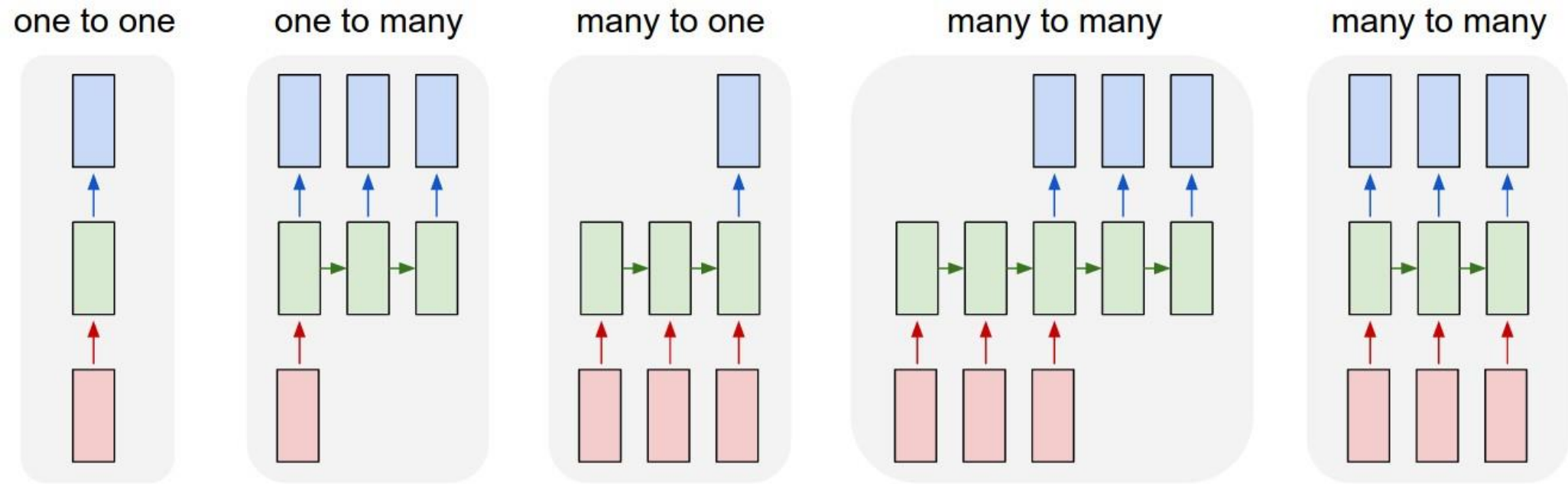
Рекуррентные нейронные сети

Анализ последовательностей

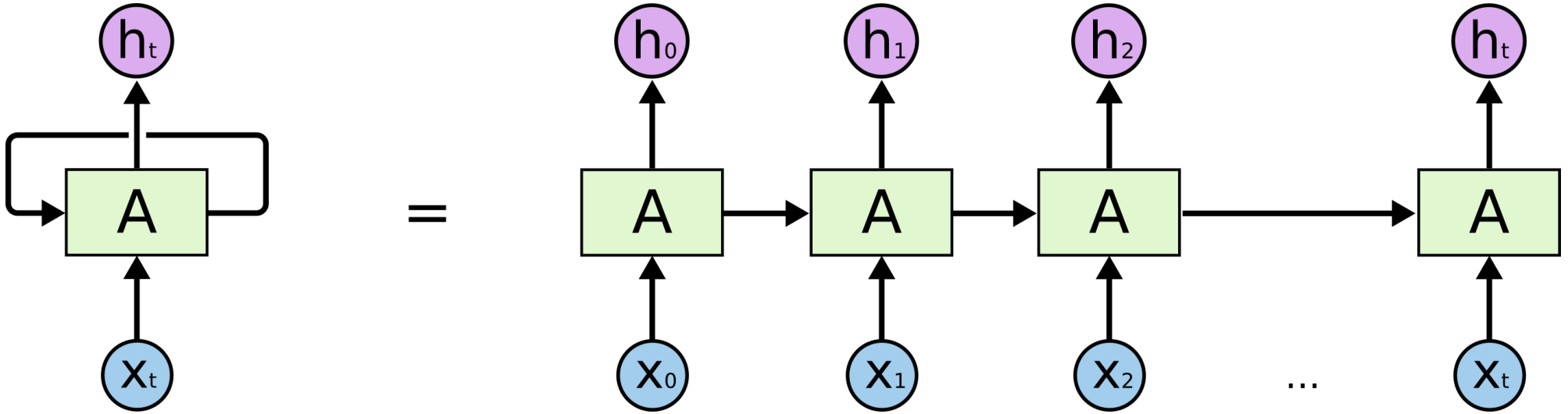
1. Временные ряды
2. Тексты
3. Аудио и видеопоследовательности



