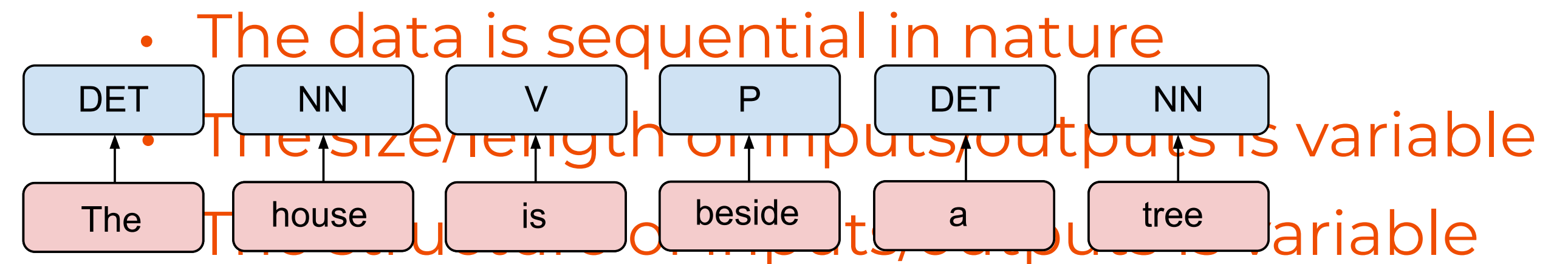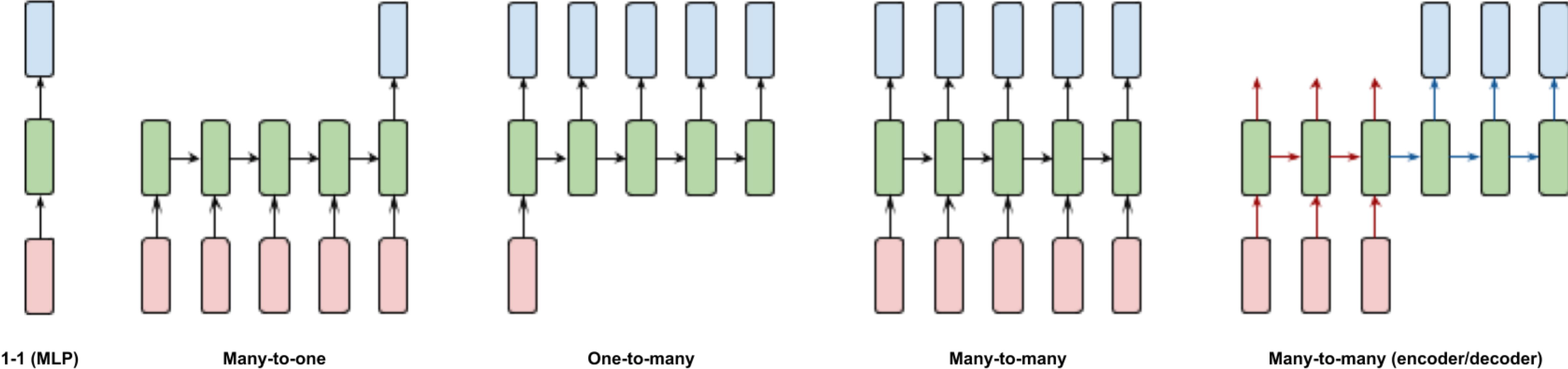# Sequence Modelling Tasks

Typical sequence modelling tasks could include:
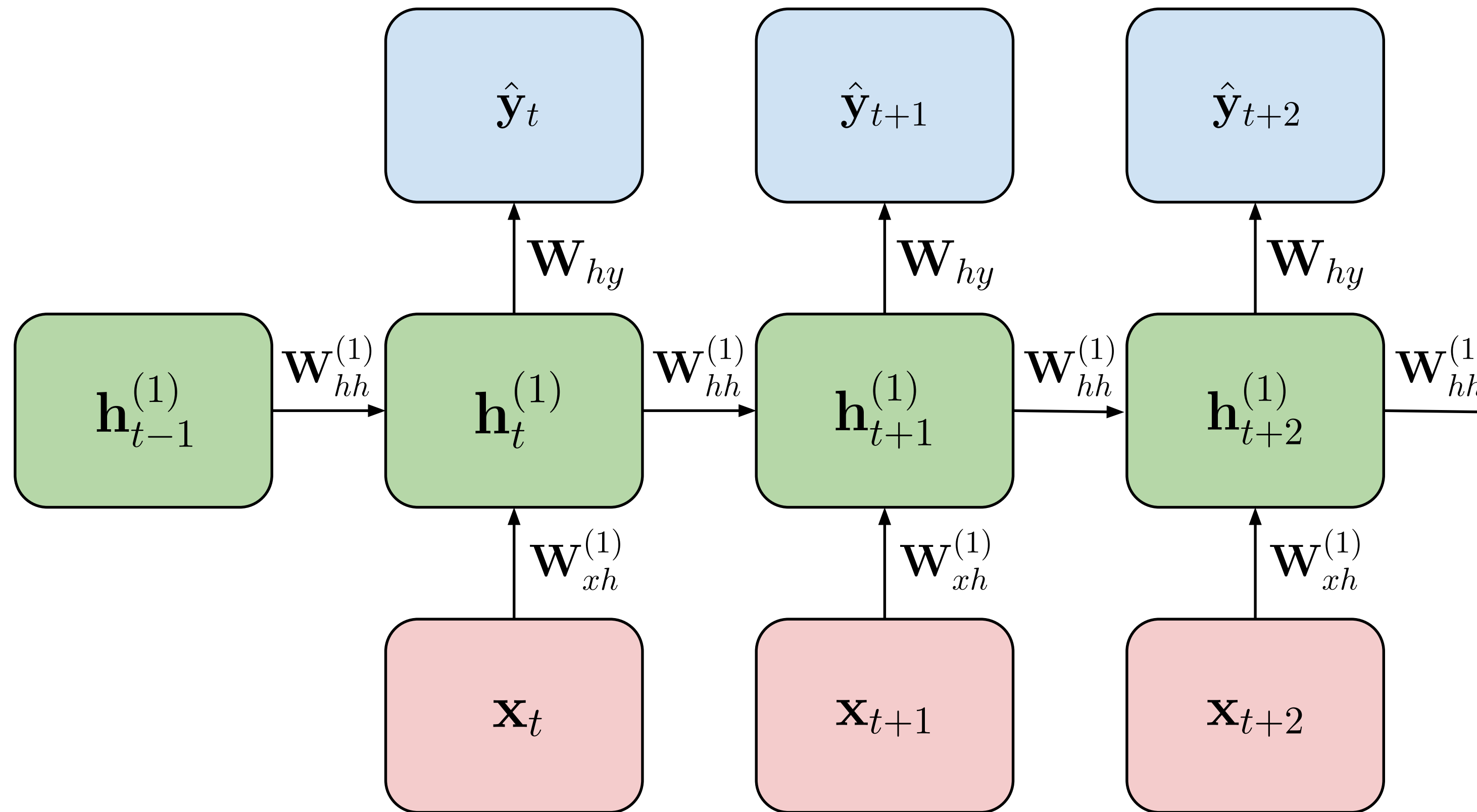
- Text sentiment analysis

