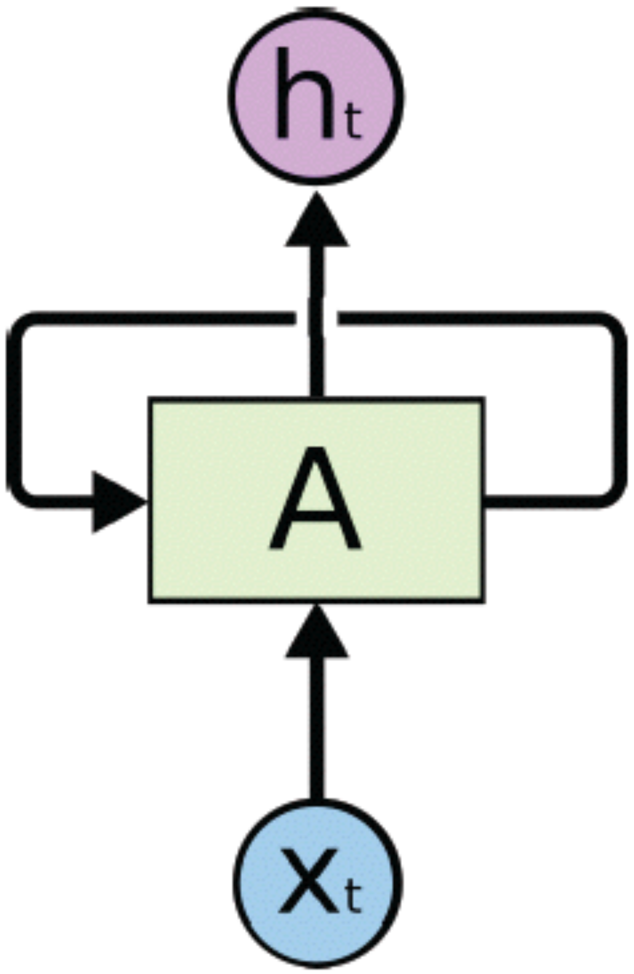
CS 6140: Machine Learning

Ehsan Elhamifar

eelhami@ccs.neu.edu

Recurrent NN

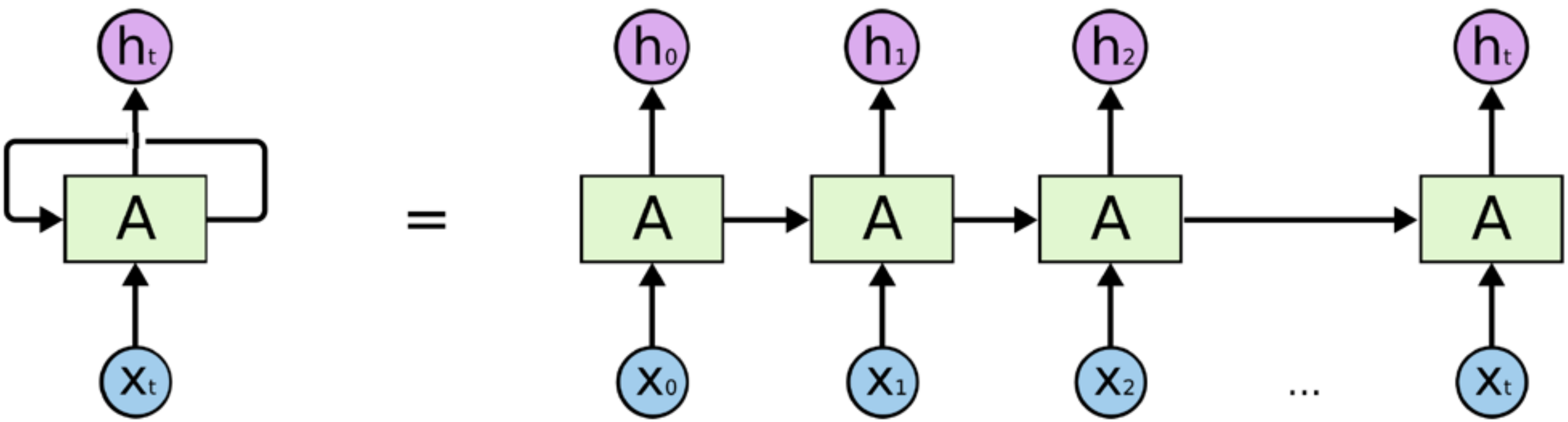
- Speech recognition, language modeling, translation, image captioning



Recurrent Neural Networks have loops.

