New Journal of Physics
The open-access journal for physics

Information transmission with spiking Bayesian neurons

Timm Lochmann and Sophie Denève[1]
Group for Neural Theory, Département d'études cognitives, Ecole Normale Supérieure, 775005 Paris, France
E-mail: sophie.deneve@ens.fr

New Journal of Physics **10** (2008) 055019 (19pp)
Received 26 September 2007
Published 29 May 2008
Online at http://www.njp.org/
doi:10.1088/1367-2630/10/5/055019

Abstract. Spike trains of cortical neurons resulting from repeated presentations of a stimulus are variable and exhibit Poisson-like statistics. Many models of neural coding therefore assumed that sensory information is contained in instantaneous firing rates, not spike times. Here, we ask how much information about time-varying stimuli can be transmitted by spiking neurons with such input and output variability. In particular, does this variability imply spike generation to be intrinsically stochastic? We consider a model neuron that estimates optimally the current state of a time-varying binary variable (e.g. presence of a stimulus) by integrating incoming spikes. The unit signals its current estimate to other units with spikes whenever the estimate increased by a fixed amount. As shown previously, this computation results in integrate and fire dynamics with Poisson-like output spike trains. This output variability is entirely due to the stochastic input rather than noisy spike generation. As a result such a deterministic neuron can transmit most of the information about the time varying stimulus. This contrasts with a standard model of sensory neurons, the linear–nonlinear Poisson (LNP) model which assumes that most variability in output spike trains is due to stochastic spike generation. Although it yields the same firing statistics, we found that such noisy firing results in the loss of most information. Finally, we use this framework to compare potential effects of top-down attention versus bottom-up saliency on information transfer with spiking neurons.

[1] Author to whom any correspondence should be addressed.

Contents

1. Introduction

The sensors of a system enable it to perceive its environment. They empower organisms to estimate quantities like light intensity or orientation of an edge and to detect the presence of dangerous or attractive events—they inform the organism about events of relevance for making decisions and performing actions. This is a real challenge because to be useful in realistic environments such processes have to be performed online and with reasonable speed and precision. Signalling within and between different brain regions typically involves electrical impulses called spikes. For example, presentation of a visual stimulus to the eye results in a bombardment of thousands of incoming spikes to neurons in the primary visual cortex. Typically, these visual neurons will only fire a few spikes in response. How is the information about different features in the changing visual inputs contained in thousands of input spikes processed and compressed in a few output spikes?

In this study, we examine the capacity of single neuron models to (a) extract an estimate of a dynamic stimulus from its noisy synaptic input and (b) transmit this estimate with its output spike train. Furthermore, we assess whether their output mimics the strong Poisson-like variability typically found for empirically measured cortical responses [1].

We will approach this issue using ideas from information theory [2] which provides generic tools to quantify communication characteristics. In the case of neural signalling, this has revealed that specific sensory systems (e.g. cells in the visual system of the blowfly) can encode and transmit information surprisingly well [3]–[5]. Maximizing the transfer of information (INFOMAX, see [6]) while limiting the cost of the code (e.g. the number of spikes) provides a rational perspective to understand different aspects of neural processing, ranging from classical neural networks to receptive fields and learning [6]–[8].

Many studies in this line [9]–[11] used sets of static stimuli (e.g. direction of wind or orientation) for which a system can simply accumulate all the incoming sensory data over time to make a decision. More recent studies [5], [12]–[14] have taken into account the dynamic nature of realistic environments: they assumed that stimulation results from simple stochastic processes with specified temporal dynamics. For example, let the stimulus be the orientation of a grating whose angle can change over time. It is intuitively clear that to estimate the current orientation from noisy data, one can no longer simply accumulate all incoming evidence. There is a payoff between the benefit of assembling data and the cost of waiting too long because the state might change in the meantime.

Such approaches, however, focused on the question of how to optimally estimate a stimulus from observable evidence [15]–[17]. While they showed how to determine the posterior probability of the stimulus given the input history, the question of how the extracted information can optimally be encoded in output spike trains was not their main interest. Spikes were randomly generated according to a Poisson process with an output rate depending on this posterior probability. It is important to note that although this specifies the expected number of events, their timing is left uncertain. Adopting such an intrinsic noise factor represents a strong model assumption whose consequences have to be assessed, e.g. it probably impairs information transmission. Furthermore, spike generation in biological neurons can be rather unlike such a random generator [18]—it is often more reliable and better described as a deterministic threshold or bifurcation mechanism [19]–[21].

This discrepancy motivated us to investigate how extrinsic versus intrinsic sources of randomness effect information transmission in spiking neurons. We ask how much information is lost from input to output spike trains if spikes are generated by either a stochastic mechanism or by a deterministic threshold mechanism. This is of great interest as stochastic spike generators have been commonly used in the past, e.g. the inhomogeneous Poisson component in the well known linear–nonlinear Poisson (LNP) model [17, 22, 23].

We focus on a model of neurons coding for a *binary* variable [24]. This variable could correspond to the presence of a preferred stimulus (an edge with a particular orientation, sound at a particular frequency). The model assumes that neural units integrate incoming evidence over time in an optimal fashion and includes a *deterministic* output mechanism similar to a leaky integrate and fire (LIF) process.

We compare this to models that are based on the same optimal estimate of the hidden state (i.e. synaptic integration is identical) thus differing only in the way this estimate is transferred into an output stream (i.e. deterministic versus stochastic firing). Our comparisons make use of a new method to estimate a lower bound on how much information about the hidden state is contained in the preceding spike train history. This method is fast and tractable and we show its derivation from the underlying statistical model.

We demonstrate that the deterministic neuron can produce output spike trains with statistical properties matching those of stochastic spike generators. Our results, however, clearly indicate that the integrate and fire process is much more efficient in transmitting information. They, furthermore, illustrate the different effects of changing a units excitability and changing the rate of *input* to this unit. Although both factors affect the output firing rate, their impact on information transfer is quite different. We conclude by pointing out the implications of this model for understanding the effects of saliency and top-down attention on neural activity.

2. Bayesian spiking neurons

To account for the dynamic nature of the environment, we assume the inputs to the model system to be the observable of a hidden Markov model (HMM): the value of a state variable Θ is *hidden*, i.e. cannot be observed directly. However, indirect external evidence is available in form of an *observable* S, which is a random variable whose distribution is fully specified by Θ. In the case of perceptual inference, the state might refer to the hidden cause of an event and the observable evidence to the sensory data it causes.

More precisely, let $\mathbf{\Theta}$ denote a finite set representing the state space of a hidden variable and $\{\Theta_t\}_{t\in\mathbf{T}}$ be a stochastic process with t denoting a time index (e.g. $\mathbf{T}=\mathbb{Z}$ for discrete time or $\mathbf{T}=I\subset\mathbb{R}$ for continuous time).

Furthermore, let $\mathbf{S}$ be a finite set of discrete events and $\{S_t\}_{t\in\mathbf{T}}$ with $S_t\in\mathbf{S}$ for all t denote another stochastic process. Let $S_{(a,b)}:=\{S_a,\ldots,S_b\}$ denote the history of observations from time a to b and $S_{(t)}:=S_{(0,t)}$. Realizations of the random variables will be denoted in the same way using small type.

A HMM [25] is a pair of discrete time stochastic processes $\{(\Theta_t,S_t)\}_t;\ t\in\mathbf{T}$, where the hidden variable Θ_t has the Markov property and the distribution of the observable S_t depends only on Θ_t. This means for all $t_1,t_2\in\mathbf{T}$ that given the state sequence $\{\Theta_t\}_{t\in\mathbf{T}}$ observations S_{t_1} and S_{t_2} are conditionally independent.

In the scenario, we are discussing, both the hidden state Θ and S are binary variables, i.e. $\mathbf{\Theta}=\mathbf{S}=\{0,1\}$. For Θ, the two states will also be called the off-state and the on-state and for S, 0 versus 1 stands for the presence or absence of an input event, e.g. an incoming spike from a presynaptic neuron.

