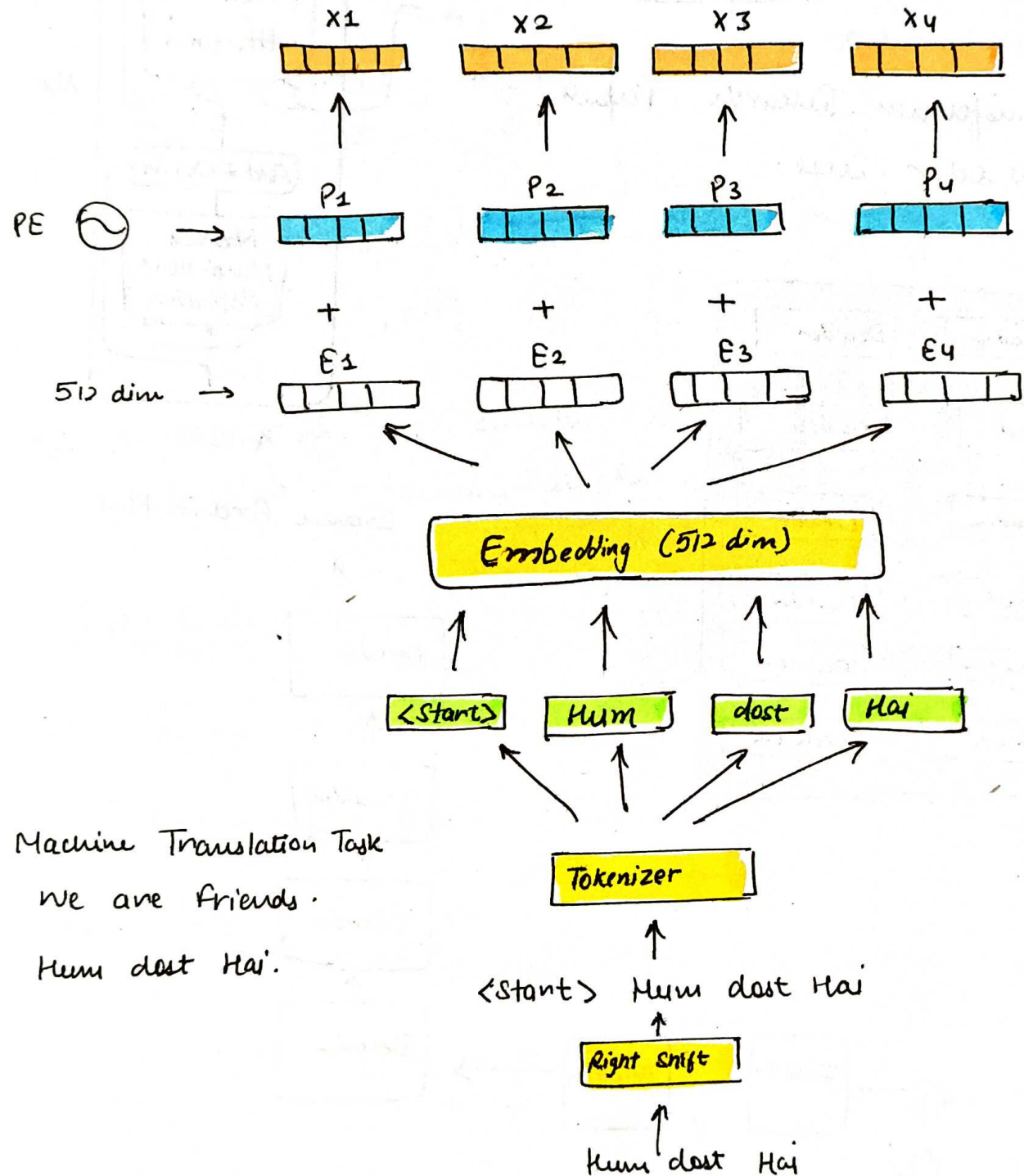


Same Architecture

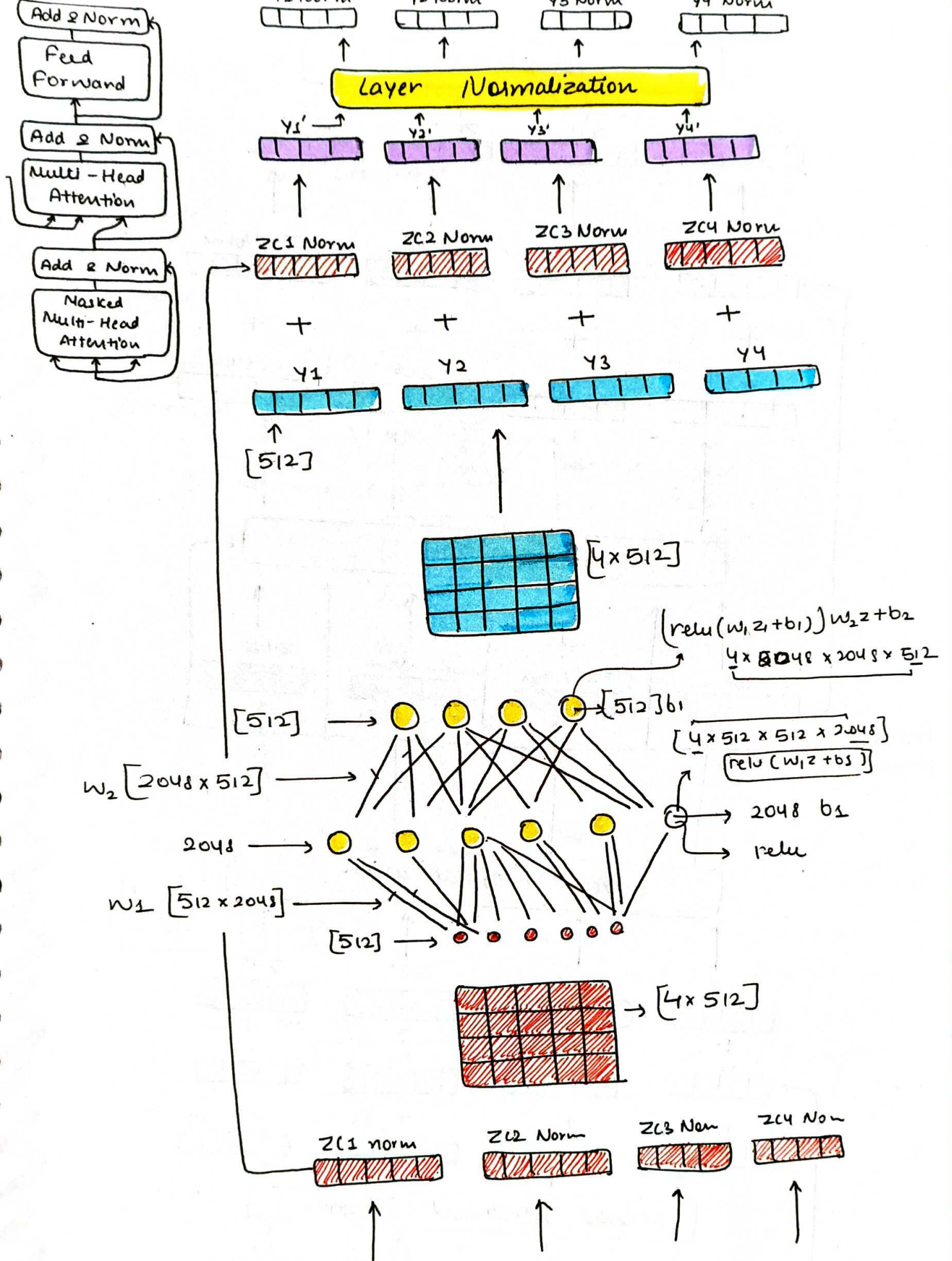




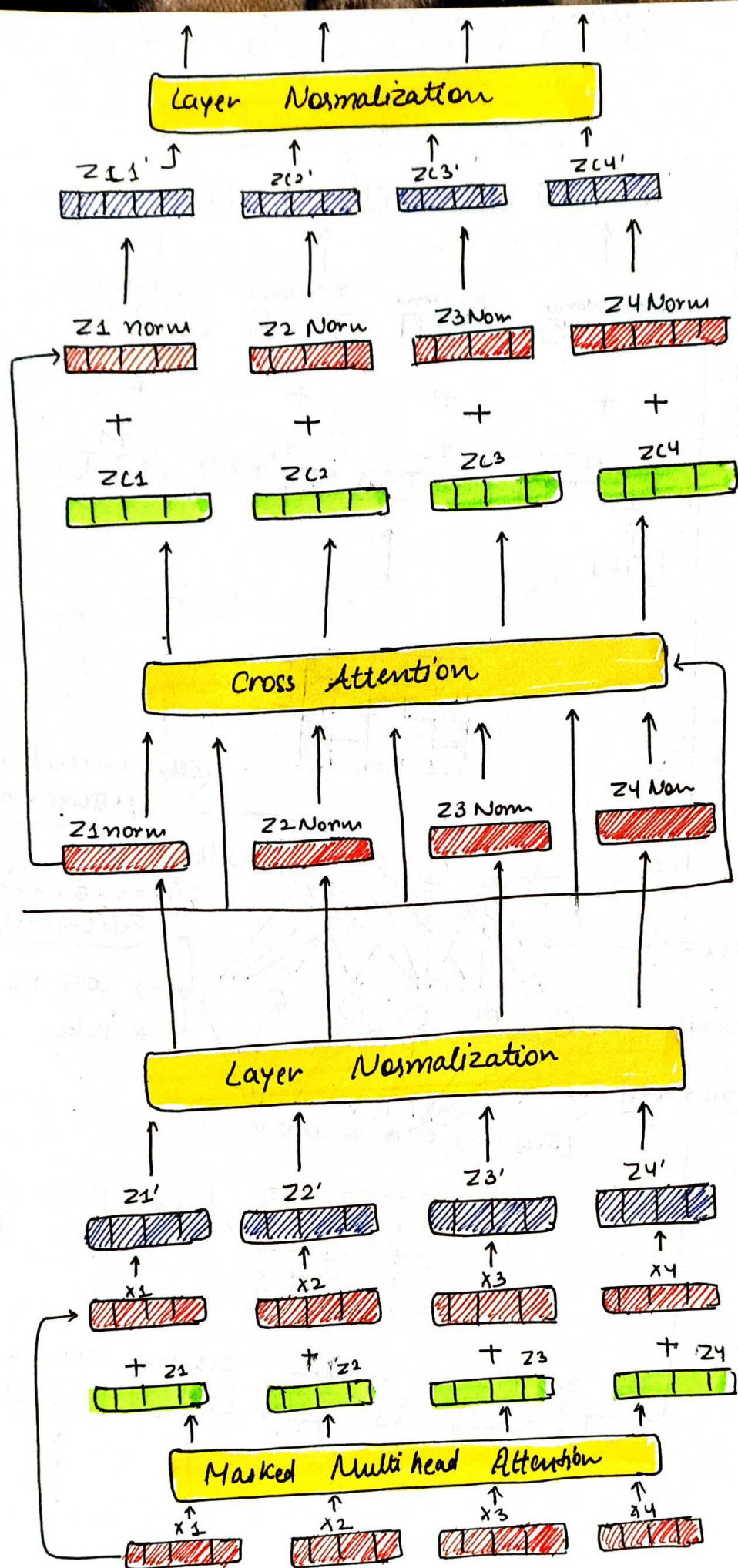
1. Shifting
2. Tokenization
3. Embedding
4. Positional Encoding







From Encoder





Output Probabilities

Eng

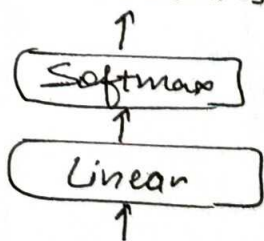
Hindi

I am good

Mai badhiya hu

we are friend

Hum dost hai



No. of unique words

represent neuron

10,000

512 x 10,000

512 dim

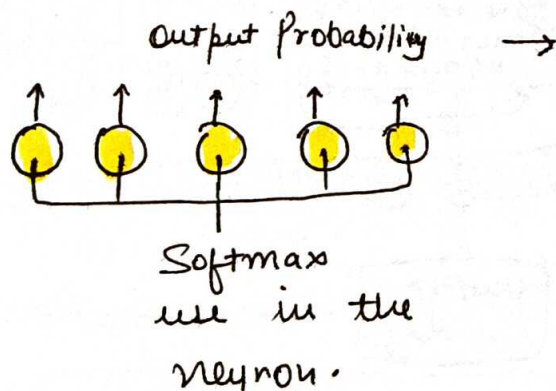
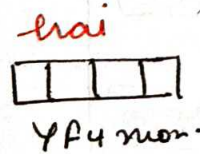
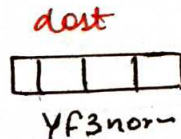
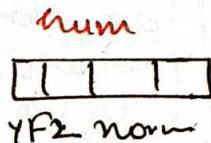
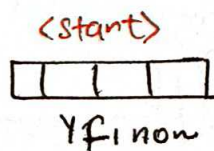
Mai badhiya hu hum

1 x 512 512 x 10,000

No. of neurons  $\rightarrow [V]$

Softmax

V is a predicted words. In Hindi sentence, how many unique words. and make vocab.



Mai  $\rightarrow 0.2$

badhiya  $\rightarrow 0.1$

⋮

hum  $\rightarrow 0.5$   $\rightarrow$  Probab is greater.

