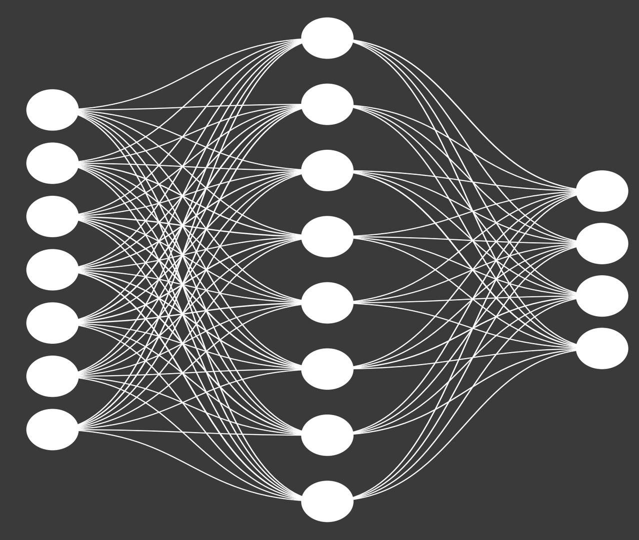
Neuroscience Meets Machine Learning

Simulating biologically accurate neurons for efficient speech recognition.

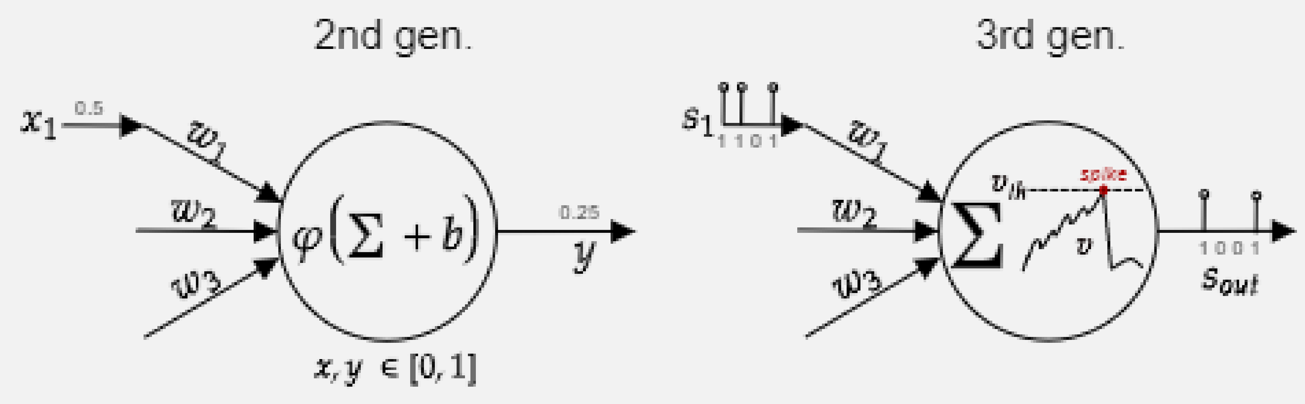
What is a Neural Network?

- A collection of neurons connected by synapses.
- Synapse weights determine strength of neuron connections.
- Weights are adjusted by algorithms such as backpropagation.
- Weights are adjusted so the network can model some system.
- The output of a neural network is called "inference“.
- Layers in recurrent neural networks feed into themselves.



What is a Spiking Neuron?

Biologically inspired spiking neurons - third generation neurons - behave very differently than the standard artificial neurons - second generation neurons.



(a) (b)

(a) Second generation neurons output a weighted sum of its inputs added to a bias. The output is transformed by an activation function and passed on to the next neurons.
(b) Third generation - neuromorphic - neurons receive spikes through weighted synapses. Spikes cause the internal voltage of the neuron to increase. When the internal voltage reaches a threshold, the neuron outputs a spike to the next neurons.

Why Spiking Neural Networks?

The human mind - 20 Watts:
- Perceives world through visual recognition.
- Guides actions through motor control.
- Extracts meaning from sound.
- Keeps ceaseless functioning of our internal systems.
- Gives rise to the phenomenon of consciousness.
- Weaves together natural language to convey meaning.

A GPU - 400 Watts:
- Runs a large language model.
- Produces AI slop and hallucinates.

