6

Dynamic Network Models

Synaptic plasticity is believed to be the fundamental building block of learning and memory in the brain (Dayan and Abbott, 2001). Its study is of crucial importance to understanding the activity and function of neural circuits. With innovations in neural recording technology providing access to the simultaneous activity of increasingly large populations of neurons, statistical models are promising tools for formulating and testing hypotheses about the dynamics of synaptic connectivity. Advances in optical techniques (Packer et al., 2012; Hochbaum et al., 2014), for example, have made it possible to simultaneously record from and stimulate large populations of synaptically connected neurons. Armed with statistical tools capable of inferring time-varying synaptic connectivity, neuroscientists could test competing models of synaptic plasticity, discover new learning rules at the mono-synaptic and network level, investigate the effects of disease on synaptic plasticity, and potentially design stimuli to modify neural networks.

Despite the popularity of autoregressive models for spike data, like the GLM (Paninski, 2004; Truccolo et al., 2005; Pillow et al., 2008), relatively little work has attempted to model the time-varying nature of neural interactions. Here we model interaction weights as a dynamical system governed by parametric synaptic plasticity rules. Building on the work of preceding chapters, we show how synaptic plasticity rules can be modeled as dynamics rules that govern how weights evolve in an activity-dependent manner. In doing so, we imbue the weights with a biophysical interpretation that we explicitly avoided in previous chapters. We discuss when this interpretive leap is warranted.

To perform inference in this model, we use particle Markov chain Monte Carlo (pMCMC) (Andrieu et al., 2010), a recently developed inference technique for time series with nonlinear dynamics. We use this new modeling framework to examine the problem of using recorded data to distinguish between proposed variants of spike-timing-dependent plasticity (STDP) learning rules. On synthetic data generated from the biophysical simulator NEURON, we show that we can recover the weight trajectories, the pattern of connectivity, and the underlying learning rules.

A Biophysical Interpretation of the GLM

The nonlinear autoregressive models of the previous chapter treat the spike count, $s_{t,n}$, as a random variable whose distribution depends on a nonnegative firing rate, $\lambda_{t,n}$. The firing rate is modeled as a nonlinear function of an activation, $\psi_{t,n}$, which is taken to be a linear function of the spike history. This linear-nonlinear cascade is often called a generalized linear model (GLM) (Paninski, 2004; Truccolo et al., 2005).

From a biophysical perspective, the activation can be thought of as analogous to the cell's membrane potential. The nonlinearity that links the activation to the firing rate approximates the spiking threshold of the neuron. When the membrane potential exceeds the spiking threshold potential of the cell, $\lambda_{t,n}$ rises to reflect the rate of the cell's spiking, and when the membrane potential decreases below the spiking threshold, $\lambda_{t,n}$ decays to zero.

As before, we model the activation, or membrane potential, as a linear function of the spike history,

$$\psi_{t,n} = \psi_n^{(0)} + \sum_{n=1}^{N} \sum_{d=1}^{D} h_{n' \to n}[d] \cdot s_{t-d,n'}. \tag{6.1}$$

where $\psi_n^{(0)}$ is now the resting potential and $h_{n' \to n}[d]$ is a post-synaptic potential that preceding spikes on neuron n' induce on the membrane potential of neuron n at lag d.

From this semi-biophysical perspective it is clear that one shortcoming of the models de-

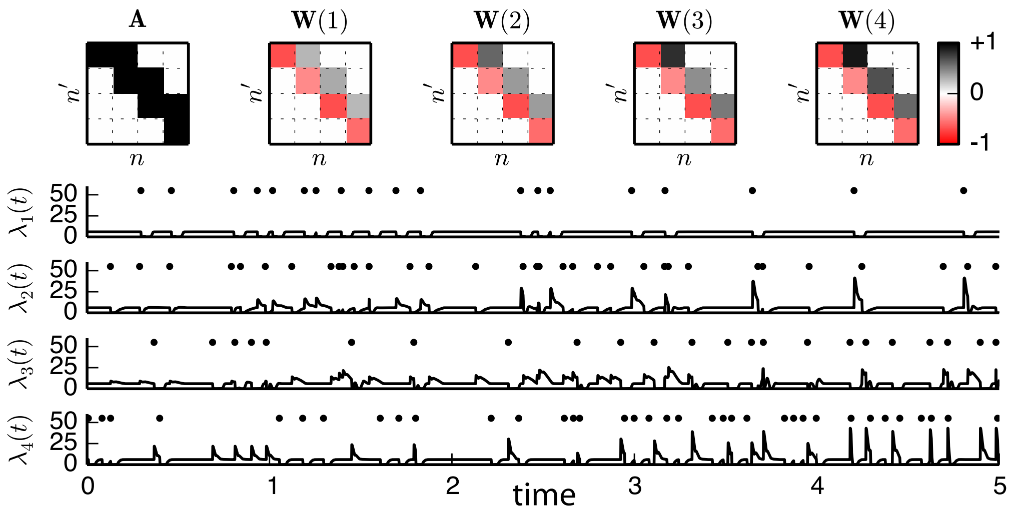

Figure 6.1: A simple network of four sparsely connected neurons whose synaptic weights are changing over time. Here, the neurons have inhibitory self connections to mimic refractory effects, and are connected via a chain of excitatory synapses, as indicated by the nonzero entries $a_{1\to 2}$, $a_{2\to 3}$, and $a_{3\to 4}$. The corresponding weights of these synapses are strengthening over time (darker entries in $\boldsymbol{W}$), leading to larger impulse responses in the firing rates and a greater number of induced post-synaptic spikes (black dots), as shown below.

veloped thus far is that they do not account for time-varying connectivity, despite decades of research showing that changes in synaptic weight occur over a variety of time scales and are the basis of many fundamental cognitive processes. This absence is due, in part, to the fact that this direct biophysical interpretation is not warranted in most traditional experimental regimes, e.g., in multi-electrode array (MEA) recordings where electrodes are relatively far apart. However, as high resolution optical recordings grow in popularity, this assumption must be revisited; this is a central motivation for the model developed in this chapter.

There have been a few efforts to incorporate dynamics into the GLM. Stevenson and Koerding (2011) extended the GLM to take inter-spike intervals as a covariates and formulated a generalized bilinear model for weights. Eldawlatly et al. (2010) modeled the time-varying parameters of a GLM using a dynamic Bayesian network (DBN). However, neither of these approaches accommodate the breadth of synaptic plasticity rules present in the literature. For example, parametric STDP models with hard bounds on the synaptic weight are not congruent with the convex optimization techniques used by (Stevenson and Koerding, 2011), nor are they naturally expressed in a DBN. Here we model time-varying synaptic weights as a potentially nonlinear dynamical system and perform inference using particle MCMC.

Nonstationary, or time-varying, models of synaptic weights have also been studied outside the context of GLMs. For example, Petreska et al. (2011) applied hidden switching linear dynamical systems models to neural recordings. This approach has many merits, especially in traditional MEA recordings where synaptic connections are less likely and nonlinear dynamics are not necessarily warranted. Outside the realm of computational neuroscience and spike train analysis, there exist a number of dynamic statistical models, such as the dynamic generalized linear models of West et al. (1985). However, the types of models we are interested in for studying synaptic plasticity are characterized by domain-specific transition models and sparsity structure, and until recently, the tools for effectively performing inference in these models have been limited.

A Sparse Time-Varying Generalized Linear Model

In order to capture the time-varying nature of synaptic weights, we extend the standard GLM by first factoring the impulse responses in the firing rate of (6.1) into a product of three terms:

$$h_{n'\to n}[d,t] \triangleq a_{n'\to n} \cdot w_{n'\to n}[t] \cdot \hbar_{n'\to n}[d]. \tag{6.2}$$

Here, $a_{n'\to n} \in \{0,1\}$ is a binary random variable indicating the presence of a direct synapse from neuron n' to neuron n, $w_{n'\to n}[t] \in \mathbb{R}$ is a non stationary synaptic "weight" trajectory associated with the synapse, and $\hbar_{n'\to n}[d]$ is a nonnegative, normalized impulse response, i.e. $\sum_{d=1}^{D} \hbar_{n'\to n}[d] \cdot \Delta t = 1$. Requiring $\hbar_{n'\to n}[d]$ to be normalized gives meaning to the synaptic weights: otherwise w would only be defined up to a scaling factor. For simplicity, we assume $\hbar[d]$ does not change over time, that is, only the amplitude and not the duration of the PSPs is time-varying. This restriction could be adapted in future work.

As in previous chapters, we model the normalized impulse responses as a linear combination of basis functions. In order to enforce the normalization of $\hbar[d]$, however, we use a *convex* combination of normalized, nonnegative basis functions. That is,

$$\hbar_{n'\to n}[d] \equiv \sum_{b=1}^{B} \theta_{n'\to n}^{(b)}\, \phi_b[d],$$

where $\sum_{d=1}^{D} \phi_b[d] \cdot \Delta t = 1$ and $\sum_{b=1}^{B} \theta_{n'\to n}^{(b)} = 1$.

