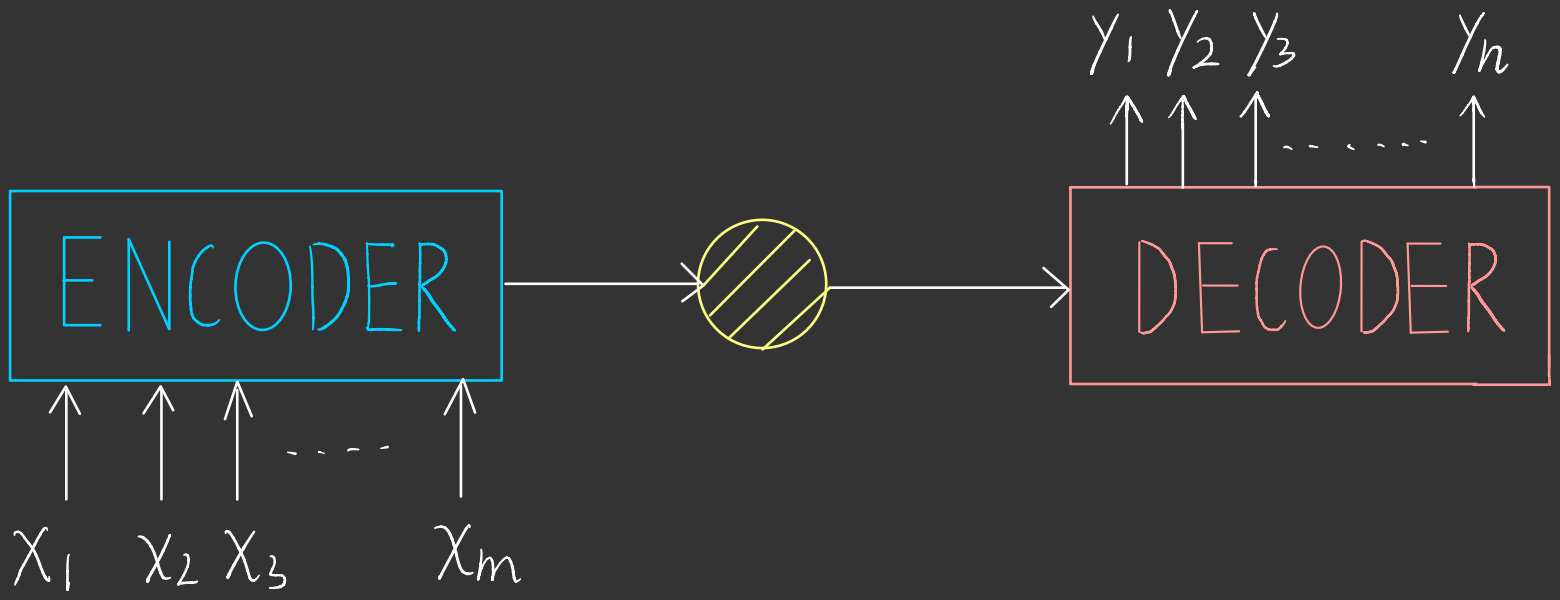
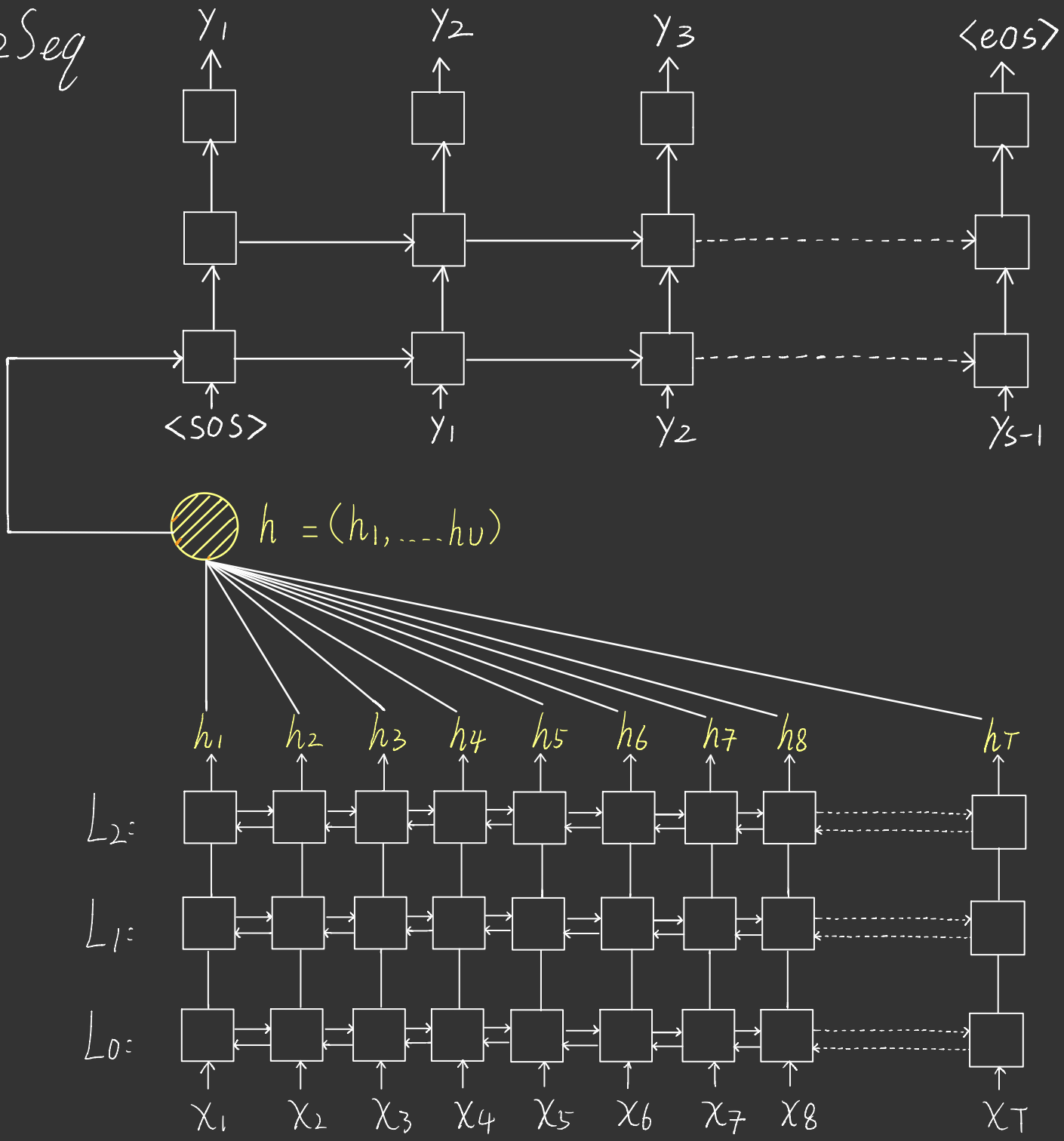
