Featured in Physics

Optimal Noise-Canceling Networks

Henrik Ronellenfitsch,[1,*] Jörn Dunkel,[1,†] and Michael Wilczek[2,‡]
[1]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA
[2]Max Planck Institute for Dynamics and Self-Organization, 37077 Göttingen, Germany

(Received 22 July 2018; revised manuscript received 3 October 2018; published 16 November 2018)

Natural and artificial networks, from the cerebral cortex to large-scale power grids, face the challenge of converting noisy inputs into robust signals. The input fluctuations often exhibit complex yet statistically reproducible correlations that reflect underlying internal or environmental processes such as synaptic noise or atmospheric turbulence. This raises the practically and biophysically relevant question of whether and how noise filtering can be hard wired directly into a network's architecture. By considering generic phase oscillator arrays under cost constraints, we explore here analytically and numerically the design, efficiency, and topology of noise-canceling networks. Specifically, we find that when the input fluctuations become more correlated in space or time, optimal network architectures become sparser and more hierarchically organized, resembling the vasculature in plants or animals. More broadly, our results provide concrete guiding principles for designing more robust and efficient power grids and sensor networks.

DOI: 10.1103/PhysRevLett.121.208301

Fluctuations fundamentally limit the function and efficiency of physical [1] and biological [2,3] networks across a wide spectrum of scales. Important examples range from atmospheric turbulence [4,5] affecting large telescope arrays [6], wind farms [7–11], and power grids [12–16] to neuronal noise in the auditory [17,18] and visual [19,20] cortices, and extrinsic and intrinsic fluctuations [21] in gene expression pathways [22,23]. Over the last decades, remarkable progress has been made in the development and understanding of noise-suppression strategies [24,25] and their limits [2,26] in physical [11,27,28] and biological [17,19,29] networks. Classical adaptive noise filtering [30–32] utilizes active control [33,34], and networks can be optimized for active controllability [35–37] and/or transport efficiency [38–43]. Still lacking at present are generic design principles for the construction of optimal passive noise-canceling networks (NCNs). While passive noise reduction has been demonstrated for single oscillators [44], it is not yet well understood how the architecture and efficiency of optimal NCNs depends on the input correlations and cost constraints in natural and man-made systems. Deciphering these dependencies can yield more robust sensory network and power grid designs and may also help clarify the role of noise-reduction in biological network evolution.

Correlated input fluctuations can have profound biomedical or technological consequences in hierarchical network structures. For instance, the detection neurons of the retina are subject to correlated fluctuations [45] which are passed on to the visual cortex where input noise has been shown to affect neural processing [19]. Similarly, deficient noise cancellation in dysfunctional auditory subnetworks has been proposed as a potential cause of tinnitus [17,18]. Another conceptually related problem of rapidly increasing importance is the feed-in of spatiotemporally correlated power fluctuations from solar and wind farms into multinational power grids [5,7,9,11–14,46–49]. These examples raise the general question of the extent to which efficient noise cancellation can be hard wired into a network's architecture if the signal fluctuations have known statistics.

Here, we show both analytically and numerically for generic oscillator networks [12,46,50–52] that it is indeed possible to design optimized weighted network topologies capable of suppressing "colored" fluctuations [9,53] as typically present in biological and engineered systems. In stark contrast to the widely studied problem of optimal synchronization [27,54–63], our results imply that optimal NCNs harness desynchronization to reduce fluctuations globally. Importantly, NCNs operate purely passively, canceling out a substantial fraction of the input fluctuations without requiring active smoothing—the network itself acts as the filter. As a general principle, we find that the more correlated fluctuating inputs are in space or time, the sparser and the more hierarchically organized the NCN will be. Interestingly, the best-performing networks are often reminiscent of leaf venation or animal vasculature, supporting the view that robustness against fluctuations has been an evolutionary factor [39,64]. The mathematical analysis below thus provides detailed guidance for how to use biomimetic network topologies to improve noise robustness in engineered grids and sensor networks.

To investigate noise cancellation in a broadly applicable setting, we consider a generic model of N spatially distributed, nonlinearly coupled second-order phase

 © 2018 American Physical Society

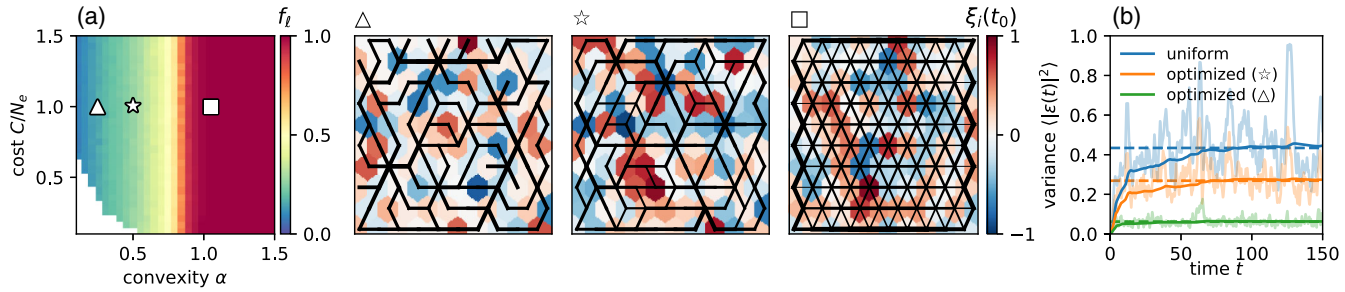

FIG. 1. Even for spatially incoherent white noise input $\sigma, \tau \to 0$, optimal NCNs exhibit a nontrivial sparse topology independent of the nonlinear steady state. (a) The fraction of loops $f_\ell = N_\ell/N_{\text{tri}}$, where N_ℓ is the number of loops in the optimized network and N_{tri} is the number of loops in a triangular grid, measures the topology of optimal networks. Each of the 30×30 pixels in the cost-convexity phase diagram is an average over 15 optimal networks obtained for different uniformly random initial B_{ij}. In the white domain, no solutions to Eq. (2) were found. The NCN topology f_ℓ is effectively independent of C. Panels △, ☆, □ show examples of optimal NCNs with different sparsities, with edge thicknesses proportional to B_{ij}^α. Backgrounds show one instance of the spatial feed-ins $\xi_i(t_0)$ normalized to $(-1, 1)$. (b) Time-averaged variance $\langle|\boldsymbol{\varepsilon}(t)|^2\rangle$ and instantaneous variances $|\boldsymbol{\varepsilon}(t)|^2$ (faint) obtained from numerical solutions of Eq. (1) on uniform and optimized network topologies for $\alpha = 0.25$ (△) and $\alpha = 0.5$ (☆) with edge cost $C = 1$ and centered inputs. Analytically predicted variances (dashed) agree with the simulations.

oscillators, with phase angles $\delta_i(t)$ at each network node i, governed by

$$\ddot{\delta}_i = -\gamma\dot{\delta}_i + \sum_{j=1}^{N} B_{ij}\sin(\delta_i - \delta_j) + P_i(t), \tag{1}$$

where γ is a damping coefficient. The oscillator couplings are symmetric, $B_{ij} = B_{ji}$, and $P_i(t)$ is the fluctuating net signal or power input at site i. Equation (1) has been successfully applied to describe the dynamics of power grids [50]. The Kuramoto model [51,54] is recovered in the overdamped limit, for which all subsequently derived results remain valid after a transformation of parameters (Supplemental Material [65]). The fluctuating inputs can be decomposed as $P_i(t) = \bar{P}_i + \xi_i(t)$, where $\xi_i(t)$ are the fluctuations around the constant mean $\bar{P}_i$. Because Eq. (1) is invariant under a constant shift $\delta_i \to \delta_i + c$, it is possible to split off the irrelevant dynamics of the mean $(1/N)\sum_j \delta_j$ (Supplemental Material [65]). As a result, only the centered inputs $\bar{P}_i^c = \bar{P}_i - (1/N)\sum_j \bar{P}_j$ and $\xi_i^c = \xi_i - (1/N)\sum_j \xi_j$ are relevant. Adopting this mean-centered frame of reference from now on, we write $\delta_i(t) = \bar{\delta}_i + \varepsilon_i(t)$ for constant average phase angles $\bar{\delta}_i$ and fluctuations $\varepsilon_i(t)$. Assuming that the angle fluctuations $\varepsilon_i(t)$ are small and linearizing around $\bar{\delta}_i$, we obtain the coupled set of equations,

$$0 = \sum_{j=1}^{N} B_{ij}\sin(\bar{\delta}_i - \bar{\delta}_j) + \bar{P}_i^c \tag{2}$$

$$\ddot{\varepsilon}_i = -\gamma\dot{\varepsilon}_i + \sum_{j=1}^{N} [B_{ij}\cos(\bar{\delta}_i - \bar{\delta}_j)](\varepsilon_i - \varepsilon_j) + \xi_i^c(t). \tag{3}$$

The zeros of the nonlinear algebraic Eq. (2) correspond to fixed points of Eq. (1). Our main goal here is to use Eq. (3) to derive and characterize optimal couplings B_{ij} that minimize the total fluctuation variance $\langle|\boldsymbol{\varepsilon}(t)|^2\rangle$, where the vector $\boldsymbol{\varepsilon}(t)$ has components $\varepsilon_i(t)$, the total instantaneous variance is the norm $|\boldsymbol{\varepsilon}(t)|^2$, and $\langle\cdot\rangle$ denotes a time average. The optimal network connectivity B_{ij} will depend on the statistics of the input fluctuations, encoded in the elements $R_{ij}(t, t') = \langle\xi_i(t)\xi_j(t')\rangle$ of the covariance matrix R.

