Using Sinusoidally-Modulated Noise as a Surrogate for Slow-Wave Sleep to Accomplish Stable Unsupervised Dictionary Learning in a Spike-Based Sparse Coding Model

Yijing Watkins[1], Edward Kim[3], Andrew Sornborger[1] and Garrett T. Kenyon[1,2]
Los Alamos National Laboratory[1]
New Mexico Consortium[2]
Drexel University[3]
twatkins@lanl.gov,ek826@drexel.edu,sornborg@lanl.gov,gkeynon@lanl.gov

Abstract

Sparse coding algorithms have been used to model the acquisition of V1 simple cell receptive fields as well as to accomplish the unsupervised acquisition of features for a variety of machine learning applications. The Locally Competitive Algorithm (LCA) provides a biologically plausible implementation of sparse coding based on lateral inhibition. LCA can be reformulated to support dictionary learning via an online local Hebbian rule that reduces predictive coding error. Although originally formulated in terms of leaky integrator rate-coded neurons, LCA based on lateral inhibition between leaky integrate-and-fire (LIF) neurons has been implemented on spiking neuromorphic processors but such implementations preclude local online learning. We previously reported that spiking LCA can be expressed in terms of predictive coding error in a manner that allows for unsupervised dictionary learning via a local Hebbian rule but the issue of stability has not previously been addressed. Here, we use the Nengo simulator to show that unsupervised dictionary learning in a spiking LCA model can be made stable by incorporating epochs of sinusoidally-modulated noise that we hypothesize are analogous to slow-wave sleep. In the absence of slow-wave sleep epochs, the $|L|_2$ norm of individual features tends to increase over time during unsupervised dictionary learning until the corresponding neurons can be activated by random Gaussian noise. By inserting epochs of sinusoidally-modulated Gaussian noise, however, the $|L|_2$ norms of any activated neurons are down regulated such that individual neurons are no longer activated by noise. Our results suggest that slow-wave sleep may act, in part, to ensure that cortical neurons do not "hallucinate" their target features in pure noise, thus helping to maintain dynamical stability.

1. Introduction

Spiking neural networks (SNNs) are computational models that mimic biological neural networks. Compared with artificial neural networks (ANN), SNNs incorporate integrate-and-fire dynamics that increase both algorithmic and computational complexity. The justification for such increased complexity is two-fold: First, by using dedicated, potentially analog, circuit elements to instantiate individual neurons and by exploiting the low-bandwidth event-based communication enabled by SNNs, such networks can be implemented in extremely low-power neuromorphic hardware [2], enabling real-time remote applications that depend on scavenged power sources such as solar recharge. Second, there is evidence that biological neural circuits utilize spike timing to transmit information more rapidly and to dynamically bind distributed features via synchronous oscillations [10] [14] [13] [5]. The potential for mimicking the dynamics of biological neural networks in fast, low-power neuromorphic processors has motivated several efforts to develop such devices [3] [4] [8] [1] [12].

To fully exploit the potential of neuromorphic hardware, it is likely that such devices must be able to learn from their environment in a manner similar to biological neural systems. In particular, these devices must be able to learn, in an unsupervised manner, how to infer representations that support subsequent processing tasks. The Locally Competitive Algorithm (LCA) describes a dynamical neural network that uses only local synaptic interactions between non-spiking leaky-integrator neurons to infer sparse representations of input stimuli [9]. Unsupervised dictionary learning using convolutional LCA [11] has been used to infer sparse representations that support a number of signal processing tasks [17][16][7][6].

As originally implemented, unsupervised learning with non-spiking LCA utilizes non-local computations, specifically transpose and normalization operations performed globally on the entire weight matrix, and further requires signed outputs in order to represent the sparse reconstruction error. In our previous work [17], we showed that unsupervised dictionary learning can be accomplished using a modified Spiking LCA (S-LCA) that employs only local computations and uses only unsigned spiking output from all neurons. Although our previous results provide a proof-of-concept for how neuromorphic processors can be configured so as to self-organize in response to natural environmental stimuli without sacrificing efficiency, the issue of stability was not addressed. In this work, we use the Nengo simulator to show that unsupervised dictionary learning in a spiking LCA model can be made stable by incorporating epochs of sinusoidally-modulated noise that we hypothesize are analogous to slow-wave sleep.

