Oscillatory recurrent gated neural integrator circuits (ORGaNICs), a unifying theoretical framework for neural dynamics

David J. Heeger[a,b,1] and Wayne E. Mackey[a,b]

[a]Department of Psychology, New York University, New York, NY 10003; and [b]Center for Neural Science, New York University, New York, NY 10003

Contributed by David J. Heeger, August 27, 2019 (sent for review July 17, 2019; reviewed by Brent Doiron and Michael M. Halassa)

Working memory is an example of a cognitive and neural process that is not static but evolves dynamically with changing sensory inputs; another example is motor preparation and execution. We introduce a theoretical framework for neural dynamics, based on oscillatory recurrent gated neural integrator circuits (ORGaNICs), and apply it to simulate key phenomena of working memory and motor control. The model circuits simulate neural activity with complex dynamics, including sequential activity and traveling waves of activity, that manipulate (as well as maintain) information during working memory. The same circuits convert spatial patterns of premotor activity to temporal profiles of motor control activity and manipulate (e.g., time warp) the dynamics. Derivative-like recurrent connectivity, in particular, serves to manipulate and update internal models, an essential feature of working memory and motor execution. In addition, these circuits incorporate recurrent normalization, to ensure stability over time and robustness with respect to perturbations of synaptic weights.

computational neuroscience | recurrent neural network | normalization | working memory | motor control

Neuroscience research on working memory has largely focused on sustained delay-period activity (1–4). A large body of experimental research has measured sustained activity in prefrontal cortex (PFC) and/or parietal cortex during delay periods of memory-guided saccade tasks (5–9) and delayed-discrimination and delayed match-to-sample tasks (10–13). Most of the models of working memory, based on neural integrators (see *SI Appendix*, Figs. S1–S3 for a primer on neural integrators), are aimed to explain sustained delay-period activity or to explain behavioral phenomena associated with sustained activity (14, 15).

Working memory, however, involves much more than simply holding a piece of information online. In cognitive psychology, the idea of working memory includes manipulating online information dynamically in the context of new sensory input. For example, understanding a complex utterance (with multiple phrases) often involves disambiguating the syntax and/or semantics of the beginning of the utterance based on information at the end of the sentence. Doing so necessitates representing and manipulating long-term dependencies, that is, maintaining a representation of the ambiguous information, and then changing that representation when the ambiguity is resolved. In addition, there are a variety of experimental results that are difficult to reconcile with sustained activity and neural integrator models. Some (if not the majority of) neurons either exhibit sequential activity such that activity is handed off from one neuron to the next during a delay period with each individual neuron being active only transiently (16–21) or they exhibit complex dynamics during delay periods (21–27). Complex dynamics (including oscillations) are evident also in the combined activity (e.g., as measured with local field potentials) of populations of neurons (28, 29). We hypothesize that these complex dynamics serve a purpose, to manipulate working memory representations.

Models of perceptual decision making, like working memory models, are also based on simple neural integrators. Specifically, perceptual decision making has been proposed to involve integration of noisy sensory information (30–34), a simple form of manipulation, in which neurons literally sum sensory-evoked activity over a period of time. However, a more general theoretical framework for representing and manipulating long-term dependencies is lacking.

Motor preparation and execution, analogous to working memory, involves maintaining a neural representation of a motor plan while manipulating that representation to generate the desired movement dynamics. Neural circuits and systems subserving motor preparation and execution exhibit analogous sustained and sequential activity phenomena (35–39), and there are analogous challenges reconciling these phenomena with neural integrator models.

Long short-term memory units (LSTMs) are machine-learning (ML) algorithms that represent and manipulate long-term dependencies (40). LSTMs are a class of recurrent neural networks. A number of variants of the basic LSTM architecture have been developed and tested for ML applications including language modeling, translation, and speech recognition (41–45). In these and other tasks, the input stimuli contain information across multiple timescales, but the ongoing presentation of stimuli makes it difficult to correctly combine that information over time. This is analogous to the problem of representing and manipulating long-term dependencies mentioned above in working memory, decision making, and motor control. An LSTM handles this problem by updating its internal state over time with a pair of gates: The update gate selects which part(s) of the current input to process, and

Significance

Oscillatory recurrent gated neural integrator circuits (ORGaNICs) are a family of recurrent neural circuits that can simulate a wide range of neurobiological phenomena, various of which have each been explained by different models. This theoretical framework can be used to simulate neural activity with complex dynamics, including sequential and traveling waves of activity. When used to model cognitive processing in the brain, these circuits can both maintain and manipulate information during a working memory task. When used to model motor control, these circuits convert spatial patterns of premotor activity to temporal profiles of motor control activity. ORGaNICs offer a conceptual framework; rethinking cortical computation in these terms should have widespread implications, motivating a variety of experiments.

Author contributions: D.J.H. and W.E.M. designed research; D.J.H. performed research; and D.J.H. wrote the paper.

Reviewers: B.D., University of Pittsburgh; and M.M.H., Massachusetts Institute of Technology.

The authors declare no competing interest.

Published under the PNAS license.

Data deposition: The MATLAB code for recreating the simulation results is available at https://archive.nyu.edu/handle/2451/60439.

[1]To whom correspondence may be addressed. Email: david.heeger@nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1911633116/-/DCSupplemental.

First published October 21, 2019.

Downloaded at Boost Library, New York University on November 6, 2019

the reset gate selectively deletes part(s) of the current output. The gates are computed at each time step from the current inputs and outputs. This enables LSTMs to maintain a representation of some of the inputs, until needed, and then to manipulate that representation based on inputs that come later in time.

Here, we introduce a theoretical framework for neural dynamics that is a generalization of and a biophysically plausible implementation of LSTMs. We show that these circuits simulate key phenomena of working memory, including both maintenance and manipulation, and both sequential and sustained activity. We also show that the exact same circuits (with the same synaptic weights) simulate key phenomena of motor control. Preliminary versions of this work, along with further details and mathematical derivations, were posted on preprint servers (46, 47). MATLAB code for recreating the simulation results is available at https://archive.nyu.edu/handle/2451/60439 (48).

Results

ORGaNICs. We begin by describing the basic architecture of oscillatory recurrent gated neural integrator circuits (ORGaNICs). The following subsections elaborate this basic architecture and demonstrate that this architecture can subserve a variety of functions including working memory and motor control.

An example ORGaNICs circuit is depicted in Fig. 1. The neural responses of a population of neurons are modeled as dynamical processes that evolve over time. The output responses depend on an input drive (a weighted sum of the responses of a population of input neurons) and a recurrent drive (a recurrent weighted sum of their own responses). The time-varying output responses are represented by a vector $\mathbf{y} = (y_1, y_2, \ldots, y_j, \ldots, y_N)$, where the subscript j indexes different neurons in the population (boldface lowercase letters denote vectors and boldface uppercase letters denote matrices.) The time-varying inputs are represented by another vector $\mathbf{x} = (x_1, x_2, \ldots, x_j, \ldots, x_M)$. The output responses are also modulated by 2 populations of time-varying modulators, recurrent modulators **a** and input modulators **b**. (We use the term "modulator" to mean a multiplicative computation regardless of whether or not it is implemented with neuromodulators.) The recurrent and input modulators are analogous, respectively, to the reset and input gates in LSTMs. The modulators depend on the inputs and outputs. So, there are 2 nested recurrent circuits: 1) recurrent drive: the output responses depend on the recurrent drive, which depends on a weighted sum of their own responses, and 2) multiplicative modulators: the output responses are modulated (multiplicatively) by the responses of 2 other populations of neurons (the modulators), which also depend on the output responses.

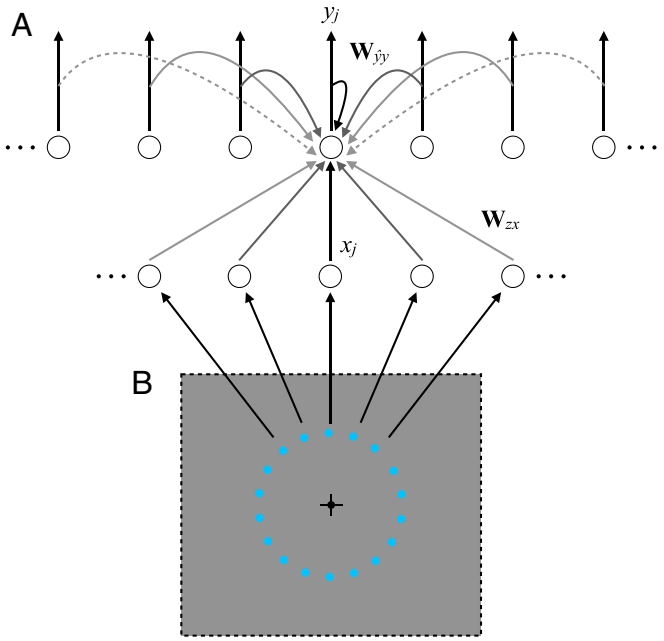

Fig. 1. ORGaNICs architecture. (*A*) Diagram of connections in an example ORGaNIC. Solid lines/curves are excitatory (positive weights) and dashed curves are inhibitory (negative weights). Gray scale represents strength of connections (weight magnitude). Only a few of the input-drive connections and recurrent-drive connections are shown to minimize clutter. (*B*) Oculomotor delayed response task. Black cross-hair, fixation point. Black circle, eye position at the beginning of a trial. Blue circles, possible target locations, each of which evokes an input.

Specifically, neural responses are modeled by the following dynamical systems equation:

$$\tau_y \frac{dy_j}{dt} = -y_j + \left(\frac{b_j^+}{1+b_j^+}\right) z_j + \left(\frac{1}{1+a_j^+}\right) \hat{y}_j, \qquad [1]$$

$$\mathbf{z} = \mathbf{W}_{zx}\mathbf{x},$$

$$\hat{\boldsymbol{y}} = \mathbf{W}_{\hat{y}y}\mathbf{y},$$

$$\mathbf{r} = \mathbf{W}_{ry}\mathbf{y},$$

$$a_j^+ \geq 0 \text{ and } b_j^+ \geq 0.$$

Eq. **1** can be implemented with a simplified biophysical (equivalent electrical circuit) model of pyramidal cells (see *SI Appendix* and ref. 46 for details). The variables (**y**, $\hat{\mathbf{y}}$, **x**, **z**, **a**, **b**, and **r**) are each functions of time, for example $\mathbf{y}(t)$, but we drop the explicit dependence on t to simplify the notation. The responses **y** depend on an input drive **z**, which is computed as a weighted sum of inputs **x**. The encoding weight matrix (also called the embedding matrix) $\mathbf{W}_{zx}$ is an $N \times M$ matrix of weights where N is the number of neurons in the circuit and M is the number of inputs to the circuit. The rows of $\mathbf{W}_{zx}$ are the response fields of the neurons. The responses **y** also depend on a recurrent drive $\hat{\mathbf{y}}$, which is computed as a weighted sum of the responses **y**. The recurrent weight matrix $\mathbf{W}_{\hat{y}y}$ is an $N \times N$ matrix. For the example circuit depicted in Fig. 1, the recurrent weights have a center-surround architecture in which the closest recurrent connections are excitatory and the more distant ones are inhibitory, and the circuit exhibits sustained activity (discussed below). For other choices of the recurrent weight matrix, the circuit can exhibit stable, ongoing oscillations, sequential activity, or traveling waves of activity (discussed below). The recurrent drive and input drive are modulated, respectively, by 2 other populations of neurons: the recurrent modulators **a** and the input modulators **b**. The superscript + is a rectifying output nonlinearity. Half-wave rectification is the simplest form of this rectifying nonlinearity, but other output nonlinearities could be substituted, for example sigmoid, exponentiation, half-squaring (49), normalization (50, 51), and so on. The value of τ_y is the intrinsic time constant of the neurons. Finally, the output responses are multiplied by a readout matrix $\mathbf{W}_{ry}$, where **r** is the readout (not depicted in the figure).

The time-varying values of the modulators **a** and **b** determine the state of the circuit by controlling the recurrent gain and effective time constant. During periods of time when both a_j and b_j are large (e.g., $a_j = b_j \gg 1$), the response time courses are dominated by the input drive, so the responses exhibit a short effective time constant. When both a_j and b_j are small ($\sim$0), the responses are dominated by the recurrent drive, so the responses exhibit a long effective time constant. When a_j is large and b_j is small, the recurrent drive is shut down (like the reset gate in an LSTM). A leaky neural integrator corresponds to a special case in which $a_j = b_j$ is constant over time (see *SI Appendix* for a primer on neural integrators).

Downloaded at Bobst Library, New York University on November 6, 2019

The modulators are themselves dynamical systems that depend on the inputs and outputs:

$$\tau_a \frac{d\mathbf{a}}{dt} = -\mathbf{a} + \mathbf{W}_{ax}\mathbf{x} + f(\mathbf{y}), \quad [2]$$

$$\tau_b \frac{d\mathbf{b}}{dt} = -\mathbf{b} + \mathbf{W}_{bx}\mathbf{x}.$$

The values of τ_a and τ_b are the intrinsic time constant of the modulator neurons. The recurrent modulator **a** depends on a function of the output responses $f(\mathbf{y})$, to incorporate recurrent normalization (*Robustness via Normalization* and *SI Appendix*). ORGaNICs (by analogy with LSTMs) use the modulators to encode a dynamically changing state. The modulators depend on the current inputs and the current outputs, which in turn depend on past inputs and outputs, so the state depends on the current inputs and past context. The modulators can be controlled separately for each neuron so that each neuron can have a different state (different values for a_j and b_j) at each instant in time. In the example that follows, however, all of the neurons in the circuit shared the same state, but that state changed over time.

ORGaNICs are inherently a nonlinear dynamical system because the input drive and the recurrent drive are each multiplied by nonlinear functions of the modulators (Eq. **1**) and because the recurrent modulator depends nonlinearly on the output responses (Eq. **2**). However, there are circumstances when these equations can be analyzed as a linear system, specifically when the modulators are constant over time, because the only remaining nonlinearity is due to the normalization which simply acts to rescale the responses.

There is considerable flexibility in the formulation of ORGaNICs, with different variants corresponding to different hypothesized neural circuits (*SI Appendix*). In one such variant, each of the modulators can depend on both the inputs and the outputs, unlike Eq. **2** in which only a depends on the output responses. In another variant, the 2 modulators have analogous effects such that larger values of a increase the gain of the recurrent drive, unlike Eq. **1** in which larger values of a decrease the gain of the recurrent drive. In yet another variant, the 2 modulators are coordinated to govern balance between input drive and recurrent drive.

The following subsections describe some examples of ORGaNICs. We begin with a simplified example of a sustained activity circuit, then modify the recurrent weights to simulate sequential activity and traveling waves, and then add multiple recurrent terms for manipulation. Simulated neural responses shown in the figures are intended to exhibit qualitative aspects of neurophysiological phenomena, that is, the models have not (yet) been optimized to replicate published data by tuning or fitting the model parameters. The weights in the various weight matrices were prespecified (not learned) for each of the simulations in this paper (although ORGaNICs are compatible with modified versions of ML algorithms; see *SI Appendix*).

Sustained Activity. We used ORGaNICs to simulate sustained activity during a memory-guided saccade task (Fig. 2), using the circuit depicted in Fig. 1*A*. In this task, a target is flashed briefly while a subject is fixating the center of a screen (Fig. 1*B*). After a delay period of several seconds, the fixation point disappears, cueing the subject to make an eye movement to the remembered location of the target.

The modulators in the simulation were constant during each successive phase of the behavioral task. Many experimental protocols in behavioral neuroscience comprise a sequence of distinct phases (including the oculomotor delayed response task; see below figures for more examples). The behavioral cues built into the experimental protocol set the state of the modulators via $\mathbf{W}_{ax}$ and $\mathbf{W}_{bx}$ in Eq. **2**, and the state changed from one phase to the next. During each phase, the modulators were constant and

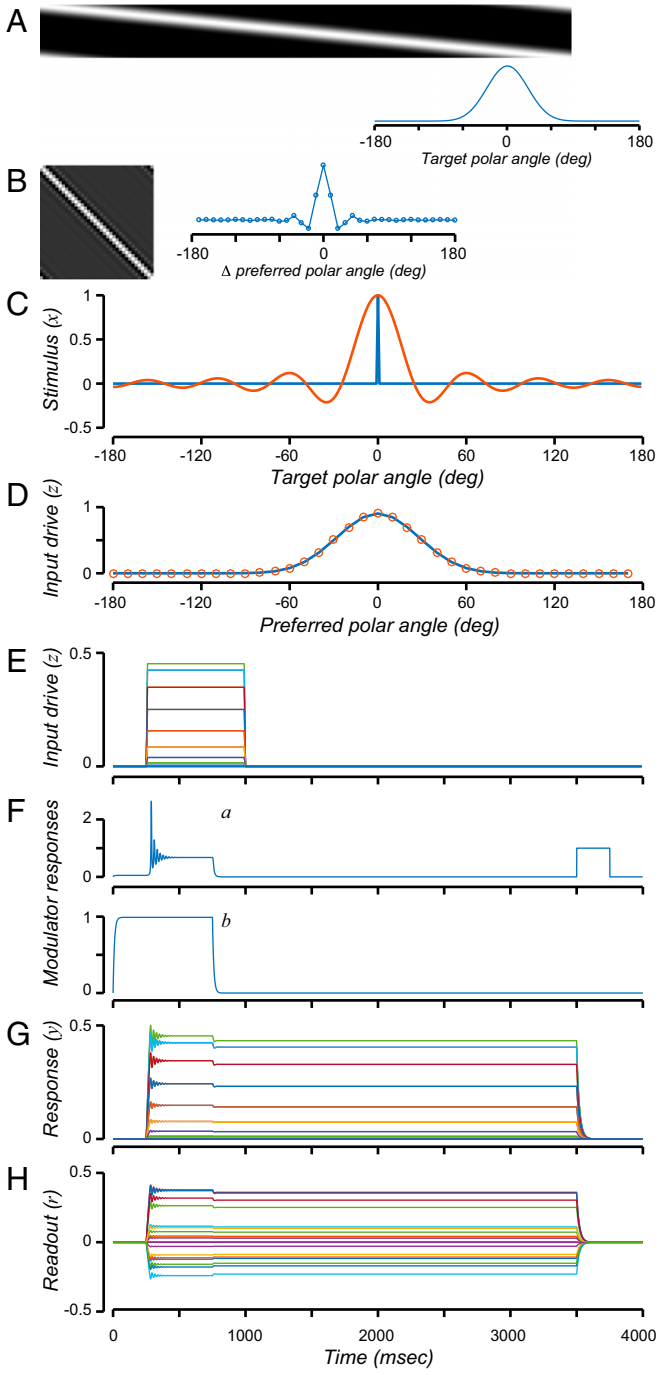

Fig. 2. Sustained activity. (*A*) Encoding matrix ($\mathbf{W}_{zx}$), each row of which corresponds to a neuron's response field. Graph, response field corresponding to the middle row of the matrix. (*B*) Recurrent weight matrix ($\mathbf{W}_{yy}$), each row of which corresponds to the recurrent synaptic weights from other neurons in the population. Graph, recurrent weights corresponding to the middle row of the matrix. (*C*) Input stimulus and reconstructed stimulus. Blue, input stimulus (**x**) corresponding to target location. Orange, reconstructed stimulus, computed as a weighted sum of the reconstructed input drive (*D*). (*D*) Input drive and reconstructed input drive. Blue, input drive (**z**) to each neuron as a function of that neuron's preferred target location. Orange, reconstructed input drive, computed as a weighted sum of the readout (*H*). (*E*) Input drive (**z**) over time. Each color corresponds to a different neuron. (*F*) Modulator responses. Top row, *a*. Bottom row, *b*. (*G*) Output responses (**y**). Each color corresponds to a different neuron. (*H*) Readout (**r**). Each color corresponds to a different component of the readout.

Downloaded at Bobst Library, New York University on November 6, 2019

the circuit reduced to a linear dynamical system, making it mathematically tractable.

Each neuron in the simulation responded selectively to target location, each with a different preferred polar angle (i.e., saccade direction) in the visual field (Figs. 1*B* and 2*A*), all with the same preferred radial position (i.e., saccade amplitude). We ignored saccade amplitude for this simulation, but it would be straightforward to replicate the circuit for each of several saccade amplitudes. The input drive $\mathbf{z}$ to each neuron, consequently, depended on target location and the time course of the target presentation (Fig. 2 *D* and *E*). The recurrent weights $\mathbf{W}_{\hat{y}y}$ were chosen to have a center-surround architecture; each row of $\mathbf{W}_{\hat{y}y}$ had a large positive value along the diagonal (self-excitation), flanked by smaller positive values, and surrounded by small negative values (Fig. 2*B*). All neurons in the circuit shared the same pair of modulators ($a_j = a$ and $b_j = b$), that is, all of the neurons had the same state at any given point in time. The input to the circuit comprised not only the target presentation but also the time courses of 2 cues, one of which indicated the beginning of the trial (at time 0 ms) and the other of which indicated the end of the delay period (at time 3,500 ms). The response time courses of the modulators followed the 2 cues (Fig. 2*F*), by setting appropriate values in the weight matrices $\mathbf{W}_{ax}$ and $\mathbf{W}_{bx}$.

