A theory of multi-task computation and task selection

Owen Marschall[*1,2], David G. Clark[†1,2], and Ashok Litwin-Kumar[‡1,2]

[1]Kavli Institute for Brain Science, Columbia University, New York, NY 10027
[2]Zuckerman Institute, Columbia University, New York, NY 10027

April 15, 2026

Abstract

Neural activity during the performance of a stereotyped behavioral task is often described as low-dimensional, occupying only a limited region in the space of all firing-rate patterns. This region has been referred to as the "neural manifold" associated with a task. More recently, recordings of neural activity in animals challenged to perform multiple tasks have suggested that each task is associated with a different low-dimensional manifold. What connectivity structures underlie this flexibility in neural dynamics, and how is interference between the dynamics associated with different tasks avoided? We develop a theoretical model for multi-task computation in nonlinear recurrent neural networks whose connectivity is constructed as a weighted sum of many low-rank components, each encoding the dynamics associated with a different task. The model demonstrates that interference between different tasks' dynamics limits flexible multi-tasking and can lead to chaotic fluctuations. However, small modulations of a network's effective connectivity overcome this interference. We derive the conditions that enable such task selection and characterize both single-neuron and population statistics in task-selected and unselected states. The model reveals the requirements for a single network to produce distinct dynamics confined to distinct neural manifolds and suggests circuit mechanisms that support this capability. Using the model, we propose different hypotheses for explaining the origin of high-dimensional neural activity in large-scale recordings.

Introduction

Adaptive behavior requires organisms to perform different neural computations at different times or in different contexts. For example, both spontaneous and skilled, practiced movements involve sequencing discrete motor computations [1, 2]. Similarly, cognitive tasks require operations on task-relevant variables that may differ across task, condition, or epoch [3–7]. Numerous studies support the proposal that individual neural computations are associated with coordinated dynamics of interconnected populations of neurons [8], implying that the effective dynamical system governing these neurons' activity must rapidly

[*]om2382@columbia.edu (OM)
[†]Present address: Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, MA 02138, USA
[‡]a.litwin-kumar@columbia.edu (ALK)

change when a new computation is to be performed. Little is known about the circuit mechanisms that underlie these dynamical switches.

During execution of a single task, population activity is typically confined to a low-dimensional neural manifold [9, 10]. Mixed selectivity—the observation that individual neurons often participate in multiple distinct tasks [3, 5, 11, 12], although not always [13]—implies that these manifolds are distributed across many neurons rather than aligned to a small subset. This distributed structure is reproduced in recurrent neural networks trained to perform multiple tasks [4, 14]. Recent work shows that the location and orientation of the manifold can change depending on the task being performed [2, 15–18]. Thus, switching between manifolds, each distributed across the population, may underlie the rapid changes in neural dynamics and behavior required for switching tasks [2].

In another line of work, neural activity has been observed to be high-dimensional during periods of spontaneous, uninstructed behavior [19, 20], in contrast to observations of low-dimensional activity during the performance of stereotyped tasks [21, 22]. These observations may be compatible with the network switching among many distinct manifolds, each associated with a different self-initiated behavior [1, 23]. The union of many such individually low-dimensional manifolds can be high-dimensional if they are heterogeneously oriented in neural state space, as recent experimental work suggests [2]. Alternatively, such activity may reflect unstructured, high-dimensional fluctuations, as proposed by random-network models of the spontaneous state [24, 25].

Altogether, these observations raise several theoretical questions. What forms of recurrent connectivity support many distinct task-related manifolds, with mixed selectivity to different tasks at the single-neuron level? Under what conditions do the dynamics associated with each manifold interfere or not interfere? How does a circuit selectively engage the dynamics appropriate for one task while suppressing others? What conditions lead to high-dimensional neural activity?

Prior work has addressed multi-tasking by training recurrent neural networks on multiple tasks using machine learning methods and then analyzing the solutions [3, 4, 14, 26, 27]. Such studies have shown that artificial networks reuse dynamical motifs, such as fixed points and ring attractors, when the tasks have shared component subtasks. We take a complementary approach, using a solvable network model rather than trained networks. For simplicity, we assume each "task" is an irreducible computation that can be solved by a single low-dimensional dynamical system, without compositional overlap [6].

Specifically, we propose and analyze a mathematically tractable model of multi-tasking neural circuits, in which distinct task-related dynamics operate on distinct low-dimensional activity manifolds. Building on low-rank network models that explain how recurrent structure shapes low-dimensional dynamics [28, 29], we model networks whose weight matrix is constructed as a sum of many low-rank single-task matrices [30]. Because the network dynamics are nonlinear, this leads to complex interactions between the dynamics associated with different tasks. We develop a dynamical mean-field theory that characterizes how tasks interfere, how spontaneous fluctuations arise when a system has learned to perform many distinct tasks, and how these dynamics manifest at both the single-neuron and population levels. Our model makes predictions about single-neuron tuning and fluctuations, trial-to-trial variability in neural activity projected onto task-related dimensions, and the dimensionality of population activity that can be examined in experimental data.

A key feature of the model is that small modulations of connectivity can lead to the dynamic selection

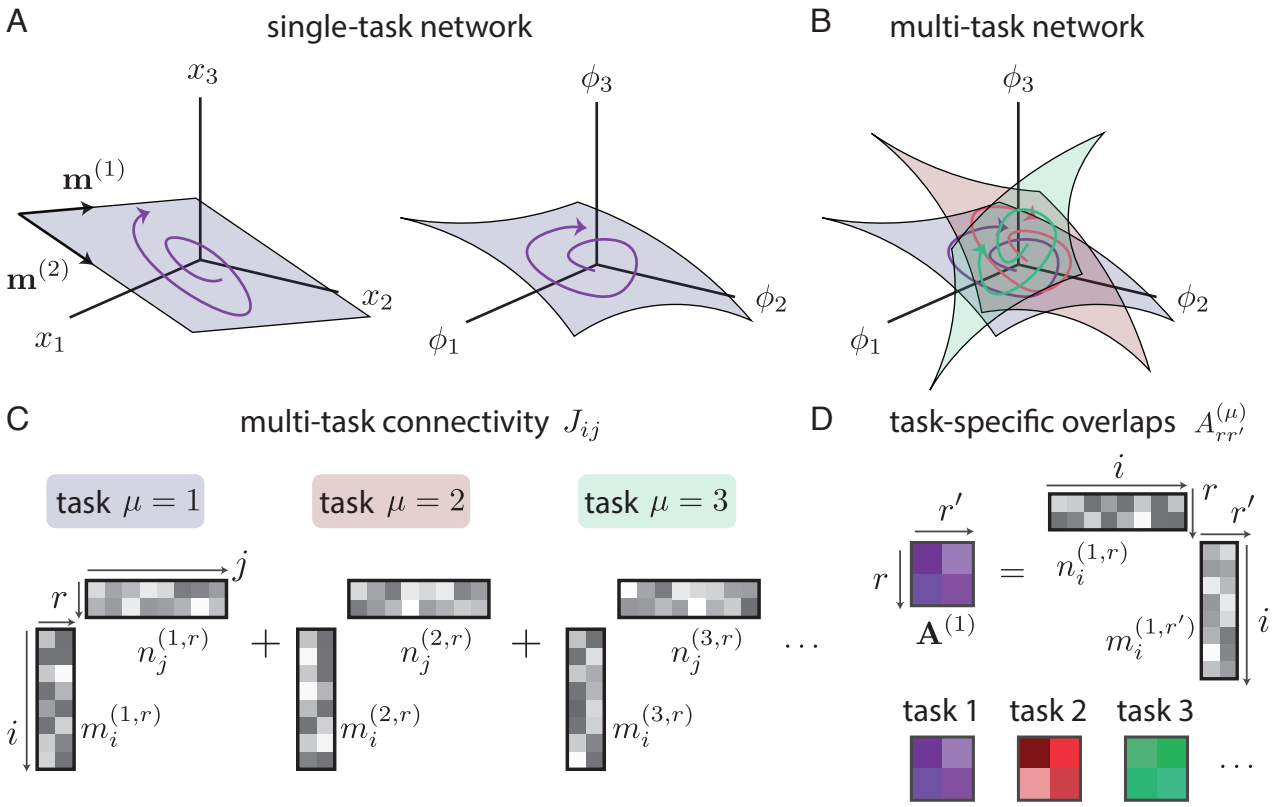

Figure 1: Single-task and multi-task network dynamics. **(A)** Left: Illustration of a state-space trajectory in the space of neuronal pre-activations (purple), confined to a task-relevant two-dimensional subspace. Such a geometry arises in rank-two networks with loading vectors $\mathbf{m}^{(1)}$, $\mathbf{m}^{(2)}$. Right: Illustration of the same trajectory in firing-rate space. Activity lies on a two-dimensional nonlinear manifold. **(B)** Schematic of a "multi-task" network that flexibly implements multiple low-dimensional dynamics. Each task is associated with a different nonlinear manifold, which collectively span more activity dimensions than any one manifold. **(C)** Schematic of multi-task network connectivity as a sum of low-rank matrices, each associated with a different task. **(D)** Each task is associated with an overlap matrix comprising pairwise inner products among the left and right loading vectors.

of a single task and the suppression of others. This provides a biologically plausible mechanism for task selection. In particular, our proposal is consistent with evidence that external inputs, for example thalamic inputs, gate and sustain motor cortical activity [31, 32]. More generally, task execution may result from coordination between recurrent dynamics and modulatory pathways. In total, we provide a theory relating connectivity, dynamics, and experimentally measurable population signatures of multi-tasking in recurrent neural circuits.

Results

We study nonlinear recurrent neural networks of N continuous-time rate neurons with currents $x_i(t)$ and firing rates $\phi_i(t)$ that evolve according to

$$\dot{x}_i(t) = -x_i(t) + \sum_{j=1}^{N} J_{ij}\phi_j(t), \tag{1}$$

with a sigmoid-shaped firing-rate function (taken to be $\phi_i(t) = \mathrm{erf}\big(\sqrt{\pi}x_i(t)/2\big)$ in simulations). The network dynamics are controlled by the $N \times N$ weight matrix $\mathbf{J}$ with elements J_{ij}. We assume that N is large, and our theoretical results are exact in the $N \to \infty$ limit. We now discuss the dynamics that occur when $\mathbf{J}$ is low-rank, before presenting our model in which many such low-rank matrices are superposed.

Low-rank networks

Many previous works have considered the case where the recurrent connectivity $\mathbf{J}$ is low-rank [28, 29, 33, 34], given by

$$J_{ij} = D\sum_{r=1}^{R} m_i^{(r)} n_j^{(r)}, \tag{2}$$

a sum over a small number $R \ll N$ of outer products of N-dimensional *loading vectors* $\mathbf{m}^{(r)}, \mathbf{n}^{(r)}$, scaled by a constant $D \geq 0$. The right loading vectors $\mathbf{n}^{(r)}$ define the subspace of neural firing rates that determines the recurrent input to the network, and this input lies in a subspace spanned by the left loading vectors $\mathbf{m}^{(r)}$ (Fig. 1A, left). When transformed by the firing-rate nonlinearity, this R-dimensional input subspace produces neural activity that lies on an R-dimensional nonlinear manifold (Fig. 1A, right).

We consider the special case of Gaussian-distributed loading-vector components with zero mean and specified covariance structure. All components are generated i.i.d. across different neurons. For a given neuron i, the left- and right-loading-vector components are sampled with nonzero covariances $\langle n_i^{(r)} m_i^{(r')} \rangle = A_{rr'}/N$ according to the entries of the $R \times R$ *overlap matrix* $\mathbf{A}$. As such, this matrix encodes the loading vectors' geometrical arrangement $\langle \mathbf{n}^{(r)} \cdot \mathbf{m}^{(r')} \rangle = A_{rr'}$ and plays a major role in determining the network's dynamics.

Activity in a low-rank network can be described by a set of *latent variables*, defined by projections onto the right loading vectors, $z_r(t) = D(\mathbf{n}^{(r)} \cdot \boldsymbol{\phi}(t))$. In the $N \to \infty$ limit, the overlap matrix determines the dynamics among the latent variables according to

$$z_r(t) = D\left\langle \phi'(t) \right\rangle \sum_{r'=1}^{R} A_{rr'} \left[L * z_{r'} \right](t), \tag{3}$$

where $L(\tau)$ is a low-pass filter with the single-neuron time constant (see Methods). The gain factor $\langle \phi'(t) \rangle$ is the population-averaged slope of the firing-rate nonlinearity. It depends on the norm of the low-pass-filtered latent variables, with larger norm reducing the gain due to the sigmoidal form of $\phi(\cdot)$ (see Methods). Eq. (3) summarizes a low-rank network's activity as an R-dimensional, nonlinear dynamical system [28, 29]. The gain factor $\langle \phi'(t) \rangle$ is the only nonlinearity in this system and thus is necessary for

nonlinear dynamical behaviors such as stable limit cycles or multistability.

Multi-task low-rank networks

Neural recordings suggest that the activity of individual circuits can switch among distinct neural manifolds that collectively span a large number of dimensions [Fig. 1B; 2, 16]. When measured across many different contexts, such activity is high-dimensional, as are measurements of dimension during spontaneous behavior [19, 20]. Such phenomena are not well modeled by low-rank networks that produce activity residing on a single low-dimensional manifold. We therefore developed a model of multi-task networks in which distinct dynamics occur on distinct neural manifolds.

In the context of our model, "task" refers to the low-dimensional dynamics associated with a particular manifold in firing-rate space, which we interpret as implementing a specific computation. Indeed, prior work in the single-task setting has shown that low-rank networks can implement experimentally inspired input–output mappings through their low-dimensional dynamics [28, 34]. Although real tasks involve both sensory inputs and behavioral outputs, we focus only on the recurrent dynamics, noting that our model could be easily extended to include task-specific inputs as in prior single-task works (see Discussion).

The network performs P different tasks, indexed by $\mu = 1, \ldots, P$. Each task is associated with a rank-R matrix of the form of Eq. (2), and the multi-task weight matrix is constructed as a weighted sum of these matrices:

$$J_{ij} = \sum_{\mu=1}^{P} D^{(\mu)} \sum_{r=1}^{R} m_i^{(\mu,r)} n_j^{(\mu,r)}. \tag{4}$$

Each task μ is therefore parameterized by R left and right loading vectors $\mathbf{m}^{(\mu,r)}$ and $\mathbf{n}^{(\mu,r)}$, and a weight $D^{(\mu)} \geq 0$ controlling the strength of the connectivity component associated with that task. The loading vectors are Gaussian and i.i.d. across neurons, as in the single-task setting. They are furthermore independent across tasks, i.e. there is zero expected overlap between loading vectors of distinct tasks. The covariance for task μ given by

$$\left\langle n_i^{(\mu,r)} m_i^{(\mu,r')} \right\rangle = A_{rr'}^{(\mu)}/N, \tag{5}$$

with the quantities $A_{rr'}^{(\mu)}$ comprising the task-specific $R \times R$ overlap matrix $\mathbf{A}^{(\mu)}$ (Fig. 1D). We refer to the subspace spanned by $\mathbf{m}^{(\mu,r)}$ for $r = 1, \ldots, R$ as *task subspace* μ, the R-dimensional space that contains the input to the network associated with task μ.

As described previously for single-task networks, the structure of overlaps between left and right loading vectors plays a major role in shaping network dynamics. The connectivity of Eq. (4) can be interpreted as that of a rank-PR network with a particular structure of overlaps among left and right loading vectors. Specifically, the overlaps between $\mathbf{m}^{(\mu,r)}$ and $\mathbf{n}^{(\nu,r')}$ are constrained to be, on average, block-diagonal, with the block entries given by $A_{rr'}^{(\mu)}$ (Fig. S1). Each of these blocks corresponds to R-dimensional dynamics associated with a single task subspace. Our model therefore describes synaptic weight matrices with strong, structured interactions among the components associated with a single task, specified by these blocks, and small, random overlap between components associated with different tasks. In terms of activity, the subspaces corresponding to different tasks are nearly, although not exactly,

orthogonal when N is large, with finite-size fluctuations in cross-task overlaps that are reduced as N grows. This blockwise constraint reduces the number of parameters needed to specify the model from P^2R^2 to PR^2 overlap-matrix elements—a dramatic reduction when P is large—while allowing the possibility of distinct tasks occupying different manifolds defined by the connectivity.

Multi-task networks exhibit winner-take-all competition across tasks

We begin by comparing single-task networks ($P = 1$) to a minimal multi-task network ($P = 2$). Network weights are set according to Eq. (4), fixing the within-task dimension $R = 2$. The $P = 1$ case reduces to the low-rank networks described by Eqs. (2) and (3). Such low-rank networks can produce a variety of different dynamics. For example, two different choices of a 2×2 overlap matrix $\mathbf{A}$ produce networks that exhibit either a stable limit cycle (Fig. 2A) or bistability (Fig. 2B). We next consider a multi-task network with $P = 2$ using a weight matrix that superposes the two matrices used above that, on their own, produce limit-cycle and fixed-point dynamics. The dynamics can be analyzed via two sets (one for each task) of latent variables $z_r^{(\mu)}(t) = D^{(\mu)}(\mathbf{n}^{(\mu,r)} \cdot \boldsymbol{\phi}(t))$ where $\mu = 1, 2$ indexes tasks and $r = 1, 2$ indexes within-task dimensions. Both sets of latent variables initially follow trajectories similar to those of the corresponding single-task networks (Fig. 2C). However, closer examination reveals that the dynamics are coupled across tasks and exhibit different asymptotic behaviors. While the dynamics in the task-1 (limit cycle) subspace are barely changed (Fig. 2A vs. Fig. 2C, left), the dynamics in the task-2 (bistable attractor) subspace lose their bistability. Trajectories are initially attracted towards one of the two previously stable fixed points but subsequently decay to the origin, once task-1 activity has grown sufficiently large in magnitude (Fig. 2B vs. Fig. 2C, right). The network therefore exhibits winner-take-all behavior, where one task dominates and suppresses others. In this example, task 1 dominates because $D^{(1)} > D^{(2)}$. Reversing this inequality causes task 2 to dominate (Fig. S2).

In a linear network, activity in distinct task subspaces evolves independently when the loading vectors for different tasks are orthogonal. Interference between tasks, defined as the influence of one task's latent variables on another's, occurs only when this orthogonality is broken. In multi-task networks with a fixed number of tasks, such interference becomes negligible as $N \to \infty$, because randomly oriented activity patterns in a high-dimensional activity space are nearly orthogonal with high probability. What, then, is the source of the winner-take-all interference observed in Fig. 2C? In nonlinear networks with finite P, Eq. (3) still describes each task's dynamics, but the gain factor now depends on the dynamics associated with all tasks [29]. Increased activity in any task subspace reduces the gain factor and suppresses the dynamics in other subspaces (see Methods). Consequently, at the steady state, the strongest task's latent variables dominate and evolve as they would in a single-task network, with all others suppressed. This form of interference persists even in the limit $N \to \infty$ and is unique to nonlinear networks that exhibit mixed selectivity and thus share neurons across tasks. In the example of Fig. 2C, the growth of task 1's oscillatory latent variables suppresses dynamics in the task-2 subspace and causes trajectories to approach the origin, eliminating bistability.

These properties are not specific to networks with our particular choice of nonlinearity, as we observe similar winner-take-all behavior for other saturating nonlinearities, as well as in networks with non-saturating nonlinearities that are stabilized through recurrent inhibition [35] (Fig. S3). However, the fact that at most one task can dominate, rather than multiple tasks dominating simultaneously, is a

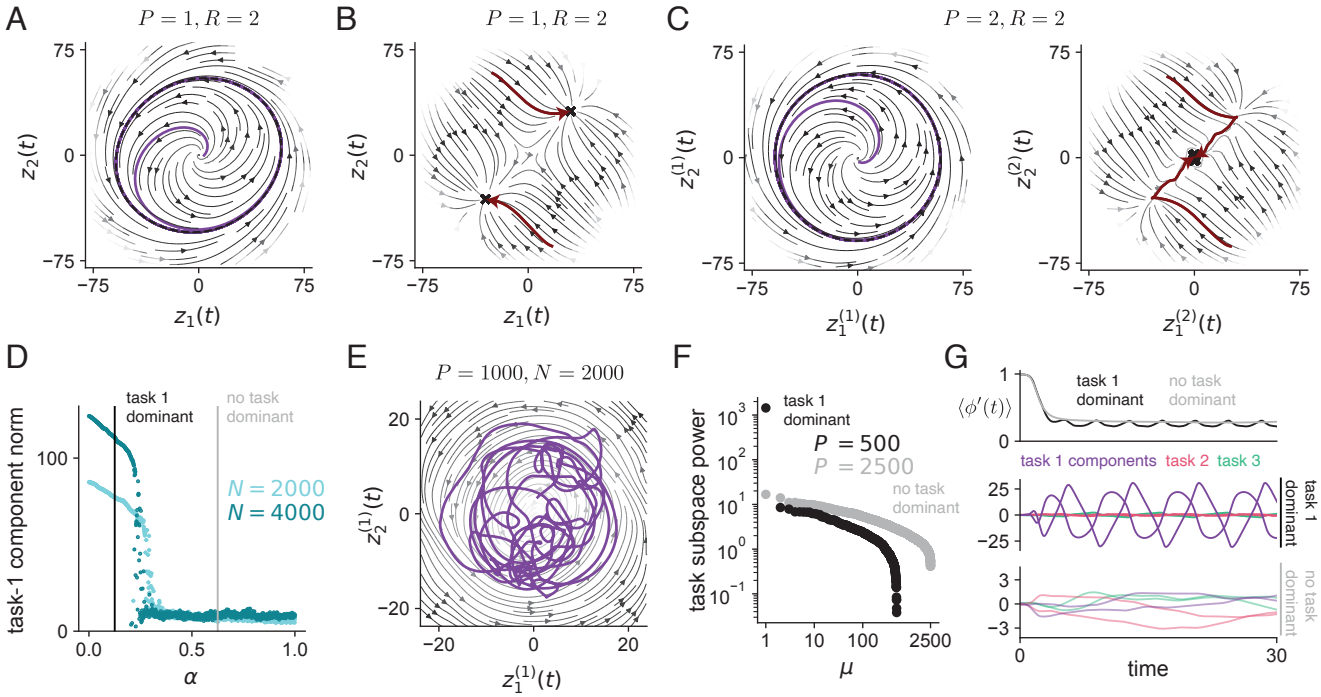

Figure 2: Dynamics of networks with varying P. (See Methods for parameter values.) **(A)** Latent-variable flow fields (black) and a latent-variable trajectory (purple) in a single-task network with $\mathbf{A}$ chosen to produce limit-cycle dynamics. Dashed line: limit cycle. **(B)** Latent-variable flow fields (black) and latent variable trajectories for two separate runs with different initial conditions (red) in a single-task network with $\mathbf{A}$ chosen to produce two stable fixed points. Black X's: final time points for all trajectories used to generate flow fields. **(C)** A two-task network with $D^{(1)} > D^{(2)}$. Left: flow fields (black) of projections $\vec{z}^{(1)}(t)$ onto task-1 subspace and a latent-variable trajectory (purple). Right: flow fields (black) of projections $\vec{z}^{(2)}(t)$ onto task-2 subspace simulated over many initial conditions, as well as two example trajectories (red). Black X's: final time points for all trajectories. **(D)** Time-averaged latent variable norm $\left\langle \|\vec{z}^{(1)}(t)\| \right\rangle_t$ for task $\mu = 1$ as a function of the number of total tasks $\alpha = P/N$ normalized by N. Same overlap parameters as (C) for $\mu = 1, 2$, and beyond $\mu > 2$ we randomly sampled $A^{(\mu)}_{rr'}$. **(E)** Same plot of task-1 dynamics as in left panel of (C) but for a network with $P = 1000$ and smaller axis scales. **(F)** Time-averaged power $\left\langle \|\vec{z}^{(\mu)}(t)/D^{(\mu)}\|^2 \right\rangle_t$ of the projections onto all task subspaces, normalized by $D^{(\mu)}$ and sorted in descending order, for a network with $P = 500$ tasks (black) and a network with $P = 2500$ tasks (gray) from panel (D). **(G)** Top: average gain as a function of time for two different networks, one with no task dominant (gray) and one with task 1 dominating (black). Middle: normalized latent-variable trajectories $z_r^{(\mu)}(t)/D^{(\mu)} = \boldsymbol{\phi}(t) \cdot \mathbf{n}^{(\mu,r)}$ for each task in the task-1-dominant network. Bottom: same as middle for the network with no task dominant; note the difference in y-axis scale.

consequence of Gaussian-distributed loadings, a simplification we use for mathematical tractability (see Discussion).