- Imagine captioning

- Question answering

- Audio synthesis

- Part-of-speech tagging

- Machine translation

- The data is sequential in nature
- The size/length of inputs/outputs is variable
- The structure of inputs/outputs is variable

DET	NN	V	P	DET	NN
The	house	is	beside	a	tree

Recurrent Neural Network Architectures

1-1 (MLP)   Many-to-one   One-to-many   Many-to-many   Many-to-many (encoder/decoder)
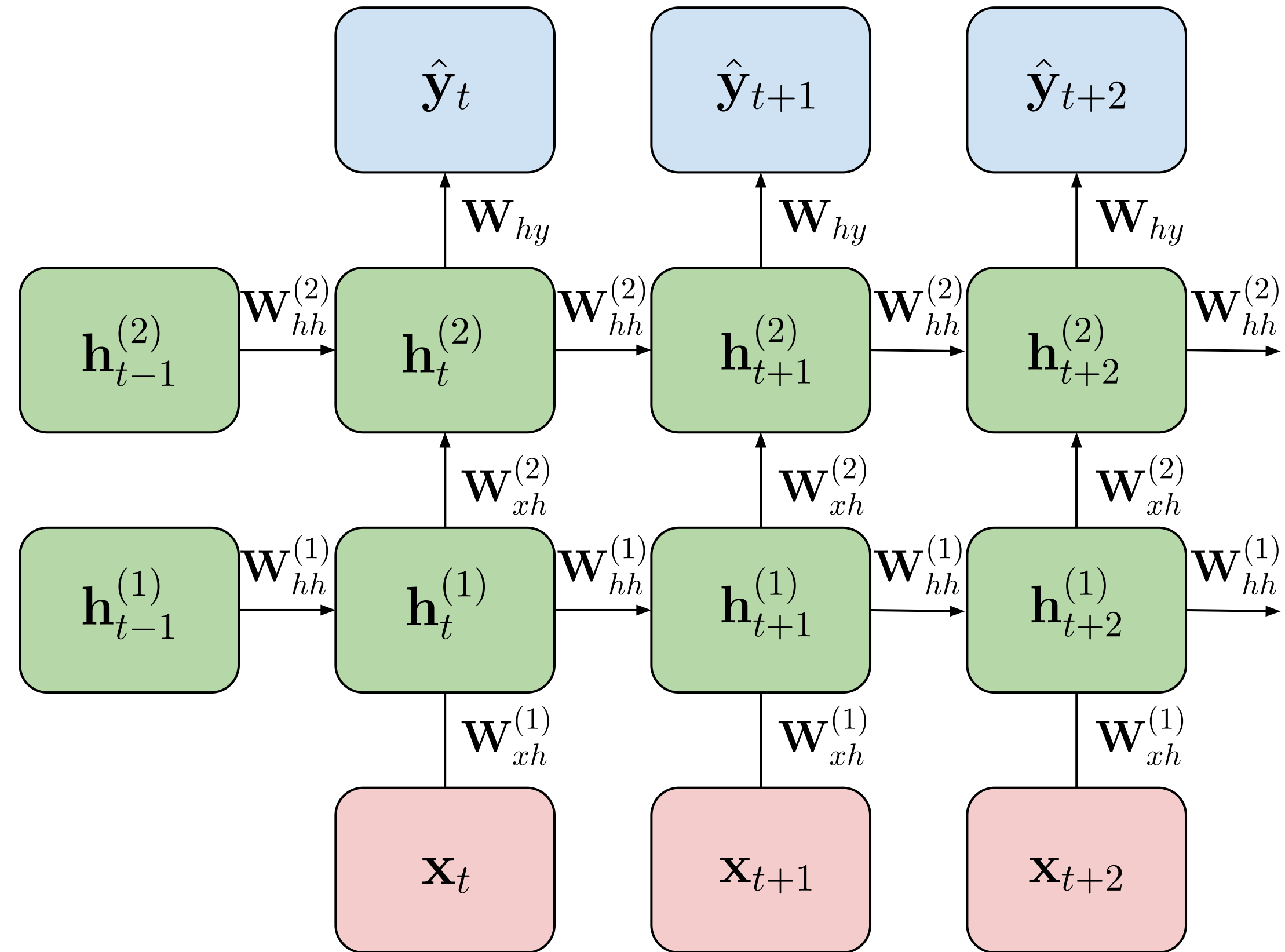
Basic RNN computation

$$\mathbf{h}_t^{(1)} = \sigma \left(\mathbf{W}_{hh}^{(1)} \mathbf{h}_{t-1}^{(1)} + \mathbf{W}_{xh}^{(1)} \mathbf{x}_t + \mathbf{b}_h^{(1)} \right)$$