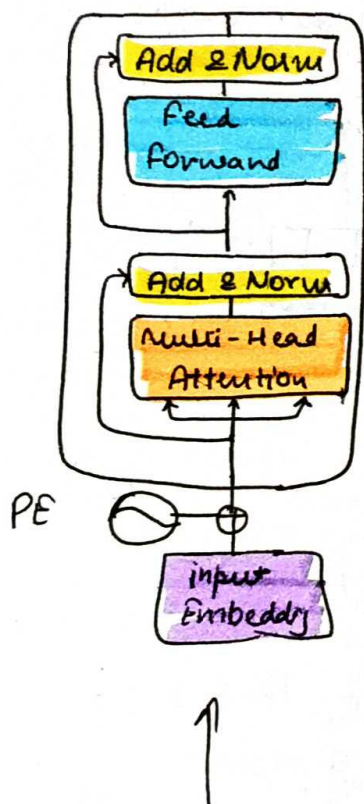
⋮

First word after <start> is "hum".

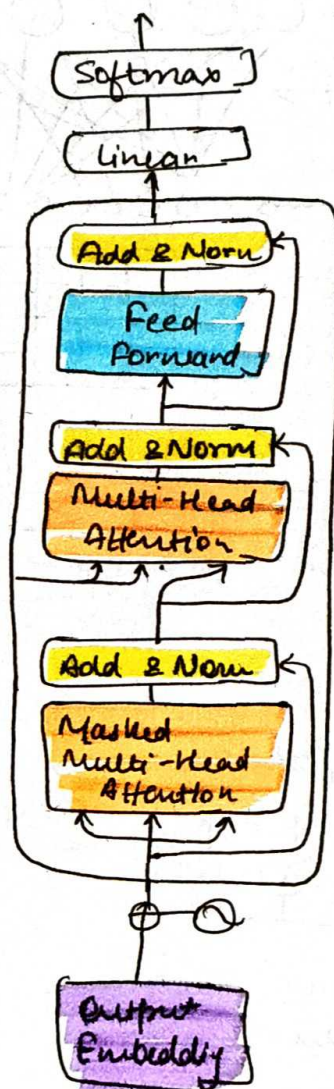
# Transformer during Inference

Training Time  $\rightarrow$  Encoder  $\rightarrow$  Same  
Decoder  $\rightarrow$  NAR

Testing Time  $\rightarrow$  Encoder  $\rightarrow$  Same  
Decoder  $\rightarrow$  AR

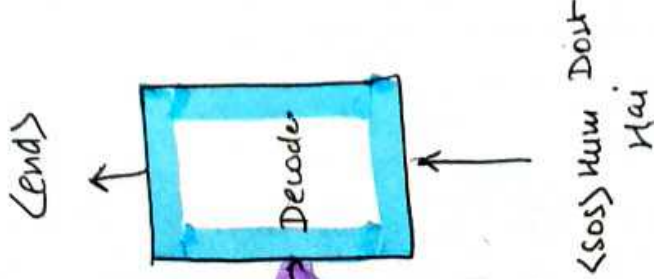
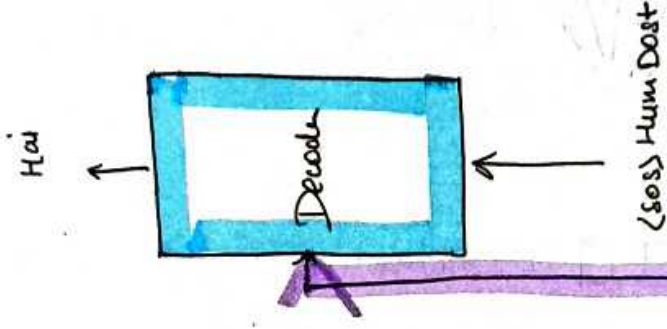
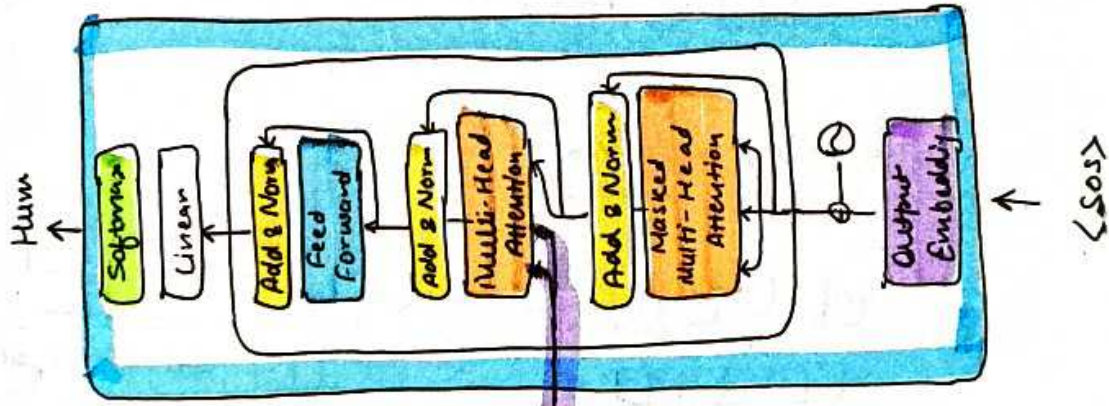
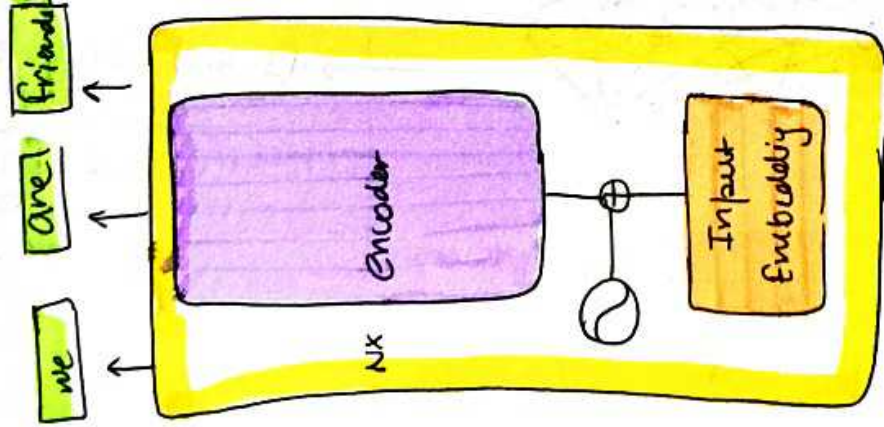