Throughout, we assume that spatiotemporal correlations factorize, although the general approach extends to the nonfactorizing case. For the time correlations, we focus on colored Ornstein-Uhlenbeck noise [53] with $R(t, t') = \hat{R}e^{-|t-t'|/\tau}/(2\tau)$. In the limit of correlation time $\tau \to 0$, white noise is recovered with $R(t, t') = \hat{R}\delta(t - t')$. For the spatial part $\hat{R} = (\hat{R}_{ij})$, we choose generic isotropic and homogeneous Gaussian covariances $\hat{R}_{ij} = e^{-|\mathbf{x}_i - \mathbf{x}_j|^2/(2\sigma^2)}$, where $\mathbf{x}_i$ is the spatial position of oscillator i and σ is a correlation length. In the limit $\sigma \to 0$, the feed-ins become incoherent with $\hat{R}_{ij} = \delta_{ij}$. The total fluctuation variance $\langle|\boldsymbol{\varepsilon}(t)|^2\rangle$ can be calculated analytically for any $\hat{R}$ in the Langevin formalism (Supplemental Material [65]),

$$\langle|\boldsymbol{\varepsilon}(t)|^2\rangle = \frac{1}{2\gamma}\mathrm{tr}\left[\left(\mathbb{1} + \frac{\tau^2}{1+\gamma\tau}L\right)^{-1}L^\dagger\hat{R}\right], \tag{4}$$

where L is the weighted graph Laplacian matrix of the network with the weights of edge (ij) given by $B_{ij}\cos(\bar{\delta}_i - \bar{\delta}_j)$, and tr is the matrix trace. The pseudoinverse $L^\dagger$ implicitly acts as a projection to center $\hat{R}$. In the white-noise limit $\tau \to 0$, Eq. (4) reduces to

$$\langle|\boldsymbol{\varepsilon}(t)|^2\rangle = \frac{1}{2\gamma}\mathrm{tr}(L^\dagger\hat{R}). \tag{5}$$

The structure of Eqs. (4) and (5) implies that, in principle, arbitrarily small variances $\langle|\boldsymbol{\varepsilon}(t)|^2\rangle$ can be achieved by choosing the B_{ij} arbitrarily large. In natural or engineered

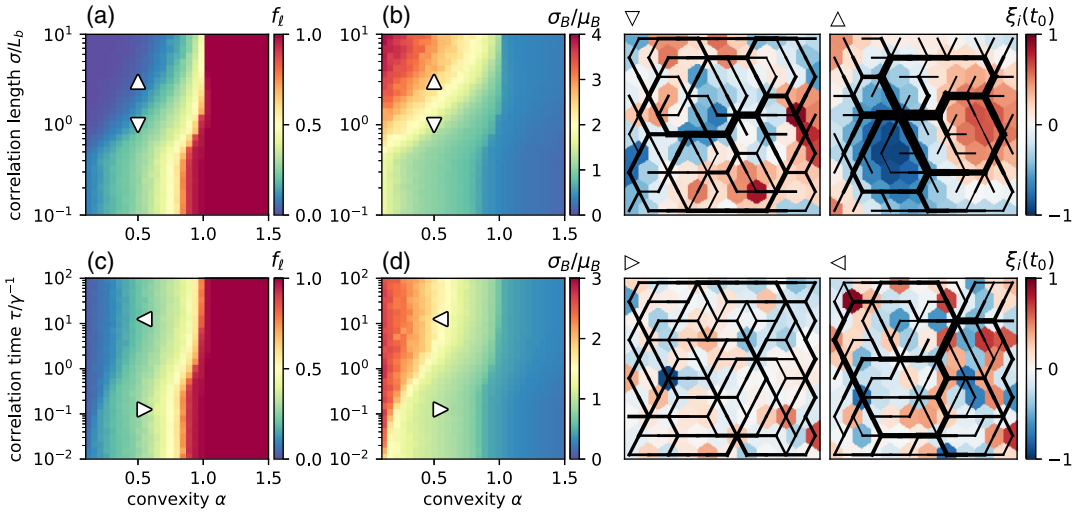

FIG. 2. Spatial and temporal input correlations lead to a similar hierarchical NCN organization despite acting through different mechanisms. (a),(b) Gaussian spatial correlations $\sigma > 0$ with temporal white noise $\tau \to 0$. The loop fractions f_ℓ in (a) show that NCN topology depends largely on α, although the transition between loopy and sparse phase shifts when the correlation scale σ approaches the mean edge length L_b. For $\sigma \gg L_b$ networks become sparser when $\alpha \sim 1$. (b) The coupling variance σ_B, normalized by the mean μ_B, indicates that nonuniform hierarchical patterns and sparsity are strongly correlated. (c),(d) Ornstein-Uhlenbeck colored noise $\tau > 0$ with spatially incoherent feed-ins $\sigma \to 0$ shows hierarchical patterns similar to those in panels (a),(b). Examples of optimal networks at the positions marked by symbols in the phase diagrams illustrate the transitions from dense uniform networks to sparse hierarchical networks with increasing spatial or temporal correlation. Each of the 30×30 pixels in (a)–(d) is an average over 15 optimal networks.

real-world networks, however, the allowed values of the B_{ij} are restricted by construction or maintenance costs. To account for this fact, we adopt here the widely used [39,40,57,64,66] cost constraint $\sum_{(ij)} B_{ij}^{\alpha} = N_e C$, where $\alpha > 0$ is a convexity parameter, C the cost per edge, and N_e the number of edges in the network. In the concave regime $\alpha < 1$, one expects sparse networks because it becomes more economical to construct a single edge with a large coupling rather than to distribute over, say, two smaller ones. Since many natural networks are sparse, and sparsity is desirable in engineering, this concave range arguably comprises the most interesting part of phase space. The cost-constrained optimization is carried out starting from a given base network and initial B_{ij}. Optimal weights are found iteratively using the method of Lagrange multipliers (Supplemental Material [65]). Weights $B_{ij} = 0$ in the final optimized network correspond to edges being pruned from the base network, and thus to changes in topology. In the case of white noise in time and close to synchrony ($\bar{\delta}_i \approx 0$), the minima have an interesting interpretation: using the eigendecomposition $\hat{R} = \sum_k \rho_k \boldsymbol{r}_k \boldsymbol{r}_k^{\top}$, one finds the defining relation $\alpha\lambda B_{ij}^{\alpha+1} = \sum_k \rho_k [B_{ij}(\varepsilon_i^{(k)} - \varepsilon_j^{(k)})]^2$, where the $\boldsymbol{\varepsilon}^{(k)}$ are steady-state angles in the presence of steady feed-ins $\boldsymbol{r}_k$. Thus, the optimal couplings are directly related to a weighted average over *local* steady state flows. In the general case, additional terms appear (Supplemental Material [65]). Armed with these analytical insights, we now turn to the numerical investigation of optimal NCNs for different input noise statistics.

As base networks, we explore planar triangular grids which are approximately realized in many biological and engineering systems such as cilia [67,68] or staggered wind farms [69]. The number of nodes is $N = 100$ and damping is fixed at $\gamma = 0.5$, following Ref. [50]. The uncentered steady feed-ins are $\bar{P}_i = \eta_i$, where the η_i are independent Gaussian random variables with zero mean and unit variance. Covariance matrices are normalized to $\mathrm{tr}(\hat{R}) = 1$, bringing steady state background and fluctuations to a similar scale. Numerical solutions of Eq. (1) were obtained using the Euler-Maruyama scheme with time step $\Delta t = 10^{-3}$. All main results remain valid for other grid geometries (Supplemental Material [65]).

Already in the simplest case, when node inputs are white noise in time ($\tau \to 0$) and spatially incoherent ($\sigma \to 0$), optimal NCNs exhibit nontrivial topologies in the sparse regime $0 < \alpha < 1$ [Fig. 1]. The fraction of loops $f_\ell = N_\ell / N_{\mathrm{tri}}$, where N_ℓ is the number of loops in the optimal network and N_{tri} is the number of loops in the full triangular grid, decreases with α [Fig. 1(a)]. This indicates that optimal NCNs become sparser for $\alpha \to 0$. The nonzero couplings in the optimized network have similar magnitude for uncorrelated inputs [Fig. 1(a),(△,☆)], and the optimal networks do not follow any symmetry of the base network. As expected, optimal networks become dense for $\alpha > 1$ [Fig. 1(a),(□)] and retain the base network topology. The nonlinear steady state, despite being fully taken into account in our optimization procedure, has little influence on the structure of optimal NCNs. Decreasing the mean

