frontiers in
COMPUTATIONAL NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 04 April 2014
doi: 10.3389/fncom.2014.00038

Stochastic variational learning in recurrent spiking networks

***Danilo Jimenez Rezende*[1,2]* *and Wulfram Gerstner*[1,2]

[1] Laboratory of Cognitive Neuroscience, School of Life Sciences, Brain Mind Institute, Ecole Polytechnique Federale de Lausanne, Lausanne, Vaud, Switzerland
[2] Laboratory of Computational Neuroscience, School of Computer and Communication Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne, Vaud, Switzerland

Edited by:
Nicolas Brunel, Centre National de la Recherche Scientifique, France

Reviewed by:
Tomoki Fukai, RIKEN Brain Science Institute, Japan
Anthony Burkitt, University of Melbourne, Australia

****Correspondence:***
Danilo Jimenez Rezende, Laboratory of Cognitive Neuroscience, School of Life Sciences, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausannne, EPFL - LCN AAB 135 (Bâtiment AAB) Station 15, CH-1015 Lausanne, Switzerland
e-mail: rezende.danilo@gmail.com

The ability to learn and perform statistical inference with biologically plausible recurrent networks of spiking neurons is an important step toward understanding perception and reasoning. Here we derive and investigate a new learning rule for recurrent spiking networks with hidden neurons, combining principles from variational learning and reinforcement learning. Our network defines a generative model over spike train histories and the derived learning rule has the form of a local Spike Timing Dependent Plasticity rule modulated by global factors (neuromodulators) conveying information about "novelty" on a statistically rigorous ground. Simulations show that our model is able to learn both stationary and non-stationary patterns of spike trains. We also propose one experiment that could potentially be performed with animals in order to test the dynamics of the predicted novelty signal.

Keywords: neural networks, variational learning, spiking neurons, synapses, action potentials

1. INTRODUCTION

Humans and animals are able to learn complex behavioral tasks and memorize events or temporally structured episodes. Most likely, learning and memory formation are intimately linked to changes in the synaptic connection strength between neurons. Long-term potentiation and depression of synapses can be induced by many different experimental protocols, and depend on voltage (Artola and Singer, 1993; Ngezahayo et al., 2000), spike-timing (Markram et al., 1997; Bi and Poo, 2001) as well as a subtle combination of timing, voltage, and frequency (Sjöström et al., 2001; Clopath et al., 2010). Spike-Timing Dependent Plasticity (STDP) has intrigued theoreticians, because it provides a local Hebbian learning rule for spiking neurons; local, here, means that the dynamics of the synapses is of the form $\frac{d}{dt}w_{ij} \propto h(post_i, pre_j)$, where pre_j is the set of pre-synaptic variables of neuron j (e.g., spike timing) and $post_i$ is the set of post-synaptic variables of neuron i (e.g., spike times and voltage) and h is an arbitrary functional.

Unsupervised learning through STDP has been repeatedly shown (Levy et al., 2001; Song and Abbott, 2001; Izhikevich et al., 2004; Morrison et al., 2007; Cateau et al., 2008; Gilson et al., 2009; Clopath et al., 2010) to yield connectivity structures that leads to non-trivial activity patterns in recurrent spiking networks.

With relation to neuroscience, unsupervised learning is most commonly related to developmental plasticity (Miller et al., 1989), formation of receptive fields (Song and Abbott, 2001) or cortical rewiring (Young et al., 2007). Indeed most early applications of unsupervised STDP concern the learning of feedforward connections and the formation of receptive fields (Gerstner et al., 1996; Kempter et al., 1999; Song et al., 2000; Song and Abbott, 2001). Unsupervised STDP will tune to the earliest spikes (Song and Abbott, 2001; Gerstner and Kistler, 2002; Guyonneau et al., 2005) and can perform Independent Component Analysis (Clopath et al., 2010; Savin et al., 2010).

On the level of behavioral neuroscience, human performance approaches in many psychophysical learning paradigms Bayes optimality, i.e., the best statistical model cannot perform better than humans do (Knill and Pouget, 2004; Körding and Wolpert, 2004). This supports the hypothesis that the brain is performing approximate inference, which implies that the brain has access to prior and posterior distribution of possible explanations of the observed data (Berkes et al., 2011).

These findings lead to the idea that the spiking activity of the brain constitutes a generative model, that is, a model of the joint distribution of percepts (observed spike trains induced by sensors) and hidden causes in the world (hidden spike train generated by neurons that are not directly affected by sensor spikes).

The ability to model hidden causes in the sensory data is important for both stationary and non-stationary situations. For stationary distributions of spike trains, hidden neurons are important to encode potential interpretations or explanations of spike patterns observed at the sensory neurons (visible neurons). For non-stationary sequences, hidden neurons are fundamental for representing the non-observed dynamics and to form long-term memories.

Various abstract Bayesian models have been proposed to account for this phenomenon (Körding and Wolpert, 2004; Deneve, 2008; Nessler et al., 2009). However, it remains an open question whether optimization in abstract Bayesian models can be translated into plausible learning rules for spiking neurons. If one considers only stationary input patterns, an explicit relation between Bayesian inference and synaptic plasticity has been suggested (Habenschuss et al., 2012). Moreover, it has been suggested recently (Pecevski et al., 2011; Shao, 2012) that spiking networks with biologically plausible dynamics can produce stationary samples from Deep Boltzmann Machines (Salakhutdinov and Hinton, 2009) and more general Bayesian networks. In the following we drop the limitations of stationary spatial patterns and consider a recurrent network of stochastically spiking neurons as a generative model of spatio-*temporal* spiking patterns.

Learning the synapses between hidden neurons in recurrent models has been recognized as a difficult problem, as these synapses only contribute indirectly to the activity of the visible neurons and as recurrence leads to the "vanishing gradient" problem (Bengio et al., 1994). Conversely, in machine learning the algorithms known to be efficient for learning graphical and recurrent models are typically non-local (Jaakkola and Jordan, 2000; Bhatnagar et al., 2007; Sutskever et al., 2008; Salakhutdinov and Hinton, 2009). That is, in order to efficiently perform parameter updates in these models, their learning rules either take into account the entire state of the model or it requires that information to propagate in a non-causal maner through the synapses.

Therefore, the locality constraints imposed by biology constitutes one of the major challenges in transforming those algorithms into biologically plausible learning rules.

Even in biology, however, certain types of non-local signals participate in the learning process, notably through neuromodulators which can convey information about the global state of the network or external information (e.g., reward or surprise) (Izhikevich, 2007; Schultz, 2008; Fremaux et al., 2010).

Here, we derive a principled learning rule for unsupervised learning in recurrent spiking networks that relies only on quantities that are locally available at the synapse: pre-synaptic activity, post-synaptic activity and a global modulatory signal.

The key innovation of our model compared to earlier studies (Brea et al., 2011; Jimenez Rezende et al., 2011) lies in the computation of a global modulating signal which is a linear superposition of local terms and can therefore be interpreted as the diffusion of a neuromodulator in the extra-cellular medium.

Furthermore our global neuromodulating signal conveys information about *novelty* or *surprise* on statistically rigorous grounds, providing interesting links with findings relating surprise and plasticity (Gu, 2002; Ranganath and Rainer, 2003; Yu and Dayan, 2005).

We show with simulations based on synthetic data that our proposed learning mechanism is capable to capture complex hidden causes behind the observed spiking patterns and is able to replicate, in its spontaneous activity, the statistics of the observed spike trains.

Finally, we provide an application of our model to a hypothetical novelty detection task where a simulated agent (e.g., a rat) is inserted into a maze with specific properties (views, rooms, topology of the maze). Our model, simulating the "brain" of this agent, successfully captures the statistical properties of this environment. We show that, after learning, our model is capable of distinguishing the original environment from another environment that differs only in its topology (relative location of the rooms). Additionally, we predict the "expected dynamics" of a neuromodulator signaling "novelty" as the agent traverses the virtual maze. We propose that this hypothetical experiment could be developed into a real experiment.

2. MATERIALS AND METHODS

2.1. NEURON MODEL

The neuron model used in our simulations is a generalized linear model (GLM) in the form of a Spike Response Model (SRM) with escape noise (Gerstner and Kistler, 2002; Jolivet et al., 2006). The spike train of a neuron j for times $t > 0$ is denoted as $X_j(t) = \sum_{t_j^f \in \left\{t_j^1,\ldots,t_j^{N_s}\right\}} \delta\left(t - t_j^f\right)$, where $\left\{t_j^1, \ldots, t_j^{N_s}\right\}$ is the set of spike timings.

We model the membrane potential of a neuron i as

$$u_i(t) = \sum_j w_{ij}\phi_j(t) + \eta_i(t), \tag{1}$$

where $\eta_j(t) = -\eta_0 \int_0^t dse^{-\frac{(t-s)}{\tau_{\text{adapt}}}} X_j(s)$ is the adaptation potential, w_{ij} is the synaptic strength between neurons i and j and $\phi_j(t)$ is the potential evoked by an incoming spike from neuron j. The evoked potentials are modeled by a simple exponential filter $\phi_j(t) = \int_0^t dse^{-\frac{(t-s)}{\tau}} X_j(s)$ implemented as a differential equation

$$\dot{\phi}_j(t) = \frac{1}{\tau}(X_j(t) - \phi_j(t)), \tag{2}$$

where τ is the time constant of the membrane potential.

The spikes are generated by a conditioned Poisson process with exponential escape rate (Jolivet et al., 2006). That is, the conditional instantaneous firing intensity $\rho_i(t)$ is taken to be

$$\rho_i(t) = \rho_0 \exp\left[\frac{u_i(t) - \vartheta}{\Delta u}\right], \tag{3}$$

where ϑ and Δu are physical constants of the neuron. However, we will keep $\rho_i(t)$ as an arbitrary function of $u_i(t)$, $\rho_i(t) = g(u_i(t))$, in all our derivations and we specify the exponential form (Equation 3) only when performing the simulations (see further below). Equations (1–3) capture the simplified dynamics of a spiking neuron with stochastic spike firing.

In the following simulations we assume that two different neurons i and j can have at most two common synapses, w_{ij} and w_{ji}. The neuron model and the potentials contributing to its activation are illustrated in **Figure 1**.

In the following sections, we will first introduce the theoretical framework in which we derive our learning rule. The learning mechanism is then derived in several steps, followed by simulations showing that our model and learning rules are capable of capturing complex spatio-temporal features in the input spike trains, reproduce them in its spontaneous activity and perform statistical inference on the hidden causes of provided data.

2.2. A PRINCIPLED FRAMEWORK FOR LEARNING

In the following we consider a stochastic, fully connected network $\mathcal{M}$ composed of two sets of neurons which are functionally distinct. The first group which we call *observed* or *visible* neurons (denoted by $\mathcal{V}$) represents an ensemble of neurons which receive data in the form of spike trains. The spike trains of the visible neurons will be referred to as $X_{\mathcal{V}}$. These neurons are exposed to the external world. The second set, which we call *hidden* neurons (denoted by $\mathcal{H}$) does not directly receive data from the external world. The spike trains of the hidden neurons will be referred to as $X_{\mathcal{H}}$. Their role is to provide "compressed explanations" for the observed data. The topology of this network is illustrated in **Figure 2C**. In the absence of external drive, the spontaneous activity of the hidden neurons will contribute to the firing of the observed neurons. The spike trains of the entire network comprising both observed and hidden neurons will be indicated simply as X.