Encoding Architecture



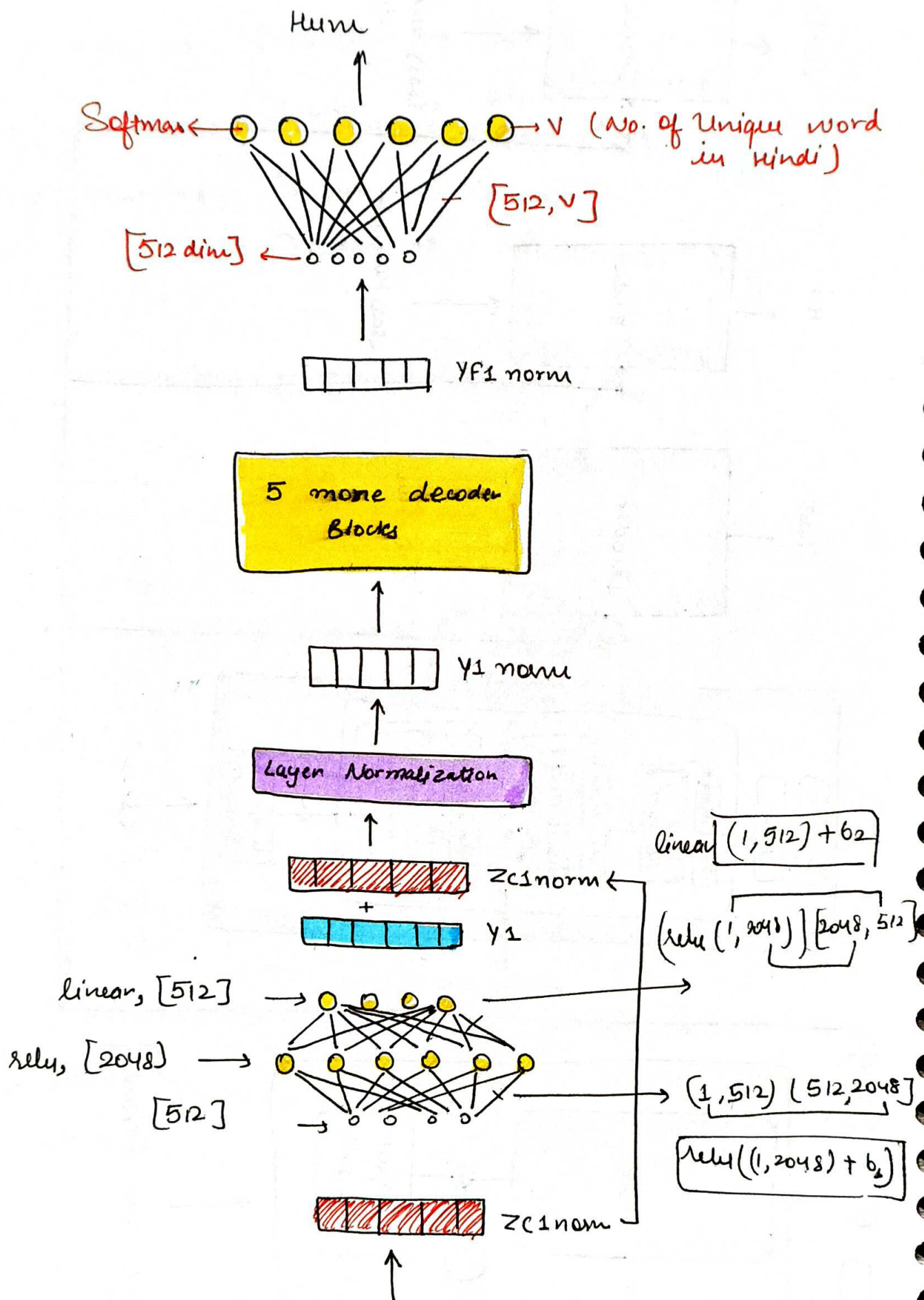
Decoder Architecture



