Biological Motivations of Artificial Neural Networks

Blake Mino

Overview

- Background
  - Neurobiology
  - Biological vs Artificial
- How We Think
  - Complexity of everyday actions
  - How we store knowledge
- Biologically Inspired Deep Learning
  - Spiking Neural Networks
  - Spiking Neural Units
  - Implementing ANN techniques on SNN systems

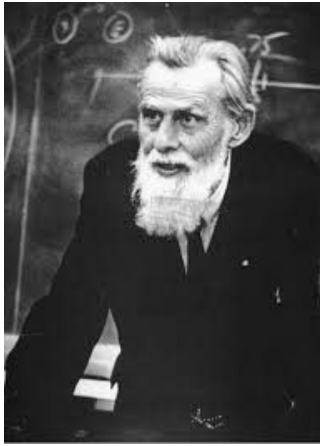
Background

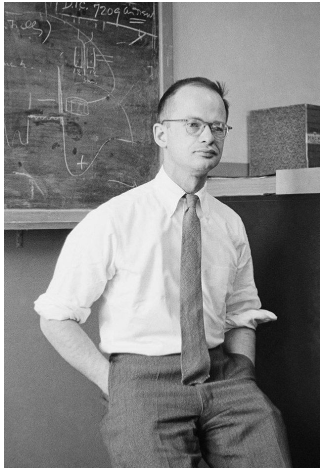
Neural Network Research

- Two types of neural network research
  - 1. Replication of biological learning
    - Essentially understanding how brains work
    - Can we understand the underlying mechanisms of learning through simulations of neuron networks?
  - 2. Creating highly effective learning machines
    - Biological realism imposes unnecessary constraints on learning

Neurobiology

- Warren McCulloch
  - Neurophysiologist and cybernetician
- Walter Pitts
  - Logician and cognitive psychologist
- 1943 Paper “A Logical Calculus of the Ideas Immanent in Nervous Activity”
  - First mathematical model of neural network
- 1947 Paper “How We Know Universals: The Perception of Auditory and Visual Forms”



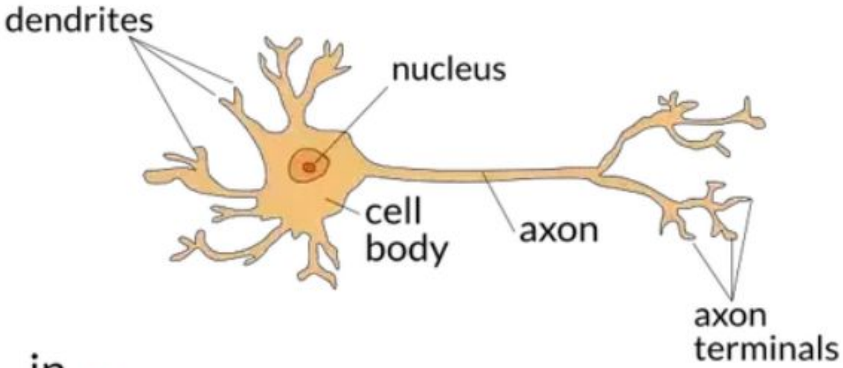


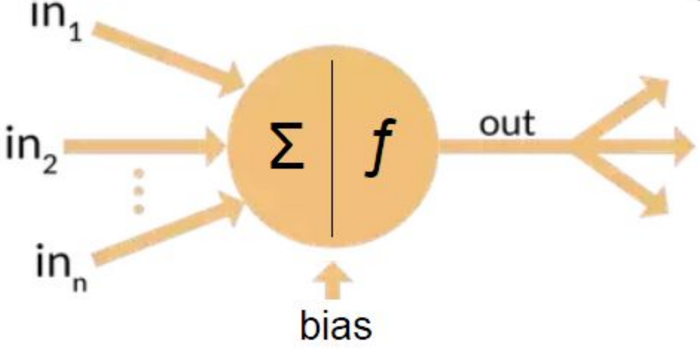
Neural Computation

- Biological Learning Systems (BLS)
  - Composed of a very complex web of interconnected neurons
  - There are, on average, 86 billion neurons in the human brain
    - Each one connected to ~10^4 others
- Artificial Neural Networks (ANN)
  - Inspired by BLS
  - Composed of a interconnected set of simple units
  - Each unit intakes several inputs and produces a single output
    - The single output then becomes the input to many other units