Our network defined in this way constitutes a generative model of spike trains with hidden neurons. In the following we interpret synaptic potentiation and depression as a form of optimization of this generative model. More precisely, we assume that synaptic plasticity (the "learning rule") is trying to increase the likelihood of the observed spike trains under the model.

In what follows, we review the calculation of the complete-data log-likelihood $\log p(X_{\mathcal{V}}, X_{\mathcal{H}})$ for a recurrent network of point-process neurons.

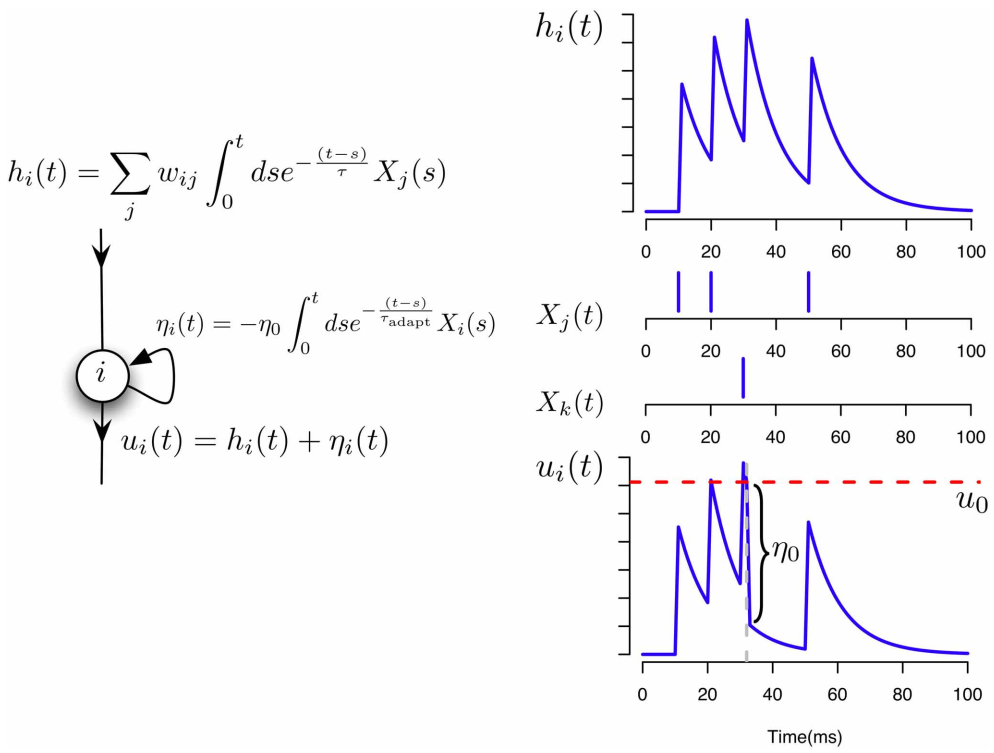

FIGURE 1 | Neuron and synaptic models. Illustration of the different contributions to the total membrane potential of a neuron *i* in our model.

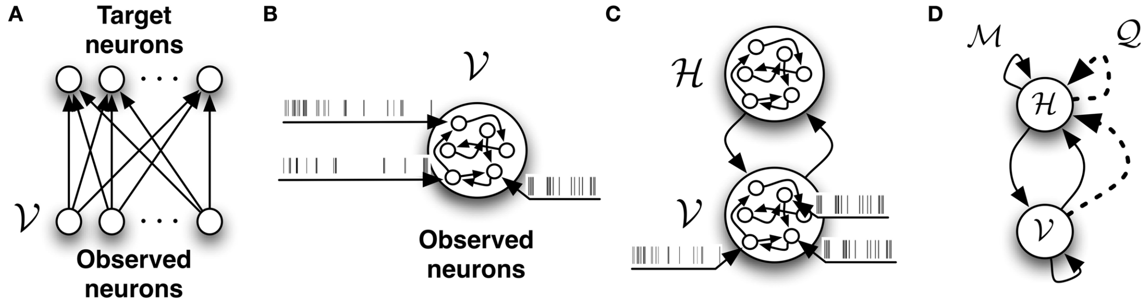

FIGURE 2 | The different network architectures discussed in this study. **(A)** Feed-forward, fully observed spiking network. It defines a conditional model of the target spiking patterns given the input patterns. **(B)** Fully connected and fully observed spiking network. It defines a generative model of the observed data. **(C)** Network decomposed in two pools of neurons: neurons which receive data or observed neurons $\mathcal{V}$ and neurons which do not receive data or hidden neurons $\mathcal{H}$. Both pools of neurons are fully connected. **(D)** The network with $\mathcal{M}$-synapses (solid links) and $\mathcal{Q}$-synapses (dashed links) which provides the infrastructure required for learning the generative model.

From Equations (1–3) it follows that the probability $P_i\left(t_i^f \in [t, t+\Delta t] | X(0\ldots t)\right)$ of producing a spike in the infinitesimal interval $[t, t+\Delta t]$ by the ith neuron conditioned on the past activity of the entire network $X(0\ldots t)$ and provided that $\rho_i(t)\Delta t \ll 1$ is given by

$$P_i\left(t_i^f \in [t, t+\Delta t] | X(0\ldots t)\right) \approx \rho_i(t)\Delta t, \tag{4}$$

and the probability $P_i\left(t_i^f \notin [t, \Delta t] | X(0\ldots t)\right)$ of producing no spike in the same interval is given by

$$P_i\left(t_i^f \notin [t, t+\Delta t] | X(0\ldots t)\right) \approx (1 - \Delta t \rho_i(t)). \tag{5}$$

Therefore, by discretizing a finite time interval $[0, T]$ into N sufficiently small bins $[t_k, t_k + \Delta t]$ with $k = 1\ldots N$ so that there is at most one spike in each bin and assuming independence between neurons within the infinitesimal bins we can write the probability $P(X(0\ldots T))$ of producing a spike train $X_i(t) = \sum_{t_i^f} \delta\left(t - t_i^f\right)$ for all neurons in the network as

$$P(X(0\ldots T)) \approx \prod_{i\in\mathcal{V}\cup\mathcal{H}} \prod_{k_i^s} \left[\rho_i\left(t_{k_i^s}^f\right)\Delta t\right] \prod_{k_i^{\text{ns}}} \left[1 - \Delta t \rho_i\left(t_{k_i^{\text{ns}}}^f\right)\right], \tag{6}$$

where k_i^s and k_i^{ns} labels the bins with one spike and bins without spike from neuron i, respectively. Taking the limit $N \to \infty$ of Equation (6) divided by the volume Δt^K, where K is the total number of spikes in the interval $[0, T]$ we obtain the probability density

$$p(X(0\ldots T)) = \prod_{i\in\mathcal{V}\cup\mathcal{H}} \left[\prod_{t_i^f} \rho\left(t_i^f\right)\right] \exp\left(-\int_0^T dt \rho_i(t)\right). \tag{7}$$

The log-likelihood corresponding to Equation (7) can be compactly written as

$$\log p(X(0\ldots T)) = \sum_{i\in\mathcal{V}\cup\mathcal{H}} \int_0^T d\tau \left[\log \rho_i(\tau) X_i(\tau) - \rho_i(\tau)\right]. \tag{8}$$

It should be stressed that the log-likelihood (Equation 8) is not a sum of independent terms, since the instantaneous firing rate of each neuron depends on the entire past activity of all the other neurons through Equations (1–3).

In the following sections we derive a plasticity rule that will attempt to maximize the log-likelihood of the observed data $\log p(X_\mathcal{V})$. For this, we briefly review the calculation of the gradient of the observation likelihood in absence of hidden neurons and then we introduce our method for approximating its gradient when there are hidden neurons.

2.3. FULLY OBSERVED NETWORK

The gradient of the log-likelihood of fully observed networks of SRM neurons and similar point-processes have been studied in detail in Paninski (2004) and Pfister et al. (2006). In what follows we review the calculation of the gradients of the data log-likelihood (Equation 8) with respect to the synaptic weights and then discuss a simulation revealing the limitations of the model.

In a network without hidden neurons as illustrated in **Figures 2A,B**, the data log-likelihood (Equation 8) reduces to a simple form

$$\log p(X_\mathcal{V}) = \sum_{k\in\mathcal{V}} \int_0^T d\tau \left[\log \rho_k(\tau) X_k(\tau) - \rho_k(\tau)\right]. \tag{9}$$

In Paninski (2004) and Pillow et al. (2004) it is shown that the optimization problem defined by the log-likelihood (Equation 9) is convex. Therefore its global maximum can be found by gradient ascent.

The gradient of Equation (9) with respect to the synaptic efficacies w_{ij} is given by

$$\nabla_{w_{ij}} \log p(X_\mathcal{V}) = \sum_{k\in\mathcal{V}} \int_0^T dt' \frac{\partial \log \rho_k(t')}{\partial w_{ij}} \left[X_k(t') - \rho_k(t')\right], \tag{10}$$

where the gradients $\frac{\partial \log \rho_k(t')}{\partial w_{ij}}$ are obtained by differentiating the firing rate function (Equation 3):

$$\frac{\partial \log \rho_k\left(t'\right)}{\partial w_{ij}} = \delta_{ki} \frac{g'\left(u_k\left(t'\right)\right)}{g\left(u_k\left(t'\right)\right)} \phi_j\left(t'\right) \tag{11}$$

We conclude that the gradient $\nabla_{w_{ij}} \log p(X_\mathcal{V})$ can be calculated in a purely local manner. More precisely, an update of the weights according to gradient ascent $\Delta w_{ij} \propto \nabla_{w_{ij}} \log p(X_\mathcal{V})$ yields a learning rule that is simply a "trace" of a product of two factors: A first factor $\phi_j\left(t'\right)$ that depends only on the presynaptic activity and a second factor $\frac{g'(u_k(t'))}{g(u_k(t'))}\left[X_k\left(t'\right) - \rho_k\left(t'\right)\right]$ that depends on the state of the postsynaptic neuron. Moreover, the gradient (Equation 10) has been shown to yield a simplified form of STDP (Pfister et al., 2006). The notion of "trace" (i.e., a temporal average of some quantity) will return for other learning rules later in this paper.

