Limits of optimal decoding under synaptic coarse-tuning

Ori Hendler,[1,2,*] Ronen Segev,[2,3,4] and Maoz Shamir[1,2,5]

[1] *Department of Physics, Ben-Gurion University of the Negev, Beer-Sheva, Israel*
[2] *School of Brain Sciences and Cognition, Ben-Gurion University of the Negev, Beer-Sheva, Israel*
[3] *Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel*
[4] *Department of Biomedical Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel*
[5] *Department of Physiology and Cell Biology, Ben-Gurion University of the Negev, Beer-Sheva, Israel*

Sensory information propagates through successive processing stages in the brain, where synaptic weight patterns between stations determine how downstream neurons decode information from upstream populations. Although optimized synaptic connectivity can enhance information transmission, it requires precise weight tuning. Recent evidence depicting substantial synaptic volatility raises two fundamental questions: How does coarse-tuning of synaptic connectivity affect information transmission? What strategies could the nervous system employ to maintain reliable communication despite synaptic fluctuations? We addressed these questions by analyzing the signal-to-noise ratio (SNR) for binary stimulus discrimination under two decoding schemes: a naïve population average and an optimized linear decoder. For the naïve decoder, we found that SNR remains largely insensitive to synaptic imprecision, since performance is already limited by correlated noise in neuronal responses. For the optimal decoder, we identified three distinct regimes, that is, weak, moderate and strong coarse-tuning. Under weak coarse-tuning, SNR^2 scales linearly with population size N. Under moderate coarse-tuning, scaling becomes sublinear. Strikingly, under strong coarse-tuning, the regime most consistent with observed neuronal heterogeneity, SNR saturates and can not be improved by recruiting larger populations. This limitation persists even when incorporating feedforward or recurrent network architectures. These findings suggest that in the biologically relevant regime of strong coarse-tuning, naïve and optimal decoders can achieve qualitatively similar performance. The analysis shows that effective readout under synaptic volatility is constrained to an invariant low-dimensional manifold aligned with the naïve decoder, potentially pointing to a fundamental principle for robust neural computation in the face of ongoing synaptic remodeling.

I. INTRODUCTION

One of the main enigmas in neuroscience is how sensory information is transmitted and decoded as it propagates downstream [1]. The profile of synaptic weights largely determines the fidelity of information transfer along the processing pathway [2–4]. Accordingly, the synaptic-weight profile can be regarded as a decoder or a readout algorithm [5, 6]. One such possible decoder is a naïve readout that assigns uniform weights to all inputs. While easy to implement, it generally results in suboptimal performance [7–11]. By contrast, an optimal decoder can leverage detailed knowledge of the population response statistics to achieve superior performance by fine-tuning its synaptic weights to the detailed statistics of the network activity [12–15]. The caveat is that synaptic weights undergo substantial remodeling.

A growing body of empirical evidence has demonstrated that synapses are highly volatile [16, 17]. In-vivo studies show that dendritic spines, which reflect synaptic strength, exhibit substantial size fluctuations over time scales ranging from a few hours to days [18–21]. In-vitro studies of cultured cortical neurons show that synaptic sizes fluctuate dramatically at a time scale of hours [22–25], with considerable fluctuations on a similar scale emerging even in the absence of neuronal activity [26–28].

Theoretical studies typically partition synaptic weights into structured and unstructured components [29–33]. The structured component captures the tuned connectivity, whereas the unstructured component represents random variability, which we refer to as the coarse-tuning of the weights, which may or may not fluctuate over time [18, 34–40].

Considerable theoretical attention has been devoted to querying how the brain can maintain functionality in the face of considerable coarse-tuning of synaptic weights. To date, however, the computational implications of the coarse-tuning of synaptic weights on the fidelity of information transfer in the brain remain largely unknown.

Here, we investigated the effect of the coarse-tuning of synaptic weights on readout accuracy in the framework of a modeling study. We begin by defining a statistical model of neuronal responses to the stimulus. We then define two decoders: a naïve decoder and an optimal one. We then apply coarse-tuning to the synaptic weights of both decoders, and investigate its effect on the decoding accuracy. Next, we consider two extensions of our readout model: a multilayer Feedforward Neural Network architecture and a Recurrent Neural Network. We show that neither can mitigate the drastic detrimental effect of coarse-tuning. The discussion centers on the implications of these findings for the theory of population coding in the brain.

* hendlero@post.bgu.ac.il

II. STATISTICAL MODEL OF THE NEURAL RESPONSE

TABLE I: Alphabetical glossary of mathematical symbols used throughout the manuscript.

Symbol	Meaning
a	Single neuron variance
c	Pairwise correlation coefficient
Δ	Disorder in random fields
g_i	Response selectivity of the i'th neuron
γ	Coarse-tuning scaling
h	Field
J	Disorder in recurrent connections
J_0	Ferromagnetic bias
κ	Coarse-tuning magnitude
μ_d	Quenched mean response to distractor
μ_g	Quenched mean of response selectivity
μ_t	Quenched mean response to target
N	Number of neurons
s	Sign of the field
σ_g^2	Quenched variance of response selectivity
T	Temperature
w_i	Weight of the i'th neuron
χ	Susceptibility

We model a population of N neurons coding for a binary stimuli $\alpha \in \{t, d\}$ (target or distractor). We assume that the neural responses $\mathbf{r}$, conditioned on stimulus α, is distributed according to multivariate Gaussian statistics with mean

$$\langle r_i \rangle_{r|s} = \mu_i^\alpha, \tag{1}$$

and covariance matrix

$$C_{ij} = a\big((1-c)\delta_{ij} + c\big), \tag{2}$$

where $\langle \cdot \rangle$ denotes averaging over neural responses for a given stimulus. The parameter a is the single neuron variance and c is the pairwise correlation coefficient. A Gaussian approximation offers a tractable and empirically-based description of neural responses [41]. The population is heterogeneous, where each neuron has distinct mean response parameter μ_i^α randomly drawn when the network is created and fixed thereafter. Trial-to-trial fluctuations thus arise from the noise structure whereas variability in μ_i^α reflects frozen heterogeneity across neurons. Different network realizations yield different parameter sets $\boldsymbol{\mu}^\alpha$, with first-order statistics

$$\ll \mu_i^t \gg = \mu_t, \quad \ll \mu_i^d \gg = \mu_d, \quad \ll g_i \gg = \mu_g, \tag{3}$$

where double brackets, $\ll \cdot \gg$, denote ensemble average over realizations. The second-order statistics are

$$\mathrm{Var}(g_i) = \sigma_g^2, \tag{4}$$

where we assume that μ_i^t and μ_i^d are uncorrelated and have equal variance $\frac{\sigma_g^2}{2}$.

We further assume that the neuronal responses exhibit stimulus selectivity, where on average, responses to the target exceed responses to the distractor. We quantify response selectivity as

$$g_i = \mu_i^t - \mu_i^d. \tag{5}$$

III. THE READOUT MODEL

To examine the accuracy at which information about a stimulus can propagate downstream, we study readout performance in the framework of a two-interval-two-alternative forced-choice task. In this paradigm, one interval presents the target stimulus and the other the distractor, with their order randomized across trials. Without loss of generality, we assume that the target is presented in the first interval for all analytical calculations. The linear readout discriminates between the two alternatives according to the sign of the field:

$$h \equiv \mathbf{w}^\top \big(\mathbf{r}^t - \mathbf{r}^d\big). \tag{6}$$

Thus, the linear readout is fully defined by its weight vector $\mathbf{w}$. The binary decision follows from the sign of this field:

$$s = \mathrm{sign}(h), \tag{7}$$

where $s = +1$ indicates "target first" (correct under our assumption) and $s = -1$ indicates "target second" (incorrect). In our Gaussian response model, the field h varies from trial to trial with mean $\langle h \rangle = \mathbf{w}^\top \big(\boldsymbol{\mu}^t - \boldsymbol{\mu}^d\big)$ and variance $\langle (\delta h)^2 \rangle = 2\,\mathbf{w}^\top \mathbf{C}\mathbf{w}$, contingent on the specific realization of $\mathbf{w}$. For a given realization, the probability of error is therefore

$$P_{\mathrm{err}} = \int_{SNR}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du, \tag{8}$$

where the signal-to-noise ratio (SNR) squared is given by

$$SNR^2 = \frac{\big(\mathbf{w}^\top \mathbf{g}\big)^2}{2\mathbf{w}^\top \mathbf{C}\mathbf{w}}, \quad signal = \langle h \rangle, \quad noise = \sqrt{\langle (\delta h)^2 \rangle}. \tag{9}$$

The probability of error is fully determined by the SNR. For large values of the SNR, a standard asymptotic expansion yields

$$P_{\mathrm{err}} = \frac{e^{-SNR^2/2}}{\sqrt{2\pi}\, SNR} \left(1 - \mathcal{O}(SNR^{-2})\right). \tag{10}$$

where the error probability decays exponentially with SNR^2 up to algebraic corrections. Accordingly, we study how different choices of the weight vector $\mathbf{w}$ impact the SNR and hence task performance, see Fig.1a.

IV. READOUT PERFORMANCE

A. Naïve Decoder Performance

We analyze the SNR as a function of different choices of readout weights, starting with the Naïve decoder. Since the response to the target is larger on average than

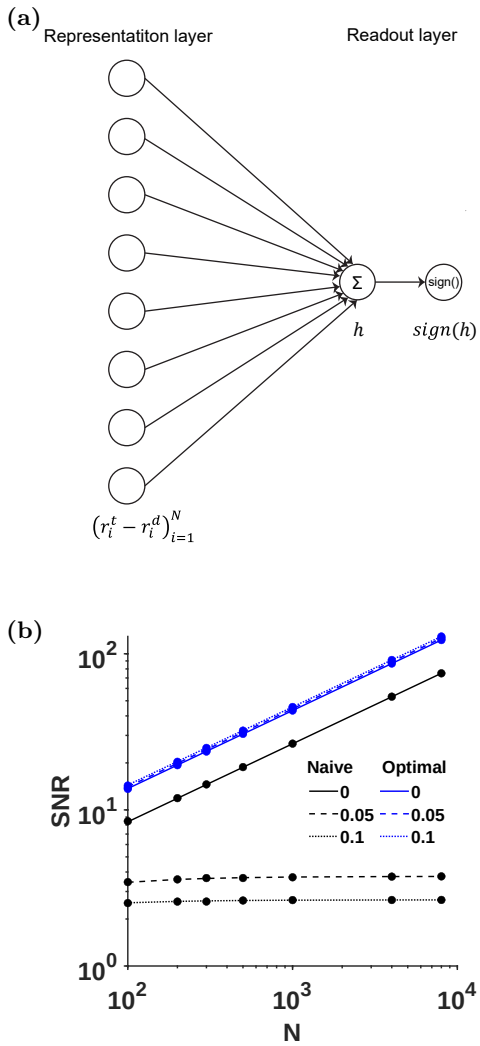

FIG. 1: **Readout model of the naïve and optimal decoders. (a)** Schematic illustration of the model architecture. **(b)** The signal-to-noise ratio (SNR) is plotted as a function of population size N for the optimal (blue) and naïve (black) decoders. Different values of the pairwise correlation coefficient $c = 0, 0.05, 0.1$ are depicted by solid, dashed, and dotted lines, respectively.

