Structural bottlenecks promote temporal coding in spiking neural networks

Zach Mobille (zmobille3@gatech.edu)
Georgia Institute of Technology
School of Mathematics
686 Cherry St NW
Atlanta, GA 30332

Usama Bin Sikandar (usama@gatech.edu)
Georgia Institute of Technology
School of Electrical and Computer Engineering
777 Atlantic Dr NW
Atlanta, GA 30332

Simon Sponberg (sponberg@gatech.edu)
Georgia Institute of Technology
School of Physics
837 State St NW
Atlanta, GA 30332

Hannah Choi (hannahch@gatech.edu)
Georgia Institute of Technology
School of Mathematics
686 Cherry St NW
Atlanta, GA 30332

Abstract

Convergent and divergent structure in the networks that comprise biological brains is found universally across many species and brain regions at various scales. Given the frequency with which this structural motif is observed, we investigate what its functional role may be. While previous theories have neglected the role of neuronal spiking, our model and analysis places this aspect at the forefront. For a suite of stimuli with different timescales, we demonstrate that bottlenecks created by network convergence have a stronger preference for spike timing codes than expansion layers created by structural divergence. Our work makes quantitative predictions concerning the relationship between a network's convergent structure and the optimal timescale it can use to encode a dynamic stimulus. These predictions suggest a connection between network architecture and information-processing capabilities hitherto unexplored, which could be confirmed experimentally in future studies.

Keywords: bottleneck; spike timing; coding; network structure; convergence; divergence; feedforward; information

Introduction

Nervous systems are networks of neurons with highly non-random structure (Turner et al., 2022). One particular structural motif that is observed in many species and brain areas is that of the bottleneck, where a large group of neurons sends signals to, or "converges onto" a much smaller group of neurons. Conversely, there are also many instances of expansion networks where a small population of neurons synapses with a much larger population; we refer to this as network "divergence". Although divergence and convergence are observed in many brain areas like cerebellum, visuomotor pathways, and olfactory systems, the computational implications of this ubiquitous network structure are only beginning to be understood (Muscinelli, Wagner, & Litwin-Kumar, 2023; Gutierrez, Rieke, & Shea-Brown, 2021). In particular, we are aware of no studies that have characterized how convergence and divergence influences information coding in networks of spiking neurons. Such a paradigm is especially curious in light of growing experimental evidence demonstrating that precise spike timing can be much more informative about sensory input (Nemenman, Lewen, Bialek, & de Ruyter van Steveninck, 2008) and motor output (Srivastava et al., 2017; Putney, Conn, & Sponberg, 2019) than spike count. Thus, our aim is to show how convergent and divergent feedforward network structure promotes or suppresses spike timing codes in a stimulus-dependent way.

Results

We focus on feedforward spiking network models with multiple layers of neurons. For each time-dependent stimulus tested, the network parameters are optimized so that the output layer encodes the stimulus provided to the input layer. This is done in a way that is agnostic to the particular coding strategy employed by the output layer, by forming an estimate of the stimulus s via:

$$\hat{s} = \alpha \hat{s}_{\text{time}} + (1-\alpha)\hat{s}_{\text{count}} \tag{1}$$

where $\hat{s}_{\text{time}}$ is an estimate based on the spike timings of the output layer and $\hat{s}_{\text{count}}$ is one based on the spike counts. We set $\alpha = 0.5$ to weigh these equally. After training the networks (Eshraghian et al., 2023) to minimize the mean-squared error between true stimulus s and $\hat{s}$ from Eq. (1), we decode the stimulus separately from each layer's population spikes binned over a range of time scales Δt to assess the extent to which that layer encodes information more with a spike timing (smaller Δt) or spike count (larger Δt) code.

In our main results, we analyze 3-layer networks, keeping the number of input neurons and the number of output neurons fixed at $N_{\text{in}} = N_{\text{out}} = 100$ while modulating the number of neurons in the middle hidden layer N_h. Stimuli with a variety of timescales are tested to evaluate what role the input pattern plays in promoting or suppressing timing v.s. count codes. Given the respective increase/decrease in dimensionality from divergence/convergence, respectively, we hypothesize that bottlenecks promote timing codes whereas expansion networks promote count codes.

Results are shown for sum-of-sines stimuli of various frequencies in Fig. 1, where we use a support vector machine

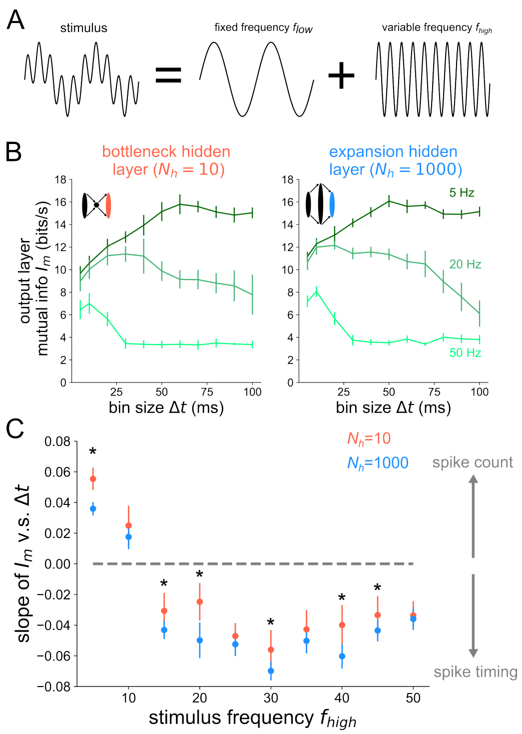

Figure 1: **Structural convergence from the hidden layer to the output layer promotes timing codes across all stimulus frequencies.** (A) Stimuli used here are sums of sines with a fixed component $f_{\text{low}} = 4$ Hz added to a variable component f_{high} (B) Mutual info I_m between the true stimulus and the decoded stimulus based on the output layer spikes binned at time resolution Δt (C) Slope of I_m v.s. Δt curves as a function of the high frequency stimulus component f_{high}. Asterisks denote where a one-sided Wilcoxon rank-sum test is significant at $p < 0.05$.

