

¹ Are single-peaked tuning curves ² tuned for speed rather than ³ accuracy?

⁴ **Movitz Lenninger^{1*}, Mikael Skoglund¹, Paweł Herman², Arvind Kumar^{2*}**

***For correspondence:**

movitzle@kth.se (ML);
arvkumar@kth.se (AK)

⁵ ¹Division of Information Science and Engineering, KTH Royal Institute of Technology,
⁶ Stockholm, Sweden; ²Division of Computational Science and Technology, KTH Royal
⁷ Institute of Technology, Stockholm, Sweden

⁸ **Abstract** According to the efficient coding hypothesis, sensory neurons are adapted to provide maximal information about the environment, given some biophysical constraints. In early visual areas, stimulus-induced modulations of neural activity (or tunings) are predominantly single-peaked. However, periodic tuning, as exhibited by grid cells, has been linked to a significant increase in decoding performance. Does this imply that the tuning curves in early visual areas are sub-optimal? We argue that the time scale at which neurons encode information is imperative to understanding the advantages of single-peaked and periodic tuning curves. Here, we show that the possibility of catastrophic (large) errors creates a trade-off between decoding time and decoding ability. We investigate how decoding time and stimulus dimensionality affect the optimal shape of tuning curves for removing catastrophic errors. In particular, we focus on the spatial periods of the tuning curves for a class of circular tuning curves. We show an overall trend for minimal decoding time to increase with increasing Fisher information, implying a trade-off between accuracy and speed. This trade-off is reinforced whenever the stimulus dimensionality is high, or there is ongoing activity. Thus, given constraints on processing speed, we present normative arguments for the existence of the single-peaked tuning organization observed in early visual areas.

²⁵ **Introduction**

²⁶ One of the fundamental problems in systems neuroscience is understanding how sensory information can be represented in the spiking activity of an ensemble of neurons. The problem is exacerbated by the fact that individual neurons are highly noisy and variable in their responses, even to identical stimuli (?). A common feature of early sensory representation is that the neocortical neurons in primary sensory areas change their average responses only to a small range of features of the sensory stimulus. For instance, some neurons in the primary visual cortex respond to moving bars oriented at specific angles (?). This observation has led to the notion of *tuning curves*. Together, a collection of tuning curves provides a possible basis for a neural code.

³⁵ ³⁶ A considerable emphasis has been put on understanding how the structure of noise and correlations affect stimulus representation given a set of tuning curves (??????). More recently, the issue ³⁷ of local and catastrophic errors, dating back to the work of Shannon (?), has been raised in the ³⁸ context of neuroscience (e.g., ??). Intuitively, local errors are small estimation errors that depend ³⁹ on the trial-by-trial variability of the neural responses and the local shapes of the tuning curves ⁴⁰

41 surrounding the true stimulus condition (Figure 2a bottom plot, see s_1). On the other hand, catastrophic errors are very large estimation errors that depend on the trial-by-trial variability and the global shape of the tuning curves (Figure 2a bottom plot, see s_2). While a significant effort has been put into studying how stimulus tuning and different noise structures affect local errors, less is known about the interactions with catastrophic errors. For example, *Fisher information* is a common measure of the accuracy of a neural code (?????). The Cramér-Rao bound states that a lower limit of the minimal mean squared error (MSE) for any unbiased estimator is given by the inverse of Fisher information (?). Thus, increasing Fisher information reduces the lower bound on MSE. However, because Fisher information can only capture local errors, the true MSE might be considerably larger in the presence of catastrophic errors (???), especially if the available decoding time is short (??).

52
 53 A curious observation is that the tuning curves of early visual areas predominately use single-peaked firing fields, whereas grid cells in the entorhinal cortex are known for their periodically distributed firing fields (?). It has been shown that the multiple firing locations of grid cells increase the precision of the neural code compared to single-peaked tuning curves (???). This raises the question of why periodic firing fields are not a prominent organization of early visual processing, too?

59
 60 The theoretical arguments in favor of periodic tuning curves have mostly focused on local errors under the assumption that catastrophic errors are negligible (?). However, given the response variability, it takes a finite amount of time to accumulate a sufficient number of spikes to decode the stimulus. Given that fast processing speed is a common feature of visual processing (????), it is crucial that each neural population in the processing chain can quickly produce a reliable stimulus-evoked signal. Therefore, the time required to produce signals without catastrophic errors will likely put fundamental constraints on any neural code, especially in early visual areas.

67
 68 Here, we contrast Fisher information with the minimal decoding time required to remove catastrophic errors (i.e., the time until Fisher information becomes a reasonable descriptor of the MSE).
 69 We base the results on the maximum likelihood estimator for uniformly distributed stimuli (i.e.,
 70 the maximum a posteriori estimator) using populations of tuning curves with different numbers
 71 of peaks. We show that minimal required decoding time tends to increase with increasing Fisher
 72 information in the case of independent Poissonian noise to each neuron. This suggests a trade-off
 73 between the decoding accuracy of a neural population and the speed by which it can produce a
 74 reliable signal. Furthermore, we show that the difference in minimal decoding time grows with the
 75 number of jointly encoded stimulus features and in the presence of spontaneous (non-specific)
 76 activity. Thus, single-peaked tuning curves require shorter decoding times and are more robust to
 77 spontaneous activity than periodic tuning curves. Finally, we exemplify the issue of large estimation
 78 errors and periodic tuning in simple spiking neural network model tracking either a step-like
 79 stimulus change or a continuously time-varying stimulus.

81 Results

82 Shapes of tuning curves, Fisher information, and catastrophic errors

83 To enable a comparison between single-peaked and periodic (multi-peaked) tuning curves, we consider circular tuning curves responding to a D-dimensional stimulus $\mathbf{s} \in [0, R]^D$ according to

$$f_i(\mathbf{s}) = a_i \prod_{j=1}^D \exp \left(\frac{1}{w} \left(\cos \left(2\pi \frac{1}{\lambda_i R} (s_j - s'_{i,j}) \right) - 1 \right) \right) + b \quad (1)$$

85 where a_i is the peak amplitude of the stimulus-related tuning curve i , w is a width scaling parameter, λ_i defines the spatial period (relative to the stimulus range, R) of neuron i , $s'_{i,j}$ determines

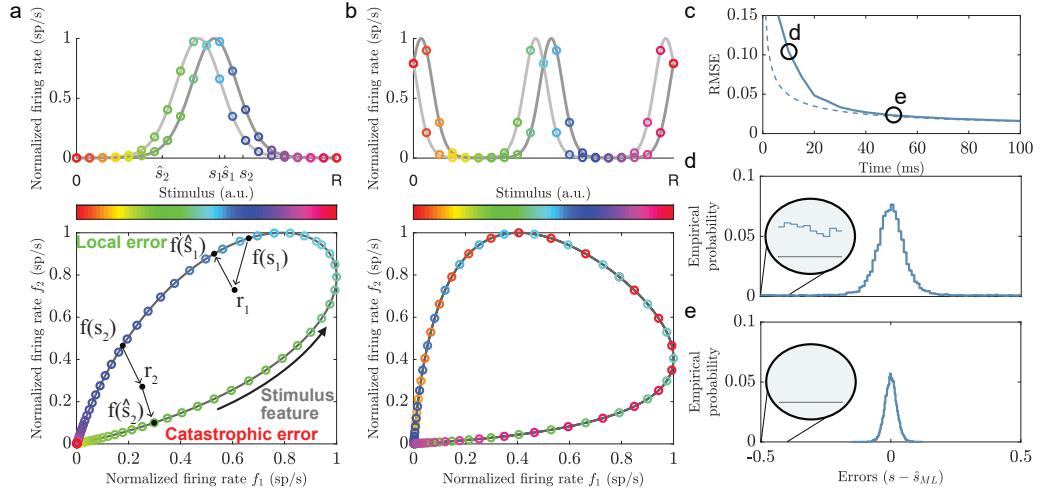


Figure 1. a) Top: A two-neuron system encoding a single variable using single-peaked tuning curves ($\lambda = 1$). Bottom: The tuning curves create a one-dimensional activity trajectory embedded in a two-dimensional neural activity space (black trajectory). Decoding the two stimulus conditions, s_1 and s_2 , illustrates the two types of estimation errors that can occur due to trial-by-trial variability, local (\hat{s}_1) and catastrophic (\hat{s}_2). b) Same as in a) but for periodic tuning curves ($\lambda = 0.5$). Notice that the stimulus conditions are intermingled and that the stimulus can not be determined from the firing rates. c) Time evolution of the root mean squared error (RMSE, solid line) and the Cramér-Rao bound (dashed line) for a population of single-peaked tuning curves ($N = 600$, $w = 0.3$, $a = 16 \text{ sp/s}$, $b = 1 \text{ sp/s}$). For about 50 ms the RMSE is significantly larger than the predicted lower bound. d) The empirical error distributions for the time point indicated in c), where the RMSE strongly deviates from the predicted lower bound. Inset: there is a non-zero empirical error probability spanning the entire stimulus domain. e) Same as in d) for the time point when the RMSE roughly converges to the Cramér-Rao bound. Notice the absence of large estimation errors.

the location of the peak(s) in the j :th stimulus dimension, and b determines the amount of spontaneous activity (see Fig 1a-b, top panels). The parameters are kept fixed for each neuron, thus ignoring any effect of learning or plasticity. In simulations, the stimulus domain was set to $[0, 1]^D$ for simplicity, although this choice does not qualitatively affect the results. Furthermore, to avoid any boundary effects, we assume that the stimulus has periodic boundaries (i.e., 0 and R is the same stimulus condition) and adjust any decoded value to lie within the stimulus domain (e.g., $s_{ML} = R + 0.1 \pmod{R} = 0.1$, etc.).

We assume that the stimulus is uniformly distributed across its domain and that its dimensions are independent of each other. This can be seen as a worst-case scenario as it maximizes the entropy of the stimulus. In a single trial, we assume that the number of emitted spikes for each neuron is conditionally independent and follows a Poisson distribution given some stimulus-dependent rate $f_i(s)$. Thus, the probability of observing a particular activity pattern, \mathbf{r} , in a population of N neurons given the stimulus-dependent rates is

$$p(\mathbf{r}|\mathbf{s}, T) = \prod_{i=1}^N p(r_i|T f_i(s)) = \prod_{i=1}^N \frac{(T f_i(s))^{r_i} \exp(-T f_i(s))}{r_i!}. \quad (2)$$

Given a model of neural responses, the Cramér-Rao bound provides a lower bound on the accuracy by which the population can communicate a signal as the inverse of the Fisher information. For sufficiently large populations, using the population and spike count models in equation (1) and equation (2), Fisher information is given by (for $a_i = a$ and $b = 0$ for all neurons, see ? or Methods and Materials for details)

$$J(\mathbf{s}) \approx (2\pi)^2 a \frac{TN}{R^2 w} B_0(1/w)^{D-1} B_1(1/w) \exp(-D/w) \overline{\lambda^{-2}} \quad (3)$$

106 where $\overline{\lambda^{-2}}$ denotes the sample average of the squared inverse of the (relative) spatial periods across
 107 the population, and $B_\alpha(\cdot)$ denotes the modified Bessel function of the first kind. Equation (3) (and
 108 similar expressions) suggests that populations consisting of periodic tuning curves, for which $\overline{\lambda^{-2}} \gg$
 109 1, are superior at communicating a stimulus signal than a population using tuning curves with
 110 only single peaks, where $\overline{\lambda^{-2}} = 1$. However, (inverse) Fisher information only predicts the amount
 111 of local errors for an efficient estimator. Hence, the presence of catastrophic errors (Figure 1a,
 112 bottom) can be identified by large deviations from the predicted MSE for an asymptotically efficient
 113 estimator (Figure 1c-d). Therefore, we define minimal decoding time as the shortest time required
 114 to approach the Cramér-Rao bound (Figure 1c and e).