We start from a HMM with a given time step Δ, i.e. $t=0,\Delta,2\Delta,\ldots,T$. At $t=0$ the probability of the hidden state is set as $p(\Theta_0=1):=p_{t_0}$. After that, the probability of Θ_t is determined by the transition probabilities $p^\Delta_{\mathrm{on}}:=p(\Theta_t=1|\Theta_{t-\Delta}=0)$ and $p^\Delta_{\mathrm{off}}:=p(\Theta_t=0|\Theta_{t-\Delta}=1)$ as

$$p(\Theta_t=1)=p^\Delta_{\mathrm{on}}p(\Theta_{t-\Delta}=0)+(1-p^\Delta_{\mathrm{off}})p(\Theta_{t-\Delta}=1), \tag{1}$$

for $t=\Delta,2\Delta,\ldots$. Furthermore, the probabilities of observing an event are given as $Q^\Delta_{\mathrm{on}}:=p(S_t=1|\Theta_t=1)$ and $Q^\Delta_{\mathrm{off}}:=p(S_t=1|\Theta_t=0)$, where $S_t=1$ is a shortcut denoting the fact that an event is observed in the interval $(t,t+\Delta)$. For the continuous time process in the limit of $\Delta\to 0$, the corresponding switching rates r_{on} and r_{off} describe how quickly Θ changes and are given as

$$r_{\mathrm{on}}:=\lim_{\Delta\to 0}\frac{p^\Delta_{\mathrm{on}}}{\Delta}\qquad\text{and}\qquad r_{\mathrm{off}}:=\lim_{\Delta\to 0}\frac{p^\Delta_{\mathrm{off}}}{\Delta}. \tag{2}$$

Similarly, the corresponding emission rates q_{on} and q_{off} are

$$q_{\mathrm{on}}:=\lim_{\Delta\to 0}\frac{Q^\Delta_{\mathrm{on}}}{\Delta}\qquad\text{and}\qquad q_{\mathrm{off}}:=\lim_{\Delta\to 0}\frac{Q^\Delta_{\mathrm{off}}}{\Delta}. \tag{3}$$

In the continuous time case, the observables are still conditionally independent given the state sequence and are modeled as Poisson processes whose instantaneous rates q_{off} and q_{on} are set by the state of the hidden variable. We will call this process describing the dynamics of the environmental dependencies the *generative model*. We will now discuss how neuron-like units can perform inference about the hidden state given the observable process. Figure 1(a) provides a graphical representation of the generative model (the HMM) and an illustration of

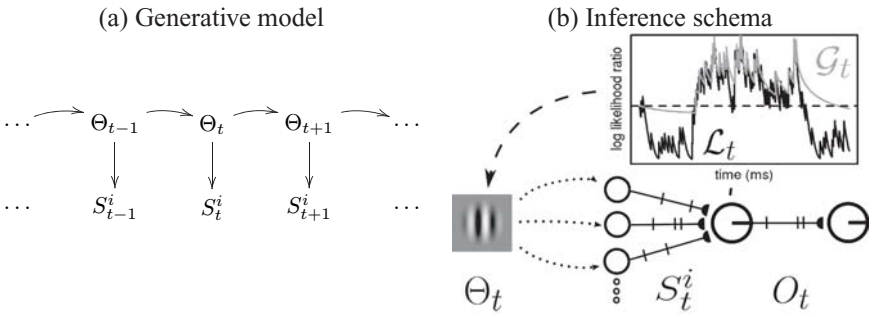

Figure 1. (a) Graphical representation of the underlying HMM. The observables S_t^i come from a switching Poisson process (SPP) whose instantaneous rate is determined by the hidden state Θ_t. (b) Schematic illustration of the inference task. Details are given in the text.

the inference task. Although the model can account for multiple input trains S_t^i, for conceptual simplicity we will present only the case of a single input train, i.e. $i = 1$. All results presented here also hold true for several input spike trains.

2.1. Inference

Let us assume that the dynamics of a single neuron-like unit can be understood as performing inference about some property in the world. As a more concrete example, we will assume that its dynamics stand in some direct relation to whether a feature of interest is present or absent. As the unit does not have direct access to this feature but receives noisy evidence in terms of incoming spikes, it proves appropriate to model this relation as a process of estimation or inference. We model the dynamics as an estimator for the (probability of the) feature being present versus absent as a function of the synaptic input the unit receives and the dynamics of the hidden process. This is the scenario illustrated in figure 1(b): the Poisson processes S^i triggered by the hidden state Θ act as the noisy inputs to a unit performing inference about the presence of Θ. In the following, we will use the shortcuts for the posterior probability $p_1 := p(\Theta_t = 1 | S_{(t)} = s_{(t)})$ and similarly for p_0. Furthermore, we will use the *logit* or *log odds* $\mathcal{L} := \mathcal{L}(t) := \log \frac{p_1}{p_0}$ of the posterior probabilities of being in state 0 or 1. The variable $\mathcal{L}$ is a function of the posterior probability of Θ given all inputs $S_{(t)}$ up to time t. It can therefore be understood as the units current knowledge about the state Θ.

It was shown in [26] how to derive the time continuous process $\mathcal{L}$ and its temporal dynamics $\dot{\mathcal{L}} = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{L}(t)$ for a set $\mathbb{I}$ of observable processes $S^i := S^i(t)$, $i \in \mathbb{I}$ that are conditionally independent given the state of Θ_t. As described in the previous section, they are assumed to be Poisson processes with rates q^i_{off} and q^i_{on} specified by the current state Θ_t. The resulting differential equation reads:

$$\dot{\mathcal{L}} = r_{\mathrm{on}}\left(1 + \mathrm{e}^{-\mathcal{L}}\right) - r_{\mathrm{off}}\left(1 + \mathrm{e}^{\mathcal{L}}\right) + \sum_{i \in \mathbb{I}} w^i \delta(s^i - 1) - \Psi, \tag{4}$$

where the weights for incoming spikes are given as $w^i := \log(q^i_{\mathrm{on}}/q^i_{\mathrm{off}})$ and $\Psi := \sum_{i\in\mathbb{I}}(q^i_{\mathrm{on}} - q^i_{\mathrm{off}})$ is a constant.

As can be seen from this equation, the inference process is fully parametrized by the pair of switching rates (r_{on}, r_{off}) and the emission rates (q^i_{on}, q^i_{off}) defined in (2) and (3). The central

point here is that this dynamics performs an optimal inference if the inputs conform with the model assumptions and the parameters (i.e. switching rates, weights and Ψ) match their true values [26]. Equation (4) specifies how evidence s_i accumulating over time should optimally be taken into account and we use this relation as a model of synaptic integration.

So far, no statement about the values of the parameters for the inference process have been made. They can be learned by standard algorithms for HMM (see e.g. [25]). Furthermore, we have shown previously that these parameters can be estimated in an online fashion, making it possible to incrementally learn them from sequentially acquired data [27]. As we focus on transmission in a well-parameterized system, we assume for the following discussion and the simulations that the true values have been successfully learned.

2.2. Output production

In [24, 26], a deterministic mechanism was suggested for transforming $\mathcal{L}$ into an output process O_t resembling the inputs S^i that determine $\mathcal{L}$. The mechanism is based on a second variable $\mathcal{G}$ with dynamics similar to those of $\mathcal{L}$:

$$\dot{\mathcal{G}} = r_{\text{on}}\left(1+\mathrm{e}^{-\mathcal{G}}\right) - r_{\text{off}}\left(1+\mathrm{e}^{\mathcal{G}}\right) + \eta\delta(O_t - 1), \tag{5}$$

where $O_t = 1$ when $\mathcal{L} > \mathcal{G} + \frac{\eta}{2}$ and zero otherwise. The resulting $\{O_t\}_{t\in\mathbf{T}}$ is again a point process and will be used to model signalling behaviour of this unit. Due to this construction, $\mathcal{G}$ is a prediction of the state Θ governed by the parameters r_{on}, r_{off} and η. It is updated every time an output event is produced and represents the estimate that is signalled to downstream units via the output spike train $\{O_t\}_{t\in\mathbf{T}}$. While the switching rates r_{on} and r_{off} parametrize how quickly this estimate is assumed to change (see section 2), η determines how much $\mathcal{L}$ and $\mathcal{G}$ have to differ to evoke an output event. Because $\mathcal{L}$ is the units estimate of the state and $\mathcal{G}$ the estimate it signalled via its output spike train, η sets the accuracy with which $\mathcal{L}$ is signalled. The size of this threshold regulates how many output events are produced in response to a given input train. We will therefore say that η controls the signal 'compression' for a unit. The resulting *output* rates for this unit during the off and on states will be denoted λ_{off} and λ_{on}, respectively.

