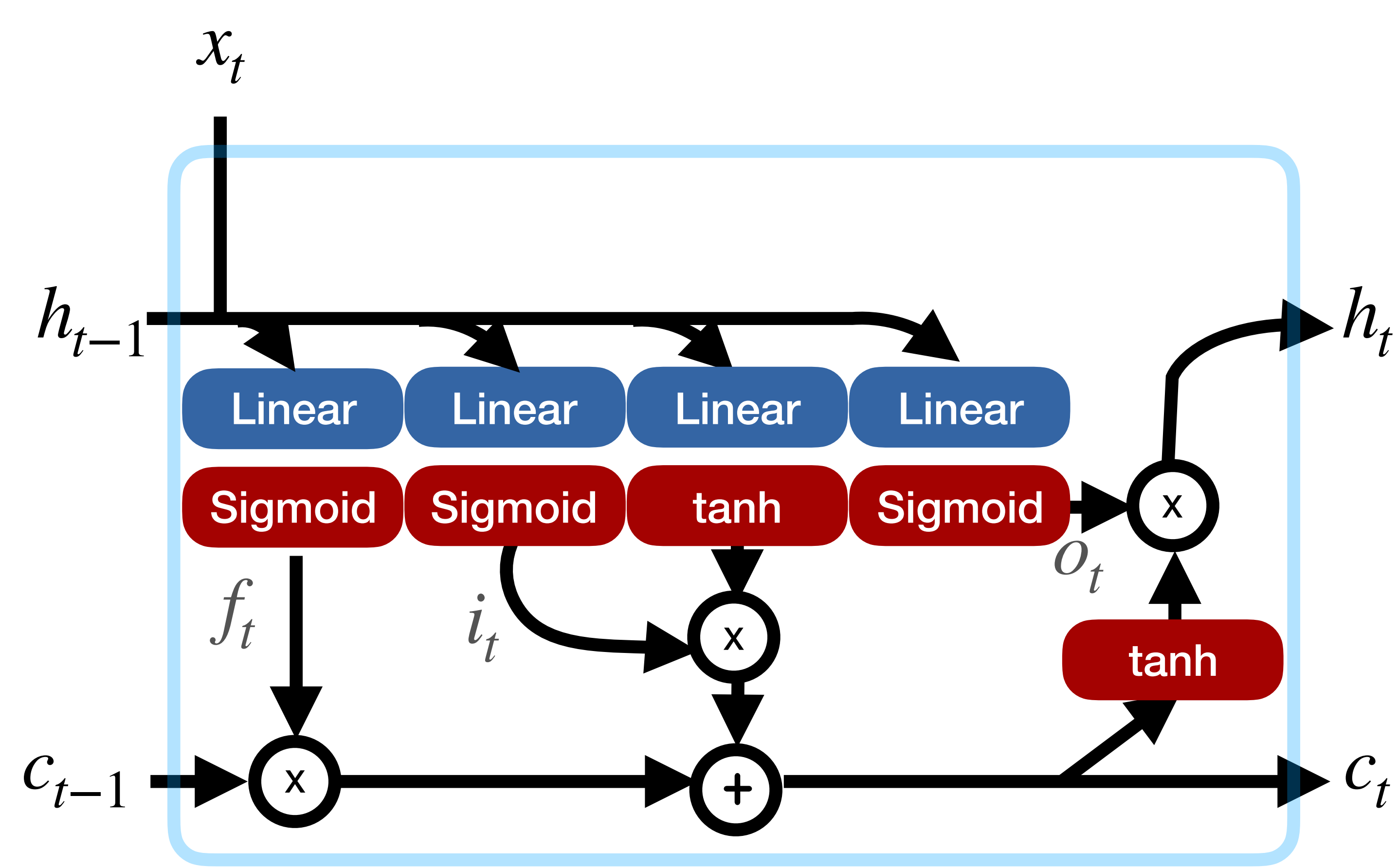
LSTMs and GRUs

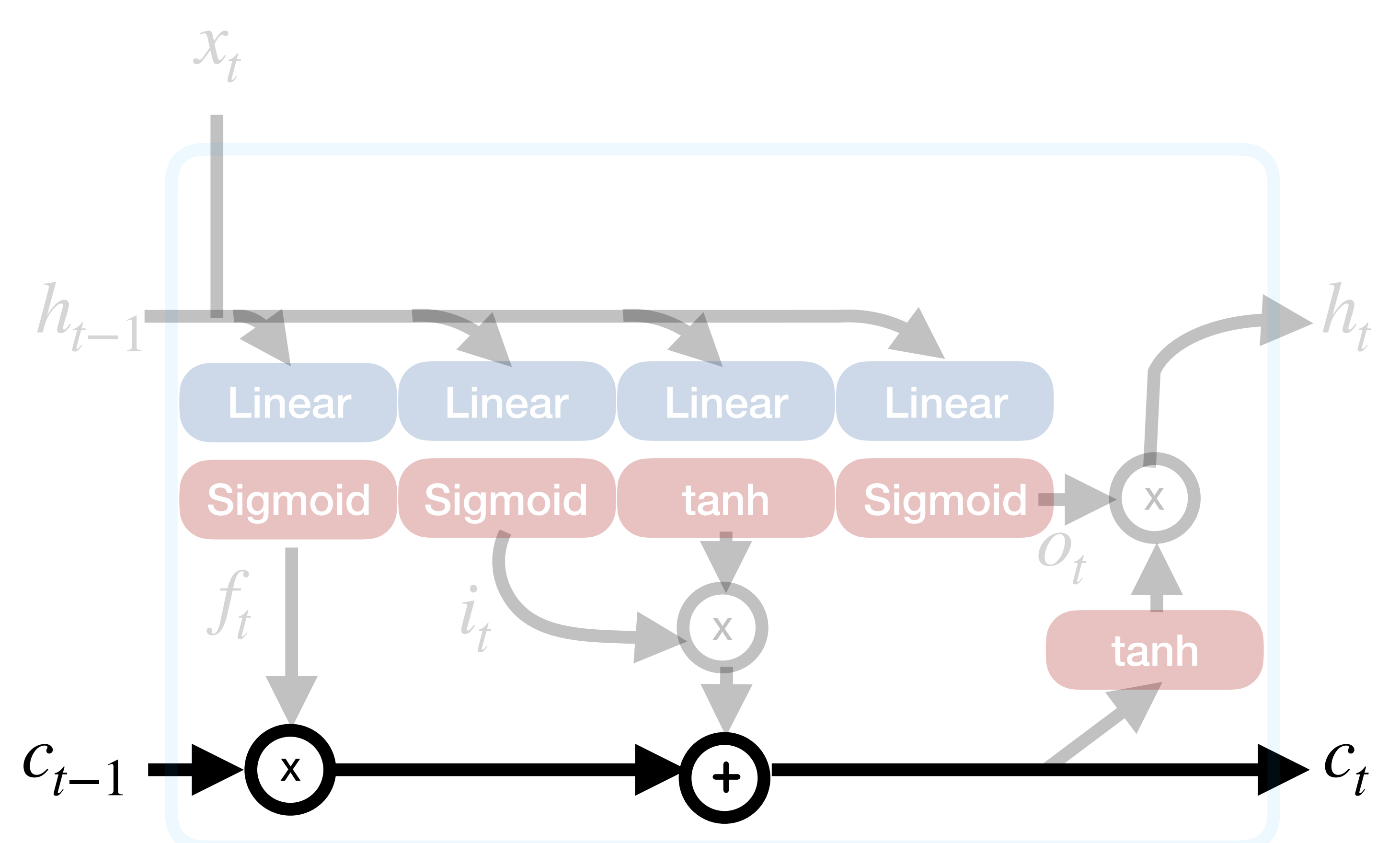
Long short-term memory

- Two recurrent connections
  - Long-term c
  - Short term h
  - Input x
  - Output h