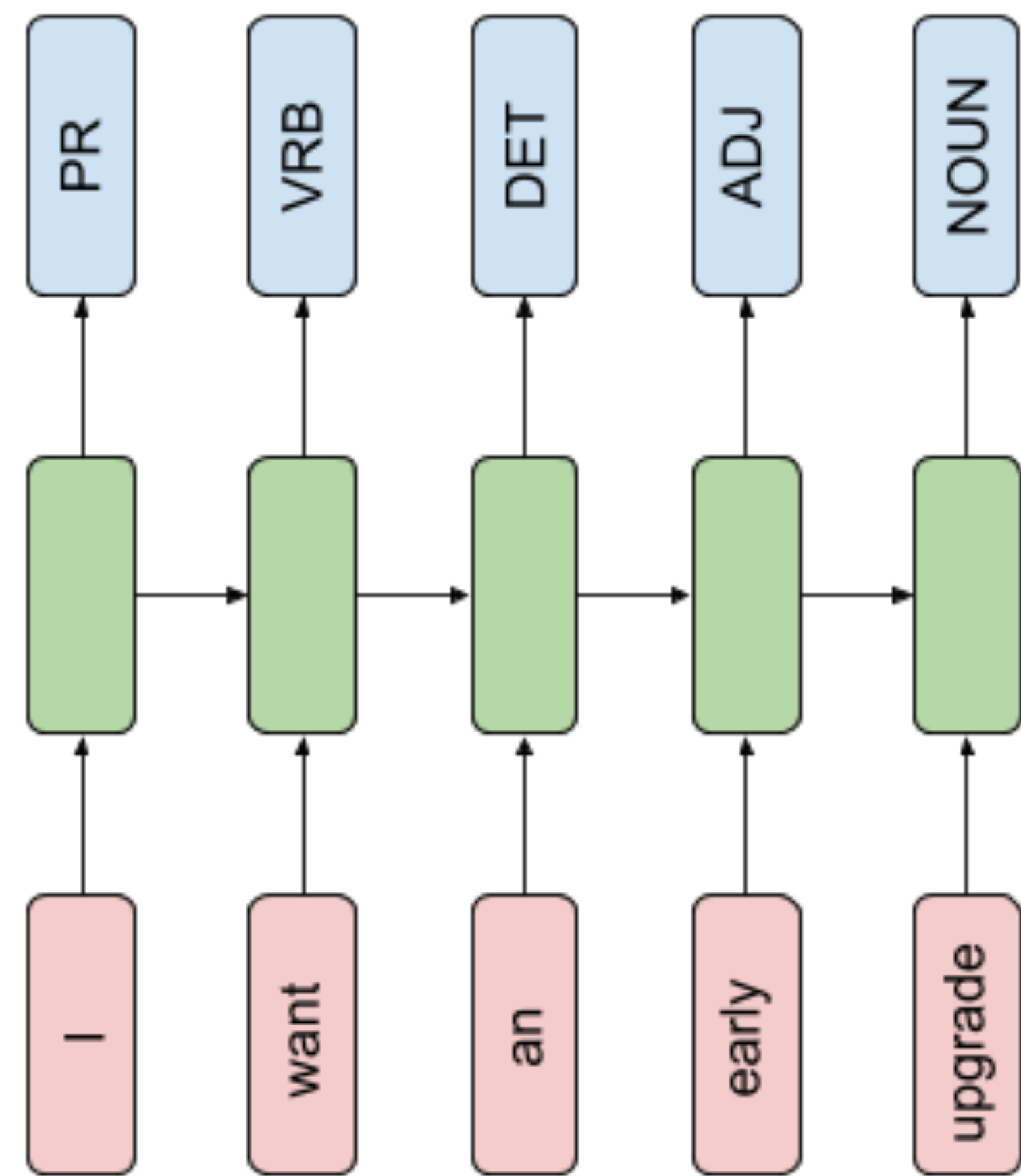$$\hat{\mathbf{y}}_t = \sigma_{out} \left(\mathbf{W}_{hy} \mathbf{h}^{(1)} + \mathbf{b}_y \right)$$

Stacked RNN
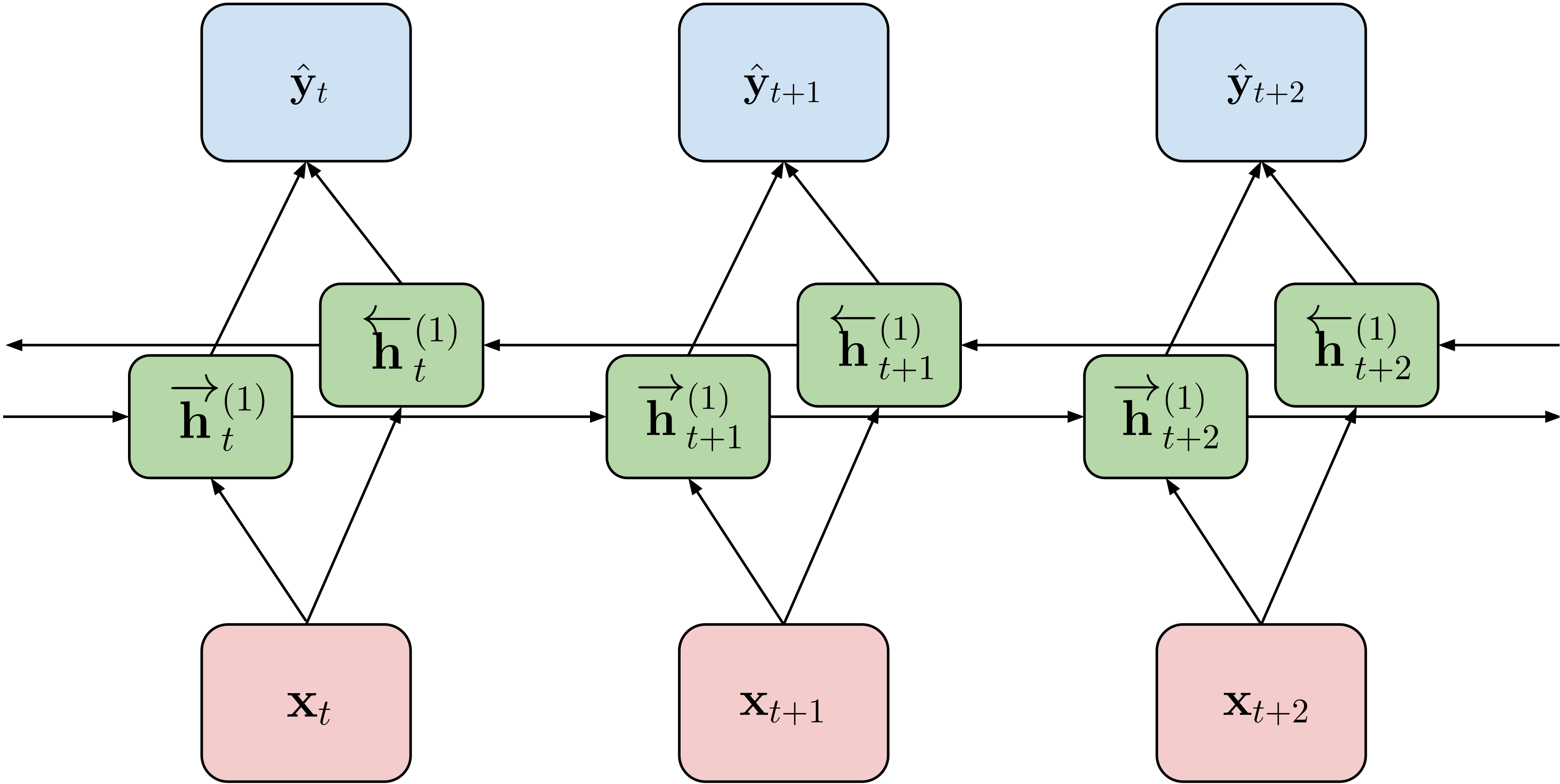

$$\mathbf{h}_t^{(k)} = \sigma\left(\mathbf{W}_{hh}^{(k)}\mathbf{h}_{t-1}^{(k)} + \mathbf{W}_{xh}^{(k)}\mathbf{h}_t^{(k-1)} + \mathbf{b}_h^{(k)}\right), \quad k = 1, \ldots, L$$

$$\hat{\mathbf{y}}_t = \sigma_{out}\left(\mathbf{W}_{hy}\mathbf{h}^{(L)} + \mathbf{b}_y\right)$$