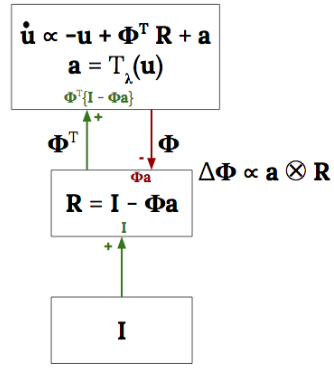


Figure 1: A non-spiking LCA model that supports unsupervised dictionary learning via a residual or sparse reconstruction error layer. See Eq. (2) and accompanying text.

2. Methods

2.1. Unsupervised Dictionary Learning with a Non-Spiking LCA

Given an overcomplete basis, non-spiking LCA [9] can be used to find a minimal set of active neurons that represent the input to some degree of fidelity. Each neuron can be thought of as a generator that adds its associated feature vector to the reconstructed input with an amplitude equal to its activation. For any particular input, the optimal sparse representation is given by a vector of neural activations that minimizes the following cost function:

$$E = \frac{1}{2}||\mathbf{I} - \{\mathbf{\Phi} * \mathbf{a}\}||_2^2 + \frac{1}{2}\lambda^2||\mathbf{a}||_0 \tag{1}$$

where $\mathbf{I}$ is the input vector and $\mathbf{\Phi}$ is a dictionary of feature kernels that are convolved with the activation coefficients $\mathbf{a}$. The feature kernels are assumed to be normalized such that $\mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{1}$. The L_0 norm $||\mathbf{a}||_0$ simply counts the number of non-zero activation coefficients. The factor λ acts as a trade-off parameter; larger λ values encourage greater sparsity (fewer non-zero coefficients) at the cost of greater reconstruction error.

LCA finds a local minimum of the cost function defined in Eq. (1) by introducing the dynamical variables (membrane potentials) $\mathbf{u}$ such that the output $\mathbf{a}$ of each neuron is given by a hard-threshold transfer function, with threshold λ, of the membrane potential: $\mathbf{a} = T_\lambda(\mathbf{u}) = H(\mathbf{u} - \lambda)\mathbf{u}$, where H is the Heaviside function[9].

For a given input, the cost function defined in equation (1) is then minimized by taking the gradient of the cost function with respect to $\mathbf{a}$ and solving the resulting set of coupled differential equations for the membrane potentials $\mathbf{u}$:

$$\dot{\mathbf{u}} \propto -\frac{\partial E}{\partial \mathbf{a}} = -\mathbf{u} + \mathbf{\Phi}^T\{\mathbf{I} - \mathbf{\Phi}T_\lambda(\mathbf{u})\} + T_\lambda(\mathbf{u}). \tag{2}$$

where the final term in Eq. (2) removes the self-interaction (a neurons does not inhibit itself).

An update rule for feature kernels can be obtained by taking the gradient of the cost function with respect to $\mathbf{\Phi}$:

$$\Delta\mathbf{\Phi} \propto -\frac{\partial E}{\partial \mathbf{\Phi}} = \mathbf{a} \otimes \{\mathbf{I} - \mathbf{\Phi}\mathbf{a}\} = \mathbf{a} \otimes \mathbf{R} \tag{3}$$

where we introduced an intermediate residual layer $\mathbf{R}$ corresponding to the sparse reconstruction error.

For non-spiking LCA, online unsupervised dictionary learning is achieved via a two step process: First, a sparse representation for a given input is obtained by integrating Eq. (2), after which Eq. (3) is evaluated to slightly reduce the reconstruction error given the sparse representation of the current input.

As illustrated in Figure 1, the weight update (3) resembles a local Hebbian learning rule for $\mathbf{\Phi}$ with pre- and post-synaptic activities $\mathbf{a}$ and $\mathbf{R}$ respectively. However, the computation of $\mathbf{\Phi}^T$ as well as the normalization constraint renders the overall dictionary learning process a non-local operation.

We have previously shown that our implementation of non-spiking LCA can be used to learn a convolutional dictionary in an unsupervised, self-organizing manner that factors a complex, high-dimensional natural image into an overcomplete set of basis vectors that capture the high-dimensional correlations in the data [11]. In the next section, we show how this implementation can be adapted to an S-LCA model that uses only local computations and unsigned spiking output.

