Modeling Laminar Recordings from Visual Cortex with Semi-Restricted Boltzmann Machines

Anonymous Author(s) Affiliation Address email

Abstract

The proliferation of high density recording techniques presents us with new challenges for characterizing the statistics of neural activity over populations of many neurons. The Ising model, which is the maximum entropy model for pairwise correlations, has been used to model the instantaneous state of a population of neurons. This model suffers from two major limitations: 1) Estimation for large models becomes computationally intractable, and 2) it cannot capture higher-order dependencies. We propose applying a more general maximum entropy model, the semi-restricted Boltzmann machine (sRBM), which extends the Ising model to capture higher order dependencies using hidden units. Estimation of large models is made practical using minimum probability flow, a recently developed parameter estimation method for energy-based models. The partition functions of the models are estimated using annealed importance sampling, which allows for comparing models in terms of likelihood. Applied to 32-channel polytrode data recorded from cat visual cortex, these higher order models significantly outperform Ising models. In addition, extending the model to spatiotemporal sequences of states allows us to predict spiking based on network history. Our results highlight the importance of modeling higher order interactions across space and time to characterize activity in cortical networks.

033 034

036 037

000 001 002

003 004

006

008 009

010

011

016 017

018

019

021

024

025

026

027

028

029

031

1 Introduction

Electrophysiology is rapidly moving towards high density recording techniques that are capable of capturing the simultaneous activity of large populations of neurons. This raises the challenge of understanding how networks encode and process information in ways that go beyond feedforward receptive field models for individual neurons. Towards this goal, modeling the distribution of network states provides a way to discover communication patterns and functional connectivity between cells.

The Ising model, originally developed in the 1920s to describe magnetic couplings in a lattice [1], 044 has been applied to modeling electrophysiological data, particularly in the retina [2], but also for cortical recordings [3]. This model treats spikes from a population of neurons binned in time as 046 binary vectors and captures pairwise correlation between them. However, the model only provides 047 a good approximation for small groups of cells [4] and is difficult to extend to capture higher-order 048 correlations. Another approach is to model each cell conditioned on the rest of the population, using Generalized Linear Models (GLM, see [5]) with terms for cross-spike couplings which capture interactions between cells. A major limitation of GLMs is that they can only be estimated efficiently 051 if they are linear in the data, so they are not easily extended to higher order interactions. While there is progress extending the GLM with additional nonlinearities, preserving convexity on subproblems, 052 this approach has so far only been applied to stimulus terms [6], but not nonlinear interactions between cells.

In this work, we attempt to apply maximum entropy models to data from the visual cortex, which is less understood than retina or LGN. Even the existence of significant correlations between cortical cells is controversial [7, 8] and the relevance of pairwise and higher order correlations remains an open question. In contrast to retinal recordings, the complex 3D structure of cortical circuits makes it virtually impossible to record from more than a tiny fraction of cells in a microcircuit. Therefore it is important to have a mechanism to deal with missing data, which we do by including latent variables in the model. The resulting model is equivalent to a semi-restricted Boltzmann machine (sRBM), which combines a restricted Boltzmann machine with additional connections between visible units.

062 The main difficulty in working with energy-based models such as this is that they cannot be nor-063 malized in closed form, as the energy function only provides probabilities up to a multiplicative 064 normalization constant. This makes exact maximum likelihood estimation intractable and often necessitates sampling or approximations. Using Minimum Probability Flow (MPF, [9]), these models 065 can be learned more efficiently without computing the intractable partition function. This allows us 066 to estimate Ising models on higher-dimensional data than is otherwise possible, and to estimate the 067 sRBM in a straightforward way. We show that the sRBM can capture the distribution of states in a 068 cortical network of tens to hundreds of cells recorded from cat visual cortex significantly better than 069 a pairwise model, giving a strong indication of structure beyond pairwise correlations. 070

2 Methods

071

072 073

074

090

093 094

098 099

100

101 102

104 105

2.1 Model and Estimation

The sRBM consists of a set of binary visible units $\mathbf{x} \in \{0, 1\}^N$ corresponding to observed neurons in the data and a set of hidden units $\mathbf{h} \in \{0, 1\}^M$ that capture higher order dependencies. Weights between visible units, corresponding to an Ising model or fully visible Boltzmann machine, capture pairwise correlations in the data. Weights between visible and hidden units, corresponding to a Restricted Boltzmann Machine (RBM, [10]), learn to describe higher order structure.

