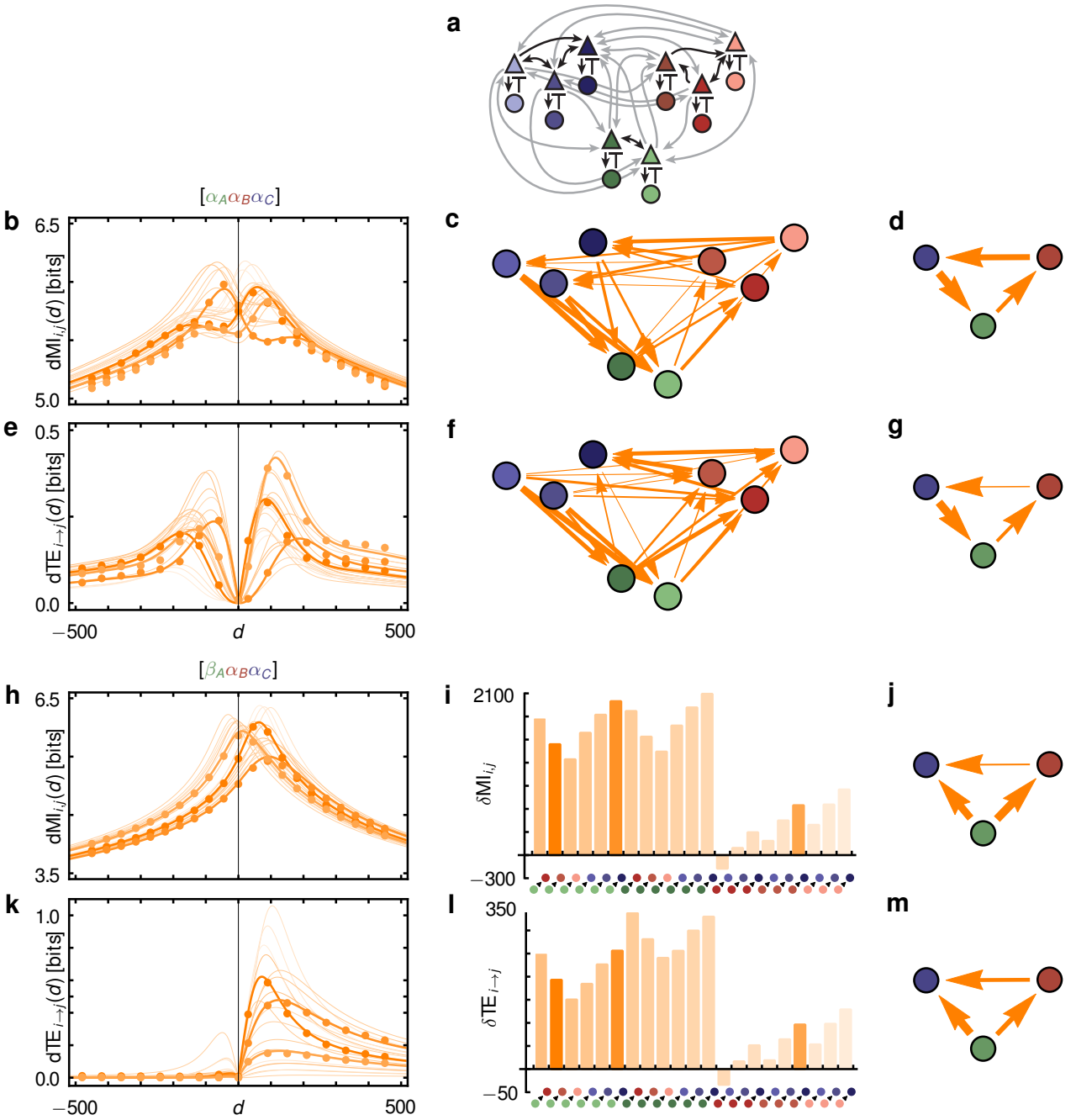
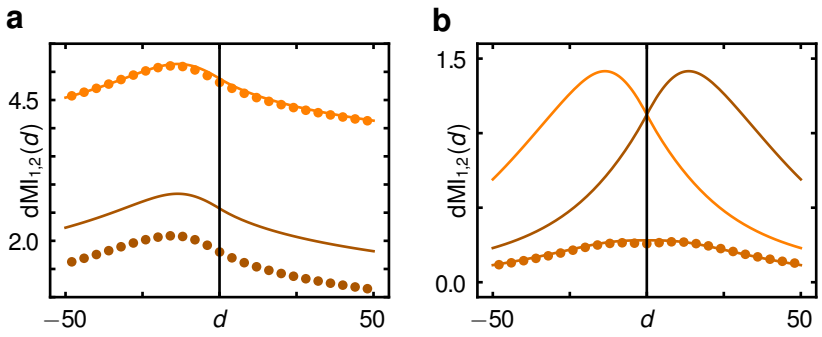
Supplementary Figure 1



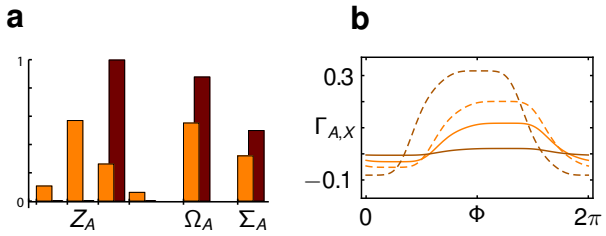
Delayed mutual Information and Transfer Entropy. a, Hierarchical network of Wilson-Cowan oscillators as in Figs. 1 and 4 of the main manuscript and Supplementary Note 8.1. **b-g,** Information sharing and flow networks for the network in a dynamical state determined by the local dynamical clusters states $\mathcal{D}_X = \alpha_X$, $X \in \{A, B, C\}$. **b,** Delayed mutual information curves dMI_{i_X,j_Y}. Shown are theoretical curves (Supplementary Note 4, Theorem 1) for all inter-cluster oscillator pairs. For three curves (bold) estimates from numerical simulations are shown (dots). **c,** effective information sharing network with link strength proportional to $\delta\mathrm{MI}_{ij}$ obtained from the $\mathrm{dMI}_{i,j}$ curves (Supplementary Note 1, (3)). **d,** Effective information sharing network with links proportional to $\delta\mathrm{MI}_{X,Y}$ obtained form the theoretical prediction (Supplementary Note 4, Corollary 3) of the hierarchically reduced system. **e-g,** as in (b-d) but for the delayed transfer entropy (Supplementary Note 4, Theorem 2 and Corollary 4). **h-m,** as in (b-g) but in a different dynamical state obtained by switching the local dynamical state in cluster A, $\mathcal{D}_A = \beta_A$. Effective information sharing and routing measures represented as bar graphs.

Supplementary Figure 2



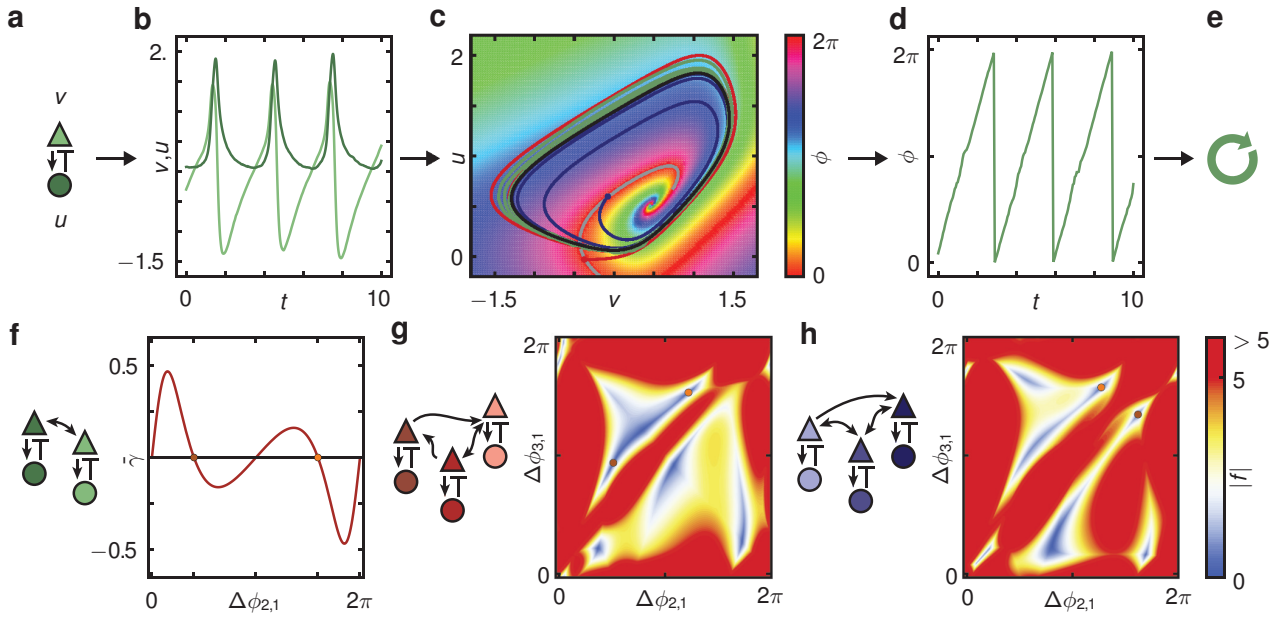
Delayed mutual information and large noise levels. **a,** Delayed mutual information for the Wilson-Cowan two-oscillator network as in Fig. 2 of the main manuscript with noise levels $\xi_{e,i} = 0.001$ and $\xi_{i,i} = 0.001$ (orange dots) and $\xi_{e,i} = 0.01$ and $\xi_{i,i} = 0.01$ (brown dots). Theoretical predictions agree well for the small noise (orange line) while systematically overestimate the delayed mutual information in the large noise regime (brown line). Directionality of the information flow determined by the shift of the peak dMI is still correctly predicted. **b,** Same as in (a) but for very large noise levels $\xi_{e,i} = 0.02$ and $\xi_{i,i} = 0.02$. No clear directionality is visible as the system fluctuates around both stable deterministic states. The actual measured dMI (beige dots) may be approximated as a weighted average (beige line, $\frac{1}{8}\left(\mathrm{MI}_{1,2}^{(\alpha)}(\mathrm{d}) + \mathrm{MI}_{1,2}^{(\beta)}(\mathrm{d})\right)$) of the dMI curves of the two stable dynamical states (orange and brown lines).

Supplementary Figure 3



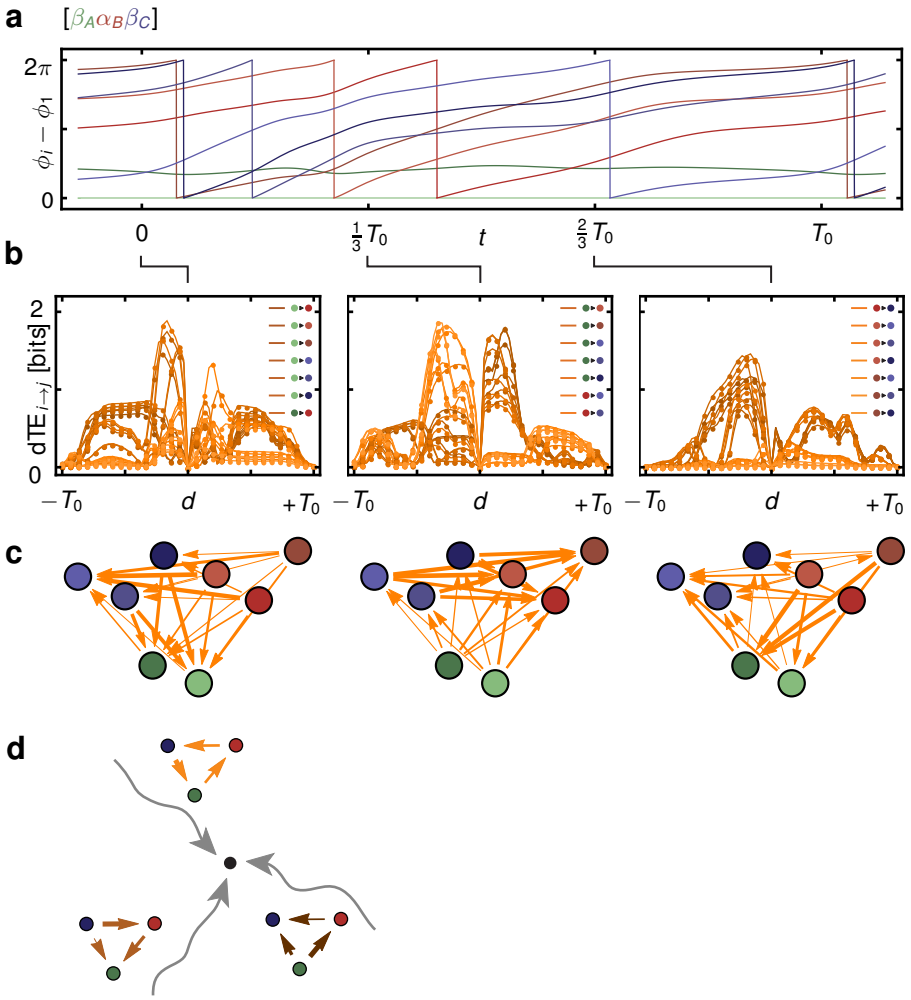
Multiple effects underlie the non-local change of information routing patterns in hierarchical networks. For the hierarchical Kuramoto phase oscillator network in Fig. 3d-f in the main manuscript the non-local effects on the information flow patterns due to changes of the weight of the local link $a_{2,3}$ arise because of combined changes of **a,** the collective cluster phase response vector $\boldsymbol{Z}_A$, the collective oscillation frequency Ω_A, effective noise levels Σ_A and **b,** the effective inter-community couplings $\Gamma_{A,B}$ (solid) and $\Gamma_{A,C}$ (dashed). In both panels orange and brown colors denote the dynamical states α and β, respectively. Generally, in phase oscillator networks the combination of these three effects contribute to the change in the information routing pattern (cf. Supplementary Note 4).

