Large Deviations of an Ergodic Synchronous Neural Network with Learning

Olivier Faugeras, James Maclaurin

Large Deviations of a Stationary Neural Network with Learning

Olivier Faugeras and James MacLaurin,

NeuroMathComp INRIA
2004 Route Des Lucioles
B.P. 93, 06902, Sophia Antipolis France
e-mail: olivier.faugeras@inria.fr; james.maclaurin@inria.fr

Abstract: In this work we determine a Large Deviation Principle (LDP) for a model of neurons interacting on a lattice $\mathbb{Z}^d$. The neurons are subject to correlated external noise, which is modelled as an infinite-dimensional stochastic integral. The probability law governing the noise is strictly stationary, and we are therefore able to find a LDP for the probability laws Π^n governing the stationary empirical measure $\hat{\mu}^n$ generated by the neurons in a cube of length $(2n+1)$. We use this LDP to determine an LDP for the neural network model. The connection weights between the neurons evolve according to a learning rule / neuronal plasticity, and these results are adaptable to a large variety of neural network models. This LDP is of great use in the mathematical modelling of neural networks, because it allows a quantification of the likelihood of the system deviating from its limit, and also a determination of which direction the system is likely to deviate. The work is also of interest because there are nontrivial correlations between the neurons even in the asymptotic limit, thereby presenting itself as a generalisation of traditional mean-field models.

MSC 2010 subject classifications: Primary 60F10; secondary 60H20,92B20,68T05,82C32.
Keywords and phrases: Large Deviations, ergodic, neural network, learning, SDE, lattice.

1. Introduction

In this paper we determine a Large Deviation Principle for a strictly stationary model of interacting processes on a lattice. We are motivated in particular by the study of interacting neurons in neuroscience, but this work ought to be adaptable to other phenomena such as mathematical finance, population genetics or insect swarms. In neuroscience, neurons form complicated networks which may be studied on many levels. On the macroscopic level, neural field equations model the density of activity per space / time. They have been very successful in understanding many phenomena in the brain, including visual hallucinations [24, 10], motion perception [31], feature selectivity in the visual cortex [38] and traveling waves [1, 25, 55, 48, 39, 28, 7]. On the microscopic level, models such as that of Hodgkin and Huxley explain the dynamics of action-potentials very accurately. One of the most important outstanding questions in mathematical neuroscience is a detailed and mathematically rigorous derivation of the macroscopic from the microscopic equations [9, 53]. In particular, perhaps two of the most difficult phenomena to model are the nature of the connection strengths between the neurons, and the stochastic noise. We will discuss these further below, but before we do this we provide a brief introduction to mean-field models of neuroscience.

Classical mean-field models are perhaps the most common method used to scale up from the level of individual neurons to the level of populations of neurons [3, 53]. For a group of neurons indexed from 1 to N, the evolution equation of a mean field model is typically of the following form (an $\mathbb{R}$-valued SDE)

$$dX_t^j = g(X_t^j)dt + \frac{1}{N}\sum_{k=1}^{N} h(X_t^k)dt + \sigma(X_t^j)dW_t^j. \tag{1}$$

We set $X_0^j = 0$. Here g is Lipschitz, h is Lipschitz and bounded, and σ is Lipschitz. (W^j) are independent Brownian Motions representing internal / external noise. Asymptoting N to ∞, we find that in the limit X^j is independent of X^k (for $j \neq k$), and each X^j is governed by the same law [52]. Since the (X^j) become more and more independent, it is meaningful to talk of their mean as being representative of the group as a whole. In reaching this limit, three crucial assumptions have been made: that the external synaptic noise is uncorrelated, that the connections between the neurons are homogeneous and that the connections are scaled by the inverse of the size of the system. We will relax each of these assumptions in our model (which is outlined in Section 4).

The noise has a large effect on the limiting behavior, but as already noted it is not necessarily easy to model. Manwani and Koch [43] distinguish three main types of noise in the brain: thermal, channel noise and synaptic noise coming from other parts of the brain. With synaptic noise in particular, it is not clear to what extent this is indeed 'noise', or whether there are correlations or neural coding that we are not yet aware of. At the very least, we expect that the correlation in the synaptic noise affecting two neurons close together should be higher than the correlation in the synaptic noise affecting two neurons a long way apart. The signal output of neurons has certainly been observed to be highly correlated [51, 50, 2]. In our model for the synaptic noise in Section 2.2, the noise is correlated, with the correlation determined by the distance between the neurons. Indeed the probability law for the noise is strictly stationary, meaning that it is invariant under shifts of the lattice, which will allow us to use ergodic theory to determine the limiting behavior.

The other major difference between the model in Section 4 and the mean field model outlined above is the model of the synaptic connections. In the study of emergent phenomena of interacting particles, the nature of the connections between the particles is often more important than the particular dynamics governing each individual [37]. One of the reasons the synaptic connections are scaled by the inverse of the number of neurons is to ensure that the mean-field equation (1) has a limit as $N \to \infty$. However this assumption, while useful, appears a little *ad hoc*. One might expect that the strength of the synaptic connections is independent of the population size, and rather the system does not 'blowup' for large populations because the strength of the connections decays with increasing distance. This is certainly the standard assumption in models of the synaptic kernel in neural field models [9]. Furthermore there is a lot of evidence that the strength of connection evolves in time through a learning rule / neural plasticity [33, 30]. We will incorporate these effects into our model of the synaptic weights, and ensure in addition that they are such that the probability law is ergodic. We note that there already exists a literature on the asymptotic analysis of such ergodically interacting diffusions, including [16, 40, 36].

The major result of this paper is a Large Deviation Principle for the neural network model in Section 4. This essentially gives the exponential rate of convergence towards the limit (see Definition 6). A Large Deviation Principle is a very useful mathematical technique which allows us to estimate finite-size deviations of the system from its limit behaviour. There has been much effort in recent years to understand such finite-size phenomena in mathematical models of neural networks - see for instance [8, 14, 54, 26]. More generally, there has already been considerable work in the Large Deviations of ergodic phenomena. Donsker and Varadhan obtained a Large Deviations estimate for the law governing the empirical process generated by a Markov Process [21]. They then determined a Large Deviations Principle for an (integer-indexed) Stationary Gaussian Process, obtaining a particularly elegant expression for the rate function using spectral theory. [15, 19, 12] obtain a Large Deviations estimate for the empirical measure generated by processes satisfying a 'hyper-mixing' condition. [4] obtain a variety of results for Large Deviations of ergodic phenomena, including one for the Large Deviations of $\mathbb{Z}$-indexed $\mathbb{R}^T$-valued stationary Gaussian processes. There also exists a literature modelling the Large Deviations and other asymptotics of weakly-interacting particle systems (see for example [17, 5, 13, 29, 34, 41, 42, 27]). These are systems of N particles, each evolving stochastically, and usually only interacting via the empirical measure.

The correlations in the noise together with the inhomogeneity of the synaptic weight model mean that the limit equation we obtain in Section 4 is not asynchronous, unlike (1) (see [35] for a discussion of (a)synchronicity). Indeed the neurons are potentially highly correlated, even in the large system limit. This means that the results of this paper would be well-suited for further investigation of stochastic resonance [11, 47, 44, 34]. Furthermore, one may obtain an LDP for the asymptotics of the synaptic weight connections $\Lambda_s^U(j,k)$ through an application of the contraction principle to Theorem 9. This would be of interest in understanding the asymptotics of the network architecture in the large size limit [56].

The paper is structured as follows. In Section 2, we define an infinite-dimensional correlated stochastic process. We use this to model the noise affecting our neural network. In Section 3 we prove an LDP for the set of laws governing the stationary empirical measure corresponding to this network. In Section 4 we outline a stochastic model of a neural network, with the neurons interacting in a stationary manner and subject to correlated stochastic noise. We determine an LDP and an expression for the rate function.

2. A class of Infinite-Dimensional Ornstein-Uhlenbeck Processes

In this section we outline a general infinite-dimensional stochastic differential equation (see Definition 3). This SDE is to be used to model the correlated noise affecting the network of neurons in Section 4. The basic aim is to ensure that the SDE is in as general a form as possible, such that its probability law is invariant under shifts of the lattice, and such that its marginal over any finite set of times is Gaussian. We will obtain a Large Deviation Principle (LDP) governing the stationary empirical measure corresponding to this SDE, and then use this LDP to obtain an LDP for our network of neurons in the next section.

2.1. Preliminaries and Definition of SDE

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, endowed with a filtration $(\mathcal{F}_t)$ $(0 \leq t \leq T)$ satisfying the usual conditions. If X is some topological space, then we denote the σ-algebra generated by the open sets by $\mathcal{B}(X)$, and the set of all probability measures on $(X, \mathcal{B}(X))$ by $\mathcal{M}(X)$. We endow $\mathcal{M}(X)$ with the topology of weak convergence.

Elements of the stochastic processes in this paper are indexed by the lattice points $\mathbb{Z}^d$: for $j \in \mathbb{Z}^d$ we write $(j(1), \ldots, j(d))$. Let $V_n \subset \mathbb{Z}^d$ be such that $j \in V_n$ if $|j(m)| \leq n$ for all $1 \leq m \leq d$. The number of elements in V_n is written as $|V_n| = (2n+1)^d$.

Definition 1. *For each $j \in \mathbb{Z}^d$, let W^j be an α-dimensional vector, for some fixed positive integer α, of independent Wiener process over the time interval $[0, T]$. $W^j_0 = 0$. W^j is independent of W^k for $j \neq k$. The probability law governing a single W^k is denoted by $P \in \mathcal{M}(\mathcal{T})$, and the joint law governing (W^j) is denoted by $P^{\mathbb{Z}^d}$.*

The following definition will be used to construct the coefficients of our stochastic differential equation. Let $\mathbb{M}^\alpha$ be the set of $\alpha \times \alpha$ matrices over $\mathbb{R}$, equipped with the maximum eigenvalue norm - for $D \in \mathbb{M}^\alpha$ we write the norm as $|D|$. We note the conjugate ${}^\dagger D$ and the conjugate transpose ${}^* D$. The state vector at each point $j \in \mathbb{Z}^d$ of our process lies in $\mathbb{R}^\alpha$, for some fixed positive integer α. We endow $\mathbb{R}^\alpha$ with the sup norm $|\cdot|$. We consider $\mathcal{T}$ to be the Banach Space $C([0, T]; \mathbb{R}^\alpha)$ with the norm

$$\|X\| = \sup_{s \in [0,T]} \{|X_s|\}.$$

Definition 2. *Let $\mathfrak{F} \subset \left(C\left([0, T]; \mathbb{M}^\alpha\right)\right)^{\mathbb{Z}^d}$ be the following Banach Algebra. For $C \in \mathfrak{F}$, we define $C^j_{t,*} := \sup_{s \in [0,t]} \left|C^j_s\right|$, and write $C^j_* := C^j_{T,*}$. We require that $\sum_{j \in \mathbb{Z}^d} C^j_* < \infty$ and define $\|C\|_{\mathfrak{F},t} = \sum_{j \in \mathbb{Z}^d} C^j_{t,*}$. The product of elements $C, D \in \mathfrak{F}$ may be written as $(CD)^j_s = \sum_{k \in \mathbb{Z}^d} C^k_s D^{j-k}_s$.*

2.2. Infinite-Dimensional Linear SDE

We now define our stationary correlated Ornstein-Uhlenbeck process and note some of its properties. To economise on space we have avoided the formalism of [49]. Instead we choose the following *moving-average* representation.