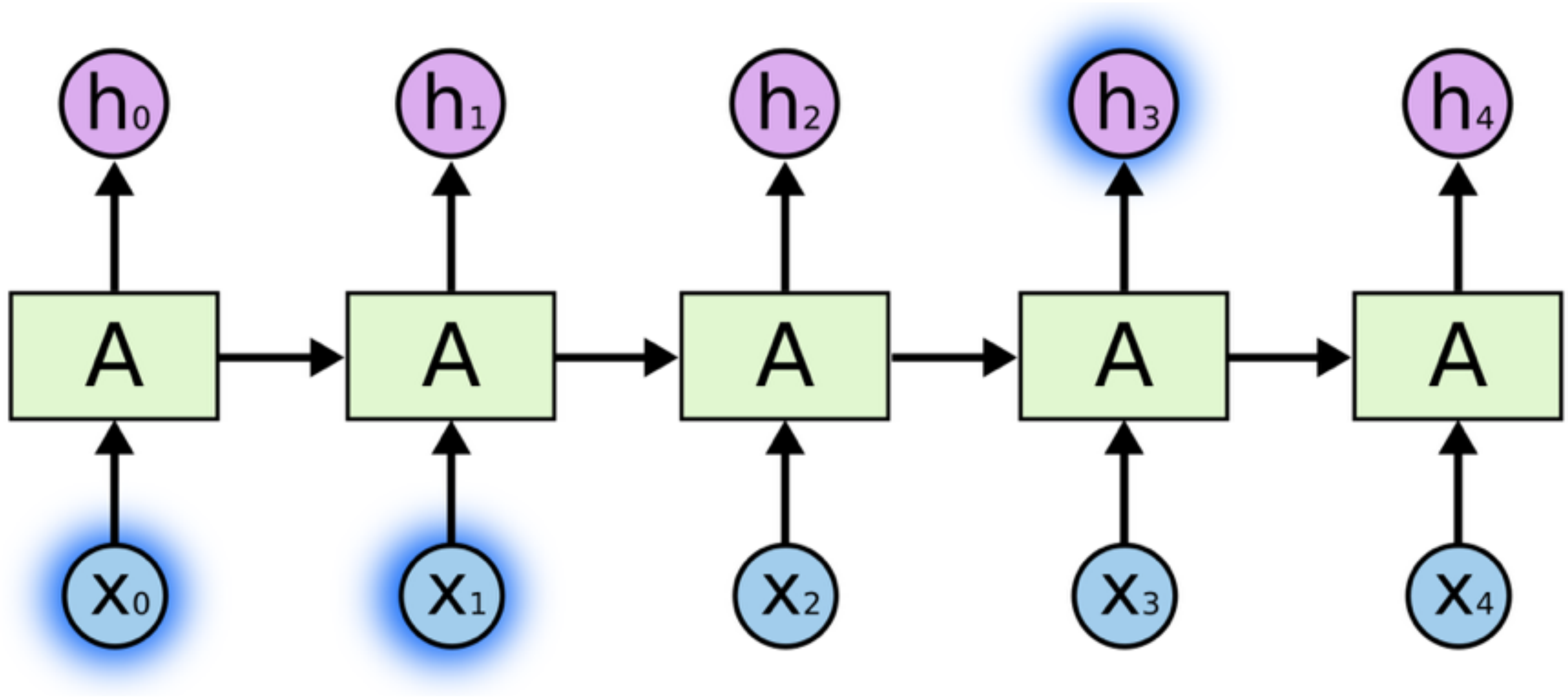
Recurrent NN

- Unrolled RNN



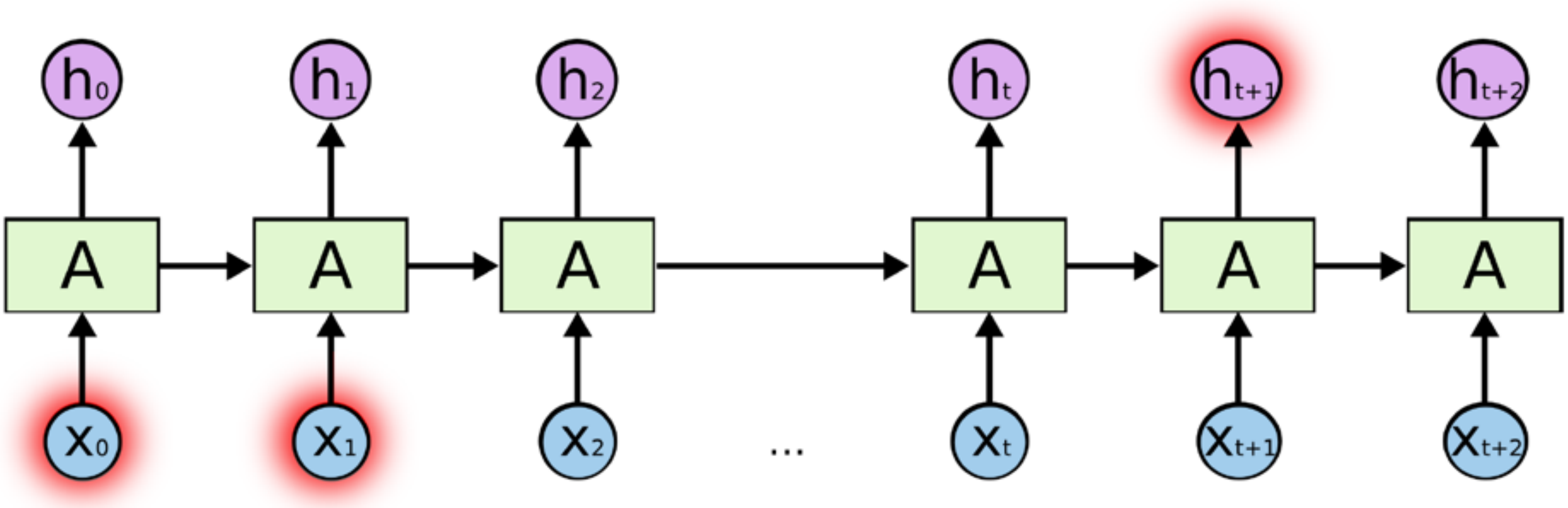
Recurrent NN

- “The clouds are in the *sky*”



Recurrent NN

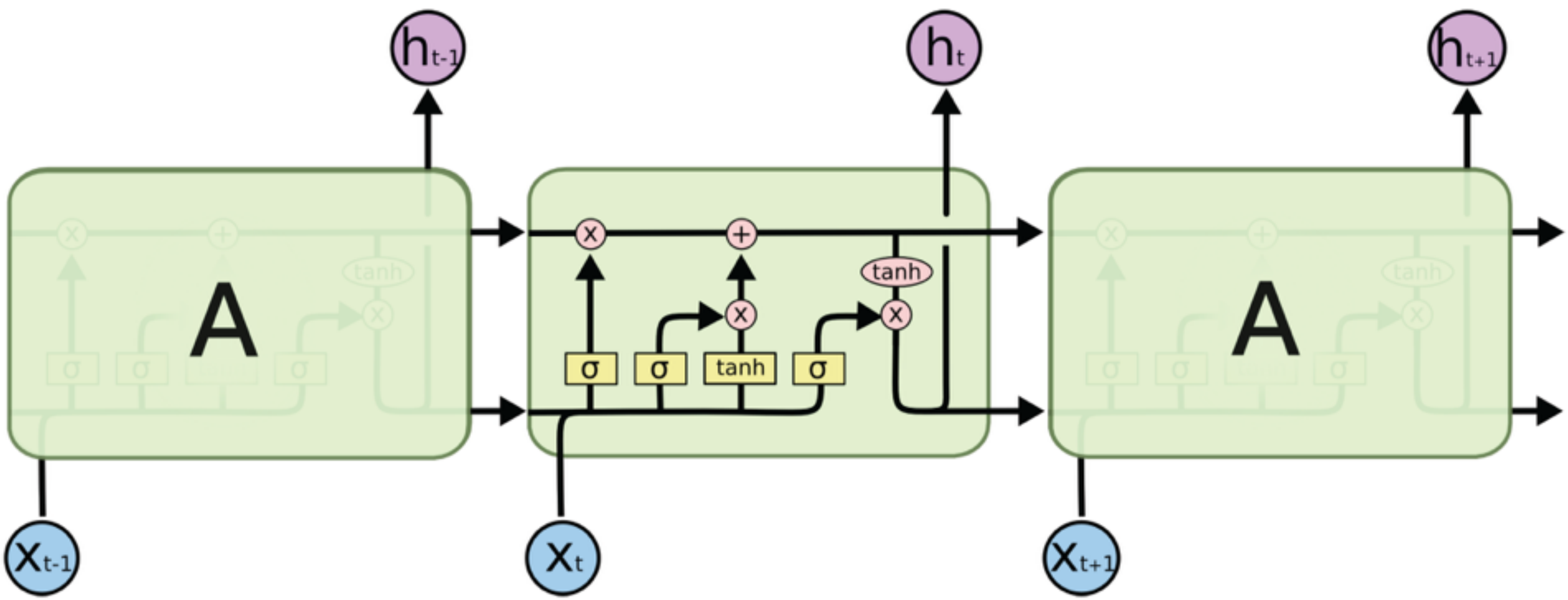
- "I grew up in France… I speak fluent *French*"