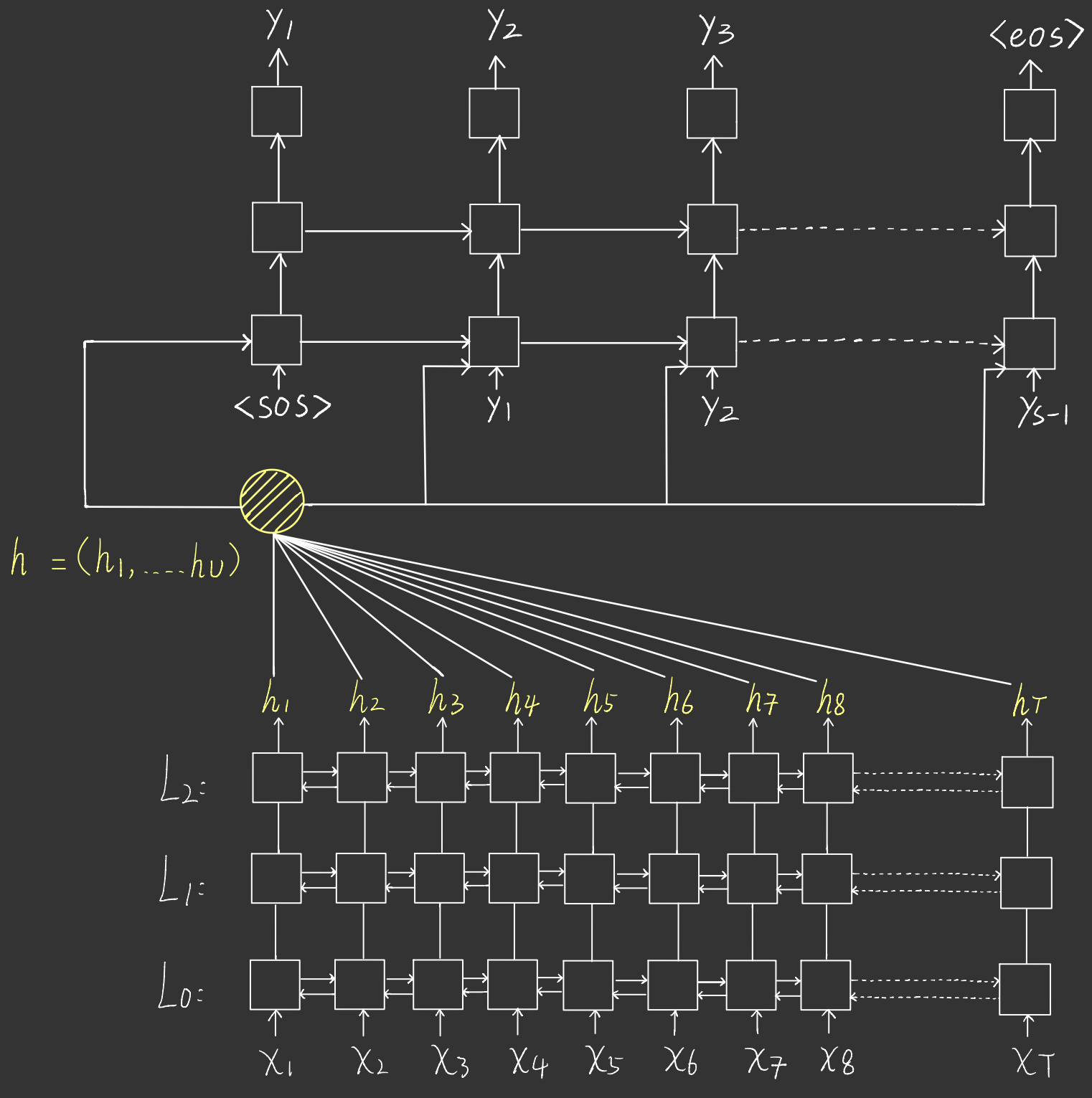
Seq2Seq