In order to expose the weakness of the fully observed model described above, we test its performance on a task which consists of learning "stair patterns" involving 3 groups of 10 visible neurons, which are probabilistically activated using low (1 Hz) and high (700 Hz) firing rates. The activations generate a sequential pattern where each group remains active for a duration drawn from a Gaussian distribution with mean 30 ms and standard deviation 10 ms (truncated at positive values), (**Figure 3A**). This benchmark is interesting firstly because it requires the formation of memories on the order of three times the membrane time constants of the single neuron dynamics and secondly because it requires the ability to learn the appropriated transition probabilities. Our simulations show that a fully observed network is not capable of learning such patterns, as can be seen by looking at the samples produced from the learned network (**Figure 3B**). However, a model with 50 hidden neurons (with the learning rule discussed further below) can learn the distribution (**Figures 3C,D**).

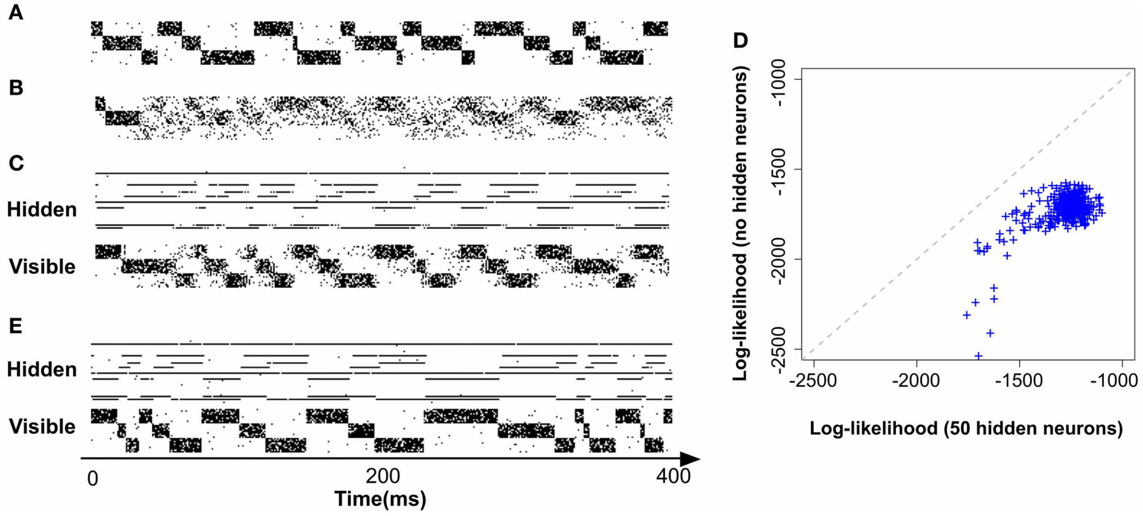

FIGURE 3 | Models with hidden neurons succeed where a fully observed model fails. Here we compare our model (online with variance reduction, see Materials and Methods) with and without hidden neurons. **(A)** Sample from the training data. **(B)** Sample activity from the visible neurons of a learned model without hidden neurons. From the generated samples we can see that a model with only visible neurons can capture some correlations present in the data (note the formation of "downward" moving blurred patterns) but fails at capturing the long-term structure of the pattern. **(C)** Sample from a learned model with 50 hidden neurons: hidden neurons' spikes (top) and visible neurons' spikes (bottom). Note that the learned hidden representation unambiguously represents the global state of the visible neurons, generating different spiking patterns for each distinc phase of the training pattern. **(D)** Relative performance of the models with (vertical axis) and without (horizontal axis) hidden neurons measured by their data log-likelihood. The blue points indicate the relative evolution of the data log-likelihood of both models during a learning session. The model with latent neurons has systematically a better performance (higher log-likelihood) than the model without hidden neurons. **(E)** Model with hidden neurons in inference mode. The activity of the hidden neurons from the $\mathcal{Q}$-network (top) forms a higher-level representation of the observed spike trains (bottom).

Moreover, even if the fully observed network could learn the data distribution it would not provide any useful representation of the data, while a network with hidden neuron naturally forms higher-level representations of the incoming data (**Figure 3E**)(top). For precise details concerning the numerical simulations and evaluations, see further below.

This simulation corroborates the intuition that a network consisting of visible neurons only is rather limited in scope. We therefore turn in the following to a more general network consisting of both visible and hidden neurons.

2.4. PARTIALLY OBSERVED NETWORK

In the following we first introduce our model that includes hidden neurons. We derive our learning rule and introduce a few modifications to improve its performance. Finally, we show with simulations that the resulting model can learn spiking patterns that couldn't be learned by a model without hidden neurons.

In a model that includes hidden neurons (that is, neurons not directly connected to the incoming data), the marginalized likelihood of the visible neurons is obtained by integrating over all possible hidden spike trains $X_{\mathcal{H}}$,

$$p(X_{\mathcal{V}}) = \int \mathcal{D}X_{\mathcal{H}}\ p(X_{\mathcal{V}}, X_{\mathcal{H}}). \tag{12}$$

In the following we introduce the variational approximation scheme for approximating Equation (12). The variational approach consists of approximating a complex distribution p by a simpler distribution q and provides a flexible generalization of the expectation-maximization (EM) algorithm (see Jaakkola and Jordan, 2000; Beal and Ghahramani, 2006).

We are interested in approximating the posterior $p(X_{\mathcal{H}}|X_{\mathcal{V}})$ by another distribution over spike trains $q(X_{\mathcal{H}}|X_{\mathcal{V}})$. We optimize the parameters of the distribution $q(X_{\mathcal{H}}|X_{\mathcal{V}})$ by minimizing the KL-divergence

$$\begin{aligned} KL(q;p) &= \int \mathcal{D}X_{\mathcal{H}} q\left(X_{\mathcal{H}}|X_{\mathcal{V}}\right) \log \frac{q(X_{\mathcal{H}}|X_{\mathcal{V}})}{p(X_{\mathcal{H}} \mid X_{\mathcal{V}})} \\ &= \int \mathcal{D}X_{\mathcal{H}} q(X_{\mathcal{H}}|X_{\mathcal{V}}) \log \frac{q(X_{\mathcal{H}}|X_{\mathcal{V}})}{p(X_{\mathcal{H}}, X_{\mathcal{V}})} + \log p(X_{\mathcal{V}}) \\ &= \left\langle \log q(X_{\mathcal{H}}|X_{\mathcal{V}}) - \log p(X_{\mathcal{H}}, X_{\mathcal{V}}) \right\rangle_{q(X_{\mathcal{H}}|X_{\mathcal{V}})} \\ &\quad + \underbrace{\log p(X_{\mathcal{V}})}_{\text{Data log-likelihood}}\ , \end{aligned} \tag{13}$$

where $\langle f(X) \rangle_p = \int \mathcal{D}Xf(x)p(x)$. The first term in Equation (13) is known in statistical physics as the *Helmholtz free energy* (Landau et al., 1980),

$$\mathcal{F} = \left\langle \log q(X_{\mathcal{H}}|X_{\mathcal{V}}) - \log p(X_{\mathcal{H}}, X_{\mathcal{V}}) \right\rangle_{q(X_{\mathcal{H}}|X_{\mathcal{V}})}. \tag{14}$$

The second term in Equation (13) is simply the data log-likelihood. Since the KL-divergence $KL(q;p)$ between two distributions q and p is non-negative (Gibbs and Su, 2002), the free energy (Equation 14) is an upper bound on the negative log-likelihood. Therefore, we can redefine the problem of maximizing the data log-likelihood $\log p(X_{\mathcal{V}})$ with respect to the parameters of the generative model p as the double optimization problem of minimizing the free energy $\mathcal{F}$ with respect to the parameters of q *and* with respect to the parameters of the original model p.

The attractiveness of such an approach for deriving biologically plausible rules comes from the fact that the distribution q is arbitrary (as long as it has the same support as the true distribution). Therefore we can either choose it in order to simplify the calculations or to improve the model's compliance with known biological constraints, such as causality and locality of the weight updates. These interesting properties of the variational approximation have been explored by a diversity of models (Dayan, 2000; Friston and Stephan, 2007; Jimenez Rezende et al., 2011; Brea et al., 2013; Nessler et al., 2013).

In what follows, we postulate that the posterior distribution of the hidden spike trains of the network $\mathcal{M}$ can be well approximated by another recurrent network of spiking neurons which we call the network $\mathcal{Q}$. This network is composed of the same neurons as the original model, except that it has different synaptic connections. Its connectivity is depicted in **Figure 2D**.

The assumption of an "inference" network $\mathcal{Q}$ is analogous to the *recognition model* introduced in Dayan (2000) for the Helmholtz machine.

Our model differs from the Helmholtz machine, as introduced in Dayan (2000), in two key aspects: (1) The Helmohtz machine is a model for stationary data, i.e., it cannot readily model temporal sequences; (2) Although the recognition network is introduced in a variational framework, the proposed learning rule is not attempting to minimize a free energy (there are different cost functions for the generative and recognition models) whereas our learning rule is explicitly attempting to minimize the free energy associated to the model.

In other words, our model consists of a *single* set of neurons with two sets of synaptic weights that can be turned on and off independently (possibly through the action of specific neuromodulators). The first set of weights, which we will refer to as $w_{ij}^{\mathcal{M}}$, parameterizes the original generative model $\mathcal{M}$. The second set of weights, which we will refer to as $w_{ij}^{\mathcal{Q}}$ parameterizes the network $\mathcal{Q}$. The topology of the network $\mathcal{Q}$ is restricted and excludes connections toward the visible neurons from hidden and other visible neurons.

In the following, all the quantities (e.g., membrane potentials and firing rates) that are computed using the parameters $w_{ij}^{\mathcal{M}}$ (i.e., with $\mathcal{Q}$ turned off) will have the superscript $\mathcal{M}$. Analogously, all the quantities that are computed using the parameters $w_{ij}^{\mathcal{Q}}$ (i.e., with $\mathcal{M}$ turned off) have the superscript $\mathcal{Q}$.

When driven solely by the observed spike trains and by the weights $w_{ij}^{\mathcal{Q}}$, the activity of the hidden neurons $X_{\mathcal{H}}$ is, by construction, an approximated sample from true the posterior distribution $p(X_{\mathcal{H}}|X_{\mathcal{V}})$ provided by the distribution $q(X_{\mathcal{H}}|X_{\mathcal{V}})$. Therefore, if we want our model to perform approximated Bayesian inference on the most likely hidden causes of an observed spike train $X_{\mathcal{V}}$ we just have to run it with the synapses $w_{ij}^{\mathcal{M}}$ turned off. Conversely if we want to produce a sample from the learned generative model, we have to run it with the synapses $w_{ij}^{\mathcal{Q}}$ turned off.

In the next sections we derive stochastic estimators of the gradients of the free energy $\mathcal{F}$ with respect to the synaptic weights $w_{ij}^{\mathcal{M}}$ and $w_{ij}^{\mathcal{Q}}$ in a biologically plausible manner.

We show that the naive gradients obtained for $w_{ij}^{\mathcal{Q}}$ are problematic since their variance grows quadratically with the size of the network. We then introduce a simple modification to reduce the variance of the obtained gradients based on techniques from reinforcement learning. Finally we modify the gradient estimators so that they turn into "on-line" parameter updates.

