$$P(z_t \mid x_{1:t}) = \alpha \, P(x_t \mid z_t) \sum_{z_{t-1}} P(z_t \mid z_{t-1}) P(z_{t-1} \mid x_{1:t-1})$$

word spoken at time t

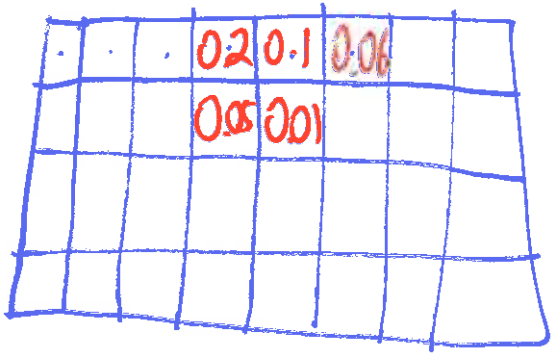
prob of hearing sound x_t if word at time t is z_t

distribution over possible words at time $t-1$

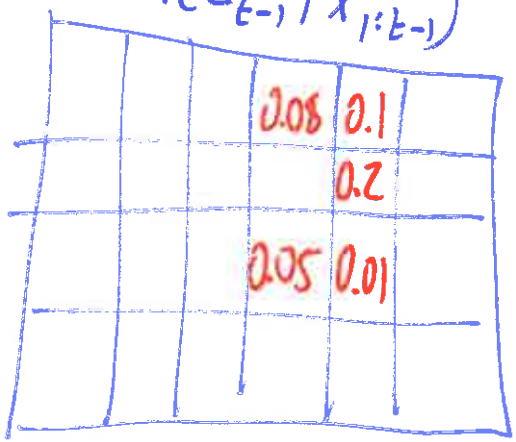
chances of changing from word at time $t-1$ to word at time t