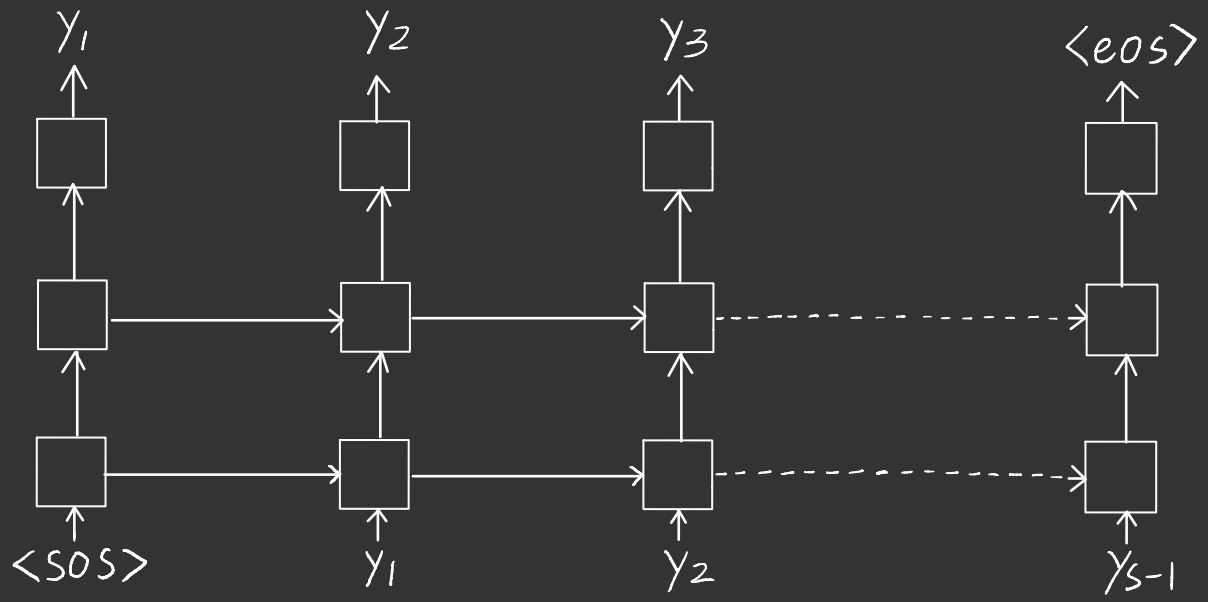


Vanilla Seq2Seq




y1
y2
y3
<eos>
<sos>
y1
y2
ys-1
h = (h1, ... hU)
h1
h2
h3
h4
h5
h6
h7
h8
hT
L2:
L1:
L0:
x1
x2
x3
x4
x5
x6
x7
x8
xT