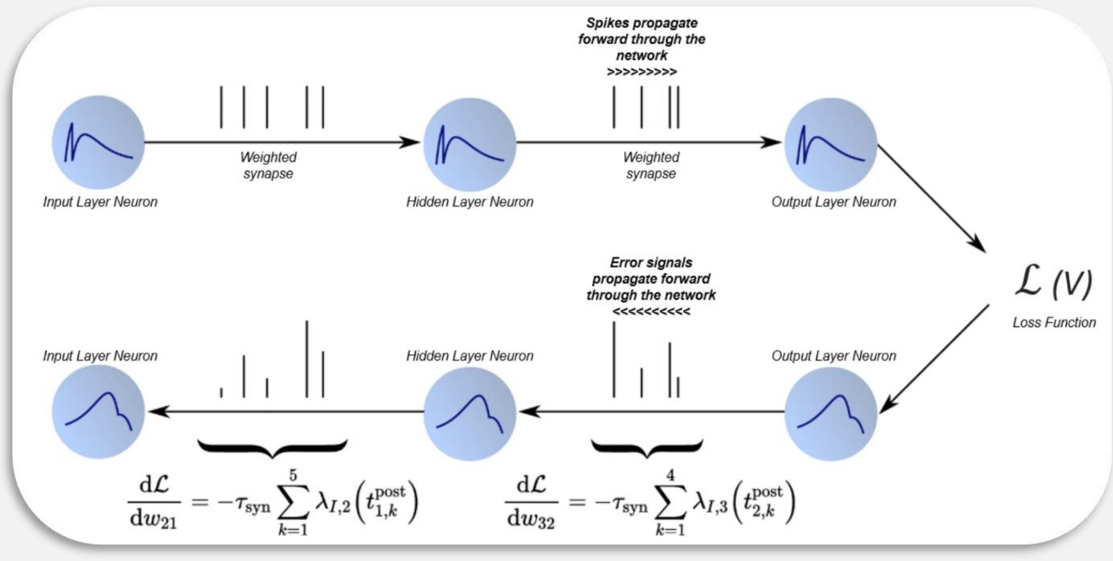
The global power consumption of AI increasing exponentially, spiking neural networks offer an efficient alternative. Studies have shown that running AI inference on neuromorphic platforms uses 100x less energy and runs 50x faster than 2nd generation equivalents.

Challenges to Training Spiking Neural Networks.

- Less researched.
- Non-differentiable spikes.
- Harder to parallelise.

Training Spiking Neural Networks Using "Eventprop*“*.

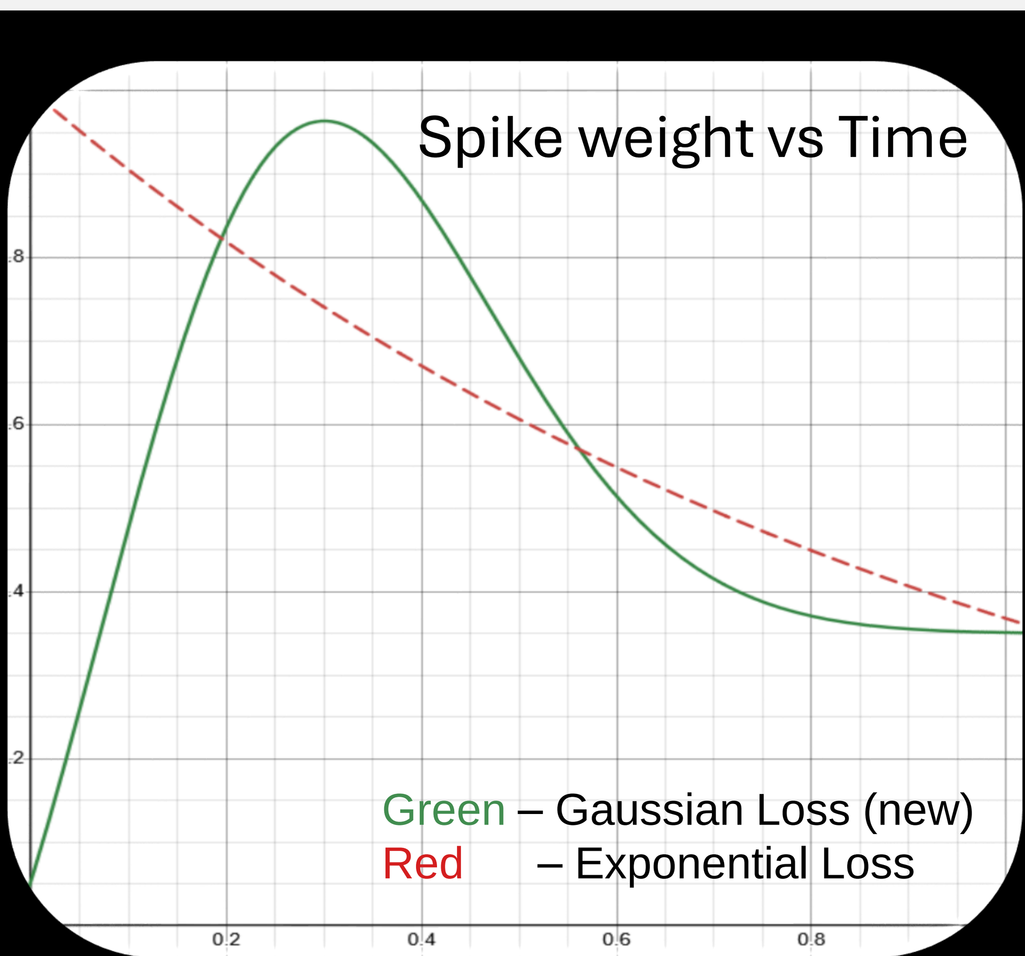
- A novel method for training SNNs.
- An application of adjoint method from optimisation theory.
- Calculates the exact gradient, despite non-continuity of spikes.
- Applies precise backpropagation.
- 4x faster and uses 3x less memory than top competitors.
- State-Of-The-Art performance for spiking speech dataset - SHD.

