“The Hamiltonian Brain” reviewed

Simon Hugo
Master MVA
hugo.simon@telecom-paris.fr

Abstract

Due to ambiguity in the inputs and intrinsic unpredictability in the environment, our brain operates in a state of high uncertainty. Bayesian inference is known to be the best approach in such conditions and it seems, according to behavioral and neural observations, that the brain frequently approximate it in numerous of its functions. We showcase Excitatory-Inhibitory Hamiltonian Monte Carlo network (EI HMC), a relatively simple algorithm introduced in "The Hamiltonian brain" article, that performs highly efficient Bayesian inference for Gaussian Scale Mixture (GSM), a generative model for natural images. This network displays several observed properties of cortical activities such as oscillations and balanced activations. Our personal contribution is to provide a broader insight on the EI HMC network, especially on observation reconstruction, reaction time asymmetry and transient firing rate increases, applying it on more realistic images.

1 Introduction

1.1 Reasoning under uncertainty

Uncertainty plagues neural computation. Our brain operates in the face of substantial uncertainty due to ambiguity in the inputs, and inherent unpredictability in the environment. In such condition, reasoning under uncertainty is crucial for survival. Most efficient way known to do so may be Bayesian inference i.e. representing as a probability distribution the ability of all considered scenarios to describe an actual state of the world, and update it according to the laws of probability, see Figure 1. Moreover, some human behaviors are known to be consistent with Bayesian inference in many sensory, motor and cognitive tasks (Gopnik et al. 2004. Thus it motivates the idea that neuron populations implement Bayesian inference in order to both explain what neurons compute and give functional meaning to some empirical observations.

Figure 1: Principle of Bayesian inference. Prior is updated through data likelihood into posterior.

1.2 How to implement Bayesian inference ?

However, traditional models of neural computation neglect recovering whole posterior distribution and instead favor circuit dynamics that seek for single best explanations of their inputs, such as Maximum a posteriori (MAP) or Mean a posteriori (MMSE). But if for instance the probability of an animal being unsafe is lower than being safe, then Maximum a posteriori would conclude animal is harmless. Thus, in the case of Figure 2 of bi-modal or heavy-tailed posterior, relying only on single best explanations instead of **probable size effect** easily prove fatal.

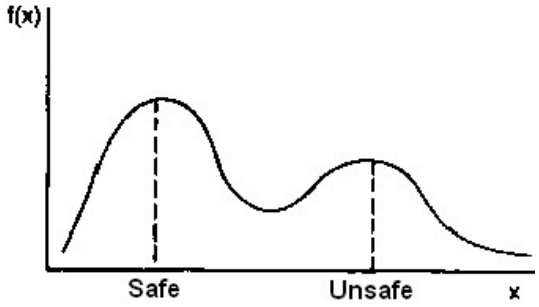


Figure 2: Under such bi-modal posterior distribution, MAP estimator would conclude to safeness with fatal consequences.

Secondly, current neural models don't capture the rich dynamics of cortical responses. Mainly, they don't display **prominent intrinsic oscillations** of neural activities and **large transient changes in response to stimulus onset** observed in V1 and other cortical areas.

Finally, they **typically violate Dales's principle** by having neurons with both excitatory and inhibitory outputs. And though some excitatory-inhibitory (EI) neural network models have been shown to capture some of these elements of cortical dynamics, they have seldom been connected to any specific computation, much alone probabilistic inference.

In the following, we define a statistical generative model for natural visual scenes to infer from. We then describe an HMC-based EI neural network that sample the posterior of this generative model. As our personal contribution, simulations finally illustrate this sampler ability to reproduce experimentally observed cortical dynamics properties on realistic face-like data.