the response to the distractor, the naïve approach consists of assigning equal weights to all neurons:

$$w_i^{\text{naïve}} = \frac{1}{N}. \tag{11}$$

The $noise^2$ (see Eq.(9)) of the naïve decoder is $\mathcal{O}(N^0)$ and does not depend on the specific realization of quenched disorder, see Appendix A. The $signal^2$ of the naïve decoder is a random variable that depends on the specific realization of the system; i.e., the choice of $\{g_i\}_{i=1}^N$. The $signal^2$ has a mean of $\mathcal{O}(N^0)$ and variance $\mathcal{O}(N^{-1})$ across different realizations. Thus, for large populations ($N \to \infty$), the $signal^2$ of a typical realization is equal to its quenched mean, up to fluctuations that vanish in the limit of large N. This property of a random variable whose fluctuations vanish relative to its mean is referred to as self-averaging. In the limit of large populations, a self-averaging variable can be replaced by its quenched mean. The quenched mean of the squared signal-to-noise ratio for the naïve decoder is (see Appendix A):

$$\ll SNR^2(\mathbf{w}^{\text{naïve}}) \gg\approx \begin{cases} \frac{N\mu_g^2}{2a} & \text{, if } c = 0 \\ \frac{N\mu_g^2}{2a}\left(\frac{1}{(1-c)+\frac{N}{N_{\text{eff}}}}\right) & \text{, if } c > 0 \end{cases} \tag{12}$$

where we define $N_{\text{eff}} \equiv \frac{1}{c}$. When response fluctuations of different neurons are uncorrelated, $c = 0$, the SNR^2 scales linearly with the population size, N, $\ll SNR^2(\mathbf{w}^{\text{naïve}}) \gg\propto NI_0$, where $I_0 = \frac{\mu_g^2}{2a}$ is the (quenched mean of the) SNR^2 of the single neuron, Fig.1b (solid black line). For $c > 0$, the SNR^2 scales linearly with the population size, $\ll SNR^2(\mathbf{w}^{\text{naïve}}) \gg\propto NI_0$, for small populations, $N \ll N_{eff}$. In the limit of large populations, $N \gg N_{eff}$, the SNR^2 of the naïve readout saturates to $\ll SNR^2(\mathbf{w}^{\text{naïve}}) \gg\propto N_{eff}I_0$, Fig.1b. Note that due to self-averaging, the error bars denoting the standard deviation of the SNR^2 across different realizations of the quenched disorder are barely visible.

B. Optimal Decoder Performance

The optimal weights that maximize the probability of correct discrimination are given by:

$$\mathbf{w}^{\text{opt}} = \mathbf{C}^{-1}\mathbf{g}. \tag{13}$$

In our model; i.e., Eq. 2, this yields:

$$w_i^{\text{opt}} = \frac{1}{Na(1-c)}\left(g_i - \frac{Nc}{(1-c)+Nc}\bar{g}\right), \tag{14}$$

where $\bar{g} = \frac{1}{N}\sum_{j=1}^N g_j$ is the population-average response selectivity. In the limit of a large population ($N \to \infty$) the weights are given by

$$w_i^{\text{opt}} \approx \frac{1}{Na(1-c)}\left(g_i - \bar{g}\right). \tag{15}$$

Hence, the optimal decoder uses the heterogeneity to extract information from a subspace that is orthogonal to the space with large noise correlations. We use this approximation, Eq. 15, for the analysis below. The SNR^2

of the optimal decoder (Eq. (14)) is given by (see Appendix B):

$$\ll SNR^2(\mathbf{w}^{\text{opt}}) \gg = \frac{\big(1+(N-2)c\big)\sigma_g^2 + \big(1-c\big)\mu_g^2}{2a(1-c)\Big(\frac{1+c(N-1)}{N}\Big)}. \tag{16}$$

In a homogeneous population; i.e., $\sigma_g^2 = 0$, the optimal readout is the naïve readout and its SNR is limited by the noise correlations. In a heterogeneous population, $\sigma_g^2 \neq 0$, the SNR of the optimal readout is superior to that of the naïve one. In particular its SNR^2 scales linearly with the population size, the leading term: $\ll SNR^2(\mathbf{w}^{\text{opt}}) \gg \propto \frac{N\sigma_g^2}{2a(1-c)}$ (for $c \neq 0$), see Fig.1b (blue lines). However, these results rely on the ability of the system to finely tune the readout weights to the specific realization of the response heterogeneity; e.g., Eq. 14.

V. COARSE-TUNED READOUT MODEL

Up to now, the quenched variability was only applied to the neuronal responses (i.e., $\{g_i\}_{i=1}^N$), but not to the synaptic weights. To model synaptic volatility we introduce additive quenched fluctuations to the weights,

$$\tilde{w}_i = \frac{w_i}{\sqrt{N}\|\mathbf{w}\|} + \xi_i, \quad \xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \kappa^2 N^{\gamma-1}\right), \tag{17}$$

where the fluctuations ξ_i are independent and identically distributed Gaussian variables, with zero mean and variance $\kappa^2 N^{\gamma-1}$. The parameter κ governs the magnitude of the fluctuation; i.e., the coarse-tuning magnitude. The parameter γ determines the scaling of the fluctuation variance with the population size, N; i.e., the coarse-tuning scaling.

Thus, the weight $\tilde{w}_i$ is expressed as the sum of a structured component of order $\mathcal{O}(N^{-1})$ and an unstructured component that scales like $\mathcal{O}(N^{\gamma})$. This decomposition into structured and unstructured components is a classic feature of theoretical models, such as the Hopfield model [34], the Sherrington–Kirkpatrick (SK) model [42], and spin-glass models of neural networks [37, 43], as well as more recent works [29, 30, 44, 45]. In these models, the structured component of the weight vector $\mathbf{w}$ typically scales as $\mathcal{O}(N^{-1})$, whereas the unstructured component scales as $\mathcal{O}(N^{-\frac{1}{2}})$, which in our notation corresponds to $\gamma = 0$.

VI. IMPACT OF COARSE-TUNING ON READOUT PERFORMANCE

The analysis of the effects of coarse-tuning on the SNR of the naïve and optimal readouts (below) identifies three regimes: weak ($\gamma = -1$), moderate ($-1 < \gamma < 0$) and strong ($\gamma = 0$) coarse-tuning. For weak coarse-tuning ($\gamma = -1$), the *signal* is self-averaging for both readouts, with mean $\mathcal{O}(N^0)$ and variance of order $\mathcal{O}(N^{-1})$, see Appendix C, D. Fig.2a presents the quenched distribution of the *signal* for weak coarse-tuning across the different realizations of response selectivity, $\{g_i\}_{i=1}^N$, and weights, $\{\tilde{w}_i\}_{i=1}^N$, where each row corresponds to a different population size, N. As shown in the figure, the variability of the *signal* vanishes with increasing N. For weak coarse-tuning, the $noise^2$ of the naïve decoder is self-averaging, Fig.2b (left column). However, the $noise^2$ of the optimal decoder is not self-averaging, Fig.2b (right column), with mean $\mathcal{O}(N^{-1})$ and variance $\mathcal{O}(N^{-2})$. Nevertheless, for both decoders, the SNR is well-approximated by $\ll SNR \gg \approx \frac{\ll signal \gg}{\sqrt{\ll noise^2 \gg}}$, Fig.2c.

Similar behavior is found for moderate coarse-tuning. Specifically, the $signal^2$ of both readouts is self-averaging. The $noise^2$ is self-averaging for the naïve decoder (Appendix C), but not for the optimal decoder (Appendix D). In the strong coarse-tuning regime ($\gamma = 0$), neither the *signal* nor the *noise* are self-averaging. Nevertheless, $\ll SNR \gg$ is well-approximated by the ratio of the averaged *signal* to the square root of the averaged $noise^2$; i.e., $\ll SNR \gg \approx \frac{\ll signal \gg}{\sqrt{\ll noise^2 \gg}}$. The quenched mean $\ll SNR \gg$ provides a fair approximation of the typical SNR. We therefore use the *signal* and $noise^2$ to quantify the SNR below.

We find that for large N the quenched mean of the SNR of the naïve and of the optimal readouts is well-approximated by (see Appendix C, D):

$$\ll SNR(\tilde{\mathbf{w}}^{\text{naïve}}) \gg \approx \frac{\mu_g}{\sqrt{2a}\sqrt{\frac{1+(N-1)c}{N} + \kappa^2 N^{\gamma}}}, \tag{18}$$

$$\ll SNR(\tilde{\mathbf{w}}^{\text{opt}}) \gg \approx \frac{\sigma_g}{\sqrt{2a}\sqrt{\frac{(1-c)}{N} + \kappa^2 N^{\gamma}}}. \tag{19}$$

As shown in Eqs. (18)-(19) the nominator is independent of the population size and the scaling of the denominator with N depends on γ.

In the weak coarse-tuning regime, $\gamma = -1$, the SNR^2 of the optimal decoder scales linearly with N. The slope of the linear scaling is determined by the magnitude of the quenched fluctuations, κ (see Fig.3a). For moderate coarse-tuning, $-1 < \gamma < 0$, the SNR^2 grows with the population size, albeit in a sublinear manner, scaling as $N^{-\gamma}$, Fig.3b. For strong coarse-tuning, $\gamma = 0$, the SNR saturates to an asymptotic value, $\frac{\sigma_g}{\sqrt{2a\kappa^2}}$, in the limit of large N, Fig.3b and Fig.3e.

The SNR of the naïve decoder saturates in all coarse-tuning regimes. For weak and moderate coarse-tuning, $-1 \leq \gamma < 0$, the asymptotic value of the naïve SNR is determined by the noise correlations, $\lim_{N\to\infty} SNR = \frac{\mu_g}{\sqrt{2ac}}$, and is independent of the quenched disorder, Fig.3c and Fig.3d. In the case of strong coarse-tuning, $\gamma = 0$, the SNR of the naïve readout saturates to $\lim_{N\to\infty} SNR = \frac{\mu_g}{\sqrt{2a(c+\kappa^2)}}$, Fig.3e.

