



HAL
open science

Spike based inference in a network with divisive inhibition

Sophie Denève, Timm Lochmann, Udo Ernst

► **To cite this version:**

Sophie Denève, Timm Lochmann, Udo Ernst. Spike based inference in a network with divisive inhibition. Deuxième conférence française de Neurosciences Computationnelles, "Neurocomp08", Oct 2008, Marseille, France. <hal-00331564>

HAL Id: hal-00331564

<https://hal.science/hal-00331564v1>

Submitted on 17 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

SPIKE BASED INFERENCE IN A NETWORK WITH DIVISIVE INHIBITION

Sophie Denève
Group for Neural Theory
ENS Paris
deneve@ens.fr

Timm Lochmann
Group for Neural Theory
ENS Paris
lochmann@ens.fr

Udo Ernst
Institute for Theoretical Physics
University of Bremen
udo@neuro.uni-bremen.de

ABSTRACT

Lateral inhibition has been found in many brain areas related to processing of sensory information. We examine the functional properties of a specific type of such inhibitory connectivity. Our analysis is grounded on a generative model (GM) describing the statistical relation between objects in the world (also called hidden states) and the neural response they evoke. For a set of assumptions regarding how information is processed in single cells and the assumptions contained in the generative model, we show that divisive inhibition enables a network to approximate optimal inference. We test the goodness of this approximation via numerical simulations. In these, we compare how well a sequence of hidden states can be reconstructed from the network activity. We show how this inference about the states can be formulated in the terminology of Hidden Markov Models (HMMs). This allows us to compare the networks performance to solutions obtained from standard HMM inference algorithms. We find divisive inhibition to clearly improve network performance and the resulting estimates being close to optimal. We discuss the results and mention further properties and extensions of our approach.

KEY WORDS

Artificial Neural Networks, Probabilistic Models and Methods, Neural Coding, Population Coding.

1 Introduction

Perceptual systems enable an organism to respond quickly and reliably in a changing environment. To this end, the underlying neural hardware processes signals in a way that can be interpreted as a kind of inference: it is trying to find out from noisy data what is going on in the outer world. Information from multiple channels (e.g. inputs from multiple afferents) has to be combined and signaled in an efficient way. For example, neurons early in the visual stream receive spike trains coming from other neurons and this type of input data is noisy and only stochastically related to what caused them (e.g. visual patterns presented to the eyes). To get a systematic understanding of how this information is processed and passed on, it is helpful to specify the statistical relationships between causes and resulting input patterns. This is concisely done in terms of *generative models* (GM) [1, 2]. Having specified the relationships in

the incoming data, one can ask how neural structures make use of them to perform the inference task. Because most signals reaching a cell could stem from multiple different configurations of causes, it is a central part of this inference task to separate different sources and to infer the most probable configuration.

The motivation for our work is that many studies characterizing the response of cells in the visual cortex found lateral connectivity resembling divisive inhibition (for a review see [3]). Together with modeling studies [4, 5, 6, 7, 8], this has provided insights into the functional role of the type of normalization induced by divisive inhibition. To our knowledge, however, a model for including inhibition in the case of online inference is lacking up to now. We therefore introduce a GM and a network performing online inference on the inputs from this model. We show that for the connections between the units in the network, this dictates a specific type of divisive inhibition. We then illustrate how one can compare this model to optimal inference with a HMM for the same setting. We present results from numerical simulations to assess the goodness of the approximation. We find that inhibition clearly improves performance close to optimal and we discuss the implications of our findings.

2 The generative model

We describe the statistical relationships between stimulus information in the environment (e.g. light impinging on the retina, acoustic stimulation of the ear, concentrations of molecules effecting the olfactory system) and their causes in a GM. Especially in the visual domain, objects may overlap in space and occlude each other. We model the consequences of this fact with a “noisy OR” relation as specified below in equation (1).

In this scenario, a set $\{H_j(t)\}_{j, j \in 1, \dots, N}$ of time dependent variables codes for the absence or presence of the corresponding hidden causes over time. If cause j is present at time t , we write $h_j(t) = 1$ and $h_j(t) = 0$ else.

As an organism does not have direct access to these causes, we model them via their influence on a set of M sensory channels $\{S_i(t)\}_{i, i \in 1, \dots, M}$ with $s_i(t) \in \{0, 1\}$.

The task of a sensory system can be thought to infer the probabilities $p(h_j^t | \mathbf{S}^{0 \rightarrow t}) := p(H_j(t) = h_j(t) | S_1(0 \rightarrow t) = s_1, S_2(0 \rightarrow t) = s_2, \dots, S_M(0 \rightarrow t) = s_m)$ of the

causes being present at time t given the history of inputs from all sensory channels $s_i := s_i(0 \rightarrow t)$.