2 Methodology

2.1 The Gaussian Scale Mixture model

The Gaussian Scale Mixture (GSM) model is a widely used generative model that captures some fundamental higher-order statistical properties of natural images (see Wainwright et al. 1999). It is defined by 1 observed image $\mathbf{x}$, 2 latent variables $\mathbf{u}$ and z, a noise variance $\sigma^2_{\mathbf{x}}$, and some edge-like features $\mathbf{A}$. Its graphical model is given by Figure 3.A and its sampling procedure is

$$
\begin{aligned}
&\mathbf{u} \sim \mathcal{N}(0, C) \quad \text{where} \quad C = (1 - \sigma^2_{\mathbf{x}})(\mathbf{A}^\top \mathbf{A})^{-1} \\
&z \sim |\mathcal{N}(0, 1)| \\
&\mathbf{x} \mid \mathbf{u}, z \sim \mathcal{N}\left(z\mathbf{A}\mathbf{u}, \sigma^2_{\mathbf{x}}\mathbf{I}\right)
\end{aligned}
$$

$|\mathcal{N}(0, 1)|$ is the truncated under 0 univariate normal random variable i.e. the absolute value of a normally distributed random variable. Note also that C describes prior covariance of $\mathbf{u}$ and is fitted to whitened data.

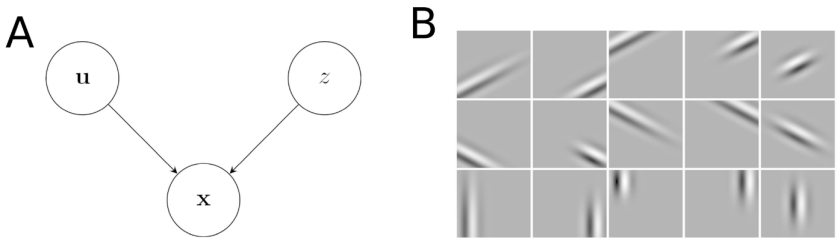


Figure 3: **A.** GSM graphical model. **B.** example of edge-like features **A**.

Interestingly, it turns out latent variable **u** is known to account for *stationary* responses of V1 neurons. More precisely, the posterior mean of **u** matches the across-trial average responses of simple cells in V1. Conversely, latent variable z may account for complex-cell activations.

2.2 MCMC framework

Markov Chain Monte Carlo (MCMC) methods are a class of methods for sampling from probability distributions. They are based on sequentially sampling new points from previous ones such that the actual dynamic follows a Markov chains that have the target distribution as stationary laws. The higher the mixing rate of the Markov chain, the faster the inference.

One classical example of MCMC is the Metropolis-Hastings (MH) algorithm proposed by Metropolis et al. 1949 and illustrated in 4. It consists in proposing a new sample in the neighborhood of the current one, and accept it with some probability that depends on this sample probability for the target distribution.

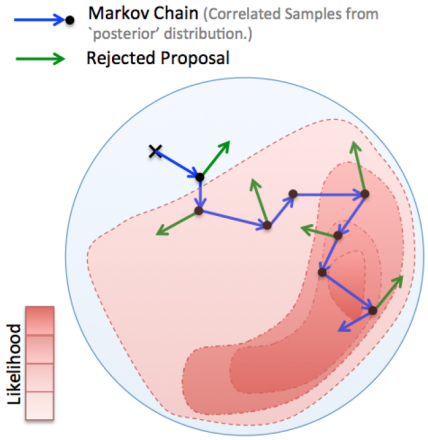


Figure 4: MCMC as Metropolis-Hastings.

2.3 Hamiltonian approach

Hamiltonian Monte Carlo (HMC) is a variation of MH algorithm that holds on a simple relevant idea proposed by Duane et al. 1987. Let's say we want to sample from the target density $\boldsymbol{q} \mapsto f(\boldsymbol{q})$. Then one can introduce a latent space $\boldsymbol{q}$ as momentum and treat the whole as an Hamiltonian system.

Formally, we apply MH algorithm on the phase state $\boldsymbol{s} = (\boldsymbol{q}, \boldsymbol{p})$, and the new proposal is given by the phase state at some time T, applying on current phase state the following Hamiltonian dynamic

$$\dot{\boldsymbol{q}} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}}, \quad \dot{\boldsymbol{p}} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}} \quad \text{with for instance} \quad \mathcal{H}(\boldsymbol{q}, \boldsymbol{p}) = -\ln f(\boldsymbol{q}) + \frac{1}{2}\|\boldsymbol{p}\|^2_{M^{-1}},$$

This chain admits the stationary joint distribution

$$(\boldsymbol{q}, \boldsymbol{p}) \mapsto f(\boldsymbol{q})\,\mathcal{N}(\boldsymbol{p} \mid 0, M)$$

which by marginalizing over $\boldsymbol{p}$ allows us to sample from the target density f.