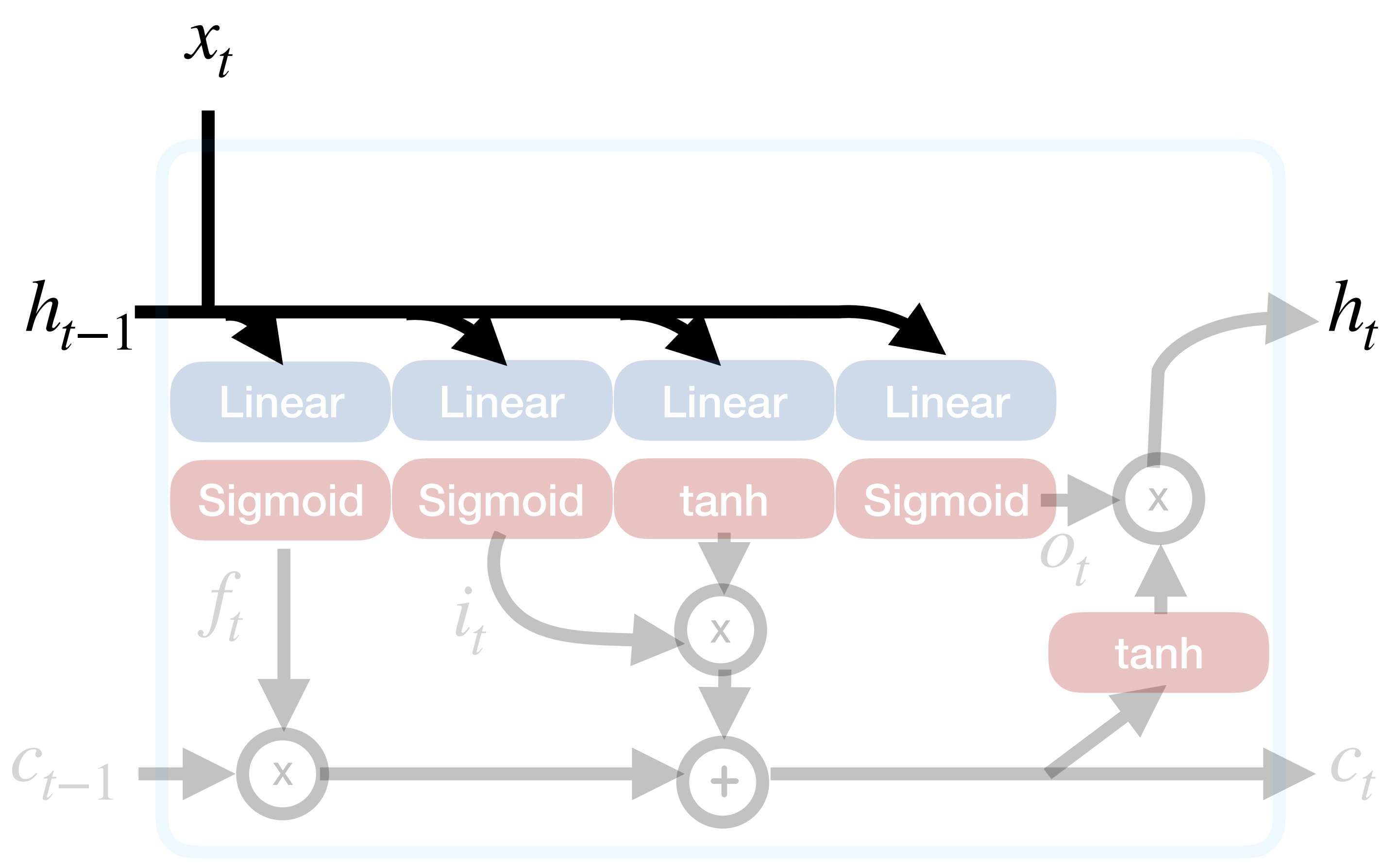
Cell state

- Cell state c
  - Only multiplication and addition
  - Shortcut
    - Similar to ResNets