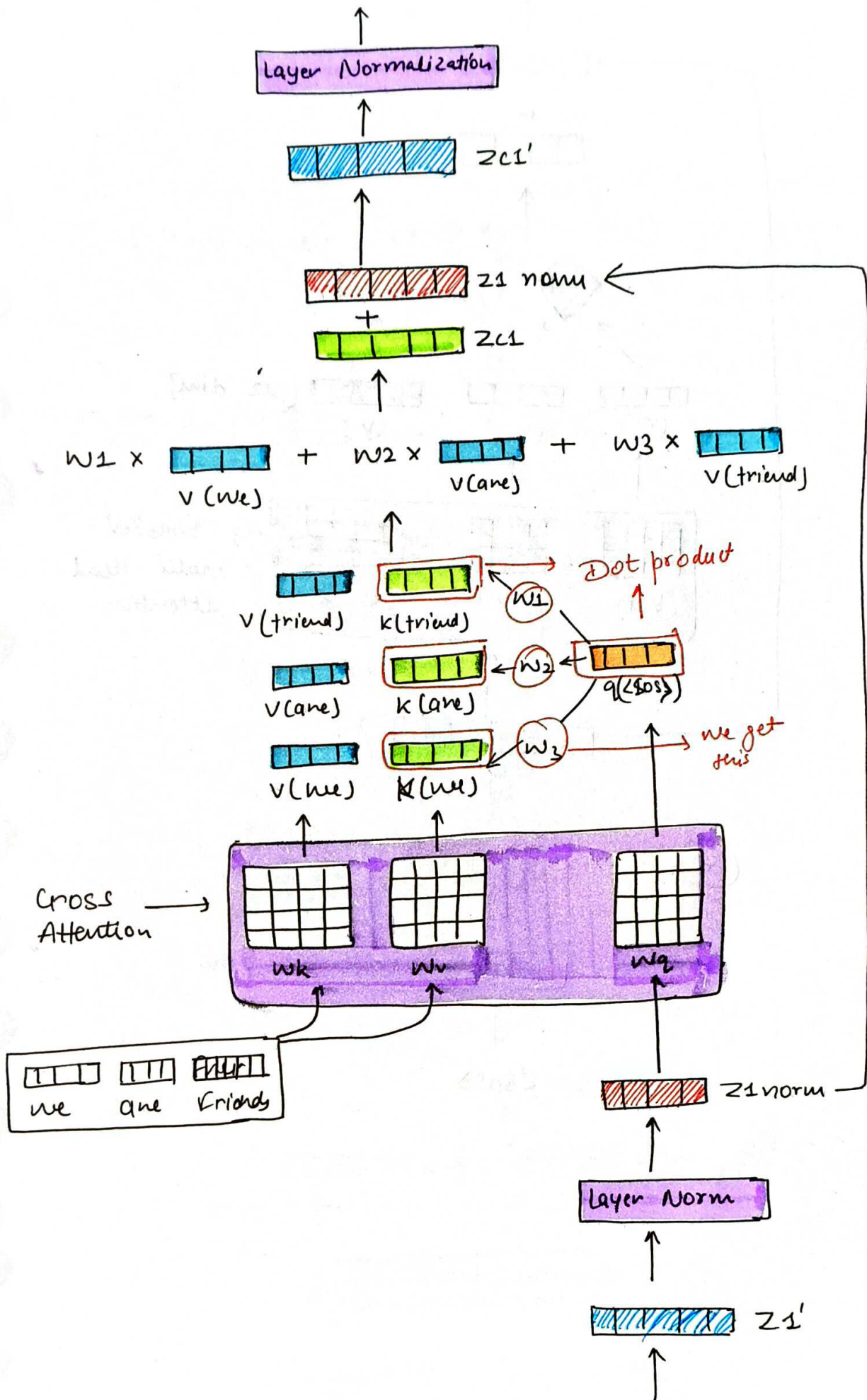


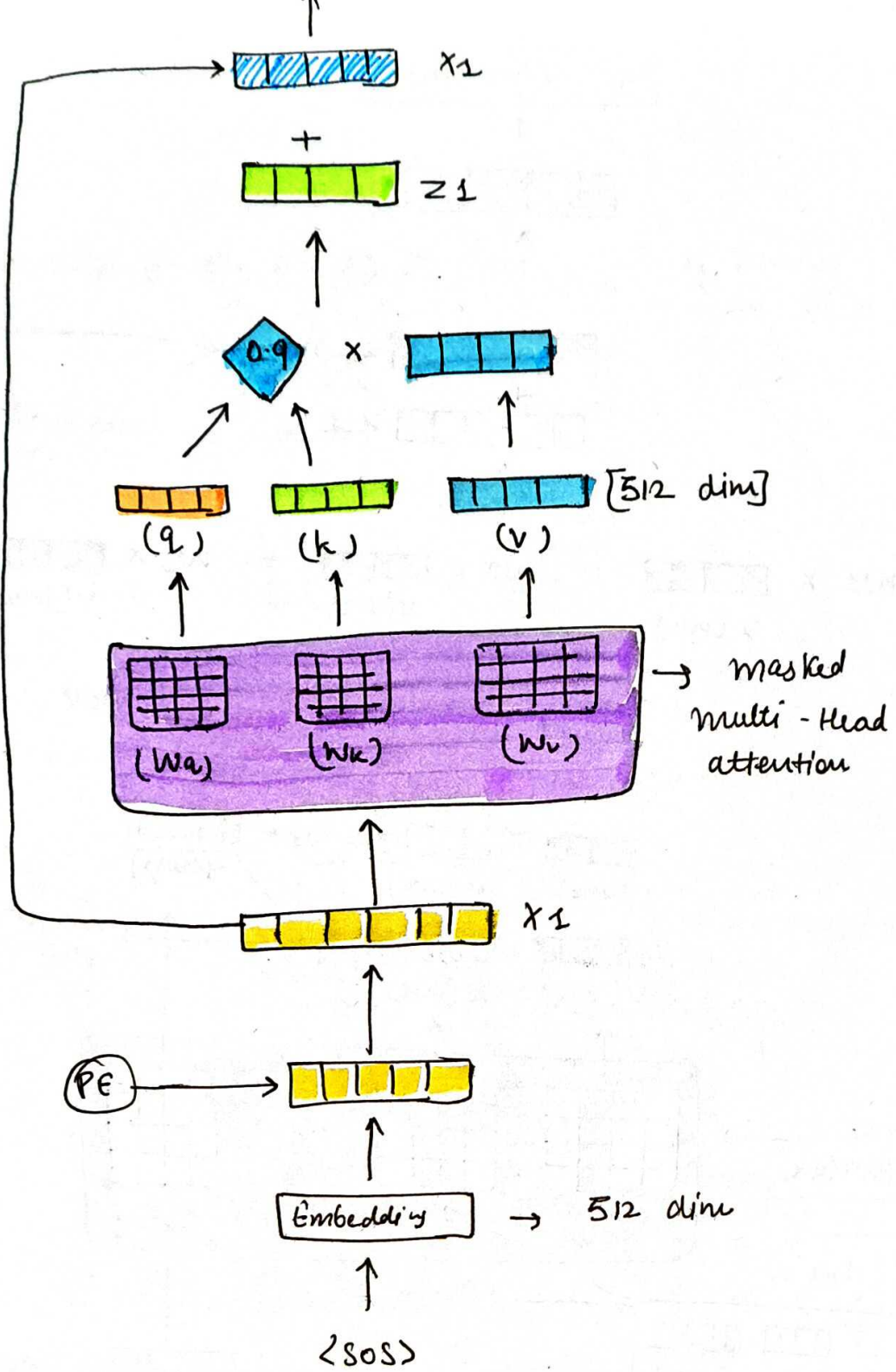
[We are friends]