Supplementary Figure 4



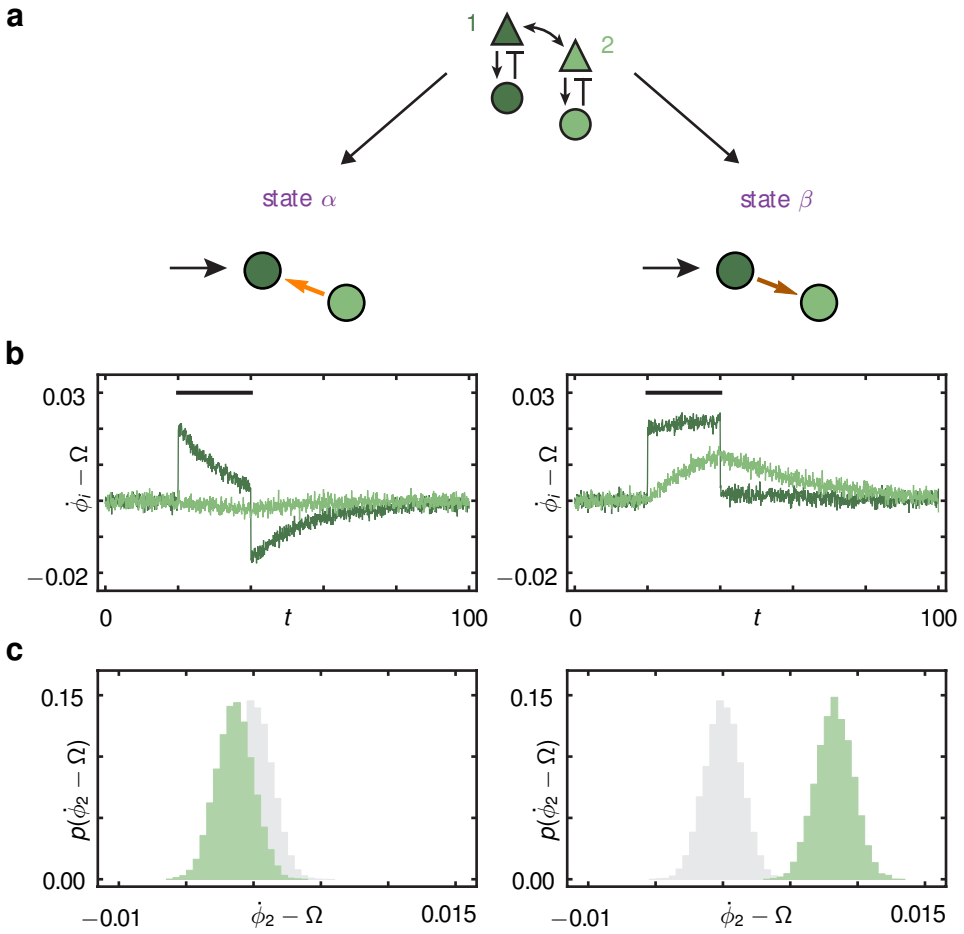
Phase reduction and multi-stability in Wilson-Cowan neuronal oscillators. a, The Wilson-Cowan neuronal oscillator [1, 2] is a mean-field description of the neuronal activity of mutually coupled excitatory (triangle) and inhibitory (square) populations of neurons with average effective membrane potential v and u, respectively (Supplementary Note 8.1). **b,** Sample stochastic trajectories of $v\,(t)$ (light green) and $u\,(t)$ (dark green) with oscillatory dynamics. **c,** Phase plane dynamics and phase field. Phase portrait shows the stochastic trajectory of (b) (green) together with the deterministic limit cycle (black). Colored background indicates the phase field $\phi\,(v,u)$ determined using the method described in Supplementary Section 8.2.2. Starting the deterministic dynamics on an iso-phase line (gray) the trajectories will converge towards the same phase on the limit cycle. Two examples are shown in blue and red. **d,** Phase $\phi\,(v(t),u(t))$ signal obtained from mapping the stochastic trajectory in (b) onto the phase field. **e,** The result is a noisy phase oscillator. **f,** Left: Sub-network A of two coupled identical Wilson-Cowan oscillators isolated from the hierarchical network in Fig. 1b in the main manuscript. Right: anti-symmetric part $\bar{\gamma}\,(\Delta\phi_{2,1}) = \gamma\,(\phi_2-\phi_1)-\gamma\,(\phi_1-\phi_2)$ of the coupling function obtained via (23) in the phase reduction process (cf. 3.2). The two zeros with negative slopes give rise to two stable phase locked states α_A (orange dot) and β_A (brown dot), i.e. multi-stability. **g,** Isolated sub-network B of Fig. 1b in the main manuscript (left) and norm $|\boldsymbol{f}|$ of the reduced vector field $\boldsymbol{f}$ for the phase differences. Zeros with non-positive eigenvalues of the Jacobian matrix give rise to stable phase-locked states α_B (orange dot) and β_B (brown dot). **h,** same as in (g) for the sub-network C in Fig. 1b in the main manuscript. Again the network only has two stable phase-locked states α_C (orange dot) and β_C (brown dot).

Supplementary Figure 5



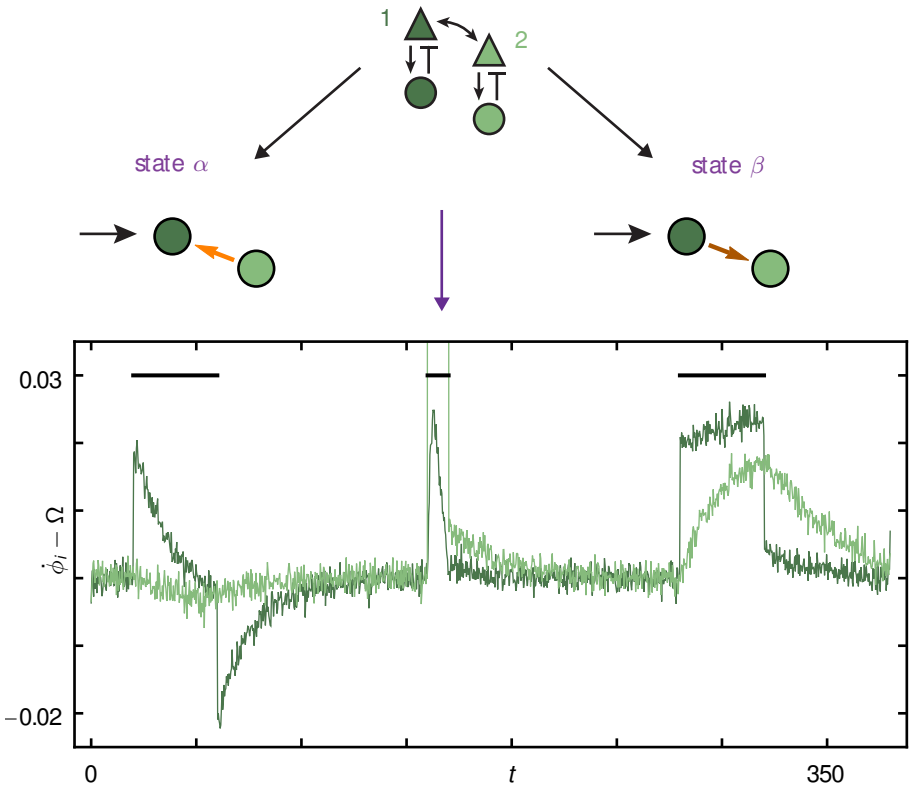
Time dependent information routing patterns in periodic dynamical states and transients. **a,** Deterministic periodic dynamics of the Wilson-Cowan network in Fig. 4 of the main manuscript with period T_0. The combination of local dynamical states is $\mathcal{D} = [\beta_A \alpha_B \beta_C]$. Extracted phases from the full dynamics are shown. **b,** Delayed transfer entropies $\mathrm{dTE}_{i\to j}\,(d,t)$ between all inter-cluster pairs of nodes in the network. Dots show a subset of numerically calculated values at different delays, lines are linear interpolations between all calculated points. Delayed transfer entropies are calculated after a transient time $t = 2T_0$ starting from initial conditions given by the deterministic periodic orbit at starting times $t_0 = 0, \frac{1}{3}T_0, \frac{2}{3}T_0$ respectively (left to right). Negative delays are used to indicate the reverse flow, i.e. $\mathrm{dTE}_{i\to j}\,(-d,t) = \mathrm{dTE}_{j\to i}\,(d,t)$. **c,** Information routing patterns for the three different time points inferred from (b). Arrow width represents the absolute value of $\delta\mathrm{TE}_{\mathrm{i,j}}$, (7), with the arrow pointing towards the dominant information routing direction. **d,** Information routing control via transients (schematic). Different displacements of the dynamical state of a three node network away from a fixed point (black point) induce different transients (gray arrows) along which different information patterns arise (network insets).

Supplementary Figure 6



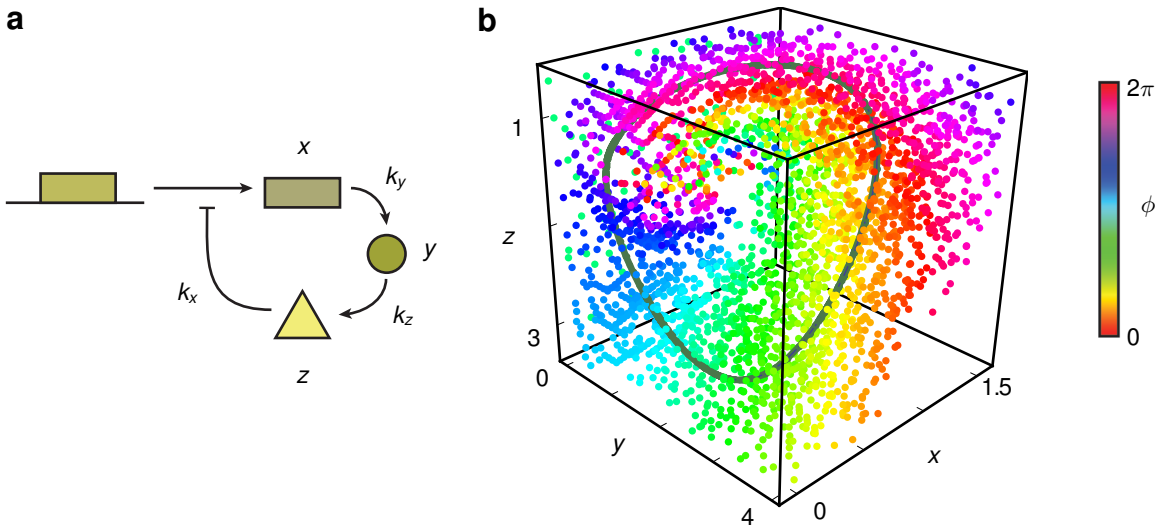
Dynamic signal transmission via dynamical reference state switching. a, Effective connectivity networks for the Wilson-Cowan network in Fig. 2 of the main manuscript for dynamical states α (left) and β (right). An additional input at times $20 \leq t \leq 40$ is applied to oscillator 1 (black arrow) that increases its intrinsic oscillation frequency ω_1. **b,** Instantaneous rotation frequencies, $\dot{\phi}_i$, shifted by the collective oscillation frequency Ω for both dynamical states. Black bars indicate stimulus to oscillator 1. **c** Probability distribution of $\dot{\phi}_2 - \Omega$, estimated from 10000 samples at $t = 10 \approx d^*$ after stimulus onset. For state α (light green) it is almost indistinguishable from the input free stationary distribution (gray), while around the state β the distributions separate in accordance with the information routing patterns in a and Fig. 2 of the main manuscript.

Supplementary Figure 7



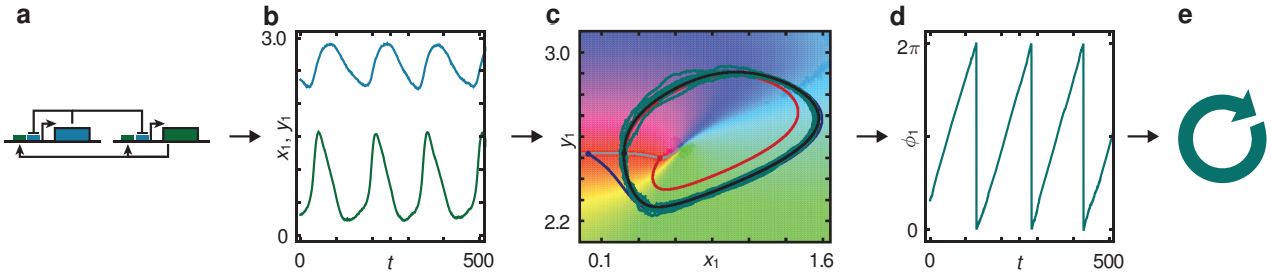
Auto-regulated signal transmission via dynamical reference state switching. Network as in Fig. 2 of the main manuscript and Supplementary Fig. 6. The inputs (black lines) consists of two signals into oscillator 1 delivered at times $20 \leq t \leq 60$ and $280 \leq t \leq 320$ in addition to a strong pulse at $160 \leq t \leq 170$ (purple arrow). Rotation frequencies $\dot{\phi}_i$ shifted by the collective oscillation frequency Ω for both dynamical states show state specific responses. While in state α the input is not seen by oscillator 2 the strong second pulse switched the system and made the third pulse visible at oscillator 2. The system auto-regulates its IRP by responding to specific characteristic input signals (here the middle strong pulse). Generally, if such strong pulses are not directly part of the input signal they can be provided by a second network that provides the pulse to trigger a switch in the information routing pattern whenever necessary for the computation.

Supplementary Figure 8



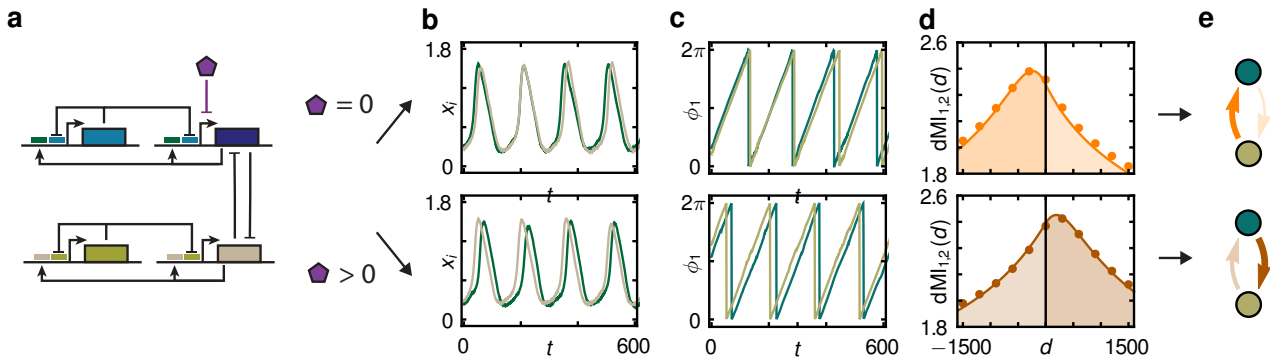
Goodwin model for oscillatory gene regulation. a, A single oscillator consists of a gene (underlined rectangle) that is transcribed into mRNA (rectangle) that has concentration x within the cell. With a rate k_y it is translated into an enzyme (disk) with concentration y that further facilitates the production of a protein (triangle) with concentration z. The protein suppresses the transcription of x. In total this results in a negative nonlinear feedback loop that generates oscillatory dynamics (cf. (Supplementary Note 8).2). **b,** Phase space showing the deterministic stable limit cycle solution of the Goodwin oscillator (solid line) as well as the phase field $\phi(x, y, z)$ estimated numerically. Dots show a small subset from the raster of the numerically estimated phases in color code.

Supplementary Figure 9



Phase reduction of a synthetic gene oscillator. **a,** A synthetic gene regulatory network as described in Supplementary Note 8.2.2. **b,** The network produces stochastic oscillations. **c,** Phase portrait showing the stochastic trajectory of (b) (turquoise) together with the deterministic limit cycle (black). Colored background indicates the phase field ϕ. Starting the deterministic dynamics on an iso-phase line (gray) the trajectories will converge towards the same phase on the limit cycle. Two examples are shown in blue and red. **d,** Phase signal obtained from mapping the stochastic trajectory on the phase field. **e,** The result is a noisy phase oscillator.

Supplementary Figure 10



Flexible information routing in a synthetic gene regulatory network. **a,** Two fast, robust genetic oscillators (blue, beige) coupled via mutual repression (gray) as described in Supplementary Note 8.2.2. **b,** Stochastic trajectories $x_i\,(t)$ in the absence (top) and presence (bottom) of a protein that represses translation (purple in (a)). **c,** phase signal determined using the method described in Supplementary Note 8.2.2 and Supplementary Fig. 9 for both traces in (b). **d,** delayed mutual information $\mathrm{dMI}_{1,2}\,(d)$ between the phase signals of the two oscillators. Dots: numerical estimation. Lines: theoretical prediction as described in the main manuscript and derived in detail in Supplementary Note 4. The peak in the dMI curve switches from left to right indicating an effective reversal of information flow between the two oscillators. **e,** Effective information flow networks show the reversal in information flow directionality. Arrow thickness corresponds to the area $\mathrm{MI}_{i\to j}$ (2) under the dMI curve for positive and negative delays respectively (shaded areas in (d)).

Supplementary Note 1
Delayed Mutual Information and Transfer Entropy

Our study aims at identifying mechanisms for flexible information routing in network dynamical systems. We here first focus on calculating the delayed mutual information (dMI) [3, 4] as a measure of how information is shared between different oscillators in the network. For a network with stochastic time dependent activity $x_i(t)$ at node $i \in \{1, \ldots, N\}$ the dMI is between the signal $x_{i,t} := x_i(t)$ at time t and a signal $x_{j,t+d} := x_j(t+d)$ shifted by a time delay d, is the Kulback-Leibler divergence between the joint probability distribution $p(x_{i,t}, x_{j,t+d})$ and the product of the marginals $p(x_{i,t})\, p(x_{j,t+d})$ if both signals where independent [5]:

$$\begin{aligned} \mathrm{dMI}_{i,j}(d) &= \int p(x_{i,t}, x_{j,t+d}) \log\left(\frac{p(x_{i,t}, x_{j.t+d})}{p(x_{i,t})\, p(x_{j,t+d})}\right) \mathrm{d}x_{i,t}\mathrm{d}x_{j,t+d} \quad . \\ &= H(x_{j,t+d}) - H(x_{j,t+d}|x_{i,t}) \end{aligned} \tag{1}$$

where $H(x_{j,t+d})$ is the entropy of $x_{j,t+d}$ and $H(x_{j,t+d}|x_{i,t})$ are the conditional entropy of $x_{j,t+d}$ conditioned on $x_{i,t}$.

