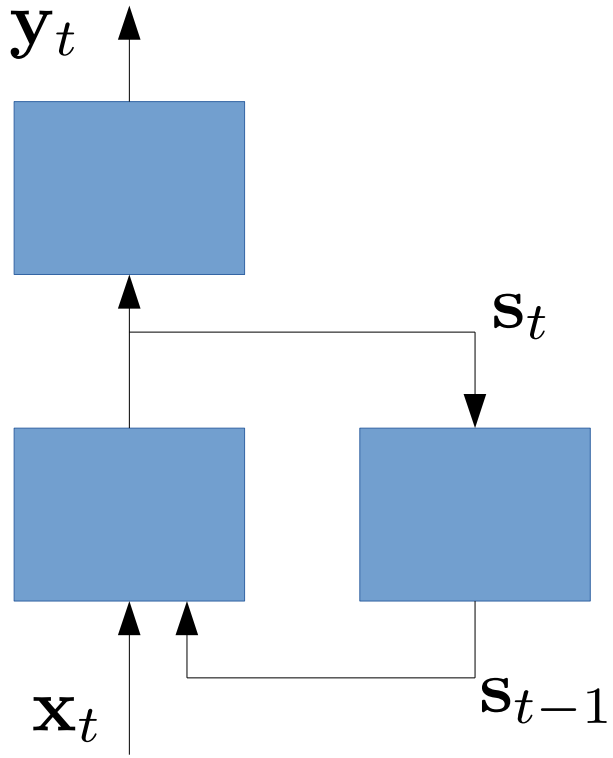
TEK 5040/9040 Recurrent Neural Networks (RNNs)

Narada Warakagoda

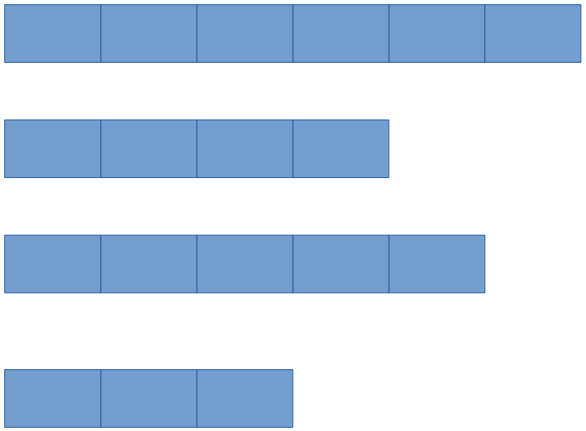
What is a Recurrent Neural Network (RNN)?

- RNN is a neural network with a feedback loop.
- Suitable for processing sequential/serial data



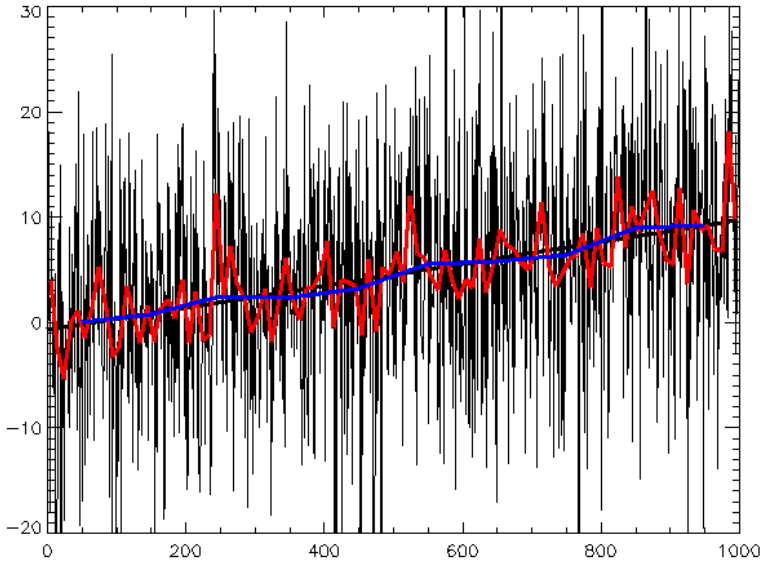
Why process data serially?