Bidirectional RNN
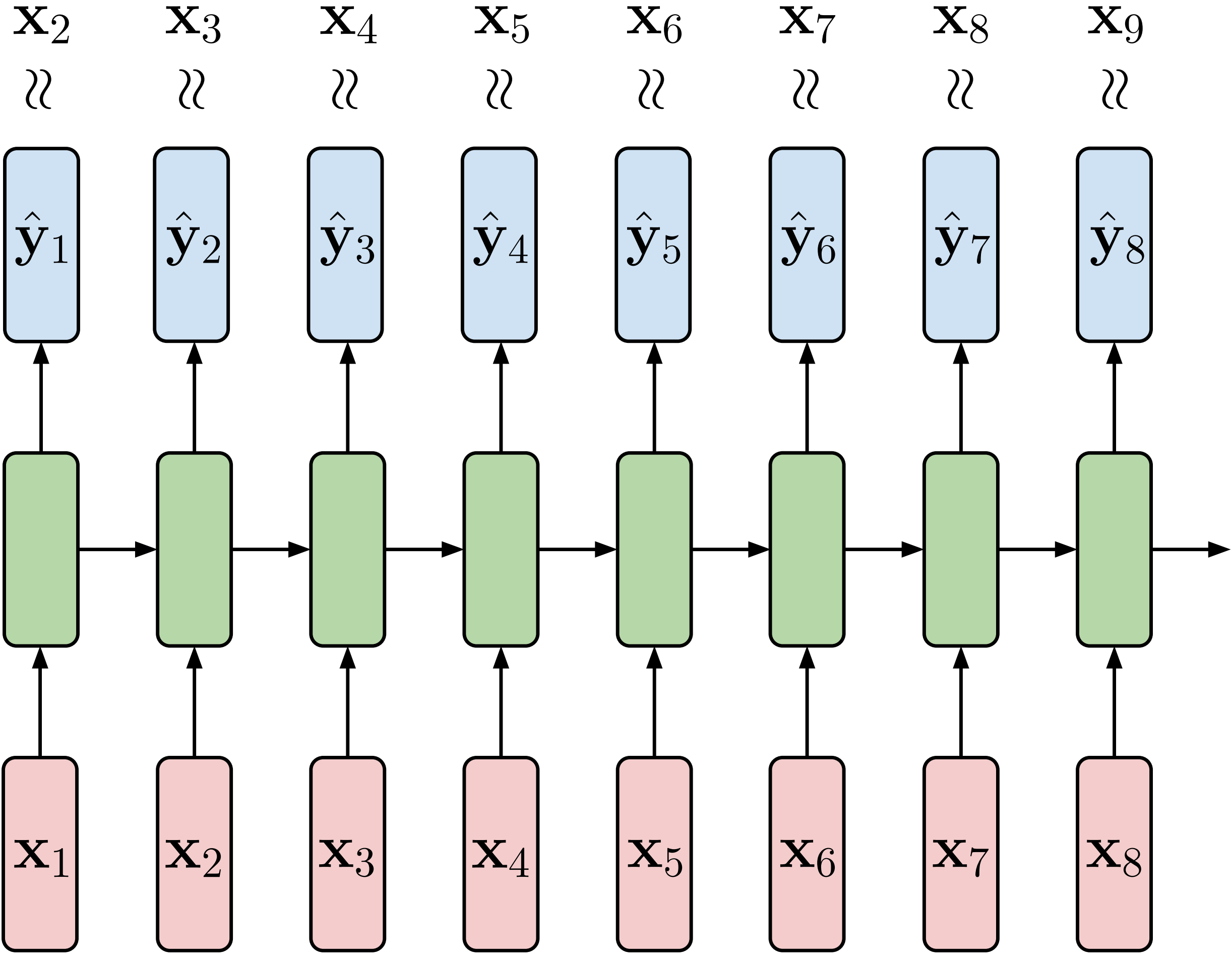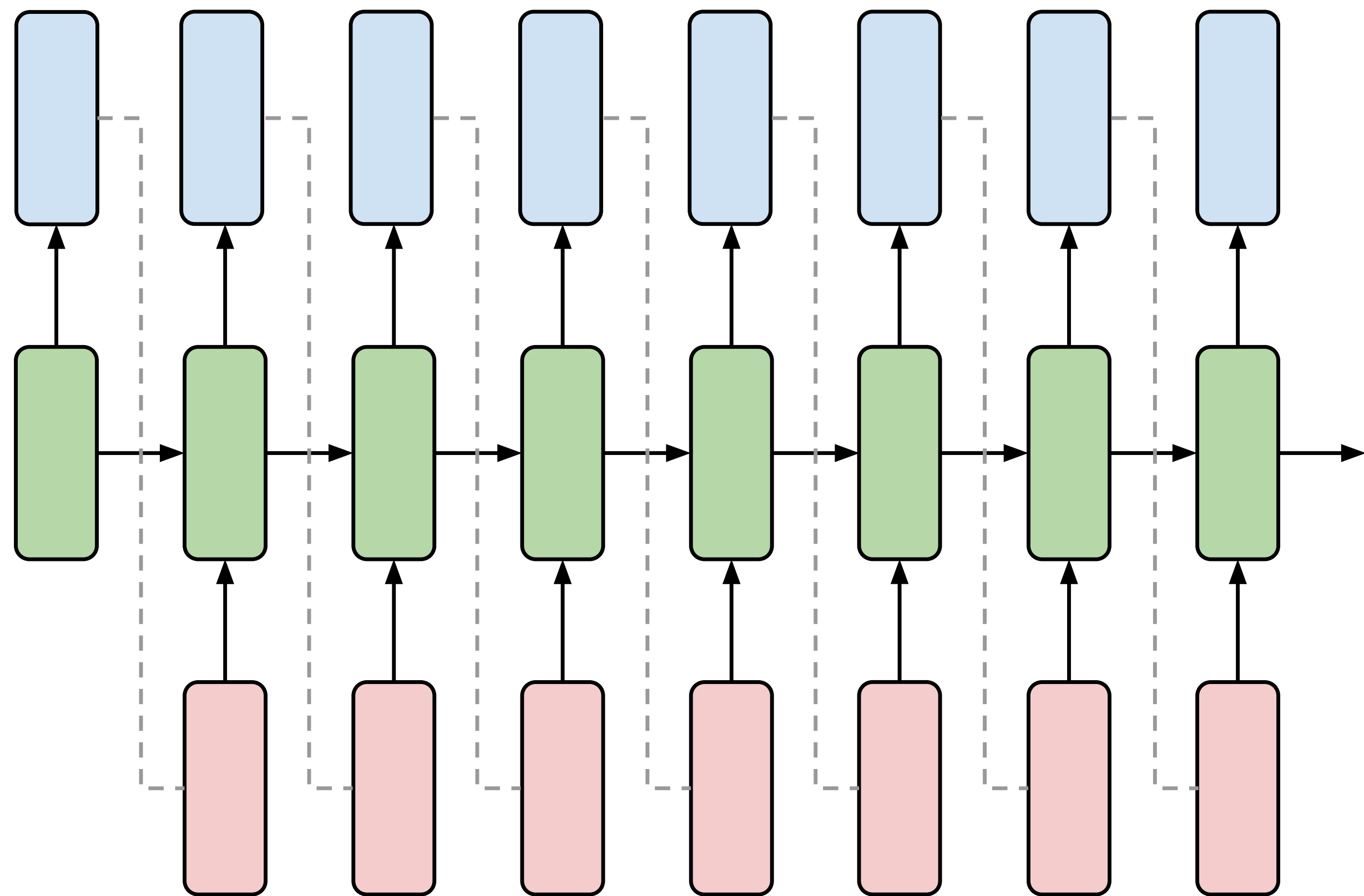
Bidirectional RNN

Schuster, M. & Paliwal, K. K. (1997), "Bidirectional Recurrent Neural Networks", *IEEE Transactions on Signal Processing*, **45** (11), 2673-2681.