Typically only the full shape of the delayed mutual information curve $\mathrm{dMI}_{i,j}(d)$ for all delays d is informative about an effective information flow direction. Indeed, consider a two dimensional system, where the first sub-systems physically drives the second one but not vice versa. Then due this drive $\mathrm{dMI}_{1,2}(d)$ will be positive for some positive delays $d > 0$. However, non-zero values for negative delays $d < 0$ are also possible if there are temporal correlations within the first signal. Thus, only the asymmetry around $d = 0$ of the full dMI curve is indicative of an information flow direction. To quantify this asymmetry we define the integrated delayed mutual information

$$\mathrm{MI}_{\mathrm{i}\to\mathrm{j}} = \int_0^\infty \mathrm{dMI}_{\mathrm{i,j}}(\delta)\, \mathrm{d}\delta \tag{2}$$

and use the difference

$$\delta\mathrm{MI}_{\mathrm{i,j}} = \mathrm{MI}_{i\to j} - \mathrm{MI}_{\mathrm{j}\to\mathrm{i}} \tag{3}$$

as an indicator for the information sharing directionality. If $\delta\mathrm{MI}_{\mathrm{i,j}} > 0$ there is more information shared from i to j than in the other direction, and vice versa. We represent this differences in graphs by a directed edge of weight $\Delta\mathrm{MI}_{i,j} = \delta\mathrm{MI}_{i\to j}$ if $\delta\mathrm{MI}_{\mathrm{i,j}} > 0$ and $\Delta\mathrm{MI}_{i,j} = 0$ otherwise. The resulting graph is termed information routing pattern (IRP).

In addition to measuring information sharing via the delayed mutual information we use the transfer entropy introduced in [6] to quantify information routing. By conditioning transition probabilities on past signals this measure is able to distinguish between shared information due to a common history or correlated common input signals from an actual transfer of information due to asymmetric driving [6].

The delayed transfer entropy (dTE) between signals $x_{i,t}$ and $x_{j,t+d}$ for a time delay d is defined by:

$$\begin{aligned} \mathrm{dTE}_{i\to j}(d) &= H(x_{j,t+d}|x_{j,t}) - H(x_{j,t+d}|x_{j,t}, x_{i,t}) \\ &= \iiint p(x_{i,t}, x_{j,t}, x_{j,t+d}) \log\left(\frac{p(x_{j,t+d}|x_{j,t}, x_{i,t})}{p(x_{j,t+d}|x_{j,t})}\right) dx_{i,t}dx_{j,t}dx_{j,t+d} \end{aligned} \tag{4}$$

which we can rewrite as

$$\mathrm{dTE}_{i\to j}(d) = \iiint p(x_{i,t}, x_{j,t}, x_{j,t+d}) \log\left(\frac{p(x_{i,t}, x_{j,t}, x_{j,t+d})\, p(x_{j,t})}{p(x_{j,t}, x_{j,t+d})\, p(x_{i,t}, x_{j,t})}\right) dx_{i,t}dx_{j,t}dx_{j,t+d} \tag{5}$$

Similarly to equations (2) and (3) we quantify asymmetries in the transfer entropy via

$$\mathrm{TE}_{\mathrm{i}\to\mathrm{j}} = \int_0^\infty \mathrm{dTE}_{\mathrm{i,j}}(\delta)\, \mathrm{d}\delta \tag{6}$$

and use the difference

$$\delta\mathrm{TE}_{\mathrm{i,j}} = \mathrm{TE}_{i\to j} - \mathrm{TE}_{\mathrm{j}\to\mathrm{i}} \tag{7}$$

as an indicator for the information flow directionality.

Supplementary Note 2
Flexible Information Routing on Top of Dynamical Reference States

To understand how bits of information from external or locally computed signals can be specifically distributed through the network or to it's downstream components on top of an underlying collective dynamical state we first consider a generic stochastic dynamical system that evolves in time t according to Equation (1) in the main manuscript, i.e.

$$\frac{d}{dt}\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varsigma}\boldsymbol{\xi} \tag{8}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ denotes the variables of the network nodes, $\boldsymbol{f}$ describes the intrinsic dynamics of the network, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_K)$ is a stochastic process representing the external inputs carrying information to be routed through the network, and $\boldsymbol{\varsigma}$ is a $K \times N$ matrix that couples the inputs to individual node's dynamics. We focus on a deterministic reference state

$$\boldsymbol{x}_t^{(\mathrm{ref})} = \boldsymbol{F}^{(\mathrm{ref})}(\boldsymbol{x}, t) \tag{9}$$

solving (8) in the absence of signals ($\xi = 0$). Here $\boldsymbol{F}^{(\mathrm{ref})}(\boldsymbol{x}, t)$ is the deterministic flow in time s starting at $\boldsymbol{x}$. For notational clarity, in the following we denote time dependence of variables in subscripts, i.e. $\boldsymbol{x}(t) = \boldsymbol{x}_t$.

To calculate the delayed mutual information or transfer entropy according to Equation (2) and (7) in the main manuscript we first seek to calculate the joint distribution $p(\boldsymbol{x}_{t+d}, \boldsymbol{x}_t)$ and rewrite it in terms of the marginal $p(\boldsymbol{x}_t)$ and transition probability $p(\boldsymbol{x}_{t+d}|\boldsymbol{x}_t)$ as

$$p(\boldsymbol{x}_{t+d}, \boldsymbol{x}_t) = p(\boldsymbol{x}_{t+d}|\boldsymbol{x}_t)\, p(\boldsymbol{x}_t) \tag{10}$$

Assuming that the stochastic signals $\boldsymbol{\varsigma}\boldsymbol{\xi}$ are small enough to keep the system fluctuating around the reference trajectory we use the small noise approximation to condition our calculations on this specific reference trajectory. For uncorrelated white noise signal sources $\boldsymbol{\xi}$, i.e. $\langle \xi_k(t)\, \xi_l(s) \rangle = \delta_{kl}\delta(t-s)$ we find [7]

$$p(\boldsymbol{x}_{t+d}|\boldsymbol{x}_t) = \mathcal{N}_{\boldsymbol{x}_{t+d}^{(\mathrm{ref})}, \boldsymbol{\Sigma}_d\left[\boldsymbol{x}^{(\mathrm{ref})}\right]}(\boldsymbol{x}_{t+d}) \tag{11}$$

where $\mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x})$ denotes a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

$$\mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{12}$$

and

$$\boldsymbol{\Sigma}_d\left[\boldsymbol{x}^{(\mathrm{ref})}\right] = \int_0^d e^{\int_s^d \boldsymbol{G}\left(\boldsymbol{x}^{(\mathrm{ref})}, r\right) dr} \boldsymbol{\varsigma}\boldsymbol{\varsigma}^{\mathrm{T}} e^{\int_s^d \boldsymbol{G}^{\mathrm{T}}\left(\boldsymbol{x}^{(\mathrm{ref})}, r\right) dr} ds \tag{13}$$

with

$$\boldsymbol{G}\left(\boldsymbol{x}^{(\mathrm{ref})}, r\right) = \boldsymbol{D}\boldsymbol{f}\left(\boldsymbol{x}^{(\mathrm{ref})}(t+r)\right) \quad . \tag{14}$$

where $\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x})$ is the Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{x}$ and we denoted the transpose of $\boldsymbol{x}$ by $\boldsymbol{x}^{\mathrm{T}}$. We also assumed

$$\boldsymbol{G}\left(\boldsymbol{x}^{(\mathrm{ref})}, r\right) \boldsymbol{G}\left(\boldsymbol{x}^{(\mathrm{ref})}, s\right) = \boldsymbol{G}\left(\boldsymbol{x}^{(\mathrm{ref})}, s\right) \boldsymbol{G}\left(\boldsymbol{x}^{(\mathrm{ref})}, r\right) \tag{15}$$

for an explicit solution to exists. In this way we obtain two information theoretic measures $\mathrm{dMI}_{i,j}\left(d, t, \boldsymbol{x}^{(\mathrm{ref})}\right)$ and $\mathrm{dTE}_{i\to j}\left(d, t, \boldsymbol{x}^{(\mathrm{ref})}\right)$ that in general depend on the lag time d, the starting distribution $p(\boldsymbol{x}_t)$ at time t, and, most importantly, on the reference trajectory $\boldsymbol{x}^{(\mathrm{ref})}$ through expression (11). We note that this dependence is still present in case (15) is not satisfied. We precisely exploit this dependence on the reference trajectory to achieve flexible information routing.

For a network with a single attractor $\mathcal{A}$ and a stationary distribution the information theoretic expressions become t independent but will still strongly depend on the underlying deterministic invariant dynamics $\boldsymbol{x}^{(\mathrm{ref})}$ in $\mathcal{A}$ and the time lag d. For systems with multiple deterministic attractors the white

noise is capable of inducing transitions between the attractors. For small noise levels in comparison to the signal strength needed to induce a transition the expected switching time to a different attractor $\mathcal{A}'$ becomes arbitrarily large and the system can exhibit a 'pseudo stationary distribution' for the fluctuation dynamics around one particular attractor. In this situation the information theoretic expressions again become independent of t but will still strongly depend on the underlying deterministic dynamics $\boldsymbol{F}$ in the vicinity of $\mathcal{A}$. Information encoded in the fluctuations around such an attracting dynamical state is thus shared and transferred differently depending on which of the different attractors the system was prepared or pushed to (cf. also Supplementary Note 4-7 below). The dependence of dMI and dTE on the starting time in systems with a non-stationary distribution can be further exploited to obtain time-dependent information routing patterns (cf. Supplementary Note 6).

Supplementary Note 3
Phase Description of Coupled Stochastic Oscillator Networks

3.1 Phase Reduction of Limit Cycle Oscillators

Oscillatory dynamics are naturally separated into an amplitude and a phase. If the attraction towards a limit cycle is sufficiently strong, deviations in the amplitudes decay fast and the dynamics may be reduced to a phase description [8].

Consider a dynamical system

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) \tag{16}$$

with $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{f}$ a sufficiently smooth n-dimensional vector field that has a stable limit cycle solution $\boldsymbol{x}_{\circlearrowleft}(t) = \boldsymbol{x}_{\circlearrowleft}(t+T)$ of period T. We may introduce a phase $\phi \in [0, 2\pi)$ on the limit cycle solution such that $\dot{\phi} = \omega$ where $\omega = 2\pi/T$. This phase description can be extended to the basin of attraction of the limit cycle by assigning to each point $\boldsymbol{x}_0$ a scalar phase $\phi(\boldsymbol{x}_0)$ such that $\lim_{t\to\infty} \phi(\boldsymbol{x}(t)) - (\omega t + \phi(\boldsymbol{x}_0)) = 0$ for the solution $\boldsymbol{x}(t)$ of (16) starting at $\boldsymbol{x}(0) = \boldsymbol{x}_0$. By definition

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\phi(\boldsymbol{x}(t)) &= \omega \\ &= \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}) \cdot \dot{\boldsymbol{x}} = \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}) \cdot f(\boldsymbol{x}) \end{aligned} \tag{17}$$

Introducing a small perturbation $\boldsymbol{r}$ of order ε into the vector field, i.e. substituting $\boldsymbol{f}(\boldsymbol{x})$ with $\boldsymbol{f}(\boldsymbol{x}) + \varepsilon \boldsymbol{r}(\boldsymbol{x}, t)$ one obtains

$$\begin{aligned} \frac{d}{dt}\phi &= \omega + \varepsilon \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}) \cdot \boldsymbol{r}(\boldsymbol{x}, t) \\ &= \omega + \varepsilon \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}_{\circlearrowleft}(\phi)) \cdot \boldsymbol{r}(\boldsymbol{x}_{\circlearrowleft}(\phi), t) + \mathcal{O}\left(\varepsilon^2\right) \end{aligned} \tag{18}$$

where in the second step the contributions up to first order in ε give a closed form expression for the phase dynamics. Here

$$\boldsymbol{z}(\phi) = \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x}_{\circlearrowleft}(\phi)) \tag{19}$$

is the phase response curve (PRC), an n-dimensional vector that determines the oscillators linear response in it's phase variable to brief perturbations in the n different coordinates applied at the oscillator's phase ϕ. The system's time evolution thus has been reduced to a one dimensional phase model. We note that this reduction process stays valid in the presence of white noise signals if the attraction towards the limit cycle is strong and the noise sources are idealizations of physical input signals [9] that have positive correlations on small time scales [10].

We use the adjoint method [8, 11, 12] to numerically determine the phase response curves $\boldsymbol{z}$. If no analytical expression for the phase field $\phi(\boldsymbol{x})$ was available, we determined it numerically. Therefore we integrated the dynamical system from a sufficiently long but fixed time with initial conditions $\boldsymbol{x}_0$ on a equally spaced raster around the limit cycle solution. From the final point we determined the closest point on the limit cycle and its phase ϕ to obtain an estimate of the phase $\phi(\boldsymbol{x}_0)$ on the grid. For arbitrary points $\boldsymbol{x}_0$ we then used multi-dimensional linear interpolation between the nearest grid points to obtain $\phi(\boldsymbol{x}_0)$ (cf. Supplementary Figs. 4, 8 and 9). All simulations where performed

using a self written dynamical systems software package combining the symbolic strength of Wolfram Mathematica 9.0 with the numerical speed of c++. For stochastic integration we used a second order stochastic Runge-Kutta algorithm [13].

3.2 Stochastic Phase Equations for Weakly Coupled Stochastic Oscillators

Consider a network of $N \in \mathbb{N}$ coupled oscillators evolving according to

$$\dot{\boldsymbol{x}}_i = \boldsymbol{f}_i(\boldsymbol{x}_i) + \sum_j \boldsymbol{g}_{i,j}(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_k \boldsymbol{h}_{i,k}(\boldsymbol{x}_i)\,\zeta_k \tag{20}$$

with $\boldsymbol{x}_i \in \mathbb{R}^{n_i}$, $i \in \{1,\ldots,N\}$, $\boldsymbol{f}_i$ smooth local dynamics, $\boldsymbol{g}_{i,j}$ smooth coupling functions, $\boldsymbol{h}_{i,k}$ n_i-dimensional vectors modeling the impact of random processes ζ_k with zero mean. We assume that each node i in the uncoupled deterministic case ($\boldsymbol{g}_{i,j} = 0$, $\boldsymbol{h}_{i,k} = 0$) has a strongly attracting limit cycle solution generated by the local dynamics $\boldsymbol{f}_i$. If in the full system the coupling and noise are sufficiently weak, the phase reduction method outlined in the previous section can be applied to each oscillator [12]. This results in a phase description of the full network (20) of the form

$$\dot{\phi}_i = \omega_i + \sum_j \boldsymbol{z}_i(\phi_i) \cdot \boldsymbol{g}_{i,j}(\phi_i, \phi_j) + \sum_k \boldsymbol{z}_i(\phi_i) \cdot \boldsymbol{h}_{i,k}(\phi_i)\,\zeta_k \quad . \tag{21}$$

As before, ϕ_i is the phase of oscillator i, $\boldsymbol{z}_i(\phi)$ its n_i-dimensional phase response curve and the constant oscillation frequency is given by $\omega_i = \boldsymbol{z}(\phi_i) \cdot \boldsymbol{f}(x_i(\phi_i))$.