Larger information gap: more difficult

Long Short Term Memory (LSTM) Nets

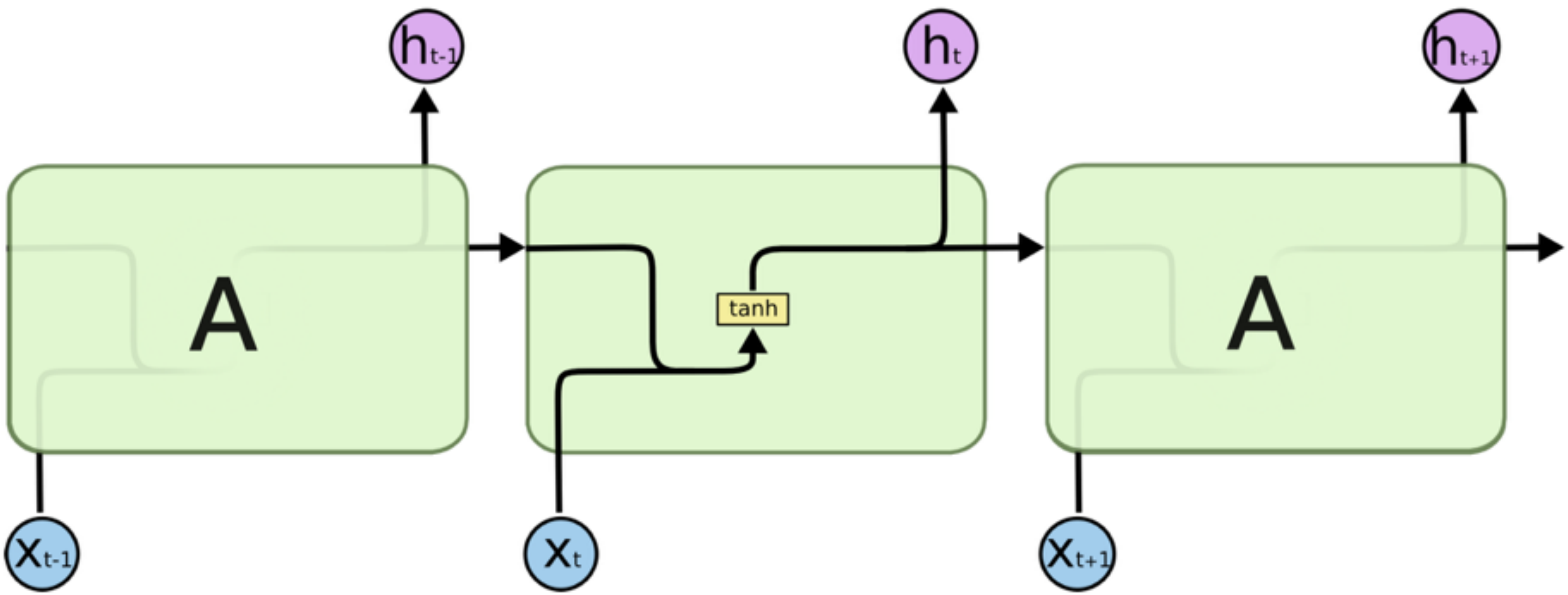
- Introduced by Hochreiter and Schmidhuber (1997)
- Capable of learning long-term dependencies



The repeating module in an LSTM contains four interacting layers.