The Ising model with visible-visible coupling weights $\mathbf{J} \in \mathcal{R}^{N \times N}$ and biases $\mathbf{b} \in \mathcal{R}^N$ has an energy function

$$E_{\rm I}(\mathbf{x}) = \mathbf{x}^T \mathbf{J} \mathbf{x} + \mathbf{b}^T \mathbf{x},\tag{1}$$

with associated probability distribution $p_{I}(\mathbf{x}) = \frac{1}{Z_{I}} \exp[-E_{I}(\mathbf{x})]$, where the typically intractable normalization constant, or partition function, $Z_{I} = \sum_{\{\mathbf{x}'\}} \exp[-E_{I}(\mathbf{x}')]$ consists of a sum over all 2^{N} system states.

The RBM with visible-hidden coupling weights $\mathbf{W} \in \mathcal{R}^{N \times M}$ and hidden and visible biases $\mathbf{b}_v \in \mathcal{R}^N$ and $\mathbf{b}_h \in \mathcal{R}^M$ has an energy function

$$E_{\rm R}(\mathbf{x}, \mathbf{h}) = \mathbf{x}^T \mathbf{W} \mathbf{h} + \mathbf{b}_v^T \mathbf{x} + \mathbf{b}_h^T \mathbf{h}, \qquad (2)$$

with associated probability distribution $p_R(\mathbf{x}, \mathbf{h}) = \frac{1}{Z_R} \exp\left[-E_R(\mathbf{x}, \mathbf{h})\right]$. The latent variables **h** can be analytically marginalized out of the distribution to obtain

$$p(\mathbf{x}) = \int d\mathbf{h} \, p(\mathbf{x}, \mathbf{h}) = \frac{1}{Z_R} \exp\left[-E_{\mathrm{R}}(\mathbf{x})\right],\tag{3}$$

where the energy for the marginalized distribution over \mathbf{x} (sometimes referred to in machine learning literature as the free energy) is

$$E_{\mathrm{R}}(\mathbf{x}) = -\sum_{i} \log(1 + \mathbf{w}_{i}^{T}\mathbf{x} + b_{h,i}) + \mathbf{b}_{v}^{T}\mathbf{x},$$
(4)

where \mathbf{w}_i are rows of the matrix \mathbf{W} . The energy function for an sRBM combines the Ising model and RBM energy terms,

$$E_{s}(\mathbf{x}, \mathbf{h}) = \mathbf{x}^{T} \mathbf{J} \mathbf{x} + \mathbf{x}^{T} \mathbf{W} \mathbf{h} + \mathbf{b}_{v}^{T} \mathbf{x} + \mathbf{b}_{h}^{T} \mathbf{h}.$$
(5)

As with the RBM, it is straightforward to marginalize over the hidden units for an sRBM,

$$p_{\rm s}\left(\mathbf{x}\right) = \frac{1}{Z_{\rm s}} \exp\left[-E_{\rm s}(\mathbf{x})\right],\tag{6}$$

106
107
$$E_{s}(\mathbf{x}) = \mathbf{x}^{T} \mathbf{J} \mathbf{x} - \sum_{i} \log(1 + \mathbf{w}_{i}^{T} \mathbf{x} + b_{h,i}) + \mathbf{b}_{v}^{T} \mathbf{x}.$$
(7)

A model based on the sRBM has previously been applied to natural image patches [11], with the parameters estimated using contrastive divergence (CD, [12]).