2.5. STOCHASTIC GRADIENTS

The complete data log-likelihood $\mathcal{L}^{\mathcal{M}}$ of the model $\mathcal{M}$ and $\mathcal{L}^{\mathcal{Q}}$ of the model $\mathcal{Q}$ are given by

$$\begin{aligned}\mathcal{L}^{\mathcal{M}} &= \log p(X_{\mathcal{V}}, X_{\mathcal{H}}) \\ &= \sum_{i\in\mathcal{V}\cup\mathcal{H}} \int_0^T d\tau \left[\log \rho_i^{\mathcal{M}}(\tau) X_i(\tau) - \rho_i^{\mathcal{M}}(\tau)\right] \quad (15)\end{aligned}$$

and

$$\begin{aligned}\mathcal{L}^{\mathcal{Q}} &= \log q(X_{\mathcal{H}}|X_{\mathcal{V}}) \\ &= \sum_{i\in\mathcal{H}} \int_0^T d\tau \left[\log \rho_i^{\mathcal{Q}}(\tau) X_i(\tau) - \rho_i^{\mathcal{Q}}(\tau)\right] \quad (16)\end{aligned}$$

respectively.

The free energy (Equation 14) corresponding to the log-likelihoods (Equations 15, 16) is given by

$$\mathcal{F} = \left\langle \mathcal{L}^{\mathcal{Q}} - \mathcal{L}^{\mathcal{M}} \right\rangle_{q(X_{\mathcal{H}}|X_{\mathcal{V}})}. \quad (17)$$

In the following we simplify the notation and write $\langle \bullet \rangle_q$ instead of $\langle \bullet \rangle_{q(X_{\mathcal{H}}|X_{\mathcal{V}})}$. We wish to write the learning equations for both $w_{ij}^{\mathcal{M}}$ and $w_{ij}^{\mathcal{Q}}$ as simple gradient descent on the free energy:

$$\dot{w}_{ij}^{\mathcal{M}} = -\mu^{\mathcal{M}} \nabla_{w_{ij}^{\mathcal{M}}} \mathcal{F} \quad (18)$$

$$\dot{w}_{ij}^{\mathcal{Q}} = -\mu^{\mathcal{Q}} \nabla_{w_{ij}^{\mathcal{Q}}} \mathcal{F}, \quad (19)$$

where $\mu^{\mathcal{M}}$ and $\mu^{\mathcal{Q}}$ are learning rates for the networks $\mathcal{M}$ and $\mathcal{Q}$, respectively. The exact gradients $\nabla_{w_{ij}^{\mathcal{M}}}\mathcal{F}$ and $\nabla_{w_{ij}^{\mathcal{Q}}}\mathcal{F}$ are difficult to evaluate analytically since we cannot compute the required expectations. Therefore we resort to unbiased stochastic approximations of those gradients.

The calculation of the gradient $\nabla_{w_{ij}^{\mathcal{M}}}\mathcal{F}$ is analogous to the fully observed case and is given by

$$\begin{aligned}\nabla_{w_{ij}^{\mathcal{M}}}\mathcal{F} &= \nabla_{w_{ij}^{\mathcal{M}}} \left\langle \mathcal{L}^{\mathcal{Q}} - \mathcal{L}^{\mathcal{M}} \right\rangle_q \\ &= -\left\langle \nabla_{w_{ij}^{\mathcal{M}}} \mathcal{L}^{\mathcal{M}} \right\rangle_q \\ &\approx -\nabla_{w_{ij}^{\mathcal{M}}} \hat{\mathcal{L}}^{\mathcal{M}}, \quad (20)\end{aligned}$$

where $\hat{\mathcal{L}}^{\mathcal{M}}$ is a point-estimate of the complete data log-likelihood of our generative model obtained by computing the required traces from a single simulation of the network $\mathcal{Q}$ in the interval [0, T]. Similarly, the stochastic gradient with respect to $w_{ij}^{\mathcal{Q}}$ is given by

$$\begin{aligned}\nabla_{w_{ij}^{\mathcal{Q}}}\mathcal{F} &= \left\langle \nabla_{w_{ij}^{\mathcal{Q}}}\mathcal{L}^{\mathcal{Q}}\right\rangle_q + \left\langle\left(-\mathcal{L}^{\mathcal{M}}+\mathcal{L}^{\mathcal{Q}}\right)\nabla_{w_{ij}^{\mathcal{Q}}}\mathcal{L}^{\mathcal{Q}}\right\rangle_q \\ &= \left\langle\left(-\mathcal{L}^{\mathcal{M}}+\mathcal{L}^{\mathcal{Q}}\right)\nabla_{w_{ij}^{\mathcal{Q}}}\mathcal{L}^{\mathcal{Q}}\right\rangle_q \\ &\approx \left(-\hat{\mathcal{L}}^{\mathcal{M}}+\hat{\mathcal{L}}^{\mathcal{Q}}\right)\nabla_{w_{ij}^{\mathcal{Q}}}\hat{\mathcal{L}}^{\mathcal{Q}} \\ &= \hat{\mathcal{F}}\,\nabla_{w_{ij}^{\mathcal{Q}}}\hat{\mathcal{L}}^{\mathcal{Q}},\end{aligned} \tag{21}$$

where we have used the fact that $\left\langle\nabla_{w_{ij}^{\mathcal{Q}}}\mathcal{L}^{\mathcal{Q}}\right\rangle_q = \nabla_{w_{ij}^{\mathcal{Q}}}\left\langle\exp\mathcal{L}^{\mathcal{Q}}\right\rangle_q = \nabla_{w_{ij}^{\mathcal{Q}}}1 = 0$ and $\hat{\mathcal{F}}$ is the point-estimate of the free energy,

$$\begin{aligned}\hat{\mathcal{F}} &= \hat{\mathcal{L}}^{\mathcal{Q}} - \hat{\mathcal{L}}^{\mathcal{M}} \\ &= \int_0^T d\tau \sum_{i\in\mathcal{H}}\left[\log\rho_i^{\mathcal{Q}}(\tau)X_i(\tau) - \rho_i^{\mathcal{Q}}(\tau)\right] \\ &\quad - \int_0^T d\tau \sum_{i\in\mathcal{V}\cup\mathcal{H}}\left[\log\rho_i^{\mathcal{M}}(\tau)X_i(\tau) - \rho_i^{\mathcal{M}}(\tau)\right] \\ &= \int_0^T d\tau\mathcal{F}_\tau,\end{aligned} \tag{22}$$

where we have defined the "instantaneous free energy" $\mathcal{F}_\tau$ as

$$\begin{aligned}\mathcal{F}_\tau = &\sum_{i\in\mathcal{H}}\left[\log\rho_i^{\mathcal{Q}}(\tau)X_i(\tau) - \rho_i^{\mathcal{Q}}(\tau)\right] \\ &- \sum_{i\in\mathcal{V}\cup\mathcal{H}}\left[\log\rho_i^{\mathcal{M}}(\tau)X_i(\tau) - \rho_i^{\mathcal{M}}(\tau)\right].\end{aligned} \tag{23}$$

Note that during learning the activity of the visible neurons is driven purely by the observed spike trains while the activity of the hidden neurons and related quantities (e.g., $\rho_i^{\mathcal{Q}}$ and $\rho_i^{\mathcal{M}}$) is driven by the $\mathcal{Q}$-network. Expanding the remaining gradients in Equations (20, 21) using the chain rule we obtain the "batch mode" learning equations

$$\dot{w}_{ij}^{\mathcal{M}}(T) \approx \mu^{\mathcal{M}}\int_0^T dt\frac{g'\left(u_i^{\mathcal{M}}(t)\right)}{g\left(u_i^{\mathcal{M}}(t)\right)}\left[X_i(t)-\rho_i^{\mathcal{M}}(t)\right]\phi_j(t) \quad \forall\, i,j\in\mathcal{V}\cup\mathcal{H}, \tag{24}$$

$$\dot{w}_{ij}^{\mathcal{Q}}(T) \approx -\mu^{\mathcal{Q}}\hat{\mathcal{F}}\int_0^T dt\frac{g'\left(u_i^{\mathcal{Q}}(t)\right)}{g\left(u_i^{\mathcal{Q}}(t)\right)}\left[X_i(t)-\rho_i^{\mathcal{Q}}(t)\right]\phi_j(t) \quad \forall\, i\in\mathcal{H}, j\in\mathcal{V}\cup\mathcal{H}. \tag{25}$$

Note that the learning rule (Equation 24) is the same as in the fully observed network case in Equation (10). The learning rule (Equation 25) for the network $\mathcal{Q}$ is similar, but contains an additional modulation factor, the point estimate of the free energy $\hat{\mathcal{F}}$, which appears as a global signal that modulates the learning of the network $\mathcal{Q}$. Since $\hat{\mathcal{F}}$ provides a lower bound on the data log-likelihood, the free energy measures how much the recent history of observed spike trains "fits" the generative model defined by the network $\mathcal{M}$. The assumption of a globally available signal conveying information about *reward* or *surprise* is standard in the reinforcement learning literature.

The naive stochastic gradients (Equations 24, 25) are not efficient in practice. Even though they constitute unbiased estimators of the true gradients, their *variance is prohibitively high*. We address this problem below.

2.6. REDUCING THE VARIANCE OF THE GRADIENTS

Stochastic gradients of the form Equation (25) have been extensively studied in the reinforcement learning literature and several approaches with different levels of complexity have been proposed for reducing their variance (Munos, 2005; Bhatnagar et al., 2007). In the following, we sketch some theoretical arguments for why gradients of the form of Equation (25) should be expected to scale badly with the size of the network.

Let N be the number of neurons in the network and T be the time window on which we compute the relevant quantities. Equation (25) contains the free energy $\hat{\mathcal{F}}$ which is, according to Equation (22), a sum of traces (integrals on the interval [0, T]) for each neuron. Therefore, in a weak-coupling scenario, the estimator (Equation 25) is the integral on the interval [0, T] of a sum of N weakly correlated terms in the free energy, that we assume to have some typical variance σ_0^2 and mean m. Under these assumptions, the variance of the gradient (Equation 25) scales approximately as

$$Var\left[\dot{w}_{ij}^{\mathcal{Q}}\right] \propto \left[\mu^{\mathcal{Q}}\right]^2\left(N\times T\times\sigma_0^2 + N^2\times T^2\times m^2\right),$$

That is, as the size of the network grows, the variance of the gradient (Equation 25) grows with the square of the number of neurons. A naive solution would be to decrease the learning rate as $\mu^{\mathcal{Q}}\propto 1/N$ but this would make learning too slow for larger networks.