In summary, in weak and moderate levels of coarse-tuning, $-1 \leq \gamma < 0$, the optimal readout is superior to

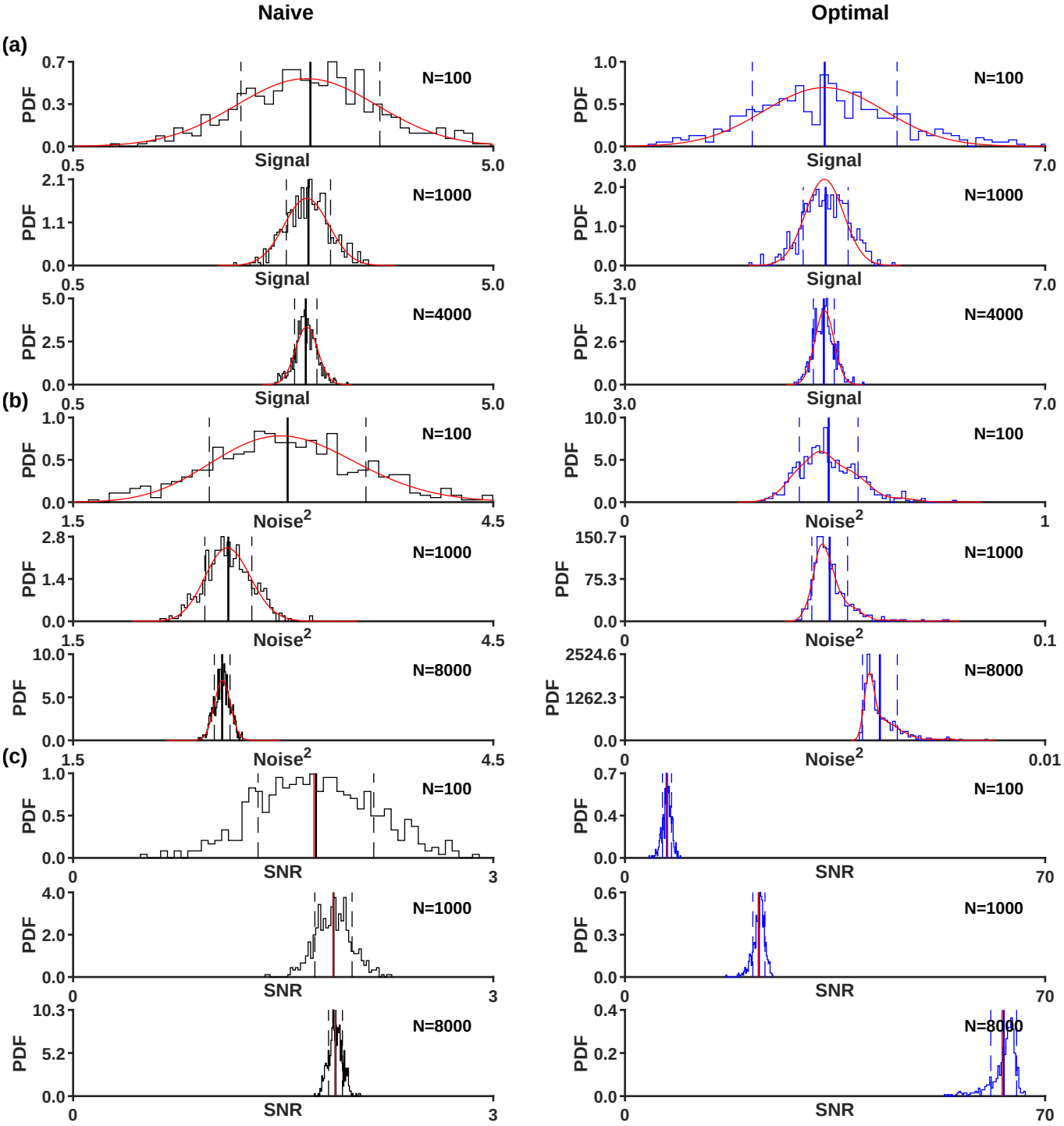

FIG. 2: **Histograms of signal, noise^2, and SNR.** Distributions of the naïve (black) and optimal (blue) decoders across population sizes $N \in \{100, 1000, 4000\}$ under coarse-tuning with $\kappa = 1$ and $\gamma = -1$. **(a)** *signal* distribution. **(b)** $noise^2$ distribution. **(c)** SNR distribution. Solid vertical lines indicate the mean; dashed lines indicate the ± 1 standard deviation. Red curves show analytical predictions derived in the main text.

that of the naïve readout, and its SNR increases with the population size. In the strong coarse-tuning regime, $\gamma = 0$, the SNR of both readouts saturate to a finite limit. Moreover, in this case, the performance of the naïve readout can be superior to that of the optimal readout, depending on the choice of parameters. This raises the question of whether the performance of the optimal readout in the case of strong coarse-tuning can be im-

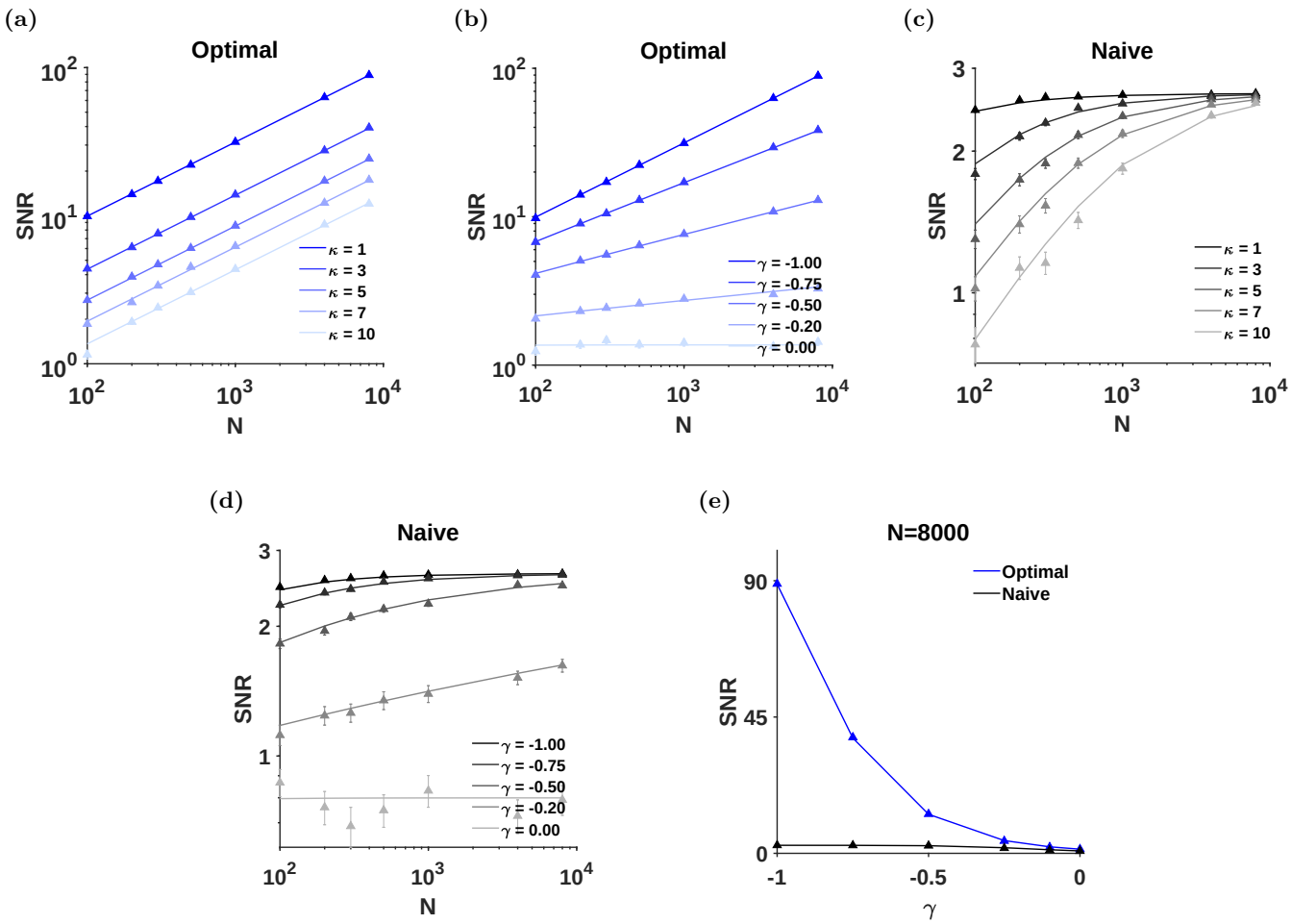

FIG. 3: ***SNR* scaling with population size under coarse-tuning. (a)** SNR versus population size N for the optimal decoder, with varying magnitudes of coarse-tuning, $\kappa \in \{1, 3, 5, 7, 10\}$ at fixed $\gamma = -1$. **(b)** SNR versus N for the optimal decoder, with varying scalings of coarse-tuning, $\gamma \in \{-1, -0.75, -0.5, -0.2, 0\}$ at fixed $\kappa = 1$. **(c,d)** Same as (a,b), but for the naïve decoder. **(e)** SNR versus γ at fixed population size $N = 8000$ and $\kappa = 1$. In all panels, the color gradient from dark to light indicates increased coarse-tuning.

proved by considering a more elaborate readout structure incorporating more layers and recurrent connectivity.

VII. READOUT LAYER AND FURTHER PROCESSING

A. Feedforward Neural Network

To incorporate an additional processing layer, we treat the binary decision variable s (Eq. 7) as a single neuron in the processing layer of N neurons, $\{s_i\}_{i-1}^N$ with $s_i = \text{sign}(h_i)$ (see Fig.4a), where h_i is the local field of neuron i and is given by

$$h_i = \sum_j \tilde{w}_{i,j}(r_j^t - r_j^d), \tag{20}$$

As above, we assume that the feedforward weight matrix, $\tilde{\mathbf{w}} = \mathbf{w} + \boldsymbol{\xi}$, is a sum of two terms. The structured component, $\mathbf{w}$, is identical for all neurons, $w_{i,j} = w_{i',j}$ $(\forall i, i')$. The second term, $\boldsymbol{\xi}$, is a random matrix representing the quenched noise of the synaptic weights. The elements $\{\xi_{i,j}\}$ are assumed to be i.i.d. Gaussian random variables with zero mean and $\ll \xi_{i,j}\xi_{i',j'} \gg = \delta_{i,i'}\delta_{j,j'}\kappa^2 N^{\gamma-1}$.

Given the synaptic weights, $\mathbf{w}$, the local fields, $\{h_i\}$, are random variables with a mean over trials $\langle h_i \rangle = \sum_j \tilde{w}_{i,j} g_j$. We denote the average taken over both trials and realizations by $[\cdot]$, yielding $[h_i] = \sigma_g \equiv h_0$. The covariance, $\text{cov}(h_i, h_k) = [h_i h_k] - h_0^2$, is given by (see Appendix E)

$$\text{cov}(h_i, h_k) = \frac{a(1-c)}{N} + N^\gamma \kappa^2 (\mu_g^2 + \sigma_g^2 + a)\delta_{ik}. \tag{21}$$

The first term on the right-hand side of Eq. 21 arises from the fact that all neurons in the processing layer

receive identical structured inputs. This term represents the residual noise present in the optimal readout even under perfect fine-tuning of the weights.

In what follows, we neglect this contribution and approximate the distribution of the random fields, $\{h_i\}$, by independent Gaussian variables with

$$\Pr(h_i) = \frac{1}{\sqrt{2\pi}\Delta} \exp\left(-\frac{(h_i - h_0)^2}{2\Delta^2}\right). \tag{22}$$

where, $\Delta^2 = N^\gamma \kappa^2(\mu_g^2 + \sigma_g^2 + a)$.