- Limited computational capability
- Limited storage capability
- Need a response immediately
- More efficient to divide task into sub-tasks
- Data sequences have arbitrary/variable lengths



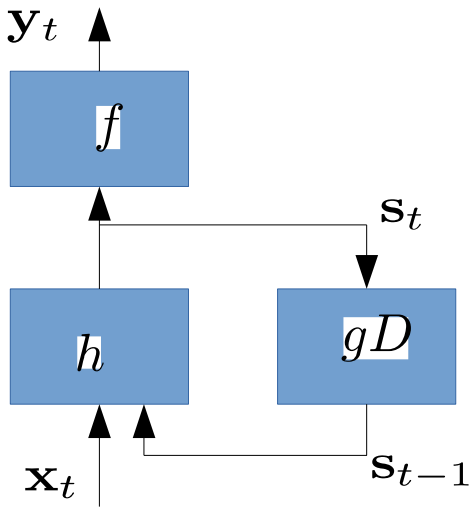
Example Applications

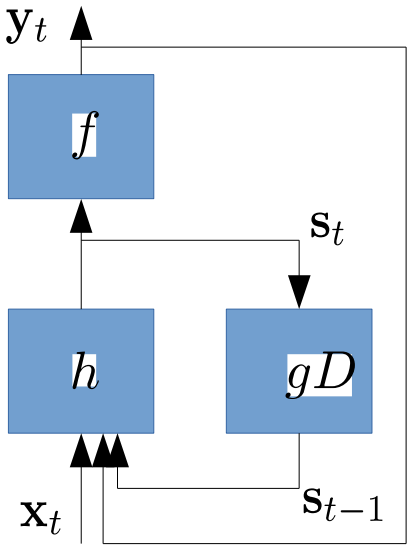
- Language modeling
- Sentiment analysis
- Machine translation
- Image captioning
- Time series analysis
- Speech/audio analysis
- Etc. etc..

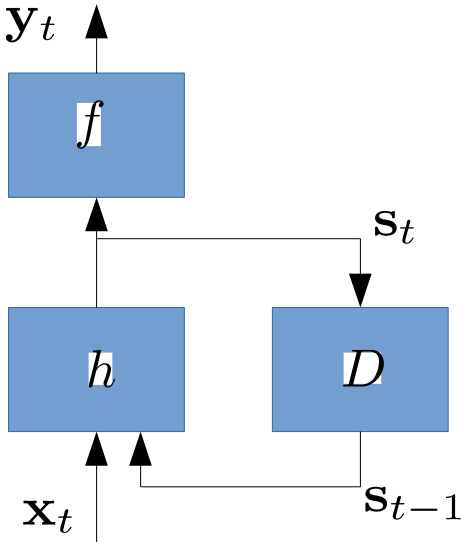


RNN Cell Models

- Multiple ways to introduce feedback





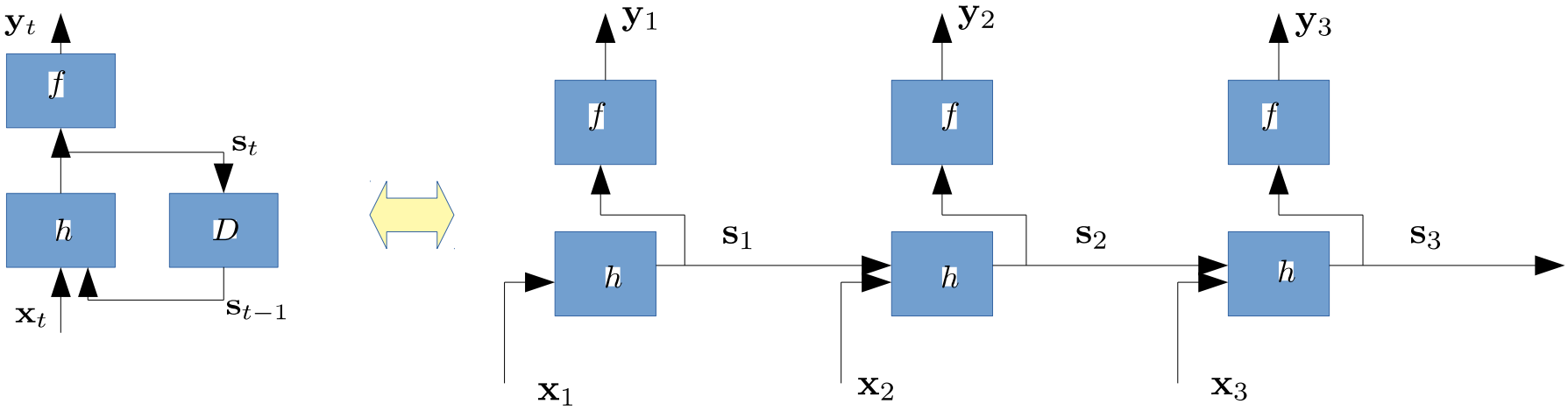


h	State transition function
f	Measurement function
gD	Feedback function with delay
D	Delay only

