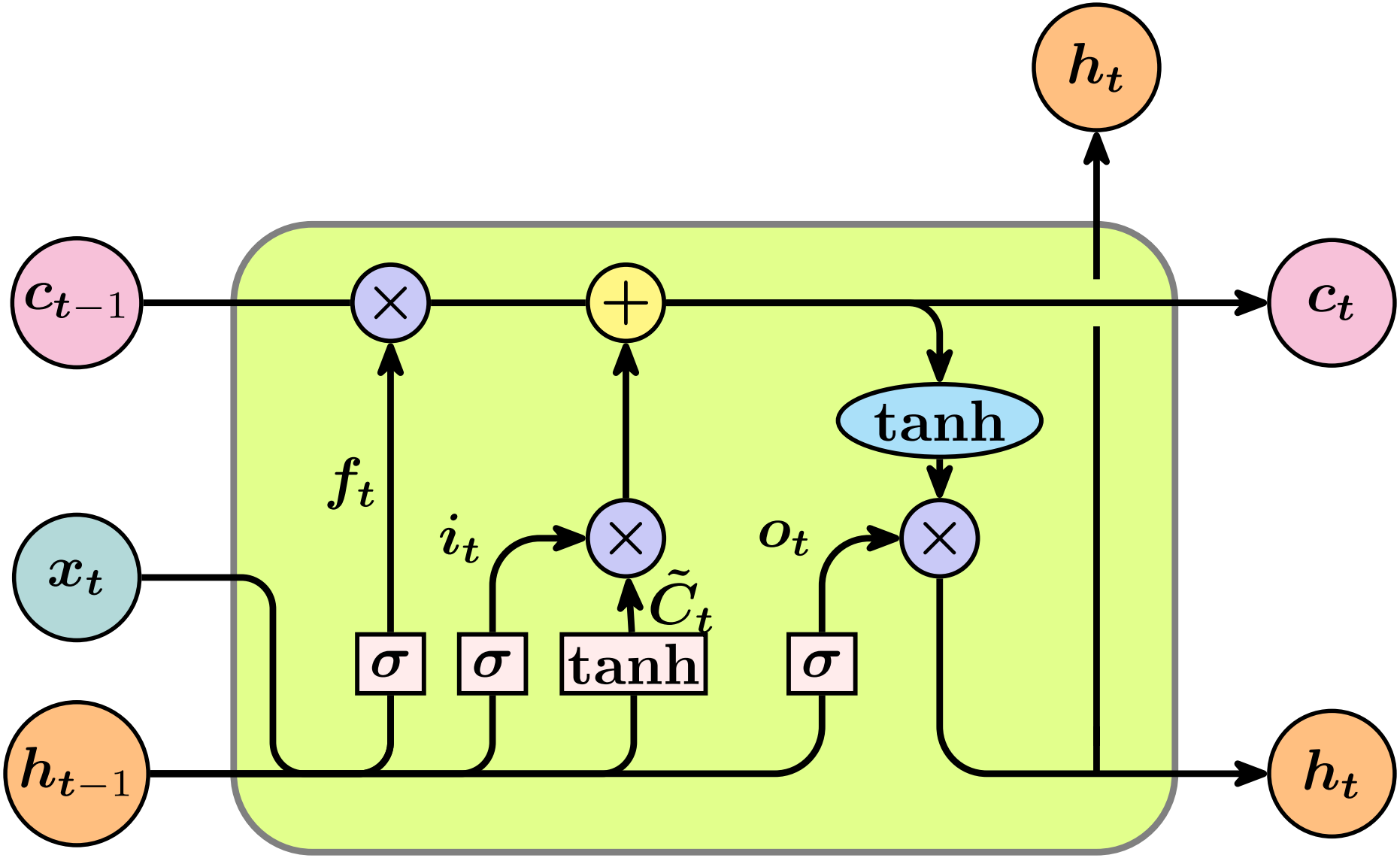