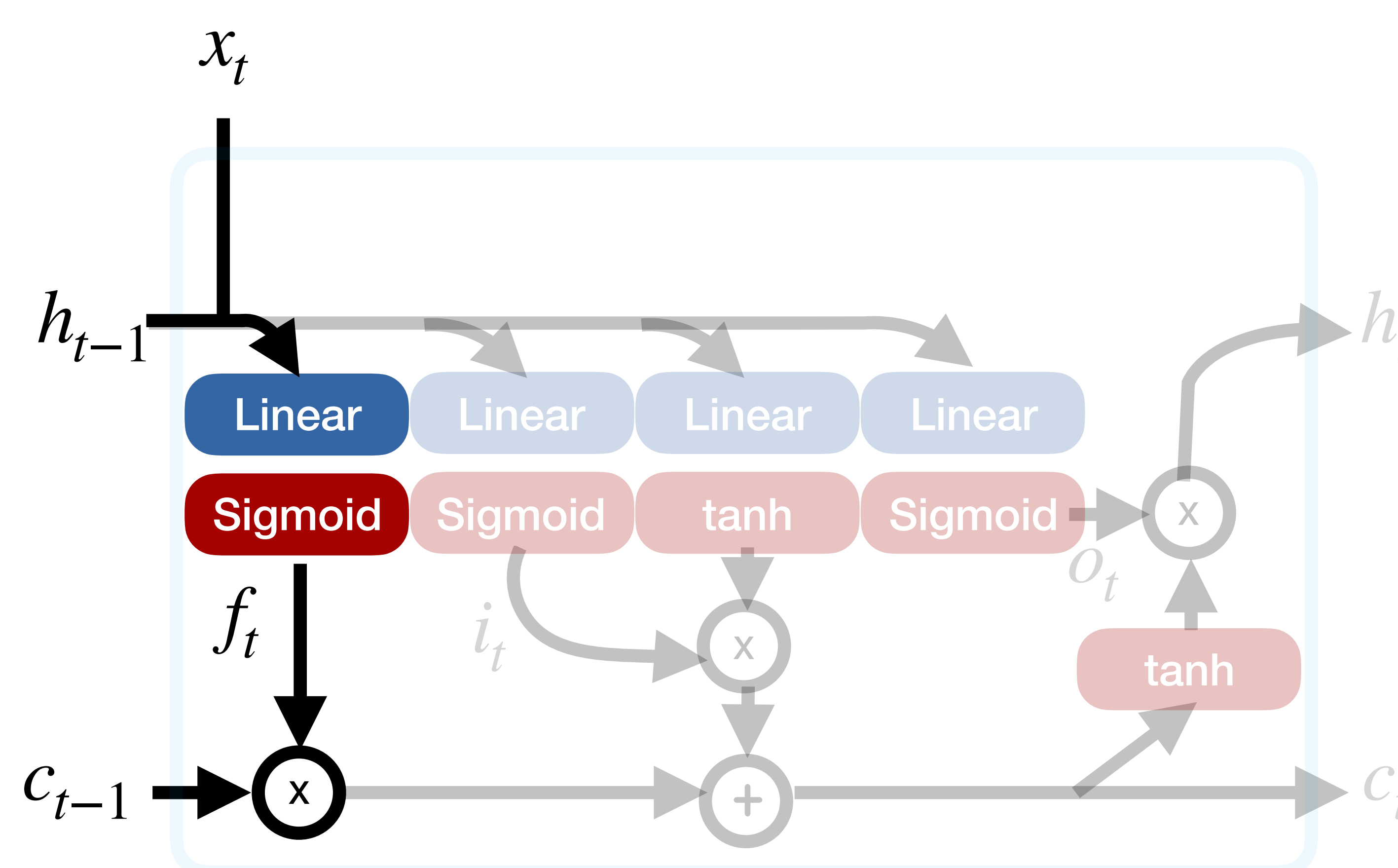
Input

- Input
  - x
  - Previous h



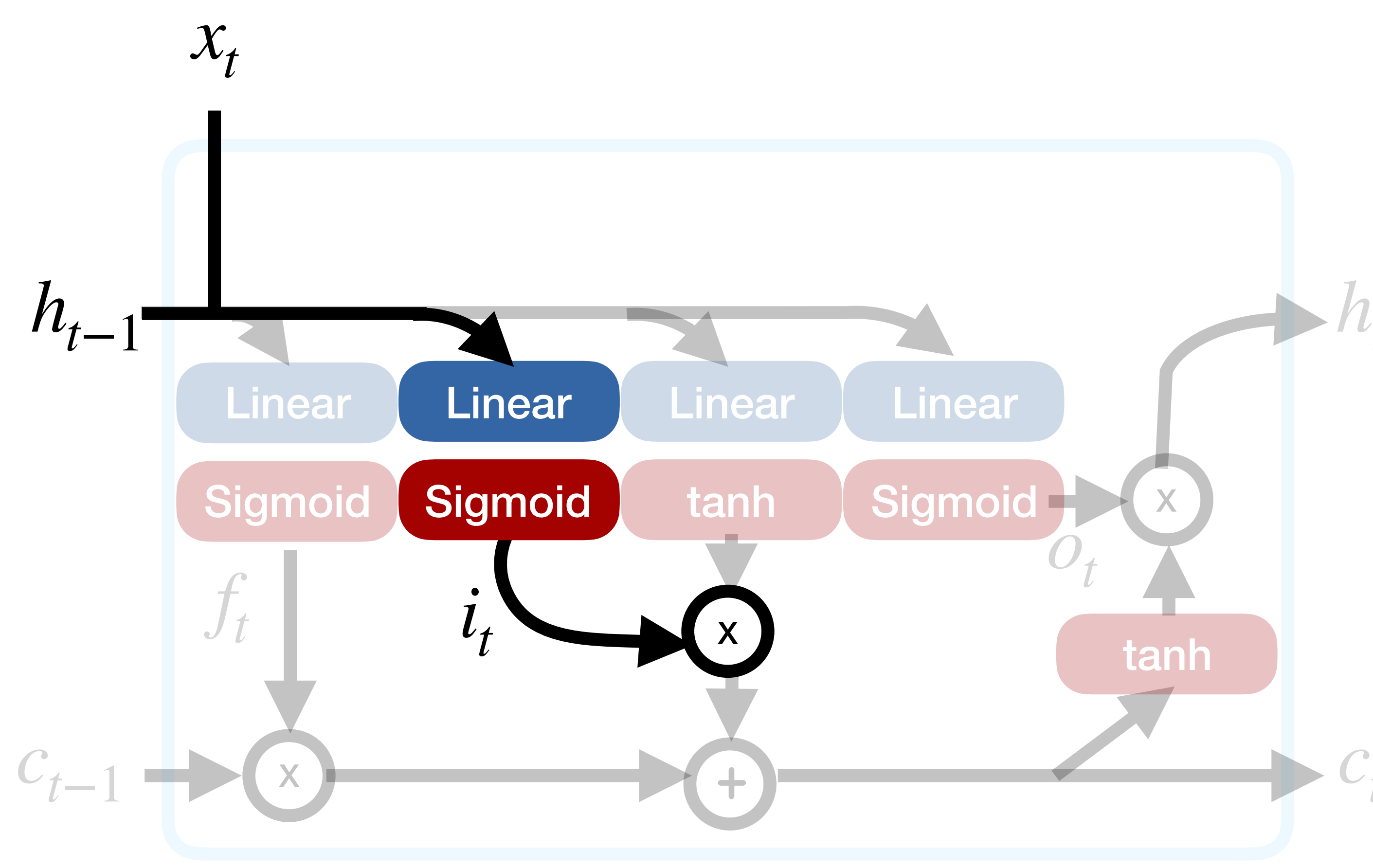