Instead of CD or other methods based on sampling, we train the models using Minimum Probability
 Flow (MPF, [9]), a recently developed estimation method for energy based models. MPF works by
 minimizing the KL divergence between the data distribution and the distribution which results from
 moving slightly away from the data distribution towards the model distribution. This KL divergence
 will be uniquely zero in the case where the model distribution is identical to the data distribution.
 While CD is a stochastic heuristic for parameter update, MPF provides a deterministic and easy
 to evaluate objective function. Second order gradient methods can therefore be used to speed up
 optimization considerably. The MPF objective function

$$K = \sum_{\mathbf{x}\in\mathcal{D}}\sum_{\mathbf{x}'\notin\mathcal{D}}g\left(\mathbf{x},\mathbf{x}'\right)\exp\left(\frac{1}{2}\left[E(\mathbf{x})-E(\mathbf{x}')\right]\right)$$
(8)

measures the flow of probability out of data states x into neighboring non-data states x', where the connectivity function $g(\mathbf{x}, \mathbf{x}') = g(\mathbf{x}', \mathbf{x}) \in \{0, 1\}$ identifies neighboring states, and \mathcal{D} is the list of data states. We consider the case where the connectivity function $g(\mathbf{x}, \mathbf{x}')$ is set to connect all states which differ by a single bit flip,

$$g(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & H(\mathbf{x}, \mathbf{x}') = 1\\ 0 & \text{otherwise} \end{cases},$$
(9)

where $H(\mathbf{x}, \mathbf{x}')$ is the Hamming distance between \mathbf{x} and \mathbf{x}' . See the supplementary material for a derivation of the MPF objective function and gradients for the sRBM, RBM, and Ising models. In all experiments, minimization of K was performed by LBFGS.

To prevent overfitting, all models were estimated with an L_1 sparseness penalty on the coupling parameters. This was done by adding a term of the form $\lambda \sum_{i,j} |J_{ij}| + \lambda \sum_{j,k} |W_{jk}|$ to the objective function, summing over the absolute values of the elements of both the visible and hidden weight matrices. The optimal sparseness λ was chosen by cross-validating the log-likelihood on a holdout set.

137 Since MPF learning does not give an estimate of the partition function, we use Annealed Importance Sampling (AIS, [13]) to compute normalized probabilities. AIS is a sequential Monte Carlo method 138 that works by gradually morphing a distribution with a known normalization constant (in our case 139 a uniform distribution over \mathbf{x}) into the distribution of interest. Normalizing the distribution allows 140 us to compute the log likelihood of the Ising and Boltzmann machine models $p_{
m model}$ and compare it 141 to the likelihood of an independent firing rate model $p_{ind}(\mathbf{x}) = \prod_i (r_i x_i + (1 - r_i)(1 - x_i))$ with 142 firing rates r_i for individual cells *i*. The independent rate model does not have free parameters, and 143 provides a useful reference point. We express the log likelihood as a gain in information rate over 144 this baseline. The excess log likelihood is defined in terms of sample expectations as 145

$$\mathcal{L} = \frac{1}{N} \sum_{\mathbf{x} \in \mathcal{D}} \left(\log_2 p_{\text{model}}(\mathbf{x}) - \sum_i \log_2 \left(r_i x_i + (1 - r_i)(1 - x_i) \right) \right).$$
(10)

Logarithms with base 2 gives the likelihood in units of bits per time bin, dividing by the bin size it can be expressed as an information rate in bits/s. This provides a quantitative way to do model comparison.

2.2 Recordings

118 119

120

126 127 128

146 147 148

152 153

154

Data were recorded from anesthetized cat visual cortex in response to a custom set of full field natural movie stimuli. Movies of 8-30 minutes duration were captured at 300 frames per second with a minimum of scene changes and presented with 150 Hz frame rate to avoid artifacts such as cells synchronizing to the frame rate or evoked potentials due to sudden luminance changes.