$P(z_t \mid x_{1:t})$

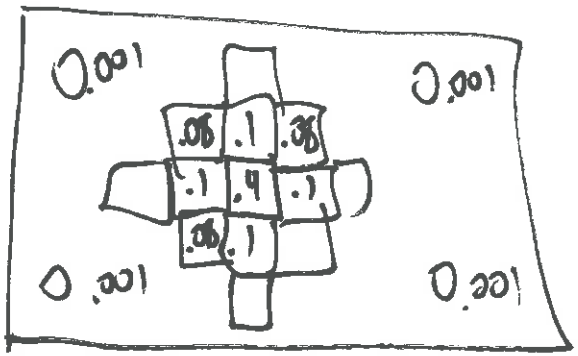
missing



$P(z_{t-1} \mid x_{1:t-1})$



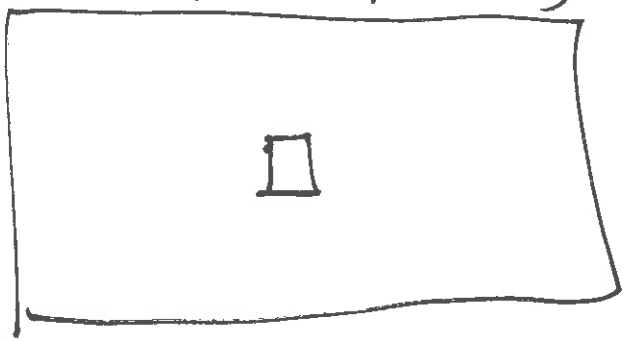
z_{t-1}



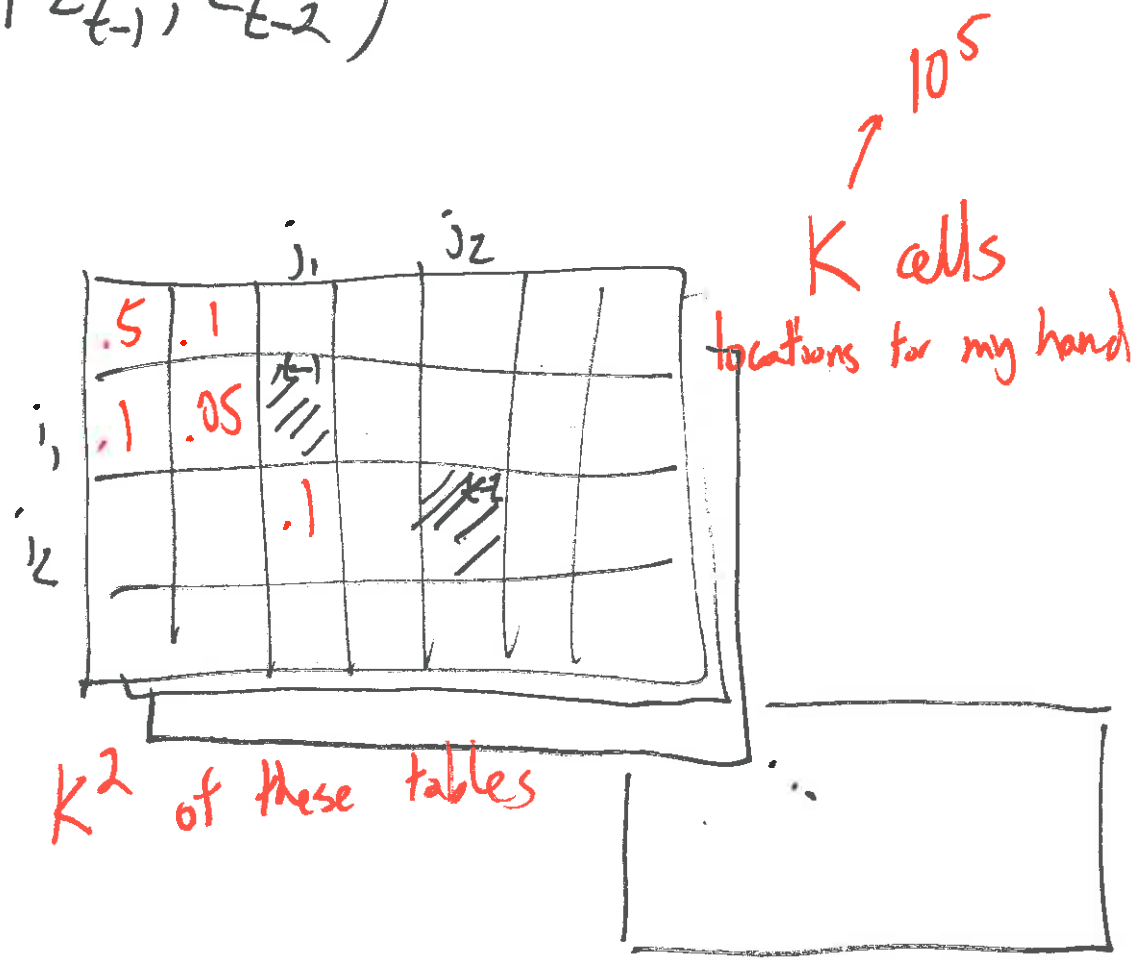
K-1

$P(z_t \mid z_{t-1}, z_{t-2})$

$z_{t-1} + (z_{t-1} - z_{t-2})$

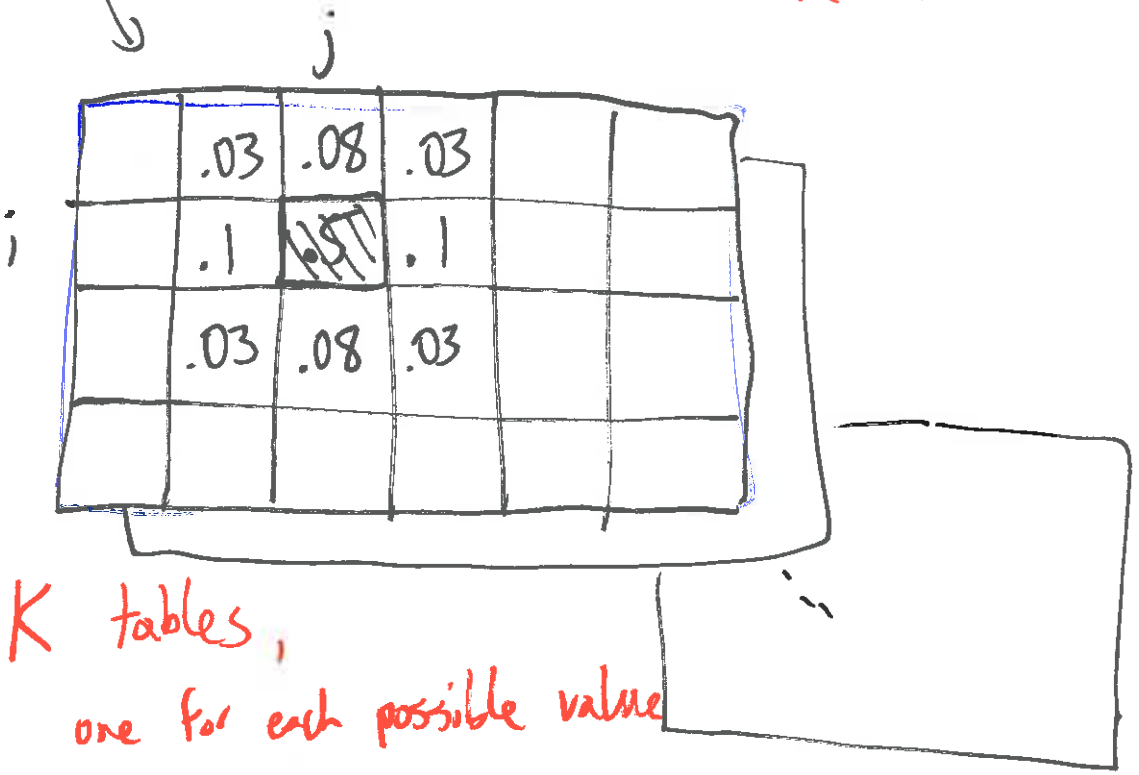


$$P(z_t \mid z_{t-1}, z_{t-2})$$



