ADAPTIVE COMPRESSED SENSING – A NEW CLASS OF SELF-ORGANIZING CODING MODELS FOR NEUROSCIENCE

William K. Coulter, Christopher J. Hillar**, Guy Isley*, Friedrich T. Sommer**

*: University of California, Berkeley, 156 Stanley Hall, MC# 3220, Berkeley, CA 94720-3220
**: Mathematical Sciences Research Institute, 17 Gauss Way, Berkeley, CA 94120
wcoulter@berkeley.edu, chillar@msri.org, gisley@gmail.com, fsommer@berkeley.edu

ABSTRACT

Sparse coding networks, which utilize unsupervised learning to maximize coding efficiency, have successfully reproduced response properties found in primary visual cortex [1]. However, conventional sparse coding models require that the coding circuit can fully sample the sensory data in a one-to-one fashion, a requirement not supported by experimental data from the thalamo-cortical projection. To relieve these strict wiring requirements, we propose a sparse coding network constructed by introducing synaptic learning in the framework of compressed sensing. We demonstrate a new model that evolves biologically realistic, spatially smooth receptive fields despite the fact that the feedforward connectivity subsamples the input and thus the learning must rely on an impoverished and distorted account of the original visual data. Further, we demonstrate that the model could form a general scheme of cortical communication: it can form meaningful representations in a secondary sensory area, which receives input from the primary sensory area through a "compressing" cortico-cortical projection. Finally, we prove that our model belongs to a new class of sparse coding algorithms in which recurrent connections are essential in forming the spatial receptive fields.

***Index Terms*—** adaptive coding, biological system modeling, random codes, image coding, nonlinear circuits

1. INTRODUCTION

Guided by the early ideas on efficient sensory coding [2, 3], self-organizing network models for sparse coding have been critical in understanding how essential response properties, such as orientation selectivity, are formed in sensory areas through development and experience [1, 4]. There is now a wealth of such models, all based on a set of similar connectivity patterns: a neuron receives feedforward drive from the afferent input and competes with other neurons in the network through mainly inhibitory lateral connections, see [5, 1, 6, 7, 8]. Some of these models are capable of reproducing the response properties in primary visual cortex quantitatively, for instance, the network model proposed in [6] that implements an algorithm called optimized orthogonal matching pursuit [9].

While these models match physiological data quite impressively, their correspondence to the anatomical connectivity in cortex is problematic. According to the models, the neurons must have access to the full data, for instance, to all pixels of an image patch. Many models even suggest that each neuron has the feedforward wiring in place so that the synaptic structure in the feedforward path can match the receptive field exactly. It is unclear if the development of the thalamic projections into V1 can reach such connection density and microscopic precision – even though thalamic receptive fields do match the receptive fields of monosynaptically connected V1 cells with some precision [10, 11]. Here, we explore learning schemes for neural representations that relieve these requirements on the feedforward wiring. In addition, we assess the ability of these learning schemes to account for learning in cortico-cortical projections, for which it has been established that only a fraction of local cells in the origin area send fibers to a target area [12]. Therefore, conventional sparse coding at the receiver end can not work.

To construct sparse coding networks with less restrictive wiring conditions we build on *compressed sensing* or *compressed sampling*, a method originally developed for data compression by subsampling. The decompression step in these algorithms has a close similarity to sparse coding models and thus these methods can form a framework for developing a new class of neural networks for self-organizing neural representations in cortical areas. Specifically, we explore the hypothesis of a generic scheme of cortical communication in which each cortical area unwraps subsampled input data into a sparse code to perform local computations and then sends a subsampled version of its local representation to other cortical areas.

2. ADAPTIVE COMPRESSED SENSING

Conventional sparse coding is governed by the objective function

$$E(\mathbf{x},\mathbf{a},\mathbf{\Psi}) = \frac{1}{2}||\mathbf{x}-\mathbf{\Psi}\mathbf{a}||^2 + S(\mathbf{a}). \tag{1}$$

Here $\mathbf{x} \in \mathbb{R}^m$ is the input data, $\mathbf{\Psi}$ is a real $m \times n$ matrix whose columns form a dictionary for constructing the input, and $\mathbf{a}$ is a coefficient vector for this reconstruction: $\tilde{\mathbf{x}} = \mathbf{\Psi}\mathbf{a}$. The function $S(\mathbf{a})$ is a sparseness constraint that penalizes neural activity and forces the coefficient vector to be sparse.