# Detailed Architecture of first decoder using <ss>



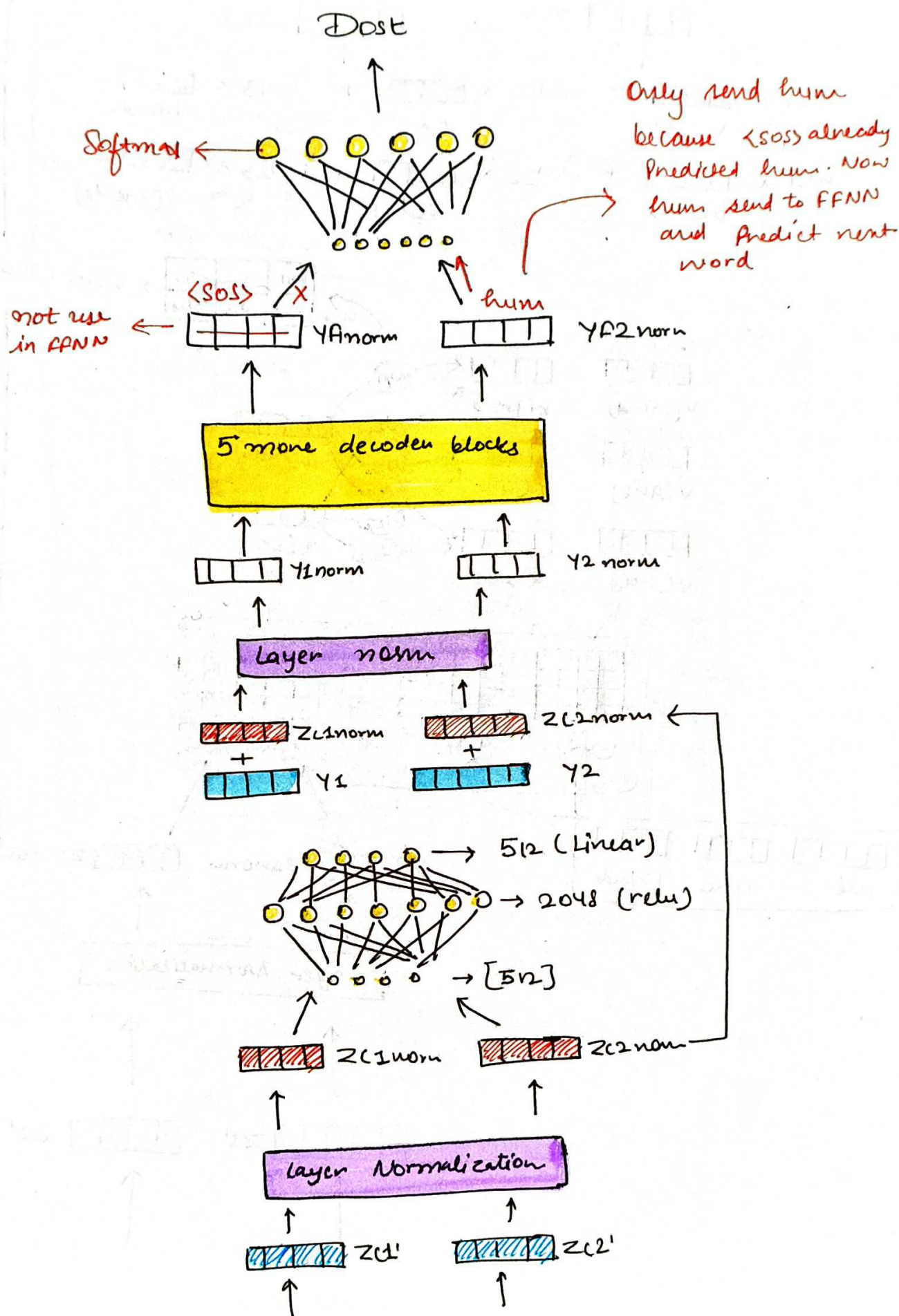


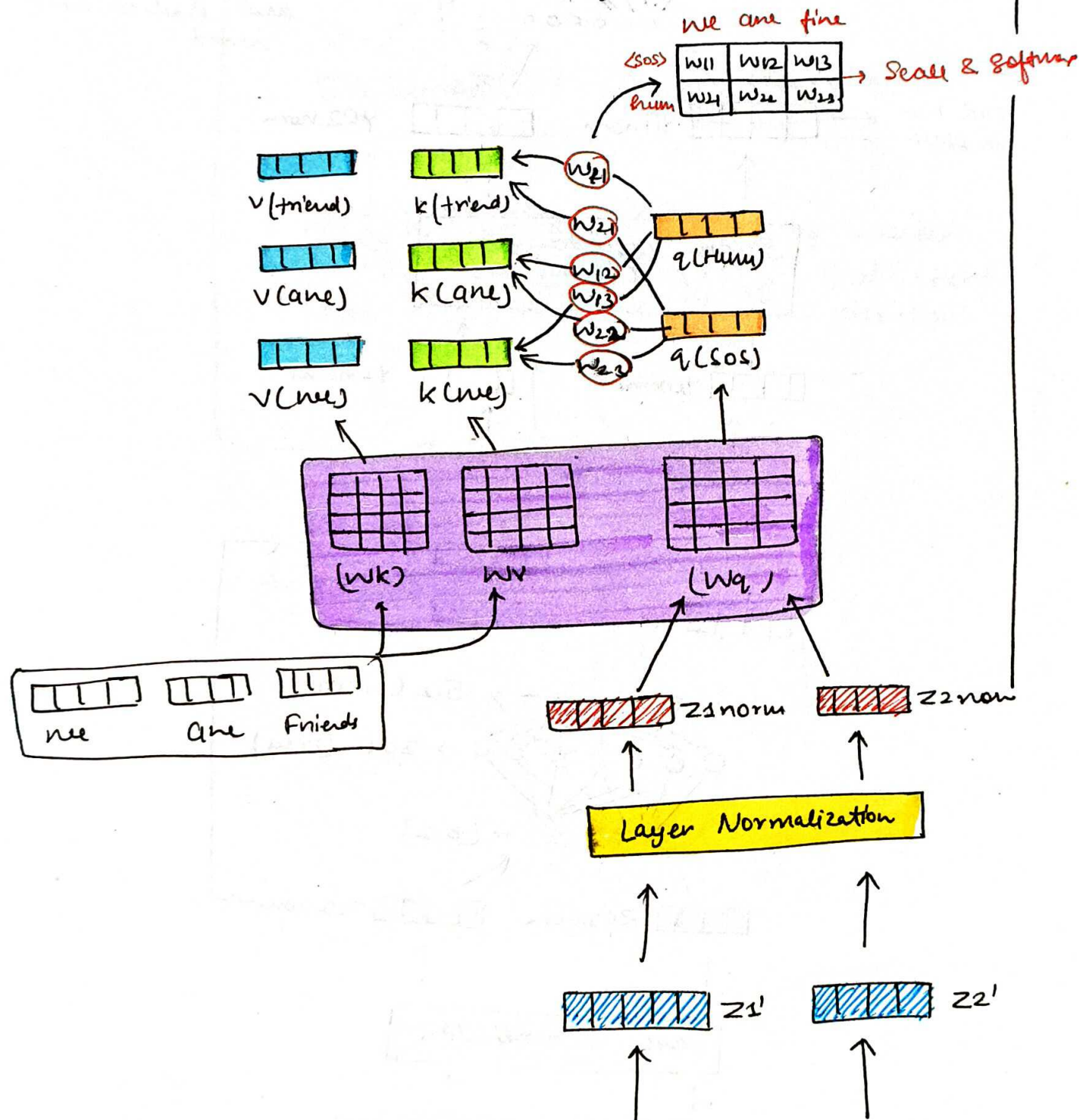
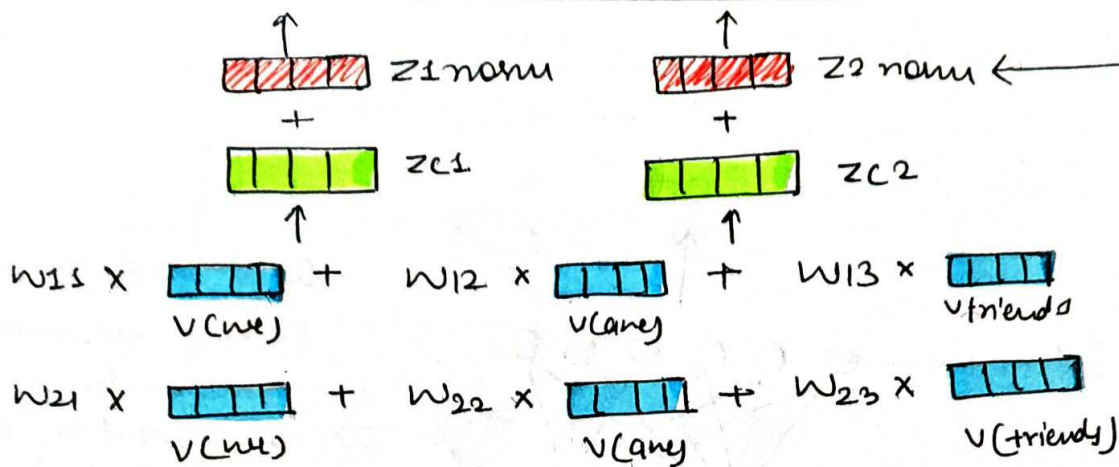