The gain by doing so is that Hamiltonian dynamic proposes moves to distant states which maintain high acceptance probability due to approximate energy conserving property. HMC reduces the correlation between successive samples and yields high mixing rate compared to classical MH as illustrated in Figure 5.

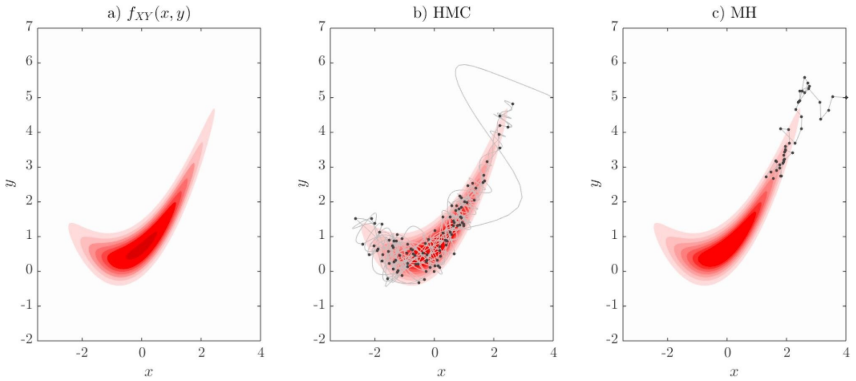


Figure 5: HMC yields higher mixing rate and acceptance ratio than MH.

2.4 Implementing HMC

The main contribution of Aitchison et al. 2014 is to provide the following Excitatory-Inhibitory (EI) network

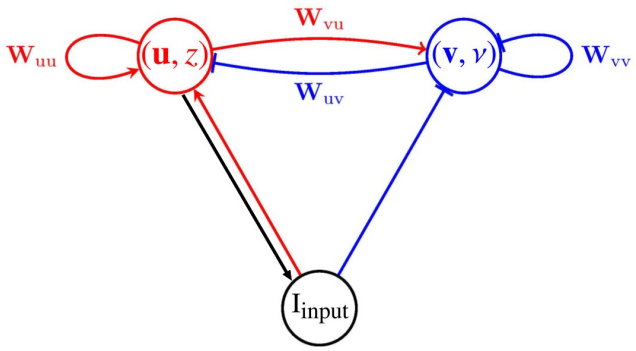


Figure 6: EI Hamiltonian network architecture.

$\mathbf{u}$ and z are our GSM latent variables and we introduce $\mathbf{v}$ and ν their momentum. They obey the following coupled Stochastic Differential Equation (SDE)

$$\dot{\mathbf{u}} = \frac{1}{\tau}\left(\mathbf{W}_{uu}\mathbf{u} - \mathbf{W}_{uv}\mathbf{v} + \frac{1}{2}\tau\rho^2\mathbf{I}_{input}\right) + \rho\boldsymbol{\eta}_u \qquad \dot{z} = \frac{1}{\tau}\left(W_{zz}z - W_{zv}\nu + \frac{1}{2}\tau\rho^2 I_{input}\right) + \rho\eta_z$$

$$\dot{\mathbf{v}} = \frac{1}{\tau}\left(\mathbf{W}_{vu}\mathbf{u} - \mathbf{W}_{vv}\mathbf{v} - \mathbf{I}_{input}\right) + \rho\boldsymbol{\eta}_v \qquad \dot{\nu} = \frac{1}{\tau}\left(W_{vz}z - W_{vv}\nu - I_{input}\right) + \rho\eta_\nu$$

where

$$\mathbf{I}_{input} = \frac{z}{\sigma_{\mathbf{x}}^2}\mathbf{A}^T(\mathbf{x} - z\mathbf{A}\mathbf{u}) - \mathbf{C}^{-1}\mathbf{u}$$

where

$$I_{input} = \frac{1}{\sigma_{\mathbf{x}}^2}(\mathbf{A}\mathbf{u})^T(\mathbf{x} - z\mathbf{A}\mathbf{u}) - z$$

where $\tau, \rho, \mathbf{W}$ are parameters and $\eta_\cdot$ is Brownian motion.