This circuit was capable of maintaining a representation of target location during the delay period with sustained activity (Fig. 2*G*). The responses followed the input drive initially (compare Fig. 2 *E* and *G*) because the value of the input modulator was set to $b = 1$ (via $\mathbf{W}_{bx}$ in Eq. **2**) by the cue indicating the beginning of the trial. The value of b then switched to be small (= 0, corresponding to a long effective time constant) before the target was extinguished, so the output responses exhibited sustained activity (Fig. 2*G*). Finally, the value of the recurrent modulator was set to $a \approx 1$ (via $\mathbf{W}_{ax}$ in Eq. **2**) by the cue indicating the end of the trial, causing the output responses to be extinguished.

The dynamics of the responses, during the delay period, depended on the eigenvalues and eigenvectors of the recurrent weight matrix $\mathbf{W}_{\hat{y}y}$. In this particular example circuit, the recurrent weight matrix (Fig. 2*B*) was a symmetric 36 × 36 matrix ($n = 36$ was the number of neurons in the circuit, that is, each of $\mathbf{y}$ and $\mathbf{z}$ were 36-dimensional vectors). For this particular recurrent weight matrix, 19 of the eigenvalues were equal to 1, and the others had values less than 1. There is, of course, nothing special about these numbers; the circuit could include any number of neurons with any number of eigenvalues equal to 1, but providing these details makes it easier to visualize and understand. The critical issue is that the weight matrix was scaled so that the largest eigenvalues were equal to 1. (It is of course unrealistic for a biological circuit to have such precisely tuned synaptic weights but we show below that the circuit is robust with respect to the precise tuning because of the built-in normalization). The corresponding eigenvectors defined an orthonormal coordinate system (or basis) for the responses. The responses during the delay period (when $b = 0$) were determined entirely by the projection of the initial values (the responses at the very beginning of the delay period) onto the eigenvectors. Eigenvectors with corresponding eigenvalues equal to 1 were sustained throughout the delay period. Those with eigenvalues less than 1 decayed to zero (smaller eigenvalues decayed more quickly). Those with eigenvalues greater than 1 would have been unstable, growing without bound (which is why the weight matrix was scaled so that the largest eigenvalues = 1). This example circuit had a representational dimensionality $d = 19$, because the recurrent weight matrix had 19 eigenvalues = 1. The neural activity in this circuit was a 19-dimensional continuous attractor during the delay period. It could, in principle, maintain the locations and contrasts of up to 19 targets, or it could maintain a 19-dimensional pattern of inputs.

The input drive and target location were reconstructed from the responses, at any time during the delay period (Fig. 2 *C* and *D*). To do so, the responses were first multiplied by a readout matrix. The readout matrix $\mathbf{W}_{ry} = \mathbf{V}^t$ was a 19 × 36 matrix, where the rows of $\mathbf{V}^t$ were computed from the eigenvectors of the recurrent weight matrix $\mathbf{W}_{\hat{y}y}$. Specifically, $\mathbf{V}$ was an orthonormal basis for the 19-dimensional subspace spanned by the eigenvectors of $\mathbf{W}_{\hat{y}y}$ with corresponding eigenvalues = 1. The resulting readout (Fig. 2*H*), at any time point, was then multiplied by a decoding (or reconstruction) matrix (*SI Appendix*). The result was a perfect reconstruction of the input drive (Fig. 2*D*, orange) up to a scale factor (because of normalization), and an approximate reconstruction of the input stimulus (Fig. 2*C*, orange) with a peak at the target location. The reconstruction of the input stimulus was imperfect because the response fields were broadly tuned for polar angle. Regardless, we do not mean to imply that the brain attempts to reconstruct the stimulus from the responses. The reconstruction merely demonstrates that the responses and readout implicitly represent the target location. The encoding matrix $\mathbf{W}_{zx}$ was a 36 × 360 matrix ($M = 360$ was the number of polar angle samples in the input stimulus). The response fields (i.e., the rows of the encoding weight matrix $\mathbf{W}_{zx}$) were designed based on the same eigenvectors. Doing so guaranteed that the input drive was reconstructed perfectly from the responses at any time during the delay period (Fig. 2*D*; see *SI Appendix* for derivation).

Robustness via Normalization. The sustained activity circuit, as described above, depended on precisely tuned synaptic weights. The recurrent weight matrices were scaled so that the eigenvalues were no greater than 1. For a linear recurrent circuit with eigenvalues greater than 1, the responses are unstable, growing without bound during a delay period. This is a well-known problem for recurrent neural networks (52–54).

ORGaNICs solve this problem by incorporating normalization. The normalization model was initially developed to explain stimulus-evoked responses of neurons in primary visual cortex (V1) (50) but has since been applied to explain neural activity in a wide variety of neural systems (51). The model's defining characteristic is that the response of each neuron is divided by a factor that includes a weighted sum of activity of a pool of neurons. The model predicts and explains many well-documented physiological phenomena, as well as their behavioral and perceptual analogs.

The simulated neural circuits used the recurrent modulator a to provide normalization via feedback. The recurrent modulator determined the amount of recurrent gain; it was a particular nonlinear function of the responses: $f(\mathbf{y})$ in Eq. **2** (see *SI Appendix* for details). For an input drive $\mathbf{z}$ that was constant for a period of time, the output responses achieved a stable state in which they were normalized (see *SI Appendix* for derivation):

$$|y_j|^2 = \frac{|z_j|^2}{\sigma^2 + \sum |z_j|^2}. \qquad [3]$$

The responses were proportional to the input drive when the amplitude of the input drive was small (i.e., when the sum of the squared input drives was $\ll \sigma^2$). The responses saturated (i.e., leveled off) when the amplitude of the input drive was large ($\gg \sigma^2$). The value of σ (the semisaturation constant) determined the input drive amplitude that achieved half the maximum response. Despite saturation, the relative responses were maintained (see *SI Appendix* for derivation):

$$\frac{|y_j|^2}{\sum |y_j|^2} = \frac{|z_j|^2}{\sum |z_j|^2}. \qquad [4]$$

That is, the normalized responses represented a ratio between the input drive to an individual neuron and the amplitude of the input drive summed across all of the neurons. Consequently, the responses of all neurons saturated together (at the same input

Downloaded at Bobst Library, New York University on November 6, 2019

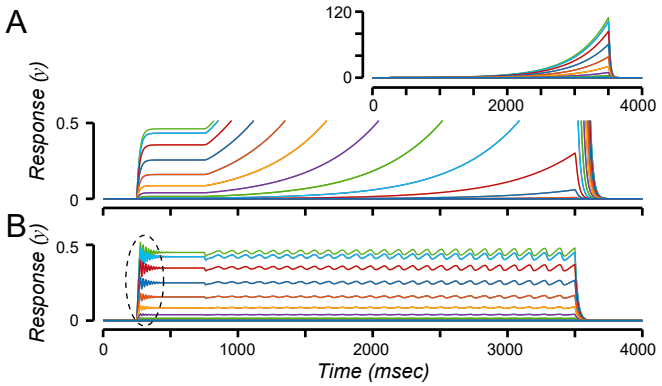

Fig. 3. Normalization. (*A*) Output responses (**y**), corresponding to the sustained activity circuit depicted in Figs. 1 and 2, but with the recurrent weight matrix scaled by a factor of 1.02. Each color corresponds to a different neuron. (*Inset*) Full range of responses on an expanded (240×) ordinate. (*B*) Output responses with normalization. Dashed oval, high frequency, coherent, synchronized oscillations following target onset.

drive amplitude) even though some neurons responded strongly to the input whereas others did not.

Recurrent normalization made the circuit robust with respect to imperfections in the recurrent weight matrix (Fig. 3). Without normalization, responses depended critically on fine tuning. For example, we used the sustained activity circuit (Figs. 1 and 2), but with $f(\mathbf{y}) = 0$ so that normalization was disabled, and we scaled the recurrent weight matrix by a factor of 1.02. The responses were unstable, growing without bound (Fig. 3*A*). Including normalization automatically stabilized the activity of the circuit (Fig. 3*B*). The increases in activity evoked by the recurrent weight matrix (with largest eigenvalues = 1.02) were countered by normalization such that the total activity in the circuit was roughly constant over time ($||\mathbf{y}||^2 \sim 1$). The ratios of the responses were maintained (Eq. **4**), enabling an accurate readout, throughout the delay period. Analogous results were obtained with the other example circuits described below, including those that exhibited oscillatory and sequential dynamics, because the normalization depends on the squared norm of the responses, which was constant over time during the delay period for each of these example circuits. The stability of the normalized responses did not depend on fine-tuning any of the other synaptic weights in the circuit; perturbing those synaptic weights by random values within ±5% yielded virtually identical simulated responses and the responses were stable even when those synaptic weights were perturbed by random values ranging from 0.5× to 2× (see *SI Appendix* for details). We have also implemented a generalization of this recurrent normalization circuit in which each neuron's response can be normalized by an arbitrary (nonnegative) weighted sum of the other neurons in the circuit.

The normalized responses exhibited high-frequency oscillations following target onset that were synchronized across all of the neurons in the circuit (Fig. 3*B*, dashed oval). There are 2 nested recurrent circuits in ORGaNICs: 1) the recurrent drive and 2) the multiplicative modulators. The high-frequency oscillations emerged because of the inherent delay in the second of these recurrent circuits (i.e., because of the multiplicative modulator underlying normalization). The oscillation frequency depended on the membrane time constants. For the time constants used for Fig. 3, the responses exhibited oscillations in the gamma frequency range. Different intrinsic time constants yielded different oscillation frequencies. The oscillation frequency would have depended also on axon length if we were to include conduction delays.

The responses exhibited lower-frequency oscillations during the delay period (Fig. 3*B*). These lower-frequency oscillations emerged because of the recurrent drive in combination with normalization; the recurrent weight matrix was scaled to have eigenvalues greater than 1, which drove the responses to increase over time, but this increase was countered by normalization. These oscillations were synchronized so the ratios of the responses were maintained (Eq. **4**), enabling an accurate readout, despite the oscillations.

Sequential Activity. ORGaNICs can be used to generate delay-period activity with complex dynamics, including sequential activity and traveling waves of activity, in addition to sustained activity, and the same theoretical framework was used to analyze them. The key idea is that the recurrent weight matrix can have complex-valued eigenvectors and eigenvalues. One way for this to happen is when the recurrent weights and output responses are complex-valued (*SI Appendix*, Fig. S4). The complex-number notation is just a notational convenience (*SI Appendix*). Another way to generate complex dynamics is for the recurrent weight matrix to be real-valued but asymmetric, such that the responses are real-valued but the eigenvectors are eigenvalues are complex-valued.

One such example circuit was designed to generate sequential activity (Fig. 4). In this example circuit, there were again 36 neurons with the same response fields as in the preceding example (Fig. 2*A*). The modulators were also the same as in the preceding example, including recurrent normalization. The recurrent weight matrix was real-valued but asymmetric (Fig. 4*A*).

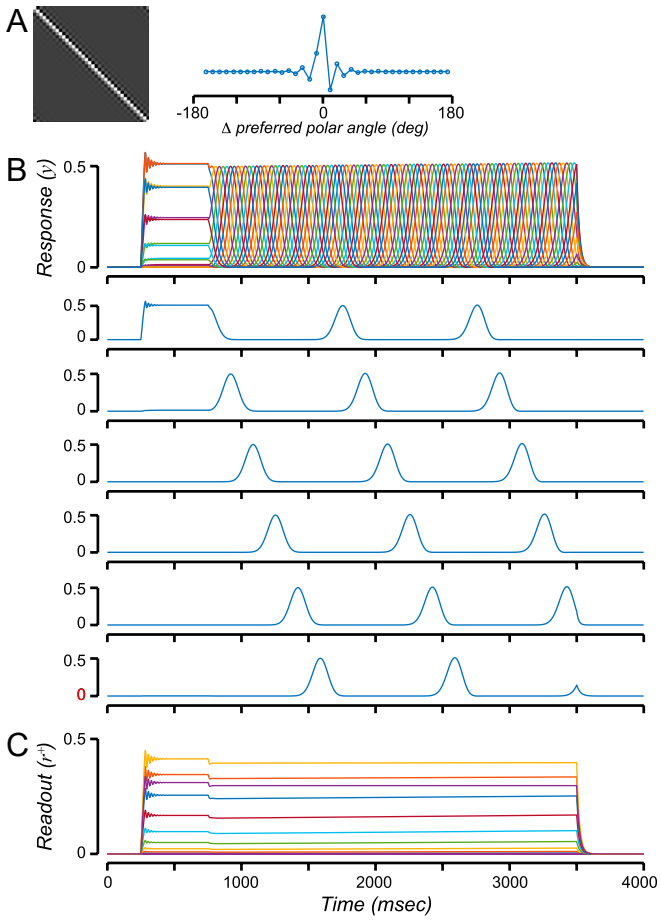

Fig. 4. Sequential activity. (*A*) Recurrent weight matrix ($\mathbf{W}_{yy}$). Graph, recurrent weights corresponding to the middle row of the matrix. (*B*) Output responses (**y**). Each color corresponds to a different neuron. Successive rows, responses of a few example neurons. (*C*) Readout ($\mathbf{r}^+$). Each color corresponds to a different component of the readout.

Downloaded at Bobst Library, New York University on November 6, 2019

Because of the asymmetry, the eigenvectors and eigenvalues of the recurrent weight matrix were complex-valued, and the output responses exhibited oscillatory dynamics (Fig. 4*B*). The recurrent weight matrix was designed so that the recurrent connectivity depended on the spatial derivative of the neural activity (55), that is, the difference in activity between nearby neurons (*SI Appendix*). Consequently, the activity was handed off from one neuron to the next during the delay period, analogous to a synfire chain (56–59), but with activity that continuously tiled time (60).

Despite the complex dynamics, the readout was constant over time (Fig. 4*C*). The readout matrix was again, as for the preceding sustained activity circuit (Fig. 2), computed as a unitary basis for the subspace spanned by the eigenvectors of $\mathbf{W}_{yy}$ with corresponding eigenvalues that had real parts = 1. However, the readout was computed as $\mathbf{r}^{+} = |\mathbf{W}_{ry}\,\mathbf{y}|$, that is, the modulus (square root of the sum of squares of real and imaginary parts) of a weighted sum of the responses. Consequently, this circuit was capable of maintaining some (but not all) information about the input during the delay period. Unlike the preceding example, it was not possible to reconstruct the input drive from the readout at arbitrary points in time during the delay period. A linear reconstruction (like that used for the preceding example) generated a copy of the input drive that shifted over time like a traveling wave (*SI Appendix*, Fig. S5). That is, the information maintained during the delay period was sufficient for discriminating some inputs (e.g., 2 targets with different contrasts or 2 pairs of targets with different spacings) but incapable of discriminating between other inputs (e.g., a single target of the same contrast presented at 2 different locations).

Motor Preparation and Motor Control. ORGaNICs are also capable of generating signals, like those needed to execute a complex sequence of movements (e.g., speech, bird song, backside double McTwist 1260 on a snowboard out of the halfpipe). Some actions are ballistic (open loop), meaning that they are executed with no sensory feedback during the movement. Others are closed loop, meaning that the movements are adjusted on the fly based on sensory feedback. ORGaNICs evoke patterns of activity over time that may underlie the execution of both open- and closed-loop movements.

An example of open-loop control (Fig. 5) was implemented using the sequential activity circuit described above, but with a different readout. The encoding matrix and the recurrent matrix were identical to those in the sequential activity circuit. The modulators were also the same as in the preceding examples, including recurrent normalization. The readout was different, simply summing the components, $r^{\Sigma} = \boldsymbol{\Sigma}\ \mathrm{Re}(\mathbf{W}_{ry}\,\mathbf{y})$. Different spatial patterns of inputs led to different temporal dynamics of the responses. When the input was chosen to drive a particular eigenvector (i.e., because the input drive was orthogonal to the other eigenvectors), then the readout during the period of motor execution (same as the delay period in the preceding example circuits) was a 1-Hz sinusoid (Fig. 5*A*). When the input was chosen to drive another eigenvector, then the readout was an 8-Hz sinusoid (Fig. 5*C*). A linear sum of these inputs evoked a readout that was proportional (because of normalization) to the linear sum of the readouts (Fig. 5*D*).

How are these temporal profiles of activity generated? Each eigenvector of the recurrent weight matrix is associated with a basis function, a pattern of activity across the population of neurons and over time. Each basis function is a complex exponential (i.e., comprising sine and cosine), the frequency of which is specified by the imaginary part of the corresponding eigenvalue:

$$\omega_i = \left(\frac{1{,}000}{2\pi\tau_y}\right)\mathrm{Im}(\lambda_i). \qquad [5]$$

The value of λ_i is the imaginary part of the i^{th} eigenvalue of the recurrent weight matrix, and ω_i is the corresponding oscillation frequency (in hertz). The factor of 1,000 is needed because the time constant τ_y is presumed to be specified in milliseconds but the oscillation frequency is specified in hertz (cycles per second). The responses exhibit an oscillating traveling wave (Fig. 5*B*); the response of any individual neuron oscillates over time and the entire pattern of activity across the population of neurons shifts over time (Fig. 5*B*, orange – yellow – purple – green – cyan – red). For inputs corresponding to different eigenvectors, the responses

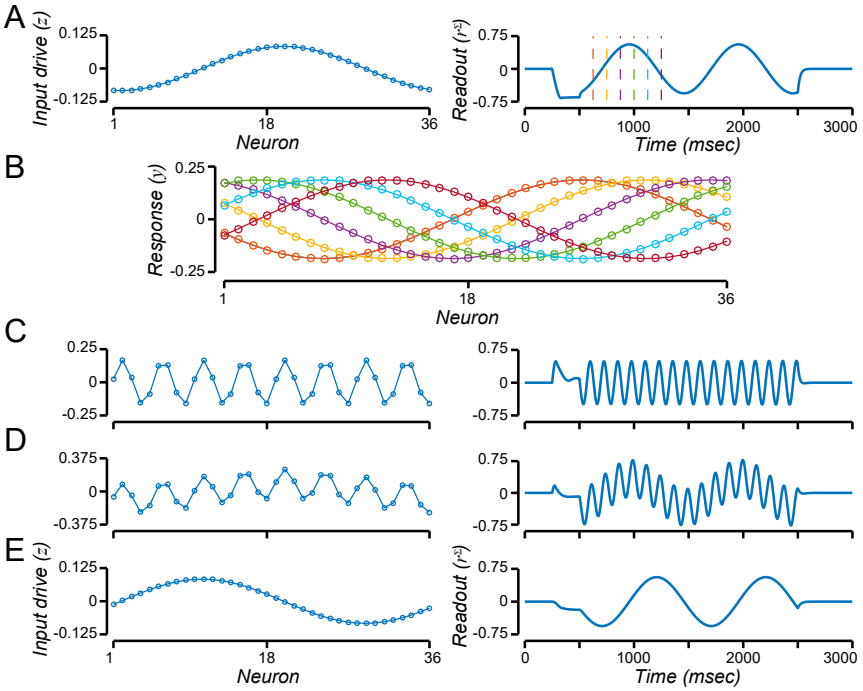

Fig. 5. Motor preparation and motor control. (*A*) Input drive and readout corresponding to input that drives only the 1-Hz component of the recurrent weight matrix. (*A*, *Left*) Input drive ($\mathbf{z}$), spatial pattern activity across the 36 neurons during the premotor time period (250 to 500 ms). (*A*, *Right*) Readout ($\mathbf{r}^{\Sigma}$) over time. Vertical dashed lines, times corresponding to curves in *B*. (*B*) Responses exhibit an oscillating traveling wave of activity. Different colors correspond to different time points, indicated in *A*. (*C*) Input drive and readout corresponding to the 8-Hz component of the recurrent weight matrix. Same format as *A*. (*D*) Summing the inputs from *A* and *C* evokes the sum of the responses. (*E*) Input drive from *A* is shifted in space, generating a readout that is shifted in time.