It is important to note here that once the inputs $S^i(t)$ and the parameters of the inference process are given, this output is fully determined by the value for η and contains no stochastic component. As described in more detail in [26], this mechanism is similar to a LIF model (see e.g. [19]) with a membrane potential equal to $V := \mathcal{L} - \mathcal{G}$: the synaptic inputs are integrated with a 'leak' (controlled by the transition rates r_{on} and r_{off}). A spike is fired when the threshold $\eta/2$ is reached, at which time the membrane potential is reset to $-\eta/2$. Figure 2(a) illustrates the output mechanism and (b) its similarity to a LIF model. We will examine how much more information can be transmitted using the TB process by comparing it to similar models with stochastic output production. These stochastic models are based on exactly the same inference process $\mathcal{L}$ and differ only in signalling, i.e. the way the internal estimate $\mathcal{L}$ is transferred into a sequence of output events.

The inhomogeneous Poisson process (IPP) forms a part of the LNP model [22] which has been suggested as a model for firing behaviour of sensory neurons and is widely used to account for the stochasticity of the neural response. Its instantaneous firing rate $\lambda(t)$ is determined by an internal estimate of the hidden state computed from $\mathcal{L}(t)$. We use the corresponding posterior probabilities $p_1 = [1+\exp(-\mathcal{L})]^{-1}$ and $p_0 = 1 - p_1$ of the state being on or off to determine this rate as $\lambda(t) = \beta_i(p_0(t)\lambda_{\text{off}} + p_1(t)\lambda_{\text{on}})$—i.e. we interpolate between the rates resulting from

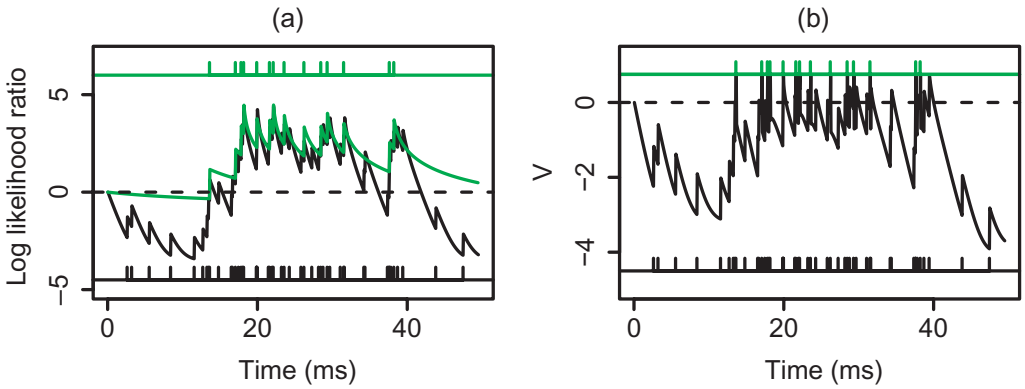

Figure 2. Illustration of the threshold-based (TB) mechanism for output generation. (a) Temporal evolution of $\mathcal{L}$ (black line) and $\mathcal{G}$ (green line) on an example trial. (b) Same trial re-plotted for a 'membrane potential' $V := \mathcal{L} - \mathcal{G}$ (black line) and a fixed threshold (green line). This neuron behaves as a LIF model. The lower black horizontal line indicates the input spike train S_t, the upper green line the output spike train O_t (see also [26]).

the threshold mechanism. For the simulations, we choose β_i such that the expected number of output events generated by this process equals the one obtained by the threshold mechanism. In spite of its conceptual shortcomings, similar models have often been used due to their analytical tractability [15, 22, 23].

Another process is a simplistic model for unreliable synaptic transmission (UST) of input events: outputs are produced only at times of input events and the probability, whether such an event at time t is transferred or filtered out by the unit depends on $\mathcal{L}(t)$. The transmission probability p_{tr} for an input event is given as $p_{\text{tr}} = \beta_{\text{f}}[1 + \exp(-\mathcal{L})]^{-1}$ and assumes that events are transmitted with a higher probability if the unit estimates the hidden state to be 1 rather than 0. β_{f} was again chosen such that the expected numbers of output events matches the one for the TB mechanism.

In the last model, we consider is a SPP with the same rates λ_{off} and λ_{on} as the output train from the TB mechanism in the off and on state. Note that this mechanism assumes knowledge of the hidden state Θ. It is therefore useful only as a simple conceptual model to compare against but cannot be considered as a realistic model for neural information transfer.

2.2.1. Comparison of firing statistics for different firing mechanisms. Figure 3(a) illustrates the strategy to compare the different output mechanisms. Output rate decreases with increasing compression parameter (i.e. increasing η for TB, decreasing mean emission rate for IPP and decreasing mean p_{tr} for UST) for all mechanisms and the chosen parameter settings result in similar output rates for the different mechanisms.

For the following simulations, we therefore used the same 'compression' when comparing the transmission capacity of different models. These methods make the models as similar with respect to their output rates as possible and leave the specific way of how spikes are generated as the main difference between them.

Figure 3(b) shows the output rates as a function of the input firing rates. Here, the input rates q_{on} and q_{off} are scaled by a constant factor α. Lines refer to the different output mechanisms

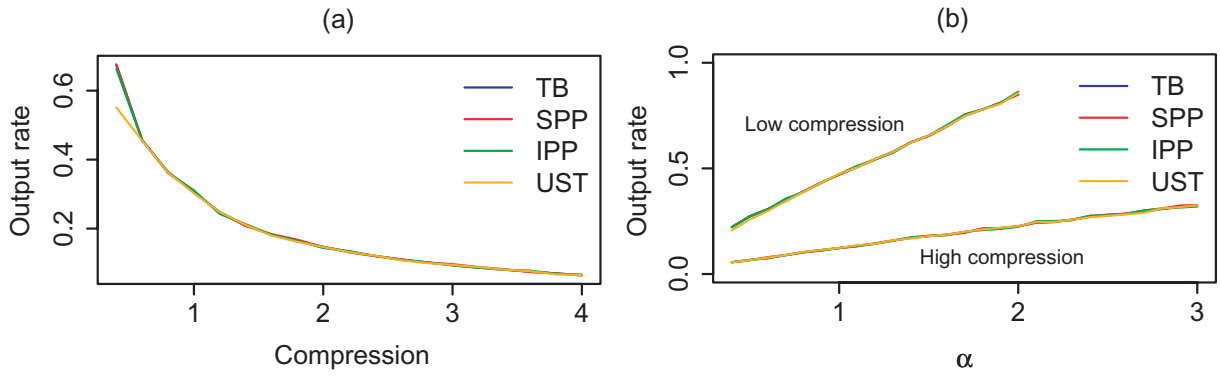

Figure 3. Comparison of output mechanisms. (a) Output rates resulting for the different mechanisms for parameter settings specifying equal compression rate. (b) Output rate increases linearly with increasing input rate. Shown are results for low compression ($\eta = 0.6$, upper lines) and high compression ($\eta = 2.4$, lower lines). α ranged from 0.4 to 2 for low and from 0.4 to 3 for high compression.

and the results for two different values of compression are shown. Note that this relation is approximately linear over a wide range of scalings and is the same for all spike generation mechanisms.