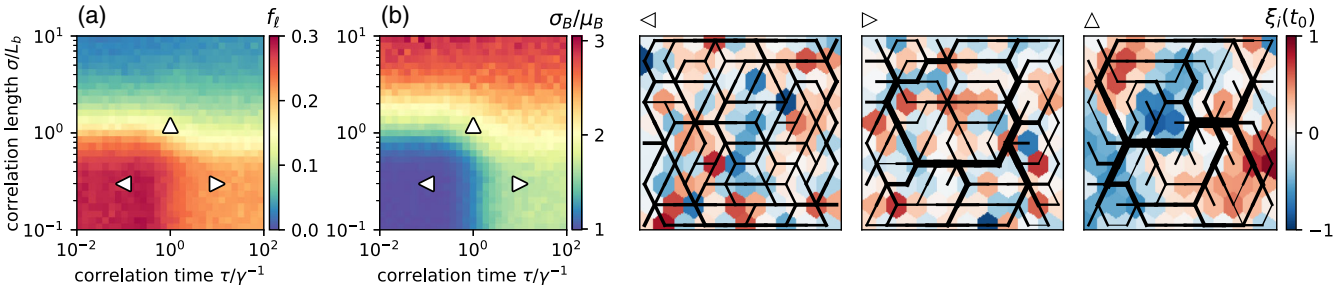

FIG. 3. Combining spatial and temporal correlations leads to three qualitatively distinct NCN phases in the (τ, σ) plane. (a) The loop density f_ℓ characterizes the three phases as follows: Short correlation times τ and short correlation lengths σ favor highly reticulate redundant networks (◁), large τ and small σ lead to a moderate reticulation (▷), whereas large τ and large σ selects low reticulation (△). (b) The coupling spread σ_B/μ_B indicates a similar division of the (τ, σ)-phase plane: Low τ, σ lead to highly uniform networks (◁), high τ and low σ lead to networks with an intermediate coupling variability (▷), and high τ, σ lead to strongly hierarchical networks with large spread in the couplings B_{ij} (△). The three phases are separated approximately by the lines $\tau/\gamma^{-1} \sim 1$ and $\sigma/L_b \sim 1$. Each pixel in the 30×30 plots (a),(b) is an average over 15 optimal networks; $\alpha = 0.5$, $C = 1$ in all panels.

coupling through the cost C pushes the NCNs towards the regime lacking solutions of Eq. (2), but causes no significant changes in topology apart from an overall scaling of the couplings, even very close to the transition [Fig. 1(a)]. Simulations of the full nonlinear Eq. (1) on the identified sparse NCNs confirm a significant noise reduction compared to uniform weights, in quantitative agreement with the predictions of the linear model [dashed and solid lines in Fig. 1(b)]. In general, the linear approximation is accurate as long as the noise is small compared to a worst-case uniform distribution $\varepsilon_i \in [-\pi, \pi)$ (Supplemental Material [65]). Since the optimal topologies show little dependence on the nonlinear steady state [Fig. 1(a)], it suffices to focus on the synchronized limit $\bar{\delta}_i = 0$ and $C = 1$ when considering correlated noisy inputs in the remainder. The existence of nontrivial optimal NCN topologies even for uncorrelated inputs is remarkable, and may already have practical applications.

Even more interesting hierarchical NCN structures arise when the input noise becomes correlated [Fig. 2]. The optimal couplings settle into nonuniform patterns containing loopy backbone structures with treelike branches, reminiscent of plant [70,71], fungal [72], or animal [73] vasculature [Fig. 2(△)]. To dissect the effects of correlations, we first consider fluctuating inputs that are still uncorrelated in time ($\tau \to 0$) but have a finite correlation length $\sigma > 0$. Our numerical analysis shows that the topology of optimal NCNs changes as σ is varied relative to the mean edge length L_b, the latter defining the natural resolution scale for a network. As expected, for $\sigma \ll L_b$, we find the same NCN topology as for incoherent inputs [Figs. 1 and 2]. In contrast, when σ becomes comparable to or larger than L_b, the optimal NCNs become significantly sparser for $0 < \alpha < 1$ [Fig. 2(a)]. This transition is accompanied by the gradual emergence of a hierarchical network structure, reflected by an increased standard deviation σ_B of the optimal coupling parameters B_{ij} relative to their mean μ_B [Fig. 2(b)]. Thus, NCNs for spatially correlated white noise develop hierarchical sparse architectures as the correlation length σ increases.

These observations can be rationalized by noting that in the limit of large σ, we have $\hat{R} \sim D$, where D is the matrix of squared Euclidean distances between oscillators. The rank of D is at most the dimension d of the embedding space [74]. Therefore, the objective Eq. (5) becomes equivalent to an average over at most d steady-state inputs. For networks with a single nonfluctuating input, it is known that the optimal topology is a maximally sparse tree [38]. Since $d = 2$ in our case, the optimal NCNs are close to such trees. This argument holds for any sufficiently well-behaved $\hat{R} = f(D/\sigma^2)$ that depends on the node distances via a scale parameter. The emergence of the hierarchical structure follows from the earlier stated fact that couplings become proportional to a mean flow, which in a treelike topology of steady inputs accumulates as the network graph is traversed upstream from a leaf node. Remarkably, for large σ, the optimal NCNs often exhibit spontaneous symmetry breaking by approximately realizing *rooted* trees, in which a hierarchical backbone emanates from one or two central nodes [Fig. 2(△)] even though no such distinguished node(s) were initially prescribed.

Interestingly, colored noise with nonvanishing correlation time $\tau > 0$ but no spatial coherence ($\sigma \to 0$) has qualitatively similar effects on the network structure. When τ is larger than the damping timescale γ^{-1}, optimal NCNs also become sparser and more hierarchically patterned [Figs. 2(c),2(d) and (◁, ▷)]. The origin of sparsity is now different because $\hat{R}$ is almost full rank for $\sigma \to 0$, and related to the large-τ asymptotic behavior of the objective, $\langle |\boldsymbol{\varepsilon}(t)|^2 \rangle \sim \mathrm{tr}[(L^\dagger)^2 \hat{R}]/(2\tau)$. Although the objective does not scale homogeneously with C anymore, only the transition between the different NCN topologies changes (Supplemental Material [65]).

Lastly, combining spatial and temporal correlations, the (τ, σ) plane subdivides into three distinct phases [Figs. 3(a), 3(b)]. For $\sigma \ll L_b$ and $\tau \ll \gamma^{-1}$, optimal NCNs are highly dense and uniform [Fig. 3(◁)]. For $\sigma \ll L_b$ but $\tau \gg \gamma^{-1}$, NCNs exhibit intermediate sparsity and hierarchical patterning [Fig. 3(▷)]. For $\sigma \gg L_b$, NCNs become generally sparse and hierarchically patterned with little dependence on τ [Fig. 3(△)], although the transition between the different NCNs topologies is shifted to smaller σ when $\tau \gg \gamma^{-1}$.

To conclude, the above analytical and numerical results show that noise cancellation can be hard wired into weighted network topology for both uncorrelated and correlated input fluctuations. As a general rule, the more correlated the input fluctuations, the sparser and more hierarchically ordered the optimal networks become. Previous work [1,51] has demonstrated the applicability of the underlying phase oscillator framework to a myriad of physical and biological systems, from neuronal networks [56,75] and ciliary carpets [76–78] to renewable energy farms and power grids [9,12,14,50]. One can therefore expect that the above ideas and results have conceptual and practical implications for most, if not all, of these systems.

This work was supported by an Edmund F. Kelly Research Award (J. D.) and a James S. McDonnell Foundation Complex Systems Scholar Award (J. D.) and the Max Planck Society (M. W.).