Long Short Term Memory (LSTM) Nets

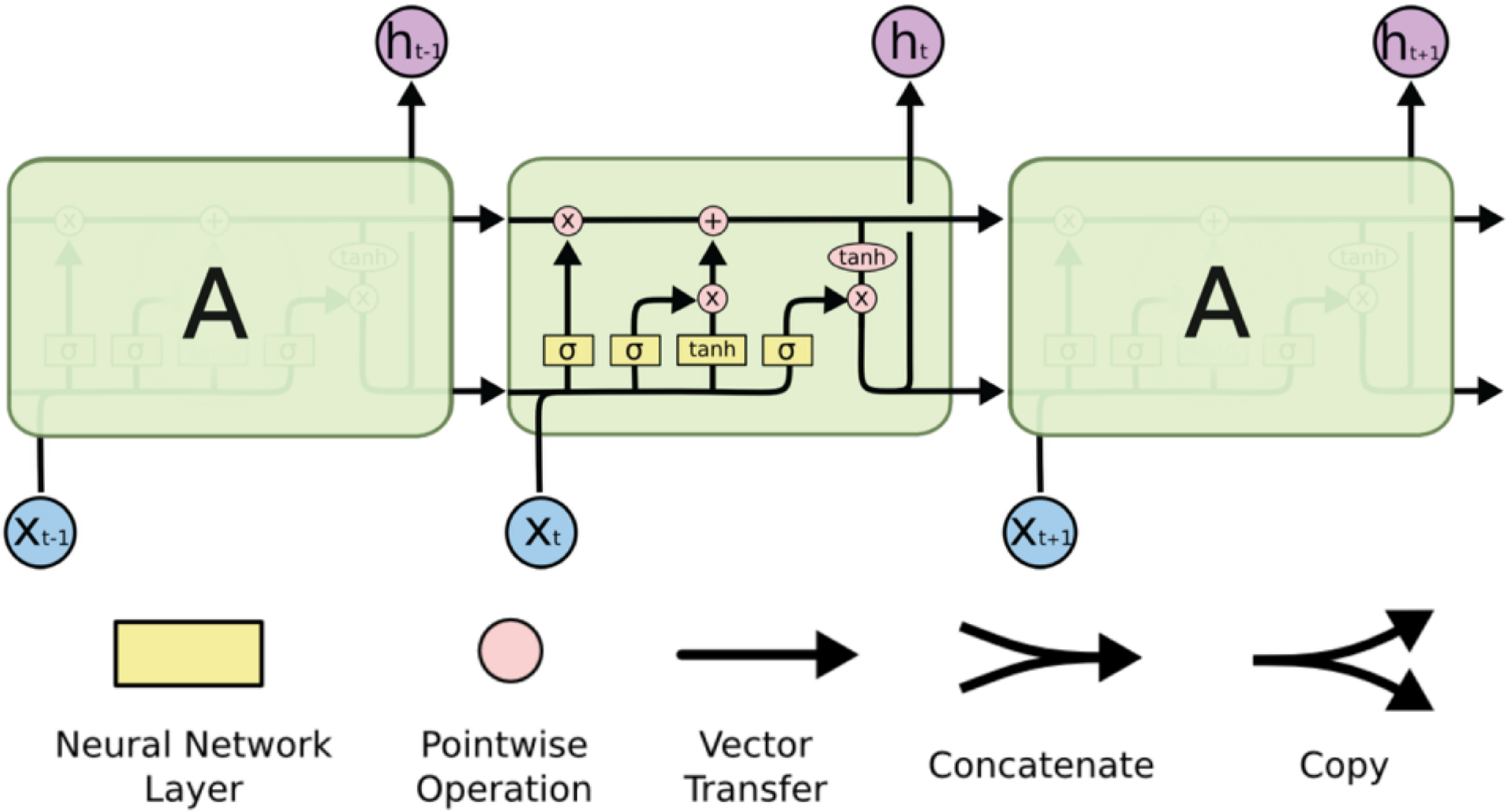
- Introduced by Hochreiter and Schmidhuber (1997)



The repeating module in a standard RNN contains a single layer.