Under the approximation of Eq. 22 the random fields, $\{h_i\}$, are i.i.d. random variables with a variance scaling as N^γ. This variability in the random fields induces heterogeneity in the responses of the neurons in the processing layer, $\{s_i\}$. In the weak to moderate coarse-tuning regimes, $-1 \leq \gamma < 0$ this heterogeneity vanishes as the population size, N grows. We therefore focus on the strong coarse-tuning regime, $\gamma = 0$, where heterogeneity in both the random fields and the neuronal responses remains $\mathcal{O}(N^0)$. Note that in the above approximation, the trial-to-trial fluctuations of responses of neurons in the processing layer, $\{s_i\}$, are uncorrelated. Hence, we set $\gamma = 0$.

To estimate the stimulus from the neural responses, $\{s_i\}$, we need to define a readout. A natural choice is a linear readout: $\sum_k c_k s_k$, where c_k are the optimal weights for the processing layer. Consistency requires us to assume strong coarse-tuning in the readout weights from the processing layer as well. Consequently, adding feedforward layers cannot overcome the limiting effect of strong coarse-tuning. Below we consider the addition of recurrent connections to the processing layer.

B. Recurrent Neural Network

Can recurrent connectivity overcome the quenched noise in the case of strong coarse-tuning? To address this question we incorporate recurrent interactions between neurons in the processing layer. Thus, we consider the network to comprise three layers: (i) a representation layer, $\{r_i\}_{i=1}^N$, (ii) a single intermediate processing layer with neurons, $\{s_i\}_{i=1}^N$, which are recurrently connected, and (iii) a decision unit denoted by $\hat{m}$, Fig.4b.

The representation layer $\mathbf{r} = \{r_i\}_{i=1}^N$ encodes the binary stimulus as previously, and has feedforward connections to the processing layer. Each processing neuron receives an i.i.d. quenched random local field $\{h_i\}$, distributed according to Equation (22), which reflects the input from $\mathbf{r}$. Crucially, unlike the purely feedforward architecture, the processing neurons could now interact via recurrent connections.

Each neuron in the processing layer is modeled as an Ising spin, $s_i \in \{\pm 1\}$ for $i \in \{1 \dots N\}$. The network evolves according to Glauber dynamics at temperature $\frac{1}{\beta}$, governed by the Hamiltonian:

$$\mathcal{H} = -\sum_{i,j} J_{i,j} s_i s_j - \sum_i h_i s_i. \tag{23}$$

The recurrent connections $\{J_{i,j}\}$, are taken to be i.i.d. quenched Gaussian random variables with:

$$\Pr(J_{i,j}) = \sqrt{\frac{N}{2\pi J^2}} \exp\left(-\frac{N(J_{i,j} - \frac{J_0}{N})^2}{2J^2}\right), \tag{24}$$

where $\frac{J_0}{N}$ is the mean strength of the recurrent connections and $\frac{J}{\sqrt{N}}$ is the standard deviation of the recurrent connections. The parameter J represents strong quenched disorder in the recurrent connections, similar to the strong coarse-tuning of the feedforward connectivity with $\gamma = 0$. Thus, the stochastic response of the Recurrent Neural Network neurons is naturally mapped to the Sherrington-Kirkpatrick (SK) model [42] with random fields [46, 47]. For simplicity, we assume that the trial-to-trial (thermal) fluctuations are more pronounced than the quenched noise, $T > J$. In this regime, one of the central order parameters characterizing the system is its magnetization (per spin), $m = \frac{1}{N}\sum_i s_i$. This system has been studied extensively, and in our regime of interest its phase diagram (for $h_0 = 0$) is characterized by an ordered (ferromagnetic), $m > 0$, and unordered (paramagnetic), $m = 0$, phase, see [46–49] and appendix G.

When the recurrent connectivity exhibits a ferromagnetic bias, $J_0 > 0$, the Recurrent Neural Network amplifies the signal in the structured component, h_0, of the input fields. Figs. 4c & 4d illustrate this susceptibility, which quantifies the change in the Recurrent Neural Network response, m, to small changes in the input signal, $\chi = \frac{dm}{dh_0}$. As shown, the susceptibility is maximal near the phase transition (marked by the black line). Comparing Figs. 4c & 4d, the main effect of the quenched disorder in the random fields, Δ, is that to obtain high signal amplification for larger values of Δ a stronger ferromagnetic bias, J_0, is required. In the ordered phase, the structured component of the input fields acts as an ordering field that chooses the direction of the spontaneous magnetization, Figs. 4e & 4f. As in the case of susceptibility, larger values of Δ require a stronger ferromagnetic bias, J_0, to obtain the same level of magnetization.

In the regions of the phase diagram in which the amplification is large, most of the signal resides in a two dimensional subspace spanned by the input fields, h, and the uniform direction, Figs. 4g & 4h. This takes us back to our initial problem: extracting the signal requires some degree of tuning of the readout weights.

VIII. DISCUSSION

Here we investigated how coarse-tuning constrains the accuracy of population codes. We focused on the scaling

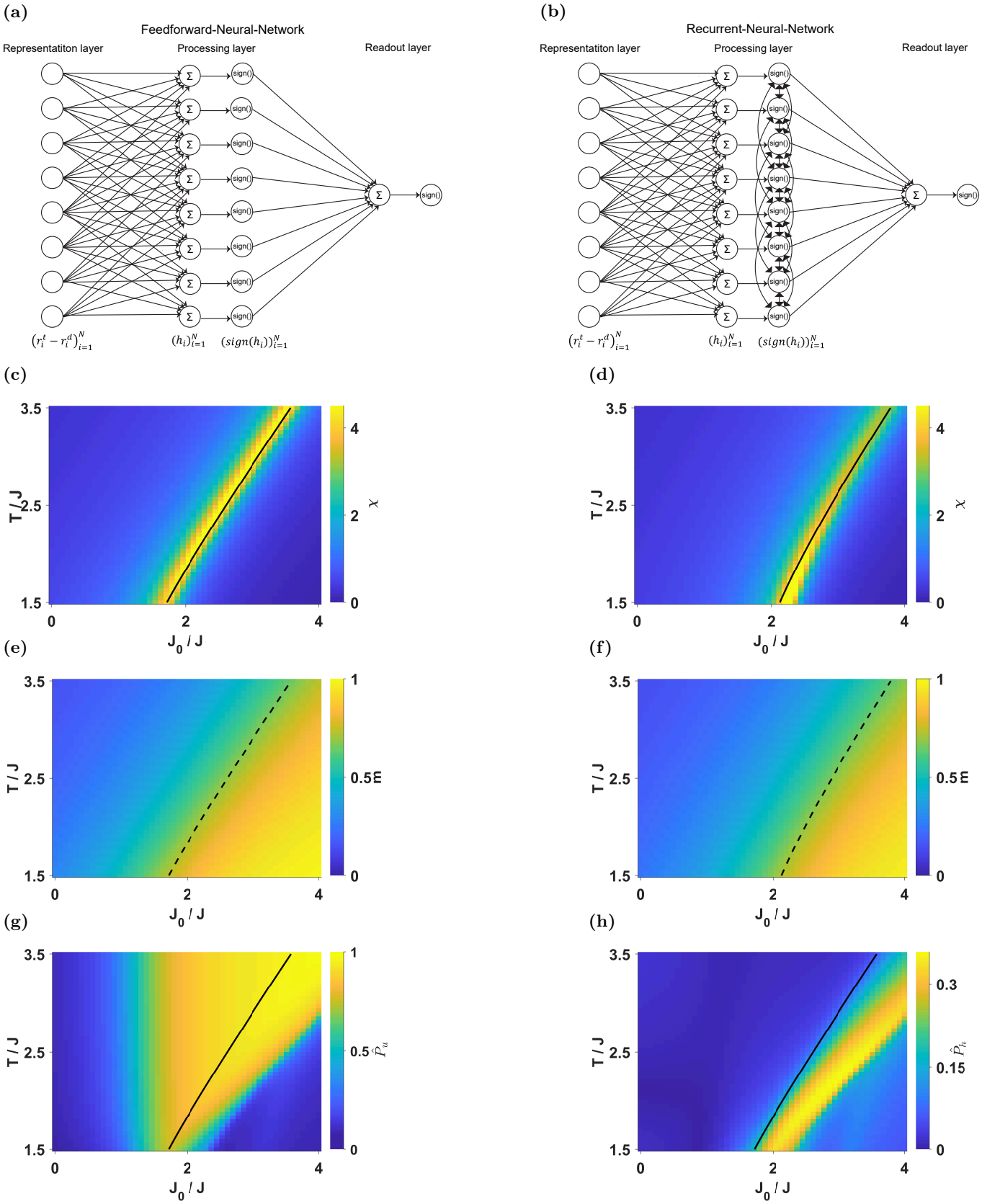

FIG. 4: **Coarse-tuning of the Feedforward Neural Network and the Recurrent Neural Network. (a,b)** Schematic illustration of the (a) Feedforward Neural Network and (b) Recurrent Neural Network architectures. **(c,d)** Susceptibility in the T/J–J_0/J parameter space for (c) $\Delta = 0.5$ and (d) $\Delta = 1$. The solid black line marks the phase transition. **(e,f)** Magnetization heat maps for (e) $\Delta = 0.5$ and (f) $\Delta = 1$. The dashed black line indicates the phase transition for the corresponding $h_0 = 0$ case. **(g,h)** Projection of the signal onto (g) the uniform direction and (h) the random local field.

of the signal-to-noise ratio (SNR) with population size for two readout algorithms: a naïve readout and an optimal readout. We found that the SNR of the naïve readout was largely insensitive to coarse-tuning, since its performance is already limited by noise correlations in the neural responses. In contrast, the SNR of the optimal readout exhibited distinct scaling regimes: it grew linearly with population size for weak coarse-tuning, crossed over to sublinear growth at intermediate coarse-tuning, and ultimately saturated for strong coarse-tuning.

Consequently, in the strong coarse-tuning regime, the SNR of both readout algorithms converged to finite limits, albeit for different underlying reasons, for large N. The asymptotic SNR of the optimal readout was governed by the magnitude of tuning fluctuations, quantified by κ, whereas the asymptotic performance of the naïve readout was dominated by noise correlations in the population activity. Finally, we showed that this fundamental limitation cannot be alleviated by adding further processing layers or by introducing recurrent connectivity.

How can our theory account for the high levels of accuracy observed in neural systems? One possibility is that the central nervous system operates in the weak to moderate coarse-tuning regimes, in which arbitrarily high accuracy can, in principle, be attained by an optimal readout given a sufficiently large population. However, we showed that in these regimes, the intrinsic heterogeneity of neuronal responses vanishes. This conflicts with extensive experimental evidence of persistent response heterogeneity [50, 51], which cast doubt on the hypothesis that neural circuits operate predominantly in the weak to moderate coarse-tuning regimes.

