Large Deviation Approach to Random Recurrent Neuronal Networks: Rate Function, Parameter Inference, and Activity Prediction

Alexander van Meegen,[1] Tobias Kühn,[1, 2, 3] and Moritz Helias[1, 2]
[1]Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany
[2]Department of Physics, Faculty 1, RWTH Aachen University, Aachen, Germany
[3]Laboratoire de Physique de l'ENS, CNRS, Paris, France
(Dated: September 21, 2020)

Statistical field theory captures collective non-equilibrium dynamics of neuronal networks, but it does not address the inverse problem of searching the connectivity to implement a desired dynamics. We here show for an analytically solvable network model that the effective action in statistical field theory is identical to the rate function in large deviation theory; using field theoretical methods we derive this rate function. It takes the form of a Kullback-Leibler divergence and enables data-driven inference of model parameters and Bayesian prediction of time series.

PACS numbers:

Introduction.– Biological neuronal networks are systems with many degrees of freedom and intriguing properties: their units are coupled in a directed, non-symmetric manner, so that they typically operate outside thermodynamic equilibrium [1, 2]. Even more challenging is the question how this non-equilibrium dynamics is used to perform information processing. A rigorous understanding is still deficient but imperatively needed to shed light into their remarkable abilities of computation.

The primary method to study neuronal networks has been mean-field theory [3–8]. Its field-theoretical basis has been exposed only recently [9, 10] which promises highly efficient schemes to study effects beyond the mean–field approximation. However, to understand the parallel and distributed information processing performed by neuronal networks, the study of the forward problem – from the microscopic parameters of the model to its dynamics – is not sufficient. One additionally faces the inverse problem of determining the parameters of the model given a desired dynamics and thus function. Formally, one needs to link statistical physics with concepts from information theory and statistical inference.

We here expose a tight relation between statistical field theory of neuronal networks, large-deviation theory, information theory, and inference. To this end, we generalize the probabilistic view of large deviation theory, which yields rigorous results for the leading order behavior in the network size N [11, 12], to arbitrary single unit dynamics and transfer functions. We then show that the central quantity of large deviation theory, the rate function, is identical to the effective action in statistical field theory. Having rendered this comprehensive picture, a third relation is found: Bayesian inference and prediction are naturally formulated within this framework, spanning the arc to information processing. Concretely, we discover a method of parameter inference from transient data and perform Bayes-optimal prediction of the time-dependent network activity.

Model.- We consider random networks of N nonlinearly interacting units $x_i(t)$ driven by an external input $\xi_i(t)$. The dynamics of the units are governed by the stochastic differential equation

$$\dot{x}_i(t) = -\nabla U(x_i(t)) + \sum_{j=1}^{N} J_{ij}\phi(x_j(t)) + \xi_i(t). \qquad (1)$$

In the absence of recurrent and external inputs, the units undergo an overdamped motion in a potential $U(x)$. The J_{ij} are independent and identically Gaussian-distributed random coupling weights with zero mean and variance $\langle J_{ij}^2\rangle = g^2/N$ where the coupling strength g controls the heterogeneity of the weights. The time-varying external inputs $\xi_i(t)$ are independent Gaussian white-noise processes with zero mean and correlation functions $\langle\xi_i(t_1)\xi_j(t_2)\rangle = 2D\delta_{ij}\delta(t_1-t_2)$. The model corresponds to the one studied in [4] if the external input vanishes, $D = 0$, the potential is quadratic, $U(x) = \frac{1}{2}x^2$, and the transfer function is sigmoidal, $\phi(x) = \tanh(x)$; for $D = \frac{1}{2}$, $U(x) = -\log(A^2 - x^2)$, and $\phi(x) = x$ it corresponds to the one in [11], which is inspired by the dynamical spin glass model of [13].

Field theory.- The field-theoretical treatment of Eq. (1) employs the Martin–Siggia–Rose–de Dominicis–Janssen path integral formalism [14–17]. We denote the expectation over paths across different realizations of the noise ξ as

$$\langle\cdot\rangle_{\mathbf{x}|\boldsymbol{J}} \equiv \left\langle\langle\cdot\rangle_{\mathbf{x}|\boldsymbol{J},\boldsymbol{\xi}}\right\rangle_{\boldsymbol{\xi}} = \int\mathcal{D}\mathbf{x}\int\mathcal{D}\tilde{\mathbf{x}}\cdot e^{S_0(\mathbf{x},\tilde{\mathbf{x}})-\tilde{\mathbf{x}}^{\mathrm{T}}\boldsymbol{J}\phi(\mathbf{x})},$$

where $\langle\cdot\rangle_{\mathbf{x}|\boldsymbol{J},\boldsymbol{\xi}}$ integrates over the unique solution of Eq. (1) given one realization $\boldsymbol{\xi}$ of the noise (see Appendix A). Here, $S_0(\mathbf{x},\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^{\mathrm{T}}(\dot{\mathbf{x}} + \nabla U(\mathbf{x})) + D\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}$ is the action of the uncoupled neurons. We use the shorthand notation $\mathbf{a}^{\mathrm{T}}\mathbf{b} = \sum_{i=1}^{N}\int_0^T \mathrm{d}t\, a_i(t)b_i(t)$.

For large N, the system becomes self-averaging, a property known from many disordered systems with large numbers of degrees of freedom: the collective behavior is stereotypical, independent of the realization J_{ij}. This holds for observables of the form $\sum_{i=1}^{N}\ell(x_i)$, where ℓ is an arbitrary functional of a single unit's trajectory. It is therefore convenient to introduce the scaled cumulant–generating functional

$$W_N(\ell) \coloneqq \frac{1}{N}\ln\left\langle\left\langle e^{\sum_{i=1}^{N}\ell(x_i)}\right\rangle_{\mathbf{x}|\boldsymbol{J}}\right\rangle_{\boldsymbol{J}}, \qquad (2)$$

where the prefactor $1/N$ makes sure that W_N is an intensive quantity, reminiscent of the bulk free energy [18]. In fact, we will show that the N-dependence vanishes in the limit $N\to\infty$ because the system decouples.

Performing the average over $\boldsymbol{J}$ and introducing the auxiliary field $C(t_1,t_2) \coloneqq g^2N^{-1}\sum_{i=1}^{N}\phi(x_i(t_1))\,\phi(x_i(t_2))$ as well as the conjugate field $\tilde{C}$, we can write W_N as [9, 19]

$$W_N(\ell) = \frac{1}{N}\ln\int\mathcal{D}C\int\mathcal{D}\tilde{C}\,e^{-\frac{N}{g^2}C^{\mathrm{T}}\tilde{C}+N\,\Omega_\ell(C,\tilde{C})}, \qquad (3)$$
$$\Omega_\ell(C,\tilde{C}) \coloneqq \ln\int\mathcal{D}x\int\mathcal{D}\tilde{x}\,e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^{\mathrm{T}}C\tilde{x}+\phi^{\mathrm{T}}\tilde{C}\phi+\ell(x)}.$$

The effective action is defined as the Legendre transform of $W_N(\ell)$,

$$\Gamma_N(\mu) \coloneqq \int\mathcal{D}x\,\mu(x)\,\ell_\mu(x) - W_N(\ell_\mu), \qquad (4)$$

where ℓ_μ is determined implicitly by the condition $\mu = W_N'(\ell_\mu)$ and the derivative $W_N'(\ell)$ has to be understood as a generalized derivative, the coefficient of the linearization akin to a Fréchet derivative [20].