Note that contrary to what authors might suggest, $\mathbf{I}_{\text{input}}$ is not an exogenous input but depends on $(\mathbf{u}, z)$ and thus is not fully Excitatory-Inhibitory and linear because of, among others, the $z\mathbf{Au}$ term.

By setting parameters to the right range of values, this EI network then implements HMC for the GSM posterior distribution! It is actually a noised version of HMC where proposal acceptance is replaced by Brownian noise. The SDE still converges toward the target distribution but its discretized version might differ a bit. Also, when $\mathbf{W}$ is set to 0, this noised HMC reduces to Langevin Monte Carlo algorithm.

3 Results

3.1 Face-like features

Aitchison et al. 2014 use GSM with edge-like features as Gabor filters to capture statistics of natural image that easily implementable, but they concede a not very realistic looking. In our experiments, we use face-like features created by Non-Negative Matrix Factorization using classical Multiplicative Update (MU) algorithm on a face dataset. Figure 7 shows a sample of these face features where one might recognize cheeks, chin, eyebrows, periorbital skin, nose and forehead.

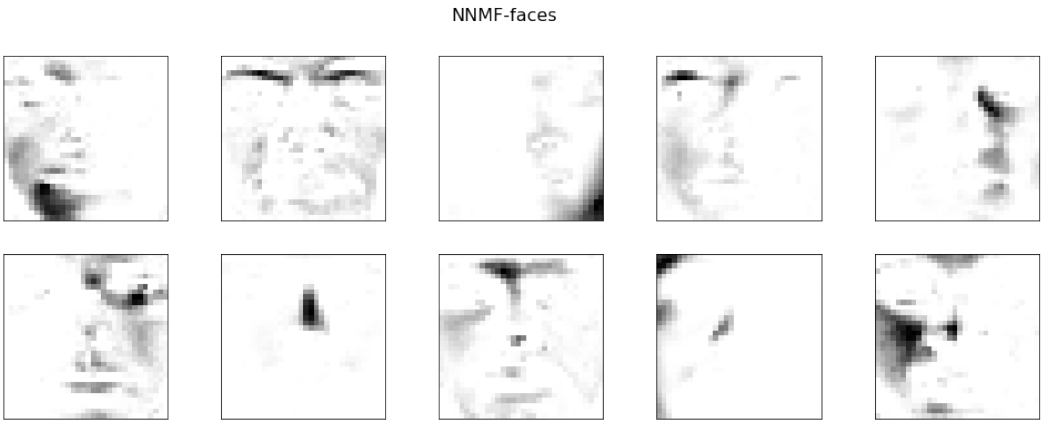


Figure 7: Face-like features obtained from NNMF of a face dataset.

3.2 Stimulus definition

Recall the GSM generative model being

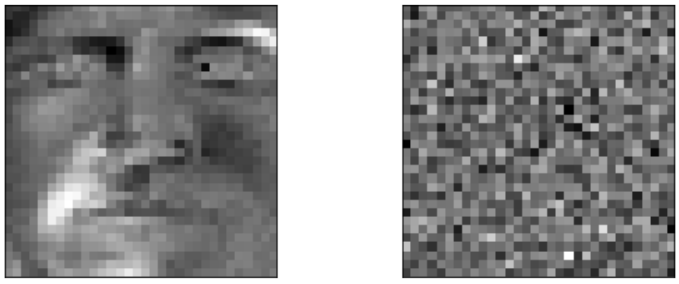
$$\begin{aligned}\mathbf{u} &\sim \mathcal{N}(0, C)\\ z &\sim |\mathcal{N}(0,1)|\\ \mathbf{x} \mid \mathbf{u}, z &\sim \mathcal{N}\left(z\mathbf{Au}, \sigma_{\mathbf{x}}^2\mathbf{I}\right)\end{aligned}$$

$\mathbf{u}$ acts then as an underlying signal selecting features to show, and z is acting as stimulation switch. Indeed, if z is close to 0, then the variance $\sigma_{\mathbf{x}}^2$ is great compared to $z\mathbf{Au}$ thus the image is almost a white noise as in figure 8b, there is no stimulation. And conversely, if z is large, variance become negligible and the returned image is almost $z\mathbf{Au}$ as in figure 8a. We define it as stimulus.

