

LSTMs & GRUs

COSC 410: Applied Machine Learning

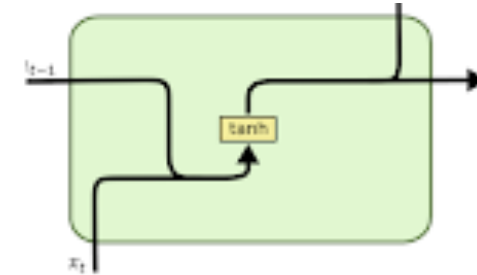
Spring 2022

Prof. Apthorpe

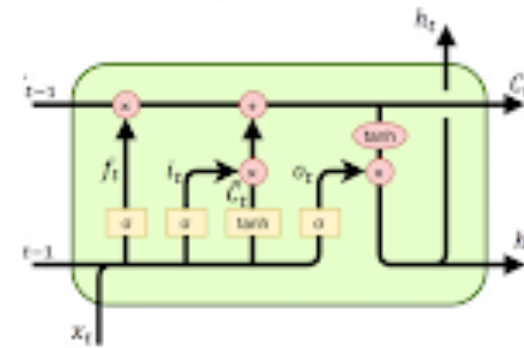
Outline

- RNN Memory Problem
- LSTM Cells
- GRU Cells
- Stacked RNNs
- Bidirectional RNNs
- RNNs in Keras

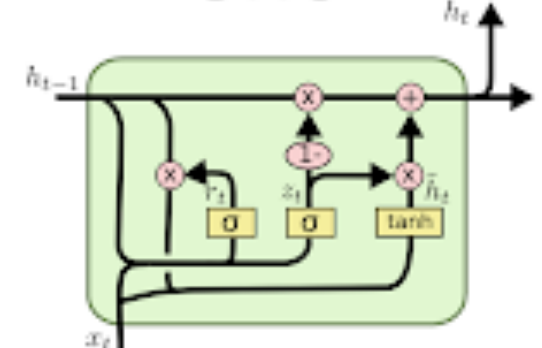
Simple RNN



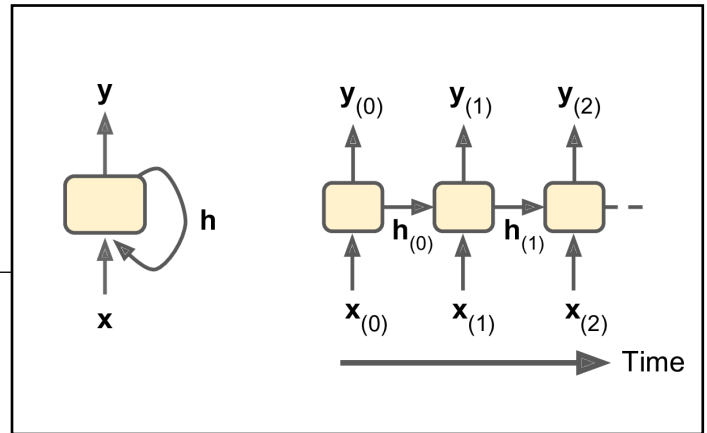
LSTM



GRU



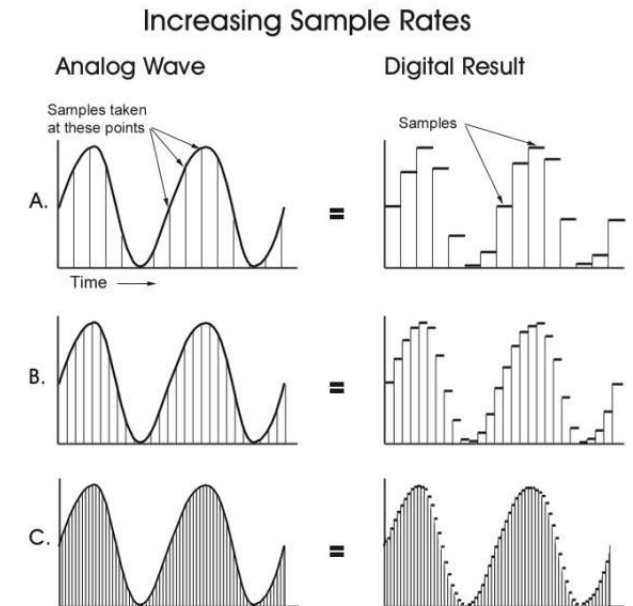
Challenges of Simple RNNs



- **Limited memory**

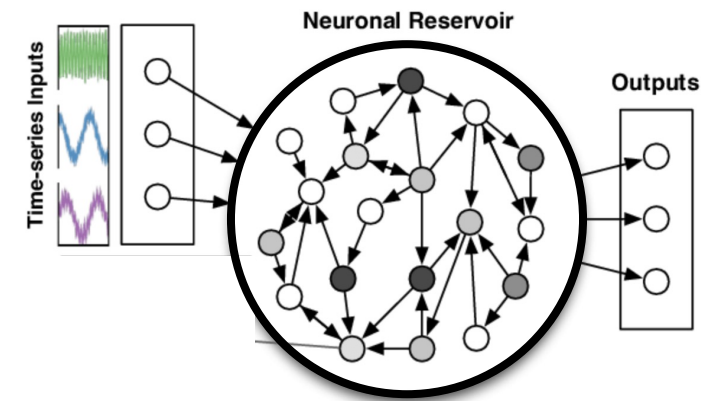
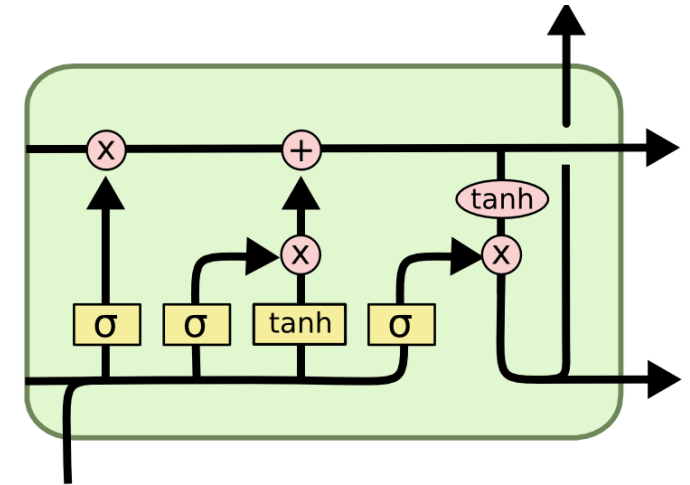
- Input at step t only affects output up to step $t+n$ **for small n**
- Influence of previous inputs **decays** over successive steps

- Long sequences (with many steps)
need networks with longer memory to model
 - High-fidelity audio, long text documents, etc.