Definition 3. *Let $A, C \in \mathfrak{F}$ and $\mathfrak{a} \in L^2([0,T];\mathbb{R}^\alpha)$. We define $Z = (Z^j)$ to be a solution in $\mathcal{T}^{\mathbb{Z}^d}$ to the infinite dimensional stochastic differential equation: for all $j \in \mathbb{Z}^d$,*

$$Z_t^j = \int_0^t \sum_{k\in\mathbb{Z}^d} A_s^{k-j} Z_s^k + \mathfrak{a}(s)ds + \sum_{k\in\mathbb{Z}^d}\int_0^t C_s^{k-j} dW_s^k. \tag{2}$$

By a solution, we mean that $Z \in \mathcal{T}^{\mathbb{Z}^d}$, satisfies (2) $\mathbb{P}$-almost surely and $E\left[\left\|Z^0\right\|^2\right] < \infty$. Existence is proven in Theorem 4, and conditions guaranteeing uniqueness are specified in Lemma 5. We note that the law of any solution is strictly stationary (a property we define at the start of Section 3). In the formalism of [49] this is a strong solution, where $Z, W \in \mathcal{H}_\lambda$ ($\mathcal{H}_\lambda$ is defined in (5)).

Definition 4. *Let $\mathfrak{A}$ be the following subset of $\mathbb{R}^{\mathbb{Z}^d}$. For $(\lambda^j) \in \mathfrak{A}$, we require that $\lambda^j > 0$ for all $j \in \mathbb{Z}^d$ and $\sum_{j\in\mathbb{Z}^d}\lambda^j < \infty$. We term $\mathfrak{A}$ the set of all admissible weights.*

If B is some Banach space, then we let $B^{\mathbb{Z}^d}$ be the product space (indexed over $\mathbb{Z}^d$), endowed with the product topology (that is, the topology generated by the cylinder sets). We let B^{V_j} be the product space indexed by points in V_j, endowed with the product topology. We note the projection operator $\pi^j : B^{\mathbb{Z}^d} \to B^{V_j}$. If $\|\cdot\|_B$ is the norm on B and $(\lambda^j) \in \mathfrak{A}$, then we define $B_\lambda^{\mathbb{Z}^d} = \{b \in B^{\mathbb{Z}^d} : \sum_{j\in\mathbb{Z}^d}\lambda^j \left\|b^j\right\|_B < \infty\}$. We define the following metric on $B_\lambda^{\mathbb{Z}^d}$, noting that the induced topology corresponds with the product topology,

$$d_B^\lambda(b_1, b_2) = \sum_{j\in\mathbb{Z}^d}\lambda^j \left\|b_1^j - b_2^j\right\|_B. \tag{3}$$

For any $C \in \mathfrak{F}$, we define the stochastic integral $Y_t = \int_0^t C_s dW_s$, where $Y = (Y^j)_{j\in\mathbb{Z}^d}$ is a $\mathcal{T}^{\mathbb{Z}^d}$-value random variable, to be

$$Y_t^j = \sum_{k\in\mathbb{Z}^d}\int_0^t C_s^{k-j} dW_s^k. \tag{4}$$

For any $(\lambda^j) \in \mathfrak{A}$, this may easily be shown to define a $\mathcal{H}_\lambda$-valued martingale, where $\mathcal{H}_\lambda$ is the weighted Hilbert Space subset of $(\mathbb{R}^\alpha)^{\mathbb{Z}^d}$ such that the following inner product converges for all $a, b \in (\mathbb{R}^\alpha)^{\mathbb{Z}^d}$

$$\langle a, b\rangle_\lambda = \sum_{j\in\mathbb{Z}^d}\lambda^j a^j \cdot b^j. \tag{5}$$

Indeed we could have alternatively defined this integral using the formalism of Da Prato and Zabczyk [49], whereby (W^j) is considered as a trace-class Q-Wiener Process in $\mathcal{H}_\lambda$.

Lemma 1. *For any $(\lambda^j) \in \mathfrak{A}$, the integral (4) is in $\mathcal{T}_\lambda^{\mathbb{Z}^d}$, $\mathbb{P}$-almost-surely. The marginal of Y_t over a finite set of times $\{t_0, \ldots, t_m\}$ defines a strictly stationary Gaussian sequence over $((\mathbb{R}^\alpha)^{(m+1)})^{\mathbb{Z}^d}$ with zero mean. The $\alpha \times \alpha$ covariance matrix of Y_s^j and Y_t^k is*

$$E\left[Y_s^{j\dagger} Y_t^k\right] = \int_0^{s\wedge t} \sum_{l\in\mathbb{Z}^d} C_r^{l\dagger} C_r^{l+j-k} dr, \tag{6}$$

where we recall that ${}^\dagger C_r^l$ denotes the conjugate.

Proof. The fact that $Y \in \mathcal{T}_\lambda^{\mathbb{Z}^d}$ almost surely follows from the monotone convergence theorem and the finiteness of

$$\sum_{j\in\mathbb{Z}^d} \lambda^j E\left[\left\|Y^j\right\|^2\right] = \sum_{j\in\mathbb{Z}^d} \lambda^j E\left[\sup_{r\in[0,T]} \sup_{1\le\gamma\le\alpha} (Y_r^{j,\gamma})^2\right]$$
$$\le 4 \sum_{j\in\mathbb{Z}^d} \lambda^j \int_0^T \sum_{l\in\mathbb{Z}^d} \mathrm{Trace}(C_r^{l\dagger} C_r^l) dr < \infty.$$

In simplifying the above we have also used (6) and [49, Theorem 3.8]. The Gaussian property of the marginals is standard [6]. □

For each t, (C_t^j) are the coefficients of an absolutely converging Fourier Series. That is, for $\theta \in [-\pi, \pi]^d$, we may define

$$\tilde{C}_s(\theta) = \sum_{j\in\mathbb{Z}^d} \exp\left(-i\langle j, \theta\rangle\right) C_s^j$$
$$C_s^j = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi]^d} \exp\left(i\langle j, \theta\rangle\right) \tilde{C}_s(\theta) d\theta.$$

Remark 2. *The spectral density satisfies, for all $\theta \in [-\pi, \pi]^d$*

$$\sum_{k\in\mathbb{Z}^d} E\left[Y_s^{0\dagger} Y_t^k\right] \times \exp\big(-i\langle k, \theta\rangle\big) := \tilde{Y}_{s\wedge t}(\theta) = \int_0^{s\wedge t} \tilde{C}_r(\theta)^* \tilde{C}_r(\theta) dr.$$

Of course one may always choose C such that $\tilde{C}_r(\theta)$ is the Hermitian square root of $\tilde{C}_r(\theta)^* \tilde{C}_r$ - the law of Y remains the same.

Before we may prove the existence of a solution, we require some preliminary definitions. Let $\Phi(t) \in \mathfrak{F}$ be the solution to the following ordinary differential equation over $\mathfrak{F}$. We stipulate that

$$\Phi^0(0) = \mathrm{Id} \quad \text{and for } k \neq 0 \quad \Phi^{\mathrm{k}}(0) = 0 \tag{7}$$

$$\frac{d}{dt}\Phi^j(t) = \sum_{k\in\mathbb{Z}^d} A_t^k \Phi^{j-k}(t). \tag{8}$$

We let $\Psi \in \mathfrak{F}$ be the unique solution of

$$\Psi^0(0) = \mathrm{Id} \quad \text{ and for } k \neq 0 \quad \Psi^{\mathrm{k}}(0) = 0 \tag{9}$$

$$\frac{d}{dt}\Psi^j(t) = -\sum_{k\in\mathbb{Z}^d} \Psi^k(t) A_t^{j-k}. \tag{10}$$

The first part of the following lemma follows from the theory of ODEs in Banach spaces. The two identities (11)-(12) follows from the application of Fourier transforms to (8) and (10).

Lemma 3. *There exists a unique solution $\Phi \in \mathfrak{F}$ to (7)-(8) and a unique solution $\Psi \in \mathfrak{F}$ to (9)-(10). Moreover $\Psi(t) = \Phi(t)^{-1}$. For all $\theta \in [-\pi,\pi]^d$, $\tilde{\Phi}(t,\theta) := \sum_{j\in\mathbb{Z}^d} \exp(-i\langle j,\theta\rangle)\Phi^j(t)$ is the unique solution to the following matrix-valued ODE*

$$\frac{d}{dt}\tilde{\Phi}(t,\theta) = \tilde{A}_t(\theta)\tilde{\Phi}(t,\theta), \tag{11}$$

with $\tilde{\Phi}(0,\theta) = \mathrm{Id}$ for all θ. Similarly $\tilde{\Psi}(t,\theta)$ is the unique solution to the matrix-valued ODE $\Psi(0,\theta) = \mathrm{Id}$ and

$$\frac{d}{dt}\tilde{\Psi}(t,\theta) = -\tilde{\Psi}(t,\theta)\tilde{A}_t(\theta). \tag{12}$$

One may prove the following through direction substitution into (2).

Theorem 4. *A solution of (2) is, for all $j \in V_n$*

$$Z_t^j = \tilde{\Phi}(t,0)\int_0^t \tilde{\Psi}(s,0)\mathfrak{a}(s)ds + \sum_{k,l\in\mathbb{Z}^d} \Phi^{k-j}(t)\int_0^t (\Psi(s)C_s)^{l-k} dW_s^l. \tag{13}$$

This satisfies (2) $\mathbb{P}$-almost-surely. The mean of Z_t^j is independent of j and given by

$$m_t^Z := \tilde{\Phi}(t,0)\int_0^t \tilde{\Psi}(s,0)\mathfrak{a}(s)ds.$$

The covariance $V_{s,t}^Z(k-j)$ is given by

$$V_{s,t}^Z(k-j) := E\left[\left(Z_s^j - m_s^Z\right)^\dagger \left(Z_t^k - m_t^Z\right)\right] = \sum_{m,p,l\in\mathbb{Z}^d} \Phi(s)^{m-j}\left(\int_0^{s\wedge t} (\Psi(r)C_r)^{l\dagger}(\Psi(r)C_r)^{l+m-p}dr\right)^\dagger \Phi(t)^{k-p}.$$

We denote the law in $\mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$ of (Z^j) by $\mathfrak{P}$. The law governing Z^{V_n} is denoted by $\mathfrak{P}^{V_n}$.