Forget gate

- Forget gate f
  - Clears cell state



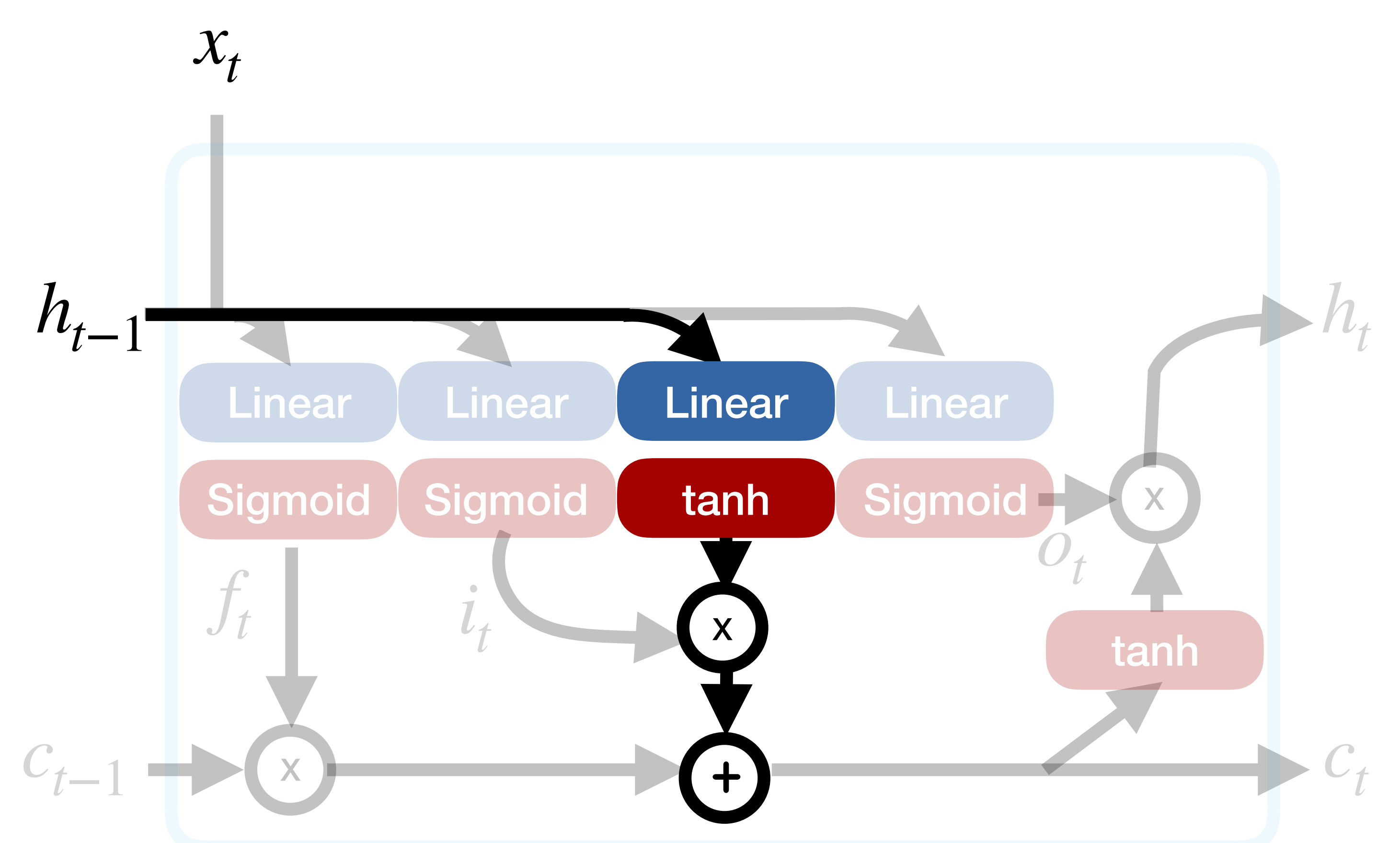
Input gate

- Input gate i