Biological vs Artificial Neurons

- Biological neurons
  - Signals received from the dendrites, processed in the nucleus, and sent out through the axon
  - Some signals are more important, and therefore trigger neurons to fire faster
  - Connections can grow stronger, weaker, break, and form
- Artificial neurons
  - Receive inputs, process weighted inputs through some function, and send the result out





Biological vs Artificial Computing

- Switching time
  - Fastest neuron switching time is $\sim 10^{-3}$ s
  - Computer switching speed $\sim 10^{-10}$ s
    - Much faster than humans
- Human information processing speed
  - Humans can make certain decisions in 10^{-1} s
    - For example: visually recognizing your mother
  - Given neuron switching speed, there cannot be more than a few hundred steps in that time
    - Implies information processing abilities of biological systems are highly parallelized distributed over many neurons

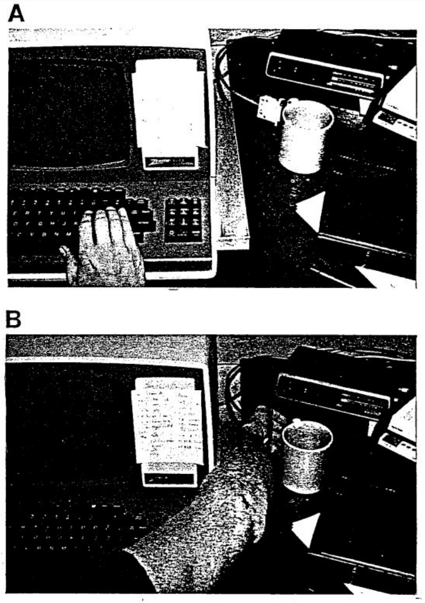
How We Think

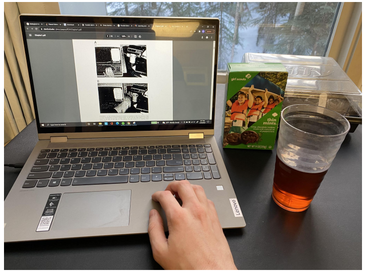
Why do we want to mimic the Brain?

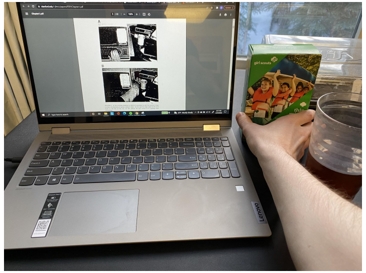
- Are people smarter than computers?
  - Better at problem solving
  - Better at perceiving objects and relationships
  - Better at understanding language and context
  - People can simultaneously consider many pieces of information and constraints, even if they are ambiguous
  - Worse at computation

Example: Reaching and Grasping

- Incredibly simple to us, but in reality requires lots of different considerations
  - Position of object, our current posture, other objects we are already holding, obstacles in the way, the size, shape, and weight of the object

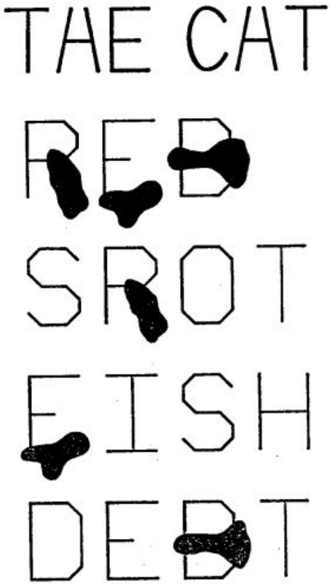






Example: Syntax and Semantics

- Consider the sentences:
  - “I saw the Grand Canyon flying to New York.”
  - “I saw the sheep grazing in the field.”
- Consider these words:



Understanding from Multiple Sources of Knowledge

- It appears that we, as humans, know a lot about a large number of different standard situations. It is suggested that we store this knowledge in terms of structures known as scripts, frames, or schemata
  - These structures form the basis of comprehension
- Most situations cannot be comprehended from just a single script, rather they are understood from the interplay between many different sources of information
  - Example: Child's birthday party at a restaurant