The binary random variables $a_{n'\to n}$, which can be collected into an $N \times N$ binary matrix $\boldsymbol{A}$, model the connectivity of the synaptic network. Similarly, the collection of weight trajectories $\{\{w_{n'\to n}[t]\}\}_{n',n}$, which we will collectively refer to as $\boldsymbol{W}[t]$, model the time-varying synaptic weights. This factorization is often called a *spike-and-slab* prior (Mitchell and Beauchamp, 1988), and it allows us to separate our prior beliefs about the structure of the synaptic network from those about the evolution of synaptic weights. For example, in the most general case we might incorporate the probabilistic network models of previous chapters as prior distributions for $\boldsymbol{A}$, but here we limit ourselves to the simplest network model, the independent Bernoulli, or Erdős-Rényi model. Under this model, each $a_{n'\to n}$ is an independent identically distributed Bernoulli random variable with sparsity parameter ρ.

Figure 6.1 illustrates how the adjacency matrix and the time-varying weights are integrated into the GLM. Here, a four-neuron network is connected via a chain of excitatory synapses, and the synapses strengthen over time due to an STDP rule. This is evidenced by the increasing amplitude of the impulse responses in the firing rates. With larger synaptic weights comes an increased probability of post-synaptic spikes, shown as black dots in the figure. In order to model the dynamics of the time-varying synaptic weights, we turn to a rich literature on synaptic plasticity and learning rules.

Learning Rules for Time-Varying Synaptic Weights

Decades of research on synapses and learning rules have yielded a plethora of models for the evolution of synaptic weights (Caporale and Dan, 2008). In most cases, this evolution can be written as a dynamical system,

$$\boldsymbol{W}[t+1] = \boldsymbol{W}[t] + \ell\left(\boldsymbol{W}[t],\, \boldsymbol{S}_{\le t}, \boldsymbol{\vartheta}\right) + \epsilon(\boldsymbol{W}[t], \boldsymbol{\vartheta})$$

where ℓ is a potentially nonlinear *learning rule* that determines how synaptic weights change as a function of previous spiking, $\boldsymbol{S}_{\leq t}$. This framework encompasses rate-based rules such as the Oja rule (Oja, 1982) and timing-based rules such as STDP and its variants. The additive noise, $\epsilon(\boldsymbol{W}[t], \boldsymbol{\vartheta})$, need not be Gaussian, and many models require truncated noise distributions.

Following biological intuition, many common learning rules factor into a product of simpler functions. For example, STDP (defined below) updates each synapse independently such that the learning rule for $w_{n' \to n}$ only depends on the current weight, $w_{n' \to n}[t]$, and the pre- and post-synaptic spike history, $\boldsymbol{S}_{\leq t}$. Biologically speaking, this means that plasticity is local to the synapse. More sophisticated rules allow dependencies among the columns of $\boldsymbol{W}$. For example, the incoming weights to neuron n may depend upon one another through normalization, as in the Oja rule (Oja, 1982), which scales synapse strength according to the total strength of incoming synapses.

Extensive research in the last fifteen years has identified the relative spike timing between the pre- and post-synaptic neurons as a key component of synaptic plasticity, among other factors such as mean firing rate and dendritic depolarization (Feldman, 2012). STDP is therefore one of the most prominent learning rules in the literature today, with a number of proposed variants based on cell type and biological plausibility. In the experiments to follow, we will make use of two of these proposed variants. First, consider the canonical STDP rule with a "double-exponential" function parameterized by $\boldsymbol{\vartheta} = \{\tau_-, \tau_+, A_-, A_+\}$ (Song et al., 2000), in which the effect of a given pair of pre-synaptic and post-synaptic spikes on a weight may be written:

$$\ell\left(w_{n' \to n}[t], \boldsymbol{S}_{\leq t}; \boldsymbol{\vartheta}\right) = \ell_+(\boldsymbol{S}_{\leq t}, A_+, \tau_+) - \ell_-(\boldsymbol{S}_{\leq t}, A_-, \tau_-), \tag{6.3}$$

where,

$$\begin{aligned}
\ell_+(\boldsymbol{S}_{\leq t}, A_+, \tau_+) &= s_{t,n} \sum_{t'=1}^{t} s_{t',n'} \cdot A_+ \cdot e^{(t-t')/\tau_+}, \\
\ell_-(\boldsymbol{S}_{\leq t}, A_-, \tau_-) &= s_{t,n'} \sum_{t'=1}^{t} s_{t',n} \cdot A_- \cdot e^{(t-t')/\tau_-}.
\end{aligned}$$

This rule states that weight changes only occur at the time of pre- or post-synaptic spikes, and that the magnitude of the change is a nonlinear function of inter-spike intervals.

A slightly more complicated model known as the multiplicative STDP rule extends this by bounding the weights above and below by W_{max} and W_{min}, respectively (Morrison et al., 2008). Then, the magnitude of the weight update is scaled by the distance from the threshold:

$$\begin{aligned}\ell\left(w_{n'\to n}[t], \boldsymbol{S}_{\leq t}, \boldsymbol{\vartheta}\right) = \tilde{\ell}_{+}(\boldsymbol{S}_{\leq t}, A_{+}, \tau_{+})\ (W_{\mathsf{max}} - w_{n'\to n}[t]),\\ -\ \tilde{\ell}_{-}(\boldsymbol{S}_{\leq t}, A_{-}, \tau_{-})\ (w_{n'\to n}[t] - W_{\mathsf{min}}).\end{aligned} \tag{6.4}$$

Here, by setting $\tilde{\ell}_{\pm} = \min(\ell_{\pm}, 1)$, we enforce that the synaptic weights always fall within $[W_{\mathsf{min}}, W_{\mathsf{max}}]$. With this rule, it often makes sense to set W_{min} to zero.

Similarly, we can construct an additive, bounded model which is identical to the standard additive STDP model except that weights are thresholded at a minimum and maximum value. In this model, the weight never exceeds its set lower and upper bounds, but unlike the multiplicative STDP rule, the proposed weight update is independent of the current weight except at the boundaries. In the canonical STDP model it is sensible to use Gaussian noise, but in the bounded multiplicative model we use truncated Gaussian noise to respect the hard upper and lower bounds on the weights. Note that this noise is dependent upon the current weight, $w_{n'\to n}[t]$.

The nonlinear nature of this rule, which arises from the multiplicative interactions among the parameters, $\boldsymbol{\vartheta} = \{A_{+}, \tau_{+}, A_{-}, \tau_{-}, W_{\mathsf{max}}, W_{\mathsf{max}}\}$, combined with the potentially non-Gaussian noise models, pose substantial challenges for inference. However, the computational cost of these detailed models is counterbalanced by dramatic expansions in the flexibility of the model and the incorporation of *a priori* knowledge of synaptic plasticity. These learning models can be interpreted as strong regularizers of models that would otherwise be highly under-determined, as there are N^2 weight trajectories and only N spike trains. In the next section we will leverage powerful new techniques for Bayesian inference in order to capitalize on these expressive models of synaptic plasticity.

Inference via Particle MCMC

The traditional approach to inference in the standard GLM is penalized maximum likelihood estimation. For a model with Poisson observations and a nonlinear link function, $g : \mathbb{R} \to \mathbb{R}_+$, the log likelihood is,

$$\log p(\boldsymbol{S} \mid \boldsymbol{\Lambda}) = \sum_{n=1}^{N} \sum_{t=1}^{T} -\lambda_{t,n} \, \Delta t + s_{t,n} \, \log \lambda_{t,n} \tag{6.5}$$

$$= \sum_{n=1}^{N} \sum_{t=1}^{T} -g(\psi_{t,n}) \, \Delta t + s_{t,n} \, \log g(\psi_{t,n}) \tag{6.6}$$

and the log likelihood of a population of non-interacting spike trains is simply the sum of each of the log likelihoods for each neuron. The likelihood depends upon the network through the definition of the activation given in Eq. 6.1 and Eq. 6.2.

Due to the potentially nonlinear and non-Gaussian nature of STDP, these existing techniques are not applicable here. Instead we use particle MCMC (Andrieu et al., 2010), a powerful technique for inference in time series. Particle MCMC samples the posterior distribution over weight trajectories, $\boldsymbol{W}[t]$, the adjacency matrix $\boldsymbol{A}$, and the model parameters $\boldsymbol{\theta}^{(n \to n')}$ and $\boldsymbol{\vartheta}$, given the observed spike trains, by combining particle filtering with MCMC. We represent the conditional distribution over weight trajectories with a set of discrete particles, $\{\boldsymbol{W}^{(p)}\}_{p=1}^{P}$. Each particle represents a sequence of weight matrices, $\boldsymbol{W}^{(p)} \in \mathbb{R}^{N \times N \times T}$, and has an associated nonnegative *particle weight* $v_T^{(p)}$. Note that the particle weights are *not* the same as the synaptic weights. Together, these define an atomic distribution over weight trajectories,

$$p(\boldsymbol{W}) \approx \frac{\sum_{p=1}^{P} v_T^{(p)} \, \delta_{\boldsymbol{W}^{(p)}}(\boldsymbol{W})}{\sum_{p=1}^{P} v_T^{(p)}}, \tag{6.7}$$

where $\delta_{w*}(w)$ is the Dirac delta function located at w^*.