3.3 Stimulus vs. no-stimulus response

We run our EI HMC network under no-stimulus and stimulus input, with $\dim(\mathbf{u}) = 15$, and we compare it to EI Langevin (setting $\mathbf{W}$=0).

One can see on Figure 9 that under no-stimulus input, EI HMC has trouble to approach the true $\mathbf{u}$ value but fastly identify the true z for being close to 0. This is because under the GSM model, if a



(a) Stimulus (b) No-stimulus

noisy image is shown, it is far more probable this is due to low value of z than because of a noisy signal $\mathbf{u}$. But then the model infer $z\mathbf{Au}$ is close to 0 so $\mathbf{u}$ is highly uncertain. Put differently, the network is able to know that it does not know, and that it is due to inactivation of the perception rather than noisy signal.

Conversely, under stimulus input, EI HMC fastly sample from low uncertainty posterior on both $\mathbf{u}$ and z. This is because under the GSM model, if a non-noise image is observed, it can only be because z is large, variance $\sigma_{\mathbf{x}}^2$ is negligible, then it can rather precisely infer the right $\mathbf{u}$.

Compared to EI Langevin, the inference is way faster and does present oscillations in neural response, 2 considerations that are actually linked. This is the introduction of a momentum latent variable and the dynamic into the induced phase space, that is responsible of these oscillations and allows HMC to perform better than Langevin.

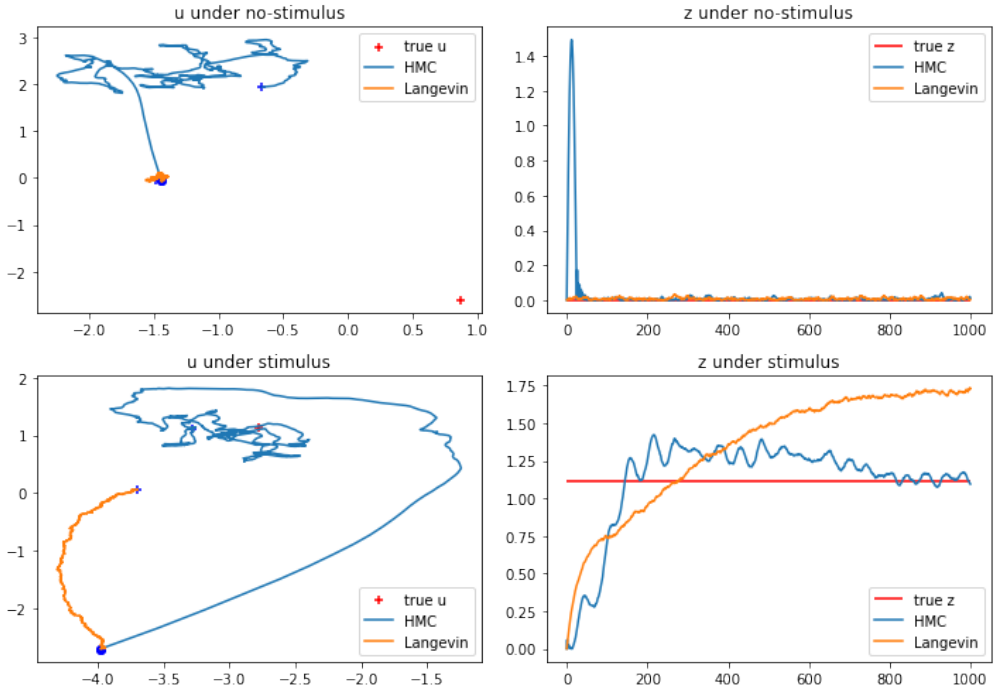
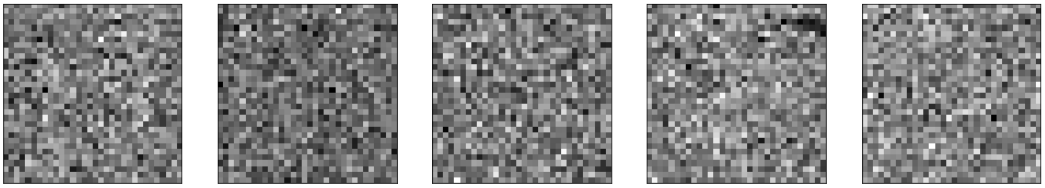


Figure 9: EI HMC and EI Langevin behaviors under no-stimulus and stimulus input. Are shown only 2 random dimensions of $\mathbf{u}$.