 +  +  = 

Parallel Distributed Processing

- Multiple pieces of information constraining others and being constrained by them
- PDP models assume that information processing takes place through the interactions of a large number of simple processing elements called units
  - In one example the units could stand for possible hypotheses about things like the letters in a display or the syntactic roles of words in a sentence
    - Activations = strengths associated with the different possible hypotheses
    - Interconnections = constraints known to exist between the hypotheses

Biologically Inspired Deep Learning

Biologically Inspired Neural Dynamics

- Improvements to ANN
    - Most advances in the field since the original model for ANN have been largely non-biologically inspired
        - Backpropagation for example
        - Many improvements have led to human-like performance of ANN for certain tasks, but requiring more power than the ~20 W required by the human brain
- Understanding the brain
    - Using more biologically realistic dynamics
        - Spiking Neural Networks (SNN)
    - SNNs are more power efficient
- Bridging the gap
    - Applying deep learning advances from ANN research to an SNN system

Spiking Neural Networks

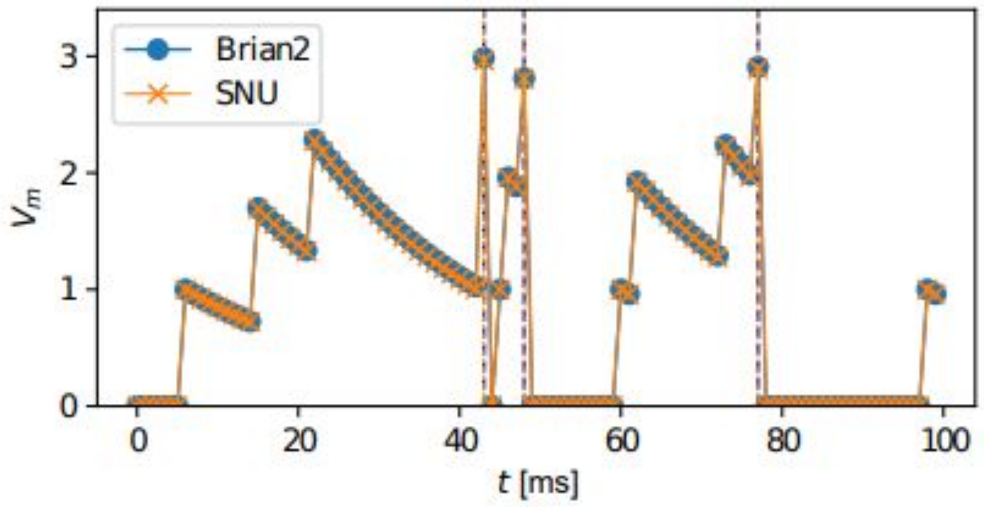
- Spikes = sparse, asynchronous voltage pulses
  - Used to compute and propagate information
- Spike-Timing Dependent Plasticity (STDP)
  - Biologically inspired learning rule for correlation detection
- Increased parallelism and reduced energy consumption
- While successful in a few specific applications, they lack adaptability for training architectures for other applications
  - How to adapt them for more common machine learning tasks?
- Recent research focussing on using STDP to perform backpropagation or otherwise deriving backpropagation for SNN

Modeling Spiking

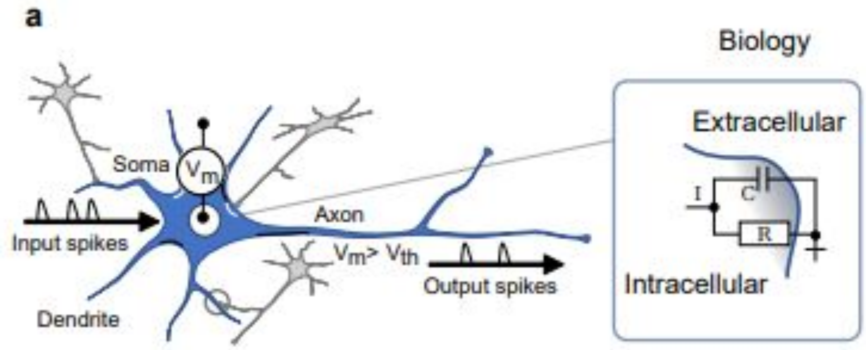
- Leaky Integrate-and-Fire (LIF) neurons
  - $\tau \frac{dV_m(t)}{dt} = -V_m(t) + RI(t)$ or the discretized version:
  - $V_{m,t+1} = \frac{\Delta T}{C} I_t + V_{m,t}(1 - \frac{\Delta T}{\tau})$
  - Where V_m is the membrane potential, *R* is the resistance of the soma, *C* is the capacitance of the cell soma, τ = *RC* is the time constant of the neuron, and *I(t)* is the incoming current from the synapses
- A neuron receives spikes, modulates them by weights to provide input current to soma