Chaotic fluctuations in networks with large P

Higher cortical areas participate in many distinct tasks, so our primary focus is on networks with large P. We therefore simulate networks with increasing P. We configure the overlap matrix of task 1 to produce a stable limit cycle as before, while the overlap matrices of other tasks are randomly generated. We also set $D^{(1)}$ to be large enough that task 1 would dominate any other task in isolation.

As P increases, the average magnitude of the projection of neural activity onto the task-1 subspace first decreases gradually, then drops sharply (Fig. 2D). This drop corresponds to a transition from a state in which task 1 is dominant and exhibits a stable limit cycle to a state in which no task is dominant. The point at which this transition occurs corresponds to a fixed ratio of number of tasks to number of neurons, denoted by $\alpha = P/N$ (Fig. 2D, light vs. dark cyan).

In an example network with large P and no task dominant, the task-1 latent variables fluctuate near the origin and are smaller in magnitude than those of networks with smaller P (Fig. 2E, purple trajectory). However, signatures of the task's oscillatory structure remain visible in an average flow field computed over many samples of network activity (Fig. 2E, black). As we will show later, the irregular trajectories of latent variables in networks with large P are driven by chaotic fluctuations that arise from many small overlaps among distinct tasks' subspaces, an effect that is vanishing for small P but significant for large P. In this state, activity is distributed across task subspaces (Fig. 2F, gray), rather than residing primarily in a dominant-task subspace (Fig. 2F, black). Additionally, the gain factor equilibrates to a time-independent value (Fig. 2G top, gray), rather than being dynamically entrained to any one task's latent variables (Fig. 2G top, black).

The simulations shown in Fig. 2 demonstrate that nonlinear multi-task networks exhibit interference. When P is small, interference leads to winner-take-all dynamics in which one fixed task always dominates, regardless of the initial condition. As P is increased, interference additionally leads to chaotic fluctuations that eventually suppress task-related dynamics. Unlike previous models in which chaos arises due to unstructured, random connectivity [36], or weight matrices with both unstructured and low-rank components [28], in our model all components of the synaptic weight matrix are task-related (Eq. 4). This suggests that chaotic fluctuations, which are often proposed as an explanation for irregular neural activity recorded *in vivo*, can arise solely as a consequence of learning many tasks (see Discussion).

Modulation of effective connectivity enables task selection

The results so far show that multi-task networks with fixed weights of the form of Eq. 4 cannot flexibly transition between distinct task-related activity patterns. Instead, at most one task dominates the winner-take-all dynamics (Fig. 2). We investigated mechanisms to circumvent this limitation. One possibility is to provide an input aligned with the task subspace of the desired task. Surprisingly, for networks with Gaussian loading vectors, this strategy does not enable task selection (Fig. S4) because such inputs modulate the gain factor by an equal amount across all tasks [29].

We therefore hypothesize that a mechanism that modulates $D^{(\mu)}$, the strength of the component of

the weight matrix associated with task μ, could enable selective activation of a desired task's dynamics. Such a modulation can be interpreted as increasing the effective recurrent strength of one task-specific latent dynamical system. Targeted gain control in which inputs lead to changes in effective connectivity through a nonlinearity, neuromodulatory signals, or short-term synaptic plasticity are potential biological substrates (see Discussion).

We develop an analytical theory that characterizes the dynamical regimes of multi-task networks with large P and predicts when task selection due to modulation of $D^{(\mu)}$ is successful. The theory applies to networks for which P and N are large, but with a fixed ratio $\alpha = P/N$. It also assumes that R, the dimension of the manifold associated with a single task, is fixed. This scaling is motivated by the hypothesis that neural circuits perform many distinct tasks, each of which is associated with a low-dimensional manifold (Fig. 3A).

The theory reveals that modulation of $D^{(\mu)}$ indeed supports selection of task-related dynamics in multi-task networks defined as Eq. 4. These connectivity modulations can place the dynamics of such networks into one of four distinct regimes. In the first, weights are small in magnitude and no activity is produced, so we do not consider this case further. In the second, the *spontaneous* state, chaotic fluctuations occur and no task is dominant (as in Fig. 2E). The other two states are the chaotic and nonchaotic *task-selected* states. The network transitions from the spontaneous state to a state in which task μ is selected if $D^{(\mu)}$ is increased sufficiently (Fig. 3B). Of course, task selection could be trivially achieved by setting $D^{(\mu)}$ to zero for all but one task, or by increasing it for a single task so dramatically that it dominates the entire weight matrix. However, our model requires only small increases in $D^{(\mu)}$ to produce large changes in task-related dynamics (see Discussion).

The simulations thus far have depicted the transition from a task-selected state to the spontaneous state as α is varied (Fig. 2D). Henceforth, we consider transitions from the spontaneous state to task-selected states due to modulation of $D^{(\mu)}$ for fixed α. Such a transition can be induced for any task by modulating its associated $D^{(\mu)}$ value (assuming the task does not involve purely decaying dynamics; see Methods). Thus, there are many distinct task-selected states, for which activity is large in the selected task's subspace and small in the others (Fig. 3C,D). This contrasts with networks without modulation of $D^{(\mu)}$, in which winner-take-all dynamics do not permit the selection of different tasks (Fig. 2).

Theory of the spontaneous state

The spontaneous state occurs when two conditions on the connectivity are met. First, for the network not to be quiescent, the average synaptic weight magnitude, $\langle J_{ij}^2 \rangle$, must be sufficiently large (see Methods). This is similar to the condition for spontaneous activity to occur in networks with independent and identically distributed weights [36] and requires that $\alpha > 0$, i.e. the number of tasks indeed scales with the network size. The second condition is that there must be no task μ whose strength $D^{(\mu)}$ is so large that it dominates the other tasks, leading to a task-selected state.

Here we describe latent-variable dynamics, and in later sections, we describe single-neuron activity (the theory describes both; see Methods). In the spontaneous state, the latent variables $z_r^{(\mu)}(t)$ $(r = 1, \ldots, R)$

for each task ($\mu = 1, \ldots, P$) follow

$$z_r^{(\mu)}(t) = \eta_r^{(\mu)}(t) + D^{(\mu)} \langle \phi' \rangle \sum_{r'=1}^{R} A_{rr'}^{(\mu)} \left[F * z_{r'}^{(\mu)} \right](t), \tag{6}$$

where $\eta_r^{(\mu)}(t)$ is a temporally correlated Gaussian noise process and $F(\tau)$ is a temporal filter. Eq. (6) describes the dynamics in previous simulations when no task is dominant (e.g. Fig. 2E; or Fig. 2G, bottom).

Eq. (6) can be compared to Eq. (3), which describes a single-task network (Fig. 2A). The first difference is the presence of an additional term $\eta_r^{(\mu)}(t)$, which represents interference due to effective noise from the many small overlaps with other tasks' dynamics. The second is that the gain factor $\langle \phi' \rangle$ is time-independent. This reflects the fact that, in the spontaneous state, the gain is not entrained to any single task's latent variables. Instead, it is determined by an average over many vanishingly correlated latent variables of different tasks, producing a time-independent quantity (Fig. 2G, black vs. gray). Finally, the linear filter $F(\tau)$ is in general different from that of Eq. (3); it depends on the statistics of the ensemble of $D^{(\mu)}$ and $\mathbf{A}^{(\mu)}$ (see Methods).

Therefore, in the spontaneous state, the dynamics of $z_r^{(\mu)}(t)$, for any task μ, are noise-driven and linear. As such, nonlinear dynamics such as asymptotically stable limit cycles or multistability cannot occur. Nevertheless, the spontaneous-state dynamics within a single task subspace are related to those that would be produced if the task were dominant (Fig. 2A,B). For example, the latent variables associated with limit-cycle dynamics exhibit an oscillatory component (Fig. S5A). Similarly, the latent variables associated with bistability exhibit a slow activity mode along the same axis that contains the fixed points in a single-task network (Fig. S5A). Thus, in the spontaneous state, latent variables for all tasks simultaneously follow noisy, linearized, and weakened versions of their task-selected dynamics.

Theory of the task-selected state

When a specific task μ's strength $D^{(\mu)}$ is increased sufficiently, the linear dynamics of the spontaneous state (Eq. 6) become unstable, leading to rapid growth of the task's latent variables. When this occurs, the gain factor $\langle \phi'(t) \rangle$ is no longer time-independent and instead is dynamically entrained to task μ's latent variables. This change amounts to the selection of the nonlinear dynamics associated with these variables and the suppression of unselected task dynamics (Fig. 3A). Just above this value of $D^{(\mu)}$, task-selected dynamics coexist with chaotic fluctuations. Increasing $D^{(\mu)}$ further leads to a second transition, beyond which the dominant task completely suppresses chaotic fluctuations, creating a nonchaotic task-selected state. Previous studies have identified related transitions in other network models [28, 37–40].

How does the activity of single neurons and of task-related latent variables change across these transitions? Increasing $D^{(1)}$ or $D^{(2)}$ (Fig. 3B) beyond the first transition threshold—to place the network in the corresponding chaotic task-selected state—leads to the associated latent variables growing in magnitude (from $\mathcal{O}(1)$ in the spontaneous state to $\mathcal{O}(\sqrt{N})$), while latent variables for any non-selected task remain $\mathcal{O}(1)$ (Fig. 3C,D). In the chaotic task-selected state just above the transition, the activity of each neuron exhibits a combination of tuning to the selected task's latent variables and chaotic fluctuations, with the two effects contributing to single-neuron activity on the same order (Fig. 3E, darker gray). In contrast,

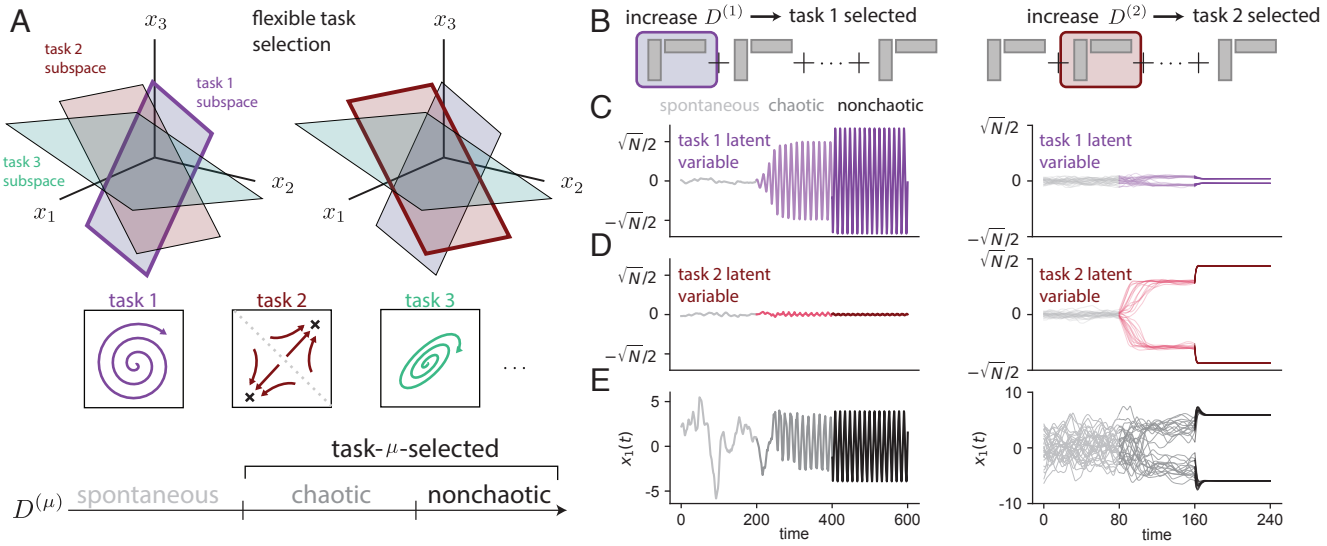

Figure 3: Task-selected states. **(A)** Top: Illustration of task selection as promoting activity in the chosen task's subspace, which contains task-specific dynamics. Bottom: Phase diagram as a function of the task strength $D^{(\mu)}$, comprising the spontaneous, chaotic task-selected, and nonchaotic task-selected states. **(B)** Schematic of task selection at the level of connectivity: modulating the strength of the chosen task's connectivity component through the associated $D^{(\mu)}$ value. **(C)** Trajectory of a task-1 latent variable (normalized by $D^{(\mu)}$), $z_1^{(1)}(t)/D^{(1)}$, when task-1 limit-cycle dynamics are selected (left) vs. when task-2 bistable dynamics are selected (right), in all 3 regimes. To illustrate bistability, multiple trajectories with different initial conditions are plotted on the right. **(D)** Same as (C) but for a task-2 latent variable $z_1^{(2)}(t)/D^{(2)}$. **(E)** Same as (C) and (D) but for an example neuron's activity.

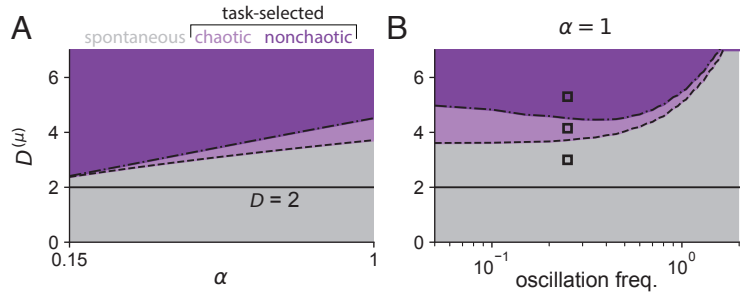

Figure 4: Phase diagrams for networks generated as a sum of single-task components associated with oscillatory dynamics (see Methods). **(A)** Network state as a function of normalized number of tasks $\alpha = P/N$ and the strength of one task $D^{(\mu)}$. All other tasks have $D^{(\nu\neq\mu)} = 2$ (solid line). Dashed line: Boundary between spontaneous and chaotic task-selected state for task μ. Dot-dashed line: Boundary between chaotic and nonchaotic task-selected state for task μ. **(B)** Same as (A) but for fixed $\alpha = 1$ and instead varying the frequency of the limit cycle task that is selected. Black boxes: Parameter values whose associated single-neuron properties are examined in Fig. 5.

the selected task's latent variables are noiseless to leading order, as single-neuron fluctuations average out in the projection onto the task subspace. Thus, the chaotic task-selected state is characterized by variability at the single-neuron level but not at the level of task-relevant variables. Increasing $D^{(1)}$ or $D^{(2)}$ further to enter the nonchaotic task-selected state suppresses this variability at the single-neuron level as well (Fig. 3E, black).

The value of $D^{(\mu)}$ required for task selection depends on task μ's structure, captured by its overlap matrix $\mathbf{A}^{(\mu)}$, as well as on the number and structure of other tasks (see Methods). Fig. 4 illustrates these dependencies for an analytically tractable class of networks in which each task corresponds to the generation of a stable oscillation with a different frequency (see Methods). A larger value of $D^{(\mu)}$ is required for task selection when the number of tasks is increased (horizontal slices in Fig. 4A; compare with Fig. 2D). The dependence of the critical value on the overlap matrix of the selected task reduces, in this case, to a dependence on the selected task's oscillation frequency (Fig. 4B).

Single-neuron statistics in spontaneous and task-selected states

Thus far, we have focused on describing latent variables defined by projections of network activity onto task-specific subspaces. The theory also describes activities of individual neurons (see Methods). In the spontaneous state, the activity of a given neuron i is described by

$$\dot{x}_i(t) = -x_i(t) + \eta_i^x(t) + [G * \phi_i](t), \tag{7}$$

where $\eta_i^x(t)$ is a temporally correlated Gaussian noise process and $G(\tau)$ is a filter that produces a nonlinear self-coupling of each neuron to its firing rate. Each neuron's activity fluctuates with negligible tuning to any single task's latent variables (Fig. S5A). The statistics of these fluctuations are homogeneous across neurons. The theory allows analytic calculation of the correlation functions of $\eta_r^{(\mu)}(t)$ and $\eta^x(t)$ and the kernels $F(\tau)$ and $G(\tau)$. This leads to predictions for the single-task and single-neuron correlation functions, $C_{rr'}^{(\mu)}(\tau)$ and $C^{\phi}(\tau)$, respectively, which we verify with simulations (Fig. S5).

In task-selected states, the equation describing single-neuron activity is similar to Eq. (7), with an

additional term corresponding to the neuron's tuning to the dominant task's latent variables:

$$\dot{x}_i(t) = -x_i(t) + \eta_i^x(t) + [G * \phi_i](t) + \sum_{r=1}^{R} m_i^{(\mu^*,r)} z_r^{(\mu^*)}(t), \tag{8}$$

where μ^* is the index of the dominant task. The $\eta_i^x(t)$ term is still a Gaussian noise process, but its kernel is not time-translation-invariant and instead depends on the dynamics of the dominant task (see Methods).

In task-selected states, neurons are statistically heterogeneous, following different trajectories depending on their tuning to the dominant task's latent variables (Fig. 5A). This heterogeneity implies that the extent to which single-neuron activity can be explained by one task's latent variables varies across neurons in the population. The mean of this distribution is low in the spontaneous state, higher in the chaotic task-selected state, and highest in the nonchaotic task-selected state (Fig. 5B).

The average autocovariance function $C^x(\tau) = \langle x_i(t)x_i(t+\tau)\rangle_{t,i}$ illustrates the differences between the temporal statistics of single-neuron firing rates in the spontaneous and task-selected states (Fig. 5C). The first transition occurs when task-related structure, in this case oscillations, becomes visible at the single-neuron level (Fig. 5C, left vs. middle). The second transition occurs when all firing-rate variance is accounted for by dominant-task activity, or equivalently, in the case of a limit-cycle task, when $C^x(0) = C^x(T)$ for an oscillation with period T (Fig. 5C, middle vs. right).

Dimensionality of spontaneous and task-selected states

To analyze the structure of population activity in multi-task networks, we also calculate its dimension. We quantify dimension using the participation ratio, a measure of linear dimensionality that approximates the number of orthogonal directions in activity space that are explored by the network's dynamics. This quantity has been the subject of numerous experimental and theoretical studies [37, 41–45].

In the spontaneous state, the dimension scales with the network size N, as previously observed in networks with random connectivity (43; Fig. 5D, left). Although this scaling produces dimension values that are large, they are always much lower than the ambient dimension N due to weak correlations among neurons [45], a common observation in neural data [22, 46, 47]. At the other extreme, in the nonchaotic task-selected state, the dimension is determined solely by the dominant task's dynamics and is therefore proportional to R (Fig. 5D, right). Between these extremes, in the chaotic task-selected state, the dimension is larger than this limiting value by a factor that does not grow with N and can be predicted by the theory (Fig. 5D, middle; see Methods). This increase arises from chaotic fluctuations in directions orthogonal to the dominant task's subspace. The magnitudes of these fluctuations depend on the values of $D^{(\mu)}$ for both the selected and unselected tasks (Fig. S6).

These results suggest that experimental measurements of neural dimensionality should be interpreted in the context of the underlying network state. In particular, the observed dimension may depend on which task-selected states are present, the statistics of transitions between these states, and the dimension of the manifolds associated with each individual task. Consistent with previous studies of neural activity during specific behavioral tasks [21, 48, 49], our model predicts that the dimension of activity is low when activity is measured within a single task-selected state, even when fluctuations are present at the

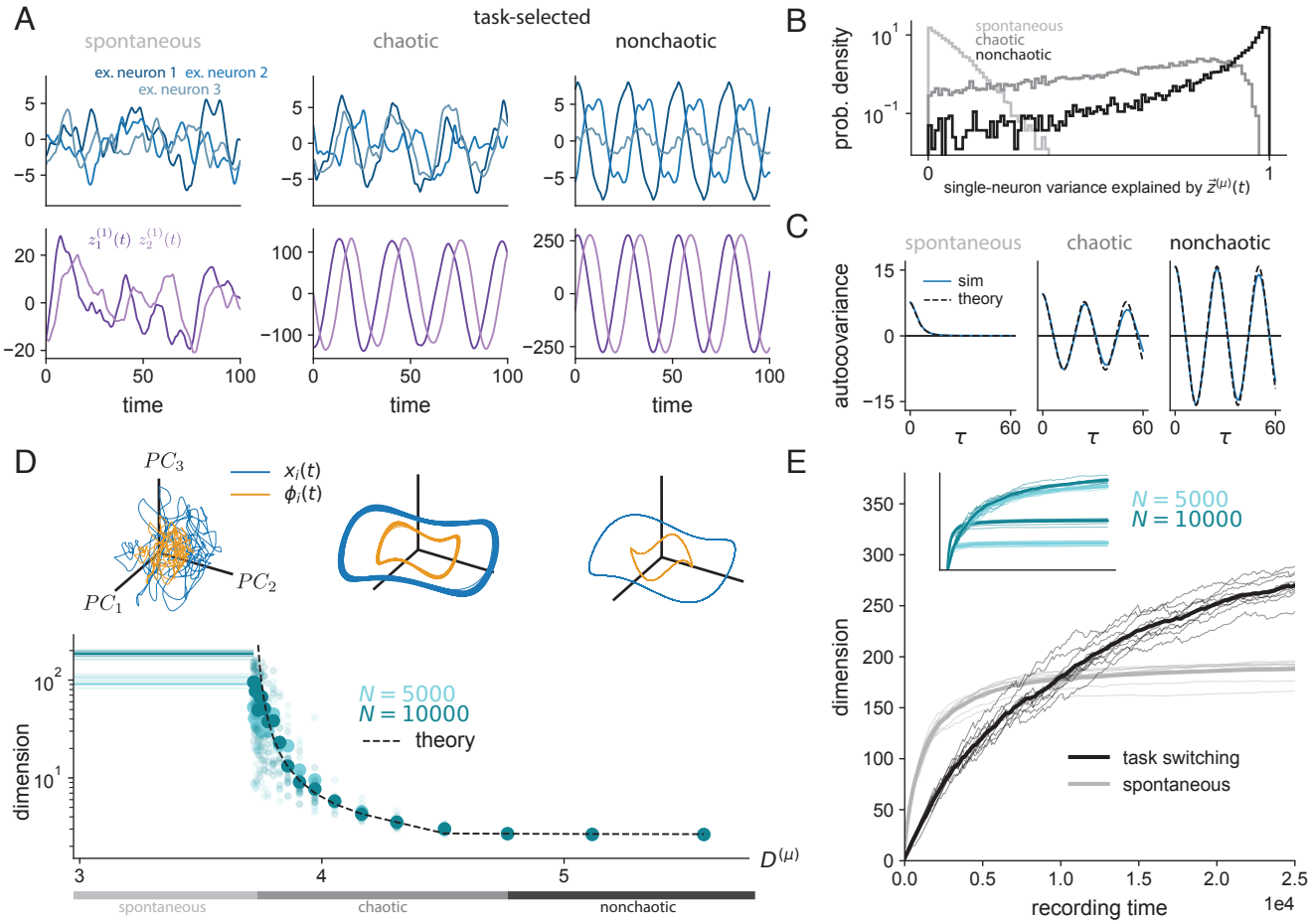

Figure 5: Single-neuron and population properties of a multi-task networks. Parameters are the same as Fig. 4 (see Methods). **(A)** Example neurons (top) and latent variables for the selected task (bottom) across network states. **(B)** Histogram R^2 values for linear regression of $x_i(t)$ against task-μ latent-variable activity. **(C)** Average single-neuron autocovariance functions $C^x(\tau) = \langle x_i(t)x_i(t+\tau)\rangle_{t,i}$. **(D)** Top: Principal component trajectories across same network states from (A), (B), and (C). For currents $x_i(t)$ and firing rates $\phi_i(t)$, principal component axes and projections onto these axes are computed separately, with $\phi_i(t)$ projections doubled for visibility. Axis scales: 150 (left), 300 (middle), 500 (right). Bottom: Linear embedding dimension computed as participation ratio of the firing rates $\phi_i(t)$, as a function of $D^{(\mu)}$. Light dots: individual network simulations. Heavy dots: averages over networks. **(E)** Firing-rate dimension as a function of recording time for a network in the spontaneous state and switching across task-selected states. For the latter, one of 150 tasks is randomly selected for 80 time constants, followed by 3 time constants of spontaneous activity, and this process is repeated. Thin lines: individual network simulations. Thick lines: average. Inset: The same plot but for networks with different N, showing dependence on N in the spontaneous state.

single-neuron level. An experimentally observed increase beyond this value can arise through two distinct mechanisms. First, the network may transition into the spontaneous state, whose dimension scales with network size, consistent with analyses of large-scale neural recordings [20]. Second, the dimension of activity may increase through the sequential engagement of multiple task-selected states, each associated with a different low-dimensional manifold.