In the deterministic case ($\boldsymbol{h}_{i,k} = 0$) it is convenient and common to exploit the weak coupling assumption further and average these equations [8, 11, 12]. However, we are not aware of a general stochastic analog that leads to a reduced but still random dynamical system. Here, we therefore assume that the system has a collective deterministic state in which it rotates with an average frequency $\Omega = \frac{2\pi}{T}$. If we transform to coordinates $\varphi_i = \phi_i - \Omega t$, the small coupling and noise ensures that the evolution of φ_i is slow compared to the collective oscillation. Its average change over one period may then be approximated as

$$\begin{aligned} \frac{d}{dt}\varphi_i &= \omega_i - \Omega + \sum_j \frac{1}{T}\int_0^T \boldsymbol{z}_i(\varphi_i + \Omega t) \cdot \boldsymbol{g}_{i,j}(\varphi_i + \Omega t, \varphi_j + \Omega t)\,\mathrm{d}t \\ &\quad + \sum_k \frac{1}{T}\int_0^T \boldsymbol{z}_i(\varphi_i + \Omega t) \cdot \boldsymbol{h}_{i,k}(\varphi_i + \Omega t)\,\zeta_k \mathrm{d}t \end{aligned} \tag{22}$$

As in the deterministic situation the first coupling terms only depends on the phase differences and may be written as $\sum_j \gamma_{ij}(\phi_i - \phi_j)$ where

$$\gamma_{ij}(\varphi_i - \varphi_j) = \gamma_{ij}(\phi_i - \phi_j) = \frac{1}{2\pi}\int_0^{2\pi} \boldsymbol{z}_i(\phi_i - \phi_j + \psi) \cdot \boldsymbol{g}_{i,j}(\phi_i - \phi_j + \psi, \psi)\,\mathrm{d}\psi \quad . \tag{23}$$

The second sum in (22) still consists of a superposition of random processes that depend on φ_i. We can simplify it further, if we assume that the oscillators receive uncorrelated noise sources, i.e $\boldsymbol{h}_{i,k}(\phi) = \delta_{ik}\boldsymbol{h}_i(\phi)$ that have correlations in time that are small with respect to the period T. The integrated noise then becomes a Gaussian process w_i with zero mean and phase-independent variances

$$\varsigma_i^2 = \frac{1}{2\pi}\int_0^{2\pi} \boldsymbol{z}_i(\psi)\,\boldsymbol{h}_i(\psi)\,\boldsymbol{h}_i^{\mathrm{T}}(\psi)\,\boldsymbol{z}_i^{\mathrm{T}}(\psi)\,\mathrm{d}\psi \tag{24}$$

We thus arrive at the averaged but still stochastic evolution equations

$$\frac{d}{dt}\phi_i = \omega_i + \sum_j \gamma_{ij}(\phi_i - \phi_j) + \varsigma_i \xi_i \tag{25}$$

with $\xi_i = \frac{dw_i}{dt}$. These stochastic phase equations [14] will serve as our starting point for the study of control mechanisms for information flows in complex network dynamical systems. A general method of stochastic averaging that results in reduced but still stochastic equations and a treatment for correlated inputs will be discussed in detail elsewhere.

Supplementary Note 4
Information Routing in Phase Oscillator Networks

Here we study general networks of coupled stochastic phase oscillators and give a detailed derivation of the theoretical results presented in the main text. In particular, for a network of coupled phase oscillators we derive expressions for the delayed mutual information (dMI) and transfer entropy (dTE) between two phase signals as a function of an underlying deterministic dynamical state and the network topology. We follow the general calculations outlined in Supplementary Note 2.

4.1 Delayed Mutual Information in Phase Oscillator Networks

4.1.1 Calculation of Delayed Mutual Information in Phase Oscillator Networks

Networks of Stochastic Phase Oscillators Consider a network of N phase oscillators $i \in \{1, \ldots, N\}$ with phases ϕ_i and intrinsic oscillation frequencies ω_i evolving according to the stochastic Kuramoto differential equations [8, 14] (cf. also Supplementary Note 3.2)

$$d\phi_i = \left(\omega_i + \sum_j \gamma_{ij} \left(\phi_i - \phi_j \right) \right) \mathrm{d}t + \sum \varsigma_{ik} \mathrm{d}w_k \tag{26}$$

with coupling functions $\gamma_{ij}\left(\phi_i - \phi_j\right)$ and external inputs that are modeled as independent Wiener processes w_k. We formally write $\xi_k = \frac{dw_k}{dt}$ such that $\left\langle \xi_k\left(t\right) \xi_l\left(s\right) \right\rangle = \delta_{kl} \delta\left(t - s\right)$ and the matrix $\boldsymbol{\varsigma} = \left(\varsigma_{ik}\right)$ in (26) describes the covariances of the inputs.

Calculation of the Probability Densities To calculate the delayed mutual information $\mathrm{MI}_{i,j}\left(d\right)$ between the phase signals $\phi_{i,t} := \phi_i\left(t\right)$ and $\phi_{j,t} := \phi_j\left(t\right)$ we derive the joint probability distribution $p\left(\phi_{i,t}, \phi_{j,t+d}\right)$ and then insert it into (1). To this end we first calculate the joint probability distribution $p\left(\boldsymbol{\phi}_t, \boldsymbol{\phi}_{t+d}\right)$ for the full phase vectors $\boldsymbol{\phi}_t = \left(\phi_{1,t}, \ldots, \phi_{N,t}\right)$. We use the identity $p\left(\boldsymbol{\phi}_t, \boldsymbol{\phi}_{t+d}\right) = p\left(\boldsymbol{\phi}_t\right) p\left(\boldsymbol{\phi}_{t+d} | \boldsymbol{\phi}_t\right)$ to split the calculation into the calculation of the transition probability $p\left(\boldsymbol{\phi}_{t+d} | \boldsymbol{\phi}_t\right)$ and the stationary probability density $p\left(\boldsymbol{\phi}_t\right)$.

A general analytical solution for the probability densities is not feasible such that we here focus on networks with a phase-locked state. Information routing via general dynamical states are considered in Supplementary Note 6.

Phase-Locked States We assume that the system (26) in the noiseless case (i.e. $\varsigma_{ik} = 0$) exhibits a stable phase-locked state with constant phase differences $\Delta\phi_{i,j}^{(\mathrm{ref})} = \phi_i - \phi_j$ and a collective oscillation frequency Ω, i.e. for all $i \in \{1, \ldots, N\}$ we have

$$\Omega = \omega_i + \sum_j \gamma_{ij} \left(\Delta\phi_{i,j}^{(0)} \right) \tag{27}$$

In particular, a solution to the deterministic dynamics is given by

$$\phi_i^{(\mathrm{ref})}\left(t\right) = \Omega t + \Delta\phi_{i,1}^{(\mathrm{ref})} \tag{28}$$

where w.l.o.g. we set $\phi_1\left(0\right) = 0$ due to the rotational symmetry of (26).

Calculation of the Transition Probability Density To calculate $p\left(\boldsymbol{\phi}_{t+d} | \boldsymbol{\phi}_t\right)$ it is convenient to first introduce new coordinates

$$\varphi_i = \phi_i - \phi_i^{(\mathrm{ref})} \tag{29}$$

such that (26) becomes

$$\mathrm{d}\boldsymbol{\varphi} = \boldsymbol{f}\left(\boldsymbol{\varphi}\right) \mathrm{d}t + \boldsymbol{\varsigma} \mathrm{d}\boldsymbol{w} \tag{30}$$

where $f_i\left(\boldsymbol{\varphi}\right) := \omega_i + \sum_j \gamma_{ij} \left(\varphi_i - \varphi_j + \Delta\phi_{i,j}^{(\mathrm{ref})} \right) - \Omega$. As before, this system is invariant under uniform phase rotations as it only depends on the phase differences $\varphi_i - \varphi_j$.

Assuming that the noise levels ς_{ik} are small these phase differences stay small close to the stable phase-locked state. Due to the nature of the white noise process strong deviations may occur that push the dynamics towards other stable dynamical states in the phase space of the system. The expected time for such a switching event becomes arbitrarily large when the noise amplitude is decreased. We are interested in the impact of a single dynamical state onto the information routing patterns within the network. We thus assume that the noise is small enough so that no intrinsic switching occurs on time scales relevant for the system to perform its communication functions. As discussed above, the small noise expansion [7] conveniently conditions our analysis to this situation. We therefore replace $\varsigma_{ik} \to \varepsilon\varsigma_{ik}$, write $\varphi_i = \varphi_i^{(0)} + \varepsilon\varphi_i^{(1)} + \dots$ and expand (30) in ε. When equating terms of zero order, $\mathcal{O}\left(\varepsilon^0\right)$, we recover the defining equations for the phase-locked reference state (27) and (28). The first order approximation is a multivariate Ornstein-Uhlenbeck process

$$\mathrm{d}\boldsymbol{\varphi}^{(1)} = \boldsymbol{G}\boldsymbol{\varphi}^{(1)}\mathrm{d}t + \boldsymbol{\varsigma}\mathrm{d}\boldsymbol{w} \tag{31}$$

with coupling matrix

$$G_{ij} = \begin{cases} -\gamma'_{ij}\left(\Delta\phi_{i,j}^{(\mathrm{ref})}\right) & \text{for } i \neq j \\ \sum_j \gamma_{ij}\left(\Delta\phi_{i,j}^{(\mathrm{ref})}\right) & \text{for } i = j \end{cases} \quad . \tag{32}$$

Its solution is given by [7]

$$\boldsymbol{\varphi}^{(1)}(t) = \exp(\boldsymbol{G}t)\,\boldsymbol{\varphi}^{(1)}(0) + \int_0^t \exp\left(\boldsymbol{G}\left(t-t'\right)\right)\boldsymbol{\varsigma}\mathrm{d}\boldsymbol{w}(t) \quad . \tag{33}$$

We define matrices

$$\boldsymbol{M}_t := \exp(\boldsymbol{G}t) \quad , \quad \boldsymbol{L}_t := \exp(\boldsymbol{G}t)\,\boldsymbol{\varsigma}\boldsymbol{\varsigma}^{\mathrm{T}}\exp\left(\boldsymbol{G}^{\mathrm{T}}t\right) \tag{34}$$

and

$$\boldsymbol{H}_t := \int_0^t \boldsymbol{L}_{t'}\mathrm{d}t' \quad . \tag{35}$$

The transition probability of starting from a phase vector $\boldsymbol{\varphi}_0^{(1)}$ at time $t = 0$ and ending in a state $\boldsymbol{\varphi}_t^{(1)}$ at time t is then given by

$$p\left(\boldsymbol{\varphi}_t^{(1)}|\boldsymbol{\varphi}_0^{(1)}\right) = \mathcal{N}_{\boldsymbol{M}_t\boldsymbol{\phi}_0^{(1)},\boldsymbol{H}_t}\left(\boldsymbol{\varphi}_t^{(1)}\right) \tag{36}$$

where $\mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(x)$ denotes a multivariate normal distribution (12). Note that up to $\mathcal{O}\left(\varepsilon^2\right)$ this is also the transition probability $p(\boldsymbol{\phi}_{t+d}|\boldsymbol{\phi}_t)$. Combining (36) with (28) and (29) we obtain

$$p(\boldsymbol{\phi}_{t+d}|\boldsymbol{\phi}_t) = \mathcal{N}_{\boldsymbol{M}_t\boldsymbol{\phi}_t,\boldsymbol{H}_t}(\boldsymbol{\phi}_t - \Omega t) \tag{37}$$

Stationary Distribution To derive the stationary distribution $p(\boldsymbol{\phi}_t)$ we utilize that the system is invariant under uniform phase rotations. It is therefore convenient to transform to new coordinates

$$\tilde{\boldsymbol{\varphi}} = (\bar{\varphi}, \boldsymbol{\delta\varphi}) = \boldsymbol{O}\boldsymbol{\varphi} \tag{38}$$

where $\boldsymbol{O}$ is an orthogonal matrix ($\boldsymbol{O}^{\mathrm{T}}\boldsymbol{O} = \boldsymbol{O}\boldsymbol{O}^{\mathbf{T}} = \mathbf{1}$) with $O_{1j} = \frac{1}{\sqrt{N}}$ such that $\bar{\boldsymbol{\varphi}}$ is an average like phase and $\boldsymbol{\delta\varphi} = (\delta\varphi_2, \dots, \delta\varphi_N)$ encodes the phase differences. Note that any phase difference $\varphi_i - \varphi_j$ can be expressed in terms of the $\boldsymbol{\delta\varphi}$ only as

$$\varphi_i - \varphi_j = \sum_{k=2}^{N} o_k^{ij}\delta\varphi_k \tag{39}$$

where

$$o_k^{ij} = (O_{ki} - O_{kj}) \tag{40}$$

In the new coordinates the dynamics (26) read

$$d\tilde{\boldsymbol{\varphi}} = \tilde{\boldsymbol{f}}(\boldsymbol{\delta\varphi})\,dt + \boldsymbol{B}d\boldsymbol{w} \tag{41}$$

where $\boldsymbol{B} = \boldsymbol{O}\varsigma$ and $\tilde{\boldsymbol{f}}(\boldsymbol{\delta\varphi}) := \boldsymbol{O}\boldsymbol{f}\left(\boldsymbol{O}^{\mathrm{T}}\tilde{\boldsymbol{\varphi}}\right)$ only depends on $\boldsymbol{\delta\varphi}$. Hence, the dynamics of the $\delta\varphi$ decouple from $\bar{\varphi}$ and the equation for $\bar{\varphi}$ itself can be solved formally as l

$$\bar{\varphi}(t) = \bar{\varphi}(0) + \int_0^t \tilde{f}_1(\boldsymbol{\delta\varphi})\, dt + \sum_k B_{1k} w_k(t) \tag{42}$$

By transforming this solution back to the original coordinates one observes that the collective rotation of the system is driven by three additive parts: the deterministic rotation Ω of the phase locked state, a drive due to the deviations of the oscillators from the phase locked state and a noise term being an average like combination of the input noises to each oscillator.

To proceed we again exploit the small noise assumption which ensures that the phase differences $\boldsymbol{\delta\varphi}$ stay small. In first order approximation we may thus linearize the $\boldsymbol{\delta\varphi}$-subsystem in (41) which gives

$$d\boldsymbol{\delta\varphi} = \boldsymbol{\delta G}\ \boldsymbol{\delta\varphi}\ dt + \boldsymbol{\delta B}\ d\boldsymbol{w} \tag{43}$$

where the $\boldsymbol{\delta G}$ is a $(N-1)\times(N-1)$ matrix defined by the equation

$$\boldsymbol{O}\boldsymbol{G}\boldsymbol{O}^{\mathrm{T}} = \begin{pmatrix} 0 & \bar{\boldsymbol{g}}^{\mathrm{T}} \\ \boldsymbol{0} & \boldsymbol{\delta G} \end{pmatrix} \tag{44}$$

and the $N-1\times N$ matrix $\boldsymbol{\delta B}$ has entries $\delta B_{i,j} = B_{i+1,j}$. Note that in (44) $\bar{\boldsymbol{g}}$ is a vector encoding the linear impact of deviations from the phase-locked state onto the evolution of the average phase.

Due to the stability of the phase locked state, $\boldsymbol{\delta G}$ only has strictly negative eigenvalues, so there is a stationary solution to the Ornstein-Uhlenbeck process (43) of the form

$$\boldsymbol{\delta\varphi}(t) = \int_{-\infty}^t \exp\left(\boldsymbol{\delta G}\left(t-t'\right)\right)\boldsymbol{\delta B} d\boldsymbol{w}\left(t'\right) \tag{45}$$

and the phase deviations $\boldsymbol{\delta\varphi}$ are Gaussian distributed with zero mean and covariances

$$\boldsymbol{H}_\delta = \int_{-\infty}^0 \exp\left(-\boldsymbol{\delta G}t\right)\boldsymbol{\delta B}\boldsymbol{\delta B}^{\mathrm{T}}\exp\left(-\boldsymbol{\delta G}^{\mathrm{T}}t\right)dt \tag{46}$$

i.e.

$$p_{\mathrm{st}}(\boldsymbol{\delta\varphi}) = \mathcal{N}_{\boldsymbol{0},\boldsymbol{H}_\delta}(\boldsymbol{\delta\varphi}) \quad . \tag{47}$$

Further, using equation (42) $\bar{\varphi}(t)$ has a Gaussian distribution with variance proportional to t for each realization of $\boldsymbol{\delta\varphi}(t)$. In particular, in the limit $t\to\infty$ this distribution on $\mathbb{R}$ becomes flat. Thus regarding $\bar{\varphi}$ as a phase variable on the circle its distribution becomes uniform which is also a direct consequence of the rotational phase symmetry of (26).