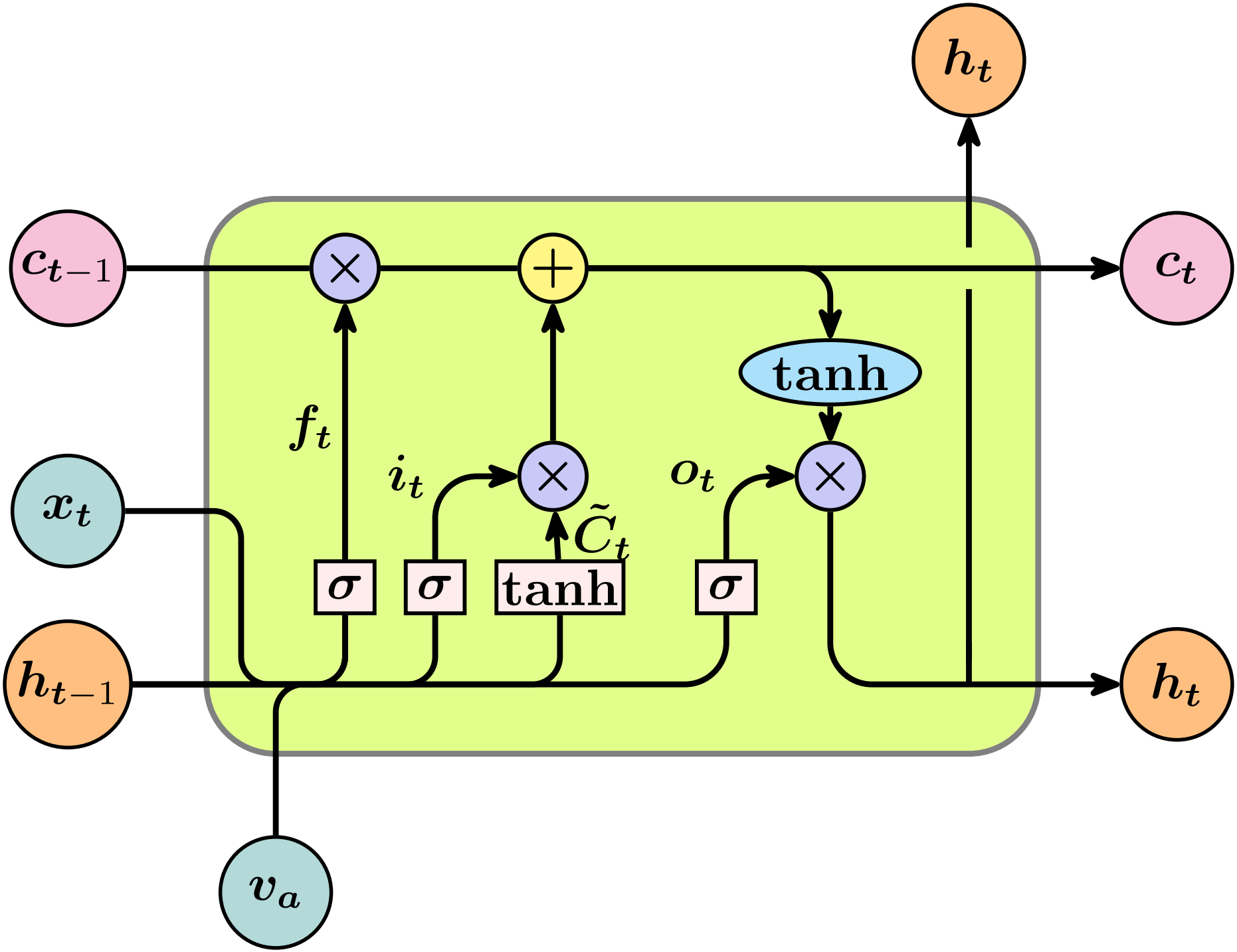