To simplify the analysis, we assume that the system has learnt all the relevant statistics needed for this task (see [9] for how this can be done). These are the temporal dynamics of the hidden causes specified via the probabilities $p(h_j^t | h_j^{t-\Delta t})$ for cause j to switch from one state to the other (i.e. $0 \rightarrow 1$ or $1 \rightarrow 0$) and the *emission probabilities* $p(s_i^t = 1 | h_j^t)$ to receive an input in channel i corresponding to object j being present.

The ‘‘Noisy-OR’’ generative model now specifies the probability of observing no event in channel i as the probability of none of the present causes h_j (nor a background cause h_0) having caused it:

$$p(s_i^t = 0 | h^t) = (1 - p_0) \prod_j (1 - h_j^t p_{ij}) \quad (1)$$

where p_{ij} are the probabilities that object j causes an observation in pixel i . Furthermore, h_0 models the effect of background noise and other constant factors (i.e. we assume $h_0(t) = 1$ for all t).

This scenario translates nicely to the setting with continuous time in which the different input sources S_i can be described as switching Poisson processes with emission rates $\lambda_i(t)$ which are determined by the configuration $h := h(t) := [h_1(t), h_2(t), \dots, h_N(t)]$ of the hidden variables at time t . We now use the fact that for a Poisson process with rate λ the probability of observing one event in a short time interval Δt is approximately $\Delta t \lambda$. Therefore $p_0 \approx \Delta t \lambda_0(t)$ and $p(s_i^t = 1 | h^t) \approx \Delta t \lambda_i(t)$. Because we assume that the observations in channel i due to the presence of hidden cause j are given as $p_{ij} \approx \Delta t q_{ij}$ and $p_0 \approx \Delta t q_0$, this yields

$$p(s_i^t = 1 | h^t) \approx 1 - (1 - \Delta t q_0) \prod_j (1 - h_j^t \Delta t q_{ij}) \quad (2)$$

which can be written as

$$p(s_i^t = 1 | h^t) \approx \Delta t \left(q_0 + \sum_j h_j^t q_{ij} \right) + O(\Delta t^2). \quad (3)$$

This specifies the relation between the variables s_i^t which are observable by the system and the parameters q_{ij} describing the hidden causes.

2.1 Online inference for a single cause

Thanks to the probabilistic description in equation (3), it is possible to describe what the system can infer about the hidden state configurations h^t given the noisy observations $\mathbf{S}^{0 \rightarrow t}$. We use the log posterior probability ratio L_j^t to describe the process of inference for hidden cause j :

$$L_j^t := \log \frac{p(h_j^t = 1 | \mathbf{S}^{0 \rightarrow t})}{p(h_j^t = 0 | \mathbf{S}^{0 \rightarrow t})}. \quad (4)$$

The temporal dynamics $\dot{L}_j := dL_j^t/dt$ for this variable describing optimal online inference were derived in [10]. Let us assume a single cause H_j with switching rates r^{on} and r^{off} where r^{on} is defined as the limit of $p(h_j^t = 1 | h_j^{t-\Delta t} = 0)/\Delta t$ for $\Delta t \rightarrow 0$ and analogously for r^{off} . Furthermore, assume that the presence or absence of the cause determines the emission rates λ_i^{on} and λ_i^{off} of a set of Poisson channels \mathbf{S}_i that are conditionally independent given the state of the hidden cause. This means that the probability of observing an event in the different channels is fully described by the hidden state at a given point in time. Under these assumptions, the dynamics of L read:

$$\dot{L}_j = \Phi(L_j^t) + \sum_i w_i \delta(s_i^t - 1) - \Theta \quad (5)$$

with $w_i := \log(\lambda_i^{\text{on}}/\lambda_i^{\text{off}})$ and $\Theta := \sum_i (\lambda_i^{\text{on}} - \lambda_i^{\text{off}})$. $\delta(\cdot)$ is Dirac’s delta function and the function Φ

$$\Phi(L_j^t) := r^{\text{on}}(1 + e^{-L_j^t}) - r^{\text{off}}(1 + e^{L_j^t}) \quad (6)$$

describes the underlying birth-death process of the hidden cause, i.e. how quickly it changes its state.