$\mathbf{x}_t$	Current input
$\mathbf{y}_t$	Current output
$\mathbf{s}_t$	Current state

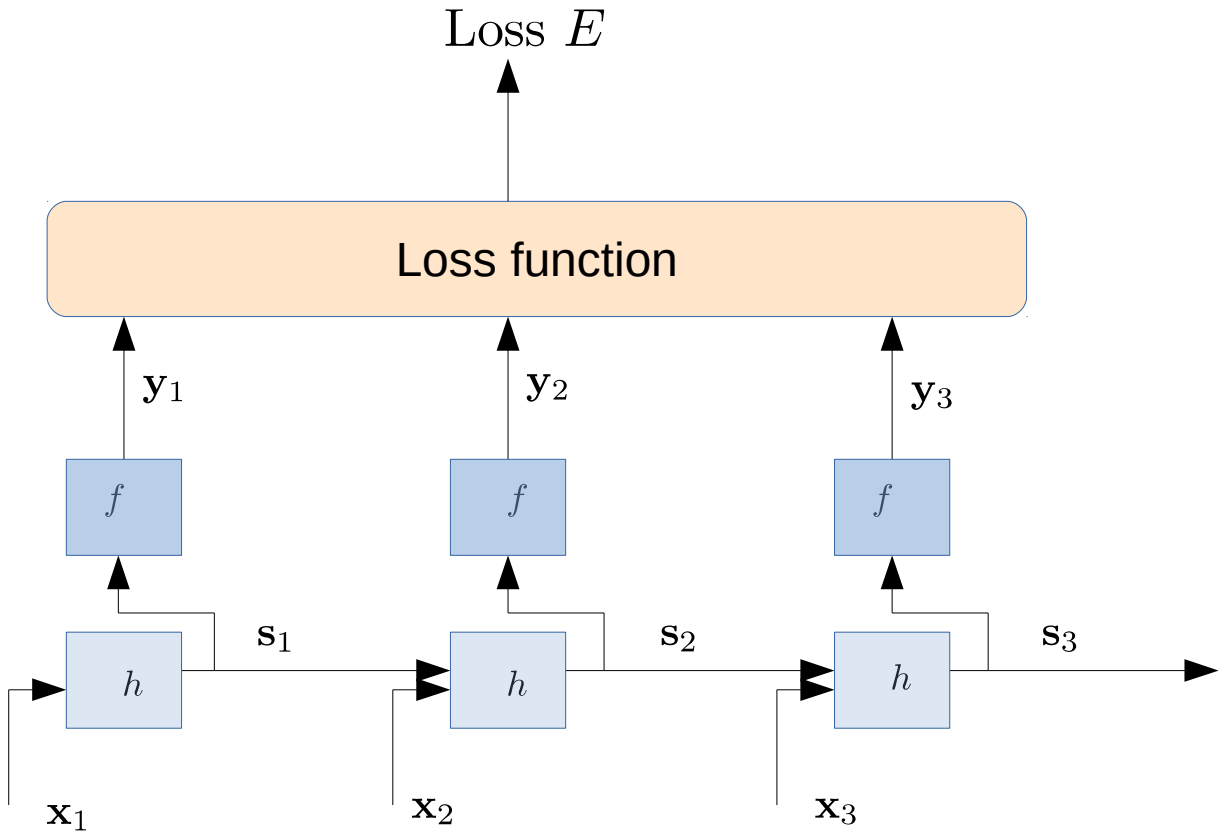
RNN Cell Unfolding

- Show the cell values at each time step



$$\mathbf{s}_t = h(\mathbf{x}_t, \mathbf{s}_{t-1}) = \tanh(\mathbf{W}_h[\mathbf{x}_t, \mathbf{s}_{t-1}] + \mathbf{b}_h)$$