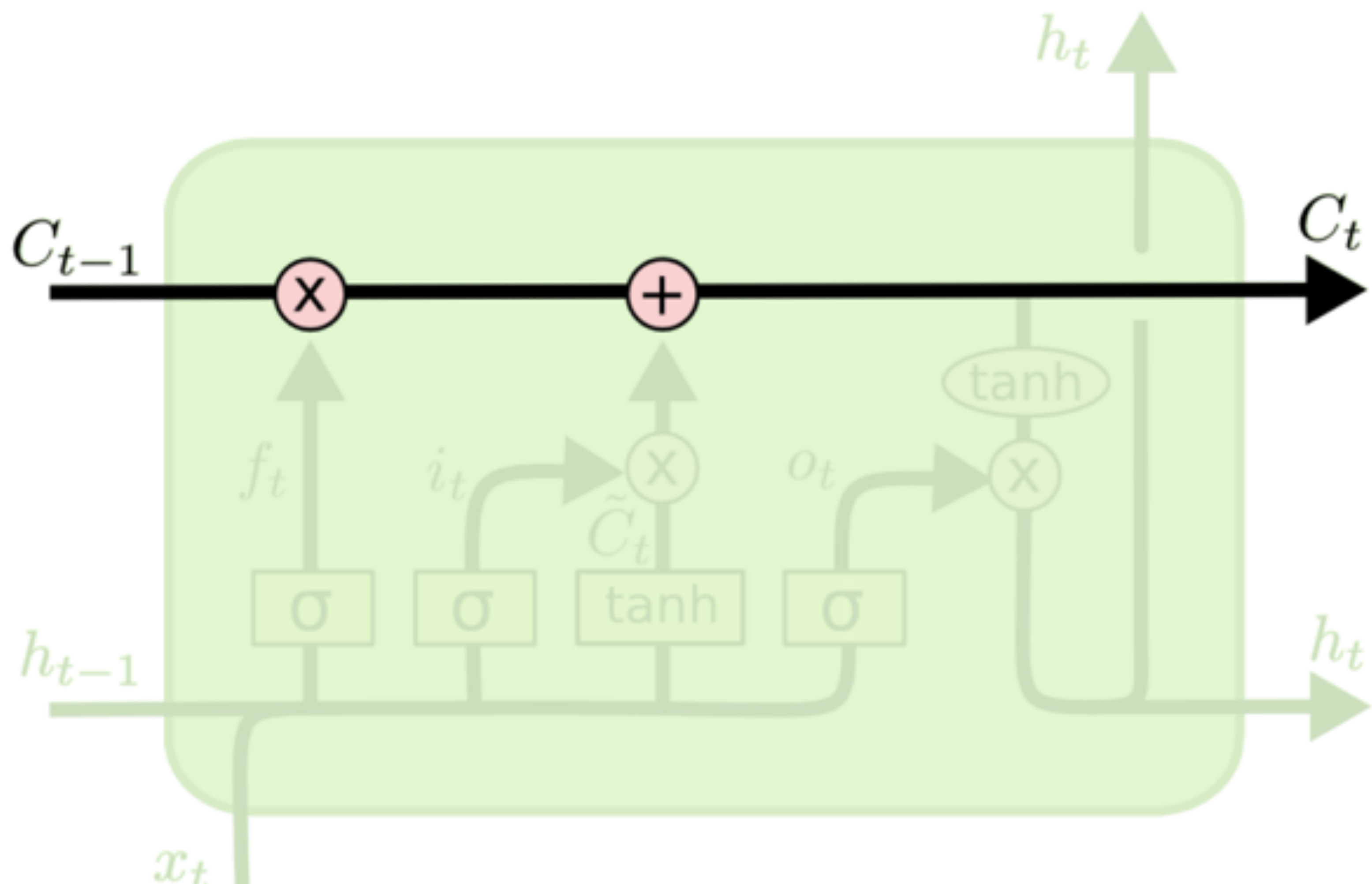
Long Short Term Memory (LSTM) Nets

- Capable of learning long-term dependencies



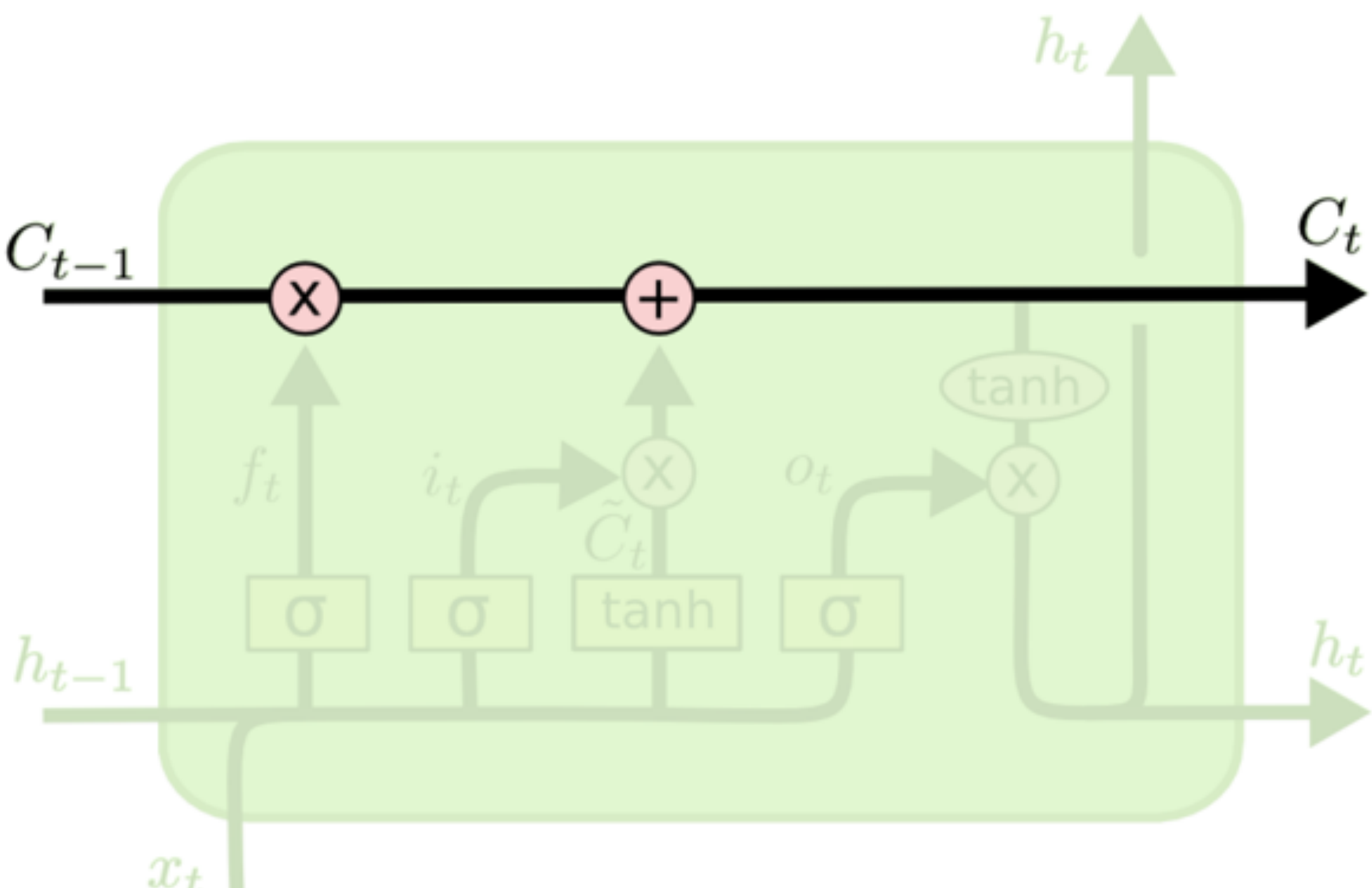
LSTM

- LSTM Key: cell state (horizontal line running through top)
- Conveyor belt with minor linear interactions (info can stay unchanged)

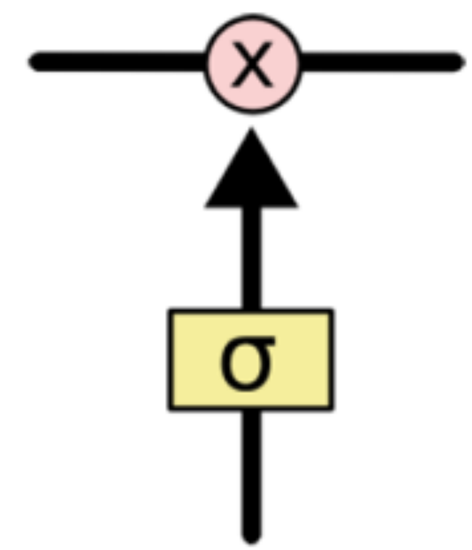


LSTM

- LSTM Key: cell state (horizontal line running through top)
- Conveyor belt with minor linear interactions (info can stay unchanged)