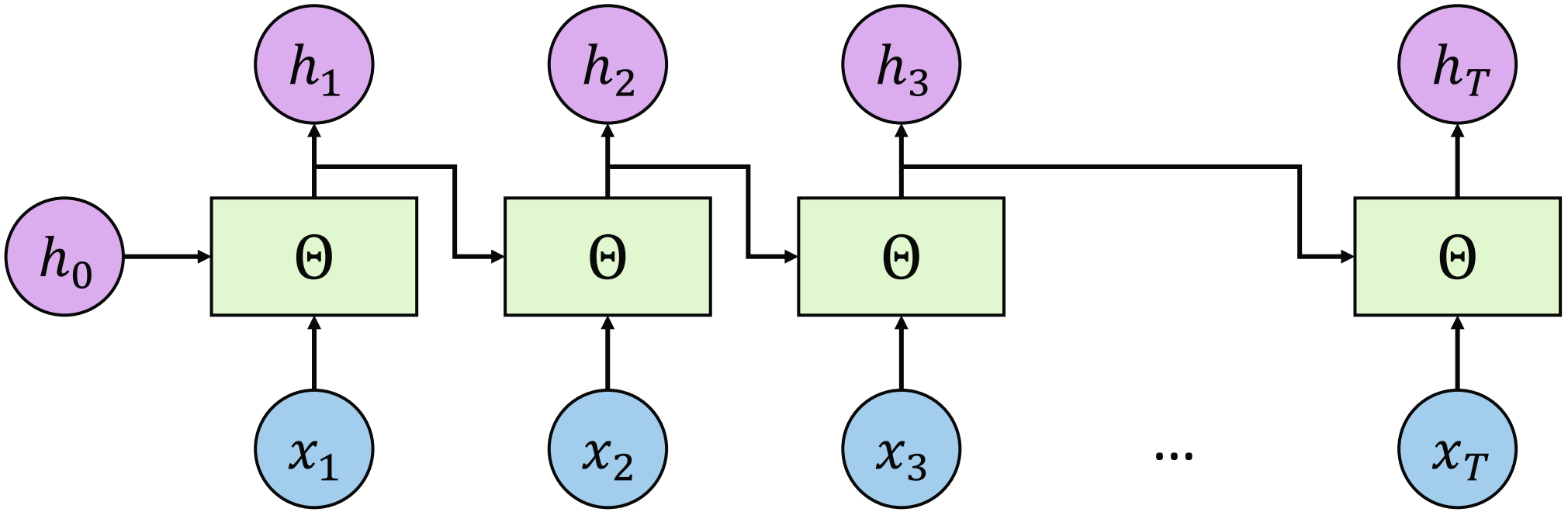
Рекуррентная нейронная сеть

Вход сети: $\{x_0, x_1, x_2, \dots, x_t\}$ – последовательность объектов;

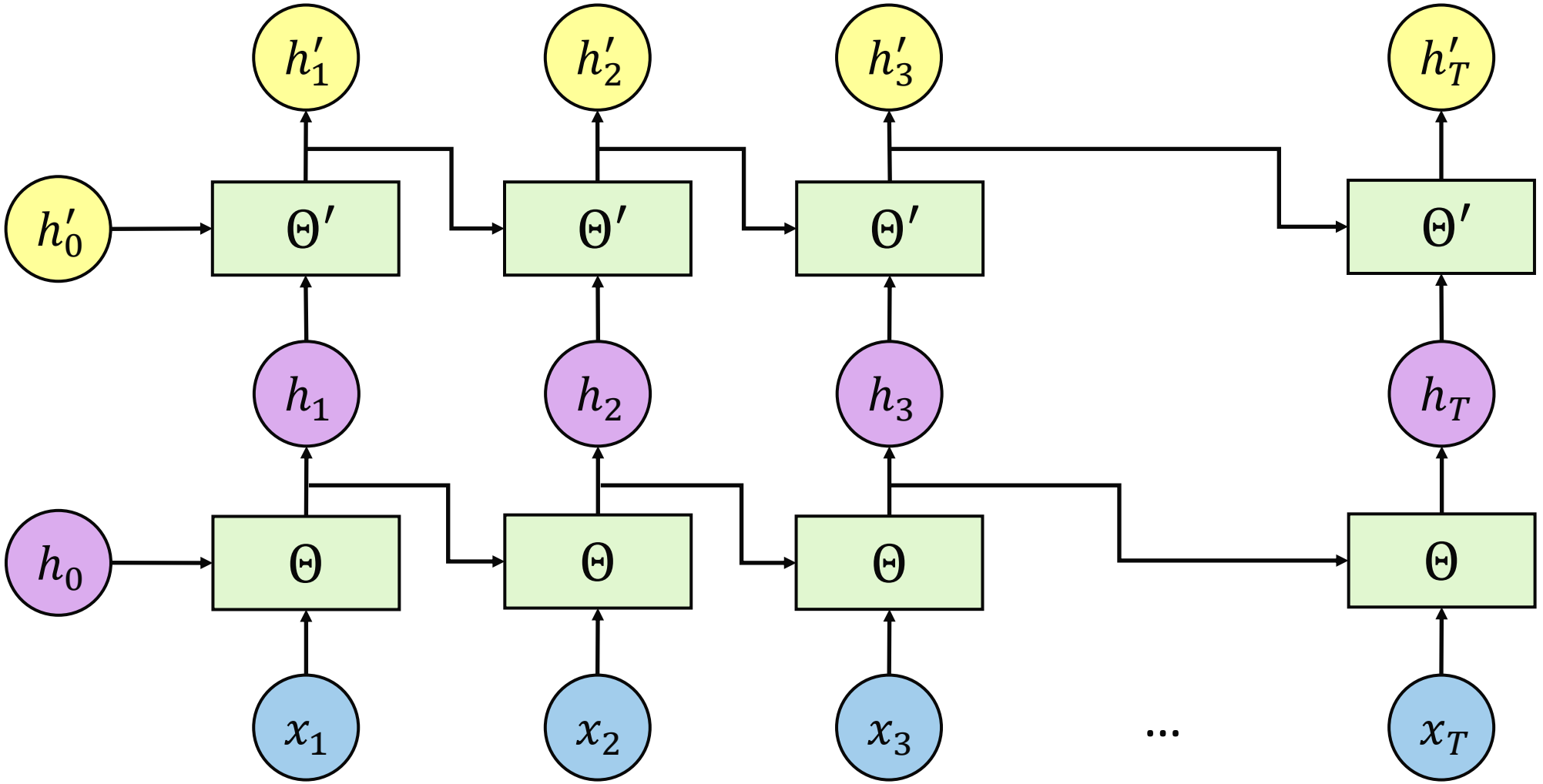
Скрытое состояние: $\{h_0, h_1, h_2, \dots, h_t\}$.