Training

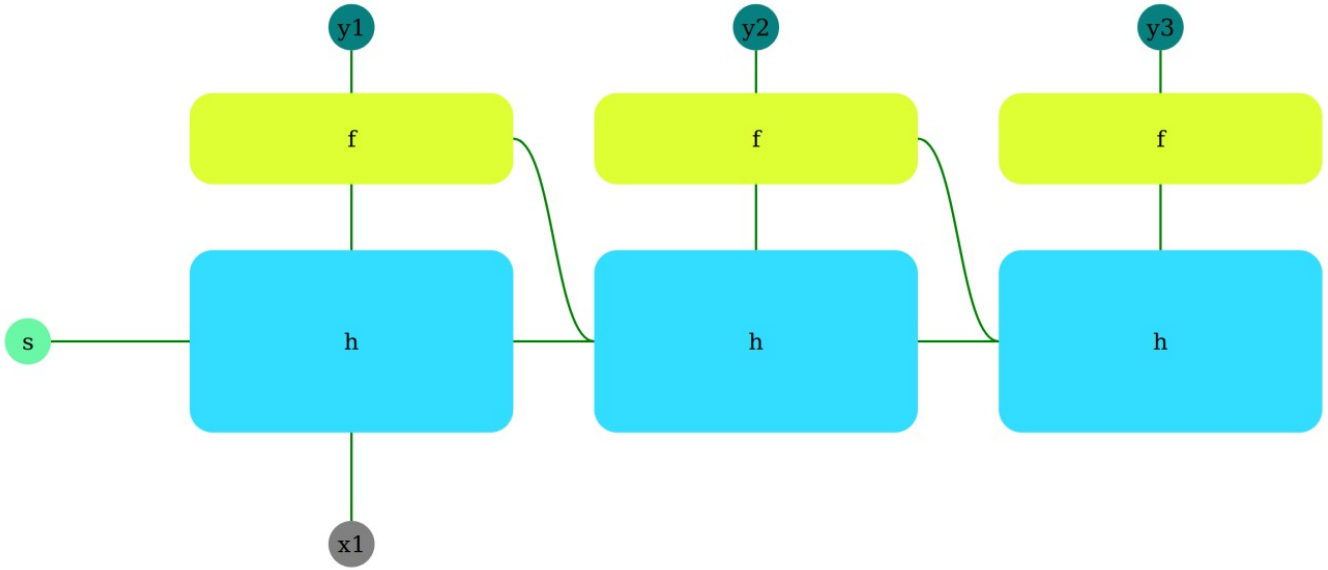

Loss E
Loss function
$\mathbf{y}_1$
$\mathbf{y}_2$
$\mathbf{y}_3$
f
f
f
h
h
h
$\mathbf{s}_1$
$\mathbf{s}_2$
$\mathbf{s}_3$
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$

RNN input/output configurations

- Multiple input- Single output
- Single input – Multiple output
- Single input – Single output
- No input – Multiple output
- Multiple input- Multiple output
  - #inputs = #outputs
  - #inputs != #outputs