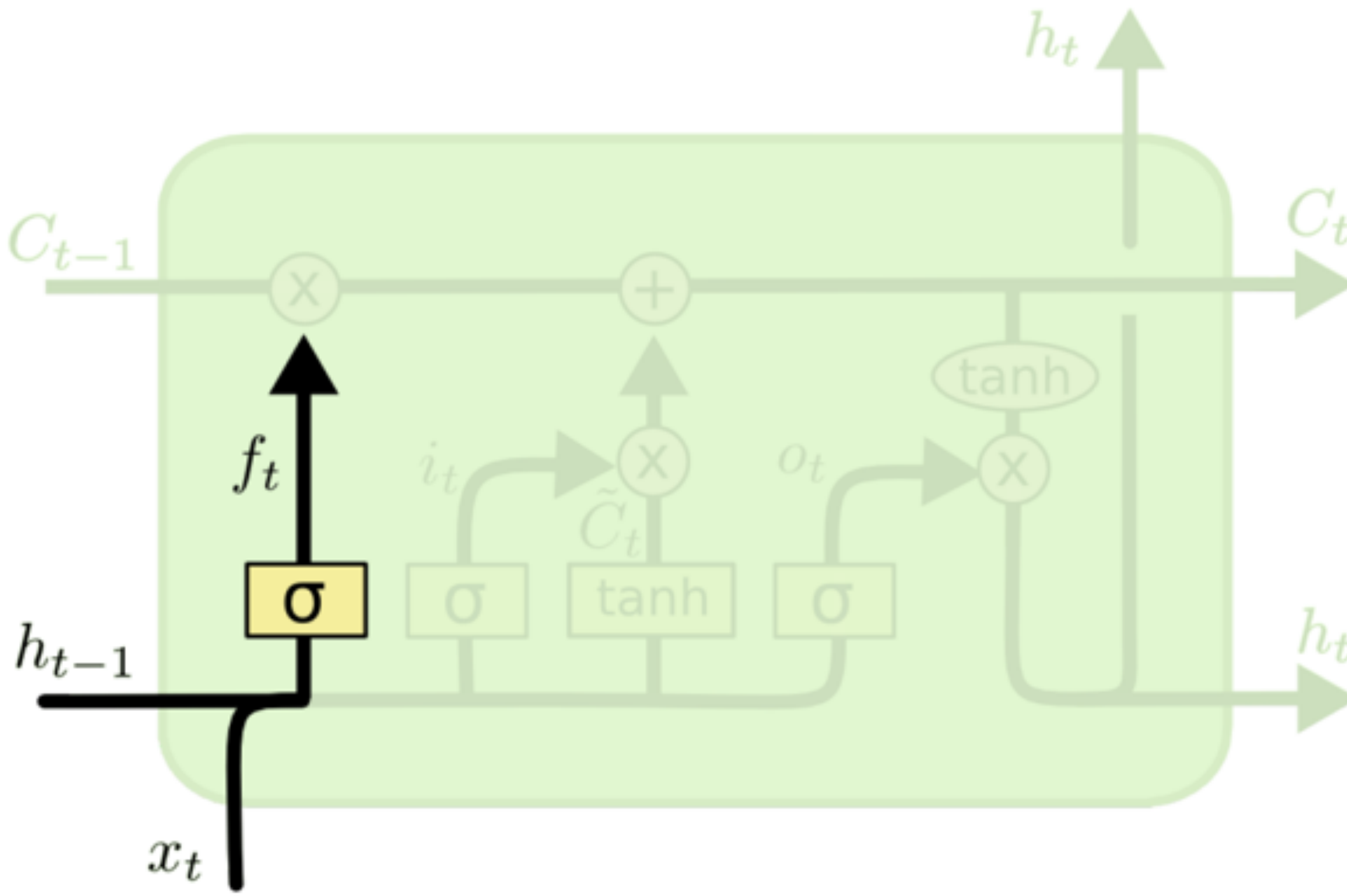


Gate: How much info let through



LSTM

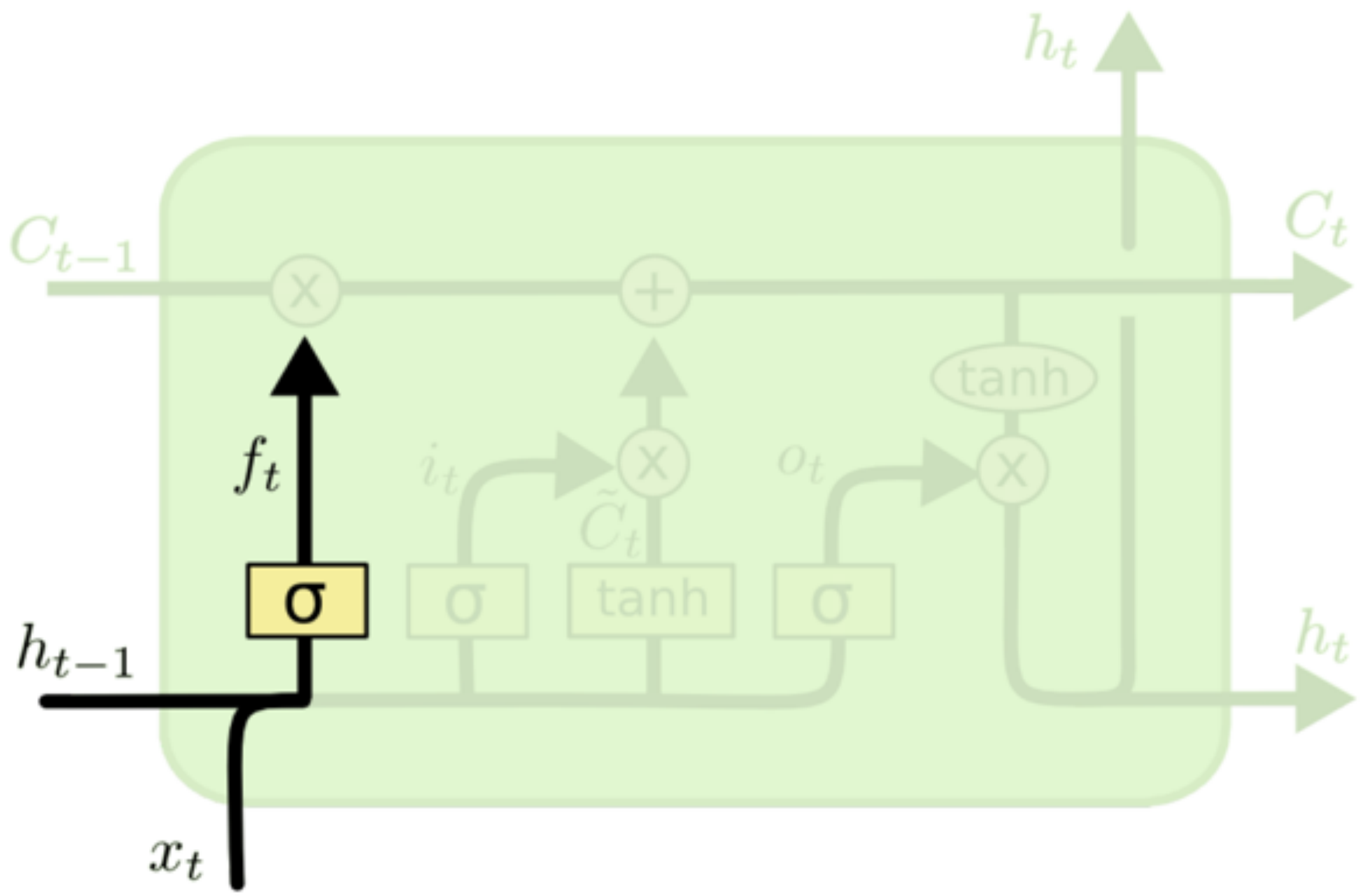
- Three types of gates
- Forget gate layer: what information to forget