Downloaded at Bobst Library, New York University on November 6, 2019

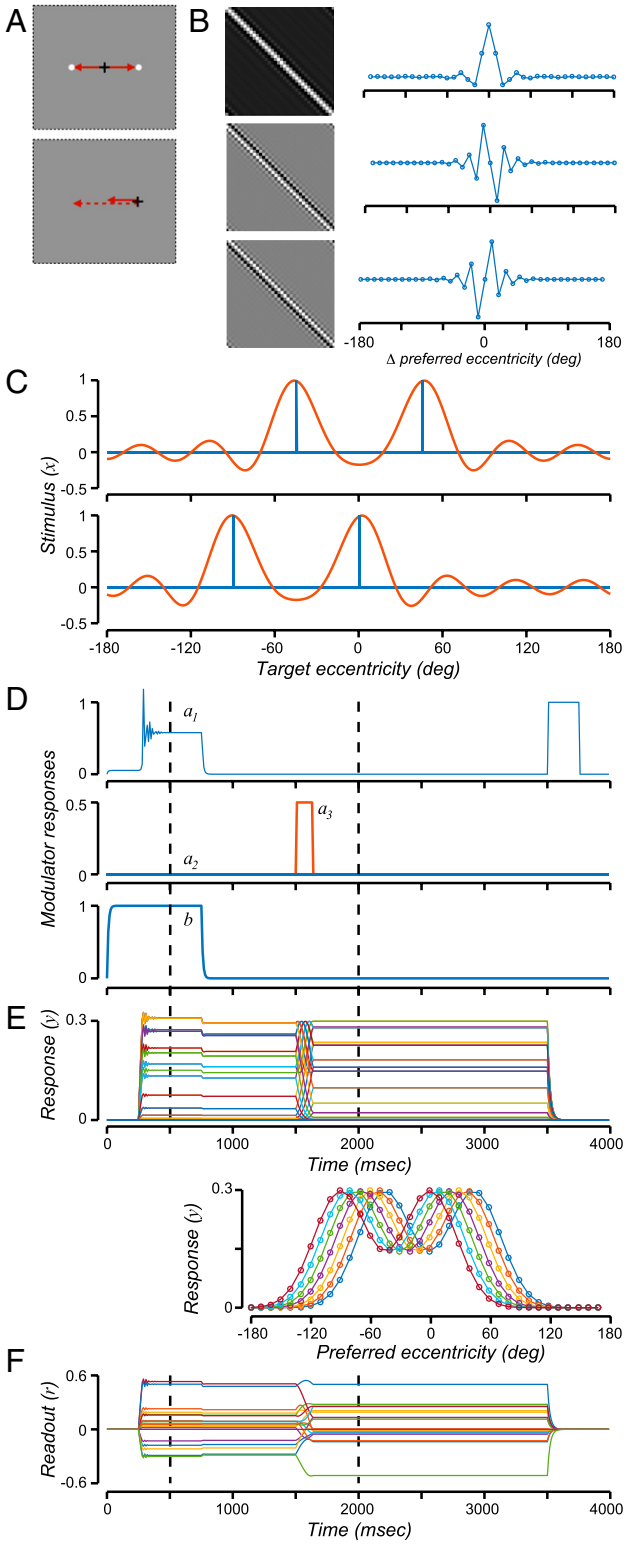

Fig. 6. Spatial updating. (*A*) Double-step saccade task. (*A*, *Top*) Targets presented. (*A*, *Bottom*) After eye movement to target 1. White dots, targets. Black cross-hairs, eye position. Solid red lines, planned eye movements without updating. Dashed red line, planned eye movement after updating. (*B*) Recurrent weight matrices. (*Top*) Recurrent weight matrix corresponding to modulator a_1 for maintaining a representation of the target locations. (*Middle* and *Bottom*) Recurrent weight matrices corresponding to modulators a_2 and a_3 for updating the representation with leftward and rightward eye movements, respectively. (*C*) Input stimulus and reconstructed stimulus. Blue, input stimulus (**x**) corresponding to the 2 target positions. Orange, reconstructed stimulus, computed as a weighted sum of the readout (*F*). (*Top*) Before eye movement to target 1. (*Bottom*) After eye movement to target 1. (*D*) Modulator responses. (*Top*) a_1. (*Middle*) a_2 (blue) and a_3 (orange). (*Bottom*) *b*. (*E*) Output responses (**y**). (*Top*) Time course of activity, with different colors corresponding to different neurons. (*Bottom*) Responses for each of several time points (different colors correspond to different time points) while updating the neural representation of the target locations. (*F*) Readout (**r**). Dashed vertical lines in *D*–*F* correspond to the snapshots in *A*.

oscillate at correspondingly different frequencies (Fig. 5*C*). The frequencies of the various components corresponding to each of the eigenvalues, for this particular recurrent weight matrix, included a number of other frequencies in addition to the 1- and 8-Hz components shown in the figure. Motor control signals with any arbitrary phase, for each of the frequency components, can be generated by shifting the input drive (Fig. 5*E*). That way, all combinations of amplitudes, frequencies, and phases can be generated just by changing the spatial pattern of premotor activity, with a fixed, linear readout. This dovetails with experimental evidence demonstrating that the function of motor preparation is to set the initial conditions that generate the desired movement (61–63), and that complex movements are based on a library of motor primitives (64, 65).

The readout for open-loop control is, in general, a linear sum of the responses r^{Σ}. The readout matrix for short-term memory, in the preceding sustained activity circuit (Fig. 2), comprised eigenvectors of the recurrent weight matrix to ensure that the input was recovered during the delay period. However, recovering the input is not the goal for open-loop control. Rather, a sum of the (co)sinusoidal basis functions was used to generate motor control signals for ballistic (open-loop) movements.

ORGaNICs may also generate more complicated control signals. The basis functions are damped oscillators when the modulators are greater than 0 but equal to one another ($a = b$) and constant over time, and when the input is constant over time. If the input is time-varying, then the responses depend on a linear combination of the inputs and the basis functions, and the responses may be used for closed-loop control. If the modulators are also time-varying, and different for each neuron, then the responses may exhibit a wide range of dynamics, with the capability (by analogy with LSTMs) of solving relatively sophisticated tasks (see Introduction for references).

Manipulation: Spatial Updating. A simulation of the double-step saccade task illustrates how ORGaNICs can both maintain and manipulate information over time (Fig. 6). In this task, 2 targets are shown while a subject is fixating the center of a screen (Fig. 6 *A*, *Upper*). A pair of eye movements are then made in sequence to each of the 2 targets. Eye movements are represented in the brain using retinotopic, that is, eye-centered, coordinates (Fig. 6 *A*, *Upper*, red lines). Consequently, after making the first eye movement, the plan for the second eye movement must be updated (Fig. 6 *A*, *Lower*; the solid red line copied from the upper panel no longer points to the second target). This is done by combining a representation of the target location with a copy of the neural signals that control the eye muscles (i.e., corollary discharge) to update the planned eye movement (Fig. 6 *A*, *Lower*, dashed red line).

The example circuit in Fig. 6 received 2 types of inputs: 1) the target locations at the beginning of the trial (Fig. 6 *C*, *Top*, blue) and 2) a corollary discharge of the impending eye movement. The targets were assumed to be along the horizontal meridian of the visual field. There were again 36 neurons, but unlike the preceding examples each neuron responded selectively to a different eccentricity along the horizontal meridian of the visual field (i.e., degrees of visual angle away from fixation), not different polar angles around fixation at a fixed eccentricity. The encoding matrix $\mathbf{W}_{zx}$ was analogous to that in the preceding examples, but the neurons were selective for target eccentricity

Downloaded at Bobst Library, New York University on November 6, 2019

instead of polar angle. Readout and reconstruction were the same as for the sustained activity circuit (Fig. 2).

What distinguishes this example circuit from the preceding examples is that there were 3 recurrent weight matrices (Fig. 6*B*), the first for maintaining a representation of the target locations (Fig. 6 *B*, *Top*), the second for changing the representation with leftward eye movements (Fig. 6 *B*, *Middle*), and the third for changing the representation with rightward eye movements (Fig. 6 *B*, *Bottom*). As in the preceding examples, the modulators were the same for each neuron in the circuit. Consequently, we can modify Eq. **1**:

$$\tau_y \frac{d\mathbf{y}}{dt} = -\mathbf{y} + \left(\frac{b^+}{1+b^+}\right)\mathbf{z} + \left(\frac{1}{1+a_1^+}\right)\hat{\mathbf{y}}_1 + \left(\frac{a_2^+}{1+a_2^+}\right)\hat{\mathbf{y}}_2 + \left(\frac{a_3^+}{1+a_3^+}\right)\hat{\mathbf{y}}_3, \quad [6]$$

$$\hat{\mathbf{y}}_k = \mathbf{W}_{\hat{y}_k \mathbf{y}}\mathbf{y},$$

where the subscript k indexes over the 3 recurrent weight matrices. The first recurrent weight matrix was identical to that in the sustained activity circuit (Fig. 2*B*). The second recurrent weight matrix was a discrete approximation to the derivative of the responses (*SI Appendix*), and third was the negative derivative matrix (i.e., the second and third recurrent matrices differed from one another by a factor of −1). To accommodate 2 dimensions of eye movements, the input drive would depend on 2-dimensional response fields tiling the visual field, and the recurrent drive would depend on 5 recurrent weight matrices, one to maintain the current eye position, a pair for the horizontal component of movements, and another pair for the vertical component (or likewise a pair for the polar angle component of movements and another pair for the radial component).

The modulators were used to encode and update a representation of the target locations (Fig. 6*D*). As in the preceding examples, the responses followed the input drive at the beginning of the simulated trial because the input modulator was set to $b = 1$ (via $\mathbf{W}_{bx}$ in Eq. **2**) by the cue indicating the beginning of the trial. The value of b then switched to be small (= 0) before the targets were extinguished, so the output responses exhibited sustained activity that represented the original target locations (Fig. 6 *C*, *Top*, orange). The modulator a_1 was responsible for recurrent normalization, as in the preceding example circuits. The modulator a_3 was nonzero for a period of time beginning just prior to the eye movement (Fig. 6 *D*, *Middle*, orange). The amplitude of a_3 and duration of time during which it was nonzero determined the magnitude of updating, that is, corresponding to the amplitude of the impending saccade (for an eye movement in the opposite direction, the amplitude of a_2, instead of a_3, would have been nonzero). Finally, the value of the recurrent modulator was set to $a_1 \approx 1$ (via $\mathbf{W}_{ax}$ in Eq. **2**) by the cue indicating the end of the trial, causing the output responses to be extinguished.

The output responses exhibited a traveling wave of activity across the topographic map of target locations during the period of time when the neural representation of the targets was updated (Fig. 6*E*). The readout (Fig. 6*F*) encoded the 2 target locations, both before and after updating. The readout and decoding matrices were identical to those in the sustained activity circuit (Fig. 2). Preceding the eye movement, the original target locations were reconstructed from the readout (Fig. 6 *C*, *Top*, orange curve). After the eye movement, the updated target locations were reconstructed (Fig. 6 *C*, *Bottom*).

Manipulation: Time Warping and Time Reversal. A challenge for models of motor control is to generate movements at different speeds, for example playing a piece of piano music, generating speech (66), or generating birdsong (67) at different tempos. Likewise, a challenge for models of sensory processing is that perception must be tolerant with respect to compression or dilation of temporal signals, for example listening to fast vs. slow speech (68). A possible mechanism for time warping is to scale the time constants of the neurons (69), all by the same factor, which scales the oscillation frequencies by the inverse of that scale factor (Eq. **5**). A fixed value for the scale factor would handle linear time rescaling in which the entire input (and/or output) signal is compressed or dilated accordingly. A neural circuit might compute a time-varying value for the scale factor, based on the inputs and/or outputs, to handle time-varying time warping.

Here, we offer a different mechanism for time warping (also time reversal), making use of the modulators. An example open-loop motor control circuit was implemented that enabled time warping and time reversal (Fig. 7). The encoding matrix and the recurrent matrix were identical to those in the spatial updating example (Fig. 6). The a_1 and b modulators were also the same as in the spatial updating example, but the time courses of the other 2 modulators a_2 and a_3 were different (Fig. 7*A*). The readout was the same as that in the motor control circuit (Fig. 5), summing across the components r^{Σ}. The input was chosen to drive all of the eigenvectors with randomly chosen amplitudes and phases. Different values of the a_2 and a_3 modulators generated control signals that were time-warped and/or time-reversed. Increasing the modulator response from 1 to 5/3 caused the readout to increase in tempo by 25% (compare Fig. 7 *B* and *C*); tempo was proportional to $a_2/(1 + a_2)$. A time-varying modulator generated time-varying time warping. The circuit exhibited these phenomena because the responses exhibited oscillating traveling waves (Fig. 5*B*). The readout was a sum of these traveling waves, and the speed of the traveling waves was controlled by the modulators (*SI Appendix*). When a_3 (instead of a_2) was nonzero, the readout was time reversed (compare Fig. 7 *B* and *D*) because the traveling waves of activity moved in the opposite direction.

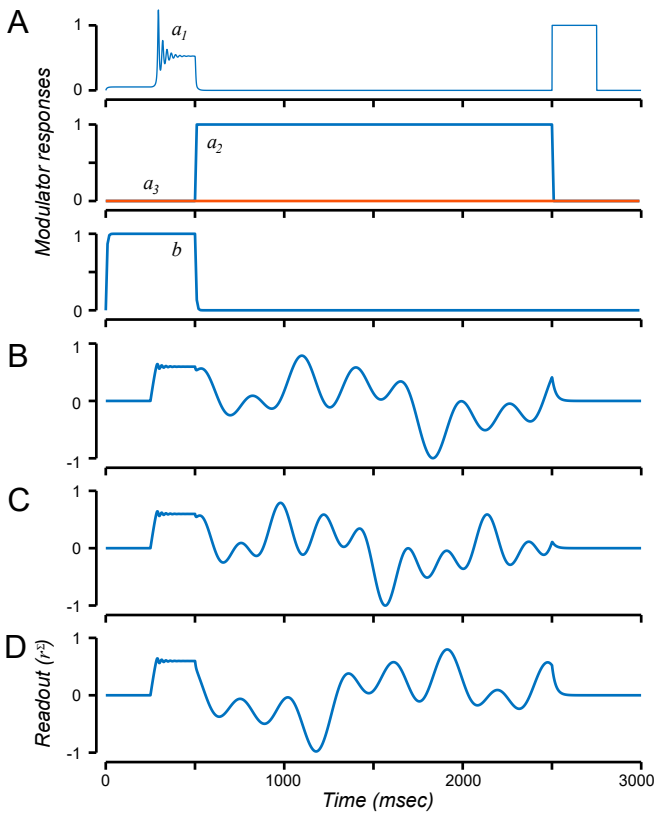

Fig. 7. Time warping and time reversal. (*A*) Modulator responses. (*B*) Readout for $a_2 = 1$ and $a_3 = 0$. (*C*) Time-warped readout for $a_2 = 5/3$ and $a_3 = 0$. (*D*) Time-reversed readout for $a_2 = 0$ and $a_3 = 1$.

Downloaded at Bobst Library, New York University on November 6, 2019

Discussion

We developed a theoretical framework for neural dynamics called ORGaNICs and applied it to simulate key phenomena of working memory and motor control. We demonstrated the following results. 1) Working memory: ORGaNICs can simulate delay-period activity with complex dynamics, including sequential activity and traveling waves of activity, to maintain and manipulate information over time. Derivative-like recurrent connectivity, in particular, generated traveling waves of activity. We propose that these traveling waves play a role in circuit function to manipulate and update internal models. 2) Motor control: The exact same circuits (with the same synaptic weights) were used to generate signals with complex motor dynamics, by converting spatial patterns of premotor activity to temporal profiles of motor control activity. Different spatial patterns of premotor activity evoked different motor control dynamics. These circuits were controlled to manipulate (e.g., time-warp) the motor dynamics. 3) Normalization: Recurrent normalization, via the recurrent modulator, ensured stability over time and robustness with respect to perturbations of synaptic weights. 4) Mechanism: ORGaNICs can be implemented with a simplified biophysical (equivalent electrical circuit) model of pyramidal cells (see *SI Appendix* and ref. 46). There is considerable flexibility in the formulation of ORGaNICs, with different variants corresponding to different hypothesized neural circuits (*SI Appendix*). We demonstrated all of the above results with 2 circuits; the first circuit generated the simulation results in Figs. 2–5 and the second one generated Figs. 6 and 7, noting that the first circuit is equivalent to a special case of the second one.

Because they are generalizations of LSTMs, ORGaNICs can solve tasks that are much more sophisticated than the typical delayed-response tasks used in most cognitive psychology and neuroscience experiments. Indeed, although this is not an ML paper, we note that ORGaNICs may offer computational advantages compared to varieties of LSTMs that are commonly used in ML applications (see *SI Appendix* and ref. 46).

This theoretical framework, of course, includes components previously proposed in the computational/theoretical neuroscience literature, and the ML literature, that have achieved some of the same goals (70–87). However, with ORGaNICs we show that a single unified circuit architecture captures key neurophysiological phenomena associated with sensory, cognitive, and motor functions, each of which has been modeled separately in the previously published literature. Unlike linear recurrent neural networks, the modulators in ORGaNICs introduce nonlinearities (analogous to the gates in LSTMs) that can perform multiple functions including handling long-term dependencies and providing robustness via normalization (discussed below). Unlike most nonlinear recurrent neural nets, ORGaNICs are mathematically tractable, making it possible to derive concrete, quantitative predictions that can be fit to experimental measurements. The theory is tractable when the modulators are constant, that is, during each successive phase of a behavioral task. In addition, the responses of the normalization circuit follow the normalization equation (Eq. **3**) exactly, so that this circuit makes predictions that are identical to those of the normalization model, thereby preserving all of the desirable features of that model, which has been fit to hundreds of experimental datasets. In classic work on neural fields (88–90), by contrast, diverse patterns of activity are accomplished by biasing a nonlinear network to different operating points, each having a different solution that can be approximated by local linearization. Here, we start with a linear dynamical system that is fully tractable, characterized by the eigenvalues and eigenvectors of the linear system, but also limited to only those patterns of activity that can be expressed as linear sums of the eigenvectors. We circumnavigate this limitation with the modulators that shape solutions to dynamically change the eigenstructure of the linear system; for each choice of values for the modulators, we have a different linear system. Unlike black-box ML approaches, ORGaNICs provide insight; for example, we understand exactly when and how it is possible to reconstruct an input by reading out the responses during the delay period of a working memory task and how to generate motor control signals with complex dynamics (see *SI Appendix* for derivations). ML algorithms are particularly useful for computing solutions to optimization problems (e.g., model fitting via gradient descent), and we plan to use ML implementations of ORGaNICs to fit experimental data. ML approaches can also provide inspiration for neuroscience theories (and vice versa), like the links presented here between ORGaNICs and LSTMs. Left open in the current paper is how the weights in the various weight matrices emerge through development and/or learning. We engineered the weights to demonstrate the computational capabilities of this theoretical framework and to illustrate that the theory can reproduce neurobiological phenomena (although ORGaNICs are compatible with modified versions of ML algorithms; see *SI Appendix*). Some of the previously published literature (cited above) focuses on learning. However, having the right circuit architecture is a prerequisite for developing an accurate model of learning.

We propose that ORGaNICs can serve as a unifying theoretical framework for neural dynamics, a canonical computational motif based on recurrent amplification, gated integration, reset, and controlling the effective time constant. Rethinking cortical computation in these terms should have widespread implications, some of which are elucidated in the paragraphs that follow (see also *SI Appendix*).

Sustained delay-period activity and sequential activity are opposite sides of the same coin. ORGaNICs, a straightforward extension of leaky neural integrators and neural oscillators, provide a unified theoretical framework for sustained activity (Fig. 2), oscillatory activity (*SI Appendix*, Fig. S4), and sequential activity (Fig. 4), just by changing the recurrent weight matrix. Indeed, ORGaNICs can switch between these different behaviors. The spatial updating circuit, for example, exhibits sustained activity during the delay periods and sequential activity coincident with the eye movement (Fig. 6). The modulators a_2 and a_3 do the job of toggling between sustained and sequential. We assert that complicated dynamics is the norm, to support manipulation as well as maintenance (e.g., Fig. 6).

ORGaNICs can be used to generate motor control signals, with the very same circuits used to model working memory, just by changing the readout. The circuits convert spatial patterns of input (premotor) activity to temporal profiles of output (motor control) activity. Different spatial patterns of premotor activity evoke motor control outputs with different temporal response dynamics (e.g., as in Figs. 5 and 7), and the modulators provide a means for manipulating (time warping and time reversal) the dynamics (Fig. 7).

ORGaNICs are applicable also to models of sensory integration (e.g., integrating corollary discharge in Fig. 6) and sensory processing (e.g., with normalization as in Fig. 3). ORGaNICs may be stacked in layers such that the inputs to one ORGaNIC are the outputs from one or more other ORGaNICs. Particular stacked architectures encompass convolutional neural nets (i.e., deep nets) as a special case: specifically, when the encoding/embedding weight matrices are convolutional and when the modulators are large ($a_j = b_j \gg 0$) such that the output responses from each layer are dominated by the input drive to that layer. Consequently, working memory, motor control, sensory processing (including prediction over time; see *SI Appendix* and ref. 46), and possibly other cognitive functions (in addition to working memory, such as cognitive control, for example controlling attention) may all share a common canonical computational foundation.