115 Periodic tuning curves and stimulus ambiguity

116 To understand why the minimal decoding time can differ with different spatial periods, consider
 117 first the problem of stimulus ambiguity that can arise with periodic tuning curves. If all tuning
 118 curves in the population share the same relative spatial period, λ , then the stimulus-evoked re-
 119 sponds can only provide unambiguous information about the stimulus in the range $[0, \lambda R]$. Be-
 120 yond this range, the response distributions are no longer unique. Thus, single-peaked tuning
 121 curves ($\lambda = 1$) provide unambiguous information about the stimulus. Periodic tuning curves ($\lambda < 1$),
 122 on the other hand, require the use of tuning curves with two or more distinct spatial periods to
 123 resolve the stimulus ambiguity (??). In the following, we assume the tuning curves are organized
 124 into discrete modules, where all tuning curves within a module share spatial period (Figure 1b)
 125 mimicking the organization of grid cells (?). For convenience, assume that $\lambda_1 > \lambda_2 > \dots > \lambda_L$ where
 126 L is the number of modules. Thus, the first module provides the most coarse-grained resolution of
 127 the stimulus interval, and each successive module provides an increasingly fine-grained resolution.
 128 It has been suggested that a geometric progression of spatial periods, such that $\lambda_i = c \lambda_{i-1}$ for some
 129 spatial factor $0 < c \leq 1$, may be optimal for maximizing the resolution of the stimulus while reduc-
 130 ing the required number of neurons (??). However, trial-by-trial variability can still cause stimulus
 131 ambiguity and catastrophic errors - at least for short decoding times, as we show later, even when
 132 using multiple modules with different spatial periods.

133 Minimal decoding times in populations with two modules

134 How does the choice of spatial periods impact the minimal decoding time? To get some intuition,
 135 we first consider the case of using only two different spatial scales. From the perspective of a prob-
 136 abilistic decoder (??), assuming that the stimulus is uniformly distributed, the maximum likelihood
 137 (ML) estimator is Bayesian optimal (and asymptotically efficient). The maximum likelihood estima-
 138 tor aims at finding the stimulus condition which is the most likely cause of the observed activity, \mathbf{r} ,
 139 or

$$\hat{s}_{ML} = \arg \max_s p(\mathbf{r}|s), \quad (4)$$

140 where $p(\mathbf{r}|s)$ is called the likelihood function. The likelihood function equals the probability of ob-
 141 serving the observed neural activity, \mathbf{r} , assuming that the stimulus condition was s . In the case
 142 of independent Poisson spike counts (or at least independence across modules), each module
 143 contributes to the joint likelihood function $p(\mathbf{r}|s)$ with individual likelihood functions, Q_1 and Q_2
 144 (?). Thus, the joint likelihood function can be seen as the product of the two individual likelihood
 145 functions, where each likelihood is $\lambda_i R$ -periodic

$$p(\mathbf{r}|s) = Q_1(s)Q_2(s). \quad (5)$$

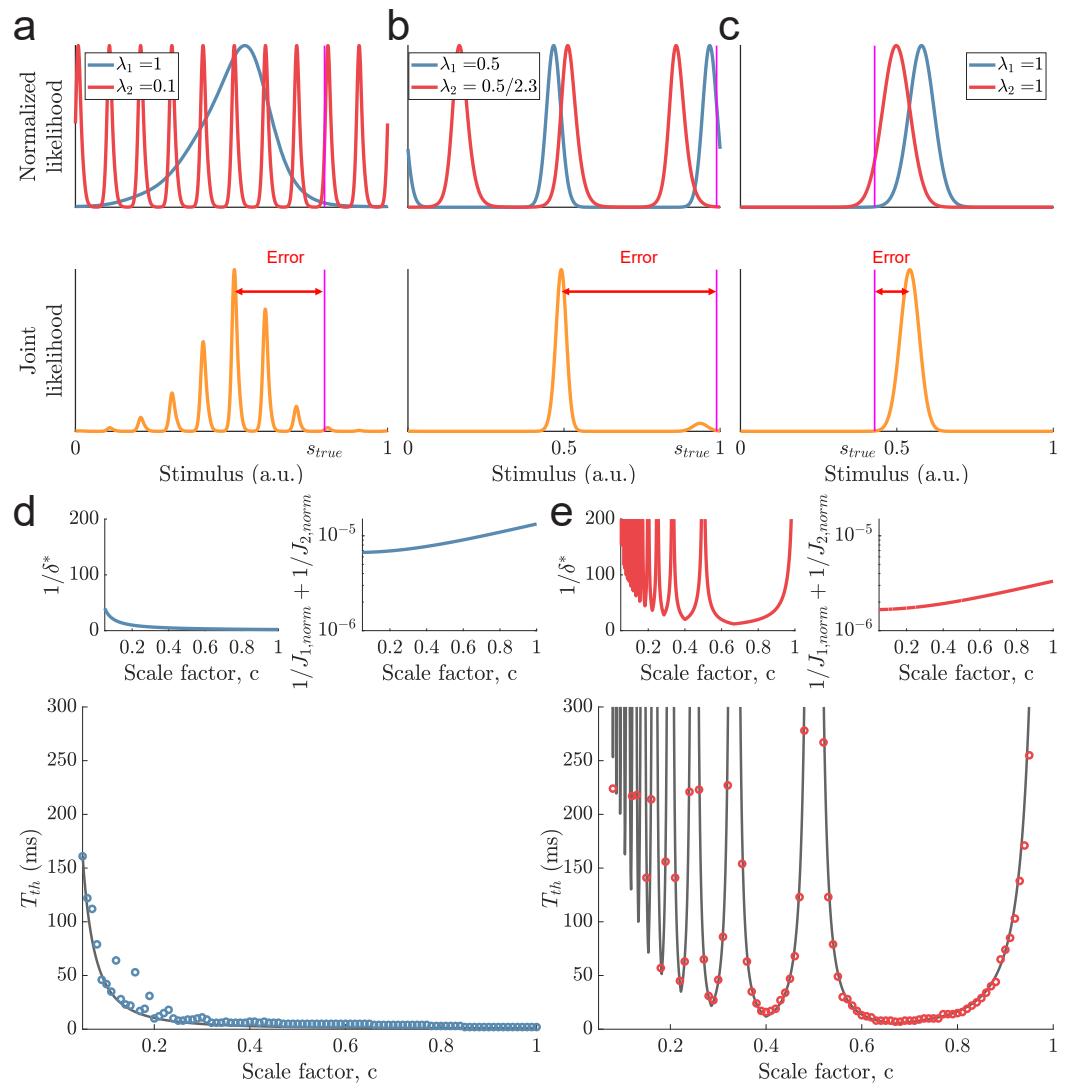


Figure 2. a) Top: Sampled individual likelihood functions of two modules with very different spatial periods. Bottom: The sampled joint likelihood function for the individual likelihood functions in the top panel. b-c) Same as in a) but for spatial periods that are similar but not identical and for a single-peaked population, respectively. d) Bottom: The dependence of the scale factor c on the minimal decoding time for $\lambda_1 = 1$. Blue circles indicate the simulated minimal decoding times, and the black line indicates the estimation of the minimal decoding times according to equation (7), with $p_{error} = 10^{-4}$. Top left: The predicted value of $1/\delta^*$. Top right: The inverse of the Fisher information. e) Same as d) but for $\lambda_1 = 1/2$. The parameters used in d-e) are: population size $N = 600$, number of modules $L = 2$, scale factors $c = 0.05 - 1$, width parameter $w = 0.3$, average evoked firing rate $f_{stim} = 20 \exp(-1/w) B_0(1/w)$ (sp/s), spontaneous activity $b = 0$ (sp/s), and threshold factor $\alpha = 2$.

Figure 2—figure supplement 1. Top: Sampled likelihood functions of two modules with $\lambda_1 = 1$ and $\lambda_2 = 1/2.3$. Bottom: The joint likelihood function is shifted across the periodic boundary. Such shifts across the periodic boundary can become more pronounced when λ_2 is slightly below a multiple of λ_1 .

Figure 2—figure supplement 2. Same as Figure 2d-e) but using threshold factor $\alpha = 1.2$. Notice the deviations from the predicted minimal decoding time for $\lambda_1 = 1$ when c is slightly below $1/2, 1/3, 1/4, \dots$, etc.

146 In this sense, each module provides its own ML-estimate of the stimulus, $s_{ML}^{(1)} = \arg \max_s Q_1(s)$ and
 147 $s_{ML}^{(2)} = \arg \max_s Q_2(s)$. Because of the periodicity of the tuning curves, there can be multiple modes
 148 for each of the likelihoods (e.g., Fig 2c and d, top panels). For the largest mode of the joint like-
 149 lihood function to also be centered close to the true stimulus condition, the distance δ between
 150 $s_{ML}^{(1)}$ and $s_{ML}^{(2)}$ must be smaller than between any other pair of modes of Q_1 and Q_2 . Thus, to avoid
 151 catastrophic errors, δ must be smaller than some largest allowed distance δ^* which guarantees this
 152 relation (see equations (11-16) for calculation of δ^* assuming the stimulus is in the middle of the
 153 domain).

154
 155 As δ varies from trial to trial, we limit the probability of the decoder experiencing catastrophic
 156 errors to some small error probability p_{error} by imposing that

$$\Pr(|\delta| > \delta^*) < p_{error}. \quad (6)$$

157 Assuming that the estimation of each module becomes efficient before the joint estimation, equa-
 158 tion (6) can be reinterpreted as a lower bound on the required decoding time before the estimation
 159 based on the joint likelihood function becomes efficient

$$T_{th} > 2 \left(\frac{\text{erfinv}(1 - p_{error})}{\delta^*} \right)^2 \left(\frac{1}{J_{1,norm}} + \frac{1}{J_{2,norm}} \right), \quad (7)$$

160 where $\text{erfinv}(\cdot)$ is the inverse of the error function and $J_{k,norm}$ refers to the time-normalized Fisher
 161 information of module k (see Methods and Materials for derivation). Thus, the spatial periods of
 162 the modules influence the minimal decoding time by determining: (1) the largest allowed distance
 163 δ^* between the estimates of the modules and (2) the variances of the estimations by the inverse of
 164 their respective Fisher information.