3.4 Stimulus reconstruction

Additionnaly, we may ask how the reconstructed observation from inferred variables looks like. Figure 10 show images generated from $\mathbf{u}$ and z at 5 different steps of the previously plotted sampling process. We clearly see it is able to reconstruct rapidly the observed image under stimulus and noise under no-stimulus.

(a) Stimulus reconstruction throughout sampling.



(b) No-stimulus reconstruction throughout sampling.

Figure 10: Reconstructed observations from inferred variables.

Note that in practice, decision is not made from a single sample at a given step but a distribution of samples across time. This is the frequency among samples that estimates posterior probability and allows to consider uncertainty.

3.5 Stimulus onset and stimulus offset

We are especially interested activity changes under stimulus onset and offset. Figure 11 shows EI HMC and EI Langevin responses under stimulus onset (at $t = 333ms$) then offset (at $t = 666ms$). One aspect Aitchison et al. 2014 does not emphasize is the clear asymmetry in time responses for deciding whether there is or not stimulus i.e. inferring the z value. Classical decision making models such as drift-diffusion to bound processes (Duncan 1986, Bogacz et al. 2006) admits equal reaction times among alternatives, that is not always observed in practice. In our case, the asymmetry is naturally induced by the generative model we infer from.

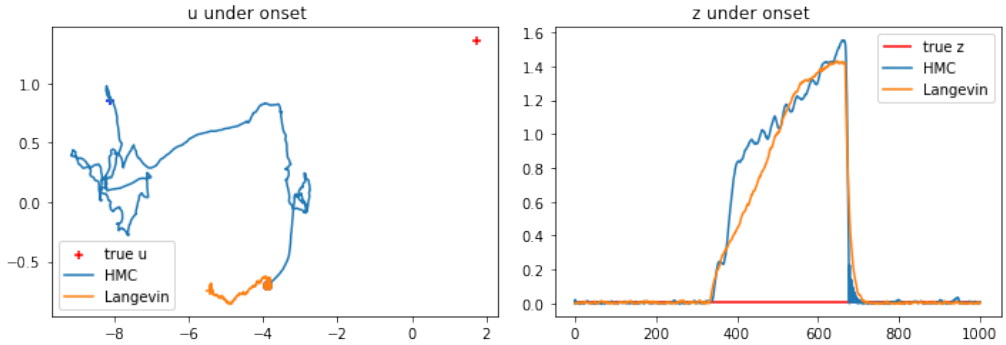


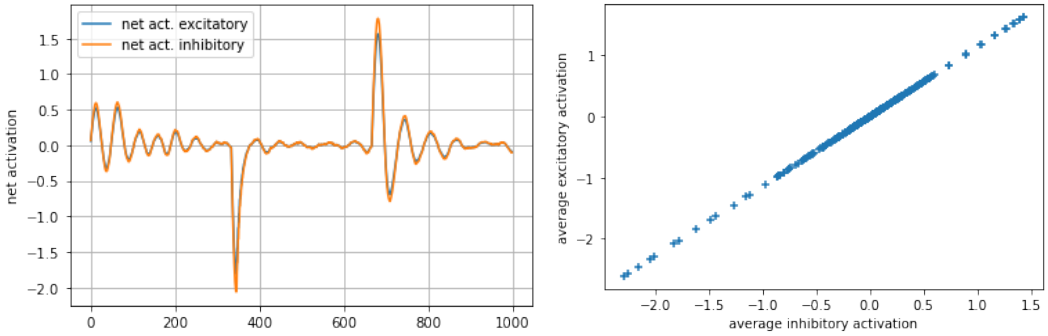
Figure 11: EI HMC and EI Langevin behaviors under stimulus onset and offset. Are shown only 2 random dimensions of $\mathbf{u}$.

3.6 Excitation-Inhibition balance

Finally, we may look at neural excitation and inhibition activities, and how they balance. In our EI network, $\mathbf{W}_{uu}\mathbf{u} - \mathbf{W}_{uv}\mathbf{v}$ and $\mathbf{W}_{vu}\mathbf{u} - \mathbf{W}_{vv}\mathbf{v}$ are respectively activation of excitatory cells and inhibitory cells. For stable sampling from the posterior, we expect excitation and inhibition to be quasi-balanced.