Derivative-like recurrent connectivity (55) simulates sequential activity and traveling waves of activity (Figs. 4–7), and we propose that these traveling waves play a particular role in circuit function. Weight matrices with derivative-like weights are a mainstay of feed-forward models of sensory processing (91, 92), but the contribution of derivative-like weights in recurrent connectivity has been underappreciated. Traveling waves are ubiquitous in cortical activity, but their functional role has remained a mystery (93). We used recurrent weight matrices based on derivatives (i.e., the difference in activity between nearby neurons) to evoke traveling waves of activity that functioned to support manipulation. The traveling

Downloaded at Bobst Library, New York University on November 6, 2019

waves served to transform spatial patterns of premotor activity to temporal patterns of motor control activity (Figs. 5 and 7) or to update internal models (working memory representations) whether or not there was an overt movement (Fig. 6).

Why do some neural circuits exhibit sustained activity while others exhibit sequential activity, and what are the relative advantages or disadvantages of each? Sustained activity circuits are useful for short-term memory (i.e., maintenance), but not for other cognitive functions that require manipulation and control. For sustained-activity circuits, a simple linear readout of the responses can be used to reconstruct the input drive (and to approximately reconstruct the input stimulus), at any point in time during a delay period (Fig. 2). In addition, sustained-activity circuits are likely to be more robust than sequential-activity circuits, because all of the components share the same dynamics. Sequential-activity circuits, on the other hand, offer much more flexibility. The same circuit, with the same fixed recurrent weight matrix and the same fixed encoding matrix, can support multiple different functions just by changing the readout. For example, the sequential-activity circuit (Fig. 4) and the motor-control circuit (Fig. 5) were identical except for the readout. For the sequential-activity circuit (Fig. 4), a (nonlinear) modulus readout generated an output that was constant over time (i.e., to support maintenance). For the motor-control circuit (Fig. 5), a linear readout was used to generate control signals as sums of (co)sinusoidal basis functions with various different frequencies and phases. Likewise, the spatial-updating circuit (Fig. 6) and the time-warping/time-reversal circuit (Fig. 7) were identical. This circuit can be used to perform working memory (maintenance and manipulation), and the same circuit (without changing the encoding or recurrent weights) can be used to execute movements with complex dynamics. One way to implement this, for example, would be to have 2 different brain areas with stereotypical intrinsic circuitry (i.e., identical recurrent weights) that support 2 different functions with different readouts. Indeed, there is experimental evidence that different brain areas support different functions with similar circuits, for example parietal areas underlying working memory maintenance and PFC areas underlying motor planning (94). Alternatively, the output from a single circuit could innervate 2 different brain areas, one of which performs the first readout and the other of which performs the second readout, or a single brain area might switch between 2 different readouts (e.g., using a gating mechanism analogous to the modulators in ORGaNICs), corresponding to different behavioral states, without changing the intrinsic connectivity within the circuit. This makes biological sense. Rather than having to change everything (the encoding matrix, the recurrent matrix, the modulators, and the readout), you need only change one thing (the readout matrix) to enable a wide variety of functions. This is not possible with recurrent weight matrices that exhibit sustained activity, simply because there is only a single mode of dynamics (constant over time).

The modulators perform multiple functions and can be implemented with a variety of circuit, cellular, and synaptic mechanisms. The time-varying values of the modulators determine the state of the circuit by controlling the recurrent gain and effective time constant of each neuron in the circuit. The multiple functions of the modulators include normalization (Fig. 3), maintenance (Figs. 2–7), controlling pattern generators (Figs. 5 and 7), gated integration/updating (Fig. 6), time warping and time reversal (Fig. 7), reset (Figs. 2–7), controlling the effective time constant (*SI Appendix*, Fig. S1), controlling the relative contributions of bottom-up versus top-down connections (95), representing and weighting the reliability of sensory evidence (likelihood) and internal model (prior, expectation) for inference, prediction over time, and multisensory integration (95). ORGaNICs may have multiple recurrent weight matrices, each multiplied by different recurrent modulators, to perform combinations of these functions (Eq. **6** and Figs. 6 and 7). Some of the modulator functions need to be fast and selective (e.g., normalization), likely implemented in local circuits. A variety of mechanisms have been hypothesized for adjusting the gain of local circuits (96–98). Some modulator functions might depend on thalamocortical loops (20, 99–101). Other modulator functions are relatively nonselective and evolve relatively slowly over time and may be implemented with neuromodulators (102–105).

Recurrent normalization, as implemented with ORGaNICs (Fig. 3), is consonant with the idea that normalization operates via recurrent amplification, that is, that weak inputs are strongly amplified but that strong inputs are only weakly amplified. Several hypotheses for the recurrent circuits underlying normalization have been proposed (50, 51, 96, 106–108), but most of them are inconsistent with experimental observations suggesting that normalization is implemented via recurrent amplification (109–114). ORGaNICs offer a family of dynamical systems models of normalization, each of which comprises coupled neural integrators to implement normalization via recurrent amplification (*SI Appendix*). When the input drive is constant over time, the circuit achieves an asymptotic stable state in which the output responses follow the normalization equation exactly (Eq. **3**).

There is a critical need for developing behavioral tasks that animal models are capable of learning, and that involve both maintaining and manipulating information over time. ORGaNICs (and LSTMs) manage long-term dependencies between sensory inputs at different times, using a combination of gated integration and reset. Typical delayed-response tasks like the memory-guided saccade task are appropriate for studying what psychologists call "short-term memory," but they are weak probes for studying working memory (115–118), because those tasks do not involve manipulation of information over time. Behavioral tasks that are popular in studies of decision making involve integration of noisy sensory information (30, 32) or integration of probabilistic cues (119). Variants of these tasks (31, 34) might be used to test the gated integration and reset functionality of ORGaNICs. The antisaccade task (120–123) and the double-step saccade task (124–126) might also be used, with delay periods, to test the theory and to characterize how cortical circuits manage long-term dependencies.

Finally, the theory motivates a variety of experiments, some examples of which are as follows. First, the theory predicts that the modulators change the effective time constant and recurrent gain of a PFC neuron. Experimental evidence suggests that the modulatory responses are computed in the thalamus (2, 20, 99). Consequently, manipulating the responses of these thalamic neurons (e.g., via optogenetics) should have a particular impact on both the time constant and recurrent gain of cortical neurons. Second, the specific biophysical implementation (*SI Appendix*, Fig. S6) predicts that the soma and basal dendrites share input drive, but with opposite sign. This would, of course, have to be implemented with inhibitory interneurons. Third, the theory predicts that that neural activity underlying motor control and working memory is normalized. Normalization might be measured in motor cortex by comparing activity when making each of 2 simple movements vs. the combination of those movements simultaneously, or by comparing activity in one subpopulation of neurons with and without optogenetic stimulation of a separate subpopulation of neurons. Normalization might be measured in working memory circuits by comparing activity when maintaining one item versus multiple items during a delay period (127–129). Fourth, following previous research (130), a model based on ORGaNICs may be fit to behavioral and neurophysiological measurements of working memory. Trial-to-trial variability of behavioral performance during a working memory task has been shown to be linked with trial-to-trial variability in delay-period activity. These data might be fit by adding noise to the responses and/or synaptic weights, leading to drift in activity ratios during a delay period. Fifth, as noted above, variants of sensory integration tasks might be used to test the gated integration and reset functionality of ORGaNICs, and variants of the antisaccade and double-step saccade tasks might also be used, with delay periods, to characterize how cortical circuits manage long-term dependencies.

ACKNOWLEDGMENTS. We thank Mike Landy, Eero Simoncelli, Charlie Burlingham, Gyuri Buzsaki, Mike Long, Kenway Louie, Roozbeh Kiani, E. J. Chichilnisky, and Lindsey Glickfeld for comments and discussion.

Downloaded at Bobst Library, New York University on November 6, 2019

1. C. F. Jacobsen, Functions of frontal association area in primates. *Arch. Neurol. Psychiatry* **33**, 558–569 (1935).
2. J. M. Fuster, G. E. Alexander, Neuron activity related to short-term memory. *Science* **173**, 652–654 (1971).
3. J. M. Fuster, Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *J. Neurophysiol.* **36**, 61–78 (1973).
4. C. Constantinidis *et al*., Persistent spiking activity underlies working memory. *J. Neurosci.* **38**, 7020–7028 (2018).
5. J. W. Gnadt, R. A. Andersen, Memory related motor planning activity in posterior parietal cortex of macaque. *Exp. Brain Res.* **70**, 216–220 (1988).
6. S. Funahashi, C. J. Bruce, P. S. Goldman-Rakic, Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
7. P. S. Goldman-Rakic, Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).
8. C. Constantinidis, G. V. Williams, P. S. Goldman-Rakic, A role for inhibition in shaping the temporal flow of information in prefrontal cortex. *Nat. Neurosci.* **5**, 175–180 (2002).
9. D. Schluppeck, C. E. Curtis, P. W. Glimcher, D. J. Heeger, Sustained activity in topographic areas of human posterior parietal cortex during memory-guided saccades. *J. Neurosci.* **26**, 5098–5108 (2006).
10. E. K. Miller, C. A. Erickson, R. Desimone, Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154–5167 (1996).
11. R. Romo, C. D. Brody, A. Hernández, L. Lemus, Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399**, 470–473 (1999).
12. C. R. Hussar, T. Pasternak, Memory-guided sensory comparisons in the prefrontal cortex: Contribution of putative pyramidal cells and interneurons. *J. Neurosci.* **32**, 2747–2761 (2012).
13. M. J. Goard, G. N. Pho, J. Woodson, M. Sur, Distinct roles of visual, parietal, and frontal motor cortices in memory-guided sensorimotor decisions. *eLife* **5**, e13764 (2016).
14. X. J. Wang, Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *J. Neurosci.* **19**, 9587–9603 (1999).
15. A. Compte, N. Brunel, P. S. Goldman-Rakic, X. J. Wang, Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* **10**, 910–923 (2000).
16. E. H. Baeg *et al*., Dynamics of population code for working memory in the prefrontal cortex. *Neuron* **40**, 177–188 (2003).
17. S. Fujisawa, A. Amarasingham, M. T. Harrison, G. Buzsáki, Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat. Neurosci.* **11**, 823–833 (2008).
18. E. Pastalkova, V. Itskov, A. Amarasingham, G. Buzsáki, Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322–1327 (2008).
19. C. D. Harvey, P. Coen, D. W. Tank, Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
20. L. I. Schmitt *et al*., Thalamic amplification of cortical connectivity sustains attentional control. *Nature* **545**, 219–223 (2017).
21. M. Lundqvist, P. Herman, E. K. Miller, Working memory: Delay activity, yes! Persistent activity? Maybe not. *J. Neurosci.* **38**, 7013–7019 (2018).
22. C. D. Brody, A. Hernández, A. Zainos, R. Romo, Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb. Cortex* **13**, 1196–1207 (2003).
23. C. K. Machens, R. Romo, C. D. Brody, Flexible control of mutual inhibition: A neural model of two-interval discrimination. *Science* **307**, 1121–1124 (2005).
24. M. Shafi *et al*., Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* **146**, 1082–1108 (2007).
25. D. A. Markowitz, C. E. Curtis, B. Pesaran, Multiple component networks support working memory in prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11084–11089 (2015).
26. D. Kobak *et al*., Demixed principal component analysis of neural population data. *eLife* **5**, e10989 (2016).
27. J. D. Murray *et al*., Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 394–399 (2017).
28. B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, R. A. Andersen, Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* **5**, 805–811 (2002).
29. M. Lundqvist *et al*., Gamma and beta bursts underlie working memory. *Neuron* **90**, 152–164 (2016).
30. M. N. Shadlen, W. T. Newsome, Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
31. J. I. Gold, M. N. Shadlen, The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J. Neurosci.* **23**, 632–651 (2003).
32. B. W. Brunton, M. M. Botvinick, C. D. Brody, Rats and humans can optimally accumulate evidence for decision-making. *Science* **340**, 95–98 (2013).
33. R. Kiani, A. K. Churchland, M. N. Shadlen, Integration of direction cues is invariant to the temporal gap between them. *J. Neurosci.* **33**, 16483–16489 (2013).
34. B. A. Purcell, R. Kiani, Hierarchical decision processes that operate over distinct timescales underlie choice and changes in strategy. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E4531–E4540 (2016).
35. E. Aksay, G. Gamkrelidze, H. S. Seung, R. Baker, D. W. Tank, In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nat. Neurosci.* **4**, 184–193 (2001).
36. R. H. Hahnloser, A. A. Kozhevnikov, M. S. Fee, An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).
37. A. A. Kozhevnikov, M. S. Fee, Singing-related activity of identified HVC neurons in the zebra finch. *J. Neurophysiol.* **97**, 4271–4283 (2007).
38. D. Z. Jin, N. Fujii, A. M. Graybiel, Neural representation of time in cortico-basal ganglia circuits. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19156–19161 (2009).
39. M. A. Long, D. Z. Jin, M. S. Fee, Support for a synaptic chain model of neuronal sequence generation. *Nature* **468**, 394–399 (2010).
40. S. Hochreiter, J. Schmidhuber, Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
41. A. Graves, Generating sequences with recurrent neural networks. arXiv:1308.0850 (4 August 2013).
42. A. Graves, A.-r. Mohamed, G. Hinton, "Speech recognition with deep recurrent neural networks" in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2013), pp. 6645–6649.
43. K. Cho *et al*., Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv:1406.1078 (3 June 2014).
44. A. Graves, G. Wayne, I. Danihelka, Neural Turing machines. arXiv:1410.5401 (20 October 2014).
45. I. Sutskever, O. Vinyals, Q. V. Le, "Sequence to sequence learning with neural networks" in *Advances in Neural Information Processing Systems* (MIT Press, 2014), vol. 2, pp. 3104–3112.
46. D. J. Heeger, W. E. Mackey, ORGaNICs: A theory of working memory in brains and machines. arXiv:1803.06288 (16 March 2018).
47. D. J. Heeger, W. E. Mackey, ORGaNICs: A canonical neural circuit computation. bioRxiv:10.1101/506337 (26 December 2018).
48. D. J. Heeger, Supplemental Material for "Oscillatory recurrent gated neural integrator circuits (ORGaNICs), a unifying theoretical framework for neural dynamics." NYU Faculty Digital Archive. https://archive.nyu.edu/handle/2451/60439. Deposited 10 October 2019.
49. D. J. Heeger, Half-squaring in responses of cat striate cells. *Vis. Neurosci.* **9**, 427–443 (1992).
50. D. J. Heeger, Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992).
51. M. Carandini, D. J. Heeger, Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2011).
52. H. S. Seung, How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13339–13344 (1996).
53. A. A. Koulakov, S. Raghavachari, A. Kepecs, J. E. Lisman, Model for a robust neural integrator. *Nat. Neurosci.* **5**, 775–782 (2002).
54. C. D. Brody, R. Romo, A. Kepecs, Basic mechanisms for graded persistent activity: Discrete attractors, continuous attractors, and dynamic representations. *Curr. Opin. Neurobiol.* **13**, 204–211 (2003).
55. K. Zhang, Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *J. Neurosci.* **16**, 2112–2126 (1996).
56. M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge University Press, 1991).
57. M. Abeles, H. Bergman, E. Margalit, E. Vaadia, Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.* **70**, 1629–1638 (1993).
58. E. Bienenstock, A model of neocortex. *Network* **6**, 179–224 (1995).
59. M. Herrmann, J. Hertz, A. Prügel-Bennett, Analysis of synfire chains. *Network* **6**, 403–414 (1995).
60. E. M. Izhikevich, Polychronization: Computation with spikes. *Neural Comput.* **18**, 245–282 (2006).
61. M. M. Churchland, J. P. Cunningham, M. T. Kaufman, S. I. Ryu, K. V. Shenoy, Cortical preparatory activity: Representation of movement or first cog in a dynamical machine? *Neuron* **68**, 387–400 (2010).
62. K. V. Shenoy, M. Sahani, M. M. Churchland, Cortical control of arm movements: A dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–359 (2013).
63. A. A. Russo *et al*., Motor cortex embeds muscle-like commands in an untangled population response. *Neuron* **97**, 953–966.e8 (2018).
64. K. A. Thoroughman, R. Shadmehr, Learning of action through adaptive combination of motor primitives. *Nature* **407**, 742–747 (2000).
65. S. F. Giszter, Motor primitives–New data and future questions. *Curr. Opin. Neurobiol.* **33**, 156–165 (2015).
66. M. A. Long *et al*., Functional segregation of cortical regions underlying speech timing and articulation. *Neuron* **89**, 1187–1193 (2016).
67. M. A. Long, M. S. Fee, Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* **456**, 189–194 (2008).
68. Y. Lerner, C. J. Honey, M. Katkov, U. Hasson, Temporal scaling of neural responses to compressed and dilated natural speech. *J. Neurophysiol.* **111**, 2433–2444 (2014).
69. R. Gütig, H. Sompolinsky, Time-warp-invariant neuronal processing. *PLoS Biol.* **7**, e1000141 (2009).
70. S. Grossberg, Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Netw.* **1**, 17–61 (1988).
71. B. A. Olshausen, C. H. Anderson, D. C. Van Essen, A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* **13**, 4700–4719 (1993).
72. R. C. O'Reilly, M. J. Frank, Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.* **18**, 283–328 (2006).
73. G. Mongillo, O. Barak, M. Tsodyks, Synaptic theory of working memory. *Science* **319**, 1543–1546 (2008).
74. M. S. Goldman, Memory without feedback in a neural network. *Neuron* **61**, 621–634 (2009).