Particle Filtering

Particle filtering (Andrieu et al., 2003) or *sequential Monte Carlo* is a method of inferring a distribution over weight trajectories by iteratively propagating forward in time and re-weighting according to how well the new samples explain the data. We build up the collection of weight trajectories iteratively, one bin at a time. We start by sampling the initial synaptic weights from the prior distribution,

$$\boldsymbol{W}^{(p)}[1] \sim p(\boldsymbol{W}[1] \mid \boldsymbol{\vartheta}),$$

computing their likelihoods, $\alpha_1^{(p)} = p(\boldsymbol{s}_1 \mid \boldsymbol{A}, \boldsymbol{W}^{(p)}[1], \{\boldsymbol{\theta}_{n \to n'}\}, \{\psi_n^{(0)}\})$, and initializing the particle weights to $v_1^{(p)} = \alpha_1^{(p)}$. Then, we iteratively proceed, updating the synaptic weights according to the learning rule and updating the particle weights according to the likelihood of the spikes.

That is, for $t = 2, \ldots, T$, we perform the following steps:

1. Sample the next synaptic weight given the weight in the preceding time bin, the learning rule, the spike history, and the global parameters,

$$\boldsymbol{W}^{(p)}[t] \sim p(\boldsymbol{W}[t] \mid \boldsymbol{W}^{(p)}[t-1], \boldsymbol{S}_{\leq t-1}, \boldsymbol{\vartheta}).$$

2. Compute the likelihood of the current spikes,

$$\alpha_t^{(p)} = p(\boldsymbol{s}_t \mid \boldsymbol{A}, \boldsymbol{W}^{(p)}[t], \{\boldsymbol{\theta}_{n \to n'}\}, \{\psi_n^{(0)}\}, \boldsymbol{S}_{\leq t})$$

3. Update the particle weight according to the likelihood of the current spikes,

$$v_t^{(p)} \leftarrow v_{t-1}^{(p)} \cdot \alpha_t^{(p)}.$$

This is, in fact, a special case of *sequential importance sampling* where our synaptic weights are sampled from the model's learning rule.

The problem with this simple algorithm is that the weights will rapidly decay to zero as particles drift from the regions of high likelihood. To counteract this effect, we often introduce a fourth step in which we *resample* the particles with replacement according to their weights.

4. Sample new particle indices with replacement according to the current weights,

$$p' \sim \text{Discrete}\left(\left[\frac{v_t^{(1)}}{\sum_p v_t^{(p)}}, \ldots, \frac{v_t^{(P)}}{\sum_p v_t^{(p)}}\right]\right),$$

   and then replace the weight trajectories with those of the new particle indices,

$$\boldsymbol{W}^{(p)}[1, \ldots, t] \leftarrow \boldsymbol{W}^{(p')}[1, \ldots, t].$$

5. Once we have resampled the particle indices, we can reset the weights.

$$v_t^{(p)} \leftarrow \frac{1}{P}.$$

This is called *sequential importance resampling*. At the end of T time steps we are left with a weighted set of synaptic weight trajectories that approximates the conditional distribution over synaptic weights for given global model parameters.

Collapsed Gibbs Sampling of $\boldsymbol{A}$ and $\boldsymbol{W}[t]$

The particle weights also provide an unbiased estimate of the marginal likelihood of entries in the adjacency matrix, $\boldsymbol{A}$, integrating out the corresponding weight trajectory. We have,

$$\begin{aligned} p(\boldsymbol{A} \mid \boldsymbol{S}, \{\boldsymbol{\theta}_{n \to n'}\}, \{\psi_n^{(0)}\}) &\propto \sum_{t=1}^{T} \int p(\boldsymbol{A}, \boldsymbol{W}[t] \mid \boldsymbol{S}, \{\boldsymbol{\theta}_{n \to n'}\}, \{\psi_n^{(0)}\}) \, \mathrm{d}\boldsymbol{W}[t] \\ &\approx \left[\prod_{t=1}^{T} \sum_{p=1}^{P} v_t^{(p)} \alpha_t^{(p)}\right] p(\boldsymbol{A} \mid \{\boldsymbol{z}_n\}, \boldsymbol{\vartheta}). \end{aligned}$$

We can leverage this estimator in a particle marginal Metropolis-Hastings (Andrieu et al., 2010) update. First, we propose an update to $\boldsymbol{A}$, then we run a particle filter to estimate the marginal likelihood of $\boldsymbol{A}$, and accept or reject the proposal accordingly. By marginalizing out the weight trajectory, we are able to explore the space of adjacency matrices more efficiently.

Factored Learning Rules If the learning rule factors into independent updates for each $w_{n' \to n}[t]$, we can update each synapse's weight trajectory separately and reduce the particles to one-dimensional trajectories. The STDP learning rules considered in this chapter all factor in this way.

Particle MCMC

Particle filtering only yields a distribution over weight trajectories and implicitly assumes that the other parameters have been specified. Particle MCMC provides a broader inference algorithm for both weights and static parameters. The idea is to interleave *conditional* particle filtering steps that sample the weight trajectory given the current model parameters and the previously sampled weights, with traditional Gibbs updates to sample the model parameters given the current weight trajectory. This combination leaves the stationary distribution of the Markov chain invariant and allows joint inference over weights and parameters.

In our implementation, we also make use of a pMCMC variant with ancestor sampling (Lindsten et al., 2012) that significantly improves convergence. Any distribution may be used to propagate the particles forward; using the learning rule is simply the easiest to implement and understand. We have omitted a number of details in this description; for a thorough overview of particle MCMC, the reader should consult (Andrieu et al., 2010; Lindsten et al., 2012).

Evaluation

We evaluated our technique with two types of synthetic data. First, we generated data from our model, with known ground-truth. Second, we used the well-known simulator NEURON to simulate driven, interconnected populations of neurons undergoing synaptic plasticity. For comparison, we show how the sparse, time-varying GLM compares to a standard

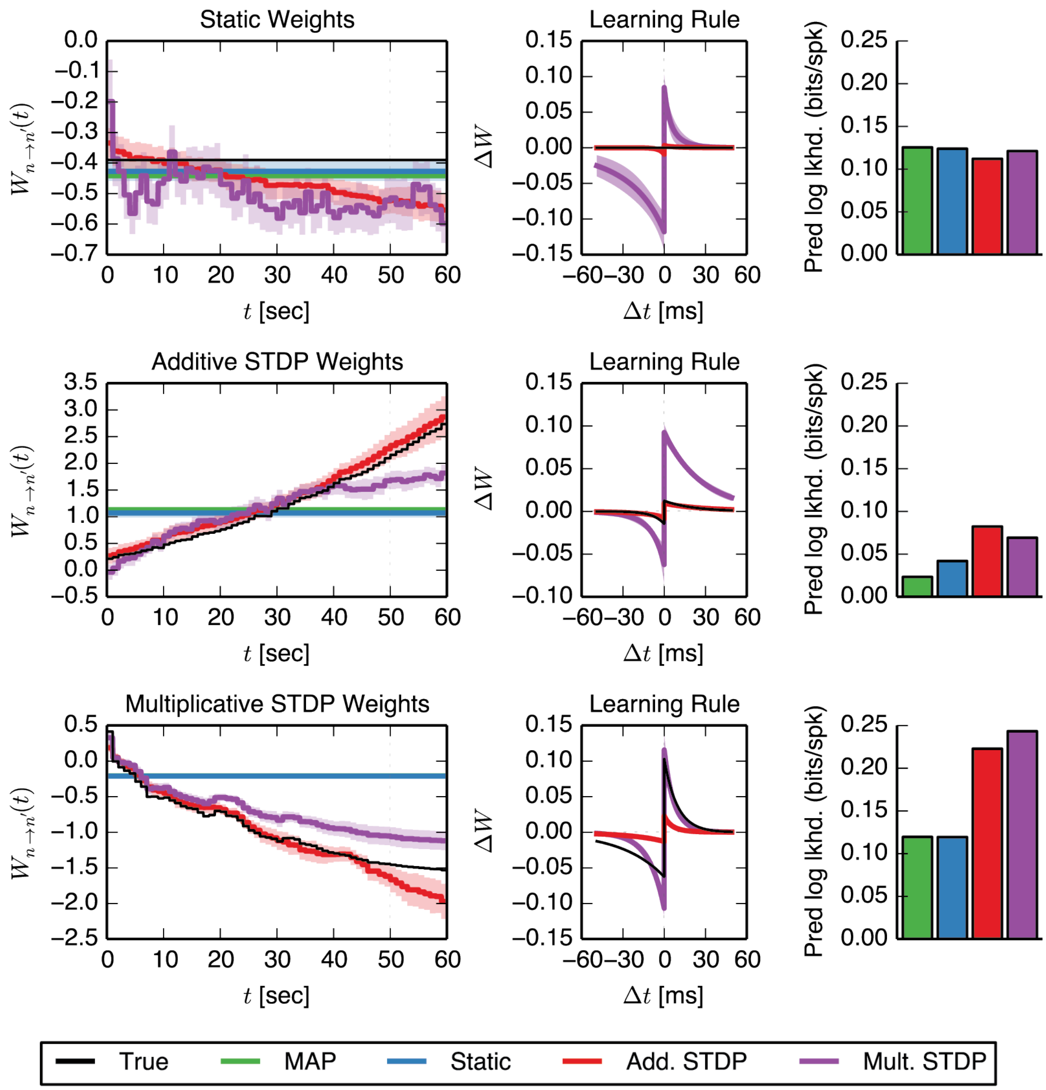

Figure 6.2: We fit time-varying weight trajectories to spike trains simulated from a GLM with two neurons undergoing no plasticity (top row), an additive, unbounded STDP rule (middle), and a multiplicative, saturating STDP rule (bottom row). We fit the first 50 seconds with four different models: MAP for an L1-regularized GLM, and fully-Bayesian inference for a static, additive STDP, and multiplicative STDP learning rules. In all cases, the correct models yield the highest predictive log likelihood on the final 10 seconds of the dataset.