Remark. *The spectral density satisfies*

$$\sum_{k\in\mathbb{Z}^d} V_{s,t}^Z(k)\exp\left(-i\langle k,\theta\rangle\right) = \tilde{\Phi}(s,\theta)\left(\int_0^{s\wedge t} \tilde{\Psi}(r,\theta)\tilde{C}_r(\theta)^*\tilde{C}_r(\theta)^*\tilde{\Psi}(r,\theta)dr\right)^*\tilde{\Phi}(t,\theta).$$

The following definition is of great use throughout the paper for 'commuting' the metric d^λ (as defined in (3)). It also allows us to find the following condition guaranteeing uniqueness.

Definition 5. *Let $\breve{\mathfrak{A}}$ be the following subset of $\mathbb{R}^{\mathbb{Z}^d}$. If $(\omega^j) \in \breve{\mathfrak{A}}$, then there exists a constant $C_\lambda > 0$ and $(\lambda^j) \in \mathfrak{A}$ such that for all $k \in \mathbb{Z}^d$, $\sum_{j\in\mathbb{Z}^d} |\omega^{k-j}|\lambda^j \leq C_\lambda \lambda^k$.*

In Lemma 14 in the Appendix we determine sufficient conditions for membership of $\breve{\mathfrak{A}}$.

Lemma 5. *Suppose that $(A^j_*) \in \breve{\mathfrak{A}}$ (as defined in Definition 5). If $Z, \hat{Z} \in \mathcal{T}^{\mathbb{Z}^d}$ are solutions to (2), with a shared probability space (i.e. for the same (W^j)), then $Z = \hat{Z}$, $\mathbb{P}$-almost-surely.*

Proof. Suppose that Z and $\hat{Z}$ are two solutions, each in $\mathcal{T}^{\mathbb{Z}^d}$. Then $Z^j_t - \hat{Z}^j_t = \int_0^t \sum_{k\in\mathbb{Z}^d} A^{k-j}_s (Z^k_s - \hat{Z}^k_s)ds$ almost-surely. Let (λ^j) satisfy the inequality in definition 5. That is, for some positive constant C_λ,

$$\sum_{k\in\mathbb{Z}^d} \lambda^j A^{k-j}_* \leq C_\lambda \lambda^k. \tag{14}$$

Using (14), we find, after multiplying by the weights λ^j and summing, that $d^\lambda_t(Z,\hat{Z}) \leq \int_0^t C_\lambda d^\lambda_s(Z,\hat{Z})ds$. Gronwall's Lemma ensures that $Z^j_t = \hat{Z}^j_t$ almost-surely. □

3. Large Deviation Principle for the infinite dimensional Ornstein-Uhlenbeck Process

We prove in this section a Large Deviation Principle for the series of laws of the *stationary empirical measures* corresponding to the SPDE defined in Definition 3 (we define these terms below). We were required to make a few extra assumptions on the coefficients A and C. The key result is in Theorem 6. In the following subsection, we determine the specific form of the rate function governing the LDP in Section 4.3 using projective limits.

For some Banach Space B, we note the shift operator $S^j : B^{\mathbb{Z}^d} \to B^{\mathbb{Z}^d}$ (for some $j \in \mathbb{Z}^d$): for $x \in B^{\mathbb{Z}^d}$, let $(S^j(x))^k = x^{j+k}$. Let $\mathcal{M}_s(B^{\mathbb{Z}^d})$ be the space of strictly stationary measures, that is $\mu \in \mathcal{M}_s(B^{\mathbb{Z}^d})$ if and only if $\mu \in \mathcal{M}(B^{\mathbb{Z}^d})$ and $\mu \circ (S^j)^{-1} = \mu$ for all $j \in \mathbb{Z}^d$.

Definition 6. *Let $(\gamma^n)_{n\in\mathbb{Z}^+} \subseteq \mathcal{M}(\mathcal{M}_s(B^{\mathbb{Z}^d}))$ be a series of probability measures, for some Banach Space B. We say that (γ^n) satisfy a Large Deviations Principle (or just LDP) if there exists a function $I : \mathcal{M}_s(B^{\mathbb{Z}^d}) \to [0,\infty)$ such that for all closed sets F,*

$$\overline{\lim_{n\to\infty}} \frac{1}{|V_n|} \log \gamma^n(F) \leq -\inf_{x\in F} I(x)$$

and for all open sets O,

$$\underline{\lim}_{n\to\infty} \frac{1}{|V_n|} \log \gamma^n(O) \geq -\inf_{x\in O} I(x).$$

We term I the rate function. If furthermore the sets $\{x : I(x) \leq t\}$ for $t \geq 0$ are compact, we say that I is a good rate function.

Definition 7. *Let B be some Banach Space. For $x \in B^{\otimes V_n}$, define $x(n) \in B^{\mathbb{Z}^d}$ to be the periodic interpolant. That is $x(n)^j = x^l$, where $l(\delta) = j(\delta) \mod (2n+1)$. The stationary empirical measure (in $\mathcal{M}_s(B^{\mathbb{Z}^d})$) is defined to be*

$$\hat{\mu}^n(x) = \frac{1}{|V_n|} \sum_{j\in V_n} \delta_{S^j \cdot x(n)}.$$

For $y \in B^{\otimes\mathbb{Z}^d}$, we write the empirical measure $\hat{\mu}^n(y) := \hat{\mu}^n(\pi^{V_n}(y))$, where the latter corresponds to the above definition.

Definition 8. *Let $\check{\mathfrak{F}}$ be the set of all $(c^j) \in \mathfrak{F}$ such that the derivative $\dot{c}^j$ exists and is in $\mathcal{T}$ for all $j \in \mathbb{Z}^d$. We also require that*

$$\sum_{j\in\mathbb{Z}^d} \left|\dot{c}^j_*\right| < \infty.$$

We note that $\check{\mathfrak{F}}$ is closed under the Banach Algebra multiplication defined in Definition 3.

Remark. *It is clear that Ψ and Φ are in $\check{\mathfrak{F}}$. If A, C satisfy the assumptions in the theorem below, then the α-mixing rate of $\mathfrak{P}$ can decay much more slowly than the hyper exponential decay in [12, Proposition 2]. See [23, Section 2.1.1] for a discussion of mixing rates.*

The main result of this section is the sequel.

Theorem 6. *Suppose that $C \in \check{\mathfrak{F}}$, and Z satisfies (13). If Π^n is the image law governing $\hat{\mu}^n(W)$, then (Π^n) satisfies an LDP with some good rate function $J : \mathcal{M}_s(\mathcal{T}^{\mathbb{Z}^d}) \to [0, \infty)$.*

Before we prove this theorem, we require some further results. Let $c \in \check{\mathfrak{F}}$, where $\check{\mathfrak{F}}$ is defined in Definition 8. For $t \in [0, T]$, let

$$X_t = \int_0^t c_r dW_r, \tag{15}$$

for independent Brownian Motions $\{W^j\}$ governed by $P^{\mathbb{Z}^d}$, as in Definition 1. Let the law of $\hat{\mu}^n(X)$ be Π^n_c.

Theorem 7. Π^n_c *satisfies a Large Deviation Principle with good rate function.*

Proof. An application of Ito's Formula yields that $P^{\mathbb{Z}^d}$-almost-surely,

$$X_t^k = \sum_{j\in\mathbb{Z}^d}\left(c_t^{j-k}W_t^j - \int_0^t \dot{c}_s^{j-k}W_s^j ds\right),$$
$$\left\|X^k\right\| \leq 2T\sum_{j\in\mathbb{Z}^d}\dot{c}_*^{j-k}\left\|W^j\right\| < \infty.$$

We obtain LDP's for successive approximations of Π_c^n. Define $\mathfrak{c}_{(m)} : \mathcal{T}^{\mathbb{Z}^d} \to \mathcal{T}^{\mathbb{Z}^d}$ as follows: $\mathfrak{c}_{(m)}(x) = \sum_{j\in V_m}\left(c_t^j x_t^{j+k} - \int_0^t \dot{c}_s^j x_s^{j+k} ds\right)$. Let $X_{(m)} = \mathfrak{c}_{(m)}(W)$, and let $Y_{(m)} = X - X_{(m)}$. We observe that $Y_{(m)}^k = \sum_{j\in\mathbb{Z}^d}\left(c_{(m),t}^{j-k}W_t^j - \int_0^t \dot{c}_{(m),s}^{j-k}W_s^j ds\right)$, where $c_{(m)}^l = c^l$ if $l \notin V_m$, and if $l \in V_m$, $c_{(m)}^l = 0$. Let $\Pi_{(m)}^n$ be the law of $\hat{\mu}^n(X_{(m)})$ and let Π_W^n be the law of $\hat{\mu}^n(W)$. Now (Π_W^n) satisfy an LDP with good rate function as a consequence of [19, Theorem 1.3]. In turn, if we define $\mathfrak{c}^{(m)} : \mathcal{M}(\mathcal{T}^{\mathbb{Z}^d}) \to \mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$ via $\mathfrak{c}^{(m)}(\mu) = \mu\circ\mathfrak{c}_{(m)}^{-1}$, we find that for fixed m, $\Pi_B^n\circ(\mathfrak{c}^{(m)})^{-1}$ satisfy an LDP with a good rate function through the Contraction Principle (since $\mathfrak{c}^{(m)}$ is continuous) [18, Theorem 4.2.1].

We may determine an LDP for $\Pi_{(m)}^n$ as follows, using a similar argument to [22, Lemma 2.1]. For $\mu, \nu \in \mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$ and $j \in \mathbb{Z}^+$, let $d_j^{\mathcal{M}}(\mu,\nu)$ be the Prohorov distance between $\pi^{V_j}(\mu)$ and $\pi^{V_j}(\nu)$. We may then define the metric $d^{\mathcal{M}}(\mu,\nu) = \sum_{j=1}^\infty 2^{-j}d_j^{\mathcal{M}}(\mu,\nu)$ on $\mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$. We observe that for any $\omega \in \mathcal{T}^{\mathbb{Z}^d}$, and positive integers l, m, n, o such that $n > m + l + o$, for all $k \in V_o$, $\pi^{V_l}\mathfrak{c}_{(m)}\circ S^k(\omega(n)) = \pi^{V_l}S^k\circ\mathfrak{c}_{(m)}(\omega(n))$. Thus, for all $x \in \mathcal{T}^{\mathbb{Z}^d}$, the absolute variation between $\pi^{V_l}\left(\hat{\mu}^n(x)\circ\mathfrak{c}_{(m)}^{-1}\right)$ and $\pi^{V_l}\hat{\mu}^n(\mathfrak{c}_{(m)}(x))$ is bounded by $\frac{1}{|V_n|}\left(|V_n| - |V_{n-m-l}|\right)$ for $n > m + l$. Since the absolute variation dominates the Prohorov Metric, we find that,

$$d^{\mathcal{M}}\left(\hat{\mu}^n(W)\circ\mathfrak{c}_{(m)}^{-1}, \hat{\mu}^n(X_{(m)})\right) \leq \frac{1}{|V_n|}\left(\sum_{l=1}^{n-m}2^{-l}\left(|V_n| - |V_{n-m-l}|\right) + \sum_{l=n-m+1}^{\infty}2^{-l}\right) \to 0,$$

as $n \to \infty$. Thus for all $\epsilon > 0$ and positive integers m

$$\overline{\lim_{n\to\infty}} P\left(d^{\mathcal{M}}\left(\hat{\mu}^n(X_{(m)}), \hat{\mu}^n(W)\circ\mathfrak{c}_{(m)}^{-1}\right) \geq \epsilon\right) = 0.$$

Accordingly, by [4, Theorem 3.11], $\Pi_{(m)}^n$ satisfies an LDP with good rate function.