In summary, we obtain $p_{\mathrm{st}}(\bar{\boldsymbol{\varphi}}) \propto p_{\mathrm{st}}(\boldsymbol{\delta\varphi}) = \mathcal{N}_{\boldsymbol{0},\boldsymbol{H}_\delta}(\boldsymbol{\delta\varphi})$ and in terms of the phases φ_i we have

$$p_{\mathrm{st}}(\boldsymbol{\varphi}) \propto \mathcal{N}_{\boldsymbol{0},\boldsymbol{H}_\delta}\left((\boldsymbol{O\varphi})_{2,\ldots,N}\right) \tag{48}$$

Joint Probability Distribution We now merge our results (36) and (48) form the previous two sections to obtain the joint distribution as

$$\begin{aligned} p(\boldsymbol{\varphi}_t, \boldsymbol{\varphi}_{t+d}) &= p(\boldsymbol{\varphi}_{t+d}|\boldsymbol{\varphi}_t)\, p_{\mathrm{st}}(\boldsymbol{\varphi}_t) \\ &\propto \mathcal{N}_{\boldsymbol{M}_d\boldsymbol{\varphi}_t,\boldsymbol{H}_d}(\boldsymbol{\varphi}_{t+d})\,\mathcal{N}_{\boldsymbol{0},\boldsymbol{H}_\delta}(\boldsymbol{\delta\varphi}_t) \end{aligned} \tag{49}$$

or more explicitly

$$p_s(\boldsymbol{\varphi}_t, \boldsymbol{\varphi}_{t+d}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\delta\varphi}_t^{\mathrm{T}}\boldsymbol{H}_\delta^{-1}\boldsymbol{\delta\varphi}_t - \frac{1}{2}\left(\boldsymbol{\varphi}_{t+d} - \boldsymbol{M}_d\boldsymbol{\varphi}_t\right)^{\mathrm{T}}\boldsymbol{H}_d^{-1}\left(\boldsymbol{\varphi}_{t+d} - \boldsymbol{M}_d\boldsymbol{\varphi}_t\right)\right) \quad . \tag{50}$$

We are interested in the marginal distribution $p_s(\varphi_{i,t}, \varphi_{j,t+d})$ and thus have to integrate out the remaining coordinates in the full joint probability distribution (50). The Gaussian integrals over the $\varphi_{k,t+d}$, $k\neq j$ yield

$$p(\boldsymbol{\varphi}_t, \varphi_{j,t+d}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\delta\varphi}_t^{\mathrm{T}}\boldsymbol{H}_\delta^{-1}\boldsymbol{\delta\varphi}_t - \frac{1}{2}\left(\varphi_{j,t+d} - \left(\boldsymbol{M}_d\boldsymbol{\varphi}_t\right)_j\right)^2\left(\boldsymbol{H}_d\right)_{jj}^{-1}\right) \tag{51}$$

Now, as we can express any phase φ_j via the phase φ_i plus a linear combination of phase differences $\boldsymbol{\delta\varphi}$ using (39) we have

$$\varphi_{j,t+d} - (\boldsymbol{M}_d \boldsymbol{\varphi}_t)_j = \varphi_{j,t+d} - \varphi_{i,t} - \left(\boldsymbol{a}_d^{j,i}\right)^T \boldsymbol{\delta\varphi}_t \tag{52}$$

where the vector $\boldsymbol{a}_d^{j,i}$ has components

$$a_{d,l}^{j,i} := \sum_k (\boldsymbol{M}_d)_{jk}\, o_l^{k,i} \tag{53}$$

and we have used $\sum_j (\boldsymbol{M}_d)_{ij} = 1$ as $\boldsymbol{G}$ has eigenvalue $\lambda_0 = 0$ with eigenvector $(1, \ldots, 1)$.

Inserting (52) into (51) we can now perform the integration over the $\boldsymbol{\delta\varphi}_t$ which results in

$$p\left(\varphi_{i,t}, \varphi_{j,t+d}\right) = \frac{1}{2\pi} \mathcal{N}_{0,\sigma_{i,j,d}^2}\left(\varphi_{j,t+d} - \varphi_{i,t}\right) \tag{54}$$

where

$$\sigma_{i,j,d}^2 = (\boldsymbol{H}_d)_{jj} + \left(\boldsymbol{a}_d^{j,i}\right)^{\mathrm{T}} \cdot \boldsymbol{H}_\delta \cdot \boldsymbol{a}_d^{j,i} \tag{55}$$

This constitutes our first main result. Via the second term this expression still depends on the coordinate transformations $\boldsymbol{O}$ which we can eliminate. Therefore we define $[[\boldsymbol{A}]]_{i,j}$ to be the matrix obtained from the matrix $\boldsymbol{A}$ by deleting its i^{th} row and j^{th} column. With this notation straightforward algebra shows

$$\boldsymbol{\delta B \delta B}^T = \left[\left[\boldsymbol{O \varsigma \varsigma}^{\mathrm{T}} \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \quad , \quad \boldsymbol{\delta G}^n = \left[\left[\boldsymbol{O G}^n \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \tag{56}$$

and thus

$$\exp\left(\boldsymbol{\delta G} t\right) = \left[\left[\boldsymbol{O} \exp\left(\boldsymbol{G} t\right) \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \tag{57}$$

For any $N \times N$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ we further have

$$\left[\left[\boldsymbol{O A O}^{\mathrm{T}}\right]\right]_{1,1} \left[\left[\boldsymbol{O B O}^{\mathrm{T}}\right]\right]_{1,1} = \left[\left[\boldsymbol{O}\left(\boldsymbol{AB} - \frac{1}{N}\boldsymbol{AJB}\right)\boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \tag{58}$$

where $\boldsymbol{J}$ is the $N \times N$ matrix of ones, i.e $J_{ij} = 1$. We have $\boldsymbol{GJ} = \boldsymbol{JG}^{\mathrm{T}} = \boldsymbol{0}$ and thus for any integers $n, m \geq 0$

$$\begin{aligned} \boldsymbol{\delta G}^n \boldsymbol{\delta B \delta B}^T \left(\boldsymbol{\delta G}^T\right)^m &= \left[\left[\boldsymbol{O G}^n \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \left[\left[\boldsymbol{O \varsigma \varsigma}^{\mathrm{T}} \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \left[\left[\boldsymbol{O}\left(\boldsymbol{G}^{\mathrm{T}}\right)^n \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \\ &= \left[\left[\boldsymbol{O G}^n \boldsymbol{\varsigma \varsigma}^T \left(\boldsymbol{G}^{\mathrm{T}}\right)^m \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \end{aligned} \tag{59}$$

and it follows that

$$\exp\left(\boldsymbol{\delta G} t\right) \boldsymbol{\delta B \delta B}^{\mathrm{T}} \exp\left(\boldsymbol{\delta G}^{\mathrm{T}} s\right) = \left[\left[\boldsymbol{O} \exp\left(\boldsymbol{G} t\right) \boldsymbol{\varsigma \varsigma}^{\mathrm{T}} \exp\left(\boldsymbol{G}^{\mathrm{T}} s\right) \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} \quad . \tag{60}$$

Hence

$$\boldsymbol{\Sigma}_\delta = \int_{-\infty}^{0} \left[\left[\boldsymbol{O} \exp\left(-\boldsymbol{G} t\right) \boldsymbol{\varsigma \varsigma}^{\mathrm{T}} \exp\left(-\boldsymbol{G}^{\mathrm{T}} t\right) \boldsymbol{O}^{\mathrm{T}}\right]\right]_{1,1} dt \tag{61}$$

where it is essential to delete the first row and column before performing the integration to ensure

convergence of the integral. Using (34) together with (53), (40) and defining $o_1^{i,k} = 0$ we obtain

$$\begin{aligned}
\left(\boldsymbol{a}_d^{j,i}\right)^{\mathrm{T}} \cdot \boldsymbol{H}_\delta \cdot \boldsymbol{a}_d^{j,i} &= \int_{-\infty}^{0} \sum_{k,m} \sum_{l,p\neq 1} (\boldsymbol{M}_d)_{jk}\, o_l^{k,i} \left(\left[\left[\boldsymbol{O}\boldsymbol{L}_{-t}\boldsymbol{O}^{\mathrm{T}} \right] \right]_{1,1} \right)_{lp} (\boldsymbol{M}_d)_{jm}\, o_p^{m,i} dt \\
&= \int_{-\infty}^{0} \sum_{k,l,m,p} (\boldsymbol{M}_d)_{jk}\, o_l^{k,i} \left(\boldsymbol{O}\boldsymbol{L}_{-t}\boldsymbol{O}^{T} \right)_{lp} o_p^{m,i} \left(\boldsymbol{M}_d \right)_{jm} dt \\
&= \int_{-\infty}^{0} \sum_{k,l,m,p,s,r} (\boldsymbol{M}_d)_{jk} \left(O_{lk} - O_{li} \right) O_{lr} L_{-t,rs} O_{ps} \left(O_{pm} - O_{pi} \right) \left(\boldsymbol{M}_d \right)_{jm} dt \\
&= \int_{-\infty}^{0} \sum_{k,m} (\boldsymbol{M}_d)_{jk} \left(L_{-t,km} - L_{-t,ki} - L_{-t,im} + L_{-t,ii} \right) \left(\boldsymbol{M}_d \right)_{jm} dt \\
&= \int_{-\infty}^{0} \left[\left(\boldsymbol{M}_d \boldsymbol{L}_{-t} \boldsymbol{M}_d^{\mathrm{T}} \right)_{jj} - 2 \left(\boldsymbol{M}_d \boldsymbol{L}_{-t} \right)_{ji} + \boldsymbol{L}_{-t,ii} \right] dt \qquad (62)
\end{aligned}$$

Note here, as before, that the sum has to be performed before the integration to ensure convergence of the integral. Using (55) we obtain the O independent expression

$$\sigma_{i,j,d}^2 = \int_0^d \left(\boldsymbol{L}_t \right)_{jj} dt + \int_0^\infty \left[\left(\boldsymbol{L}_{t+d} \right)_{jj} + \left(\boldsymbol{L}_t \right)_{ii} - 2 \left(\boldsymbol{M}_d \boldsymbol{L}_t \right)_{ji} \right] dt \quad . \qquad (63)$$

and in the original coordinates the distribution (54) is given by

$$p\left(\phi_{i,t}, \phi_{j,t+d} \right) = \mathcal{N}_{0,\sigma_{i,j,d}^2} \left(\phi_{j,t+d} - \phi_{i,t} - \Delta\phi_{j,i} - \Omega d \right) \quad . \qquad (64)$$

The above results are for the phases considered on the real line. If we identify the phases at all points modulo 2π the Gaussian distribution (64) becomes a wrapped Gaussian distribution which in accordance with the small noise approximation for small standard deviations $\sigma_{i,j,d} \ll 2\pi$ is well approximated by a van Mises distribution for circular variables [15]. It is of the form

$$\mathcal{M}_{\mu,k}\left(\phi \right) = \frac{1}{2\pi I_0\left(k \right)} \exp\left(k \cos\left(\phi - \mu \right) \right) \qquad (65)$$

where $I_n\left(k \right)$ denotes the n^{th} modified Bessel function of the first kind [15], μ is the average phase and $k = 1/\sigma^2$ a concentration parameter. We thus obtain as the final result for the joint probability distribution,

$$p\left(\phi_{i,t}, \phi_{j,t+d} \right) = \mathcal{M}_{\Delta\phi_{j,i} + \Omega d, \sigma_{d,i,j}^{-2}} \left(\phi_{j,t+d} - \phi_{i,t} \right) \qquad (66)$$

with $\sigma_{d,i,j}^2$ given in (63).

Main Theorem For a joint probability distribution of the from $p\left(\phi_1, \phi_2 \right) = \frac{1}{2\pi} \mathcal{M}_{\mu,k}\left(\phi_1 - \phi_2 \right)$ we calculate for the mutual information

$$\begin{aligned}
\mathrm{MI}_{\mathrm{vM}}\left(k \right) &:= \iint p\left(\phi_1, \phi_2 \right) \log\left(\frac{p\left(\phi_1, \phi_2 \right)}{p\left(\phi_1 \right) p\left(\phi_2 \right)} \right) d\phi_1 d\phi_2 \\
&= \frac{k I_1\left(k \right)}{I_0\left(k \right)} - \log\left(I_0\left(k \right) \right) \qquad (67)
\end{aligned}$$

Combining this with the results from the previous derivations we obtain our main result for the information flow in phase oscillator networks:

Theorem 1. *The delayed mutual information* $\mathrm{dMI}_{i,j}\left(d \right)$ *between oscillator* i *and oscillator* j *in system (26) with stochastic dynamics around a deterministic stable phase locked state (27) with phase differences* $\Delta\phi_{i,j}^{(\mathrm{ref})}$ *is given by*

$$\mathrm{dMI}_{i,j}\left(d \right) = \begin{cases} \mathrm{MI}_{\mathrm{vM}}\left(\sigma_{d,i,j}^{-2} \right) & \text{for } d \geq 0 \\ \mathrm{MI}_{\mathrm{vM}}\left(\sigma_{-d,j,i}^{-2} \right) & \text{for } d < 0 \end{cases} \qquad (68)$$

where $\sigma_{d,i,j}^2$ *is given in (63) and* $\mathrm{MI}_{\mathrm{vM}}$ *in (67).*

4.2 Transfer Entropy in Networks of Coupled Phase Oscillators

In this section we extend our results on the calculation of the delayed mutual information to the transfer entropy introduced in [6]. The delayed transfer entropy (dTE) between phase signal $\phi_{i,t}$ and $\phi_{j,t+d}$ for a time delay d is given by (4)To calculate it we use the form (5) for which we have to determine the joint distribution $p\left(\phi_{i,t},\phi_{j,t},\phi_{j,t+d}\right)$ and some of its marginals. We therefore use the result (51) and rewrite

$$\varphi_{j,t+d}-\left(\boldsymbol{M}_d\boldsymbol{\varphi}_t\right)_j=\varphi_{j,t+d}-\varphi_{j,t}-\left(\boldsymbol{a}_d^j\right)^T\boldsymbol{\delta\varphi}_t \tag{69}$$

where the vector $\boldsymbol{a}_d^j$ has $n-1$ components $a_{d,l}^j$ $(l=2,\ldots,N)$

$$a_{d,l}^j:=\sum_{k=1}^{N}\left(\boldsymbol{M}_d\right)_{jk}\left(O_{lk}-O_{lj}\right)\quad . \tag{70}$$

We thus obtain

$$\begin{aligned}p\left(\boldsymbol{\varphi}_t,\boldsymbol{\varphi}_{j,t+d}\right) &\propto \exp\left(-\frac{1}{2}\boldsymbol{\delta\varphi}_t^T\boldsymbol{\Sigma}_\delta^{-1}\boldsymbol{\delta\varphi}_t-\frac{1}{2}\left(\varphi_{j,t+d}-\varphi_{j,t}-\left(\boldsymbol{a}_d^j\right)^T\boldsymbol{\delta\varphi}\right)^2\left(\boldsymbol{\Sigma}_d\right)_{jj}^{-1}\right)\\ &= \exp\left(-\frac{1}{2}\left(\varphi_{j,t+d}-\varphi_{j,t}\right)^2\left(\boldsymbol{\Sigma}_d\right)_{jj}^{-1}\right)\times\\ &\quad\exp\left(-\frac{1}{2}\delta\varphi_t^T\left(\boldsymbol{\Sigma}_\delta^{-1}+\left(\boldsymbol{\Sigma}_d\right)_{jj}^{-1}\boldsymbol{a}_d^j\left(\boldsymbol{a}_d^j\right)^T\right)\boldsymbol{\delta\varphi}_t+\varphi_{j,t+d}\left(\boldsymbol{\Sigma}_d\right)_{jj}^{-1}\left(\boldsymbol{a}_d^j\right)^T\boldsymbol{\delta\varphi}\right)\end{aligned} \tag{71}$$

Using that

$$\Delta\varphi_{k,j}:=\varphi_k-\varphi_j=\left(\boldsymbol{e}_k^T-\boldsymbol{e}_j^T\right)\boldsymbol{O}^T\begin{pmatrix}\bar{\varphi}\\ \boldsymbol{\delta\varphi}\end{pmatrix}=\left(\boldsymbol{e}_k^T-\boldsymbol{e}_j^T\right)\boldsymbol{O}^T\begin{pmatrix}0\\ \boldsymbol{\delta\varphi}\end{pmatrix} \tag{72}$$

we see that there is an invertible $(n-1)\times(n-1)$ matrix $\boldsymbol{K}_j$ such that

$$\boldsymbol{\Delta\varphi}:=\left(\Delta\varphi_{k,j}\right)_{k\neq j}=\boldsymbol{K}_j\boldsymbol{\delta\varphi}\qquad\boldsymbol{\delta\varphi}=\boldsymbol{K}_j^{-1}\boldsymbol{\Delta\varphi} \tag{73}$$

We write $\Delta\varphi_{j,t+d,t}=\varphi_{j,t+d}-\varphi_{j,t}$ and calculate

$$p\left(\varphi_t,\varphi_{j,t+d}\right) = \exp\left(-\frac{1}{2}\Delta\varphi_{j,t+d,t}^2\left(\boldsymbol{\Sigma}_d\right)_{jj}^{-1}-\frac{1}{2}\boldsymbol{\Delta\varphi}^T\boldsymbol{R}_{j,d}^{-1}\boldsymbol{\Delta\varphi}+\boldsymbol{\Delta\varphi}_{j,t+d,t}\boldsymbol{b}_{j,d}^T\boldsymbol{\Delta\varphi}\right) \tag{74}$$

with

$$\begin{aligned}\boldsymbol{R}_{j,d}^{-1} &= \left(\boldsymbol{K}_j^{-1}\right)^T\left(\boldsymbol{\Sigma}_\delta^{-1}+\left(\boldsymbol{\Sigma}_d\right)_{jj}^{-1}\boldsymbol{a}_d^j\left(\boldsymbol{a}_d^j\right)^T\right)\boldsymbol{K}_j^{-1}\\ \boldsymbol{b}_{j.d}^T &= \left(\boldsymbol{\Sigma}_d\right)_{jj}^{-1}\left(\boldsymbol{a}_d^j\right)^T\boldsymbol{K}_j^{-1}\end{aligned} \tag{75}$$