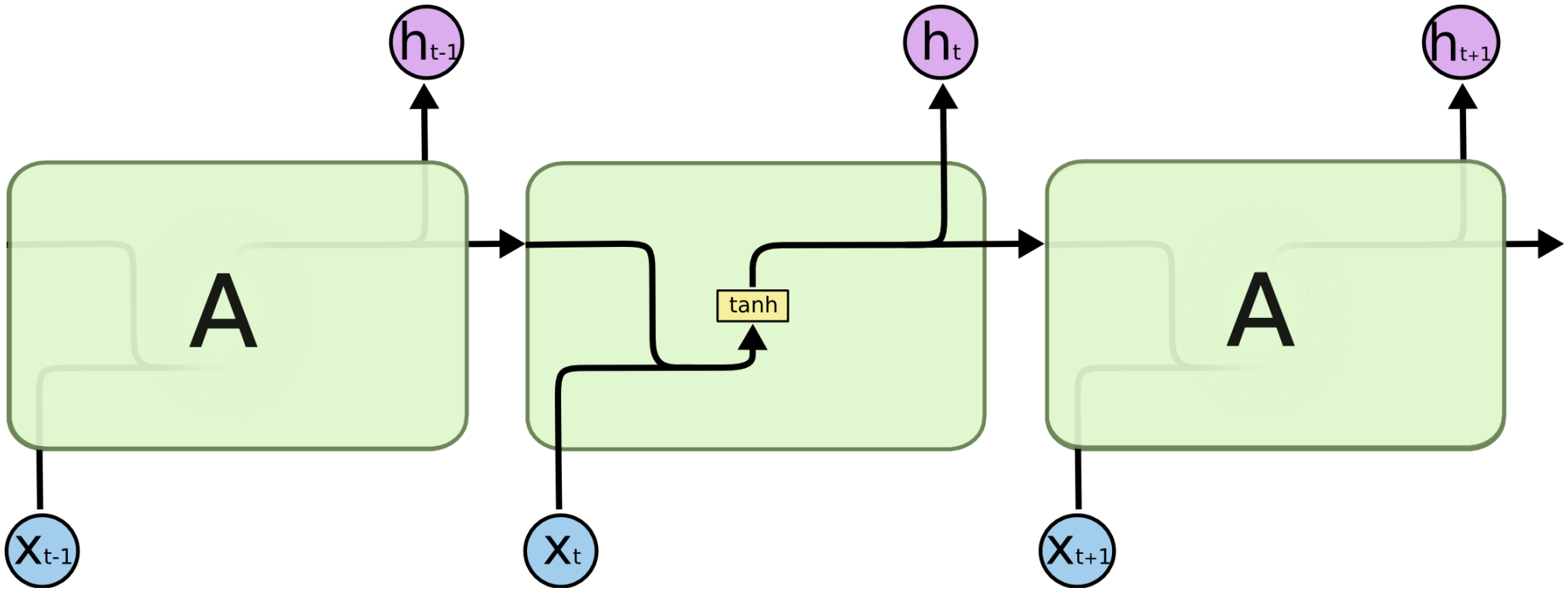
Добавление слоев



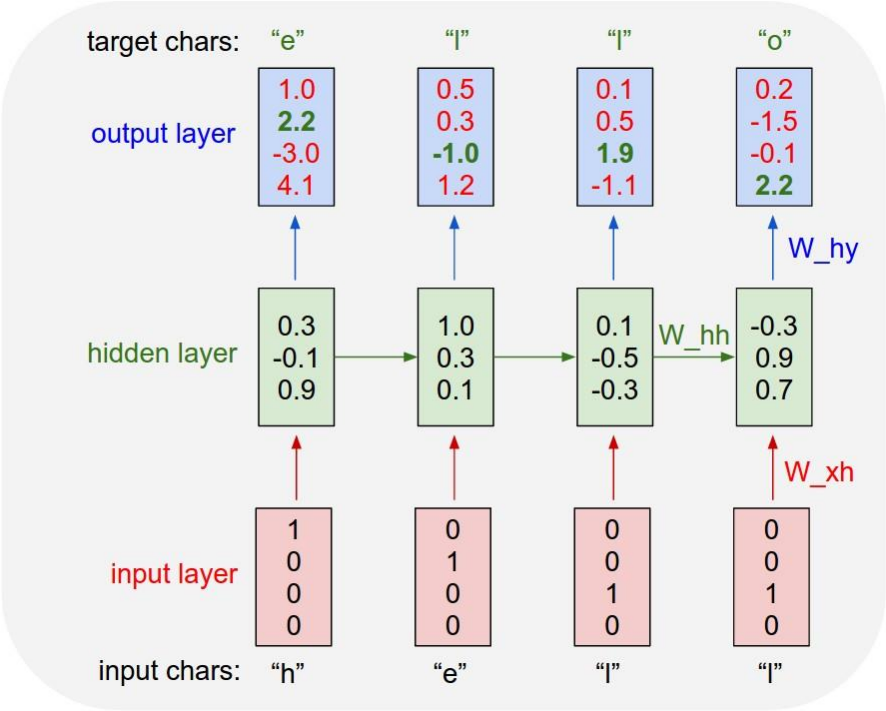
Добавление слоев



Рекуррентная нейронная сеть



Пример: генерация текста



Iteration 100

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