165
 166 To give some intuition of the approximation, if the spatial periods of the modules are very dif-
 167 ferent, $\lambda_2 \ll \lambda_1$, then there exist many peaks of Q_2 around the peak of Q_1 (Figure 2a). Additionally,
 168 there can be modes of Q_1 and Q_2 far away from the true stimulus close together. Thus, $\lambda_2 \ll \lambda_1$
 169 can create a highly multi-modal joint likelihood function where small deviations in $s_{ML}^{(1)}$ and $s_{ML}^{(2)}$ can
 170 cause a shift, or a change, of the maximal mode of the joint likelihood. To avoid this, δ^* must be
 171 small, leading to longer decoding times by equation (7). If, instead, the two modules have similar
 172 spatial periods $\lambda_2 \sim \lambda_1$, or λ_1 is close to a multiple of λ_2 , then the distance between the peaks some
 173 periods away is also small (Figure 2b), again leading to longer decoding times. Thus, assuming
 174 $\lambda_1 < 1$, both small and large scale factors c can lead to long decoding times. In other words, pe-
 175 riodic tuning suffers from the dilemma that small shifts in the individual stimulus estimates can
 176 cause catastrophic shifts in the joint likelihood function. When $\lambda_1 = 1$, however, only small scale
 177 factors c pose such problems, at least unless the stimulus is close to the periodic edge (i.e., $s \approx 0$ or
 178 $s \approx R$, see Figure 2 - Figure Supplement 1). On the other hand, compared to single-peaked tuning
 179 curves, periodic tuning generally leads to sharper likelihood functions, increasing the accuracy of
 180 the estimates once catastrophic errors are removed (e.g., compare the widths of the joint likeli-
 181 hood functions in Figure 2a-c).

182
 183 To test the approximation in equation (7), we simulated a set of populations ($N = 600$ neurons)
 184 with different spatial periods. The populations were created using identical tuning parameters ex-
 185 cept for the spatial periods, whose distribution varied across the populations, and the amplitudes,
 186 which were adjusted to ensure an equal average firing rate (across all stimulus conditions) for all
 187 neurons (see Methods and Materials for details on simulations). As described above, the spatial
 188 periods were related by a scale factor c . Different values of c were tested for the largest period
 189 being either $\lambda_1 = 1$ or $\lambda_1 = 1/2$. Furthermore, only populations with unambiguous codes over the

190 stimulus interval were included (i.e., $c \neq 1/2, 1/3, 1/4, \dots$ for $\lambda_1 = 1/2$) (?). For each population, the
 191 minimal decoding time was found by gradually increasing the decoding time until the empirical
 192 MSE was lower than twice the predicted lower bound (i.e., $\alpha = 2$, see equation (9) and Methods
 193 and Materials for details). Limiting the probability of catastrophic errors to $p_{error} = 10^{-4}$, equation
 194 (7) is a good predictor of the minimal decoding time (Figure 2d-e, bottom panels, $R^2 \approx 0.92$ and
 195 $R^2 \approx 0.95$ for $\lambda_1 = 1$ and $\lambda_1 = 1/2$, respectively). For both $\lambda_1 = 1$ and $\lambda_1 = 1/2$, the minimal decoding
 196 time increases overall with decreasing scale factor, c (see Figure 2d-e). However, especially for
 197 $\lambda_1 = 1/2$, the trend is interrupted by large peaks (Figure 2e). For $\lambda_1 = 1$, there are deviations from
 198 the predicted minimal decoding time for small scale factors, c . They occur whenever λ_2 is slightly
 199 below a multiple of $\lambda_1 = 1$, and get more pronounced when increasing the sensitivity to threshold
 200 factor $\alpha = 1.2$ (see Figure 2 - Supplement Figure 2). We believe one cause of these deviations is the
 201 additional shifts across the periodic boundary (as in Figure 2 - Supplement Figure 1) that can occur
 202 when c is just below $1/2, 1/3, 1/4, \dots$, etc.

203 **Minimal decoding times for populations with more than two modules**

204 From the two-module case above, it is clear that the choice of scale factor influences the minimal
 205 decoding time. However, equation (7) is difficult to interpret and is only valid for two module
 206 systems ($L = 2$). To approximate how the minimal decoding time scales with the distribution
 207 of spatial periods in populations with more than two modules, we extended the approximation
 208 method first introduced by ?. The method was originally used to assess the number of neurons
 209 required to reach the Cramér-Rao bound for single-peaked tuning curves with additive Gaussian
 210 noise for the ML estimator. In addition, it only considered encoding a one-dimensional stimulus
 211 variable. We adapted this method to approximate the required decoding time for stimuli with an
 212 arbitrary number of dimensions, Poisson-distributed spike counts, and tuning curves with arbitrary
 213 spatial periods. In this setting, the scaling of minimum decoding time with the spatial periods,
 214 $\lambda_1, \dots, \lambda_L$, can be approximated as (see Methods and Materials for derivation)

$$T_{th} \gg A(w) \frac{1}{aN} \frac{\exp(D/w)}{B_0(1/w)^{(D-1)}} \frac{\overline{\lambda^{-3}}^2}{\overline{\lambda^{-2}}^3} \simeq \frac{A^*(w)}{Nf_{stim}(D)} \frac{\overline{\lambda^{-3}}^2}{\overline{\lambda^{-2}}^3}, \quad (8)$$

215 where $\overline{\lambda^{-2}}$ and $\overline{\lambda^{-3}}$ indicate the sample average across the inverse spatial periods (squared or cubed,
 216 respectively) in the population, $f_{stim}(D)$ the average evoked firing rate across the stimulus domain,
 217 and $A(w)$ is a function of w (see Methods and Materials for detailed expression). The last approx-
 218 imation holds with equality whenever all tuning curves have an integer number of peaks. The
 219 derivation was carried out assuming the absence of spontaneous activity and that the amplitudes
 220 within each population are similar, $a_1 \approx \dots \approx a_N$. Importantly, the approximation also assumes the
 221 existence of a unique solution to the maximum likelihood equations. Therefore, it is ill-equipped
 222 to predict the issues of stimulus ambiguity. Thus, going back to the two module case, it cannot cap-
 223 ture the additional effects of $\lambda_2 \ll \lambda_1$ or when λ_1 is close to a multiple of λ_2 , as in Figure 2d-e. On
 224 the other hand, complementing the theory presented in equation (7), equation (8) provides a more
 225 interpretable expression of the scaling of minimal decoding time. For $c \leq 1$, the minimal decoding
 226 time, T_{th} , is expected to increase roughly linearly with decreasing scale factor, c (see equation (23)).
 227 The scaling should also be similar for different choices of λ_1 . Furthermore, keeping the amplitudes
 228 fixed and increasing the stimulus dimensionality, D , should dramatically increase the minimal de-
 229 coding time. Assuming all other parameters except D are constant, the minimal decoding time
 230 should grow roughly exponentially with the number of stimulus dimensions.

231
 232 To confirm the validity of equation (8), we simulated populations of $N = 600$ tuning curves across
 233 $L = 5$ modules, where the spatial periods were again defined using a scale factor c and the largest
 234 period, λ_1 (Figure 3a). To avoid the effects of $c \ll 1$, we limited the range of the scale factor to

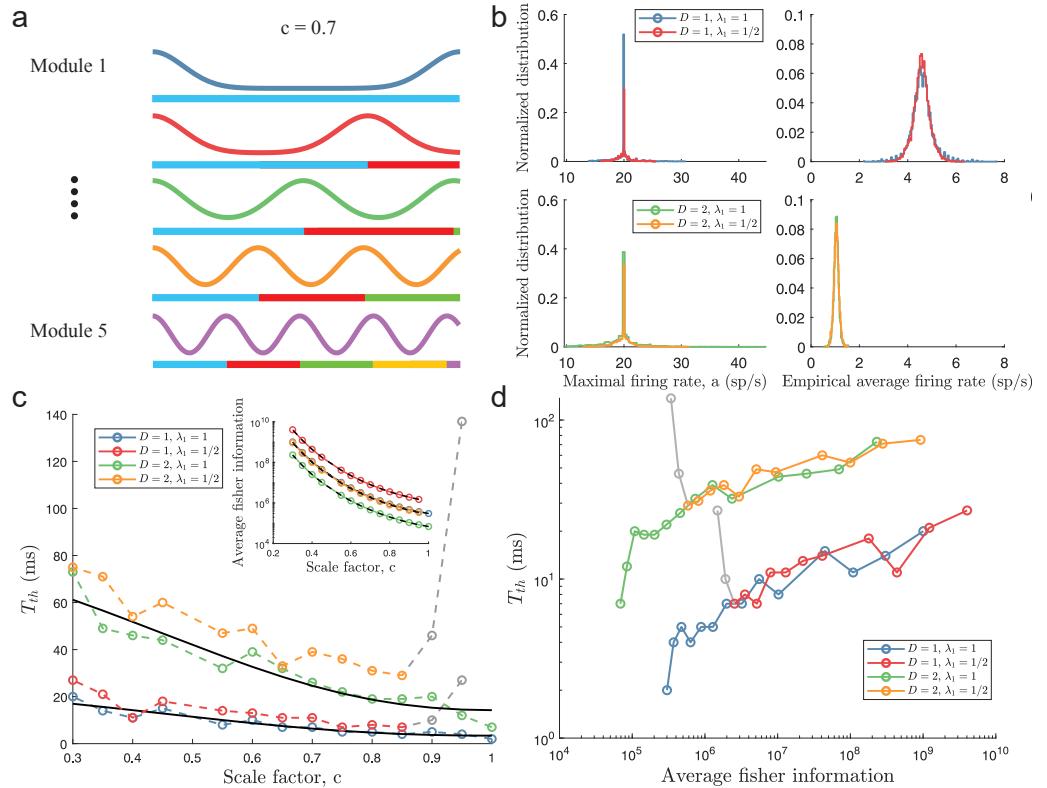


Figure 3. a) Illustration of the likelihood functions of a population with $L = 5$ modules using scale factor $c = 0.7$. b) The peak stimulus-evoked amplitudes of each neuron (left column) were selected such that all neurons shared the same expected firing rate for a given stimulus condition (right column). c) Inset: Plot of average Fisher information as a function of the scale factor c (colored lines: estimations from simulation data, black lines: theoretical approximations). Main plot: Plot of minimal decoding time as a function of scale factor c . Minimal decoding time tends to increase with decreasing grid scales (colored lines: estimated minimal decoding time from simulations, black lines: fitted theoretical predictions using equation (23)). The gray color corresponds to points with large discrepancies between the predicted and the simulated minimal decoding times. d) Plot of the average Fisher information against the minimal decoding time. Points colored in gray are the same as in panel c). The parameters used in the simulations are: population size $N = 600$, number of modules $L = 5$, scale factors $c = 0.3 - 1$, width parameter $w = 0.3$, average evoked firing rate $f_{stim} = 20 \exp(-D/w) B_0(1/w)^D$ (sp/s), spontaneous activity $b = 0$ (sp/s), and threshold factor $\alpha = 2$.

Figure 3—figure supplement 1. Replotting the minimal decoding time for $D = 2$ from Figure 3c (solid lines with circles, color code the same as in the main figure) and the predicted minimal decoding times for $D = 2$ using the data for $D = 1$ scaled by $\frac{\exp(1/w)}{B_0(1/w)}$ (dashed lines with diamond shapes, color code same as for $D = 2$). The predicted scaling of the minimal decoding time with stimulus dimensionality is in good agreement with the actual scaling.