*henrikr@mit.edu
†dunkel@mit.edu
‡michael.wilczek@ds.mpg.de

[1] B. Lindner, J. Garca-Ojalvo, A. Neiman, and L. Schimansky-Geier, Phys. Rep. **392**, 321 (2004).
[2] W. Bialek, Annu. Rev. Biophys. Biophys. Chem. **16**, 455 (1987).
[3] L. S. Tsimring, Rep. Prog. Phys. **77**, 026601 (2014).
[4] J. Wyngaard, *Turbulence in the Atmosphere* (Cambridge University Press, Cambridge, England, 2010).
[5] P. Milan, M. Wächter, and J. Peinke, Phys. Rev. Lett. **110**, 138701 (2013).
[6] A. Wootten and A. R. Thompson, Proc. IEEE **97**, 1463 (2009).
[7] P. Sørensen, A. D. Hansen, and P. A. C. Rosas, J. Wind Eng. Ind. Aerodyn. **90**, 1381 (2002).
[8] C. Luo, H. Banakar, B. Shen, and B.-T. Ooi, IEEE Trans. Energy Convers. **22**, 341 (2007).
[9] L. J. Lukassen, R. J. A. M. Stevens, C. Meneveau, and M. Wilczek, Wind Energy **21**, 474 (2018).
[10] Z. Chen, in *Transmission and Distribution Conference and Exhibition, Asia and Pacific, 2005 IEEE/PES* (IEEE, New York, 2005), p. 1, doi: 10.1109/TDC.2005.1547160.
[11] J. O. G. Tande, Wind Energy **6**, 281 (2003).
[12] K. Schmietendorf, J. Peinke, and O. Kamps, Eur. Phys. J. B **90**, 222 (2017).
[13] B. Schäfer, C. Beck, K. Aihara, D. Witthaut, and M. Timme, Nat. Energy **3**, 119 (2018).
[14] T. Nesti, A. Zocca, and B. Zwart, Phys. Rev. Lett. **120**, 258301 (2018).
[15] H. Haehne, J. Schottler, M. Waechter, J. Peinke, and O. Kamps, Europhys. Lett. **121**, 30001 (2018).
[16] T. Coletta, B. Bamieh, and P. Jacquod, Proceedings of CDC '18 (to be published).
[17] D. D. Ridder, S. Vanneste, N. Weisz, A. Londero, W. Schlee, A. B. Elgoyhen, and B. Langguth, Neurosci. Biobehav. Rev. **44**, 16 (2014).
[18] A. M. Leaver, L. Renier, M. A. Chevillet, S. Morgan, H. J. Kim, and J. P. Rauschecker, Neuron **69**, 33 (2011).
[19] B. B. Averbeck, P. E. Latham, and A. Pouget, Nat. Rev. Neurosci. **7**, 358 (2006).
[20] I. Kanitscheider, R. Coen-Cagli, and A. Pouget, Proc. Natl. Acad. Sci. U.S.A. **112**, E6973 (2015).
[21] A. Hilfinger and J. Paulsson, Proc. Natl. Acad. Sci. U.S.A. **108**, 12167 (2011).
[22] M. Thattai and A. Van Oudenaarden, Proc. Natl. Acad. Sci. U.S.A. **98**, 8614 (2001).
[23] J. Paulsson, Nature (London) **427**, 415 (2004).
[24] L. O. Chua and L. Yang, IEEE Circuits Syst. Mag. **35**, 1273 (1988).
[25] S. Moon and J.-N. Hwang, in *Acoustics, Speech, and Signal Processing, 1993, ICASSP-93, 1993 IEEE International Conference on* Vol. 1 (IEEE, New York, 1993), p. 573, doi: 10.1109/ICASSP.1993.319183.
[26] I. Lestas, G. Vinnicombe, and J. Paulsson, Nature (London) **467**, 174 (2010).
[27] B. Li and K. Y. Michael Wong, Phys. Rev. E **95**, 012207 (2017).
[28] M. Tyloo, T. Coletta, and P. Jacquod, Phys. Rev. Lett. **120**, 084101 (2018).
[29] C. Zechner, G. Seelig, M. Rullan, and M. Khammash, Proc. Natl. Acad. Sci. U.S.A. **113**, 4729 (2016).
[30] R. Bucy, J. Math. Anal. Appl. **20**, 1 (1967).
[31] B. Widrow, J. Glover, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, J. Eugene Dong, and R. Goodlin, Proc. IEEE **63**, 1692 (1975).
[32] P. M. Clarkson, *Optimal and Adaptive Signal Processing*, 1st ed. (CRC Press, New York, 1993).
[33] I. Klickstein, A. Shirin, and F. Sorrentino, Nat. Commun. **8**, 15145 (2017).
[34] G. Yan, G. Tsekenis, B. Barzel, J.-J. Slotine, Y.-Y. Liu, and A.-L. Barabási, Nat. Phys. **11**, 779 (2015).
[35] Y.-D. Xiao, S.-Y. Lao, L.-L. Hou, and L. Bai, Phys. Rev. E **90**, 042804 (2014).
[36] M. Liang, S. Jin, D. Wang, and X. Zou, Eur. Phys. J. B **89**, 186 (2016).
[37] W.-X. Wang, X. Ni, Y.-C. Lai, and C. Grebogi, Phys. Rev. E **85**, 026115 (2012).
[38] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, Phys. Rev. Lett. **84**, 4745 (2000).
[39] E. Katifori, G. J. Szöllősi, and M. O. Magnasco, Phys. Rev. Lett. **104**, 048704 (2010).
[40] S. Bohn and M. O. Magnasco, Phys. Rev. Lett. **98**, 088702 (2007).
[41] D. Segrè, D. Vitkup, and G. M. Church, Proc. Natl. Acad. Sci. U.S.A. **99**, 15112 (2002).

[42] M. Durand, Phys. Rev. E **73**, 016116 (2006).
[43] B. W. Andrews, T.-M. Yi, and P. A. Iglesias, PLoS Comput. Biol. **2**, e154 (2006).
[44] E. Kenig, M. C. Cross, R. Lifshitz, R. B. Karabalin, L. G. Villanueva, M. H. Matheny, and M. L. Roukes, Phys. Rev. Lett. **108**, 264102 (2012).
[45] P. Ala-Laurila, M. Greschner, E. J. Chichilnisky, and F. Rieke, Nat. Neurosci. **14**, 1309 (2011).
[46] B. Schäfer, M. Matthiae, X. Zhang, M. Rohden, M. Timme, and D. Witthaut, Phys. Rev. E **95**, 060203 (2017).
[47] K. Schmietendorf, J. Peinke, R. Friedrich, and O. Kamps, Eur. Phys. J. Spec. Top. **223**, 2577 (2014).
[48] D. Manik, M. Rohden, H. Ronellenfitsch, X. Zhang, S. Hallerberg, D. Witthaut, and M. Timme, Phys. Rev. E **95**, 012319 (2017).
[49] M. M. Bandi, Phys. Rev. Lett. **118**, 028301 (2017).
[50] P. H. Nardelli, N. Rubido, C. Wang, M. S. Baptista, C. Pomalaza-Raez, P. Cardieri, and M. Latva-aho, Eur. Phys. J. Spec. Top. **223**, 2423 (2014).
[51] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler, Rev. Mod. Phys. **77**, 137 (2005).
[52] J. Kromer, A. Khaledi-Nasab, L. Schimansky-Geier, and A. B. Neiman, Sci. Rep. **7**, 3956 (2017).
[53] P. Hänggi and P. Jung, Adv. Chem. Phys. **89**, 239 (1994).
[54] F. A. Rodrigues, T. K. Peron, P. Ji, and J. Kurths, Phys. Rep. **610**, 1 (2016).
[55] B. C. Bag, K. G. Petrosyan, and C.-K. Hu, Phys. Rev. E **76**, 056210 (2007).
[56] C. M. Gray, J. Comput. Neurosci. **1**, 11 (1994).
[57] T. Tanaka and T. Aoyagi, Phys. Rev. E **78**, 046210 (2008).
[58] M. Fazlyab, F. Dörfler, and V. M. Preciado, Automatica **84**, 181 (2017).
[59] P. S. Skardal, D. Taylor, and J. Sun, Phys. Rev. Lett. **113**, 144101 (2014).
[60] M. Fardad, F. Lin, and M. R. Jovanovic, IEEE Trans. Autom. Control **59**, 2457 (2014).
[61] D. Kelly and G. A. Gottwald, Chaos **21**, 025110 (2011).
[62] M. Brede, Phys. Lett. A **372**, 2618 (2008).
[63] J. H. Meng and H. Riecke, Sci. Rep. **8**, 6949 (2018).
[64] D. Hu and D. Cai, Phys. Rev. Lett. **111**, 138701 (2013).
[65] See Supplemental Material http://link.aps.org/supplemental/10.1103/PhysRevLett.121.208301 for detailed derivations, connection to the Kuramoto model, detailed description of the numerical methods, analysis of noise-canceling improvement, validity of the linear model, dependence of the topology on correlation time, and analysis of square grid topologies.
[66] H. Ronellenfitsch and E. Katifori, Phys. Rev. Lett. **117**, 138301 (2016).
[67] A. J. Hudspeth, Neuron **59**, 530 (2008).
[68] K. Drescher, R. E. Goldstein, and I. Tuval, Proc. Natl. Acad. Sci. U.S.A. **107**, 11171 (2010).
[69] R. J. A. M. Stevens and C. Meneveau, Annu. Rev. Fluid Mech. **49**, 311 (2017).
[70] L. Sack and C. Scoffoni, New Phytol. **198**, 983 (2013).
[71] H. Ronellenfitsch, J. Lasser, D. C. Daly, and E. Katifori, PLoS Comput. Biol. **11**, e1004680 (2015).
[72] L. Heaton, B. Obara, V. Grau, N. Jones, T. Nakagaki, L. Boddy, and M. D. Fricker, Fungal Biol. Rev. **26**, 12 (2012).
[73] M. Fruttiger, Invest. Ophthalmol. Visual Sci. **43**, 522 (2002).
[74] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, IEEE Signal Process. Mag. **32**, 12 (2015).
[75] Y. Penn, M. Segal, and E. Moses, Proc. Natl. Acad. Sci. U.S.A. **113**, 3341 (2016).
[76] T. Niedermayer, B. Eckhardt, and P. Lenz, Chaos **18**, 037128 (2008).
[77] N. Uchida and R. Golestanian, Europhys. Lett. **89**, 50011 (2010).
[78] D. R. Brumley, M. Polin, T. J. Pedley, and R. E. Goldstein, Phys. Rev. Lett. **109**, 268102 (2012).