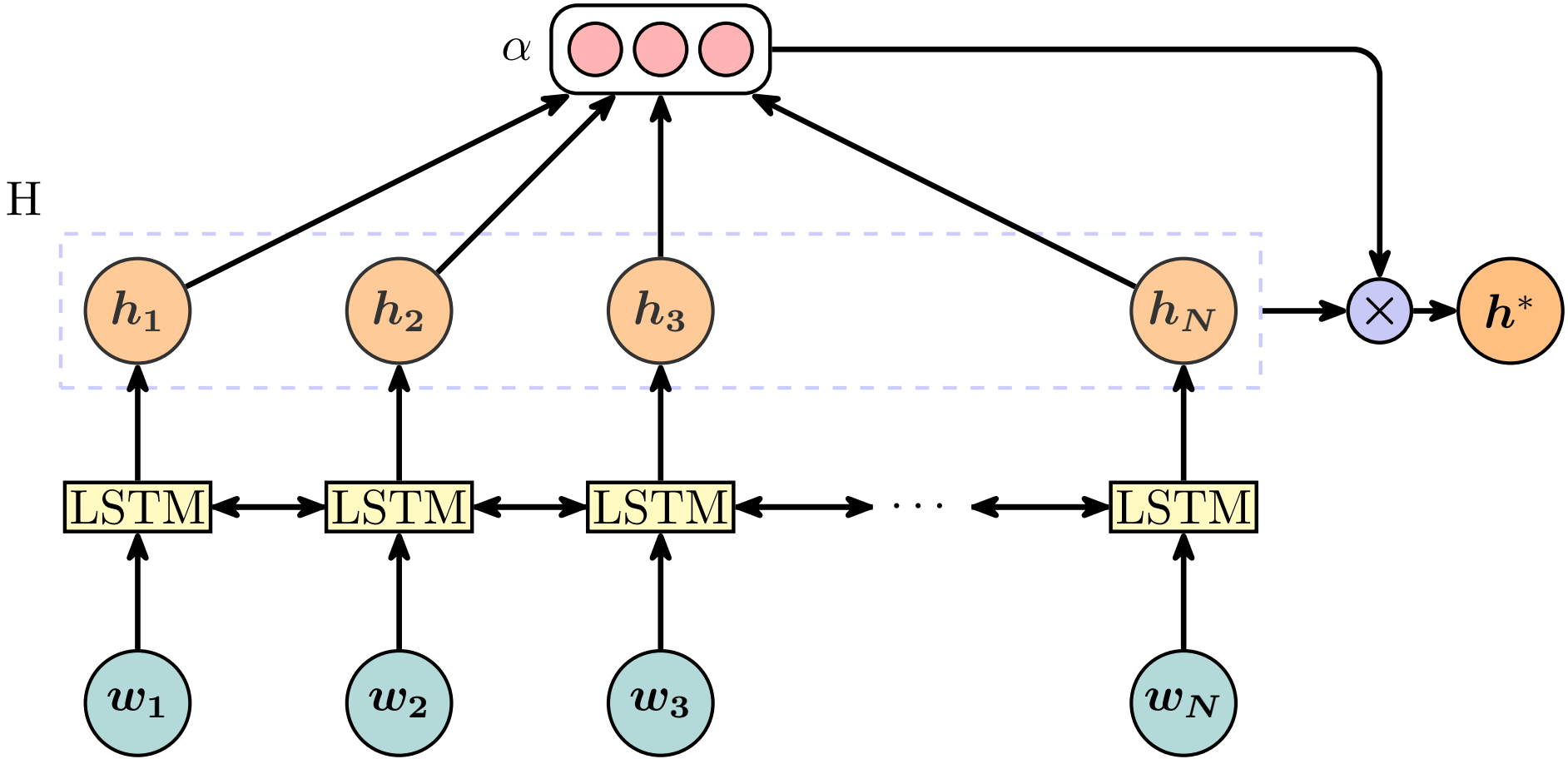