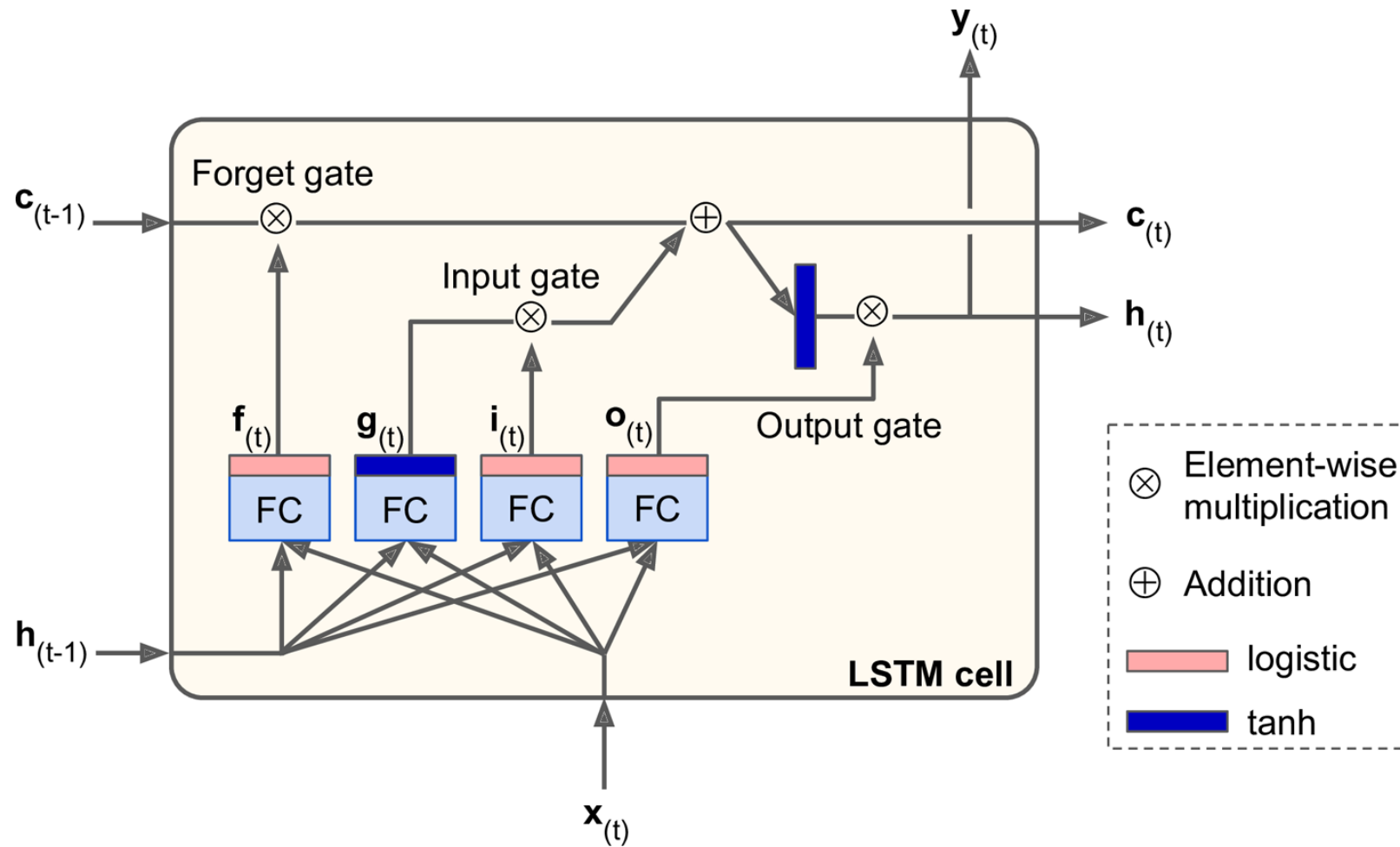
Challenges of Simple RNNs

- Techniques to increase RNN memory
 - **Hand-designed recurrent “cells”**
 - **Reservoir networks**



Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM)



Long Short-Term Memory (LSTM)