If these units are to be used as generic building blocks of a neural inference machine, their output should conform to the assumptions made by other downstream units receiving their output signals as input. For the current setting, this means that the output trains should conform to the properties of a SPP. Figures 4(a) and (b), top row, show the empirical cumulative distribution functions (ecdf) resulting from the TB mechanism during the on-state and the off-state, respectively. Plain lines show the ecdfs resulting for different values of η, while the dashed lines depict the theoretical predictions for the exponential distribution characteristic of a Poisson process with the same rate. One can see that for small η, the output ISI distribution contains more short ISIS than predicted by the exponential distribution of the same rate (see dashed lines); the process is more 'bursty'. On the contrary, for large η, there are less short ISIs than expected for a Poisson process: spike trains are more 'regular' than expected. For intermediate values of η, the output processes are all similar to the Poisson process. In all cases, however, the deviation from a 'pure' Poisson process is small and largely within the range of experimentally observed output distributions [1, 28].

Figures 4(c) and (d) show the same statistics for the different alternative models equalized with respect to number of output spikes produced by the TB mechanism at $\eta = 2$. As for the threshold mechanism, their ISI distribution is close to the exponential distribution actually assumed by the inference mechanism. We conclude from figures 3 and 4 that the threshold mechanism and the stochastic firing models used for comparison are well matched with respect to firing rate and ISI distribution of their output spike trains.

3. Methods for measuring information transfer

The previous section suggests how one can divide the problem of optimal information transmission into two parts: inference and transmission. While the first is about extracting an accurate estimate of the (probability of the) stimulus from the noisy synaptic input, the second

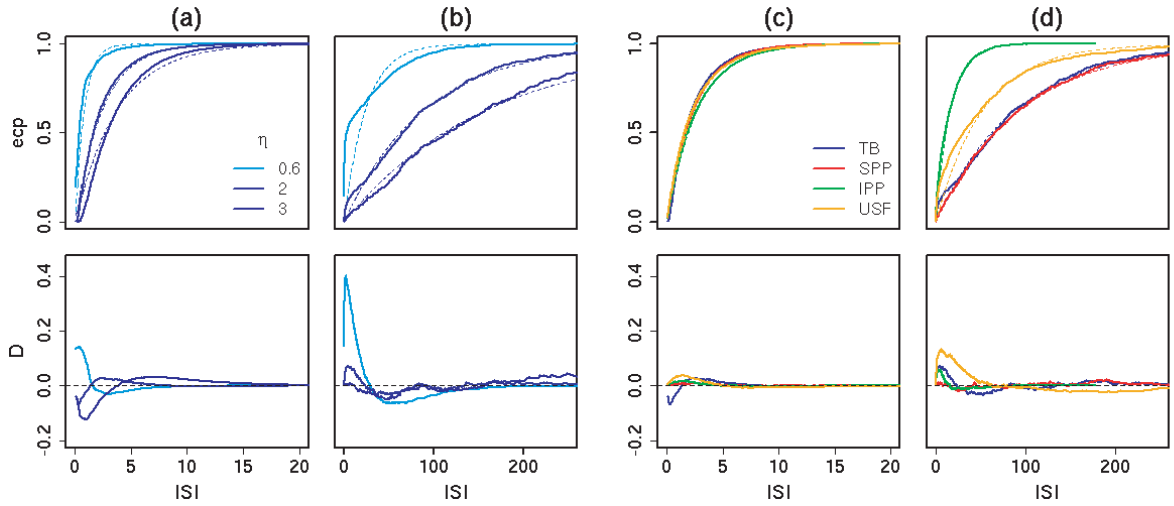

Figure 4. Ecdf for the Bayesian spiking neuron. Top row: ecdfs for three different values of η (plain lines) and cumulative distribution function (cdf) for the exponential distributions with the same output rate (dashed lines) in the on-state (a) and the off-state (b). Y-axis in the top row show the empirical cumulative probability (ecp) for the corresponding ISI given on the x-axis. (c) and (d): same as (a) and (b) but for the different output models corresponding to $\eta = 2$. Bottom row: differences between the ecdf and the cdfs shown in the top row. Parameter settings were $r_{\text{off}} = 0.05$, $r_{\text{on}} = 0.03$, $q_{\text{off}} = 0.5$ and $q_{\text{on}} = 1.5$. ISI lengths are given in ms and the graphs are based on virtual 500 s with $\Delta = 0.05$ ms.

part asks how this estimate can be signalled in an optimal fashion. We will show that it is useful to distinguish the two different variables about which information can be transmitted: the binary *hidden state* Θ and the hidden state *estimate*, measured by the logit function $\mathcal{L}$. We now describe a way to estimate how much information about these two variables is contained in the outputs of the model and how we can use these estimates to characterize processing in a single neuron.

3.1. General protocol

The method we use to estimate information transfer is summarized in figure 5. A single unit receives the observables (input spikes) S_t from a hidden Markov process as inputs. The dynamics of its internal estimate $\mathcal{L}$ can be interpreted as performing optimal inference about the hidden state Θ (see equation (4)). Based on this estimate, depending on the specific output mechanism described in section 2.2, an output spike train O_t is produced. For the TB mechanism, this is determined by equation (5).

From this output spike train, we can extract another estimate of the hidden state, $\hat{\mathcal{L}}$, by again treating O_t as observations from another HMM with the same hidden state sequence θ. This will not correspond to exact inference, since the output spike train is not, strictly speaking, a Poisson process with a switching rate (except for SPP). However, this is not a bad approximation of the output statistics, as shown in the previous section. Note that whereas the probability of the observable S_t given the state (i.e. q_{on} and q_{off}) are known, this is not the case for the output process O_t. However, in the present case, we can determine the missing parameters, λ_{on} and

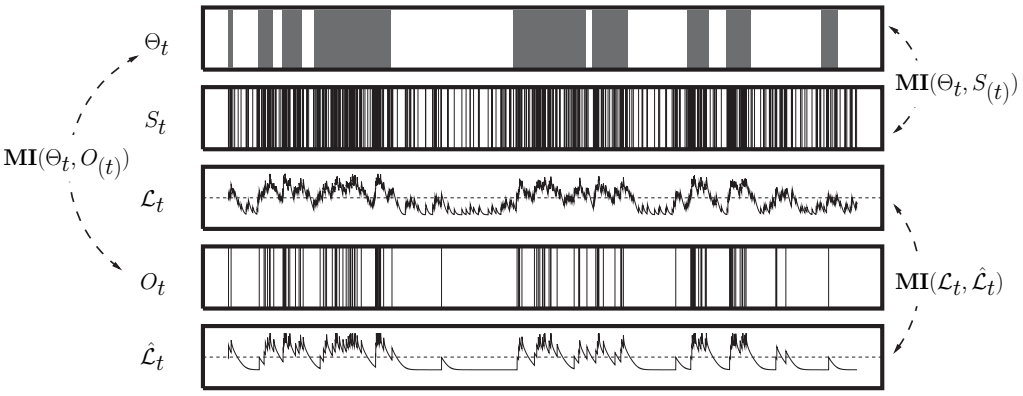

Figure 5. Illustration of the variables used to estimate information transfer. From top to bottom, a realization sequence of the hidden state Θ_t, the input train S_t, the internal estimate $\mathcal{L}_t$ the resulting output train O_t and the readout estimate $\hat{\mathcal{L}}_t$ are shown.

λ_{off}, directly by counting the number of output events when Θ is 0 or 1. $\hat{\mathcal{L}}$ is obtained from equation (4) by replacing q_{on}, q_{off} by λ_{on}, λ_{off} and S_t by O_t.

Next, we use the Monte Carlo estimator (see next section) to measure the information about Θ_t contained in the observed history $S_{(t)}$ up to time t, denoted $MI(\theta_t, S_{(t)})$. In the same way, we use the Monte Carlo estimator to compute (lower bound on) the information contained in the output spike train $O_{(t)}$, denoted $MI(\theta_t, O_{(t)})$. Comparing these two quantities, we can assess how efficiently information is transmitted by this unit. In the next section, we describe how we estimate these quantities using the internal estimates $\mathcal{L}$, computed from the input spikes and $\hat{\mathcal{L}}$, obtained from the output spikes.