Of course, in EI Langevin, $\mathbf{W}$ always 0 so activations are balanced because they are not. As for EI HMC, Figure 12 shows net excitation and net inhibition can be highly variable but still quite balanced. This is coherent with sensory-evoked cortical activities empirically shown by Okun et al. 2008.

Crucially, Figure 12a also shows transient increases in firing rates upon stimulus onset and offset which fits with numerous empirical observations (Müller et al. 2001).



(a) excitatory and inhibitory inputs among sampling.

(b) excitatory inputs vs. inhibitory inputs.

Figure 12: Excitation-inhibition balance of z component.

4 Conclusion and perspectives

In what precedes, we introduced EI HMC, a relatively simple algorithm that performs highly efficient Bayesian inference for GSM, a generative model for natural images. This algorithm can be implemented by an (almost) EI network and thus complies with Dales's principle. Moreover, its behavior displays 4 qualitative properties observed in empirical measures and that are not fully supported by other classical models. First, it shows prominent intrinsic and cortical oscillations which would be due to phase space exploration and thus are given functional explanation. Second, there is asymmetry in the process of inferring presence or absence of a stimulus. Third, it has balanced excitation and inhibition. Fourth, there are transient increases in firing rates upon stimulus onset and offset. Our personal contribution was to provide a broader insight on the EI HMC network, especially on observation reconstruction, reaction time asymmetry and transient firing rate increases, applying it on more realistic images.

However, EI HMC architecture mainly rely on GMS model. But despite its accurate modeling of many statistical properties, and even with more realistic features such as face-like features, GMS might not capture some natural images notions such as object and obstruction. Thus one may ask how EI HMC should be adapted to other more expressive generative models. Furthermore, it is not clear how non-linearities appearing in $\mathbf{I}_{\text{input}}$ should be implemented by actual neurons. Finally, it would also be important to understand how local learning rules might be able to set up weights needed for implementing EI HMC.

References

Aitchison, Laurence et al. (2014). *The Hamiltonian brain: efficient probabilistic inference with excitatory-inhibitory neural circuit dynamics*. DOI: 10.48550/ARXIV.1407.0973. URL: https://arxiv.org/abs/1407.0973.

Bogacz, Rafal et al. (Nov. 2006). "The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks." In: *Psychological review* 113, pp. 700–65. DOI: 10.1037/0033-295X.113.4.700.

Duane, Simon et al. (1987). "Hybrid Monte Carlo". In: *Physics Letters B* 195.2, pp. 216–222. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(87)91197-X. URL: https://www.sciencedirect.com/science/article/pii/037026938791197X.

Duncan, Luce R. (1986). *Response times : their role in inferring elementary mental organization*. eng. Oxford psychology series. New York Oxford: Oxford University Press Clarendon Press.

Gopnik, Alison et al. (Feb. 2004). "A Theory of Causal Learning in Children: Causal Maps and Bayes Nets". In: *Psychological review* 111, pp. 3–32. DOI: 10.1037/0033-295X.111.1.3.

Metropolis, Nicholas et al. (1949). "The Monte Carlo Method". In: *Journal of the American Statistical Association* 44.247, pp. 335–341.

Müller, James R. et al. (2001). "Information Conveyed by Onset Transients in Responses of Striate Cortical Neurons". In: *Journal of Neuroscience* 21.17, pp. 6978–6990. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.21-17-06978.2001. eprint: https://www.jneurosci.org/content/21/17/6978.full.pdf. URL: https://www.jneurosci.org/content/21/17/6978.

Okun, Michael et al. (June 2008). "Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities". In: *Nature neuroscience* 11, pp. 535–7. DOI: 10.1038/nn.2105.

Wainwright, Martin J et al. (1999). "Scale Mixtures of Gaussians and the Statistics of Natural Images". In: *Advances in Neural Information Processing Systems*. Ed. by S. Solla et al. Vol. 12. MIT Press. URL: https://proceedings.neurips.cc/paper/1999/file/6a5dfac4be1502501489fc0f5a24b667-Paper.pdf.