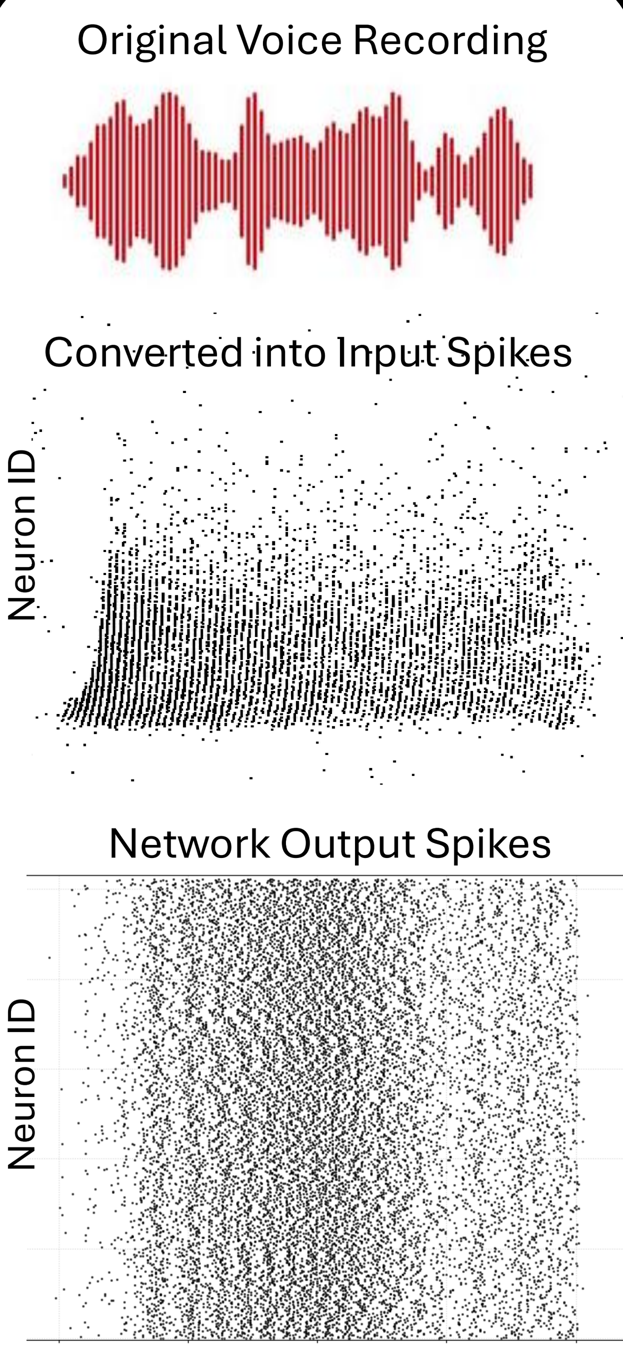


The Need for a Novel Loss Function.

- Current loss function has achieved state-of-the-art performance on spiking speech dataset - SHD.
- Current loss function fixes inefficiencies caused by spikes occurring near the end of the recording.
- However, it emphasises very early spikes the most. I hypothesised that this hinders learning rate since spikes happening instantaneously to the recording starting are caused by noise.

$$\mathcal{L}_{sum-exp} = -\frac{1}{N_{batch}} \sum_{m=1}^{N_{bathc}} log\left(\frac{exp(\int_0^T e^{-t/T} V_{l(m)}^m(t)dt)}{\sum_{k=1}^{N_{out}} exp(\int_0^T e^{-t/T} V_k^m(t)dt)}\right)$$