$P(z_t \mid z_{t-1} = j)$

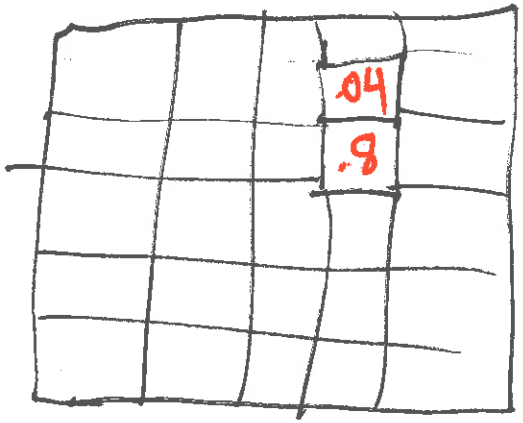
K cells

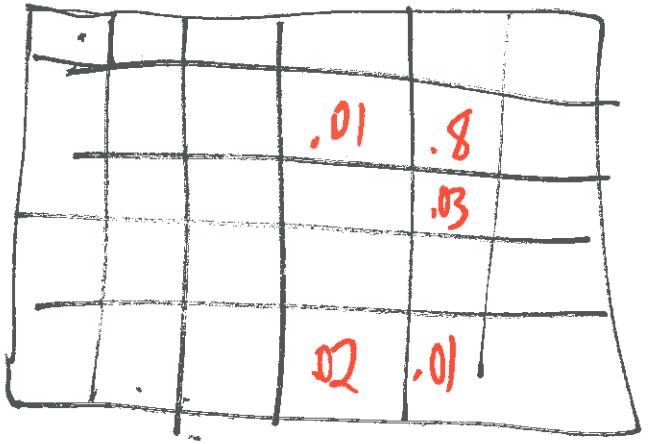


$K \cdot (K-1)$

K tables,

one for each possible value of z_{t-1}

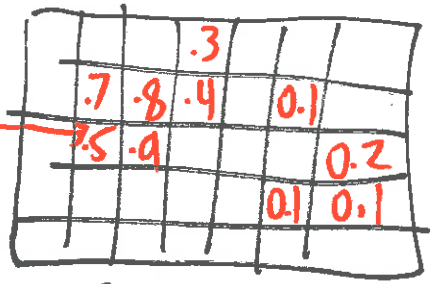




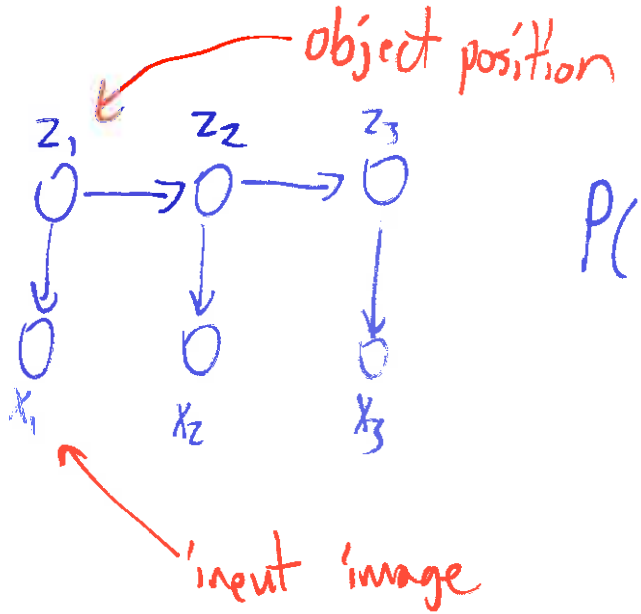
$$\sum_{z_{t-1}} P(z_t \mid z_{t-1}) \cdot P(z_{t-1} \mid x_{1:t-1})$$