GLM with a group LASSO prior on the impulse response coefficients for which we can perform efficient MAP estimation.

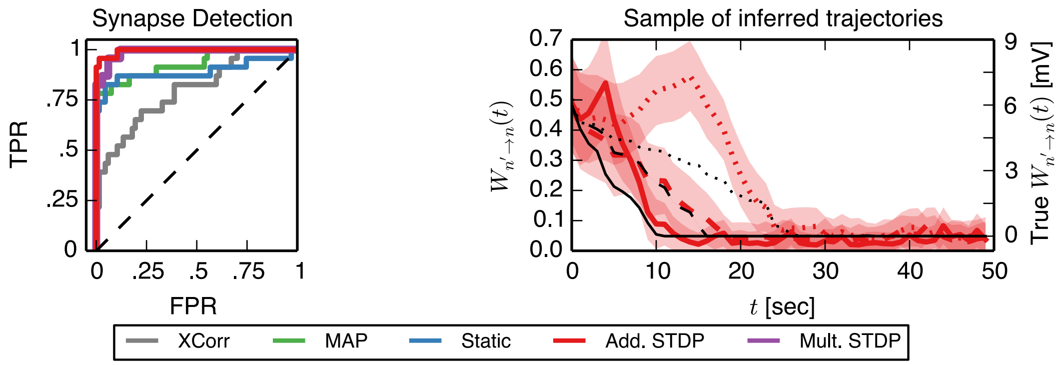

Figure 6.3: Evaluation of synapse detection on a 60 second spike train from a network of 10 neurons undergoing synaptic plasticity with a saturating, additive STDP rule, simulated with NEURON. The sparse, time-varying GLM with an additive rule outperforms the fully-Bayesian model with static weights, MAP estimation with L1 regularization, and simple thresholding of the cross-correlation matrix.

GLM-based Simulations

As a proof of concept, we study a single synapse undergoing a variety of synaptic plasticity rules and generating spikes according to a GLM. The neurons also have inhibitory self-connections to mimic refractory effects. We tested three synaptic plasticity mechanisms: a static synapse (i.e., no plasticity), the unbounded, additive STDP rule given by Equation 6.3, and the bounded, multiplicative STDP rule given by Equation 6.4. For each learning rule, we simulated 60 seconds of spiking activity at 1kHz temporal resolution, updating the synaptic weights every 1s. The baseline firing rates were normally distributed with mean 20Hz and standard deviation of 5Hz. Correlations in the spike timing led to changes in the synaptic weight trajectories that we could detect with our inference algorithm.

Figure 6.2 shows the true and inferred weight trajectories, the inferred learning rules, and the predictive log likelihood on ten seconds of held-out data for each of the three ground truth learning rules. When the underlying weights are static (top row), MAP estimation and static learning rules do an excellent job of detecting the true weight whereas the two time-varying models must compensate by either setting the learning rule as close to zero as possible, as the additive STDP does, or setting the threshold such that the weight trajectory is nearly constant, as the multiplicative model does. Note that the scales of the additive and multiplicative learning rules are not directly comparable since the weight updates in the multiplicative case are modulated by how close the weight is to the threshold. When the underlying weights vary (middle and bottom rows), the static models must compromise with an intermediate weight. Though the STDP models are both able to capture the qualitative trends, the correct model yields a better fit and better predictive power in both cases.

In terms of computational cost, our approach is clearly more expensive than alternative approaches based on MAP estimation or MLE. We developed a parallel implementation of our algorithm to capitalize on conditional independencies across neurons, i.e. for the additive and multiplicative STDP rules we can sample the weights $\boldsymbol{W}_{*\to n}$ independently of the weights $\boldsymbol{W}_{*\to n'}$. On the two neuron examples we achieve upward of 2 iterations per second (sampling all variables in the model), and we run our model for 1000 iterations. Convergence of the Markov chain is assessed by analyzing the log posterior of the samples, and typically stabilizes after a few hundred iterations. As we scale to networks of ten neurons, our running time quickly increases to roughly 20 seconds per iteration, which is mostly dominated by slice sampling the learning rule parameters. In order to evaluate the conditional probability of a learning rule parameter, we need to sample the weight trajectories for each synapse. Though these running times are nontrivial, they are not prohibitive for networks that are realistically obtainable for optical study of synaptic plasticity.

Biophysical simulations

Using the biophysical simulator NEURON, we performed two experiments. First, we considered a network of 10 sparsely interconnected neurons (28 excitatory synapses) undergoing synaptic plasticity according to an additive STDP rule. Each neuron was driven independently by a hidden population of 13 excitatory neurons and 5 inhibitory neurons connected to the visible neuron with probability 0.8 and fixed synaptic weights averaging 3.0 mV. The visible synapses were initialized close to 6.0 mV and allowed to vary between 0.0 and 10.5 mV. The synaptic delay was fixed at 1.0 ms for all synapses. This yielded a mean firing rate of 10 Hz among visible neurons. Synaptic weights were recorded every 1.0 ms. These parameters were chosen to demonstrate interesting variations in synaptic strength, and as we transition to biological applications it will be necessary to evaluate the sensitivity of the model to these parameters and the appropriate regimes for the circuits under study.

We began by investigating whether the model is able to accurately identify synapses from spikes, or whether it is confounded by spurious correlations. Figure 6.3 shows that our approach identifies the 28 excitatory synapses in our network, as measured by ROC curve (Add. STDP AUC=0.99), and outperforms static models and cross-correlation. In the sparse, time-varying GLM, the probability of an edge is measured by the mean of $\boldsymbol{A}$ under the posterior, whereas in the standard GLM with MAP estimation, the likelihood of an edge is measured by area under the impulse response.

Looking into the synapses that are detected by the time-varying model and missed by the static model, we find an interesting pattern. The improved performance comes from synapses that decay in strength over the recording period. Three examples of these synaptic weight trajectories are shown in the right panel of Figure 6.3. The time-varying model assigns over 90% probability to each of the three synapses, whereas the static model infers less than a 40% probability for each synapse.

Finally, we investigated our model's ability to distinguish various learning rules by looking at a single synapse, analogous to the experiment performed on data from the GLM. Figure 6.4 shows the results of a weight trajectory for a synapse under additive STDP with a strict threshold on the excitatory post-synaptic current. The time-varying GLM with an additive model captures the same trajectory, as shown in the left panel. The GLM weights have been linearly rescaled to align with the true weights, which are measured in millivolts. Furthermore, the inferred additive STDP learning rule, in particular the time constants and relative amplitudes, perfectly match the true learning rule.

These results demonstrate that a sparse, time-varying GLM is capable of discovering synaptic weight trajectories, but in terms of predictive likelihood, we still have insufficient evidence to distinguish additive and multiplicative STDP rules. By the end of the training period, the weights have saturated at a level that almost surely induces post-synaptic spikes. At this point, we cannot distinguish two learning rules which have both reached saturation. This motivates further studies that leverage this probabilistic model in an optimal experimental design framework, similar to recent work by Shababo et al. (2013), in order to conclusively test hypotheses about synaptic plasticity.

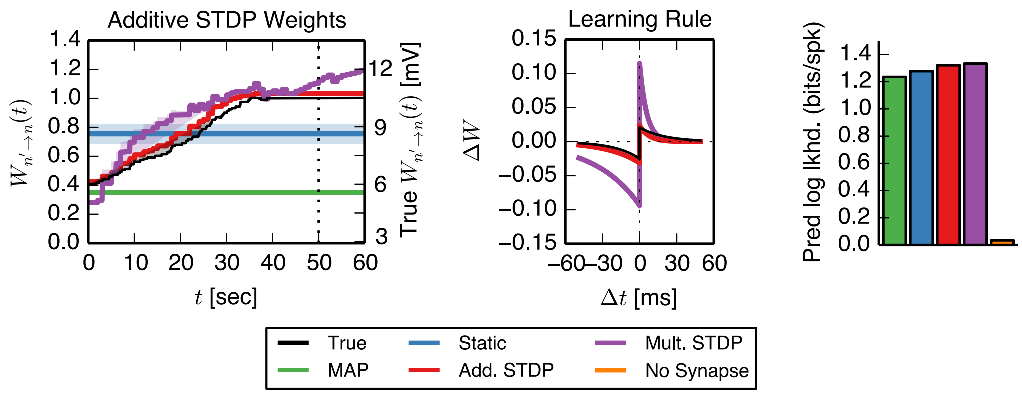

Figure 6.4: Analogously to Figure 6.2, a sparse, time-varying GLM can capture the weight trajectories and learning rules from spike trains simulated by NEURON. Here an excitatory synapse undergoes additive STDP with a hard upper bound on the excitatory post-synaptic current. The weight trajectory inferred by our model with the same parametric form of the learning rule matches almost exactly, whereas the static models must compromise in order to capture early and late stages of the data, and the multiplicative weight exhibits qualitatively different trajectories. Nevertheless, in terms of predictive log likelihood, we do not have enough information to correctly determine the underlying learning rule. Potential solutions are discussed in the main text.