Attention

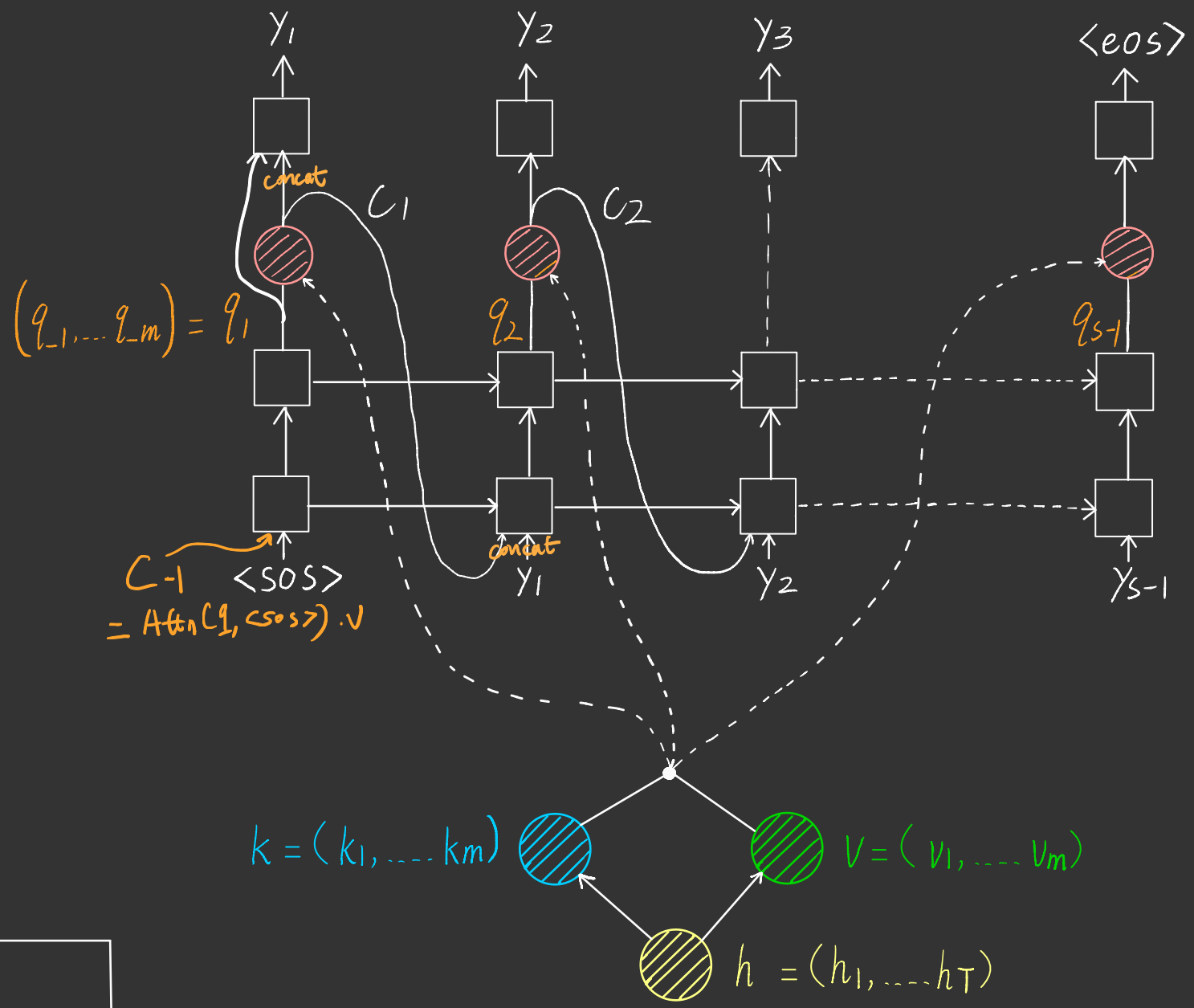


$h = (h_1, \ldots h_T)$

h_1	h_2	h_3	h_4	h_5	.....	h_T