Alternatively, high decoding accuracy may be achieved in the strong coarse-tuning regime provided that the asymptotic SNR is sufficiently large. In this regime, the saturation of the SNR implies that pooling information beyond an effective population size, N_{eff}, yields little additional benefits. Hence, the readout is robust to neuronal loss as long as the remaining population exceeds N_{eff} neurons.

Synaptic weights are inherently volatile, even in the absence of neuronal activity [26–28]. As a result, noise in synaptic strengths is expected to accumulate over time. A growing body of experimental evidence indicates that the response statistics of the representation layer can also evolve over time, a phenomenon commonly referred to as *representational drift* [52–56]. Together, these sources of variability are expected to degrade the performance of any fixed readout over time. Ongoing learning can partially counteract this degradation by continuously adjusting synaptic weights toward their target values. The interplay between stochastic synaptic fluctuations and corrective learning dynamics should thus produce a steady-state distribution of synaptic weights centered around their fine-tuned values. In the present work, we did not explicitly model the stochastic dynamics of synaptic plasticity; instead, quenched disorder was used to represent this steady-state distribution. A detailed investigation of the coupled dynamics of synaptic volatility and learning is beyond the scope of this study and is left for future work. Nevertheless, we believe that the present framework provides a foundational step in this direction.

With respect to learning, it is important to note that the naïve and optimal readouts make markedly different assumptions concerning the capabilities of the central nervous system. Learning the optimal readout generally requires a supervised learning mechanism, whereas the naïve readout can be acquired through comparatively simple forms of plasticity, such as homeostatic regulation. In particular, in our model, any homeostatic plasticity rule that induces independent fluctuations of synaptic weights around a common nonzero mean is sufficient to implement the naïve readout.

It is often assumed that the central nervous system operates near optimality [57]. However, because the performance of the naïve readout is comparable to that of the optimal readout and because the naïve readout can be implemented through substantially simpler learning mechanisms, the naïve readout remains a viable alternative. This reasoning is supported by two other potentially counterintuitive considerations.

The first is the invariant-manifold hypothesis which posits that the brain maintains functional stability despite ongoing synaptic volatility because the set of synaptic configurations that support a given computation forms a manifold (or low-dimensional subspace). Synaptic fluctuations are therefore substantial along this manifold, because they are strongly suppressed in directions orthogonal to it, thereby preserving function [33, 36, 53, 58–63]. In the context of our work, functionality was measured as the ability to extract the signal from the neuronal responses and estimate the stimulus. The signal was embedded by the response selectivity vector, $\mathbf{g}$ defined in Eq. 5. This vector can be decomposed into a fixed uniform component and zero-mean quenched fluctuations. The uniform direction, corresponding to the naïve readout, therefore spans the invariant manifold.

The second consideration arises from comparisons between theoretical readout performance and psychophysical measurements. Previous work has suggested that the naïve readout may be more consistent with empirical observations than the optimal readout [14]. Mendelson and Shamir [64] estimated the asymptotic accuracy of a naïve decoder (specifically, the population vector for angular estimation) and found performance levels comparable to those observed in psycho-physical experiments [65, 66]. Taken together, these observations raise the possibility that naïve readout mechanisms may play a role in neural decoding.

ACKNOWLEDGMENTS

This work was supported by the Israel Science Foundation (ISF) under Grant No. 624/22 and Grant

No. 824/21. The authors declare no competing interests.

Appendix A: SNR^2 of the fine-tuned naïve decoder

Substituting the naïve choice of weights, $\boldsymbol{w} = \frac{\mathbf{1}}{N}$, into Eq. 9, yields

$$signal = \boldsymbol{w}^\top \boldsymbol{g} = \frac{1}{N}\sum_{i=1}^{N} g_i, \tag{A1}$$

$$noise^2 = 2\boldsymbol{w}^\top \mathbf{C}\boldsymbol{w} = \frac{2}{N^2}\sum_{i,j} C_{ij} = \frac{2aN[1+(N-1)c]}{N^2}, \tag{A2}$$

where we use in Eq. A2 the correlation matrix of Eq. 2. The *signal* is a random variable because it depends on the specific realization of the neuronal response selectivity, $\{g_i\}$. Since the g_i's are i.i.d. random variables the *signal* is a self-averaging quantity with $signal \sim \mathcal{N}\Big(\mu_g + \frac{\sigma_g^2}{N},\ \frac{\sigma_g^2}{N}\Big)$; thus, in the limit of large N we can ignore the quenched variability of the *signal* and take it to be deterministic with $signal = \mu_g$. From Eq. 2, the $noise^2$ has no quenched fluctuations and is $\mathcal{O}(N^0)$. Thus, the SNR^2 of the naïve readout is

$$\ll SNR^2(\mathbf{w}^{\text{naïve}}) \gg \approx \frac{N\mu_g^2}{2a[1+(N-1)c]} \tag{A3}$$

Appendix B: SNR^2 of fine-tuned Optimal decoder

Substituting the optimal weights, $\boldsymbol{w}^{\text{opt}} = \mathbf{C}^{-1}\boldsymbol{g}$, into Eq. 9, yields

$$SNR^2(\mathbf{w}^{\text{opt}}) = \frac{\boldsymbol{g}^\top \mathbf{C}^{-1}\boldsymbol{g}}{2}. \tag{B1}$$

For our choice of correlation structure, Eq. A2, its inverse is given by

$$\mathbf{C}^{-1} = \frac{1}{a(1-c)}\mathbf{I} - \frac{c}{a(1-c)\big(1-c+Nc\big)}\mathbf{1}, \tag{B2}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is an all ones matrix. Thus, the SNR^2 is a quenched random variable:

$$SNR^2(\mathbf{w}^{\text{opt}}) = \frac{1}{2a(1-c)}\sum_i g_i^2 - \frac{c\left(\sum_i g_i\right)^2}{a(1-c)\big(1-c+Nc\big)}, \tag{B3}$$

with mean

$$\ll SNR^2(w^{\text{opt}}) \gg = \frac{\Big(\big(1+(N-2)c\big)\sigma_g^2 + \big(1-c\big)\mu_g^2\Big)}{2a(1-c)\Big(\frac{1+c(N-1)}{N}\Big)}. \tag{B4}$$

In the limit of large N, $SNR^2(w^{opt}) \approx A_1 + A_2$, where $A_1 = \frac{N}{2a(1-c)}\bar{g}^2$, $A_2 = \frac{N}{2a(1-c)}\bar{g}^2$, and $\bar{x} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i$ denotes spatial averaging of x. As both A_1 and A_2 are self averaging, so is $SNR^2(w^{opt})$.

Appendix C: The effect of coarse-tuning on the naïve decoder

The *signal*, s, of the coarse-tuned naïve decoder, $s = \bar{g} + \boldsymbol{\xi}^\top \boldsymbol{g}$, is a random variable with quenched mean of

$$\ll s \gg = \mu_g. \tag{C1}$$

Denoting the quenched fluctuations of the *signal* by $\Delta s = s - \ll s \gg$, the quenched variance of the *signal*, $\ll (\Delta s)^2 \gg$ is given by

$$\ll (\Delta s)^2 \gg = \frac{\sigma_g^2}{N} + \kappa^2 N^\gamma(\mu_g^2 + \sigma_g^2). \tag{C2}$$

The squared *noise*, n^2, of the naïve decoder is now a random variable

$$n^2 = 2\boldsymbol{w}^\top \mathbf{C}\boldsymbol{w} = \frac{2}{N^2}\sum_{i,j} C_{ij} + 2I_1 + 2I_2, \tag{C3}$$

with

$$I_1 = \boldsymbol{\xi}^\top \mathbf{C}\boldsymbol{\xi}, \qquad I_2 = \frac{2}{N}\sum_{i,j} C_{ij}\,\xi_j. \tag{C4}$$

The quenched mean of the squared *noise* is given by

$$\ll n^2 \gg = \frac{2a\,[1+(N-1)c]}{N} + 2a\,\kappa^2 N^\gamma. \tag{C5}$$

The quenched fluctuations of the squared *noise*, $\Delta n^2 = n^2 - \ll n^2 \gg$, can be written as the sum of two terms: $\Delta n^2 = \Delta I_1 + I_2$, where

$$\begin{aligned}\ll (\Delta I_1)^2 \gg &= Na^2 \ll (\xi_i^2 - \ll \xi_i^2 \gg)^2 \gg \\ &\quad + 2N(N-1)(ac)^2 \ll \sum_{i\neq j}\sum_{i'\neq j'} \xi_i\xi_j\xi_{i'}\xi_{j'} \gg \end{aligned} \tag{C6}$$

$$= 2a^2\kappa^4 N^{2\gamma-1}\left(1+(N-1)c^2\right), \tag{C7}$$

$$\ll (\Delta I_2)^2 \gg = \frac{4}{N^2}\sum_{i,j,i',j'} C_{ij}C_{i'j'} \ll \xi_j\xi_{j'} \gg \approx 4(ac)^2\kappa^2 N^\gamma. \tag{C8}$$

Thus, the quenched variance of n^2 is

$$\ll (\Delta n^2)^2 \gg \approx 8a^2\kappa^4 N^{2\gamma-1}(1+c^2(N-1)) + 16(a\kappa c)^2 N^\gamma. \tag{C9}$$

Appendix D: Optimal readout under coarse-tuning

The quenched mean of the *signal* of the coarse-tuned optimal readout is

$$\ll s \gg = \sigma_g. \tag{D1}$$

The quenched fluctuations of the *signal* are given by:

$$\Delta s \equiv s - \ll s \gg = \frac{\boldsymbol{g}^\top(\boldsymbol{g} - \bar{\boldsymbol{g}})}{N\sigma_g} - \sigma_g + \sum_i \xi_i g_i. \tag{D2}$$

where, assuming large N, we approximated $\mathbf{w}^{\rm opt} \propto \mathbf{g}-\bar{\mathbf{g}}$.