  - Allows state update



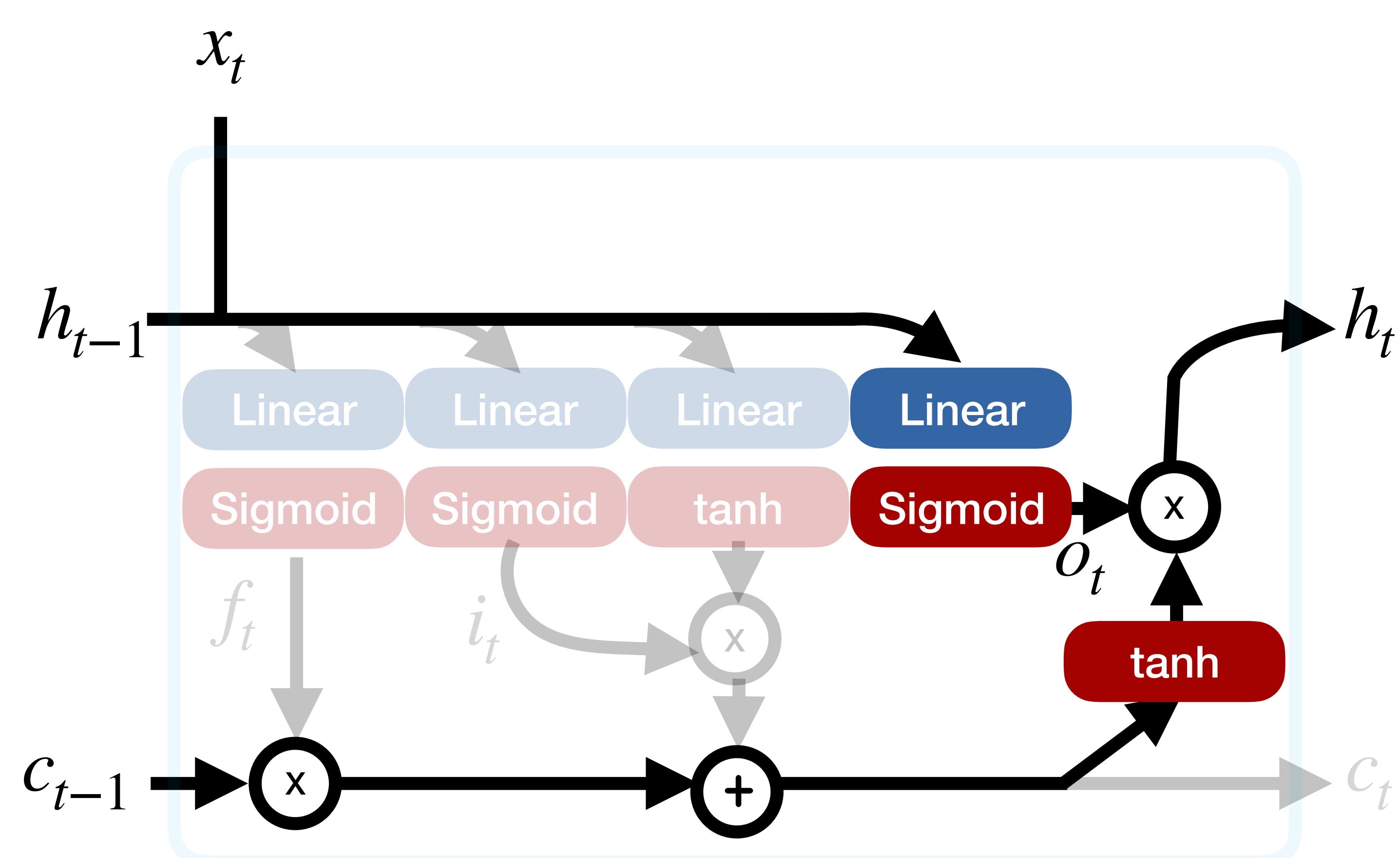
State update

- State update
  - tanh



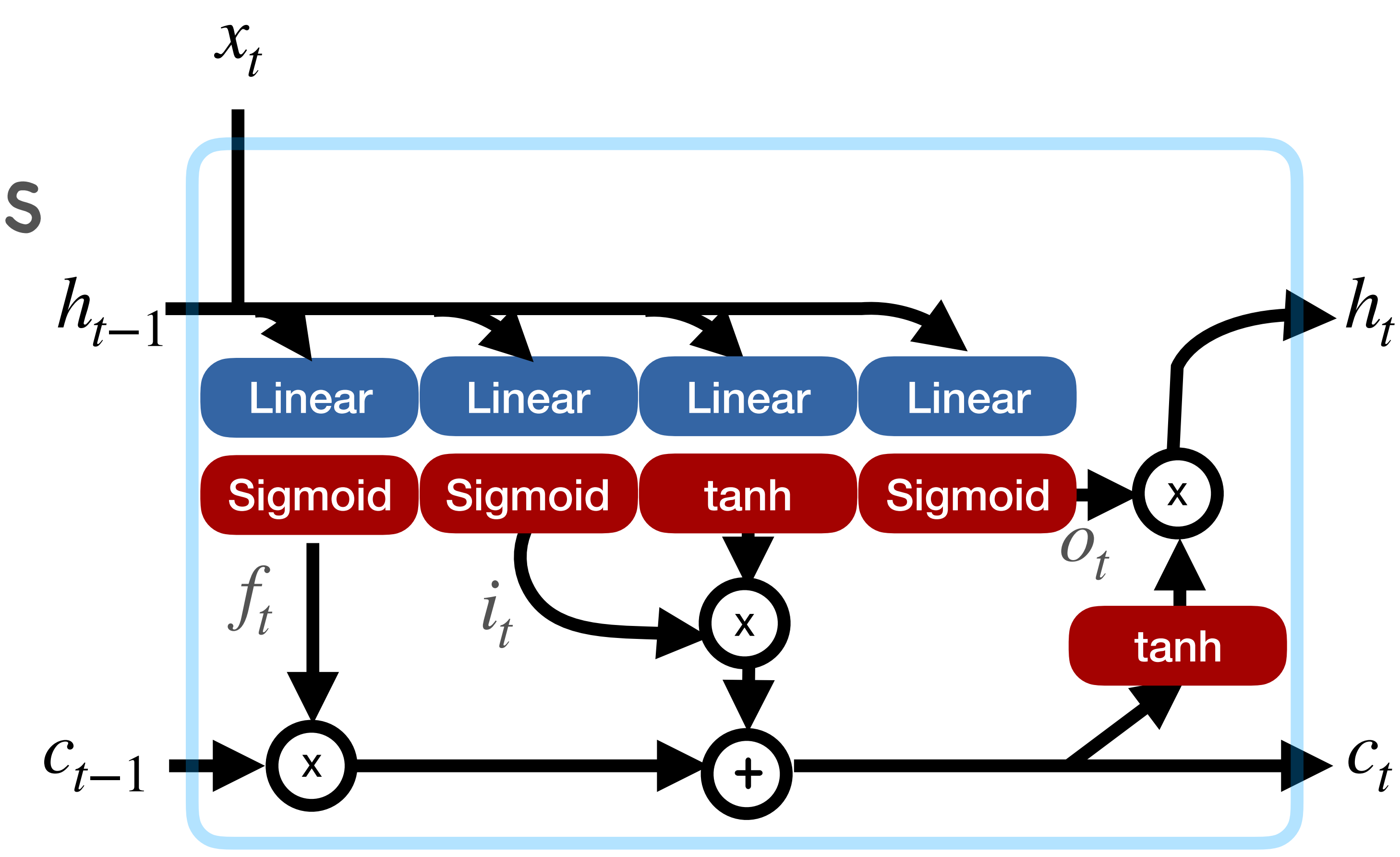