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

LSTM

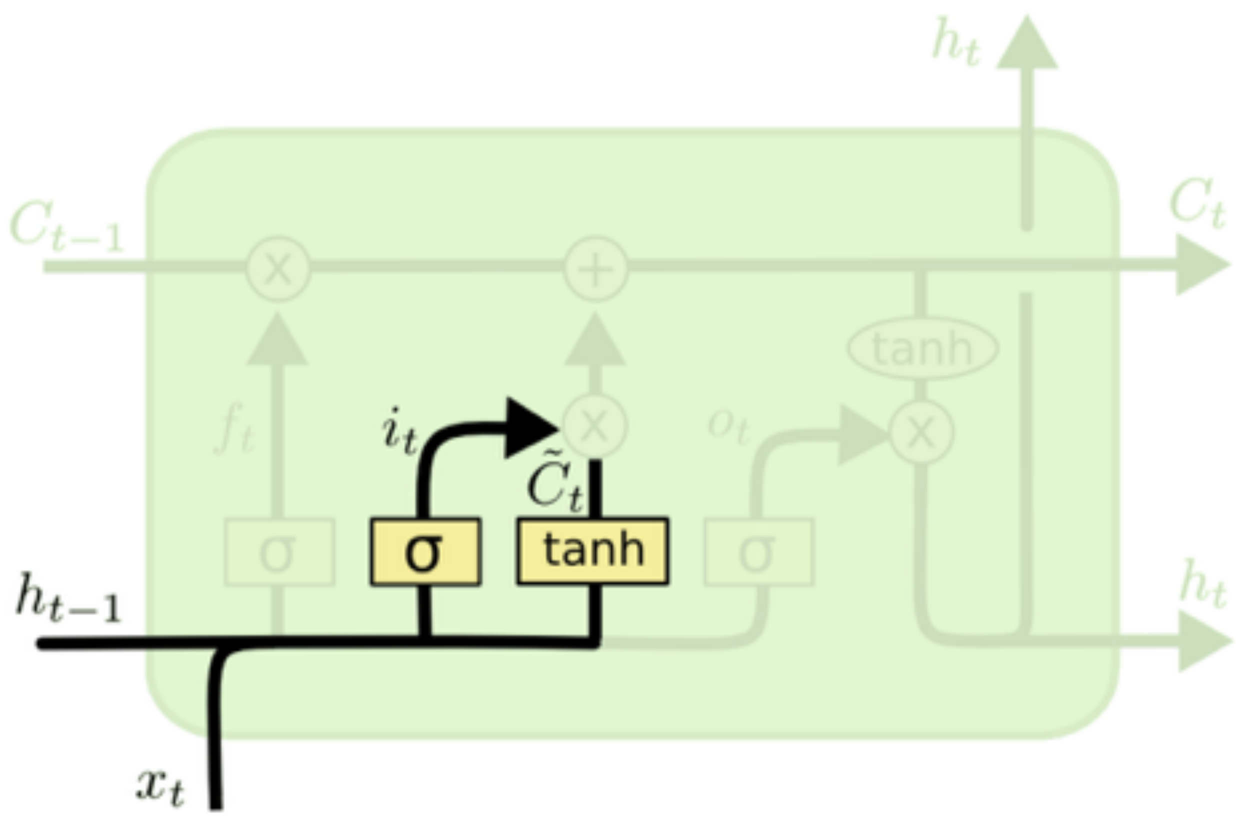
- Forget gate layer: what information to forget
- Keep gender of the current subj, until see a new subj.