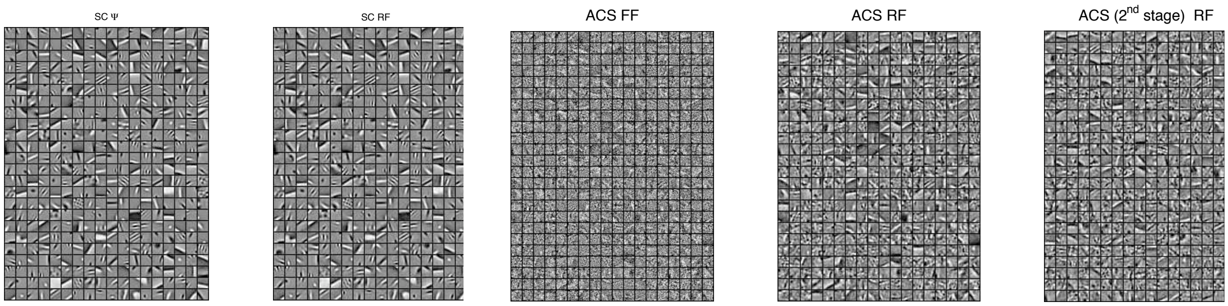
For a given input $\mathbf{x}$, the sparse coding operation is given by an energy minimization

$$\mathbf{a}(\mathbf{x}) := \arg\min_a E(\mathbf{x},\mathbf{a},\mathbf{\Psi}) \in \mathbb{R}^n. \tag{2}$$

We adapt the dictionary to the data by minimizing $E(\mathbf{x},\mathbf{a}(\mathbf{x}),\mathbf{\Psi})$ from Eq. 1 and Eq. 2 with respect to $\mathbf{\Psi}$. Using gradient descent for the adaptation yields a Hebbian synaptic learning rule for the $\mathbf{\Psi}$ components [1, 6].

Compressed sensing is a technique for data compression using a random projection matrix $\mathbf{\Phi}$ to compress the data $\mathbf{x} \in \mathbb{R}^m$ to $\mathbf{\Phi}\mathbf{x} \in \mathbb{R}^k$ with $k < m$. The decompression uses energy minimization (2)

Hillar is supported under an NSA Young Investigators Grant and an MSRI postdoctoral fellowship



(a) Feedforward weights and receptive fields of sparse coding circuit [6]. The patterns look indistinguishable to the eye and can be proven to be the same (see Thm 1).

(b) Feedforward weights (ACS FF) and receptive fields (ACS RF) of adaptive compressed sensing circuit. The plot "ACS (2nd stage) RF" depicts receptive fields learned in a cascaded secondary sensory area receiving $\mathbf{\Phi_2 a(x)}$ as the input.

Fig. 1. Comparison of Feedforward weights (FF) and Receptive Fields (RF) of various models.

of an error-based energy function similar to Eq. 1:

$$E(\mathbf{x}, \mathbf{a}, \mathbf{\Phi}) = \frac{1}{2}||\mathbf{\Phi x} - \mathbf{\Phi \Psi a}||^2 + S(\mathbf{a}). \tag{3}$$

The original data is reconstructed as $\tilde{\mathbf{x}} = \mathbf{\Psi}\ \mathbf{a}(\mathbf{\Phi x})$. In conventional compressed sensing, a fixed dictionary $\mathbf{\Psi}$ is chosen. The decompression can be shown to work if (i) a dictionary $\mathbf{\Psi}$ is used in which the data can be sparsely represented, (ii) matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are incoherent, and (iii) the dimension of data compression k is larger than the sparsity of the data [13, 14].

