1
Introduction

Neuroscience is undergoing a data-driven revolution. In laboratories around the world, optical tools are illuminating thousands of neurons at a time (Kerr and Denk, 2008). For some organisms, we can now monitor large fractions of the neurons in the brain (Ahrens et al., 2013; Prevedel et al., 2014; Lemon et al., 2015). Not long ago, this immense recording capability would have been mind-boggling to the average practitioner poking around in the dark, hoping to record from a handful of neurons. Complementing these recording capabilities, recent advances in connectomics — the mapping of synaptic connectivity (Sporns et al., 2005) — are providing a more complete atlas of the wiring diagram that underlies neural activity (Lichtman et al., 2008; Helmstaedter et al., 2013; Oh et al., 2014). Likewise, sophisticated monitoring and processing methods (e.g. Berman et al., 2014; Wiltschko et al., 2015) are providing rich measurements of the natural behavior that this activity gives rise to.

These unprecedented technological capabilities are leading to a fundamental paradigm shift. The scientific process of proposing hypotheses, testing them against experimental observations, and revising them accordingly, is becoming increasingly data-driven. Rather than collecting data to test a specific hypothesis, we can now collect large-scale recordings in relatively unstructured experimental setups (e.g. from freely behaving animals) and use statistical structure in the data to guide us toward hypotheses. Of course, there is no "free

lunch" — we must still specify what types of structure to look for. For example, when we cluster neuronal data we are hypothesizing that there are different types of neurons, and when we apply principal components analysis to neural population recordings we are assuming that neural activity reflects a low-dimensional latent state. As we move toward this more data-driven paradigm, we are presented with a challenge: how do we formulate models that capture our intuitions and hypotheses about neural computation, how do we fit these models to large scale recordings, and how do we revise our models based on what we have learned?

This thesis builds a sequence of modeling and inference tools to address this challenge. We build tools for analyzing *neural spike trains*, which are sequences of discrete events in time, and we use probabilistic models to compose flexible hypotheses that allow us to *discover structure* in the dynamics of spike trains. The workhorse of this process is a set of *Bayesian methods*, statistical inference algorithms that take in a neural recording and output a distribution over parameters and variables of our model. We begin by describing each of these these in turn.

Neurons, Spikes, and Computation

The human brain is a three pound ball of densely packed cells. It is soft, with a consistency like that of tofu. Amazingly, from this curd-like mass of cells, our mental faculties arise. About half of the roughly 170 billion cells in our brain are neurons, widely believed to be the fundamental units of computation (Dayan and Abbott, 2001). Neurons are electrically excitable cells that contain an ionic soup of charged atoms like sodium, potassium, and calcium, which together maintain a difference in electrical potential between the inside and the outside of the cell membrane (Kandel et al., 2000). In most neurons, when the membrane potential is excited above a certain threshold, a cascade of ion channels opens and closes, causing the membrane potential to undergo a characteristic action potential, or "spike."

The first challenge of deciphering neural computation is understanding how information is encoded in patterns of coordinated spiking activity across neural populations. To first approximation, spikes are discrete events in time. That is, in a short enough window

of time, a neuron either does or does not fire an action potential. Much progress has been made in understanding how sensory neurons digitize continuous signals and how the activity of motor neurons innervates muscles and drives behavior (Rieke et al., 1999), yet how internal states, plans, goals, decisions, and thoughts are encoded remains largely a mystery.

The second challenge is understanding how this information is transformed over time. The neurons in our brain are connected by about 10^{14} synapses (Kandel et al., 2000). When a pre-synaptic neuron fires an action potential, neurotransmitters are released that then bind with receptors on the post-synaptic neuron, causing its ion channels to open and allowing current to flow into or out of the post-synaptic cell. These currents induce brief changes in the membrane potential of the downstream cell called post-synaptic potentials (PSPs). Depending on the direction of current, the PSP will be either excitatory, making the downstream neuron more likely to spike, or inhibitory, suppressing post-synaptic spiking. The probability of neurotransmitter release and the number of post-synaptic receptors together determine the "strength" of the synapse, which is manifest in the amplitude of the PSP (Cowan et al., 2003).