Note that W_N and Γ_N are, respectively, generalizations of a cumulant–generating functional and of the effective action [21] because both map a functional (ℓ or μ) to the reals. For the choice $\ell(x) = j^{\mathrm{T}}x$, we recover the usual cumulant–generating functional of the single unit's trajectory (see Appendix D) and the corresponding effective action.

Rate function.- Self–averaging corresponds to a sharply peaked distribution of an observable over realizations of $\boldsymbol{J}$—because the distribution is very narrow, the observable always attains the same value. However, this can only hold for observables averaged over all units, reminiscent of the central limit theorem. Therefore, it is natural to restrict the observables to network–averaged ones; formally, this can be achieved without loss of generality using the empirical measure

$$\mu(y) \coloneqq \frac{1}{N}\sum_{i=1}^{N}\delta(x_i - y), \qquad (5)$$

since $\frac{1}{N}\sum_{i=1}^{N}\ell(x_i) = \int\mathcal{D}y\,\mu(y)\ell(y)$. Of particular interest is the leading–order exponential behavior of the distribution of empirical measures $P(\mu) = \langle\langle P(\mu\,|\,\mathbf{x})\rangle_{\mathbf{x}|\boldsymbol{J}}\rangle_{\boldsymbol{J}}$ across realizations of $\boldsymbol{J}$ and $\boldsymbol{\xi}$ described by the rate function [see e.g. 22]

$$H(\mu) \coloneqq -\lim_{N\to\infty}\frac{1}{N}\ln P(\mu). \qquad (6)$$

For large N, the probability of an empirical measure that does not correspond to the minimum $H'(\bar{\mu}) = 0$ is exponentially suppressed. Put differently, the system is self–averaging and the statistics of any network–averaged observable can be obtained using $\bar{\mu}$.

Similar to field theory, it is convenient to introduce the scaled cumulant–generating functional of the empirical measure. Because $\frac{1}{N}\sum_{i=1}^{N}\ell(x_i) = \int\mathcal{D}y\,\mu(y)\ell(y)$ holds for an arbitrary functional $\ell(x_i)$ of the single unit's trajectory x_i, Eq. (2) has the form of the scaled cumulant–generating functional for μ at finite N.

Using a saddle-point approximation for the integrals over C and $\tilde{C}$ in Eq. (3) (see Appendix A), we get

$$W_\infty(\ell) = -\frac{1}{g^2}C_\ell^{\mathrm{T}}\tilde{C}_\ell + \Omega_\ell(C_\ell,\tilde{C}_\ell). \qquad (7)$$

Both C_ℓ and $\tilde{C}_\ell$ are determined self-consistently by the saddle-point equations $C_\ell = g^2\,\partial_{\tilde{C}}\Omega_\ell(C,\tilde{C})\big|_{C_\ell,\tilde{C}_\ell}$ and $\tilde{C}_\ell = g^2\,\partial_C\Omega_\ell(C,\tilde{C})\big|_{C_\ell,\tilde{C}_\ell}$ where ∂_C denotes a partial functional derivative.

From the scaled cumulant–generating functional, Eq. (7), we obtain the rate function via a Legendre transformation [23]: $H(\mu) = \int\mathcal{D}x\,\mu(x)\ell_\mu(x) - W_\infty(\ell)$ with ℓ_μ implicitly defined by $\mu = W_\infty'(\ell_\mu)$. Comparing with Eq. (4), we observe that the rate function is equivalent to the effective action: $H(\mu) = \lim_{N\to\infty}\Gamma_N(\mu)$. The equation $\mu = W_\infty'(\ell_\mu)$ can be solved for ℓ_μ to obtain a closed expression for the rate function viz. effective action (see Appendix B)

$$H(\mu) = \int\mathcal{D}x\,\mu(x)\ln\frac{\mu(x)}{\langle\delta(\dot{x}+\nabla U(x)-\eta)\rangle_\eta}, \qquad (8)$$

where η is a zero–mean Gaussian process with a correlation function that is determined by $\mu(x)$,

$$C_\eta(t_1,t_2) = 2D\,\delta(t_1-t_2) + g^2\int\mathcal{D}x\,\mu(x)\,\phi(x(t_1))\phi(x(t_2)). \qquad (9)$$

For $D = \frac{1}{2}$, $U(x) = -\log(A^2 - x^2)$, and $\phi(x) = x$, Eq. (8) can be shown to be a equivalent to the mathematically rigorous result obtained in the seminal work by Ben Arous and Guionnet (see Appendix C).

The rate function Eq. (8) takes the form of a Kullback-Leibler divergence. Thus, it possesses a unique minimum at

$$\bar{\mu}(x) = \langle\delta(\dot{x}+\nabla U(x)-\eta)\rangle_\eta, \qquad (10)$$

which corresponds to the well-known self-consistent stochastic dynamics that is obtained in field theory

[4, 9, 10, 19]. Note that the correlation function of the effective stochastic input η at the minimum depends self-consistently on $\bar{\mu}(x)$ through Eq. (9).

Parameter Inference.– Thus far, we considered the network–averaged statistics of the activity for given statistics of the connectivity and the external input. The rate function opens the way to address the inverse problem: given the network–averaged activity statistics, encoded in the corresponding empirical measure μ, what are the statistics of the connectivity and the external input, i.e. g and D?

We determine the parameters using maximum likelihood estimation. Using Eq. (6) and Eq. (8), the likelihood of the parameters is given by

$$\ln P(\mu\,|\,g, D) \simeq -NH(\mu\,|\,g, D),$$

where $\simeq$ denotes equality in the limit $N \to \infty$ and we made the dependence on g and D explicit. The maximum likelihood estimation is given by the minimum with respect to the parameters g and D of the Kullback–Leibler divergence on the right hand side. Only the cross entropy $-\int \mathcal{D}x\,\mu(x) \ln \langle \delta(\dot{x} + \nabla U(x) - \eta)\rangle$, by Eq. (9), depends on the parameters, thus maximizing the likelihood estimation is equivalent to minimizing the cross entropy. The derivative of the cross entropy by the parameters yields

$$\partial_a \ln P(\mu\,|\,g, D) \simeq -\frac{N}{2}\mathrm{tr}\left((C_0 - C_\eta)\frac{\partial C_\eta^{-1}}{\partial a}\right),$$

where we used $\langle \delta(\dot{x} + \nabla U(x) - \eta)\rangle_\eta = Z^{-1}\exp(-S(x))$ with $S(x) = \frac{1}{2}(\dot{x} + \nabla U(x))^{\mathrm{T}} C_\eta^{-1}(\dot{x} + \nabla U(x))$ and normalization $Z = \sqrt{\det(2\pi C_\eta)}$ (see Appendix F), defined $C_0(t_1, t_2) \equiv \int \mathcal{D}x\,\mu(x)\left(\dot{x}(t_1) + \nabla U(x(t_1))\right)\left(\dot{x}(t_2) + \nabla U(x(t_2))\right)$, and abbreviated $a \in \{g, D\}$. The derivative vanishes for $C_0 = C_\eta$. Assuming stationarity, in Fourier domain this condition reads

$$\mathcal{S}_{\dot{x}+\nabla U(x)}(f) = 2D + g^2 \mathcal{S}_{\phi(x)}(f), \tag{11}$$

where $\mathcal{S}_X(f)$ denotes the network–averaged power spectrum of the observable X. Using non–negative least squares [24], Eq. (11) allows a straightforward inference of g and D (see Fig. 1).