The quenched variance reads:

$$\begin{aligned} \ll (\Delta s)^2 \gg &= \frac{1}{N^2\sigma_g^2} \ll \left((\boldsymbol{g}-\bar{\boldsymbol{g}})^\top \boldsymbol{g} - N\sigma_g^2\right)^2 \gg \\ &+ \ll \Big(\sum_{i=1}^N \xi_i g_i\Big)^2 \gg . \end{aligned} \tag{D3}$$

The contribution of the first term on the right hand side of Eq. D3 can be neglected for large N. Thus, the quenched variance of the *signal* is

$$\ll (\Delta s)^2 \gg = \kappa^2 N^\gamma (\mu_g^2 + \sigma_g^2). \tag{D4}$$

The squared *noise* is given by

$$n^2 = 2\boldsymbol{w}^\top \mathbf{C}\boldsymbol{w} = I_3 + I_4 + I_5 \tag{D5}$$

$$I_3 = \frac{2}{N^2\sigma_g^2}(\boldsymbol{g}-\bar{\boldsymbol{g}})^\top \mathbf{C}(\boldsymbol{g}-\bar{\boldsymbol{g}}) \tag{D6}$$

$$I_4 = 2\boldsymbol{\xi}^\top \mathbf{C}\boldsymbol{\xi} \tag{D7}$$

$$I_5 = \frac{4}{N\sigma_g}(\boldsymbol{g}-\bar{\boldsymbol{g}})^\top \mathbf{C}\boldsymbol{\xi} \tag{D8}$$

and its quenched mean is

$$\begin{aligned} \ll n^2 \gg &= \frac{2}{N^2\sigma_g^2} \ll (\boldsymbol{g}-\bar{\boldsymbol{g}})^\top \mathbf{C}(\boldsymbol{g}-\bar{\boldsymbol{g}}) \gg \\ &+ 2 \ll \boldsymbol{\xi}^\top \mathbf{C}\boldsymbol{\xi} \gg \\ &= \frac{2a(1-c)}{N} + 2a\kappa^2 N^\gamma. \end{aligned} \tag{D9}$$

To compute the quenched variance of n^2, note that I_3 is the *noise* term of the fine-tuned optimal readout; hence, is self-averaging (see B). Now

$$\ll (\Delta I_4)^2 \gg = 8a^2\kappa^4 N^{2\gamma-1}\left(1 + c^2(N-1)\right), \tag{D10}$$

and

$$\begin{aligned} \ll (\Delta I_5)^2 \gg &= \frac{16a^2(1-c)^2}{N^2\sigma_g^2} \ll \left(\boldsymbol{\xi}^\top(\boldsymbol{g}-\bar{\boldsymbol{g}})\right)^2 \gg \\ &= 16a^2(1-c)^2\kappa^2 N^{\gamma-2}. \end{aligned} \tag{D11}$$

where we use in Eq. D11 the fact the $(\mathbf{g}-\bar{\mathbf{g}})$ is an eigenvector of $\mathbf{C}$ with eigenvalue of $a(1-c)$. Thus, the quenched variance of n^2 is given by

$$\begin{aligned} \ll (\Delta n^2)^2 \gg \approx & 8a^2\kappa^4 N^{2\gamma-1}\left(1 + c^2(N-1)\right) \\ &+ 16(a\kappa)^2(1-c)^2 N^{\gamma-2}. \end{aligned} \tag{D12}$$

Appendix E: Feedforward neural network

We consider a processing layer of N neurons with local fields

$$h_i = \sum_{j=1}^N \tilde{w}_{ij}(\mathbf{r}^t - \mathbf{r}^d) \tag{E1}$$

where,

$$\begin{aligned} \tilde{w}_{ij} &= \frac{w_{ij}}{\sqrt{N}\,\|\mathbf{w}\|} + \xi_{ij}, \\ \ll \xi_{ij} \gg &= 0, \qquad \ll \xi_{ij}\xi_{i'j'} \gg = \kappa^2 N^{\gamma-1}\delta_{ii'}\delta_{jj'}. \end{aligned} \tag{E2}$$

The mean of the local field is

$$[h_i] = \frac{\ll \mathbf{w}^\top \mathbf{g} \gg}{\sqrt{N}\|\mathbf{w}\|} = \frac{\ll (\mathbf{g}-\bar{\mathbf{g}})^\top \mathbf{g} \gg}{\sqrt{N}\|\mathbf{w}\|} \equiv h_0, \tag{E3}$$

where $[\cdot]$ denotes averaging with respect to both the trial-to-trial variability and the quenched disorder. The covariance of the local fields is given by

$$\begin{aligned} \mathrm{cov}(h_i, h_k) &= [h_i h_k] - h_0^2 \\ &= \big[(h_i - \langle h_i\rangle)(h_k - \langle h_k\rangle)\big] \\ &+ \big[(\langle h_i\rangle - [h_i])(\langle h_k\rangle - [h_k])\big]. \end{aligned} \tag{E4}$$

The first term on the right hand side of Eq. E4 yields

$$\begin{aligned} [(h_i - \langle h_i\rangle)(h_k - \langle h_k\rangle)] &= [\mathbf{w}C\mathbf{w}] + aN^\gamma\kappa^2\delta_{ik} \\ &\approx \frac{a(1-c)}{N} + aN^\gamma\kappa^2\delta_{ik}, \end{aligned} \tag{E5}$$

where for large N we approximated $\mathbf{w}C\mathbf{w} \approx \frac{1}{N^2\sigma_g^2}(\mathbf{g}-\bar{\mathbf{g}})^\top C(\mathbf{g}-\bar{\mathbf{g}}) = \frac{a(1-c)}{N}$. The second term yields

$$[(\langle h_i\rangle - [h_i])(\langle h_k\rangle - [h_k])] = N^\gamma\kappa^2(\mu_g^2 + \sigma_g^2)\delta_{ik}. \tag{E6}$$

Thus, we obtain

$$\mathrm{cov}(h_i, h_k) = \begin{cases} \dfrac{a(1-c)}{N} + N^\gamma\kappa^2(a + \mu_g^2 + \sigma_g^2), & i = k, \\ \dfrac{a(1-c)}{N}, & i \neq k. \end{cases} \tag{E7}$$

Appendix F: Sherrington-Kirkpatrick model

The Recurrent Neural Network model of section VII B naturally maps to the Sherrington-Kirkpatrick model with random fields. This model has been extensively studied; see e.g., [42, 46, 47, 67]. For completeness we summarize the central results that are relevant to the current study.

We consider a system of N Ising spins, $\{s_i\}_{i=1}^N$, governed by the Hamiltonian in Eq. 23 at thermal equilibrium with inverse temperature β. The fields, $\{h_i\}$, are i.i.d. quenched Gaussian random variables, with the probability density

$$\Pr(h_i) = \frac{1}{\sqrt{2\pi}\Delta}\exp\left(-\frac{(h_i - h_0)^2}{2\Delta^2}\right). \tag{F1}$$

The recurrent connections, $\{J_{i,j}\}$, are i.i.d. quenched Gaussian random variables with

$$\Pr(J_{i,j}) = \sqrt{\frac{N}{2\pi J^2}}\exp\left(-\frac{N(J_{i,j} - \frac{J_0}{N})^2}{2J^2}\right). \tag{F2}$$

To appreciate the effect of recurrent connectivity it is insightful to first consider the case of $J = 0$. At thermal equilibrium the local magnetization, $m_i = \langle s_i \rangle$, is given by:

$$m_i = \tanh\big(\beta(J_0 m + h_i)\big), \tag{F3}$$

where m is the mean magnetization, $m = \frac{1}{N}\sum_i m_i$, which serves as an order parameter of the system. Thus, the self-consistent equation for m is

$$m = \int \mathcal{D}z \tanh\big(\beta(J_0 m + h_0 + \Delta z)\big), \tag{F4}$$

where $\mathcal{D}z \equiv \frac{dz}{\sqrt{2\pi}} \exp\frac{-z^2}{2}$ is the standard Gaussian measure. In the absence of external fields, $h_0 = 0$ and $\Delta = 0$, the paramagnetic state, $m = 0$, is always a solution of Eq. (F4). When the recurrent connectivity, J_0, is stronger than some critical value, $J_0^{\text{crit}} = \frac{1}{\beta}$, the paramagnetic solution loses its stability and the system settles into one of the two solutions of the ferromagnetic phase. In this case, even a small ordering field, $|h_0| \ll 1$, will choose the sign of the ferromagnetic solution. Thus, in this case, the recurrent connectivity offers amplification of its input.

To understand the effect of random fields, $\Delta > 0$, we first examine the case $J = 0$ and $h_0 = 0$. In this case, due to the symmetry of Eq. (F4), a solution with $m = 0$ always exists. However, since $\Delta > 0$ the local magnetization cannot be zero. Thus, the system is characterized by an additional order parameter:

$$q = \int \mathcal{D}z \tanh^2\big(\beta(J_0 m + h_0 + \Delta z)\big), \tag{F5}$$

the Edward Anderson order parameter [67].

To find the the ferromagnetic transition (see [49]), one sets the mean of the external field to zero, $h_0 = 0$, and, assuming a second order phase transition, expands Eq. F4 in small m. Using

$$\begin{aligned}\tanh(x+y) = \tanh x &\\ &+ y(1-\tanh^2 x)\\ &- y^2 \tanh x(1-\tanh^2 x)\\ &+ \frac{y^3}{3}\big((1-\tanh^2 x)(3\tanh^2 x - 1)\big)\\ &- \frac{y^4}{3}\big(\tanh x\big(1-\tanh^2 x\big)\big(3\tanh^2 x - 2\big)\big)\\ &+ y^5(-\tanh^6 x + 2\tanh^4 x - \frac{17}{15}\tanh^2 x\\ &+ \frac{2}{15}) + O(y^6),\end{aligned} \tag{F6}$$

one obtains

$$\begin{aligned} m = (\beta J_0 m) &\int \mathcal{D}z\,\big(1-\tanh^2[\beta\Delta z]\big)\\ &+ \frac{(\beta J_0 m)^3}{3}\int \mathcal{D}z\,\Big(1-\tanh^2[\beta\Delta z]\Big)\Big(3\tanh^2[\beta\Delta z]-1\Big)\\ &+ (\beta J_0 m)^5 \int \mathcal{D}z\,\Big(-\tanh^6[\beta\Delta z] + 2\tanh^4[\beta\Delta z]\\ &- \tfrac{17}{15}\tanh^2[\beta\Delta z] + \tfrac{2}{15}\Big)\\ &+ O(m^7). \end{aligned} \tag{F7}$$

Equivalently, denoting $q_{2k} \equiv \int \mathcal{D}z \tanh^{2k}(\beta\Delta z)$,

$$\begin{aligned} m = \beta J_0(1-q_2)m + (\beta J_0)^3\left(-q_4 + \tfrac{4}{3}q_2 - \tfrac{1}{3}\right)m^3\\ + (\beta J_0)^5\left(-q_6 + 2q_4 - \tfrac{17}{15}q_2 + \tfrac{2}{15}\right)m^5 + O(m^7). \end{aligned} \tag{F8}$$

A second-order phase transition to the ferromagnetic phase occurs when the coefficient of the linear term in m in Eq. F8 changes sign

$$\beta_c J_{0,c}\big(1 - q(\beta_c, J_{0,c})\big) = 1. \tag{F9}$$

Note that under the assumption of Gaussian random fields, the coefficient of the cubic term in m in Eq. F8 is always positive; see [68]. Accordingly, a transition to the ferromagnetic phase can always be induced for sufficiently large J_0. However, in the presence of random fields ($\Delta > 0$), the transition is shifted to larger values of the recurrent coupling J_0. Thus, fluctuations in the local fields ($\Delta > 0$) attenuate the amplification of the ordering field h_0 by the recurrent interactions.