Training RNNs: Generative Models
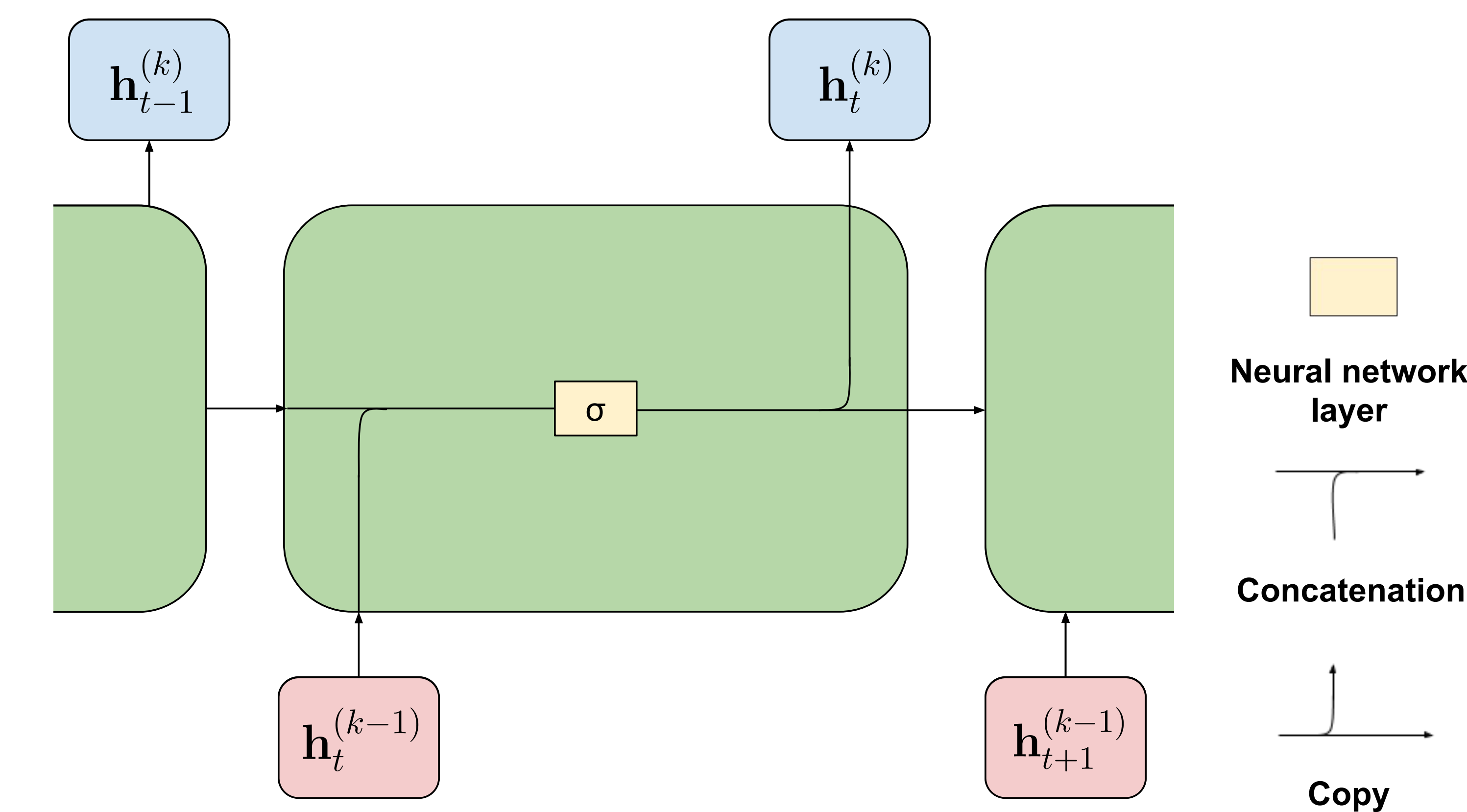
Training RNNs: Generative Models

Long Short Term Memory (LSTM)

The LSTM is a more sophisticated type of recurrent cell.

- Typical RNN implementations use the LSTM
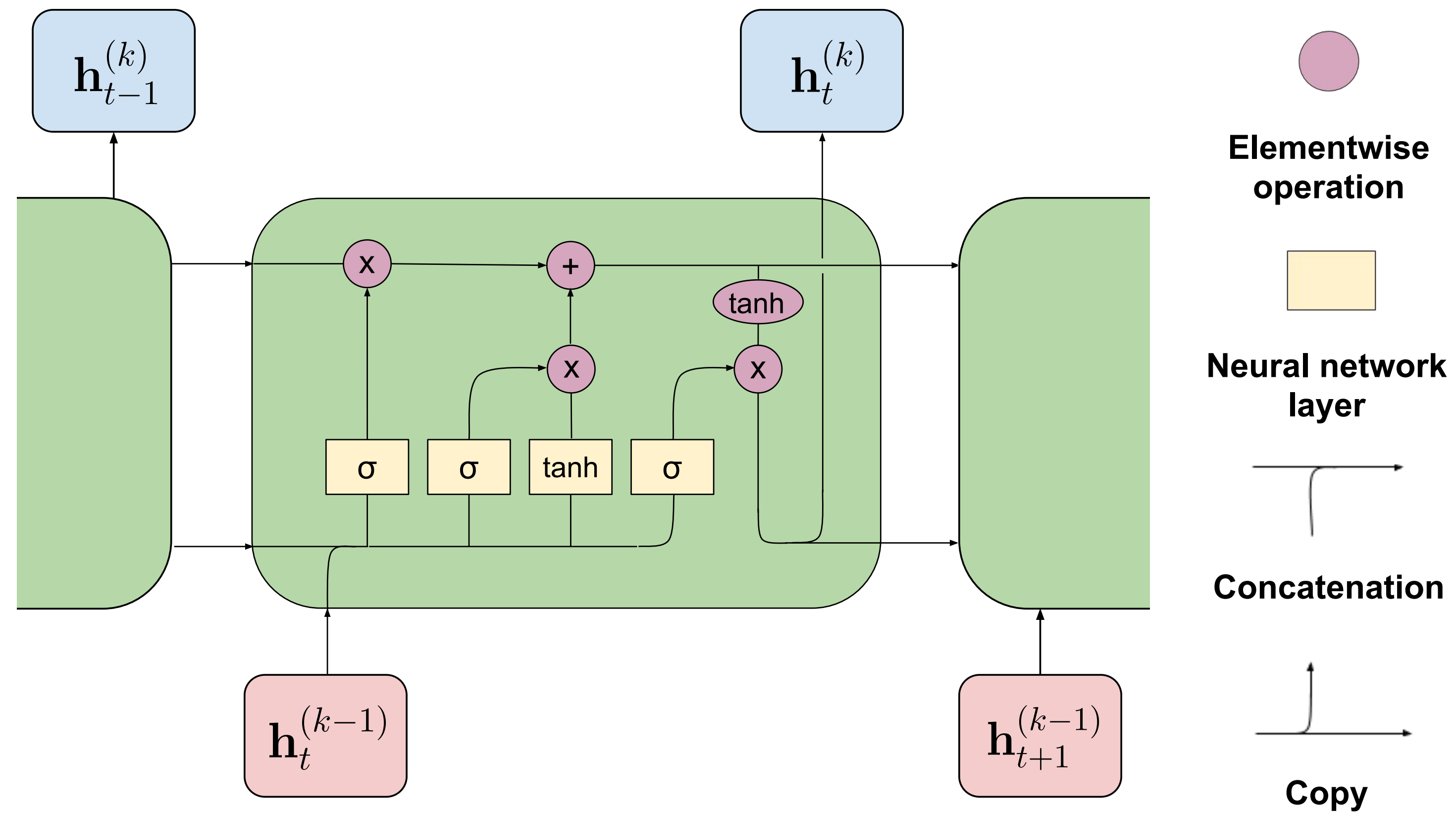
- Mitigates the problem of vanishing gradients