Model Comparison.– Parameter estimation allows us to determine the statistical properties of the recurrent connectivity g and the external input D. However, this leaves the potential U and the transfer function ϕ unspecified. We determine U and ϕ using model comparison techniques [25].

We consider two options to obtain U and ϕ: comparing the mean squared error in Eq. (11) for the inferred parameters and comparing the likelihood of the inferred parameters. For the latter option, we can use the rate function from Eq. (6) and Eq. (8). Given two choices U_i,

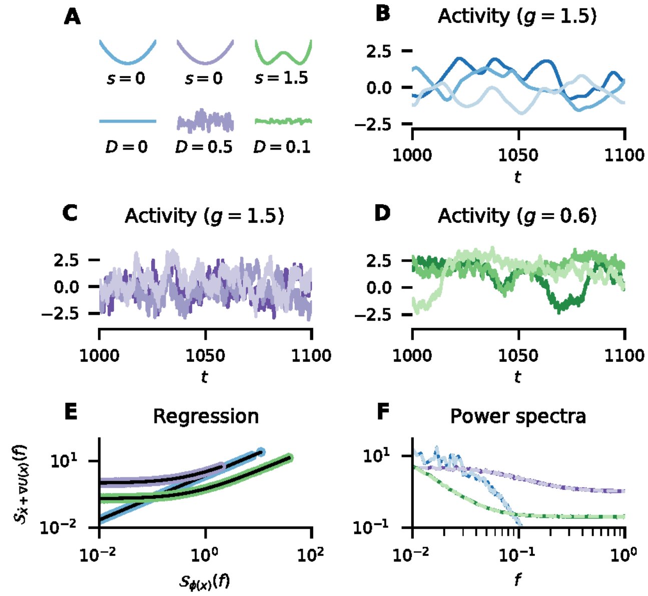

Figure 1: Maximum likelihood parameter estimation in networks with different single-unit potentials $U(x) = \frac{1}{2}x^2 + s\ln\cosh x$ and external noise D (**A**). Activity of three randomly chosen units (different hues) for various coupling strengths g as indicated in the title with s and D color coded (**B**-**D**). Parameter estimation via non-negative least squares based on Eq. (11) (**E**), and the power spectra on the left–(dark, solid lines) and right–hand–sides (light, dashed lines) of Eq. (11) for the inferred parameters (**F**). Further parameters: $N = 5000$, temporal discretization $dt = 10^{-2}$, simulation time $T = 1000$, time-span discarded to reach steady state $T_0 = 100$.

ϕ_i, $i \in \{1, 2\}$, with corresponding inferred parameters $\hat{g}_i$, $\hat{D}_i$, we have

$$\ln \frac{P(\mu\,|\,U_1, \phi_1, \hat{g}_1, \hat{D}_1)}{P(\mu\,|\,U_2, \phi_2, \hat{g}_2, \hat{D}_2)} \simeq -N(H_1 - H_2) \tag{12}$$

with $H_i \equiv H(\mu\,|\,U_i, \phi_i, \hat{g}_i, \hat{D}_i)$. The difference $H_1 - H_2$ equals the difference of the minimal cross entropies for the respective choices U_i, ϕ_i. Assuming an infinite observation time, this difference can be expressed as an integral that is straightforward to evaluate numerically (see Appendix G).

To illustrate the procedure, we consider the potential

$$U(x) = \frac{1}{2}x^2 - s\ln\cosh x,$$

which is bistable for $s > 1$ [5] and determine s using the mean squared error and the cross entropy difference (see Fig. 2). Parameter estimation yields estimates $\hat{g}$ and $\hat{D}$ that depend on s (Fig. 2**A**,**B**). The mean squared error displays a clear minimum at the true value $s = 1.5$ (Fig. 2**C**) whereas the maximal cross entropy occurs at a value larger than $s = 1.5$ (Fig. 2**D**). The latter effect

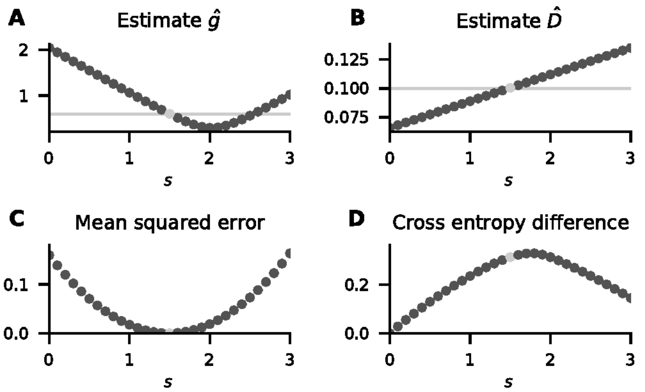

Figure 2: Model comparison in a network with single-unit potential $U(x) = \frac{1}{2}x^2 - s \ln\cosh x$. Maximum likelihood estimates of $\hat{g}$ and $\hat{D}$ for given choices of s (**A**, **B**). True values of g and D indicated as gray lines; estimates at the true value $s = 1.5$ indicated as gray symbols. Mean squared error between left– and right–hand–side of Eq. (11) for given s (**C**). Cross entropy difference between model with $s = 0$ and with given s (**D**). Further parameters as in Fig. 1.

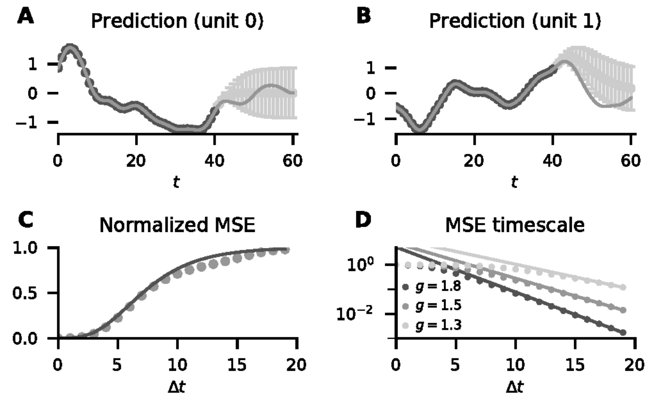

Figure 3: Prediction of the future network activity (**A**,**B**). Full trajectory of two arbitrary units shown as gray solid curves, training data as black dots, and prediction $\hat{x}$ with error bars determined by $\sigma_{\hat{x}}$ as gray symbols. The network–averaged mean squared error, indicated as gray symbols, is well captured by the predicted variance $\sigma_{\hat{x}}^2$, shown as solid curve (**C**). The error increases on half of the timescale of the autocorrelation function: $1 - \sigma_{\hat{x}}^2/\sigma_x^2$, shown as symbols, decreases asymptotically as $\mathcal{C}\exp(-2\Delta t/\tau_c)$, indicated as lines (**D**). Parameters: $g = 1.5$, $s = D = 0$; further parameters as in Fig. 1.

arises because the cross entropy is dominated by the parameter estimates, thus the mean squared error provides a more reliable criterion in this case.

