35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

A Variational Nonparametric Bayesian Approach for Inferring Rat Hippocampal Population Codes

Zhe Chen, *Senior Member, IEEE*, and Matthew A. Wilson

***Abstract*— Rodent hippocampal population codes represent important spatial information of the environment during navigation. Several computational methods have been developed to uncover the neural representation of spatial topology embedded in rodent hippocampal ensemble spike activity. Here we extend our previous work and propose a nonparametric Bayesian approach to infer rat hippocampal population codes. Specifically, we develop an infinite hidden Markov model (iHMM) and variational Bayes (VB) inference method to analyze rat hippocampal ensemble spike activity. We demonstrate the effectiveness of our approach using an open field navigation example and discuss the significance/implications of our results.**

***Index Terms*— Spatial representation, hippocampal population codes, infinite hidden Markov model, variational Bayes**

I. Introduction

The hippocampus plays an important role in spatial navigation and episodic memory. Spatial representation of the environment is pivotal for navigation in rodents. Unlike the topographic map with complete metric information, a topological map contains only relative ordering or connectivity information between spatial locations and is invariant to orientation or deformation. An important question is that how neurons downstream of the hippocampus can infer representations of space from hippocampal spike activity without *a priori* place field information (namely, without the measurement of spatial correlate). Several reports have dedicated to mathematical analysis of this problem [6], [7]. We have previously developed a probabilistic approach to uncover the neural representation of spatial topology embedded in rodent hippocampal population codes [2], [3]. Here we extend the previous work and consider a variational nonparametric Bayesian approach for statistical inference. Specifically, we reformulate the finite-state hidden Markov model (HMM) into an infinite HMM (iHMM) within a nonparametric Bayesian framework, and derive an efficient variational Bayesian (VB) inference algorithm. The variational nonparametric Bayesian inference was motivated from the work of [9], which appeared faster and more efficient than other sampling-based inference methods [13], [14]. For our population Poisson firing model, it is also important to point out that the VB formulation with conjugate priors implicitly accommodates an overdispersed Poisson model known as the negative binomial distribution. The negative binomial distribution can be viewed as a continuous mixture of Poisson distributions, where the mixing distribution of the Poisson rate is a gamma distribution.

We test the new approach with rat hippocampal data recorded in a two-dimensional (2D) open field environment. From the derived spatial topology graph, we quantify the animal's behavior in terms of the number of uniquely visited states and the mean state density. We also discuss the practicality of topographically-constrained codes in spatial representation and its implication for robotic localization and mapping.

Supported by NIH Grant RO1-MH061976 (to M.A.W.) and an Early Career Award from the Mathematical Biosciences Institute, Ohio State University (to Z.C.). The authors are with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. (Email: zhechen@mit.edu)

II. Methods

A. Previous Probabilistic Model

We have previously used a finite m-state HMM to characterize the population spiking activity from a population of C hippocampal place cells [2], [3]. We assumed that (i) the animals spatial location during locomotion, modeled as a latent state process, followed a first-order discrete-state Markov chain $\mathcal{S} = S_{1:T} \equiv \{S_t\} \in \{1, \ldots, m\}$, and (ii) the spike counts of individual place cells at time t, conditional on the hidden state S_t, followed a Poisson probability with their respective tuning curve functions $\mathbf{\Lambda} = \{\lambda_c\}$. Essentially, we employed a Markov-driven population Poisson firing model with the following probability distributions

$$S_{t-1} \rightarrow S_t \sim P_{S_{t-1}S_t} \quad (1)$$

$$y_{c,t}|S_t = j \sim Poisson\Big(y_{c,t}|\lambda_c(j)\Big) \quad (2)$$

$$\log p(\mathbf{y}_{1:T}|\mathcal{S}, \boldsymbol{\theta}) = \sum_{c=1}^{C}\sum_{t=1}^{T} \log p(y_{c,t}|S_t, \boldsymbol{\theta}) \quad (3)$$