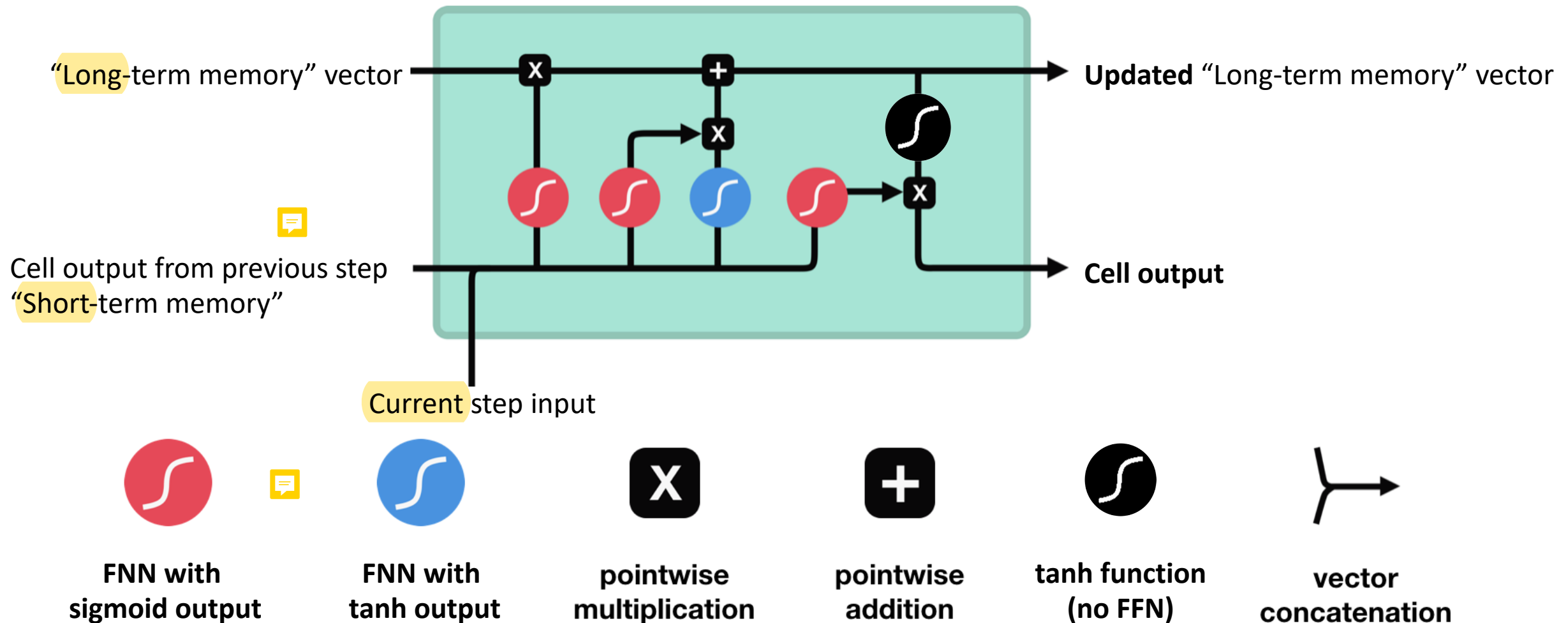
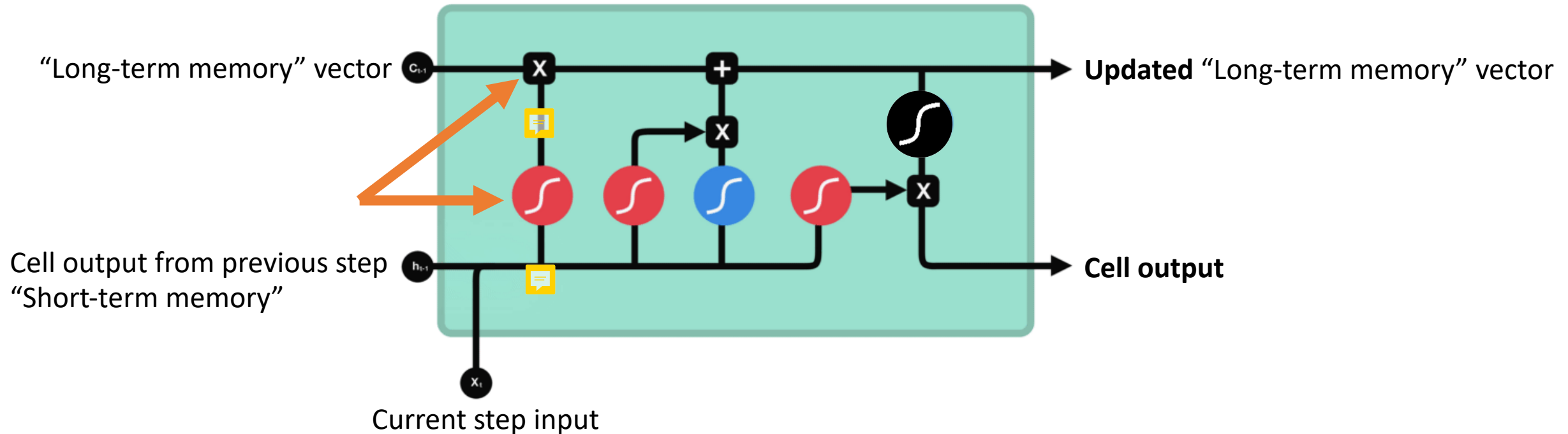


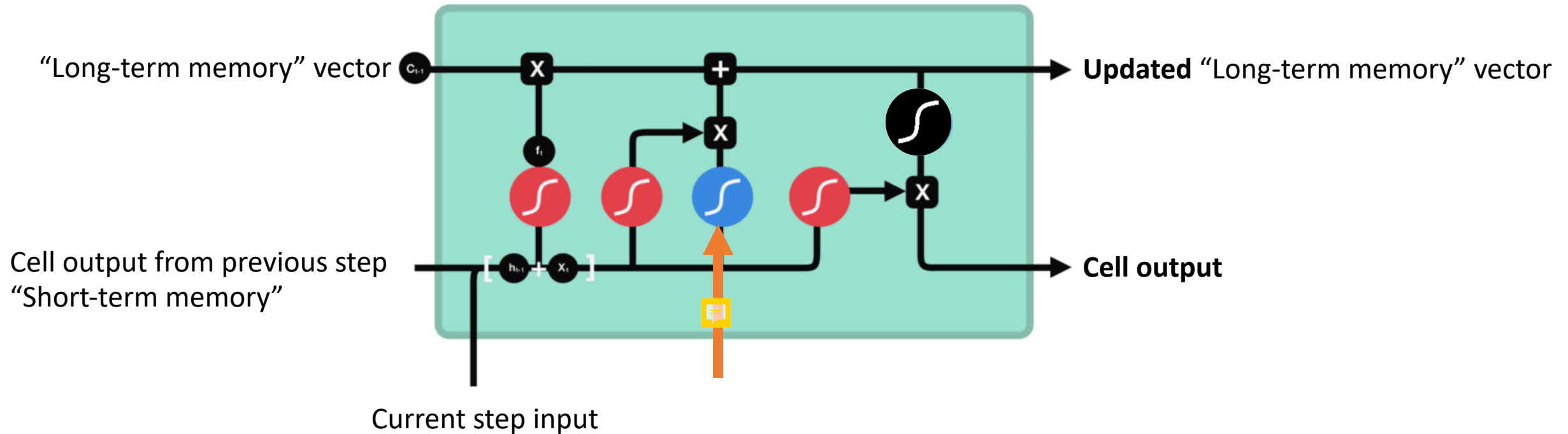
Figure & animations from Michael Phi

Long Short-Term Memory (LSTM)