Besides the primary question of how much we can learn from the observed history about the *current state* we will also discuss how much information is transmitted about the internal estimate $\mathcal{L}$. This second question is crucial because it may be as important to communicate the current knowledge about Θ as Θ itself. Transmitting small fluctuations in probability with high fidelity may be important for later processing stages (see section 5). We approximate the mutual information between the two continuous random variables $\mathcal{L}_t$ and $\hat{\mathcal{L}}_t$ by quantizing these variables. We checked the convergence behaviour for increasing number of bins (data not shown) and note that other measures like correlation or mutual information using adaptive binsizes [29] yield very similar results.

Next, we describe the Monte Carlo sampling method used to estimate the information about Θ contained in the input and output spike trains.

3.2. An estimator based on Monte Carlo sampling

For two discrete random variables X and Y with probability mass functions $P(X)$ and $P(Y)$, the entropy $H(X)$ is defined as

$$H(X) := -\sum_{x \in \mathcal{X}} p(X = x) \log p(X = x). \tag{6}$$

The mutual information then can be defined as:

$$I(X, Y) := H(X) - H(X|Y), \tag{7}$$

where $H(X|Y)$ is the conditional entropy

$$H(X|Y) := \sum_{y\in\mathcal{Y}} p(Y=y)H(X|y) \tag{8}$$

and $H(X|y)$ is the entropy of the conditional distribution $P(X|Y=y)$. Applying these concepts to estimate the mutual information between the hidden state and its preceding observation history, a straightforward approach might be to compute the mutual information $I(\Theta_t, S_{(t-\Delta,t)})$ between sequences of observables and the hidden state for increasing Δ. In the end, we are interested in how much information can be decoded from the whole history preceding time t, i.e. $S_{(t)}$.

As these sequences become increasingly long, this poses a problem when estimating this quantity numerically as we need to keep track of binary sequences of length n whose number increase exponentially. Unfortunately, estimating the mutual information using formula (7) gives a systematically biased information estimate if the sample is too small to cover the space well [29, 30]. Due to the exponential growth of possible sequences, assembling large enough samples to cover the space of possible sequences sufficiently is impractical even for simulation studies.

In the present case, we can use the Markov property and the finite temporal dependencies between the observations (due to the nonzero switching rates r_{on} and r_{off}) to directly estimate the mutual information contained in the spike trains using Monte Carlo sampling.

The mutual information between the state Θ_t and its preceding observation history $S_{(t)}$ can be written as

$$\begin{aligned} I(\Theta_t, S_{(t)}) &:= \sum_{\theta_t\in\mathbf{\Theta}_t}\sum_{s_{(t)}\in\mathbf{S}_{(t)}} p(\Theta_t=\theta_t, S_{(t)}=s_{(t)}) \log \frac{p(\Theta_t=\theta_t, S_{(t)}=s_{(t)})}{p(\Theta_t=\theta_t)p(S_{(t)}=s_{(t)})} \\ &= H(\Theta_t) + \sum_{\theta_t, s_{(t)}} p(\Theta_t=\theta_t, S_{(t)}=s_{(t)}) \log p(\Theta_t=\theta_t|S_{(t)}=s_{(t)}) \\ &= H(\Theta_t) - H(\Theta_t|S_{(t)}). \end{aligned}$$

We can now rewrite the conditional entropy $H(\Theta_t|S_{(t)})$ as:

$$\begin{aligned} H(\Theta_t|S_{(t)}) &= -\sum_{\theta_t, s_{(t)}} p(\Theta_t=\theta_t, S_{(t)}=s_{(t)}) \log p(\Theta_t=\theta_t|S_{(t)}=s_{(t)}) \\ &= -\left\langle \log p(\Theta_t=\theta_t|S_{(t)}=s_{(t)})\right\rangle_{P(\mathbf{\Theta}_t, S_{(t)})}. \end{aligned}$$

Where the angular brackets $\langle\cdot\rangle_P$ denote the average over the distribution P.

Using the fact that $\hat{p}(\theta_t=1|s_{(t)}) := [1+\exp(-\mathcal{L})]^{-1}$ is the posterior probability of the state, we define the estimator of the conditional entropy as:

$$\hat{H}(\Theta_t|S_{(t)}) = -\left\langle \log \hat{p}(\Theta_t=\theta_t|S_{(t)}=s_{(t)})\right\rangle_{P(\mathbf{\Theta}_t, S_{(t)})}. \tag{9}$$

Although the estimated posterior probability $\hat{p}(\theta_t=1|s_{(t)})$ is an optimal estimate and involves no approximation, we use the notation $\hat{p}(\cdot)$ here to distinguish between the probability estimate resulting from sampling (by drawing sequences from the HMM) and the estimated posterior probability $\hat{p}(\theta_t=1|s_{(t)})$ obtained via $\mathcal{L}$.

As $H(\Theta_t|S_{(t)})$ is the same for all t and the dependence of the estimate on past observations decreases quickly with time (data not shown) we take the mean over a long sequence of the HMM instead of averaging over multiple histories. Estimating the entropy of the marginal

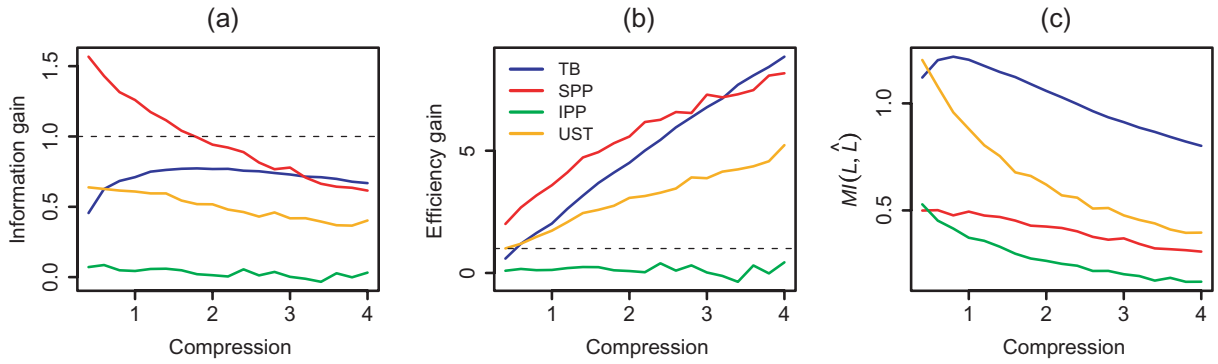

Figure 6. Monte Carlo estimates of (a) gain of information about Θ contained in the output trains compared to the input train, (b) gain in encoding efficiency and (c) mutual information between $\mathcal{L}$ and $\hat{\mathcal{L}}$.

distribution of Θ_t is unproblematic as the relative frequencies for Θ being 0 or 1 converge quickly to their asymptotic values.

Although the output processes O_t are not guaranteed to be Poisson, we use the same estimator $\hat{p}(\theta_t|o_{(t)})$ based on $\hat{\mathcal{L}}$ as an approximation to the true posterior probability $p(\theta_t|o_{(t)})$ to get an estimate $\hat{H}(\Theta_t|O_{(t)})$ of the conditional entropy. Due to Jensen's inequality, this estimate is an upper bound on the true entropy. Thus, the mutual information measured using the Monte Carlo sampling method is a lower bound on the information truly contained in the output spike train.

4. Information efficiency of the threshold and the stochastic mechanisms

We will use two quantities to characterize information transfer: the *information gain* and the *efficiency gain*. They are defined as follows:

$$\text{information gain: } \frac{MI(\Theta, O_{(t)})}{MI(\Theta, S_{(t)})}, \qquad \text{efficiency gain: } \frac{MI(\Theta, O_{(t)})/\bar{\lambda}}{MI(\Theta, S_{(t)})/\bar{q}}, \tag{10}$$

where $\bar{q}$ and $\bar{\lambda}$ are the mean input and output firing rates of the unit. Whereas the first quantity measures the absolute *proportion* of information still contained in the output train, the latter allows to assess how much more efficiently this information is encoded in the output spike train than in the input train. It tells us how much more information is encoded *per spike*.

