Many models even suggest that each neuron has the feedforward access to the full data, for instance, to all pixels of an image patch. Conversely, their correspondence to the anatomical connectivity in cortex optimized orthogonal matching pursuit [9].

properties in primary visual cortex quantitatively, for instance, the 7, 8]. Some of these models are capable of reproducing the response network through mainly inhibitory lateral connections, see [5, 1, 6, a set of similar connectivity patterns: a neuron receives feedforward experience [1, 4]. There is now a wealth of such models, all based on understanding how essential response properties, such as orientation selectivity, are formed in sensory areas through development and

organizing network models for sparse coding have been critical in

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 orientational 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 [3] [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

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

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

For a given input \( x \), the sparse coding operation is given by an energy minimization

\[ a(x) := \arg \min_a E(x, a, \Psi) \in \mathbb{R}^n. \]  

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

Compressed sensing is a technique for data compression using a random projection matrix \( \Phi \) to compress the data \( x \in \mathbb{R}^m \) to \( \Phi x \in \mathbb{R}^k \) with \( k < m \). The decompression uses energy minimization [2]
Learning is executed by gradient descent on $\Theta$ of an error-based energy function similar to Eq. 1:

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

The original data is reconstructed as $\tilde{x} = \Psi a(\Phi x)$. In conventional compressed sensing, a fixed dictionary $\Psi$ is chosen. The decompression can be shown to work if (i) a dictionary $\Psi$ is used in which the data can be sparsely represented, (ii) matrices $\Phi$ and $\Psi$ 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 [6,6], we introduce adaptive compressed sensing (ACS), an adaptive version of compressed sensing governed by:

$$E(x, a, \Phi, \Theta) = \frac{1}{2} ||\Phi x - \Theta a||^2 + \lambda ||a||_{L_0}$$

$$= -x^T \Phi^T \Theta a + \frac{1}{2} a^T \Theta^T \Theta a + \lambda ||a||_{L_0} + \text{const.}$$

Learning is executed by gradient descent on $\Theta$ in exactly the same fashion as in conventional sparse coding, e.g. [1,16]. 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 $\Theta$, 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, $S(a) = \lambda ||a||_{L_0}$, 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

$$\frac{\partial E'}{\partial a_i} = -(x^T \Phi^T \Theta)_{i} + (\Theta^T \Theta a)_{i},$$

see [6] for further detail. In the neural network for the ACS method the feedforward weights are $FF := \Phi^T \Theta$ and the competitive feedback weights are $FB := -\Theta^T \Theta$. Note that if $\Phi$ is the identity matrix, ACS coincides with conventional sparse coding for which the corresponding neural network would be defined by $FF = \Phi$ and $FB = -FF^T FF$ [1,6]. The important difference between the two wiring schemes is that the feedforward weights of ACS sub-sample and mix the original data. Thus, coding and weight adaptation 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 $12 \times 12$ pixels, making the dimension of the data $m = 144$. For ACS, we used a sampling matrix $\Phi$ that downsampled the original data to $k = 60$ dimensions. All coding circuits contained $n = 432$ neurons, thereby producing representations of the original data $a \in \mathbf{R}^n$ that were three times over-complete. 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 sub-sampled version of the sparse code generated in the primary visual area $\Phi_2 a(x) \in \mathbf{R}^k$ and produces a sparse code $a_2 \in \mathbf{R}^n$, 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 $\Theta$ matrix. Note that computing the data dictionary from $\Theta$ requires an ill-posed step of matrix factorization: $\Theta = \Phi \Psi$. 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 $x \in \mathbf{R}^m$ as

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

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.
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 \( x \) is a column vector of random variables on a measure space \( \Omega \) with probability measure \( \mu \), the matrix \( RF \) will be an approximation to the correlation of \( x \) and \( a(x) \) (which is assumed to have zero mean): \( \text{Cor}(x, a) = \int x a(x)^{\top} d\mu \). 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 \( RF = \text{Cor}(x, a) \). We are interested in calculating the necessary relationships between the quantities \( RF, FF, FB, \Phi, \text{and } \Theta \). (Recall: \( FF = \Phi^\top \Theta \) and \( FB = -\Theta^\top \Theta \).

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

\[
\Phi RF = \int \Phi x a(x)^{\top} d\mu = \Theta \int a(x) a(x)^{\top} d\mu = \Theta D.
\]

In particular, this implies the following.

**Theorem 1** If \( \Phi \) 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 \( \Phi \) 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 1 below. We omit here the proofs.

**Theorem 2** If \( FF, \Phi, \text{and } \Theta \) are nonzero and \( \Theta \Theta^\top \) is invertible, then with \( C = \frac{||u(FF^\top FF)||}{||u(\Theta \Theta^\top)||} \), we have:

\[
\min_{t} ||RF - t FF|| \geq C \cdot \min_{t} ||tI - \Phi \Phi^\top||.
\]

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

**Corollary 1** If \( \Phi \Phi^\top \) 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 \( \Phi \) 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 [13].

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 successful in reproducing physiological responses in primary sensory regions but 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 has some extent of 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 that 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 representation. The simulation results prove 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 Braebenberg’s idea of the pump of thought [13]. 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