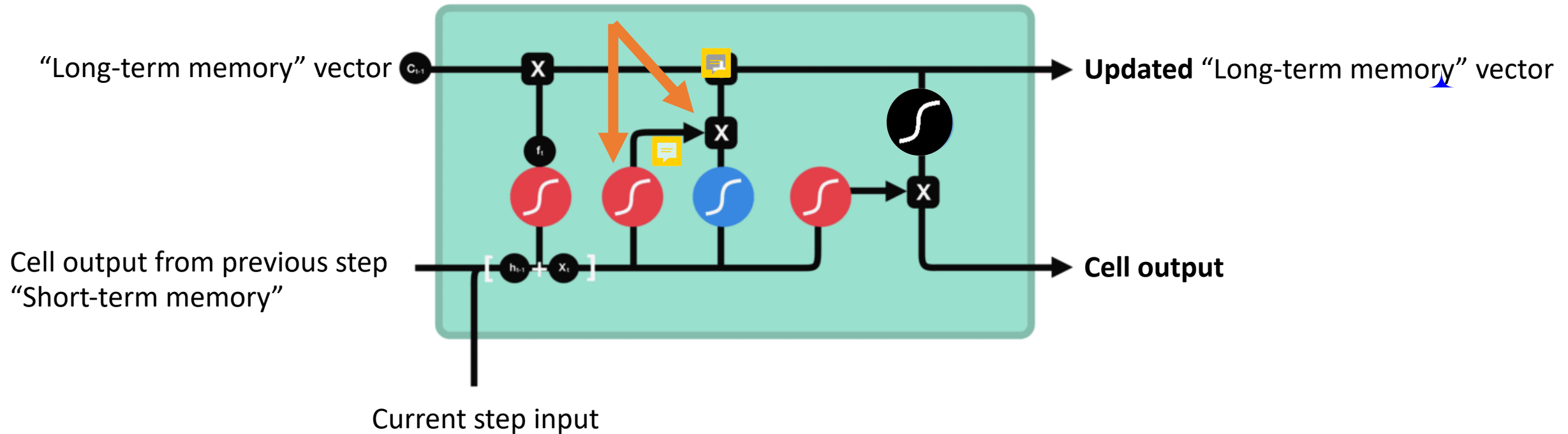
- **“Forget” network & gate** determine which indices of long-term memory to keep
 - Output of sigmoid activation acts like a Boolean mask on the long-term memory vector

Long Short-Term Memory (LSTM)



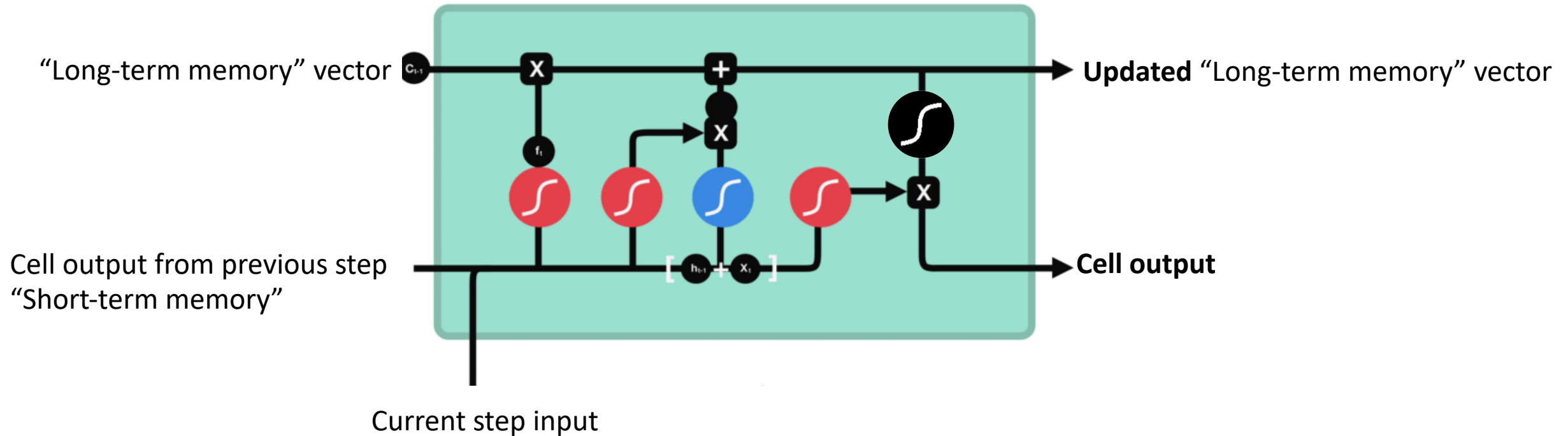
- **“Main” network** processes current step & short-term memory

Long Short-Term Memory (LSTM)



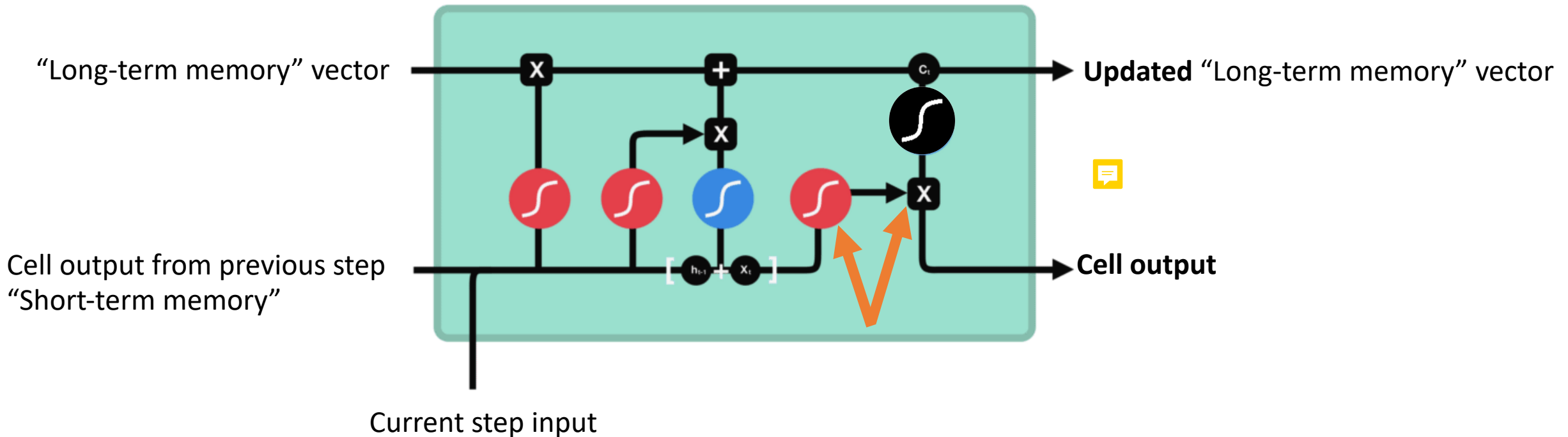
- **"Input" network & gate** determine which indices of main network output should be added to long-term memory