2.2. S-LCA Unsupervised Dictionary Learning

Our S-LCA model for unsupervised dictionary learning is shown in Figure 2, where the superscripts indicate the different layers (e.g. $\mathbf{a}^\mathbf{I}$ denotes the input spikes). We replace the non-spiking leaky-integrator model of Eq. 2 with

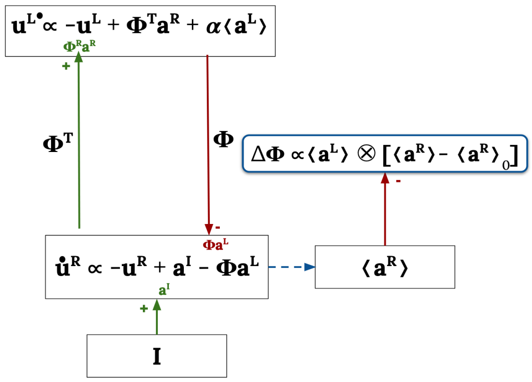


Figure 2: S-LCA with unsupervised dictionary learning. $\mathbf{\Phi}^T$ now denotes a separate set of connections that are initialized to be equal to $\mathbf{\Phi}$ and obey a numerically identical learning rule with pre and post synaptic factors flipped.

a leaky integrate-and-fire (LIF) model consisting of a membrane potential, $\mathbf{u}$, and a binary spiking output, $\mathbf{a}$:

$$\dot{\mathbf{u}} \propto -\mathbf{u} + \mathbf{u}_{input} \tag{4}$$

$$\mathbf{a} = \begin{cases} 0 & \mathbf{u} < \lambda \\ 1; \mathbf{u} \to 0 & \mathbf{u} \geq \lambda \end{cases} \tag{5}$$

where λ again plays the role of a threshold that controls the level of sparsity and $\mathbf{u}_{input}$ is the sum of the input received from connected neurons. When $\mathbf{u}$ crosses the threshold λ a spike is generated and u is reset to zero.

As with non-spiking LCA, the residual layer $\mathbf{R}$ in S-LCA is driven by the difference between the input and the reconstructed input generated by the LCA layer, which for S-LCA is given by $\mathbf{a^I} - \mathbf{\Phi a^L}$. Values of $\langle\mathbf{a^R}\rangle$ above and below the target baseline firing rate $\langle\mathbf{a^R}\rangle_0$ encode positive and negative errors, respectively. Eq. 4 for the residual layer $\mathbf{R}$ then becomes

$$\dot{\mathbf{u}}^{\mathbf{R}} \propto -\mathbf{u}^{\mathbf{R}} + \mathbf{a}^{\mathbf{I}} - \mathbf{\Phi}\mathbf{a}^{\mathbf{L}}. \tag{6}$$

We augment Eq. 6 with a firing condition analogous to Eq. 5 with $\lambda \to \lambda^{\mathbf{R}}$.

The input to the sparse coding layer $\mathbf{L}$ in Figure 2 is denoted by $\mathbf{u}^{\mathbf{L}}_{input}$, given by,

$$\mathbf{u}^{\mathbf{L}}_{input} = \mathbf{\Phi}^{\mathbf{T}}\mathbf{a}^{\mathbf{R}}. \tag{7}$$

Likewise, in Figure 2, the LCA layer $\mathbf{L}$ is again an LIF layer whose equation of motion is given by:

$$\dot{\mathbf{u}}^{\mathbf{L}} \propto -\mathbf{u}^{\mathbf{L}} + \mathbf{\Phi}^{\mathbf{T}}\mathbf{a}^{\mathbf{R}} + \alpha\langle\mathbf{a}^{\mathbf{L}}\rangle, \tag{8}$$

where $\alpha\langle\mathbf{a^L}\rangle$ is a self interaction involving the low-pass filtered (trace) of the spiking output that helps to maintain firing activity when the residual error is zero. The constant $\alpha = 0.75$ and the trace time constant 500 msec were determined empirically to produce improved sparse reconstructions without ringing or overshoot. We again augment Eq. 8 with a firing condition analogous to Eq. 5 with $\lambda \to \lambda^{\mathbf{L}}$.