- The LSTM has a 'memory' cell to retain information over many time steps

- Uses a gating system to allow information in and out of the cell state

Basic RNN

$$\mathbf{h}_t^{(k)} = \sigma \left(\mathbf{W}_{hh}^{(k)} \cdot [\mathbf{h}_{t-1}^{(k)}, \mathbf{h}_t^{(k-1)}] + \mathbf{b}_h^{(k)} \right), \; k = 1, \ldots, L$$

LSTM

Hochreiter, S. and Schmidhuber, J. (1997), "Long short-term memory", *Neural Computation*, **9** (8), 1735–1780

LSTM
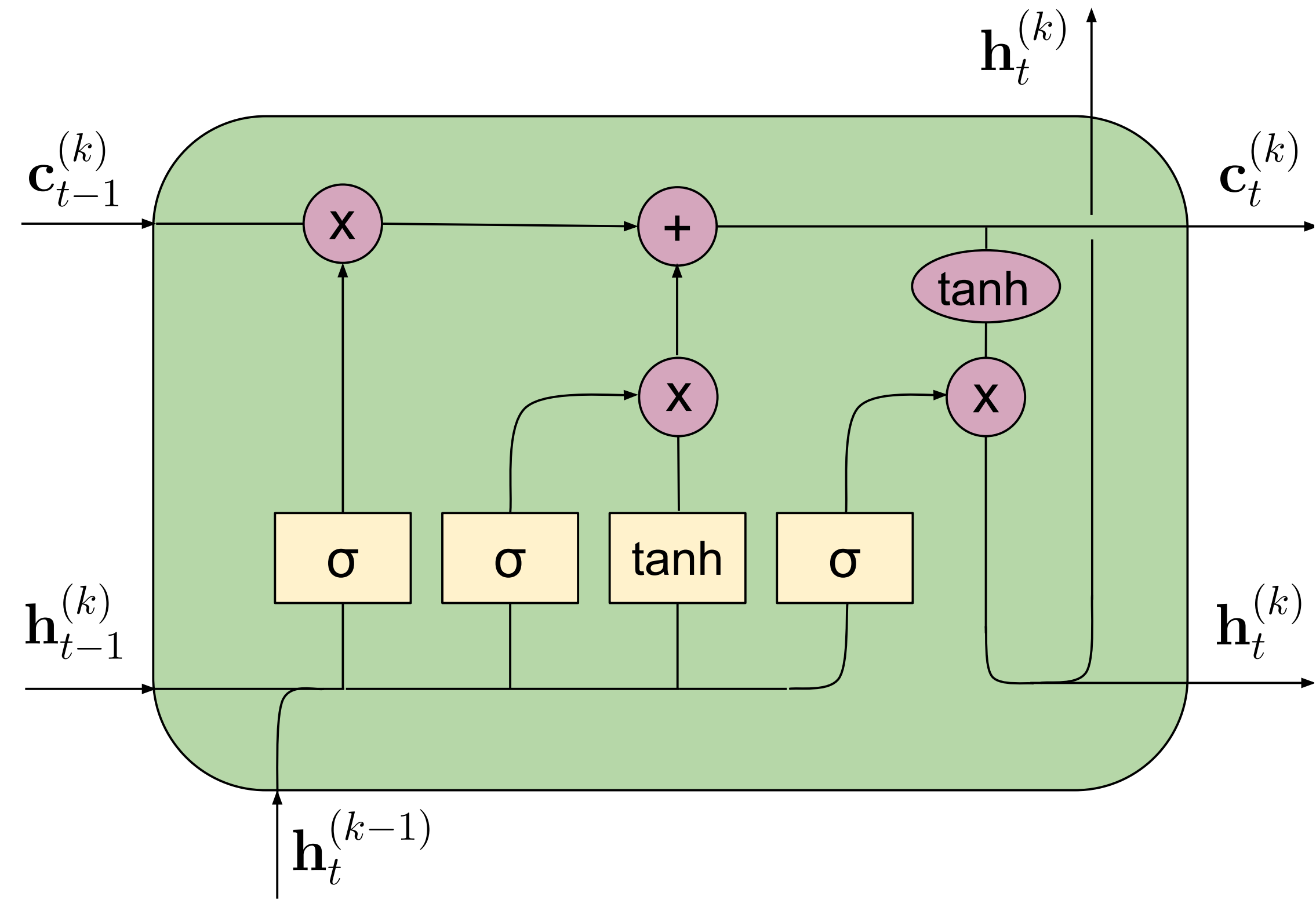
LSTM: Forget Gate

$$\mathbf{f}_t^{(k)} = \sigma \left(\mathbf{W}_f^{(k)} \cdot [\mathbf{h}_t^{(k-1)}, \mathbf{h}_{t-1}^{(k)}] + \mathbf{b}_f^{(k)} \right)$$

$$\mathbf{c}_t^{(k)} \leftarrow \mathbf{f}_t^{(k)} \odot \mathbf{c}_{t-1}^{(k)}$$