Figure 6 shows the estimated *information gain* (a) and the *efficiency gain* (b) for different values of compression. One can observe that the absolute amount encoded is highest for the SPP depicted as the red line. This is reasonable since for this mechanism the model applied by the Monte Carlo estimator is truthful, i.e. matches the true generative model. However, its underlying assumption cannot be fulfilled—it assumes that the hidden state is perfectly known to the unit producing the output. In a realistic setting, however, the state has to be inferred from noisy evidence (see section 2.2).

Compared to the threshold mechanism (blue line), which conserves most of the information contained in the input, the IPP (green line) performs strikingly badly. While the information gain is not highly sensitive to the compression (a), the signalling efficiency increases monotonically for the depicted range (b) and is much larger than 1 for the threshold mechanism except for all but very small values of η. Thus, the information contained in each output spike is much higher than the information contained in an input spike.

This indicates that if signalling *almost* the maximum amount of information is good enough and output events are costly to produce, a higher threshold is a good strategy to make signalling more efficient while keeping metabolic costs low and loss of precision at a reasonable level.

Even if most information about Θ can be transmitted with high threshold and few output spikes, there is a cost in terms of transmitting short or subtle events, a trade-off between precision and cost. Figure 6(c) shows the mutual information between the two estimates $\mathcal{L}$ and $\hat{\mathcal{L}}$ (see section 3.1). This quantifies how much information about the analog quantity $\mathcal{L}$ is contained in the output. It is different from information about the binary state Θ as it tells about the certainty with which the unit assumes the state to be 1.

In this case also, the threshold mechanism is by far the most efficient. However, one can see that the amount of output information about the internal estimate decreases with increasing compression for all mechanisms. In other words, the output tracks the changes in the probability of Θ with less accuracy for higher compression. Short stimulus presentation, for example, will not be transmitted since they do not last long enough to raise $\mathcal{L}$ to the threshold and thus to generate an output spike.

This result clearly shows that a stochastic firing mechanism has an extremely detrimental effect on information transmission capacity (compare the blue curves and the green curves in figure 6). Using the instantaneous estimate of the stimulus to draw a new stochastic process results in a loss of almost all the information contained in the input. This underlines the intuition that if the state Θ is unknown, the TB mechanism provides a quite efficient strategy to transmit all that is known about Θ.

On the other hand, this deterministic neuron behaves like an IPP and does not contain more information about the stimulus than a neuron whose firing rate would directly depend on the *true* state (compare the blue curves and the red curves in figures 6(a) and (b)). However, it represents more accurately the current state of knowledge given by the internal estimate $\mathcal{L}$ (figure 6(c)). This suggests that the Bayesian LIF neuron does not fundamentally change the nature of the code (i.e. it can be interpreted as rate code) but ensures that information contained in the large number of weakly informative synaptic events is efficiently transmitted in a much smaller number of output spikes, avoiding the additional randomness that a stochastic firing mechanism would imply.

In the rest of the result section, we will concentrate on the threshold mechanism and consider how it could be used to understand the consequence of modulating the firing rate of a neuron by acting either on its input or on its threshold η. This has interesting implications for the differential effect of saliency and top-down attention on neural processing and firing behaviour.

5. Input scaling versus modulation of threshold

As can be seen from figure 3, the number of output events can be changed either by changing η or the input rate. Having examined the influence of η in the previous section, we will now show how changes in the input rates q^i_{on} and q^i_{off} affect the information transmission. Intuitively, the more different the two rates become, the stronger the signal and the easier the task to decide whether the hidden state is on or off. This intuition bears out in the discriminability for the static scenario with two Poisson emission probabilities (for some simple calculations, see [31]). The following results specify the analogous relation for the dynamic case of the HMM examined in the previous section.

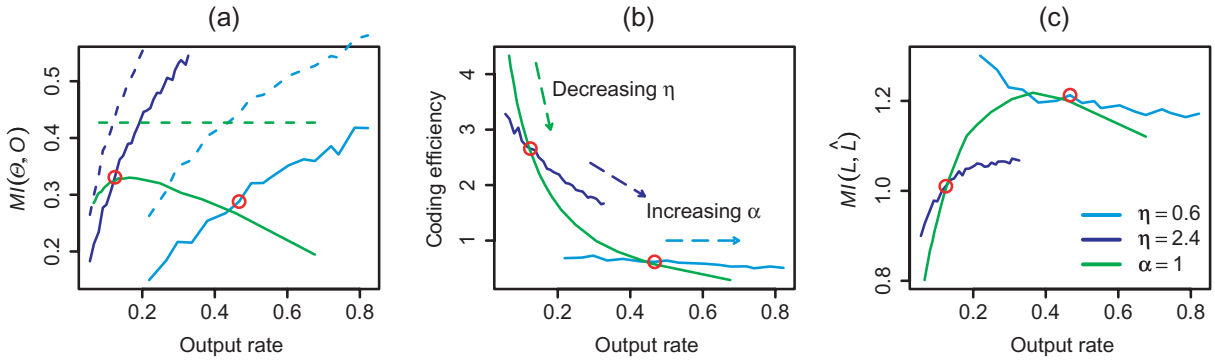

Figure 7. Effect of scaling input rates: (a) mutual information between the outputs and the hidden state (plain lines), dashed lines depict information $MI(\Theta, S)$ available to the unit, (b) the efficiency of coding measured as $MI(\Theta, O)/\bar{\lambda}$ and (c) the mutual information between $\mathcal{L}$ and $\hat{\mathcal{L}}$. α varied between 0.4 and 3 for $\eta = 2.4$ and from 0.4 to 2 for $\eta = 0.6$ and the red circles point at simulations with the same parameter settings.

We change the input rates q^i_{on} and q^i_{off} by multiplying both of them by some value α. While changes in η modify the behaviour of the unit, changing α corresponds to changing the input. This also changes the parameters of the unit (it changes the value of Ψ in equation (4)). Making the link to neurophysiological data, these different changes can be interpreted as changes in saliency (e.g. luminance contrast of a visual stimulus) versus changes in excitability (e.g. how strong the synaptic input has to be to cause a spike) of the unit. Both have been found to multiplicatively scale firing rates [32] and changes in η represent one way to model effects of attention. While it has been noted that the two modifications have rather similar effects on the outputs of cells [33], we ask for the differences in terms of information transmission. While changing η cannot change the information contained in the *input*, scaling α certainly does. It is therefore interesting to see how this affects the amount of information contained in the outputs of the unit.

Figure 7 compares the effect of changing η with comparable changes in α. The graphs show (a) the amount of information contained in the output of the unit for different output rates depending on the settings of α and η. The lines depict changes in one parameter, when the other is fixed (green line: $\alpha = 1,\ 0.4 \leqslant \eta \leqslant 4$; light blue: $\eta = 0.6,\ 0.4 \leqslant \alpha \leqslant 2$; dark blue: $\eta = 2.4,\ 0.4 \leqslant \alpha \leqslant 3$). For the green curve, increases in output rate are due to decreases in η, whereas for the blue curves, increases in output rate correspond to increases in α.

The information contained in the input is illustrated by the dashed blue lines in (a) and shows the simple fact that increasing α increases the available information. Figure 7(a) also shows that increasing the output rate via scaling monotonically increases the information in the output, too. For fixed scaling (dashed green line, $\alpha = 1$), the information available in the input is the same for all values of η as it is independent of the output mechanism. Figure 7(b) shows $MI(\Theta, O)/\bar{\lambda}$, i.e. the mutual information between the hidden state and the output train divided by the mean output rate. As output rate can be taken as a rough estimate for metabolic cost, this provides a measure of signalling efficiency: it describes how much information per spike is transmitted. One can see that both decreasing η and increasing α decrease efficiency. Together with (a) this means that although the amount of information contained in the output spike train

increases with increasing firing rate, efficiency drops. The green line in (b) further justifies our interpretation of η as a way to set the 'signal compression': decreasing this parameter decreases the efficiency of signal encoding and increases firing rate.