Unsupervised dictionary learning can be used to update the weight matrices $\mathbf{\Phi}$ given only information locally available at each synapse. To compute the weight updates, we introduce the low-pass filtered spike trains, or instantaneous firing rates or traces, of the LIF neurons in the residual layer $\overline{\langle\mathbf{a^R}\rangle} = \langle\mathbf{a^R}\rangle - \langle\mathbf{a^R}\rangle_0$ computed relative to the target baseline firing rate of the residual layer. The firing rates of the LIF neurons in the sparse coding layer $\langle\mathbf{a^L}\rangle$ are likewise represented as low-pass filtered versions of the corresponding spike trains. In terms of these local firing rates, the update of $\mathbf{\Phi}$ is given by a local Hebbian learning rule:

$$\Delta\Phi \propto \langle\mathbf{a^L}\rangle \otimes \overline{\langle\mathbf{a^R}\rangle}. \tag{9}$$

The traces time in the local Hebbian learning rule are 100 msec.

3. Results

3.1. Instability of unsupervised learning without sleep

In our previous work [17], we showed that unsupervised dictionary learning can be accomplished using a modified Spiking LCA (S-LCA) that employs only local computations and uses only unsigned spiking output from all neurons. Although our previous work provide a proof-of-concept for how neuromorphic processors can be configured so as to self-organize in response to natural environmental stimuli without sacrificing efficiency, the issue of stability was not addressed.

As indicated in In Figure 3 and Figure 4, in the absence of slow-wave sleep epochs, the $|L|_2$ norm of individual features tends to increase over time during unsupervised dictionary learning until the corresponding neurons become continuously activated. The features learned in the absence of sleep epochs are initially reasonable (Figure 5) but ultimately become meaningless as the system becomes unstable.

3.2. Stability of learning with slow-wave sleep

As illustrated in In Figure 6, we used sinusoidally-modulated noise that we hypothesize is analogous to slow-wave sleep. Shown is the norm of the stimulus over one slow-wave sleep cycle. Nothing about the model or learning rule changes during slow-wave sleep. The only difference between ”wakefulness” and slow-wave sleep is the input stimuli to the system (see Figure 9 and Figure 10). During wakefulness, the input stimuli consist of natural gray-scale images which can be sparsely reconstructed from learnable

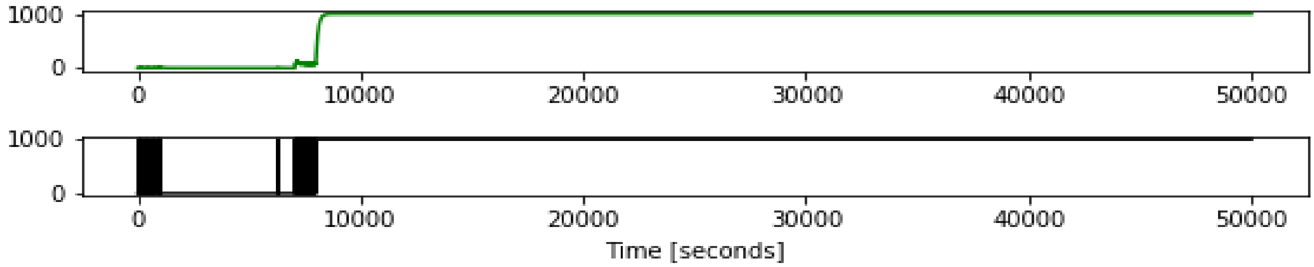


Figure 3: Spikes (bottom) and spike trace (top) from a representative S-LCA neuron without sleep. In the absence of slow-wave sleep epochs, the $|L|_2$ norm of individual features tends to increase over time during unsupervised dictionary learning until the corresponding neurons are continuously activated. Individual spikes cannot be resolved at this level of resolution.