Activity Prediction.– Parameter inference and model comparison allow us to fully determine the model from data. Using this information, we can proceed to a functional aspect: predicting the future activity of a unit of the recurrent neuronal network from the knowledge of its recent past.

If the potential of the model is quadratic, $U(x) \propto \frac{1}{2}x^2$, the measure $\bar{\mu}$ that minimizes the rate function corresponds to a Gaussian process. For Gaussian processes, it is possible to perform Bayes–optimal prediction only based on its correlation function [25, 26]. Denoting the correlation function of the process as C_x, the prediction is given by

$$\hat{x} = \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{x} \tag{13}$$

with $K_{ij} = C_x(t_i, t_j)$, $k_i = C_x(t_i, \hat{t})$, and $x_i = x(t_i)$. Here $\hat{t}$ denotes the timepoint of the prediction and $\{t_i\}$ a set of timepoints where the activity is known. The predicted value $\hat{x}$ itself is Gaussian distributed with variance

$$\sigma_{\hat{x}}^2 = \kappa - \mathbf{k}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{k} \tag{14}$$

where $\kappa = C_x(\hat{t}, \hat{t})$. The variance $\sigma_{\hat{x}}^2$ quantifies the uncertainty associated with the prediction $\hat{x}$.

Given the inferred parameters, we determine the self-consistent autocorrelation function C_x using Eq. (9) and Eq. (10). We use this self-consistent autocorrelation function to predict the future activity of two arbitrary units using Eq. (13) and Eq. (14) (Fig. 3**A**,**B**). To quantify the error, we calculate the network–averaged mean squared error relative to the variance obtained from the self–consistent autocorrelation function $C_x(0)$ and compare it to the variability predicted by Eq. (14), yielding a good correspondence (Fig. 3**C**). The timescale of the error is determined by half of the timescale of the autocorrelation function (see Appendix G). We plot $1 - \sigma_{\hat{x}}^2/\sigma_x^2$ against an exponential decay $\mathcal{C}\exp(-2\tau/\tau_c)$, where $C_x(\tau)/C_x(0) \sim \exp(-\tau/\tau_c)$, and find a very good agreement (Fig. 3**D**). Since τ_c diverges for $g \to 1+$ (cf. [4]), the timescale of the error diverges as well.

Discussion.– In this Letter, we found a tight link between the field theoretical approach to neuronal networks and its counterpart based on large deviation theory. We obtained the rate function of the empirical measure for the widely used and analytically solvable model of a recurrent neuronal network [4] by field theoretical methods. This rate function generalizes the seminal result by Ben Arous and Guionnet [11, 12] to arbitrary potentials and transfer functions. Intriguingly, our derivation elucidates that the rate function is identical to the effective action and takes the form of a Kullback–Leibler divergence, akin to Sanov's theorem for sums of independent and identically distributed random variables [22, 23]. The rate function can thus be interpreted as a distance between an empirical measure, for example given by data, and the activity statistics of the network model. This result allows us to address the inverse problem: 1) Inferring the parameters of the connectivity and external input from a set of trajectories. 2) Determining the potential and the transfer function. 3) Predicting future activity in a Bayes–optimal way.

The exposed link between the effective action defined within statistical field theory and the rate function, central to large deviation theory, opens the door to applying established field-theoretical techniques, such as the loopwise expansion [21], to obtain systematic corrections beyond the mean-field limit. Such sub–exponential corrections to the rate function are important for small or sparse networks with non–vanishing mean connectivity, to explain correlated neuronal activity, and to study information processing in finite-size networks with realistically limited resources. More generally, the link allows the systematic derivation of results using field theory and to subsequently prove them in a mathematically rigorous manner within the large deviation framework.

Acknowledgments.– We are grateful to Olivier Faugeras and Etienne Tanré for helpful discussions on LDT of neuronal networks, to Anno Kurth for pointing us to the Fréchet derivative and to Alexandre René, David Dahmen, Kirsten Fischer, and Christian Keup for feedback on an earlier version of the manuscript. This work was partly supported by the Helmholtz young investigator's group VH-NG-1028, European Union Horizon 2020 grant 785907 (Human Brain Project SGA2).

Appendix

A. Scaled Cumulant Generating Functional

Here, we derive the scaled cumulant generating functional and the saddle-point equations. The first steps of the derivations are akin to the manipulations presented in [9, 19], thus we keep the presentation concise. We interpret the stochastic differential equations governing the network dynamics in the Itô convention. Using the Martin–Siggia–Rose–de Dominicis–Janssen path integral formalism, the expectation $\langle\cdot\rangle_{\mathbf{x}|\boldsymbol{J}}$ of some arbitrary functional $G(\mathbf{x})$ can be written as

$$\begin{aligned}\langle\langle G(\mathbf{x})\rangle_{\mathbf{x}|\boldsymbol{J},\boldsymbol{\xi}}\rangle_{\boldsymbol{\xi}} &= \int\mathcal{D}\mathbf{x}\,\langle\delta(\dot{\mathbf{x}}+\nabla U(\mathbf{x})+\boldsymbol{J}\phi(\mathbf{x})+\boldsymbol{\xi})\rangle_{\boldsymbol{\xi}}\,G(\mathbf{x})\\ &= \int\mathcal{D}\mathbf{x}\int\mathcal{D}\tilde{\mathbf{x}}\,e^{S_0(\mathbf{x},\tilde{\mathbf{x}})-\tilde{\mathbf{x}}^{\mathrm{T}}\boldsymbol{J}\phi(\mathbf{x})}G(\mathbf{x})\end{aligned}$$

where we used the Fourier representation $\delta(x) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}e^{\tilde{x}x}d\tilde{x}$ in every timestep in the second step and defined the action

$$S_0(\mathbf{x},\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^{\mathrm{T}}(\dot{\mathbf{x}}+\nabla U(\mathbf{x}))+D\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}.$$

An additional average over realizations of the connectivity $\boldsymbol{J}\overset{\text{i.i.d.}}{\sim}\mathcal{N}(0,N^{-1}g^2)$ only affects the term $-\tilde{\mathbf{x}}^{\mathrm{T}}\boldsymbol{J}\phi(\mathbf{x})$ in the action and results in

$$\langle e^{-\tilde{\mathbf{x}}^{\mathrm{T}}\boldsymbol{J}\phi(\mathbf{x})}\rangle_{\boldsymbol{J}} = \int\mathcal{D}C\int\mathcal{D}\tilde{C}\,e^{-\frac{N}{g^2}C^{\mathrm{T}}\tilde{C}+\frac{1}{2}\tilde{\mathbf{x}}^{\mathrm{T}}C\tilde{\mathbf{x}}+\phi(\mathbf{x})^{\mathrm{T}}\tilde{C}\phi(\mathbf{x})},$$

where we introduced the network–averaged auxiliary field

$$C(u,v) = \frac{g^2}{N}\sum_{i=1}^{N}\phi(x_i(u))\phi(x_i(v))$$

via a Hubbard–Stratonovich transformation. The average over the connectivity and the subsequent Hubbard–Stratonovich transformation decouple the dynamics across units; afterwards the units are only coupled through the global fields C and $\tilde{C}$.

