The influence of spike minimisation on classification performance and predictive-coding properties in a multi-compartmental energy SNN

Melle Bolding, 15302741

Lab report for Neural Dynamics and Deep Learning

December 20, 2024

Abstract

In this study, I investigate the influence of the addition of a spike loss term on classification performance and predictive-coding properties in the multicompartamental energy SNN of [1]. Using the MNIST dataset, I found that the addition of a minimal spike loss term ($\alpha_{spike} = 0.01$) slightly improved classification and generative performance of both control and energy-optimised models. However, no conclusive evidence was found that indicated that the addition of a minimal or moderate spike loss term induced predictive-coding properties in control models.

Contents

1 Introduction

The adult human brain consumes approximately 20 watts [2]. Given the computational architecture and capabilities, this amount of energy expenditure is remarkably small. This indicates that optimisation of energy consumption is an important property of the human brain. In comparison, most artificial neural networks that attempt to replicate the capacities of the brain expend significantly more energy. However, minimal energy consumption is a hallmark feature of a more biologically plausible class of neural networks, spiking neural networks (SNNs) [3].

A currently influential theoretical framework for understanding the hierarchical processing of sensory information in the brain is known as predictive-coding (PC). This framework postulates that the brain builds a generative world-model that is able, through hierarchically organized brain areas, to predict sensory input. Two empirical characteristics of PC are generative features and a difference in response to expected and unexpected stimuli [4].

Returning to energy consumption in the brain, the main caloric cost is the propagation of action potentials (i.e. spikes). Specifically, most energy is spent through the activation of ion pumps in order to restore ion gradients across neuronal membranes. Furthermore, a significant amount of energy is spent by the release and recycling of neurotransmitters at synapses [5]. Both processes are directly related to the transmission of action potentials.

In [1], the energy-optimised model uses the inter-compartmental voltage difference as energy loss term (see **2 Methods**). It finds that this term induces PC properties and additionally reduces the number of spikes substantially. However, classification accuracy significantly decreases as compared to the control model (from 98.18% to 97.59%). This raises the question whether this drop in performance is due to the reduced number of spikes or whether this is inherent to the energy-optimised model. Introducing a spike loss term to the control model would equal the number of spikes and enable a fair comparison between models. Furthermore, as argued before, a more biologically plausible way of optimising energy in the brain would be to penalise spiking directly.

In this work, I investigate the effect of a simple spike loss term on classification performance and PC properties of both the control and energy-optimised models of [1]. Using the MNIST dataset, I find that the addition of a minimal spike loss term ($\alpha_{spike} = 0.01$) slightly improves classification and generative performance of both control and energy-optimised models. However, no conclusive evidence is found that indicates that the addition of a spike loss term induces predictive-coding properties in *control* models.

2 Methods

The addition of a spike loss term

2.1 Original model

In [1], a simple multi-compartment spiking neuron model that resembles a pyramidal cell of the human cortex is used (Figure 1). A single neuron contains both a somatic and dendritic (apical tuft) compartment. The dendritic compartment takes inputs from higher areas in the network. Meanwhile, the soma integrates feedforward information. The reservoir potential of dendritic compartment unidirectionally affects the reservoir potential of the somatic compartment. The spiking mechanics are modeled by the adaptive leaky-integrate-and-fire (ALIF) model. The fully connected network is composed of three layers of spiking neurons, without lateral connections, and an non-spiking output layer.

The network is trained on MNIST handwritten digit classification (60,000 training and 10,000 test samples) using a combination of the online learning algorithm forward propagation through time (FPTT) and surrogate gradients. The loss function that is optimised throughout the training is given by Eq.1.

$$\mathcal{L}_t = \mathcal{L}_{clf,t} + \alpha_{reg}\mathcal{L}_{reg,t} + \alpha_E \mathcal{L}_{E,t}, \qquad (1)$$