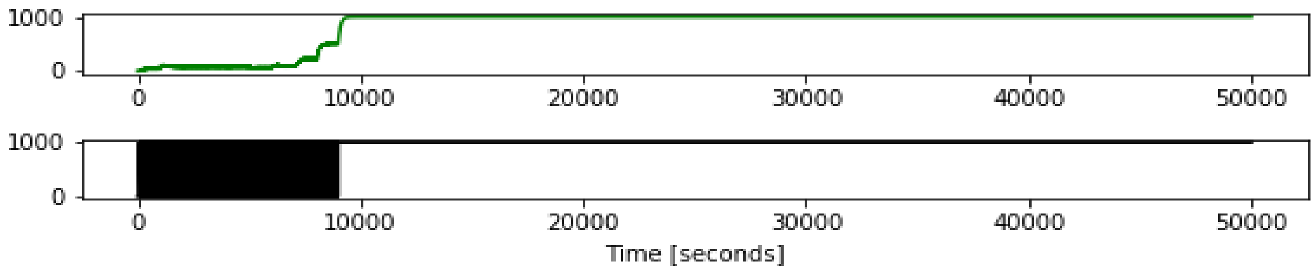
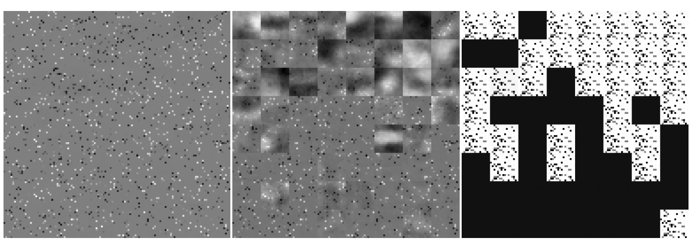


Figure 4: Representative residual layer neuron spikes (bottom) and spike trace (top) without sleep. As the gain of the feedback loop increases, the residual layer becomes persistently active.



(a) Before training (b) Some training (c) More training

Figure 5: Top 64 most active dictionary elements without slow-wave sleep.

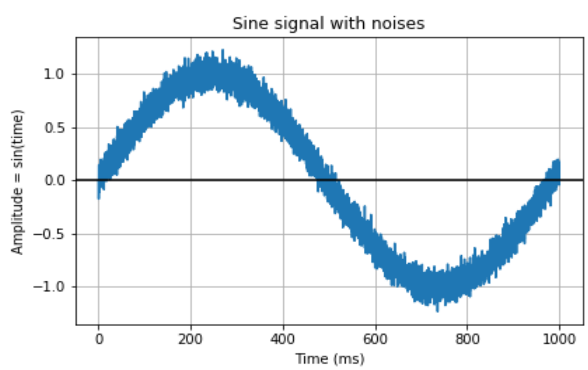


Figure 6: Sinusoidally-modulated noise that we hypothesize is analogous to slow-wave sleep.

features. During slow-wave sleep, however, the stimulus consists of random Gaussian noise that is not learnable.

As indicated in Figure 7 and Figure 8, by inserting epochs of sinusoidally-modulated Gaussian noise, the $|L|_2$ norms of the feature vectors associated with neurons activated during slow-wave sleep are down regulated such that individual neurons are no longer persistently active. In particular, neurons dynamically adjust their gain during slow-wave sleep so as not to be activated by random Gaussian noise. Instead, as in the representative S-LCA neuron depicted, activity is only sparsely induced by a limited subset of natural stimuli. On occasion, S-LCA neurons are activated by sinusoidally-modulated Gaussian noise but any such activations come to be repressed over successive slow-wave sleep cycles. Our approach is consistent with evidence that long term potentiation changes occur during wakefulness due to a net increase in synaptic weights, and slow wave sleep is necessary to promote generalized depression of synapses [15].

Example reconstructions of a natural stimulus over the

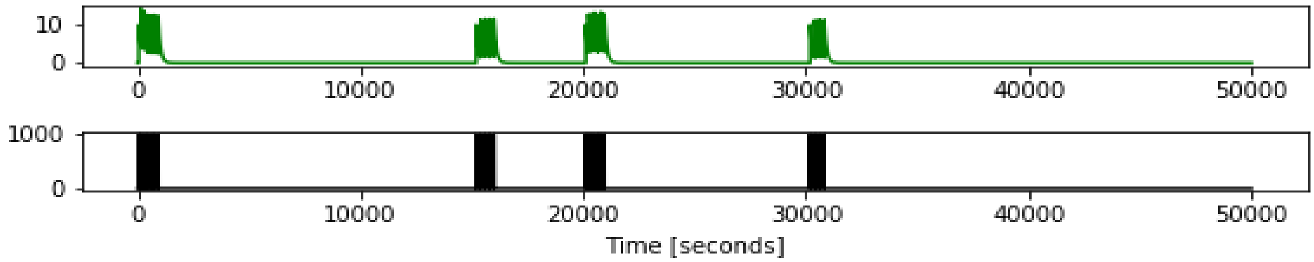


Figure 7: Representative S-LCA neuron spikes (bottom) and spike trace (top) with slow-wave sleep. By inserting epochs of sinusoidally-modulated Gaussian noise, the $|L|_2$ norms of any neurons activated during slow-wave sleep are down regulated such that individual neurons are no longer activated by random Gaussian noise. As a result, neurons are sparsely activated by only a subset of natural stimuli and rarely by random noise.