Long Short-Term Memory (LSTM)



- **Update** long-term memory vector

Long Short-Term Memory (LSTM)

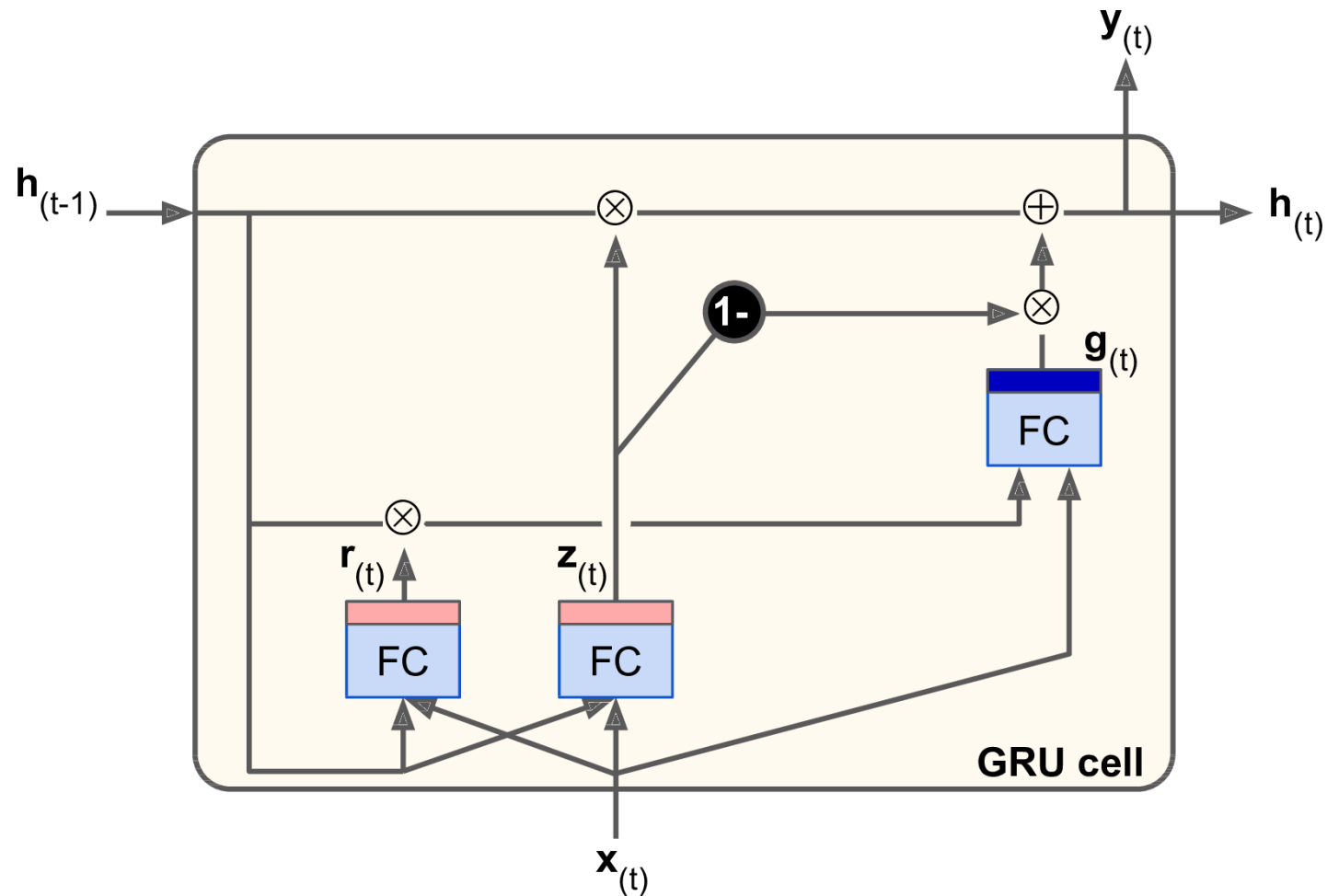


- **"Output" network & gate** determine which indices of updated long-term memory vector should be the cell's primary output → used as the **current step prediction label** and/or the **new** short-term memory vector



Gated Recurrent Units (GRU)

Gated Recurrent Units (GRU)



Gated Recurrent Units (GRU)