Figure 7(c) shows the influence of α and η on the transmission of information about the internal estimator $\mathcal{L}$. While changing α has a relatively minor effect, changing the excitability η has a much more dramatic effect on the capacity of the spike train to faithfully represent fluctuations in the *certainty* about Θ (in contrast to Θ itself). Thus, the main benefit of increasing the excitability of the neuron (i.e. increasing fire rate by decreasing η), is to represent small changes in probability more faithfully. This suggests that top-down attention particularly affects events that are hard to detect (i.e. events that only affect the probability of a feature by a small amount), rather than the total amount of information available about the attended stimuli. In particular, it affects the minimum duration of a stimulus signalled by at least one output spike (data not shown). Note, however, that there is an optimum beyond which increasing the excitability actually has a detrimental effect.

6. Discussion

This study examines information transmission properties of a model neural unit in dynamic environments. We first specified the problem such a unit faces: to perform inference about changing and not directly observable events from noisy evidence. We then described a model that (a) optimally uses such evidence in an online fashion and (b) specifies in a deterministic way how to transmit the acquired knowledge in form of spike trains. At this point, we described alternative output mechanisms with which we compared this model. They are based on the same optimal inference process but contain an intrinsically random component in their output mechanism.

Quantifying information transmission between spike sequences and the stimulus is difficult for technical reasons. We suggest a solution based on optimal decoding that allows to estimate a lower bound for this information. Similar methods based on HMM and Kalman filters had previously been used to analyze neural recordings in a dynamical setting [34, 35]. The results of our simulations show that both the amount of information transmitted and the efficiency with which this is done depend on a single parameter in the model which sets the level of signal compression. Furthermore, the stochastic output mechanisms transmit much less information for all parameter settings.

6.1. Measuring information transfer about a dynamical variable

Neural processing can be understood as a kind of communication between different cells or brain areas. To signal efficiently, maximizing the information contained in spike trains while limiting the number of costly spikes is essential. In this study, we used a HMM and Monte Carlo sampling to estimate the information about a dynamic binary hidden cause at a specific point in time contained in the spike train observed up to this time. This differs from classical studies analyzing the capacity of rate codes or temporal codes for static stimuli [11], [36]–[38]. Other related studies used either different generative models (e.g. with continuous variables and linear dynamics) [34] or did not focus on the specifics of spike generation [35].

Our method for estimating information is closer to the previous work studying tracking of dynamic variables [39] or information transfer by the H1 neuron in the blowfly [5]. The work

by Rieke *et al* [39] also measured the quality of reconstruction for a continuously changing stimulus from spike trains, albeit using different noise models and the Wiener filter theory.

More general statistical methods allowing to estimate, e.g. dynamic receptive fields from spike trains were presented in [39]. By focusing on the specific case of binary variables, our framework yields a simpler specification of the inference dynamics. The model is described in terms of two differential equations suggesting that intrinsic properties of synaptic integration and spike generation actually reflect the dynamics governing inference in time.

6.2. Comparison to other models of spike generation

Closely related models were presented in [15, 16]. These models focus on inference in a discrete state space, implemented in simple dynamical systems like a network of LIF neurons. However, rather than providing a deterministic criterion, the internal estimate determines the *probability* of producing an output—which corresponds to an IPP. Our results indicate that such a rate-based mechanism is sub-optimal.

Our framework may provide an important extension of the previous studies characterizing the input–output relationship of cells in the sensory processing pathways. In analogy to the previous studies estimating spatio-temporal receptive fields for the LNP model [23], our model first filters the synaptic input, then uses a nonlinear transfer function to generate spikes. Thus, the value of $\mathcal{L}$ at a given point in time can be interpreted as the output of a highly nonlinear temporal filter. But whereas the linear filters are usually determined via reverse correlation to minimize a certain error criterion, $\mathcal{L}$ is motivated by arguments from optimal inference for the underlying generative model. Also, in contrast to the LNP model, we use a deterministic spike-generation rule in the form of an adaptive threshold. This may provide a better description of sensory processing in early visual pathways. Thus, retinal ganglion-cell activities are better described as LIF neurons than LNP cells, with the addition of an H-current that inhibits the neuron after each output spike [40]. A similar 'H-current' is implemented by the jump of η in the $\mathcal{G}$-threshold in our model. However, in contrast with the phenomenological models based on LNP and integrate and fire units, an approach based on optimal statistical inference predicts the shape of the filter and the H-current directly from first principles. Thus, this model suggests a way to link the measured neural processing properties with the 'natural' stimulus to which the cell is exposed. Indeed, the filter and transfer functions of retinal and V1 neurons appear to be dynamically adapted to the statistics on their stimuli [41]–[43]. This comparison will be the subject of future studies.

6.3. Saliency and top-down attention

We use this framework to analyze the different ways in which bottom-up saliency and top-down attention might influence processing. Our findings regarding the effects of input scaling and signal compression allow to understand the effects of salience and attention on neural responses with respect to their impact on information transmission about dynamic stimuli. They complement the previous studies based on rate codes that examined the influence of tuning curve shape (e.g. [44]) assuming Poisson distributed spike counts.

On the one hand, we found that scaling of input rates ($\sim$ saliency) increases the amount of information available to the system (see figure 7, dashed lines). For example, changes in luminance contrast make tasks like stimulus discrimination and inference easier [45, 46]. On the other hand, changes of the internal compression parameter η ($\sim$attention) clearly cannot change

the information available to the unit. They can only affect how well a unit transmits the available information about the environment and its estimate thereof. Our results show a difference between the transmission of information about the internal estimate $\mathcal{L}$ and information about the hidden state Θ. While we found that the influence of input scaling and threshold has rather different effects on information about Θ (see figure 7(a)), decrease in compression associated with top-down attention results in a much more faithful representation about precise temporal fluctuation of the *estimate* (i.e. probability of presence) of the stimulus.

Furthermore, adjusting the compression level not only changes the firing rate, but also affects the statistics of the output as illustrated by the distribution of inter-event times (see figure 4). The output process is more bursty for small thresholds and with increasing η becomes more regular than a Poisson process of the same rate. This explains why there is no free lunch here and performance can be improved only moderately. By simply amplifying the output signal, information that should have been gained by increasing the rates is, in fact, lost by introducing auto-correlations of the spike train. A closer examination of these correlations will be necessary to draw a reliable conclusion regarding the effect of attention on information encoded in experimentally recorded spike trains. Similar conclusions have been reached regarding effects of attention on the tuning curves of cortical neurons [47].

6.4. Further work

The model is formulated for a simplistic setting: a binary hidden state and Poisson emission processes. Further analytical work should extend this to more realistic settings dealing with larger discrete or continuous spaces for the hidden state (cf [13]). Furthermore, the strong assumption of Poisson observables should be relaxed, a straightforward extension being Gamma-distributed ISIs, as suggested by neurophysiological data [28].

While the current study used fixed settings for the parameters of the hidden process, it is intuitively clear that they influence the inference process, too. Lowering the switching rates r_{on} and r_{off} makes inference about the hidden state easier. Together with the uncertainty intrinsic in the noisy observation process, this stochasticity of the switching cause ultimately sets the limit on how accurately the state at time t can be estimated [48]. Further work will elucidate this aspect in more detail as it is central to understand how perceptual systems can be tuned to detect the onset of events more quickly and to monitor changes in their environment. While related approaches [49] have started to tackle these questions, there remain many conceptual problems to be solved.