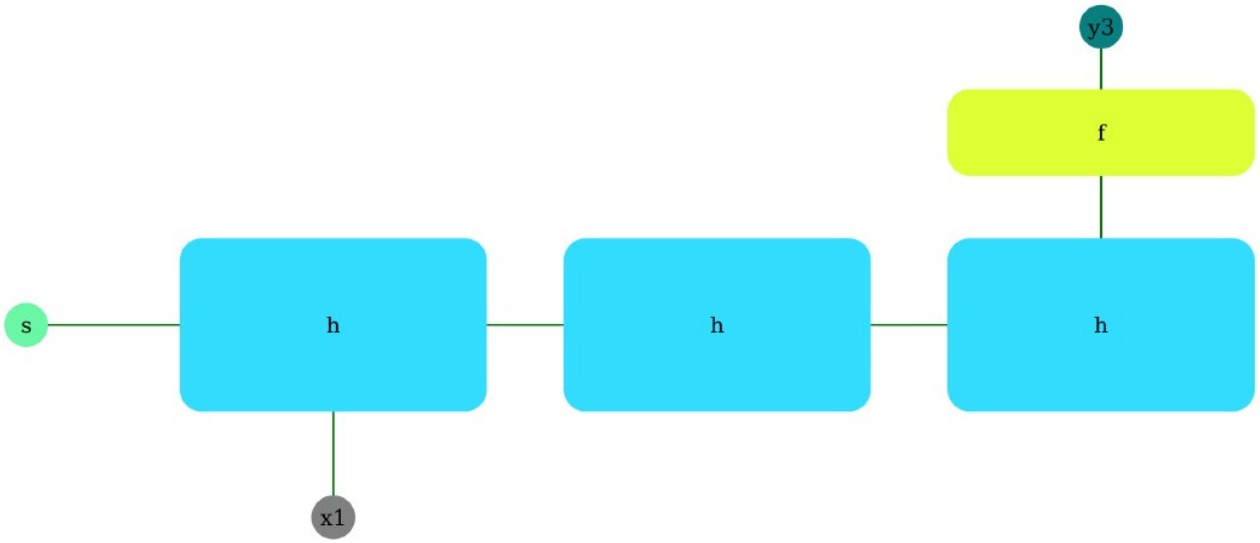
Single input-Multiple output

- Conditional sequence generation
  - Eg: Image captioning



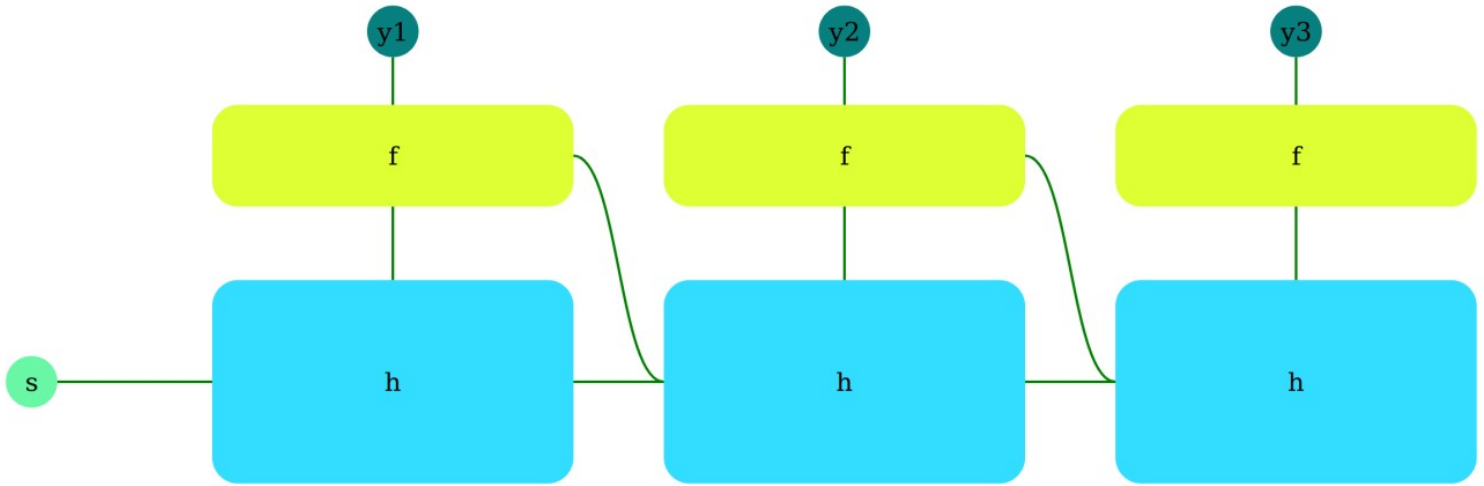
Single input- Single output

- Theoretically possible, but unusual



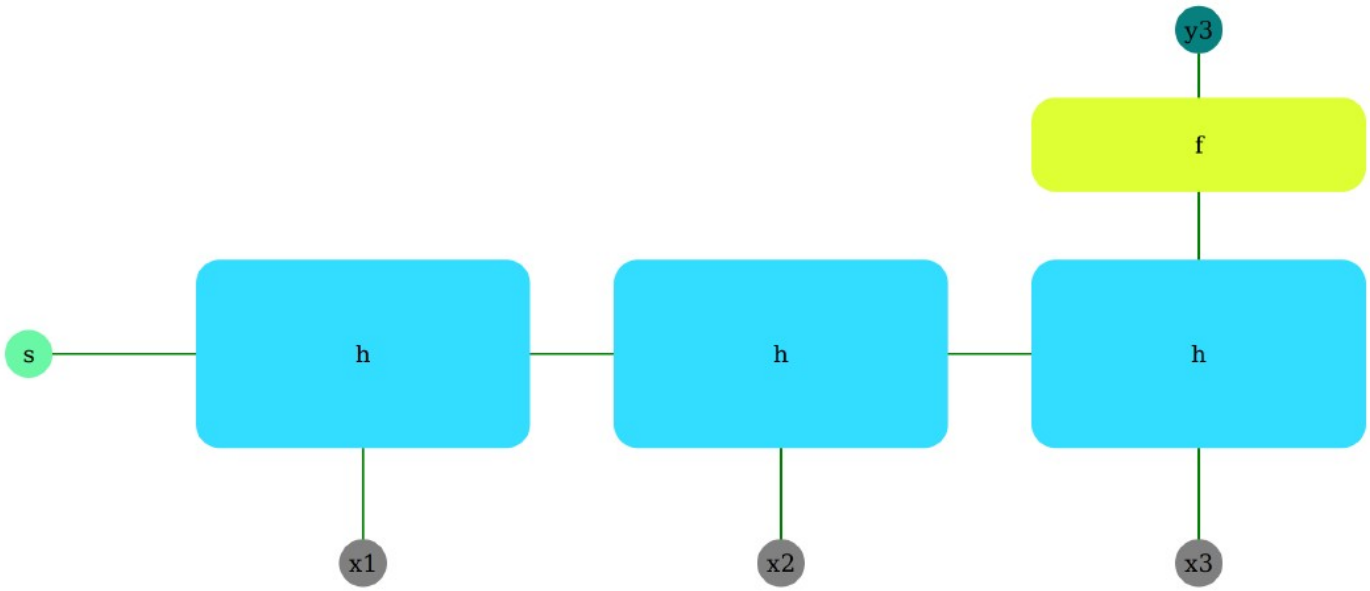
No input – Multiple output