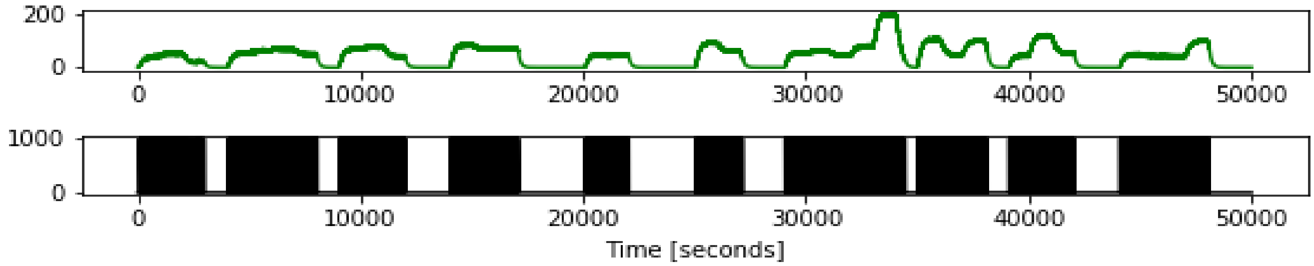


Figure 8: Representative residual neuron spikes (bottom) and spike trace (top) with slow-wave sleep. Although the trace does not always tend to the baseline value indicting zero residual, we observe that average residual does diminish with continued unsupervised training.

course of unsupervised training with slow-wave sleep are shown in Figure 11 and Figure 12. Corresponding learned features are shown in Figure 13. As dictionary elements converge toward more meaningful features, the corresponding reconstructions become more accurate. Computational constraints prevented us from continuing these experiments until the dictionary was fully converged. However, including epochs of sinusoidally-modulated noise hypothesized to be analogous to slow-wave sleep produced stable sparse reconstructions that steadily improved with further unsupervised training whereas the system became dynamically unstable in the absence of slow-wave sleep even when using identical random weight initialization parameters and learning rates.

3.3. Conclusion

Although our previous work [17] provide a proof-of-concept for how neuromorphic processors can be configured so as to self-organize in response to natural environmental stimuli without sacrificing efficiency, the issue of stability was not addressed. In this work, we use the Nengo simulator to show that unsupervised dictionary learning in a spiking LCA model can be made stable by incorporating epochs of sinusoidally-modulated noise that we hypothesize are analogous to slow-wave sleep.

References

[1] Filipp Akopyan, Jun Sawada, Andrew Cassidy, Rodrigo Alvarez-Icaza, John Arthur, Paul Merolla, Nabil Imam, Yutaka Nakamura, Pallab Datta, Gi-Joon Nam, et al. Truenorth: Design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(10):1537–1557, 2015.

[2] K. Boahen. A neuromorph's prospectus. *Computing in Science Engineering*, 19(2):14–28, Mar. 2017.

[3] Mike Davies, Narayan Srinivasa, Tsung-Han Lin, Gautham Chinya, Yongqiang Cao, Sri Harsha Choday, Georgios Dimou, Prasad Joshi, Nabil Imam, Shweta Jain, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1):82–99, 2018.

[4] Simon Friedmann, Johannes Schemmel, Andreas Grübl, Andreas Hartel, Matthias Hock, and Karlheinz Meier. Demonstrating hybrid learning in a flexible neuromorphic hardware system. *IEEE transactions on biomedical circuits and systems*, 11(1):128–142, 2017.

Figure 9: Sinusoidally-modulated slow-wave sleep with the noise amplitude set to 0.

Figure 10: Sinusoidally-modulated slow-wave sleep with the noise amplitude set to 1.