Figure 3—figure supplement 2. a-c) Minimal decoding times as a function of scale factor, c , for various width parameters w . All other parameters are as in Figure 3. As in Figure 3c, there is a trend of increasing minimal decoding time with decreasing scale factor c . However, the range of minimal decoding times decreases with decreasing widths (for $D = 1$).

Figure 3—figure supplement 3. Same as Figure 3 in the main text, but simulated using threshold factor $\alpha = 1.2$. For each T , the MSE is evaluated based on 15000 random stimulus samples.

Figure 3—figure supplement 4. Same as Figure 3 in the main text, but simulated using another criterion on minimal decoding time. Instead of comparing the empirical MSE to the predicted lower bound, we here compared the full empirical distribution of errors to the predicted Gaussian distribution provided by the Cramér-Rao bound. The minimal decoding time is reached whenever the empirical distribution is indistinguishable from the predicted distribution using a one-sided KS test with significance level $\alpha = 0.05/j$, where j is the j :th time comparison. For each T , the MSE is evaluated based on 15000 random stimulus samples.

Figure 3—figure supplement 5. Plot of the mean spike counts (summed over the population) required to remove catastrophic errors for the populations in Figure 3. Each circle indicates the minimal spike count for a single population with a constant scale factor encoding either a 1-dimensional (x-axis) or a 2-dimensional stimulus (y-axis). Blue circles indicate $\lambda_1 = 1$ and red circles $\lambda_1 = 1/2$. Being on the grey line corresponds to having the same required spike count for both stimulus cases.

235 $0.3 \leq c \leq 1$. The upper bound on c was kept (for $\lambda_1 = 1$) to include entirely single-peaked populations. Again, the assumption of homogeneous amplitudes in equation (8) was dropped in simulations (Figure 3b, left column) to ensure that the average firing rate across the stimulus domain is equal for all neurons (Figure 3b, right column). This had little effect on Fisher information, where the theoretical prediction was based on the average amplitudes across all populations with the same λ_1 and stimulus dimensionality D (see Figure 3c, inset). As before, Fisher information grows with decreasing scale factor c and with decreasing spatial period λ_1 . As expected, increasing the stimulus dimensionality decreases Fisher information if all other parameters are kept constant. On the other hand, the minimal decoding time also increases with decreasing spatial periods and further increases with stimulus dimensionality (Figure 3c). The increase in decoding time between $D = 1$ and $D = 2$ is also very well predicted by equation (8), at least for $c > 0.5$ (Figure 3 - Figure Supplement 1). In these simulations, the choice of width parameter was compatible with experimental data (?), but similar trends were found for a range of different width parameters (although the effect becomes smaller for small w , see Figure 3 - Figure Supplement 2).

249

250 From equation (8), we fitted two constants, K_1 (regressor) and K_2 (intercept), using least square regression across populations sharing the same largest period, λ_1 , and stimulus dimensionality, D (see equation (23)). Within the simulated range of scale factors, the regressions provide reasonable fits for the populations with $\lambda_1 = 1$ (Figure 3c, $R^2 \approx 0.89$ and $R^2 \approx 0.90$ for $D = 1$ and $D = 2$, respectively). For the populations with $\lambda_1 = 1/2$, equation (8) becomes increasingly unable to predict the behavior of the minimal decoding time as c approaches 1 (see the red and yellow lines in Figure 3c-d). On the other hand, as was suggested above, the scaling of the minimal decoding time with c is, in fact, similar for $\lambda_1 = 1$ and $\lambda_1 = 1/2$ whenever c is less than ≈ 0.9 . As suggested by Figure 3d, there is also a strong correlation between Fisher information and minimal decoding time, again indicating a speed-accuracy trade-off. As we argue in the Discussion, the correlation between minimal decoding time and Fisher information is not simply due to a tougher requirement on MSE but reflects an important trade-off between accuracy and speed. Furthermore, very similar results are obtained when either decreasing the threshold factor to $\alpha = 1.2$ (Figure 3 - Supplement Figure 3) or changing the minimal decoding time criterion to a one-sided Kolmogorov-Smirnov test (KS-test) between the empirical distribution of errors and the Gaussian error distribution predicted by the Cramér-Rao bound (Figure 3 - Supplement Figure 4, using an ad-hoc Bonferroni-type correction for multiple sequential testing, α/j , where j is the j th time comparison and $\alpha = 0.05$ is the significance level.)

268

269 Thus, while periodic tuning curves provide lower estimation errors for long decoding times by minimizing local errors (Figure 3c, inset), a population of single-peaked tuning curves is faster at producing a statistically reliable signal by removing catastrophic errors (equation (8) and Figure 3c). Generalizing minimal decoding times to an arbitrary number of stimulus dimensions reveals that the minimal decoding time also depends on the stimulus dimensionality (Figure 3c, compare lines for $D = 1$ and $D = 2$). Interestingly, however, the approximation predicts that although minimal decoding time grows with increasing stimulus dimensionality, the minimal required spike count might be independent of stimulus dimensionality, at least for populations with integer spatial frequencies, i.e., integer number of peaks (equation (A5.2)). The populations simulated here have non-integer spatial frequencies. However, the trend of changes in the mean spike count is still just slightly below 1 (indicating that slightly fewer spikes across the population are needed with increasing D , see Figure 3 - Figure Supplement 5). Thus, as the average firing rate decreases with the number of encoded features D (Figure 3b, right column), the increase in minimal decoding time for stimuli of higher dimensionality can be largely explained by requiring a longer time to accumulate the sufficient number of spikes across the population.

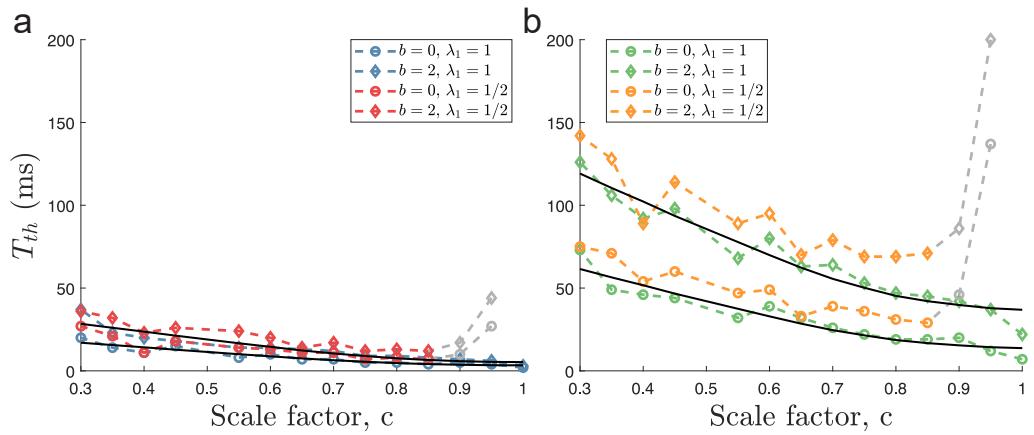


Figure 4. a) The case of encoding a one-dimensional stimulus ($D = 1$) with or without spontaneous activity at 2 sp/s (diamond and circle shapes, respectively). b) The case of a two-dimensional stimulus ($D = 2$) under the same conditions as for a). In both conditions, spontaneous activity increases the time required for all populations to produce reliable signals, but the effect is strongest for $c \ll 1$. The parameters used in the simulations are: population size $N = 600$, number of modules $L = 5$, scale factors $c = 0.3 - 1$, width parameter $w = 0.3$, average evoked firing rate $f_{stim} = 20 \exp(-D/w) B_0(1/w)^D$ (sp/s), spontaneous activity $b = 2$ (sp/s), and threshold factor $\alpha = 1.2$.

Figure 4—figure supplement 1. Same as Figure 4 in the main text, but using $\alpha = 1.2$.

Figure 4—figure supplement 2. Same as Figure 4 in the main text, but using the one-sided KS-test criterion described before (see Figure 3 - Figure Supplement 3).

284 **Effect of spontaneous activity**

285 Many cortical areas exhibit spontaneous activity, i.e., activity that is not stimulus-specific (??). Thus,
286 it is important to understand the impact of spontaneous activity on the minimal decoding time, too.
287 Unfortunately, because our approximation of the minimal decoding times did not include sponta-
288 neous activity, we relied on simulations to study the effect of such non-specific activity.

289
290 When including independent ongoing spontaneous activity at 2 spikes/s (sp/s) to all neurons for
291 the same populations as in Figure 3, minimal decoding times were elevated across all populations
292 (Figure 4). Furthermore, the minimal decoding time increased faster with decreasing c in the pres-
293 ence of spontaneous activity compared to the case without spontaneous activity (ratios of fitted
294 regressors $K_1(b = 2)/K_1(b = 0)$ using equation (23) were approximately 1.69 and 1.72 for $D = 1$ and
295 $D = 2$, respectively). Similar results are found using $\alpha = 1.2$ (Figure 4 - Figure Supplement 1) or
296 the alternative criterion on minimal decoding time based on one-sided KS-tests described earlier
297 (Figure 4 - Figure Supplement 2). Thus, spontaneous activity can have a substantial impact on the
298 time required to produce reliable signals. Figure 4 suggests that areas with spontaneous activity
299 are less suited for periodic tuning curves. Especially, the combination of multidimensional stim-
300 uli and spontaneous activity leads to much longer minimal decoding times for tuning curves with
301 small spatial periods ($c \ll 1$). For example, when encoding a two-dimensional stimulus, only the
302 populations with $\lambda_1 = 1$, $c = 1$ and $\lambda_1 = 1$, $c = 0.95$ could remove catastrophic errors in less than 40
303 ms when spontaneous activity at 2 sp/s was present. Thus, the ability to produce reliable signals at
304 high speeds severely deteriorates for periodic tuning curves in the presence of non-specific spon-
305 taneous activity.