This web of synaptic connections imbues neural circuits with complex dynamics that govern how patterns of neural activity evolve over time, along with the information they encode. It is these dynamics that actually perform computation — the transformation of an input into an output. While these dynamics have been very well studied at the level of single connections between pairs of neurons, the dynamics of large networks of interconnected neurons is at the frontier of research.

Perhaps the most fundamental aspect of intelligence is our ability to learn, to store memories, and to generalize from past experience. In neural circuits, learning is most directly associated with the process of synaptic plasticity. In response to coordinated patterns of pre- and post-synaptic spiking, synapses adapt their strength. This synaptic plasticity leads to changes in the dynamics of neural circuits. In some cases, this plasticity leads to the creation of new associations that support generalization from noisy or partial sensory input. Thus, a third challenge of deciphering neural computation is understanding the processes of learning that cause neural dynamics to evolve in an activity-dependent manner.

Our goal, as neuroscientists, is to address these three challenges — encoding, dynamics, and learning — by searching for patterns in large-scale recordings of neural spike trains,

simultaneously recorded time series of spiking activity in populations of neurons. While this type of data is very hard to collect in humans, it is widely collected in worms, fish, flies, mice, rats, monkeys, and a host of other model organisms. Still, these model organisms have taught us a great deal about the workings of human brains and intelligence. This thesis does not focus on the unique aspects of any single organism or brain region, but rather on the general challenges associated with modeling structure in time series of discrete events that are common to all neural spike trains.

Discovering Structure: Types, Features, Networks, and States

Discovering structure is primarily about finding meaningful abstractions. While neurons, spikes, and synapses are the elementary building blocks of many models of neural computation, our goal is to connect this level of detail to more abstract theories of computation. Just as our knowledge of *in silico* computation is partitioned into a hierarchy of concepts — from transistors, logic gates, pipelines and processors, to assembly code, operating systems, algorithms and programs — our knowledge of neural computation must also include a hierarchy of concepts and descriptions. Marr (1982) proposed three levels of abstraction: the computational level, which concerns the inputs and outputs of a system; the algorithmic level, which specifies the transformations between inputs and outputs; and the implementation level, which focuses on how these transformations are realized in neural substrates. From this perspective, our goal is to interpolate between neural spike trains, typically associated with implementation level descriptions, and higher level abstractions, like those hypothesized by algorithmic and computational theories. To do so, we compose our spike train models out of a set of simple motifs that appear over and over again in models of neural computation.

Discrete Types and Low-Dimensional Features

Many successes of neuroscience have come from careful cataloging of neural types and features. From the earliest investigations of Ramón y Cajal, it has been clear that the brain is made up of anatomically distinct cell types (Cajal, 1899), but further classifications have been made on the basis of functional distinctions as well. Kuffler (1953) identified the re-

sponse properties of "on" and "off" ganglion cells in the retina, characterizing the ways in which these cells respond to light. Kuffler's students, Hubel and Wiesel, carried on this work, characterizing simple and complex cells in visual cortex (Hubel and Wiesel, 1962) and eventually winning the Nobel Prize for their discoveries. These are pioneering examples of a common theme in neuroscience: clustering cells into discrete types that facilitate our understanding of neural computation. Indeed, the clustering of cells in the early visual pathway continues to this day (Macosko et al., 2015; Sanes and Masland, 2015).

Complementing these classic examples of discrete subtypes of neurons are similarly compelling examples of neurons characterized by continuous features. Most prominent among these are the place cells of the hippocampus (O'Keefe and Nadel, 1978). These cells fire selectively when an animal is in a particular location in its environment. Thus, each cell is naturally associated with a continuous position in space. As we seek to make sense of spike trains from other complex neural circuits, discrete latent types and continuous latent features form one of the core building blocks in our probabilistic modeling toolkit.