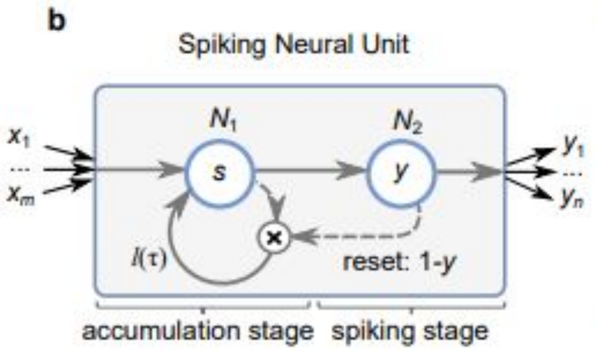
Modeling Spiking cont.

- V_m increases every time an input spike is received, then decays over time
- When $V_m > V_{th}$ an output spike is emitted and V_m is reset



Spiking Neural Units





- a: Biological neurons receive input spikes which are built up into the membrane potential V_m until it reaches the spiking threshold V_{th}. At this point the neuron emits output spikes through the axon.
- b: Spiking Neural Units (SNU) model the spiking neural dynamics using two ANN neurons to perform the integration and emission of output spikes

LIF implementation in SNUs

- LIF parameters directly relate to SNU parameters

LIF nomenclature	Correspondence	SNU nomenclature
membrane potential	$V_m = s$	unit state
spiking threshold	$V_{th} = -b$	bias
input soma current	I	
	$\Delta T/C\ I = Wx$	weighted input
membrane time constant	τ	
	$(1-\Delta T/\tau) = l(\tau)$	state decay

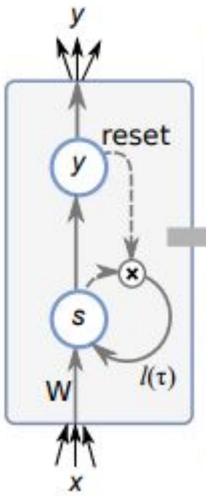
$$V_{m,t+1} = \frac{\Delta T}{C} I_t + V_{m,t}\left(1 - \frac{\Delta T}{\tau}\right)$$

$$s_t = g\left(W x_t + l(\tau) \odot s_{t-1} \odot (1 - y_{t-1})\right)$$
$$y_t = h(s_t + b),$$

- S_t is the vector of internal state variables calculated by the N_1 subunit, y_t is the output vector calculated by the N_2 subunit, g is the input activation function, h is the output activation function, W is the synaptic weight matrix, $l(\tau)$ is the self-looping weight, and $\odot$ denotes a point-wise vector multiplication
- $l(\tau)$ performs a discrete time approximation of the membrane potential decay. If it is set to $l(\tau) = 1$, then the state does not decay
  - This is the "Leaky" part of LIF

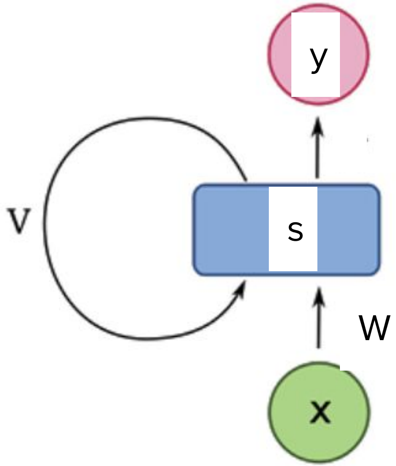
Comparing Artificial Neurons

SNU



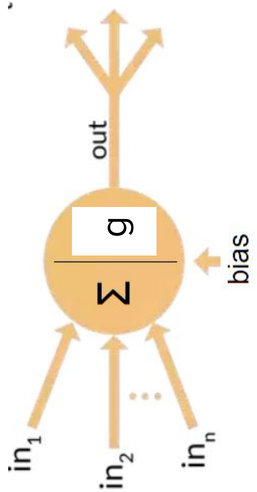
- $s_t = g(W*x_t+l(\tau)*s_{t-1}*(1-y_{t-1}))$
- $y_t = h(s_t+b)$