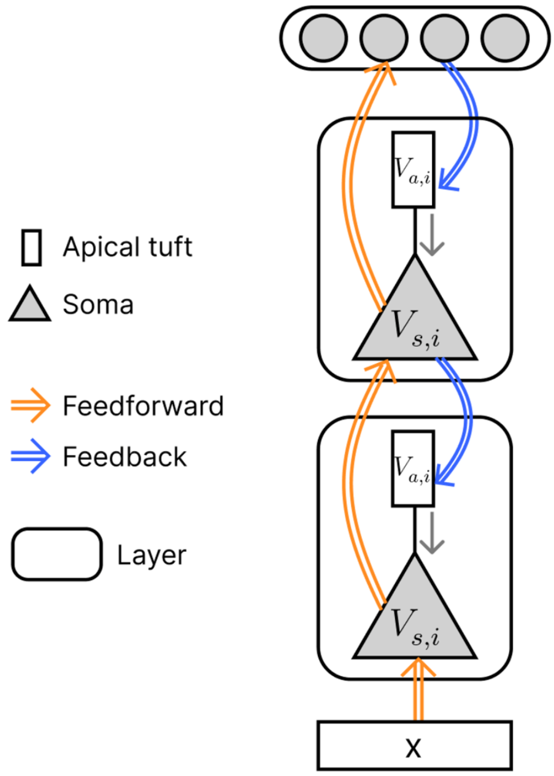


Figure 1: Schematic of a 2-layer spiking network with multi-compartmental neurons. The apical tuft and soma are shown. Additionally, the feedforward and feedback connection are shown in orange and blue respectively. Lastly, the output layer (top) is composed on non-spiking neurons that determine the predicted class.

where $\mathcal{L}_{clf,t}$ represents the task-related classification loss (negative log-likelihood), $\mathcal{L}_{reg,t}$ gives the dynamic FPTT regularizer, and $\mathcal{L}_{E,t}$ is the voltage energy term as defined by Eq.2.

$$\mathcal{L}_{E,t} = \left(\sum_{l} \sum_{i} g(V_{a,i}^{l}(t), V_{s,i}^{l}(t)) \right) / N, \quad (2)$$

where N is the total number of neurons and g the absolute difference between voltages, which is given by $g(V_{a,i}^{l}(t), V_{s,i}^{l}(t)) = |V_{a,i}^{l}(t) - V_{s,i}^{l}(t)|$.

Regarding the energy parameter α_E, the control and energy-optimised model have $\alpha_E = 0$ and $\alpha_E = 0.05$, respectively. Further initialisation values and hyperparameters can be found in [1].

2.2 Spike loss term

To determine the effect of penalising spiking, a spike term, $\alpha_{spike}\mathcal{L}_{spike}$, can be added to the overall loss term (Eq.1). This is done in a straight-forward manner,

$$\mathcal{L}_{spike} = \frac{\sum_{L=1}^{3} n_L}{BN}, \quad (3)$$

where n_L gives the number of spikes per layer and B the batch size of the dataset. Furthermore, the parameter α_{spike} is introduced to allow for different amounts of spike penalising.

3 Results

Control and energy-optimised spike-minimisation models

In this section the results of six different models are presented. These models are a combination of control and energy-optimised models with no, minimal ($\alpha_{spike} = 0.01$), and moderate ($\alpha_{spike} = 0.1$) spike minimisation.

3.1 Spike rate & energy per neuron

In Figure 2, the mean combined spike rate across the training epochs is shown for six different control and energy-optimised models. From the figure it is clear that the addition of the spike loss term indeed decreases spike rate in both the control and energy-optimised models. Furthermore, increasing α_{spike} indeed increases the amount of spike suppression. Additionally, the figure clearly shows that during training towards convergence, spike rate decreases across all models. Then, after convergence, the model settles in a stable state and spike rate remains constant. However, the exception to this is the original energy-optimised model. Here, we see a slight bump in spike rate after convergence. Moreover, the variance of this model is significantly larger than the variance of other models.