Recordings were made with single shank 32 channel polytrodes with a channel spacing of $50\mu m$, spanning all the layers of visual cortex. Individual data sets had on the order of of 25 simultaneously recorded neurons. The data was spike sorted offline using k-means clustering (KlustaKwik) with a manual cleanup step (MClust). Unless noted otherwise, spikes were binned at 20ms where bins with

Cells

10⁴

tuno2 O

10⁰

10

162

163



170

171

172

173 174 175



178

180 181

Figure 1: (a) Example data from 28 cells, 20ms bins, 3s of data total. Columns of this matrix are the input to our algorithm. (b) Power-law distribution of data patterns, log-log plot showing all data states ordered by the frequency of occurrence. The dashed line indicates states that have only been observed once. (c) Histogram of correlation coefficients between pairs of states. The data is dominated by weak positive correlations.

104

a) 3s of network activity across 28 cells in 20ms bins

50

b) Frequency of states

 10^{2}

State

ε,

Time (bin number)

 10^{3}

10²

10¹

-0.02

100

c) Pairwise correlations

0

0.02

Correlation coefficient

150

0.04

186 187

a single spike (2.4% of bins) and multiple spikes (0.9% of bins) were both treated as spiking and the rest as non-spiking bins.

To register individual recording channels with cortical layers, recording locations were reconstructed
 from Nissl-stained histological sections, and a current source density analysis in response to flashed
 full-field stimuli was performed to infer the location of cortical layer IV on the polytrode [14].

The models were estimated on a data set of 180,000 data vectors corresponding to 60 minutes of recording time. The data were split into two subsets of 90,000 data points: one a training set for parameter estimation and the other a test set for cross-validation. We also analyzed spatiotemporal patterns of data, which were created by concatenating consecutive state vectors. Concatenating 15 time bins of 6.7ms allows us to discover spatiotemporal patterns and predict spiking given the history over the last 100ms past state. These models were trained on 54,000 samples and cross-validated on a hold-out set of equal size to determine the optimal sparseness penalty.

3 Results

201 202

200

203 We estimated Ising, RBM and sRBM models for a population of cortical cells simultaneously 204 recorded across all cortical layers in a microcolumn. The data sets had 23 single units and 28 205 multiunits, from which we estimated separate models (results are qualitatively similar, single unit 206 data not shown). The population was verified to be visually responsive and the majority of cells 207 were orientation selective simple cells (20 of the multiunits, 10 of the isolated single cells). Fig. 1 a) shows an example of the 20ms binned data. Since the firing is sparse, the distribution of states 208 follows an approximate power law, which is shown in (b). Consistent with the literature [7], there 209 are weak positive correlations between most pairs of cells, as Fig. 1 c) shows. 210

The coupling matrices for all three models are shown in Fig. 2. The Ising model has many small weights that encourage positive pairwise correlations by lowering the energy of connected states being active simultaneously. The matrix consists of two dominant clusters, one with the 9 deepest cells and one with the 19 more superficial cells. This general structure is also picked up in the RBM, which has two hidden units that connect almost exclusively to superficial cells (1 and 2), and other units that connect to deep (3 and 4) or intermediate (5 and 6) cells. The sRBM has more sparse



Figure 2: Connectivity matrices of the three models estimated for 28 multiunits, and the correlation matrix of the training data. The horizontal lines indicate approximate boundaries between cortical layers V/VI, layer IV and layer II/III. (a) Correlation matrix of training data. (b) Ising model with the bias terms on the diagonal. The model has many small coupling terms that encode positive correlations. (c) The RBM coupling matrix is shown with hidden units ordered by activity from left to right, and an additional visible unit on top which corresponds to the hidden bias. (d) The sRBM weights are shown in the same way with the visible couplings on the left and hidden couplings on the right. The visible couplings are qualitatively very different from the Ising model, as much of the structure is now captured by hidden units.