In the general case of strong quenched disorder in the recurrent connections, $J > 0$, one can derive equations for the local magnetization that generalizes Eq. (F3) (see Appendix G):

$$m_i = \tanh\Big(\beta \sum_{j(\neq i)} J_{ij} m_j - \beta^2 J^2 (1-q) m_i + \beta h_i\Big). \tag{F10}$$

Within our chosen parameter regime, $\beta J < 1$, the two order parameters m and q suffice to characterize the system [67]. Averaging Eq. (F10) over space and replacing the spatial average with averaging over the quenched disorder yields the self-consistent equations for the order parameters:

$$m = \int \mathcal{D}z \tanh\Big(\beta(J_0 m + h_0 + \sqrt{\Delta^2 + J^2 q}\,z)\Big), \tag{F11}$$

$$q = \int \mathcal{D}z \tanh^2\big(\beta\big(J_0 m + h_0 + \sqrt{\Delta^2 + J^2 q}\,z\big)\big). \tag{F12}$$

Similar to random fields ($\Delta > 0$), the coupling of quenched disorder in the recurrent connections ($J > 0$) with nonzero q, suppresses amplification of the ordering field, h_0. As a result, the transition to the ferromagnetic phase is shifted to larger values of the ferromagnetic bias

J_0. Nevertheless, for sufficiently large J_0, the recurrent interactions still provide amplification of the signal, h_0.

We further map the Almeida-Thouless (AT) line [48], writing

$$\mathcal{A} = (\beta J)^2 \int \mathcal{D}z \mathrm{sech}^4\big(\beta\big(J_0 m + h_0 + \sqrt{\Delta^2 + J^2 q}\, z\big)\big). \tag{F13}$$

For $\mathcal{A} < 1$, the two order parameters in (F11) and (F12) fully characterize the system. The AT boundary is given by $\mathcal{A} = 1$; beyond it ($\mathcal{A} > 1$) the solution above is insufficient and additional techniques need to be utilized to characterize the system such as replica symmetry breaking [69–71].

Appendix G: TAP equation

Here we derive TAP equations [72] for the Sherrington-Kirkpatrick model with local random fields using the cavity method [73]. The derivation presented below extends the method of Ref. [74] to include random local fields. The extension is straightforward and is presented here for completeness. The derivation proceeds in three steps.

Step 1: adding a spin

We first extend the N-spin system by adding a single spin, s_0, at site 0, which interacts with the existing spins via couplings $\{J_{0j}\}_{j=1}^N$ drawn from the distribution in Eq. F2. The Hamiltonian of the extended $(N+1)$-spin system is given by

$$\mathcal{H}^{(N+1)} = \mathcal{H}^{(N)} - h_0 s_0, \tag{G1}$$

$$h_0 = \sum_{j=1}^N J_{0j}\, s_j + h_0^{ext}. \tag{G2}$$

where h_0 is the local field at site 0. The states of the extended system are distributed according to the Gibbs distribution with Hamiltonian $\mathcal{H}^{(N+1)}$:

$$P_{N+1}\left(\{s_i\}_{i=0}^N\right) = \frac{1}{Z_{N+1}} \exp\big[-\beta \mathcal{H}^{(N+1)}\big], \tag{G3}$$

$$Z_{N+1} = \mathrm{Tr}_{\{s_i\}_{i=0}^N} \exp\big[-\beta \mathcal{H}^{(N+1)}\big]. \tag{G4}$$

From Eq. G3 we obtain the marginal joint distribution of the cavity spin, s_0, and the cavity field, h_0,

$$P_{N+1}(h_0, s_0) = \frac{1}{Z_{N+1}} \mathrm{Tr}_{\{s_i\}_{i=1}^N} \Bigg\{ \delta\Big(h_0 - h_0^{\mathrm{ext}} - \sum_{j=1}^N J_{0j} s_j\Big) \times \exp\big[-\beta \mathcal{H}^{(N+1)}\big] \Bigg\}. \tag{G5}$$

Using Eq. G1 we can write

$$P_{N+1}(h_0, s_0) = \frac{1}{z} \exp\big[\beta h_0 s_0\big] P_N(h_0), \tag{G6}$$

where

$$P_N(h_0) = \frac{1}{Z_N} \mathrm{Tr}_{\{s_i\}_{i=1}^N} \Bigg\{ \delta\Big(h_0 - h_0^{\mathrm{ext}} - \sum_{j=1}^N J_{0j} s_j\Big) \times \exp\big[-\beta \mathcal{H}^{(N)}\big] \Bigg\}, \tag{G7}$$

and the normalization, z, is given by

$$z = \frac{Z_{N+1}}{Z_N} = \big\langle 2\cosh(\beta h_0)\big\rangle_N. \tag{G8}$$

with $\langle\cdot\rangle_N$ denoting thermal averaging with respect to the N-spin system.

Averaging s_0 and h_0 with $P_{N+1}(h_0, s_0)$ gives the standard identities

$$\langle s_0\rangle_{N+1} = \frac{\langle \sinh(\beta h_0)\rangle_N}{\langle \cosh(\beta h_0)\rangle_N}, \tag{G9}$$

$$\langle h_0\rangle_{N+1} = \frac{\langle h_0 \cosh(\beta h_0)\rangle_N}{\langle \cosh(\beta h_0)\rangle_N}. \tag{G10}$$

Step 2: statistics of the cavity field h_0

Next, we characterize the first two moments of the cavity field, h_0, in the N-spin system:

$$\langle h_0\rangle_N = \sum_{j=1}^N J_{0j}\, \langle s_j\rangle_N, \tag{G11}$$

$$\langle (\delta h_0)^2\rangle_N = \sum_{i,j=1}^N J_{0i} J_{0j}\, \langle \delta s_i\, \delta s_j\rangle_N = \sum_{i=1}^N J_{0i}^2\, \langle (\delta s_i)^2\rangle_N, \tag{G12}$$

with $\delta X = X - \langle X\rangle_N$. In the last equality of Eq. G12 we used the fact that $\langle \delta s_i \delta s_j\rangle_N$ is of order $\mathcal{O}(1/\sqrt{N})$ and is independent of J_{0j}, to approximate the sum over i and j only by the contribution of the $i = j$ terms.

Defining $q_N = \frac{1}{N}\sum_{i=1}^N \langle s_i\rangle_N^2$, and using the self-averaging of the right-hand side of Eq. G12 we obtain $\langle (\delta h_0)^2\rangle_N = J^2(1 - q_N)$. Assuming the distribution of the cavity field, h_0, in the N spin system is well-approximated by Gaussian distribution as the sum of N independent variables. Further assuming that $q_N \to q$ as $N \to \infty$, the distribution of the cavity field is given by:

$$P_N(h_0) = \frac{1}{\sqrt{2\pi J^2(1-q)}} \exp\left(-\frac{\big(h_0 - h_0^{\mathrm{ext}} - \langle h_0\rangle_N\big)^2}{2J^2(1-q)}\right). \tag{G13}$$

Step 3: obtaining the TAP equation

Substituting (G13) into the Eqs. G9 & G10 yields

$$\langle s_0\rangle_{N+1} = \tanh\big(\beta\langle h_0\rangle_N\big), \tag{G14}$$

$$\langle h_0\rangle_{N+1} = \langle h_0\rangle_N + \beta J^2(1-q)\langle s_0\rangle_{N+1}. \tag{G15}$$

Substituting $\langle h_0\rangle_N$ from Eq. G15 into Eq. G14 and replacing site 0 with a generic site i, one obtains the TAP equations:

$$m_i = \tanh\Big(\beta\sum_{j(\neq i)} J_{ij}m_j - \beta^2J^2(1-q)m_i + \beta h_i^{\text{ext}}\Big). \tag{G16}$$

Note that, in this case h_i^{ext} also exhibits quenched statistics.

The susceptibility matrix

The susceptibility matrix is defined by $\chi_{ik} \equiv \partial m_i/\partial h_k$. The diagonal element of the susceptibility matrix can be obtained using the fluctuation-dissipation theorem, $\chi = \beta\mathbf{C}$, where $\mathbf{C}$ is the correlation matrix $C_{ij} = \langle\delta s_i\delta s_j\rangle$. Since $s_i \in \{\pm1\}$, $C_{ii} = \langle s_i^2\rangle - \langle s_i\rangle^2 = 1 - m_i^2$. Thus,

$$\chi_{ii} = \beta(1-m_i^2). \tag{G17}$$

For the off-diagonal elements, we differentiate Eq. G16 with respect to the local field at site k, yielding

$$\chi_{ik} \equiv \partial m_i/\partial h_k = (1-m_i^2)\beta\Big[\delta_{ik}+\sum_j J_{ij}\chi_{jk}-\beta J^2(1-q)\chi_{ik}\Big]. \tag{G18}$$

Defining the matrix $\mathbf{D}$ to be

$$D_{ij} = \delta_{ij}\beta(1-m_i^2), \tag{G19}$$

we obtain

$$\chi_{ij} = \Big(\mathbf{D} + \beta J^2(1-q^2)\mathbf{I} - \mathbf{J}\Big)^{-1}_{ij} \qquad i \neq j. \tag{G20}$$

Figures 1-3: Neural population readout

Table II summarizes the parameter values used for the results in figures 1-3. Parameters marked with an asterisk (*) vary across figure panels and are specified explicitly in the corresponding captions.

We performed 500 realizations of the quenched disorder. In each realization, we sampled $\mu_t^i \sim \mathcal{N}(\mu_t, \sigma_t^2)$ and $\mu_d^i \sim \mathcal{N}(\mu_d, \sigma_d^2)$ for each neuron i, yielding the selectivity $g_i = \mu_t^i - \mu_d^i$. Coarse-tuning was drawn as $\xi_i \sim \mathcal{N}(0, \kappa^2 N^{\gamma+1})$. For each realization, we constructed the readout weights for the naïve and optimal decoders and computed the SNR. Finally we averaged the SNR over all realizations. The simulation results of the SNR (markers) were compared to the analytical expressions of the SNR (solid lines) derived in the main text.

TABLE II: Parameters for the neural population readout model (Figs. 1–3).

Parameter	Description	Value
N	Number of Neurons	*
a	Single-neuron variance	12
c	Pairwise correlation coefficient	*
μ_t, μ_d	Quenched mean response (t,d)	12, 9
μ_g	Quenched mean of response selectivity	3
σ_g^2	Quenched variance of response selectivity	24
κ	Coarse-tuning magnitude	*
γ	Coarse-tuning scaling	*

Figure 4: Feedforward Neural Network and Recurrent Neural Network simulations

Table III summarizes the parameter values used for the Feedforward Neural Network and Recurrent Neural Network simulations. Parameters marked with an asterisk (*) vary across figure panels and are specified explicitly in the corresponding captions. The coupling matrix and local fields were constructed according to Eqs. (22) and (24), respectively.

TABLE III: Parameters for the Feedforward Neural Network and Recurrent Neural Network simulations (Fig. 4).

Parameter	Description	Value
N	Number of neurons in processing layer	200
J	Disorder in the recurrent connections	1
h_0	Signal	*
Δ	Disorder in the random fields	*
T	Temperature	[1.5, 3.5]
J_0	Ferromagnetic bias	[0, 4]

The TAP equations, Eq. (G16), were solved using damped fixed-point iteration with mixing parameter $\alpha = 0.35$. Convergence was declared when $\max_i |m_i^{(t+1)} - m_i^{(t)}| < 10^{-10}$, with a maximum of 4000 iterations.