Downloaded at Boost Library, New York University on November 6, 2019

75. M. Lundqvist, A. Compte, A. Lansner, Bistable, irregular firing and population oscillations in a modular attractor memory network. *PLoS Comput. Biol.* **6**, e1000803 (2010).
76. M. Lundqvist, P. Herman, A. Lansner, Theta and gamma power increases and alpha/beta power decreases with memory load in an attractor network model. *J. Cogn. Neurosci.* **23**, 3008–3020 (2011).
77. S. Druckmann, D. B. Chklovskii, Neuronal circuits underlying persistent representations despite time varying activity. *Curr. Biol.* **22**, 2095–2103 (2012).
78. G. Hennequin, T. P. Vogels, W. Gerstner, Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82**, 1394–1406 (2014).
79. D. Sussillo, M. M. Churchland, M. T. Kaufman, K. V. Shenoy, A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* **18**, 1025–1033 (2015).
80. K. Rajan, C. D. Harvey, D. W. Tank, Recurrent network models of sequence generation and memory. *Neuron* **90**, 128–142 (2016).
81. R. Costa, I. A. Assael, B. Shillingford, N. de Freitas, T. Vogels, "Cortical microcircuits as gated-recurrent neural networks" in *31st Annual Conference on Neural Information Processing Systems*, U. von Luxburg *et al*., Eds. (NIPS, La Jolla, CA, 2017), vol. 1, pp. 272–283.
82. J. M. Murray, G. S. Escola, Learning multiple variable-speed sequences in striatum via cortical tutoring. *eLife* **6**, e26084 (2017).
83. V. Goudar, D. V. Buonomano, Encoding sensory and motor patterns as time-invariant trajectories in recurrent neural networks. *eLife* **7**, e31134 (2018).
84. N. Kraynyukova, T. Tchumatchenko, Stabilized supralinear network can give rise to bistable, oscillatory, and persistent activity. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 3464–3469 (2018).
85. E. Orhan, W. J. Ma, A diverse range of factors affect the nature of neural representations underlying short-term memory. bioRxiv:10.1101/244707 (13 October 2018).
86. J. P. Stroud, M. A. Porter, G. Hennequin, T. P. Vogels, Motor primitives in space and time via targeted gain modulation in cortical networks. *Nat. Neurosci.* **21**, 1774–1783 (2018).
87. C. Tallec, Y. Ollivier, Can recurrent neural networks warp time? arXiv:1804.11188 (23 March 2018).
88. S. Amari, Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**, 77–87 (1977).
89. B. Ermentrout, Neural networks as spatio-temporal pattern-forming systems. *Rep. Prog. Phys.* **61**, 353–430 (1998).
90. P. C. Bressloff, Spatiotemporal dynamics of continuum neural fields. *J. Phys. A Math. Theor.* **45**, 033001 (2011).
91. E. H. Adelson, J. R. Bergen, "The plenoptic function and the elements of early vision" in *Computational Models of Visual Processing*, M. S. Landy, J. A. Movshon, Eds. (MIT Press, 1991), pp. 3–20.
92. E. P. Simoncelli, D. J. Heeger, A model of neuronal responses in visual area MT. *Vision Res.* **38**, 743–761 (1998).
93. L. Muller, F. Chavane, J. Reynolds, T. J. Sejnowski, Cortical travelling waves: Mechanisms and computational principles. *Nat. Rev. Neurosci.* **19**, 255–268 (2018).
94. W. E. Mackey, C. E. Curtis, Distinct contributions by frontal and parietal cortices support working memory. *Sci. Rep.* **7**, 6188 (2017).
95. D. J. Heeger, Theory of cortical function. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1773–1782 (2017).
96. M. Carandini, D. J. Heeger, Summation and division by neurons in primate visual cortex. *Science* **264**, 1333–1336 (1994).
97. M. Carandini, D. J. Heeger, W. Senn, A synaptic explanation of suppression in visual cortex. *J. Neurosci.* **22**, 10053–10065 (2002).
98. F. S. Chance, L. F. Abbott, A. D. Reyes, Gain modulation from background synaptic input. *Neuron* **35**, 773–782 (2002).
99. Z. V. Guo *et al*., Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* **545**, 181–186 (2017).
100. R. V. Rikhye, A. Gilra, M. M. Halassa, Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nat. Neurosci.* **21**, 1753–1763 (2018).
101. R. V. Rikhye, R. D. Wimmer, M. M. Halassa, Toward an integrative theory of thalamic function. *Annu. Rev. Neurosci.* **41**, 163–183 (2018).
102. K. Thurley, W. Senn, H. R. Lüscher, Dopamine increases the gain of the input-output response of rat prefrontal pyramidal neurons. *J. Neurophysiol.* **99**, 2985–2997 (2008).
103. V. Varga *et al*., Fast synaptic subcortical control of hippocampal circuits. *Science* **326**, 449–453 (2009).
104. E. Marder, Neuromodulation of neuronal circuits: Back to the future. *Neuron* **76**, 1–11 (2012).
105. K. Wei *et al*., Serotonin affects movement gain control in the spinal cord. *J. Neurosci.* **34**, 12690–12700 (2014).
106. H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, D. Ferster, Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**, 578–592 (2009).
107. K. Louie, T. LoFaro, R. Webb, P. W. Glimcher, Dynamic divisive normalization predicts time-varying value coding in decision-related circuits. *J. Neurosci.* **34**, 16046–16057 (2014).
108. D. B. Rubin, S. D. Van Hooser, K. D. Miller, The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85**, 402–417 (2015).
109. H. Adesnik, M. Scanziani, Lateral competition for cortical space by layer-specific horizontal circuits. *Nature* **464**, 1155–1160 (2010).
110. X. Huang, Y. M. Elyada, W. H. Bosking, T. Walker, D. Fitzpatrick, Optogenetic assessment of horizontal interactions in primary visual cortex. *J. Neurosci.* **34**, 4976–4990 (2014).
111. J. J. Nassi, M. C. Avery, A. H. Cetin, A. W. Roe, J. H. Reynolds, Optogenetic activation of normalization in alert macaque visual cortex. *Neuron* **86**, 1504–1517 (2015).
112. T. K. Sato, B. Haider, M. Häusser, M. Carandini, An excitatory basis for divisive normalization in visual cortex. *Nat. Neurosci.* **19**, 568–570 (2016).
113. H. Adesnik, Synaptic mechanisms of feature coding in the visual cortex of awake mice. *Neuron* **95**, 1147–1159.e4 (2017).
114. K. A. Bolding, K. M. Franks, Recurrent cortical circuits implement concentration-invariant odor coding. *Science* **361**, eaat6904 (2018).
115. R. C. Atkinson, R. M. Shiffrin, "Human memory: A proposed system and its control processes" in *Psychology of Learning and Motivation*, K.W. Spence, J. T. Spence, Eds. (Elsevier, 1968), vol. 2, pp. 89–195.
116. N. Cowan, *Attention and Memory: An Integrated Framework* (Oxford University Press, 1998).
117. N. Cowan, What are the differences between long-term, short-term, and working memory? *Prog. Brain Res.* **169**, 323–338 (2008).
118. B. R. Postle, The cognitive neuroscience of visual short-term memory. *Curr. Opin. Behav. Sci.* **1**, 40–46 (2015).
119. T. Yang, M. N. Shadlen, Probabilistic reasoning by neurons. *Nature* **447**, 1075–1080 (2007).
120. P. E. Hallett, Primary and secondary saccades to goals defined by instructions. *Vision Res.* **18**, 1279–1296 (1978).
121. S. Funahashi, M. V. Chafee, P. S. Goldman-Rakic, Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature* **365**, 753–756 (1993).
122. D. P. Munoz, S. Everling, Look away: The anti-saccade task and the voluntary control of eye movement. *Nat. Rev. Neurosci.* **5**, 218–228 (2004).
123. K. Johnston, S. Everling, Neural activity in monkey prefrontal cortex is modulated by task context and behavioral instruction during delayed-match-to-sample and conditional prosaccade-antisaccade tasks. *J. Cogn. Neurosci.* **18**, 749–765 (2006).
124. G. Westheimer, Eye movement responses to a horizontally moving visual stimulus. *AMA Arch. Opthalmol.* **52**, 932–941 (1954).
125. W. Becker, R. Jürgens, An analysis of the saccadic system by means of double step stimuli. *Vision Res.* **19**, 967–983 (1979).
126. M. E. Goldberg, C. J. Bruce, Primate frontal eye fields. III. Maintenance of a spatially accurate saccade signal. *J. Neurophysiol.* **64**, 489–508 (1990).
127. W. J. Ma, W. Huang, No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *J. Vis.* **9**, 3.1–30 (2009).
128. Z. Wei, X. J. Wang, D. H. Wang, From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. *J. Neurosci.* **32**, 11228–11240 (2012).
129. S. Keshvari, R. van den Berg, W. J. Ma, No evidence for an item limit in change detection. *PLoS Comput. Biol.* **9**, e1002927 (2013).
130. K. Wimmer, D. Q. Nykamp, C. Constantinidis, A. Compte, Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat. Neurosci.* **17**, 431–439 (2014).

Downloaded at Bobst Library, New York University on November 6, 2019

Supplemental Information for

Oscillatory Recurrent Gated Neural Integrator Circuits (ORGaNICs): A Unifying Theoretical Framework for Neural Dynamics

David J. Heeger and Wayne E. Mackey

Corresponding author: David J. Heeger
Email: david.heeger@nyu.edu

Table of contents

Primer on Leaky Neural Integrators and Neural Oscillators 1

Detailed Methods and Mathematical Derivations 5

Sustained activity (Fig. 2) 5

Oscillatory activity with complex-valued recurrent weights 6

Sequential activity (Figs. 4-5) 8

Manipulation (Figs. 6-7) 9

Recurrent normalization (Fig. 3) 10

Extended Discussion 13

Biophysical implementation 13

Variations and extensions 15

Implications for neuroscience 17

Implications for machine learning 18

References 19

Primer on Leaky Neural Integrators and Neural Oscillators

A leaky neural integrator (**Fig. S1**) corresponds to a special case of ORGaNICs. For this special case, the modulators $a = b$ are equal to one another and constant over time, and the recurrent weight matrix equals the identity matrix. A simple example with a single neuron is expressed as follows:

$$\tau \frac{dy}{dt} = -y + \lambda z + (1-\lambda)y \ , \qquad \text{(S1)}$$

$$\lambda = \left(\frac{b^+}{1+b^+}\right) \text{ and } (1-\lambda) = \left(\frac{1}{1+b^+}\right) \ ,$$

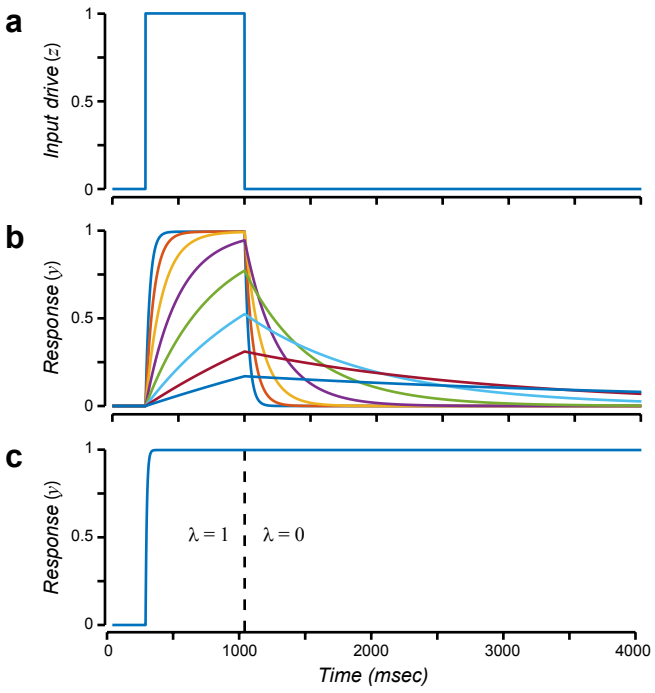

Figure S1. Leaky neural integrator (Eq. S1). a. Input drive (z) over time. **b.** Output responses (y) when the modulator (λ) is constant over time. Different colors correspond to different values of λ. **c.** Output responses (y) corresponding to time-varying modulator: $\lambda = 1$ for $t < 1000$ and $\lambda = 0$ for $t > 1000$.

i.e.,

$$\tau' \frac{dy}{dt} = -y + z \ , \qquad \text{(S2)}$$

$$\tau' = \frac{\tau_y}{\lambda} \ .$$

where τ is the intrinsic time-constant and τ' is the effective time-constant. For this simple special case, the neuron acts like a shift-invariant linear system, i.e., a recursive linear filter with an exponential impulse response function. If the input drive z is constant over time, then the responses y exhibit an exponential time course with steady state $y = z$, and time constant τ' (**Fig. S1**). This special case reveals how λ, and consequently the modulators, a and b, determine the effective time-constant of the leaky integrator.

The simple form in **Eq. S1** is problematic because it allows negative firing rates (the value of y could be positive or negative depending on the input drive). To fix that, we use a complementary pair of neurons (analogous to ON and OFF cells, e.g., in the retina, LGN, or V1) that receive complementary copies of the input drive, z and -z, and for which the firing rates are a halfwave-rectified copy of the underlying membrane potential fluctuations (optionally with a scale factor to convert from mV to spikes/sec):

$$\tau \frac{dv^+}{dt} = -v^+ + \lambda z + (1-\lambda)\left(y^+ - y^-\right) \ , \qquad \text{(S3)}$$

$$\tau \frac{dv^-}{dt} = -v^- + \lambda(-z) + (1-\lambda)\left(y^- - y^+\right) \ ,$$

where

$$\begin{aligned} y^+ &= \lfloor v^+ \rfloor = \max(v^+, 0) \\ y^- &= \lfloor v^- \rfloor = \max(v^-, 0) \\ v^+ &= y^+ - y^- \\ v^- &= y^- - y^+ \end{aligned} \ . \qquad \text{(S4)}$$

Here, v^+ and v^- are the membrane potential fluctuations (and also the recurrent drive) of the two neurons, z is the input drive, and y^+ and y^- are the output firing rates.

A leaky neural integrator with multiple neurons comprises a circuit such that the output responses of each neuron depend on a recurrent weighted sum:

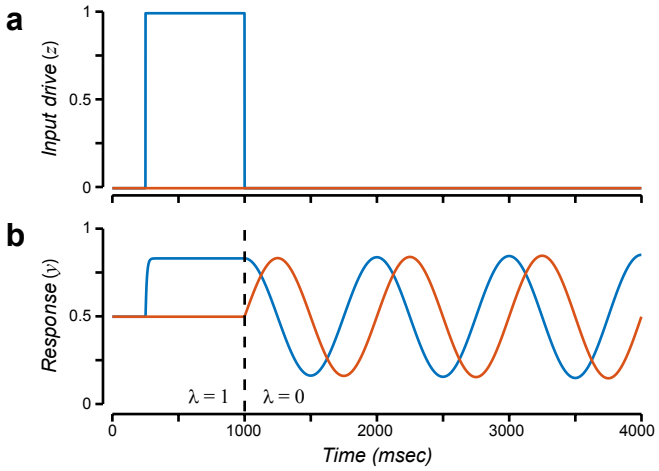

Figure S2. Neural oscillator (Eq. S7). a. Input (z) over time. Blue, real-part. Orange, imaginary part. **b.** Output responses (y) corresponding to time-varying modulator: $\lambda = 1$ for $t < 1000$ and $\lambda = 0$ for $t > 1000$. Blue, real-part. Orange, imaginary part.

$$\tau \frac{d\mathbf{y}}{dt} = -\mathbf{y} + \lambda \mathbf{z} + (1-\lambda)\hat{\mathbf{y}} \ , \qquad \text{(S5)}$$

$$\mathbf{z} = \mathbf{W}_{zx}\mathbf{x} \ ,$$

$$\hat{\mathbf{y}} = \mathbf{W}_{\hat{y}y}\mathbf{y} \ .$$

We use boldface lowercase letters to represent vectors and boldface uppercase to denote matrices. The variables ($\mathbf{y}$, $\hat{\mathbf{y}}$, $\mathbf{z}$, $\mathbf{x}$) are each functions of time. The time-varying output responses are represented by a vector $\mathbf{y} = (y_1, y_2, \ldots, y_j, \ldots, y_N)$ where the subscript j indexes different neurons in the network. The time-varying input drive is represented by another vector $\mathbf{z} = (z_1, z_2, \ldots, z_j, \ldots, z_N)$. The input stimulus is a continuous function $x(\theta,t)$ where θ parameterizes the stimulus space (e.g., polar angle and/or eccentricity in the visual field). The input stimulus is sampled to be represented by another time-varying vector $\mathbf{x} = (x_1, x_2, \ldots, x_j, \ldots, x_M)$, where $x_j(t) = x(\theta_j,t)$ and θ_j are the locations of the samples. The input drive z_j to each neuron is a weighted sum of the input $\mathbf{x}$, and the weights are given by the encoding matrix weight matrix $\mathbf{W}_{zx}$ (second line of **Eq. S5**). The recurrent drive $\hat{y}_j$ to each neuron is a weighted sum of the outputs, and the weights are given by the recurrent weight matrix $\mathbf{W}_{\hat{y}y}$ (third line of **Eq. S5**). We can use the same trick as above (**Eqs. S3-S4**) to ensure non-negative firing rates.

A neural oscillator (**Fig. S2**) corresponds to a special case of a leaky neural integrator in which the responses and recurrent weights are complex-valued. Specifically:

$$\tau \frac{dy}{dt} = -y + \lambda z + (1-\lambda)\hat{y} \ , \qquad \text{(S6)}$$

$$\hat{y} = wy = \left(1 + i2\pi\omega\tau\right)\left(\mathrm{Re}\left(y\right) + i\,\mathrm{Im}\left(y\right)\right) \ ,$$

The input drive z, responses y, and recurrent drive $\hat{y}$ are complex numbers, w is the complex-valued recurrent weight, and ω is the oscillation frequency. The complex-number notation is just a notational convenience. The complex-valued responses may be represented by a pair of neurons, and the complex-valued weight may be represented by pairs of synaptic weights that are matched to one another in the two neurons:

$$\tau \frac{d\mathbf{y}}{dt} = -\mathbf{y} + \lambda \mathbf{z} + (1-\lambda)\hat{\mathbf{y}} \ , \qquad \text{(S7)}$$

$$\hat{\mathbf{y}} = \mathbf{W}_{y\hat{y}}y = \begin{pmatrix} 1 & -2\pi\omega\tau \\ 2\pi\omega\tau & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} .$$

The values of y_1 and y_2 correspond, respectively, to the real and imaginary parts of the complex-valued responses. Intuitively, the responses oscillate because the recurrent weight matrix is a rotation matrix in the limit when the rotation angle is small, i.e., $\cos(\theta) = 1$ and $\sin(\theta) = \theta$ when $\theta \to 0$. If, at one instant in time, $\mathbf{y} = (1\ 0)^t$, then an instant later the responses will have

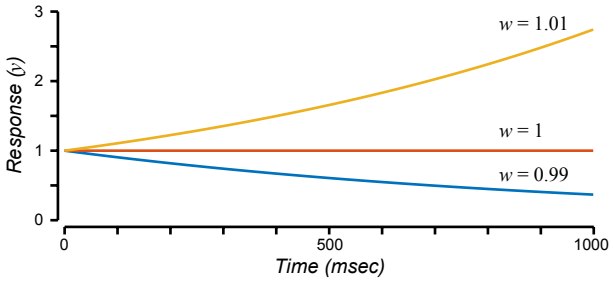

Figure S3. Stability. Responses when $\lambda = 0$ for 3 different values of the recurrent weight. Yellow, responses grow without bound when recurrent weight is greater than 1. Blue, responses decay to zero when recurrent weight is less than 1. Orange, responses are constant over time when recurrent weight is equal to 1.

changed akin to rotating slightly around the unit circle.

The dynamics of the responses depend on the recurrent weight matrix (**Fig. S3**). This is particularly important when $\lambda = 0$ (corresponding to the sustained delay-period activity in **Fig. S1** or the period of stable oscillations in **Fig. S2**). If the weights are too small then the responses decay over time. If the weights are too large then the responses grow without bound. A simple is example is given by a circuit with 3 neurons and a diagonal recurrent weight matrix:

$$\mathbf{W}_{y\hat{y}} = \begin{pmatrix} 1.01 & 0 & 0 \\ 0 & 1.0 & 0 \\ 0 & 0 & 0.99 \end{pmatrix} . \qquad \text{(S8)}$$

Rewriting **Eq. S5** for the special case of a diagonal weight matrix:

$$\tau \frac{dy_j}{dt} = -y_j + \lambda z_j + (1-\lambda)\hat{y}_j \ , \qquad \text{(S9)}$$

$$\hat{y}_j = w_j y_j \ ,$$

where the recurrent weights w_j are the elements along the diagonal of the recurrent weight matrix $\mathbf{W}_{\hat{y}y}$. When $\lambda = 0$, this equation simplifies further:

$$\tau \frac{dy_j}{dt} = -y_j + w_j y_j \ . \qquad \text{(S10)}$$

For $w_j = 1$, the responses are constant over time (the derivative in **Eq. S10** is 0). For $w_j > 1$, the response grow over time. And for $w_j < 1$, the response decay over time.

In general, for an arbitrary recurrent weight matrix, the dynamics of the responses depend on the eigenvalues and eigenvectors of the recurrent weight matrix. When the eigenvectors and eigenvalues of the recurrent weight matrix are composed of complex values, the responses exhibit oscillations. For example, the recurrent weight matrix in the neural oscillator example (**Fig. S2**) is an anti-symmetric, 2x2 matrix, with complex-valued eigenvalues and eigenvectors (**Eq. S7**). The real-parts of the eigenvalues determine stability. In this case, the real parts of the eigenvalues are equal to 1 (the weight matrix was in fact scaled so that the eigenvalues have real parts that were equal to 1). The corresponding eigenvectors define a coordinate system (or basis) for the responses. The responses during the period of stable oscillations (when $\lambda = 0$) are determined entirely by the projection of the initial values (the responses just before the input was turned off) onto the eigenvectors. Eigenvectors with corresponding eigenvalues that have real parts equal to 1 are stable. Those with eigenvalues that have real parts less than 1 decay to zero (smaller eigenvalues decay more quickly). Those with eigenvalues that have real parts greater than 1 grow without bound (which is why the weight matrix was scaled so that the largest eigenvalues = 1). The imaginary parts of the eigenvalues of the recurrent weight matrix

(in this example equal to $2\pi\omega\tau$) determine the oscillation frequencies (ω in this example).

Detailed Methods and Mathematical Derivations

Sustained activity (Fig. 2)

The recurrent weight matrix $\mathbf{W}_{\hat{y}y}$ for this circuit was a 36x36 matrix, designed based on quadrature-mirror filter wavelets (1, 2). Quadrature-mirror filters are (approximately) mutually orthogonal when shifted or scaled by factors of 2. Consequently, the rows of $\mathbf{W}_{\hat{y}y}$ were shifted copies of one another (i.e., $\mathbf{W}_{\hat{y}y}$ was convolutional), the even (or odd) rows of $\mathbf{W}_{\hat{y}y}$ were (approximately) mutually orthogonal, and the eigenvectors of the recurrent weight matrix, with corresponding eigenvalues equal to 1, spanned a 19-dimensional subspace. The non-orthogonal components, with corresponding eigenvalues < 1, rapidly decayed to zero during the delay period.

The steady-state responses during the delay period depended on the dot products of the initial responses and the eigenvectors of the recurrent weight matrix $\mathbf{W}_{\hat{y}y}$ with corresponding eigenvalues equal to 1:

$$\mathbf{p} = \mathbf{V}^t \mathbf{y}_0 , \qquad \text{(S11)}$$

$$\mathbf{y}_s = \mathbf{V}\, \mathbf{p} ,$$

where $\mathbf{y}_s$ is the vector of steady-state responses, $\mathbf{y}_0$ is the vector of initial values at the beginning of the delay period, and $\mathbf{V}$ is a 36x19 matrix. The columns of $\mathbf{V}$ (the rows of $\mathbf{V}^t$) are an orthonormal basis for those eigenvectors of the recurrent weight matrix $\mathbf{W}_{\hat{y}y}$ that have corresponding eigenvalues equal to 1, and $\mathbf{p}$ is the projection of $\mathbf{y}_0$ on $\mathbf{V}$.