2.2 Divisive inhibition for multiple causes

In a more realistic setting, observable events in a channel S_i could stem from different hidden causes H_j . This situation in which the *causal fields* of different hidden causes overlap is illustrated in figure 1. In this case, approximate

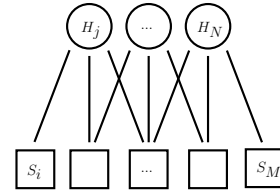


Figure 1. Generative Model for the scenario of multiple objects H_j being capable of causing sensory data in different channels S_i

inference can be implemented in a kind of mean field approach using the fact that $p_j^t := 1/(1 + e^{-L_j^t})$ and replacing the binary h_j by these probabilities. Using equation (3) one arrives at an approximate version of inference that accounts for the influence of neighboring hidden causes. The probabilities $p(s_i^t = 1 | h^t)$ for observing an event in the absence vs. presence of cause j (and possibly other causes) are now approximately given as

$$\Delta t(q_0 + \sum_{k \neq j} p_k q_{ik}) \quad \text{and} \quad \Delta t(q_0 + q_{ij} + \sum_{k \neq j} p_k q_{ik}) \quad (7)$$

With the abbreviation $A_{ij} := q_0 + \sum_{k \neq j} p_k q_{ik}$ referring to the influence of causes other than H_j , the probabilities of

observing NO event in absence vs. presence of cause j are respectively

$$1 - \Delta t A_{ij} \quad \text{and} \quad 1 - \Delta t (q_{ij} + A_{ij}). \quad (8)$$

This provides all the information necessary for the inference algorithm and in analogy to eq. (5) yields the time-discrete version

$$L_j^{t+\Delta t} \approx L_j^t + \Delta t \Phi(L_j^t) + \sum_i w_{ij} s_i^t + \sum_i b_{ij} (1 - s_i^t) \quad (9)$$

with $b_{ij} := \log\left(\frac{1 - \Delta t (q_{ij} + A_{ij})}{1 - \Delta t A_{ij}}\right)$ and the weights

$$w_{ij} := \log\left(\frac{q_{ij} + A_{ij}}{A_{ij}}\right). \quad (10)$$

Analogous to the single cell case, taking the limit of $\Delta t \rightarrow 0$ gives the continuous equation

$$\dot{L}_j^t = \Phi(L_j) + \sum_i w_{ij} \delta(s_i^t - 1) - \sum_i q_{ij}. \quad (11)$$

This realizes a specific type of *divisive inhibition*. For illustration assume $q_{ij} \ll q_0$. If we define the input weights W_{ij} as

$$W_{ij} = \log\left(\frac{q_0 + q_{ij}}{q_0}\right) \approx \frac{q_{ij}}{q_0} \quad (12)$$

equation 11 becomes

$$\dot{L}_j^t \approx \Phi(L_j) + \sum_i \frac{W_{ij}}{1 + \sum_{k \neq j} W_{ik} p_k^t} \delta(s_i^t - 1) - \sum_i q_{ij}. \quad (13)$$

In this type of divisive inhibition, the impact of an input depends on both its relation to the cause (via the weight W_{ij}) and how well other competing hypotheses can explain the incoming evidence (described by $\sum_{k \neq j} W_{ik} p_k^t$). Although equation (13) provides an intuitive understanding of the type of inhibition, it involves the additional approximation in equation (12). For the simulation we therefore use equation (9) with the weights defined in (10).

3 Numerical Evaluation

This model is easy to implement and can perform online inference in the scenario described earlier: input from multiple sensory channels is fed into a network of laterally coupled units. These inputs are spike trains and the network processes them according to equation (9). To assess the goodness of the approximations involved, we need to determine how well this task can be performed in principle. As both the dynamics of the hidden variables and the emission probabilities are known, this question can be answered using standard inference algorithms for HMMs.

3.1 Relation to optimal inference in Hidden Markov Models

We describe the relation between our scenario and HMMs for discrete time. A HMM is characterized via the set of states \mathcal{C} of its hidden process and their prior distribution π_i , the transition probabilities $T_{ij} := p(c_t = j | c_{t-1} = i)$ with $i, j \in \mathcal{C}$, the set of possible observations \mathcal{S} , and the probabilities $p(s_t = k | c_t = i)$ with $i \in \mathcal{C}$ and $k \in \mathcal{S}$. They give the probabilities of switching from one state to another and how probable it is to make an observation, given the hidden process is in a specific state at that time. As described in [11], there exist methods to estimate the hidden states given a sequence of observations and to learn the underlying parameters from such data. In particular, given an observed sequence of data s up to time t , one can compute the *forward probabilities* $p(c_t = i | s)$ that the hidden process is in state i at time t . Furthermore, the so-called Viterbi algorithm allows to efficiently determine the *most likely path*, i.e. the state sequence for which the joint probability of the observed data and estimated path is maximal.