These mechanisms make distinct predictions for how the dimensionality of activity grows with recording time (Fig. 5E). In the spontaneous state, dimensionality increases rapidly at short times and then saturates (Fig. 5E, gray). When the network sequentially switches across task-selected states, the rate of dimensionality growth depends on the amount of time spent in each task state. Even in cases in which this time is short (as in our simulations), dimensionality grows more slowly than in the spontaneous state (Fig. 5E, black).

The two mechanisms also differ in their saturation values. In the spontaneous state, network activity explores a chaotic attractor in which projections onto certain directions in neural activity space contain more variance than others, limiting dimensionality. In contrast, sequential task selection produces a sequence of trajectories that are highly decorrelated from one another. As a result, the saturation value for sequential task selection is approximately equal to the number of visited task states multiplied by the within-task dimension. This value can be either higher or lower than the spontaneous-state dimension. For a sufficiently large number of tasks, the dimensionality associated with sequential task selection can greatly exceed that of the spontaneous state (Fig. 5E, inset). Altogether, the theory demonstrates that either a fixed high-dimensional covariance matrix or state-dependent switching among different low-dimensional covariance matrices are both potential models for high-dimensional activity in neural recordings.

Discussion

We identified limitations to multi-task capabilities in nonlinear low-rank networks due to interference among tasks. Such limitations are not evident in analyses of networks constructed to perform a single task or in linear networks that lack the winner-take-all interference we described. We proposed a solution to the problem of interference that relies on modulation of effective connectivity. Networks with such a mechanism transition between a spontaneous state in which no task is selected and task-selected states in which activity lies primarily on a task-specific low-dimensional manifold. Our theory thus retains the analytic tractability of low-rank networks [28, 29, 34] while describing the geometry of neural dynamics associated with multiple tasks.

Neuronal architectures for multi-task networks

The connectivity of our model, and its modulation during task selection, may be interpreted in multiple ways. If the network is viewed as a single recurrently connected circuit, modulation of $D^{(\mu)}$ corresponds to a change in the strength of a single task-related component of the weight matrix (Fig. 6A). Such a change requires selective modulation of synapses depending on their participation in this weight matrix component. Potential biological substrates include rapid synaptic plasticity or targeted neuromodulation [51–56].

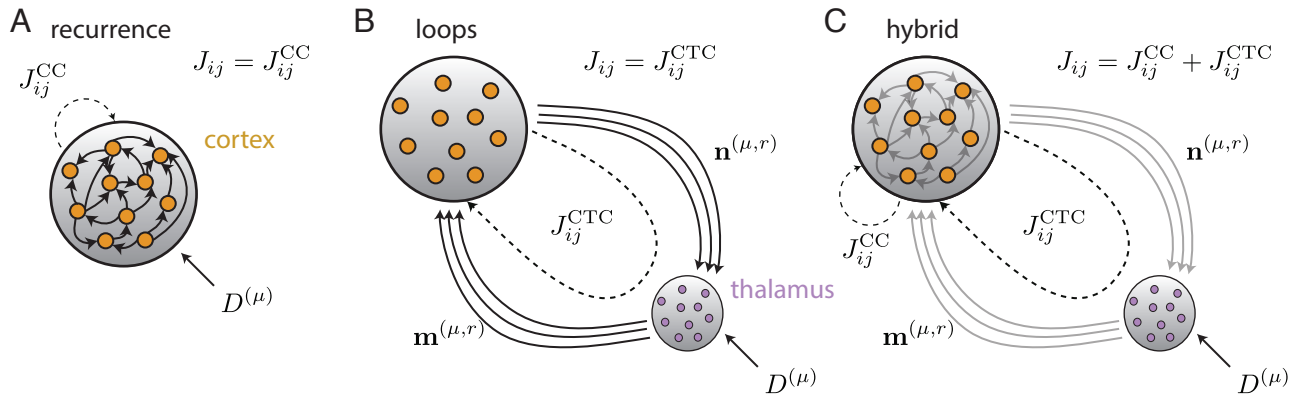

Figure 6: Illustrations of three interpretations of the network's effective connectivity. **(A)** Modulation of $D^{(\mu)}$ modifies the recurrent connections J_{ij}^{CC} of a cortical network. **(B)** Interactions between neurons are driven by effective connectivity J_{ij}^{CTC} defined by thalamocortical loops. Modulation of $D^{(\mu)}$ adjusts the gain of thalamic neurons [50]. **(C)** Model in which the effective connectivity is a combination of (A) and (B).

Alternatively, the loadings $\mathbf{m}^{(\mu,r)}$ and $\mathbf{n}^{(\mu,r)}$ may be interpreted as synaptic weights between a network and an auxiliary population of neurons, for instance thalamocortical and corticothalamic weights, respectively (Fig. 6B). In this case, task selection is accomplished through adjustment of the gain of a subset of task-associated thalamic neurons. A linear network model of such a mechanism has been proposed [50]. More generally, because only small modulations of effective connectivity are required in our nonlinear model, its weights may be interpreted as a combination of fixed corticocortical and adjustable thalamocortical synapses (Fig. 6C).

We focused on how a single network can generate many distinct sets of low-dimensional dynamics, without explicitly modeling task-related sensory inputs or how these dynamics are transformed to produce behavior. In low-rank single-task networks, task-related inputs define additional dimensions of the task subspace [28, 34]. Similarly, we expect our qualitative results on interference and winner-take-all dynamics to generalize to cases where distinct tasks operate with distinct input and output channels. Future work could examine the consequences of sharing neural modes across tasks, whether input channels, latent dynamical modes, or output channels [2, 3, 5, 7].

Related models of multi-tasking

Previous studies have examined multi-tasking in recurrent neural networks that are trained end-to-end to perform individual tasks that depend on a contextual input [3, 27], or sets of tasks commonly used in neuroscience experiments [4, 14]. These models produce diverse dynamical behaviors but their weight matrices have complex statistical structure from the training process that limits analytical tractability. It would be interesting to examine whether the tonic inputs commonly used to signal task context in trained multi-task networks effectively drive similar task-selection mechanisms via gain modulation, perhaps in task-specific populations [14, 29, 34], by shifting the operating point of the nonlinearity.

Other works have studied nonlinear models with closed-form expressions for the weight matrix, in-

cluding summations over a large number of low-rank components, as in our model. In certain cases, such connectivity leads to chaotic activity [38, 45, 55, 57, 58]. These studies considered networks that are designed to produce specific dynamical features, including fixed points [38], continuous attractors [58], and sequential activity [55, 57]. Our model, on the other hand, permits each task to be associated with any low-dimensional dynamical system described by Eq. (3), that is, any dynamical system that can be described within the framework of single-population low-rank networks [29]. Thus, although we limited our analysis to limit cycles and fixed points, our model also accommodates sequential activity and non-normal latent dynamics [59].

Modulating connectivity has been proposed in previous works. A previous study developed a linear network model of thalamocortical interactions (Fig. 6B) that sequences transient motor motifs [50]. Other models have used statistical inference to learn time-varying scaling factors that adjust low-rank connectivity components in nonlinear networks [6, 56]. Our theory provides an analytical account of such modulation and its consequences for task selection and spontaneous activity.

In biological and artificial systems, complex behaviors are often built by combining simpler behavioral components [14, 60]. In our model, each "task" should be interpreted as a minimal computational component associated with a behavior, which may only be part of what is often referred to as a task in behavioral experiments. For example, fixation, stimulus, delay, and response periods may each demand different network dynamics [6]. To retain analytic tractability, here we assumed no overlap or reuse of dynamics across tasks, but including such structure is an interesting direction for future work.

Related theoretical work suggests that selecting among low-rank connectivity components, each suited to a particular subtask, is a more effective strategy for continual learning than treating each compound behavior as its own learning problem [6]. Animal experiments indicate that complex motor control arises from sequencing different low-dimensional dynamics, each associated with a particular manifold of state space, to generate flexible behavioral sequences [2]. Our theory provides an analytically tractable mechanism for such sequential composition. In contrast, the winner-take-all dynamics of our model does not support multiple simultaneously active tasks, due to the Gaussian-distributed loadings. Such dynamics require additional circuit structure, such as multiple neuronal populations (e.g. from mixture-of-Gaussians loadings) with population-specific gain factors [29, 34]. Nonetheless, the same gain-dependent mechanisms that enable task selection and suppression in the Gaussian case may also facilitate simultaneous selection of multiple tasks—and suppression of all unselected tasks—in more complex models, which we leave to future work. Such models would additionally allow for more complex single-task dynamics beyond what can be achieved with one population [34].

Magnitude of effective connectivity modulation

Our theory shows that, in nonlinear networks, a modest modulation of effective connectivity is often sufficient for task selection. Adjustment of all values of $D^{(\mu)}$, such that each value is either zero for unselected tasks, or nonzero for the selected task, would achieve task selection trivially by completely overwriting the synaptic weight matrix each time a task is selected. In contrast, our model requires adjusting the $D^{(\mu)}$ value for only the selected task, which corresponds to changes in synaptic weights that are $\mathcal{O}(1/P)$, whereas each synaptic weight is $\mathcal{O}(1/\sqrt{P})$. Thus a small task-related connectivity component can dramatically influence a network's activity, an observation that has been made in models with both

structured low-rank and unstructured weight matrix components [28, 33, 39, 61]. Our model demonstrates that this feature holds in networks for which all connectivity components are task-related.

Moreover, the degree of modulation, even relative to $1/P$, can be made small in our model. The change in $D^{(\mu)}$ required for task selection depends on its distance from the critical value at which task selection occurs (Fig. 4B, black solid line vs. dashed). When α is small, a plausible limit in which the number of tasks is large but much smaller than the number of neurons, these distances can be decreased (Fig. 4A). Rapid reconfiguration of dynamics thus occurs with only small connectivity adjustments, as each task is close to selection. In such networks, dynamics are near the edge of stability in the spontaneous state [62], providing a functional explanation for the presence of irregular spontaneous activity.

Dimensionality and the relationship between spontaneous and evoked activity

Spontaneous activity has historically been defined as activity in the absence of sensory stimulation [63]. Classic theoretical work showed how such activity could be self-generated by a nonlinear network without external inputs [36]. However, recent high-yield recordings during periods of unstimulated, uninstructed behavior have revealed high-dimensional activity whose high-variance components are tuned to kinematic variables [19, 20], which is not well described by unstructured chaotic activity. The spontaneous state of our model has similarities to the chaotic regime of unstructured networks [36] but describes both the irregular activity of neurons and also of task-relevant latent variables. The spontaneous state of our model may explain the idiosyncratic behaviors that have been observed to accompany periods of "quiet wakefulness" [15], while the chaotic task-selected state may explain how such behaviors coexist with coherent goal-directed actions [64].

When freely exploring, rodents perform a variety of self-initiated behaviors, such as whisking or grooming [1]. Many of these behavioral "syllables" feature precise kinematic control that is likely not explained by the spontaneous state of our model. Rather, recordings during long periods of free behavior may correspond to switching among different task-selected states of our model. This predicts that individual neurons should have heterogeneous tuning to kinematic variables for an active behavioral syllable, and this tuning will change across syllables (Fig. 5B). A second prediction is that, when pooling neural activity across distinct syllables, the observed dimension should increase with the diversity of behavioral syllables observed. Conversely, dimensionality measured over many repetitions of a single behavioral syllable should be small, even if there is trial-to-trial variability at the single-neuron level (Fig. 5D). A final prediction is that the initial rate of dimension growth relative to recording time should be faster during quiet wakefulness than during behavior, and the magnitude of this effect depends on the average syllable duration (Fig. 5E). These predictions can be examined in neural recordings during free behavior [65].

Acknowledgments

We thank Manuel Beiran for discussions and for proposing the inhibition-stabilized model of Fig. S3. We thank Lea Duncker for feedback on the manuscript. OM, DGC, and ALK were supported by the Kavli Foundation and the Gatsby Charitable Foundation (grant no. GAT3708). ALK was also supported by the National Science Foundation (grant no. 2443158).

References

[1] Jeffrey E Markowitz, Winthrop F Gillis, Maya Jay, Jeffrey Wood, Ryley W Harris, Robert Cieszkowski, Rebecca Scott, David Brann, Dorothy Koveal, Tomasz Kula, et al. Spontaneous behaviour is structured by reinforcement without explicit reward. *Nature*, 614(7946):108–117, 2023.

[2] Elom A Amematsro, Eric M Trautmann, Najja J Marshall, LF Abbott, Michael N Shadlen, Daniel M Wolpert, and Mark M Churchland. Motor cortex flexibly deploys a high-dimensional repertoire of subskills. *bioRxiv*, pages 2025–09, 2025.

[3] Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, 503(7474):78–84, 2013.

[4] Guangyu Robert Yang, Madhura R Joglekar, H Francis Song, William T Newsome, and Xiao-Jing Wang. Task representations in neural networks trained to perform many cognitive tasks. *Nature neuroscience*, 22(2):297–306, 2019.

[5] Marino Pagan, Vincent D Tang, Mikio C Aoi, Jonathan W Pillow, Valerio Mante, David Sussillo, and Carlos D Brody. Individual variability of neural computations underlying flexible decisions. *Nature*, 639(8054):421–429, 2025.

[6] Haozhe Shan, Sun Minni, and Lea Duncker. Separating the what and how of compositional computation to enable reuse and continual learning. *arXiv preprint arXiv:2510.20709*, 2025.

[7] Sina Tafazoli, Flora M Bouchacourt, Adel Ardalan, Nikola T Markov, Motoaki Uchimura, Marcelo G Mattar, Nathaniel D Daw, and Timothy J Buschman. Building compositional tasks with shared neural subspaces. *Nature*, pages 1–9, 2025.

[8] Saurabh Vyas, Matthew D Golub, David Sussillo, and Krishna V Shenoy. Computation through neural population dynamics. *Annual review of neuroscience*, 43(1):249–275, 2020.

[9] Patrick T Sadtler, Kristin M Quick, Matthew D Golub, Steven M Chase, Stephen I Ryu, Elizabeth C Tyler-Kabara, Byron M Yu, and Aaron P Batista. Neural constraints on learning. *Nature*, 512(7515): 423–426, 2014.

[10] Matthew D Golub, Patrick T Sadtler, Emily R Oby, Kristin M Quick, Stephen I Ryu, Elizabeth C Tyler-Kabara, Aaron P Batista, Steven M Chase, and Byron M Yu. Learning by neural reassociation. *Nature neuroscience*, 21(4):607–616, 2018.

[11] Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D Daw, Earl K Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451): 585–590, 2013.

[12] Mark G Stokes, Makoto Kusunoki, Natasha Sigala, Hamed Nili, David Gaffan, and John Duncan. Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, 78(2):364–375, 2013.

[13] Julie J Lee, Michael Krumin, Kenneth D Harris, and Matteo Carandini. Task specificity in mouse parietal cortex. *Neuron*, 110(18):2961–2969, 2022.

[14] Laura N Driscoll, Krishna Shenoy, and David Sussillo. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nature Neuroscience*, 27(7):1349–1363, 2024.

[15] Frederic Lanore, N Alex Cayco-Gajic, Harsha Gurnani, Diccon Coyle, and R Angus Silver. Cerebellar granule cell axons support high-dimensional representations. *Nature neuroscience*, 24(8):1142–1150, 2021.

[16] Nikhilesh Natraj, Daniel B Silversmith, Edward F Chang, and Karunesh Ganguly. Compartmentalized dynamics within a common multi-area mesoscale manifold represent a repertoire of human hand movements. *Neuron*, 110(1):154–174, 2022.

[17] David A Sabatini and Matthew T Kaufman. Reach-dependent reorientation of rotational dynamics in motor cortex. *Nature Communications*, 15(1):7007, 2024.

[18] Simon Borgognon, Nicolò Macellari, Alexandra M Hickey, Matthew G Perich, Houman Javaheri, Rafael Ornelas-Kobayashi, Maude Delacombaz, Christopher Hitz, Florian Fallegger, Stéphanie P Lacour, et al. Regional specialization of movement encoding across the primate sensorimotor cortex. *Nature Communications*, 16(1):5729, 2025.

[19] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364 (6437):eaav7893, 2019.

[20] Jason Manley, Sihao Lu, Kevin Barber, Jeffrey Demas, Hyewon Kim, David Meyer, Francisca Martínez Traub, and Alipasha Vaziri. Simultaneous, cortex-wide dynamics of up to 1 million neurons reveal unbounded scaling of dimensionality with neuron number. *Neuron*, 112(10):1694–1709, 2024.

[21] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. Neural population dynamics during reaching. *Nature*, 487 (7405):51–56, 2012.

[22] Peiran Gao, Eric Trautmann, Byron Yu, Gopal Santhanam, Stephen Ryu, Krishna Shenoy, and Surya Ganguli. A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv*, page 214262, 2017.

[23] Atika Syeda, Lin Zhong, Renee Tung, Will Long, Marius Pachitariu, and Carsen Stringer. Facemap: a framework for modeling neural activity based on orofacial tracking. *Nature neuroscience*, 27(1): 187–195, 2024.

[24] Carl Van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.

[25] Alfonso Renart, Jaime De La Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex Reyes, and Kenneth D Harris. The asynchronous state in cortical circuits. *science*, 327(5965):587–590, 2010.

[26] Lea Duncker, Laura Driscoll, Krishna V Shenoy, Maneesh Sahani, and David Sussillo. Organizing recurrent network dynamics by task-computation to enable continual learning. *Advances in neural information processing systems*, 33:14387–14397, 2020.

[27] Christopher Langdon and Tatiana A Engel. Latent circuit inference from heterogeneous neural responses during cognitive tasks. *Nature neuroscience*, 28(3):665–675, 2025.

[28] Francesca Mastrogiuseppe and Srdjan Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron*, 99(3):609–623, 2018.

[29] Manuel Beiran, Alexis Dubreuil, Adrian Valente, Francesca Mastrogiuseppe, and Srdjan Ostojic. Shaping dynamics with multiple populations in low-rank recurrent networks. *Neural Computation*, 33(6):1572–1615, 2021.

[30] Brian Cheung, Alexander Terekhov, Yubei Chen, Pulkit Agrawal, and Bruno Olshausen. Superposition of many models into one. *Advances in neural information processing systems*, 32, 2019.

[31] Zengcai V Guo, Hidehiko K Inagaki, Kayvon Daie, Shaul Druckmann, Charles R Gerfen, and Karel Svoboda. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature*, 545(7653): 181–186, 2017.

[32] Britton A Sauerbrei, Jian-Zhong Guo, Jeremy D Cohen, Matteo Mischiati, Wendy Guo, Mayank Kabra, Nakul Verma, Brett Mensh, Kristin Branson, and Adam W Hantman. Cortical pattern generation during dexterous movement is input-driven. *Nature*, 577(7790):386–391, 2020.

[33] Friedrich Schuessler, Alexis Dubreuil, Francesca Mastrogiuseppe, Srdjan Ostojic, and Omri Barak. Dynamics of random recurrent networks with correlated low-rank structure. *Physical Review Research*, 2(1):013111, 2020.

[34] Alexis Dubreuil, Adrian Valente, Manuel Beiran, Francesca Mastrogiuseppe, and Srdjan Ostojic. The role of population structure in computations through neural dynamics. *Nature neuroscience*, 25(6): 783–794, 2022.

[35] Yashar Ahmadian, Daniel B Rubin, and Kenneth D Miller. Analysis of the stabilized supralinear network. *Neural computation*, 25(8):1994–2037, 2013.

[36] Haim Sompolinsky, Andrea Crisanti, and Hans-Jurgen Sommers. Chaos in random neural networks. *Physical review letters*, 61(3):259, 1988.

[37] Kanaka Rajan, LF Abbott, and Haim Sompolinsky. Stimulus-dependent suppression of chaos in recurrent neural networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 82 (1):011903, 2010.

[38] Ulises Pereira-Obilinovic, Johnatan Aljadeff, and Nicolas Brunel. Forgetting leads to chaos in attractor networks. *Physical Review X*, 13(1):011009, 2023.

[39] David G Clark and LF Abbott. Theory of coupled neuronal-synaptic dynamics. *Physical Review X*, 14(2):021001, 2024.

[40] David G Clark and Manuel Beiran. Structure of activity in multiregion recurrent neural networks. *Proceedings of the National Academy of Sciences*, 122(10):e2404039122, 2025.

[41] Ashok Litwin-Kumar, Kameron Decker Harris, Richard Axel, Haim Sompolinsky, and LF Abbott. Optimal degrees of synaptic connectivity. *Neuron*, 93(5):1153–1164, 2017.

[42] Evan S Schaffer, Neeli Mishra, Matthew R Whiteway, Wenze Li, Michelle B Vancura, Jason Freedman, Kripa B Patel, Venkatakaushik Voleti, Liam Paninski, Elizabeth MC Hillman, et al. The spatial and temporal structure of neural activity across the fly brain. *Nature communications*, 14(1):5572, 2023.

[43] David G Clark, LF Abbott, and Ashok Litwin-Kumar. Dimension of activity in random neural networks. *Physical Review Letters*, 131(11):118401, 2023.