The encoding weight matrix $\mathbf{W}_{zx}$ was a 36x360 matrix (N=36 neurons and M=360 polar angle samples, ignoring for now the stimulus components corresponding to the cues that indicated the beginning and end of each trial). The response fields (i.e., the rows of the encoding weight matrix $\mathbf{W}_{zx}$) were each designed to be one cycle of a raised cosine:

$$\psi_j(\theta) \propto \left| \cos\left(\frac{\theta - \varphi_j}{2}\right)^n \right| , \qquad \text{(S12)}$$

where θ is stimulus polar angle and φ_j is the preferred polar angle (i.e., the response field center) of the j^{th} neuron. The response field centers were evenly spaced. The value of the exponent was chosen to be $n = 18$, so that the response fields of the 36 neurons spanned a 19-dimensional subspace of polar angles, i.e., the rank of $\mathbf{W}_{zx}$ was 19, equal to the dimensionality of the subspace spanned by $\mathbf{V}$. These raised-cosine response fields have the desirable property that they are shiftable (3, 4), i.e., the response of a neuron with a response field of the same shape but shifted to any intermediate polar angle can be computed (exactly) as a weighted sum of the responses of a basis set of 18 of the 36 neurons. The response fields evenly tiled the polar angle component of the visual field (i.e., the sum of the squares of the rows of $\mathbf{W}_{zx} = 1$):

$$\sum \psi_j^2(\theta) = 1 . \qquad \text{(S13)}$$

In addition, the encoding weight matrix $\mathbf{W}_{zx}$ was designed so as to project the input stimulus onto the same subspace as that spanned by $\mathbf{V}$:

$$\mathbf{W}_{zx} = \mathbf{V}\mathbf{W}_{rx} \ , \qquad \text{(S14)}$$

$$\mathbf{y}_0 = \mathbf{W}_{zx}\mathbf{x}_0 = \mathbf{V}\mathbf{W}_{rx}\mathbf{x}_0 \ ,$$

where $\mathbf{x}_0$ was a vector (360x1) corresponding to the input stimulus. The matrix $\mathbf{W}_{rx}$ (19x360) was computed so that the rows of $\mathbf{W}_{zx}$ were the raised cosine functions described above, i.e., $\mathbf{W}_{rx} = \mathbf{V}^t \mathbf{W}_{zx}$.

The same matrix $\mathbf{V}$ was used to perform the readout. Consequently, the input stimulus could be reconstructed approximately from the steady-state responses (ignoring normalization which serves to rescale the responses and consequently the reconstruction):

$$\mathbf{r} = \mathbf{W}_{ry}\mathbf{y}_s = \mathbf{V}^t\mathbf{y}_s \ , \qquad \text{(S15)}$$

$$\hat{\mathbf{x}} = \mathbf{W}_{rx}^{\#}\mathbf{r} = \mathbf{W}_{rx}^{\#}\mathbf{W}_{ry}\mathbf{y}_s = \mathbf{W}_{rx}^{\#}\mathbf{V}^t\mathbf{y}_s = \mathbf{W}_{rx}^{\#}\mathbf{V}^t\mathbf{V}\mathbf{p} = \mathbf{W}_{rx}^{\#}\mathbf{V}^t\mathbf{V}\mathbf{V}^t\mathbf{y}_0 = \mathbf{W}_{rx}^{\#}\mathbf{V}^t\mathbf{V}\mathbf{V}^t\mathbf{V}\mathbf{W}_{rx}\mathbf{x}_0 \approx \mathbf{x}_0 \ ,$$

where the last step approximates $\mathbf{x}_0$ because $\mathbf{V}$ is an orthonormal matrix (i.e., $\mathbf{V}^t\mathbf{V} = \mathbf{I}$), and # denotes pseudo-inverse. The reconstruction matrix, the pseudo-inverse of $\mathbf{W}_{rx}$, was computed with ridge regression:

$$\mathbf{W}_{rx}^{\#} = \left(\mathbf{W}_{rx}^t\mathbf{W}_{rx} + k\mathbf{I}\right)^{-1}\mathbf{W}_{rx}^t \ , \qquad \text{(S16)}$$

where k was a small constant to ensure that the reconstruction was stable with respect to perturbations in $\mathbf{r}$.

The input drive was reconstructed exactly (up to a scale factor because of the rescaling due to normalization) from the readout:

$$\hat{\mathbf{z}} = \mathbf{V}\mathbf{r} = \mathbf{V}\mathbf{W}_{ry}\mathbf{y}_s = \mathbf{V}\mathbf{V}^t\mathbf{V}\mathbf{V}^t\mathbf{y}_0 = \mathbf{V}\mathbf{V}^t\mathbf{V}\mathbf{V}^t\mathbf{V}\mathbf{W}_{rx}\mathbf{x}_0 = \mathbf{V}\mathbf{W}_{rx}\mathbf{x} = \mathbf{W}_{zx}\mathbf{x}_0 = \mathbf{z}_0 \ , \qquad \text{(S17)}$$

where $\mathbf{z}_0$ is a vector (36x1) corresponding input drive during target presentation.

The steady-state responses (and consequently the readout) were the same even when the encoding weights also included components that were orthogonal to $\mathbf{V}$. Specifically, if the encoding weights were $\mathbf{W}_{zx} = (\mathbf{V}+\mathbf{V}_p)\mathbf{W}_{rx}$ such that $\mathbf{V}^t\mathbf{V}_p = 0$:

$$\mathbf{y}_s = \mathbf{V}\mathbf{V}^t\mathbf{y}_0 = \mathbf{V}\mathbf{V}^t\left(\mathbf{V} + \mathbf{V}_p\right)\mathbf{W}_{rx}\mathbf{x}_0 = \mathbf{V}\mathbf{V}^t\mathbf{V}\mathbf{W}_{rx}\mathbf{x}_0 = \mathbf{V}\mathbf{W}_{rx}\mathbf{x}_0 \ . \qquad \text{(S18)}$$

The behavioral cues built into the experimental protocol set the state of the modulators via $\mathbf{W}_{ax}$ and $\mathbf{W}_{bx}$, and the state changed from one phase to the next. In particular, we presumed that there were two input stimuli, in addition to the 360 possible target locations, that corresponded to the cues. The first of these stimuli was 1 from the beginning of each trial to just before target offset. The second of these stimuli was 1 for a brief period of time at the end of the trial.

Oscillatory activity with complex-valued recurrent weights

One way to generate delay-period activity with oscillatory dynamics is when the recurrent weights and output responses are complex-valued (**Fig. S4**). As noted above, the complex-valued responses may be represented by pairs of neurons, and the complex-valued weights in the recurrent weight matrix may be represented by pairs of synaptic weights (with each pair of synaptic weights shared by each pair of neurons to perform complex-number multiplication).

For the example circuit in **Fig. S4**, there were again 36 neurons with the same encoding matrix $\mathbf{W}_{zx}$ as in the preceding (sustained activity) example. The modulators were also the

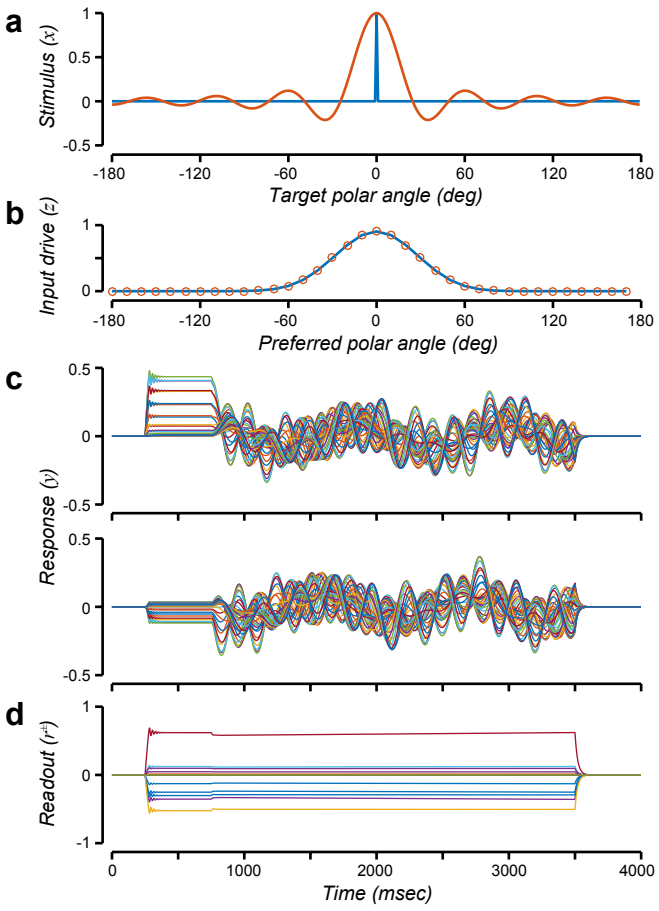

Figure S4. Oscillatory activity. a. Input stimulus and reconstructed stimulus. Blue, input stimulus ($\mathbf{x}$) corresponding to target location. Orange, reconstructed stimulus, computed as a weighted sum of the reconstructed input drive (panel b). **b.** Input drive and reconstructed input drive. Blue, input drive ($\mathbf{z}$) to each neuron as a function of that neuron's preferred target location. Orange, reconstructed input drive, computed as a weighted sum of the readout (panel d) at a randomly chosen time point during the delay period. **c.** Output responses ($\mathbf{y}$). Top row, real part of the responses. Bottom row, imaginary part of the responses. Each color corresponds to a different neuron. **d.** Readout ($\mathbf{r}^{\pm}$). Each color corresponds to a different component of the readout.

same as in the preceding example. The recurrent weights were complex-valued. The real part of the recurrent weight matrix $\mathbf{W}_{\hat{y}y}$ was the same as that in the sustained delay-period example (**Fig. 2b**), such that 19 eigenvalues had real parts equal to 1, and the real parts of the other eigenvalues were less than 1. But the imaginary part of the recurrent weight matrix was different, with random values (between 0 and $8\pi\tau_y$) for the imaginary parts of all 36 eigenvalues, corresponding to oscillation frequencies between 0 and 4 Hz (although other frequencies or combinations of frequencies could be used instead). Consequently, the activity exhibited complex dynamics (sums of oscillations) that resembled the phenomenology of delay-period activity (5-10). The circuit was a 19-dimensional (complex-valued) continuous attractor during the delay period because the recurrent weight matrix was constructed to have 19 eigenvalues with real parts equal to 1.

In spite of the complex response dynamics (**Fig. S4c**), the readout was again (as it was for the sustained activity circuit) constant over time during the delay period (**Fig. S4d**), the input drive was reconstructed exactly (up to a scale factor) from the responses at any time during the delay period (**Fig. S4b**), and target location was reconstructed approximately (**Fig. S4a**).

The readout for this example circuit was more complicated than that for the circuit that exhibited sustained delay-period activity. The readout depended not only on a weighted sum of the responses but also an estimate of the sign:

$$\mathbf{r} = \mathbf{W}_{ry}\mathbf{y} \ , \qquad \text{(S19)}$$

$$\mathbf{r}^{\pm} = \mathbf{D}(\mathbf{s})\left|\mathbf{r}\right| \ .$$

where $\mathbf{r}^{\pm}$ is the sign-corrected readout. The readout matrix $\mathbf{W}_{ry} = \mathbf{V}^t$ was a unitary basis for the eigenvectors of the recurrent weight matrix $\mathbf{W}_{\hat{y}y}$ with corresponding eigenvalues that had real parts equal to 1. The vector $\mathbf{s}$ consisted of ± 1 values to correct the sign of the readout, and $\mathbf{D}(\mathbf{s})$ was a diagonal matrix such that each element of the vector $\mathbf{s}$ was multiplied by the corresponding element of $|\mathbf{W}_{ry}\,\mathbf{y}|$.

The values of $\mathbf{s}$ were computed from the responses $\mathbf{y}$, sampled at two time points. First, the instantaneous frequency of each quadrature pair of neural responses was computed from the real- and imaginary-parts of the responses:

$$\omega_j = \frac{\phi_j(t_2) - \phi_j(t_1)}{2\pi(\Delta t)} \quad , \tag{S20}$$

$$\phi_j(t) = \tan^{-1}\left(\frac{\mathrm{Im}(r_j)}{\mathrm{Re}(r_j)}\right) \quad ,$$

where $\Delta t = t_2 - t_1$ was presumed be known, although the values of t_1 and t_2 (i.e., the times at which the responses were sampled) were presumed to be unknown. Second, the elapsed time of the delay period T and the corresponding response sign $\mathbf{s}$ were estimated by minimizing:

$$\sum_j \left| s_j \hat{r}_j - r_j \right|^2 \quad , \tag{S21}$$

$$\hat{r}_j = e^{2\pi i T \omega_j} \quad ,$$

$$s_j = \mathrm{sgn}\left(\mathrm{Re}\left(\frac{r_j}{\hat{r}_j}\right)\right) \quad ,$$

where r_j was the complex-valued response at time T, and ω_j was the instantaneous frequency (**Eq. S20**). Specifically, we sampled a large number of values of T to determine an estimate for the elapsed time that minimized the first line of **Eq. S21**. Given that estimate of T, the response sign $\mathbf{s}$ was then computed using the last two lines of **Eq. S21**. There is a unique solution for $\mathbf{s}$ when at least two of the oscillation temporal periods have no common multiples. This calculation, as written, is not neurobiologically plausible, but we anticipate that a neural net can approximate the function that transforms from $\mathbf{y}$ to $\mathbf{s}$, or from $\mathbf{y}$ to $\mathbf{r}^{\pm}$.

After correcting the sign of the readout, the input drive and input stimulus were reconstructed using the same procedure as that described above for the sustained delay-period circuit (**Eqs. S15-S17**).

Sequential activity (Figs. 4-5)

The response fields were the same as in the preceding examples (**Eqs. S12-S13**). The recurrent weight matrix was real-valued but asymmetric. It was computed from a discrete approximation to the derivative of the responses (11). This was facilitated by using raised cosines for the response fields. The raised cosine response fields have the desirable property that they are shiftable (3, 4), i.e., the response of a neuron with a response field of the same shape but shifted to any preferred response field location can be computed (exactly) as a weighted sum of the responses of a basis set of any 18 of the 36 neurons. First, we computed an upsampling matrix that (exactly) interpolated the responses of the 36 neurons to an arbitrarily large number of responses with arbitrarily closely-spaced preferred locations. Second, we computed a downsampling matrix, as the pseudo-inverse of the upsampling matrix, that subsampled the responses to the response field locations of the 36 neurons in the network. Third, we also created a Toeplitz matrix with a convolution kernel that approximated differentiation (12). Fourth, and finally, the recurrent weight matrix depended on the product of these 3 matrices:

$$\mathbf{W}_{y\hat{y}} = \mathbf{I} + \mathbf{D} \quad , \tag{S22}$$

$$\mathbf{D} = \mathbf{D}_1 \mathbf{D}_2 \mathbf{D}_3 \quad ,$$

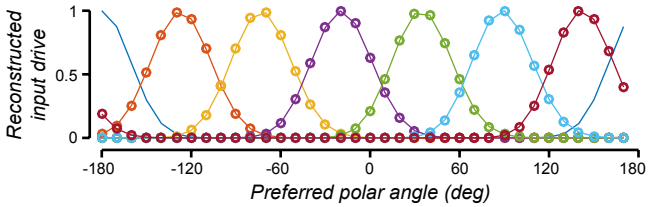

Figure S5. Linear readout and reconstruction for the sequential activity circuit (Fig. 4). Blue curve, input drive (z). Other colors, linear readout and reconstruction at evenly spaced time points during the delay period.

where $\mathbf{I}$ was the identity matrix, $\mathbf{D}_1$ was the NxM downsampling matrix, $\mathbf{D}_2$ was the MxM derivative matrix, $\mathbf{D}_3$ was the MxN upsampling/interpolation matrix, N was equal to the number of neurons (36), M was the (arbitrary) upsampling resolution (e.g., M=360).

Unlike the preceding examples, it was not possible to reconstruct the input drive from the readout at arbitrary points in time during the delay period. A linear reconstruction (like that used for the preceding examples) generated shifted copies of the input drive, that shifted over time like a traveling wave (**Fig. S5**). Consequently, this circuit was capable of maintaining some (but not all) information about the input during the delay period. A possible alternative might be to design the readout to be constant over time, even when the responses exhibit oscillatory dynamics, by ensuring that the rows of the readout weight matrix are orthogonal to the oscillatory components of the recurrent weight matrix (13). This simple idea, based on standard textbook linear algebra, however, cannot be applied to obtain a constant readout from the sequential activity in our circuit because our recurrent weight matrix was full rank.

Manipulation (Figs. 6-7)

There were three recurrent weight matrices in this example circuit, the first for maintaining a representation of the target locations, the second for updating the representation with leftward eye movements, and the third for updating the representation with rightward eye movements. The first recurrent weight matrix was identical to that in the sustained activity circuit (i.e., shifted copies of the quadrature-mirror filter kernel, see above). The second recurrent weight matrix was a discrete approximation to the derivatives of the responses ($\mathbf{D}$ in the 2nd line of **Eq. S22**), and third was the negative derivative matrix (-1 times the matrix defined by **Eq. S22**).

During the period of spatial updating illustrated in **Fig. 6** (i.e., $b = 0$, and either a_2 or $a_3 > 0$), **Eq. 6** can be rewritten:

$$\tau_y \frac{dy_j}{dt} = -s \left. \frac{dy}{d\varphi} \right|_{\varphi=\varphi_j} . \qquad \text{(S23)}$$

The value of $s = a_k / (a_k + 1)$, for k equal to either 2 or 3, was the speed of the traveling wave of activity during updating. The variable φ was a continuous parameterization of response field location (eccentricity), and $y(\varphi)$ was a continuous function for every possible response field location, interpolated from the y_j samples that represented the neural responses. The derivative $\mathrm{d}y/\mathrm{d}\varphi$, sampled at $\varphi = \varphi_j$, was the derivative of the neural responses sampled at the response field location of the j^{th} neuron. The recurrent weight matrix $\mathbf{D}$ computed this derivative; multiplying by this weight matrix was equivalent to interpolating the y_j samples to a continuous function, computing the derivative of that function, and then resampling. The duration of spatial updating (i.e., the period of time during which either a_2 or a_3 was non-zero) was proportional to the amplitude of the eye movement:

$$\theta = sT \ , \qquad \text{(S24)}$$

where θ was the location of the target stimulus and T was the duration of updating. Conse-

quently, the neural responses over time were discrete samples of a continuous function:

$$y_j(t) = y(\varphi_j, t) \ , \tag{S25}$$

$$y(\varphi, t) = y_0(\varphi - st) \ ,$$

where y_0 was the continuously interpolated neural activity at the point in time (i.e., $t = 0$) just prior to updating.

Likewise, during the period of motor execution illustrated in **Fig. 7** (i.e., $b = 0$, and either a_2 or $a_3 > 0$), the responses exhibited oscillating traveling waves. The value of $s = a_k / (a_k + 1)$, for k equal to either 2 or 3, was the speed of the traveling waves.

Recurrent normalization (Fig. 3)

ORGaNIC normalization was implemented with a dynamical system comprised of coupled neural integrators:

$$\tau_y \frac{dy_j}{dt} = -y_j + \left(\tfrac{b^+}{1+b^+}\right) z_j + \left(\tfrac{1}{1+a^+}\right) \hat{y}_j \ , \tag{S26}$$

$$\tau_b \frac{db}{dt} = -b + \mathbf{w}_{bx}^t \mathbf{x} \ , \tag{S27}$$

$$\tau_a \frac{da}{dt} = -a + \sqrt{u} + \sqrt{u} a + \mathbf{w}_{ax}^t \mathbf{x} \ , \tag{S28}$$

$$\tau_u \frac{du}{dt} = -u + u \left\| \mathbf{y} \right\|^2 + \left(\tfrac{\sigma b^+}{1+b^+}\right)^2 \text{ such that } u > \left(\tfrac{\sigma b^+}{1+b^+}\right)^2 . \tag{S29}$$

The norm of y is the sum of squares of the real and imaginary parts, summed across neurons:

$$\left\| \mathbf{y} \right\|^2 = \sum_j \left| y_j \right|^2 = \sum_j \left[\mathrm{Re}\left(y_j\right)^2 + \mathrm{Im}\left(y_j\right)^2 \right] . \tag{S30}$$

This system of equations is a special case of **Eqs. 1-2**, in which $f(\mathbf{y})$ is defined in terms of u, where the value of of u is the response of yet another modulator neuron. The input modulator b depended only on the input, and in particular was set to 1 when a cue was presented indicating the beginning of the trial, i.e., $\mathbf{w}_{bx}^t \mathbf{x} = 1$ when the cue was presented. The recurrent modulator a determined the gain of the recurrent drive. The value of a depended on the output responses (**Eqs. S28**-**S29**). The value of a also depended on the input, i.e., $\mathbf{w}_{ax}^t \mathbf{x} = 1$, such that it was ~1 at the end of the trial.