Supplemental Material: Optimal noise-canceling networks

Henrik Ronellenfitsch, Jörn Dunkel, and Michael Wilczek
(Dated: October 3, 2018)

I. CENTERED DYNAMICS

Eq. (1) from the main paper contains a freedom of re-defining $\delta_i \to \delta_i + c$ for some constant c corresponding to a reference angle. Here, we fix this freedom by introducing the new variables

$$\psi_i(t) = \delta_i(t) - \mu(t)$$
$$\mu(t) = \frac{1}{N}\sum_j \delta_j(t).$$

Taking derivatives and plugging them into Eq. (1), we find that they satisfy

$$\ddot{\psi} = -\gamma\dot{\psi} + \sum_j B_{ij}\sin(\psi_i - \psi_j) + P_i(t) - \frac{1}{N}\sum_j P_j(t) \tag{S1}$$

$$\ddot{\mu} = -\gamma\dot{\mu} + \frac{1}{N}\sum_j P_j(t), \tag{S2}$$

where we used $\sum_{i,j} B_{ij}\sin(\psi_i - \psi_j) = 0$ due to antisymmetry. Equation (S1) is equivalent to Eq. (1) but with centered inputs, and Eq. (S2) describes a stochastically forced particle with damping. We decompose the inputs into constant means and stochastic fluctuations, $P_i(t) = \bar{P}_i + \xi_i(t)$. Without fluctuations, a steady state is only possible if the constant forcing in Equation (S2) vanishes, $\sum_j \bar{P}_j = 0$. We shall assume this to be true from here on and focus on Eq. (S1), because the dynamics of the mean is independent of the weighted network topology encoded in the B_{ij}. We find the centered dynamics

$$\ddot{\psi} = -\gamma\dot{\psi} + \sum_j B_{ij}\sin(\psi_i - \psi_j) + \bar{P}_i + \xi_i(t) - \frac{1}{N}\sum_j \xi_j(t). \tag{S3}$$

Equation (S3) is again simply Eq. (1) but with *centered* fluctuations. It is only these centered fluctuations that are relevant for optimal NCNs. In vector form they can be written using the projection matrix Q as

$$Q\boldsymbol{\xi} = \left(\mathbb{1} - \frac{1}{N}J\right)\boldsymbol{\xi},$$

where $J_{ij} = 1$. Similarly, the centered correlation matrix is

$$R_c = Q\langle\boldsymbol{\xi}\boldsymbol{\xi}^\top\rangle Q = QRQ.$$

II. DERIVATION OF THE OBJECTIVE FUNCTION

In this section we derive the objective function for white noise and colored noise. Note that unlike in the main paper, for notational ease we use the inverse correlation time scale $\kappa = \tau^{-1}$. We first consider the case of pure white noise, and then generalize to Ornstein-Uhlenbeck colored noise.

A. White Noise

Here, we compute the variance of fluctuations directly in the Langevin formalism.

We consider the linearized second-order system in the centered frame from the preceding section,

$$\ddot{\boldsymbol{\varepsilon}} + \gamma\dot{\boldsymbol{\varepsilon}} + L\boldsymbol{\varepsilon} = Q\boldsymbol{\xi}(t),$$

where $\langle\boldsymbol{\xi}\rangle = \mathbf{0}$, $\langle\boldsymbol{\xi}(t)\boldsymbol{\xi}(t')^\top\rangle = QRQ\,\delta(t-t')$ is white noise input in time with spatial correlation matrix R. We can rewrite the system as first order,

$$\begin{pmatrix}\dot{\boldsymbol{\varepsilon}}\\ \dot{\boldsymbol{\nu}}\end{pmatrix} = \begin{pmatrix}0 & \mathbb{1}\\ -L & -\gamma\mathbb{1}\end{pmatrix}\begin{pmatrix}\boldsymbol{\varepsilon}\\ \boldsymbol{\nu}\end{pmatrix} + \begin{pmatrix}0\\ Q\boldsymbol{\xi}\end{pmatrix}$$
$$\Leftrightarrow \dot{\boldsymbol{y}} = M\boldsymbol{y} + \boldsymbol{u}.$$

The solution to this system can be expressed as

$$\boldsymbol{y}(t) = \exp(Mt)\boldsymbol{y}_0 + \int_0^T \exp(M(t-t'))\boldsymbol{u}(t')\,dt'.$$

The eigenvalues of M are easy to compute by explicitly writing down the eigenvector condition in block-matrix form. One obtains

$$\lambda_{i,\pm} = -\frac{\gamma}{2} \pm \sqrt{\frac{\gamma^2}{4} - \omega_i^2},$$

where the ω_i^2 are the (positive) eigenvalues of the Laplacian L. Since $\mathrm{Re}(\lambda_{i,\pm}) < 0$ except for the eigenvector of all 1's in the first block, the homogeneous solution $\exp(Mt)\boldsymbol{y}_0$ decays for large times except for a constant angular shift. In the following, we change into a frame where this shift vanishes and focus on the particular solution.

We want to compute the matrix of correlations for large times,

$$\begin{aligned}
\langle \boldsymbol{y}(t)\boldsymbol{y}(t')^\top \rangle &= \int_0^T ds \int_0^{t'} ds' \exp(M(t-s))\langle \boldsymbol{u}(s)\boldsymbol{u}^\top(s')\rangle \exp(M^\top(t'-s')) \\
&= \int_0^T ds \int_0^{t'} ds' \exp(Ms)\hat{R}\exp(M^\top s')\,\delta(t-t'-s+s') \\
&= \int_0^\infty ds \exp(M(s+(t-t')))\hat{R}\exp(M^\top s).
\end{aligned} \tag{S4}$$

We substituted $s \to t-s, s' \to t'-s'$, used the fact that

$$\langle \boldsymbol{u}(t)\boldsymbol{u}^\top(t')\rangle = \begin{pmatrix} 0 & 0 \\ 0 & Q\langle \boldsymbol{\xi}(t)\boldsymbol{\xi}^\top(t')\rangle Q \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & QRQ \end{pmatrix}\delta(t-t') = \hat{R}\,\delta(t-t'),$$

and finally took the limit of $t, t' \to \infty$ while keeping $t-t'$ fixed. Since we want to find the variance, we now set $t-t'=0$. This matrix-valued integral cannot be evaluated directly, but we can integrate by parts to obtain

$$\begin{aligned}
\langle \boldsymbol{y}(0)\boldsymbol{y}(0)^\top \rangle = E &= \int_0^\infty ds \exp(Ms)\hat{R}\exp(M^\top s) \\
&= \left[M^\dagger \exp(Ms)\hat{R}\exp(M^\top s)\right]_0^\infty - M^\dagger \int_0^\infty ds \exp(Ms)\hat{R}\exp(M^\top s)M^\top \\
&= -M^\dagger \hat{R} - M^\dagger E M^\top \\
\Leftrightarrow ME + EM^\top &= -\hat{R}.
\end{aligned} \tag{S5}$$

This matrix equation for E is called the *Lyapunov equation*, and there is no analytic expression for its solution. (Note that we used the pseudo-inverse. This is allowed because even though M has a nontrivial nullspace of dimension 1 corresponding to $(\mathbf{1}, \mathbf{0})$, this nullspace is projected out by $\hat{R}$.) The total variance of the fluctuations $\boldsymbol{\varepsilon}(t)$ is encoded in the trace of the upper-left block of E. We write the Lyapunov equation explicitly in block-form,

$$\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}\begin{pmatrix} 0 & -L \\ \mathbb{1} & -\gamma\mathbb{1} \end{pmatrix} + \begin{pmatrix} 0 & \mathbb{1} \\ -L & -\gamma\mathbb{1} \end{pmatrix}\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & -QRQ \end{pmatrix}, \tag{S6}$$

where we made the symmetric ansatz $E = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$ with $A^\top = A$ and $C^\top = C$ (Remember that E is a correlation matrix and therefore symmetric). Our goal is now to find an expression for $\mathrm{tr}(A)$. Multiplying out yields the equations

$$\begin{aligned}
B + B^\top &= 0 \\
C - AL - \gamma B &= 0 \\
C - LA - \gamma B^\top &= 0 \\
2\gamma C + LB + B^\top L &= QRQ.
\end{aligned}$$

Adding and subtracting the first and second yields

$$\begin{aligned}
C &= \frac{1}{2}(AL + LA) \\
B &= \frac{1}{2\gamma}(LA - AL).
\end{aligned} \tag{S7}$$

Plugging these into the third and taking the trace,

$$\begin{aligned} 2\gamma\frac{1}{2}(AL+LA)+\frac{1}{2\gamma}(L^2A-LAL-LAL+AL^2)&=QRQ \\ \Rightarrow \gamma(L^\dagger AL+L^\dagger LA)+\frac{1}{2\gamma}(L^\dagger L^2A-2L^\dagger LAL+L^\dagger AL^2)&=L^\dagger QRQ \\ \Rightarrow 2\gamma\,\mathrm{tr}_{\perp\mathbf{1}}(A)=\mathrm{tr}(L^\dagger QRQ)&=\mathrm{tr}(L^\dagger R). \end{aligned} \tag{S8}$$

Here, we can only take the trace over the subspace perpendicular to the vector with all ones, because that is the subspace that $L^\dagger$ projects on. Additionally, we used the fact that $QL=LQ=L$ because L is a graph Laplacian whose kernel is spanned by the vector $\mathbf{1}$ of all 1's.