[44] Aitor Morales-Gregorio, Anno C Kurth, Junji Ito, Alexander Kleinjohann, Frédéric V Barthélemy, Thomas Brochier, Sonja Grün, and Sacha J van Albada. Neural manifolds in v1 change with top-down signals from v4 targeting the foveal region. *Cell Reports*, 43(7), 2024.

[45] David G Clark, Owen Marschall, Alexander van Meegen, and Ashok Litwin-Kumar. Connectivity structure and dynamics of nonlinear recurrent neural networks. *Physical Review X*, 15(4):041019, 2025.

[46] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[47] Dagmara Panas, Hayder Amin, Alessandro Maccione, Oliver Muthmann, Mark van Rossum, Luca Berdondini, and Matthias H Hennig. Sloppiness in spontaneously active neuronal networks. *Journal of Neuroscience*, 35(22):8480–8492, 2015.

[48] Christopher J Cueva, Alex Saez, Encarni Marcos, Aldo Genovesio, Mehrdad Jazayeri, Ranulfo Romo, C Daniel Salzman, Michael N Shadlen, and Stefano Fusi. Low-dimensional dynamics for working memory and time encoding. *Proceedings of the National Academy of Sciences*, 117(37):23021–23032, 2020.

[49] Gengshuo John Tian, Ou Zhu, Vinay Shirhatti, Charles M Greenspon, John E Downey, David J Freedman, and Brent Doiron. Neuronal firing rate diversity lowers the dimension of population covariability. *bioRxiv*, 2024.

[50] Laureline Logiaco, LF Abbott, and Sean Escola. Thalamic control of cortical dynamics in a model of flexible motor sequencing. *Cell reports*, 35(9), 2021.

[51] Gianluigi Mongillo, Omri Barak, and Misha Tsodyks. Synaptic theory of working memory. *Science*, 319(5869):1543–1546, 2008.

[52] Eve Marder. Neuromodulation of neuronal circuits: back to the future. *Neuron*, 76(1):1–11, 2012.

[53] Neta Ravid Tannenbaum and Yoram Burak. Shaping neural circuits by high order synaptic interactions. *PLOS Computational Biology*, 12(8):e1005056, 2016.

[54] Jake P Stroud, Mason A Porter, Guillaume Hennequin, and Tim P Vogels. Motor primitives in space and time via targeted gain modulation in cortical networks. *Nature neuroscience*, 21(12):1774–1783, 2018.

[55] Maxwell Gillett, Ulises Pereira, and Nicolas Brunel. Characteristics of sequential activity in networks with temporally asymmetric hebbian learning. *Proceedings of the National Academy of Sciences*, 117 (47):29948–29958, 2020.

[56] Julia Costacurta, Shaunak Bhandarkar, David Zoltowski, and Scott Linderman. Structured flexibility in recurrent neural networks via neuromodulation. *Advances in Neural Information Processing Systems*, 37:1954–1972, 2024.

[57] Maxwell Gillett and Nicolas Brunel. Dynamic control of sequential retrieval speed in networks with heterogeneous learning rules. *Elife*, 12:RP88805, 2024.

[58] Francesco P Battaglia and Alessandro Treves. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Physical Review E*, 58(6):7738, 1998.

[59] Ulises Pereira-Obilinovic, Kayvon Daie, Susu Chen, Karel Svoboda, and Ran Darshan. Neural dynamics outside task-coding dimensions drive decision trajectories through transient amplification. *bioRxiv*, pages 2025–11, 2025.

[60] Vishwa Goudar, Barbara Peysakhovich, David J Freedman, Elizabeth A Buffalo, and Xiao-Jing Wang. Schema formation in a neural population subspace underlies learning-to-learn in flexible sensorimotor problem-solving. *Nature Neuroscience*, 26(5):879–890, 2023.

[61] Albert J Wakhloo, David G Clark, and LF Abbott. Associative synaptic plasticity creates dynamic persistent activity. *bioRxiv*, 2025.

[62] Robert Legenstein and Wolfgang Maass. Edge of chaos and prediction of computational performance for neural circuit models. *Neural networks*, 20(3):323–334, 2007.

[63] Lilach Avitan and Carsen Stringer. Not so spontaneous: Multi-dimensional representations of behaviors and context in sensory areas. *Neuron*, 110(19):3064–3075, 2022.

[64] Simon Musall, Matthew T Kaufman, Ashley L Juavinett, Steven Gluf, and Anne K Churchland. Single-trial neural dynamics are dominated by richly varied movements. *Nature neuroscience*, 22(10): 1677–1686, 2019.

[65] Sherry Lin, Winthrop F Gillis, Caleb Weinreb, Ayman Zeine, Samuel C Jones, Emma M Robinson, Jeffrey Markowitz, and Sandeep Robert Datta. Characterizing the structure of mouse behavior using motion sequencing. *Nature Protocols*, 19(11):3242–3291, 2024.

[66] Marcelo A Montemurro, Francisco A Tamarit, Daniel A Stariolo, and Sergio A Cannas. Out-of-equilibrium dynamics of the hopfield model in its spin-glass phase. *Physical Review E*, 62(4):5721, 2000.

[67] David G Clark. Transient dynamics of associative memory models. *arXiv preprint arXiv:2506.05303*, 2025.

[68] A Crisanti and H Sompolinsky. Path integral approach to random neural networks. *Physical Review E*, 98(6):062120, 2018.

[69] Moritz Helias, David Dahmen, Moritz Helias, and David Dahmen. Functional formulation of stochastic differential equations. *Statistical Field Theory for Neural Networks*, pages 57–67, 2020.

[70] Xuanyu Shen and Yu Hu. Covariance spectrum in nonlinear recurrent neural networks. *arXiv preprint arXiv:2508.05288*, 2025.

Methods and Materials

Dynamical mean-field theory

We now describe the full theory in greater detail, leaving derivations of the results to the Appendix. The dynamics describing the recurrent network (Eq. 1) can be interpreted as a bidirectionally coupled system of neurons and latent variables (Fig. 7A). Each latent variable is defined as a sum over neuronal firing rates, and in turn the input to each neuron can be written as a sum over latent variables:

$$z_r^{(\mu)}(t) = D^{(\mu)} \sum_i n_i^{(\mu,r)} \phi_i(t) \tag{9}$$

$$\dot{x}_i(t) = -x_i(t) + \sum_{\mu=1}^{P} \sum_{r=1}^{R} m_i^{(\mu,r)} z_r^{(\mu)}(t). \tag{10}$$

The magnitudes of these latent variables can be large ($\mathcal{O}(\sqrt{N})$, as in Fig. 2A) or small ($\mathcal{O}(1)$, as in Fig. 2E). We refer to any task whose latent variables are large as a "dominant" task, and the others as "non-dominant" tasks. This distinction is central to the mean-field description, because each neuron's activity is significantly affected by any dominant task, meaning that a finite fraction of its variance can be explained by that task's variables. This is not true for the non-dominant tasks, although they collectively provide non-negligible input as an ensemble. A task can be dominant for two reasons: 1) initial conditions or inputs pushing the state of the network to preferentially align with the task subspace, or 2) self-amplifying dynamics within the task subspace, as shown in Fig. 3. A random network state results in weak latent variables for every task.

We set aside a separate index $\mu^* = 1^*, \ldots, P^*$, with $P^* \geq 0$ small, to enumerate tasks $z_r^{(\mu^*)}(t)$ with dominant $\mathcal{O}\big(\sqrt{N}\big)$ activity, if there are any.[1] Then $\mu = 1, \ldots, P$ continues to enumerate the many non-dominant tasks with $\mathcal{O}(1)$ activity. In the $N, P \to \infty$ limit with fixed ratio $\alpha = P/N$, fixed P^*, and fixed R, we can statistically describe the trajectories of all neurons and weak-task-related latent variables as decoupled processes, rather than as solutions to the high-dimensional dynamics of Eq. (1). In this mean-field description, the activity of the ith neuron $x_i(t)$ and the rth latent variable $z_r^{(\mu)}(t)$ of the μth non-dominant task evolve as

$$\dot{x}_i(t) = -x_i(t) + \eta_i^x(t) + \sum_{\mu^*=1^*}^{P^*} \sum_{r=1}^{R} m_i^{(\mu^*,r)} z_r^{(\mu^*)}(t) + \int_{-\infty}^{t} dt' G^z(t,t') \phi_i(t') \quad \eta_i^x(t) \sim \mathcal{GP}(0, Q^z(t,t')) \tag{11}$$

$$z_r^{(\mu)}(t) = \eta_r^{(\mu)}(t) + D^{(\mu)} \sum_{r'=1}^{R} A_{rr'}^{(\mu)} \int_{-\infty}^{t} dt' S^{\phi}(t,t') z_{r'}^{(\mu)}(t') \qquad \eta_r^{(\mu)}(t) \sim \mathcal{GP}\big(0, \big(D^{(\mu)}\big)^2 C^{\phi}(t,t')\big). \tag{12}$$

In deriving the theory, several functions of time emerge that are shared by all neurons and tasks: four kernel

[1] Although in the main text we emphasized that there can be at most one dominant task, this is true for sufficiently long times (to forget initial conditions) and when there are no external inputs. Generally, initial conditions or external input drive can set multiple task-related latent variables to be large. Once inputs are released, the network eventually equilibrates to having at most one dominant task. In all cases, only finitely many tasks can be dominant since the state vector norm is at most $\sqrt{N}$.

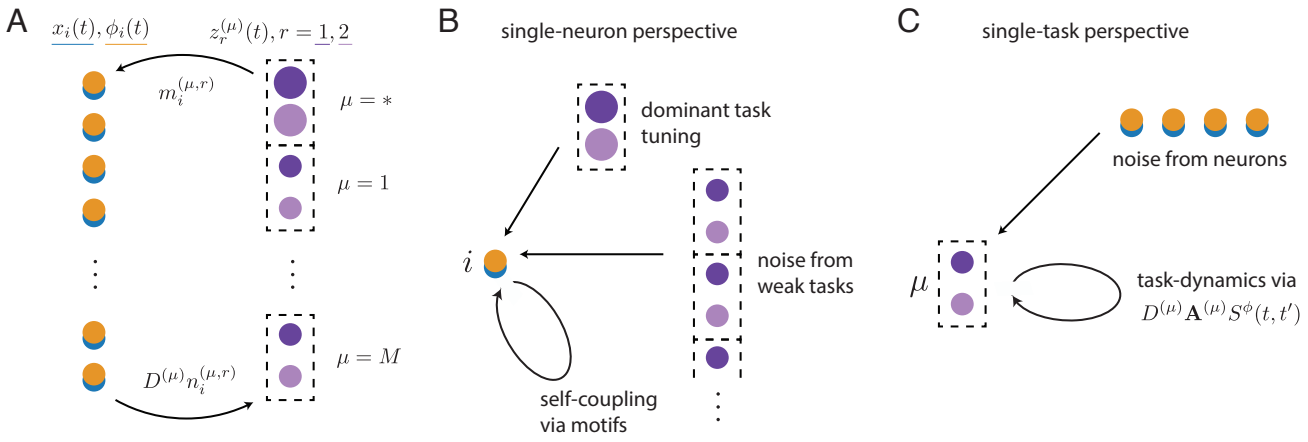

Figure 7: Conceptual illustration of the DMFT. **(A)** Cartoon illustrating an alternative interpretation of the neuron-to-neuron recurrent dynamics, as continuous-time, bidirectional interactions between a reservoir of neurons (blue: currents; yellow: firing rates) and a reservoir of latent variables (dark and light purple). Here we have a single set of dominant latent variables ($P^* = 1$), illustrated as larger purple disks. **(B)** Types of effective input a single neuron receives in the DMFT: input from dominant task latent variables with tuning determined by $m_i^{(\mu^*,r)}$ (top); summed input from weak-task latent variables, which is well described as a sample $\eta_i^x(t)$ from a Gaussian process (middle); and nonlinear effective self-interaction that arises through coherent summation of the neuron's perturbative effect on latent-variable trajectories (bottom). **(C)** Types of effective input a single weak-task latent variable receives in the DMFT: summed input from neurons, which is well described as a sample $\eta_r^{(\mu)}(t)$ from a Gaussian process (top); and effective self-interactions that are mediated by single-neuron response properties as well as task-specific parameters $D^{(\mu)}$ and $\mathbf{A}^{(\mu)}$ (bottom).

functions $C^{\phi}(t,t')$, $S^{\phi}(t,t')$, $Q^{z}(t,t')$, $G^{z}(t,t')$ that describe effective noise statistics and interactions, plus the set of dominant-task latent variable trajectories $\{z_r^{(\mu^*)}(t)\}_{\mu^*,r}$. (In the main text, we wrote $S^{\phi}(t,t') = \langle \phi'(t) \rangle F(t,t')$ and $G^{z}(t,t') = G(t,t')$ to simplify notation—see section on self-consistency conditions.) All these functions' values depend collectively on $\{(D^{(\mu)}, \mathbf{A}^{(\mu)})\}_{\mu=1}^{P}$ and α, and they can be calculated by enforcing self-consistency, as we describe later.

Each neuron i is described in Eq. (11) as leaky integration of 3 current sources (Fig. 7B):

1. A Gaussian process noise term $\eta_i^x(t)$ arising from interactions with the large reservoir of weak-task-related latent variables and capturing the effects of chaotic fluctuations. Its temporal statistics $Q^z(t,t')$ are inherited from average statistics of non-dominant-task latent variables.

2. Drive from the dominant-task latent variables $z_r^{(\mu^*)}(t)$ via the neuron-specific loadings $m_i^{(\mu^*,r)}$.

3. A nonlinear self-coupling to the neuron's *own* firing rate, which is generally a consequence of introducing task-related overlaps. The kernel mediating this coupling $G^z(t,t')$ is determined by weighted averages of the induced network connectivity motifs of all orders.

Similarly, each weakly active task μ is described in Eq. (12) as an integral equation with two input sources (Fig. 7C):

1. A Gaussian process noise term $\eta_r^{(\mu)}(t)$ that arises from the sum over many irregular neuronal activity trajectories via Eq. (9). Its temporal statistics $\left(D^{(\mu)}\right)^2 C^{\phi}(t,t')$ are determined by the average neuronal autocovariance function and scaled by $D^{(\mu)}$.

2. A coupling among task μ's latent variables mediated by $D^{(\mu)}\mathbf{A}^{(\mu)}$ and the kernel $S^{\phi}(t,t')$. The kernel $S^{\phi}(t,t')$ is determined by average impulse-response properties of the neuronal population. The average gain is a crucial factor in determining $S^{\phi}(t,t')$, since (un)saturated neurons are less (more) responsive to inputs.

The dominant-task latent variables satisfy

$$z_r^{(\mu^*)}(t) = \left\langle D^{(\mu^*)} n_i^{(\mu^*,r)} \left\langle \phi(x_i(t)) \right\rangle_{W_i^x} \right\rangle_i , \tag{13}$$

where $\langle \cdot \rangle_{W_i^x}$ indicates averaging over the random process described by Eq. (11) for a given neuron i. In general, this expression does not have a simple closed-form solution, and calculating the forward dynamics requires iterative numerical approaches to maintain self-consistency. However, in special cases where currents $x_i(t)$ are Gaussian, the dynamics of $z_r^{(\mu^*)}(t)$ essentially reduce to the simple form of Eq. (3) from the single-task network case, by Stein's Lemma.[2]

In summary, our theory is able to completely describe network activity for an arbitrary α and ensemble of task dynamics, at the level of both single neurons and task-specific subspaces. Inputs to each can be decomposed into regular, task-related signals and effectively noisy fluctuations due to chaos, with mutually determined statistics that can be analytically solved for.

[2]With a small difference: $\langle \phi'(t) \rangle$ is determined by $z_r^{(\mu^*)}(t)$ as well as a constant interference term from non-dominant tasks. But since this term has Gaussian statistics, Stein's Lemma still applies.

Note on N-scaling

Here we summarize how various structural and dynamical quantities scale with the network size N. Components of loading vectors ($\mathbf{m}^{(\mu,r)}$, $\mathbf{n}^{(\mu,r)}$ for $r = 1, \ldots, R$, $\mu = 1, \ldots, P$) are $\mathcal{O}(1/\sqrt{N})$, resulting in elements of each single-task weight matrix ($D^{(\mu)} \sum_{r=1}^{R} \mathbf{m}^{(\mu,r)} (\mathbf{n}^{(\mu,r)})^T$ for $\mu = 1, \ldots, P$) being $\mathcal{O}(1/N)$. Because the number of tasks P is $\mathcal{O}(N)$, each single-task weight matrix has zero mean, and the single-task weight matrices are sampled independently, the elements of the full weight matrix that sums over all single-task matrices is $\mathcal{O}(1/\sqrt{N})$.

Neural activities ($x_i(t)$, $\phi_i(t)$ for i = 1, ..., N) are $\mathcal{O}(1)$. Projections of neural activity onto single-task subspaces ($z_r^{(\mu)}(t) = \sum_{i=1}^{N} n_i^{(\mu,r)} \phi_i(t)$ for tasks $\mu = 1, \ldots, P$ and within-task components $r = 1, \ldots, R$) scale in two possible ways depending on whether a task μ is dominant or non-dominant. In particular, projections onto non-dominant task subspaces are $\mathcal{O}(1)$, while projections onto dominant-task subspaces are $\mathcal{O}(\sqrt{N})$. This is because, in the dominant-task case, the network state preferentially aligns with their associated subspaces, and the overall state norm $\|\boldsymbol{\phi}(t)\|$ is $\mathcal{O}(\sqrt{N})$ while the norms of loadings vectors are approximately 1. In the language of associative-memory networks that store patterns as fixed points, these forms of (non-)alignment are referred to as uncondensed or condensed.

These distinctions in scaling are visible when comparing networks of different sizes (Fig. 2D light vs. dark cyan). The activity level of dominant task $\mu = 1$ is scaled by approximately $\sqrt{2}$ when the number of neurons is doubled, but past the transition to chaos—when task 1 loses dominance—its activity level drops to a value that is the same for networks of different sizes.

Connectivity statistics

The free parameters of the model are P, R, and $\{(D^{(\mu)}, \mathbf{A}^{(\mu)})\}_{\mu=1}^{P}$. Statistics of J_{ij} can be written in terms of these parameters. The overall strength of the connectivity, which we call g_{eff}, is given by the variance of the off-diagonal elements of $\mathbf{J}$ and depends on the second moment of the task weights (but not the overlaps):

$$g_{\text{eff}}^2 = N \left\langle J_{ij}^2 \right\rangle_{\mathbf{J}} = \alpha R \left\langle D^2 \right\rangle_{D,\mathbf{A}} . \tag{14}$$

Like the analogous g parameter of [36], holding other parameters fixed while increasing this degree of "disorder" eventually effects a transition to chaotic activity, with stronger chaotic fluctuations with increasing g_{eff}. However, in this model with connection symmetries, g_{eff} does not alone determine the network dynamics.

The mean connection strength is

$$\langle J_{ij} \rangle_{\mathbf{J}} = \delta_{ij} \alpha \left\langle D \mathrm{Tr}(\mathbf{A}) \right\rangle_{D,\mathbf{A}} , \tag{15}$$

implying strong autapses $J_{ii} \sim \mathcal{O}(1)$ when the overlaps have a nonzero average trace. This homogeneous mean autapse strength $\langle J_{ii} \rangle_{\mathbf{J}}$ can be interpreted as a zeroth-order connectivity motif ρ_0. The first-order connectivity motif strength ρ_1 is the correlation between transposed elements for $i \neq j$,

$$\rho_1 = N \left\langle J_{ij} J_{ji} \right\rangle_{\mathbf{J}} = \alpha \left\langle D^2 \mathrm{Tr}(\mathbf{A}^2) \right\rangle_{D,\mathbf{A}} . \tag{16}$$

Higher-order connectivity motifs, e.g. $\rho_2 = \langle J_{ij} J_{jk} J_{ki} \rangle_{\mathbf{J}}$, arise from nonzero average traces of higher matrix

powers of $\mathbf{A}$.

If present, these symmetries will lead to effective self-coupling between each neuron and itself, whether through strong self-connections or multi-synaptic pathways. In a Neumann expansion of the the spontaneous state self-coupling kernel $G^z(\omega)$ (see section on self-consistency conditions),

$$G^z(\omega) = \alpha \left\langle D \ \mathrm{Tr}\Big(\mathbf{A}\left(\mathbf{I} - DS^\phi(\omega)\mathbf{A}\right)^{-1}\Big)\right\rangle_{D,\mathbf{A}} \tag{17}$$

$$= \alpha \sum_{k=0}^{\infty} (S^\phi(\omega))^k \left\langle D^{k+1}\mathrm{Tr}(\mathbf{A}^{k+1})\right\rangle_{D,\mathbf{A}} \tag{18}$$

$$= \alpha \sum_{k=0}^{\infty} (S^\phi(\omega))^k \rho_k, \tag{19}$$

which matches our intuition that neurons self-interact through strong network motifs, modulated by the appropriate power of the impulse-response function.

This self-coupling is responsible for non-Gaussian currents $x(t)$, which makes calculations difficult. However, designing a non-trivial ensemble of dynamics that eliminates ρ_k for every k is possible, and in such cases we get Gaussian $x(t)$.

Stability condition for $D^{(\mu)}_{\mathrm{crit},1}$

Computing the stability condition for an integral equation of the form

$$z_r(t) = \eta^{z_r}(t) + D\sum_{r'=1}^{R} A_{rr'}\left[S^\phi * z_{r'}\right](t') \tag{20}$$

requires the Laplace transform of the transfer function $\mathbf{M}^{(\mu)}(s) = (\mathbf{I} - D^{(\mu)}S^\phi(s)\mathbf{A}^{(\mu)})^{-1}$. All singularities s^* of this inversion must satisfy $\mathfrak{Re}(s^*) < 0$ for a value of $D^{(\mu)}$ to produce stable dynamics. This condition is generally difficult to calculate without a closed-form expression for $S^\phi(\tau)$, as it requires numerical estimation of the Laplace transform. If we assume vanishing connectivity motifs and recover Gaussian currents, implying $S^\phi(\tau) = \Theta(\tau)\langle\phi'\rangle e^{-\tau}$, the dynamics become equivalent to a leaky, noise-driven ODE, and the stability condition simplifies to

$$D^{(\mu)}_{\mathrm{crit},1} = \frac{1}{\langle\phi'\rangle\,\mathcal{S}^{(\mu)}}, \tag{21}$$

where $\mathcal{S}^{(\mu)} = \max\left[\mathfrak{Re}\left(\lambda\right)|\lambda \text{ eigenvalue of } \mathbf{A}^{(\mu)}\right]$ is the spectral abscissa of $\mathbf{A}^{(\mu)}$.