Building on a model by Rehn and Sommer [6], we introduce *adaptive compressed sensing (ACS)*, an adaptive version of compressed sensing governed by:

$$\begin{aligned} E(\mathbf{x}, \mathbf{a}, \mathbf{\Phi}, \mathbf{\Theta}) &= \frac{1}{2}||\mathbf{\Phi x} - \mathbf{\Theta a}||^2 + \lambda ||\mathbf{a}||_{L_0} \qquad (4) \\ &= -\mathbf{x}^\top \mathbf{\Phi}^\top \mathbf{\Theta a} + \frac{1}{2}\mathbf{a}^\top \mathbf{\Theta}^\top \mathbf{\Theta a} + \lambda ||\mathbf{a}||_{L_0} + const. \end{aligned}$$

Learning is executed by gradient descent on $\mathbf{\Theta}$ in exactly the same fashion as in conventional sparse coding, e.g. [1, 6]. Note, however, the difference between ACS and conventional sparse coding. The new algorithm (4) forms a dictionary of the compressed data, the $k \times n$ matrix $\mathbf{\Theta}$, whereas conventional sparse coding forms a dictionary of the original data, an $m \times n$ matrix. Although we use the L_0-sparseness constraint to penalize the number of active units, $S(\mathbf{a}) = \lambda ||\mathbf{a}||_{L0}$, similar schemes of adaptive compressed sensing can be realized with other types of sparseness constraints.

Network implementation of ACS: Analogous to earlier models of sparse coding, coding in ACS can be implemented in a network where each neuron i computes the gradient of the two differentiable terms in Eq. 4 as

$$\frac{\partial E'}{\partial a_i} = -(\mathbf{x}^\top \mathbf{\Phi}^\top \mathbf{\Theta})_i + (\mathbf{\Theta}^\top \mathbf{\Theta a})_i, \tag{5}$$

see [6] for further detail. In the neural network for the ACS method the *feedforward weights* are $FF := \mathbf{\Phi}^\top \mathbf{\Theta}$ and the competitive *feedback weights* are $FB := -\mathbf{\Theta}^\top \mathbf{\Theta}$. Note that if $\mathbf{\Phi}$ is the identity matrix, ACS coincides with conventional sparse coding for which the corresponding neural network would be defined by $FF = \mathbf{\Psi}$ and $FB = -FF^\top FF$ [1, 6]. The important difference between the two wiring schemes is that the feedforward weights of ACS subsample and mix the original data. Thus, coding and weight adapatation in ACS lack the full access to the original data that is available to conventional sparse coding. Remarkably, the simulation experiments described in the next section demonstrate that the neurons in the ACS network still develop biologically realistic receptive fields, despite the limited exposure to the original data.

3. SIMULATION EXPERIMENTS WITH ADAPTIVE COMPRESSED SENSING

We compared the ability of networks described in section 2 to code patches of natural scene images and form receptive fields. The images were preprocessed by "whitening," as described in [1]. The coding circuits encoded patches of 12×12 pixels, making the dimension of the data $m = 144$. For ACS, we used a sampling matrix $\mathbf{\Phi}$ that downsampled the original data to $k = 60$ dimensions. All coding circuits contained $n = 432$ neurons, thereby producing representations of the original data $a \in \mathbb{R}^n$ that were three times overcomplete. In addition to image coding in a primary sensory area, we also tested whether the ACS model could be used by a secondary sensory area (2nd stage). Our model of the 2nd stage receives a subsampled version of the sparse code generated in the primary visual area $\mathbf{\Phi_2 a(x)} \in \mathbb{R}^k$ and produces a sparse code $\mathbf{a_2} \in \mathbb{R}^n$, again with $k = 60$ and $n = 432$. Models used a coefficient $\lambda = 0.1$ in the sparseness constraint of Eq. 4.

Since the ACS model learns a dictionary of the compressed data rather than the original data, the original image cannot be reconstructed from the adapted $\mathbf{\Theta}$ matrix. Note that computing the data dictionary from $\mathbf{\Theta}$ requires an ill-posed step of matrix factorization: $\mathbf{\Theta} = \mathbf{\Phi \Psi}$. Therefore, to assess the quality of the emerging codes in the ACS model, we measured receptive fields in the trained circuit (as physiologists do from the responses of real neurons). We compute the *receptive fields* for a set I of visual stimuli $\mathbf{x} \in \mathbb{R}^m$ as

$$RF := \frac{1}{|I|} \sum_{\mathbf{x} \in I} \mathbf{x} \cdot \mathbf{a}(\mathbf{x})^\top. \tag{6}$$

Notice that RF is an $m \times n$ real matrix, the i-th column representing the receptive field of the i-th neuron.