In the following, we adopt a simple baseline removal approach to reduce the variance of our gradient estimator. That is, we simply subtract the mean $\bar{\mathcal{F}}$ of the free energy $\hat{\mathcal{F}}$, calculated as a moving average across several previous batches of length T, from the current value $\hat{\mathcal{F}}(T)$. This yields the learning rule

$$\dot{w}_{ij}^{\mathcal{Q}}(T) \approx -\mu^{\mathcal{Q}}e(T)\int_0^T dt\frac{g'\left(u_i^{\mathcal{Q}}(t)\right)}{g\left(u_i^{\mathcal{Q}}(t)\right)}\left[X_i(t)-\rho_i^{\mathcal{Q}}(t)\right]\phi_j(t) \quad \forall i\in\mathcal{H}, j\in\mathcal{V}\cup\mathcal{H}, \tag{26}$$

where we have introduced the "free energy error signal" $e(T) = \hat{\mathcal{F}}(T) - \bar{\mathcal{F}}$.

With this simple change, we can see that the variance of our gradient scales as

$$Var\left[\dot{w}_{ij}^{\mathcal{Q}}\right] \propto \left[\mu^{\mathcal{Q}}\right]^2 \left(N \times T \times \sigma_0^2\right).$$

The baseline removal trick is not a solution to the problem but it drastically reduces the variance of our estimator when the size of the network is increased.

Note that the quadratic or linear growth of the gradient's variance with the number of neurons is not just an artifact of our variational approximation. The arguments presented in this section apply to any learning rule that has the form of local Hebbian traces modulated by global signals that are composed of several weakly correlated terms. For instance, the learning rule proposed in Brea et al. (2011) also falls into this category.

2.7. ONLINE vs. BATCH LEARNING

The gradients that define our learning scheme (Equations 24, 26) are given in terms of quantities accumulated over a given time interval [0, T]. In this sense, we have derived a "batch" learning rule and we would have to wait until the end of the interval [0, T] in order to apply the changes in the parameters of our model.

However, compatibility with biology requires that we have an online version of our algorithm. This can be approximately achieved if, instead of accumulating the traces required for calculating the gradients on time interval [0, T] and applying the parameters updates in the end, we replace the traces by moving averages and apply the parameter updates at *every* time step. The modified learning rules can be formulated as follows. At each synapse, we have "Hebbian traces" $H_{ij}^{\mathcal{M},\mathcal{Q}}(t)$ that keep track of pre- and post-synaptic activity and evolve according to

$$\tau_G \dot{H}_{ij}^{\mathcal{M}}(t) = -H_{ij}^{\mathcal{M}}(t) + \frac{g'\left(u_i^{\mathcal{M}}(t)\right)}{g\left(u_i^{\mathcal{M}}(t)\right)}\left[X_i(t) - \rho_i^{\mathcal{M}}(t)\right]\phi_j(t), \quad \forall i, j \in \mathcal{V} \cup \mathcal{H} \tag{27}$$

$$\tau_G \dot{H}_{ij}^{\mathcal{Q}}(t) = -H_{ij}^{\mathcal{Q}}(t) + \frac{g'\left(u_i^{\mathcal{Q}}(t)\right)}{g\left(u_i^{\mathcal{Q}}(t)\right)}\left[X_i(t) - \rho_i^{\mathcal{Q}}(t)\right]\phi_j(t), \quad \forall i \in \mathcal{H}, j \in \mathcal{V} \cup \mathcal{H} \tag{28}$$

where we have introduced a time constant τ_G controlling the time-scale of the moving averages. Similarly, the online estimate of the error signal $e_N(T)$ is obtained by replacing time integrals in the interval [0, T] with moving averages

$$\tau_G \dot{\hat{\mathcal{F}}}(t) = -\hat{\mathcal{F}}(t) + \mathcal{F}_t. \tag{29}$$

$$\tau_{\text{baseline}} \dot{\bar{\mathcal{F}}} = -\bar{\mathcal{F}} + \hat{\mathcal{F}}(t). \tag{30}$$

That is, $\hat{\mathcal{F}}$ is a "short term" moving average of the instantaneous free energy $\mathcal{F}_t$ (Equation 23) with time-scale τ_G while $\bar{\mathcal{F}}$ is a "long term" moving average of $\mathcal{F}_t$ with a longer time-scale depending on τ_G and τ_{baseline}. Note that the "error signal" $e_N(T)$ can also be interpreted as an instantaneous surprise measure relative to the slow "background" surprise level $\bar{\mathcal{F}}$. The updates of the weights uses the Hebbian traces and fixed learning rates $\mu^{\mathcal{M}}$ and $\mu^{\mathcal{Q}}$

$$\dot{w}_{ij}^{\mathcal{M}}(t) = \mu^{\mathcal{M}} H_{ij}^{\mathcal{M}}(t) \tag{31}$$

$$\dot{w}_{ij}^{\mathcal{Q}}(t) = -\mu^{\mathcal{Q}} e_N(t) H_{ij}^{\mathcal{Q}}(t). \tag{32}$$

Thus the update of the $\mathcal{M}$-network is given by a "Hebbian" rule whereas the update of the $\mathcal{Q}$-network follows a three-factor rule with the surprise as a global factor. This architecture is illustrated in **Figure 4A** and the three-factor rule for the $\mathcal{Q}$-synapses is illustrated in **Figure 4C**. Another way of interpreting the learning rule (Equation 32) is that it is simply proportional to the covariance between the Hebbian trace $H_{ij}^{\mathcal{Q}}$ and the moving average of the free energy $\hat{\mathcal{F}}$. If they are uncorrelated, the expected change in the parameters will be zero and the synaptic weights will just perform a centered random walk.

2.8. A SIMPLIFIED MODEL

Since the variational distribution q in Equation (13) can be arbitrary, one could imagine a model simpler than the one derived in the previous sections which consists in approximating the posterior distribution of the hidden neurons directly by the forward dynamics of the generative model. In practice this approximation amounts to constraining the synaptic weights $w_{ij}^{\mathcal{Q}}$ of the $\mathcal{Q}$-network to be equal to the synaptic weights $w_{ij}^{\mathcal{M}}$ of the $\mathcal{M}$-network for $i \in \mathcal{H}$ and $j \in \mathcal{V} \cup \mathcal{H}$.

Under this constraint, the learning equations (31) and (32) reduces to

$$\dot{w}_{ij}^{\mathcal{M}}(t) = \mu^{\mathcal{M}} \begin{cases} H_{ij}^{\mathcal{M}}(t) & \text{if } i \in \mathcal{V} \\ -e_N(t) H_{ij}^{\mathcal{M}}(t) & \text{otherwise} \end{cases} \tag{33}$$

and the instantaneous free energy simplifies to a sum over the observed neurons only

$$\mathcal{F}_\tau = -\sum_{i \in \mathcal{V}} \left[\log \rho_i^{\mathcal{M}}(\tau) X_i(\tau) - \rho_i^{\mathcal{M}}(\tau)\right]. \tag{34}$$

The architecture of this model is illustrated in **Figure 4B** The idea of using the forward-dynamics as a proposal distribution for the posterior has been used in Brea et al. (2011) and Brea et al. (2013), where the proposals are then weighted by an importance-sampling scheme to better represent the true posterior distribution over the hidden activity.

Further below (see results) we show that, at least in the context of the variational learning discussed here, this approximation does not outperform our more general model.

2.9. NUMERICAL SIMULATIONS

All the simulations in this study are based on a discrete-time version of the Equations (31, 32). The spiking process is approximated by taking $1 - \exp\left[-dt\rho(t)\right]$ as the probability of producing one spike in the finite time bin $[t, t + dt]$.

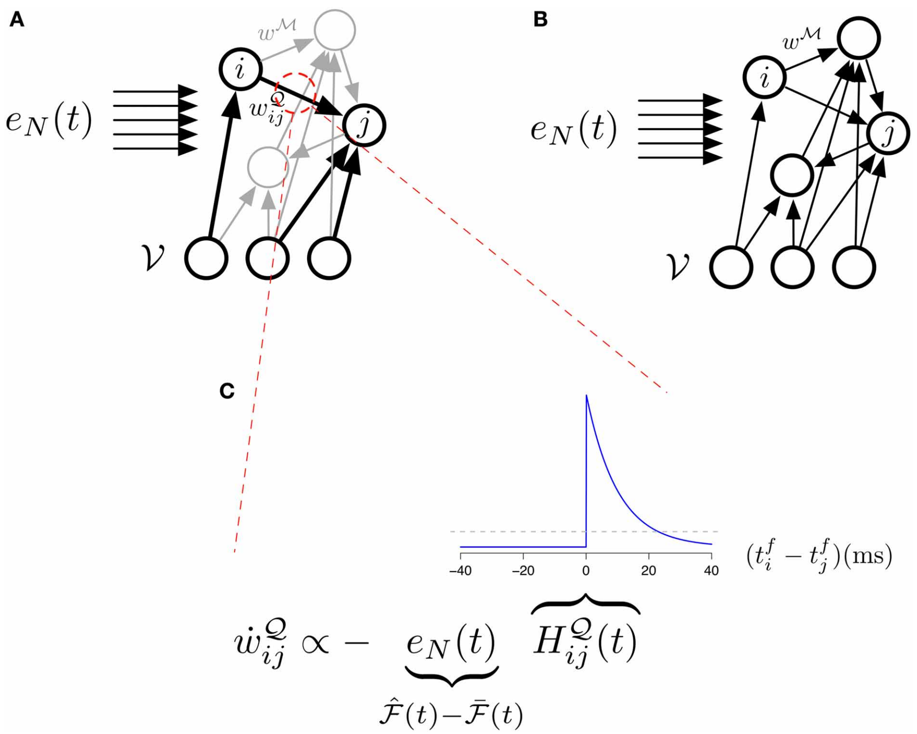
A
$w^{\mathcal{M}}$
i
$w_{ij}^{\mathcal{Q}}$
j
$e_N(t)$
$\mathcal{V}$
B
$w^{\mathcal{M}}$
i
j
$e_N(t)$
$\mathcal{V}$
C
−40
−20
0
20
40
$(t_i^f - t_j^f)(\mathrm{ms})$
$\dot{w}_{ij}^{\mathcal{Q}} \propto - \underbrace{e_N(t)}_{\hat{\mathcal{F}}(t)-\bar{\mathcal{F}}(t)} \overbrace{H_{ij}^{\mathcal{Q}}(t)}$

FIGURE 4 | The different studied architectures and three-factor STDP. **(A)** Illustration of our main model (Equations 31, 32) with its two sets of synapses $\mathcal{Q}$ (black) and $\mathcal{M}$ (gray). The "error signal" $e_N(t)$ modulates the learning of the $\mathcal{Q}$-synapses. **(B)** Illustration of the simplified model (Equations 33, 34) with its single set of synapses $\mathcal{M}$. The "error signal" $e_N(t)$ modulates the learning of the $\mathcal{M}$-synapses. **(C)** Learning in the main model's $\mathcal{Q}$-synapses is triggered by the covariance between the Hebbian trace $H_{ij}^{\mathcal{Q}}$ and the moving average of free energy $\hat{\mathcal{F}}$. The inset blue curve indicates the shape of the Hebbian trace term as a function of the time interval between pre and post synaptic spikes. The dotted gray line corresponds to $H_{ij}^{\mathcal{Q}} = 0$ (no weight change). A positive covariance will induce synaptic depression (left quadrant of blue curve) while a negative covariance will induce synaptic potentiation (right quadrant of blue curve). A null covariance would induce a centered random walk on the synaptic weights.