We can now perform the Gaussian integrals over all $\Delta\varphi_{k,j}$ with $k\neq i$ to obtain

$$p\left(\varphi_{j,t+d},\varphi_{j,t}\varphi_{i,t}\right)=\frac{1}{2\pi}\mathcal{N}_{\mathbf{0},\boldsymbol{C}}\left(\varphi_{j,t+d}-\varphi_{j,t},\varphi_{i,t}-\varphi_{j,t}\right) \tag{76}$$

where the 2×2 matrix $\boldsymbol{C}$ is defined by

$$\boldsymbol{C}^{-1}=\begin{pmatrix}\left(\boldsymbol{\Sigma}_d\right)_{j,j}^{-1}-\boldsymbol{b}_{j,d}^T\boldsymbol{R}_{j,d}\boldsymbol{b}_{j,d}+\left(\boldsymbol{R}_{j,d}\right)_{i,i}^{-1}\left(\boldsymbol{R}_{j,d}\boldsymbol{b}_{j,d}\right)_i^2 & -\left(\boldsymbol{R}_{j,d}\right)_{i,i}^{-1}\left(\boldsymbol{R}_{j,d}\boldsymbol{b}_{j,d}\right)_i\\ -\left(\boldsymbol{R}_{j,d}\right)_{i,i}^{-1}\left(\boldsymbol{R}_{j,d}\boldsymbol{b}_{j,d}\right)_i & \left(\boldsymbol{R}_{j,d}\right)_{i,i}^{-1}\end{pmatrix} \tag{77}$$

Here we also used the fact that we are considering phase variables and restricted the range of $\varphi_{j,t}\in[0,2\pi]$. Independence of $\varphi_{j,t}$ in this expression again reflects rotational symmetry of the full system. Calculation of the marginals and insertion in to (5) results in our second main result

Theorem 2. *The delayed transfer entropy* $\mathrm{dTE}_{i,j}(d)$ *between oscillator* i *and oscillator* j *in system (26) close to a deterministic stable phase locked state (27) with phase differences* $\Delta\phi_{i,j}^{(\mathrm{ref})}$ *and small noise fluctuations* ς *is given by*

$$\mathrm{dTE}_{i\to j}(d) = \frac{1}{2}\log\left(\frac{C_{11}C_{22}}{\det(\boldsymbol{C})}\right) = -\frac{1}{2}\log\left(1-\frac{C_{12}^2}{C_{11}C_{22}}\right) \tag{78}$$

where C_{ij} *are entries in the matrix* $\boldsymbol{C}$ *given by (77).*

As for the dMI, this result is not dependent on the orthogonal transformation O. To take into account the phase character of the two variables $\varphi_{j,t+d}$ and $\varphi_{i,t}$ one can view the distribution (76) as a wrapped Gaussian or approximate it with a multivariate van Mises distribution similarly to the calculations for the dMI above. Due to the small noise all three expressions are all approximately correct and we stop our derivations here. Indeed, Fig. 2 in the main manuscript and Supplementary Fig. 1 show that the theoretical prediction for the dTE in this form is in fact in good agreement with numerical estimates. The figures also show that the dTE more clearly reveals the asymmetry in the information flows in parallel to the already detected ones with the simpler delayed mutual information measures.

Note that due to rotational symmetry, the dTE in (78) is actually a mutual information for the phase difference variables $\varphi_{j,t+d}-\varphi_{j,t}$ and $\varphi_{i,t}-\varphi_{j,t}$. Interestingly, in our approximation and in contrast to the dMI the above expression for the dTE is independent of the absolute noise level which makes it a pure function of the underlying network parameters and its dynamical state. Therefore it may serve as a generic measure to characterize effective interactions in these networks.

4.3 Delayed Mutual Information and Transfer Entropy for Two Coupled Oscillators

Applying theorem 1 to a network of $N=2$ oscillators and setting $g_1 = G_{1,1}$, $g_2 = G_{2,2}$ in (32) we obtain for $i \neq j$ and $\lambda = (g_1+g_2)$

$$\sigma_{i,j,d}^2 = \frac{\xi^2}{\lambda^3}\begin{cases} d\lambda\left(g_1^2+g_2^2\right) - \lambda^2 - 2g_j^2\left(e^{\lambda d}-1\right) & \text{for } d \geq 0 \\ |d|\,\lambda\left(g_1^2+g_2^2\right) - \lambda^2 - 2g_i^2\left(e^{\lambda|d|}-1\right) & \text{for } d \leq 0 \end{cases} \tag{79}$$

from which the delayed mutual information follows via (68). Maximizing the mutual information with respect to d is equivalent to minimizing $\sigma_{i,j,d}^2$ and gives maximal shared information at

$$d^* = (g_1+g_2)^{-1}\log\left(\frac{1}{2}\left(1+\left(\frac{g_2}{g_1}\right)^2\right)\right) \tag{80}$$

if $g_1 < g_2$.

Similarly we obtain for the transfer entropy with $d>0$ the analytic expression

$$\mathrm{dTE_{i\to j}(d)} = -\frac{1}{2}\log\left(\frac{g_j^2\left(e^{\lambda d}-1\right)^2}{\left(g_1^2+g_2^2\right)\lambda d + 2g_1g_2\left(e^{\lambda d}-1\right)}+1\right) \quad . \tag{81}$$

4.4 Limits and Future Extensions

Interestingly, even though we performed several approximation steps (e.g. phase reduction, phase estimation and a small noise expansion) to arrive at our general theoretical results (Theorem 1), there is a good agreement to numerical simulations. We here discuss some limitations and future extension of our theory.

In our derivation we assumed weak coupling. Relaxation of this assumption is possible in systems of amplitude oscillators that mainly couple via the phase (e.g. as observed in [16]). In general, more complex dynamics can occur for strong coupling (e.g. [17]) which requires an extension of our results to amplitude oscillators and more complex dynamical states. In our numerical simulations we find that even in systems with non-weak coupling displaying phase locked dynamical states our theory gives

good estimates for the directionality of the information flow. For even stronger coupling and noise inputs recent generalized phase reduction techniques that use a full manifold of limit cycles generated by different but constants inputs [18] may be used together with our theory. For future studies it would be further interesting to extend our theory to amplitude oscillators where the information routing is controlled by the phase dynamics while the actual information transfer is separated to amplitudes modes.

We assumed small noise levels in our derivation to enable the small noise approximation. In Supplementary Fig. 2 we show the effect of larger noise levels on the delayed mutual information. We observe that for larger noise levels the theoretical result is systematically larger than the numerical estimation but the shape stays qualitatively similar (cf. Supplementary Fig. 2a). The strong noise makes it likely to push the system further away from the phase locked dynamics such that the linear approximations made in the small noise expansions no longer remain valid. The transients back to the phase-locked state blur the mutual information and reduce the overall height in the dMI curves. In systems with multiple phase-locked attractors very strong noise is capable of pushing the system into different attractor states from time to time and the dMI becomes a superposition of the dMIs in each phase locked state weighted by factors that depend on the average transition rates. Additional transients between these states may blur this superposition (cf. Supplementary Fig. 2b). Conditioning the time series on specific states, a separation into the different dMI curves associated with the different underlying dynamical attractors may be achieved (not shown). Analogous reasoning applies to the delayed transfer entropy measure of information flow.

4.5 Numerical Estimation of Delayed Mutual Information and Transfer Entropy

Throughout this work the delayed mutual information and transfer entropy was determined by estimating the probability distributions appearing in (1) from simulated time series. Stochastic integration of the time series was performed using a second order stochastic Runge-Kutta algorithm [13] with fixed time step Δt. For non-phase reduced systems the phases where estimated from the stochastic time series of the full system as described in Supplementary Note 3.1. For each time delay joint probability densities were estimated by binning the data using a uniform spacing in phase space. All integrals were then performed numerically on this grid. Sample sizes and bin counts where chosen large so that the result of this implementation of the maximum likelihood or naive estimator [19, 20] was within the error of the best upper bound estimator [20]. Symbol sizes in all plots are larger then this error. To check validity of the curves we also doubled sample sizes and used a finer binning restricted to five standard deviations of the phase data.

Supplementary Note 5
Information Routing in Hierarchical Networks of Phase Oscillators

Here we derive expressions for the information flow between sub-groups of phase oscillators in hierarchical (modular) networks. We first reduce each individual cluster to a meta-oscillator described by a collective phase and response function following refs. [21, 22] but additionally account for the stochastic dynamics. We show that the reduced stochastic phase oscillator model has the same functional form as the original model. We can therefore apply our general results, Theorem 1 and 2, to obtain the non-local delayed mutual information and transfer entropies between the clusters as a function of the local cluster properties and their local dynamical states.

5.1 Hierarchical Networks of Phase Oscillators

Throughout this section we consider networks of N phase oscillators as described in Supplementary Note 4.1.1 with hierarchical network structure. We assume that the oscillators are clustered into M different groups or clusters $X \in \{1, \dots, M\}$ that consists of N_X oscillators so that $\sum_X N_X = N$. We denote the i^{th}-oscillator in cluster X by i_X. In this notation, the evolution equation (26) becomes

$$d\phi_{i_X}(t) = \left[\omega_{i_X} + \sum_{j_X} \gamma_{i_X,j_X}(\phi_{i_X} - \phi_{j_X}) + \sum_{Y}\sum_{j_Y} \gamma_{i_X,j_Y}(\phi_{i_X} - \phi_{j_Y})\right] dt + \sum_k \varsigma_{i_X,k} dw_k \tag{82}$$

where the first sum on the right hand side represents the stronger intra-cluster couplings whereas the second the weaker inter-cluster connections.

For each individual cluster X we assume the existence of a phase-locked state in the noiseless dynamics ($\varsigma_{i_x,k} = 0$) with phase differences $\Delta\phi_{i_X,j_X} = \Delta\phi_{i_X} - \Delta\phi_{j_X}$ that obey

$$\omega_{i_X} + \sum_{j_X} \gamma_{i_X,j_X}(\Delta\phi_{i_X,j_X}) = \Omega_X = \text{const.} \tag{83}$$

where Ω_X is the collective cluster frequency. The dynamics of the individual oscillators may then be described by $\phi_{i_X,0}(t) = \Phi_X(t) + \Delta\phi_{i_X,1}$ where Φ_X is the collective cluster phase.

We further assume that the differences in the cluster oscillation frequencies Ω_X are of the order of the inter-cluster coupling strength so that in the full deterministic model the clusters themselves show a stable phase-locked pattern with phase differences $\Delta\Phi_{X,Y}$ and a global collective rotation frequency Ω obeying

$$\Omega = \Omega_X + \sum_Y \Gamma_{X,Y}(\Delta\Phi_{X,Y}) = \text{const.} \tag{84}$$

Finally, we assume that the noise strengths $\varsigma_{i_x,k}$ are small in comparison to the strength of attraction towards this phase-locked dynamical state.

5.2 Collective Phase Reduction

In this section we make use of the hierarchical network structure by first neglecting the inter-cluster couplings and consider each cluster separately. Using the assumption that in the noiseless system each cluster has a stable phase locked state (83) we may regard each group as a single meta-oscillator, for each of which we can perform a standard phase reduction step [8]. A cluster X is then described by its collective phase Φ_X and its collective phase response curve $\boldsymbol{Z}_X$ [21, 22]. The phase response vector satisfies the adjoint equation

$$\frac{d}{dt}\boldsymbol{Z}_X = -\boldsymbol{G}_X^{\mathrm{T}}\boldsymbol{Z}_X \tag{85}$$

where $\boldsymbol{G}_X$ is the $N_X \times N_X$ matrix

$$(\boldsymbol{G}_X)_{i_X,j_Y} = \begin{cases} -\gamma'_{i_X j_X}(\Delta\phi_{i_X,j_X}) & \text{for } i_X \neq j_X \\ \sum_{k_X} \gamma'_{i_X k_X}(\phi_{i_X,k_X}) & \text{for } i_X = j_X \end{cases} \tag{86}$$

As $\boldsymbol{G}_X$ is time independent and a Laplacian matrix we can solve (85) by choosing $\boldsymbol{Z}_X$ to be the normalized left eigenvector of $\boldsymbol{G}_X$ with eigenvalue $\lambda_{X,0} = 0$ which is constant and independent of the phase Φ_X. It is given by

$$Z_{X,i_X} = \frac{\det\left([[\boldsymbol{G}_X]]_{i_X,i_X}\right)}{\sum_{i_X}\det\left([[\boldsymbol{G}_X]]_{i_X,i_X}\right)} \quad . \tag{87}$$

So far we performed the phase reduction analysis for the deterministic situation $\varsigma_{i_x,k} = 0$. As the noise sources w_i in our model represent physical input signals they will have small non-zero correlations in time [10]. Moreover, we assumed sufficient stability of the limit cycle against amplitude perturbations. Given these two facts the above reduction results stay valid in the stochastic situation [9].

The full stochastic system (82) in the phase reduced form then becomes

$$d\Phi_X = \Omega_X + \sum_Y \boldsymbol{Z}_X^{\mathrm{T}}\boldsymbol{G}_{X,Y}(\Phi_X, \Phi_Y) + \sum_{i_X,k} Z_{X,i_X}\varsigma_{i_X,k} dw_k \tag{88}$$

where

$$\boldsymbol{G}_{X,Y}\left(\Phi_X,\Phi_Y\right)_{i_X} = \sum_{i_Y} \gamma_{i_X j_Y}\left(\Phi_X - \Phi_Y - \Delta\phi_{i_X,j_Y}\right) \quad . \tag{89}$$

Now note that in equation (88) the $\boldsymbol{Z}_X$ are constant vectors and $\boldsymbol{G}_{X,Y}$ only depends on the cluster phase differences. We therefore can write (88) in the form

$$\frac{d}{dt}\Phi_X = \Omega_X + \sum_Y \Gamma_{X,Y}\left(\Phi_X - \Phi_Y\right) + \sum_K \Sigma_{X,K}\Xi_K \tag{90}$$

with inter-cluster coupling function

$$\Gamma_{X,Y}\left(\Phi_X - \Phi_Y\right) \quad = \quad \sum_{i_X,j_Y} Z_{X,i_X}\gamma_{i_X j_Y}\left(\Phi_X + \phi_{i_X,0} - \Phi_Y - \phi_{j_Y,0}\right) \quad , \tag{91}$$

and M independent Gaussian white noise processes Ξ_K and a $M \times M$ - covariance matrix $\boldsymbol{\Sigma} = (\Sigma_{X,K})$ that satisfies

$$\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}}\right)_{X,Y} = \sum_{i_X,j_Y,k} Z_{X,i_X}\varsigma_{i_X k}\varsigma_{j_Y,k}Z_{Y,j_Y} \quad . \tag{92}$$

Note that the reduction process starting from equation (26) and leading to (90) leaves the form of the dynamical equations invariant. Hence, this reduction step may be iterated in multi-scale networks to obtain a phase description of the stochastic dynamics on every scale.

5.3 Delayed Mutual Information and Transfer Entropy between Collective Cluster Phases

The equation (90) is now formally completely analogous to (25) and we may therefore directly apply theorem 1 to obtain an expression for the delayed mutual information between the collective phase signals of the clusters:

Corollary 3. *The delayed mutual information* $\mathrm{dMI}_{X,Y}(d)$ *between the time series of the collective phases* Φ_X *and* Φ_Y *of cluster* X *and cluster* Y *in system (82) close to a phase locked state with phase differences* $\Delta\Phi_{X,Y}$ *and small noise amplitudes* $\varsigma_{i_x k}$ *is given by (68) when substituting* ϕ_i *with* Φ_X, $\Delta\phi_{i,j}$ *by* $\Delta\Phi_{X,Y}$, ω_i *with* Ω_X, $\gamma_{i,j}$ *by* $\Gamma_{X,Y}$, *and* $\varsigma_{i,k}$ *by* $\Sigma_{X,K}$.