Now, we consider the scaled cumulant generating functional of the empirical density

$$W_N(\ell) = \frac{1}{N}\ln\left\langle\left\langle e^{\Sigma_{i=1}^{N}\ell(x_i)}\right\rangle_{\mathbf{x}|\boldsymbol{J}}\right\rangle_{\boldsymbol{J}}.$$

Using the above results and the abbreviation $\phi(x)\equiv\phi$, it can be written as

$$\begin{aligned}W_N(\ell) &= \frac{1}{N}\ln\int\mathcal{D}C\int\mathcal{D}\tilde{C}\,e^{-\frac{N}{g^2}C^{\mathrm{T}}\tilde{C}+N\,\Omega_\ell(C,\tilde{C})},\\ \Omega_\ell(C,\tilde{C}) &= \ln\int\mathcal{D}x\int\mathcal{D}\tilde{x}\,e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^{\mathrm{T}}C\tilde{x}+\phi^{\mathrm{T}}\tilde{C}\phi+\ell(x)},\end{aligned}$$

where the N in front of the single–particle cumulant generating functional Ω results from the factorization of the N integrals over x_i and $\tilde{x}_i$ each; thus it is a hallmark of the decoupled dynamics. Next, we approximate the C and $\tilde{C}$ integrals in a saddle–point approximation which yields

$$W_N(\ell) = -\frac{1}{g^2}C_\ell^{\mathrm{T}}\tilde{C}_\ell+\Omega_\ell(C_\ell,\tilde{C}_\ell)+O(\ln(N)/N),$$

where C_ℓ and $\tilde{C}_\ell$ are determined by the saddle–point equations

$$\begin{aligned}C_\ell &= g^2\,\partial_{\tilde{C}}\Omega_\ell(C,\tilde{C})\big|_{C_\ell,\tilde{C}_\ell},\\ \tilde{C}_\ell &= g^2\,\partial_C\Omega_\ell(C,\tilde{C})\big|_{C_\ell,\tilde{C}_\ell}.\end{aligned}$$

Here, ∂_C denotes a partial functional derivative. In the limit $N\to\infty$, the remainder $O(\ln(N)/N)$ vanishes and the saddle–point approximation becomes exact.

B. Rate Function

Here, we derive the rate function from the scaled cumulant generating functional. According to the Gärtner-Ellis theorem [23], we obtain the rate function via the Legendre transformation

$$H(\mu) = \int\mathcal{D}x\,\mu(x)\ell_\mu(x)-W_\infty(\ell_\mu) \tag{15}$$

with ℓ_μ implicitly defined by

$$\mu = W_\infty'(\ell_\mu). \tag{16}$$

Due to the saddle–point equations, the derivative of the cumulant generating functional in Eq. (16) simplifies to $W_\infty'(\ell_\mu) = (\partial_\ell\Omega_\ell)(C_\ell,\tilde{C}_\ell)\big|_{\ell_\mu}$ where the derivative only acts on the ℓ that is explicit in $\Omega_\ell(C_\ell,\tilde{C}_\ell)$ and not on the implicit dependencies through C_ℓ, $\tilde{C}_\ell$. Thus, Eq. (16) yields

$$\mu(x) = \frac{\int \mathcal{D}\tilde{x}\, e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^\mathrm{T} C_{\ell_\mu}\tilde{x}+\phi^\mathrm{T}\tilde{C}_{\ell_\mu}\phi+\ell_\mu(x)}}{\int \mathcal{D}x \int \mathcal{D}\tilde{x}\, e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^\mathrm{T} C_{\ell_\mu}\tilde{x}+\phi^\mathrm{T}\tilde{C}_{\ell_\mu}\phi+\ell_\mu(x)}}.$$

Taking the logarithm and using $W_\infty(\ell_\mu) + \frac{1}{g^2} C_{\ell_\mu}^\mathrm{T}\tilde{C}_{\ell_\mu} = \Omega_{\ell_\mu}(C_{\ell_\mu},\tilde{C}_{\ell_\mu})$ leads to

$$\begin{aligned}\ell_\mu(x) &= \ln \frac{\mu(x)}{\int \mathcal{D}\tilde{x}\, e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^\mathrm{T} C_{\ell_\mu}\tilde{x}}} + W_\infty(\ell_\mu) \\ &\quad + \frac{1}{g^2} C_{\ell_\mu}^\mathrm{T}\tilde{C}_{\ell_\mu} - \phi^\mathrm{T}\tilde{C}_{\ell_\mu}\phi.\end{aligned}$$

Inserting $\ell_\mu(x)$ into the Legendre transformation (15) yields

$$\begin{aligned}H(\mu) &= \int \mathcal{D}x\, \mu(x) \ln \frac{\mu(x)}{\int \mathcal{D}\tilde{x}\, e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^\mathrm{T} C_{\ell_\mu}\tilde{x}}} \\ &\quad + \frac{1}{g^2} C_{\ell_\mu}^\mathrm{T}\tilde{C}_{\ell_\mu} - C_\mu^\mathrm{T}\tilde{C}_{\ell_\mu}\end{aligned}$$

with

$$C_\mu(u,v) = \int \mathcal{D}x\, \mu(x)\phi(x(u))\phi(x(v)).$$

Identifying $\mu(x)$ in the saddle–point equation

$$\begin{aligned}C_{\ell_\mu} &= g^2\, \partial_{\tilde{C}}\Omega_\ell(C,\tilde{C})\big|_{C_{\ell_\mu},\tilde{C}_{\ell_\mu}} \\ &= g^2 \frac{\int \mathcal{D}x \int \mathcal{D}\tilde{x}\, \phi\phi e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^\mathrm{T} C_{\ell_\mu}\tilde{x}+\phi^\mathrm{T}\tilde{C}_{\ell_\mu}\phi+\ell_\mu(x)}}{\int \mathcal{D}x \int \mathcal{D}\tilde{x}\, e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^\mathrm{T} C_{\ell_\mu}\tilde{x}+\phi^\mathrm{T}\tilde{C}_{\ell_\mu}\phi+\ell_\mu(x)}}\end{aligned}$$

yields

$$C_{\ell_\mu}(u,v) = g^2 \int \mathcal{D}x\, \mu(x)\phi(x(u))\phi(x(v))$$

and thus $C_{\ell_\mu} = g^2 C_\mu$. Accordingly, the last two terms in the Legendre transformation cancel and we arrive at

$$H(\mu) = \int \mathcal{D}x\, \mu(x) \ln \frac{\mu(x)}{\int \mathcal{D}\tilde{x}\, e^{S_0(x,\tilde{x})+\frac{g^2}{2}\tilde{x}^\mathrm{T} C_\mu\tilde{x}}} \tag{17}$$

where still $C_\mu(u,v) = \int \mathcal{D}x\, \mu(x)\phi(x(u))\phi(x(v))$.