We now show that $\mathbf{1}^\top A\mathbf{1}=0$. We compute directly

$$\begin{aligned} \mathbf{1}^\top A\mathbf{1}&=(\mathbf{1}^\top,0)E(\mathbf{1}^\top,0)^\top \\ &=\int_0^\infty ds(\mathbf{1}^\top,0)\exp(Ms)\hat{R}\exp(M^\top s)(\mathbf{1}^\top,0)^\top. \end{aligned}$$

It is easy to compute

$$\begin{aligned} M^\top\begin{pmatrix}\mathbf{1}\\0\end{pmatrix}&=\begin{pmatrix}0\\\mathbf{1}\end{pmatrix} \\ M^\top\begin{pmatrix}0\\\mathbf{1}\end{pmatrix}&=-\gamma\begin{pmatrix}0\\\mathbf{1}\end{pmatrix}. \end{aligned}$$

Therefore, the matrix exponential can be expanded into a series,

$$\begin{aligned} \exp(M^\top s)\begin{pmatrix}\mathbf{1}\\0\end{pmatrix}&=\begin{pmatrix}\mathbf{1}\\0\end{pmatrix}-\frac{1}{\gamma}\begin{pmatrix}0\\\mathbf{1}\end{pmatrix}\sum_{n=1}^\infty\frac{(-\gamma s)^n}{n!} \\ &=\begin{pmatrix}\mathbf{1}\\0\end{pmatrix}-\frac{1}{\gamma}\begin{pmatrix}0\\\mathbf{1}\end{pmatrix}(e^{-\gamma s}-1). \end{aligned}$$

From the structure of $\hat{R}$, we immediately obtain,

$$\hat{R}\exp(M^\top s)\begin{pmatrix}\mathbf{1}\\0\end{pmatrix}=-\frac{1}{\gamma}\begin{pmatrix}0\\QRQ\mathbf{1}\end{pmatrix}(e^{-\gamma s}-1)=0.$$

Thus $\mathbf{1}^\top A\mathbf{1}=0$, the trace over the perpendicular subspace is actually the full trace, and we obtain,

$$\langle|\boldsymbol{\varepsilon}(t)|^2\rangle=\frac{1}{2\gamma}\,\mathrm{tr}(L^\dagger R). \tag{S9}$$

B. Colored noise

We now assume that $\langle\boldsymbol{\xi}(t)\boldsymbol{\xi}(t')^\top\rangle=\frac{\kappa}{2}e^{-\kappa|t-t'|}QRQ$. (Remember that $\kappa=\tau^{-1}$ is the inverse time scale). We can express Eq. (S4) as follows, taking the long-time limits,

$$\begin{aligned} \frac{2}{\kappa}\langle\boldsymbol{y}(0)\boldsymbol{y}(0)^\top\rangle&=\int_0^\infty ds\int_0^\infty ds' e^{-\kappa|s-s'|}e^{Ms}\hat{R}e^{M^\top s'} \\ &=\int_0^\infty ds\,e^{(M-\kappa\mathbb{1})s}\hat{R}\int_0^s ds' e^{(M^\top+\kappa\mathbb{1})s'}+\int_0^\infty ds\,e^{(M+\kappa\mathbb{1})s}\hat{R}\int_s^\infty ds' e^{(M^\top-\kappa\mathbb{1})s'} \\ &=(M-\kappa\mathbb{1})^{-1}\hat{R}(M^\top+\kappa\mathbb{1})^{-1}+\left(\int_0^\infty ds\,e^{Ms}\hat{R}e^{M^\top s}\right)\left((M^\top+\kappa\mathbb{1})^{-1}-(M^\top-\kappa\mathbb{1})^{-1}\right) \\ &=(M-\kappa\mathbb{1})^{-1}\hat{R}(M^\top+\kappa\mathbb{1})^{-1}+E\left((M^\top+\kappa\mathbb{1})^{-1}-(M^\top-\kappa\mathbb{1})^{-1}\right), \end{aligned} \tag{S10}$$

where the matrix-valued integral E solves the Lyapunov equation again, Eq. (S5). In order to continue, we require some expressions for the block-wise inverses of $M\pm\kappa\mathbb{1}$. Define $S_\pm=(\pm\kappa-\gamma)(\kappa(\kappa\mp\gamma)\mathbb{1}+L)^{-1}$, then

$$\begin{aligned} (M\pm\kappa\mathbb{1})^{-1}&=\begin{pmatrix}S_\pm & -\frac{S_\pm}{\pm\kappa-\gamma}\\ \frac{LS_\pm}{\pm\kappa-\gamma} & \frac{\mathbb{1}}{\pm\kappa-\gamma}-\frac{LS_\pm}{(\pm\kappa-\gamma)^2}\end{pmatrix} \\ (M^\top\pm\kappa\mathbb{1})^{-1}&=(M\pm\kappa\mathbb{1})^{-T}. \end{aligned}$$

With this, the first term in Eq. (S10) is

$$(M-\kappa\mathbb{1})^{-1}\hat{R}(M^\top+\kappa\mathbb{1})^{-1}=\begin{pmatrix}-\frac{S_-QRQS_+}{(\kappa+\gamma)(\kappa-\gamma)} & *\\ * & *\end{pmatrix}=\begin{pmatrix}(\kappa(\kappa-\gamma)\mathbb{1}+L)^{-1}QRQ(\kappa(\kappa+\gamma)\mathbb{1}+L)^{-1} & *\\ * & *\end{pmatrix}, \tag{S11}$$

where we only computed the upper-left block because it contains the correlations of the fluctuations themselves. Next, we compute the products of E with the block inverses,

$$\begin{aligned}E(M^\top\pm\kappa\mathbb{1})^{-1}=\begin{pmatrix}A & B\\ B^\top & C\end{pmatrix}(M\pm\kappa\mathbb{1})^{-1}&=\begin{pmatrix}A & B\\ B^\top & C\end{pmatrix}\begin{pmatrix}S_\pm & \frac{LS_\pm}{\pm\kappa-\gamma}\\ -\frac{S_\pm}{\pm\kappa-\gamma} & \frac{\mathbb{1}}{\pm\kappa-\gamma}-\frac{S_\pm}{(\pm\kappa-\gamma)^2}\end{pmatrix}\\ &=\begin{pmatrix}AS_\pm-\frac{BS_\pm}{\pm\kappa-\gamma} & *\\ * & *\end{pmatrix},\end{aligned} \tag{S12}$$

where again we only computed the relevant parts and employed the decomposition of E from Eq. (S6). In order to obtain the total fluctuation variance, we need the traces over the upper-left blocks. For Eq. (S11), this is

$$\operatorname{tr}\left((\kappa(\kappa-\gamma)\mathbb{1}+L)^{-1}QRQ(\kappa(\kappa+\gamma)\mathbb{1}+L)^{-1}\right)=\operatorname{tr}\left(((L+\kappa^2\mathbb{1})^2-\kappa^2\gamma^2)^{-1}QRQ\right).$$

In order to compute the trace in Eq. (S12), we note that $[S_\pm,L]=0$ and multiply Eqns. (S8) and (S7) by $S_\pm$. Taking the trace then yields

$$\begin{aligned}\operatorname{tr}(AS_\pm)&=\frac{1}{2\gamma}\operatorname{tr}(L^\dagger QRQS_\pm)\\ \operatorname{tr}(BS_\pm)&=\frac{1}{2\gamma}\operatorname{tr}((LA-AL)S_\pm)=0.\end{aligned}$$

Finally, we obtain

$$\langle|\boldsymbol{\varepsilon}(t)|^2\rangle=\frac{\kappa}{2}\left(\operatorname{tr}\left(((L+\kappa^2\mathbb{1})^2-\kappa^2\gamma^2)^{-1}QRQ\right)+\frac{\kappa-\gamma}{2\gamma}\operatorname{tr}(L^\dagger(\kappa(\kappa-\gamma)\mathbb{1}+L)^{-1}QRQ)+\frac{\kappa+\gamma}{2\gamma}\operatorname{tr}(L^\dagger(\kappa(\kappa+\gamma)\mathbb{1}+L)^{-1}QRQ)\right).$$

This expression can be further simplified by computing the trace in the eigenbasis $\{\phi_i\}$ of L,

$$\begin{aligned}\langle|\boldsymbol{\varepsilon}(t)|^2\rangle&=\frac{\kappa}{2}\sum_{i>0}\phi_i^\top QRQ\phi_i\frac{\omega_i^2+\frac{\kappa-\gamma}{2\gamma}(\kappa(\kappa+\gamma)+\omega_i^2)+\frac{\kappa+\gamma}{2\gamma}(\kappa(\kappa-\gamma)+\omega_i^2)}{\omega_i^2(\kappa(\kappa-\gamma)+\omega_i^2)(\kappa(\kappa+\gamma)+\omega_i^2)}\\ &=\frac{\kappa(\kappa+\gamma)}{2\gamma}\sum_{i>0}\phi_i^\top R\phi_i\frac{\omega_i^2+\kappa(\kappa-\gamma)}{\omega_i^2(\kappa(\kappa-\gamma)+\omega_i^2)(\kappa(\kappa+\gamma)+\omega_i^2)}\\ &=\frac{\kappa(\kappa+\gamma)}{2\gamma}\operatorname{tr}\left(L^\dagger(L+\kappa(\kappa+\gamma)\mathbb{1})^{-1}R\right)\\ &=\frac{1}{2\gamma}\operatorname{tr}\left(L^\dagger\left[L\frac{1}{\kappa(\kappa+\gamma)}+\mathbb{1}\right]^{-1}R\right),\end{aligned} \tag{S13}$$

which corresponds to Eq. (4) in the main paper using $\kappa=1/\tau$.

C. The Kuramoto model

The linearized Kuramoto model is described by

$$\dot{\boldsymbol{\varepsilon}}+L\boldsymbol{\varepsilon}=\boldsymbol{\xi}.$$

Thus, the calculation from the preceding section still works upon replacing M by $-L$, and without decomposing into blocks. The fluctuation variance is simply $\langle|\boldsymbol{\varepsilon}(t)|^2\rangle=\operatorname{tr}\left(\langle\boldsymbol{\varepsilon}(t)\boldsymbol{\varepsilon}(t)^\top\rangle\right)$.

In the white noise case we obtain

$$\langle|\boldsymbol{\varepsilon}(t)|^2\rangle=\frac{1}{2}\operatorname{tr}\left(L^\dagger R\right),$$

and in the case of Ornstein-Uhlenbeck colored noise similarly,

$$\langle|\boldsymbol{\varepsilon}(t)|^2\rangle=\frac{\kappa}{2}\operatorname{tr}\left(L^\dagger(L+\kappa\mathbb{1})^{-1}R\right).$$

Thus, formally the results for the Kuramoto model and the swing equation are related by a replacement of variables $\gamma\to1,\kappa(\kappa+\gamma)\to\kappa$ because objective functions that differ only by a constant pre-factor have the same minimizers.