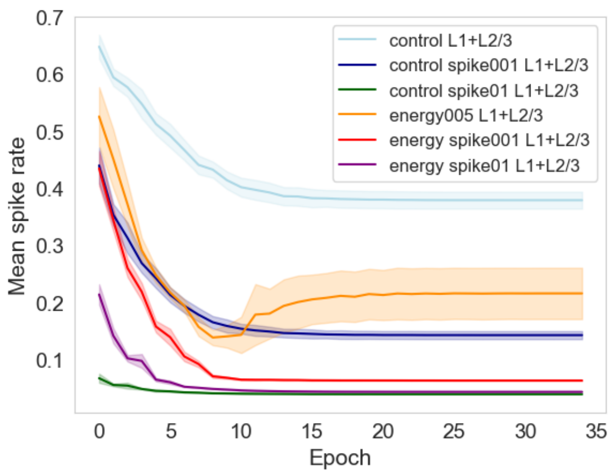
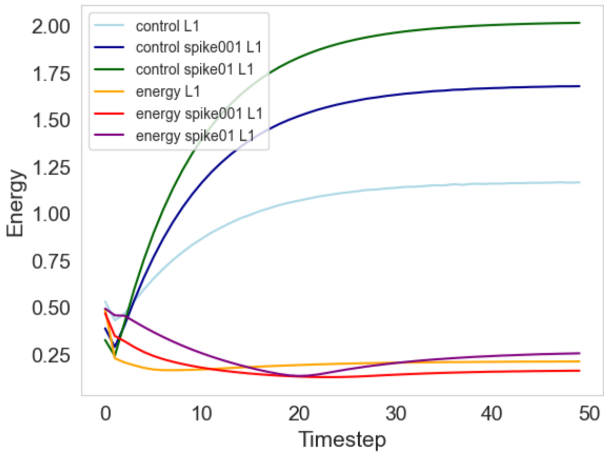
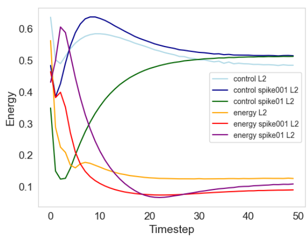


Figure 2: Mean combined spike rate (divided by network dimensions and timesteps) for six different control and energy-optimised models. On the x-axis the number of training epochs is shown. The shaded area represents the standard deviation from six different seeds per model.

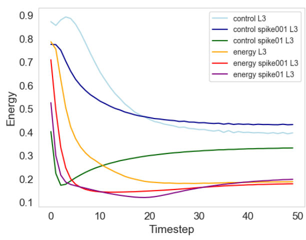
In Figure 3, the average voltage energy per neuron per layer for the different models is shown. We can see that the control model neurons have significantly higher average energy per neuron as the energy-optimised models. Furthermore, we see that the spike loss term does not have homogeneous effect on this energy term. Generally, the average voltage energy per neuron gives a reasonable proxy for generative features. Namely, when the voltage difference between apical tuft and soma is small, the model is presumably making predictions to minimise this difference. Clearly, the spike loss term does not attribute much to this.



(a) Layer 1



(b) Layer 2



(c) Layer 3

Figure 3: Average voltage energy per neuron per layer for six different control and energy-optimised models.

3.2 Classification performances

In Figure 4, the mean test accuracy for the MNIST classification task is shown for six different control and energy-optimised models. Each model was trained with six different seeds. The shaded area represents the standard deviation of these models. From the figure, we can see that the energy-optimised models converge to about 97.50% test accuracy. In comparison, the control models converge to about 98.20%. Furthermore, we see that for the energy-optimised models, the minimal and moderate spike models slightly improve accuracy about equally at convergence. They converge to about 97.65% test accuracy. For the control models, the minimal spike model performs slightly better than the control model at convergence (98.35% test accuracy). In contrast, the moderate spike model performs slightly worse at convergence (98.15% test accuracy).

Lastly, it is noted that the original energy-optimised models have a significantly larger variance than the other models. This could indicate that, without the spike loss term, the energy-optimised occasionally has trouble finding the global minimum in training. This would be supported by the heightened spike rate after convergence as previously seen in Figure 2.

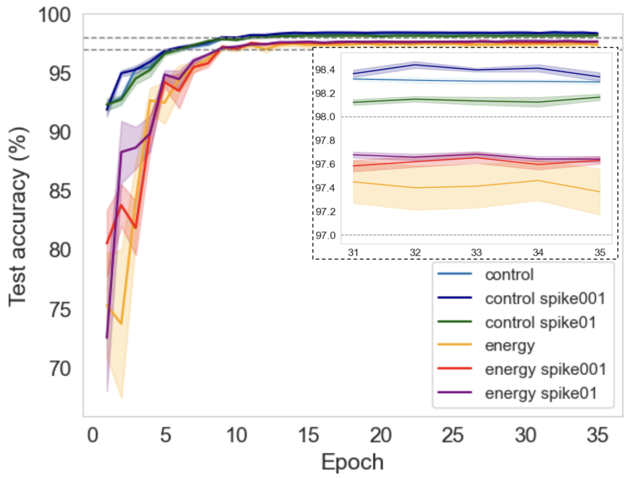


Figure 4: Mean test accuracy for the MNIST classification task for six different control and energy-optimised models. On the x-axis the number of training epochs is shown. The shaded area represents the standard deviation from six different seeds per model.