In the main text, we use the notation

$$\int \mathcal{D}\tilde{x}\, e^{S_0(x,\tilde{x})+\frac{g^2}{2}\tilde{x}^\mathrm{T} C_\mu\tilde{x}} = \langle \delta(\dot{x} + \nabla U(x) - \eta)\rangle_\eta$$

with $C_\eta = 2D\delta + g^2 C_\mu$ appearing in the rate function. Indeed, using the Martin–Siggia–Rose–de Dominicis–Janssen formalism, we have

$$\begin{aligned}\langle \delta(\dot{x} + \nabla U(x) - \eta)\rangle_\eta &= \int \mathcal{D}\tilde{x}\, e^{\tilde{x}^\mathrm{T}(\dot{x}+\nabla U(x))}\langle e^{\tilde{x}^\mathrm{T}\eta}\rangle_\eta \\ &= \int \mathcal{D}\tilde{x}\, e^{\tilde{x}^\mathrm{T}(\dot{x}+\nabla U(x))+\frac{1}{2}\tilde{x}^\mathrm{T} C_\eta\tilde{x}}\end{aligned}$$

which shows that the two notations are equivalent since $\tilde{x}^\mathrm{T}(\dot{x} + \nabla U(x)) + \frac{1}{2}\tilde{x}^\mathrm{T} C_\eta\tilde{x} = S_0(x,\tilde{x}) + \frac{g^2}{2}\tilde{x}^\mathrm{T} C_\mu\tilde{x}$ for $C_\eta = 2D\delta + g^2 C_\mu$.

C. Equivalence to Ben Arous and Guionnet (1995)

Here, we show explicitly that the rate function we obtained generalizes the rate function obtained by Ben Arous and Guionnet. We start with Theorem 4.1 in [11] adapted to our notation: Define

$$Q(x) \coloneqq \int \mathcal{D}\tilde{x}\, e^{\tilde{x}^\mathrm{T}(\dot{x}+\nabla U(x))+\frac{1}{2}\tilde{x}^\mathrm{T}\tilde{x}}$$

and

$$G(\mu) \coloneqq \int \mathcal{D}x\, \mu(x) \log\Big(\langle e^{gy^\mathrm{T}(\dot{x}+\nabla U(x))-\frac{g^2}{2}y^\mathrm{T}y}\rangle_y\Big),$$

where $\langle\cdot\rangle_y$ is the expectation value over a zero–mean Gaussian process y with $C_\mu(u,v) = \int \mathcal{D}x\, \mu(x)x(u)x(v)$, written as $\langle\cdot\rangle_y = \int \mathcal{D}y \int \mathcal{D}\tilde{y}\, (\cdot)\, e^{\tilde{y}^\mathrm{T}y+\frac{1}{2}\tilde{y}^\mathrm{T}C_\mu\tilde{y}}$. With the Kullback–Leibler divergence $D_\mathrm{KL}(\mu|Q)$, Theorem 4.1 states that the function

$$\tilde{H}(\mu) = \begin{cases} D_\mathrm{KL}(\mu|Q) - G(\mu) & \text{if } D_\mathrm{KL}(\mu|Q) < \infty \\ +\infty & \text{otherwise}\end{cases}$$

is a good rate function.

Now we relate $\tilde{H}$ to the rate function that is derived above, Eq. (17). Using the Onsager–Machlup action, we can write

$$D_\mathrm{KL}(\mu|Q) = \int \mathcal{D}x\, \mu(x) \log\frac{\mu(x)}{e^{-S_\mathrm{OM}(x)}} + \mathcal{C}$$

with $S_\mathrm{OM}(x) = \frac{1}{2}(\dot{x} + \nabla U(x))^\mathrm{T}(\dot{x} + \nabla U(x))$. Next, we transform $gy \to y$, $\tilde{y}/g \to \tilde{y}$ and solve the integral over y in $G(\mu)$:

$$\int \mathcal{D}y\, e^{-\frac{1}{2}y^\mathrm{T}y+y^\mathrm{T}(\dot{x}+\nabla U(x)+\tilde{y})} \propto e^{S_\mathrm{OM}[x]+\tilde{y}^\mathrm{T}(\dot{x}+\nabla U(x))+\frac{1}{2}\tilde{y}^\mathrm{T}\tilde{y}}.$$

The Onsager–Machlup action in the logarithm in $D_\mathrm{KL}(\mu|Q)$ and $G(\mu)$ cancel and we arrive at

$$\tilde{H}(\mu) = \int \mathcal{D}x\, \mu(x) \log\frac{\mu(x)}{\int \mathcal{D}\tilde{y}\, e^{\tilde{y}^\mathrm{T}(\dot{x}+\nabla U(x))+\frac{1}{2}\tilde{y}^\mathrm{T}(g^2C_\mu+\delta)\tilde{y}}}$$

up to an additive constant that we set to zero. Since $C_\mu(u,v) = \int \mathcal{D}x\, \mu(x)x(u)x(v)$, the rate function by Ben Arous and Guionnet is thus equivalent to Eq. (8) with $\phi(x) = x$ and $D = \frac{1}{2}$.

D. Relation to Sompolinsky, Crisanti, Sommers (1988)

Here, we relate the approach that we laid out in the main text to the approach pioneered by Sompolinsky, Crisanti, and Sommers [4] (reviewed in [9, 10]) using our notation for consistency. Therein, the starting point is the scaled cumulant–generating functional

$$\hat{W}_N(j) = \frac{1}{N}\ln\left\langle\left\langle e^{j^{\mathrm{T}}\mathbf{x}}\right\rangle_{\mathbf{x}|\boldsymbol{J}}\right\rangle_{\boldsymbol{J}},$$

which gives rise to the cumulants of the trajectories. For the linear functional

$$\ell(x) = j^{\mathrm{T}}x,$$

we have $\sum_{i=1}^N \ell(x_i) = j^{\mathrm{T}}\mathbf{x}$ and thus $W_N(j^{\mathrm{T}}x) = \hat{W}_N(j)$. Put differently, the scaled cumulant–generating functional of the trajectories $\hat{W}_N(j)$ is a special case of the more general scaled cumulant–generating functional $W_N(\ell)$ we consider in this manuscript. Of course one can start from the scaled cumulant–generating functional of the observable of interest and derive the corresponding rate function. Conversely, we show below how to obtain the rate function of a specific observable from the rate function of the empirical measure.