III. NUMERICAL OPTIMIZATION

Here we describe our optimization algorithm for the case of white noise. The Ornstein-Uhlenbeck case is similar, with a different objective function.

A. Cost-constrained optimization close to synchrony

For simplicity, let us consider the case where there are no steady state flows, $\bar{P}_i = 0$, $\bar{\delta}_i \equiv const.$

We choose to optimize for fixed cost,

$$N_e C = \sum_e B_e^\alpha,$$

where α is a parameter that can be tuned and that controls the economy of scale for the couplings. The Lagrangian is

$$\mathcal{L} = \mathrm{tr}(L^\dagger R) + \lambda \left(\sum_e B_e^\alpha - N_e C \right).$$

Taking partial derivatives and setting them to zero yields

$$\begin{aligned} \lambda \alpha B_e^{\alpha-1} &= e^\top E^\top L^\dagger R L^\dagger E e \\ \Rightarrow B_e &= c \, (B_e^2 e^\top E^\top L^\dagger R L^\dagger E e)^{\frac{1}{1+\alpha}}, \end{aligned}$$

for a constant c.

For $\alpha < 1$, the landscape is non-convex and many local minima exist. For $\alpha > 1$, the landscape is convex, and one finds a unique global minimum.

B. Cost-constrained optimization with nonzero steady-state flow

The Lagrangian is again

$$\mathcal{L} = \mathrm{tr}(L^\dagger R) - \lambda \left(\sum_e B_e^\alpha - N_e C \right).$$

The Laplacian weights are $B_e \cos(\Delta\bar{\delta}_e)$ where the difference $\Delta\bar{\delta}_e = \bar{\delta}_i - \bar{\delta}_j$ for the edge $e = (ij)$. We take the derivative with respect to the couplings and set to zero,

$$\lambda \alpha B_e^{\alpha-1} = \cos(\Delta\bar{\delta}_e) e^\top E^\top L^\dagger R L^\dagger E e - \sum_f B_f \sin(\Delta\delta_f) f^\top E^\top L^\dagger R L^\dagger E f \frac{\partial \Delta\bar{\delta}_f}{\partial B_e}. \tag{S14}$$

The derivatives of the steady state angle differences can be computed by taking derivatives of the steady state condition,

$$\begin{aligned} 0 &= \frac{\partial}{\partial B_e} \sum_j B_{ij} \sin(\bar{\delta}_i - \bar{\delta}_j) \\ \Rightarrow \frac{\partial \Delta\bar{\delta}_f}{\partial B_e} &= -f^\top E^\top L^\dagger E e \sin(\Delta\bar{\delta}_e) = -S_{ef} \sin(\Delta\bar{\delta}_e), \end{aligned}$$

where we defined the symmetric matrix $S_{ef} = e^\top E^\top L^\dagger E f$. Plugging into Eq. (S14), we obtain

$$\lambda \alpha B_e^{\alpha-1} = \cos(\Delta\bar{\delta}_e) \langle \Delta\varepsilon_e^2 \rangle + \sin(\Delta\bar{\delta}_e) \sum_f S_{ef} B_f \langle \Delta\varepsilon_f^2 \rangle_R \sin(\Delta\bar{\delta}_f). \tag{S15}$$

Here, we introduced the shorthand $\langle \Delta\varepsilon_e^2 \rangle_R = e^\top E^\top L^\dagger R L^\dagger E e$ for the average squared linearized angle difference along an edge under the correlation matrix R.

For Ornstein-Uhlenbeck correlations, a similar but more unwieldy expression holds.

C. Algorithm for cost-constrained optimization

In order to solve Eq. (S15), we use the following algorithm, based on Ref. [39].

1. Start with initial couplings $B_e^{(0)}$
2. Run a few steps of a nonlinear root finder (trust-region algorithm as implemented in the package `NLsolve.jl`, `https://github.com/JuliaNLSolvers/NLsolve.jl`.) to obtain the steady state angles $\delta_{ss}^{(0)}$
3. Compute

$$\hat{B}_e^{(i+1)} = \left((B_e^{(i)})^2 \cos(\Delta\bar{\delta}_e^{(i)}) \langle \Delta(\varepsilon_e^2)^{(i)} \rangle + (B_e^{(i)})^2 \sin(\Delta\bar{\delta}_e^{(i)}) \sum_f S_{ef}^{(i)} B_f^{(i)} \langle (\Delta\varepsilon_f^2)^{(i)} \rangle_R \sin(\Delta\bar{\delta}_f^{(i)}) \right)^{\frac{1}{1+\alpha}}$$

4. Normalize

$$B_e^{(i+1)} = C^{1/\alpha} \frac{\hat{B}_e^{(i+1)}}{\left(\sum_f (\hat{B}_f^{(i+1)})^\alpha \right)^{1/\alpha}}$$

5. Run another few iterations of a nonlinear root finder to obtain $\bar{\delta}^{(i+1)}$
6. Repeat from 3 until convergence of both the steady state angles and the B_e.

Sometimes the RHS in step 3 becomes negative in an intermediate step. In that case we set it to zero hoping to converge to a good solution later.

The number of variables in the cost-constrained optimization is given by the number of nodes N in the network, for which the non-linear root finder in step 2 and 5 solves, and by the number of edges E which are obtained by the fixed-point iteration in steps 3 and 4. In regular graphs such as the ones we consider, each node is connected by the same number of edges (except at the boundaries), such that $E = \mathcal{O}(N)$. For a regular network constructed from rows containing n nodes each in d spatial dimensions, the number of variables thus scales as $\mathcal{O}(n^d)$.

IV. IMPROVEMENT DUE TO OPTIMIZATION

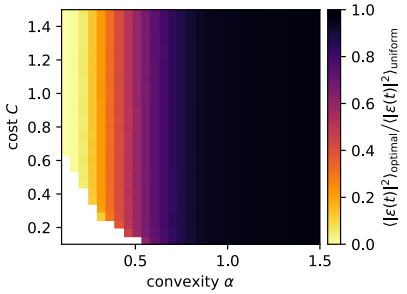

FIG. S1. Improvement in noise canceling due to optimization. We compute the ratio of the optimal network objective $\langle|\boldsymbol{\varepsilon}(t)|^2\rangle_{\text{optimal}}$ to the objective $\langle|\boldsymbol{\varepsilon}(t)|^2\rangle_{\text{uniform}}$, which is computed for uniform networks, $B_{ij} = const.$ Each pixel is an average over ratios computed for 25 optimal networks with different, uniformly random initial conditions. While the improvement is insignificant for convex, dense networks (which are almost uniform anyways), sparse networks with $0 < \alpha < 1$ provide a significant advantage.

V. TIME SERIES OF THE NON-LINEAR SWING EQUATION DYNAMICS

We solve the non-linear swing equation with stochastic feed-in as a system of coupled stochastic differential equations. In SDE form the white noise case reads,

$$\begin{aligned} d\delta_i &= \nu_i dt \\ d\nu_i &= -\gamma\nu_i dt + \sum_{j=1}^{N} B_{ij} \sin(\delta_i - \delta_j) dt + \bar{P}_i^c dt + \sum_{j=1}^{N} C_{ij} dW_j, \end{aligned}$$

where the dW_j are i.i.d. Wiener processes and $C = U\sqrt{\Sigma}$ is constructed from the singular value decomposition of the correlation matrix, $R = U\Sigma U^\top$. With this definition, the feed-ins have the desired correlation matrix $\langle C\frac{d\boldsymbol{W}(t)}{dt}\frac{d\boldsymbol{W}(t')^\top}{dt}C^\top\rangle = R\,\delta(t-t')$.

In the Ornstein-Uhlenbeck case the system of SDEs is augmented to

$$\begin{aligned} dX_i &= -\kappa X_i dt + \sqrt{2\kappa}dW_i \\ d\delta_i &= \nu_i dt \\ d\nu_i &= -\gamma\nu_i dt + \sum_{j=1}^{N} B_{ij}\sin(\delta_i-\delta_j)dt + \bar{P}_i^c dt + \sum_{j=1}^{N} C_{ij}X_j dt, \end{aligned} \tag{S16}$$

where again the dW_j are i.i.d. Wiener processes and the matrix C is defined as before. We obtain the desired feed-in correlations, $\langle C\boldsymbol{X}(t)\boldsymbol{X}(t')^\top C^\top\rangle = R\,e^{-\kappa|t-t'|}$. We employ the Julia language's `DifferentialEquations.jl` package to solve the SDEs using the Euler-Maruyama method. Eq. (S16) is replaced by the package's time step-independent distributionally correct Ornstein-Uhlenbeck process.