Discussion

Motivated by the advent of optical tools for interrogating networks of synaptically connected neurons, which make it possible to study synaptic plasticity in novel ways, we have extended the GLM to model a sparse, time-varying synaptic network, and introduced a fully-Bayesian inference algorithm built upon particle MCMC. Our initial results suggest that it is possible to infer weight trajectories for a variety of biologically plausible learning rules.

A number of interesting questions remain as we look to apply these methods to biological recordings. We have assumed access to precise spike times, though extracting spike times from optical recordings poses inferential challenges of its own. Solutions like those of Vogelstein et al. (2009) could be incorporated into our probabilistic model. Computationally, particle MCMC could be replaced with stochastic EM to achieve improved performance (Lindsten et al., 2012), and optimal experimental design could aid in the exploration of stimuli to distinguish between learning rules. Beyond these direct extensions, this work opens up potential to infer latent state spaces with potentially nonlinear dynamics and non-Gaussian noise, and to infer learning rules at the synaptic or even the network level.

References

Yashar Ahmadian, Jonathan W Pillow, and Liam Paninski. Efficient Markov chain Monte Carlo methods for decoding neural spike trains. *Neural Computation*, 23(1):46–96, 2011.

Misha B Ahrens, Michael B Orger, Drew N Robson, Jennifer M Li, and Philipp J Keller. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature Methods*, 10(5):413–420, 2013.

Laurence Aitchison and Peter E Latham. Synaptic sampling: A connection between PSP variability and uncertainty explains neurophysiological observations. *arXiv preprint arXiv:1505.04544*, 2015.

Laurence Aitchison and Máté Lengyel. The Hamiltonian brain. *arXiv preprint arXiv:1407.0973*, 2014.

David J Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598, 1981.

Charles H Anderson and David C Van Essen. Neurobiological computational systems. *Computational Intelligence Imitating Life*, pages 1–11, 1994.

Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

Michael J Barber, John W Clark, and Charles H Anderson. Neural representation of probabilistic information. *Neural Computation*, 15(8):1843–64, August 2003.

Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, 1966.

Matthew J. Beal, Zoubin Ghahramani, and Carl E. Rasmussen. The infinite hidden Markov model. *Advances in Neural Information Processing Systems 14*, pages 577–585, 2002.

Jeffrey M Beck and Alexandre Pouget. Exact inferences in a neural implementation of a hidden Markov model. *Neural Computation*, 19(5):1344–1361, 2007.

Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. Marginalization in neural circuits with divisive normalization. *The Journal of Neuroscience*, 31(43):15310–15319, 2011.

Jeffrey M Beck, Katherine A Heller, and Alexandre Pouget. Complex inference in neural circuits with probabilistic population codes and topic models. *Advances in Neural Information Processing Systems*, pages 3059–3067, 2012.

Yoshua Bengio and Paolo Frasconi. An input output HMM architecture. *Advances in Neural Information Processing Systems*, pages 427–434, 1995.

Pietro Berkes, Gergo Orbán, Máté Lengyel, and József Fiser. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013): 83–7, January 2011.

Gordon J Berman, Daniel M Choi, William Bialek, and Joshua W Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672, 2014.

Philippe Biane, Jim Pitman, and Marc Yor. Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions. *Bulletin of the American Mathematical Society*, 38(4):435–465, 2001.

Christopher M Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

David M Blei. Build, compute, critique, repeat: Data analysis with latent variable models. *Annual Review of Statistics and Its Application*, 1:203–232, 2014.

David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.

Carolyn R Block and Richard Block. *Street gang crime in Chicago*. US Department of Justice, Office of Justice Programs, National Institute of Justice, 1993.

Carolyn R Block, Richard Block, and Illinois Criminal Justice Information Authority. Homicides in Chicago, 1965-1995. ICPSR06399-v5. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], July 2005.

Charles Blundell, Katherine A Heller, and Jeffrey M Beck. Modelling reciprocating relationships with Hawkes processes. *Advances in Neural Information Processing Systems*, pages 2600–2608, 2012.

George EP Box. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)*, pages 383–430, 1980.

David H Brainard and William T Freeman. Bayesian color constancy. *Journal of the Optical Society of America A*, 14(7):1393–1411, 1997.

Kevin L Briggman, Henry DI Abarbanel, and William B Kristan. Optical imaging of neuronal populations during decision-making. *Science*, 307(5711):896–901, 2005.

David R. Brillinger. Maximum likelihood analysis of spike trains of interacting nerve cells. *Biological Cybernetics*, 59(3):189–200, August 1988.

David R Brillinger, Hugh L Bryant Jr, and Jose P Segundo. Identification of synaptic interactions. *Biological Cybernetics*, 22(4):213–228, 1976.

Michael Bryant and Erik B Sudderth. Truly nonparametric online variational inference for hierarchical Dirichlet processes. *Advances in Neural Information Processing Systems 25*, pages 2699–2707, 2012.

Lars Buesing, Johannes Bill, Bernhard Nessler, and Wolfgang Maass. Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons. *PLoS Computational Biology*, 7(11):e1002211, November 2011.

Lars Buesing, Jakob H. Macke, and Maneesh Sahani. Learning stable, regularised latent models of neural population dynamics. *Network: Computation in Neural Systems*, 23: 24–47, 2012a.

Lars Buesing, Jakob H Macke, and Maneesh Sahani. Spectral learning of linear dynamics from generalised-linear observations with application to neural population data. *Advances in Neural Information Processing Systems*, pages 1682–1690, 2012b.

Lars Buesing, Timothy A Machado, John P Cunningham, and Liam Paninski. Clustered factor analysis of multineuronal spike data. *Advances in Neural Information Processing Systems*, pages 3500–3508, 2014.

Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

Santiago Ramón Cajal. *Textura del Sistema Nervioso del Hombre y los Vertebrados*, volume 1. Imprenta y Librería de Nicolás Moya, Madrid, Spain, 1899.

Natalia Caporale and Yang Dan. Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46, 2008.

Nick Chater and Christopher D Manning. Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, 10(7):335–344, 2006.

Zhe Chen, Fabian Kloosterman, Emery N Brown, and Matthew A Wilson. Uncovering spatial topology represented by rat hippocampal population neuronal codes. *Journal of Computational Neuroscience*, 33(2):227–255, 2012.

Zhe Chen, Stephen N Gomperts, Jun Yamamoto, and Matthew A Wilson. Neural representation of spatial topology in the rodent hippocampus. *Neural Computation*, 26(1): 1–39, 2014.

Sharat Chikkerur, Thomas Serre, Cheston Tan, and Tomaso Poggio. What and where: A Bayesian inference theory of attention. *Vision Research*, 50(22):2233–2247, 2010.

Yoon Sik Cho, Aram Galstyan, Jeff Brantingham, and George Tita. Latent point process models for spatial-temporal networks. *arXiv:1302.2671*, 2013.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945, 2004.

Aaron C Courville, Nathaniel D Daw, and David S Touretzky. Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, 10(7):294–300, 2006.

Ronald L Cowan and Charles J Wilson. Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *Journal of Neurophysiology*, 71(1):17–32, 1994.

W Maxwell Cowan, Thomas C Südhof, and Charles F Stevens. *Synapses*. Johns Hopkins University Press, 2003.

Mary Kathryn Cowles and Bradley P Carlin. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91: 883–904, 1996.

John P Cunningham and Byron M Yu. Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509, 2014.

Paul Dagum and Michael Luby. Approximating probabilistic inference in Bayesian belief networks is NP-hard. *Artificial Intelligence*, 60(1):141–153, 1993.

Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: Volume I: Elementary Theory and Methods*. Springer Science & Business Media, 2 edition, 2003.

Peter Dayan and Larry F Abbott. *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press, 2001.

Peter Dayan and Joshua A Solomon. Selective Bayes: Attentional load and crowding. *Vision Research*, 50(22):2248–2260, 2010.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

Sophie Deneve. Bayesian spiking neurons I: inference. *Neural Computation*, 20(1):91–117, January 2008.

Luc Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, USA, 1986.

Christopher DuBois, Carter Butts, and Padhraic Smyth. Stochastic block modeling of relational event dynamics. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pages 238–246, 2013.

Seif Eldawlatly, Yang Zhou, Rong Jin, and Karim G Oweiss. On the use of dynamic Bayesian networks in reconstructing functional neuronal networks from spike train ensembles. *Neural Computation*, 22(1):158–189, 2010.

Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.

Sean Escola, Alfredo Fontanini, Don Katz, and Liam Paninski. Hidden Markov models for the stimulus-response relationships of multistate neural systems. *Neural Computation*, 23(5):1071–1132, 2011.

Warren John Ewens. Population genetics theory—the past and the future. In S. Lessard, editor, *Mathematical and Statistical Developments of Evolutionary Theory*, pages 177–227. Springer, 1990.

Daniel E Feldman. The spike-timing dependence of plasticity. *Neuron*, 75(4):556–71, August 2012.

Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47, 1991.

Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973.

Christopher R Fetsch, Amanda H Turner, Gregory C DeAngelis, and Dora E Angelaki. Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience*, 29(49):15601–15612, 2009.

Christopher R Fetsch, Alexandre Pouget, Gregory C DeAngelis, and Dora E Angelaki. Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, 15(1):146–154, 2012.

József Fiser, Pietro Berkes, Gergő Orbán, and Máté Lengyel. Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, 14(3):119–130, 2010.

Alyson K Fletcher, Sundeep Rangan, Lav R Varshney, and Aniruddha Bhargava. Neural reconstruction with approximate message passing (neuramp). *Advances in Neural Information Processing Systems*, pages 2555–2563, 2011.

Emily B Fox. *Bayesian nonparametric learning of complex dynamical phenomena*. PhD thesis, Massachusetts Institute of Technology, 2009.

Emily B Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. An HDP-HMM for systems with state persistence. *Proceedings of the International Conference on Machine Learning*, pages 312–319, 2008.

Jeremy Freeman, Greg D Field, Peter H Li, Martin Greschner, Deborah E Gunning, Keith Mathieson, Alexander Sher, Alan M Litke, Liam Paninski, Eero P Simoncelli, et al. Mapping nonlinear receptive field structure in primate retina at single cone resolution. *eLife*, 4:e05241, 2015.

Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, 11(2):127–38, February 2010.

Karl J Friston. Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping*, 2(1-2):56–78, 1994.

Deep Ganguli and Eero P Simoncelli. Implicit encoding of prior probabilities in optimal neural populations. *Advances in Neural Information Processing Systems*, pages 6–9, 2010.

Peiran Gao and Surya Ganguli. On simplicity and complexity in the brave new world of large-scale neuroscience. *Current Opinion in Neurobiology*, 32:148–155, 2015.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian Data Analysis*. CRC press, 3rd edition, 2013.

Stuart Geman and Donald Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741, 1984.

Felipe Gerhard, Tilman Kispersky, Gabrielle J Gutierrez, Eve Marder, Mark Kramer, and Uri Eden. Successful reconstruction of a physiological circuit with known connectivity from spiking activity alone. *PLoS Computational Biology*, 9(7):e1003138, 2013.

Samuel J Gershman, Matthew D Hoffman, and David M Blei. Nonparametric variational inference. *Proceedings of the International Conference on Machine Learning*, pages 663–670, 2012a.

Samuel J Gershman, Edward Vul, and Joshua B Tenenbaum. Multistability and perceptual inference. *Neural Computation*, 24(1):1–24, 2012b.

Sebastian Gerwinn, Jakob Macke, Matthias Seeger, and Matthias Bethge. Bayesian inference for spiking neuron models with a sparsity prior. *Advances in Neural Information Processing Systems*, pages 529–536, 2008.

Charles J Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, pages 473–483, 1992.

Walter R Gilks. *Markov Chain Monte Carlo*. Wiley Online Library, 2005.

Anna Goldenberg, Alice X Zheng, Stephen E Fienberg, and Edoardo M Airoldi. A survey of statistical network models. *Foundations and Trends in Machine Learning*, 2(2):129–233, 2010.

Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1019–1028, 2010.

Noah Goodman, Vikash Mansinghka, Daniel M Roy, Keith Bonawitz, and Joshua B Tenenbaum. Church: a language for generative models. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 220––229, 2008.

Noah D Goodman, Joshua B Tenenbaum, and Tobias Gerstenberg. Concepts in a probabilistic language of thought. Technical report, Center for Brains, Minds and Machines (CBMM), 2014.

Agnieszka Grabska-Barwinska, Jeff Beck, Alexandre Pouget, and Peter Latham. Demixing odors-fast inference in olfaction. *Advances in Neural Information Processing Systems*, pages 1968–1976, 2013.

SG Gregory, KF Barlow, KE McLay, R Kaul, D Swarbreck, A Dunham, CE Scott, KL Howe, K Woodfine, CCA Spencer, et al. The DNA sequence and biological annotation of human chromosome 1. *Nature*, 441(7091):315–321, 2006.

Thomas L Griffiths, Charles Kemp, and Joshua B Tenenbaum. Bayesian models of cognition. In Ron Sun, editor, *The Cambridge Handbook of Computational Psychology*. Cambridge University Press, 2008.

Roger B Grosse, Chris J Maddison, and Ruslan R Salakhutdinov. Annealing between distributions by averaging moments. *Advances in Neural Information Processing Systems*, pages 2769–2777, 2013.

Roger B Grosse, Zoubin Ghahramani, and Ryan P Adams. Sandwiching the marginal likelihood using bidirectional Monte Carlo. *arXiv preprint arXiv:1511.02543*, 2015.

Yong Gu, Dora E Angelaki, and Gregory C DeAngelis. Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11(10):1201–1210, 2008.

Fangjian Guo, Charles Blundell, Hanna Wallach, and Katherine A Heller. The Bayesian echo chamber: Modeling influence in conversations. *arXiv preprint arXiv:1411.2674*, 2014.

Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83, 1971.

Moritz Helmstaedter, Kevin L Briggman, Srinivas C Turaga, Viren Jain, H Sebastian Seung, and Winfried Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461):168–174, 2013.

Geoffrey E Hinton. How neural networks learn from experience. *Scientific American*, 1992.

Geoffrey E Hinton and Terrence J Sejnowski. Optimal perceptual inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1983.

Daniel R Hochbaum, Yongxin Zhao, Samouil L Farhi, Nathan Klapoetke, Christopher A Werley, Vikrant Kapoor, Peng Zou, Joel M Kralj, Dougal Maclaurin, Niklas Smedemark-Margulies, et al. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nature Methods*, 2014.

Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117 (4):500, 1952.

Peter D Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in Neural Information Processing Systems*, 20:1–8, 2008.

Matthew D Hoffman, David M Blei, Chong Wang, and John Paisley. Stochastic variational inference. *The Journal of Machine Learning Research*, 14(1):1303–1347, 2013.

Douglas N. Hoover. Relations on probability spaces and arrays of random variables. Technical report, Institute for Advanced Study, Princeton, 1979.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

Patrik O Hoyer and Aapo Hyvarinen. Interpreting neural response variability as Monte Carlo sampling of the posterior. *Advances in neural information processing systems*, pages 293–300, 2003.

Yanping Huang and Rajesh P. N. Rao. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):580–593, September 2011.

David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962.

Hemant Ishwaran and Mahmoud Zarepour. Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283, 2002.

Tomoharu Iwata, Amar Shah, and Zoubin Ghahramani. Discovering latent influence in online social activities via shared cascade Poisson processes. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 266–274, 2013.

Mehrdad Jazayeri and Michael N Shadlen. Temporal context calibrates interval timing. *Nature Neuroscience*, 13(8):1020–1026, 2010.

Mehrdad Jazayeri and Michael N Shadlen. A neural mechanism for sensing and reproducing a time interval. *Current Biology*, 25(20):2599–2609, 2015.

Matthew J Johnson. *Bayesian time series models and scalable inference*. PhD thesis, Massachusetts Institute of Technology, June 2014.

Matthew J Johnson and Alan S Willsky. Bayesian nonparametric hidden semi-Markov models. *Journal of Machine Learning Research*, 14(1):673–701, 2013.

Matthew J Johnson and Alan S Willsky. Stochastic variational inference for Bayesian time series models. *Proceedings of the International Conference on Machine Learning*, 32:1854–1862, 2014.

Matthew J Johnson, Scott W Linderman, Sandeep R Datta, and Ryan P Adams. Discovering switching autoregressive dynamics in neural spike train recordings. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2015.

Lauren M Jones, Alfredo Fontanini, Brian F Sadacca, Paul Miller, and Donald B Katz. Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences*, 104(47):18772–18777, 2007.

Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999.

Eric R Kandel, James H Schwartz, Thomas M Jessell, et al. *Principles of neural science*, volume 4. McGraw-Hill New York, 2000.

David Kappel, Stefan Habenschuss, Robert Legenstein, and Wolfgang Maass. Network plasticity as Bayesian inference. *PLoS Computational Biology*, 11(11):e1004485, 2015a.

David Kappel, Stefan Habenschuss, Robert Legenstein, and Wolfgang Maass. Synaptic sampling: A Bayesian approach to neural network plasticity and rewiring. *Advances in Neural Information Processing Systems*, pages 370–378, 2015b.

Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

Jason ND Kerr and Winfried Denk. Imaging in vivo: watching the brain in action. *Nature Reviews Neuroscience*, 9(3):195–205, 2008.