229

230

231

232

233

234

235

236

239 connectivity for both the pairwise and second order units. The pairwise units are balanced around 240 zero, showing that there are both excitatory and inhibitory pairwise connections once the strong 241 global patterns have been explained by higher order units. Positive couplings within the superficial 242 (II/III) and intermediate laminae (IV), as well as the inhibitory interactions between intermediate 243 and deep (V/VI) layers become visible. The majority of the structure though is explained by a small 244 number of hidden units that connect to groups of either deep or superficial neurons. The separation 245 agrees well with the physiological layer boundaries, which are indicated as horizontal lines in the 246 coupling matrices.

The RBM and sRBM in this comparison were both estimated with 25 hidden units, but we show only units that did not go to zero due to the sparseness constraint. Since the model with the optimal level of sparseness already turns off some of the hidden units entirely, it was unnecessary to vary the number of hidden units.

In the sRBM it is easy to distinguish features of the data explained by second order interactions, from those that require higher order interactions. As observed in [11], the effect of the pairwise couplings is similar to whitening. By having pairwise units that explain pairwise interactions in the data, the hidden units are left to capture only higher order dependencies, significantly simplifying the structure of both coupling matrices.

We compare the likelihood of the different models as a function of the sparseness parameter λ so that each model is optimized individually to give the maximum likelihood on a cross-validation dataset. Fig. 3 a) shows the comparison of the excess log likelihood over the independent model for the three different models in units of bits/s for the entire population. Both higher-order models significantly outperform the Ising model in fitting the data.

262 While the optimal sparseness parameter λ is largest for the Ising model, it is worth pointing out 263 that the RBM requires significantly fewer nonzero parameters to provide a better fit to the data. 264 In the Ising model 227/529 elements are larger than a cut-off of .001 (43%), in the RBM 191/624 265 elements are above this value (31%), and in the sRMB 285/1129 elements are active (26%). The 266 RBM provides a high log likelihood with the fewest parameters, showing the importance of higher 267 order over pairwise dependencies.

In Fig. 3 b) we compare the relative performance of Ising model with RBM and sRBM models as a function of model dimensionality. The number of parameters of the RBM and Ising model were kept the same by fixing the number of hidden units in the RBM to be equal to the number of visible units,



309 Figure 3: (a) Model comparison using likelihood gain over the firing rate model. All three models 310 outperform the independent model by about 20 bits/s for the population of 28 cells. The higher 311 order models with hidden units give a further improvement of about 2 bit/s for the population. The models profit relatively little from the L_1 regularization, as there is sufficient training data to 312 constrain the parameters (b) Comparing models of different size, we can see how the likelihood per 313 neuron increases as each neuron is modeled as part of a larger population. This effect holds both 314 for Ising and higher order models, but since the Ising model cannot capture many of the relevant 315 dependencies, the increase in likelihood levels off much more quickly. (c) Scatter plot of test data 316 set showing empirical probabilities against model probabilities for all three models. For frequent 317 states, all models perform well. For rare states, the sRBM significantly outperforms simpler models. 318

319

320

the sRBM was also estimated with a square weight matrix for the hidden layer. As expected, the
 likelihood per visible unit increases monotonically with the network size for all models, as additional
 information gained from network interactions can only improve the predictive power of the model.
 However, the curve levels off more quickly for the Ising model than for the RBM and sRBM. This



Figure 4: (a) Hidden units of spatiotemporal sRBM model. For each hidden unit, the horizontal 345 axis is time and the vertical axis visible units. (b) Likelihood gain over independent rate model 346 of the model for one cell condition on the network state. Depending on the firing rate of the cell, 347 more additional information can be encoded by the higher order models. (c) Spike prediction from network history. For a randomly selected cell, we show 5s of predicted activity given the history of 348 the network state. The inserts show model predictions at times where actual spikes occur, in each 349 case there is an elevated probability under the models. The Ising model tends to give more false 350 positives than the higher order models. 351

354

355

356

324

326

327

331

333 334

337

338

339

341

suggests that as the model size increases, higher order structure that is not well captured by the Ising model becomes increasingly important. Only for very small models with on the order of 10 units does the Ising model perform similarly to the higher order models. This observation is in agreement with [4], who noted the same limitation of Ising models for higher dimensional data from the retina.