Self-consistency conditions

For the rest of the methods section and appendices, we often drop the neuron index i and non-dominant task index μ, because in the DMFT they are thought of as ensembles from which N and P samples are taken, respectively, rather than explicit lists of elements. A given non-dominant task μ has an associated set of D, $\mathbf{A}$ values that are sampled from a joint distribution $\Pr(D, \mathbf{A})$, and a given neuron i has an associated set of normalized (to $\mathcal{O}(1)$) couplings $c = \{\overline{m}^{(\mu^*,r)}, \overline{n}^{(\mu^*,r)}\}_{\mu^* r}$ to the dominant-task latent

variables. We denote the random process for a single neuron with given couplings c as W_c^x, and the random process for a single non-dominant task with given dynamics D, $\mathbf{A}$ as $W_{D,\mathbf{A}}^{\vec{z}}$. The fully general set of self-consistency conditions are

$$S^\phi(t,t') = \left\langle \left\langle \frac{\delta\phi(t)}{\delta I(t')} \right\rangle_{W_c^x} \right\rangle_c \tag{22}$$

$$C^\phi(t,t') = \left\langle \left\langle \phi(t)\phi(t') \right\rangle_{W_c^x} \right\rangle_c \tag{23}$$

$$Q^z(t,t') = \alpha \left\langle \left\langle \sum_r z_r(t) z_r(t') \right\rangle_{W_{D,\mathbf{A}}^{\vec{z}}} \right\rangle_{D,\mathbf{A}} \tag{24}$$

$$G^z(t,t') = \alpha \left\langle D \sum_{rr'} A_{rr'}^T \left\langle \frac{\delta z_r(t')}{\delta I_{r'}(t)} \right\rangle_{W_{D,\mathbf{A}}^{\vec{z}}} \right\rangle_{D,\mathbf{A}} \tag{25}$$

$$z_r^{(\mu^*)}(t) = D^{(\mu^*)} \left\langle \bar{n}^{(\mu^*,r)} \left\langle \phi(t) \right\rangle_{W_c^x} \right\rangle_c . \tag{26}$$

Each kernel function (or order parameter) is computed as a within-process average parameterized by some value, e.g. $\langle\cdot\rangle_{W_c^x}$ or $\langle\cdot\rangle_{W_{D,\mathbf{A}}^{\vec{z}}}$, that is itself averaged over the ensemble of all such parameters, e.g. $\langle\cdot\rangle_c$ or $\langle\cdot\rangle_{D,\mathbf{A}}$.

Thus, the DMFT amounts to a bipartite ensemble of neurons and task-related latent variables with mutually referential statistics. The kernels $Q^z(t,t')$ and $G^z(t,t')$ defining the single-neuron stochastic dynamics are inherited from averages over the ensemble of latent variables, hence their superscripts referencing their definition. Conversely, the kernels $C^\phi(t,t')$ and $S^\phi(t,t')$ defining single-task-latent-variable stochastic dynamics are inherited from averages over the ensemble of neurons. Again, to avoid explaining the bipartite structure implicit in this notation in the main text, we simplified notation and replaced $S^\phi(t,t') = \langle\phi'(t)\rangle\, F(t,t')$ and $G^z(t,t') = G(t,t')$.

Spontaneous state self-consistency conditions

In the spontaneous state, the self-consistency conditions simplify tremendously. First, there are no $z_r^{(\mu^*)}(t)$ values that need to be solved for, and thus the single-neuron dynamics simplify. Second, we can assume stationarity of the kernel functions, i.e. that each kernel function depends only on $t - t'$, not t and t' separately. (This is a fairly mild additional assumption, but one that may be broken with sufficiently strong connectivity motifs [66]). Neurons are now statistically homogeneous, so we do not need to average over the ensemble of processes W_c^x and can instead consider a single representative process W^x. The non-dominant-task latent variables are linear, and their associated autocovariance functions and impulse-response functions can be written compactly in Fourier space. With all of these simplifications, the self-consistency equations take the form

$$C^\phi(\tau) = \langle\phi(t)\phi(t+\tau)\rangle_{W^x} \qquad S^\phi(\tau) = \langle\delta\phi(t)/\delta I(t-\tau)\rangle_{W^x} \tag{27}$$

$$Q^z(\omega) = \alpha \left\langle D^2 \left\|\mathbf{M}(\omega)\right\|_{\mathcal{F}}^2 \right\rangle_{D,\mathbf{A}} C^\phi(\omega) \qquad G^z(\omega) = \alpha \left\langle D \ \mathrm{Tr}\Big(\mathbf{A}\mathbf{M}(\omega)\Big) \right\rangle_{D,\mathbf{A}} , \tag{28}$$

where $\mathbf{M}(\omega) = \big(\mathbf{I} - DS^\phi(\omega)\mathbf{A}\big)^{-1}$ is the transfer function for a task with dynamics according to D, $\mathbf{A}$.

Methods for solving the DMFT

To solve the DMFT means to calculate the kernel functions $C^\phi(t,t')$, $S^\phi(t,t')$, $Q^z(t,t')$, $G^z(t,t')$, plus the set of dominant-task latent variable trajectories $\{z_r^{(\mu^*)}(t)\}_{\mu^*,r}$ for an ensemble of task dynamics in a given discretization of time.

Fully general t, t' case

Solving the theory in this most general case is in principle possible by iteratively enforcing self-consistency in sampling a large number of single-neuron trajectories for provisional guesses of each kernel function and the dominant-task latent variable trajectories. A summary of the process:

1. Start with initial guesses for $C^\phi(t,t')$, $S^\phi(t,t')$, and $\{z_r^{(\mu^*)}(t)\}_{\mu^*,r}$.

2. From $C^\phi(t,t')$ and $S^\phi(t,t')$, compute $Q^z(t,t')$ and $G^z(t,t')$ by explicitly averaging over the P non-dominant tasks in the ensemble. Since the non-dominant latent dynamics are linear, we can take advantage of having analytic expressions for their within-task cross-covariances and impulse-response functions, which are necessary ingredients for computing $Q^z(t,t')$ and $G^z(t,t')$.

3. Simulate N_{batch} samples of Eq. (11) in parallel, each requiring an independent sample of the Gaussian noise $\eta^x(t)$ (by the Cholesky decomposition of $Q^z(t,t')$), as well as a sample of the dominant-task coupling strengths $m_i^{(\mu^*,r)}$.

4. Empirically measure the single-unit autocovariance, impulse-response functions, and dominant-task latent variable trajectories based on these samples. (The autocovariance and dominant latent variables are straightforward to calculate from their definitions, while the impulse-response function requires differentiating the individual $x_i(t)$ trajectories.)

5. Smoothly update the values of $C^\phi(t,t')$, $S^\phi(t,t')$, and $\{z_r^{(\mu^*)}(t)\}_{\mu^*,r}$ based on these empirical measurements

6. Repeat 2-5 until convergence.

Details for a similar method applied to modern Hopfield networks are outlined in [67].

Although this method, as well as some other methods we present below, is numerical in nature, it still does not require ever constructing or simulating a *network* according to Eq. (1) with a synaptic weight matrix—only sampling neural activity and task-related latent variables as random processes with mutually referential statistics.

We did not ultimately present any results with this method, as it is prohibitively computationally expensive with difficult boundary conditions. However, we summarize it to the reader to emphasize that, as written, the DMFT simply encodes a set of mutually referential statistics between a reservoir of neurons and a reservoir of non-dominant latent variables, with a special set of dominant-task latent variable trajectories. In principle, a self-consistent set of such statistics can be found by iterative sampling and updating. However, calculation becomes much easier in certain special cases, as we articulate below.

Spontaneous state case

At a high-level, the general method for solving the theory for the spontaneous state is similar to the fully general method, but now we have much simpler Fourier representations of the kernel functions, rather than explicit t and t' dependence, and we do not have to worry about dominant-task latent variable trajectories. The process is simplified to:

1. Begin with an initial guess for $C^\phi(\tau)$ and $S^\phi(\tau)$.
2. Based on these values, compute $Q^z(\omega)$ and $G^z(\omega)$ in Fourier space using Eq .(28) with an explicit average over the $D^{(\mu)}$, $\mathbf{A}^{(\mu)}$ ensemble.
3. Sample N_{batch} Gaussian noise samples of $\eta_i^x(t)$ based on the kernel $Q^z(\tau)$, and compute $x_i(t)$ and $\phi_i(t)$ from Eq. (11) without dominant-task latent variables.
4. Empirically measure $C^\phi(\tau)$ and $S^\phi(\tau)$, leveraging the Furutsu-Novikov theorem for estimating $S^\phi(\tau)$: $S^\phi(\omega) = C^{\phi\eta}(\omega)/C^\eta(\omega)$, where $C^{\phi\eta}(\tau) = \langle \phi_i(t)\eta_i^x(t+\tau)\rangle_{t,i}$ is the cross-covariance of $\phi_i(t)$ and the driving noise $\eta_i^x(t)$ [43].
5. Smoothly update

$$C^\phi(\tau) \leftarrow (1-\delta_{\text{mem}})C^\phi(\tau) + \delta_{\text{mem}}\widehat{C^\phi(\tau)} \tag{29}$$

$$S^\phi(\tau) \leftarrow (1-\delta_{\text{mem}})S^\phi(\tau) + \delta_{\text{mem}}\widehat{S^\phi(\tau)} \tag{30}$$

   with memory parameter $\delta_{\text{mem}} \in (0,1)$.
6. Repeat 2-5 until convergence.

More details for a similar method (for $R = 1$ special case) can be found in [45]. This method was used for the theory fits in Fig. S5B,C,D, with $\tau_{\max} = 120$, $d\tau = 0.05$, $N_{\text{batch}} = 10^4$, $N_{\text{iter}} = 400$, and $\delta_{\text{mem}} = 0.2$.

Spontaneous state with reservoir of limit cycles

In a special case of an ensemble with constant $D^{(\mu)} = D$ and $\mathbf{A}^{(\mu)}$ i.i.d. sampled as block-Haar matrices (see Appendix), the single-unit self-coupling kernel $G^z(\tau)$ vanishes. Crucially, this means the currents $x(t)$ now have Gaussian statistics, and we can leverage closed-form Gaussian integrals in our calculations. Moreover, the kernel that determines input noise to each neuron $Q^z(\tau)$ admits a closed-form expression in Fourier space without the need for an explicit average over tasks μ. In turn, $C^x(\omega)$ has a closed-form relationship with $C^\phi(\omega)$ that is parameterized by the (constant) average gain $\langle\phi'\rangle$ which depends on $C^x(\tau = 0)$. This can be translated into a second-order ODE (Eq. 150) and solved for as in [36], with the additional step of self-consistently determining the gain.

Alternatively, we can directly solve in the Fourier domain by iteratively enforcing

$$C^x(\omega) = \frac{g_{\text{eff}}^2}{1 - (\gamma D\,\langle\phi'\rangle)^2 + \omega^2} C^\phi(\omega), \tag{31}$$

where $g_{\text{eff}}^2 = \alpha R \left\langle D^2 \right\rangle_D$ (defined in Eq. (14)), over the whole spectrum. We iterate on the following process:

1. Start with an initial guess for $C^x(\omega)$.
2. Take the inverse Fourier transform to get $C^x(\tau)$.
3. Calculate $C^\phi(\tau)$ in terms of $C^x(\tau)$ via Eq. (41) and $\langle\phi'\rangle$ in terms of $C^x(\tau=0)$ via Eq. (40).
4. Update $C^x(\omega)$ using Eq. (31) with memory parameter $\delta_{\text{mem}} \in (0,1)$.
5. Repeat steps 2-4 until convergence.

We evaluated over a temporal window of $\tau_{\max} = 200$, $d\tau = 0.05$ with $\delta_{\text{mem}} = 0.5$.

Task-selected state with reservoir of limit cycles

For notational simplicity, we assume there is one dominant oscillation, which we index by $*$. We first write down the dynamics for the two latent variables $z_r^{(*)}(t)$ of the dominant pattern with pattern strength $D^{(*)}$ and overlap $\mathbf{A}^{(*)}$, with intrinsic frequency $\omega^{(*)} = \tan(\theta^{(*)})$. By the assumption of dominance, $z_r^{(*)}(t) \sim \mathcal{O}(\sqrt{N})$, we can divide by a factor of $\sqrt{N}$, making the $\mathcal{O}(1)$ noise vanish for large N, and get

$$\frac{z_r^{(*)}(t)}{\sqrt{N}} = D^{(*)} \sum_{r'} A_{rr'}^{(*)} \int dt' S^\phi(t,t') \frac{z_{r'}^{(*)}(t')}{\sqrt{N}}, \tag{32}$$

which reduces in the case of Gaussian $x(t)$ to

$$(1+\partial_t)\left(\frac{z_r^{(*)}(t)}{\sqrt{N}}\right) = D^{(*)} \langle\phi'\rangle \sum_{r'} A_{rr'}^{(*)} \left(\frac{z_{r'}^{(*)}(t)}{\sqrt{N}}\right), \tag{33}$$

because $S^\phi(t,t') = \Theta(t-t') \langle\phi'(t)\rangle e^{-(t-t')}$. These dynamics are also consistent with the direct definition of the condensed patterns from the self-consistency conditions and applying Stein's lemma to Eq. (26), which the Gaussianity of $x(t)$ permits. Since ϕ is saturating, this system has a steady-state oscillatory solution of a particular amplitude ρ. Assuming marginal stability of the oscillation enforces constant $\langle\phi'\rangle = 1/(\gamma D^{(*)} \cos(\theta^{(*)})$, which depends on both ρ and the noise variance—though we don't yet know how this is partitioned between them. For now, we write the steady-state solution for the dominant pattern in terms of ρ as

$$\frac{z_1^{(*)}(t)}{\sqrt{N}} = \rho\cos(\omega^{(*)}t) \qquad \frac{z_2^{(*)}(t)}{\sqrt{N}} = \rho\sin(\omega^{(*)}t). \tag{34}$$

The dynamics for any one neuron $x_m(t)$ features noise drive from $\eta^x(t)$ sampled iid as a Gaussian process with stationary kernel $Q^z(\tau)$, determined like before as an average over the 2-point functions of the weak-task trajectories, plus deterministic drive from this dominant task with coupling $m^{(*,r)}$:

$$(1+\partial_t)x_m(t) = \eta^x(t) + m^{(*,1)}\sqrt{N}\rho\cos(\omega^{(*)}t) + m^{(*,2)}\sqrt{N}\rho\sin(\omega^{(*)}t) \tag{35}$$

$$= \eta^x(t) + m_1\rho\cos(\omega^{(*)}t) + m_2\rho\sin(\omega^{(*)}t) \tag{36}$$

where we have defined $m^{(*,r)}\sqrt{N} = m_r \sim \mathcal{O}(1)$ as the order-1 set of couplings between this arbitrary unit and the condensed oscillation, which encode both amplitude and relative phase information. The P subscript in $x_m(t)$ indicates this dependence on $m_1, m_2 \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.

As currently setup, this problem is reminiscent of [37], but with a few key differences:

- The noise is not due to an i.i.d. reservoir, but rather a reservoir of low-rank matrices with 2×2-block Haar overlaps. This has an effect on the temporal shape of the noise statistics (the difference between Eq. 150 and the equation for an unstructured network).
- In [37], neural inputs feature phase heterogeneity across units but uniform amplitude. Our case has heterogeneity in both amplitude and phase. This changes how we write down our average ϕ autocovariance function.
- The overall scaling of the (heterogenenous) amplitudes $\rho\|m\|$ is not a free parameter but rather self-consistently determined with $D^{(*)}$, the constant D on the uncondensed patterns, and current variance. This changes the axes on the bifurcation diagram, since $D^{(*)}$ is what we freely vary and not ρ, but once equilibrium ρ is determined, the effect on the dynamics is similar to that of the external current from [37].

We write $x_m(t) = x_m^0(t) + x_m^1(t)$ as a sum of a filtered deterministic part and a filtered noise part, the former having a closed-form expression:

$$x_m^0(t) = \frac{\rho}{\sqrt{1+(\omega^{(*)})^2}}\big(m_1\cos(\omega^{(*)}t-\theta_0) + m_2\sin(\omega^{(*)}t-\theta_0)\big), \tag{37}$$

which has autocovariance $C^{x^0}(\tau) = \tilde{\rho}^2\cos(\omega^{(*)}\tau)$ for $\tilde{\rho}^2 = \rho^2/(1+(\omega^{(*)})^2)$. The filtered noise part $x^1(t)$ has autocovariance with the same relation to $C^\phi(\omega)$ as in the spontaneous case:

$$C^{x^1}(\omega) = \frac{g_{\text{eff}}^2}{1+\omega^2-(\gamma D\langle\phi'\rangle)^2}C^\phi(\omega). \tag{38}$$

(See Appendix for a derivation of this relation.) If we can write an expression for $C^\phi(\tau)$ in terms of the statistics of $x_m^0(t)$ and $x_m^1(t)$, then we close the equations. Since they are both Gaussian, their sum is Gaussian, and $C^x(\omega) = C^{x^0}(\omega) + C^{x^1}(\omega)$ determines $C^\phi(\tau)$ in closed form (after taking the IFFT). That suggests the following iterative process for solving the DMFT:

1. For a given $D^{(*)}$ and $\omega^{(*)}$, first calculate the average gain $\langle\phi'\rangle$ set by the marginality condition of the dominant oscillation.
2. Set an initial guess $C^{x^1}(\omega)$.
3. Take the IFFT to obtain $C^{x^1}(\tau)$.
4. Invert Eq. (40) to solve for the complementary $\tilde{\rho}$ value that combines with $C^{x^1}(\tau=0)$ to give the already determined $\langle\phi'\rangle$ value.
5. Compute $C^x(\tau) = C^{x^0}(\tau) + C^{x^1}(\tau)$ in terms of our provisional $C^{x^1}(\tau)$ and $\tilde{\rho}$ values.

6. Compute $C^{\phi}(\tau)$ using Eq. (41).

7. Take the FFT of $C^{\phi}(\tau)$ to obtain $C^{\phi}(\omega)$

8. Update estimate for $C^{x^1}(\omega)$ using $C^{\phi}(\omega)$ via Eq. (38) with memory parameter $\delta_{\text{mem}} \in (0, 1)$.

9. Repeat steps 2-8 until convergence.

We evaluated over a temporal window of $\tau_{\max} = 200$, $d\tau = 0.05$ with $\delta_{\text{mem}} = 0.5$. Additionally, for a given dominant frequency $\omega^{(*)}$ value, we adjusted the value to precisely match the nearest on-bin frequency value to avoid spectral smearing.

Phase boundaries

Computing $D^{(\mu)}_{\text{crit},1}$ is straightforward from Eq. (21) and only requires solving the spontaneous state DMFT and computing $\langle \phi' \rangle$. To compute $D^{(\mu)}_{\text{crit},2}$, we used a midpoint solver to identify the value of $D^{(\mu)} \in (D^{(\mu)}_{\text{crit},1}, 1.7 \times D^{(\mu)}_{\text{crit},1})$ that results in a $C^x(\tau)$ curve with

$$\frac{C^x(0) - C^x(2\pi/\omega^{(*)})}{C^x(0)} \approx 0.02, \tag{39}$$

where 0.02 is our criterion for when the relative magnitude of fluctuations is effectively vanished.

Gaussian integrals

Any special case with Gaussian currents (e.g. the reservoir of oscillations) can take advantage of Gaussian integrals to simplify calculation. By our choice of nonlinearity $\phi(x) = \text{erf}(\sqrt{\pi}x/2)$, we can leverage certain closed-form expressions:

$$\left\langle \phi'(x_i(t)) \right\rangle_i = \int dy (2\pi)^{-1/2} e^{-y^2/2} \phi' \left(\sqrt{C^x(t,t)} y \right) = \sqrt{1 + (\pi/2) C^x(t,t)} \tag{40}$$

$$\langle \phi(x_i(t)) \phi(x_i(t+\tau) \rangle_i = \frac{2}{\pi} \sin^{-1} \left(\frac{C^x(t, t+\tau)}{2/\pi + C^x(t,t)} \right). \tag{41}$$

Whenever we consider Gaussian currents $x_i(t)$ in this work, they are always zero mean with possibly time-dependent variance $C^x(t,t)$, but more often we are in the stationary case $\tau = t - t'$ with $C^x(0) = C^x(t,t)$.

Dimensionality of task-selected state

See [45] for a detailed walkthrough of dimensionality calculations, which can in principle apply to the spontaneous ($\mathcal{O}(N)$ dimensionality) or task-selected ($\mathcal{O}(1)$ dimensionality) states. However, in practice we found poor numerical fits for dimensionality in the spontaneous state when overlaps were even modestly strong.

The dimensionality calculation simplifies tremendously for the task-selected state. We can leverage the duality between neuron-neuron covariance (used most directly to define the PR) and time-time covariance (which has identical spectrum to the neuron-neuron covariance) as first shown in [45]. For a given variable

$a \in \{x, \phi\}$, its PR can be computed in terms of its autocovariance function $C^a(\tau)$ and four-point function $\Psi^a(\tau, \tau') = \frac{1}{N}\sum_{i,j=1}^{N} C^a_{ij}(\tau)C^a_{ij}(\tau')$

$$\mathrm{PR}^a = \frac{C^a(0)^2}{\Psi^a(0,0)}. \tag{42}$$

Leveraging the duality discussed above, we can alternatively calculate $\Psi^a(\tau, \tau')$ as

$$\Psi^a(\tau, \tau') = N \left\langle C^a(t, t')C^a(t + \tau, t' + \tau') \right\rangle_{t,t'}. \tag{43}$$

For $\tau = \tau' = 0$, we're essentially computing the average square of $C^a(t, t')$ over the whole (t, t') plane. The values near the $t = t'$ diagonal contribute negligibly compared to the whole plane, so we need only examine the asymptotic off-diagonal behavior when $t - t'$ is large. Any fluctuations $\delta C^a(t, t') = C^a(t, t') - C^a(t - t')$ about the stationarized two-point function are negligible in this case, since typical values over the plane are $\mathcal{O}(1)$ because of the dominant dynamics. Thus we can estimate $\Psi^a(0, 0)$ directly from the autocovariance function $C^a(\tau)$ as

$$\Psi^a(0, 0) = \frac{1}{T}\int_{\tau}^{\tau+T} d\tau' (C^a(\tau'))^2 \tag{44}$$

where the domain of integration depends on the nature of the dominant task. For a limit-cycle task, we need only integrate over one period of an oscillation to get the average behavior over the whole plane, as long as we start past when the amplitude decay is finished (Fig. S7). For a fixed point, we need only integrate over a brief interval when the $C^a(\tau)$ has reached its asymptotic value.

Simulating networks

To validate the theoretical calculations, we simulated networks as well. We varied N depending on how much we needed to reduce finite-size effects in a given context, we always used $dt = 0.05$ and RK4 forward integration. Parameters are in Table 1 for all network simulations in the main and supplementary figures.

Software availability

Core functions for simulating networks and solving the theory are available in `https://github.com/omarschall/mft-theory`, as well as an example notebook showcasing the fits in Fig. 5C.

Additional mathematical details

Connectivity and connection to SVD

Eq. (4) can be equivalently interpreted as the product of three matrices: an $N \times PR$ matrix whose columns comprise the left loading vectors $\mathbf{m}^{(\mu,r)}$, an $PR \times PR$ diagonal matrix whose diagonal elements comprise the task weights $D^{(\mu)}$ each repeated R times, and an $PR \times N$ matrix whose rows comprise the right loading vectors $\mathbf{n}^{(\mu,r)}$ (Fig. S1A). Thus the multi-task network (Eq. 4) is of the same form as a low-rank network with rank PR (Eq. 2).