Networks and Dynamics

Networks play a central role in modern neuroscience: they enable us to reason about complex systems in terms of relationships among their constituent parts. Whether the nodes of the network represent individual neurons in a "connectome" (e.g. Sporns et al., 2005), populations of cells in a neural circuit (e.g. Felleman and Van Essen, 1991), voxels in an fMRI recording (e.g. Friston, 1994), or idealized neurons in a theoretical model (e.g. Hopfield, 1982), networks tell us about the pairwise relationships in large populations of nodes. Once we have extracted such a network, a great deal of intuition can be gleaned from its aggregate properties (Bullmore and Sporns, 2009; Newman, 2003). Moreover, the network itself can often be summarized in terms of latent types and features of nodes that govern how likely any pair of nodes is to be connected (Goldenberg et al., 2010).

We often use networks to represent the dynamics of systems. For example, an edge in the network may represent the influence that activity on one node exerts on the subsequent activity of another. In this way, networks provide a simple summary of complex dynamics. Much of this thesis is devoted to learning network representations from neural spike trains, building on a rich body of existing work on generalized linear models for neu-

ral spike trains (Paninski, 2004; Truccolo et al., 2005; Pillow et al., 2008). In these models, nodes correspond to the individual neurons in our dataset, and the edges represent a probabilistic relationship between the spiking activity of one neuron and the future firing rate of its downstream neighbors. By combining latent variable models for the network with dynamics models for its edges, this thesis will construct sophisticated, hierarchical models for dynamical spike trains.

LATENT STATES OF NEURAL POPULATIONS

Networks model dynamic data in terms of relationships between nodes, which are typically associated with individual neurons in the generalized linear model. In some cases, however, it is more natural to think about neural dynamics in terms of a latent state that evolves over time but cannot be directly observed. For example, population activity might reflect a low-dimensional, continuous latent state with smooth (Yu et al., 2009) or linear (Smith and Brown, 2003; Paninski et al., 2010) dynamics, or a discrete latent state with Markovian dynamics (e.g. Jones et al., 2007; Latimer et al., 2015). In these models, the firing rates of individual neurons are a function of the underlying latent state.

By combining these simple building blocks — latent types and features, networks, and dynamic latent states — we can express sophisticated hypotheses about the underlying structure in observed neural spike trains. However, in order to determine this structure, we need to instantiate these hypotheses in the form of a model that we can fit and evaluate. Bayesian probabilistic models and inference algorithms provide a principled means of tackling this problem.

A BAYESIAN APPROACH

Structural hypotheses specify how a set of parameters and latent variables give rise to patterns of neural activity. These hypotheses are naturally formalized in terms of probabilistic generative models, which provide an intuitive way of expressing the joint probability of parameters, latent variables, and observed data. Once formalized, we use the tools of Bayesian inference to compute the posterior distribution over parameters and latent variables implied by a given spike train and our prior beliefs, and use this posterior distribution to gain
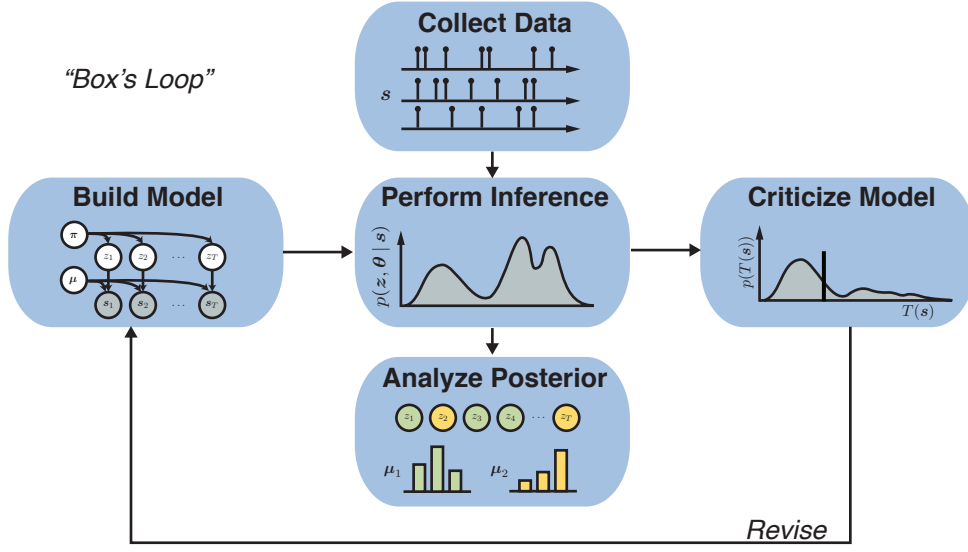
Figure 1.1: Box's loop for hypothesis driven probabilistic modeling. Adapted from Blei (2014).