3.3 Generative features

Generative features are indicative of PC properties. In [1], only the energy-optimised model shows these features. In Figure 5, the pair-wise representational similarity per layer between clamped and normal representations is shown for control and energy-optimised models. Here, the main indicator for generative features is the diagonal in layer 3. We see that the energy-optimised models (left) have this most clearly. Furthermore, minimal spike loss slightly improves the clarity of this diagonal. In contrast, the control models show no clear diagonal with the exception of a minimal signal in the moderate spike loss model.

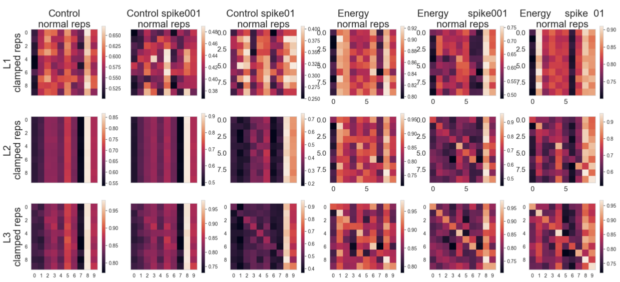


Figure 5: Pair-wise representational similarity of clamped vs normal representations for control and energy-optimised models. This is done per model layer.

In Figure 6, we aim to confirm our observations from Figure 5. Indeed, we see that the energy-optimised models show traces of digit reconstruction. However, no qualitative difference can be observed between the energy-optimised models. Moreover, despite the slight diagonal signal in the control model with moderate spike loss, there is no evidence of generative features among the control models in this figure.

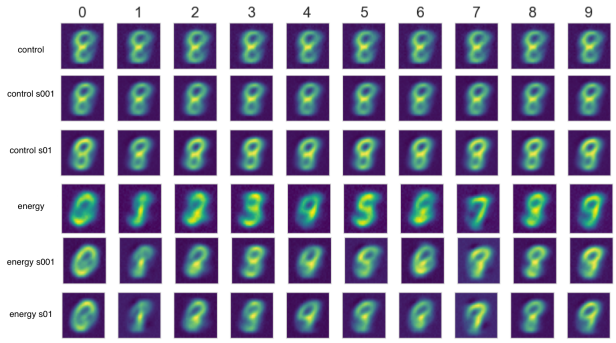


Figure 6: Decoded internal representations without input but with class clamping. Only the energy-optimised models show differences between classes.

4 Discussion & Conclusions

Firstly, we can conclude that including a simple spike loss term (Eq.3) significantly reduces the number of spikes in this SNN. Furthermore, a minimal inclusion of this term ($\alpha_{spike} = 0.01$) improves the classification and generative performances of both the control and energy-optimised model. Notably, as in both models the spike rate was greatly reduced, the efficiency of these models was greatly increased. Thus, in the light of efficiency and biological plausibility, this result encourages the use of such spike loss terms in other SNNs. Specifically, it might be interesting to determine the generalisability of this result. Additionally, it could be worthwhile to look into more involved spike loss terms. Especially looking into the manner in which such a term would be included in training and testing.

Secondly, despite a slight hint at generative features in control models with moderate spike loss, no conclusive evidence for this was found. This would indicate that the voltage energy term is indeed necessary for PC properties, a spike term is not enough. However, there remains some uncertainty whether there truly are no generative features. In future studies it might be possible to find or develop a metric that could reduce or eliminate this uncertainty.

References

[1] SM Bohte M Zhang R Chitic. "Energy Optimization Induces Predictive-coding Properties in a Multicompartment Spiking Neural Network Model". In: *bioRxiv preprint* (2024). URL: https://doi.org/10.1101/2024.01.17.575877.

[2] JH Schwartz ER Kandal. *Principles of Neural Science*. Elsevier, 1985.

[3] Yamazaki K Vo-Ho VK Bulsara D Le N. "Spiking Neural Networks and Their Applications: A Review". In: *Brain Sci. 12(7)* (2022). DOI: 10.1101/2024.01.17.575877.

[4] D Mumford. "On the computational architecture of the neocortex. II. The role of cortico-cortical loops." In: *Biological cybernetics, 66(3)* (1992).

[5] Pellerin L. Magistretti PJ. "Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging." In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 354(1387)* (1999). URL: https://doi.org/10.1098/rstb.1999.0471.