The scenario described in section 2, i.e. several sensory channels $S_i(t)$ influenced by multiple possible causes $H_i(t)$ evolving over time, can be transformed into this setting: we assign a configuration c_i to each of the 2^N possible hidden states $\vec{H}_i := [h_1(t), h_2(t), \dots, h_N(t)]$ and, in the same way, each of the possible 2^M observable patterns $\vec{S}_i := [s_1(t), s_2(t), \dots, s_M(t)]$ is referenced by an index s_i . The corresponding transition and emission probabilities specifying the corresponding HMM are therefore fully determined by the generative model given in section 2.

3.2 Performance of divisive Inhibition

We can now assess, how well the network with divisive inhibition performs compared to a “naive” network without such competitive interaction and evaluate both networks with respect to what would be the optimal solution. To this end, we specify a generative model and sample sequences of hidden states and corresponding observations. We then compare, how well the divisive inhibition network reconstructs the hidden sequences and compare its performance with the results from the Viterbi algorithm and the forward probabilities. These latter two correspond to “optimal guesses”. But whereas the Viterbi solution takes into account all information, the forward algorithm uses only information up to time t .

We evaluated the reconstruction goodness for different parameters of the generative model. We sampled r^{on} and r^{off} uniformly from the interval (0.01, 0.05). The q_{ij} were either specified to be strongly overlapping circular Gaussians evenly covering $(0, 2\pi)$ under the constraint $q^{\text{min}} < q_{ij} < q^{\text{max}}$ with q^{min} sampled from (0.1, 0.3) and q^{max} from (1.5, 2.0). We also tested the system with “random” causal fields, i.e. the q_{ij} being sampled uniformly. The results for these two cases are very similar and we describe the results from 50 simulations with $M = 7, N = 5$

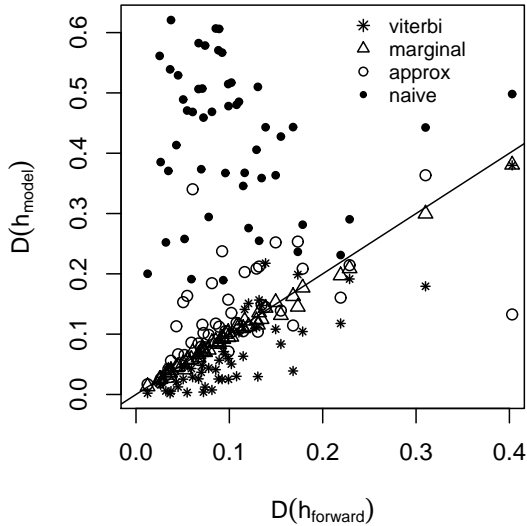


Figure 2. Goodness of reconstruction based on the forward probabilities (“forward”) vs. other models. The different models are “marginal” – marginal probabilities of the corresponding hidden cause being present computed from the forward probabilities, “approx” – the divisive inhibition network, “naive” – network based on equation (5) without divisive inhibition.

for $T = 1500$, $dt = 0.05$ with circular Gaussian q_{ij} .

To quantify our results, we use the Hamming distance between estimated state sequences $\hat{\mathbf{h}}$ and the true sequence \mathbf{h} , that is: $D(\hat{\mathbf{h}}) := \sum_{j,t} |h_j(t) - \hat{h}_j(t)|/NT$ where a sequence $\mathbf{h} := \{h_j(t)\}_{j=1,\dots,N;t=1,\dots,T}$ corresponds to the (estimated) states of all causes for all times. This distance measures the reconstruction goodness as it gives the mean probability that the system gets state $H_i(t)$ wrong. As our model (9) provides an estimate of the current posterior probabilities, it also allows for a comparison with measures accounting for the certainty of the state estimate (e.g. Kullback-Leibler divergence). This, however, does not hold for the Viterbi algorithm which focuses on the whole sequence rather than the posterior probabilities at individual times. As a first principled comparison we therefore stick with the Hamming distance as a simple measure of reconstruction goodness.

Figure 2 shows $D(\mathbf{h}_{\text{model}})$ for the different models plotted against $D(\mathbf{h}_{\text{forward}})$, the reconstruction based on the forward probabilities. $D(\mathbf{h}_{\text{forward}})$ provides a measure of how difficult the current sequence was: it represents how much information can be read off from data up to this point. If a model makes fewer errors, the corresponding points lie below the main diagonal.