Roozbeh Kiani and Michael N Shadlen. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, 324(5928):759–64, May 2009.

John F. C. Kingman. *Poisson Processes (Oxford Studies in Probability).* Oxford University Press, January 1993. ISBN 0198536933.

David C Knill and Whitman Richards. *Perception as Bayesian inference*. Cambridge University Press, 1996.

Konrad P Körding and Daniel M Wolpert. Bayesian integration in sensorimotor learning. *Nature*, 427(6971):244–7, January 2004.

Alp Kucukelbir, Rajesh Ranganath, Andrew Gelman, and David Blei. Automatic variational inference in Stan. *Advances in Neural Information Processing Systems*, pages 568–576, 2015.

Stephen W Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1):37–68, 1953.

Harold W Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

Kenneth W Latimer, Jacob L Yates, Miriam LR Meister, Alexander C Huk, and Jonathan W Pillow. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*, 349(6244):184–187, 2015.

Tai Sing Lee and David Mumford. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20(7):1434–1448, 2003.

Robert Legenstein and Wolfgang Maass. Ensembles of spiking neurons with noise support optimal probabilistic inference in a dynamically changing environment. *PLoS Computational Biology*, 10(10):e1003859, 2014.

William C Lemon, Stefan R Pulver, Burkhard Höckendorf, Katie McDole, Kristin Branson, Jeremy Freeman, and Philipp J Keller. Whole-central nervous system functional imaging in larval Drosophila. *Nature Communications*, 6, 2015.

Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.

Percy Liang, Slav Petrov, Michael I Jordan, and Dan Klein. The infinite PCFG using hierarchical Dirichlet processes. *Proceedings of Empirical Methods in Natural Language Processing*, pages 688–697, 2007.

David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

Jeff W Lichtman, Jean Livet, and Joshua R Sanes. A technicolour approach to the connectome. *Nature Reviews Neuroscience*, 9(6):417–422, 2008.

Scott W Linderman and Ryan P. Adams. Discovering latent network structure in point process data. *Proceedings of the International Conference on Machine Learning*, pages 1413–1421, 2014.

Scott W Linderman and Ryan P Adams. Scalable Bayesian inference for excitatory point process networks. *arXiv preprint arXiv:1507.03228*, 2015.

Scott W Linderman and Ryan P Johnson, Matthew Jand Adams. Dependent multinomial models made easy: Stick-breaking with the Pólya-gamma augmentation. *Advances in Neural Information Processing Systems*, pages 3438–3446, 2015.

Scott W Linderman, Christopher H Stock, and Ryan P Adams. A framework for studying synaptic plasticity with neural spike train data. *Advances in Neural Information Processing Systems*, pages 2330–2338, 2014.

Scott W Linderman, Ryan P Adams, and Jonathan W Pillow. Inferring structured connectivity from spike trains under negative-binomial generalized linear models. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2015.

Scott W Linderman, Matthew J Johnson, Matthew W Wilson, and Zhe Chen. A nonparametric Bayesian approach to uncovering rat hippocampal population codes during spatial navigation. *Journal of Neuroscience Methods*, 263:36–47, 2016a.

Scott W Linderman, Aaron Tucker, and Matthew J Johnson. Bayesian latent state space models of neural activity. *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2016b.

Fredrik Lindsten, Michael I Jordan, and Thomas B Schön. Ancestor sampling for particle Gibbs. *Advances in Neural Information Processing Systems*, pages 2600–2608, 2012.

Shai Litvak and Shimon Ullman. Cortical circuitry implementing graphical models. *Neural Computation*, 21(11):3010–3056, 2009.

James Robert Lloyd, Peter Orbanz, Zoubin Ghahramani, and Daniel M Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. *Advances in Neural Information Processing Systems*, 2012.

Wei Ji Ma and Mehrdad Jazayeri. Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, 37:205–220, 2014.

Wei Ji Ma, Jeffrey M Beck, Peter E Latham, and Alexandre Pouget. Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11):1432–8, November 2006.

David JC MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992.

Jakob H Macke, Lars Buesing, John P Cunningham, M Yu Byron, Krishna V Shenoy, and Maneesh Sahani. Empirical models of spiking in neural populations. *Advances in neural information processing systems*, pages 1350–1358, 2011.

Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

Vikash Mansinghka, Daniel Selsam, and Yura Perov. Venture: a higher-order probabilistic programming platform with programmable inference. *arXiv preprint arXiv:1404.0099*, 2014.

David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press, 1982.

Paul Miller and Donald B Katz. Stochastic transitions between neural states in taste processing and decision-making. *The Journal of Neuroscience*, 30(7):2559–2570, 2010.

T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023—-1032, 1988.

Shakir Mohamed, Zoubin Ghahramani, and Katherine A Heller. Bayesian and L1 approaches for sparse unsupervised learning. *Proceedings of the International Conference on Machine Learning*, pages 751–758, 2012.

Jesper Møller, Anne Randi Syversveen, and Rasmus Plenge Waagepetersen. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482, 1998.

Michael L Morgan, Gregory C DeAngelis, and Dora E Angelaki. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, 59(4):662–673, 2008.

Abigail Morrison, Markus Diesmann, and Wulfram Gerstner. Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, 98(6):459–478, 2008.

Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

Radford M Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.

Radford M. Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, pages 113–162, 2010.

John A Nelder and R Jacob Baker. Generalized linear models. *Encyclopedia of Statistical Sciences*, 1972.

Bernhard Nessler, Michael Pfeiffer, Lars Buesing, and Wolfgang Maass. Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity. *PLoS Computational Biology*, 9(4):e1003037, 2013.

Mark EJ Newman. The structure and function of complex networks. *Society for Industrial and Applied Mathematics (SIAM) Review*, 45(2):167–256, 2003.

Krzysztof Nowicki and Tom A B Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

Seung Wook Oh, Julie A Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M Henry, et al. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, 2014.

Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982.

John O'Keefe and Lynn Nadel. *The Hippocampus as a Cognitive Map*, volume 3. Clarendon Press, 1978.

Peter Orbanz and Daniel M Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):437–461, 2015.

Peter Orbanz and Yee Whye Teh. Bayesian nonparametric models. In *Encyclopedia of Machine Learning*, pages 81–89. Springer, 2011.

Adam M Packer, Darcy S Peterka, Jan J Hirtz, Rohit Prakash, Karl Deisseroth, and Rafael Yuste. Two-photon optogenetics of dendritic spines and neural circuits. *Nature Methods*, 9(12):1202–1205, 2012.

Liam Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243–262, January 2004.

Liam Paninski, Yashar Ahmadian, Daniel Gil Ferreira, Shinsuke Koyama, Kamiar Rahnama Rad, Michael Vidne, Joshua Vogelstein, and Wei Wu. A new look at state-space models for neural data. *Journal of Computational Neuroscience*, 29(1-2):107–126, 2010.

Andrew V Papachristos. Murder by structure: Dominance relations and the social structure of gang homicide. *American Journal of Sociology*, 115(1):74–128, 2009.

Il Memming Park and Jonathan W Pillow. Bayesian spike-triggered covariance analysis. *Advances in Neural Information Processing Systems*, pages 1692–1700, 2011.

Patrick O Perry and Patrick J Wolfe. Point process modelling for directed interaction networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2013.

Biljana Petreska, Byron Yu, John P Cunningham, Gopal Santhanam, Stephen I Ryu, Krishna V Shenoy, and Maneesh Sahani. Dynamical segmentation of single trials from population neural data. *Advances in Neural Information Processing Systems*, pages 756–764, 2011.

David Pfau, Eftychios A Pnevmatikakis, and Liam Paninski. Robust learning of low-dimensional dynamics from large neural ensembles. *Advances in Neural Information Processing Systems*, pages 2391–2399, 2013.

Jonathan W. Pillow and James Scott. Fully Bayesian inference for neural models with negative-binomial spiking. *Advances in Neural Information Processing Systems*, pages 1898–1906, 2012.

Jonathan W Pillow, Jonathon Shlens, Liam Paninski, Alexander Sher, Alan M Litke, EJ Chichilnisky, and Eero P Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999, 2008.

Eftychios A Pnevmatikakis, Daniel Soudry, Yuanjun Gao, Timothy A Machado, Josh Merel, David Pfau, Thomas Reardon, Yu Mu, Clay Lacefield, Weijian Yang, et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron*, 2016.

Nicholas G Polson, James G Scott, and Jesse Windle. Bayesian inference for logistic models using Pólya-gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013.

Ruben Portugues, Claudia E Feierstein, Florian Engert, and Michael B Orger. Whole-brain activity maps reveal stereotyped, distributed networks for visuomotor behavior. *Neuron*, 81(6):1328–1343, 2014.

Alexandre Pouget, Jeffrey M Beck, Wei Ji Ma, and Peter E Latham. Probabilistic brains: knowns and unknowns. *Nature Neuroscience*, 16(9):1170–1178, 2013.

Robert Prevedel, Young-Gyu Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrödel, Ramesh Raskar, Manuel Zimmer, Edward S Boyden, et al. Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy. *Nature Methods*, 11(7):727–730, 2014.