357 Another way to compare the different models is to show a scatter plot of probabilities under the 358 model against empirical probabilities for the test data. Fig. 3c compares the independent, Ising, 359 RBM and sRBM models to the empirical distribution. In this plot, we can also visualize the expected 360 performance range for a perfect model that assigns patterns the same probability with which they 361 are generated by the brain. This performance range follows from the counting error in a binomial 362 distribution. The 5% and 95% confidence intervals for the binomial counting error on the model 363 probabilities are shown as dotted lines, the solid line is the identity. Even though the higher order models outperform the Ising model significantly in log likelihood, all three closely approach the 364 empirical probabilities. 365

366 367

368

Spatiotemporal models 3.1

369 The same models can be used to capture spatiotemporal patterns. Consecutive network states binned at 6.7ms are concatenated in blocks of 15 time steps, for a total network dimensionality of 150 with 370 subpopulations of 10 cells. The model was cross-validated and the sparseness parameter optimized 371 in the same way as for the instantaneous model. This allows us to learn kernels that describe the 372 temporal structure of interactions between cells. Additionally, we can compute the conditional dis-373 tribution of single units given the state of the rest of the network and use this distribution to predict 374 spiking based on the network history. 375

Fig. 4a shows the learned weights of 25 hidden units for a spatiotemporal sRBM model. Each 376 subplot corresponds to one hidden unit, which connects to 10 neurons (vertical axis) across 15 time 377 steps or 100ms (horizontal axis, present to past). There are two types of patterns learned by the hidden units. One population specializes in spatial coupling across different cells at a constant time lag.
The remaining units describe smooth, long-range temporal dependencies, typically for individual
cells. Both of these populations capture higher order structure connecting many neurons that would
be hard to approximate with pairwise couplings. For instance, it is difficult for a pairwise model to
assign high probability to a group of neurons activating simultaneously without also assigning high
probability to the activation of subgroups. The temporal structure occurs at time scales of 20ms to
50ms, falling into the beta and gamma frequency range.

385 By conditioning on the state of the network, we can compute how much information about a cell 386 is captured by the model over a naive prediction based on the firing rate of a cell. This is plotted 387 in 4 b) in a similar way to excess log likelihood for the entire population in Fig. 3. Depending 388 on the firing rate of the cell, all three models significantly increase the prediction performance over the independent model. The higher order models consistently outperform the pairwise Ising model, 389 and the sRBM consistently outperforms the RBM. Fig. 4 c) provides a more intuitive look at the 390 prediction. For 5s of data we show the actual spikes and the network prediction for one cell. The 391 independent model would correspond to a constant probability over time. The inserts show three 392 randomly chosen 150ms segments where an actual spike occurred in the data. In each case there is 393 an elevation in the model's firing probability. 394

395 396

397

4 Conclusion

Our comparison of second- and higher order models shows that in cortical networks, higher order interactions play a significant role. This effect is independent of network size and holds for spatial and temporal interactions between cells.

While the present paper is the first attempt at probabilistically modeling the population response of neurons within a column of cat visual cortex, several works have characterized other neural populations with similar models. The original idea of applying Ising models to retinal data [2] has since been extended to include higher order interactions [4]. The authors claim that in the retina, small populations of cells can be modeled well using only pairwise correlations. This is not the case for cortical data, where higher order interactions dominate even small networks.

The key difference between our work and that by Ganmor and Elad [4] is that our estimation method allows us to assign probabilities to all activity patterns. Their model uses higher order coupling terms to build an accurate probabilistic model of patterns observed in the data. However, they do not assign probabilities to unobserved patterns. If the energy function from their model is used for a probabilistic model over all activity patterns, then the majority of the probability mass is assigned to states far from the data, and the resulting log likelihood is poor.