Output

- Output gate o
  - Produce an output?
- Output h
  - tanh of cell state



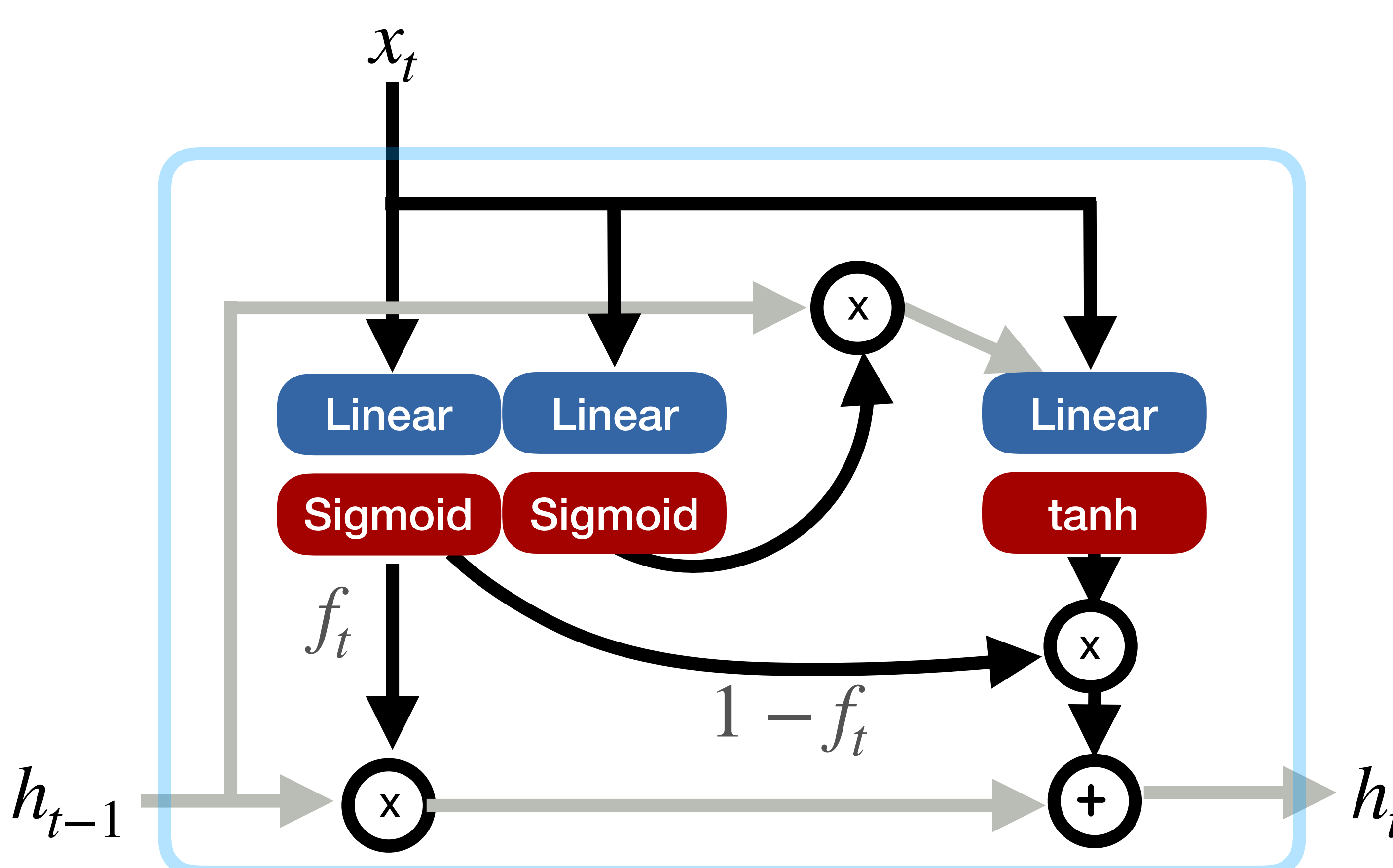
LSTMs

- Can learn to keep state for up to 100 time steps
  - Fewer vanishing gradients
  - Short cut