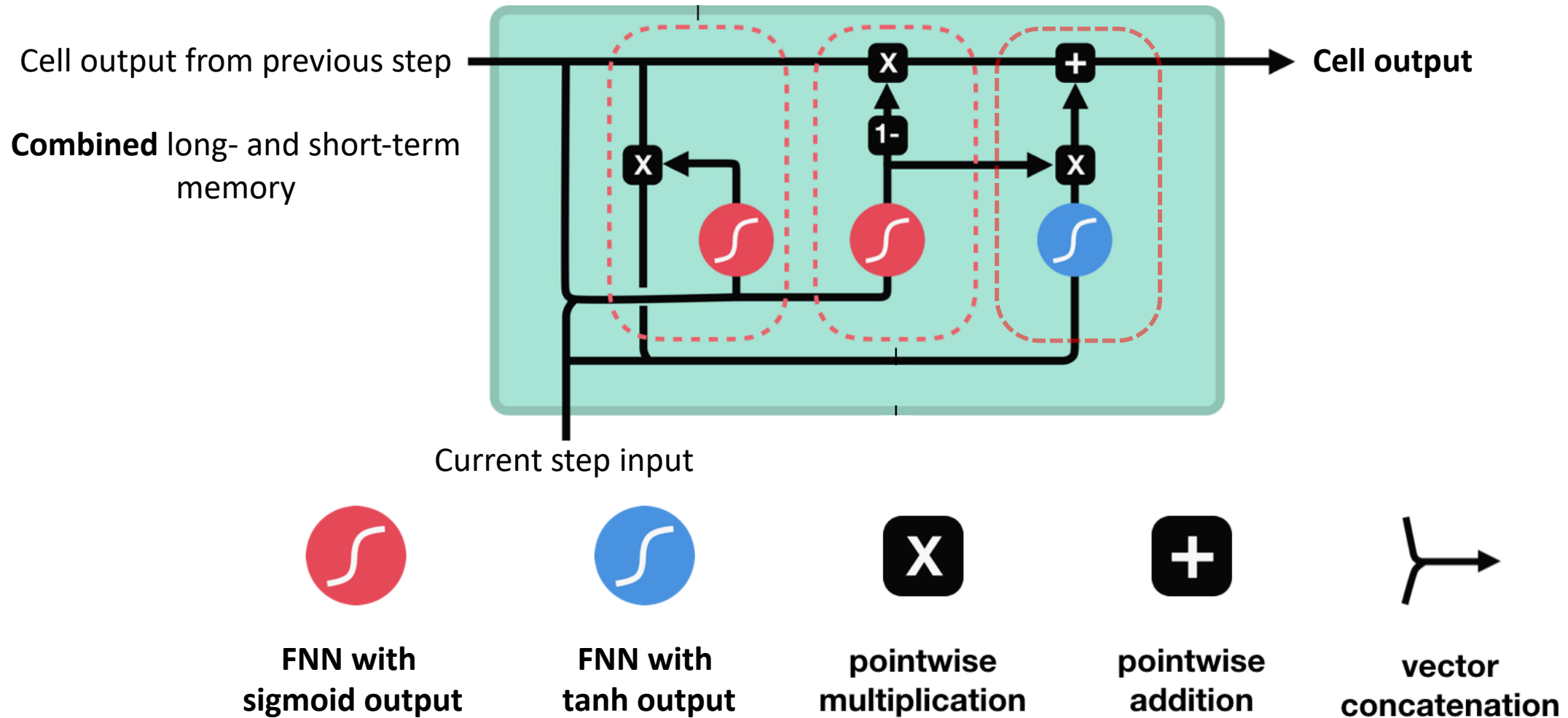
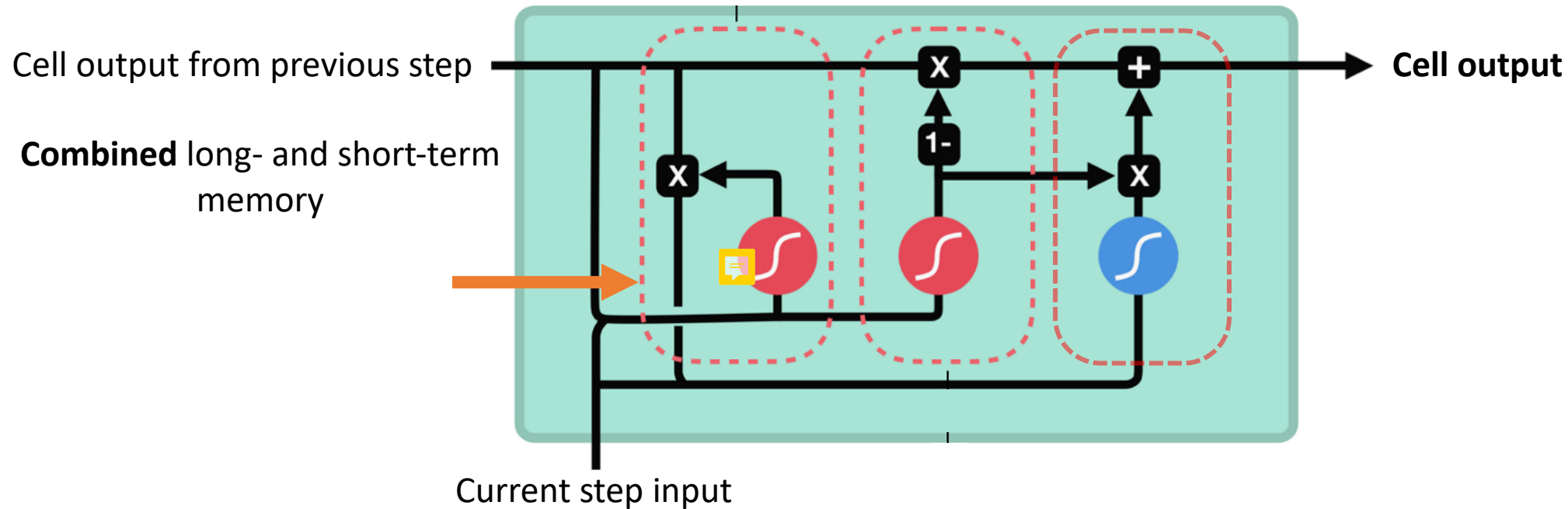


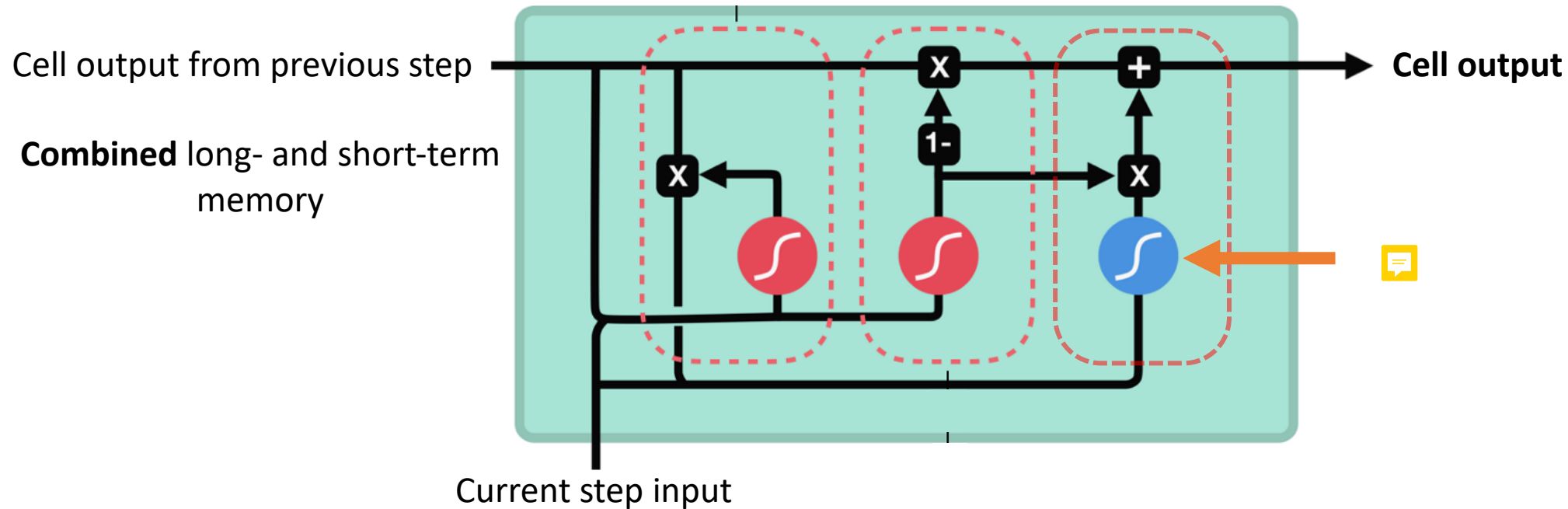
Figure & animations from Michael Phi

Gated Recurrent Units (GRU)