where $\sim$ denotes "distributed from"; $\boldsymbol{P} = \{P_{ij}\}$ denotes an m-by-m state transition probability matrix, with P_{ij} representing the transition probability from state i to j (since $\sum_{k=1}^{m} P_{ik} = 1$, each row of matrix $\boldsymbol{P}$ specifies a multinomial likelihood); $y_{c,t}$ denotes the number of spike counts from the c-th place cell within the t-th temporal bin (we used 250 ms bin size during locomotion) and $\mathbf{y}_{1:T} = \{y_{c,t}\}$ denotes time series of C-dimensional population response vector; and $Poisson(y_c|\lambda_c)$ defines a Poisson distribution with the rate parameter λ_c. Finally, $\log p(\mathbf{y}_{1:T}|\mathcal{S}, \boldsymbol{\theta})$ defines the observed data log-likelihood given the hidden state sequence $\mathcal{S}$ and all parameters $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{P}, \mathbf{\Lambda}\}$ (where $\boldsymbol{\pi} = \{\pi_i\}$ denotes a probability vector for the initial state S_1.

B. Infinite HMM (iHMM)

We have previously proposed a Bayesian deviance information criterion (DIC) to select the model size m [2],

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

[3]. Here we extend the finite-state HMM to an iHMM, which is a nonparametric Bayesian extension of the HMM with an infinite number of hidden states [1]. Namely, the iHMM treats the priors via a stochastic process. Instead of imposing a Dirichlet prior distribution onto the finite state transition matrix $\boldsymbol{P}$, we use an infinite-dimensional generalization of the Dirichlet distribution: Dirichlet process, for the infinite-length multinomial distribution. When the number of hidden states m reaches infinity, the hierarchical prior will approach a hierarchical Dirichlet process [13]. Specifically, we construct the Dirichlet process via a "stick-breaking process" [9], and assume the following prior for the i-th row of transition matrix $\boldsymbol{P}$ [9]

$$\tilde{P}_{ij} \sim Beta(1, \alpha_P), \quad P_{ij} = \tilde{P}_{ij} \prod_{k=1}^{j-1} \Big(1 - \tilde{P}_{ik}\Big) \tag{4}$$

where $\sum_{j=1}^{\infty} P_{ij} = 1$, and $Beta(a, b)$ defines a beta distribution with two positive shape parameters a and b. A similar formulation also applies to the element of $\boldsymbol{\pi} = \{\pi_i\}$

$$\tilde{\pi}_i \sim Beta(1, \alpha_\pi), \quad \pi_i = \tilde{\pi}_i \prod_{k=1}^{i-1} \Big(1 - \tilde{\pi}_k\Big) \tag{5}$$

The name "stick-breaking" comes from the interpretation of π_i as the length of the piece of a unit-length stick assigned to the i-th value. After the first $i-1$ values having their portions assigned, the length of the remainder of the stick is broken according to a sample $\tilde{\pi}_i$ from a beta distribution, and $\tilde{\pi}_i$ indicates the portion of the remainder to be assigned to the i-th value. In Eq. 5, the smaller α_π, the less (in a statistical sense) of the stick will be left for subsequent values.

C. Overdispersed Poisson Model

To account for overdispersion of discrete data (i.e., the sample variance exceeds the sample mean), the negative binomial (NB) distribution is often used to characterize overdispersed Poisson firing of neurons. The negative binomial distribution is a continuous mixture of Poisson distributions (i.e., a compound probability distribution) where the mixing distribution of the Poisson rate is a gamma distribution. In other words, the NB distribution is viewed as a gamma-Poisson (mixture distribution): a $Poisson(\lambda)$ distribution whose rate λ is itself a gamma random variable $Gamma(r, \frac{p}{1-p})$ $(r > 0, 0 < p < 1)$. By setting $\lambda = \frac{rp}{1-p}$ or $p = \frac{\lambda}{r+\lambda}$, it is known that

$$Poission(\lambda) = \lim_{r \to \infty} NB\Big(r, \frac{\lambda}{r + \lambda}\Big) \tag{6}$$

where the NB mean and variance statistics are $\frac{rp}{1-p}$ and $\frac{rp}{(1-p)^2}$, respectively. In the Bayesian framework, if we assign the rate parameter with a gamma conjugate prior, the posterior will be a gamma distribution [2] and the posterior predictive (upon marginalizing the parameters) will have a form of NB distribution [10]. Therefore, even a Poisson firing likelihood model is used, the Bayesian formulation implicitly accommodates the overdispersion phenomenon, consequently improving the robustness of the firing rate model.