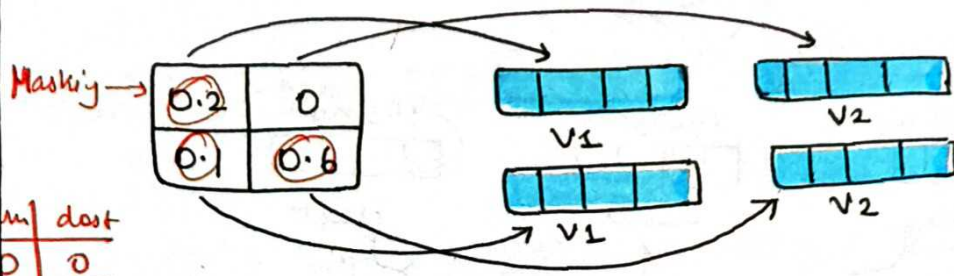
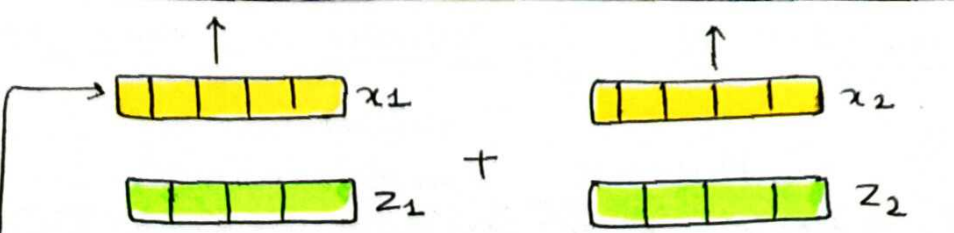


# Detailed Architecture of Second Decoder using <eos> hum







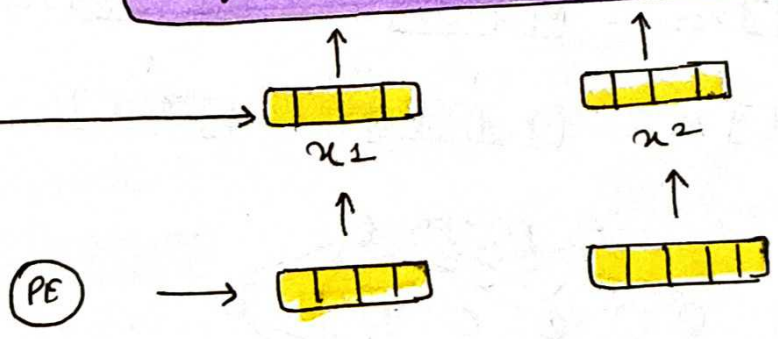
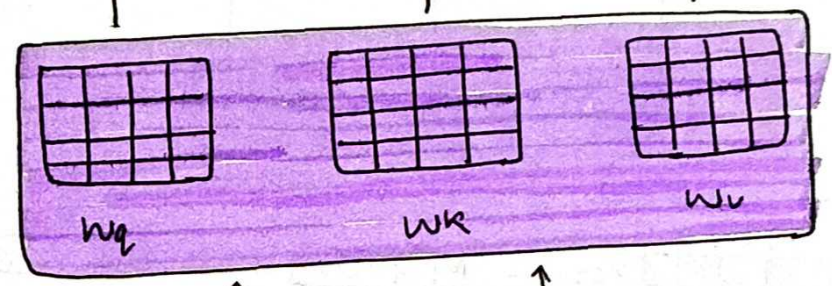
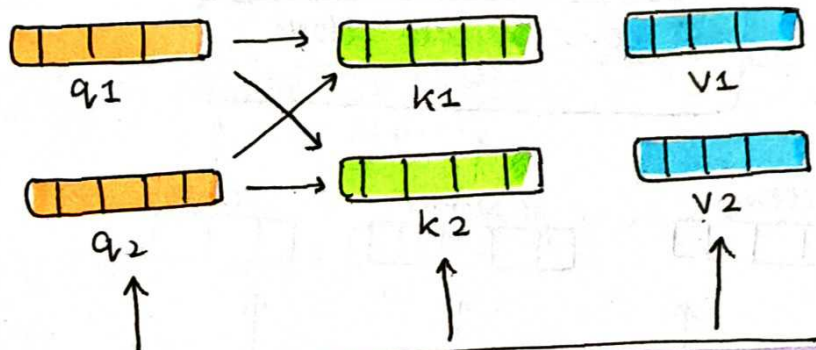


[Masking]