Just as the Singular Value Decomposition (SVD) decomposes a matrix as a weighted sum of outer products, our model *generates* a matrix as a weighted sum of outer products (Fig. S1A). The loadings

$\mathbf{m}^{(\mu,r)}$ ($\mathbf{n}^{(\mu,r)}$) play the role of the left (right) singular vectors, except they are not perfectly orthonormal but rather orthonormal up to $\mathcal{O}(1/\sqrt{N})$ noise, i.e. $\mathbf{m}^{(\mu,r)} \cdot \mathbf{m}^{(\nu,r')} = \delta_{\mu\nu}\delta_{rr'} + \mathcal{O}(1/\sqrt{N})$. The unrolled index $a = 1, \ldots, M \times R$ corresponding to (μ, r) lists the "singular values" $D^{(a)}$ for each of the $P \times R$ modes (Fig. S1B). For a general matrix, these could be any non-negative values, but here we constrain them to be constant within R blocks, i.e. $D^{(\mu,r)} = D^{(\mu,r')} \equiv D^{(\mu)}$. For an arbitrary matrix, the overlaps between left and right singular vectors are generally unconstrained up to basic geometric consistency, and similarly we are free to generate connectivity through a nearly arbitrary pattern of overlaps $O_{ab} = \left\langle \mathbf{n}^{(a)} \cdot \mathbf{m}^{(b)} \right\rangle$ (Fig. S1C). We consider the special case of overlap matrices with an $R \times R$ block-diagonal structure (Fig. S1C), because keeping R fixed and allowing $P \to \infty$ makes the model mathematically tractable. However, the set of overlaps as specified by $A^{(\mu)}_{rr'}$ is arbitrary[3]. This formulation is equivalent to the "random-mode model" with overlaps from [45], except the symmetries have a block-diagonal, rather than merely diagonal (equivalently $R = 1$), structure.

Full details of small-P case

The small-P case is equivalent to the single-task network case, because if P is small then $R \times P$ is also small. In the large-N limit, there is zero direct cross-talk between latent variables from distinct tasks, and Eq. (3) factorizes into self-contained systems of size R

$$z^{(\mu)}_r(t) = D^{(\mu)} \left\langle \phi'(t) \right\rangle \sum_{r'=1}^{R} A^{(\mu)}_{rr'} \left[L * z^{(\mu)}_{r'} \right](t) \tag{45}$$

up to their sharing of the gain factor via the Gaussian integral of ϕ' with variance given by the squared norm of all the low-pass-filtered latent variables $\kappa^{(\mu)}_r(t) = \left[L * z^{(\mu)}_r \right](t)$:

$$\left\langle \phi'(t) \right\rangle = \int dy \; (2\pi)^{-1/2} e^{-y^2/2} \phi' \left(y \sqrt{\frac{1}{N} \sum_{\mu r} \left(\kappa^{(\mu)}_r(t) \right)^2} \right), \tag{46}$$

which can be evaluated by Eq. (40). Here $L(\tau) = \Theta(\tau) e^{-\tau}$ is a low-pass filter, a closed-form special case of the filter $F(\tau)$ that we can leverage with $G(\tau) = 0$, as in particular occurs when $\alpha = 0$ (i.e. small P) but in other cases as well.

Winner-take-all intuition

We present intuition for why there must always be a dominant task in this case. For a task μ in this setting to be dominant, its linearized dynamics must be unstable about the origin, so that it grows until $z^{(\mu)}_r(t)$ is large enough to influence the nonlinearity $\langle \phi'(t) \rangle$ and self-stabilize. (This necessarily requires $z^{(\mu^*)}_r(t) \sim \mathcal{O}(\sqrt{N})$ because of the N scaling in the argument of ϕ'.) The dynamics are marginally stable when

$$\left\langle \phi'(t) \right\rangle D^{(\mu)} \mathcal{S}^{(\mu)} = 1, \tag{47}$$

[3]Up to the constraint that the overall $2R \times 2R$ loadings covariance matrix, containing the n-n block $\left\langle n^{(\mu,r)}_i n^{(\mu,r')}_i \right\rangle = N^{-1}\delta_{rr'}$ and m-m block $\left\langle m^{(\mu,r)}_i m^{(\mu,r')}_i \right\rangle = N^{-1}\delta_{rr'}$ as well as the n-m blocks given by $A^{(\mu)}_{rr'}$, must be positive definite.

where $\mathcal{S}^{(\mu)} = \max\left[\mathfrak{Re}\left(\lambda\right)|\lambda \text{ eigenvalue of } \mathbf{A}^{(\mu)}\right]$. If no task even achieves $D^{(\mu)}\mathcal{S}^{(\mu)} \geq 1$ with maximal gain $\langle\phi'(t)\rangle = 1$, then the network is in the quiescent state, with all activity decaying to the origin rapidly. To avoid this trivial case, we assume at least one task is sufficiently strong to avoid the quiescent state. Consider the "strongest" such task, i.e. the one with highest value of $D^{(\mu)}\mathcal{S}^{(\mu)}$. To keep itself balanced, this task must set $\langle\phi'(t)\rangle \sim \left(D^{(\mu)}\mathcal{S}^{(\mu)}\right)^{-1}$ so that its dynamics are either marginally stable, or alternate phases of growth and decay cyclically (e.g. an elliptical limit cycle). Since all other tasks share this gain factor and have by definition weaker intrinsic strength, each will have a strictly weaker recurrent drive and therefore feature contracting dynamics. This is true even in cases where multiple tasks are fine-tuned to have the same strength—finite-sampling effects will always result in one task being marginally stronger than another. However, the more fine-tuned two tasks are to have the same strength, the more likely particular network realizations can feature long transient periods of activity where tasks appear "co-dominant," until one eventually takes over.

What makes Gaussian loadings in particular lead to only one task dominating, is the fact that, in such networks, there is only one "population" in the sense of [29]. Thus there is only one gain factor, which equilibrates to a value that marginally stabilizes the strongest dynamics and suppresses all weaker ones, as described above. In a network with multiple populations, multiple gain factors could in principle separately equilibrate to values that marginally stabilize two different tasks' dynamics—a trivial example would be adjoining two completely unconnected single-task networks that we call one network. However, nontrivial examples with overlap and interaction among populations also exist, which we leave to future work.

Fixed-point task dominance

In the main text, we showed an example where the limit cycle dynamics dominate the fixed-point dynamics. Here we swap the $D^{(\mu)}$ values such that the bistable dynamics dominate, and the activity in the task-1 subspace decays to the origin (Fig. S2A,B). The two matrices generating the limit-cycle and fixed-point dynamics, respectively,

$$\mathbf{A}^{(1)} = \begin{pmatrix} 0.8 & -0.4 \\ 0.4 & 0.8 \end{pmatrix}, \qquad \mathbf{A}^{(2)} = \begin{pmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}, \tag{48}$$

were chosen such that they have identical spectral abscissa of $\mathcal{S}^{(1)} = \mathcal{S}^{(2)} = 0.8$, so that the dominant task is whichever simply has larger $D^{(\mu)}$.

Supplemental Figures and Tables

Table 1: Summary of network parameters used in simulations across figures. In all cases, $R = 2$.

Panels	N	P or α	$D^{(\mu)}$	Notes on $A^{(\mu)}$ / sampling
Fig. 2A	2000	$P = 1$	$D = 2.2$	$\mathbf{A} = \begin{pmatrix} 0.8 & 0.4 \\ -0.4 & 0.8 \end{pmatrix}$ (limit cycle)
Fig. 2B	2000	$P = 1$	$D = 2$	$\mathbf{A} = \begin{pmatrix} 0.5 & 0.3 \\ 0.3 & 0.5 \end{pmatrix}$ (bistable)
Fig. 2C	2000	$P = 2$	$D^{(1)} = 2.2,\ D^{(2)} = 2$	$\mathbf{A}^{(1)}$ limit cycle, and $\mathbf{A}^{(2)}$ bistable
Fig. 2D,E,F	varied	varied	$D^{(1)} = 3,\ D^{(\mu>1)} = 2$	$\mathbf{A}^{(1)}$ limit cycle; $\mathbf{A}^{(2)}$ bistable; $A_{rr'}^{(\mu>2)}$ sampled i.i.d. $\mathcal{N}(0, 0.5)$ with rejection of non–positive-semidefinite draws
Fig. 2G	4000	$\alpha = 0.5$	$D^{(1)} = 2$ or 9; $D^{(\mu>1)} = 2$	$\mathbf{A}^{(1)}$ limit cycle; $\mathbf{A}^{(2)}$ bistable; $A_{rr'}^{(\mu>2)}$ sampled i.i.d. $\mathcal{N}(0, 0.6)$ with rejection of non–positive-semidefinite draws
Fig. 3C,D,E	5000	$\alpha = 0.5$	$D^{(\mu)} = 2$ by default; $D^{(1)}$ modulated from $2 \to 3.75 \to 8$; $D^{(2)}$ from $2 \to 4.2 \to 8$	$\mathbf{A}^{(1)}$ limit cycle; $\mathbf{A}^{(2)}$ bistable; $A_{rr'}^{(\mu>2)}$ sampled i.i.d. $\mathcal{N}(0, 0.6)$ with rejection of non–positive-semidefinite draws
Fig. 4A	4000	varied	$D^{(\nu\neq\mu)} = 2$, and $D^{(\mu)}$ varied, where μ corresponds to limit cycle of frequency $\omega = 0.25$.	$\mathbf{A}^{(\nu\neq\mu)} = \gamma \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ with $\theta \sim \mathrm{Unif}(0, 2\pi)$ (see Appendix)
Fig. 4B	4000	$\alpha = 1$	$D^{(\mu)}$ varied on y-axis, and μ varied on x-axis by as the intrinsic frequency $\omega = \tan\theta$ for $\mathbf{A}^{(\mu)}$.	same as above
Fig. 5A,B,C	8000	$\alpha = 1$	same $D^{(\mu)}$ as Fig. 4 modulated from $2 \to 4.15 \to 5.3$ (black boxes),	same as Fig. 4
Fig. 5D	varied	$\alpha = 1$	same $D^{(\mu)}$ modulated	same as Fig. 4
Fig. 5E	varied	$\alpha = 1$	same as Fig. 4	
Fig. S2	2000	$P = 2$	$D^{(1)} = 2$, $D^{(2)} = 2.2$	same as Fig. 2C
Fig. S3A	2000	$P = 2$	$D^{(1)} = 3.5$, $D^{(2)} = 4$ (or swapped)	same as Fig. 2C
Fig. S3B	2000	$P = 2$	$D^{(1)} = 3.3$, $D^{(2)} = 3$ (or swapped)	same as Fig. 2C
Fig. S4	5000	$P = 2$	$D^{(\mu)} = 2$ or $2 \to 2.75 \to 3.5 \to 4.25$	same as Fig. 3
Fig. S5A-D	8000	$P = 2$	$D^{(\mu)} = 2$	same as Fig. 3
Fig. S6	10000	$P = 2$	$D^{(\nu\neq\mu)} = 2 \to 4 \to 8$	same as Fig. 4

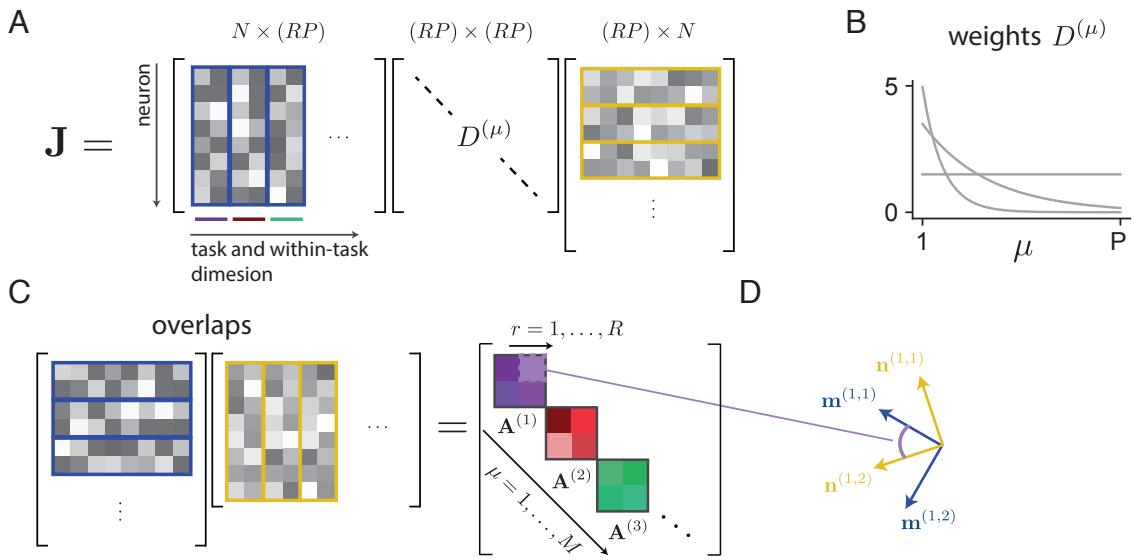

Figure S1: More detailed illustrations of connectivity structure. **(A)** Alternative interpretation of the sum in Fig. 1C as one large matrix product, with left loadings organized in one matrix as columns (blue) and right loadings in another as rows (yellow), with scalings as diagonal elements. The blue or yellow rectangular boundaries group the loading vectors into tasks, so $r = 1, 2$ indexes vectors within a block, while $\mu = 1, \ldots, P$ indexes rectangular blocks. **(B)** Example connectivity strength spectra $D^{(\mu)}$, which can generally be dispersed or concentrated among tasks. **(C)** Illustration of the overall overlap matrix for all loading vectors, grouped by task by blue or yellow rectangular boundaries, as the matrix product of the left loading vectors as rows of one matrix (blue) with the right loading vectors as columns of one matrix (yellow). **(D)** Geometric interpretation of an entry of the overlap matrix as the degree of alignment of the associated left and right loading vectors. The $\mathbf{m}$'s (and $\mathbf{n}$'s) are shown to be orthogonal among themselves.

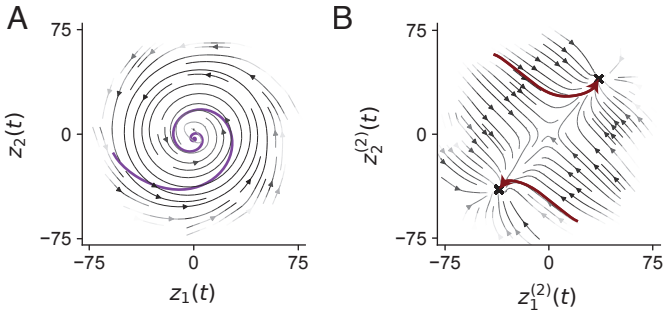

Figure S2: Dynamics for the same network as Fig. 2C but with $D^{(1)} = 2$, $D^{(2)} = 2.2$. **(A)** Purple: activity in the task-1 subspace for a single run of the dynamics. Black: flow fields over many runs. **(B)** Red: activity in the task-2 subspace over two runs with different initial conditions. Black: average flow fields over many runs.

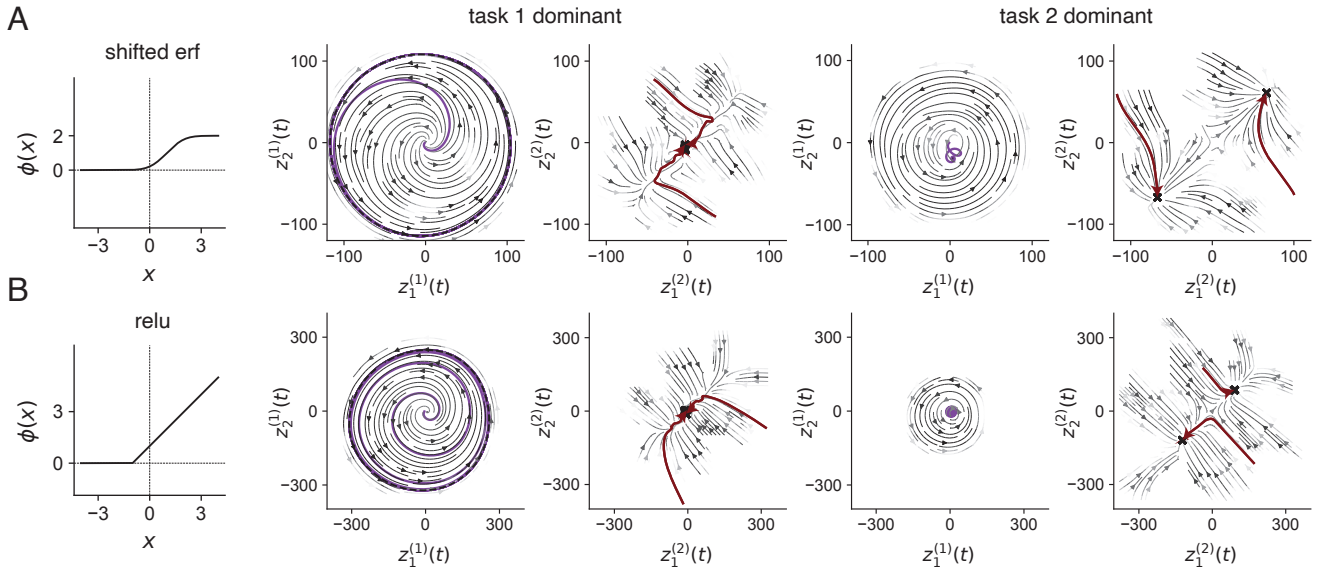

Figure S3: Similar simulations as in Fig. 2, but for networks with different nonlinearities. **(A)** A network using a diagonally shifted error function, by 1 in each direction, such that outputs are always positive, and it requires positive current to activate a neuron. **(B)** A network using a non-saturating nonlinearity, in particular a relu function (shifted leftward by 1). To ensure global stability, an additional term $-g_I/N$ is added to each J_{ij}, with $g_I = 2$.

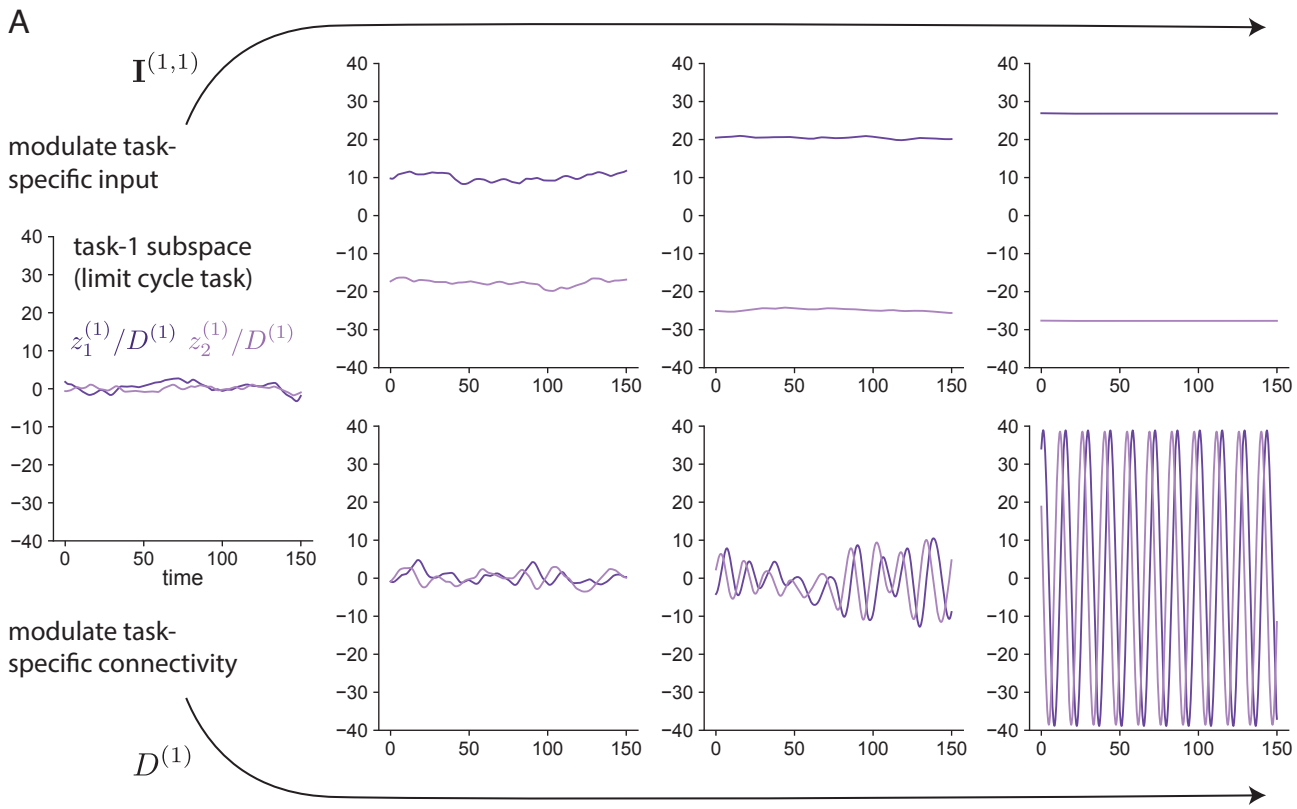

Figure S4: Comparing modulation of task-specific input drive and task-specific connectivity, for task $\mu = 1$ in the same network shown in Fig. 3. **(A)** Left: spontaneous-state activity in the task-1 subspace, featuring weak, noisy oscillations. Top: modulating the additive input drive $I_i^{(1,1)} = I_0\sqrt{N}m_i^{(1,1)}$ to each neuron i in the direction of the first task-1 left loading vector. Values of I_0 are (left to right) $0, 0.5, 1, 1.5$. Bottom: modulating the strength of the task-1 connectivity component. Values of $D^{(1)}$ are (left to right) $2, 2.75, 3.5, 4.25$.

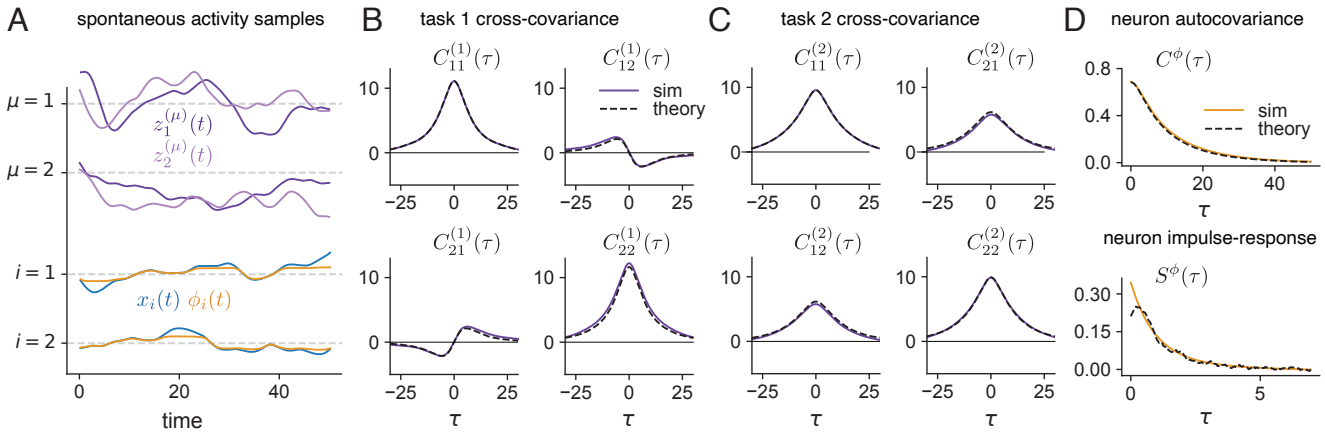

Figure S5: Dynamics in the spontaneous state. **(A)** Example activity traces for a network with $\alpha = 0.5$; $D^{(\mu)} = 2$ for all μ; $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$ as in Fig. 2C, with $A^{(\mu)}_{rr'} \sim \mathcal{N}(0, 0.6)$ i.i.d. with rejection sampling for positive-definiteness violations; and $N = 8000$. Top: latent activity for a limit-cycle task ($\mu = 1$) and a fixed-point task ($\mu = 2$) in the spontaneous state. Bottom: current and firing-rate traces for two example neurons in the network. **(B)** Lagged cross-covariance functions for $\vec{z}^{(1)}(t)$, i.e. $\left\langle z^{(1)}_r(t) z^{(1)}_{r'}(t-\tau)\right\rangle_t$. Top left: $r = 1$, $r' = 1$. Top right: $r = 1, r' = 2$, etc. Solid: simulation. Dashed: theory. **(D)** Single-neuron autocovariance function $C^{\phi}(\tau) = \left\langle \phi_i(t)\phi_i(t+\tau)\right\rangle_t$ (top) and impulse-response function $S^{\phi}(\tau) = \langle \phi' \rangle\, F(\tau)$ (bottom) and for the same network. Solid: simulation. Dashed: theory.