The values of the parameters used in this study are reported in **Table 1**.

The initial synaptic weight of both $\mathcal{M}$ and $\mathcal{Q}$ were sampled from a Gaussian distribution with mean zero and standard deviation of 0.01.

For all experiments, the training data consists of binary arrays with ones indicating spikes and zeros indicating no-spike in the corresponding time bin. The training data-sets are organized in batches of 200 ms which are sequentially presented to the model. During our training sessions, each learning epoch corresponds to 500 presentations of data batches. In other words, each epoch corresponds to a total of 100 s of spiking data. During the learning phase, the visible neurons are exactly driven by the data spike trains, that is, they are forced to spike or to not spike in the exact same way as the data samples at each time bin. During the spontaneous activity phase the networks are running without any external drive.

The log-likelihood of test data was estimated by an importance sampling procedure. Given a generative model with density $p(x_v, x_h)$ over observed variables x_v and hidden variables x_h, importance sampling allow us to estimate the density $p(x_v)$ of a data point x_v as

$$p(x_v) = \left\langle p(x_v|x_h) \right\rangle_{p(x_h)} = \left\langle p(x_v|x_h) w(x_h, x_v) \right\rangle_{q(x_h)}, \quad (35)$$

Table 1 | Parameters used in the simulations.

Parameter	Description	Value
dt	Time discretization interval	1 ms
τ	Membrane potential time constant	10 ms
η_0	Adaptation potential strength	0.1 mV
τ_{adapt}	Adaptation potential time-scale	10 ms
ρ_0	Firing rate scale	1 kHz
ϑ	Firing threshold	0 mV
Δu	Firing sensitivity to the membrane potential	1 mV
$\mu^{\mathcal{M}}$	Learning rate for the model network $\mathcal{M}$	0.00001
$\mu^{\mathcal{Q}}$	Learning rate for the recognition network $\mathcal{Q}$	0.00001
τ_G	Time scale for the moving averages of the gradients and free energy	10 ms
τ_{baseline}	Time scale for the moving average of mean free energy	100 ms

where $q(x_h|x_v)$ is an arbitrary distribution with same support as $p(x_h)$ and $w(x_h, x_v) = p(x_h)/q(x_h|x_v)$ is the importance weight. The Equation (35) can be rewritten in terms of the point-estimated of the free energy (Equation 22) as follows

$$\begin{aligned} p(x_v) &= \langle p(x_v|x_h)w(x_h, x_v)\rangle_{q(x_h|x_v)} \\ &= \langle \exp[\log p(x_v|x_h) + \log p(x_h) - \log q(x_h|x_v)]\rangle_{q(x_h|x_v)} \\ &= \left\langle \exp -\hat{\mathcal{F}}(x_v, x_h)\right\rangle_{q(x_h|x_v)}, \end{aligned} \tag{36}$$

where $\hat{\mathcal{F}}(x_v, x_h)$ is a point-estimator of the free energy.

Using Equation (36) we estimate the log-likelihood of a observed spike-train by sampling several times from the network $\mathcal{Q}$ and taking the average of the exponentiated point-estimator of the free energy (Equation 22). For our applications, we have generated 500 samples of duration 100 ms from the network $\mathcal{Q}$ per estimation.

3. RESULTS

The learning rules derived above (see Materials and Methods) come in four different variants. First, a batch-based naive gradient descent rule. Second, a variant that reduces the variance of the gradient estimator. Third an online version of the variance-reduced rule that is biologically more plausible than a batch rule. Finally, a simplified version of the model where the $\mathcal{Q}$-network is merged with the $\mathcal{M}$-network in a specific manner.

3.1. VARIANCE REDUCED RULE vs. NAIVE RULE

Naive gradient descent on the free energy yields the batch learning rule (Equation 25). Tested on the stairs-patterns (**Figure 3A**) using a network of 30 visible neurons and 30 hidden neurons generates very slow learning. In **Figure 5A** we show that the learning rule with the variance reduction (Equation 26) performs substantially better than the naive gradient (Equation 25). Both networks have the same number of visible and hidden neurons, but the first network is trained with the gradient given in Equation (25) whereas the second network is trained with the gradient defined in Equation (26). The data log-likelihood of both networks is approximated by importance sampling after every learning epoch.

In **Figure 5A**, along the horizontal axis we plot, across several epochs the log-likelihood of the model using the learning

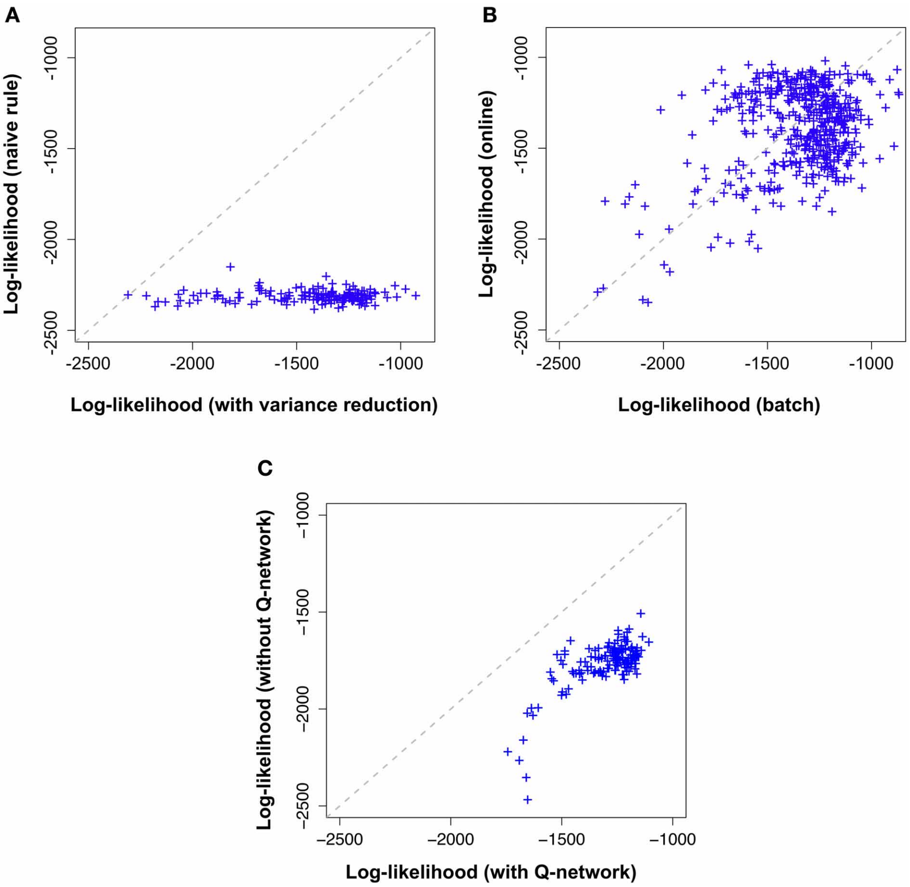

FIGURE 5 | Variance reduction and $\mathcal{Q}$-network are important while on-line approximation is not. Comparison of the different flavors of the proposed model across 100,000 learning epochs on the "stairs pattern" task. Shown log-likelihoods where estimated by importance sampling every 500 epochs during learning. Likelihood values (crosses) further to the right (high log-likelihood) correspond to the end of learning while the cross on the diagonal marks the beginning of learning (epoch 1). **(A)** Model with variance reduction (horizontal axis) achieves a higher log-likelihood than a naive model without variance reduction (vertical axis). **(B)** Batch model with variance reduction(horizontal axis) and on-line model with variance reduction (vertical axis) exhibit a similar evolution of log-likelihoods. **(C)** Online-model with variance reduction (horizontal axis) performs better than the simplified model (without the $\mathcal{Q}$-network) with variance reduction defined by Equations (33, 34) (vertical axis).

rule with variance reduction and along the vertical axis the log-likelihood of the model without variance reduction (i.e., the naive batch gradient rule). We find that the log-likelihood of the variance reduced model is consistently above the log-likelihood of the naive model, indicating an increase in learning speed of more than a factor 100.

3.2. ONLINE ALGORITHM vs. VARIANCE REDUCED BATCH ALGORITHM

Online algorithms are biologically more plausible than batch algorithms which require storage of intermediate results. Here we show that the online learning rules (Equations 31, 32) induce only minor impairments of the performance compared to the batch rules (Equations 24, 26).

Both versions of the model were trained on the same "stair pattern" task and the results are shown in **Figure 5B**. Along the horizontal axis we indicate the log-likelihood of the batch model and along the vertical axis that of the online model, across several epochs of learning. Both performances are strongly correlated, indicating that the online approximation does not introduce any major impairment in the model for this particular dataset.

3.3. FORWARD DYNAMICS vs. THE $\mathcal{Q}$-NETWORK

Here we show that the simplified model defined by the Equations (33, 34) does not reach the performance of the more general model defined by Equations (31, 32).

Both versions of the model were trained on the same "stair pattern" task and the results are shown in **Figure 5C**. Along the horizontal axis we indicate the log-likelihood of our model with $\mathcal{Q}$-network and along the vertical axis that of the simplified model, across several epochs of learning. Both performances are correlated. However, the simplified model has clearly lower log-likelihoods than the complete model. This result suggests that the forward dynamics of the generative model may provide a poor approximation to the true posterior distribution of the hidden neurons compared to having an independently parameterized inference network.

3.4. HIDDEN REPRESENTATIONS AND INFERENCE

To show that our model is not only learning a prior that represents the data but can also form interesting hidden representations and perform inference on the hidden explanations of the incoming data, we use the $\mathcal{Q}$-network of a model with 50 hidden neurons, trained on the stairs dataset **Figure 3A**. The activity of the hidden neurons of the model during sampling (or "dreaming") are shown on **Figure 3C** (top). As we can see just by visually inspecting the activity of the hidden neurons, they form an unambiguous representation of the activity of the visible neurons in this simple example. Conversely, by running the model on "inference mode" (i.e., with the synapses $w^{\mathcal{Q}}$ activated) can also see that the model is capable of performing inference on the causes of the incoming data **Figure 3E**.

3.5. THE ROLE OF THE NOVELTY SIGNAL

The resulting online rule given in Equation (32) for is a Hebbian-type plasticity rule modulated by a global novelty signal $e_N(t)$ where $e_N(t)$ is a measure of surprise relative to a slow moving average of the free energy. We repeat the learning rule for the $\mathcal{Q}$-network from Equation (32),

$$\dot{w}^{\mathcal{Q}}_{ij}(t) = -\mu^{\mathcal{Q}} e_N(t) H^{\mathcal{Q}}_{ij}(t), \tag{37}$$

where $H^{\mathcal{Q}}_{ij}(t)$ is a synaptic trace that keeps track of "Hebbian" coincidence between pre- and post-synaptic activity.