306
307 This result has an intuitive explanation. The amount of catastrophic errors depends on the prob-
308 ability that the trial variability reshapes the neural activity to resemble the possible activities for
309 a distinct stimulus condition (see Figure 1a). From the analysis presented above, periodic tuning
310 curves have been suggested to be more susceptible to such errors. Adding spontaneous activity

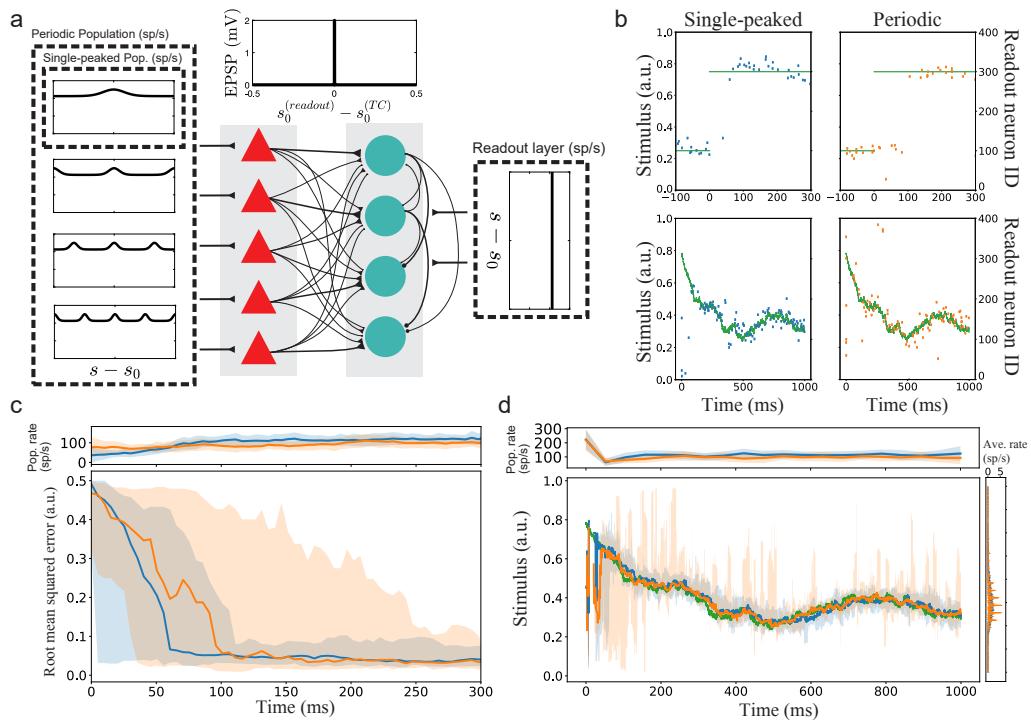


Figure 5. Implications for a simple spiking neural network with sub-optimal readout. a) Illustration of the spiking neural networks (SNNs). b) Example of single trials. Top row: Two example trials for step-like change in stimulus (green line). The left and right plots show the readout activity for the single-peaked (blue) and periodic SNNs (orange), respectively. Note that the variance around true stimulus is larger for the single-peaked SNN (i.e., larger local errors) but that there are fewer very large errors than for the periodic SNN. Bottom row: Same as for the top row but with a continuously time-varying stimulus. c) Bottom: The median RMSE (thick lines) over all trials in a sliding window (length 50 ms) for the single-peaked (blue) and periodic (orange) SNNs. The shadings correspond to the regions between the 5th and 95th percentiles. Top: The instantaneous population firing rates of the readout layers and the standard deviations (same color code as in the bottom panel). d) Bottom left: The median estimated stimulus across trials in a sliding window (length 10 ms) for the single-peaked (blue) and periodic (orange) SNNs. Shaded areas again correspond to the regions between the 5th and 95th percentiles. The true stimulus is shown in green. Bottom right: the average firing rate of each neuron, arranged according to the preferred stimulus condition. Top: The instantaneous population firing rates of the readout layers and the standard deviations. See Methods and Materials for simulation details and Tables (1) - (3) for all parameters used in the simulation.

311 does not reshape the stimulus-evoked parts of the tuning curves but only increases the trial-by-
 312 trial variability. Thus, by this reasoning, it is not surprising that the systems which are already
 313 more susceptible also are even more negatively affected by the increased variability induced by
 314 spontaneous activity. The importance of Figure 4 is that even spontaneous activity as low as 2 sp/s
 315 can have a clearly visible effect on minimal decoding time.

316 Implications for a simple spiking neural network with sub-optimal readout

317 Until this point, the arguments about minimal decoding time have relied on rate-based tuning
 318 curves encoding static stimuli. To extend beyond static stimuli and to exemplify the role of decod-
 319 ing time for spiking neurons, we simulated simple two-layer feed-forward spiking neural networks
 320 to decode time-varying stimulus signals. The first layer ($N_1 = 500$) corresponds to the tuning curves
 321 (without connections between the simulated neurons). The stimulus-specific tuning of the Poisso-
 322 nian inputs to these neurons is either fully single-peaked, creating a population of single-peaked
 323 tuning curves, or periodic with different spatial periods, creating a population of periodic tuning
 324 curves (Figure 5a, see Methods and Materials for details). The second layer instead acts as a read-
 325 out layer ($N_2 = 400$, allowing a weak convergence of inputs from the first layer). This layer receives

326 both stimulus-specific excitatory input from the first layer and external non-specific Poissonian
327 excitation (corresponding to background activity). The connection strength between the first and
328 second layers depends on the difference in preferred stimulus conditions between the pre- and
329 post-synaptic neurons. Such connectivity could, for example, be obtained by unsupervised Hebb-
330 bian learning. Because the tuning curves in the first layer can be periodic, they can also connect
331 strongly to several readout neurons. We introduced lateral inhibition among the readout neu-
332 rons (without explicitly modeling inhibitory neurons) to create a winner-take-all style of dynamics.
333 The readout neurons with large differences in preferred stimulus inhibit each other more strongly.
334 Decoding is assumed to be instantaneous and based on the preferred stimulus condition of the
335 spiking neuron in the readout layer. However, to compare the readouts, we averaged the stimulus
336 estimates in sliding windows.

337

338 We tested two different types of time-varying stimuli: (1) a step-like change from $s = 0.25$ to $s = 0.75$
339 (Figure 5b top row, green trace) and (2) a continuously time-varying stimulus drawn from an Orn-
340 stein-Uhlenbeck process (Figure 5b bottom row, green trace; see Methods and Materials). In the
341 case of a step-like stimulus change, the readout layer for the single-peaked population required
342 a shorter time to switch states than the periodic network (Figure 5c). The shorter switching time
343 is consistent with the hypothesis that single-peaked tuning curves have shorter minimal decoding
344 times than periodic tuning curves. In these simulations, the difference is mainly due to some neu-
345 rons in the first layer of the periodic network responding both before and after the step change.
346 Thus, the correct readout neurons (after the change) must compensate for the hyper-polarization
347 built up before the change and the continuing inhibitory input from the previously correct readout
348 neurons (which still get excitatory inputs, too). Note that there are only minor differences in the
349 population firing rates between the readout layers, confirming that this is not a consequence of
350 different excitation levels but rather of the structures of excitation.

351

352 The continuously time-varying stimulus could be tracked well by both networks. However, aver-
353 aging across trials shows that SNNs with periodic tuning curves have larger sporadic fluctuations
354 (Figure 5d). This suggests that decoding with periodic tuning curves has difficulties in accurately
355 estimating the stimulus without causing sudden, brief periods of large errors. To make a statisti-
356 cal comparison between the populations, we investigated the distributions of root mean squared
357 error (RMSE) across trials. In both stimulus cases, there is a clear difference between the network
358 with single-peaked tuning curves and the network with periodic ones. For the step-like change in
359 stimulus condition, a significant difference in RMSE arises roughly 100 ms after the stimulus change
360 (Figure 6a). For the time-varying stimulus, using single-peaked tuning curves also results in signifi-
361 cantly lower RMSE compared to a population of periodic tuning curves (Figure 6b, RMSE calculated
362 across the entire trial).

363 Discussion

364 Several studies have suggested that periodic tuning creates an unparalleled precise neural code
365 by minimizing local errors (????). Nevertheless, despite this advantage of periodic tuning, single-
366 peaked tuning curves are widespread in early sensory areas and especially in the early visual sys-
367 tem. There is a long history of studying information representation using rate-based tuning curves.
368 Still, the effect of spatial periodicity and catastrophic errors on the required decoding time has not
369 been addressed. Here, we showed that the possibility of catastrophic estimation errors (Figure 2a)
370 introduces the possibility that different shapes of tuning curves can have different minimal decod-
371 ing times.

372

373 But why study the minimal decoding time rather than simply compare the time evolution of the
374 MSE? Comparing MSE directly between populations can be a misleading measure of reliability if
375 the distributions of errors are qualitatively different. That is, if the amounts of local errors are

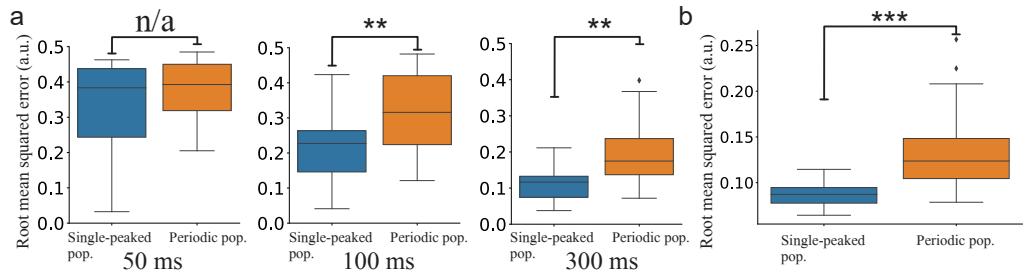


Figure 6. Statistical comparison of the SNN models. a) Step-like change: Comparison between the distributions of accumulated RMSEs at different decoding times ($p = 0.4, 9.0 \cdot 10^{-4}$, and $8.7 \cdot 10^{-5}$, respectively). b) OU-stimulus: The distributions of RMSE across trials for the two SNNs ($p = 4.3 \cdot 10^{-8}$). All statistical comparisons in a) and b) were based on two-sample Kolmogorov-Smirnov (KS) tests using 30 trials per network.

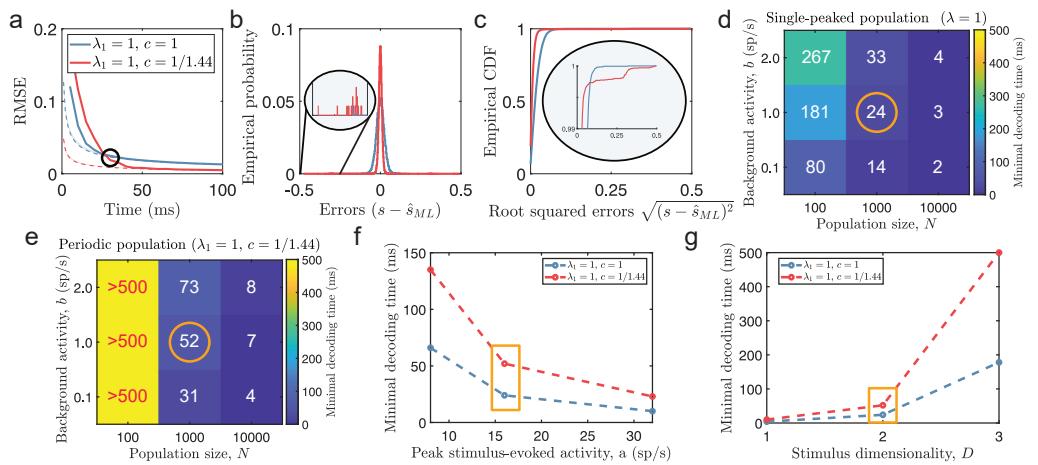


Figure 7. a) Time evolution of RMSE (solid lines) for two populations (blue: $\lambda_1 = 1, c = 1$, red: $\lambda_1 = 1, c = 1/1.44$). Dashed lines indicate the lower bound predicted by Cramér-Rao, and the shaded areas indicate the region between the lower bound and the limit set for minimal decoding times. The black circle indicates the point where the periodic population has become optimal in terms of MSE. b) The empirical distribution of errors for the time indicated by the black circle in a). The single-peaked population (blue) has a wider distribution of errors centered around 0 compared to the periodic population (red), as suggested by having a higher MSE. Inset: Zooming in on rare error events reveals that while the periodic population has a narrower distribution of errors around 0, it also has occasional errors across large parts of the stimulus domain. c) The empirical CDF of the errors for the same two populations as in b). Inset: a zoomed-in version (last 1%) of the empirical CDF highlights the heavy-tailed distribution of errors for the periodic population. d-e) Minimal decoding time for different combinations of population sizes (N) and levels of spontaneous activity (b) for the single-peaked population (d) and the periodic population (e). f) Minimal decoding time as a function of average stimulus-evoked firing rate (x-axis re-scaled to the corresponding peak amplitude, a , for single-peaked tuning curves for easier interpretation). $a = 8, 16, 32$ sp/s, respectively. g) Minimal decoding time as a function of stimulus dimensionality. The parameters in a-g) are, unless indicated on the axes, set according to the orange circles and rectangles in d-g). The width parameter is $w = 0.3$ and threshold factor $\alpha = 1.2$.