To obtain the LDP for (Π_c^n), it suffices to prove (39) in Lemma 15. Consider an arbitrary $b > 0$. Then, making use of Lemma 8, for constants $\kappa_1, \kappa_2 > 0$,

$$E\Big[\exp\big(b\sum_{k\in V_n}\Big\|Y^k_{(m)}\Big\|\big)\Big] \le E\Big[\exp\big(b\sum_{j\in\mathbb{Z}^d}\|W^j\|\big(\sum_{k\in V_n}2T\dot{c}^{j-k}_{(m),*}\big)\big)\Big]$$
$$\le \exp\Big(2T\kappa_1 b\sum_{j\in\mathbb{Z}^d,k\in V_n}\dot{c}^{j-k}_{(m),*} + 4T^2\kappa_2 b^2\sum_{j\in\mathbb{Z}^d}\big(\sum_{k\in V_n}\dot{c}^{j-k}_{(m),*}\big)^2\Big)$$
$$\le \exp\left(2NT\kappa_1 b\sum_{j\in\mathbb{Z}^d}\dot{c}^{j}_{(m),*} + 4NT\kappa_2 b^2\big(\sum_{j\in\mathbb{Z}^d}\dot{c}^{j}_{(m),*}\big)^2\right).$$

Since $\sum_{j\in\mathbb{Z}^d}\dot{c}^j_{(m),*} \to 0$ as $m\to\infty$, we obtain (39). □

Proof of Theorem 6.

Proof. We assume for notational ease that $\mathfrak{a} = 0$. Once we have proved the LDP for the case $\mathfrak{a} = 0$, the LDP for the case $\mathfrak{a} \neq 0$ follows immediately through a Contraction Principle. Define $X_t = \int_0^t (\Psi(r)C_r)dW_r$ and

$$Z^k_{(m)} = \sum_{j\in V_m}\Phi(t)^j X_t^{j+k}. \tag{16}$$

Let $\Phi_{(m)}$ denote the continuous map $\mathcal{T}^{\mathbb{Z}^d} \to \mathcal{T}^{\mathbb{Z}^d}$ corresponding to the above, i.e. such that $Z_{(m)} = \Phi_{(m)}(Y)$. Let the law governing $\hat{\mu}^n(Z_{(m)})$ be $\Pi^n_{Z,(m)}$ and the law governing $\hat{\mu}^n(X)$ be Π^n_X. From Theorem 7, (Π^n_X) satisfy an LDP with a good rate function. Using similar reasoning to the proof of Theorem 7, we may deduce that $(\Pi^n_{Z,(m)})$ satisfy an LDP with good rate function. Let $Y_{(m)} = Z - Z_{(m)}$. To prove the LDP for (Π^n), through Lemma 15 it suffices to prove that for all $\mathfrak{b} > 0$,

$$\lim_{m\to\infty}\overline{\lim_{n\to\infty}}\frac{1}{|V_n|}\log E\left[\exp\left(\mathfrak{b}\sum_{j\in V_n}\Big\|Y^j_{(m)}\Big\|\right)\right] = 0. \tag{17}$$

Through's Ito's Lemma, we find that

$$Y^k_{(m),t} = \sum_{j\notin V_m}\Phi^j(t)\sum_{p\in\mathbb{Z}^d}\left((\Psi C)^{p-j-k}_t W^p_t - \int_0^t (\dot{\Psi C})^{p-j-k}_s W^p_s ds\right).$$

Observe that $\left|(\Psi C)^{p-j-k}_t W^p_t - \int_0^t (\dot{\Psi C})^{p-j-k}_s W^p_s ds\right| \le 2T(\dot{\Psi C})^{p-j-k}_* \|W^p\|$.

Hence

$$E\left[\exp\left(\mathfrak{b}\sum_{k\in V_n}\left\|Y^k_{(m)}\right\|\right)\right] \leq E\left[\exp\left(\mathfrak{b}\sum_{p\in\mathbb{Z}^d}W^p_*\sum_{k\in V_n,j\notin V_m}\Phi^j_* 2T(\dot{\Psi C})^{p-j-k}_*\right)\right]$$

$$\leq \exp\left(\kappa_1\mathfrak{b}\sum_{p\in\mathbb{Z}^d,k\in V_n,j\notin V_m}2T(\dot{\Psi C})^{p-j-k}_*\Phi^j_*+\right.$$

$$\left.4\kappa_2 T\mathfrak{b}^2\sum_{p\in\mathbb{Z}^d}\Big(\sum_{k\in V_n,j\notin V_m}\Phi^j_*(\dot{\Psi C})^{p-j-k}_*\Big)^2\right)$$

$$\leq \exp\left(N\kappa_1\mathfrak{b}\sum_{p\in\mathbb{Z}^d}2T(\dot{\Psi C})^p_*\sum_{j\notin V_m}\Phi^j_*\right.$$

$$\left.+4\kappa_2 T\mathfrak{b}^2\sum_{p\in\mathbb{Z}^d}\sum_{k,l\in V_n}\sum_{j,h\notin V_m}\Phi^j_*\Phi^h_*(\dot{\Psi C})^{p-j-k}_*(\dot{\Psi C})^{p-j-l}_*\right)$$

$$\leq \exp\left(N\kappa_1\mathfrak{b}\sum_{p\in\mathbb{Z}^d}2T(\dot{\Psi C})^p_*\sum_{j\notin V_m}\Phi^j_* + 4N\kappa_2 T\mathfrak{b}^2\Big(\sum_{k\in\mathbb{Z}^d}(\dot{\Psi C})^k_*\Big)^2\Big(\sum_{j\notin V_m}\Phi^j_*\Big)^2\right).$$

Since $\left(\sum_{j\notin V_m}\Phi^j_*\right)\to 0$ as $m\to\infty$, (17) is satisfied. □

Lemma 8. *Suppose that $\upsilon_t = \int_0^t g(s)dW_s$, for $t\in[0,T]$, $g\in L^2[0,T]$ and W a standard Wiener Process. Let $\Sigma = \int_0^T g(s)^2 ds$ and suppose that $\Sigma\leq\Sigma_*$ for some $\Sigma_*>0$. Then there exist positive constants κ_1 and κ_2 such that $E\left[\exp\left(\|\upsilon\|\right)\right]\leq \exp\left(\kappa_1\Sigma+\kappa_2\Sigma^2\right)$.*

Proof. For $j\geq 2$, we have from Doob's Martingale Inequality that

$$E\left[\|\upsilon\|^j\right]\leq\left(\frac{j}{j-1}\right)^j\sup_{s\in[0,T]}E\left[|\upsilon_s|^j\right]=\left(\frac{j}{j-1}\right)^j E\left[|\upsilon_T|^j\right]. \tag{18}$$

Now υ_T is Gaussian, which means that for $k>1$,

$$E\left[|\upsilon_T|^k\right]=\Sigma^k 2^{\frac{k}{2}}\Gamma\left(\frac{k+1}{2}\right)\pi^{-\frac{1}{2}}. \tag{19}$$

Furthermore, from the Burkholder-Davis-Gundy inequality, we have that

$$E\left[\|\upsilon\|\right]\leq\kappa_1\Sigma, \tag{20}$$

for some constant κ_1. Thus, making use of the monotone convergence theorem,

$$E\left[\exp\left(\|\upsilon\|\right)\right] = E\left[\sum_{j=0}^{\infty}\frac{1}{\Gamma(j+1)}\left(\|\upsilon\|\right)^{j}\right]$$
$$\leq 1+\kappa_2\Sigma+\sum_{j=2}^{\infty}\left(\frac{j}{j-1}\right)^{j}\Sigma^{j}2^{\frac{j}{2}}\Gamma\left(\frac{j+1}{2}\right)\Gamma\left(j+1\right)^{-1}\pi^{-\frac{1}{2}}. \tag{21}$$

It is an identity that

$$\Gamma\left(\frac{j+1}{2}\right)\Gamma\left(\frac{j}{2}+1\right) = 2^{-j}\sqrt{\pi}\Gamma(j+1).$$

We choose κ_2 such that for all $k \geq 1$ and $0 < \Sigma \leq \Sigma_*$,

$$\left(\frac{2k}{2k-1}\right)^{2k}\Sigma^{2k}2^{-k}\Gamma(k+1)^{-1}+$$
$$\left(\frac{2k+1}{2k}\right)^{2k+1}\Sigma^{2k+1}2^{-\frac{2k+1}{2}}\Gamma\left(k+\frac{3}{2}\right)^{-1} \leq \kappa_2^{k}\Sigma^{2k}\Gamma(k+1)^{-1}. \tag{22}$$

We thus obtain from (21) that

$$E\left[\exp\left(|\upsilon|\right)\right] \leq 1+\kappa_1\Sigma+\sum_{j=2}^{\infty}\Sigma^{2j}\kappa_2^{j}\Gamma(j+1)^{-1}$$
$$\leq (1+\kappa_1\Sigma)\exp(\kappa_2\Sigma^2) \leq \exp\left(\kappa_1\Sigma+\kappa_2\Sigma^2\right).$$

□

4. Large Deviations of a Stationary Model of Interacting Neurons

In this section we determine a Large Deviation Principle (LDP) for a network of $(2n+1)^d$ neurons as $n \to \infty$. The $(2n+1)^d$ neurons are indexed over V_n, with correlated noise Z, as stated in Theorem 4. The synaptic connection $\Lambda^U(j,k)$ between neurons j and k is independent of the the size of the network, whereas in mean-field models the synaptic connections are usually scaled via the inverse of the size of the network. The system converges as $n \to \infty$ because the connection strength decays as j and k become further apart. The connection strength can evolve in a learning manner. The chief result is in Theorem 9 and in Remark 4.1 we note some examples of the models satisfying the conditions of this theorem. We consider the state space for each neuron at a particular time s to be in $\mathbb{R}^\alpha$ for some fixed integer α. The fundamental evolution equation governing the network of neurons is written in (23). However before we proceed further, we require some preliminary definitions.