The Hebbian trace

$$\tau_G \dot{H}^{\mathcal{Q}}_{ij}(t) = -H^{\mathcal{Q}}_{ij}(t) + \frac{g'\left(u^{\mathcal{Q}}_i(t)\right)}{g\left(u^{\mathcal{Q}}_i(t)\right)}\left[X_i(t) - \rho^{\mathcal{Q}}_i(t)\right]\phi_j(t) \tag{38}$$

is activated by the product of a voltage-dependent post-synaptic factor $\frac{g'\left(u^{\mathcal{Q}}_i(t)\right)}{g\left(u^{\mathcal{Q}}_i(t)\right)}\left[X_i(t) - \rho^{\mathcal{Q}}_i(t)\right]$ and a presynaptic EPSP caused by presynaptic spike arrival.

It is known that the brain is able to detect novelty and to broadcast novelty related signal across large brain regions (Gu, 2002; Ranganath and Rainer, 2003). In order to illustrate the dynamics of the novelty signal in our model we describe a task which could be transformed into a real animal experiment.

We place an agent (e.g., a rat) in a maze and let it explore it. During the exploration, the agent will learn the topology of the maze through a combination of visual and proprioceptive information. If the agent is suddenly transported to another maze with many similar but a few different parts we expect this change to trigger some novelty signal in the brain of the agent whenever it encounters the parts of the new maze that differ from the learned maze. In order for the agent's brain to detect a change in the environment, it must first learn a sufficiently accurate model of the environment. We hypothesized that our recurrent neuronal network learns the structure of the environment and at the same time provides an online surprise signal when it encounters a "novel" situation. Such a novelty signal could be compared to recordings of different neuromodulators.

We created two virtual mazes composed of 16 "rooms" arranged as a square lattice where only neighboring rooms are accessible from each other: a first one which the agent will learn, the target maze, indicated in **Figure 6A** (left) and the test maze **Figure 6A** (right) which is similar to the first but with a few, randomly chosen rooms, replaced as indicated in **Figure 6A** by the yellow squares. Note that the only way to detect the difference between the two mazes in this simulation is to actually learn the exact connectivity graph of the rooms, because the rooms are themselves identical in both mazes.

The views corresponding to each room were generated by randomly choosing images of handwritten digits from the MNIST dataset (LeCun and Cortes, 2010). The MNIST images are 28 × 28 gray-scale images of handwritten digits. We converted the pixel-values to firing rates in the range [0.01 Hz, 9 Hz]. To keep the simulations simple, time was considered in abstract units for these simulations (with time steps of 100 ms instead of 1 ms).

For the "brain" of the agent we used a recurrent binary network with 30 hidden neurons and 28 × 28 visible neurons. In order to train this network, data batches where produced by recoding the activity of the visible neurons while the agent performs random trajectories of 100 time-steps in the target maze.

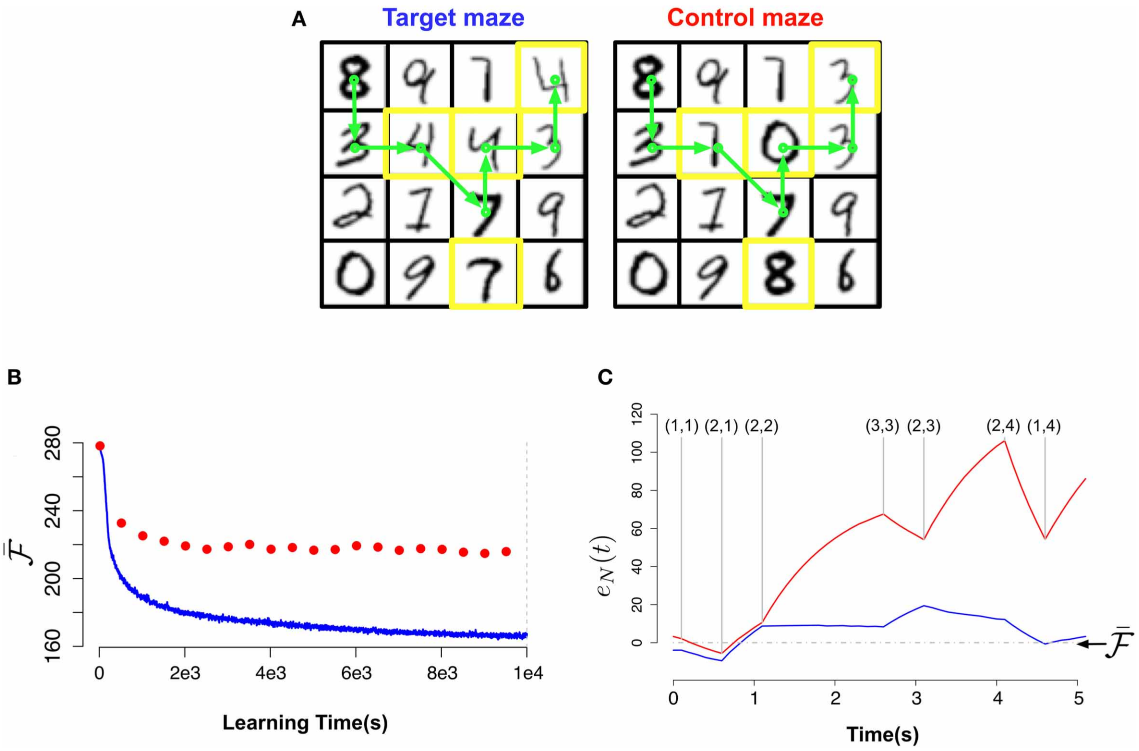

FIGURE 6 | Novelty-gated Hebbian learning. (A) The target maze (left) and test maze (right). Each one of the 16 rooms in the maze is represented by a random MNIST digit. Transitions are only possible between neighbor rooms. The test maze was generated from the target one by randomly replacing a few rooms (indicated with yellow borders). A sample trajectory is shown with the green arrows. **(B)** Evolution of the slow moving average of the free energy $\bar{\mathcal{F}}$ as a function of the amount of observed data for the target maze (blue) and the mean free energy of the same model when "teleported" to the control maze (red) every 500 s. **(C)** Fine-grained temporal details of the fast free energy traces $e_N(t)$ while traversing the sample trajectory for the target maze (blue) and for the test maze (red). Visited rooms are indicated by the annotated coordinates at each change-point. During the intervals between two consecutive change-points the agent remains in the same room.

Each learning epoch corresponds to 500 presentations of these data-batches to our model.

In **Figure 6B** we plot the "slow" moving average of the free energy $\bar{\mathcal{F}}$ as a function of the learning time during the exploration phase for the target maze (blue curve) and for the test maze (red dots). The reported free energy for the control maze was measured every 5000 s for a path of length 50 s. As we can see, in the beginning of the learning the model is unable to distinguish between both mazes (both have high free energy). But as the model learns the target maze, it successfully identifies the test maze as "unfamiliar" (attributing low free energy to the target maze and much higher free energy to the test maze).

In **Figure 6C** we plot for both mazes the free energy error signal $e_N(t)$ for the sample trajectory shown in **Figure 6A**. From **Figure 6C** (blue) we can see that $e_N(t)$ fluctuates around zero for the learned maze but deviates largely from zero for the test maze.

This result suggests that if animals have a neurophysiological correlate of the "free energy error signal" $e(T)$ introduced in Equation (26) we should look for activity bursts when the animals traverses unexpected situations (e.g., when traversing the test maze from position (2, 1) to position (2, 2)).

Moreover we should expect a substantial increase in the variance of the changes in synaptic weights when moving from a learned maze to an unfamiliar maze due to the change in the baseline of the surprise levels.

4. DISCUSSION

We have proposed an alternative to the learning algorithms previously proposed in Brea et al. (2011) and Jimenez Rezende et al. (2011) for learning a generative model of spike trains defined by recurrent spiking networks. Our new model combines techniques from variational learning and reinforcement learning to derive a new efficient synaptic plasticity rule.

The resulting (see Materials and Methods) online rule for synapses is a Hebbian-type trace modulated by a global novelty signal. The Hebbian terms are traces of products of pre-synaptic terms (EPSPs) and post-synaptic terms. Similar gradients have been studied in Pfister et al. (2006) where they are found to yield STDP-like dynamics. Importantly, the global modulating signal, Equation (23) is a *linear superposition* of terms locally computed by each neuron so we can interpret it as the diffusion of a neuromodulator in the extra-cellular medium.

The original feature of the proposed model is that it uses an auxiliary recurrent spiking network in order to approximate the posterior distribution of the hidden spiking activity given the observed spike trains in an on-line manner. Using this auxiliary network allowed us to derive a learning rule which is based on local gradients modulated by slow non-local factors conveying information about "novelty."

We have shown that naive stochastic gradients derived in such a framework are not viable in practice due to their high variance (which may grow quadratically with the number of hidden neurons). Deriving viable learning rules thus requires finding low-variance and unbiased estimators of the gradients defined in Equations (24, 25). In this paper we have only reduced this problem as our learning rule (Equation 26) still has a variance that grows (linearly) with the number of hidden neurons.

Our proposed learning algorithm, has potential applications for finding functional networks from recorded neurophysiological data. Since it can learn a recurrent spiking network that approximatively "explains" the data. Taking into account external currents injected into the network would be straightforward.

We also provide an on-line neural estimator of novelty/surprise and make experimentally testable predictions about its dynamics. The estimator is a quantity that naturally emerges from the statistical principles behind our framework instead of being an *ad hoc* quantity.

From the biological literature, it is known that novelty or surprise is, at least partly, encoded in neuromodulators such as ACh (Ranganath and Rainer, 2003; Yu and Dayan, 2005). Moreover, neuromodulators are known to affect synaptic plasticity (Gu, 2002). We suggested a hypothetical animal experiment that could test predictions concerning a novelty signal and its potential relation to plasticity.

5. RELATED WORK

The framework in which we derive our model is very general and has been used in different ways in previous works (Dayan, 2000; Friston and Stephan, 2007; Jimenez Rezende et al., 2011; Brea et al., 2013; Nessler et al., 2013). The uniqueness of the present works relies on the combination of methods from variational learning and reinforcement learning yielding a learning rule that is both biologically plausible and efficient.

The main difference between our model and models that exploit more analytical properties of the variational approximation (explicitly computing expectations and covariances terms in the free energy) (Friston and Stephan, 2007; Jimenez Rezende et al., 2011) is that, although these analytical approximations may provide gradient estimators with much lower variance and typically yield more scalable algorithms, these methods are intrinsically non-local and further approximations are required in order to obtain a biologically plausible learning rule.

An interesting learning algorithm for recurrent spiking networks which approximates the gradients of the KL-divergence (Equation 13) using an importance sampling technique has been proposed in Brea et al. (2011) and Brea et al. (2013). Their algorithm does not use an auxiliary network to approximate the activity of the hidden dynamics conditioned on the observed spike trains. Instead, the generative forward dynamics of the model is used as a proposal distribution which is then weighted by an importance-sampling approximation.