weight: α_1 α_2 α_3 α_4 α_5 ..... α_T

Eg. context $= \sum_i \alpha_i h_i$



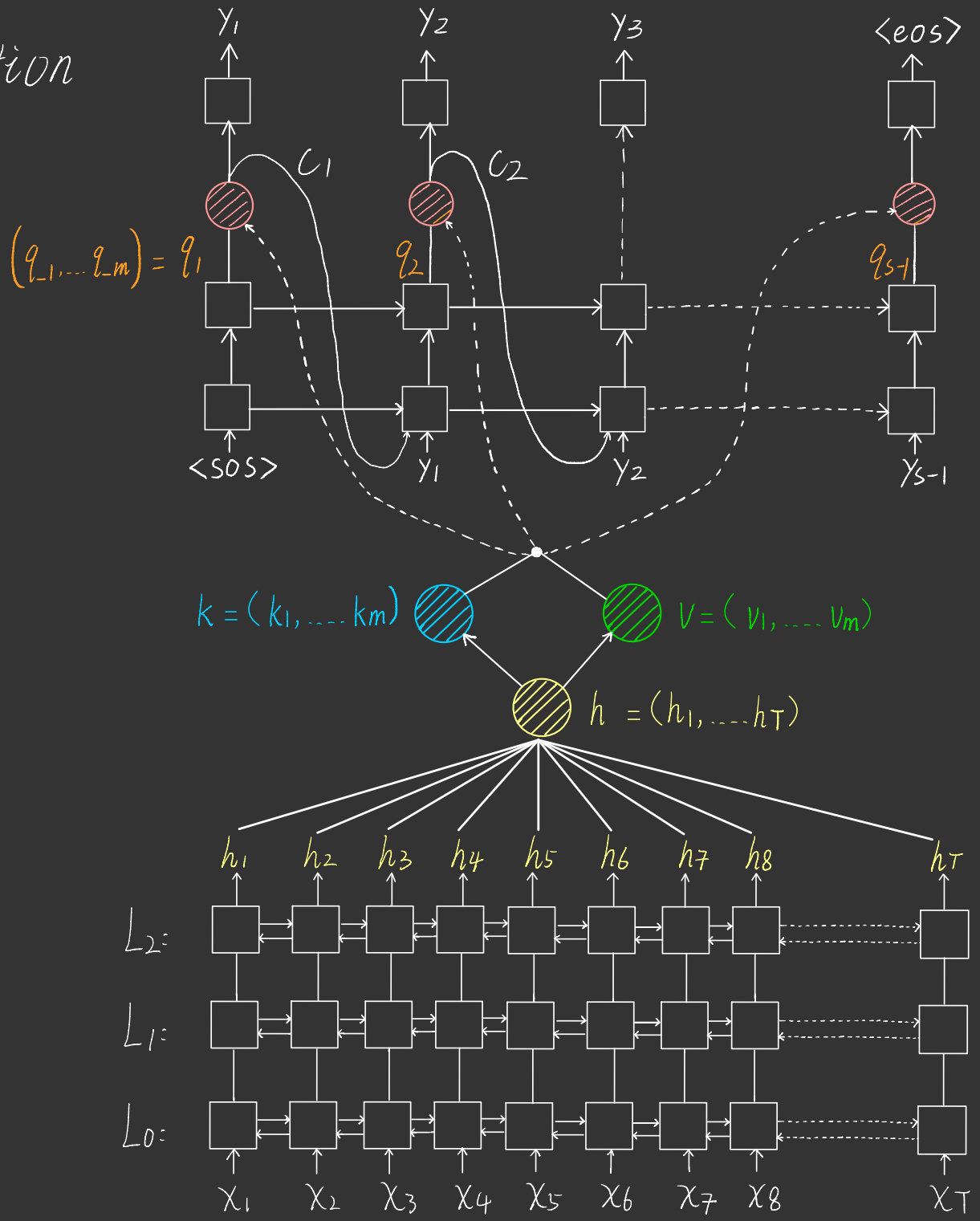