h_t
c_{t-1}
c_t
tanh
f_t
i_t
o_t
$\tilde{C}_t$
x_t
σ
σ
tanh
σ
h_{t-1}
h_t

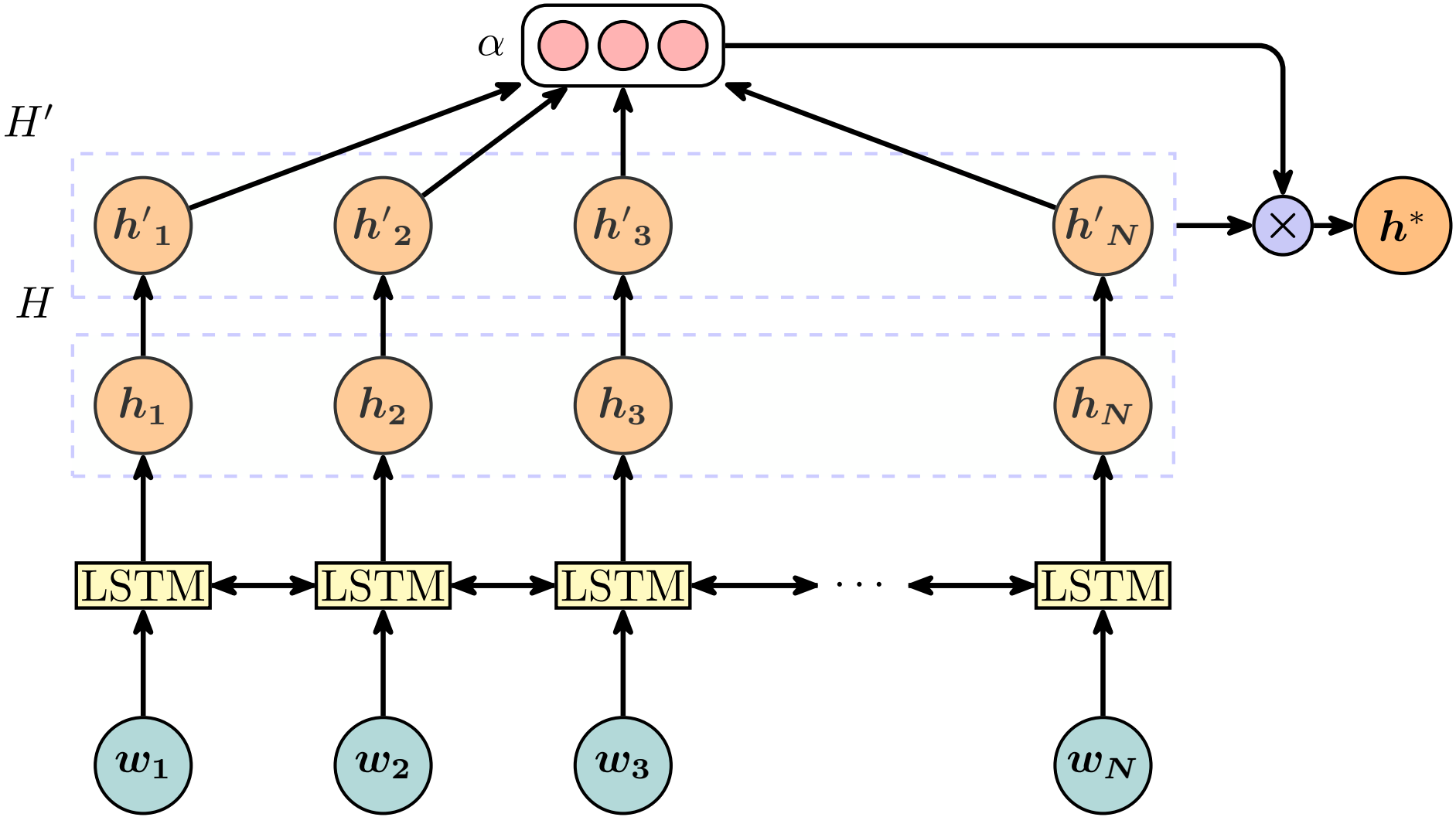

h_t
c_{t-1}
c_t
tanh
f_t
i_t
o_t
$\tilde{C}_t$
x_t
σ
σ
tanh
σ
h_{t-1}
h_t
v_a

$$
\begin{aligned}
X &= \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \\
f_t &= \sigma\left(W_f \cdot X + b_f\right) \\
i_t &= \sigma\left(W_i \cdot X + b_i\right) \\
o_t &= \sigma\left(W_o \cdot X + b_o\right) \\
\tilde{C}_t &= \tanh\left(W_c \cdot X + b_c\right) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{C}_t \\
h_t &= o_t \odot \tanh\left(c_t\right)
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
X &= \begin{bmatrix} h_{t-1} \\ x_t \\ v_a \end{bmatrix} \\
f_t &= \sigma\left(W_f \cdot X + b_f\right) \\
i_t &= \sigma\left(W_i \cdot X + b_i\right) \\
o_t &= \sigma\left(W_o \cdot X + b_o\right) \\
\tilde{C}_t &= \tanh\left(W_c \cdot X + b_c\right) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{C}_t \\
h_t &= o_t \odot \tanh\left(c_t\right)
\end{aligned}
\tag{2}
$$