413 In contrast to these maximum entropy models, which have the goal to describe the distribution 414 of states without making any predictions about connectivity or function, generalized linear models 415 (GLMs) have been established as successful stimulus-response models for primary visual cortex. By including the response of other neurons in addition to stimulus terms to predict the response of a cell 416 in the population, GLMs have also been used to capture network interactions. This has been shown 417 to greatly improve the performance over a feedforward stimulus model even in the retina [5], and 418 is presumably of even greater importance in cortex. By computing conditional distributions on the 419 network state we have shown that the maximum entropy population model can be used in a similar 420 way for spike prediction, opening the possibility for future comparisons of the relative performance 421 of these two model classes. 422

423 424

425

426

427

428

429

430

431

References

- [1] Ernst Ising. Beitrag zur theorie des ferromagnetismus. Zeitschrift für Physik A, 31(1):253–258, 1925.
- [2] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007– 12, April 2006.
- [3] MT Schaub. The Ising decoder: reading out the activity of large neural ensembles. *Journal of computational neuroscience*, 2012.

432	[4]	Elad Ganmor, Ronen Segev, and Elad Schneidman. Sparse low-order interaction network
433		underlies a highly correlated and learnable neural population code. <i>Proceedings of the National</i>
434		Academy of Sciences of the United States of America, 108(23):9679–84, June 2011.
436	[5]	Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, E J
437		Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. <i>Nature</i> , 454(7207):995–9, August 2008.
438	[6]	D. Butts, C. Weng, J. Jin, JM. Alonso, and L. Paninski. Temporal Precision in the Visual
439 440		Pathway through the Interplay of Excitation and Stimulus-Driven Suppression. <i>Journal of Neuroscience</i> 31(31):11313–11327 August 2011
441	[7]	Alexander S Ecker Dhilipp Barans, Georgias Kaliris, Matthias Bathga, Nikos K Lagathatis
442 443	[/]	and Andreas S Tolias. Decorrelated neuronal firing in cortical microcircuits. <i>Science (New York, N.Y.)</i> , 327(5965):584–7, January 2010.
444	[8]	Alfonso Repart I De la Rocha Peter Bartho Liad Hollender N Parga Alex Reves and K D
445 446	[0]	Harris. The asynchronous state in cortical circuits. <i>Science</i> , 327(5965):587–590, 2010.
447	[9]	Jascha Sohl-Dickstein. Minimum Probability Flow Learning. ICML, 2011.
448 449	[10]	GE Hinton and Terrence J Sejnowski. Learning and relearning in Boltzmann machines. In <i>Foundations of Neural Computation</i> , MIT Press, Cambridge, Mass., 1986.
450	[11]	Simon Osindero and Geoffrey Hinton Modeling image patches with a directed hierarchy of
451	[11]	Markov random fields. <i>Advances in neural information processing</i> , pages 1–8, 2008.
452	[12]	Geoffrey E Hinton Training products of experts by minimizing contrastive divergence Neural
453	[12]	<i>computation</i> , 14(8):1771–800, August 2002.
454	[13]	RM Neal Annealed importance sampling Technical report University of Toronto Toronto
455	[10]	2001.
450	[14]	U Mitzdorf. Current source-density method and application in cat cerebral cortex: investigation
458		of evoked potentials and EEG phenomena. <i>Physiological reviews</i> , 65(1):37–100, January 1985.
459		
460		
461		
461 462		
461 462 463		
461 462 463 464		
461 462 463 464 465		
461 462 463 464 465 466		
461 462 463 464 465 466 467		
461 462 463 464 465 466 466 467 468		
461 462 463 464 465 466 466 467 468 469 470		
461 462 463 464 465 466 467 468 469 470 471		
461 462 463 464 465 466 467 468 469 470 471 472		
461 462 463 464 465 466 467 468 469 470 471 472 473		
461 462 463 464 465 466 467 468 469 470 471 472 473 474		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483		
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484		