- Generation of sequence samples (Auto-regressive)
  - Eg: Text generation



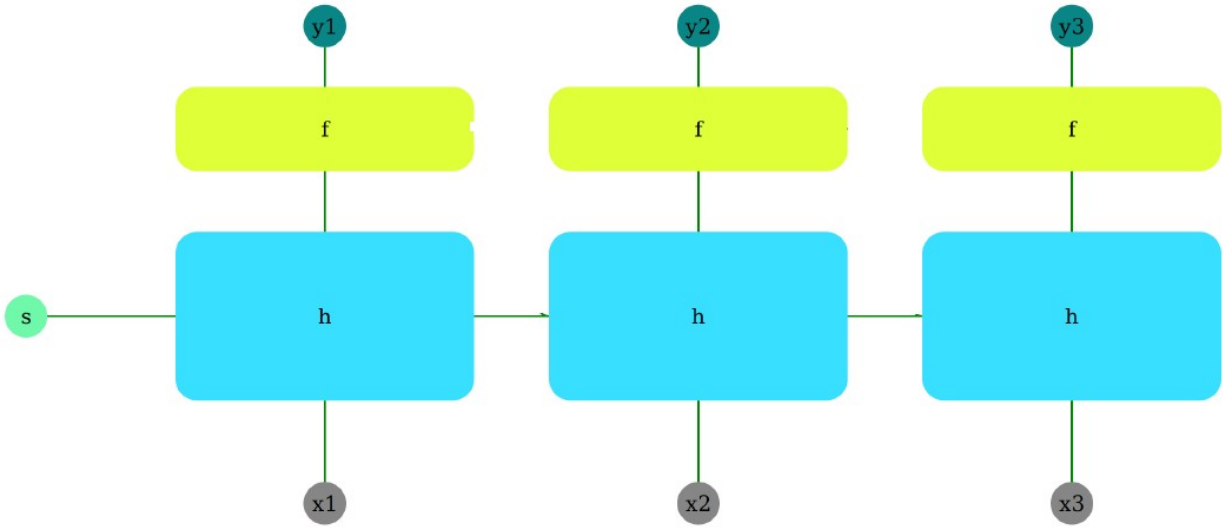
Multiple input- Single output

- Sequence classification
  - Eg; Sentiment analysis



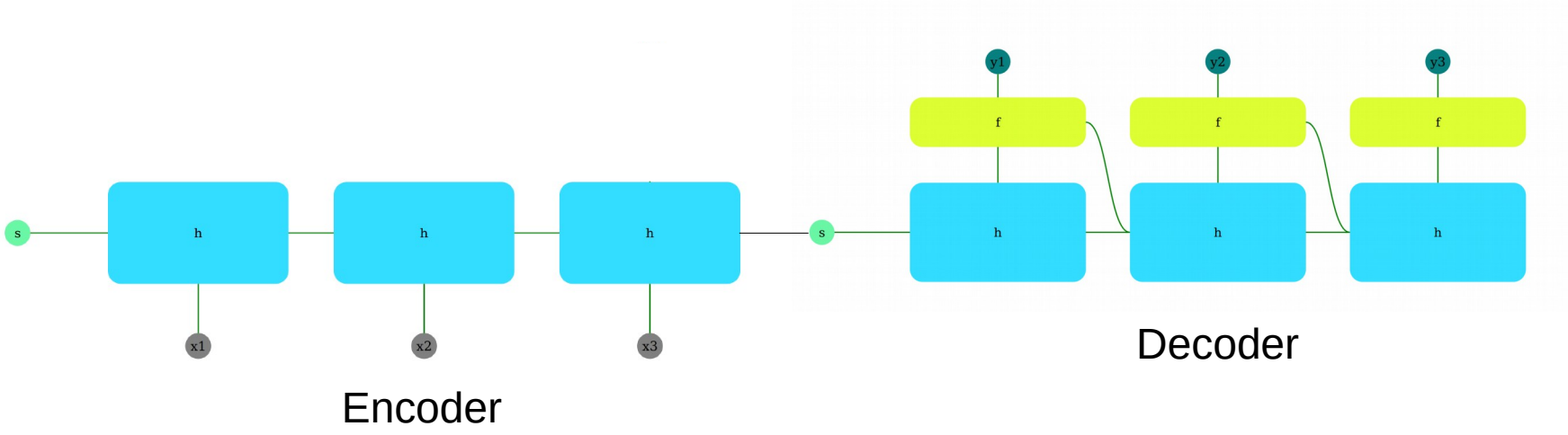
Multiple input – Multiple output (#inputs = #outputs)