The stability of the normalized responses did not depend on fine tuning of the synaptic weights in the circuit. Including normalization automatically stabilized the activity of the circuit when the recurrent drive was scaled by a few percent (**Fig. 3b**). The stability of the normalized responses did not depend on fine tuning any of the other synaptic weights in the circuit (i.e., any of the other terms in **Eqs. S27**-**S29**). We performed simulations perturbing the synaptic weights represented by each of the terms in **Eqs. S27**-**S29**. Multiplying by random values within ±5% yielded virtually identical simulated responses and the responses were stable even when those terms were multiplied by random values ranging from 0.5x to 2x.

The sustained activity and oscillatory activity circuits can achieve a stable state in which the

output responses are normalized:

$$|y_j|^2 = \frac{|z_j|^2}{\sigma^2 + \sum |z_j|^2} \ . \tag{S31}$$

To derive this result, we restrict the analysis to when $\mathbf{w}_{ax}^t \mathbf{x} = 0$ (noting that $\mathbf{w}_{ax}^t \mathbf{x}$ was non-zero only at the end of the trial after the delay period), and when a and b are both ≥ 0 (noting that this will generally be the case in the stable state), and we write the stable state for each of **Eqs. S26**, **S28**, and **S29**:

$$\frac{d|y_j|}{dt} = 0 \quad \text{and} \quad |y_j| = |\hat{y}_j| \tag{S32}$$

$$\begin{aligned}
y_j &= \left(\tfrac{b}{1+b}\right) z_j + \left(\tfrac{1}{1+a}\right) \hat{y}_j \\
y_j &= \left(\tfrac{b}{1+b}\right) z_j + \left(\tfrac{1}{1+a}\right) y_j \\
|y_j|^2 &= \left(\tfrac{b}{1+b}\right)^2 |z_j|^2 + \left(\tfrac{1}{1+a}\right)^2 |y_j|^2 \\
\left(\tfrac{a}{1+a}\right)^2 |y_j|^2 &= \left(\tfrac{b}{1+b}\right)^2 |z_j|^2 \\
|y_j|^2 &= \left(\tfrac{1+a}{a}\right)^2 \left(\tfrac{b}{1+b}\right)^2 |z_j|^2
\end{aligned}$$

$$\begin{aligned}
\frac{da}{dt} &= 0 \\
a &= \sqrt{u} + \sqrt{u}a \\
a &= (1+a)\sqrt{u} \\
u &= \left(\tfrac{a}{1+a}\right)^2
\end{aligned} \tag{S33}$$

$$\begin{aligned}
\frac{du}{dt} &= 0 \\
u &= u||\mathbf{y}||^2 + \left(\tfrac{\sigma b}{1+b}\right)^2 \\
u\left(1 - ||\mathbf{y}||^2\right) &= \left(\tfrac{\sigma b}{1+b}\right)^2
\end{aligned} \tag{S34}$$

Combining the last line of **Eq. S32** with the last line of **Eq. S33**:

$$|y_j|^2 = \left(\tfrac{1}{u}\right)\left(\tfrac{b}{1+b}\right)^2 |z_j|^2 \ . \tag{S35}$$

Combining **Eq. S35** with the last line of **Eq. S34**:

$$\begin{aligned}
|y_j|^2 &= \left(1 - ||\mathbf{y}||^2\right)\left(\tfrac{1+b}{\sigma b}\right)^2 \left(\tfrac{b}{1+b}\right)^2 |z_j|^2 \\
|y_j|^2 &= \left(\tfrac{1}{\sigma}\right)^2 \left(1 - ||\mathbf{y}||^2\right) |z_j|^2
\end{aligned} \tag{S36}$$

Summing both sides:

$$\|\mathbf{y}\|^2 = \left(\tfrac{1}{\sigma}\right)^2 \left(1 - \|\mathbf{y}\|^2\right) \|\mathbf{z}\|^2 \ , \tag{S37}$$

and simplifying:

$$\begin{aligned} \sigma^2 \|\mathbf{y}\|^2 &= \|\mathbf{z}\|^2 \left(1 - \|\mathbf{y}\|^2\right) \\ \sigma^2 \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2 \|\mathbf{y}\|^2 &= \|\mathbf{z}\|^2 \\ \left(\sigma^2 + \|\mathbf{z}\|^2\right) \|\mathbf{y}\|^2 &= \|\mathbf{z}\|^2 \\ \|\mathbf{y}\|^2 &= \frac{\|\mathbf{z}\|^2}{\sigma^2 + \|\mathbf{z}\|^2} \end{aligned} \tag{S38}$$

Comparing **Eq. S37** with the last line of **Eq. S38**:

$$\left(\tfrac{1}{\sigma}\right)^2 \left(1 - \|\mathbf{y}\|^2\right) = \frac{1}{\sigma^2 + \|\mathbf{z}\|^2} \ . \tag{S39}$$

Substituting **Eq. S39** into the last line of **Eq. S36** yields the desired result (**Eq. S31**).

To show that the ratios of the responses equal the ratios of the input drives, we combine **Eq. S31** with the last line of **Eq. S38**:

$$\frac{|y_j|^2}{\|\mathbf{y}\|^2} = \left(\frac{|z_j|^2}{\sigma^2 + \|\mathbf{z}\|^2}\right)\left(\frac{\sigma^2 + \|\mathbf{z}\|^2}{\|\mathbf{z}\|^2}\right) = \frac{|z_j|^2}{\|\mathbf{z}\|^2} \ . \tag{S40}$$

For the sequential activity circuit, the stable state is different:

$$\frac{d\|y_j\|}{dt} = 0 \quad \text{and} \quad \|y_j\| = \|\hat{y}_j\| \ , \tag{S41}$$

i.e.,

$$\|y_j\|^2 = \left(\tfrac{1}{1+b}\right)\|z_j\|^2 + \left(\tfrac{1}{1+a}\right)\|\hat{y}_j\|^2 \tag{S42}$$

$$\|y_j\|^2 = \left(\tfrac{1+a}{a}\right)^2 \left(\tfrac{b}{1+b}\right)^2 \|z_j\|^2 \ ,$$

which yields:

$$\|\mathbf{y}\|^2 = \frac{\|\mathbf{z}\|^2}{\sigma^2 + \|\mathbf{z}\|^2} \ . \tag{S43}$$

The circuit expressed by **Eqs. S26-S29** is but one example of an ORGaNIC normalization model. There is, in fact, a family of similar dynamical systems models, each of which comprises coupled neural integrators to implement normalization via recurrent amplification. All of the models in this family can achieve the same stable state (**Eq. S31**), but the various different models in this family imply different circuits with different dynamics. For example, one such variant avoids the square-root operation in **Eq. S28** by adding another type of neuron to the circuit:

$$\tau_y \frac{dy_j}{dt} = -y_j + \left(\frac{b^+}{1+b^+}\right) z_j + \left(\frac{1}{1+a^+}\right) \hat{y}_j \ , \quad \text{(S44)}$$

$$\tau_b \frac{db}{dt} = -b + \mathbf{w}_{bx}^t \mathbf{x} \ , \quad \text{(S45)}$$

$$\tau_a \frac{da}{dt} = -a + v + va + \mathbf{w}_{ax}^t \mathbf{x} \ , \quad \text{(S46)}$$

$$\tau_v \frac{dv}{dt} = -v + \frac{u}{v} \text{ such that } v > 0 \ , \quad \text{(S47)}$$

$$\tau_u \frac{du}{dt} = -u + u\left\|\mathbf{y}\right\|^2 + \left(\frac{\sigma b^+}{1+b^+}\right)^2 \text{ such that } u > \left(\frac{\sigma b^+}{1+b^+}\right)^2 . \quad \text{(S48)}$$

It is straightforward to show that **Eqs. S44-S48** have the same stable state as **Eqs. S26-29**; setting **Eqs. S46** and **S47** to zero and solving for u yields the last line of **Eq. S33**.

In another variant, we replaced $1/(1+a^+)$ in **Eq. 1** with $2a'/(1+a')$, in which $0<a'<1$. In the original formulation, the activity of the modulator $a^+ = 0$ during a delay period and non-zero during reset. But in this alternative formulation, the modulator $a' = 1$ during a delay period and zero during reset:

$$\tau_y \frac{dy_j}{dt} = -y_j + \left(\frac{b^+}{1+b^+}\right) z_j + \left(\frac{2a'}{1+a'}\right) \hat{y}_j \ , \quad \text{(S49)}$$

$$\tau_b \frac{db}{dt} = -b + \mathbf{w}_{bx}^t \mathbf{x} \ , \quad \text{(S50)}$$

$$\tau_a \frac{da}{dt} = -a - \sqrt{u} - \sqrt{u}a - \mathbf{w}_{ax}^t \mathbf{x} + 1 \ , \quad \text{(S51)}$$

$$\tau_u \frac{du}{dt} = -u + u\left\|\mathbf{y}\right\|^2 + \left(\frac{\sigma b^+}{1+b^+}\right)^2 \text{ such that } u > \left(\frac{\sigma b^+}{1+b^+}\right)^2 . \quad \text{(S52)}$$

We have also developed and implemented a generalization of this recurrent normalization circuit in which each neuron's response can be normalized by an arbitrary (non-negative) weighted sum of the other neurons in the circuit.

Extended Discussion

Biophysical implementation

A possible starting point for a biophysical implementation of ORGaNICs is depicted in **Fig. S6** (although we are open to the possibility of alternative mechanistic implementations). The key idea is that the two terms corresponding to the input drive and recurrent drive are computed in separate dendritic compartments of a cortical pyramidal cell, so that each compartment can be modulated independently. The model comprises 3 compartments for the soma, the apical dendrite, and the basal dendrite. Each compartment is an RC circuit with a variable-conductance resistor and a variable current source. The capacitors represent the electrical capacitance of the neural membrane. The two fixed-conductance resistors (R_a and R_b) represent the resistances between the compartments (i.e., along the dendritic shafts). The membrane potentials v_s, v_a, and v_b correspond to the soma, the apical dendrite, and the basal

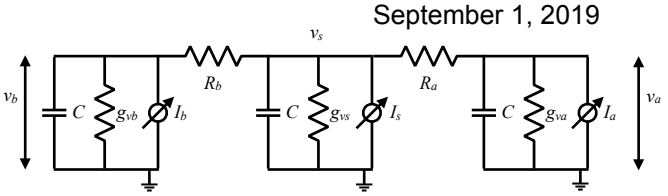

Figure S6. Biophysical implementation. Electrical-circuit model of a pyramidal cell with separate RC circuit compartments for the soma, the apical dendritic tree, and the basal dendritic tree. g_{va}, g_{vb}, shunting synapses represented by variable conductance resistors. I_a, I_b, synaptic input represented by variable current sources. v_s, somatic membrane potential. v_a, v_b, membrane potentials, respectively, in apical and basal dendrites. R_a, R_b, fixed-conductance resistors between compartments (i.e., along the dendritic shafts). C, membrane capacitance.

dendrite, respectively. Some of the synaptic inputs are modeled as conductance changes (g_{va}, g_{vb}, g_{vs}) while others are approximated as currents (I_s, I_a, I_b). To implement ORGaNICs with this pyramidal-cell model, we specified the synaptic inputs (I_s, I_a, I_b, g_{va}, g_{vb}, and g_{vs}) to each neuron in terms of its input drive (y), recurrent drive ($\hat{y}$), and modulators (a and b). We also presumed that the output firing rate of a neuron was well-approximated by halfwave rectification of the membrane potential, and that negative values (corresponding to hyperpolarization of the membrane potential v_s) were represented by a separate neuron that received the complementary synaptic inputs (identical for g_{va} and g_{vb}, and opposite in sign for I_s, I_a, and I_b), analogous to ON and OFF cells (e.g., in the retina, LGN, or V1). Then the steady-state value for the somatic membrane-potential (i.e., when the synaptic inputs are constant) is:

$$gv_s = \left(\frac{b^+}{1+b^+}\right) z + \left(\frac{1}{1+a^+}\right) \hat{y} \qquad \text{(S53)}$$

$$g = g_{vs} + \frac{a^+}{R_a\left(1+a^+\right)} + \frac{b^+}{R_b\left(1+b^+\right)} \ .$$

where g is the total synaptic conductance (for derivation and detailed description of implementation, see 14).

This is identical to the steady-state response of ORGaNICs (compare **Eq. S53** with **Eq. 1**) when the total somatic conductance is g=1. There are a variety of combinations of the various parameters for which the total somatic conductance is approximately equal to 1 (14). Two particular interesting special cases correspond to when the modulators are both on (i.e., equal to 1, such the responses are dominated by the input drive), and when the modulators are both off (i.e., equal to 0, during a delay period). ORGaNICs with multiple recurrent drive terms (see below) can be implemented with separate dendritic compartments, each corresponding to a recurrent weight matrix and recurrent modulator.

This is, of course, a simplified model of pyramidal cells, but plausible nonetheless. First, it has been argued that shunting inhibition does not yield division (15); *in vivo* neurons are rarely at their resting potential because of spontaneous background activity so a shunting synapse (assumed to be a chloride channel) predominantly causes hyperpolarization rather than division. However, exact division can be implemented with two synaptic conductances, one excitatory and one inhibitory, that increase (and decrease) in proportion (16). And there is experimental evidence that cortical circuits are capable of divisive suppression (17, 18). Second, there is no leak conductance in the dendrites. We can think of $g_{vs} = 1$ as the somatic leak conductance, but the model (as expressed above) has no dendritic leak conductances. If we were to add dendritic leak conductances, then the responses would decay during a delay period even when the modulators =0. One could compensate for this decay by scaling the recurrent weight matrix to have eigenvalues larger than one. Third, the input drive and recurrent drive in the model are mediated by synaptic currents, not conductance changes. A large body of the computational neuroscience literature likewise approximates synaptic inputs by current

sources. A push-pull arrangement of synaptic inputs can act like a current source (16). Doing so necessitates a high level of spontaneous activity so that increases in excitation are met with equal decreases in inhibition, and vice versa. But spontaneous activity in most cortical areas, although non-zero, is generally low. Instead, synaptic inputs could approximate current sources when the membrane potential remains far from the (excitatory and inhibitory) synaptic reversal potentials.

Variations and extensions

There is considerable flexibility in the formulation of ORGaNICs, with different variants corresponding to different hypothesized neural circuits (e.g., by a change of variables that might necessitate a different cell type or by including an additional cell type that performs an intermediate step in the computation). In fact, there is a family of dynamical systems models, each of which comprises coupled neural integrators, with similar functionality.

As noted above, there is a family of dynamical systems models, each of which comprises coupled neural integrators to implement normalization via recurrent amplification. Each of the various different models in this family imply different circuits with different dynamics.

A variant of ORGaNICs is capable of prediction over time (14, 19). Information processing in the brain is dynamic; dynamic and predictive processing is needed to control behavior in sync with or in advance of changes in the environment. Without prediction, behavioral responses to environmental events will always be too late because of the lag or latency in sensory and motor processing. Prediction is a key component of theories of motor control and in explanations of how an organism discounts sensory input caused by its own behavior (20-22). Prediction has also been hypothesized to be essential in sensory and perceptual processing (23-25), and in navigation and the planning of action sequences (26).

ORGaNICs build on Heeger's Theory of Cortical Function (TCF) (19) that offers a framework for understanding how the brain accomplishes three key functions: (i) inference: perception is nonconvex optimization that combines sensory input with prior expectation; (ii) exploration: inference relies on neural response variability to explore different possible interpretations; (iii) prediction: inference includes making predictions over a hierarchy of timescales. TCF has a single modulator for all of the neurons in each layer whereas ORGaNICs may have a separate pair of modulators, a_j and b_j, for each neuron y_j. ORGaNICs also have a more general form for the recurrent weight matrix. But TCF includes a feedback drive across the layers of a stacked architecture, in addition to the input drive and recurrent drive. In some states (depending on the values of the modulators), neural responses are dominated by the feedforward drive and TCF is identical to a conventional feedforward model (e.g., deep net), thereby preserving all of the desirable features of those models. In other states, TCF is a generative model that constructs a sensory representation from an abstract representation, like memory recall. In still other states, TCF combines prior expectation with sensory input, explores different possible perceptual interpretations of ambiguous sensory inputs, and predicts forward in time. A variant of ORGaNICs may be stacked (like TCF) to include feedback connections and the capability of a generative model, but with greater flexibility and computational power because of the general form for the recurrent weight matrix, and because there may be a separate pair of modulators for each output neuron.

There are various formulations for how to compute the modulators. According to **Eq. 2**, the recurrent modulators are a nonlinear function of the responses (see above for details), but could instead be linear sums of the responses: $\mathbf{W}_{ay}\,\mathbf{y}$ and $\mathbf{W}_{by}\,\mathbf{y}$. Other variants (particularly for complex-valued responses) compute linear sums of the modulus of the responses, $\mathbf{W}_{ay}\,|\mathbf{y}|$, or

linear sums of the various readouts: $\mathbf{W}_{ay}\,\mathbf{r}$, $\mathbf{W}_{ay}\,\mathbf{r}^{+}$, $\mathbf{W}_{ay}\,\mathbf{r}^{\pm}$. And likewise for the input modulators. According to **Eq. 1**, there is a separate pair modulators a_j and b_j for each neuron, but this need not be the case. Subpopulations of neurons might share some modulators. For example, all 36 neurons shared a single pair of modulators in the various example circuits above. Another option would be to have a number of basis modulators that are shared:

$$a_j = \sum_k^K w_{jk} a_k \quad , \tag{S54}$$

where a_{k} are the responses of the basis modulators, w_{jk} are weights, and the number of basis modulators K is less then the number of neurons N. And likewise for the input modulators b_j. This idea is commensurate with an analysis of experimental data demonstrating a small number of separate time-varying modulatory signals, shared across a population neurons (27). It also dovetails with the observation that some of the modulators are computed via thalamocortical loops (28-31), because the number of neurons in a thalamic nucleus is much smaller than the number of neurons in the corresponding cortical area. In yet another variant, the two modulators are coordinated to govern balance between input drive and recurrent drive, like λ in **Eq. S5**.

Optionally, the encoding and readout weights may have components that are orthogonal to the eigenvectors of the recurrent weight matrix (see above for details), and non-zero spontaneous firing rates can be accommodated by adding offsets to each neuron's input drive (14). For example, if there is a non-zero offset added to the recurrent drive, then the corresponding component of the responses will reflect the elapsed time interval since the beginning of a delay period.

Left open is how to determine the weights in the various weight matrices: the encoding matrix ($\mathbf{W}_{zx}$), the recurrent weight matrix ($\mathbf{W}_{\hat{y}y}$), the readout matrix ($\mathbf{W}_{ry}$), and the modulator weight matrices ($\mathbf{W}_{ax}$, $\mathbf{W}_{bx}$). Some of the weight matrices (e.g., the recurrent weights) might be prespecified, emerging during development, resulting in stereotypical connectivity within a local circuit. A supervised learning approach would estimate the weights via gradient descent (i.e., back propagation), given target values for the response time-courses (or the readout time-courses). To use back propagation, the recurrent weight matrix needs to be rescaled after each gradient step so that to the eigenvalue with the largest real part is no greater than one. But the brain is unlikely to have access to such target values sampled over time. Another approach would be an unsupervised learning algorithm based on minimizing prediction error over time (19).

The readouts were different for each of the example circuits. For the sustained activity circuits (**Figs. 2** and **6**), linear readouts reconstructed the input drive (and the target location). For the circuit exhibiting oscillatory dynamics (**Fig. S4**), a (sign-corrected) nonlinear readout was used to reconstruct. For the motor control circuits (**Figs. 5** and **7**), a different linear readout was used to convert spatial patterns of input (premotor) activity to temporal profiles of output (motor control) activity. We did not attempt to reconstruct the input stimulus from the readout, because recovering the input is not the goal for motor control. Likewise, we did not reconstruct the input from the sequential activity circuit (**Fig. 4**); The modulus readout enabled the output to be constant over time (i.e., supporting maintenance), but this readout was not capable of reconstructing the input drive. We do not, in general, mean to imply that the brain attempts to reconstruct an input stimulus from the responses. Reconstruction is useful for short-term memory, but not for other cognitive functions. Rather, we hypothesize that the brain relies on a

set of canonical (nonlinear) neural computations, repeating them across brain regions so that each stage of computation transforms the information it receives (32-36). The readout is part of this nonlinear transformation from one stage of processing to the next.

ORGaNICs can be combined with so-called “activity-silent” models of working memory (37-40). Some electrophysiological studies have found that activity can be modest, or even entirely absent, during a working memory delay period (7, 38, 41). According to “activity-silent” models of working memory, information is maintained via short-term synaptic plasticity (STSP). This idea is complementary to, not mutually exclusive of persistent (sustained, sequential, or oscillating) activity. Specifically, the third line of **Eq. 1** can be modified to include STSP:

$$\hat{\mathbf{y}} = \mathbf{W}_{\hat{y}y}\mathbf{D}(\mathbf{g})\mathbf{y} \quad . \tag{S55}$$

The value of g_j is the synaptic gain of the j^{th} neuron, presumed to depend on the recent history of that neuron’s firing rate (37, 42).