LSTM: Input and Content Gates

$$\tilde{\mathbf{c}}_t^{(k)} = \tanh\left(\mathbf{W}_c^{(k)} \cdot [\mathbf{h}_t^{(k-1)}, \mathbf{h}_{t-1}^{(k)}] + \mathbf{b}_c^{(k)}\right)$$
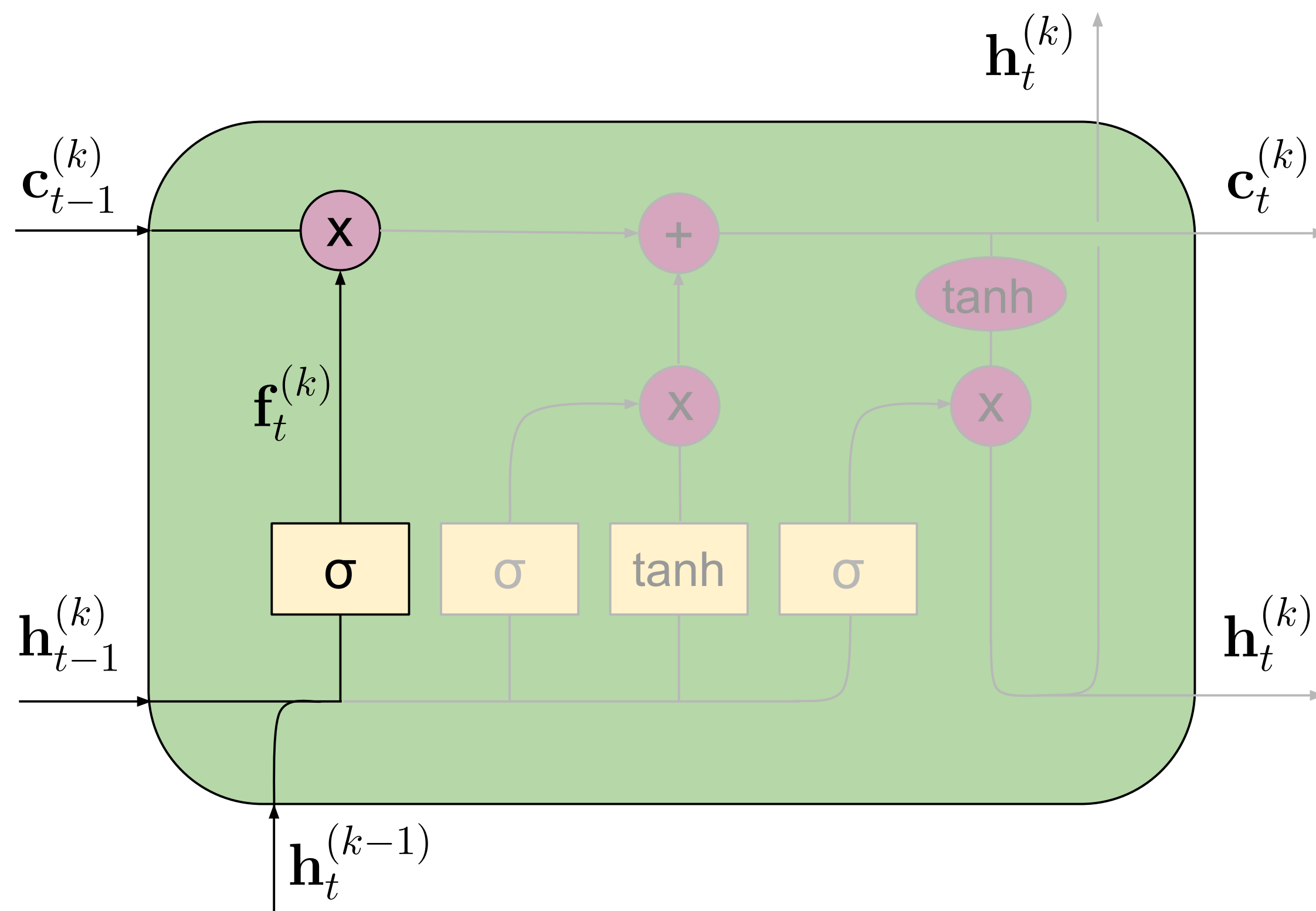
$$\mathbf{i}_t^{(k)} = \sigma\left(\mathbf{W}_i^{(k)} \cdot [\mathbf{h}_t^{(k-1))}, \mathbf{h}_{t-1}^{(k)}] + \mathbf{b}_i^{(k)}\right)$$

$$\mathbf{c}_t^{(k)} \leftarrow \mathbf{c}_t^{(k)} + \mathbf{i}_t^{(k)} \odot \tilde{\mathbf{c}}_t^{(k)}$$

LSTM: Output Gate

$$\mathbf{o}_t^{(k)} = \sigma\left(\mathbf{W}_o^{(k)} \cdot [\mathbf{h}_t^{(k-1)}, \mathbf{h}_{t-1}^{(k)}] + \mathbf{b}_o^{(k)}\right)$$

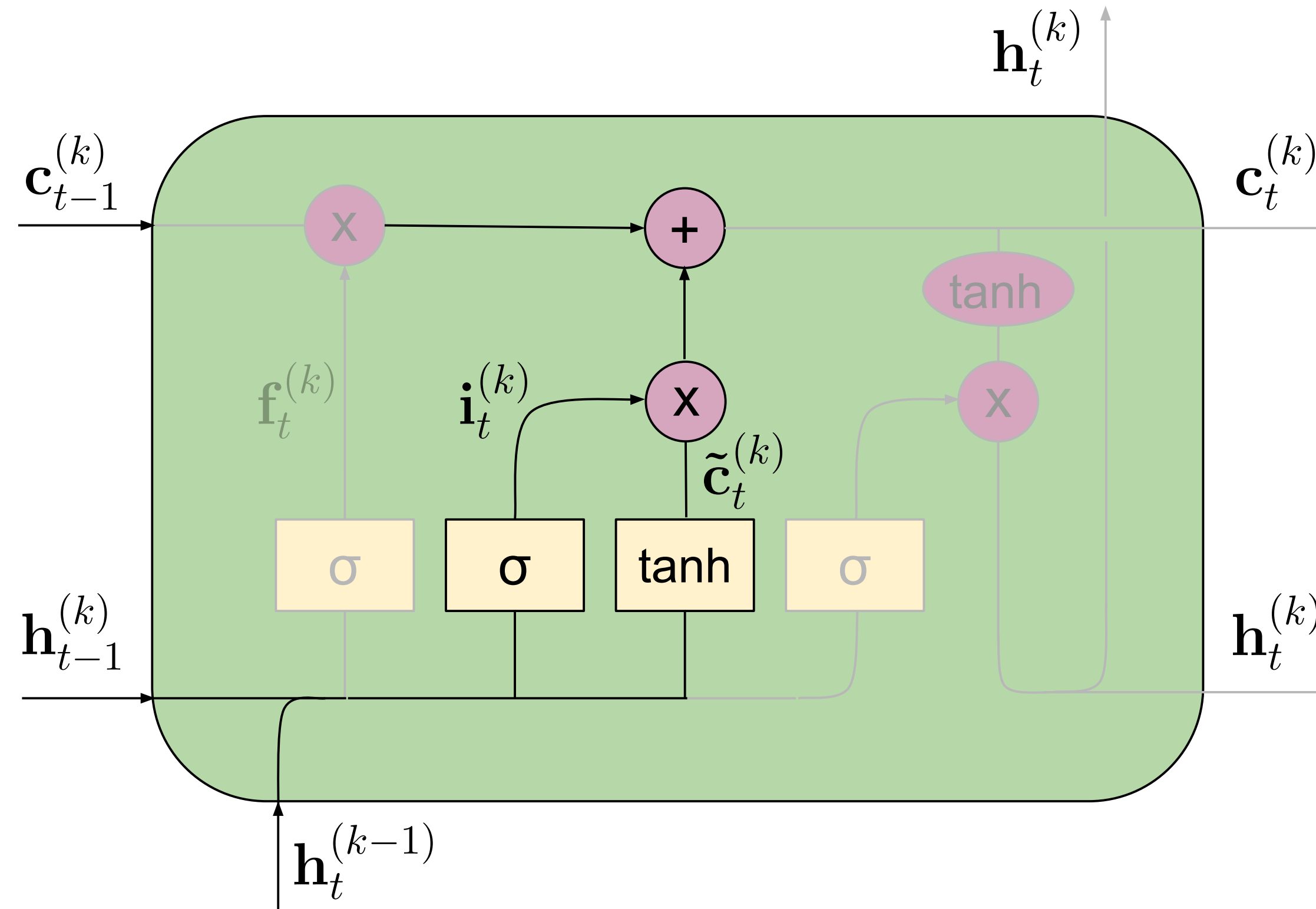$$\mathbf{h}_t^{(k)} = \mathbf{o}_t^{(k)} \odot \tanh(\mathbf{c}_t^{(k)})$$