$P(z_{t-1} \mid x_{1:t-1})$

$P(x_t \mid z_t = \text{"}\,\text{"})$



$P(x_t \mid z_t)$

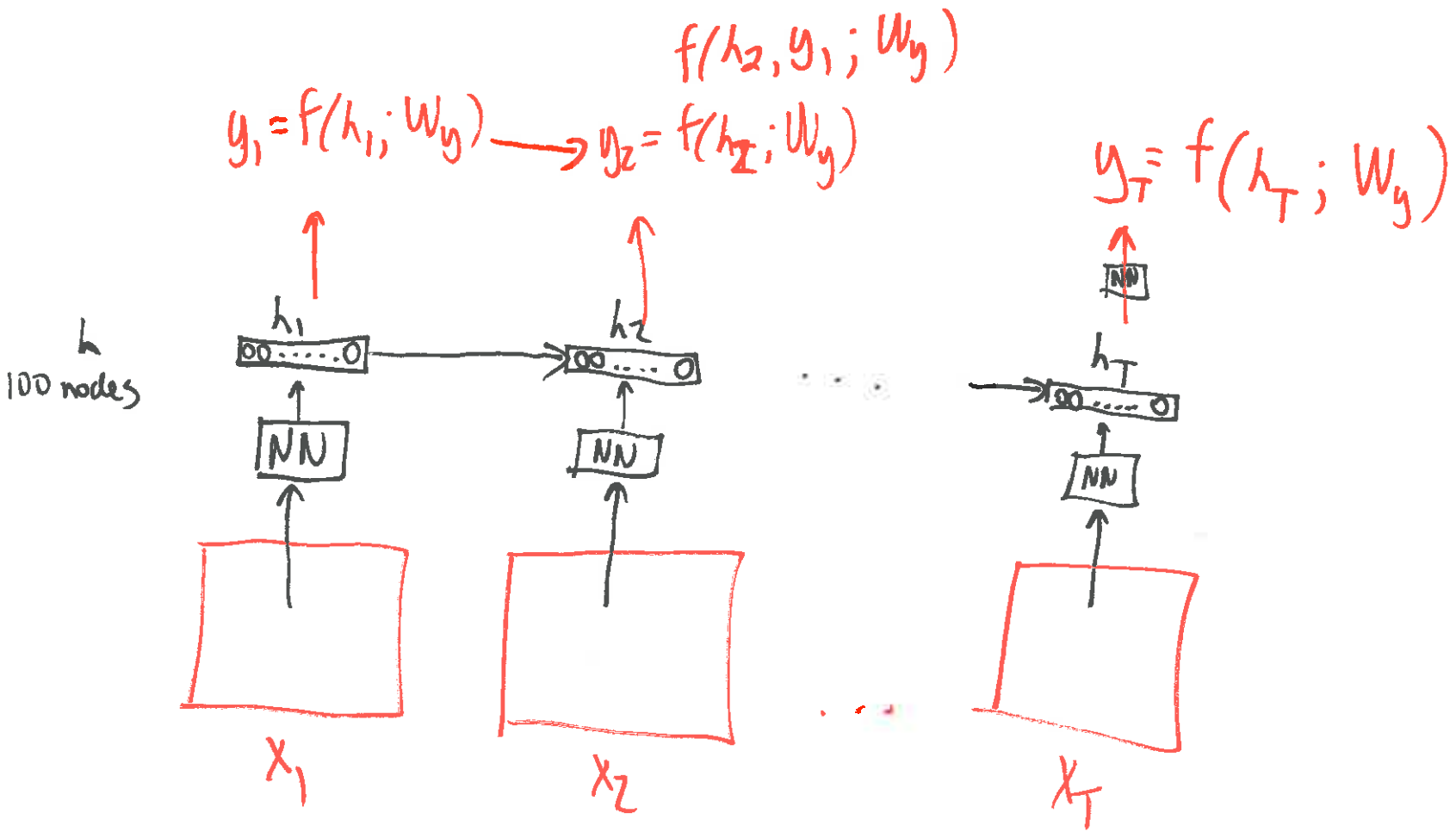


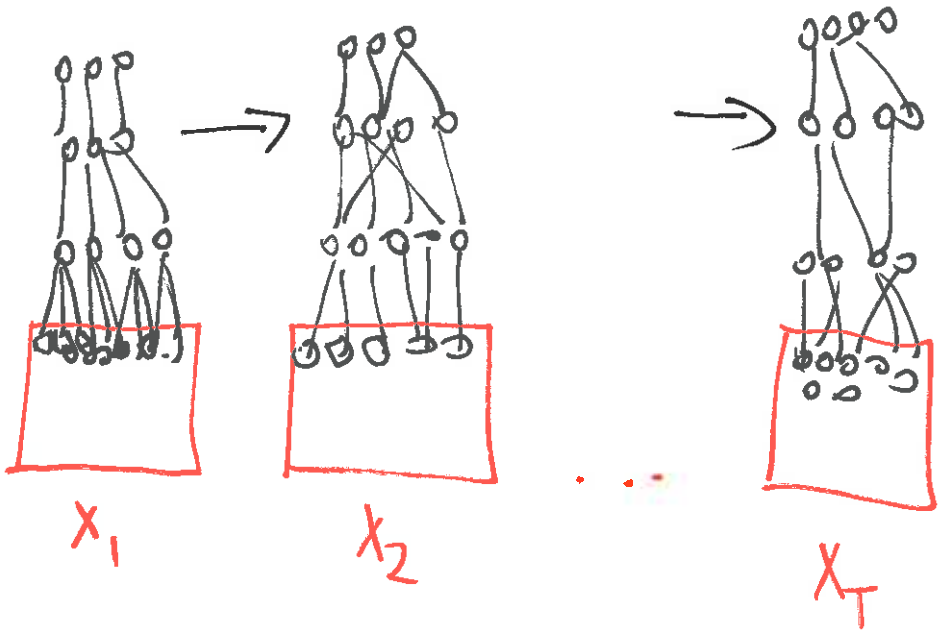