D. Variational Bayesian (VB) Inference

The VB inference can be derived in a similar fashion as in [2]. The objective of VB inference is to maximize the marginal log-likelihood or its lower bound

$$\begin{aligned}
\log p(\mathbf{y}_{1:T}) &= \log \int d\boldsymbol{\pi} \int d\boldsymbol{P} \int d\boldsymbol{\Lambda} \sum_{\mathcal{S}} p(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}) \\
&\quad \times p(\mathbf{y}_{1:T}, \mathcal{S} | \boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}) \\
&\geq \int d\boldsymbol{\pi} \int d\boldsymbol{P} \int d\boldsymbol{\Lambda} \sum_{\mathcal{S}} q(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \mathcal{S}) \\
&\quad \times \log \frac{p(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}) p(\mathbf{y}_{1:T}, \mathcal{S} | \boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda})}{q(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \mathcal{S})} \\
&= \Big\langle \log p(\mathbf{y}_{1:T}, \mathcal{S}, \boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}) \Big\rangle_q + \mathcal{H}_q(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \mathcal{S}) \equiv \mathcal{F}(q)
\end{aligned} \tag{7}$$

where $p(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda})$ denotes the parameter prior distribution, $p(\mathbf{y}_{1:T}, \mathcal{S} | \boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda})$ defines the complete data likelihood, and $q(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \mathcal{S})$ is called the variational posterior distribution which approximates the joint posterior of the unknown state and parameter $p(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \mathcal{S} | \mathbf{y}_{1:T})$. The term $\mathcal{H}_q$ represents the entropy of the variational posterior distribution q, and $\mathcal{F}$ is referred to as the free energy.

Note that although the state space is infinite, the posteriors will only have large transition probabilities in a finite number of states while all others will be nearly equal to zero. Similar to the nonparametric VB formulation in [9], two approximations are made: first we assume a factorial form of variational posterior distribution $q(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \mathcal{S}) = q(\boldsymbol{\pi}) q(\boldsymbol{P}) q(\boldsymbol{\Lambda}) q(\mathcal{S}) \approx p(\boldsymbol{\pi}, \boldsymbol{P}, \boldsymbol{\Lambda}, \mathcal{S} | \mathbf{y}_{1:T})$; second, we only compute the probabilities of m states of the infinite large state space (where m is called the "truncation level of stick-breaking"). Note that the truncation level is an approximation of the infinite state space, which is statistically different from the assumption of a finite m-state space [9].

We use a beta prior for the stick-breaking probability (Eqs. 4 and 5) and a gamma conjugate prior for the Poisson rate parameter $\lambda_c(j)$ (see [2]). In the VB-M step, for $i, j = 1, \ldots, m$, we update the variational parameter posteriors (in the same conjugate family) as follows

$$\begin{aligned}
q(\tilde{\pi}_i) &= Beta(a_{\pi_i}, b_{\pi_i}) \\
&= Beta\Big(1 + q(S_1 = i), \alpha_\pi + q(S_1 > i)\Big) \\
q(\tilde{P}_{ij}) &= Beta(a_{P_{ij}}, b_{P_{ij}}) \\
&= Beta\Big(1 + \sum_{t=2}^{T} q(S_{t-1} = i, S_t = j), \\
&\qquad \alpha_P + \sum_{t=2}^{T} q(S_{t-1} = i, S_t > j)\Big)
\end{aligned}$$

where (a_{π_i}, b_{π_i}) and $(a_{P_{ij}}, b_{P_{ij}})$ are the hyperparameters of the beta distributions for $\tilde{\pi}_i$ and $\tilde{P}_{ij}$, respectively. The variational posterior $q(\boldsymbol{\Lambda})$ has a nearly identical form as derived previously [2].

In the VB-E step, we update the variational state posterior

$$\begin{aligned}\log q(\mathcal{S}) &= -\log Z(\mathbf{y}_{1:T}) + \big\langle \log p(\mathcal{S}, \mathbf{y}_{1:T}|\boldsymbol{\theta})\big\rangle_{q(\boldsymbol{\pi})q(\boldsymbol{P})q(\boldsymbol{\Lambda})} \\ &= -\log Z(\mathbf{y}_{1:T}) + \sum_{i=1}^{m} \hat{S}_{1,i}\big\langle \log \pi_i\big\rangle_{q(\boldsymbol{\pi})} \\ &\quad + \sum_{t=2}^{T}\sum_{i=1}^{m}\sum_{j=1}^{m} \hat{S}_{t-1,i}\hat{S}_{t-1,j}\big\langle \log P_{ij}\big\rangle_{q(\boldsymbol{P})} \\ &\quad + \sum_{t=1}^{T}\sum_{c=1}^{C}\sum_{i=1}^{m} \hat{S}_{t,i}\big\langle -\lambda_{ic} + y_{c,t}\log\lambda_{ic}\big\rangle_{q(\boldsymbol{\Lambda})}\end{aligned}$$