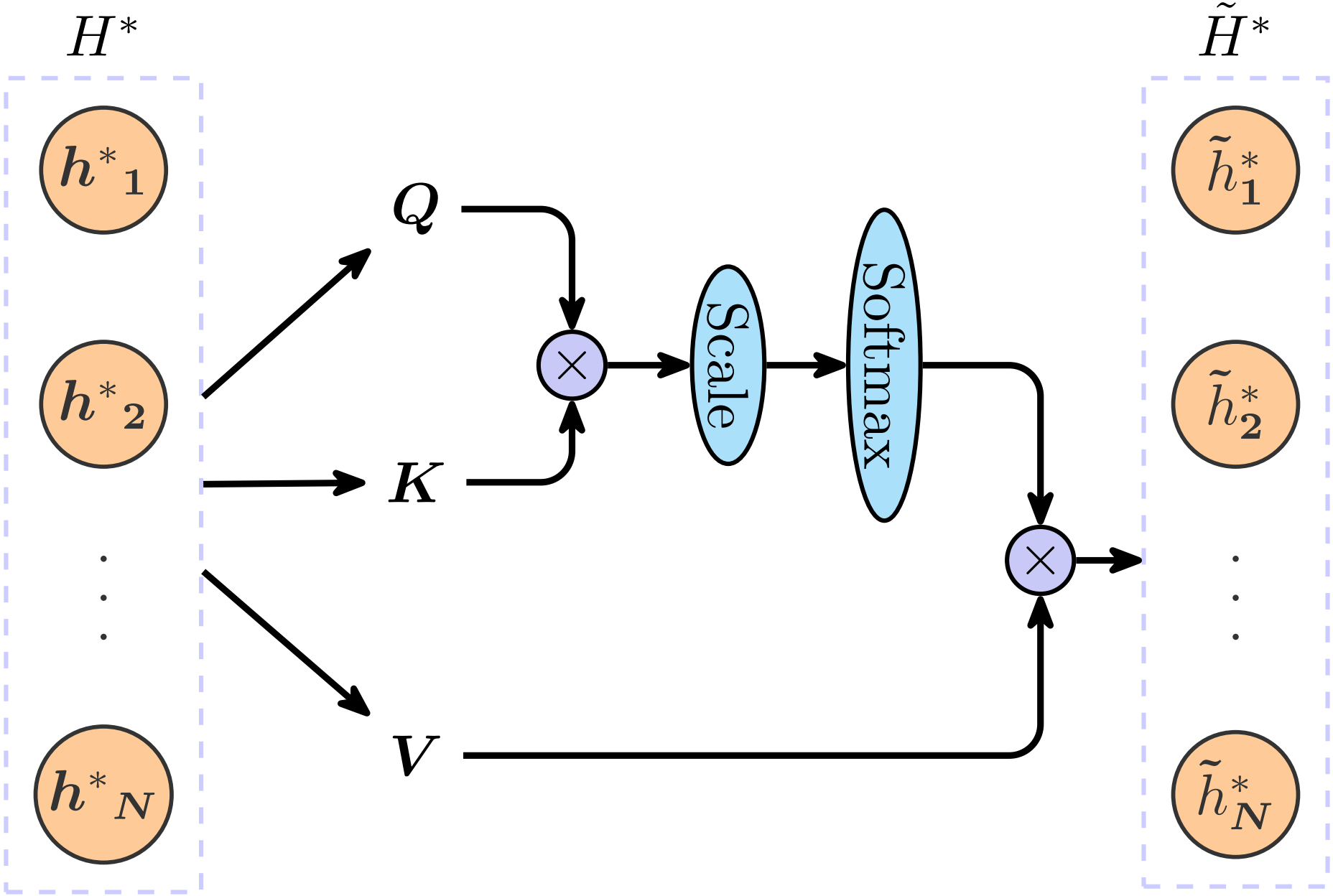

α
H
h_1
h_2
h_3
h_N
h^*
LSTM
LSTM
LSTM
LSTM
w_1
w_2
w_3
w_N

$$
\begin{aligned}
M &= \tanh\left(\begin{bmatrix} W_h H \\ W_v v_a \otimes e_N \end{bmatrix}\right) \\
\alpha &= \mathrm{softmax}\left(w^T M\right) \\
h^* &= H\alpha^T
\end{aligned}
\tag{3}
$$