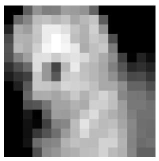
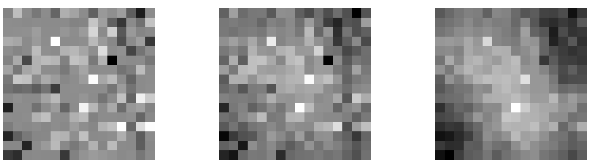
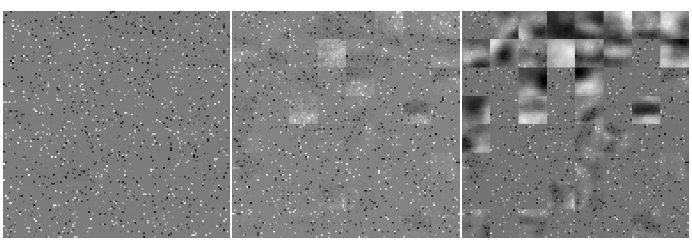


Figure 11: Cifar-10 Input.



(a) Before training (b) Some training (c) More training

Figure 12: Reconstructions with sleep.



(a) Before training (b) Some training (c) More training

Figure 13: Top 64 most active dictionary elements with Sleeps.

[5] Garrett T Kenyon. Extreme synergy: Spatiotemporal correlations enable rapid image reconstruction from computer-generated spike trains. *Journal of vision*, 10(3):21–21, 2010.

[6] Sheng Y Lundquist, Melanie Mitchell, and Garrett T Kenyon. Sparse coding on stereo video for object detection. In *workshop on Learning with Limited Labeled Data: Weak Supervision and Beyond, NIPS 2017*. NIPS, 2017.

[7] Sheng Y Lundquist, Dylan M Paiton, Peter F Schultz, and Garrett T Kenyon. Sparse encoding of binocular images for depth inference. In *Image Analysis and Interpretation (SSIAI), 2016 IEEE Southwest Symposium on*, pages 121–124. IEEE, 2016.

[8] E. Painkras, L. A. Plana, J. Garside, S. Temple, S. Davidson, J. Pepper, D. Clark, C. Patterson, and S. Furber. Spinnaker: A multi-core system-on-chip for massively-parallel neural net simulation. In *Proceedings of the IEEE 2012 Custom Integrated Circuits Conference*, pages 1–4, Sept 2012.

[9] Christopher J. Rozell, Don H. Johnson, Richard G. Baraniuk, and Bruno A. Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural Comput.*, 20(10):2526–2563, Oct. 2008.

[10] Rufin Van Rullen and Simon J Thorpe. Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural computation*, 13(6):1255–1283, 2001.

[11] Pete Schultz, Dylan Paiton, Wei Lu, and Garrett Kenyon. Replicating kernels with a short stride allows sparse reconstructions with fewer independent kernels, 2014.

[12] Patrick M. Sheridan, Fuxi Cai, Chao Du, Wen Ma, Zhengya Zhang, and Wei Lu. Sparse coding with memristor networks. *Nature Nanotechnology*, 12, 05 2017.

[13] Greg J Stephens, Sergio Neuenschwander, John S George, Wolf Singer, and Garrett T Kenyon. See globally, spike locally: oscillations in a retinal model encode large visual features. *Biological cybernetics*, 95(4):327–348, 2006.

[14] Simon Thorpe, Arnaud Delorme, and Rufin Van Rullen. Spike-based strategies for rapid processing. *Neural networks*, 14(6-7):715–725, 2001.

[15] Giulio Tononi and Chiara Cirelli. Sleep and synaptic homeostasis: a hypothesis. *Brain research bulletin*, 62(2):143–150, 2003.

[16] Yijing Watkins, Oleksandr Iaroshenko, Mohammad R. Sayeh, and Garrett T. Kenyon. Image compression: Sparse coding vs. bottleneck autoencoders. In *2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, SSIAI '18, 2018.

[17] Yijing Watkins, Austin Thresher, David Mascarenas, and Garrett T. Kenyon. Sparse coding enables the reconstruction of high-fidelity images and video from retinal spike trains. In *Proceedings of the International Conference on Neuromorphic Systems*, ICONS '18, pages 8:1–8:5, New York, NY, USA, 2018. ACM.