Contraction Principle

Here, we relate the rather general rate function of the empirical measure $H(\mu)$ to the rate function of a particular observable $I(C)$. As an example, we choose the correlation function

$$C(u,v) = \frac{g^2}{N}\sum_{i=1}^N \phi(x_i(u))\phi(x_i(v))$$

because it is a quantity that arises naturally during the Hubbard–Stratonovich transformation. The generic approach to this problem is given by the contraction principle [23]:

$$I(C) = \inf_{\mu\,\mathrm{s.t.}\,C=g^2\int\mathcal{D}x\,\mu(x)\phi\phi} H(\mu).$$

Here, the infimum is constrained to the empirical measures that give rise to the correlation function C, i.e. those that fulfill $C(u,v) = g^2\int\mathcal{D}x\,\mu(x)\phi(x(u))\phi(x(v))$. Writing $H(\mu)$ as the Legendre transform of the scaled cumulant–generating functional, $H(\mu) = \inf_\ell[\int\mathcal{D}x\,\mu(x)\ell(x) - W_\infty(\ell)]$, the empirical measure only appears linearly. Using a Lagrange multiplier $k(u,v)$, the infimum over μ leads to the constraint $\ell(x) = g^2\phi^{\mathrm{T}}k\phi$ and we arrive at

$$I(C) = \inf_k[k^{\mathrm{T}}C - W_\infty(g^2\phi^{\mathrm{T}}k\phi)].$$

Once again, we see how to relate $W_N(\ell)$ to a specific observable—this time for the choice $\ell(x) = g^2\phi^{\mathrm{T}}k\phi$.

Up to this point, the discussion applies to any observable. For the current example, we can proceed a bit further. With the redefinition $\tilde{C} + g^2k \to \tilde{C}$, we get

$$W_\infty(g^2\phi^{\mathrm{T}}k\phi) = \mathrm{extr}_{C,\tilde{C}}\left[-\frac{1}{g^2}C^{\mathrm{T}}\tilde{C} + C^{\mathrm{T}}k + \Omega_0(C,\tilde{C})\right],$$
$$\Omega_0(C,\tilde{C}) = \ln\int\mathcal{D}x\int\mathcal{D}\tilde{x}\,e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^{\mathrm{T}}C\tilde{x}+\phi^{\mathrm{T}}\tilde{C}\phi},$$

which made Ω_0 independent of k. Now we can take the infimum over k, leading to

$$I(C) = \mathrm{extr}_{\tilde{C}}\left[\frac{1}{g^2}C^{\mathrm{T}}\tilde{C} - \Omega_0(C,\tilde{C})\right]. \tag{18}$$

The remaining extremum gives rise to the condition

$$C = g^2\frac{\int\mathcal{D}x\int\mathcal{D}\tilde{x}\,\phi\phi e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^{\mathrm{T}}C\tilde{x}+\phi^{\mathrm{T}}\tilde{C}\phi}}{\int\mathcal{D}x\int\mathcal{D}\tilde{x}\,e^{S_0(x,\tilde{x})+\frac{1}{2}\tilde{x}^{\mathrm{T}}C\tilde{x}+\phi^{\mathrm{T}}\tilde{C}\phi}},$$

i.e. a self–consistency condition for the correlation function.

As a side remark, we mention that the expression in the brackets of Eq. (18) is the joint effective action for C and $\tilde{C}$, because for $N\to\infty$, the action equals the effective action. This result is therefore analogous to the finding that the effective action in the Onsager–Machlup formalism is given as the extremum of its counterpart in the Martin–Siggia–Rose–de Dominicis–Janssen formalism [27, eq. (24)]. The only difference is that here, we are dealing with second order statistics and not just mean values. The origin of this finding is the same in both cases: we are only interested in the statistics of the physical quantity (the one without tilde, x or C, respectively). Therefore we only introduce a source field (k in the present case) for this one, but not for the auxiliary field, which amounts to setting the source field of the latter to zero. This is translated into the extremum in Eq. (18) over the auxiliary variable [27, appendix 5].

E. Log–Likelihood Derivative

Here, we calculate the derivatives of the log–likelihood with respect to the parameters g and D. In terms of the rate function, we have

$$\partial_a \ln P(\mu\,|\,g,D) \simeq -N\partial_a H(\mu\,|\,g,D)$$

where a denotes either g or D. The parameters appear only in the cross entropy

$$\partial_a H(\mu) = -\int\mathcal{D}x\,\mu(x)\partial_a\ln\langle\delta(\dot{x}+\nabla U(x)-\eta)\rangle_\eta$$

through the correlation function $C_\eta(u,v) = 2D\delta(u-v) + g^2\int\mathcal{D}x\,\mu(x)\phi(x(u))\phi(x(v))$. Above, we showed that

$$\langle\delta(\dot{x}+\nabla U(x)-\eta)\rangle_\eta = \int\mathcal{D}\tilde{x}\,e^{\tilde{x}^{\mathrm{T}}(\dot{x}+\nabla U(x))+\frac{1}{2}\tilde{x}^{\mathrm{T}}C_\eta\tilde{x}}.$$

Because $\tilde{x}$ is at most quadratic in the exponent, the integral is solvable and we get

$$\langle\delta(\dot{x}+\nabla U(x)-\eta)\rangle_\eta = \frac{e^{-\frac{1}{2}(\dot{x}+\nabla U(x))^{\mathrm{T}}C_\eta^{-1}(\dot{x}+\nabla U(x))}}{\sqrt{\det(2\pi C_\eta)}}.$$

Note that the normalization $1/\sqrt{\det(2\pi C_\eta)}$ does not depend on the potential U. Now we can take the derivatives of $\ln\langle\delta(\dot{x}+\nabla U(x)-\eta)\rangle_\eta$ and get

$$\partial_a \ln\langle\delta(\dot{x}+\nabla U(x)-\eta)\rangle_\eta = -\frac{1}{2}(\dot{x}+\nabla U(x))^{\mathrm{T}}\frac{\partial C_\eta^{-1}}{\partial a}(\dot{x}+\nabla U(x)) - \frac{1}{2}\partial_a \mathrm{tr}\ln C_\eta$$

where we used $\ln\det C = \mathrm{tr}\ln C$. With this, we arrive at

$$\partial_a H(\mu) = \frac{1}{2}\mathrm{tr}\left(C_0\frac{\partial C_\eta^{-1}}{\partial a}\right) + \frac{1}{2}\mathrm{tr}\left(\frac{\partial C_\eta}{\partial a}C_\eta^{-1}\right)$$

where the integral over the empirical measure gave rise to $C_0 = \int \mathcal{D}x\,\mu(x)(\dot{x}+\nabla U(x))(\dot{x}+\nabla U(x))$ and we used $\partial_a \ln C = \frac{\partial C}{\partial a}C^{-1}$. Finally, using $\frac{\partial C}{\partial a}C^{-1} = CC^{-1}\frac{\partial C}{\partial a}C^{-1} = C\frac{\partial C^{-1}}{\partial a}$, we get

$$\partial_a \ln P(\mu\,|\,g,D) \simeq -\frac{N}{2}\mathrm{tr}\left((C_0 - C_\eta)\frac{\partial C_\eta^{-1}}{\partial a}\right)$$

as stated in the main text.