³⁷⁶ different, lower MSE does necessarily not imply lower amounts of catastrophic errors. This is illus-
³⁷⁷ trated in Figures 7a-c, comparing a single-peaked population with a periodic population using the
³⁷⁸ suggested optimal scale factor $c \approx 1/1.44$ for a two-dimensional stimulus (?). Thus, a comparison
³⁷⁹ of MSE only becomes valid once the minimal decoding times have been met. Here we assume that
³⁸⁰ catastrophic errors should strongly affect the usability of a neural code. Therefore, we argue that
³⁸¹ the first criterion for any rate-based neural code should be to satisfy its constraint on decoding
³⁸² time to avoid catastrophic errors.

³⁸³

³⁸⁴ The emerging question is whether there is a trade-off between the accuracy (i.e., Fisher informa-
³⁸⁵ tion) of a neural code and the minimal required decoding time for single-peaked and periodic
³⁸⁶ tuning. Fisher information has already been shown to be a poor descriptor of MSE when decoding
³⁸⁷ time is short (?), but it is not known whether the minimal decoding time changes with the shapes
³⁸⁸ of tuning curves. The answer is yes. We found that minimal decoding time increased with decreas-
³⁸⁹ ing the spatial periods of tuning curves (Figure 3c), suggesting a trade-off between accuracy and
³⁹⁰ speed for populations of tuning curves. We show that our results are valid for a range of popula-
³⁹¹ tion sizes (Figure 7d-e), spontaneous (Figure 7d-e) and evoked activities (Figure 7f), and stimulus
³⁹² dimensions (Figure 7g). We used the more conservative threshold factor $\alpha = 1.2$ to capture all
³⁹³ the nuances w.r.t. the level of spontaneous activity even for large population sizes. In simulated
³⁹⁴ networks with spiking neurons, we showed that the use of periodic tuning curves increased the
³⁹⁵ chances of large estimation errors, leading to longer times before switching "states" (Figure 5c)
³⁹⁶ and difficulties tracking a time-varying, one-dimensional stimulus (Figure 5d). Experimental data
³⁹⁷ suggest that minimal decoding times can be very short, of the order of tens of milliseconds, reflect-
³⁹⁸ ing that a considerable part of the information contained in firing rates over long periods is present
³⁹⁹ in short sample periods, too (?). Moreover, as the tuning curves in this study all have equal average
⁴⁰⁰ firing rates, we can reinterpret the minimal decoding time in terms of the prominence of the first
⁴⁰¹ spikes. The first few spikes have been shown to carry significant amounts of task information in
⁴⁰² both visual (?), olfactory (?), and somatosensory areas (?). In our simulations, tens of spikes carry
⁴⁰³ enough information to produce a reliable stimulus estimate free of catastrophic errors (Figure 3 -
⁴⁰⁴ Figure Supplement 4). As with decoding time, single-peaked tuning curves also need fewer spikes
⁴⁰⁵ to produce reliable signals. Thus, the speed-accuracy trade-off can be reinterpreted as a trade-off
⁴⁰⁶ between being accurate and efficient.

⁴⁰⁷

⁴⁰⁸ The notion of a speed-accuracy trade-off is further strengthened when considering high-dimensional
⁴⁰⁹ stimuli that demand longer minimal decoding times. Natural stimuli typically have higher dimen-
⁴¹⁰ sionality than those used in animal experiments. Many sensory neurons are tuned to multiple
⁴¹¹ features of the external stimulus, creating mixed selectivity of features (e.g., ?). For neurons re-
⁴¹² sponding to task-related variables, mixed selectivity has been shown to enable linear separability
⁴¹³ and to improve discriminability (?). For continuous stimulus estimations, mixed selectivity has
⁴¹⁴ also been proposed to decrease MSE when decoding time is limited (?). However, to remove catas-
⁴¹⁵ trofic errors, which, as we have argued, is not necessarily synonymous with lower MSE, the ex-
⁴¹⁶ponential increase in minimal decoding time could easily lead to very long decoding times. Thus,
⁴¹⁷ minimal decoding time should set a bound on the number of features a population can jointly en-
⁴¹⁸code reliably. In addition, neurons in sensory areas often exhibit a degree of non-specific activity
⁴¹⁹ (?). Introducing spontaneous activity to the populations in our simulations further amplified the
⁴²⁰ differences in minimal decoding times (Figure 4). Thus, for jointly encoded stimuli, especially in ar-
⁴²¹eas with high degrees of spontaneous activity, a population of single-peaked tuning curves might
⁴²² be the optimal encoding strategy for rapid and reliable communication.

⁴²³

⁴²⁴ To conclude, we provide normative arguments for the single-peaked tuning of early visual ar-
⁴²⁵eas. Rapid decoding of stimulus is crucial for the survival of the animals. Consistent with this,
⁴²⁶ animals and humans can process sensory information at impressive speeds. For example, the hu-

man brain can generate differentiating event-related potentials to go/no-go categorization tasks
 using novel complex visual stimuli in as little as 150 milliseconds (?). These "decoding" times do
 not decrease for highly familiar objects, suggesting that the speed of visual processing cannot be
 reduced by learning (?). Given constraints on low latency communication, it is crucial that each
 population can quickly produce a reliable signal. In this regard, single-peaked tuning curves are
 superior to periodic ones. The fact that early visual areas exhibit spontaneous activity and encode
 multi-dimensional stimuli further strengthens the relevance of the differences in minimal decod-
 ing times. We note that these results might extend beyond the visual areas, too. Constraints on
 decoding time are likely important for any early sensory area. As alluded to above, the first spikes
 seem to carry significant amounts of information in various sensory areas. Although this study
 focused on tuning curves encoding continuous variables, catastrophic errors can occur in systems
 with discrete tuning curves, too, and could therefore be an important consideration for any early
 sensory area. In addition, hippocampal place cells involved in spatial navigation (??) are known
 for their single-peaked tuning (but see ?). The interesting observation in this context is that place
 cells produce reliable signals faster than their input signals from the medial entorhinal cortex with
 a combination of single- and multi-peaked tuning (?).

In general, our work highlights that minimum decoding time is an important attribute and should
 be considered while evaluating candidate neural codes. Our analysis suggests that decoding of
 high-dimensional stimuli can be prohibitively slow with rate-based tuning curves. Experimental
 data on the representation of high-dimensional stimuli is rather scant as relatively low-dimensional
 stimuli are typically used in experiments (e.g., oriented bars). Our work gives a compelling reason
 to understand whether and how biological brains can reliably encode high-dimensional stimuli at
 behaviorally relevant time scales.

Methods and Materials

Minimal decoding times - Simulation protocols

To study the dependence of decoding time T on MSE for populations with different distributions of
 spatial frequencies, we simulated populations of synthetic tuning curves (equation (1)). The stimulus
 was chosen to be circular with range $[0, 1]^D$ to avoid boundary effects. The preferred stimulus
 conditions s' were sampled independently from a random uniform distribution over $[0, 1]$ (inde-
 pendently and uniformly for each stimulus dimension). The preferred locations s' were shared
 across all populations to ensure equal comparison. The amplitude, a_i , of each neuron was tuned
 according to equation (A1.4) to ensure equal average firing rate across the stimulus domain for all
 neurons. In each trial, a stimulus $s \in [0, 1]^D$ was also independently sampled from a uniform distri-
 bution over $[0, 1]^D$. The spike counts for each neuron were then sampled according to equation (2).

Minimal decoding time was defined as the minimal decoding time for which the neural popula-
 tion approximately reaches the Cramér-Rao bound. To estimate the reaction time in simulations,
 we incrementally increased decoding time T (using 1 ms increments, starting at $T = 1$ ms) until

$$\overline{MSE(T, \lambda)} \leq \alpha \cdot \overline{\text{diag}((J(T, \lambda))^{-1})}. \quad (9)$$

As the ML estimator is asymptotically efficient (attaining the Cramér-Rao bound in the limit of infi-
 nite data), the factor α in equation (9) was added as a relaxation. Note that the mean here refers to
 the mean across stimulus dimensions (for multi-dimensional stimuli) and that $\text{diag}(\cdot)$ refers to tak-
 ing the diagonal elements from the inverse of the Fisher information matrix, $(J(T, \lambda))^{-1}$. For a given
 decoding time T , the estimation of MSE was done by repeatedly sampling random stimulus condi-
 tions (from a uniform distribution), sampling a noisy response to the stimulus (Poisson distributed
 spike counts), and then applying maximum likelihood estimation (see section 'Implementation of

473 maximum likelihood estimator' for details on implementation). In Figures 2 and 7, 15000 stimulus
 474 conditions were drawn for each T , and in Figure 3, stimulus conditions were repeatedly drawn
 475 until the two first non-zero digits of the MSE were stable for 1000 consecutive trials. However, in
 476 controls not presented here, we could not see any significant difference between these two sam-
 477 pling approaches. Because the Fisher information matrix J was analytically estimated only in the
 478 special case without spontaneous activity, it was approximated in simulations by the element-wise
 479 average across 10000 randomly sampled stimulus conditions (also uniformly distributed), where
 480 each element was calculated according to equation (A2.5) or (A2.6) given a random stimulus trial.