4.1. Outline of Model

We outline our finite model of $(2n+1)^d$ stationary interacting neurons indexed over V_n. Let $Z^j_s := (Z^{1,j}_s, \ldots, Z^{\alpha,j}_s)$, $j \in \mathbb{Z}^d$, be a vector of correlated Ornstein-Uhlenbeck Processes (as described in Theorem 6). The evolution equation is, for $j \in V_n$,

$$U^j_t = U_{ini} + \int_0^t \mathfrak{g}_s(U^j_s) + \sum_{k \in V_n} \Lambda^U_s(j, (j+k) \mod V_n) ds + \int_0^t \mathcal{E}(U^j_s) dZ^j_s. \quad (23)$$

Here $(j+k) \mod V_n := l \in V_n$, such that $(j(p)+k(p)) \mod (2n+1) = l(p)$ for all $1 \leq p \leq d$. Thus one may think of the neurons as existing on a torus. U_{ini} is some constant. $\Lambda^U_s(j,l) : \mathcal{T} \times \mathcal{T} \to \mathbb{R}^\alpha$ is a function of U^j and U^l, possibly depending on the past history. It models the effect of presynaptic neuron l on postsynaptic neuron j - see Remark 4.1 for an example of a model for Λ^U. It must satisfy the inequality (25) outlined below. We note that $\mathcal{E}(X) \in \mathbb{M}^\alpha$, such that $\mathcal{E}(X)$ is diagonal, and the p^{th} diagonal element is given by $\mathcal{E}^p(X^p)$), where each $\mathcal{E}^p : \mathbb{R} \to \mathbb{R}$ is assumed to possess a derivative which is continuous and bounded.

Let $u^p : \mathbb{R} \to \mathbb{R}$ $(1 \leq p \leq \alpha)$ satisfy the differential equation

$$\frac{du^p}{dx} = \mathcal{E}^p(u^p).$$

Such a u^p exists as a standard result in ordinary differential equation theory, for any initial condition $u^p(0)$. It does not matter which initial condition is chosen, as long as the criteria below are satisfied. We define $\phi : \mathbb{R}^\alpha \to \mathbb{R}^\alpha$ to be $\phi(x)^p = u^p(x^p)$. Let $\psi : \mathcal{T} \to \mathcal{T}$ be such that $\psi(U)_s = \phi(U_s)$. We also consider ψ to be defined as a map $\mathcal{T}^{V_n} \to \mathcal{T}^{V_n}$ and $\mathcal{T}^{\mathbb{Z}^d} \to \mathcal{T}^{\mathbb{Z}^d}$ by applying the above map to each element. Each of these maps is continuous for the respective topology. We let $\phi^p : \mathbb{R}^\alpha \to \mathbb{R}$ be the obvious projection.

We assume that the following function compositions $\mathbb{R}^\alpha \to \mathbb{R}$ are uniformly Lipschitz continuous (for all $1 \leq p \leq \alpha$ and $s \in [0,T]$),

$$\frac{\mathfrak{g}^p_s \circ \phi}{\mathcal{E}^p \circ \phi^p} \quad , \quad \dot{\mathcal{E}}^p \circ \phi^p. \quad (24)$$

We require that for some $(\Lambda^{max}(j)) \in \check{\mathfrak{A}}$ (see Definition 5 for a definition of $\check{\mathfrak{A}}$), and all $j, k \in \mathbb{Z}^d$

$$\left| \mathcal{E}(\phi(X^j_s))^{-1} \Lambda^{\psi(X)}_s(j,k) - \mathcal{E}(\phi(Y^j_s))^{-1} \Lambda^{\psi(Y)}_s(j,k) \right| \leq \Lambda^{max}(k-j) \left(\min\left(\left\| \mathrm{X^j - Y^j} \right\|_{\mathrm{s}}, 1 \right) + \min\left(\left\| \mathrm{X^k - Y^k} \right\|_{\mathrm{s}}, 1 \right) \right). \quad (25)$$

Here, for $x \in C([0,T];\mathbb{R}^\alpha)$, we have the semi-norm $\|x\|_t = \sup_{s\in[0,t]} |x_s|$, and $|\cdot|$ is the sup-norm on $\mathbb{R}^\alpha$. The weights must also respect the stationarity, i.e. $\Lambda_s^{S^p\cdot Y}(j,k) = \Lambda_s^Y(p+j,p+k)$, for all $p,j,k \in \mathbb{Z}^d$.

Let Π^n be the law of the empirical measure $\hat{\mu}^n(U)$ generated by the solution U of (23) (refer to Definitions 6 and 7 for an explanation of these terms). We determine an expression for the rate function J^U of the following theorem in Theorem 13.

Theorem 9. *The series of laws (Π^n) satisfy an LDP with some good rate function J^U.*

Proof. We claim that the solution U of (23) satisfies $U = \psi(\breve{U})$, where $\breve{U}$ satisfies (28) for the definitions of f, g_s, h outlined as follows. We define $g_s = (g_s^1,\ldots,g_s^p)$, where

$$g_s^p(X) = \frac{\mathfrak{g}_s^p(\phi(X))}{\mathcal{E}^p(\phi^p(X))} - \dot{\mathcal{E}}^p(\phi^p(X)) \sum_{l\in\mathbb{Z}^d}\sum_{1\le q\le\alpha} C_s^{l,pq} C_s^{l,pq}.$$

We define $\breve{\Lambda}_s^X(j,k) := (\mathcal{E}(\phi(X_s^j)))^{-1}\Lambda_s^{\psi(X)}(j,k)$. The identity $U = \psi(\breve{U})$ then follows as a direct consequence of Ito's Lemma.

The induced map $\mathcal{M}(\mathcal{T}^{\mathbb{Z}^d}) \to \mathcal{M}(\mathcal{T}^{\mathbb{Z}^d}) : \mu \to \mu\circ\psi^{-1}$ is clearly also continuous. It is proved in Theorem 10 below that $(\underline{\underline{\Pi}}^n)$ satisfy an LDP. Since $\hat{\mu}^n(U) = \hat{\mu}^n(\breve{U})\circ\psi^{-1}$, a contraction principle dictates that (Π^n) also satisfy an LDP with good rate function.

□

Remark. *We make some comments on the sorts of models satisfying the above conditions. We assume that $\mathcal{E}^p = 1$ for all p for ease. The theory of the previous section may be adapted to a stochastic version of the Hodgkin-Huxley model, as outlined in [3] (for example). It may also be applied to the Fitzhugh-Nagumo model (as outlined in [3] for example) - although care must be taken with the internal dynamics term.*

We now outline a possible model for the synaptic connections. We decompose the effect of neuron k on j into the following product. We write $\Lambda_t^U(j,k) = \bar{\Lambda}_t^U(j,k)\mathfrak{h}(U_t^j)\circ \mathfrak{f}(U_t^k)$. Here $\mathfrak{h},\mathfrak{f}$ are bounded and Lipschitz continuous functions $\mathbb{R}^\alpha\to\mathbb{R}^\alpha$, and $\circ$ is the Hadamard product. $\bar{\Lambda}_t^U(j,k)$ represents the strength of the synaptic connection from presynaptic neuron k to postsynaptic neuron j, whose time evolution may occur as a result of neuronal plasticity or learning. An example of such a model is the following classical Hebbian Learning model (refer to [32] for a more detailed description, and in particular Equation 10.6). Suppose that the maximal connection strength between neurons j and $j+k$ is given by $\Lambda^{max}(k)$, where $(\Lambda^{max}(k)) \in \breve{\mathfrak{A}}$. We assume that the 'activity' of neuron j at time t is given as $\mathfrak{v}(U_t^j)$. Here $\mathfrak{v}:\mathbb{R}^\alpha\to\mathbb{R}$ is Lipschitz continuous, positive and bounded. The evolution equation is

$$\frac{d}{dt}\bar{\Lambda}_t^U(j,k) = \Lambda^{corr}\left(\Lambda^{max}(k-j) - \bar{\Lambda}_t^U(j,k)\right)\mathfrak{v}(U_t^j)\mathfrak{v}(U_t^k) - \Lambda^{dec}\bar{\Lambda}_t(j,k). \tag{26}$$

Here $\Lambda^{corr},\Lambda^{dec}$ are non-negative constants (if we let them be zero then we obtain weights which are constant in time). Initially, we stipulate that

$$\bar{\Lambda}_0(j,k) := \Lambda^{ini}(k-j) \le \Lambda^{max}(k-j), \tag{27}$$

where $\Lambda^{ini}(l) \ge 0$ are constants stipulating the initial strength of the weights.

Other nonlocal learning rules are possible: for a neuroscientific motivation see for example [46, 45, 30]. In brief, one may assume that the synaptic connection $\bar{\Lambda}^U(j,k)$ *is a function of* $\{U^l\}_{l-j\in V_m \text{ or } l-k\in V_m}$, *for some fixed* $m>0$. *We must then redefine the state variable at index point* $j\in\mathbb{Z}^d$ *to be the states of all the neurons in the cube centred at* j *and of side length* $(2m+1)$.

Finally, we note that more general initial conditions, for example a spatially-stationary Gaussian process, would be possible.

4.2. Large Deviation Principle for the Neural Network With Constant Diffusion Coefficient

We prove an LDP for the following simplified system. This LDP is needed to prove Theorem 9. Let $Z^j_s := (Z^{1,j}_s,\dots,Z^{\alpha,j}_s)$, $j\in\mathbb{Z}^d$, be a vector of correlated Ornstein-Uhlenbeck Processes (as described in Theorem 6). We consider a finite-dimensional equation for $j\in V_n$ of the form

$$\breve{U}^j_t = \breve{U}_{ini} + \int_0^t g_s(\breve{U}^j_s) + \sum_{k\in V_n}\breve{\Lambda}^{\breve{U}}_s(j,(j+k)\mathrm{mod V_n})\mathrm{ds} + \mathrm{Z^j_t}. \tag{28}$$

Here $g_s:\mathbb{R}^\alpha\to\mathbb{R}^\alpha$ are uniformly Lipschitz for all $s\in[0,T]$ with Lipschitz constant g_L. $\breve{U}_{ini}$ is a constant. We make the following assumptions on the weights. We require that $\breve{\Lambda}^U_s(j,k)$ is a function of U^j and U^k only, and that $\breve{\Lambda}^{S^p\cdot Y}_s(j,k)=\breve{\Lambda}^Y_s(p+j,p+k)$, for all $p,j,k\in\mathbb{Z}^d$. We assume for all $U,V\in\mathcal{T}^{\mathbb{Z}^d}$, $j,k\in\mathbb{Z}^d$ and $s\in[0,T]$, there exists $(\breve{\Lambda}^{max}(j))\in\breve{\mathfrak{A}}$ (see Definition 5) such that

$$\left|\breve{\Lambda}^U_s(j,k)-\breve{\Lambda}^V_s(j,k)\right|\leq \breve{\Lambda}^{max}(k-j)\left(\min\left(\left\|\mathrm{U^j-V^j}\right\|_\mathrm{s},1\right)+\min\left(\left\|\mathrm{U^k-V^k}\right\|_\mathrm{s},1\right)\right). \tag{29}$$