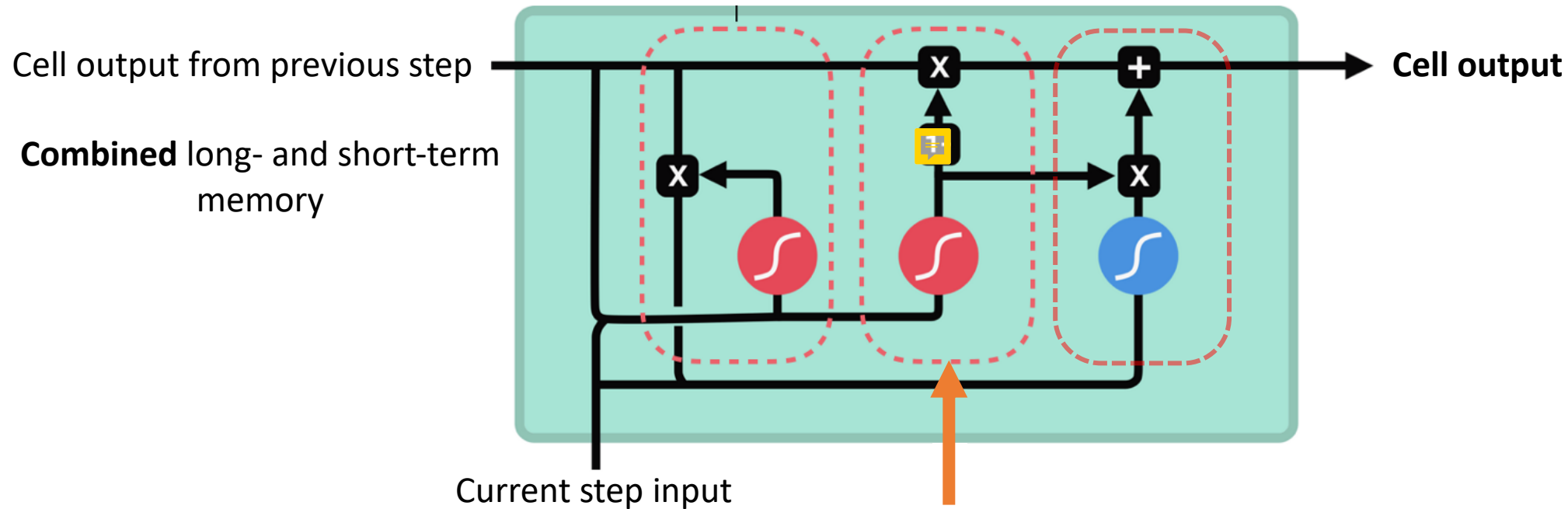
- **“Reset” network & gate** determine which indices of combined memory vector should be included in main network input

Gated Recurrent Units (GRU)



- **“Main” network** processes current step & selected memory values

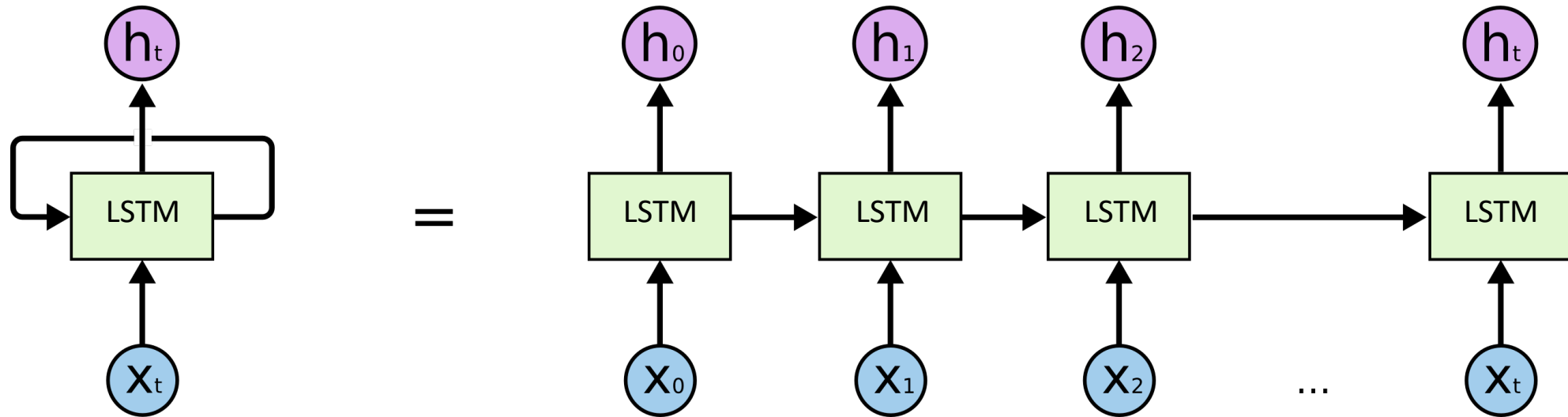
Gated Recurrent Units (GRU)



- **“Update” network & gate** determine which indices of main network output should be stored in combined memory vector. The previous memory values at those indices are cleared