Similarly we can use theorem 2 to obtain:

Corollary 4. *The delayed transfer entropy* $\mathrm{dTE}_{X,Y}(d)$ *between the time series of the collective phases* Φ_X *and* Φ_Y *of cluster* X *and cluster* Y *in system (82) close to a phase locked state with phase differences* $\Delta\Phi_{X,Y}$ *and small noise amplitudes* $\varsigma_{i_x,k}$ *is given by (78) with the substitutions as in corollary (3).*

We remark that these results together with the invariance of the dynamical equations under the stochastic phase reduction process can be used to resolve the information flow on every scale in a hierarchical network of coupled oscillators close to a phase-locked state. In Supplementary Fig. 1 the dMI and dTE measures calculated here and in the previous section are shown for the hierarchical Wilson-Cowan network of Figs. 1 and 2 of the main manuscript. Supplementary Fig. 3 shows that local changes within a cluster affect multiple cluster properties such as its collective phase response and also the effective cluster inter-actions. Together these changes all determine the new global dynamical state and the global information routing pattern.

Supplementary Note 6
Time Dependent Information Routing Patterns

Our analysis in Supplementary Note 4 can be extended to non-phase locked collective states. Using the expressions for the delayed mutual information (1) and transfer entropy (5) we obtain two information theoretic measures $\mathrm{dMI}_{i,j}(d,t)$ and $\mathrm{dTE}_{i\to j}(d,t)$ that in general now not only depend on d but also on

the initial state x_t as visible from the general approach presented in Supplementary Note 2. As analytic solutions to these expressions are not feasible in this situation we refer to numerical evaluation.

A example for a information routing pattern generated by a non-phase-locked dynamical state is shown in Supplementary Fig. 5a-c. Here the neuronal network from Fig. 4 in the main manuscript was used with the local dynamical states of the three clusters in the configuration $\mathcal{D} = [\beta_A \alpha_B \beta_C]$ that led to a stable limit cycle solution in the deterministic case (Supplementary Fig. 5a). We observe dynamical changes in the information routing patterns as determined by the delayed transfer entropy $\mathrm{dTE}_{i\to j}(d,t)$ along the limit cycle solution (Supplementary Fig. 5b,c). Using phase reduction on the local dynamical states result in the effective three node networks in Fig. 4c of the main manuscript. In the simulations we used 1.2×10^6 sample trajectories with time step $\Delta t = 0.1$ and 50 equally spaced bins in each phase coordinate to estimate of the joint probabilities entering the dTE.

We note that our theory can also be extended using an instant time information flow measure [23, 24]. It is derived from basic principles to measure information flow in dynamical systems and resolves the instantaneous time dependence. However, unlike dMI or dTE it does not provide any relevant time scales for the information flow itself. Details and discussion on the time dependence and time resolve measures of information routing will be presented elsewhere.

Supplementary Note 7
Dynamical State Dependent Communication

The results of the previous sections are concerned with abstract dynamic information routing measures that provide insight into the routability of information in complex networks. Here we provide a concrete example how a signal can be en- an decoded to achieve dynamical state dependent signal transmission. In general, the concrete realization of the en- and decoding schemes will differ and will be problem as well as system dependent.

For the network from Fig. 2 in the main manuscript the speed of the phase variable $\dot{\phi}_i$ is suitable to encode the transmitted signals between the two nodes. Both, neuronal units as well as gene regulatory circuits have been shown to be able to detect the speed of change in a signal, and, in particular, to effectively apply a threshold to it [25, 26, 27]. In Supplementary Fig. 6 the network is prepared in either of the dynamical states α (left) or β (right). While in state α a signal delivered to oscillator 1 is not propagated in a detectable way to oscillator 2, the signal is detected in oscillator 2 when the network is in the state β. Both findings are in accordance with the predicted information routing patterns.

Finally, in Supplementary Fig. 7 the two oscillator network from Fig. 2 of the main manuscript is prepared in the dynamical state α (left). While in state α a signal delivered to oscillator 1 is not propagated in a detectable way to oscillator 2, a stronger pulse in the input that signals the need to change the IRP switches the network into state β and now signals delivered to oscillator 1 are detected by oscillator 2. The switching signal could be either part of the input to the network or provided by a second network that detects the need to switch the IRP. Pulse or burst like stimuli are common in neuronal networks [27] and, interestingly, in gene regulatory circuits as well [28].

Supplementary Note 8
Model Descriptions

In the previous sections we developed a theory for the delayed mutual information in phase oscillator networks. Here we give details on the example networks we use in the main manuscript to illustrate the broad applicability of our theory and to investigate control mechanisms for information routing.

8.1 Wilson-Cowan Equations For Networks of Neuronal Populations

Collective neuronal oscillations are frequently observed in many parts of the nervous system, ranging from primary sensory circuits, through local cortical networks to larger inter-areal formations [29, 30, 31]. On a population level the dynamics of these networks may be described in terms of mean field equations for the firing rates of neuronal sub-populations [1, 32]. To induce external signals as additive

noise we here focus on a variant by Grannan et al. [2] where the equations are rephrased in terms of average membrane-potential like variables.

We consider networks of N neuronal groups $i \in \{1, \ldots, N\}$, each described by the average membrane potential variables v_i and u_i for the excitatory and inhibitory sub-population respectively, evolving according to

$$\begin{aligned} \tau \frac{d}{dt} v_i &= -v_i + g_{\mathrm{ee}} g(v_i) - g_{\mathrm{ie}} g(u_i) + i_{\mathrm{e},i} + \sum_j g_{i,j} g(v_j) + \rho_{\mathrm{e},i} \zeta_{\mathrm{e},i} \\ \tau \frac{d}{dt} u_i &= -u_i + g_{\mathrm{ei}} g(v_i) - g_{\mathrm{ii}} g(u_i) + i_{\mathrm{i},i} + \rho_{\mathrm{i},i} \zeta_{\mathrm{i},i} \quad . \end{aligned} \tag{93}$$

Here τ is the membrane time scale. Within each group the excitatory synaptic connections to the inhibitory and to the excitatory sub-group are denoted by g_{ei} and g_{ee}, while g_{ii} and g_{ie} denote the inhibitory projections. The different groups are coupled via long range excitatory connections with strength $g_{i,j}$. The firing rate of group i is obtained from the potential v_i via

$$g(v_i) = \frac{1}{1 + \exp(-4\beta(v_i - v_0))} \tag{94}$$

External signals are modeled as independent white noise processes $\zeta_{\mathrm{e},i}$ and $\zeta_{\mathrm{i,i}}$ with strengths $\rho_{\mathrm{e,i}}$ and $\rho_{\mathrm{i},i}$.

For the simulations that lead to Figs. 1, 2 and 4 in the main manuscript and Supplementary Figs. 1 and 4 we used parameters as in [2]. In particular, $g_{\mathrm{ee}} = 15$, $g_{\mathrm{ei}} = 15$, $g_{\mathrm{ii}} = 5$, $g_{\mathrm{ie}} = 12$, $\beta = 1.1$, $v_0 = 1$, $i_{\mathrm{i},i} = 0$, $i_{\mathrm{e},i} = 0.75$ that give rise to a stable limit cycle solution (cf. Supplementary Fig. 4b,c). For present links (indicated as arrows between the triangles in each network graph) the coupling strength within each cluster was $g_{i,j} = 0.1$ while between clusters $g_{i,j} = 0.015$. A stochastic integration time step of $\Delta t = 0.01$ with noise levels $\rho_{\mathrm{e},i} = 0.001$ and $\rho_{\mathrm{i},i} = 0.001$ were used.

The scalar phase field $\phi_i(v_i, u_i)$ was determined on a grid of 200×200 points in the relevant phase space region $(v_i, u_i) \in [-2.1, -2] \times [-0.3, 2.5]$ with trajectories of duration $t = 10000$. The result is shown in Supplementary Fig. 4.

Stable phase-locked states where determined numerically using (27). In particular, by concentrating on the phase differences $\Delta\phi_{i,1}$ we obtain from (26) in the noiseless case

$$\frac{d}{dt}\Delta\phi_{i,1} = \omega_i - \omega_1 + \sum_j \gamma_{i,j}(\Delta\phi_{i,1} - \Delta\phi_{j,1}) - \sum_j \gamma_{1,j}(-\Delta\phi_{j,1}) \tag{95}$$

For two identical oscillators we obtain

$$\frac{d}{dt}\Delta\phi_{1,2} = \gamma(\Delta\phi_{1,2}) - \gamma(-\Delta\phi_{1,2}) = \bar{\gamma}(\Delta\phi_{1,2}) \tag{96}$$

which shows that zeros of the anti-symmetric part of the coupling function $\bar{\gamma}$ with negative slope give rise to stable phase locked states. Interestingly, for two coupled Wilson-Cowan oscillators we find two such solutions giving rise to intrinsic multi-stable dynamics in these neuronal networks (cf. Supplementary Fig. 4f). In general we determined the phase-locked states by numerically finding zeros of the norm of the vector field on the right hand side of (95), i.e.

$$\left\| \frac{d}{dt}\boldsymbol{\Delta\phi} \right\|^2 = \sum_i \left(\omega_i - \omega_1 + \sum_j \gamma_{i,j}(\Delta\phi_{i,1} - \Delta\phi_{j,1}) - \sum_j \gamma_{1,j}(-\Delta\phi_{j,1}) \right)^2 \tag{97}$$

with negative eigenvalues of the Jacobian to ensure stability. For the two sub-networks with three oscillators of the full modular network used in Fig. 1 and 4 in the main manuscript we find two multi-stable states each (cf. Supplementary Fig. 4g and h). Phase-locking in the full network was determined by using the second phase-reduction step on the clusters (cf. (5.2)) and repeating the above procedure for every combination of local phase-locked states in each cluster. Phase-locking was validated by direct simulation of the phase reduced and full equations.

Besides the two three-oscillator networks used in the main article, we also investigated phase-locking dynamics in the three remaining strongly connected network motives [33] of three oscillators: The ring network only exhibits a single phase-locked state, the fully connected network shows a large variety of multi-stable phase-locked and periodic states, while three stable states are found for the remaining network, showing that multi-stability is a generic phenomena in these networks of neuronal oscillators.

For the numerical calculations of the delayed mutual information we used time series of duration $t = 4 \times 10^6$ and 4000×4000 bins for estimation the joint portability distributions. The transfer entropy was estimated by estimating the joint probability density $p\left(\phi_{j,t+d}, \phi_{j,t}, \phi_{i,t}\right)$ with the same methods as for the dMI but using $500 \times 500 \times 500$ bins. Switching between the different stable dynamical states can be induced for example by a stimulus to one of the oscillators in a single clusters using input current pulses of strengths 2 and duration $\Delta t = 0.5$.

8.2 Coupled Biochemical Oscillator Networks

Rhythmic activity of protein expressions in cells is a widely observed phenomenon [34, 35, 36], including circadian rhythms [37, 38, 39], the cell cycle [40], regulatory mechanisms during development and growth [41, 42], and in synthetic gene networks [43, 44, 45, 46, 47, 48]. The phase of these oscillations directly encodes the expression levels of the proteins involved in the oscillation and thus information about the current state of the cell. To illustrate the mechanisms for the control of information routing in these networks we consider a network of coupled generic biochemical Goodwin oscillators as well as a model of a synthetic gene regulatory network of fast and tunable oscillators [44, 47].

8.2.1 Network of Coupled Goodwin Oscillators

The model network consists of mutually coupled bio-chemical Goodwin oscillators [49, 50, 34, 36] that were introduced to study enzyme kinetics and successfully describe aspects of circadian oscillations such as in Neurospora [51]. Oscillations arise due to a nonlinear and inhibitory feedback mechanism: A gene is transcribed into mRNA that is translated into an enzyme, which then catalyzes the formation of a protein that represses the initial translation process (cf. Supplementary Fig. 8). Denoting the concentrations of mRNA, enzyme and protein of the i-th oscillator as x_i, y_i and z_i respectively, the system evolves in time t according to [36]

$$\begin{aligned}
\frac{d}{dt}x_i &= k_{x,i}\frac{1}{1+z_i^p} - d_{x,i}x_i + \sum_j c_{i,j}\left(t\right) + \rho_{x,i}\zeta_{x,i} \\
\frac{d}{dt}y_i &= k_{y,i}x_i - d_{y,i}\frac{y_i}{K_{m,i}+y_i} + \rho_{y,i}\zeta_{y,i} \\
\frac{d}{dt}z_i &= k_{z,i}y_i - d_{z,i}\frac{z_i}{K_{n,i}+z_i} + \rho_{z,i}\zeta_{z,i}
\end{aligned} \tag{98}$$

Here $k_{x,i}$,$k_{y,i}$ and $k_{z,i}$ are the rates for translation, transcription and catalysis, and $d_{x,i}$ is the degradation rate of x_i. Enzymes and proteins are assumed to be degraded by Michaelis-Menten kinetics [50] with constants $K_{m,i}$ and $K_{n,i}$. Repression of translation is modeled via a Hill-function [52] with cooperativity parameter p for the auto repression. Other oscillators j couple to oscillator i via their proteins z_j that also repress translation of x_i with cooperativity parameter l_i, Hill-constant $K_{c,i}$ and coupling strength $a_{i,j}$ [53], i.e.

$$c_{i,j}\left(t\right) = \frac{a_{i,j}}{K_{c,i} + z_j^l\left(t\right)} \quad . \tag{99}$$

All concentrations receive external signals modeled as uncorrelated white noise processes $\zeta_{x,i}$, $\zeta_{y,i}$ and $\zeta_{z,i}$ with $\left\langle \zeta_{a,i}\left(t\right), \zeta_{b,j}\left(s\right)\right\rangle = \delta_{a,b}\delta_{i,j}\delta\left(t-s\right)$ and noise levels $\rho_{x,i}$,$\rho_{y,i}$ and $\rho_{z,i}$. For the example in Fig. 1 of the main manuscript we choose two oscillators $i \in \{1,2\}$ with constants adapted from [36]. In particular, $p = 4$, $k_{x,i} = 0.2 + (i-1)0.1$, $d_{x,i} = 0.1$, $k_{y,1} = 0.2$, $d_{y,i} = 0.1$, $K_{m,i} = 0.01$, $k_{z,i} = 0.05$, $d_{z,i} = 1$, $K_{n,i} = 20$. Coupling constants where $a_{1,2} = a_{2,1} = 0.015$, $a_{1,1} = a_{2,2} = 0$, $K_{c,i} = 1$ and $l_i = 4$.

The collective dynamics was changed and subsequently the effective directionality in information routing was reversed by using $d_{x,2} = 0.12$ instead of $d_{x,2} = 0.1$ in the second row of Fig. 1a-d in the

main manuscript. Thus local changes to one oscillator are capable of controlling the information to and from another oscillatory components of the gene network.

For numerical integrations we used a time step $\Delta t = 0.1$. The phase field $\phi_i(x_i, y_i)$ was estimated as described in Supplementary Section 3.1 using a grid of 100^3 initial conditions in the phase space region $(x_i, y_i, z_i) \in [0, 1.8] \times [0, 4.6] \times [0.2, 4.4]$ and deterministic trajectories of duration $t = 4000$ (cf. Supplementary Fig. 8b). The delayed mutual information was estimated using stochastic trajectories of duration $t = 10^7$ and noise strengths $\rho_{x,i} = \rho_{y,i} = \rho_{z,i} = 0.001$. The theoretical predictions were obtained by semi-analytically determining the phase response curves, rotation frequencies, coupling functions and effective noise levels as described in Supplementary Note 3.2 together with numerically solving for a stable phase locked state using (27) and then using the theoretical results in Supplementary Note 4.

8.2.2 Network of Synthetic Gene Oscillators

We here consider a network of synthetic gene regulatory oscillators, in which similarly to the previous example information routing between oscillators may be steered by acting only locally on one of them. In the simplest example, we consider two mutually coupled genetic oscillators where the dynamics of each individual oscillator is based on an artificially engineered robust and tunable genetic oscillator proposed in [44] and experimentally realized in [47].