Lawrence R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Adrian E Raftery and Steven Lewis. How many iterations in the Gibbs sampler? *Bayesian Statistics*, pages 763–773, 1992.

Rajesh Ranganath, Sean Gerrish, and David M Blei. Black box variational inference. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 33:275––283, 2014.

Rajesh P. N. Rao. Bayesian computation in recurrent neural circuits. *Neural Computation*, 16(1):1–38, January 2004.

Rajesh P. N. Rao. Neural models of Bayesian belief propagation. In *Bayesian brain: Probabilistic approaches to neural computation*, pages 236–264. MIT Press Cambridge, MA, 2007.

Rajesh P. N. Rao and Dana H Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1): 79–87, January 1999.

Danilo J Rezende, Daan Wierstra, and Wulfram Gerstner. Variational learning for recurrent spiking networks. *Advances in Neural Information Processing Systems*, pages 136–144, 2011.

Fred Rieke, David Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes: exploring the neural code*. MIT press, 1999.

Christian Robert and George Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.

Dan Roth. On the hardness of approximate reasoning. *Artificial Intelligence*, 82(1):273–302, 1996.

Maneesh Sahani. *Latent variable models for neural data analysis*. PhD thesis, California Institute of Technology, 1999.

Maneesh Sahani and Peter Dayan. Doubly distributional population codes: simultaneous representation of uncertainty and multiplicity. *Neural Computation*, 2279:2255–2279, 2003.

Joshua R Sanes and Richard H Masland. The types of retinal ganglion cells: current status and implications for neuronal classification. *Annual Review of Neuroscience*, 38:221–246, 2015.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4: 639–650, 1994.

Ben Shababo, Brooks Paige, Ari Pakman, and Liam Paninski. Bayesian inference and online experimental design for mapping neural microcircuits. *Advances in Neural Information Processing Systems*, pages 1304–1312, 2013.

Vahid Shalchyan and Dario Farina. A non-parametric Bayesian approach for clustering and tracking non-stationarities of neural spikes. *Journal of Neuroscience Methods*, 223: 85–91, 2014.

Lei Shi and Thomas L Griffiths. Neural implementation of hierarchical Bayesian inference by importance sampling. *Advances in Neural Information Processing Systems*, 2009.

Yousheng Shu, Andrea Hasenstaub, and David A McCormick. Turning on and off recurrent balanced cortical activity. *Nature*, 423(6937):288–293, 2003.

Jack W Silverstein. The spectral radii and norms of large dimensional non-central random matrices. *Stochastic Models*, 10(3):525–532, 1994.

Aleksandr Simma and Michael I Jordan. Modeling events with cascades of Poisson processes. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, 2010.

Eero P Simoncelli. Optimal estimation in sensory systems. *The Cognitive Neurosciences, IV*, 2009.

Anne C Smith and Emery N Brown. Estimating a state-space model from point process observations. *Neural Computation*, 15(5):965–91, May 2003.

Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012.

Sen Song, Kenneth D Miller, and Lawerence F Abbott. Competitive Hebbian learning through spike-timing-dependent synaptic plasticitye. *Nature Neuroscience*, 3(9):919–26, September 2000. ISSN 1097-6256.

Daniel Soudry, Suraj Keshri, Patrick Stinson, Min-hwan Oh, Garud Iyengar, and Liam Paninski. Efficient "shotgun" inference of neural connectivity from highly sub-sampled activity data. *PLoS Computational Biology*, 11(10):1–30, 10 2015. doi: 10.1371/journal.pcbi.1004464.

Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: a structural description of the human brain. *PLoS Computational Biology*, 1(4):e42, 2005.

Olav Stetter, Demian Battaglia, Jordi Soriano, and Theo Geisel. Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. *PLoS Computational Biology*, 8(8):e1002653, 2012.

Ian Stevenson and Konrad Koerding. Inferring spike-timing-dependent plasticity from spike train data. *Advances in Neural Information Processing Systems*, pages 2582–2590, 2011.

Ian H Stevenson, James M Rebesco, Nicholas G Hatsopoulos, Zach Haga, Lee E Miller, and Konrad P Körding. Bayesian inference of functional connectivity and network structure from spikes. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 17(3):203–213, 2009.

Alan A Stocker and Eero P Simoncelli. Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 9(4):578–85, April 2006.

Yee Whye Teh and Michael I Jordan. Hierarchical Bayesian nonparametric models with applications. *Bayesian Nonparametrics*, pages 158–207, 2010.

Yee Whye Teh, Michael I Jordan, Matthew J Beal, and David M Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

Joshua B Tenenbaum, Thomas L Griffiths, and Charles Kemp. Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10(7):309–318, 2006.

Joshua B Tenenbaum, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285, 2011.

Luke Tierney and Joseph B Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986.

Wilson Truccolo, Uri T. Eden, Matthew R. Fellows, John P. Donoghue, and Emery N. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology*, 93(2):1074–1089, 2005. doi: 10.1152/jn.00697.2004.

Philip Tully, Matthias Hennig, and Anders Lansner. Synaptic and nonsynaptic plasticity approximating probabilistic inference. *Frontiers in Synaptic Neuroscience*, 6(8), 2014.

Srini Turaga, Lars Buesing, Adam M Packer, Henry Dalgleish, Noah Pettit, Michael Hausser, and Jakob Macke. Inferring neural population dynamics from multiple partial recordings of the same neural circuit. *Advances in Neural Information Processing Systems*, pages 539–547, 2013.

Leslie G Valiant. *Circuits of the Mind*. Oxford University Press, Inc., 1994.

Leslie G Valiant. Memorization and association on a realistic neural model. *Neural Computation*, 17(3):527–555, 2005.

Leslie G Valiant. A quantitative theory of neural computation. *Biological Cybernetics*, 95 (3):205–211, 2006.

Jurgen Van Gael, Yunus Saatci, Yee Whye Teh, and Zoubin Ghahramani. Beam sampling for the infinite hidden Markov model. *Proceedings of the International Conference on Machine Learning*, pages 1088–1095, 2008.

Michael Vidne, Yashar Ahmadian, Jonathon Shlens, Jonathan W Pillow, Jayant Kulkarni, Alan M Litke, EJ Chichilnisky, Eero Simoncelli, and Liam Paninski. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. *Journal of Computational Neuroscience*, 33(1):97–121, 2012.

Joshua T Vogelstein, Brendon O Watson, Adam M Packer, Rafael Yuste, Bruno Jedynak, and Liam Paninski. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophysical Journal*, 97(2):636–655, 2009.

Joshua T Vogelstein, Adam M Packer, Timothy A Machado, Tanya Sippy, Baktash Babadi, Rafael Yuste, and Liam Paninski. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, 104(6):3691–3704, 2010.

Hermann von Helmholtz and James Powell Cocke Southall. *Treatise on Physiological Optics: Translated from the 3rd German Ed.* Optical Society of America, 1925.

Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

Yair Weiss, Eero P Simoncelli, and Edward H Adelson. Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6):598–604, 2002.

Mike West, P Jeff Harrison, and Helio S Migon. Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association*, 80(389):73–83, 1985.

John G White, Eileen Southgate, J Nichol Thomson, and Sydney Brenner. The structure of the nervous system of the nematode Caenorhabditis elegans: the mind of a worm. *Philosophical Transactions of the Royal Society of London: Series B (Biological Sciences)*, 314:1–340, 1986.

Louise Whiteley and Maneesh Sahani. Attention in a Bayesian framework. *Frontiers in Human Neuroscience*, 6, 2012.

Alexander B Wiltschko, Matthew J Johnson, Giuliano Iurilli, Ralph E Peterson, Jesse M Katon, Stan L Pashkovski, Victoria E Abraira, Ryan P Adams, and Sandeep Robert Datta. Mapping sub-second structure in mouse behavior. *Neuron*, 88(6):1121–1135, 2015.

Jesse Windle, Nicholas G Polson, and James G Scott. Sampling Pólya-gamma random variates: alternate and approximate techniques. *arXiv preprint arXiv:1405.0506*, 2014.

Frank Wood and Michael J Black. A nonparametric Bayesian alternative to spike sorting. *Journal of Neuroscience Methods*, 173(1):1–12, 2008.

Frank Wood, Jan Willem van de Meent, and Vikash Mansinghka. A new approach to probabilistic programming inference. *arXiv preprint arXiv:1507.00996*, 2015.

Tianming Yang and Michael N Shadlen. Probabilistic reasoning by neurons. *Nature*, 447 (7148):1075–80, June 2007.

Byron M. Yu, John P. Cunningham, Gopal Santhanam, Stephen I. Ryu, Krishna V. Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Journal of Neurophysiology*, 102:614–635, 2009.

Alan Yuille and Daniel Kersten. Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10(7):301–308, 2006.

Richard S Zemel, Peter Dayan, and Alexandre Pouget. Probabilistic interpretation of population codes. *Neural Computation*, 10(2):403–30, February 1998.

Ke Zhou, Hongyuan Zha, and Le Song. Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 16, 2013.

Mingyuan Zhou, Lingbo Li, Lawrence Carin, and David B Dunson. Lognormal and gamma mixed negative binomial regression. *Proceedings of the International Conference on Machine Learning*, pages 1343–1350, 2012.