Iteration 300

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

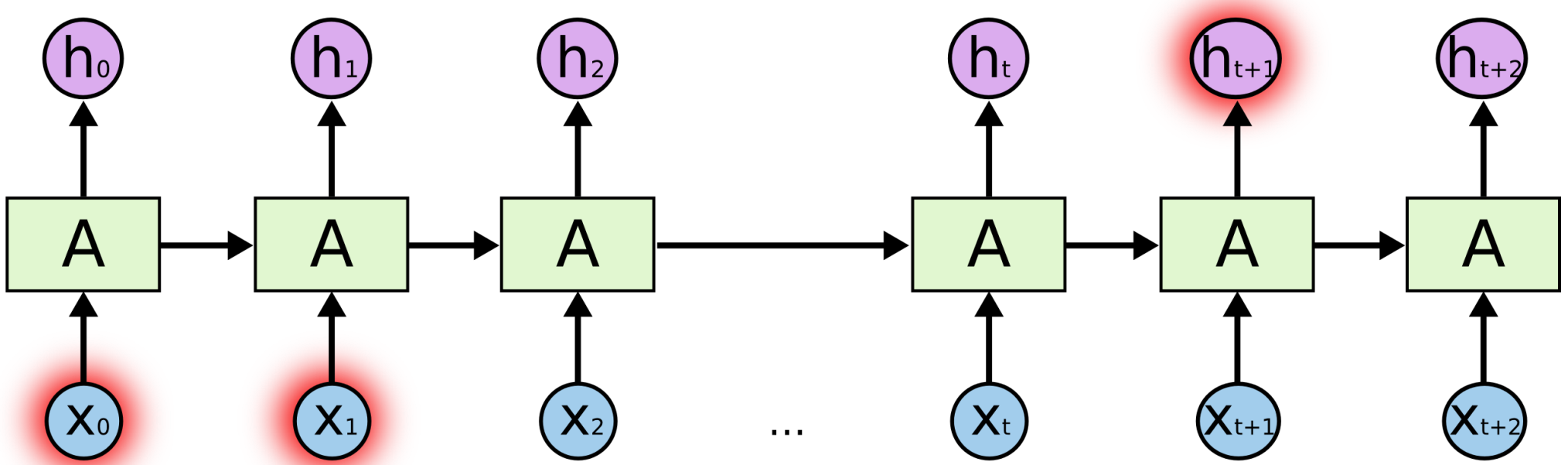
Iteration 700

```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

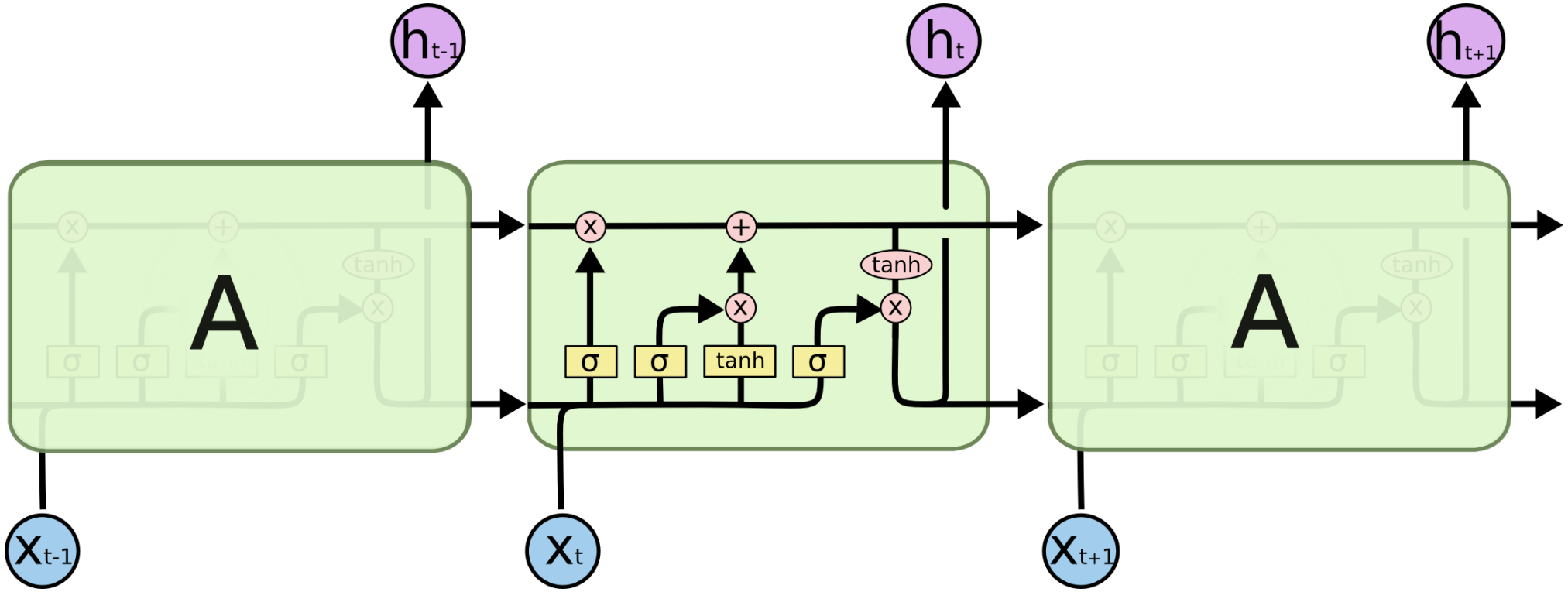
Iteration 2000