insight into the structure of the data and potential shortcomings of our theories.

Recent decades have witnessed an explosion of interest in Bayesian machine learning methods, and the development of both theoretical and analytical tools have led to the widespread adoption of these techniques (Bishop, 2006; Murphy, 2012). Nevertheless, performing Bayesian inference in sophisticated models will always be a challenge. The art of probabilistic modeling lies in balancing two objectives: designing models that capture the details of the generative processes underlying our data, while simultaneously ensuring that these models admit efficient inference algorithms. To achieve this balance, we often look for model-specific structure that we can exploit during inference. This thesis is largely dedicated to developing algorithms that capitalize on the properties of models in order to efficiently scale to large recordings.

Figure 1.1 outlines the process of model building. We begin by collecting data, in this case, large-scale recordings of neural spike trains, and building a probabilistic model that captures our intuitions and hypotheses about structure in that data. Given these two ingredients, we perform Bayesian inference to compute the posterior distribution over structural variables and parameters. Visualizing, analyzing, and exploring this posterior often lead to

new insights and ways to think about generative processes that can explain further structure in the data. We can also criticize the model, formally, by assessing goodness-of-fit, evaluating predictive likelihoods, and performing posterior predictive checks (Gelman et al., 2013). This leads to new hypotheses, which in turn lead to new models and repetition of this process. Thus, model building is an interative process in which we are constantly adapting and refining hypotheses. Blei (2014), from whom this figure has been adapted, calls this "Box's loop," a reference to the foundational work of Box (1980).

We are certainly not the first to adopt a Bayesian approach to modeling of neural data. Indeed, probabilistic methods (e.g. Brillinger et al., 1976; Brillinger, 1988; Paninski, 2004; Truccolo et al., 2005; Pillow et al., 2008), particularly Bayesian methods (e.g. Sahani, 1999; Rieke et al., 1999; Yu et al., 2009; Park and Pillow, 2011; Macke et al., 2011), have played a major role in advancing our understanding of spike trains. We build on this foundation and specifically focus on expanding the set of models, combining standard point process and spike count models with hierarchical prior distributions that capture latent structure. To go along with these more sophisticated models, we also derive efficient[*] Bayesian inference algorithms that take advantage of clever data augmentation strategies. Together, these expand the repertoire of probabilistic models that we may leverage when looking for structure in neural data.

Summary of Contributions

This thesis consists of a sequence of probabilistic models and inference algorithms that capture increasingly sophisticated structure in neural spike trains. While these models build upon one another, layering in additional structural hypotheses and incorporating more and more detail, the chapters can also be read independently without severe loss of clarity, especially by those who are more versed in machine learning.