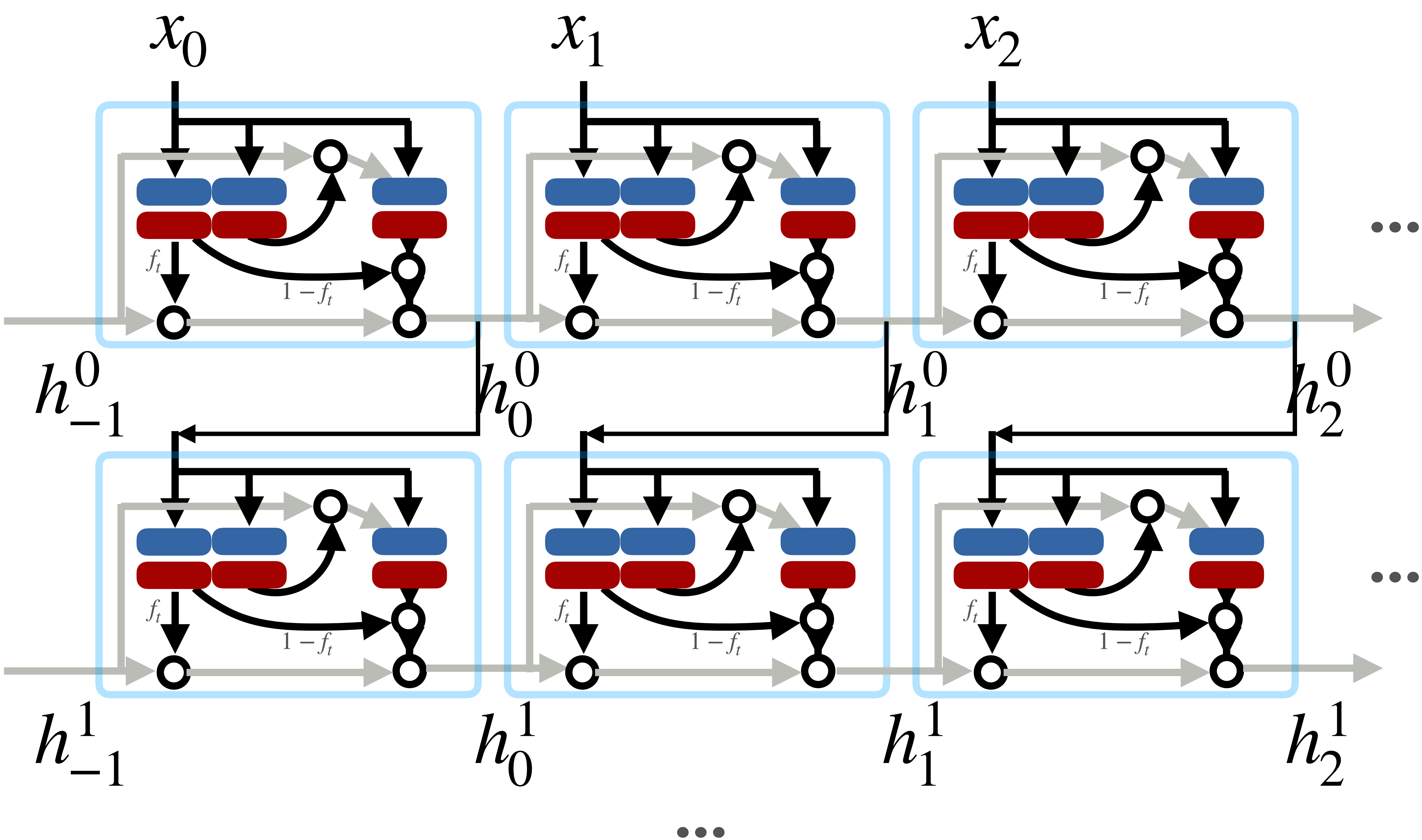


Gated Recurrent Units

- Simpler LSTM
  - Single state
  - Fewer gates
  - Similar performance



LSTM/GRU Networks



LSTM / GRU applications

- Hand writing synthesis
- Natural language processing
- Image generation

Image source: Demo by Alex Graves
http://www.cs.toronto.edu/~graves/

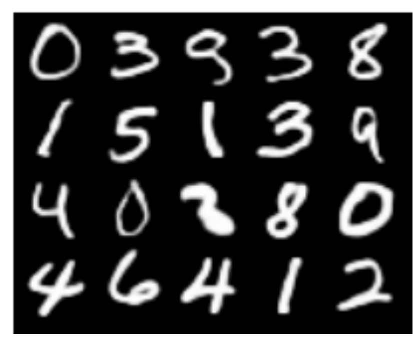


Image source: Gregor et al., https://arxiv.org/pdf/1502.04623.pdf

- Generating Sequences With Recurrent Neural Networks, Graves, arXiv 2013
- Sequence to Sequence Learning with Neural Networks, Sutskever et al., NIPS 2014
- DRAW: A Recurrent Neural Network For Image Generation, Gregor et al., ICML 2015