Figure 2 shows the feedforward weights and the receptive fields of the different coding circuits. While the feedforward weights and receptive fields in Fig 1(a) are very similar for sparse coding, they are markedly different for ACS in Fig 1(b). Interestingly, while subsampling makes the feedforward weights somewhat amorphous and noisy, the resulting receptive fields of ACS are smooth and resemble the receptive fields of sparse coding. When used in a secondary sensory area (2nd stage), ACS forms response properties that are similar to those in the primary sensory area, though the response properties differ on a neuron-by-neuron basis.

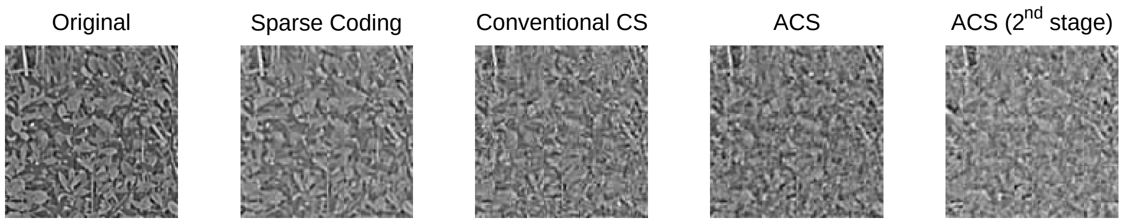


Fig. 2. Original image and reconstruction using the different methods on 12 by 12 image patches. For reconstructing the images from the representations formed by ACS (in the first and 2nd stage), we used the receptive fields.

To assess how well the sparse codes describe the original input, we computed image reconstructions. For sparse coding, we used the basis functions $\mathbf{\Psi}$; for ACS, the receptive fields RF. Fig 2 shows that ACS forms representations in the primary and secondary area that can be used for reconstruction, although the quality of reconstructions obtained from conventional sparse coding is not achieved.

Fig 3 compares the reconstruction quality of conventional compressed sensing (using the basis functions that were adapted to the original data) and adaptive compressed sensing. The mean reconstruction qualities do not differ, though ACS performs with lower variance over the set of input patches we tested.

These simulation results suggest that ACS is able to form representations of sensory data that convey its essential structure although the coding network receives only a subsampled version of the data. In the next section, we investigate the mathematical differences between conventional sparse coding and adaptive compressed sensing.

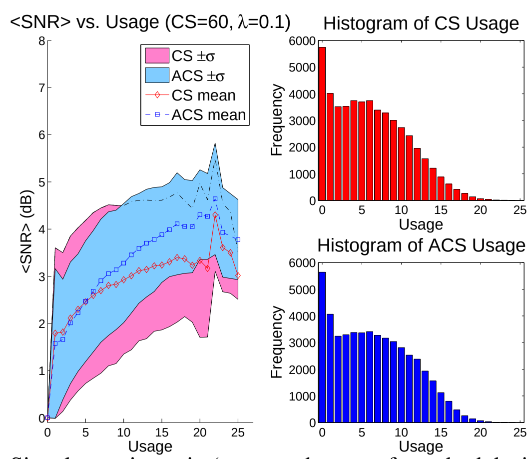


Fig. 3. Left: Signal-to-noise ratio (mean and range of standard deviation) in reconstructions with conventional compressed sensing (red) and ACS (blue). Means do not differ significantly, but the performance of ACS has smaller variance. Right: Histograms of the number of neurons used per reconstruction are similar for both methods.

4. DIFFERENCES BETWEEN ACS AND CONVENTIONAL SPARSE CODING MODELS

In this section we derive two theorems to establish that ACS defines a class of sparse coding algorithms whose properties differ qualitatively from those of conventional sparse coding models.

4.1. Receptive fields and feedforward weights coincide in conventional sparse coding networks

Assuming that $\mathbf{x}$ is a column vector of random variables on a measure space Ω with probability measure μ, the matrix RF will be an approximation to the correlation of $\mathbf{x}$ and $\mathbf{a}(\mathbf{x})$ (which is assumed to have zero mean): $\mathrm{Cor}(\mathbf{x}, \mathbf{a}) = \int \mathbf{x}\mathbf{a}(\mathbf{x})^\top d\mu$.

The strong law of large numbers guarantees that given enough samples, the matrix RF will be close to the integral above. For this reason, we will assume that $RF = \mathrm{Cor}(\mathbf{x}, \mathbf{a})$. We are interested in calculating the necessary relationships between the quantities RF, FF, FB, $\mathbf{\Phi}$, and $\mathbf{\Theta}$. (Recall: $FF = \mathbf{\Phi}^\top \mathbf{\Theta}$ and $FB = -\mathbf{\Theta}^\top \mathbf{\Theta}$.)