RNN Unit



- $s_t = g(W*x_t+V*s_{t-1})$
- $y_t = f(s_t)$

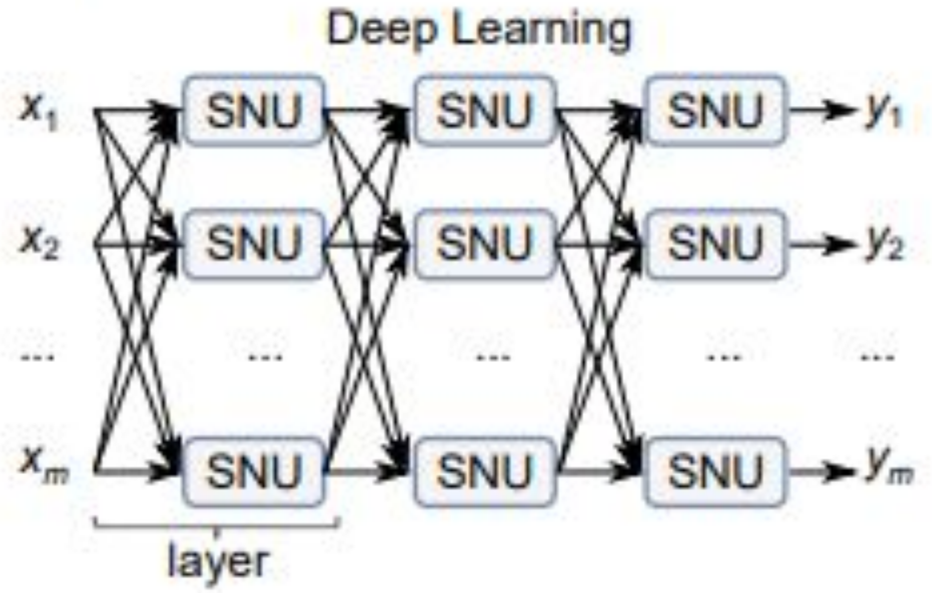
ANN Unit



- $y = g(W*x)$

SNUs in ANN frameworks

- SNUs can be applied to construct deep learning architectures

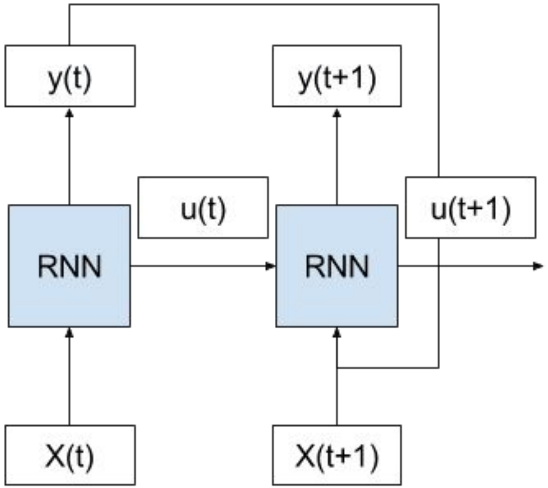


Soft SNUs

- Traditionally, information in SNNs is transmitted via binary spike values, all or nothing
    - Input data is binarized, and the step activation function of subunit N_2 is used to determine binary output
- soft SNUs (sSNU) generalize the dynamics to non-spiking ANNs
    - Input data no longer required to be binarized
    - N_2's activation function can be continuous
    - Retains temporal integration dynamics
    - Allows for variable membrane potential reset depending on the magnitude of the output spike