attention module

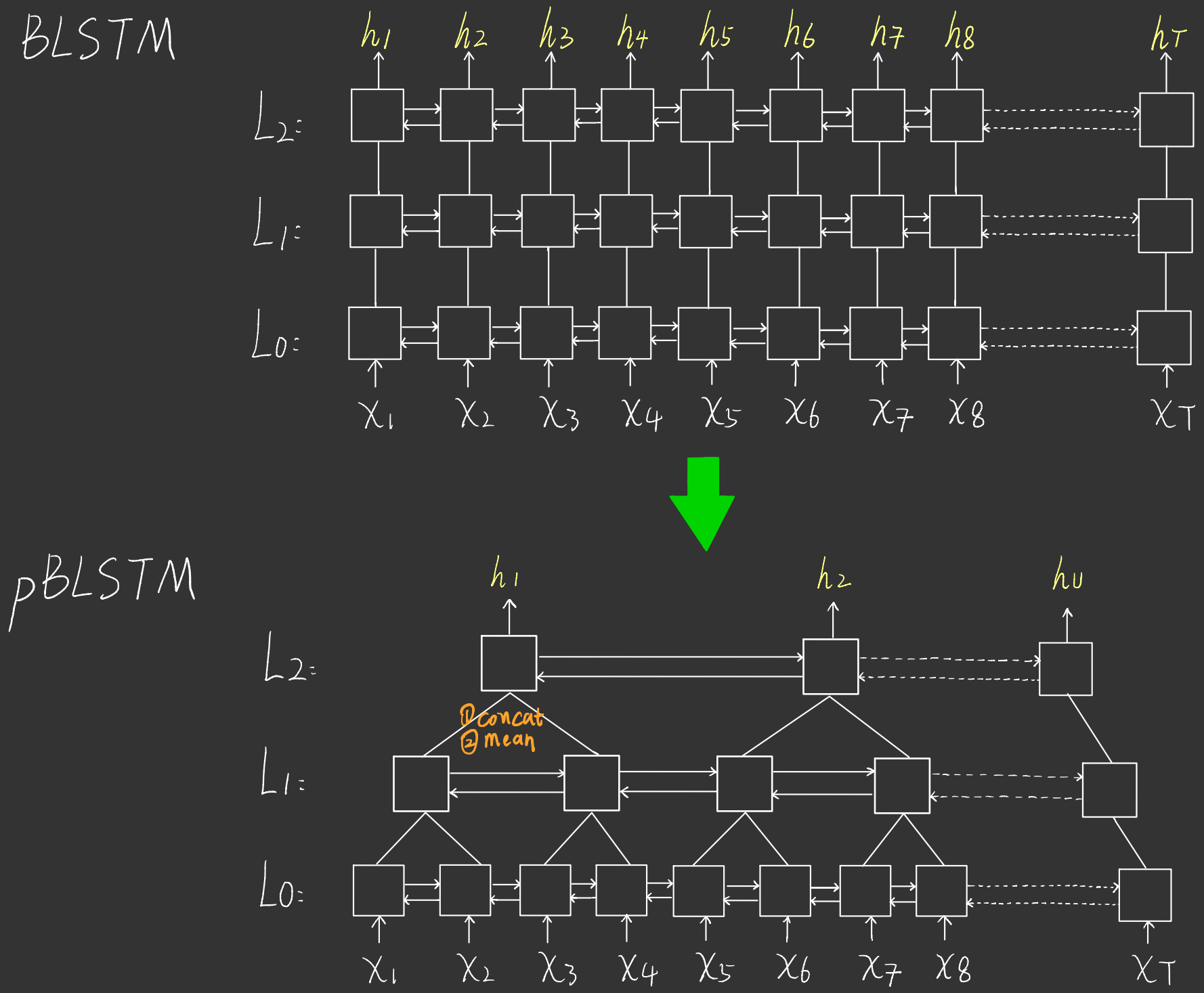
② $c = \sum cos\_simi(q, k) \cdot v$

scaled dot-product