	$\langle \text{SOS} \rangle$	Hum
$\langle \text{SOS} \rangle$	X	X
Hum	X	X

	$\langle \text{SOS} \rangle$	Hum	dest
$\text{SOS}$	X	0	0
Hum	X	X	0
dest	X	X	X

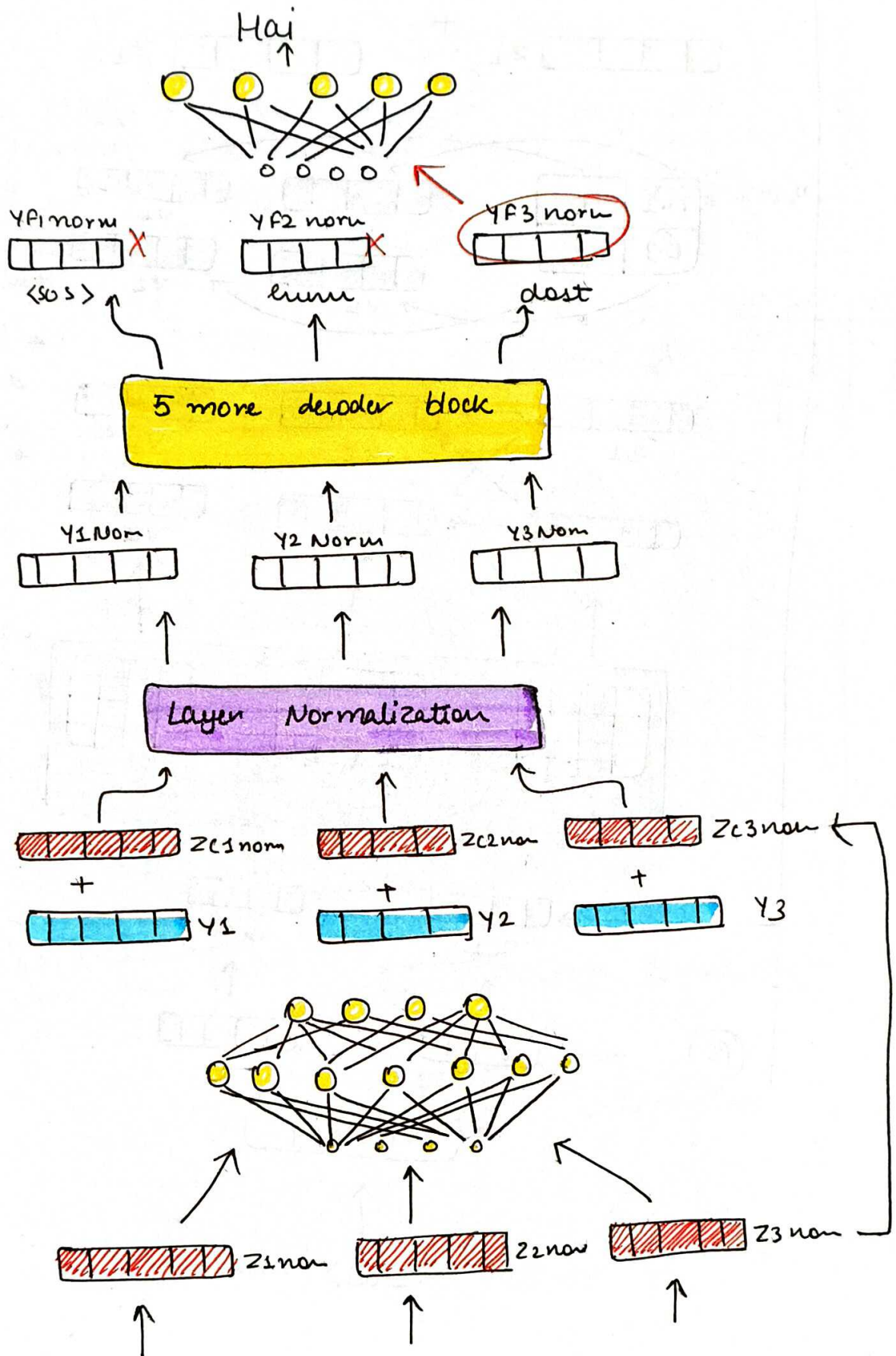
\* In  $\langle \text{SOS} \rangle$ , we don't know hum.  
 \* Hum know  $\langle \text{SOS} \rangle$  and hum.



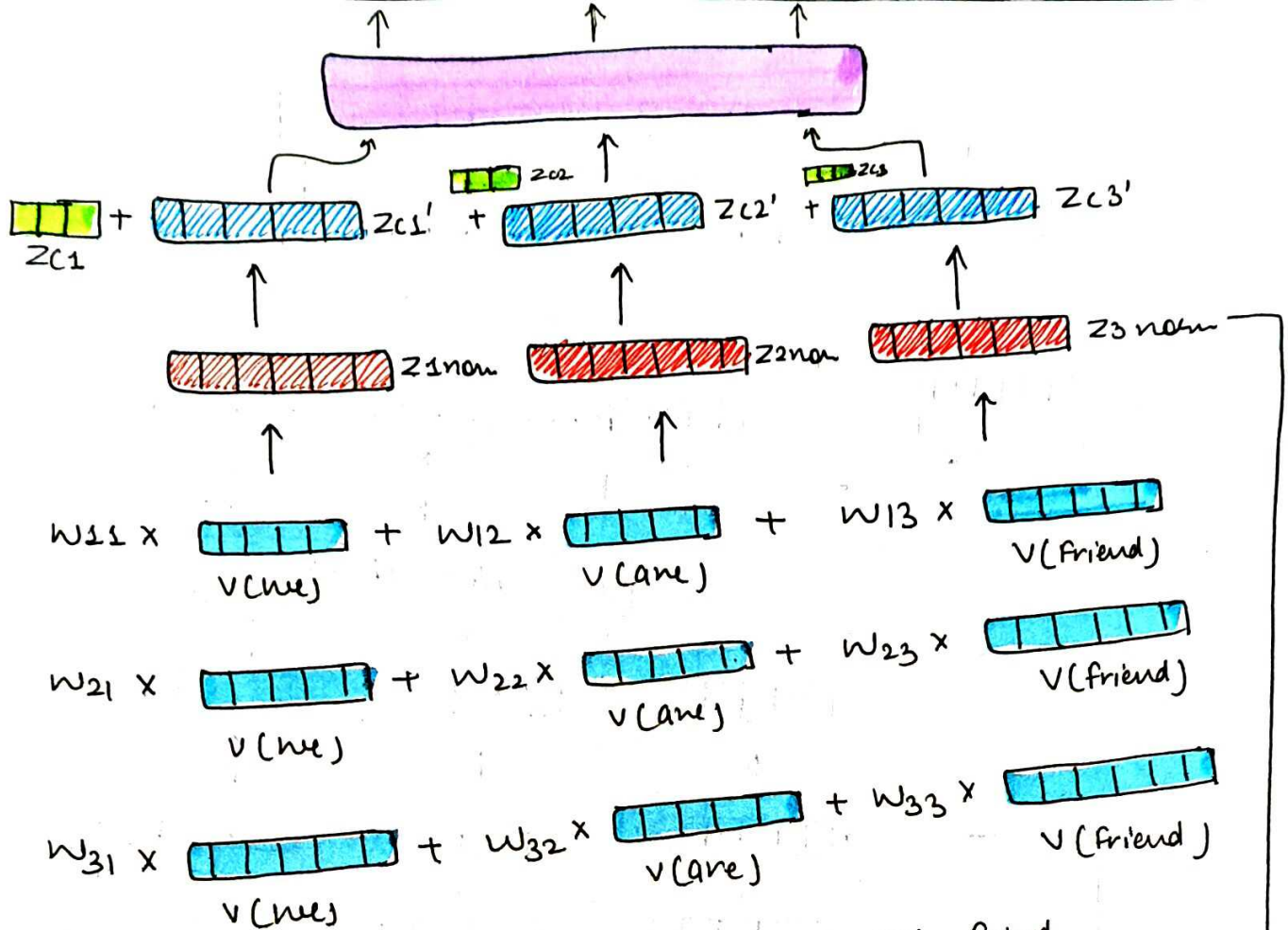
Embedding

$\langle \text{SOS} \rangle$  Hum

# Detailed Architecture of third decoder (sos) from das







wee   ane   friend

$\langle \text{sos} \rangle$	$w_{11}$	$w_{12}$	$w_{13}$
hum	$w_{21}$	$w_{22}$	$w_{23}$
dest	$w_{31}$	$w_{32}$	$w_{33}$