LSTM

$$\mathbf{f}_t^{(k)} = \sigma\left(\mathbf{W}_f^{(k)} \cdot [\mathbf{h}_t^{(k-1)}, \mathbf{h}_{t-1}^{(k)}] + \mathbf{b}_f^{(k)}\right)$$

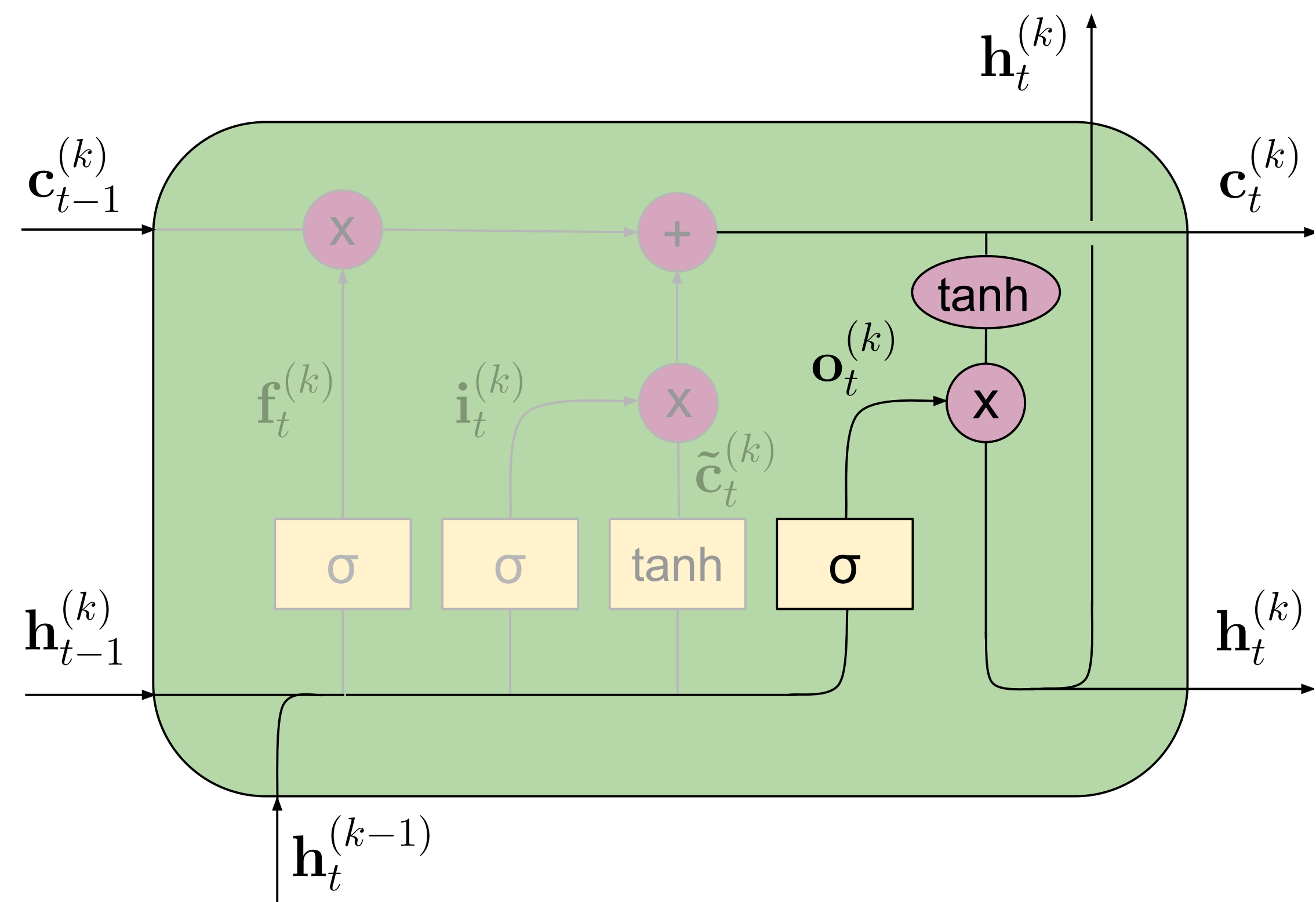$$\mathbf{c}_t^{(k)} \leftarrow \mathbf{f}_t^{(k)} \odot \mathbf{c}_{t-1}^{(k)}$$

$$\tilde{\mathbf{c}}_t^{(k)} = \tanh\left(\mathbf{W}_c^{(k)} \cdot [\mathbf{h}_t^{(k-1)}, \mathbf{h}_{t-1}^{(k)}] + \mathbf{b}_c^{(k)}\right)$$

$$\mathbf{i}_t^{(k)} = \sigma\left(\mathbf{W}_i^{(k)} \cdot [\mathbf{h}_t^{(k-1))}, \mathbf{h}_{t-1}^{(k)}] + \mathbf{b}_i^{(k)}\right)$$

$$\mathbf{c}_t^{(k)} \leftarrow \mathbf{c}_t^{(k)} + \mathbf{i}_t^{(k)} \odot \tilde{\mathbf{c}}_t^{(k)}$$

$$\mathbf{o}_t^{(k)} = \sigma\left(\mathbf{W}_o^{(k)} \cdot [\mathbf{h}_t^{(k-1)}, \mathbf{h}_{t-1}^{(k)}] + \mathbf{b}_o^{(k)}\right)$$

$$\mathbf{h}_t^{(k)} = \mathbf{o}_t^{(k)} \odot \tanh(\mathbf{c}_t^{(k)})$$