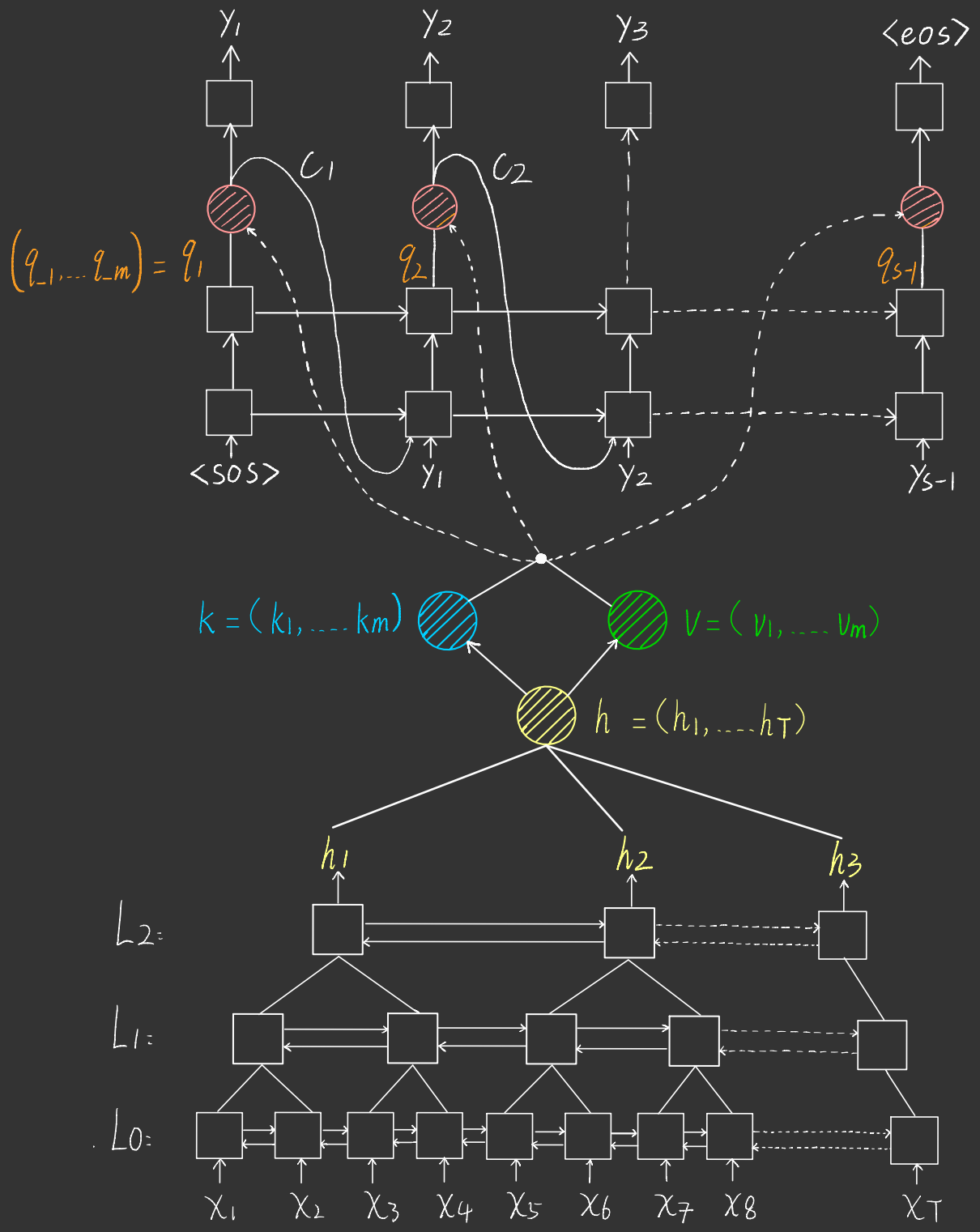
① $\sum Softmax\left(\frac{q k^T}{\sqrt{d_k}}\right) \cdot V$