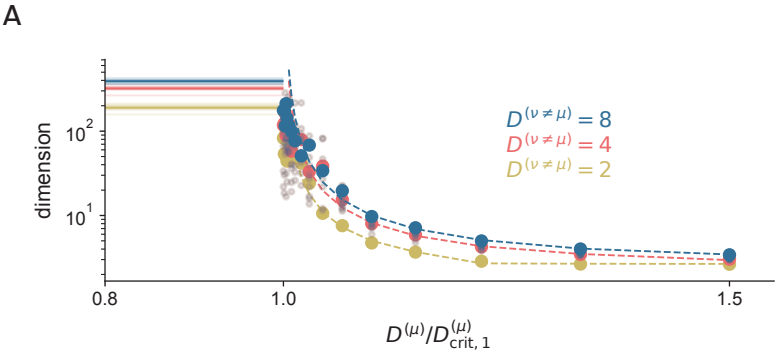

Figure S6: **(A)** Participation ratio of firing rates for networks with variable $D^{(\nu)}$ values for all non-selected tasks $\nu \neq \mu$, as $D^{(\mu)}$ is varied relative to its critical point for task selection. Light dots: individual network simulations. Heavy dots: averages over networks. Dashed line: theory.

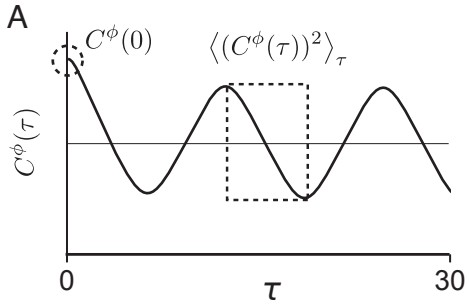

Figure S7: Cartoon showing how to read off the dimensionality of the task-selected state from the single-unit autocovariance function. **(A)** Black: Theoretically calculated autocovariance function $C^{\phi}(\tau)$ from Fig. 3F (middle). The value at $\tau = 0$ at the average squared value over one period after full amplitude decay are the relevant quantities for computing the PR. The latter captures asymptotic behavior of the autocovariance function away from $\tau = 0$, which is needed for Eq. (44).

Appendix: Derivations of key mathematical results

Path integral for spontaneous state

This section takes advantage of mathematical techniques which are outlined in e.g. [68] or [69]. Similar derivations can be found in [45]. In this section, we derive the DMFT for the spontaneous state, assuming no dominant patterns. We start by writing the Martin-Siggia-Rose-Janssen-de Dominicis path integral for the partition function $Z(\mathbf{J})$ of the network dynamics:

$$Z(\mathbf{J}) \propto \int \mathcal{D}\mathbf{x}\,\mathcal{D}\hat{\mathbf{x}}\,\exp\left[i\sum_i \int dt\,\hat{x}_i(t)\Big(T[x_i](t) - \sum_j J_{ij}\phi_j(t)\Big)\right], \tag{49}$$

currently keeping the weights as shorthand J_{ij} and using $T = 1 + \partial_t$ as notational convenience. We now compute the average

$$\bar{Z} = \int \prod_{\mu,r,i} dm_i^{(\mu,r)} \prod_{\mu,r,i} dn_i^{(\mu,r)}\,\Pr\Big(\{m_i^{(\mu,r)}\}_{r,\mu,i}, \{n_i^{(\mu,r)}\}_{r,\mu,i}\Big)\,Z(\mathbf{J}).$$

The probability density $\Pr(\cdots)$ for a given set of loadings[4] for each neuron i and each task μ,

$$\vec{l}_i^{(\mu)} = \Big(m_i^{(\mu,1)}, \ldots, m_i^{(\mu,R)}, n_i^{(\mu,1)}, \cdots n_i^{(\mu,R)}\Big) \in \mathbb{R}^{2R},$$

is

$$\Pr\Big(\{m_i^{(\mu,r)}\}_{r,\mu,i}, \{n_i^{(\mu,r)}\}_{r,\mu,i}\Big) \propto \exp\left[-\frac{1}{2}\sum_i\sum_\mu \vec{l}_i^{(\mu)T}\mathcal{P}^{(\mu)}\vec{l}_i^{(\mu)}\right], \tag{50}$$

since each neuron's loadings are independent of the others', and loadings for different tasks are also independent. Within a task for a given neuron, the loadings are distributed as a multivariate Gaussian with precision matrix

$$\mathcal{P}^{(\mu)} = \mathrm{inv}\Big(\mathbf{\Sigma}^{(\mu)}\Big).$$

We define a set of $R \times P$ latent variables

$$z_r^{(\mu)}(t) = D^{(\mu)}\sum_i n_i^{(\mu,r)}\,\phi_i(t),$$

and their definition-enforcing conjugates $\hat{z}_r^{(\mu)}(t)$. Then our disorder-averaged partition function is

$$\bar{Z} \propto \int \mathcal{D}\mathbf{x}\,\mathcal{D}\hat{\mathbf{x}}\,\mathcal{D}\mathbf{z}\,\mathcal{D}\hat{\mathbf{z}}\,d\vec{\mathbf{l}}\,\exp\Bigg[-\frac{1}{2}\sum_{i\mu}\vec{l}_i^{(\mu)T}\mathcal{P}^{(\mu)}\vec{l}_i^{(\mu)} + i\sum_{\mu r}\hat{z}_r^{(\mu)}\,z_r^{(\mu)} + i\sum_i \hat{x}_i T[x_i]$$
$$- i\sum_{i\mu r} m_i^{(\mu,r)}\,\hat{x}_i\,z_r^{(\mu)} - i\sum_{i\mu r} D^{(\mu)}\,n_i^{(\mu,r)}\,\phi_i\,\hat{z}_r^{(\mu)}\Bigg]. \tag{51}$$

[4] In general, vector-arrow notation imply a fixed-dimension vector like $\vec{l}_i^{(\mu)} \in \mathbb{R}^{2R}$, while bolded notation implies a vector with extensive (i.e. N-scaling) dimension, e.g. $\mathbf{m}^{(\mu,r)} \in \mathbb{R}^N$.

where suppressed time indices imply time integrals over a shared dummy time index. We can rewrite this as a quadratic in terms of $\vec{l}_i^{(\mu)}$ by defining the vector

$$\vec{b}_{i\mu} = -i\Big(\hat{x}_i\, z_1^{(\mu)}, \ldots, \hat{x}_i\, z_R^{(\mu)},\, D^{(\mu)}\, \phi_i\, \hat{z}_1^{(\mu)}, \ldots, D^{(\mu)}\, \phi_i\, \hat{z}_R^{(\mu)}\Big) \in \mathbb{R}^{2R},$$

allowing us to perform a matrix version of completing the square for each $i\mu$:

$$\bar{Z} \propto \int \mathcal{D}\mathbf{x}\, \mathcal{D}\hat{\mathbf{x}}\, \mathcal{D}\mathbf{z}\, \mathcal{D}\hat{\mathbf{z}}\, d\vec{\mathbf{l}} \exp\left[\sum_{i\mu}\Big(-\frac{1}{2}\vec{l}_i^{(\mu)T}\mathcal{P}^{(\mu)}\vec{l}_i^{(\mu)} + \vec{l}_i^{(\mu)T}\vec{b}_{i\mu}\Big) + i\sum_{\mu r}\hat{z}_r^{(\mu)}\, z_r^{(\mu)} + i\sum_i \hat{x}_i T[x_i]\right] \tag{52}$$

$$= \int \mathcal{D}\mathbf{x}\, \mathcal{D}\hat{\mathbf{x}}\, \mathcal{D}\mathbf{z}\, \mathcal{D}\hat{\mathbf{z}}\, d\vec{\mathbf{l}} \exp\left[\sum_{i\mu}\Big(-\frac{1}{2}\Big(\vec{l}_i^{(\mu)} - \mathbf{\Sigma}^{(\mu)}\Big)^T\mathcal{P}^{(\mu)}\Big(\vec{l}_i^{(\mu)} - \mathbf{\Sigma}^{(\mu)}\Big) + \frac{1}{2}\vec{b}_{i\mu}^T\mathbf{\Sigma}^{(\mu)}\vec{b}_{i\mu}\Big)\right. \tag{53}$$

$$\left. + i\sum_{\mu r}\hat{z}_r^{(\mu)}\, z_r^{(\mu)} + i\sum_i \hat{x}_i T[x_i]\right] \tag{54}$$

$$\propto \int \mathcal{D}\mathbf{x}\, \mathcal{D}\hat{\mathbf{x}}\, \mathcal{D}\mathbf{z}\, \mathcal{D}\hat{\mathbf{z}} \exp\left[i\sum_{\mu r}\hat{z}_r^{(\mu)}\, z_r^{(\mu)} + i\sum_i \hat{x}_i T[x_i] + \frac{1}{2}\sum_{i\mu}\vec{b}_{i\mu}^T\mathbf{\Sigma}^{(\mu)}\vec{b}_{i\mu}\right]. \tag{55}$$

Recall that $\mathbf{\Sigma}^{(\mu)}$ has a block diagonal structure with the identity along each of the two $R \times R$ principal blocks. Then we can rewrite this quadratic term as three terms, reintroducing the time indices briefly:

$$\vec{b}_{i\mu}^T\mathbf{\Sigma}^{(\mu)}\vec{b}_{i\mu} = -\frac{1}{N}\sum_r \int\int dt\, dt' \Big[\hat{x}_i(t)\hat{x}_i(t')\, z_r^{(\mu)}(t) z_r^{(\mu)}(t') + \big(D^{(\mu)}\big)^2\, \phi_i(t)\phi_i(t')\, \hat{z}_r^{(\mu)}(t)\hat{z}_r^{(\mu)}(t')\Big] \tag{56}$$

$$- \frac{2}{N}\sum_{rr'} D^{(\mu)} A_{rr'}^{(\mu)} \int\int dt\, dt'\, z_r^{(\mu)}(t)\hat{z}_{r'}^{(\mu)}(t')\, \hat{x}_i(t)\, \phi_i(t'). \tag{57}$$

The first two terms are easy to deal with and imply a reciprocal relationship between the two-point function of the latent variables $z_r^{(\mu)}(t)$ and how it appears in the noise driving $x_i(t)$, and the two-point function of the units $\phi_i(t)$ and how it appears in the noise driving $z_r^{(\mu)}(t)$. In the random-mode-model without symmetry [45], these are the only terms, and they lead to a straightforward solution. The third term, which arises due to the latent weight symmetries encoded in the off-diagonal block $A_{rr'}^{(\mu)}$ of $\mathbf{\Sigma}^{(\mu)}$, is more difficult to deal with. We have to split the expression into the $t > t'$ and $t < t'$ half-planes. In the former, we can write $z_r^{(\mu)}(t)\hat{z}_{r'}^{(\mu)}(t') = i\delta z_r^{(\mu)}(t)/\delta I_{r'}^{(\mu)}(t')$, and in the latter we can write $\phi_i(t')\hat{x}_i(t) = i\delta\phi_i(t')/\delta I_i(t)$, where $\delta/\delta I_i(t)$ represents the functional derivative with respect to external current. (In the flipped cases, each is zero by causality.)

Then we define a set of fields and their conjugates:

$$C^\phi(t,t') = \frac{1}{N}\sum_i \phi_i(t)\phi_i(t') \qquad \text{conjugate: } \hat{Q}^z(t,t'), \tag{58}$$

$$Q^z_r(t,t') = \frac{1}{N}\sum_\mu z^{(\mu)}_r(t) z^{(\mu)}_r(t') \qquad \text{conjugate: } \hat{C}^\phi_r(t,t'), \tag{59}$$

$$S^\phi(t,t') = \frac{1}{N}\sum_i \frac{\delta\phi_i(t)}{\delta I_i(t')} \qquad \text{conjugate: } \hat{G}^z(t,t'), \tag{60}$$

$$G^z_{rr'}(t,t') = \frac{1}{N}\sum_\mu D^{(\mu)}\, A^{(\mu)T}_{rr'} \frac{\delta z^{(\mu)}_r(t)}{\delta I^{(\mu)}_{r'}(t')} \qquad \text{conjugate: } \hat{S}^\phi_{rr'}(t,t'). \tag{61}$$

Plugging in these definitions, we have

$$\bar{Z} \propto \int \mathcal{D}\mathbf{x}\,\mathcal{D}\hat{\mathbf{x}}\, \exp\left[i\sum_i \hat{x}_i T[x_i] - \frac{1}{2}\sum_i\sum_r \hat{x}_i\, Q^z_r\, \hat{x}_i - \frac{i}{2}\sum_i\sum_{rr'} \hat{x}_i\, G^z_{rr'}\, \phi_i \right] \tag{62}$$

$$\times \int \mathcal{D}\mathbf{z}\,\mathcal{D}\hat{\mathbf{z}}\, \exp\left[i\sum_{\mu r} \hat{z}^{(\mu)}_r\, z^{(\mu)}_r - \frac{1}{2}\sum_\mu (D^{(\mu)})^2 \sum_r \hat{z}^{(\mu)}_r\, C^\phi\, \hat{z}^{(\mu)}_r - \frac{i}{2}\sum_{\mu rr'} D^{(\mu)} A^{(\mu)T}_{rr'}\, \hat{z}^{(\mu)}_{r'} S^\phi z^{(\mu)}_r \right]. \tag{63}$$

Note how the interaction term has a half in the neuron path integral and a half in the latent path integral, with kernels defined as population-averaged susceptibilities in the other picture. In other words, a given neuron interacts with itself by a weighted average of latent-latent impulse responses, and a given set of same-task latent variables interact among themselves by an average of single-neuron impulse responses.

Then, integrating over the definitions of these fields and decoupling noninteracting single-neuron and single-task path integrals, we obtain definition-enforcing terms and a global N scaling outside an action that is purely a function of the fields:

$$\bar{Z} \propto \int \mathcal{D}C^\phi\, \mathcal{D}\{\hat{C}^\phi_r\}\, \mathcal{D}S^\phi\, \mathcal{D}\{\hat{S}^\phi_{rr'}\}\, \mathcal{D}\{Q^z_r\}\, \mathcal{D}\hat{Q}^z\, \mathcal{D}\{G^z_{rr'}\}\, \mathcal{D}\hat{G}^z \tag{64}$$

$$\times \exp\Big(-NS\Big[C^\phi, \{\hat{C}^\phi_r\}, S^\phi, \{\hat{S}^\phi_{rr'}\}, \{Q^z_r\}, \hat{Q}^z, \{G^z_{rr'}\}, \hat{G}^z\Big]\Big) \tag{65}$$

with

$$S[C^\phi, \{\hat{C}^\phi_r\}, S^\phi, \{\hat{S}^\phi_{rr'}\}, \{Q^z_r\}, \hat{Q}^z, \{G^z_{rr'}\}, \hat{G}^z] = -\frac{1}{2}C^\phi\,\hat{Q}^z - \frac{1}{2}\sum_r Q^z_r\,\hat{C}^\phi_r + \frac{1}{2}S^\phi\,\hat{G}^z + \frac{1}{2}\sum_{rr'} G^z_{rr'}\,\hat{S}^\phi_{rr'} \tag{66}$$

$$- \log W^x[\{Q^z_r\}, \hat{Q}^z, \{G^z_{rr'}\}, \hat{G}^z] \tag{67}$$

$$- \alpha\left\langle \log W^{\vec{z}}_{D,\mathbf{A}}\Big[C^\phi, \{\hat{C}^\phi_r\}, S^\phi, \{\hat{S}^\phi_{rr'}\}\Big]\right\rangle_{D,\mathbf{A}}. \tag{68}$$

Here,

$$W^x = \int \mathcal{D}x\, \mathcal{D}\hat{x}\, \exp\left[i\hat{x}T[x] - \frac{1}{2}\sum_r \hat{x}\, Q^z_r\, \hat{x} - \frac{1}{2}\phi\, \hat{Q}^z\, \phi - \frac{i}{2}\sum_{rr'} \hat{x}\, G^z_{rr'}\, \phi - \frac{i}{2}\phi\, \hat{G}^z\, \hat{x} \right], \tag{69}$$

$$W^{\vec{z}}_{D,\mathbf{A}} = \int \mathcal{D}\vec{z}\, \mathcal{D}\hat{z}\, \exp\left[i\sum_r \hat{z}_r\, z_r - \frac{1}{2}\sum_r z_r\, \hat{C}^\phi_r\, z_r - \frac{1}{2}D^2 \sum_r \hat{z}_r\, C^\phi\, \hat{z}_r \right. \tag{70}$$

$$\left. - \frac{i}{2}D\sum_{rr'} A^T_{rr'}\, \hat{z}_{r'}\, S^\phi\, z_r - \frac{i}{2}D\sum_{rr'} A^T_{rr'}\, z_r\, \hat{S}^\phi_{rr'}\, \hat{z}_{r'} \right]. \tag{71}$$

We differentiate with respect to each of the $R^2 + R + 2$ fields and their conjugates:

$$\frac{\delta S}{\delta C^\phi(t,t')} = -\hat{Q}^z(t,t') + \alpha \left\langle D^2 \sum_r \left\langle \hat{z}_r(t)\hat{z}_r(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{72}$$

$$\frac{\delta S}{\delta \hat{C}^\phi_r(t,t')} = -Q^z_r(t,t') + \alpha \left\langle \left\langle z_r(t) z_r(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{73}$$

$$\frac{\delta S}{\delta S^\phi(t,t')} = \hat{G}^z(t,t') + i\alpha \left\langle D \sum_{rr'} A^T_{rr'} \left\langle \hat{z}_{r'}(t) z_r(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{74}$$

$$\frac{\delta S}{\delta \hat{S}^\phi_{rr'}(t,t')} = G^z_{rr'}(t,t') + i\alpha \left\langle D\, A^T_{rr'} \left\langle z_r(t) \hat{z}_{r'}(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{75}$$

$$\frac{\delta S}{\delta Q^z_r(t,t')} = -\hat{C}^\phi_r(t,t') + \left\langle \hat{x}(t)\hat{x}(t') \right\rangle_{W^x}, \tag{76}$$

$$\frac{\delta S}{\delta \hat{Q}^z(t,t')} = -C^\phi(t,t') + \left\langle \phi(t)\phi(t') \right\rangle_{W^x}, \tag{77}$$

$$\frac{\delta S}{\delta G^z_{rr'}(t,t')} = \hat{S}^\phi_{rr'}(t,t') + i\left\langle \hat{x}(t)\phi(t') \right\rangle_{W^x}, \tag{78}$$

$$\frac{\delta S}{\delta \hat{G}^z(t,t')} = S^\phi(t,t') + i\left\langle \phi(t)\hat{x}(t') \right\rangle_{W^x}. \tag{79}$$

Solving the saddle-point system yields

$$\hat{Q}^z_\star(t,t') = \hat{C}^\phi_{\star,r}(t,t') = 0, \tag{80}$$

$$S^\phi_{\star,rr'}(t,t') = S^\phi_\star(t',t) = \left\langle \frac{\delta\phi(t')}{\delta I(t)} \right\rangle_\star, \tag{81}$$

$$C^\phi_\star(t,t') = \left\langle \phi(t)\phi(t') \right\rangle_\star, \tag{82}$$

$$Q^z_{\star,r}(t,t') = \alpha \left\langle \left\langle z_r(t) z_r(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{83}$$

$$G^z_{\star,rr'}(t,t') = \alpha \left\langle D\, A^T_{rr'} \left\langle \frac{\delta z_r(t)}{\delta I_{r'}(t')} \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{84}$$

$$\hat{G}^z_\star(t,t') = \alpha \left\langle D \sum_{rr'} A^T_{rr'} \left\langle \frac{\delta z_r(t')}{\delta I_{r'}(t)} \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{85}$$

so that

$$\hat{G}^z_\star(t,t') = \sum_{rr'} G^z_{\star,rr'}(t',t). \tag{86}$$

Then, going back into the single-site and latent processes, we obtain a ϕ process driven by noise term $\eta^x \sim \mathcal{GP}\Big(0, Q^z_\star\Big)$ where $Q^z_\star = \sum_r Q^z_{\star,r}$, plus self-interactions via self-coupling kernel $\hat{G}^z_\star$. However, the process for the latent variables is more complex and involves R^2 self-couplings. Altogether, we have

$$T[x](t) = \eta^x(t) + \int_{-\infty}^t dt' \hat{G}^z_\star(t,t')\phi(x(t')) \qquad \eta^x \sim \mathcal{GP}(0, Q^z_\star(t,t')) \tag{87}$$

$$z_r(t) = \eta^{z_r}(t) + D \sum_{r'=1}^R A_{rr'} \int_{-\infty}^t dt' S^\phi_\star(t,t') z_{r'}(t') \qquad \eta^{z_r} \sim \mathcal{GP}(0, D^2 C^\phi_\star(t,t')). \tag{88}$$

From now on, we drop the $\star$ notation and assume any kernel as written is the saddle-point solution. Further, we assume that all kernels depend only on $t-t'$, allowing us to leverage the convolution theorem. In Fourier space, we solve for $z_r(\omega)$ and get

$$z_r(\omega) = \eta^{z_r}(\omega) + D \sum_{r'} A_{rr'}\, S^\phi(\omega)\, z_{r'}(\omega), \tag{89}$$

$$\implies \vec{z}(\omega) = \Big(\mathbf{I} - D S^\phi(\omega)\mathbf{A}\Big)^{-1} \vec{\eta}^z(\omega) \equiv \mathbf{M}(\omega)\, \vec{\eta}^z(\omega), \tag{90}$$

$$\implies Q^z_r(\omega) = \alpha \left\langle \langle z_r(\omega) z_r(\omega)^* \rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}} = \alpha \left\langle \left\langle \sum_{r'} M_{rr'}(\omega)\, \eta^{z_{r'}}(\omega) \sum_{r''} M_{rr''}(\omega)^*\, \eta^{z_{r''}}(\omega)^* \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}} \tag{91}$$

$$= \alpha \left\langle D^2 \Big[\mathbf{M}(\omega)\mathbf{M}^\dagger(\omega)\Big]_{rr} \right\rangle_{D,\mathbf{A}} C^\phi(\omega). \tag{92}$$

Then using Furutsu–Novikov we get an expression for $\hat{G}^z(\omega)$. We summarize together with $Q^z(\omega)$ as

$$Q^z(\omega) = \alpha \left\langle D^2 \, \|\mathbf{M}(\omega)\|_{\mathcal{F}}^2 \right\rangle_{D,\mathbf{A}} C^\phi(\omega), \tag{93}$$

$$\hat{G}^z(\omega) = \alpha \left\langle D \, \mathrm{Tr}\Big(\mathbf{A}\mathbf{M}(\omega)\Big) \right\rangle_{D,\mathbf{A}}. \tag{94}$$