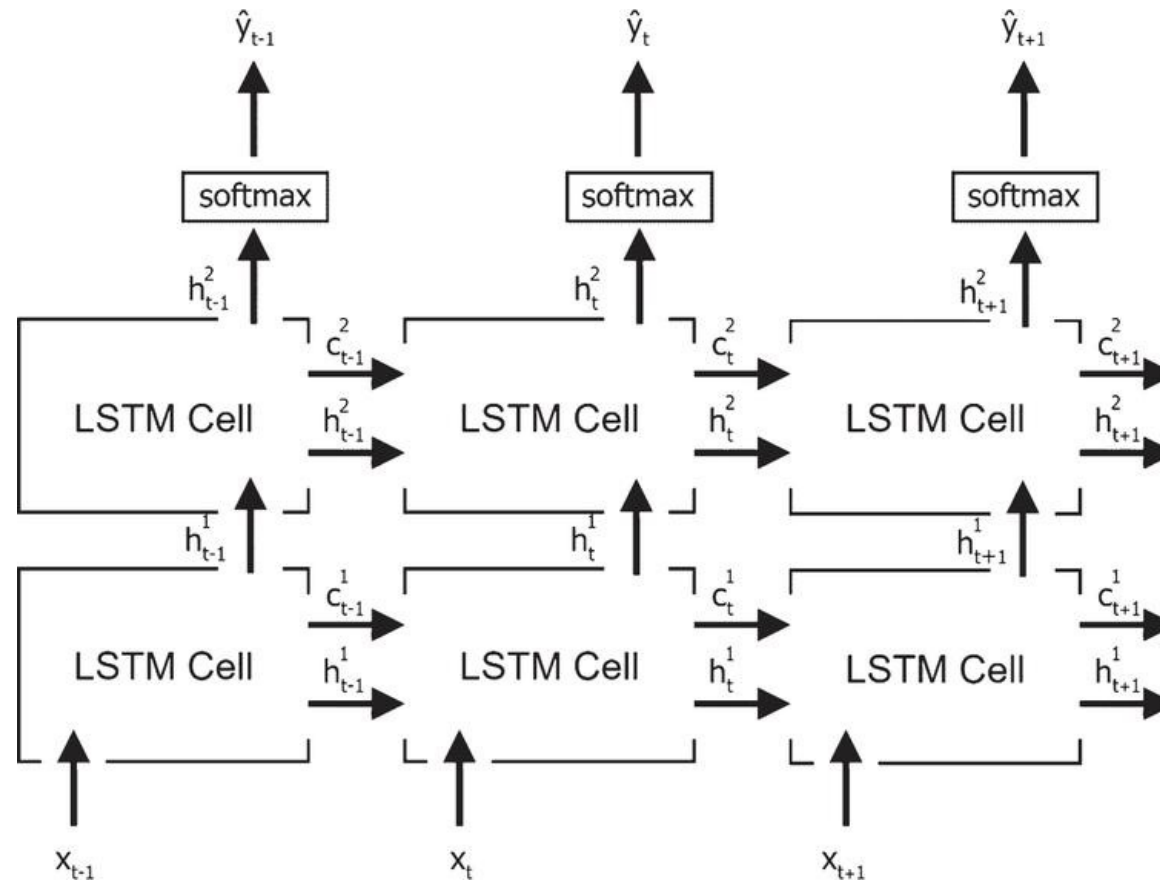
RNN Architectures

Single-Unit LSTM or GRU Network



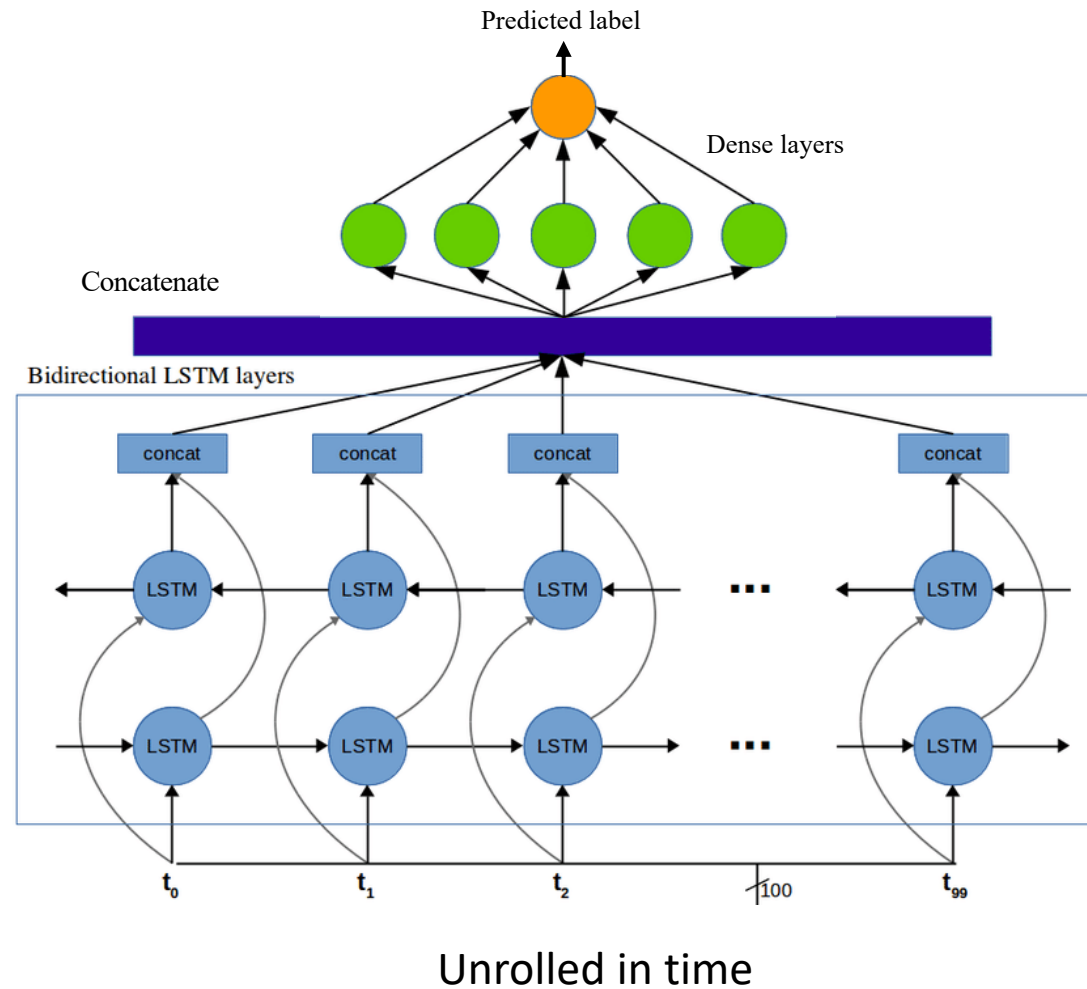
Unrolled in time

Stacked LSTMs & GRUs



Unrolled in time

Bidirectional LSTMs & GRUs



RNNs in Keras

RNN.ipynb

Questions?