Seq2Seq + Attention

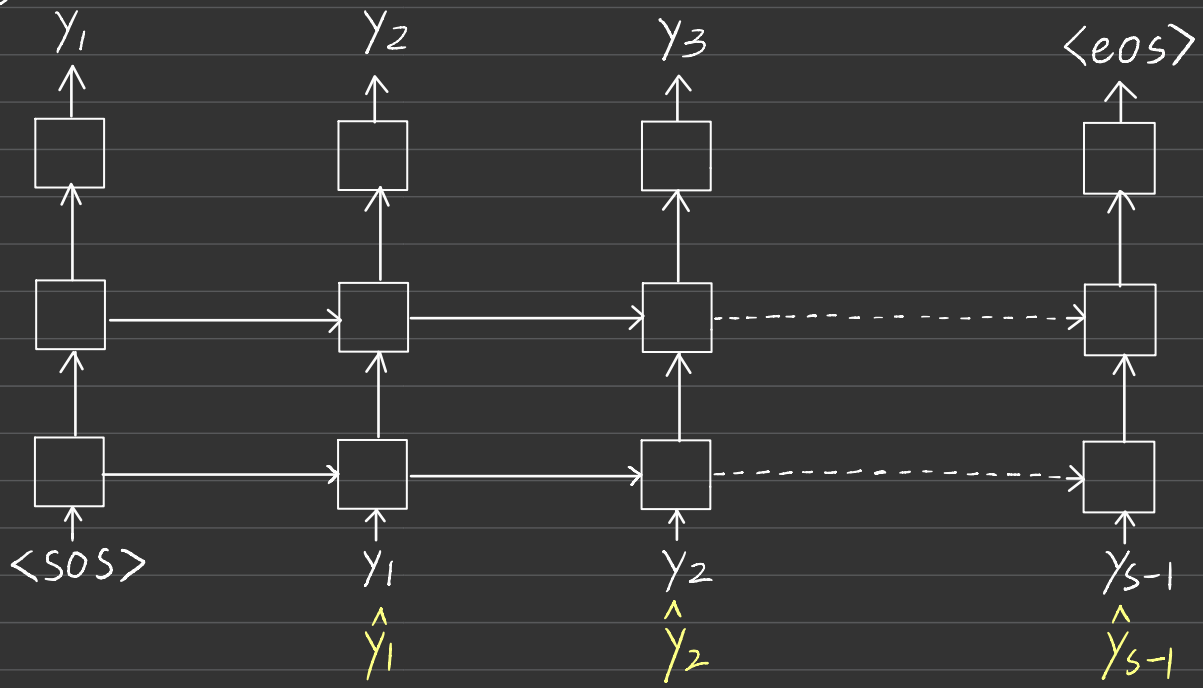



BLSTM
h1 h2 h3 h4 h5 h6 h7 h8 hT
L2:
L1:
L0:
x1 x2 x3 x4 x5 x6 x7 x8 xT
pBLSTM
h1 h2 hU
L2:
L1:
L0:
① concat
② mean
x1 x2 x3 x4 x5 x6 x7 x8 xT

LAS


y1
y2
y3
<eos>
c1
c2
(q_1,... q_m) = q1
q2
qs-1
<sos>
y1
y2
ys-1
k = (k1, ... km)
V = (v1, ... vm)
h = (h1, ... hT)
h1
h2
h3
L2:
L1:
L0:
x1
x2
x3
x4
x5
x6
x7
x8
xT

Teacher forcing:

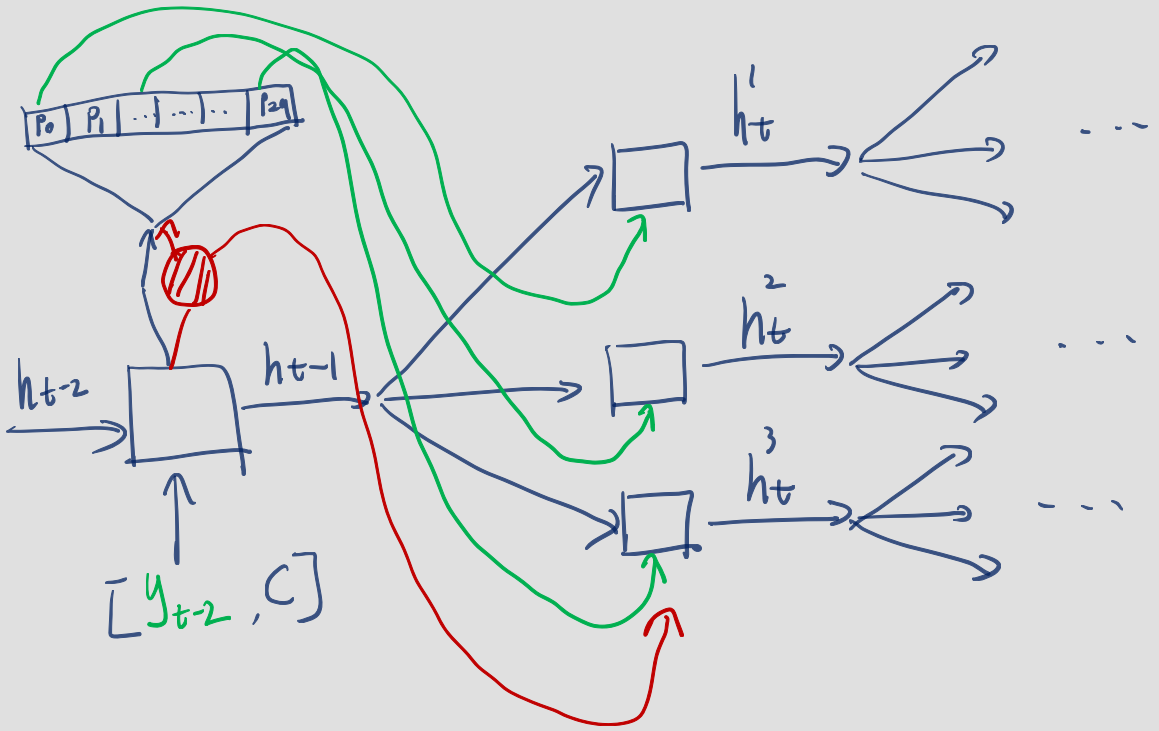


~~Gumble noise~~

Data Augmentation

LSTM Beam Decoder

Beam width = 3



$$P(\text{decoded}) = P(y_0 \mid h_{-1}, y_{-1}, C_{-1}) \cdot P(y_1 \mid h_0, y_0, C_0) \cdot \cdots$$

$$= \pi_i P(y_i \mid h_{i-1}, y_{i-1}, C_{i-1})$$

$$= \Sigma_i \log P(y_i \mid h_{i-1}, y_{i-1}, C_{i-1})$$