We write $\breve{\Lambda}^{max}_{sum}:=\sum_{k\in\mathbb{Z}^d}\breve{\Lambda}^{max}(k)$. Let $(\lambda^j)\in\mathfrak{A}$ be a set of admissible weights with $\lambda_{sum}=\sum_{j\in\mathbb{Z}^d}\lambda^j$. The weights are assumed to satisfy the requirement of Definition 5, i.e. for some positive constant C_λ

$$\sum_{j\in\mathbb{Z}^d}\lambda^j\breve{\Lambda}^{max}(k-j)\leq C_\lambda\lambda^k. \tag{30}$$

Let $\underline{\underline{\Pi}}^n$ be the law of the empirical process $\hat{\mu}(\breve{U})$ corresponding to the solution of (28). The infinite-dimensional limit equation, towards which the finite-dimensional system converges, is

$$\breve{U}^j_t=\breve{U}_{ini}+\int_0^t g_s(\breve{U}^j_s)+\sum_{k\in\mathbb{Z}^d}\breve{\Lambda}^{\breve{U}}_s(j,k)ds+Z^j_t. \tag{31}$$

For $Z\in\mathcal{T}^{\mathbb{Z}^d}_\lambda$ we define $\mathfrak{K}(Z)\in\mathcal{T}^{\mathbb{Z}^d}$ to be the unique solution (proved in the lemma just below) of (31). The major result of this section is

Theorem 10. *$\underline{\underline{\Pi}}^n$ satisfies a Large Deviation Principle with good rate function $\breve{J}(\mu) := J(\mathfrak{K}^{-1}(\mu))$. Here J is the rate function in Theorem 6.*

Proof. This follows from [20, Exercise 2.1.20(ii)] and Lemma 12 below. □

The rest of this section is dedicated to the proof of Lemma 12. We define the metric d_t^λ on $(C([0,t];\mathbb{R}^\alpha))_\lambda^{\mathbb{Z}^d}$ to be

$$d_t^\lambda(X,Y) = \sum_{j\in\mathbb{Z}^d} \lambda^j \left\| X^j - Y^j \right\|_t. \tag{32}$$

Here, $(C([0,t];\mathbb{R}^\alpha))_\lambda^{\mathbb{Z}^d}$ is the subspace of $(C([0,t];\mathbb{R}^\alpha))^{\mathbb{Z}^d}$ where the above sum is non infinite. We write d^λ if $t = T$. We let $d^{\lambda,\mathcal{M}}$ denote the Prohorov-Levy metric induced on $\mathcal{M}(\mathcal{T}_\lambda^{\mathbb{Z}^d})$ by d^λ. For $m \in \mathbb{Z}^+$ and $(Z^j) \in \mathcal{T}_\lambda^{\mathbb{Z}^d}$, we define $\mathfrak{K}^m(Z) \in \mathcal{T}^{\mathbb{Z}^d}$ to be the unique solution of

$$\breve{U}_t^j = \breve{U}_{ini} + \int_0^t g_s(\breve{U}_s^j) + \sum_{k\in V_m} \breve{\Lambda}_s^{\breve{U}}(j,j+k)ds + Z_t^j.$$

Lemma 11. *$\mathfrak{K}(Z)$ and $\mathfrak{K}^m(Z)$ are well-defined (for all m), and Lipschitz continuous. They are in $\mathcal{T}_\lambda^{\mathbb{Z}^d}$ almost surely.*

Proof. It suffices to prove this for $\mathfrak{K}$, because $\mathfrak{K}^m$ is clearly a special case.

We prove the existence as follows using Picard Iteration. Fix $S \leq T$ and define, for $t \leq S$,

$$\Gamma(\breve{U})_t^j = \breve{U}_{ini} + \int_0^t g_s(\breve{U}_s^j) + \sum_{k\in\mathbb{Z}^d} \breve{\Lambda}_s^{\breve{U}}(j,k)ds + Z_t^j.$$

We note that $\Gamma(\breve{U}) \in \mathcal{T}_\lambda^{\mathbb{Z}^d}$ because

$$\sum_{m\in\mathbb{Z}^d} \lambda^m \left\| \Gamma(\breve{U})^m \right\| \leq \breve{U}_{ini}\lambda_{sum} + Tg_L d^\lambda(\breve{U},0) \\ + d^\lambda(Z,0) + 2T\alpha\lambda_{sum}\breve{\Lambda}_{sum}^{max}.$$

We obtain the following bound using (29) and the fact that
$\sum_{j\in\mathbb{Z}^d} \lambda^j \left| \breve{\Lambda}^{max}(k) \right| \leq C_\lambda \lambda^k$. We find that, for some $S \in (0,T]$,

$$d_S^\lambda\left(\Gamma(\breve{U}),\Gamma(\breve{V})\right) \leq \left(g_L + \alpha\left(\breve{\Lambda}_{sum}^{max} + C_\lambda\right)\right) S d_S^\lambda(\breve{U},\breve{V}).$$

For S small enough, the above map is contractive, and therefore there exists a unique limit by the Banach Fixed Point Theorem. We may repeat this process over the interval $[S,2S]$, and keep going until we have a unique limit over $\mathcal{T}_\lambda^{\mathbb{Z}^d}$. The uniqueness is proved using Gronwall's Lemma, very similarly to that which follows further below.

We now prove the Lipschitz continuity. Let $\breve{U} = \mathfrak{K}(Z)$ and $\breve{V} = \mathfrak{K}(\hat{Z})$. Then, for $t > 0$,

$$d_t^\lambda(\breve{U},\breve{V}) \leq d_t^\lambda\left(Z,\hat{Z}\right) + \int_0^t g_L d_s^\lambda(\breve{U},\breve{V}) + \sum_{j,k\in\mathbb{Z}^d} \lambda^j |\breve{\Lambda}_s^{\breve{U}}(j,j+k) - \breve{\Lambda}_s^{\breve{V}}(j,j+k)| ds. \quad (33)$$

We bound the above using (29). Thus

$$d_t^\lambda\left(\breve{U},\breve{V}\right) \;\leq\; g_L d_t^\lambda(\breve{U},\breve{V}) + \alpha\left(\breve{\Lambda}_{sum}^{max} + C_\lambda\right)\int_0^t d_s^\lambda(\breve{U},\breve{V})ds + d_t^\lambda\left(Z,\hat{Z}\right).$$

By Gronwall's Inequality,

$$d_t^\lambda(\breve{U},\breve{V}) \leq d_t^\lambda(Z,\hat{Z})\exp\left(t\left(g_L + \alpha\left(\breve{\Lambda}_{sum}^{max} + C_\lambda\right)\right)\right). \quad (34)$$

This gives us the required Lipschitz Continuity (on substituting $t = T$). □

For $\mu \in \mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$, we define (in a slight abuse of notation)

$$\mathfrak{K}^m(\mu) = \mu \circ (\mathfrak{K}^m)^{-1} \quad \text{and} \quad \mathfrak{K}(\mu) = \mu\circ\mathfrak{K}^{-1}.$$

We notice that $\mathfrak{K}^m$, $\mathfrak{K}$ and their inverses map stationary measures to stationary measures.

Lemma 12. *Let $\breve{U} \in \mathcal{T}^{V_n}$ satisfy (28) for some $Z \in \mathcal{T}^{V_n}$. We have that*

$$\mathfrak{K}^n\left(\hat{\mu}^n(Z)\right) = \hat{\mu}^n(\breve{U}). \quad (35)$$

We note that we may also write $\breve{U} = \pi^{V_n}\mathfrak{K}^n(Z(n))$. There exist constants (β_m) $(m \in \mathbb{Z}^+)$, with $\beta_m \to 0$ as $m \to \infty$, such that for all $W \in \mathcal{T}_\lambda^{\mathbb{Z}^d}$, $\mu \in \mathcal{M}(\mathcal{T}_\lambda^{\mathbb{Z}^d})$,

$$d^\lambda\left(\mathfrak{K}(Z),\mathfrak{K}^m(Z)\right) \leq \beta_m,$$
$$d^{\lambda,\mathcal{M}}\left(\mathfrak{K}(\mu),\mathfrak{K}^m(\mu)\right) \leq \beta_m.$$

Proof. For the solution $\breve{U}$ of (28) generated by some Z, we notice that $\mathfrak{K}^n(Z(n)) = \breve{U}(n)$, where these are the periodic interpolants as in Definition 7. We also notice that, for all $k \in \mathbb{Z}^d$,

$$\mathfrak{K}^n(Z(n))^k = \mathfrak{K}^n(Z(n))^{k \mod V_n}$$
$$\mathfrak{K}^n(S^k\cdot(Z(n))) = S^k\cdot\mathfrak{K}^n\left(Z(n)\right).$$

These considerations allow us to conclude (35).

Let $\breve{U} = \mathfrak{K}(Z)$ and $\breve{V} = \mathfrak{K}^m(Z)$. Similarly to (33), we find that

$$d_t^\lambda(\breve{U},\breve{V}) \leq 2t\alpha\sum_{j\in\mathbb{Z}^d, k\notin V_m}\lambda^j\breve{\Lambda}^{max}(k) + \int_0^t g_L d_s^\lambda(\breve{U},\breve{V}) + \sum_{j,k\in\mathbb{Z}^d}\lambda^j|\breve{\Lambda}_s^{\breve{U}}(j,j+k) - \breve{\Lambda}_s^{\breve{V}}(j,j+k)|ds.$$

Applying Gronwall's Inequality, we find that (analogously to (34), and its preceding bounds)

$$d_T^{\lambda}(\breve{U},\breve{V}) \leq 2T\alpha\lambda_{sum}\sum_{k\notin V_m}\breve{\Lambda}^{max}(k)\times \exp\left(T\left(g_L+\alpha\left(\breve{\Lambda}^{max}_{sum}+C_{\lambda}\right)\right)\right).$$

We may take β_m to be the right-hand side of the above.

The last identity in the lemma follows directly from this, as a consequence of the definition of the Prohorov Metric. □

4.3. The Specific Form of the Rate Function in Theorem 9

Let $\sigma = [t_1,\ldots,t_m]$ be a finite series of times in $[0,T]$, such that $t_i < t_{i+1}$. We term σ a partition, and denote the set of all such partitions by $\mathbf{J}$. Let $|\sigma| = \sup_{1\leq i\leq m-1}|t_{i+1}-t_i|$. Let $\pi_\sigma : \mathcal{T} \to (\mathbb{R}^\alpha)^m := \mathcal{T}_\sigma$ be the obvious projection; we naturally extend the definition of $\pi_\sigma : \mathcal{T}^{\mathbb{Z}^d} \to ((\mathbb{R}^\alpha)^m)^{\mathbb{Z}^d}$. We write $\pi_\sigma^{\mathcal{M}} : \mathcal{M}(\mathcal{T}^{\mathbb{Z}^d}) \to \mathcal{M}(\mathcal{T}_\sigma^{\mathbb{Z}^d})$ for the corresponding projection of measures.