We begin with a brief overview of point processes, probabilistic modeling, and Bayesian inference in Chapter 2. This lays the foundation for the methodological contributions in

---

[*]We do not mean efficient in the formal, complexity-theoretic sense. Indeed, approximate Bayesian inference is, in the worst case, NP-hard (Dagum and Luby, 1993; Roth, 1996). We simply mean that the individual iterations of our algorithm can be performed in low-order polynomial time and that these algorithms yield good empirical results in a reasonable, if not provably polynomial, number of iterations.

the remainder of this thesis. We also introduce common notation for spike trains, firing rates, and latent variables of the model. Readers who are well versed in probabilistic modeling may skip this chapter without much harm.

Chapter 3 introduces our first probabilistic model for neural spike trains — a combination of continuous time Hawkes processes and probabilistic network models. This is based on Linderman and Adams (2014). However, since this is the first model of the thesis, we provide a more thorough discussion of network models, Hawkes processes, and the steps in deriving an efficient Markov chain Monte Carlo (MCMC) inference algorithm. This is also the only chapter that will apply these techniques to spike trains from domains other than neuroscience, specifically, from finance and criminology.

Chapter 4 continues the focus on Hawkes processes and networks, but here we break from the continuous time models of Chapter 3 and begin working in discrete time. The remainder of the thesis continues in this vein. This chapter also introduces a scalable variational inference algorithm that leverages the discrete model structure. This is based on Linderman and Adams (2015).

Chapter 5 reconsiders the linear interactions of Hawkes processes and develops a nonlinear autoregressive model for neural spike trains. Again, we use probabilistic network models as prior distributions over the pattern of functional interaction, and in doing so, show that interesting representations of neural populations can be learned directly from the data. The key to efficient inference is the Pólya-gamma augmentation — a recently developed method for performing Bayesian inference in models with linear Gaussian structure and discrete observations (Polson et al., 2013). A preliminary version of this work was presented by Linderman et al. (2015).

Chapter 6 continues the network theme, but here we begin to transition from static models to ones with dynamic latent states. Specifically, we adopt a mechanistic view of the network and treat its connections as actual synapses. This leads to a consideration of the evolution of synaptic weights over time. We construct a framework for modeling arbitrary synaptic plasticity rules and fitting them with particle MCMC. This chapter is based on Linderman et al. (2014).

Chapter 7 departs from network models and instead considers dynamical discrete state space models, specifically hidden Markov models (HMM), for neural data. We address

a major challenge in applying these well-known models, namely, selecting the number of latent states. By introducing a hierarchical Dirichlet process (HDP) prior and both an MCMC and a variational inference algorithm, we develop a Bayesian nonparametric solution to this problem. While this combination is relatively well studied, these models are notoriously sensitive to hyperparameter settings. Thus, we introduce a variety of methods for selecting hyperparameters and perform a thorough comparison on real and synthetic data. This is based on Linderman et al. (2016a).

Building on the discrete state space models of the preceding chapter, Chapter 8 develops efficient Bayesian inference algorithms for switching linear dynamical systems (SLDSs) with discrete count observations. The Pólya-gamma augmentation once again makes Bayesian inference surprisingly easy. Given these auxiliary variables, a host of tools for inference in Gaussian models is at our disposal. We then turn to a major problem, that of model comparison. We derive a novel sampling algorithm for the Pólya-gamma distribution that renders annealed importance sampling (AIS) simple and efficient for a broad class of models, including the SLDSs of this chapter. A preliminary version of this work was presented in Linderman et al. (2016b).

Chapter 9 takes a step in a radical direction. We build on the probabilistic models of the preceding chapters, but here we combine these models with a top-down theory of neural computation. Specifically, we consider the repercussions of the "Bayesian brain" hypothesis, namely, that neural circuits are performing approximate Bayesian inference in a probabilistic model of the world. We make predictions about the patterns of neural spike trains that would be expected from such a circuit, and show how the methods of previous chapters can be leveraged to reverse engineer probabilistic models from observations of neural spike trains. This work is necessarily speculative, but we believe it suggests a promising path forward as we seek to reconcile computational and algorithmic theories of neural computation with large scale recordings.

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.

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 block-structures. *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.