Whereas certain properties might be easy to derive from the posterior estimate by simply looking at the corresponding subsequences of $\mathcal{L}$, this is more difficult for others. For example, estimating the time when the hidden variable was 0 or 1 during a given time interval is easy while determination of the *history of the state*, i.e. the most likely subsequence for the hidden state sequence which caused a specific observed sequence is more difficult. Similar problems form the main interest of signal detection theory and the analysis of HMM. That is why results from this field should prove to be of central importance to understand further aspects of the underlying inference task perceptual systems have to deal with.

Acknowledgments

This work has been supported by the BACS consortium grant FP6-IST-027140 and the BIND Marie Curie Team of Excellence grant MECT-CT-2005-024831. We thank M Remme, M Boerlin and the anonymous reviewers for their comments on the manuscript.

References

[1] Tolhurst D, Movshon J and Dean A 1983 The statistical reliability of signals in single neurons in cat and monkey visual cortex *Vision Res.* **23** 775–85

[2] Shannon C and Weaver W 1949 *The Mathematical Theory of Communications* (Champaign, IL: University of Illinois Press)

[3] Borst A and Theunissen F 1999 Information theory and neural coding *Nat. Neurosci.* **2** 947–57

[4] Meister M and Berry M 1999 The neural code of the retina *Neuron* **22** 435–50

[5] Rieke F, Warland D, de Ruyter van Steveninck R and Bialek W 1997 *Spikes: Exploring the Neural Code* (Cambridge, MA: MIT Press)

[6] Linsker R 1988 Self organisation in a perceptual network *Comput. Mag.* **21** 105–17

[7] Bell A and Sejnowski T 1997 The 'independent components' of natural scenes are edge filters *Vision Res.* **37** 3327–38

[8] Toyoizumi T, Aihara K and Gerstner W 2005 Generalized Bienenstock-Cooper-Munro rule for spiking neurons that maximizes information transmission *Proc. Natl. Acad. Sci. USA* **102** 5239–44

[9] Geisler W 1989 Sequential ideal-observer analysis of visual discriminations *Psychological Rev.* **96** 267–314

[10] Golden J and Shadlen M 2001 Neural computations that underlie decisions about sensory stimuli *Trends Cogn. Sci.* **5** 10–6

[11] Theunissen F and Miller J 1991 Representation of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons *J. Neurophys.* **66** 1690–703

[12] Brenner N, Bialek W and de Ruyter van Steveninck 2000 Adaptive rescaling maximizes information transmission *Neuron* **26** 695–702

[13] Huys Q, Zemel R, Natarajan R P and Dayan P 2007 Fast population coding *Neural Comput.* **19** 460–97

[14] Rao R and Ballard D 1997 Dynamic model of visual recognition predicts neural response properties in the visual cortex *Neural Comput.* **9** 721–63

[15] Beck J and Pouget A 2007 Exact inferences in a neural implementation of a hidden Markov model *Neural Comput.* **19** 1344–61

[16] Rao R 2004 Bayesian computation in recurrent neural circuits *Neural Comput.* **16** 1-38

[17] Rao R 2005 Hierarchical Bayesian inference in networks of spiking neurons *Advances in Neural Information Processing Systems 17* ed L K Saul, Y Weiss and L Bottou (Cambridge, MA: MIT Press) pp 1113–20

[18] Mainen Z and Sejnowski T 1995 Reliability of spike timing in neocortical neurons *Science* **268** 1503–6

[19] Dayan P and Abbott L 2001 *Theoretical Neuroscience* (Cambridge, MA: MIT Press)

[20] Gerstner W and Kistler W 2002 *Spiking Neuron Models: Single Neurons, Populations and Plasticity* (Cambridge: Cambridge University Press)

[21] Gutkin B and Ermentrout G 1998 Dynamics of membrane excitability determine interspike interval variability: a link between spike generation mechanisms and cortical spike train statistics *Neural Comput.* **10** 1047–65

[22] Carandini M, Demb J, Mante V, Tolhurst D, Dan Y, Olshausen B, Gallant J and Rust N 2005 Do we know what the early visual system does? *J. Neurosci.* **25** 10577–97

[23] Chichilnisky E 2001 A simple white noise analysis of neuronal light responses *Network* **12** 199–213

[24] Denève S 2005 Bayesian inference in spiking neurons *Advances in Neural Information Processing Systems 17* ed L K Saul, Y Weiss and L Bottou (Cambridge, MA: MIT Press) pp 353–60

[25] Rabiner L 1989 A tutorial on hidden Markov models and selected applications in speech recognition *Proc. IEEE* **77** 257–86

[26] Denève S 2008 Bayesian spiking neurons I: inference *Neural Comput.* **20** 91–117

[27] Mongillo G and Denève S 2008 On-line learning with hidden Markov models *Neural Comput.* at press doi:10.1162/neco.2008.10-06-351

[28] Troy J and Robson J 1992 Steady discharges of *x* and *y* retinal ganglion cells of cat under photopic illuminance *Visual Neurosci.* **9** 535–53
[29] Golomb D, Hertz J, Panzeri S, Treves A and Richmond B 1997 How well can we estimate the information carried in neuronal responses from limited samples? *Neural Comput.* **9** 649–65
[30] Paninski L 2003 Estimation of entropy and mutual information *Neural Comput.* **15** 1191–254
[31] Gautrais J and Thorpe S 1998 Rate coding versus temporal order coding: a theoretical approach *BioSystems* **48** 57–65
[32] McAdams C and Maunsell J 1999 Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4 *J. Neurosci.* **19** 431–41
[33] Treue S 2004 Perceptual enhancement of contrast by attention *Trends Cogn. Sci.* **8** 435–7
[34] Feng J and Brown D 2004 Decoding input signals in time domain—a model approach *J. Comput. Neurosci.* **16** 237–49
[35] Paninski L, Shoham S, Fellows M, Hatsopoulos N and Donoghue J 2004 Superlinear population encoding of dynamic hand trajectory in primary motor cortex *J. Neurosci.* **24** 8551–61
[36] Oram M, Wiener M, Lestienne R and Richmond B 1999 Stochastic nature of precisely timed spike patterns in visual system neuronal responses *J. Neurophysiol.* **81** 3021–33
[37] Richmond B and Optican L 1990 Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. II. Information transmission *J. Neurophysiol.* **64** 370–80
[38] Stein R 1967 The information capacity of nerve cells using a frequency code *Biophys. J.* **7** 797–826
[39] Eden U T, Frank L M, Barbieri R, Solo V and Brown E N 2004 Dynamic analysis of neural encoding by point process adaptive filtering *Neural Comput.* **16** 971–98
[40] Pillow J, Paninski L, Uzzell V, Simoncelli E and Chichilnisky E 2005 Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model *J. Neurosci.* **25** 11003–13
[41] Fairhall A, Lewen G, Bialek W and de Ruyter van Steveninck R 2001 Efficiency and ambiguity in an adaptive neural code *Nature* **412** 787–92
[42] Hosoya T, Baccus S and Meister M 2005 Dynamic predictive coding by the retina *Nature* **436** 71–7
[43] Smirnakis S, Berry M, Warland D, Bialek W and Meister M 1997 Adaptation of retinal processing to image contrast and spatial scale *Nature* **386** 69–73
[44] Nakahara H, Wu S and Amari S-I 2001 Attention modulation of neural tuning through peak and base rate *Neural Comput.* **13** 2031–47
[45] Cameron E, Tai J and Carrasco M 2002 Covert attention affects the psychometric function of contrast sensitivity *Vision Res.* **42** 949–67
[46] Sclar G and Freeman R 1982 Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast *Exp. Brain Res.* **46** 457–61
[47] Pouget A, Denève S, Ducom J-C and Latham P 1999 Narrow versus wide tuning curves: what's best for a population code? *Neural Comput.* **11** 85–90
[48] Rezaeian M 2006 Hidden Markov process: a new representation, entropy rate and estimation entropy *Preprint* cs.IT/0606114
[49] Yu A J 2007 Optimal change-detection and spiking neurons *Advances in Neural Information Processing Systems 19* ed L K Saul, Y Weiss and L Bottou (Cambridge, MA: MIT Press) pp 1545–52