481 **Implementation of maximum likelihood estimator**

482 Given some noisy neural responses, \mathbf{r} , the maximum likelihood estimator (MLE) chooses the stim-
 483 ulus condition which maximizes the likelihood function, $\hat{\mathbf{s}}_{ML} = \arg \max_{\mathbf{s}} \mathcal{L}(\mathbf{r}, \mathbf{s}) = \arg \max_{\mathbf{s}} \prod_{i=1}^N p(r_i | \mathbf{s})$.
 484 A common approach is to instead search for the maximum of the log-likelihood function (the
 485 logarithm is a monotonic function and therefore preserves any maxima/minima). The stimulus-
 486 dependent terms of the log-likelihood can then be expressed as

$$\log p(\mathbf{r} | \mathbf{s}) \propto \mathcal{V}(\mathbf{r}; \mathbf{s}) = \sum_{i=1}^N r_i \log(T f_i(\mathbf{s})) - T f_i(\mathbf{s}).$$

487 Unfortunately, the log-likelihood function is not guaranteed to be concave, and finding the stimulus
 488 condition $\hat{\mathbf{s}}_{ML}$ which maximizes the log-likelihood function is not trivial (a non-convex optimiza-
 489 tion problem). To overcome this difficulty, we combined grid-search with the Nelder–Mead method,
 490 an unconstrained non-linear program solver (implemented using MATLAB's built-in function *fmin-
 491 search*, <https://www.mathworks.com/help/matlab/ref/fminsearch.html>). Grid search was used to
 492 find a small set of starting points with the largest log-likelihood values (in simulations, the four
 493 stimulus conditions with the largest log-likelihood values were used). The true stimulus condition
 494 \mathbf{s}^* was always added into the set of starting points regardless of the log-likelihood value of that con-
 495 dition (yielding a total of 5 starting points). Then the Nelder–Mead method was used with these
 496 starting points to find a set of (possibly local) maxima. Thus, this approach does not overestimate
 497 the amount of threshold distortion but can potentially miss some global estimation errors instead.
 498 Given an estimated stimulus $\hat{\mathbf{s}}_{ML}$, the error was then evaluated along each stimulus dimension
 499 independently

$$\epsilon^2 = ((s_1 - \hat{s}_{ML,1})^2, \dots, (s_D - \hat{s}_{ML,D})^2).$$

500 **Spiking network model**

501 **Stimuli**

502 As in the previous simulations, we assumed that the stimulus domain was a circular stimulus de-
 503 fined between $[0, 1)$. We simulated the responses to two different types of stimuli, (1) a step-like
 504 change in stimulus condition from $s = 0.25$ to $s = 0.75$ and (2) a stimulus drawn from a modified
 505 Ornstein–Uhlenbeck process

$$\frac{ds_t}{dt} = -\frac{s_t}{\tau_s} + \sqrt{\frac{2\sigma_s^2}{\tau_s}} \xi_s \quad (\text{mod } 1).$$

506 For parameter values, see Table 1.

507 **Network model**

508 The spiking networks were implemented as two-layer, feed-forward networks using LIF neurons
 509 with (dirac) delta synapses. The dynamics of the membrane potential for the neurons in the first
 510 layer were described by

$$\frac{dV_i}{dt} = -\frac{V_i - V_{rest}}{\tau_{mem}} + \sum_k J_E \delta(t - t_k),$$

511 where V_i is the voltage of neuron i , τ_{mem} the membrane time constant, t_k the timing of the k th
 512 input spike to neuron i , and J_E the induced EPSP. The neurons in the first layer were constructed
 513 to correspond to either single-peaked or periodic tuning curves. Two networks were tested, one
 514 network where the first layer corresponds to single-peaked tuning curves and a second network
 515 corresponding to periodic tuning curves (with $L = 4$ modules). For each neuron i in module j
 516 in the first layer, the input was drawn from independent Poisson point processes with stimulus-
 517 dependent rates $f_i^{(j)}(s(t))$

$$f_i^{(j)}(s(t)) = a \exp\left(\frac{1}{w}\left(\cos\left(\frac{2\pi}{\lambda_j}(s(t) - s_i^{(j)})\right) - 1\right)\right) + b.$$

518 Here, the constants a and b were chosen such that the baseline firing rate was slightly above zero
 519 and the maximal firing rate was slightly below 20 sp/s (see Table 3 for all network related para-
 520 meter values). Because of the choice of λ_j , the modulation strengths of the inputs were such that the
 521 average input to each neuron was equal. For each module in the first layer, the preferred locations
 522 $s_i^{(j)}$ were equidistantly placed across $[0, \lambda_j]$.

523
 524 Similarly, for the second layer, the membrane potential was described by

$$\frac{dV_i}{dt} = -\frac{V_i - V_{rest}}{\tau_{mem}} + \sum_{j \in [1, \dots, N_1]} \sum_k J_{EE}(\Delta_{i,j}) \delta(t - t_k^{(j)} - d) + \sum_{j \in [1, \dots, N_2]} \sum_k J_I(\Delta_{i,j}) \delta(t - t_k^{(j)} - d) + \sum_k J_E \delta(t - t_k),$$

525 where $J_{EE}(\Delta_{i,j})$ and $J_I(\Delta_{i,j})$ are synapse-specific EPSPs/IPSPs which depends on the difference in
 526 preferred tuning $\Delta_{i,j}$ between the pre- and post-synaptic neurons (see equation (10)), $t_k^{(j)}$ the timing
 527 of the k th spike of pre-synaptic neuron j , and d the delay (see Table 2 for parameter values). The
 528 neurons in the second layer were only tuned to a single preferred stimulus location each, equidis-
 529 tantly placed across $[0, 1]$. Whenever a spike occurred in the first layer, it elicited EPSPs with a delay
 530 of 1.5 ms in all neurons in the second layer. The size of the EPSPs depended on the difference in
 531 preferred tuning ($\Delta_{i,j}$) between the pre- and post-synaptic neurons

$$J_{EE}(\Delta_{i,j}) = \exp\left(\frac{1}{w_{ro}}(\cos(2\pi\Delta) - 1)\right) J_{EE}. \quad (10)$$

532 Here J_{EE} determines the maximal EPSP (mV), and the constant w_{ro} was chosen such that the full
 533 width at half maximum of the EPSP kernels tiled the stimulus domain without overlap. Note that
 534 for periodically tuned neurons in the first layer (i.e., with multiple preferred locations), the $\Delta_{i,j}$ were
 535 determined by the smallest difference in preferred tuning across the multiple preferred locations.

536
 537 As for the excitatory neurons in the first layer, whenever a spike occurred in the second layer,
 538 it elicited IPSPs with a delay of 1.5 ms in all other neurons in the second layer. Again, the size of
 539 the IPSPs depended on the difference in preferred tuning ($\Delta_{i,j}$) between the two neurons, but this
 540 time according to

$$J_I(\Delta_{i,j}) = -|\sin(\pi\Delta)| J_I.$$

541 Thus, the range of inhibition was much broader compared to the excitation.

542 Evaluating decoding performance

543 We assumed that the decoder was instantaneously based on the neuron index of the firing neuron
544 in the readout layer. let $\Phi(t_k)$ denote a function that provides the index of the neuron firing at time
545 t_k . Given the equidistant distribution of preferred locations for the readout neurons, the stimulus
546 is instantaneously decoded by mapping the neuron identity to the interval $[0, 1]$

$$\hat{s}(t_k) = \frac{\Phi(t_k)}{N_2},$$

547 where N_2 is the number of neurons in the readout layer. For both stimulus cases, the decoding
548 performance was evaluated using (1) the distribution of RMSE (Figure 5d) or estimated stimulus
549 conditions (Figure 5e) in a sliding window or (2) the distributions of accumulated RMSE (Figure 6).

550 Parameters

Table 1. Parameters and parameter values for O-U stimulus.

Parameters	Parameter values
τ_s (s)	0.5
σ_s	0.1

Table 2. Parameters and parameter values for LIF neurons.

Parameters	Parameter values
Membrane time constant, τ_{memb} (ms)	20
Threshold memb. potential, V_{th} (mV)	20
Reset memb. potential (mV)	10
Resting potential, V_0 (mV)	0
Refractory period, τ_{rp} (ms)	2

Table 3. Spiking network parameters and parameter values.

Parameters	Parameter values
Number of neurons 1st layer, N_1	500
Number of neurons 2nd layer, N_2	400
Maximal stimulus-evoked input rate, a (sp/s)	750
Baseline input rate, b (sp/s)	4250
Spatial periods, λ_j	[1] or [1, 2, 3, 4]
Width parameter, w	0.3
Width parameter (readout layer), w_{ro}	$\frac{(\pi/N_2)^2}{2\log(2)}$
Input EPSP (1st layer), J_E (mV)	0.2
Maximal EPSP (2nd layer), J_{EE} (mV)	2
Maximal IPSP (2nd layer), J_{II} (mV)	2
Synaptic delays, d (ms)	1.5

551 Simulation tools

552 All the simulations were done using code written in MATLAB and Python (using Brian2 simulator
553 ?). The simulation code will be made available at https://github.com/movitzle/Short_Decoding_Time
554 upon publication of the manuscript.

555 Approximating minimal decoding time in 2-module systems

556 To gain an understanding of the interaction of two modules with different spatial periods, consider
557 the likelihood function as a product of the likelihood functions of the two modules individually

$$p(\mathbf{r}|s) = Q_1(s)Q_2(s).$$

558 Using the Laplace approximation, each of these functions can be approximated as a periodic sum
559 of Gaussians (?). Assuming that each module becomes efficient before the joint likelihood, we only
560 focus on the largest, periodically occurring, peaks

$$Q_i(s) \approx \hat{Q}_i(\mathbf{r}^{(i)}|s) = A_i \sum_{n_i=-K_i}^{K_i} \exp\left(-\frac{\Sigma_i}{2}(s - (s_{ML}^{(i)} + n_i \lambda_i R))^2\right),$$

561 where $\mathbf{r}^{(i)}$ denotes the activity pattern of module i , $s_{ML}^{(i)}$ the maximal peak closest to the true stim-
562 ulus condition, s_0 , and K_i is large enough for $\hat{Q}_i(\mathbf{r}^{(i)}|s)$ to cover the entire stimulus range $[0, R]$. The
563 approximation can be seen as "rolling out" the stimulus domain from $[0, R]$ to \mathbb{R} . Therefore, to
564 neglect the impact of the stimulus boundary, we assume that the stimulus is in the middle of the
565 stimulus domain and $K_1 = \lceil \frac{R}{2\lambda_1} \rceil$ and $K_2 = \lceil \frac{R}{2\lambda_2} \rceil$. Furthermore, assuming that each module is
566 efficient, the width of the Gaussians can be approximated as

$$\Sigma_i \approx -\frac{d^2}{ds^2} \log Q_i(s) \approx J_i(s),$$

567 where $J_i(s) \approx J_i$ is the Fisher information of module i . The joint likelihood function can thus be
568 approximated as

$$p(\mathbf{r}|s) \approx \hat{Q}_1(\mathbf{r}^{(1)}|s)\hat{Q}_2(\mathbf{r}^{(2)}|s) = \\ = A_1 A_2 \sum_{n_1=-K_1}^{K_1} \exp\left(-\frac{J_1}{2}(s - (s_{ML}^{(1)} + n_1 \lambda_1 R))^2\right) \sum_{n_2=-K_2}^{K_2} \exp\left(-\frac{J_2}{2}(s - (s_{ML}^{(2)} + n_2 \lambda_2 R))^2\right).$$

569 As the likelihood functions depend on the particular realization of the spike counts, the distance
570 between the modes of the respective likelihoods closest to the true stimulus condition s_0 , $\delta_{0,0} =$
571 $s_{ML}^{(1)} - s_{ML}^{(2)}$, is a random variable. Note that in the Results section, $\delta_{0,0}$ is simply referred to as δ for
572 clarity.