Overall, the Viterbi algorithm performs best. This is expected as it uses information from the whole sequence to estimate the state at time t . The model named “marginal” results from projecting the forward probabilities from the

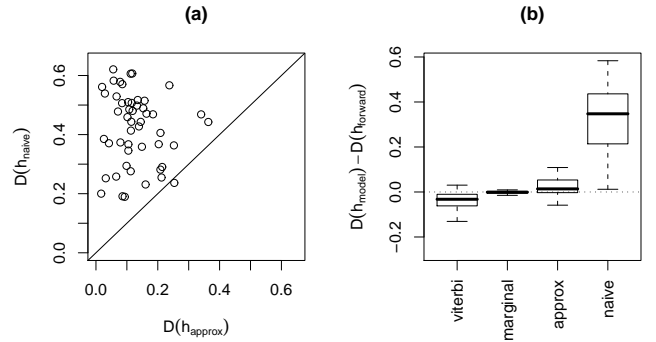


Figure 3. (a) Comparison of divisive inhibition (approx) with the naive network. (b) Difference between forward model and the other estimates.

2^N states to the N marginals and deciding for each state whether it is present or absent. Its predictions are very similar to the the most probable configurations ($\mathbf{h}_{\text{forward}}$). In comparison with the other algorithms, the divisive inhibition network ($\mathbf{h}_{\text{approx}}$) performs significantly better than the naive network ($\mathbf{h}_{\text{naive}}$) which does not account for overlapping causal fields. Although not optimal, its performance is comparable to the predictions derived from the forward probabilities. Figure 3(a) accentuates the comparison between the two network models by plotting $D(\mathbf{h}_{\text{naive}})$ against $D(\mathbf{h}_{\text{approx}})$. One can see that the naive network performs worse most of the time. Figure 3 further illustrates the performance of the different models. The box-plots show the distributions for the difference between the forward and the other models, i.e. how much worse the models are than the forward model.

4 Conclusion

We introduced a generative model accounting for the characteristic statistics found in natural environments. It represents the “noisy OR” relation describing ambiguous situations in which observations could stem from different causes. Based on this relation, we derived a specific type of divisive inhibition for neural networks to approximate optimal inference. We tested its performance in numerical simulations and our results make 2 points: (a) the proposed connectivity enables a good approximation of optimal online inference from spike trains. In this respect, the model provides a functional interpretation of lateral inhibition. Furthermore, it provides (b) an efficient way to perform online inference in the described setting.

Our study extends previous work on inference with single units [10] and the principle of output decorrelation [4] to the case of online processing with multiple units. This is necessary to deal with signal ensembles capturing the central statistical characteristics of realistic perceptual environments. In comparison to other approaches based on

generative models like ICA [6] or Maximal Causes Analysis [12], it differs in its focus on inference in dynamic environments.

While the current contribution deals only with inference, the model naturally incorporates an efficient mechanism to signal the online estimate with output spikes [10, 13]. In this way, the model accounts for further central properties of receptive field measurements like contrast dependent sharpening. Similar to [8], our approach anchors the model in the stimulus statistics but furthermore explains aspects of spike generation and temporal processing.

Acknowledgements: Thanks to Boris Gutkin for helpful discussions. This work was supported by grants from BACS (FP6-IST-027140) and BIND / Marie Curie (MECT-CT-2005-024831).

References

- [1] G. Hinton and Z. Ghahramani. Generative models for discovering sparse distributed representations. *Phil Trans R Soc Lond B*, 352(1358):1177–1190, 1997.
- [2] B. Olshausen and D. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.
- [3] M. Carandini, J. Demb, V. Mante, D. Tolhurst, Y. Dan, B. Olshausen, J. Gallant, and N. Rust. Do we know what the early visual system does? *Journal of Neuroscience*, 25(46):10577–10597, 2005.
- [4] M. Srinivasan, S. Laughlin, and A. Dubs. Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc Lond B Biol Sci*, 216(1205):427–459, 1982.
- [5] D. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–197, 1992.
- [6] A. Bell and T. Sejnowski. The “independent components” of natural scenes are edge filters. *Vision Research*, 37(23):3327–3338, 1997.
- [7] E. Simoncelli and D. Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38(5):743–761, 1998.
- [8] O. Schwartz and E. Simoncelli. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8):819–825, 2001.
- [9] G. Mongillo and S. Denève. On-line learning with hidden Markov models. *Neural Computation*, 20:1706–1716, 2008.
- [10] S. Denève. Bayesian spiking neurons I: Inference. *Neural Computation*, 20:91–117, 2008.
- [11] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *proceedings of the IEEE*, vol 77, pages 257–286, 1989.
- [12] J. Luecke and M. Sahani. Maximal causes for non-linear component extraction. *Journal of Machine Learning Research*, in press.
- [13] T. Lochmann and S. Denève. Information transmission with spiking Bayesian neurons. *New Journal of Physics*, 10:article id: 055019, 2008.