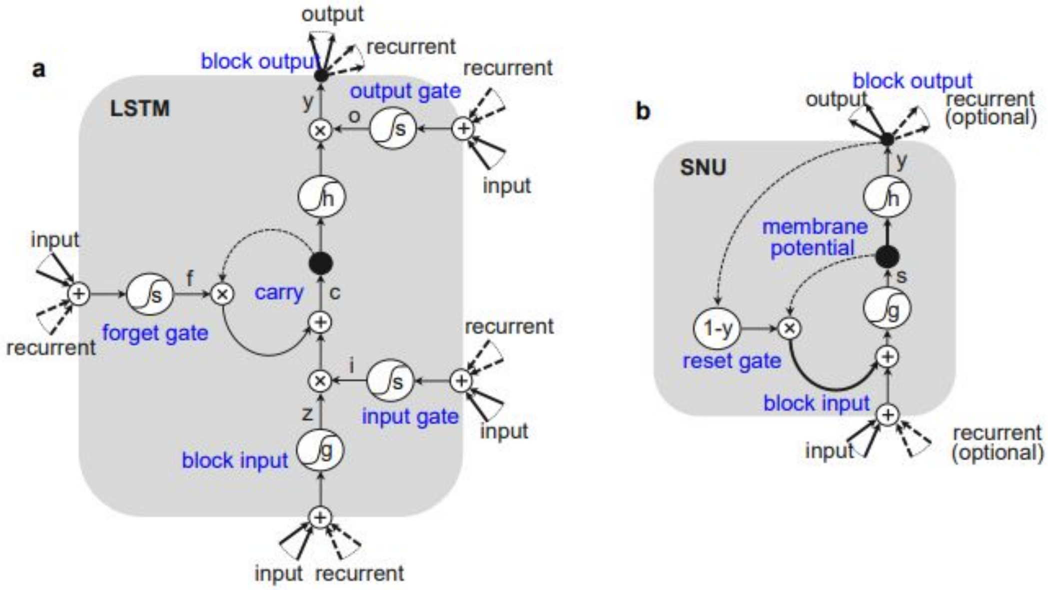
Recurrent Neural Networks

- Unlike Traditional ANNs which primarily intake static data, RNNs intake data with a temporal context.
- Several variations
    - Long-Short-Term Memory (LSTM)
    - Gated Recurrent Units (GRU)
- Unfolding over time
    - Visualizing the passing of information between time steps



Spiking Generalizations of ANN

- SNUs are comparable to Long-Short-Term Memory (LSTM) units
  - Both maintain temporal context
  - Both have recurrent loops