In our setup, the data $\mathbf{x}$ are assumed to come from a sparse number k of independent causes (nonzero values in $\mathbf{a}$). Moreover, the method of recovering $\mathbf{a}(\mathbf{x})$ from a particular $\mathbf{x}$ is assumed to be exact (or near exact) in solving Eq. 2 and *independently distributed*; that is, $\mathbf{\Phi x} = \mathbf{\Theta}\ \mathbf{a}(\mathbf{x})$ and we have: $D = \int \mathbf{a}(\mathbf{x})\mathbf{a}(\mathbf{x})^\top d\mu$, in which D is an $n \times n$ diagonal matrix. One now calculates:

$$\mathbf{\Phi} RF = \int \mathbf{\Phi x}\ \mathbf{a}(\mathbf{x})^\top d\mu = \mathbf{\Theta} \int \mathbf{a}(\mathbf{x})\mathbf{a}(\mathbf{x})^\top d\mu = \mathbf{\Theta} D. \quad (7)$$

In particular, this implies the following.

Theorem 1 *If $\mathbf{\Phi}$ is the identity, then the receptive fields are scalar multiples of the feedforward weights.*

4.2. Feedback co-shapes receptive fields in the ACS model

In the compressive sensing regime, the matrix $\mathbf{\Phi}$ is no longer the identity but instead a compressive sampling matrix. In this case, the receptive fields are almost never scalar multiples of the feedforward weights. A precise analytic relationship is given by the following theorem. As an important consequence, we obtain the qualitative interpretation found in Theorem 4 below. We omit here the proofs.

Theorem 2 *If FF, $\mathbf{\Phi}$, and $\mathbf{\Theta}$ are nonzero and $\mathbf{\Theta}\mathbf{\Theta}^\top$ is invertible, then with $C = \frac{|\mathrm{tr}(RF^\top FF)|}{||FF||^2 \cdot ||\mathbf{\Phi}|| \cdot ||\mathbf{\Theta}^\top (\mathbf{\Theta}\mathbf{\Theta}^\top)^{-1}||}$, we have:*

$$\min_t ||RF - tFF|| \geq C \cdot \min_t ||tI - \mathbf{\Phi}\mathbf{\Phi}^\top||,$$

What is important here is not the technical statement of Theorem 2, but rather the following qualitative versions.

Corollary 1 *If $\mathbf{\Phi}\mathbf{\Phi}^\top$ is not (close to) a scalar multiple of the identity, then RF is not (close to) a scalar multiple of FF.*

Theorem 3 *If the feedback weights are not a scalar multiple of $FF^\top FF$, then RF is not a scalar multiple of FF.*

Finally, we remark that in the compressive sensing regime $k \ll n$ and $\mathbf{\Phi}$ is a random matrix; thus, the hypothesis of the previous results are satisfied generically:

Theorem 4 *In adaptive compressed sensing, the receptive fields are almost surely not scalar multiples of the feedforward weights.*

5. DISCUSSION AND CONCLUSIONS

We have proposed *adaptive compressed sensing (ACS)*, a new scheme of learning under compressed sensing that forms a dictionary adapted to represent the compressed data optimally. The coding and learning scheme of ACS can be formulated as a neural network, building on an earlier sparse coding model [6]. Our model learns in the weights of the coding circuit while keeping the random projection fixed, as opposed to a previous suggestion which optimizes the compression performance by learning in the random projection [15].

Our study focuses on the application of ACS to understand how cortical regions in ascending sensory pathways can analyze and represent signals they receive through thalamo-cortical or cortico-cortical connections. Conventional sparse coding theories were succcessful in reproducing physiological responses in primary sensory regions but they require exact matches between feedforward connections and receptive field patterns of cortical neurons (see Theorem 1 and Fig 1(a) for an example). Although it has been shown that thalamocortical wiring is to some extent specific [10, 11], exact matches between feedforward circuitry and receptive fields are not supported by experimental data. In addition, a recent quantitative study of cortico-cortical projections suggests that the number of fibers reaching a target area can only be a fraction of the local neurons in the area of origin [12].