$$P(z_3 \mid x_{1:3})$$

y_1 y_2 y_3 y_4

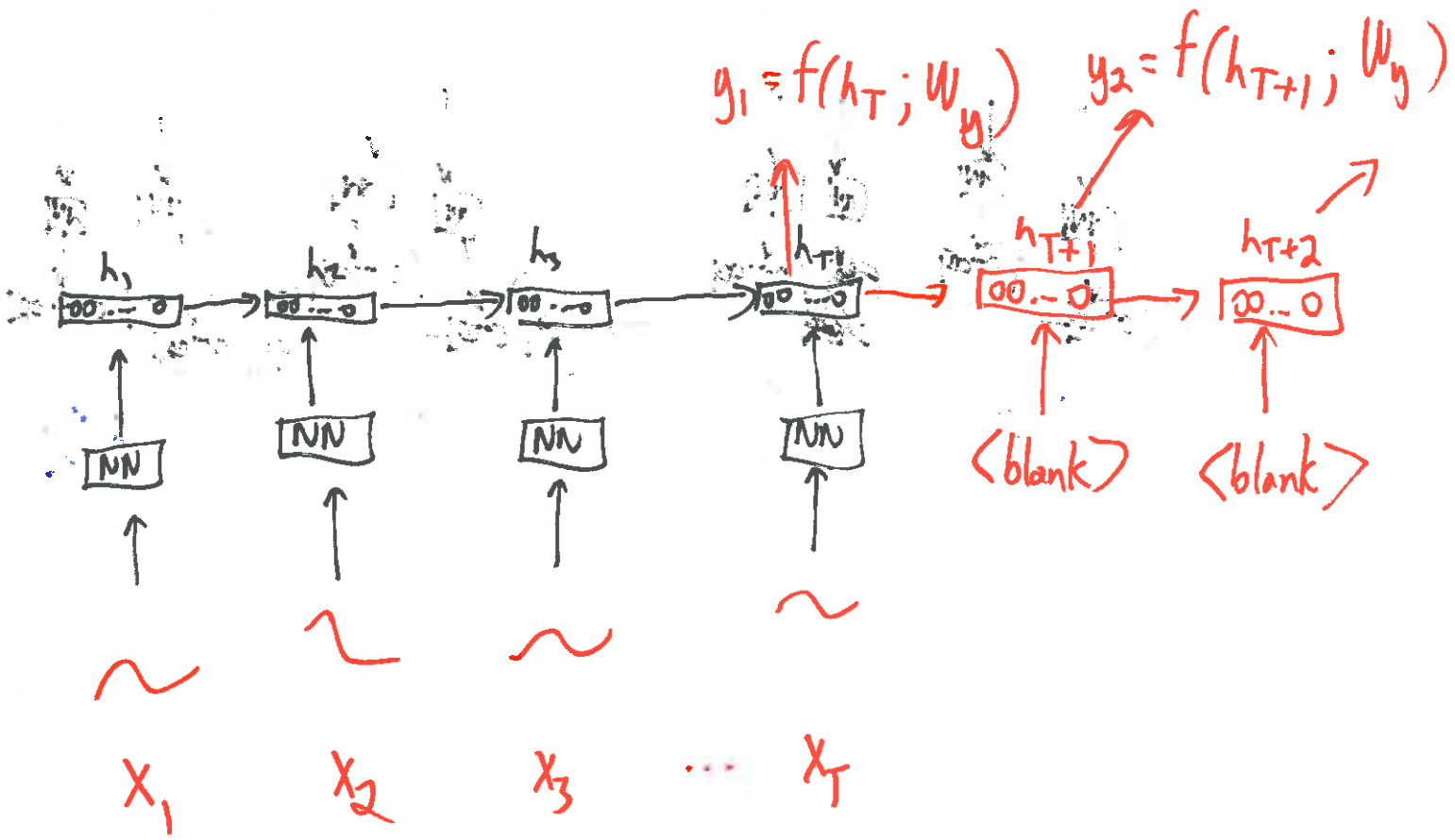
The mar stands outside

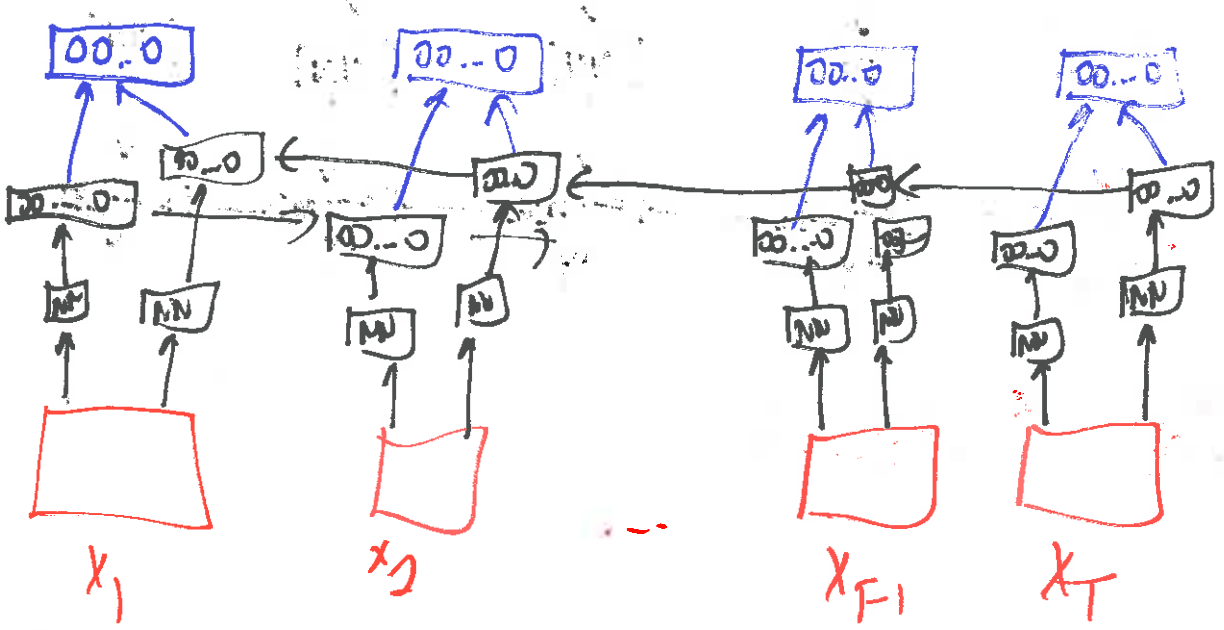
y = "sports video"


$f(h_2, y_1; W_y)$
$y_1 = f(h_1; W_y)$
$y_2 = f(h_2; W_y)$
$y_T = f(h_T; W_y)$
h
100 nodes
h_1
h_2
h_T
NN
NN
NN
NN
X_1
X_2
X_T


X_1
X_2
X_T

Speech recognition




x_1
x_2
x_{T-1}
x_T