with

$$\begin{aligned}\big\langle \log \pi_i\big\rangle_{q(\boldsymbol{\pi})} &= \psi(a_{\pi_i}) - \psi(a_{\pi_i} + b_{\pi_i}) \\ &\quad + \sum_{k=1}^{i-1}\Big[\psi(b_{\pi_k}) - \psi(a_{\pi_k} + b_{\pi_k})\Big] + C_1 \\ \big\langle \log P_{ij}\big\rangle_{q(\boldsymbol{P})} &= \psi(a_{P_{ij}}) - \psi(a_{P_{ij}} + b_{P_{ij}}) \\ &\quad + \sum_{k=1}^{i-1}\Big[\psi(b_{P_{ik}}) - \psi(a_{P_{ik}} + b_{P_{ik}})\Big] + C_2.\end{aligned}$$

where ψ is the *diagamma function*, C_1 and C_2 are normalizing constants such that $\sum_i q(\pi_i) = 1$ and $\sum_j P_{ij} = 1$. The computation of $\big\langle -\lambda_{ic} + y_{c,t}\log\lambda_{ic}\big\rangle_{q(\boldsymbol{\Lambda})}$ is the same as [2].

Finally, we compute the free energy $\mathcal{F}(q(\boldsymbol{\pi})q(\boldsymbol{P})q(\boldsymbol{\Lambda})q(\mathcal{S}))$ by iteratively minimizing four equations of the Kullback-Leibler (KL) divergence between the posteriors and the priors until it reaches a local optimal solution.

III. Experimental Data and Results

Experiments were conducted under the supervision of the Massachusetts Institute of Technology Committee on Animal Care and followed the NIH guidelines. The microdrive arrays containing between 11 and 24 tetrodes were implanted above the right dorsal hippocampus of male Long-Evans rats. The tetrodes were slowly lowered into the brain reaching the cell layer of CA1 two to four weeks following the date of surgery. Recorded spikes were manually clustered and sorted to obtain single units using a custom software (XClust, M.A.W.). Data from one animal's recording session were used here for illustration purpose. Once 37 stable hippocampal units were obtained, the animal was allowed to freely foraged in an open field environment (radius: 60 cm). To identify the period of rodent locomotion during spatial navigation, we used a velocity threshold (>12 cm/s) to select the RUN epochs and merged them together. The animal's RUN trajectory (about 7.5-min recording) and spatial occupancy map are shown in Fig. 1.

A. State Transition Matrix and Topology Graphs

We started with a large value of the truncation level: $m = 169$ for the iHMM. Upon VB inference we recovered the posterior and posterior mean of the state sequence and state transition matrix of the iHMM. As expected, many rows of the estimated matrix $\hat{\boldsymbol{P}}$ shrank to zeros (Fig. 2, left). We empirically selected a nonzero submatrix with size 82-by-82 from $\hat{\boldsymbol{P}}$, thresholded (zeroing values below 0.01) and renormalized it (denoted by $\hat{\boldsymbol{P}}_{new}$). By varying the initial m value between 100 and 180, the "effective" state dimensionality remained similar (between 78 and 84). This result appeared very robust (i.e., low variance) compared to the Bayesian DIC model selection (for an independent fixed 81-state HMM, results not shown).

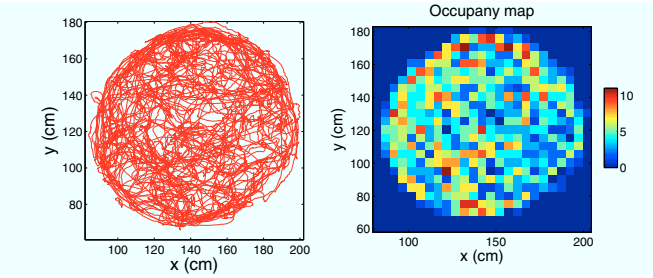

Fig. 1. Rat's behavioral trajectory (left) and spatial occupancy (right, 5cm bin size) in the open field environment.