```
"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
```

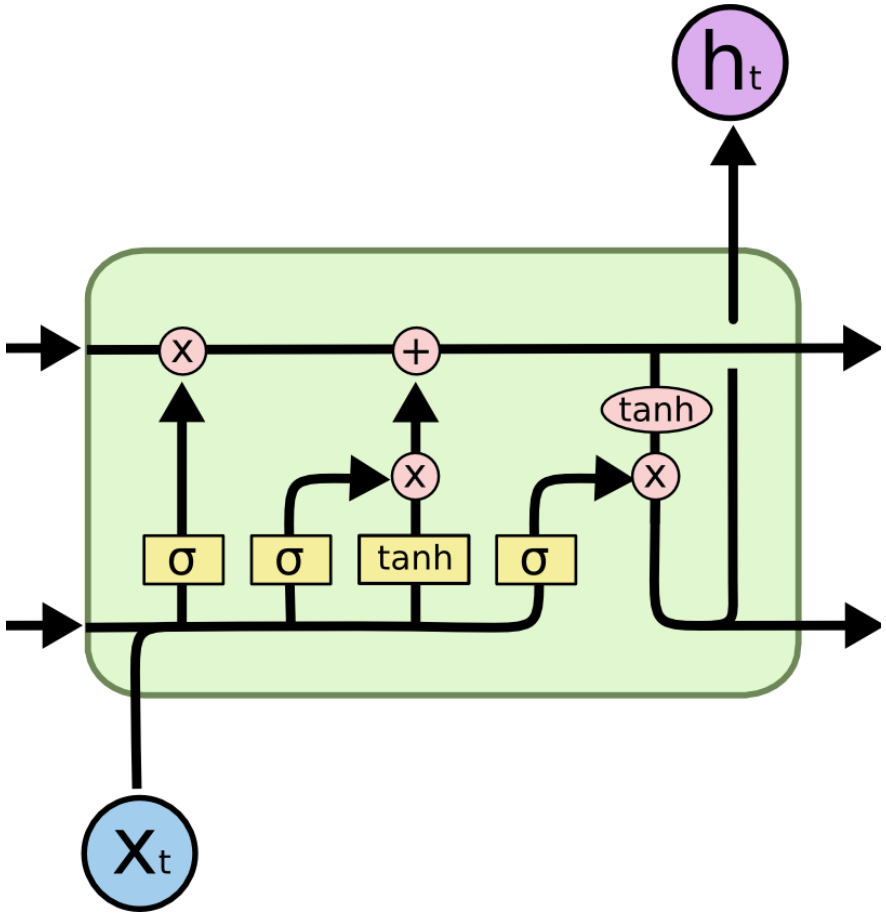
Проблема длинных зависимостей



Long Short-Term Memory (LSTM)



Long Short-Term Memory (LSTM)



Forget gate:

$$f_t = \sigma\left(\Theta_f \cdot [h_{t-1}, x_t] + b_f\right)$$

Input gate:

$$i_t = \sigma(\Theta_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(\Theta_C \cdot [h_{t-1}, x_t] + b_c)$$

Cell update:

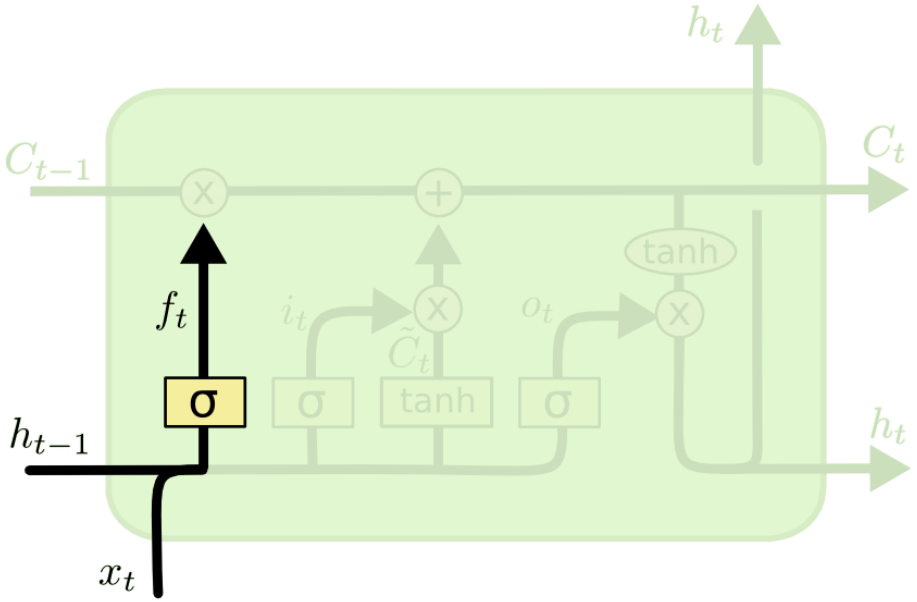
$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$

Output gate:

$$o_t = \sigma(\Theta_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

Long Short-Term Memory (LSTM)



Forget gate:

$$f_t = \sigma\left(\Theta_f \cdot [h_{t-1}, x_t] + b_f\right)$$

Input gate:

$$i_t = \sigma(\Theta_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(\Theta_C \cdot [h_{t-1}, x_t] + b_c)$$