- Sequence transformation
  - Eg: noise removal of a time series

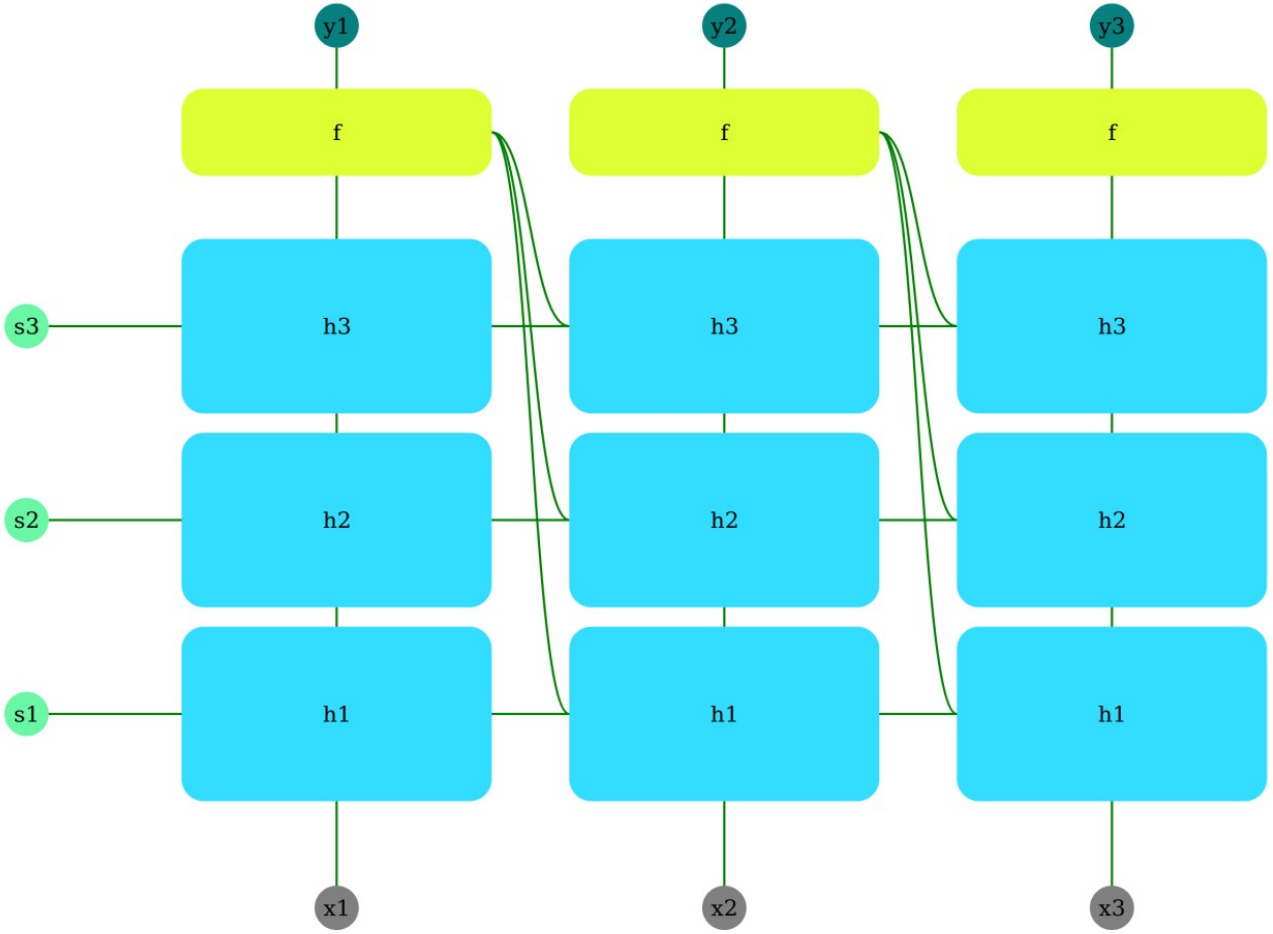


Multiple input- Multiple output (#inputs != #outputs)