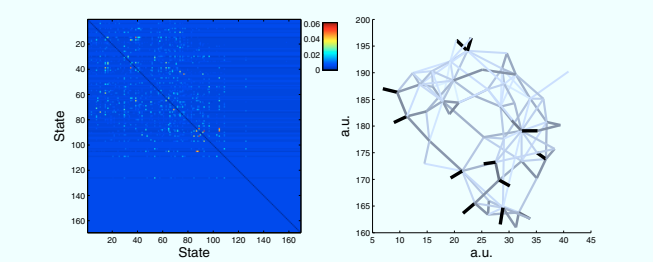

Fig. 2. Left: Estimated state transition matrix of iHMM for $m = 169$. Upon VB inference, many matrix rows were nearly zero. Right: Inferred spatial topology graph based on the a nonzero submatrix (82-by-82) of the left panel (upon proper thresholding and renormalization).

We further visualized the spatial topology by an undirected graph, in which the nodes represent the states. The presence of the edges between the nodes indicates that two nodes were connected in space, and the strength of the edge between two nodes (shown by the color darkness) is proportional to the transition probability value between two states. Specifically, we fed $\hat{\boldsymbol{P}}_{new}$ to a custom graph-drawing force-based algorithm [2] to produce a 2D topology graph (Fig. 2, right). Note that the topology graph is dimensionless (a.u.) and is invariant to scaling and rotation.

For the purpose of result assessment, we plotted the state space map (Fig. 3*A*, which showed the median value of the spatial position that each state represented) as well as the estimated animal's spatial trajectories (Fig. 3*B,C*). In the illustrated example, the mean (median) of the estimation errors across the complete RUN trajectory in the x and y-axes were 11.51 (8.81) cm and 10.76 (8.59) cm, respectively.

B. Quantification of Topological Codes

From the derived topology graph, we characterized the activeness of the rat's RUN behavior in time (Fig. 4, left) by counting the number of uniquely visited nodes (normalized by the total number of the nodes). Given the state space

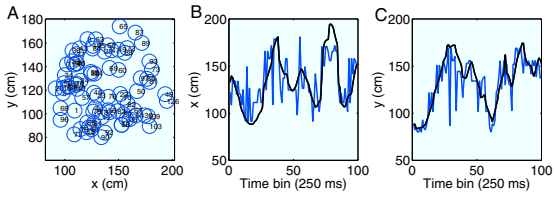

Fig. 3. ***A***, Estimated state space map, where the median value of the spatial position that each state (circled number) represented was shown. From the state space map and state sequence, we inferred the animal's trajectory. ***B***,***C***, Comparison between the actual (black) and estimated (blue) animal's RUN trajectory (based on the state space map in ***A***). The mean (median) estimation errors across the complete RUN trajectory in the x and y-axes were 11.51 (8.81) cm and 10.76 (8.59) cm, respectively.

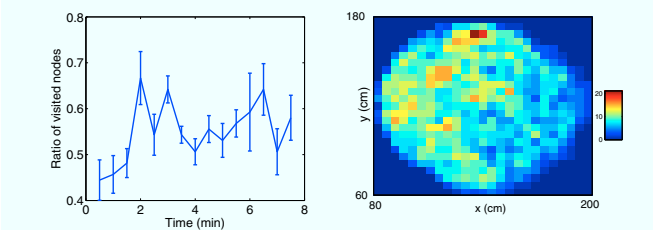

Fig. 4. Left: The ratio of uniquely visited states in time (error bar shows the the standard deviation from 10 Monte Carlo runs). Right: The heat map of the state density (mean 5.4, median 6, unit: 25cm^{-2}).

map, we also assessed the state density for every discretized spatial location (bin size 5 cm) of the environment. The state density measures the degree of ambiguity about the number of unique states required to represent a specific spatial location. Similar to the analysis in [3], we illustrated the state density by a heap map (Fig. 4, right), with a mean (median) value of 5.4 (6). Note that the state density unit is 25cm^{-2}. Comparing the heat map with the animal's occupancy map (Fig. 1, right), it appeared that these two maps were correlated (Pearson's correlation $\rho = 0.60, P < 10^{-7}$), and the area with a high occupancy was more likely to have a high state density.