Cell update:

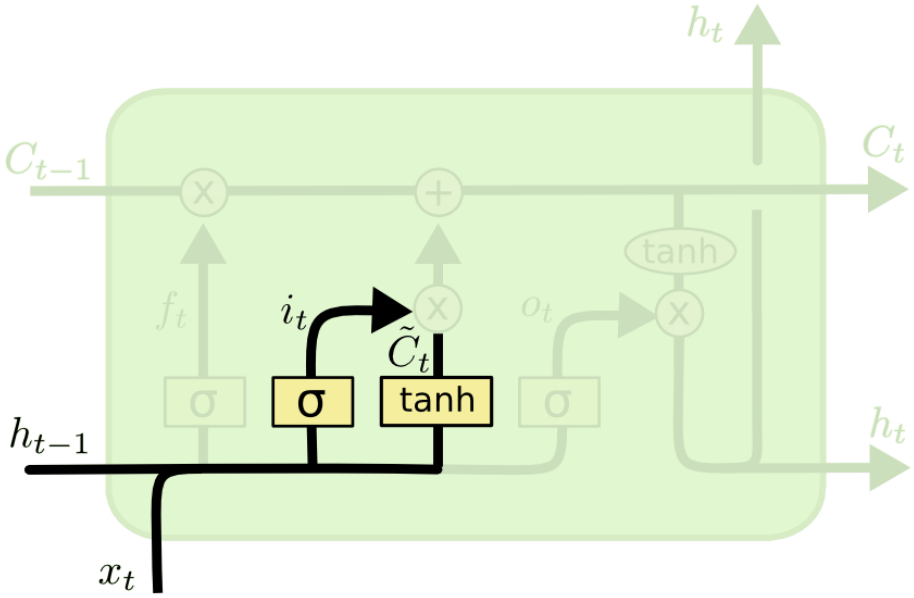
$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$

Output gate:

$$o_t = \sigma(\Theta_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

Long Short-Term Memory (LSTM)



Forget gate:

$$f_t = \sigma(\Theta_f \cdot [h_{t-1}, x_t] + b_f)$$

Input gate:

$$i_t = \sigma(\Theta_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(\Theta_C \cdot [h_{t-1}, x_t] + b_c)$$

Cell update:

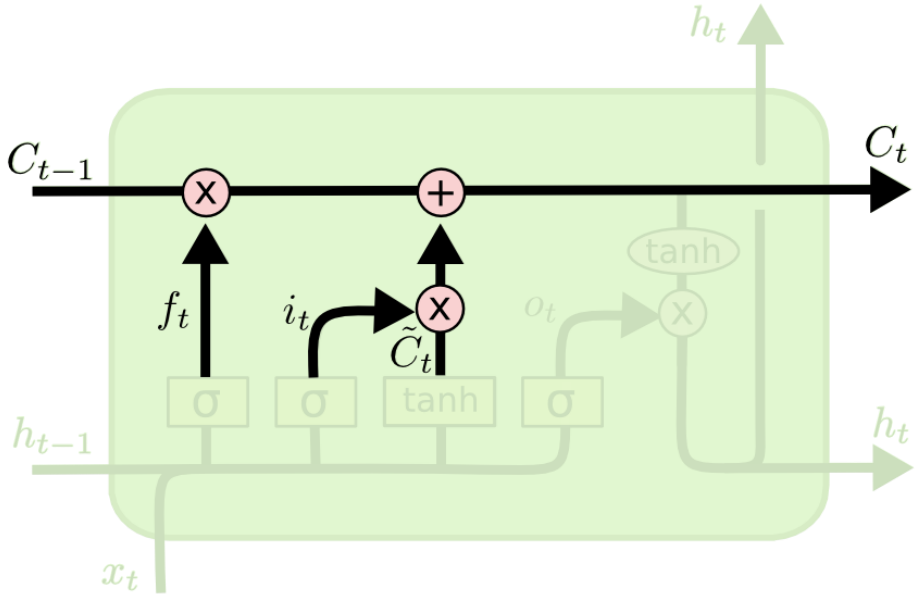
$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$

Output gate:

$$o_t = \sigma(\Theta_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

Long Short-Term Memory (LSTM)



Forget gate:

$$f_t = \sigma\left(\Theta_f \cdot [h_{t-1}, x_t] + b_f\right)$$

Input gate:

$$i_t = \sigma(\Theta_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(\Theta_C \cdot [h_{t-1}, x_t] + b_c)$$

Cell update:

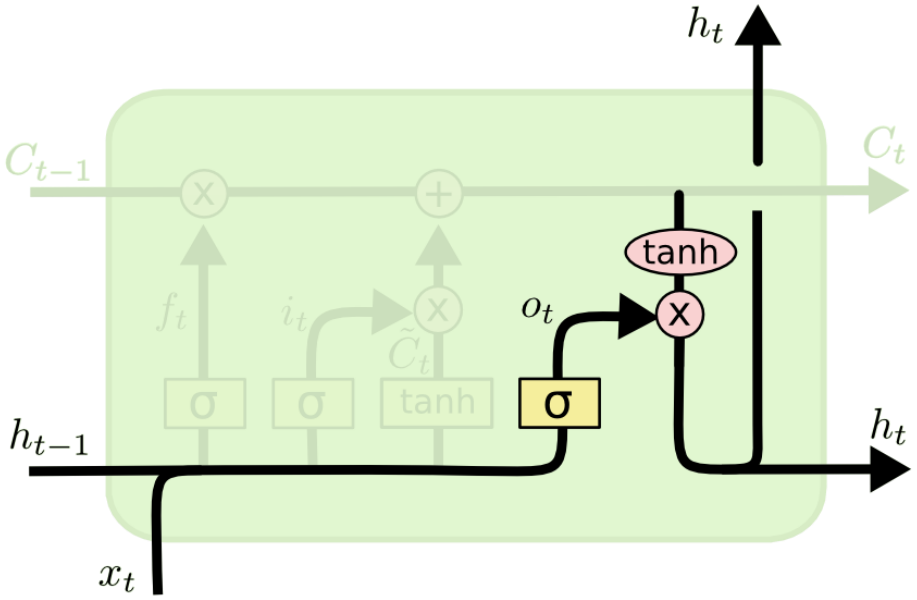
$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$

Output gate:

$$o_t = \sigma(\Theta_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

Long Short-Term Memory (LSTM)



Forget gate:

$$f_t = \sigma(\Theta_f \cdot [h_{t-1}, x_t] + b_f)$$

Input gate:

$$i_t = \sigma(\Theta_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(\Theta_C \cdot [h_{t-1}, x_t] + b_c)$$

Cell update:

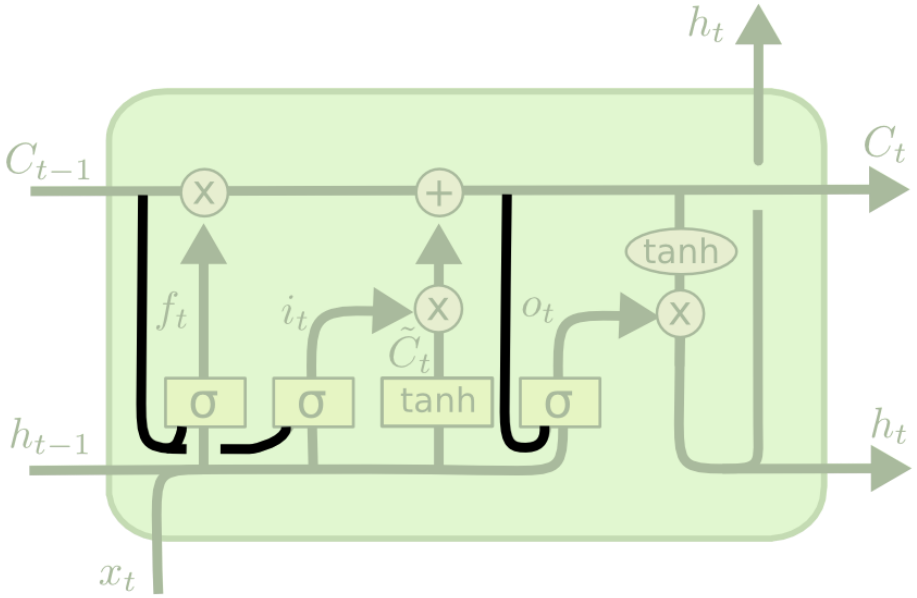
$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t$$

Output gate:

$$o_t = \sigma(\Theta_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

LSTM with “peephole connections”



Forget gate:

$$f_t = \sigma\left(\Theta_f \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] + b_f\right)$$

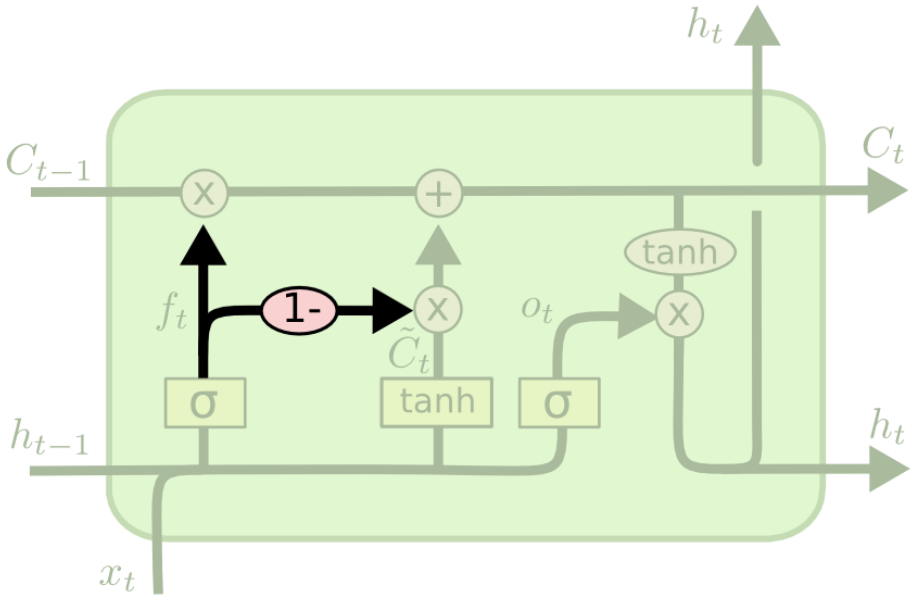
Input gate:

$$i_t = \sigma(\Theta_i \cdot [\boldsymbol{C_{t-1}}, h_{t-1}, x_t] + b_i)$$

Output gate:

$$o_t = \sigma(\Theta_o \cdot [\boldsymbol{C_t}, h_{t-1}, x_t] + b_o)$$

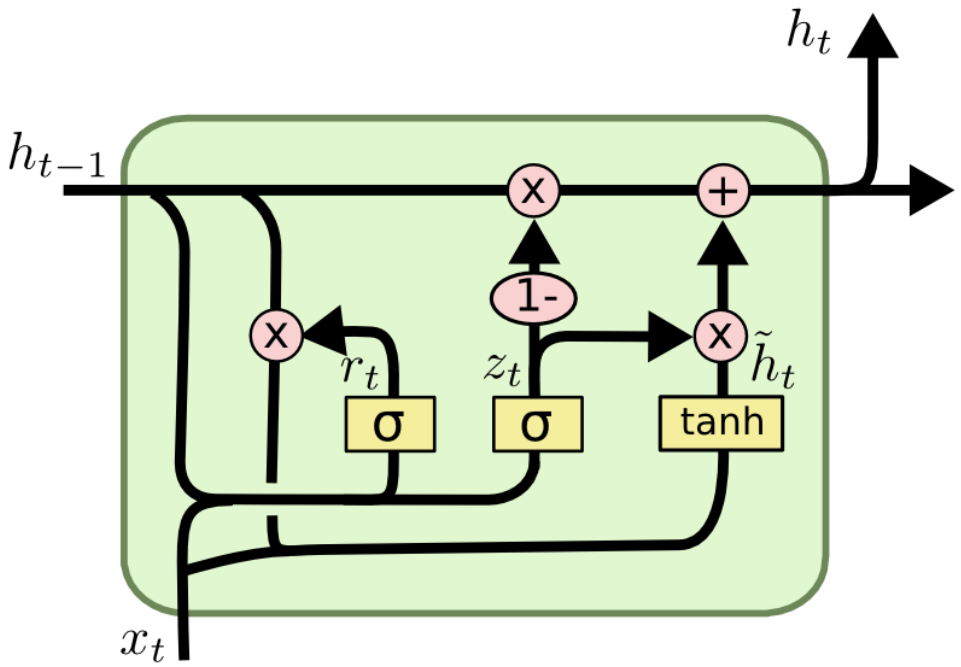
LSTM with coupled forget and input gates



Cell update:

$$C_t = f_t \times C_{t-1} + (1 - f_t) \times \tilde{C}_t$$

Gated Recurrent Unit (GRU)



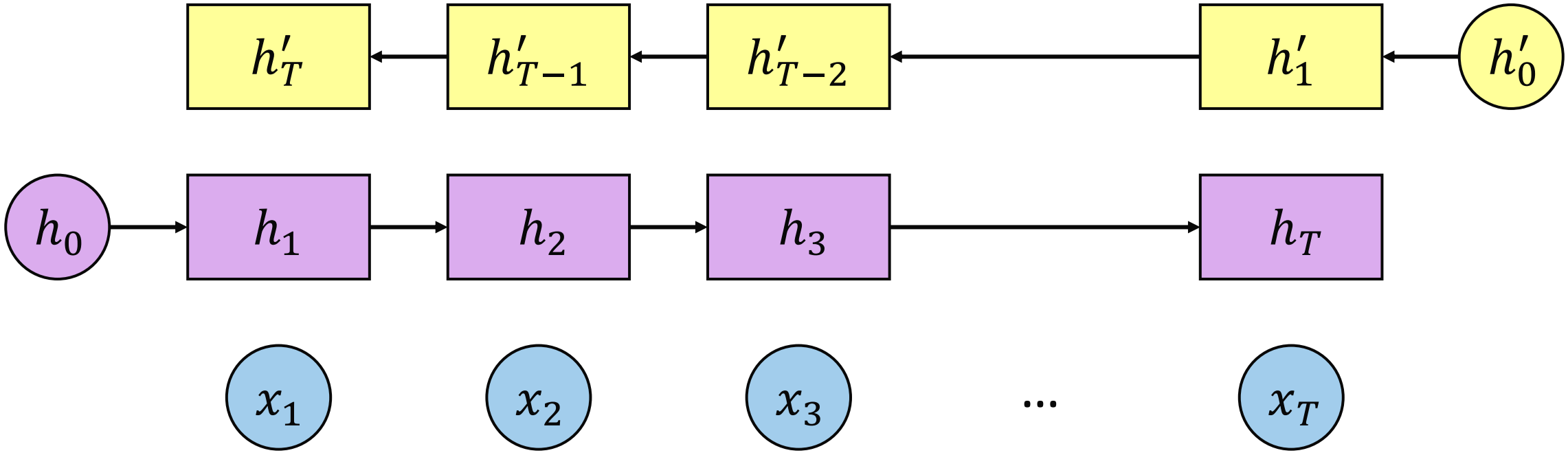
$$z_t = \sigma(\Theta_z \cdot [h_{t-1}, x_t] + b_z)$$

$$r_t = \sigma(\Theta_r \cdot [h_{t-1}, x_t] + b_r)$$

$$\tilde{h}_t = \tanh(\Theta \cdot [r_t \times h_{t-1}, x_t] + b)$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t$$

Bidirectional RNN



Машинный перевод