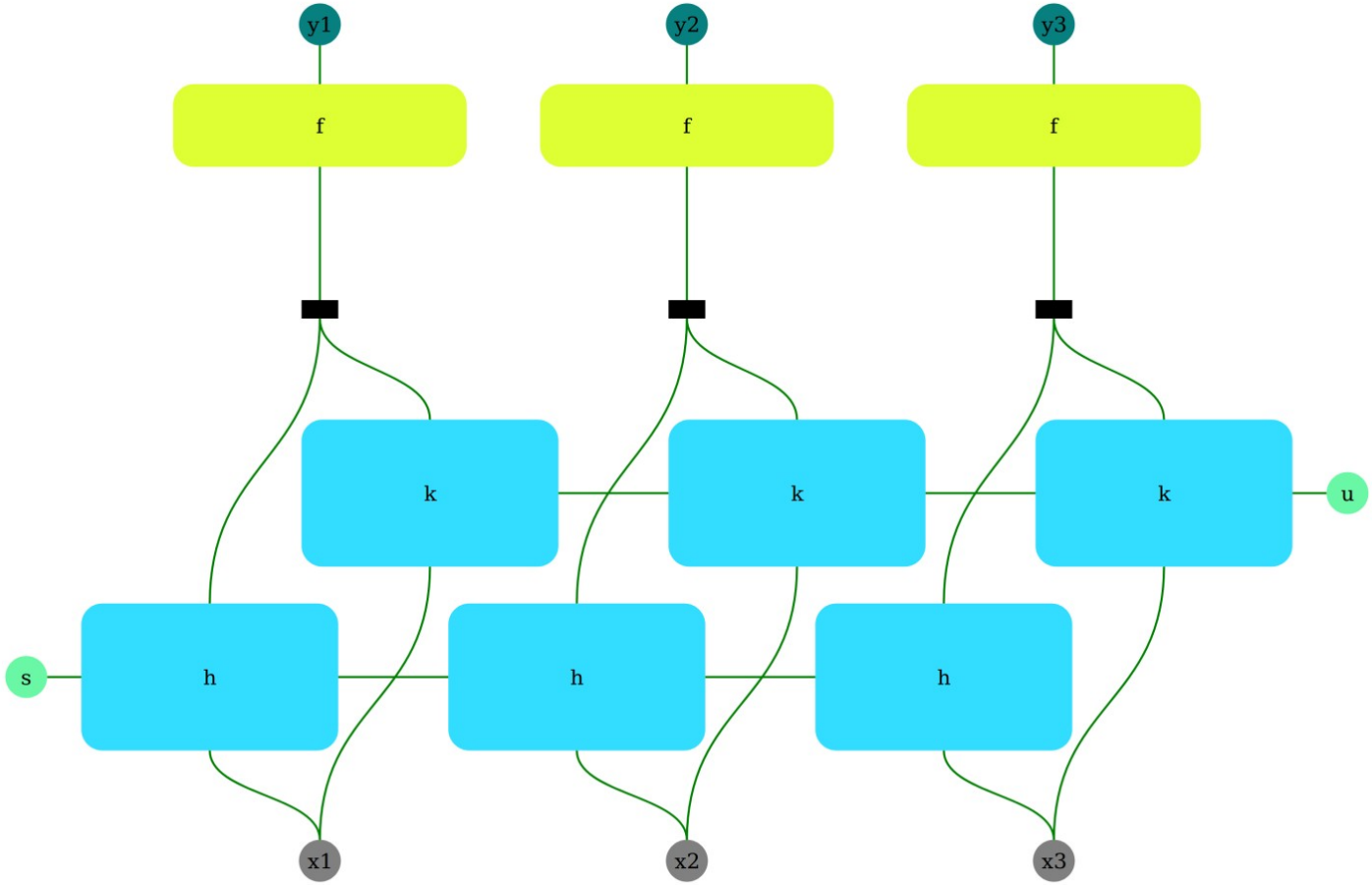
- Unrestricted sequence transformation
- Encoder-Decoder architecture
  - Eg: Machine translation



RNN Configurations I (stacked RNNs)



RNN Configurations II (Bidirectional single layer RNNs)



RNN Configurations III (Bidirectional multi-layer RNNs)