Path integral for task-selected state

We pick up at Eq. (55) and explicitly separate the μ index into a set of P^* dominant and $P - P^* \approx P$ weak tasks. We perform the disorder integration only over the weak tasks' couplings and keep the average partition function depending on the particular realization of the dominant couplings:

$$\bar{Z}[\{m_i^{(\mu^*,r)}, n_i^{(\mu^*,r)}\}] \propto \int \mathcal{D}\mathbf{x}\, \mathcal{D}\hat{\mathbf{x}}\, \mathcal{D}\mathbf{z}\, \mathcal{D}\hat{\mathbf{z}} \exp \Bigg[i \sum_i \hat{x}_i T[x_i] - i\sqrt{N} \sum_{i\mu^* r} m_i^{(\mu^*,r)} \, \hat{x}_i \, z_r^{(\mu^*)} \tag{95}$$

$$+ i \sum_{\mu r} \hat{z}_r^{(\mu)} z_r^{(\mu)} - i\sqrt{N} \sum_{i\mu^* r} D^{(\mu^*)} \, n_i^{(\mu^*,r)} \, \phi_i \, \hat{z}_r^{(\mu^*)} \tag{96}$$

$$- \frac{1}{2N} \sum_{\mu i r} \hat{x}_i z_r^{(\mu)} \hat{x}_i z_r^{(\mu)} \tag{97}$$

$$- \frac{1}{2N} \sum_{\mu i r} \left(D^{(\mu)}\right)^2 \phi_i \hat{z}_r^{(\mu)} \phi_i \hat{z}_r^{(\mu)} \tag{98}$$

$$- \frac{1}{N} \sum_{\mu i r r'} D^{(\mu)} A_{rr'}^{(\mu)T} z_r^{(\mu)} \hat{x}_i \hat{z}_{r'}^{(\mu)} \phi_i \Bigg] \tag{99}$$

We now define an expanded set of order parameters

$$C^\phi(t,t') = \frac{1}{N} \sum_i \phi_i(t)\phi_i(t') \qquad \text{conjugate: } \hat{Q}^z(t,t'), \tag{100}$$

$$Q_r^z(t,t') = \frac{1}{N} \sum_\mu z_r^{(\mu)}(t) z_r^{(\mu)}(t') \qquad \text{conjugate: } \hat{C}_r^\phi(t,t'), \tag{101}$$

$$S^\phi(t,t') = \frac{1}{N} \sum_i \frac{\delta \phi_i(t)}{\delta I_i(t')} \qquad \text{conjugate: } \hat{G}^z(t,t'), \tag{102}$$

$$G_{rr'}^z(t,t') = \frac{1}{N} \sum_\mu D^{(\mu)} \, A_{rr'}^{(\mu)T} \frac{\delta z_r^{(\mu)}(t)}{\delta I_{r'}^{(\mu^*)}(t')} \qquad \text{conjugate: } \hat{S}_{rr'}^\phi(t,t') \tag{103}$$

$$z_r^{(\mu^*)}(t) = \frac{1}{\sqrt{N}} D^{(\mu^*)} \sum_i n_i^{(\mu^*,r)} \phi_i(t) \qquad \text{conjugate: } \hat{z}^{(\mu^*)}(t) \tag{104}$$

For the dominant tasks, their activity is no longer a dynamical or stochastic variable but rather a trajectory that is subject to the saddle-point condition of the overall dynamics. Note that we have rescaled the usual definition of the task latents by a factor of $1/\sqrt{N}$ so that all order parameters are $\mathcal{O}(1)$. We insert

definitions of order parameters when possible and separate into integrals over $\mathbf{x}$'s and $\mathbf{z}$'s.

$$\bar{Z}^x[\{m_i^{(\mu^*,r)}, n_i^{(\mu^*,r)}\}] \propto \int \mathcal{D}\mathbf{x}\,\mathcal{D}\hat{\mathbf{x}} \exp\Bigg[i\sum_i \hat{x}_i T[x_i] - i\sqrt{N}\sum_{i\mu^* r} m_i^{(\mu^*,r)} \hat{x}_i z_r^{(\mu^*)} - i\sqrt{N}\sum_{i\mu^* r} D^{(\mu)} n_i^{(\mu^*,r)} \phi_i \hat{z}_r^{(\mu^*)} \tag{105}$$

$$- \frac{1}{2}\sum_i \sum_r \hat{x}_i Q_r^z \hat{x}_i - i\sum_i \sum_{rr'} \hat{x}_i G_{rr'}^z \phi_i \Bigg] \tag{106}$$

$$\tag{107}$$

$$\bar{Z}^z \propto \int \mathcal{D}\mathbf{z}\,\mathcal{D}\hat{\mathbf{z}} \exp\Bigg[i\sum_{\mu r} \hat{z}_r^{(\mu)} z_r^{(\mu)} - \frac{1}{2}\sum_\mu \left(D^{(\mu)}\right)^2 \sum_r \hat{z}_r^{(\mu)} C^\phi \hat{z}_r^{(\mu)} - i\sum_\mu D^{(\mu)} \sum_{rr'} \hat{z}_{r'}^{(\mu)} S^\phi z_r^{(\mu)} \Bigg] \tag{108}$$

$$\tag{109}$$

Then we can construct one path integral over the fields as a function of the realizations of the normalized dominant couplings $c = \{\overline{m}_i^{(\mu^*,r)}, \overline{n}_i^{(\mu^*,r)}\}$:

$$\bar{Z}[\{m_i^{(\mu^*,r)}, n_i^{(\mu^*,r)}\}] \propto \int \mathcal{D}C^\phi\,\mathcal{D}\{\hat{C}_r^\phi\}\,\mathcal{D}S^\phi\,\mathcal{D}\{\hat{S}_{rr'}^\phi\}\,\mathcal{D}\{Q_r^z\}\,\mathcal{D}\hat{Q}^z\,\mathcal{D}\{G_{rr'}^z\}\,\mathcal{D}\hat{G}^z \mathcal{D}\{z_r^{(\mu^*)}\}\mathcal{D}\{\hat{z}_r^{(\mu^*)}\} \tag{110}$$

$$\times \exp\Big(-NS\Big[C^\phi, \{\hat{C}_r^\phi\}, S^\phi, \{\hat{S}_{rr'}^\phi\}, \{Q_r^z\}, \hat{Q}^z, \{G_{rr'}^z\}, \hat{G}^z, \{z_r^{(\mu^*)}\}, \{\hat{z}_r^{(\mu^*)}\}\Big]\Big) \tag{111}$$

with

$$S[\text{all fields}] = -\frac{1}{2}C^\phi\,\hat{Q}^z - \frac{1}{2}\sum_r Q_r^z\,\hat{C}_r^\phi \tag{112}$$

$$+\frac{1}{2}S^\phi\,\hat{G}^z + \frac{1}{2}\sum_{rr'} G_{rr'}^z\,\hat{S}_{rr'}^\phi + i\sum_{\mu^* r} \hat{z}_r^{(\mu^*)} z_r^{(\mu^*)} \tag{113}$$

$$-\Big\langle \log W_c^x[\{Q_r^z\}, \hat{Q}^z, \{G_{rr'}^z\}, \hat{G}^z, \{z_r^{(\mu^*)}\}, \{\hat{z}_r^{(\mu^*)}\}]\Big\rangle_c \tag{114}$$

$$-\alpha\Bigg\langle \log W_{D,\mathbf{A}}^{\vec{z}}\Big[C^\phi, \{\hat{C}_r^\phi\}, S^\phi, \{\hat{S}_{rr'}^\phi\}\Big]\Bigg\rangle_{D,\mathbf{A}}. \tag{115}$$

Here,

$$W_c^x = \int \mathcal{D}x\,\mathcal{D}\hat{x}\,\exp\Bigg[i\hat{x}T[x] - i\hat{x}\sum_{\mu^* r} \bar{m}^{(\mu^*,r)} z_r^{(\mu^*)} - i\phi\sum_{\mu^* r} D^{(\mu^*)} \bar{n}^{(\mu^*,r)} \hat{z}_r^{(\mu^*)} \tag{116}$$

$$-\frac{1}{2}\sum_r \hat{x}\,Q_r^z\,\hat{x} - \frac{1}{2}\phi\,\hat{Q}^z\,\phi - \frac{i}{2}\sum_{rr'} \hat{x}\,G_{rr'}^z\,\phi - \frac{i}{2}\phi\,\hat{G}^z\,\hat{x}\Bigg] \tag{117}$$

$$W^{\vec{z}}_{D,\mathbf{A}} = \int \mathcal{D}\vec{z}\mathcal{D}\hat{z}\, \exp\Bigg[i\sum_r \hat{z}_r\, z_r - \frac{1}{2}\sum_r z_r\, \hat{C}^{\phi}_r\, z_r - \frac{1}{2}D^2 \sum_r \hat{z}_r\, C^{\phi}\, \hat{z}_r \tag{118}$$

$$- \frac{i}{2}D\sum_{rr'} A^T_{rr'}\, \hat{z}_{r'}\, S^{\phi}\, z_r - \frac{i}{2}D\sum_{rr'} A^T_{rr'}\, z_r\, \hat{S}^{\phi}_{rr'}\, \hat{z}_{r'} \Bigg]. \tag{119}$$

where we have defined $\bar{m}^{(\mu^*,r)} = \sqrt{N} m^{(\mu^*,r)}$ (and similarly $\bar{n}^{(\mu^*,r)}$) as the $\mathcal{O}(1)$ normalized couplings of an arbitrary unit to the dominant patterns. Then differentiating with respect to each order parameter, we get

$$\frac{\delta S}{\delta C^{\phi}(t,t')} = -\hat{Q}^z(t,t') + \alpha \left\langle D^2 \sum_r \left\langle \hat{z}_r(t)\hat{z}_r(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{120}$$

$$\frac{\delta S}{\delta \hat{C}^{\phi}_r(t,t')} = -Q^z_r(t,t') + \alpha \left\langle \left\langle z_r(t) z_r(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{121}$$

$$\frac{\delta S}{\delta S^{\phi}(t,t')} = \hat{G}^z(t,t') + i\alpha \left\langle D \sum_{rr'} A^T_{rr'} \left\langle \hat{z}_{r'}(t) z_r(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{122}$$

$$\frac{\delta S}{\delta \hat{S}^{\phi}_{rr'}(t,t')} = G^z_{rr'}(t,t') + i\alpha \left\langle D\, A^T_{rr'} \left\langle z_r(t) \hat{z}_{r'}(t') \right\rangle_{W^{\vec{z}}_{D,\mathbf{A}}} \right\rangle_{D,\mathbf{A}}, \tag{123}$$

$$\frac{\delta S}{\delta Q^z_r(t,t')} = -\hat{C}^{\phi}_r(t,t') + \left\langle \left\langle \hat{x}(t)\hat{x}(t') \right\rangle_{W^x_c} \right\rangle_c, \tag{124}$$

$$\frac{\delta S}{\delta \hat{Q}^z(t,t')} = -C^{\phi}(t,t') + \left\langle \left\langle \phi(t)\phi(t') \right\rangle_{W^x_c} \right\rangle_c \tag{125}$$

$$\frac{\delta S}{\delta G^z_{rr'}(t,t')} = \hat{S}^{\phi}_{rr'}(t,t') + i \left\langle \left\langle \hat{x}(t)\phi(t') \right\rangle_{W^x_c} \right\rangle_c \tag{126}$$

$$\frac{\delta S}{\delta \hat{G}^z(t,t')} = S^{\phi}(t,t') + i \left\langle \left\langle \phi(t)\hat{x}(t') \right\rangle_{W^x_c} \right\rangle_c \tag{127}$$

$$\frac{\delta S}{\delta z^{(\mu^*)}_r(t)} = i\hat{z}^{(\mu^*)}_r(t) + i \left\langle \left\langle \hat{x}(t) \right\rangle_{W^x_c} \bar{m}^{(\mu^*,r)} \right\rangle_c \tag{128}$$

$$\frac{\delta S}{\delta \hat{z}^{(\mu^*)}_r(t)} = i z^{(\mu^*)}_r(t) + i \left\langle \left\langle \phi(t) \right\rangle_{W^x_c} D^{(\mu^*)} \bar{n}^{(\mu^*,r)} \right\rangle_c \tag{129}$$

We get the same self-consistency conditions as before, except now the single-site properties are averages over heterogeneous units, where the particular sample of $c = \{\bar{m}^{(\mu^*,r)}, \bar{n}^{(\mu^*,r)}\}_{\mu^*,r}$ for that unit determines its dynamics and properties, plus two additional conditions:

$$\hat{z}^{(\mu^*)}_r(t) = 0 \tag{130}$$

$$z^{(\mu^*)}_r(t) = D^{(\mu^*)} \left\langle \bar{n}^{(\mu^*,r)} \left\langle \phi(t) \right\rangle_{W^x_c} \right\rangle_c. \tag{131}$$

Then we have our final mean-field theory

$$T[x](t) = \eta^x(t) + \sum_{\mu^* r} \bar{m}^{(\mu^*,r)} z_r^{(\mu^*)}(t) + \int_{-\infty}^{t} dt' G^z(t,t')\phi(t') \qquad \eta^x(t) \sim \mathcal{GP}(0, Q^z(t,t')) \tag{132}$$

$$z_r(t) = \eta^{z_r}(t) + D \sum_{r'} A_{rr'} \int_{-\infty}^{t} dt' S^\phi(t,t') z_{r'}(t') \qquad \eta^{z_r}(t) \sim \mathcal{GP}(0, D^2 C^\phi(t,t')). \tag{133}$$

In the main text a factor of $\sqrt{N}$ gets traded between $\bar{m}^{(\mu^*,r)}$ and the non-normalized dominant pattern $z_r^{(\mu^*)}(t)$.

DMFT simplifications for ensemble of limit-cycle tasks

Consider the following process for generating $R \times R$ matrices, where R must be even. We start by sampling $R/2$ independent 2×2 rotation matrices matrices via $\theta_l \sim U(0, 2\pi)$ for $l = 1, \ldots, R/2$:

$$\mathbf{R}_l = \begin{pmatrix} \cos\theta_l & -\sin\theta_l \\ \sin\theta_l & \cos\theta_l \end{pmatrix}, \tag{134}$$

and arrange them in a block-diagonal structure. We then take a random $R \times R$ Haar orthogonal matrix $\mathbf{B}$ and intermix the entries, but preserving the traces of all matrix powers. We finally scale the result by $\gamma < 1$ (e.g. $\gamma = 0.99$) to ensure that $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{I} \end{pmatrix}$ is positive definite. Then

$$\mathbf{A} = \gamma \mathbf{B} \left(\bigoplus_{l=1}^{R/2} \mathbf{R}_l \right) \mathbf{B}^T \tag{135}$$

has a few desirable properties as a process for randomly generating a matrix. (For simplicity, the main text considers just $R = 2$, $\mathbf{B} = \mathbf{I} \implies \mathbf{A} = \gamma \mathbf{R}_1$, although its useful properties extend to this more general case.) First, for any $k \geq 1$, we have

$$\left\langle \mathrm{Tr}(\mathbf{A}^k) \right\rangle = \left\langle \mathrm{Tr} \left(\gamma^k \bigoplus_{l=1}^{R/2} \mathbf{R}_l^k \right) \right\rangle = \gamma^k \sum_{l=1}^{R/2} \left\langle \mathrm{Tr} \left(\mathbf{R}_l^k \right) \right\rangle = \gamma^k \sum_{l=1}^{R/2} \left\langle \mathrm{Tr} \left(\mathbf{R}_l \right) \right\rangle = 0, \tag{136}$$

using the fact that $\mathbf{R}_l^k$ has the same distribution as $\mathbf{R}_l$. The γ factor conveniently also reduces the SEM when we effectively "measure" the mean trace matrix power for finite samples of such matrices in simulated networks, which can be large for large powers k.

For independently sampled D, this property eliminates the single-unit self-coupling by setting $G^z(t,t') = 0$, but without resorting to a trivial distribution of $\mathbf{A} = \mathbf{0}$ as is necessary and sufficient in the $R = 1$ case. We also have a closed-form equation for $Q^z(\omega)$, which reads

$$Q^z(\omega) = \alpha \left\langle D^2 \left\| \mathbf{M}(\omega) \right\|_{\mathcal{F}}^2 \right\rangle_{D,\mathbf{A}} C^\phi(\omega). \tag{137}$$

where $\mathbf{M}(\omega) = (\mathbf{I} - DS^\phi(\omega)\mathbf{A})^{-1}$. Under these assumptions about the distribution of $\mathbf{A}$, we can calculate

$Q^z(\omega)$. First, we note that the similarity transformation pops out of the entire $\mathbf{M}(\omega)$ matrix:

$$\mathbf{M}(\omega) = \left(\mathbf{I} - DS^\phi \mathbf{A}\right)^{-1} = \left(\mathbf{B}\mathbf{B}^T - \gamma DS^\phi \mathbf{B} \bigoplus_{l=1}^{R/2} \mathbf{R}_l \mathbf{B}^T\right)^{-1} = \left(\mathbf{B}\left(\mathbf{I} - \gamma DS^\phi(\omega) \bigoplus_{l=1}^{R/2} \mathbf{R}_l\right)\mathbf{B}^T\right)^{-1} \tag{138}$$

$$= \mathbf{B}\left(\mathbf{I} - \gamma DS^\phi(\omega) \bigoplus_{l=1}^{R/2} \mathbf{R}_l\right)^{-1} \mathbf{B}^T \implies \|\mathbf{M}(\omega)\|_{\mathcal{F}} = \left\|\left(\mathbf{I} - \gamma DS^\phi(\omega) \bigoplus_{l=1}^{R/2} \mathbf{R}_l\right)^{-1}\right\|_{\mathcal{F}}, \tag{139}$$

since the Frobenius norm is invariant under orthogonal similarity transformations. Next we apply the matrix inversion on each block separately:

$$\left(\mathbf{I} - \gamma DS^\phi(\omega) \bigoplus_{l=1}^{R/2} \mathbf{R}_l\right)^{-1} = \bigoplus_{l=1}^{R/2} (\mathbf{I}_{2\times 2} - \gamma DS^\phi(\omega)\mathbf{R}_l)^{-1}. \tag{140}$$

Each component matrix that gets inverted $\mathbf{I}_{2\times 2} - \gamma DS^\phi(\omega)\mathbf{R}_l$ has eigenvalues $1 - \gamma DS^\phi(\omega)e^{\pm i\theta_l}$, so their inverses have eigenvalues $1/(1 - \gamma DS^\phi(\omega)e^{\pm i\theta_l})$. Since this matrix is clearly normal, its Frobenius norm squared is simply the sum of the eigenvalues' squared norms. Putting it all together and assuming D is sampled independently from $\mathbf{A}$, we have

$$\left\langle D^2 \|\mathbf{M}(\omega)\|_{\mathcal{F}}^2 \right\rangle_{D,\mathbf{A}} = \left\langle \sum_{l=1}^{R/2} \left(\frac{D^2}{|1 - \gamma DS^\phi(\omega)e^{i\theta_l}|^2} + \frac{D^2}{|1 - \gamma DS^\phi(\omega)e^{-i\theta_l}|^2} \right) \right\rangle_{D,\{\theta_l\}} \tag{141}$$

$$= R \left\langle \frac{D^2}{|1 - \gamma DS^\phi(\omega)e^{i\theta}|^2} \right\rangle_{D,\theta} = R \left\langle D^2 \frac{1}{2\pi} \int_0^{2\pi} d\theta \frac{1}{|1 - \gamma DS^\phi(\omega)e^{i\theta}|^2} \right\rangle_D \tag{142}$$

$$= R \left\langle \frac{D^2}{1 - |\gamma DS^\phi(\omega)|^2} \right\rangle_D. \tag{143}$$

The last line requires assuming $|\gamma DS^\phi(\omega)| < 1$, which is a fairly safe assumption for a distribution of D with no outliers. Then, in Fourier space, the statistics of the effective neuron noise follow

$$Q^z(\omega) = \alpha R \left\langle \frac{D^2}{1 - |\gamma DS^\phi(\omega)|^2} \right\rangle_D C^\phi(\omega). \tag{144}$$

With these simplifications, we have now a bipartite mean-field theory

$$(1 + \partial_t)x(t) = \eta^x(t) \tag{145}$$

$$z_r(t) = \eta^{z_r}(t) + D\left\langle \phi' \right\rangle \sum_{r'} A_{rr'} \int_{-\infty}^{t} dt' e^{-(t-t')} z_{r'}(t'), \tag{146}$$

because we know $S^\phi(\tau) = \Theta(\tau)\left\langle \phi' \right\rangle e^{-\tau}$ and $G^z(\tau) = 0$. We can differentiate the second equation and get

$$(1 + \partial_t)z_r(t) = \dot{\eta}^{z_r}(t) + \eta^{z_r}(t) + D\left\langle \phi' \right\rangle \sum_{r'} A_{rr'} z_{r'}(t'), \tag{147}$$

exploiting the fact that the derivative of $S^\phi(\tau)$ is negative of itself, producing a leak term. With this representation of the weak-task latent variable dynamics, it is much more straightforward to determine the stability condition on D: it reduces to whether the largest real part among eigenvalues of $D\langle\phi'\rangle\mathbf{A}$ is below or above 1.

Let's examine the single-site picture and solve for the temporal dynamics of the $x(t)$ autocovariance:

$$(1+\omega^2)C^x(\omega) = \alpha R\left\langle \frac{D^2}{1-|\gamma D S^\phi(\omega)|^2}\right\rangle_D C^\phi(\omega) \tag{148}$$

$$\implies C^x(\omega) = \alpha R\left\langle \frac{D^2}{1-(\gamma D\langle\phi'\rangle)^2+\omega^2}\right\rangle_D C^\phi(\omega). \tag{149}$$

Now let's make yet a further assumption, that D is in fact a constant such that $g_{\text{eff}}^2 = \alpha R D^2$ rather than following some distribution—this is a common choice for the non-selected tasks. Then we have

$$\partial_\tau^2 C^x(\tau) = \left(1-\frac{\gamma^2\nu}{\alpha R}\right)C^x(\tau) - g_{\text{eff}}^2 C^\phi(\omega), \tag{150}$$

which is exactly the second-order ODE describing $C^x(\tau)$ for an unstructured network [36], except the $C^x(\tau)$ term is scaled by $1-\frac{\gamma^2\nu}{\alpha R}$ instead of 1, where $\nu \equiv (g_{\text{eff}}\langle\phi'\rangle)^2$ can be thought of as the effective slope for an equivalent linear network [43, 70].

The average gain depends exclusively on $C^x(0)$ (via a Gaussian integral). However, to solve for $C^x(0)$ by the Sompolinsky potential-energy trick in this case requires knowing the average gain already, because of its appearance in ν. Therefore we solve for $C^x(0)$ for a provisional value of $\langle\phi'\rangle$, which we iteratively update as we solve for $C^x(0)$ again with our new value of $\langle\phi'\rangle$, until it stops changing. This value of $\langle\phi'\rangle$ then determines the stability for an arbitrary task, based on the spectral abscissa of $D\langle\phi'\rangle\mathbf{A}$. However, the purely spectral methods we describe earlier are more effective for small α.