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

LSTM

- Input gate layer: information to add to the cell state
- Candidate states

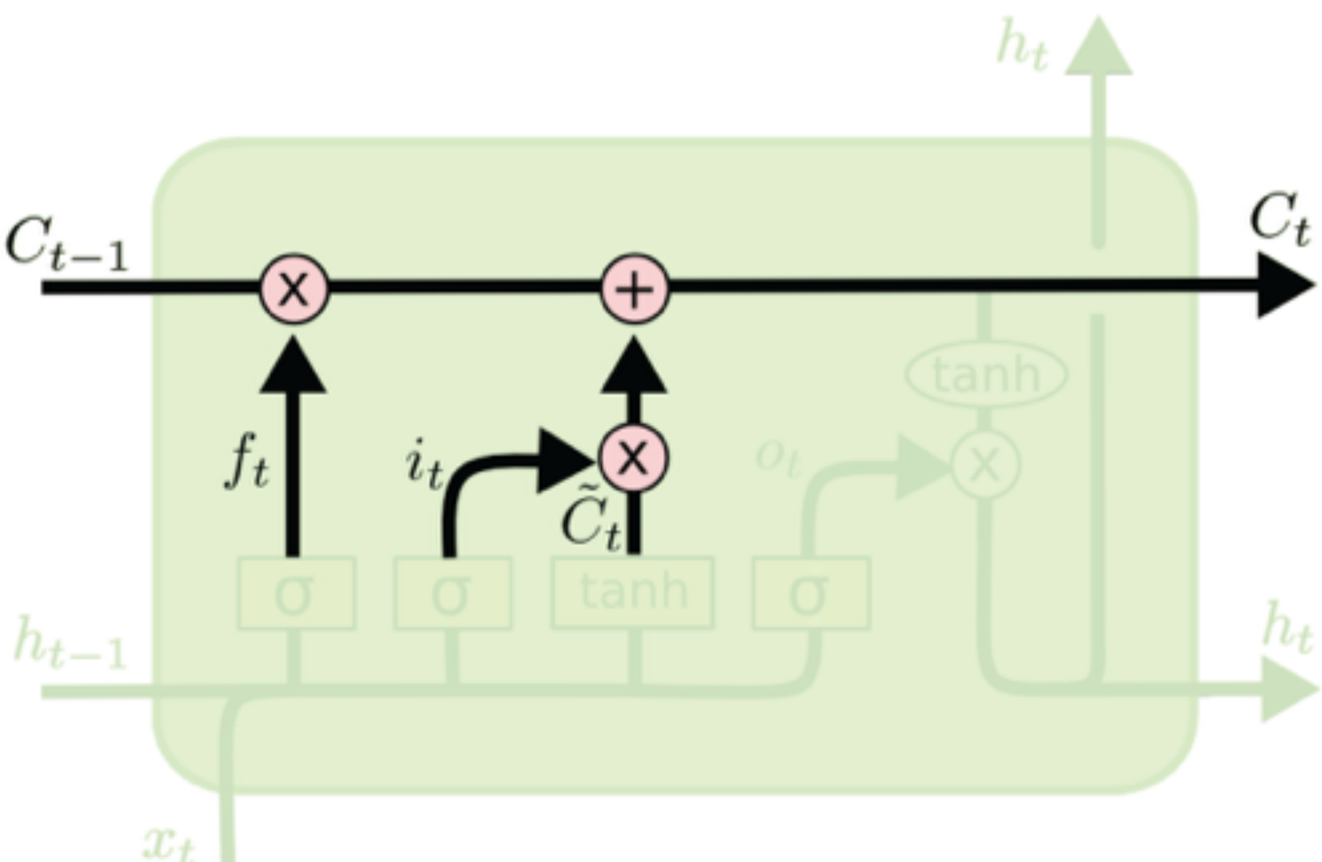


$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \ + \ b_i\right)$$

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

LSTM

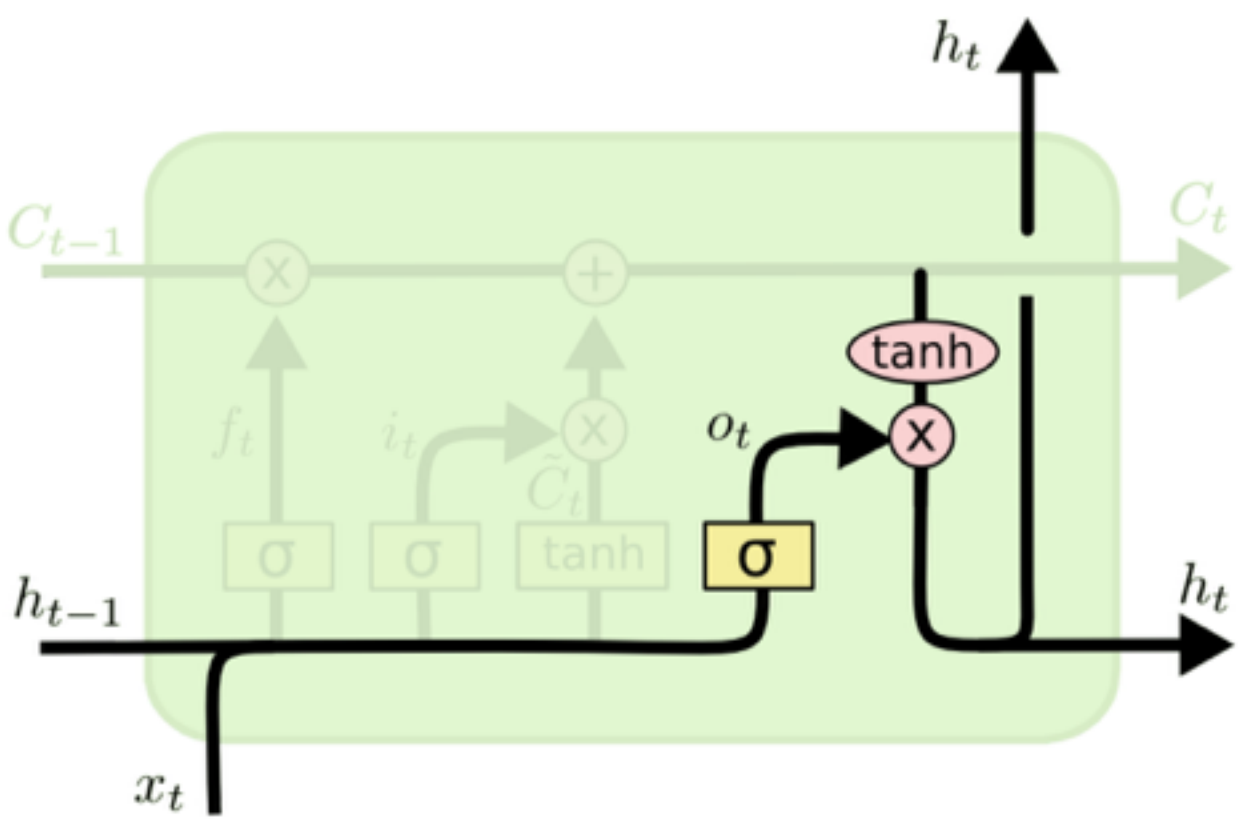
- Update cell state using forget and input gates
- Remove the gender of old subj, add the gender of new subj.



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

LSTM

- Output gate (tanh: to transform info into [-1,+1], sigm: modulation)
- Saw a new subj: then output info about being singular or plural (so that we know what verb to use if needed)



$$o_t = \sigma\left(W_o\ [h_{t-1}, x_t]\ +\ b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$