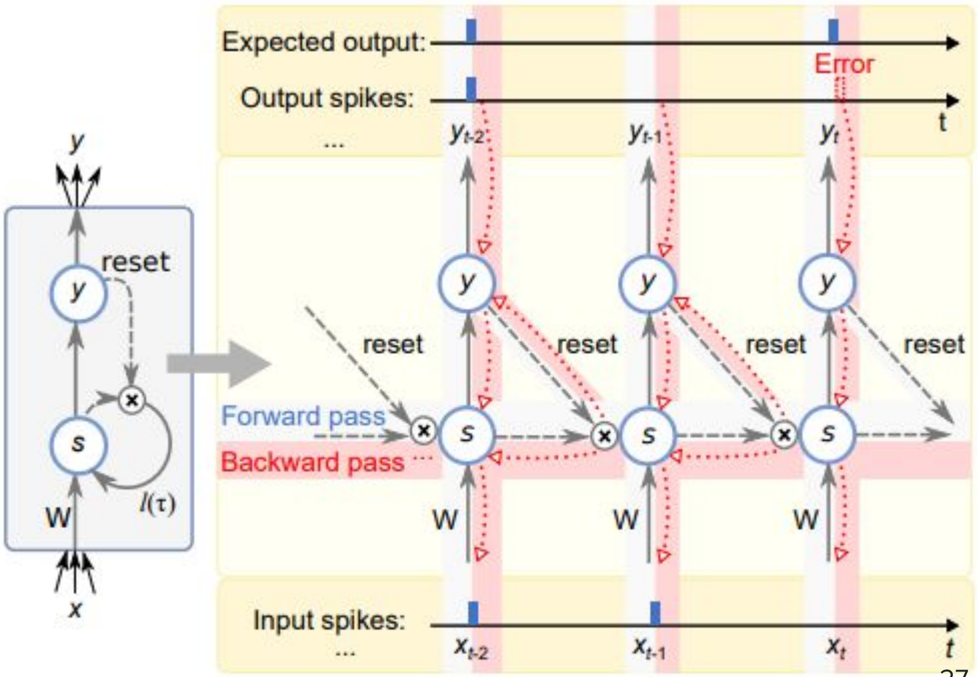


d

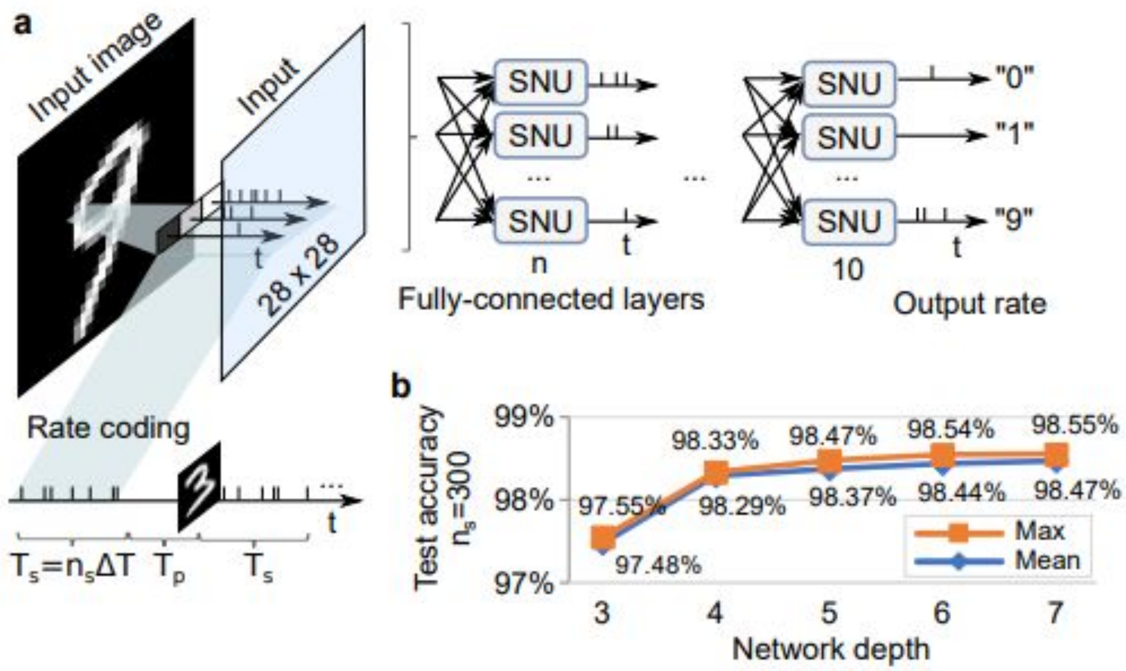
Neuron type	# of parameters
ANN neuron	$(m+1)n$
RNN neuron	$(m+1)n+n^2$
LSTM	$4((m+1)n+n^2)$
GRU	$3((m+1)n+n^2)$
SNU	$(m+1)n$
recurrent SNU	$(m+1)n+n^2$

Training with Backpropagation Through Time

- While a SNU-based network is feed-forward, the individual units use self-looping recurrent connections
  - Can be trained with Backpropagation Through Time (BPTT)
- SNU structure is unfolded over time
- Only involves local state of the neuron
  - RNNs require unfolding of all units in a layer



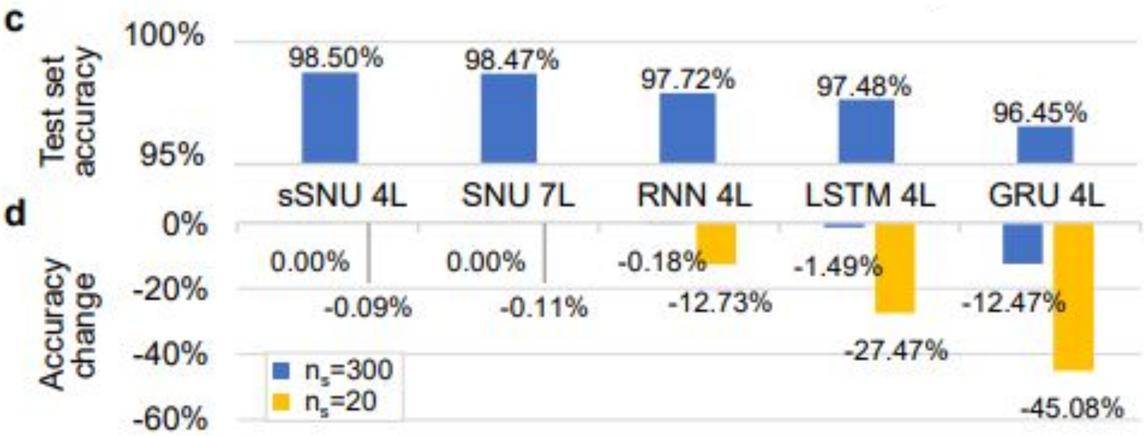
Handwritten Digit Recognition



- Input data coded using rate-coding
  - Greyscale info of each pixel conveyed via firing rate of input spikes
  - Generated for period T_s which corresponds to n_s discrete-time steps
- Depth increases accuracy