We then define the instantaneous mean angle by the integral

$$\bar{\delta}_i(t) = \frac{1}{t}\int_0^t \delta_i(t')dt',$$

where we evaluate the integral numerically from the simulation time series. From this, the numerical fluctuations and fluctuation variances are

$$\begin{aligned} \varepsilon_i(t) &= \delta_i(t) - \bar{\delta}_i(t) \\ \langle|\boldsymbol{\varepsilon}(t)|^2\rangle &= \frac{1}{t}\int_0^t \left|\boldsymbol{\delta}(t) - \bar{\boldsymbol{\delta}}(t)\right|^2 dt'. \end{aligned}$$

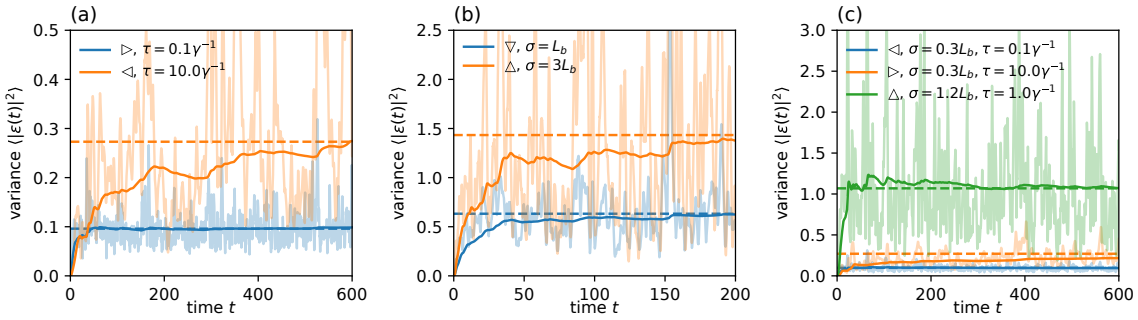

FIG. S2. Time series of the non-linear swing equation dynamics for the same networks as in the main paper, Fig. 2. (a) Ornstein-Uhlenbeck colored noise and spatially incoherent feed-in. (b) Gaussian spatially coherent feed-in with temporal white noise. (c) Spatio-temporally correlated feed-in.

In addition to the time series for white noise and spatially incoherent noise shown in the main paper, Fig. 1, here we also show time series for the optimal networks from Figs. 2 and 3 of the main paper (see Fig. S2). It can be seen clearly that higher correlations also lead to larger fluctuation variances.

A. Validity of the linear model

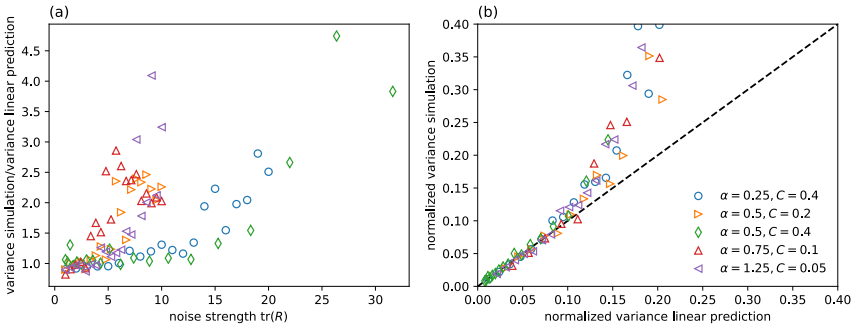

FIG. S3. Validity of the linear optimized model for white noise. (a) We compare optimal networks at various values of convexity α and cost C by plotting the ratio of the total fluctuation variance computed as a time average from fully non-linear simulations run until a time $t = 200$ and the prediction from the linear model. For each 8×8 triangular network, we vary the total noise strength $\mathrm{tr}(R)$, where R is spatially incoherent. The linear approximation is good for small noise strengths. (b) Normalizing the variance by the worst case variance, $N\pi^2/3$, computed by assuming uniformly distributed fluctuations ε_i on $[-\pi, \pi)$. The linear prediction is adequate until the predicted variance reaches approximately 10% of the worst case variance.

In order to test the validity of the linear model, we perform simulations of the fully nonlinear swing equation in the white noise, spatially incoherent case for various values of C and α. For each combination of parameters, we scale the total noise variance $\mathrm{tr}(\hat{R})$ until the linear prediction and simulations begin to disagree (see Fig. S3 (a)). By rescaling the total noise variances $\langle|\boldsymbol{\varepsilon}(t)|^2\rangle \to \langle|\boldsymbol{\varepsilon}(t)|^2\rangle/(N\pi^2/3)$, where $N\pi^2/3$ is the worst case variance, we see that the linear model is accurate up to $\approx 10\%$ of the worst case variance (see Fig. S3 (b)).

VI. DEPENDENCE OF ORNSTEIN-UHLENBECK TOPOLOGY ON COST

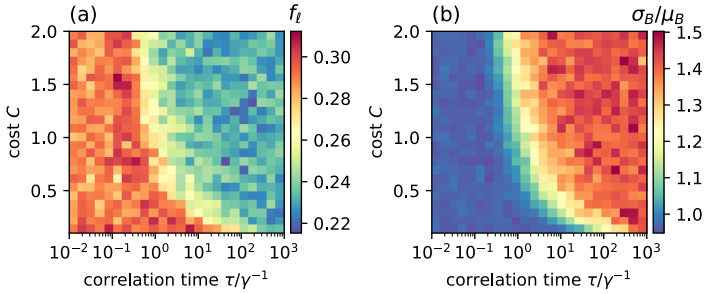

FIG. S4. Dependence of the topology of Ornstein-Uhlenbeck optimized networks on the cost parameter. We show phase space of the loop density and coupling spread for 8×8 triangular networks at fixed $\alpha = 0.5$ and for spatially incoherent noise. Very low cost networks with $C \ll 1$ (which have small couplings) stay uniform and dense even at longer correlation times τ. For higher cost networks with $C > 1$ with larger couplings the transition to sparsity and hierarchical organization occurs for smaller τ.

Unlike for the white noise case, the Ornstein-Uhlenbeck noise variance Eq. (S13) is not homogeneous upon rescaling the cost parameter $C \to sC$, even in the well-synchronized limit $\bar{\delta}_i \approx 0$. Therefore, unlike for white noise, the optimal networks depend on C. Fig. S4 shows the phase space of optimal networks as a function of cost and correlation time. We see that while for small C the transition between topologies shifts towards larger s, the topologies themselves remain unchanged (as quantified by f_ℓ and σ_B/μ_B).

VII. PHASE SPACES AND OPTIMAL NETWORKS FOR SQUARE GRIDS

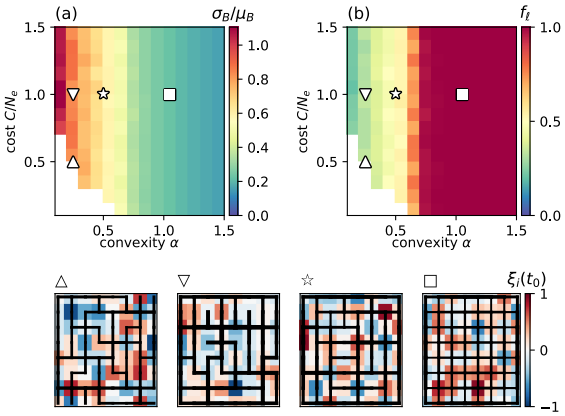

FIG. S5. Topology phase space and optimal networks for 8×8 square grids for white noise in time and incoherent spatial feed-in. Each pixel in the 15×15 phase space is an average over 5 optimal networks.

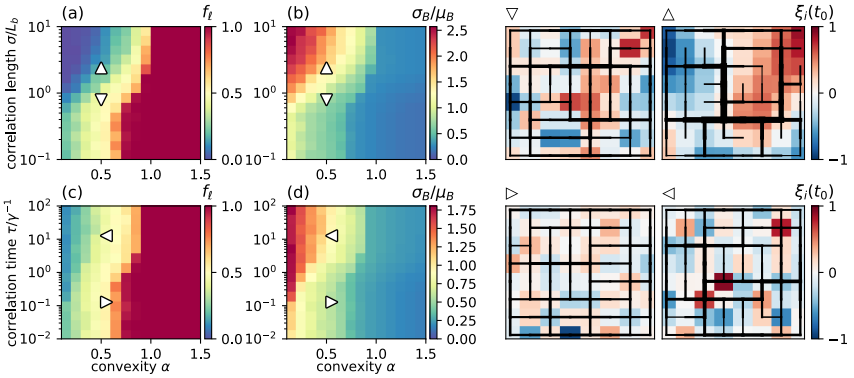

FIG. S6. Topology phase space and optimal networks for 8×8 square grids in the well-synchronized limit at $C = 1$. (a,b) White noise in time and Gaussian correlated spatial feed-in. (c,d) Ornstein-Uhlenbeck noise in time and spatially incoherent feed-in. Each pixel in the 15×15 phase space is an average over 5 optimal networks.

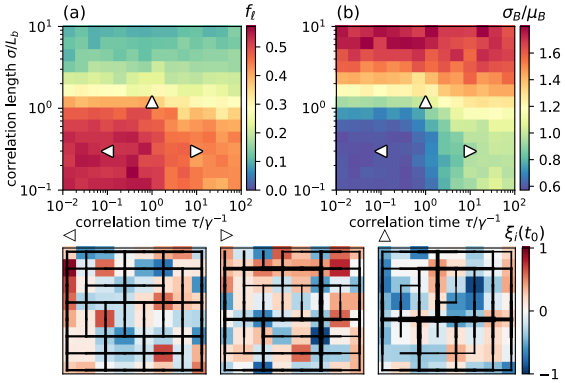

FIG. S7. Topology phase space and optimal networks for 8×8 square grids with spatio-temporal correlated feed-in in the well-synchronized limit at $C = 1$, $\alpha = 0.5$. Each pixel in the 15×15 phase space is an average over 5 optimal networks.