Let Π_σ^n be the law of the projection $\pi_\sigma\hat{\mu}^n(U) \in \mathcal{M}(\mathcal{T}_\sigma^{\mathbb{Z}^d})$. Since this is a continuous map, we find through the contraction principle that (Π_σ^n) satisfies an LDP with a good rate function J_σ. We note that [18, Theorem 4.5.10] could be used to find an expression for J_σ.

Theorem 13. $J(\mu) = \sup_{\sigma\in\mathbf{J}} J_\sigma(\mu)$. *For any series of partitions* $\sigma^{(m)} \subseteq \sigma^{(m+1)}$ $(m \in \mathbb{Z}^+)$ *such that* $|\sigma^{(m)}| \to 0$ *as* $m \to \infty$, $J(\mu) = \lim_{m\to\infty} J_{\sigma^{(m)}}(\mu)$.

Proof. We denote the weak topology of $\mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$ by τ_w. We may equip $\mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$ with the projective limit topology τ_{proj}. This is generated by open sets of the form $\pi_\sigma^{-1}(A)$, where A is open in the weak topology of $\mathcal{M}(\mathcal{T}_\sigma^{\mathbb{Z}^d})$. Observe that the σ-algebra generated by this topology is the same as the σ-algebra $\mathcal{B}(\mathcal{M}(\mathcal{T}^{\mathbb{Z}^d}))$ generated by the weak topology. We notice also that the map $U \to \hat{\mu}^n(U)$ is continuous under the projective limit topology. Now the Projective Limit theorem [17, Thm 3.3] yields that (Π^n) satisfy an LDP with good rate function J_{proj}, when $\mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$ is equipped with the projective limit topology. Furthermore $J_{proj}(\mu) = \sup_{\sigma\in\mathbf{J}} J_\sigma(\mu)$. Since (Π^n) are exponentially tight with respect to the weak topology on $\mathcal{M}(\mathcal{T}^{\mathbb{Z}^d})$ (since they satisfy an LDP with good rate function), the Inverse Contraction Principle [18, Thm 4.2.4] together with the uniqueness of the rate function [18, Lemma 4.1.4] yields that the rate functions J and J_{proj} must be the same. Indeed we could have repeated the above argument with the projective limit topology generated by $\pi^{-1}_{\sigma^{(m)}}(A)$, where A is open in $\mathcal{M}(\mathcal{T}_{\sigma^{(m)}}^{\mathbb{Z}^d})$ and $\sigma^{(m)}$ satisfy the conditions in the theorem. This would yield the second result. □

5. Appendix

The Lemma below gives a sufficient condition for membership of $\breve{\mathfrak{A}}$ in Definition 5.

Lemma 14. *Suppose that $(\omega^j) \in \mathbb{R}^{\mathbb{Z}^d}$ satisfies $\omega^j \leq K_\omega \prod_{1\leq\delta\leq d} |j(\delta)|^{-\kappa}$ for some $\kappa > 1$ and $K_\omega > 0$ (we understand that $0^{-\kappa} = 1$). Then $(\omega^j) \in \breve{\mathfrak{A}}$, that is, there exist $(\lambda^j) \in \mathfrak{A}$ and a constant C_ω such that for all $k \in \mathbb{Z}^d$*

$$\sum_{j\in\mathbb{Z}^d} \lambda^j |\omega^{k-j}| \leq C_\omega \lambda^k. \tag{36}$$

Proof. Let $k \in \mathbb{Z}^d$. For $j \in \mathbb{Z}^d$, define $\lambda^j = \prod_{1\leq a\leq d} |j(a)|^{-\kappa}$, noting that $(\lambda^j) \in \mathfrak{A}$. We need to find a bound for the following, holding uniformly for all k,

$$\begin{aligned}\sum_{j\in\mathbb{Z}^d} (\lambda^k)^{-1}\lambda^j\omega^{k-j} &\leq K_\omega \sum_{j\in\mathbb{Z}^d}\prod_{1\leq a\leq d} \frac{|k(a)-j(a)|^{-\kappa}}{|k(a)|^{-\kappa}} |j(a)|^{-\kappa} \\ &\leq K_\omega \prod_{1\leq a\leq d}\sum_{j\in\mathbb{Z}^d} \frac{|k(a)-j(a)|^{-\kappa}}{|k(a)|^{-\kappa}} |j(a)|^{-\kappa}. \end{aligned} \tag{37}$$

It may be seen that it suffices to find an upper bound for the following, holding uniformly for all $\mathbb{Z} \ni p \neq 0$,

$$\sum_{m\in\mathbb{Z}-\{p\}} \left|1-\frac{m}{p}\right|^{-\kappa} |m|^{-\kappa}. \tag{38}$$

(Note that we have excluded the $m = p$ term from the above for notational ease. This is obviously bounded) From the symmetry of the above equation, we may assume that $p \geq 1$. In fact we may assume that $p \geq 2$, as the above sum clearly converges if $p = 1$. Now

$$\sum_{m\leq 0} \left|1-\frac{m}{p}\right|^{-\kappa} |m|^{-\kappa} \leq \sum_{m\leq 0} |m|^{-\kappa} < \infty.$$

Similarly $\sum_{m\geq 2p} \left|1-\frac{m}{p}\right|^{-\kappa} |m|^{-\kappa} < \infty$. Note also that

$$\sum_{p+1\leq m\leq 2p-1} \left|1-\frac{m}{p}\right|^{-\kappa} |m|^{-\kappa} \leq \sum_{1\leq m\leq p-1} \left|1-\frac{m}{p}\right|^{-\kappa} |m|^{-\kappa}.$$

Now let $q = \frac{p}{2}$ if p is even, or $\frac{p+1}{2}$ if p is odd. It is clear that

$$\sum_{m=1}^{q} \left|1-\frac{m}{p}\right|^{-\kappa} |m|^{-\kappa} \leq \left(\frac{1}{3}\right)^{-\kappa} \sum_{m=1}^{q} m^{-\kappa} \leq \left(\frac{1}{3}\right)^{-\kappa} \sum_{m=1}^{\infty} m^{-\kappa}.$$

Finally,

$$\sum_{m=q+1}^{p-1} \left|1 - \frac{m}{p}\right|^{-\kappa} |m|^{-\kappa} = \sum_{m=q+1}^{p-1} \left(\frac{m}{p}\right)^{-\kappa} (p-m)^{-\kappa}$$
$$\leq 2^{\kappa} \sum_{m=q+1}^{p-1} (p-m)^{-\kappa} \leq 2^{\kappa} \sum_{j=1}^{\infty} j^{-\kappa}.$$

That is, we have found the required bound for (38). This yields the bound for (37). □

The following Lemma is an adaptation of [4, Theorem 4.9] to $\mathbb{Z}^d$.

Lemma 15. *Let $(\lambda^k) \in \mathfrak{A}$, such that $\sum_{k\in\mathbb{Z}^d} \lambda^k = 1$. Suppose that for $m \in \mathbb{Z}^+$, $Y_{(m)} \in \mathcal{T}_\lambda^{\mathbb{Z}^d}$ is a strictly stationary random sequence, such that the law of $\hat{\mu}^n(Y_{(m)})$ is $\Pi^n_{(m)}$. Suppose that for each m, $(\Pi^n_{(m)})$ satisfies an LDP with good rate function. Suppose that $X = Y_{(m)} + Z_{(m)}$ for some stationary sequence $Z_{(m)}$, and let the law of $\hat{\mu}^n(X)$ be Π^n_X. If there exists a constant $\kappa_3 > 0$ such that for all $\mathfrak{b} > 0$*

$$\overline{\lim_{m\to\infty}}\,\overline{\lim_{n\to\infty}} \frac{1}{|V_n|} \log E\left[\exp\left(\mathfrak{b} \sum_{j\in V_n} \left\| Z^j_{(m)} \right\| \right)\right] < \kappa_3, \tag{39}$$

then (Π^n_X) satisfies an LDP with good rate function.

Proof. Let $d^{\lambda,\mathcal{M}}$ be the Prohorov Metric on $\mathcal{M}(\mathcal{T}_\lambda^{\mathbb{Z}^d})$ induced by d^λ. It suffices, thanks to [18, Theorem 4.2.16, Ex 4.2.29], to prove that for all $\delta > 0$,

$$\lim_{m\to\infty} \overline{\lim_{n\to\infty}} \frac{1}{(2n+1)^d} \log \mathbb{P}\left(d^{\lambda,\mathcal{M}}(\hat{\mu}^n(X), \hat{\mu}^n(Y_{(m)})) > \delta\right) = -\infty. \tag{40}$$

For $x \in \mathcal{T}_\lambda^{\mathbb{Z}^d}$, write $|x|_\lambda := d^\lambda(x, 0)$. Let $B \in \mathcal{B}(\mathcal{T}_\lambda^{\mathbb{Z}^d})$. Let $B^\delta = \{x \in \mathcal{T}_\lambda^{\mathbb{Z}^d} : d^\lambda(x,y) \leq \delta$ for some $y \in B\}$ be the closed blowup, and $B(\delta)$ be the closed blowup of $\{0\}$. Then

$$\hat{\mu}^n(X)(B) = \frac{1}{|V_n|} \sum_{j\in V_n} 1_B\left(S^j Y_{(m)}(n) + S^j Z_{(m)}(n)\right)$$
$$\leq \frac{1}{|V_n|} \sum_{j\in V_n} \left\{1_B\left(S^j Y_{(m)}(n) + S^j Z_{(m)}(n)\right) 1_{B(\delta)}(S^j Z_{(m)}(n)) \right.$$
$$\left. +1_{B(\delta)^c}\left(S^j Z_{(m)}(n)\right)\right\}$$
$$\leq \frac{1}{|V_n|} \sum_{j\in V_n} 1_{B^\delta}(S^j Y_{(m)}(n)) + \frac{1}{|V_n|} \#\{j \in V_n : |S^j Z_{(m)}(n)|_\lambda > \delta\}$$
$$\leq \hat{\mu}^n(Y_{(m)})(B^\delta) + \frac{1}{|V_n|} \#\left\{j \in V_n : |S^j Z_{(m)}(n)|_\lambda > \delta\right\}.$$