to decode the stimulus from the population spiking of neurons comprising the output layer. The mutual information I_m between the decoded stimulus and true stimulus is shown in Fig. 1B as a function of Δt for two types of networks: the bottleneck network with $N_h = 10$ and an expansion network with $N_h = 1000$. In both networks, output layer information about slow stimuli (i.e. 5 Hz) is maximized at large values around $\Delta t = 50$ or 60 ms, whereas faster stimuli (i.e. 50 Hz) are optimally encoded at smaller timescales around $\Delta t = 10$ ms. The preference for timing v.s. count codes is quantified in Fig. 1C with the slope of the I_m v.s. Δt curves by finding the best line fit. Across all stimulus frequencies tested, the slopes are more negative in the expansion network than the bottleneck network, indicating that structural convergence from the hidden layer to the output layer promotes temporal spike coding in the output layer.

While the results of Fig. 1 address how the coding strategy of populations of neurons post-synaptic to structural expansion v.s. compression differ, they do not indicate what coding strategy is optimal in the bottleneck/expansion layer itself.

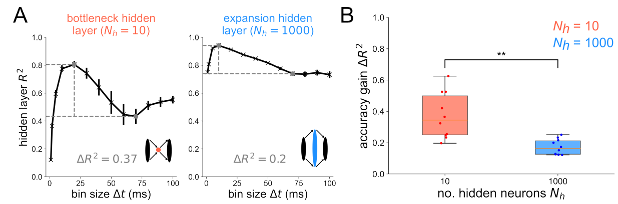

Figure 2: **Bottlenecks have more to gain from temporal codes than expansion layers. Stimulus used here is a 4 Hz + 20 Hz sum of sines.** (A) Decoding accuracy from the hidden layer spikes as a function of Δt. Gray dots denote points used to compute accuracy gain. (B) Accuracy gain of the temporal code over the count code when reconstructing the stimulus based on spikes from the hidden layer, for bottleneck and expansion networks. One-sided Wilcoxon rank-sum test $p = 5.8 \times 10^{-4}$.

In Fig. 2 we show results from our decoding analysis of the hidden layer for both the bottleneck and expansion network. Fig. 2A shows the decoding accuracy R^2 from the hidden layer as a function of Δt for both networks. In either case, the maximum R^2 occurs at a small value of $\Delta t = 10$ or 20 ms and a local minimum occurs around $\Delta t = 70$ ms. The difference in R^2 between these points is the accuracy gained by using a temporal code over a count code, and is plotted in Fig. 2B for $N_h = 10$ and $N_h = 1000$. From these results it is clear that, although higher accuracy is achieved by the expansion layer across all Δt's, the gain in accuracy from using a temporal code is significantly higher in the bottleneck than in the expansion layer.

Discussion

Our results show that temporal coding of a variety of stimuli is promoted by structural network convergence. Past studies have considered models of neural networks where the activity of units is static or smooth, and are thus limited to the context of continuous firing rate codes (Muscinelli et al., 2023; Gutierrez et al., 2021). In reality, biological neurons communicate with each other at the discrete times defined by their action potentials. The model developed and analyzed here reflects this reality, thus revealing a novel structure-function relationship in spiking neural networks. This work makes concrete predictions about how the relative information in spike timing v.s. count is modulated by structural divergence and convergence, and could be tested in experimental models with similar divergence/convergence ratios.

Acknowledgments

Research was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number T32GM142616 (Z.M.), the Alfred P. Sloan Foundation Fellowship in Neuroscience (H.C.), and the Air Force Office of Scientific Research grant FA9550-22-1-0315 (S.S. and H.C.).

References

Eshraghian, J. K., Ward, M., Neftci, E. O., Wang, X., Lenz, G., Dwivedi, G., ... Lu, W. D. (2023). Training spiking neural networks using lessons from deep learning. *Proceedings of the IEEE*.

Gutierrez, G. J., Rieke, F., & Shea-Brown, E. T. (2021). Nonlinear convergence boosts information coding in circuits with parallel outputs. *Proceedings of the National Academy of Sciences*, *118*(8), e1921882118.

Muscinelli, S. P., Wagner, M. J., & Litwin-Kumar, A. (2023). Optimal routing to cerebellum-like structures. *Nature neuroscience*, *26*(9), 1630–1641.

Nemenman, I., Lewen, G. D., Bialek, W., & de Ruyter van Steveninck, R. R. (2008). Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS computational biology*, *4*(3), e1000025.

Putney, J., Conn, R., & Sponberg, S. (2019). Precise timing is ubiquitous, consistent, and coordinated across a comprehensive, spike-resolved flight motor program. *Proceedings of the National Academy of Sciences*, *116*(52), 26951–26960.

Srivastava, K. H., Holmes, C. M., Vellema, M., Pack, A. R., Elemans, C. P., Nemenman, I., & Sober, S. J. (2017). Motor control by precisely timed spike patterns. *Proceedings of the National Academy of Sciences*, *114*(5), 1171–1176.

Turner, N. L., Macrina, T., Bae, J. A., Yang, R., Wilson, A. M., Schneider-Mizell, C., ... others (2022). Reconstruction of neocortex: Organelles, compartments, cells, circuits, and activity. *Cell*, *185*(6), 1082–1100.