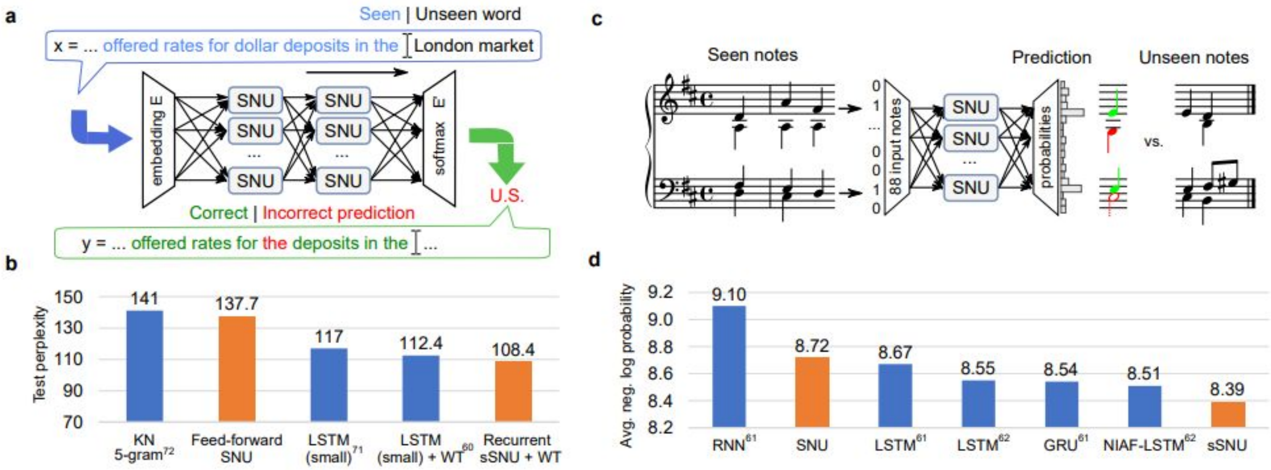
Handwritten Digit Recognition cont.

- sSNU-based network performed the best
- SNU-based networks experienced least loss in accuracy when presented with a continuous, uninterrupted stream of images



Language and Music Prediction

- Language prediction
  - Predicting the next word based off of previous words
- Music prediction
  - Predicting the next note based off of previous notes



Summary

- Goals of neural computation
  - Understand how the brain actually works
  - Understand a new style of computation
    - Inspired by neurons and how their connections change and adapt
    - Parallel processing over sequential computation
    - Should be good at things brains are good at
      - Vision
    - Should be bad at things brains are bad at
      - Quick calculations
- How humans think
  - Many parameters that constrain others while simultaneously being constrained by them
  - Highly parallelized
- SNNs and ANNs
  - SNUs can act as a novel ANN unit and benefit from deep learning techniques

Nengo

- Python package for creating and testing neural networks
  - Can model SNNs as well
- https://www.nengo.ai/

Resources

- Srihari, Sargur. "ML 5.0 Neural Networks: Biological Motivation." *YouTube*, uploaded by Sargur Srihari, 21 Oct. 2020, https://www.youtube.com/watch?v=bU7CTi4EXeE&ab_channel=SargurSrihari
- Rumelhart, David E., and MacClelland, James L. *Parallel Distributed Processing Explorations in the Microstructure of Cognition*. MIT Pr, 1986
- Woźniak, S., Pantazi, A., Bohnstingl, T. et al. Deep learning incorporating biologically inspired neural dynamics and in-memory computing. Nat Mach Intell 2, 325–336 (2020). https://doi.org/10.1038/s42256-020-0187-0
- Eliasmith, Chris. "Spiking Neural Networks for More Efficient AI Algorithms." *YouTube*, uploaded by WaterlooAI, 31 Jan. 2020, https://www.youtube.com/watch?v=PeW-TN3P1hk