The susceptibility matrix χ, Eq. (G18), was computed from the converged TAP solution. The dominant eigenvector of χ, $\mathbf{v}_{\max}$ was obtained using MATLAB's eigs function with tolerance 10^{-10}; if this fails to converge, a full eigendecomposition was performed instead.

To analyze the structure of the dominant mode, Fig. 4g & 4h, we projected the unit-normalized eigenvector $\hat{\mathbf{v}} = \mathbf{v}_{\max}/\|\mathbf{v}_{\max}\|$ onto the subspace spanned by $\hat{\mathbf{u}} = \mathbf{1}/\|\mathbf{1}\|$ (uniform direction) and $\hat{\mathbf{h}} = \mathbf{h}/\|\mathbf{h}\|$ (normalized field). The projection coefficients (c_u, c_h) were obtained by solving

$$\begin{pmatrix} 1 & \hat{\mathbf{u}}\cdot\hat{\mathbf{h}} \\ \hat{\mathbf{u}}\cdot\hat{\mathbf{h}} & 1 \end{pmatrix}\begin{pmatrix} c_u \\ c_h \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{u}}\cdot\hat{\mathbf{v}} \\ \hat{\mathbf{h}}\cdot\hat{\mathbf{v}} \end{pmatrix}. \tag{G21}$$

The residual $\boldsymbol{\delta} = \hat{\mathbf{v}} - c_u\hat{\mathbf{u}} - c_h\hat{\mathbf{h}}$ is orthogonal to both basis vectors.

The phase transition for $h_0 = 0$, denoted in Fig. 4c-f by a black line, was computed as follows. For each temperature, we first solved Eq. (F12) for the Edwards-Anderson parameter q. The critical J_0/J is then given by

$$\frac{J_0}{J} = \frac{T/J}{\langle \mathrm{sech}^2(\beta J\sqrt{q + (\Delta/J)^2}\, z)\rangle_z}, \tag{G22}$$

where $z \sim \mathcal{N}(0,1)$ and the expectation was evaluated using Gauss-Hermite quadrature with 80 nodes. All heat maps were computed on a 50×50 grid for a single disorder realization.

[1] M. N. Shadlen and W. T. Newsome, Current opinion in neurobiology **4**, 569 (1994).
[2] L. F. Abbott, Quarterly reviews of biophysics **27**, 291 (1994).
[3] E. Salinas and L. F. Abbott, Journal of Neuroscience **15**, 6461 (1995).
[4] H. S. Seung and H. Sompolinsky, Proceedings of the national academy of sciences **90**, 10749 (1993).
[5] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner, Science **233**, 1416 (1986).
[6] E. Salinas and L. Abbott, Journal of computational neuroscience **1**, 89 (1994).
[7] E. Zohary, M. N. Shadlen, and W. T. Newsome, Nature **370**, 140 (1994).
[8] M. N. Shadlen, K. H. Britten, W. T. Newsome, and J. A. Movshon, Journal of neuroscience **16**, 1486 (1996).
[9] L. F. Abbott and P. Dayan, Neural computation **11**, 91 (1999).
[10] H. Sompolinsky, H. Yoon, K. Kang, and M. Shamir, Physical Review E **64**, 051904 (2001).
[11] O. Hendler, R. Segev, and M. Shamir, PLOS Computational Biology **21**, e1013092 (2025).
[12] M. Shamir and H. Sompolinsky, Neural computation **18**, 1951 (2006).
[13] B. B. Averbeck and D. Lee, Journal of neurophysiology **95**, 3633 (2006).
[14] M. Shamir, Current opinion in neurobiology **25**, 140 (2014).
[15] H. P. Snippe and J. J. Koenderink, Biological cybernetics **67**, 183 (1992).
[16] J. L. Chen, K. L. Villa, J. W. Cha, P. T. So, Y. Kubota, and E. Nedivi, Neuron **74**, 361 (2012).
[17] M. Cane, B. Maco, G. Knott, and A. Holtmaat, Journal of Neuroscience **34**, 2075 (2014).
[18] Y. Loewenstein, A. Kuras, and S. Rumpel, Journal of Neuroscience **31**, 9481 (2011).
[19] T. Pfeiffer, S. Poll, S. Bancelin, J. Angibaud, V. V. K. Inavalli, K. Keppler, M. Mittag, M. Fuhrmann, H. Steffens, B. Schmid, and et al., eLife **7**, e34700 (2018).
[20] Y. Zuo, A. Lin, P. Chang, and W. B. Gan, Neuron **46**, 181 (2005).
[21] J. T. Trachtenberg, B. E. Chen, G. W. Knott, G. Feng, J. R. Sanes, E. Welker, and K. Svoboda, Nature **420**, 788 (2002).
[22] A. Minerbi, R. Kahana, L. Goldfeld, M. Kaufman, S. Marom, and N. E. Ziv, PLoS biology **7**, e1000136 (2009).
[23] L. Hazan and N. E. Ziv, Journal of Neuroscience **40**, 2828 (2020).
[24] N. Ren, G. Wei, A. Ghanbari, and I. H. Stevenson, Journal of Neuroscience **42**, 8608 (2022).
[25] J. Matz, A. Gilyan, A. Kolar, T. McCarvill, and S. R. Krueger, Proceedings of the National Academy of Sciences **107**, 8836 (2010).
[26] N. E. Ziv and N. Brenner, Trends in neurosciences **41**, 89 (2018).
[27] N. Yasumatsu, M. Matsuzaki, T. Miyazaki, J. Noguchi, and H. Kasai, Journal of Neuroscience **28**, 13592 (2008).
[28] L. Hazan and N. E. Ziv, Journal of Neuroscience **40**, 2828 (2020).
[29] J. Kadmon, J. Timcheck, and S. Ganguli, Advances in neural information processing systems **33**, 16677 (2020).
[30] J. Timcheck, J. Kadmon, K. Boahen, and S. Ganguli, PLoS computational biology **18**, e1010593 (2022).
[31] A. Shomar, L. Geyrhofer, N. E. Ziv, and N. Brenner, PLoS computational biology **13**, e1005668 (2017).
[32] A. Statman, M. Kaufman, A. Minerbi, N. E. Ziv, and N. Brenner, PLoS computational biology **10**, e1003846 (2014).
[33] G. Shimizu, K. Yoshida, H. Kasai, and T. Toyoizumi, Current opinion in neurobiology **70**, 34 (2021).
[34] J. J. Hopfield, Proceedings of the National Academy of Sciences **79**, 2554 (1982).
[35] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, Cambridge, UK, 1989).
[36] L. Susman, N. Brenner, and O. Barak, Nature communications **10**, 4441 (2019).
[37] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Physical Review A **32**, 1007 (1985).
[38] H. Sompolinsky, A. Crisanti, and H. J. Sommers, Physical Review Letters **61**, 259 (1988).
[39] J. Barral and A. D Reyes, Nature neuroscience **19**, 1690 (2016).
[40] J.-n. Teramae and T. Fukai, Proceedings of the IEEE **102**, 500 (2014).
[41] A. Roxin, N. Brunel, D. Hansel, G. Mongillo, and C. Van Vreeswijk, Journal of Neuroscience **31**, 16217 (2011).
[42] D. Sherrington and S. Kirkpatrick, Physical review letters **35**, 1792 (1975).
[43] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Physical Review Letters **55**, 1530 (1985).
[44] J. Kadmon and H. Sompolinsky, Physical Review X **5**, 041030 (2015).
[45] J. Aljadeff, M. Stern, and T. Sharpee, Physical Review Letters **114**, 088101 (2015).
[46] R. Soares, F. Nobre, and J. De Almeida, Physical Review B **50**, 6151 (1994).
[47] A. Bray and M. Moore, Journal of Physics C: Solid State

Physics **13**, 419 (1980).
[48] J. R. de Almeida and D. J. Thouless, Journal of Physics A: Mathematical and General **11**, 983 (1978).
[49] T. Schneider and E. Pytte, Physical Review B **15**, 1519 (1977).
[50] L. Lim, D. Mi, A. Llorca, and O. Marín, Neuron **100**, 294 (2018).
[51] D. L. Ringach, R. M. Shapley, and M. J. Hawken, Journal of neuroscience **22**, 5639 (2002).
[52] Y. Ziv, L. D. Burns, E. D. Cocker, E. O. Hamel, K. K. Ghosh, L. J. Kitch, A. E. Gamal, and M. J. Schnitzer, Nature neuroscience **16**, 264 (2013).
[53] L. N. Driscoll, N. L. Pettit, M. Minderer, S. N. Chettih, and C. D. Harvey, Cell **170**, 986 (2017).
[54] D. Deitch, A. Rubin, and Y. Ziv, Current biology **31**, 4327 (2021).
[55] D. Khatib, A. Ratzon, M. Sellevoll, O. Barak, G. Morris, and D. Derdikman, Neuron **111**, 2348 (2023).
[56] N. Geva, D. Deitch, A. Rubin, and Y. Ziv, Neuron **111**, 2357 (2023).
[57] K. Friston, Nature reviews neuroscience **11**, 127 (2010).
[58] M. E. Rule, T. O'Leary, and C. D. Harvey, Current opinion in neurobiology **58**, 141 (2019).
[59] M. E. Rule, A. R. Loback, D. V. Raman, L. N. Driscoll, C. D. Harvey, and T. O'Leary, elife **9**, e51121 (2020).
[60] A. Dubreuil, A. Valente, M. Beiran, F. Mastrogiuseppe, and S. Ostojic, Nature neuroscience **25**, 783 (2022).
[61] L. N. Driscoll, L. Duncker, and C. D. Harvey, Current opinion in neurobiology **76**, 102609 (2022).
[62] N. Sherf and M. Shamir, Physical Review Research **7**, 013113 (2025).
[63] G. Socolovsky and M. Shamir, bioRxiv , 2025 (2025).
[64] O. P. Mendels and M. Shamir, Frontiers in Computational Neuroscience **12**, 12 (2018).
[65] D. Andrews, Vision Research **7**, 975 (1967).
[66] M. Dick and S. Hochstein, Perception & Psychophysics **46**, 227 (1989).
[67] S. F. Edwards and P. W. Anderson, Journal of Physics F: Metal Physics **5**, 965 (1975).
[68] A. Aharony, Physical Review B **18**, 3318 (1978).
[69] A. Bray and M. Moore, Journal of Physics C: Solid State Physics **12**, 79 (1979).
[70] A. Blandin, M. Gabay, and T. Garel, Journal of Physics C: Solid State Physics **13**, 403 (1980).
[71] G. Parisi, Physical Review Letters **43**, 1754 (1979).
[72] D. J. Thouless, P. W. Anderson, and R. G. Palmer, Philosophical Magazine **35**, 593 (1977).
[73] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, Vol. 9 (World Scientific Publishing Company, 1987).
[74] M. Shamir and H. Sompolinsky, Physical Review E **61**, 1839 (2000).