Thus

$$
\begin{aligned}
&\mathbb{P}\left(d^{\lambda,\mathcal{M}}(\hat{\mu}^n(X), \hat{\mu}^n(Y_{(m)})) > \delta\right) \\
&\qquad \le \mathbb{P}\left(\frac{1}{|V_n|}\#\left\{j \in V_n : |S^j Z_{(m)}(n)|_\lambda > \delta\right\} > \delta\right) \\
&\qquad \le \mathbb{P}\left(\sum_{j\in V_n} |S^j Z_{(m)}(n)|_\lambda > |V_n|\delta^2\right) \\
&\qquad \le \exp\left(-\mathfrak{b}|V_n|\delta^2\right) E\left[\exp\left(b\sum_{j\in V_n} |S^j Z_{(m)}(n)|_\lambda\right)\right] \\
&\qquad \le \exp\left(-\mathfrak{b}|V_n|\delta^2\right) E\left[\exp\left(b\sum_{j\in V_n, k\in\mathbb{Z}^d} \lambda^k \left\|Z_{(m)}(n)^{j+k}\right\|\right)\right]
\end{aligned}
$$

for an arbitrary $\mathfrak{b} > 0$. By Jensen's Inequality,

$$
\begin{aligned}
&\exp\left(-\mathfrak{b}|V_n|\delta^2\right) E\left[\exp\left(\mathfrak{b}\sum_{j\in V_n, k\in\mathbb{Z}^d} \lambda^k \left\|Z_{(m)}(n)^{j+k}\right\|\right)\right] \\
&\qquad \le \exp\left(-\mathfrak{b}|V_n|\delta^2\right) \sum_{k\in\mathbb{Z}^d} \lambda^k E\left[\exp\left(\mathfrak{b}\sum_{j\in V_n} \left\|Z_{(m)}(n)^{j+k}\right\|\right)\right] \\
&\qquad = \exp\left(-\mathfrak{b}|V_n|\delta^2\right) E\left[\exp\left(\mathfrak{b}\sum_{j\in V_n} \left\|Z_{(m)}(n)^{j}\right\|\right)\right],
\end{aligned}
$$

by the stationarity of $Z_{(m)}$ and the fact that $\sum_{k\in\mathbb{Z}^d} \lambda^k = 1$. We may thus infer, using (39), that

$$
\lim_{m\to\infty} \overline{\lim_{n\to\infty}} \frac{1}{|V_n|} \log \mathbb{P}\left(d^{\lambda,\mathcal{M}}(\hat{\mu}^n(X), \hat{\mu}^n(Y_{(m)})) > \delta\right) \le -\mathfrak{b}\delta^2 + \kappa_3.
$$

Since $\mathfrak{b}$ is arbitrary, we may take $\mathfrak{b} \to \infty$ to obtain (40). □

References

[1] S.-I. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):77–87, June 1977.

[2] Bruno B. Averbeck, Peter E. Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature reviews. Neuroscience*, 7(5):358–366, May 2006.

[3] Javier Baladron, Diego Fasoli, Olivier Faugeras, and Jonathan Touboul. Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and Fitzhugh-Nagumo neurons. *The Journal of Mathematical Neuroscience*, 2(1), 2012.

[4] John R. Baxter and Naresh C. Jain. An approximation condition for large deviations and some applications. In Vitaly Bergulson, editor, *Convergence in Ergodic Theory and Probability*. Ohio State University Mathematical Research Institute Publications, 1993.

[5] G. Ben-Arous and A. Guionnet. Large deviations for langevin spin glass dynamics. *Probability Theory and Related Fields*, 102(4):455–509, 1995.

[6] V.I. Bogachev. *Gaussian Measures*. American Mathematical Society, 1998.

[7] Paul Bressloff. *Waves in Neural Media*. Lectures Notes on Mathematical Modelling in the Applied Sciences. Springer, 2014.

[8] P.C. Bressloff. Stochastic neural field theory and the system-size expansion. *SIAM J. Appl. Math*, 70:1488–1521, 2009.

[9] P.C. Bressloff. Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical*, 45, 2012.

[10] P.C. Bressloff, J.D. Cowan, M. Golubitsky, P.J. Thomas, and M.C. Wiener. What Geometric Visual Hallucinations Tell Us about the Visual Cortex. *Neural Computation*, 14(3):473–491, 2002.

[11] Nicolas Brunel and David Hansel. How noise affects the synchronization properties of recurrent networks of inhibitory neurons. *Neural Computation*, 2006.

[12] W. Bryc and A. Dembo. Large deviations and strong mixing. In *Annales de l'IHP Probabilités et statistiques*, volume 32, pages 549–569. Elsevier, 1996.

[13] A. Budhiraja, P. Dupuis, and Fischer M. Large deviation properties of weakly interacting processes via weak convergence methods. *Annals of Probability*, 40(1):74–102, 2012.

[14] M.A. Buice, J.D. Cowan, and C.C. Chow. Systematic fluctuation expansion for neural network activity equations. *Neural computation*, 22(2):377–426, 2010.

[15] T. Chiyonobu and S. Kusuoka. The large deviation principle for hypermixing processes. *Probability Theory and Related Fields*, 78:627–649, 1988.

[16] J. Cox, Klaus Fleischmann, and Andreas Greven. Comparison of interacting diffusions and an application to their ergodic theory. *Probability Theory and Related Fields*, 105:513–528, 1996.

[17] Donald Dawson and Jurgen Gartner. Large deviations from the mckean-vlasov limit for weakly interacting diffusions. *Stochastics*, 20, 1987.

[18] A. Dembo and O. Zeitouni. *Large deviations techniques*. Springer, 1997. 2nd Edition.

[19] J.D. Deuschel, D.W. Stroock, and H. Zessin. Microcanonical distributions for lattice gases. *Communications in Mathematical Physics*, 139, 1991.

[20] Jean-Dominique Deuschel and Daniel W. Stroock. *Large Deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, 1989.

[21] M.D. Donsker and S.R.S Varadhan. Asymptotic evaluation of certain markov process expectations for large time, iv. *Communications on Pure and Applied Mathematics*, XXXVI:183–212, 1983.

[22] M.D. Donsker and S.R.S. Varadhan. Large deviations for stationary Gaussian processes. *Commun. Math. Phys.*, 97:187–210, 1985.

[23] Paul Doukhan. *Mixing: Properties and Examples.* Springer-Verlag, 1994.
[24] GB Ermentrout and J.D. Cowan. Temporal oscillations in neuronal nets. *Journal of mathematical biology*, 7(3):265–280, 1979.
[25] GB Ermentrout and JB McLeod. Existence and uniqueness of travelling waves for a neural network. In *Royal Society(Edinburgh), Proceedings*, volume 123, pages 461–478, 1993.
[26] Diego Fasoli. *Attacking the Brain with Neuroscience: Mean-Field Theory, Finite Size Effects and Encoding Capability of Stochastic Neural Networks.* PhD thesis, Université Nice Sophia Antipolis, September 2013.
[27] Olivier Faugeras and James MacLaurin. A large deviation principle and an analytical expression of the rate function for a discrete stationary gaussian process. Technical report, ArXiV: http://arxiv.org/abs/1311.4400, November 2013.
[28] Gregory Faye. Existence and stability of traveling pulses in a neural field equation with synaptic depression. *SIAM Journal of Applied Dynamical Systems*, 12, 2013.
[29] Markus Fischer. On the form of the large deviation rate function for the empirical measures of weakly interacting systems. *Archiv*, 2012.
[30] Mathieu Galtier. *A mathematical approach to unsupervised learning in recurrent neural networks.* PhD thesis, ParisTech, December 2011.
[31] M.A. Geise. *Neural Field Theory for Motion Perception.* Kluwer Academic Publishing, 1999.
[32] W. Gerstner and W. Kistler. *Spiking Neuron Models.* Cambridge University Press, 2002.
[33] W. Gerstner and W. M. Kistler. Mathematical formulations of hebbian learning. *Biological Cybernetics*, 87:404–415, 2002.
[34] G Giacomin, E Luçon, and Poquet. Coherence stability and effect of random natural frequencies in populations of coupled oscillators. *Archiv*, 2011.
[35] Iris Ginzburg and Haim Sompolinsky. Theory of correlations in stochastic neural networks. *Physical Review E*, 50(4), 1994.
[36] A. Greven and F. Den Hollander. Phase transitions for the long-time behavior of interacting diffusions. *The Annals of Probability*, 35(4), 2007.
[37] Hermann Haken. *Information and Self-Organization.* Springer, 2006.
[38] D. Hansel and H. Sompolinsky. *Methods in Neuronal Modeling, From Ions to Networks*, chapter Modeling Feature Selectivity in Lo- cal Cortical Circuits. MIT Press, 1998.
[39] Z.P. Kilpatrick and P.C. Bressloff. Effects of synaptic depression and adaptation on spatiotemporal dynamics of an excitatory neuronal network. *Physica D: Nonlinear Phenomena*, 239(9):547–560, 2010.
[40] Thomas M. Liggett. *Interacting Particle Systems.* Springer Berlin Heidelberg, 2005.
[41] Eric Luçon. Quenched limits and fluctuations of the empirical measure for plane rotators in random media. *Electronic Journal of Probability*, 2012.
[42] Eric Luçon and W. Stannat. Mean field limit for disordered diffusions with singular interactions. *Archiv*, 2013.
[43] Amit Manwani and Christof Koch. Detecting and estimating signals in

noisy cable structures i: Neuronal noise sources. *Neural Computation*, 1999.

[44] Mark D. McDonnell and Lawrence M. Ward. The benefits of noise in neural systems: bridging theory and experiment. *Nature Reviews: Neuroscience*, 12, 2011.

[45] K.D. Miller and D.J.C. MacKay. The role of constraints in hebbian learning. *Neural Comp*, 6:100–126, 1996.

[46] Erkki Oja. A simplified neuron model as a principal component analyzer. *J. Math. Biology*, 15:267–273, 1982.

[47] Srdjan Ostojic, Nicolas Brunel, and Vincent Hakim. Synchronization properties of networks of electrically coupled neurons in the presence of noise and heterogeneities. *Journal of Computational Neuroscience*, 26, 2009.

[48] D.J. Pinto and G.B. Ermentrout. Spatially structured activity in synaptically coupled neuronal networks: 1. traveling fronts and pulses. *SIAM Journal on Applied Mathematics*, 62(1):206–225, 2001.

[49] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, 1992.

[50] E. Schneidman, M.J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[51] Haim Sompolinsky, Hyoungsoo Yoon, Kukjin Kang, and Maoz Shamir. Population coding in neuronal systems with correlated noise. *Physical Review E*, 64, 2001.

[52] Alain-Sol Sznitman. Topics in propagation of chaos. In Donald Burkholder, Etienne Pardoux, and Alain-Sol Sznitman, editors, *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989*, volume 1464 of *Lecture Notes in Mathematics*, pages 165–251. Springer Berlin / Heidelberg, 1991. 10.1007/BFb0085169.

[53] J. Touboul. The propagation of chaos in neural fields. *The Annals of Applied Probability*, 24(3), 2014.

[54] Jonathan Touboul and Bard Ermentrout. Finite-size and correlation-induced effects in mean-field dynamics. *J Comput Neurosci*, 31(3):453–484, 2011.

[55] Henry Tuckwell. Analytical and simulation results for the stochastic spatial fitzhugh-nagumo model neuron. *Neural Computation*, 2008.

[56] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393, 1998.