We have tested if ACS could serve as a computational model for how cortical areas can form a representation of data received through afferent projections that subsample the activity pattern in the previous stage. We demonstrate that ACS can form representation of visual data, though, unlike in conventional sparse coding models, the coding circuit receives only a subsampled version of the original data. Further, we have demonstrated that the algorithm is stackable in a hierarchy. The sparse code formed by ACS in a primary sensory area, when sent through another compressing projection can be decoded in a secondary sensory area into another meaningful visual representation. The simulation results proof the concept that the ACS model can serve as a generic building block in a communication scheme between cortical areas. The scheme consists of repeated cycles of compression and expansion. Specifically, a sparse local representations is compressed, sent through cortico-cortical projections and expanded to sparse local representations at the receiver end, reminiscent to Braitenberg's idea of the pump of thought [18]. The scheme of ACS suggests that representations in the brain can be sparse [19, 20] and dense [21, 22], with the type of code being lamina-specific. Regarding the still debated role of recurrent circuitry in producing orientation selectivity (e.g., [23, 24, 25]), ACS suggests that if the input subsamples the data then feedback in shaping the receptive fields becomes essential for coding efficiency.

6. REFERENCES

[1] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–608, 1996.

[2] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61:183 – 193, 1954.

[3] H. B. Barlow. *Understanding natural vision*. Springer-Verlag, Berlin, 1983.

[4] Anthony J. Bell and Terrence J. Sejnowski. The independent components of natural images are edge filters. *Vision Research*, 37:3327–3338, 1997.

[5] P. Földiák. Forming sparse representaions by local anti–Hebbian learning. *Biological Cybernetics*, 64:165–170, 1990.

[6] M Rehn and F. T. Sommer. A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *Journal of Computational Neuroscience*, 22:135–146, 2007.

[7] C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A Olshausen. Sparse coding via thresholding and local competition in neural circuits. *Neural Computation*, 20:2526–2563, 2008.

[8] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for l_1-minimization with applications to compressed sensing. *SIAM J. Imaging Sciences*, 1(1):143 – 168, 2008.

[9] L. Rebollo-Neira and D. Lowe. Optimised orthogonal matching pursuit approach. *IEEE Signal Processing Letters*, 9:137–140, 2002.

[10] B. Chapman, K. R. Zahs, and M. P. Stryker. Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column in ferret visual cortex. *J. of Neurosience*, 11(5):1347–1358, 1991.

[11] C. R. Reid and J.-M. Alonso. Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378:281–284, 1995.

[12] A. Schüz, D. Chaimow, D. Liewald, and M. Dortemann. Quantitative aspects of corticocortical connections: a tracer study in the mouse. *Cerebral Cortex*, 16:1474–1486, 2006.

[13] E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 58(8):1207–1223, 2006.

[14] E. Candes and M. Wakin. An indroduction to compressed sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[15] Y. Weiss, H. S. Chan, and W. T. Freeman. Learning compressed sensing. *Allerton Conference Proceedings*, http://www.cs.huji.ac.il/ weiss:1 – 9, 2007.

[16] G. Palm. *Neural Assemblies*. Springer, Berlin, 1982.

[17] A. Lansner. Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations. *Trends in Neurosciences*, 32(3):178–186, 2009.

[18] V. Braitenberg. Cell assemblies in the cerebral cortex. In R. Heim and G. Palm, editors, *Theoretical Approaches to Complex Systems*. Springer, 1977.

[19] M. R. DeWeese and A. M. Zador. Non-gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. *J. Neuroscience*, 26(47):12206 – 12218, 2006.

[20] T. Hromadka, M. R. Deweese, and A. M. Zador. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.*, 6(1):16–24, 2008.

[21] M.N. Shadlen and W.T. Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4:569–579, 1994.

[22] M.N. Shadlen and W.T. Newsome. Is there a signal in the noise? [comment]. *Current Opinion in Neurobiology*, 5:248–250, 1995.

[23] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154, 1962.

[24] J. A. Hirsch, J.-M. Alonso, R. C. Reid, and L. M. Martinez. Synaptic integration in striate cortical simple cells. *Journal of Neuroscience*, 18:9517 – 9528, 1998.

[25] D. Ferster and K. D. Miller. Neural mechanisms of orientation selectivity in the visual cortex. *Annual Review of Neuroscience*, 23:441–471, 2000.