Implications for neuroscience

We propose that ORGaNICs can serve as a unifying theoretical framework for neural dynamics, a canonical computational motif based on recurrent amplification, gated integration, reset, and controlling the effective time constant. Rethinking cortical computation in these terms should have widespread implications, some of which are elucidated in the *Discussion*. Some additional implications are as follows.

1) Sequential activity and internal models. ORGaNICs can be used to maintain and manipulate information over time. One of our example circuits encoded a representation of a pair of targets and then updated the representation (of the remembered locations) coincident with an eye movement (**Fig. 6**). This process of updating the representation of a visual target has been called remapping (43-45). The responses during the update exhibited a traveling wave of activity (i.e., sequential activity), commensurate with measurements of neural activity during remapping (46). We can think of this as an internal model (also called a forward model) of the movement. Internal models are hypothesized to predict the result of a movement plan and to explain how an organism discounts sensory input caused by its own behavior (20-22, 47). Our simulation results are also reminiscent of an ostensibly different class of phenomena: internally-generated sequential activity in parietal cortex, prefrontal cortex, and hippocampus during navigation, motor planning, and episodic recall (26, 42, 48, 49). We hypothesize that these forms of sequential activity can be explained by the same gated integration computation that we used for spatial updating, i.e., updating an internal model.

2) Dimensionality-reduction. Sensory, cognitive, and motor functions depend on the interactions of many neurons. It has become increasingly popular to use dimensionality-reduction algorithms to analyze measurements of neural activity recorded simultaneously in large numbers of neurons (50, 51). These data-analysis algorithms presume that the neural activity over time is confined to a low-dimensional manifold, much lower than the number of neurons recorded. The simplest of these algorithms use principal components analysis or factor analysis, thereby presuming that the neural activity is confined to a linear subspace. The simulated responses of ORGaNICs are indeed confined to a linear subspace at any moment in time, but the subspace changes over time with changes in state. The responses are in the subspace of the encoding matrix when the modulators are both large. The responses are in the subspace of the eigenvectors of the recurrent weight matrix (with corresponding eigenvalues that have real parts =1) when the input modulator b=0 and when $\mathbf{W}_{ax}\,\mathbf{x} = 0$ for long enough such that the other components decay. The responses can be in any of an infinite number of subspaces for intermediate

values of the modulators and with non-zero inputs.

3) Oscillations and brain state. We found that changes in state are associated with measurable changes in response dynamics (**Fig. 3b**). Simulated responses exhibited high frequency oscillations following target onset, and lower frequency oscillations during the delay period, that were synchronized across all neurons in the circuit. The change in oscillation frequency corresponded to a change in state, induced by changing the modulator responses. This is commensurate with experimental observations that different brain states are associated with particular oscillation frequencies (52-61). The high-frequency oscillations emerged in our simulations because of the multiplicative modulator underlying normalization; the oscillation frequency depended on the membrane time constants, and would have depended also on axon length if we were to include conduction delays. The emergence of high-frequency oscillations dovetails with observations that gamma oscillations are linked to normalization (62-68). But this multiplicative (i.e., nonlinear) process is different from most previous models of gamma oscillations (65, 69-74), which fundamentally rely on linear recurrent neural oscillators (see above for a primer on neural oscillators).

4) E:I balance. The neurons in each of the example circuits in this paper all had the same values for the time constants. That can be generalized, e.g., so that inhibitory neurons have a slower time constant (and consequently delayed responses) compared to excitatory neurons (14). If different neurons in the circuit have different time constants, then the value(s) of the time constant(s), in combination with the eigenvalues of the recurrent weight matrix, determine whether or not there is sustained activity or oscillations, whether the oscillations are stable or decay, and the frequencies of the oscillations. Consequently, the stability of ORGaNICs (and related neural integrator circuits) depends on a combination of the recurrent weight matrix and the relative values of the intrinsic time constants. A differential delay between excitatory and inhibitory neurons can compensate for an imbalance in the excitatory versus inhibitory synaptic weights, and vice versa.

Implications for machine learning

1) Go complex. Simple harmonic motion is everywhere. For many machine learning (ML) applications (e.g., speech processing, music processing, analyzing human movement), the dynamics of the input signals may be characterized with damped oscillators, in which the amplitudes, frequencies and phases of the oscillators may change over time. The complex-valued weights and responses in ORGaNICs are well-suited for these kinds of signals. Likewise, we propose using damped-oscillator basis functions as a means for predicting forward in time (14, 19). There has been relatively little focus on complex-valued recurrent neural networks (75-89), and even less on complex-valued LSTMs (90-94).

2) Stability and vanishing gradients. To ensure stability and to avoid exploding gradients during learning, the recurrent weight matrix may be rescaled so that the eigenvalue with the largest real part is no larger than 1. This rescaling may be added as an extra step during learning after each gradient update. Doing so should help to avoid vanishing gradients by using halfwave rectification instead of a sigmoidal output nonlinearity (93).

3) Normalization. Incorporating normalization can make the computation robust with respect to imperfections in the recurrent weight matrix (**Fig. 3**). Normalization maintains the ratios of the responses (**Eq. 4**), unlike sigmoids or other static output nonlinearities (also called transfer functions) that are typically used in ML systems.

4) Multiple recurrent terms. The modulators in ORGaNICs, analogous to the input and reset

gates in LSTMs, perform multiple functions (see *Discussion*). ORGaNICs may have multiple recurrent weight matrices, each multiplied by different recurrent modulators, to perform combinations of these functions (**Eq. 6** and **Figs. 6-7**). This is unlike an LSTM that has only a single recurrent weight matrix and a single reset gate.

5) Time warping. ORGaNICs offer a means for time warping (**Fig. 7**). Invariance with respect to compression or dilation of temporal signals (e.g., fast vs. slow speech) is a challenge for many ML applications (95). ML systems typically attempt to circumvent this problem by learning models with every possible tempo. ORGaNICs might be applied to solve this problem much more efficiently, eliminating redundancy and increasing generalization, with less training.

6) Neuromorphic implementation. Given the biophysical (equivalent electrical circuit) implementation of ORGaNICs (**Fig. S6**), it may be possible to design and fabricate analog VLSI ORGaNICs chips. Analog circuitry may be more energy-efficient in comparison to representing and processing information digitally (96, 97). Such an analog electrical-circuit may be configured to download various parameter settings (e.g., the weight matrices), computed separately offline.

References

1. E. P. Simoncelli (1988) Orthogonal sub-band image transforms, MSE thesis. in *MIT Dept. of Electrical Engineering and Computer Science,* .
2. E. P. Simoncelli, E. H. Adelson, "Subband Transforms" in Subband Image Coding, J. W. Woods, Ed. (Kluwer Academic Publishers, Norwell, MA, 1990), pp. 143-192.
3. W. T. Freeman, E. H. Adelson, The design and use of steerable filters. *IEEE Pattern Analysis and Machine Intelligence* **13**, 891-906 (1991).
4. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, D. J. Heeger, Shiftable multi-scale transforms. *IEEE Transactions on Information Theory, Special Issue on Wavelets* **38**, 587-607 (1992).
5. C. D. Brody, A. Hernandez, A. Zainos, R. Romo, Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb Cortex* **13**, 1196-1207 (2003).
6. C. K. Machens, R. Romo, C. D. Brody, Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* **307**, 1121-1124 (2005).
7. M. Shafi *et al.*, Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* **146**, 1082-1108 (2007).
8. D. A. Markowitz, C. E. Curtis, B. Pesaran, Multiple component networks support working memory in prefrontal cortex. *Proc Natl Acad Sci U S A* **112**, 11084-11089 (2015).
9. D. Kobak *et al.*, Demixed principal component analysis of neural population data. *Elife* **5** (2016).
10. J. D. Murray *et al.*, Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc Natl Acad Sci U S A* **114**, 394-399 (2017).
11. K. Zhang, Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci* **16**, 2112-2126 (1996).
12. E. P. Simoncelli, "Design of multi-dimensional derivative filters" in International Conference on Image Processing. (1994), vol. I, pp. 790-793.
13. S. Druckmann, D. B. Chklovskii, Neuronal Circuits Underlying Persistent Representations Despite Time Varying Activity. *Current Biology* **22**, 2095-2103 (2012).
14. D. J. Heeger, W. E. Mackey, ORGaNICs: A Theory of Working Memory in Brains and Machines. *arXiv preprint arXiv:1803.06288* (2018).

15. G. R. Holt, C. Koch, Shunting inhibition does not have a divisive effect on firing rates. *Neural Comput* **9**, 1001-1013 (1997).

16. M. Carandini, D. J. Heeger, Summation and division by neurons in primate visual cortex. *Science* **264**, 1333-1336 (1994).

17. N. R. Wilson, C. A. Runyan, F. L. Wang, M. Sur, Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature* **488**, 343-348 (2012).

18. B. A. Seybold, E. A. K. Phillips, C. E. Schreiner, A. R. Hasenstaub, Inhibitory Actions Unified by Network Integration. *Neuron* **87**, 1181-1192 (2015).

19. D. J. Heeger, Theory of cortical function. *Proc Natl Acad Sci U S A* **114**, 1773-1782 (2017).

20. D. M. Wolpert, Z. Ghahramani, M. I. Jordan, An internal model for sensorimotor integration. *Science* **269**, 1880-1882 (1995).

21. M. Kawato, Internal models for motor control and trajectory planning. *Curr Opin Neurobiol* **9**, 718-727 (1999).

22. T. B. Crapse, M. A. Sommer, Corollary discharge across the animal kingdom. *Nat Rev Neurosci* **9**, 587-600 (2008).

23. W. Bialek, I. Nemenman, N. Tishby, Predictability, complexity, and learning. *Neural Computation* **13**, 2409-2463 (2001).

24. R. Nijhawan, Visual prediction: psychophysics and neurophysiology of compensation for time delays. *Behav Brain Sci* **31**, 179-198; discussion 198-239 (2008).

25. S. E. Palmer, O. Marre, M. J. Berry, 2nd, W. Bialek, Predictive information in a sensory population. *Proc Natl Acad Sci U S A* **112**, 6908-6913 (2015).

26. E. Pastalkova, V. Itskov, A. Amarasingham, G. Buzsaki, Internally generated cell assembly sequences in the rat hippocampus. *Science* **321**, 1322-1327 (2008).

27. N. C. Rabinowitz, R. L. Goris, M. Cohen, E. P. Simoncelli, Attention stabilizes the shared gain of V4 populations. *Elife* **4**, e08998 (2015).

28. Z. V. Guo *et al.*, Maintenance of persistent activity in a frontal thalamocortical loop. *Nature* **545**, 181-186 (2017).

29. L. I. Schmitt *et al.*, Thalamic amplification of cortical connectivity sustains attentional control. *Nature* **545**, 219-223 (2017).

30. R. V. Rikhye, A. Gilra, M. M. Halassa, Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nat Neurosci* **21**, 1753-1763 (2018).

31. R. V. Rikhye, R. D. Wimmer, M. M. Halassa, Toward an Integrative Theory of Thalamic Function. *Annual review of neuroscience* **41**, 163-183 (2018).

32. D. J. Heeger, E. P. Simoncelli, J. A. Movshon, Computational models of cortical visual processing. *Proc Natl Acad Sci U S A* **93**, 623-627 (1996).

33. E. P. Simoncelli, D. J. Heeger, A model of neuronal responses in visual area MT. *Vision Res* **38**, 743-761 (1998).

34. M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex. *Nat Neurosci* **2**, 1019-1025 (1999).

35. M. Riesenhuber, T. Poggio, Neural mechanisms of object recognition. *Curr Opin Neurobiol* **12**, 162-168 (2002).

36. M. Carandini, D. J. Heeger, Normalization as a canonical neural computation. *Nat Rev Neurosci* **13**, 51-62 (2011).

37. G. Mongillo, O. Barak, M. Tsodyks, Synaptic theory of working memory. *Science* **319**, 1543-1546 (2008).

38. O. Barak, M. Tsodyks, R. Romo, Neuronal population coding of parametric working memory. *J Neurosci* **30**, 9424-9430 (2010).

39. O. Barak, M. Tsodyks, Working models of working memory. *Curr Opin Neurobiol* **25**, 20-24 (2014).

40. M. G. Stokes, 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci* **19**, 394-405 (2015).

41. G. Rainer, E. K. Miller, Timecourse of object-related neural activity in the primate prefrontal cortex during a short-term memory task. *Eur J Neurosci* **15**, 1244-1254 (2002).

42. S. Fujisawa, A. Amarasingham, M. T. Harrison, G. Buzsaki, Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat Neurosci* **11**, 823-833 (2008).

43. J. R. Duhamel, C. L. Colby, M. E. Goldberg, The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* **255**, 90-92. (1992).

44. H. M. Rao, J. P. Mayo, M. A. Sommer, Circuits for presaccadic visual remapping. *J Neurophysiol* **116**, 2624-2636 (2016).

45. A. C. Marino, J. A. Mazer, Saccades Trigger Predictive Updating of Attentional Topography in Area V4. *Neuron* **98**, 429-438 e424 (2018).

46. X. Wang *et al.*, Perisaccadic Receptive Field Expansion in the Lateral Intraparietal Area. *Neuron* **90**, 400-409 (2016).

47. M. I. Jordan, D. E. Rumelhart, Forward models: Supervised learning with a distal teacher. *Cognitive Science* **16**, 307-354 (1992).

48. C. J. MacDonald, K. Q. Lepage, U. T. Eden, H. Eichenbaum, Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron* **71**, 737-749 (2011).

49. C. D. Harvey, P. Coen, D. W. Tank, Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62-68 (2012).

50. J. P. Cunningham, B. M. Yu, Dimensionality reduction for large-scale neural recordings. *Nat Neurosci* **17**, 1500-1509 (2014).

51. C. Pandarinath *et al.*, Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat Methods* **15**, 805-815 (2018).

52. W. Klimesch, EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res Brain Res Rev* **29**, 169-195 (1999).

53. N. Kopell, G. B. Ermentrout, M. A. Whittington, R. D. Traub, Gamma rhythms and beta rhythms have different synchronization properties. *Proc Natl Acad Sci U S A* **97**, 1867-1872 (2000).

54. A. K. Engel, P. Fries, W. Singer, Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* **2**, 704-716 (2001).

55. B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, R. A. Andersen, Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat Neurosci* **5**, 805-811 (2002).

56. J. Csicsvari, B. Jamieson, K. D. Wise, G. Buzsaki, Mechanisms of gamma oscillations in the hippocampus of the behaving rat. *Neuron* **37**, 311-322 (2003).

57. G. Buzsaki, A. Draguhn, Neuronal oscillations in cortical networks. *Science* **304**, 1926-1929 (2004).

58. G. Buzsaki, *Rhythms of the Brain* (Oxford University Press, 2006).

59. A. P. Maurer, B. L. McNaughton, Network and intrinsic cellular mechanisms underlying theta phase precession of hippocampal neurons. *Trends Neurosci* **30**, 325-333 (2007).

60. S. Palva, J. M. Palva, New vistas for alpha-frequency band oscillations. *Trends Neurosci* **30**, 150-158 (2007).

61. M. Lundqvist *et al.*, Gamma and Beta Bursts Underlie Working Memory. *Neuron* **90**, 152-164 (2016).

62. J. A. Henrie, R. Shapley, LFP power spectra in V1 cortex: the graded effect of stimulus contrast. *Journal of neurophysiology* **94**, 479-490 (2005).

63. M. A. Gieselmann, A. Thiele, Comparison of spatial integration and surround suppression characteristics in spiking activity and the local field potential in macaque V1. *The European journal of neuroscience* **28**, 447-459 (2008).

64. S. Ray, J. H. Maunsell, Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* **9**, e1000610 (2011).

65. X. Jia, D. Xing, A. Kohn, No consistent relationship between gamma power and peak frequency in macaque primary visual cortex. *J Neurosci* **33**, 17-25 (2013).

66. S. Ray, A. M. Ni, J. H. Maunsell, Strength of gamma rhythm depends on normalization. *PLoS Biol* **11**, e1001477 (2013).

67. D. Hermes, K. J. Miller, B. A. Wandell, J. Winawer, Stimulus Dependence of Gamma Oscillations in Human Visual Cortex. *Cereb Cortex* **25**, 2951-2959 (2015).

68. M. W. Self *et al.*, The Effects of Context and Attention on Spiking Activity in Human Early Visual Cortex. *PLoS Biol* **14**, e1002420 (2016).

69. M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, B. L. McNaughton, Paradoxical effects of external modulation of inhibitory interneurons. *J Neurosci* **17**, 4382-4388 (1997).

70. H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, D. Ferster, Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron* **62**, 578-592 (2009).

71. K. Kang, M. Shelley, J. A. Henrie, R. Shapley, LFP spectral peaks in V1 cortex: network resonance and cortico-cortical feedback. *Journal of computational neuroscience* **29**, 495-507 (2010).

72. X. J. Wang, Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol Rev* **90**, 1195-1268 (2010).

73. G. Buzsaki, X. J. Wang, Mechanisms of gamma oscillations. *Annual review of neuroscience* **35**, 203-225 (2012).

74. D. Xing *et al.*, Stochastic generation of gamma-band activity in primary visual cortex of awake and anesthetized monkeys. *J Neurosci* **32**, 13873-13880a (2012).

75. B. Widrow, J. McCool, M. Ball, The complex LMS algorithm. *Proceedings of the IEEE* **63**, 719-720 (1975).

76. C. Li, X. Liao, J. Yu, Complex-valued recurrent neural network with IIR neuron model: training and applications. *Circuits, Systems and Signal Processing* **21**, 461-471 (2002).

77. S. L. Goh, D. P. Mandic, An augmented extended Kalman filter algorithm for complex-valued recurrent neural networks. *Neural Computation* **19**, 1039-1055 (2007).

78. D. P. Mandic, V. S. L. Goh, Complex valued nonlinear adaptive filters: noncircularity, widely linear and neural models (John Wiley & Sons, 2009), vol. 59.

79. T. Nitta, Complex-Valued Neural Networks: Utilizing High-Dimensional Parameters: Utilizing High-Dimensional Parameters (IGI Global, 2009).

80. W. Zhou, J. M. Zurada, Discrete-time recurrent neural networks with complex-valued linear threshold neurons. *IEEE Transactions on Circuits and Systems II: Express Briefs* **56**, 669-673 (2009).

81. H.-G. Zimmermann, A. Minin, V. Kusherbaeva (2011) Historical consistent complex valued recurrent neural network. in *International Conference on Artificial Neural Networks* (Springer), pp 185-192.

82. A. Hirose, *Complex-valued neural networks* (Springer Science & Business Media, 2012), vol. 400.

83. J. Hu, J. Wang, Global stability of complex-valued recurrent neural networks with time-delays. *IEEE Transactions on Neural Networks and Learning Systems* **23**, 853-865 (2012).

84. A. Minin, A. Knoll, H.-G. Zimmermann, Complex valued recurrent neural network: From architecture to training. *Journal of Signal and Information Processing* **3**, 192 (2012).

85. T. Fang, J. Sun, Stability of complex-valued recurrent neural networks with time-delays. *IEEE Transactions on Neural Networks and Learning Systems* **25**, 1709-1713 (2014).

86. Z. Zhang, C. Lin, B. Chen, Global stability criterion for delayed complex-valued recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems* **25**, 1704-1708 (2014).

87. W. Gong, J. Liang, J. Cao, Matrix measure method for global exponential stability of complex-valued recurrent neural networks with time-varying delays. *Neural Networks* **70**, 81-89 (2015).

88. R. Rakkiyappan, J. Cao, G. Velmurugan, Existence and uniform stability analysis of fractional-order complex-valued neural networks with time delays. *IEEE Transactions on Neural Networks and Learning Systems* **26**, 84-97 (2015).

89. A. M. Sarroff, V. Shepardson, M. A. Casey, Learning representations using complex-valued nets. *arXiv preprint arXiv:1511.06351* (2015).

90. M. Arjovsky, A. Shah, Y. Bengio (2016) Unitary evolution recurrent neural networks. in *International Conference on Machine Learning*, pp 1120-1128.

91. I. Danihelka, G. Wayne, B. Uria, N. Kalchbrenner, A. Graves, Associative long short-term memory. *arXiv preprint arXiv:1602.03032* (2016).

92. S. Wisdom, T. Powers, J. Hershey, J. Le Roux, L. Atlas (2016) Full-capacity unitary recurrent neural networks. in *Advances in Neural Information Processing Systems*, pp 4880-4888.

93. L. Jing *et al.*, Gated orthogonal recurrent units: On learning to forget. *arXiv preprint arXiv: 1706.02761* (2017).

94. C. Jose, M. Cisse, F. Fleuret, Kronecker Recurrent Units. *arXiv preprint arXiv:1705.10142* (2017).

95. C. Tallec, Y. Ollivier (2018) Can Recurrent Neural Networks warp time? in *International Conference on Learning Representations*.

96. C. Mead, Neuromorphic Electronic Systems. *Proceedings of the Ieee* **78**, 1629-1636 (1990).

97. R. Sarpeshkar, Analog versus digital: extrapolating from electronics to neurobiology. *Neural Comput* **10**, 1601-1638 (1998).