C. Robustness of the Population Firing Model

As discussed in Section II.C, given a gamma conjugate prior, the posterior predictive distribution of Poisson firing model has a form of negative binomial distribution. To assess the robustness of the population firing model, at every 250-ms temporal bin we added additional non-Poissonian noise to the observed population spike counts by drawing from a negative binomial distribution (Eq. 6, with $r = 1, p = 0.5$, mean 1; or equivalently, 4 Hz noise). Upon repeating the analysis, we found that the results of effective state dimensionality, mean state density, and topography graph were similar. This suggests that the Poisson firing model is robust to overdispersion within the VB inference framework.

IV. Discussion

We have demonstrated that the HMM integrated with VB inference is capable of uncovering the neural representation of spatial topology from rat hippocampal population codes [2], [3]. The current work extends it to the iHMM by using a variational nonparametric Bayesian approach, our results validate the robustness of the new approach. There have been debates regarding the spatial representation of hippocampal codes: topographic vs. topological. A pure hippocampal topographic code implies a precise characterization of the place fields with respect to the full environment. Such a code would be optimal for decoding, but it may not be efficient for encoding especially given a large environment and small sample size. In contrast, an abstract topological code with a small number of topographic constraints may achieve a good tradeoff between decoding accuracy and encoding efficiency (results shown in [3]). Such topographic constraints may be environmental landmarks. From a computational perspective, it is therefore more efficient to decode an animal's spatial location based on a "fuzzy" semi-topographic map without storing the place fields of neuronal ensembles [3], [4].

In terms of engineering applications, our results also provide insight into the so-called "simultaneous localization and mapping" (SLAM) problem in mobile robots [5]. In this context, the sensor observations consist of noisy odometry measurements (obtained from laser-range finder or camera data) and the objective is to establish a spatial representation that allows correcting the internal model of the dynamic environment. In a graph-oriented approach [11], the nodes of pose graphs represent the poses of the robot at discrete time steps and are connected by constraints. Such spatial maps consist of both topographic and topological representations.

References

[1] M. J. Beal, Z. Ghahramani and C. E. Rasmussen, "The infinite hidden Markov model," *Advances in Neural Information Processing Systems, 14*, pp. 577-585. Cambridge, MA: MIT Press, 2002.

[2] Z. Chen, F. Kloosterman, E. N. Brown and W. A. Wilson, "Uncovering hidden spatial topology represented by hippocampal population neuronal codes," *J. Comp. Neurosci.*, vol. 33, pp. 227–255, 2012.

[3] Z. Chen, S. N. Gomperts, J. Yamamoto and W. A. Wilson, "Neural representation of spatial topology in the rodent hippocampus," *J. Neurophysiol.*, under review.

[4] Z. Chen, F. Kloosterman, S. Layton and W. A. Wilson, "Transductive neural decoding of unsorted neuronal spikes of rat hippocampus," in *Proc. IEEE EMBC12*, pp. 1310-1313, 2012.

[5] N. Cuperlier, M. Quoy and P. Gaussier, "Neurobiologically inspired mobile robot and planning," *Frontiers in Neurorobotics*, 1:4 doi: 10.3389/neuro.12/004, 2007.

[6] C. Curto and V. Itskov, "Cell groups reveal structure of stimulus space," *PLoS Comput. Biol.*, vol. 4, e1000205, 2008.

[7] Y. Dabaghian, F. Memoli, L. M. Frank and G. Carlsson, "A topological paradigm for hippocampal spatial map formatioin using persistent homology," *PLoS Comput. Biol.*, vol. 8, e1002581, 2012.

[8] T. J. Davidson, F. Kloosterman and W. A. Wilson, "Hippocampal replay of extended experience," *Neuron*, vol. 63, pp. 497–507, 2009.

[9] N. Ding and Z. Ou, "Variational nonparametric Bayesian hidden Markov model," in *Proc. IEEE ICASSP'10*, pp. 2098-2101, 2010.

[10] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (2nd ed). CRC Press, 2003.

[11] M. Kaess *et al.*, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Int. J. Robotics Research* vol. 31, pp. 217–236, 2012.

[12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press.

[13] Y. W. Teh, M. I. Jordan, M. J. Beal and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Stat. Assoc.*, vol. 101, pp. 1566–1581, 2006.

[14] J. van Gael, Y. Saatci, Y. W. Teh and Z. Ghahramani, "Beam sampling for the infinite hidden Markov model," in *Proc. 25th Int. Conf. Machine Learning*, pp. 1088–1095, 2008.