F. Cross Entropy Difference

Here, we express the cross entropy difference

$$H_1 - H_2 := H(\mu\,|\,U_1,\phi_1,\hat{g}_1,\hat{D}_1) - H(\mu\,|\,U_2,\phi_2,\hat{g}_2,\hat{D}_2)$$

in a form that can be evaluated numerically. Using the rate function, we get

$$H_1 - H_2 = \int \mathcal{D}x\,\mu(x)\ln\frac{\langle\delta(\dot{x}+\nabla U_2(x)-\eta_2)\rangle_{\eta_2}}{\langle\delta(\dot{x}+\nabla U_1(x)-\eta_1)\rangle_{\eta_1}}$$

with $C_{\eta_i} = 2\hat{D}_i\delta + \hat{g}_i^2\int\mathcal{D}x\,\mu(x)\phi_i\phi_i$. Again, we use

$$\langle\delta(\dot{x}+\nabla U(x)-\eta)\rangle_\eta = \frac{e^{-\frac{1}{2}(\dot{x}+\nabla U(x))^{\mathrm{T}}C_\eta^{-1}(\dot{x}+\nabla U(x))}}{\sqrt{\det(2\pi C_\eta)}}$$

to arrive at

$$\begin{aligned} H_1 - H_2 = &-\frac{1}{2}\mathrm{tr}\left(C_1 C_{\eta_1}^{-1}\right) - \frac{1}{2}\mathrm{tr}\ln C_{\eta_1} \\ &+\frac{1}{2}\mathrm{tr}\left(C_2 C_{\eta_2}^{-1}\right) + \frac{1}{2}\mathrm{tr}\ln C_{\eta_2} \end{aligned}$$

with $C_i = \int\mathcal{D}x\,\mu(x)(\dot{x}+\nabla U_i(x))(\dot{x}+\nabla U_i(x))$. For stationary correlation functions over infinite time intervals, we can evaluate the traces as integrals over the power spectra:

$$\begin{aligned} \mathrm{tr}(AB^{-1}) &\propto \int_{-\infty}^{\infty}\frac{\tilde{A}(f)}{\tilde{B}(f)}df, \\ \mathrm{tr}\ln A &\propto \int_{-\infty}^{\infty}\ln(\tilde{A}(f))df. \end{aligned}$$

With this, we get

$$\begin{aligned} H_1 - H_2 \propto &-\frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathcal{S}_{\dot{x}+\nabla U_1(x)}(f)}{2\hat{D}_1 + \hat{g}_1^2\mathcal{S}_{\phi_1(x)}(f)}df \\ &-\frac{1}{2}\int_{-\infty}^{\infty}\ln(2\hat{D}_1 + \hat{g}_1^2\mathcal{S}_{\phi_1(x)}(f))df \\ &+\frac{1}{2}\int_{-\infty}^{\infty}\frac{\mathcal{S}_{\dot{x}+\nabla U_2(x)}(f)}{2\hat{D}_2 + \hat{g}_2^2\mathcal{S}_{\phi_2(x)}(f)}df \\ &+\frac{1}{2}\int_{-\infty}^{\infty}\ln(2\hat{D}_2 + \hat{g}_2^2\mathcal{S}_{\phi_2(x)}(f))df. \end{aligned}$$

Accordingly, the cross entropy difference can be evaluated with integrals over the respective power spectra that can be obtained using Fast Fourier Transformation.

G. Timescale of Prediction Error

We here relate the timescale of the prediction error to the timescale of the autocorrelation function $C_x(\tau)/C_x(0) \sim \exp(-\tau/\tau_c)$. The predicted variance in the continuous time limit is determined by the corresponding limit of Eq. (14),

$$\sigma_{\hat{x}}^2 = C_x(\hat{t},\hat{t}) - \int_0^T\int_0^T C_x(\hat{t},u)C_x^{-1}(u,v)C_x(v,\hat{t})\,du\,dv,$$

where T denotes the training interval. Writing $\hat{t} = T + \tau$ and approximating $C_x(T+\tau,u) \approx C_x(T,u)e^{-\tau/\tau_c}$, we get

$$\sigma_{\hat{x}}^2 \approx C_x(\hat{t},\hat{t}) - e^{-2\tau/\tau_c}C_x(T,T),$$

where we used $\int_0^T C_x^{-1}(u,v)C_x(v,T)\,dv = \delta(u-T)$. Using stationarity $C_x(u,v) = C_x(v-u)$, we arrive at

$$\sigma_{\hat{x}}^2/\sigma_x^2 \approx 1 - e^{-2\tau/\tau_c}$$

where $C_x(0) = \sigma_x^2$. Thus, for large τ, the timescale of the prediction error is given by $\tau_c/2$.

[1] M. I. Rabinovich, P. Varona, A. I. Selverston, and H. D. Abarbanel, Rev. Mod. Phys. **78**, 1213 (2006).

[2] H. Sompolinsky, Physics Today **41**, 70 (1988).

[3] S.-I. Amari, Systems, Man and Cybernetics, IEEE Transactions on **SMC-2**, 643 (1972), ISSN 2168-2909.

[4] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Phys. Rev. Lett. **61**, 259 (1988).

[5] M. Stern, H. Sompolinsky, and L. F. Abbott, Phys. Rev. E **90**, 062710 (2014).

[6] J. Kadmon and H. Sompolinsky, Phys. Rev. X **5**, 041030 (2015).

[7] J. Aljadeff, M. Stern, and T. Sharpee, Phys. Rev. Lett. **114**, 088101 (2015).

[8] A. van Meegen and B. Lindner, Phys. Rev. Lett. **121**, 258302 (2018).

[9] J. Schuecker, S. Goedeke, and M. Helias, Phys Rev X **8**, 041029 (2018).

[10] A. Crisanti and H. Sompolinsky, Phys Rev E **98**, 062120 (2018).

[11] G. B. Arous and A. Guionnet, Probability Theory and Related Fields **102**, 455 (1995), ISSN 1432-2064.

[12] A. Guionnet, Probability Theory and Related Fields **109**, 183 (1997).

[13] H. Sompolinsky and A. Zippelius, Phys. Rev. Lett. **47**, 359 (1981).

[14] P. Martin, E. Siggia, and H. Rose, Phys. Rev. A **8**, 423 (1973).

[15] H.-K. Janssen, Zeitschrift für Physik B Condensed Matter **23**, 377 (1976).

[16] C. Chow and M. Buice, J Math. Neurosci **5**, 1 (2015).

[17] J. A. Hertz, Y. Roudi, and P. Sollich, Journal of Physics A: Mathematical and Theoretical **50**, 033001 (2017).

[18] N. Goldenfeld, *Lectures on phase transitions and the renormalization group* (Perseus books, Reading, Massachusetts, 1992).

[19] J. Schuecker, S. Goedeke, D. Dahmen, and M. Helias, arXiv (2016), 1605.06758 [cond-mat.dis-nn].

[20] M. S. Berger, *Nonlinearity and Functional Analysis* (Elsevier, 1977), 1st ed., ISBN 9780120903504.

[21] J. Zinn-Justin, *Quantum field theory and critical phenomena* (Clarendon Press, Oxford, 1996).

[22] M. Mezard and A. Montanari, *Information, physics and computation* (Oxford University Press, 2009).

[23] H. Touchette, Physics Reports **478**, 1 (2009).

[24] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (SIAM, 1995).

[25] D. J. MacKay, *Information theory, inference and learning algorithms* (Cambridge university press, 2003).

[26] G. Matheron, Economic Geology **58**, 1246 (1963).

[27] J. Stapmanns, T. Kühn, D. Dahmen, T. Luu, C. Honerkamp, and M. Helias, Phys. Rev. E **101**, 042124 (2020).