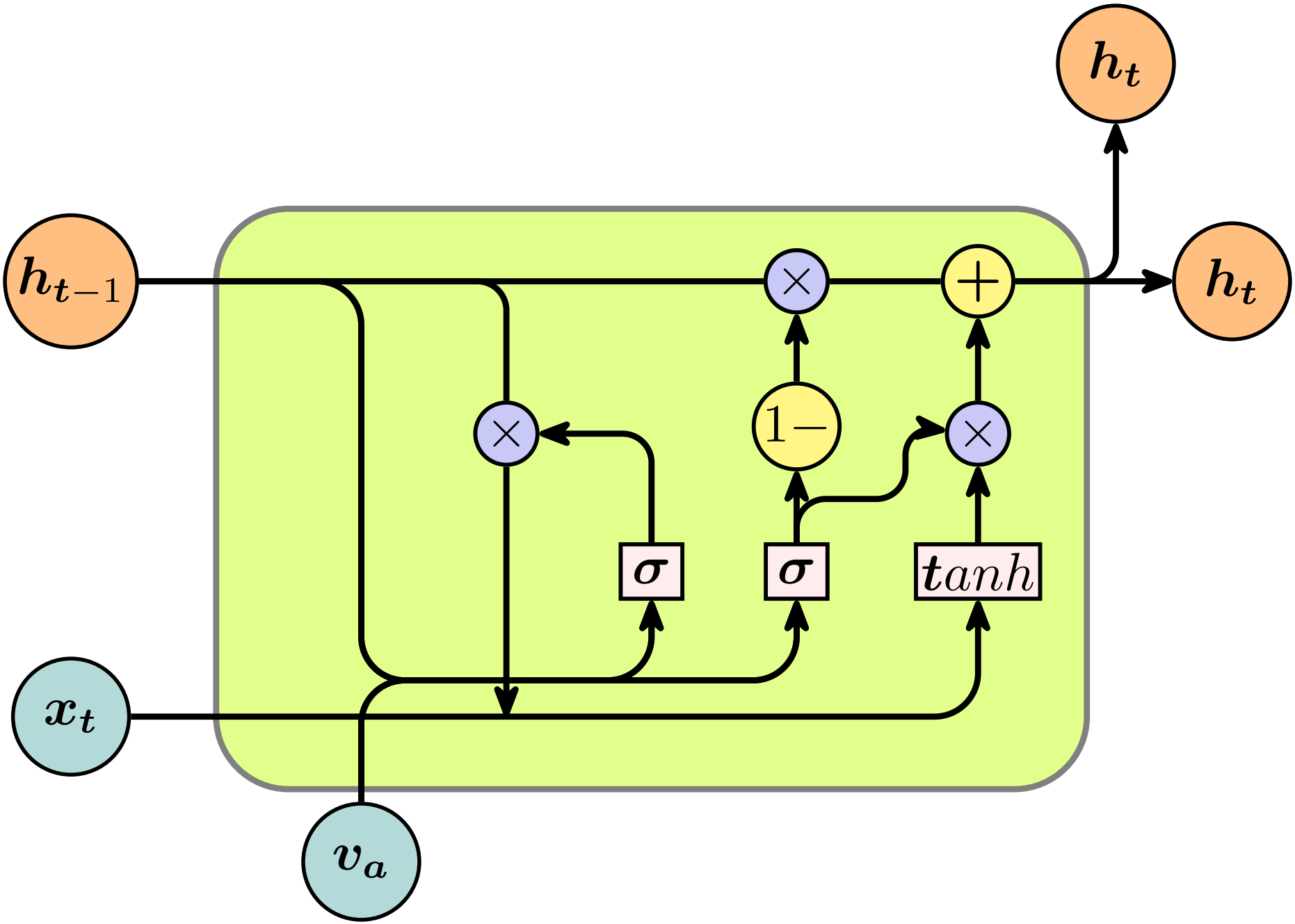

α
H′
H
h′1
h′2
h′3
h′N
h*
h1
h2
h3
hN
LSTM
LSTM
LSTM
· · ·
LSTM
w1
w2
w3
wN

$$
\begin{aligned}
v_a^{'} &= v_a + p_a \\
H^{'} &= H \ + \ P \\
\alpha &= \text{softmax}\left(\frac{\left(W_v v_a^{'}\right)^T \left(W_h H^{'}\right)}{\sqrt{d_a}}\right) \\
h^* &= H\alpha^T
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
Q &= H^* W_Q \\
K &= H^* W_k \\
V &= H^* W_V \\
\tilde{H}^* &= \text{softmax}\left(\frac{Q \cdot K^\top}{\sqrt{d}}\right) \cdot V
\end{aligned}
\tag{5}
$$


H^*
$\tilde{H}^*$
h^*_1
h^*_2
h^*_N
Q
K
V
Scale
Softmax
$\tilde{h}^*_1$
$\tilde{h}^*_2$
$\tilde{h}^*_N$

$$
\begin{aligned}
r_t &= \sigma\left(W_r \cdot \begin{bmatrix} h_{t-1} \\ x_t \\ v_a \end{bmatrix} + b_r\right) \\
z_t &= \sigma\left(W_z \cdot \begin{bmatrix} h_{t-1} \\ x_t \\ v_a \end{bmatrix} + b_z\right) \\
\tilde{h_t} &= \tanh\left(W \cdot \begin{bmatrix} r_{t*}h_{t-1} \\ x_t \\ v_a \end{bmatrix}\right) \\
h_t &= (1 - z_t) \cdot h_{t-1} + z_t \cdot h_t
\end{aligned}
\tag{6}
$$


h_t
h_{t-1}
h_t
$1-$
σ
σ
$tanh$
x_t
v_a