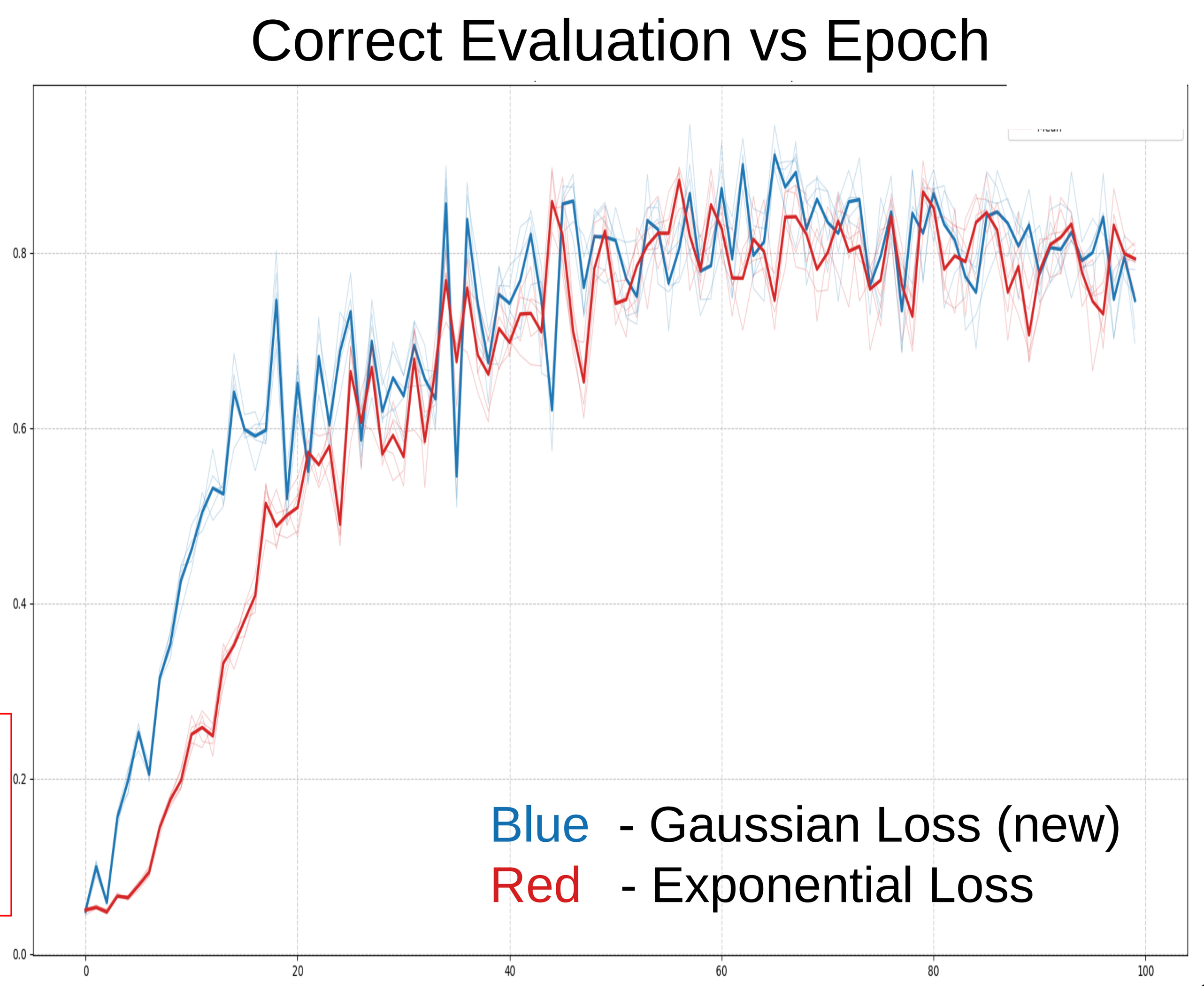




Deriving and Implementing a Novel Loss Function

- I implemented a Gaussian-based loss function, aiming to minimise the effect of noise caused by initial spikes at the output.
- I derived the mathematical Eventprop scheme necessary to propagate the error signals and gradient backwards in the network.
- My loss function greatly increased the learning rate over the previous state-of-the-art method.

$$\mathcal{L}_{sum-gaussian} = -\frac{1}{N_{batch}} \sum_{m=1}^{N_{bathc}} log\left(\frac{exp(\int_0^T o + (t-\mu)e^{-\frac{(t-\mu)^2}{2\sigma^2}} V_{l(m)}^m(t)dt)}{\sum_{k=1}^{N_{out}} exp(\int_0^T o + (t-\mu)e^{-\frac{(t-\mu)^2}{2\sigma^2}} V_k^m(t)dt)}\right)$$



Rihards Klotins

Supervisor: Dorian Florescu
2025

UNIVERSITY OF BATH