Another important family of algorithms proposed in Habenschuss et al. (2012) and Nessler et al. (2013), heavily relies on approximating recurrent neural networks with soft winer-takes-all (WTA) dynamics. One advantage of assuming a WTA dynamics is that the computation of the gradients of the KL-divergence (Equation 13) greatly simplifies, yielding a simple local learning rule. A limitation of their approach is that the model does not take into account the full temporal dynamics during inference.

FUNDING

Financial support was provided by the Swiss National Science Foundation (SystemsX) as well as by the European Research Council (Grant Agreement no. 268 689), and the European Community's Seventh Framework Program (Grant Agreement no. 269921, BrainScaleS).

ACKNOWLEDGMENTS

The authors would like to thank Guillaume Hennequin, Johanni Brea, and Jean-Pascal Pfister who provided valuable comments on the text.

REFERENCES

Artola, A., and Singer, W. (1993). Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci.* 16, 480–487. doi: 10.1016/0166-2236(93)90081-V

Beal, M. J., and Ghahramani, Z. (2006). Variational bayesian learning of directed graphical models with hidden variables. *Bayesian Anal.* 1, 793–832. doi: 10.1214/06-BA126

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* 5, 157–166. doi: 10.1109/72.279181

Berkes, P., Orban, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87. doi: 10.1126/science.1195870

Bhatnagar, S., Sutton, R., Ghavamzadeh, M., and Lee, M. (2007). Incremental natural actor-critic algorithms. *Adv. Neural Inf. Process. Syst.* 20, 105–112.

Bi, G., and Poo, M. (2001). Synaptic modification of correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.* 24, 139–166. doi: 10.1146/annurev.neuro.24.1.139

Brea, J., Senn, W., and Pfister, J. (2011). Sequence learning with hidden units in spiking neural networks. *Adv. Neural Inf. Process. Syst.* 24, 1422–1430.

Brea, J., Senn, W., and Pfister, J.-P. (2013). Matching recall and storage in sequence learning with spiking neural networks. *J. Neurosci.* 33, 9565–9575. doi: 10.1523/JNEUROSCI.4098-12.2013

Cateau, H., Kitano, K., and Fukai, T. (2008). Interplay between a phase respone curve and spike-timing dependent plasticity leading to wireless clustering. *Phys. Rev. E* 77:051909. doi: 10.1103/PhysRevE.77.051909

Clopath, C., Busing, L., Vasilaki, E., and Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based spike-timing-dependent-plasticity with homeostasis. *Nat. Neurosci.* 13, 344–352. doi: 10.1038/nn.2479

Dayan, P. (2000). "Helmholtz machines and wake-sleep learning," in *Handbook of Brain Theory and Neural Network*, ed M. A. Arbib (Cambridge, MA: MIT Press), 44.

Deneve, S. (2008). Bayesian spiking neurons i: inference. *Neural Comput.* 20, 91–117. doi: 10.1162/neco.2008.20.1.91

Fremaux, N., Sprekeler, H., and Gerstner, W. (2010). Functional requirements for reward-modulated spike-timing-dependent plasticity. *J. Neurosci.* 40, 13326–13337. doi: 10.1523/JNEUROSCI.6249-09.2010

Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature* 383, 76–78. doi: 10.1038/383076a0

Gerstner, W., and Kistler, W. K. (2002). *Spiking Neuron Models. Single Neurons, Populations, Plasticity*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511815706

Gibbs, A. L., and Su, F. E. (2002). On choosing and bounding probability metrics. *Int. Stat. Rev.* 70, 419. doi: 10.2307/1403865

Gilson, M., Burkitt, A., Grayden, D., Thomas, D., and van Hemmen, J. L. (2009). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks iv: structuring synaptic pathways among recurrent connections. *Biol. Cybern.* 27, 427–444. doi: 10.1007/s00422-009-0346-1

Gu, Q. (2002). Neuromodulatory transmitter systems in the cortex and their role in cortical plasticity. *Neuroscience* 111, 815–835. doi: 10.1016/S0306-4522(02)00026-X

Guyonneau, R., VanRullen, R., and Thorpe, S. (2005). Neurons tune to the earliest spikes through stdp. *Neural Comput.* 17, 859–879. doi: 10.1162/0899766053429390

Habenschuss, S., Bill, J., and Nessler, B. (2012). Homeostatic plasticity in bayesian spiking networks as expectation maximization with posterior constraints. *Adv. Neural Inf. Process. Syst.* 25, 782–790.

Izhikevich, E., Gally, J., and Edelman, G. (2004). Spike-timing dynamics of neuronal groups. *Cereb. Cortex* 14, 933–944. doi: 10.1093/cercor/bhh053

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cereb. Cortex* 17, 2443–2452. doi: 10.1093/cercor/bhl152

Jaakkola, T., and Jordan, M. (2000). Bayesian parameter estimation via variational methods. *Stat. Comput.* 10, 25–37. doi: 10.1023/A:1008932416310

Jimenez Rezende, D., Wierstra, D., and Gerstner, W. (2011). Variational learning for recurrent spiking networks. *Adv. Neural Inf. Process. Syst.* 24, 136–144.

Jolivet, R., Rauch, A., Lüscher, H.-R., and Gerstner, W. (2006). Predicting spike timing of neocortical pyramidal neurons by simple threshold models. *J. Comput. Neurosci.* 21, 35–49. doi: 10.1007/s10827-006-7074-5

Kempter, R., Gerstner, W., and van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Phys. Rev. E* 59, 4498–4514. doi: 10.1103/PhysRevE.59.4498

Knill, D., and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural codding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007

Körding, K., and Wolpert, D. (2004). Bayesian integration in sensorimotor learning. *Nature* 427, 244–247. doi: 10.1038/nature02169

Landau, L. D., and Lifshitz, E. (1980). *Statistical Physics*. Elsevier Science and Technology Books.

LeCun, Y., and Cortes, C. (2010). *MNIST Handwritten Digit Database*. Available online at: http://yann.lecun.com/exdb/mnist/

Levy, N., Horn, D., Meilijson, I., and Ruppin, E. (2001). Distributed synchrony in a cell assembly of spiking neurons. *Neural Netw.* 14, 815–824. doi: 10.1016/S0893-6080(01)00044-2

Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postysnaptic AP and EPSP. *Science* 275, 213–215. doi: 10.1126/science.275.5297.213

Miller, K., Keller, J. B., and Stryker, M. P. (1989). Ocular dominance column development: analysis and simulation. *Science* 245, 605–615. doi: 10.1126/science.2762813

Morrison, A., Aertsen, A., and Diesmann, M. (2007). Spike-timing dependent plasticity in balanced random networks. *Neural Comput.* 19, 1437–1467. doi: 10.1162/neco.2007.19.6.1437

Munos, R. (2005). "Geometric variance reduction in Markov chains: application to value function and gradient estimation," in *AAAI'05 Proceedings of the 20th National Conference on Artificial Intelligence*, Vol. 2, (Pittsburgh, PA: AAAI Press), 1012–1017. Available online at: http://dl.acm.org/citation.cfm?id=1619410.1619495

Nessler, B., Pfeiffer, M., Buesing, L., and Maass, W. (2013). Bayesian computation emerges in generic cortical microcircuits through spike-timing-dependent plasticity. *PLoS Comput. Biol.* 9:e1003037. doi: 10.1371/journal.pcbi.1003037

Nessler, B., Pfeiffer, M., and Maass, W. (2009). Stdp enables spiking neurons to detect hidden causes of their inputs. *Response* 22, 1–9.

Ngezahayo, A., Schachner, M., and Artola, A. (2000). Synaptic activation modulates the induction of bidirectional synaptic changes in adult mouse hippocampus. *J. Neurosci.* 20, 2451–2458. Available online at: http://www.biomedsearch.com/nih/Synaptic-activity-modulates-induction-bidirectional/10729325.html

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network* 15, 243–262. doi: 10.1088/0954-898X/15/4/002

Pecevski, D., Buesing, L., and Maass, W. (2011). Probabilistic inference in general graphical models through sampling in stochastic networks of spiking neurons. *PLOS Comput. Biol.* 7:e1002294. doi: 10.1371/journal.pcbi.1002294

Pfister, J., Toyoizumi, T., Barber, D., and Gerstner, W. (2006). Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural Comput.* 18, 1318–1348. doi: 10.1162/neco.2006.18.6.1318

Pillow, J., Paninski, L., and Simoncelli, E. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire model. *Adv. Neural Inf. Process. Syst.* 16, 1311–1318.

Ranganath, C., and Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* 4, 193–202. doi: 10.1038/nrn1052

Salakhutdinov, R., and Hinton, G. E. (2009). "Deep Boltzmann machines," in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS)*, Vol. 5, eds David A. Van Dyk and M. Welling (Clearwater Beach, FL), 448–455. Available online at: http://www.jmlr.org/proceedings/papers/v5/salakhutdinov09a.html

Savin, C., Joshi, P., and Triesch, J. (2010). Independent component analysis in spiking neurons. *PLOS Comput. Biol.* 6:e1000757. doi: 10.1371/journal.pcbi.1000757

Schultz, W. (2008). A neural substrate of prediction and reward. *Science* 275, 1593–1599. doi: 10.1126/science.275.5306.1593

Shao, L. Y. (2012). Linear-nonlinear-poisson neurons can do inference on deep boltzmann machines. *CoRR*. Arxiv.org/abs/1210.8442.

Sjöström, P., Turrigiano, G., and Nelson, S. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* 32, 1149–1164. doi: 10.1016/S0896-6273(01)00542-6

Song, S., and Abbott, L. (2001). Column and map development and cortical remapping through spike-timing dependent plasticity. *Neuron* 32, 339–350. doi: 10.1016/S0896-6273(01)00451-2

Song, S., Miller, K., and Abbott, L. (2000). Competitive Hebbian learning through spike-time-dependent synaptic plasticity. *Nat. Neurosci.* 3, 919–926. doi: 10.1038/78829

Sutskever, I., Hinton, G. E., and Taylor, G. W. (2008). The recurrent temporal restricted boltzmann machine. *Adv. Neural Inf. Process. Syst.* 21, 1601–1608.

Young, J., Waleszczyk, W., Wang, C., Calford, M., Dreher, B., and Obermayer, K. (2007). Cortical reorganization consistent with spike timing–but not correlation-dependent plasticity. *Nat. Neurosci.* 10, 887–895. doi: 10.1038/nn0807-1073c

Yu, A. J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron* 46, 681–692. doi: 10.1016/j.neuron.2005.04.026

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 November 2013; accepted: 17 March 2014; published online: 04 April 2014.

Citation: Jimenez Rezende D and Gerstner W (2014) Stochastic variational learning in recurrent spiking networks. Front. Comput. Neurosci. ***8****:38. doi: 10.3389/fncom.2014.00038*

This article was submitted to the journal Frontiers in Computational Neuroscience.