Each individual oscillator $i \in \{1, 2\}$ is described by (rescaled) concentrations of two proteins, x_i and y_i. While protein x_i facilitates the synthesis of itself and y_i, y_i is repressing its own translation and that of x_i (cf. Supplementary Fig. 9a). In the uncoupled case we consider the dynamical equations as proposed in [44], table I and equation (1). We additionally assume that oscillator i is coupled to j via a diffusive term (cf. Supplementary Fig. 10)

$$c_{i,j}(t) = a_{i,j}\left(x_j(t) - x_i(t)\right) \tag{100}$$

Such a coupling may be realized by an additional reaction $x_1 \rightleftharpoons x_2$ with reaction rates $a_{1,2}$ and $a_{2,1}$. The reduced dynamics of this network then reads

$$\begin{aligned} \dot{x}_i &= \frac{1 + x_i^2 + \alpha_i \beta_i x_i^4}{\left(1 + x_i^2 + \beta_i x_i^4\right)\left(1 + y_i^4\right)} - \gamma_{x,i} x_i + c_{i,j}(t) + \rho_{x,i}\zeta_{x,i} \\ \tau_{y,i}\dot{y}_i &= \frac{1 + x_i^2 + \alpha_i \sigma_i x_i^4}{\left(1 + x_i^2 + \beta_i x_i^4\right)\left(1 + y_i^4\right)} - \gamma_{y,i} y_i + \rho_{x,i}\zeta_{x,i} \end{aligned} \tag{101}$$

where α_i is the increase in transcription rate due to binding of x_i to one of the operator sites of the promoter and β_i the affinity for an x_i dimer to bind to one of the operator sites relative to the other, $\tau_{y,i}$ is the time scale of the dynamics of protein y_i. The parameters $\gamma_{x,i}$ and $\gamma_{y,i}$ describe the effective decay of the proteins. The parameter $\gamma_{y,i}$ can be influenced by changing the concentration of proteins that degrade or permanently bind to y_i [44, 47] (cf. Supplementary Fig. 10a). All concentrations receive white noise processes $\zeta_{x,i}$ and $\zeta_{y,i}$ with $\langle \zeta_{a,i}(t), \zeta_{b,j}(s) \rangle = \delta_{a,b}\delta_{i,j}\delta(t - s)$ and noise levels $\rho_{x,i}$ and $\rho_{y,i}$.

For the examples in Supplementary Figs. 9 and 10 we choose parameter as in [44], in particular $\alpha_i = 11$, $\sigma_i = 2$, $\tau_{y,i} = 5$, $\gamma_{x,i} = 0.105$ and $\gamma_{y,2} = 0.027$. We regard $\gamma_{y,1}$ as a parameter that may be controlled externally and that locally influences the oscillation properties of oscillator $i = 1$. For the examples we choose $\gamma_{y,1} = 0.026$ or $\gamma_{y,1} = 0.029$. Noise levels were $\rho_{x,i} = \rho_{y,i} = 0.002$. As visible from Supplementary Fig. 10 this change is capable of reversing the information routing between the two oscillators.

We used a time step $\Delta t = 0.1$ for numerical integration. Deterministic trajectories of duration $t = 10000$ on a grid of 120^2 points equally spaced in relevant phase space region $(x_i, y_i) \in [-0.25, 2.1] \times [2.25, 3.1]$ was used to estimate the phase field numerically as described in Supplementary Note 3.1. The delayed mutual information was calculated using a time series of duration $t = 10^7$ and a histogram of $10^4 \times 10^4$ equidistant bins in each phase coordinate. The theoretical predictions were obtained by semi-analytically determining the phase response curves, rotation frequencies, coupling functions and effective noise levels as described in Supplementary Note 3.2 together with numerically solving for a stable phase locked state using (27).

8.3 Networks of Stuart-Landau Oscillators

A very generic mechanism to generate oscillations is via a Hopf bifucation. In this section we consider a generic network of coupled Stuart-Landau oscillators. Using center manifold reduction and normal form theory any system of deterministic weakly coupled oscillators each close to Hopf bifurcation may be transformed into this form [12]. Here we present a stochastic extension of this result.

The network consists of N stochastic Stuart-Landau oscillators each described by a two dimensional state (x_i, y_i) that evolve according to

$$\begin{aligned} dx_i &= \left(a_i^\rho x_i - a_i^\iota y_i - \left(b_i^\rho x_i - b_i^\iota y_i\right)\left(x_i^2 + y_i^2\right) + \sum_j c_{ij}^\rho x_j - c_{ij}^\iota y_j \right) dt + \rho_{x,i} dw_{x,i} \\ dy_i &= \left(a_i^\iota x_i + a_i^\rho y_i - \left(b_i^\iota x_i + b_i^\rho y_i\right)\left(x_i^2 + y_i^2\right) + \sum_j c_{ij}^\iota x_j + c_{ij}^\rho y_j \right) dt + \rho_{y,i} dw_{y,i} \end{aligned} \tag{102}$$

where a_i^ρ, a_i^ι, b_i^ρ and b_i^ι are real parameters capturing the properties of the oscillatory dynamics of node i and and c_{ij}^ρ and c_{ij}^ι are real coupling constants. As before $w_{\alpha,i}$ are independent unit variance Wiener processes and $\rho_{x,i}$ and $\rho_{y,i}$ the standard deviations of the noise. To keep calculations simple we here focus on uncorrelated inputs to the individual oscillators and assume $\rho_{x,i} = \rho_{y,i} = \rho_i$.

For this model a full analytic stochastic phase reduction is possible. The scalar phase field as introduced in Supplementary Note 3.1 is given by the analytical expression

$$\phi_i = \arctan\left(x_i, y_i\right) - b_i \log\sqrt{x_i^2 + y_i^2} \tag{103}$$

which gives rise to phase equations of the form

$$d\phi_i = \left(\omega_i + \sum_{j\neq i}^{n} c_{ij}^r \frac{r_{0,j}}{r_{0,i}} \left[\sin\left(\phi_j - \phi_i + c_{ij}^\varphi\right) - b_i \cos\left(\phi_j - \phi_i + c_{ij}^\varphi\right) \right] \right) dt + \varsigma_{\phi,i} dw_{\phi,i} \tag{104}$$

with $r_{i,0}^2 = \frac{a_i^\rho}{b_i^\rho}$ and $\omega_i = a_i^\iota - a_i^\rho b_i$, $b_i = \frac{b_i^\iota}{b_i^\rho}$ and $\sigma_{\phi,i}^2 = \rho_i^2\left(1 + b_i^2\right)/r_{0,i}^2$. Full details of this derivation and the more generic case of correlated input noises will be published elsewhere.

In the example in Fig. 3a-c of the main manuscript we used parameters $a_i^\rho = 1$,$a_i^\iota = 3$, $b_i^\rho = 2$, $b_i^\iota = 0.5$, $c_{i,j}^\varphi = 2$, $c_{i,j}^\rho = 0.02$ for intra-cluster links and $c_{i,j}^\varphi = 0.004$ for inter-cluster links. The noise level was $\rho_i = 0.001$. For the lower panels in Fig. 3b,c in the main manuscript we changed a_1^ι to $a_1^\iota = 3.008$ for the oscillator indicated by an arrow in panel a. Phase-locked states where determined numerically using equation (27). To estimate the delayed mutual information numerically we used integration steps $\Delta t = 0.1$, sample trajectories in the stationary state of length $t = 10^6$ and 3000×3000 bins in phase space. Panels a,b in Fig. 3 in the main manuscript highlighting the transition in information routing directionality were calculated using parameter values from $a_1^\iota = 2.9965$ to $a_1^\iota = 3.01$ in steps of 0.0005. Arrows in the effective networks represent the quantified information routing as in (2).

8.4 Networks of Kuramoto Phase-Oscillators

For the Kuramoto network shown in the lower part of Fig. 3 in the main manuscript we use equation (26) with $\gamma_{i,j}(\phi) = a_{i,j}\left(0.6 - \sin\left(\phi - 0.8\sin(\phi)^4 + 1\right)\right)$ and $\omega_i = 1.075$ for oscillators in the first cluster and $\omega_i = 1.0$ otherwise. For present links in the graph the coupling strength between clusters was $a_{i,j} = 0.1$ or $a_{i,j} = 0.2$ with equal probability and $a_{i,j} = 1$ or $a_{i,j} = 2$ within the clusters. By changing a local link indicated by an arrow in Fig. 3f in the main manuscript from 0.3 to 1.7 the non-local information sharing pattern is changed. Thus local link changes are capable of controlling information routing in the full network.

Density plots in the lower panels in Fig. 3 in the main manuscript illustrate the discrete switching like change in the directionality of information routing between the clusters due to changes in a single

local link. The plots were constructed using steps of 0.05 in the coupling weight. After determining the phase-locked states numerically for each weight we used Corollary 3 to determine the delayed mutual information. Weights of the effective information routing networks were determined using (2).

Supplementary References

[1] Wilson, H.R. and Cowan, J.D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**, 1–24 (1972).

[2] Grannan, E.R., Kleinfeld, D., and Sompolinsky, H. Stimulus-dependent synchronization of neuronal assemblies. *Neural Comput.* **5**, 550–569 (1993).

[3] Shaw, R. Strange attractors, chaotic behavior, and information flow. *Zeitschrift für Naturforsch.* **36**, 80 (1981).

[4] Vastano, J. and Swinney, H. Information transport in spatiotemporal systems. *Phys. Rev. Lett.* 1773–1776 (1988).

[5] Shannon, C.E. and Weaver, W. *The mathematical theory of communication.*, ((University of Illinois Press, Urbana, IL.1949).

[6] Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–4 (2000).

[7] Gardiner, G.W. Handbook of Stochastic Methods (1983).

[8] Kuramoto, Y. *Chemical Oscillations, Waves and Turbulence* (1984).

[9] Teramae, J.n., Nakao, H., and Ermentrout, G.B. Stochastic Phase Reduction for a General Class of Noisy Limit Cycle Oscillators. *Phys. Rev. Lett.* **102**, 194102 (2009).

[10] van Kampen, N.G. *Stochastic Processes in Physics and Chemistry*, (Elsevier2007), 3 edition.

[11] Ermentrout, G. and Kopell, N. Multiple pulse interactions and averaging in systems of coupled neural oscillators. *J. Math. Biol.* **29**, 195–217 (1991).

[12] Hoppenstaedt, F.C. and Izhikevich, E.M. *Weakly connected neural networks*, volume 126 of *Applied Mathematical Sciences*, (Springer, New York1997).

[13] Honeycutt, R. Stochastic runge-kutta algorithms. I. White noise. *Phys. Rev. A* **45**, 600–603 (1992).

[14] Acebrón, J., Bonilla, L., and Vicente, C. The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* **77**, 137–185 (2005).

[15] Bronstein, I.N., Semendjaev, K.A., Musiol, G., and Mühlig, H. *Taschenbuch der Mathematik*, (Verlag Harry Deutsch, Frankfurt am Main, Germany1999).

[16] Rosenblum, M., Pikovsky, A., and Kurths, J. Phase synchronization of chaotic oscillators. *Phys. Rev. Lett.* **76**, 1804–1807 (1996).

[17] Ermentrout, G. and Kopell, N. Oscillator death in systems of coupled neural oscillators. *SIAM J. Appl. Math.* **50**, 125–146 (1990).

[18] Kurebayashi, W., Shirasaka, S., and Nakao, H. Phase description of limit-cycle oscillators subject to strong perturbations. In *Eng. Chem. Complexity, 7th Int. Conf.* (2013).

[19] Strong, S., Koberle, R., de Ruyter van Steveninck, R., and Bialek, W. Entropy and Information in Neural Spike Trains. *Phys. Rev. Lett.* **80**, 197–200 (1998).

[20] Paninski, L. Estimation of Entropy and Mutual Information. *Neural Comput.* **15**, 1191–1253 (2003).

[21] Kawamura, Y., Nakao, H., Arai, K., Kori, H., and Kuramoto, Y. Collective Phase Sensitivity. *Phys. Rev. Lett.* **101**, 024101 (2008).

[22] Kori, H., Kawamura, Y., Nakao, H., Arai, K., and Kuramoto, Y. Collective-phase description of coupled oscillators with general network structure. *Phys. Rev. E* **80**, 036207 (2009).

[23] Liang, X. and Kleeman, R. Information Transfer between Dynamical System Components. *Phys. Rev. Lett.* **95**, 244101 (2005).

[24] Majda, a.J. and Harlim, J. Information flow between subspaces of complex dynamical systems. *Proc. Natl. Acad. Sci.* **104**, 9558–9563 (2007).

[25] Sorre, B., Warmflash, A., Brivanlou, A.H., and Siggia, E.D. Encoding of temporal signals by the TGF-β pathway and implications for embryonic patterning. *Dev. Cell* **30**, 334–42 (2014).

[26] Handbook of Brain Microcircuits, (Oxford University Press2010).

[27] Dayan, P. and Abbott, L.F. *Theoretical Neuroscience* (2001).

[28] Purvis, J.E., Karhohs, K.W., Mock, C., Batchelor, E., Loewer, a., and Lahav, G. p53 Dynamics Control Cell Fate. *Science* **336**, 1440–1444 (2012).

[29] Gray, C., König, P., Engel, A., and Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).

[30] Glass, L. Synchronization and rhythmic processes in physiology. *Nature* **410**, 277–284 (2001).

[31] Buzsaki, G. *Rythms of the Brain*, (Oxford University Press2006).

[32] Wilson, H.R. and Cowan, J.D. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* **13**, 55–80 (1973).

[33] Sporns, O. and Kötter, R. Motifs in brain networks. *PLoS Biol.* **2**, e369 (2004).

[34] Goldbeter, A. Computational approaches to cellular rhythms. *Nature* **420**, 238–45 (2002).

[35] Kruse, K. and Jülicher, F. Oscillations in cell biology. *Curr. Opin. Cell Biol.* **17**, 20–6 (2005).

[36] Novák, B. and Tyson, J.J. Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9**, 981–91 (2008).

[37] Hastings, M. Circadian clockwork: two loops are better than one. *Nat. Rev. Neurosci.* **1**, 3–6 (2000).

[38] Bell-Pedersen, D., Cassone, V.M., Earnest, D.J., Golden, S.S., Hardin, P.E., Thomas, T.L., and Zoran, M.J. Circadian Rhythms from Multiple Oscillators: Lessons from diverse Organisms. *Nat. Rev. Drug Discov.* **4**, 121–30 (2005).

[39] Zhang, E.E. and Kay, S.a. Clocks not winding down: unravelling circadian networks. *Nat. Rev. Mol. Cell Biol.* **11**, 764–76 (2010).

[40] Charvin, G., Cross, F.R., and Siggia, E.D. Forced periodic expression of G1 cyclins phase-locks the budding yeast cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6632–7 (2009).

[41] Dequéant, M.L., Glynn, E., Gaudenz, K., Wahl, M., Chen, J., Mushegian, A., and Pourquié, O. A complex oscillating network of signaling genes underlies the mouse segmentation clock. *Science* **314**, 1595–8 (2006).

[42] Moreno-Risueno, M.a., Van Norman, J.M., Moreno, A., Zhang, J., Ahnert, S.E., and Benfey, P.N. Oscillating gene expression determines competence for periodic Arabidopsis root branching. *Science* **329**, 1306–11 (2010).

[43] Elowitz, M.B. and Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–8 (2000).

[44] Hasty, J., Dolnik, M., Rottschäfer, V., and Collins, J. Synthetic Gene Network for Entraining and Amplifying Cellular Oscillations. *Phys. Rev. Lett.* **88**, 148101 (2002).

[45] McMillen, D., Kopell, N., Hasty, J., and Collins, J.J. Synchronizing genetic relaxation oscillators by intercell signaling. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 679–84 (2002).

[46] Ullner, E., Zaikin, A., Volkov, E., and García-Ojalvo, J. Multistability and Clustering in a Population of Synthetic Genetic Oscillators via Phase-Repulsive Cell-to-Cell Communication. *Phys. Rev. Lett.* **99**, 148103 (2007).

[47] Stricker, J., Cookson, S., Bennett, M.R., Mather, W.H., Tsimring, L.S., and Hasty, J. A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–9 (2008).

[48] Mondragón-Palomino, O., Danino, T., Selimkhanov, J., Tsimring, L., and Hasty, J. Entrainment of a population of synthetic genetic oscillators. *Science* **333**, 1315–9 (2011).

[49] Goodwin, B.C. An Entrainment Model for Timed Enzyme Synthesis in Bacteria. *Nature* **5022**, 479–481 (1966).

[50] Bliss, R.D., Painter, P.R., and Marr, a.G. Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.* **97**, 177–93 (1982).

[51] Ruoff, P., Vinsjevik, M., Monnerjahn, C., and Rensing, L. The Goodwin model: simulating the effect of light pulses on the circadian sporulation rhythm of Neurospora crassa. *J. Theor. Biol.* **209**, 29–42 (2001).

[52] Griffiths, J. Mathematics of cellular control processes I. Negative Feedback to One Gene. *J. Theor. Biol* **20**, 202–208 (1968).

[53] Wagner, A. Circuit topology and the evolution of robustness in two-gene circadian oscillators. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 11775–80 (2005).