573
574 The joint likelihood distribution $p(\mathbf{r}|s)$ has its maximal peak close to the true stimulus condition
575 s_0 if $\delta_{0,0}$ is the smallest distance between any pairs of peaks of Q_1 and Q_2 (see equation (A3.1) for
576 details). Assuming that both modules provide efficient estimates, the distance $\delta_{0,0}$ can be approxi-
577 mated as a normally distributed random variable

$$\delta_{0,0} = s_{ML}^{(1)} - s_{ML}^{(2)} = (s_{ML}^{(1)} - s_0) - (s_{ML}^{(2)} - s_0) \sim \mathcal{N}\left(0, \frac{1}{T}(J_{1,norm}^{-1} + J_{2,norm}^{-1})\right), \quad (11)$$

578 where $J_{k,norm}$ refers to the time-normalized Fisher information of module k . Thus, as the decoding
579 time T increases, the variance of $\delta_{0,0}$ decreases. Hence, it is necessary for the decoding time T to
580 be large enough such that it is rare for $\delta_{0,0}$ not to be the smallest distance between any pair of
581 peaks. Similarly, the distance between the other pair of peaks in Q_1 and Q_2 within the stimulus
582 range becomes

$$\delta_{n_1, n_2} = (s_{ML}^{(1)} + n_1 \lambda_1 R) - (s_{ML}^{(2)} + n_2 \lambda_2 R) = \\ = \delta_{0,0} + R(n_1 \lambda_1 - n_2 \lambda_2), \quad (12)$$

583 where $n_1 \in \{-K_1, \dots, K_1\}$ and $n_2 \in \{-K_2, \dots, K_2\}$ are indexing the different Gaussians as before. Thus,
584 the threshold point for catastrophic error is if there is another pair of modes with the same distance
585 between them, i.e.,

$$|\delta_{0,0}| = |\delta_{n_1, n_2}| = |\delta_{0,0} + R(n_1\lambda_1 - n_2\lambda_2)|, \quad (13)$$

586 for some n_1 and n_2 belonging to the index sets as above. Thus, to avoid catastrophic errors, it is
587 necessary that

$$|\delta_{0,0}| \leq |\delta_{0,0} + R(n_1\lambda_1 - n_2\lambda_2)|, \quad (14)$$

588 for all $n_1 \in \{-K_1, \dots, K_1\}$ and $n_2 \in \{-K_2, \dots, K_2\}$. By solving equation (14), and taking into account
589 that $\delta_{0,0}$ can be either positive or negative, we get the maximally allowed displacement

$$\delta^* = \min_{n_1, n_2 : (n_1, n_2) \neq (0,0), n_1 \in \{-K_1, \dots, K_1\}, n_2 \in \{-K_2, \dots, K_2\}} \frac{1}{2} |R(n_1\lambda_1 - n_2\lambda_2)|. \quad (15)$$

590 Note that for $\lambda_1 = 1$, all n_1 represent the same mode (but one full rotation R away). Thus, we limit
591 the search such that not both $\lambda_1 n_1 \geq 1$ and $\lambda_2 n_2 \geq 1$. Assuming that the period of the second module
592 is a scaling of the first module, $\lambda_2 = c\lambda_1$, the above equation becomes

$$\delta^* = \min_{n_1, n_2 : (n_1, n_2) \neq (0,0)} \frac{1}{2} |R\lambda_1(n_1 - n_2c)|. \quad (16)$$

593 Note that stimulus ambiguity can never be resolved if $\delta_{n_1, n_2} = \delta_{0,0}$ for some pair $(n_1, n_2) \neq (0,0)$, which
594 is analogous to the condition in ?.

595
596 To limit the probability of catastrophic estimation errors from the joint distribution to some small
597 error probability p_{error} , the following should hold

$$Pr(|\delta_{0,0}| > \delta^*) < p_{error}.$$

598 Because $\delta_{0,0} \sim \mathcal{N}(0, J_1^{-1} + J_2^{-1})$, we have

$$Pr(|\delta_{0,0}| > \delta^*) = 1 - \text{erf}\left(\frac{\delta^*}{\sqrt{2}\sigma}\right) < p_{error},$$

599 where $\text{erf}(\cdot)$ is the error-function and $\sigma = \sqrt{J_1^{-1} + J_2^{-1}}$. By rearranging the terms and using equation
600 (A2.4), we can obtain a lower bound on the required decoding time

$$T_{th} > 2 \left(\frac{\text{erf}^{-1}(1 - p_{error})}{\delta^*} \right)^2 \left(\frac{1}{J_{1,norm}} + \frac{1}{J_{2,norm}} \right),$$

601 where $J_{i,norm}$ is the time-normalized Fisher information of module i . Note that δ^* can easily be found
602 using an exhaustive search according to equation (15) or equation (16).

603 Approximating minimal decoding time

604 To approximate the order by which the population reaction time scales with the distribution of spa-
 605 tial periods and the stimulus dimensionality, we extended the approximation method introduced
 606 by ?. The key part of the approximation method is to use a Taylor series to reason about which
 607 conditions must hold for the distribution of errors to be normally distributed with a covariance
 608 equal to the inverse of the Fisher information matrix. Note that this approximation assumes the
 609 existence of a unique solution to the maximum likelihood equations, thus it is not applicable to
 610 ambiguous neural codes (e.g., $c = 1/2, 1/3, 1/4, \dots$, etc.).

611

612 First, let's recollect the Taylor series with Lagrangian reminder for a general function g

$$g(x + \delta) = g(x) + g'(x)\delta + \frac{1}{2}g''(x^*)\delta^2,$$

613 where x^* is somewhere on the interval $[x, x + \delta]$. Thus, in the multivariate case, the derivative in the
 614 j:th direction of the log-likelihood function for stimulus condition $\hat{s}_{ML} = \hat{s}$ can be rewritten using a
 615 Taylor series with Lagrangian reminder as

$$\begin{aligned} \left. \frac{\partial}{\partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=\hat{\mathbf{s}}} &= \left. \frac{\partial}{\partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^\circ} + \sum_{l=1}^D \left. \frac{\partial^2}{\partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^\circ} (\hat{s}_l - s_l^\circ) + \\ &+ \frac{1}{2} \sum_{l=1}^D \sum_{m=1}^D \left. \frac{\partial^3}{\partial s_m \partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^*} (\hat{s}_l - s_l^\circ)(\hat{s}_m - s_m^\circ), \end{aligned}$$

616 for all $k \in \{1, \dots, D\}$ where s° is the true stimulus condition and s^* is a stimulus point between s° and
 617 \hat{s} .

618

619 If the estimated stimulus is close to the true stimulus, then the quadratic order terms are small.
 620 If so, the variance of $(\hat{s} - s^\circ)$ converges towards $\mathcal{N}(0, J^{-1})$ (in distribution), where J is the Fisher
 621 information matrix (?). However, if the estimated stimulus is not close to the true stimulus, then
 622 the quadratic terms are not negligible. Therefore, when T is sufficiently large, and the variance of
 623 the estimation follows the Cramér-Rao bound, the following should hold for all $k \in \{1, \dots, D\}$

$$\left| \sum_{l=1}^D \left. \frac{\partial^2}{\partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^\circ} (\hat{s}_l - s_l^\circ) \right| \gg \left| \frac{1}{2} \sum_{l=1}^D \sum_{m=1}^D \left. \frac{\partial^3}{\partial s_m \partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^*} (\hat{s}_l - s_l^\circ)(\hat{s}_m - s_m^\circ) \right|.$$

624 In this regime, we make the following term-wise approximations

$$\left. \frac{\partial^2}{\partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^\circ} \approx \mathbb{E} \left[\left. \frac{\partial^2}{\partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^\circ} \right] = -J_{k,l}(s^\circ) = -J_{k,l},$$

625 and

$$\left. \frac{\partial^3}{\partial s_m \partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^*} \approx \mathbb{E} \left[\left. \frac{\partial^3}{\partial s_m \partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right|_{\mathbf{s}=s^*} \right] = M_{k,l,m}(s^*), \quad (17)$$

626 which gives

$$\left| \sum_{l=1}^D J_{k,l} (\hat{s}_l - s_l^\circ) \right| \gg \left| \frac{1}{2} \sum_{l=1}^D \sum_{m=1}^D M_{k,l,m}(s^*) (\hat{s}_l - s_l^\circ)(\hat{s}_m - s_m^\circ) \right|. \quad (18)$$

627 Because $M_{k,l,m} \approx 0$ unless $k = l = m$ (see equations (A4.1-A4.2)) and using an upper bound for
 628 $M_{k,k,k}(s^*) \leq M_k^*$ (see equation (A4.3)), equation (18) simplifies to

$$\left| \sum_{l=1}^D J_{k,l}(\hat{s}_l - s_l^\circ) \right| \gg \left| \frac{1}{2} M^*(\hat{s}_k - s_k^\circ)^2 \right|.$$

629 Furthermore, because $J(s)$ is a diagonal matrix (see equation (A2.7)), we have

$$\left| J_{k,k}(\hat{s}_k - s_k^\circ) \right| \gg \left| \frac{1}{2} M^*(\hat{s}_k - s_k^\circ)^2 \right|.$$

630 Next, by taking the square of the absolute values, we obtain

$$J_{k,k}^2(\hat{s}_k - s_k^\circ)^2 \gg \frac{1}{4} \bar{M}^{*2} \left((\hat{s}_k - s_k^\circ)^2 \right)^2. \quad (19)$$

631 Because we assumed that N and T are sufficiently large to meet the Cramér-Rao bound, we have
632 that

$$(\hat{s}_k - s_k^\circ)(\hat{s}_l - s_l^\circ) \sim \{\bar{J}^{-1}\}_{k,l}. \quad (20)$$

633 Inserting equation (20) into equation (19) gives

$$J_{k,k}^2 \{J^{-1}\}_{k,k} \gg \frac{1}{4} M_k^{*2} \left(\{J^{-1}\}_{k,k} \right)^2, \quad (21)$$

634 or, equivalently,

$$1 \gg \frac{1}{4} M_k^{*2} \{J^{-1}\}_{k,k}^3 = \frac{1}{4} \frac{M_k^{*2}}{\{J\}_{k,k}^3}.$$

635 By approximating the term M_k^* (see equation (A4.3)) and using the expression for Fisher information
636 (equation (A2.4)), the expression for population reaction times can be obtained as

$$T_{th} \gg A(w) \frac{1}{aN} B_0 \left(\frac{1}{w} \right)^{(D-1)} \exp \left(\frac{D}{w} \right) \frac{\overline{\lambda}^{-2}}{\overline{\lambda}^{-3}}, \quad (22)$$

637 where $A(w)$ is a function of w . Lastly, by casting equation (22) in terms of the scale factor c , and
638 fitting using (for example) least square regression, we obtain

$$T_{th} \approx K_1 A(w) \frac{1}{aM} \frac{\exp(D/w)}{B_0(1/w)^{(D-1)}} \frac{\left(\sum_{j=0}^{L-1} c^{-3j} \right)^2}{\left(\sum_{j=0}^{L-1} c^{-2j} \right)^3} + K_2, \quad (23)$$

639 where M is the number of neurons per module, and K_1 and K_2 are constants. Note that in the
640 simulations, w is fixed and $A(w)$ can therefore be incorporated into K_1 .

641 Acknowledgments

642 We thank the reviewers for the helpful comments to improve the manuscript.

644
645 **Tuning curves and spike count model**
646

647 In the paper, we study the representation of a multidimensional stimulus $s = (s_1, \dots, s_D)$. For
 648 simplicity, it is assumed that the range of the stimulus in each dimension is equal, such that
 $s_j \in [0, R]$ for all $j \in \{1, \dots, D\}$. Note that this assumption does not qualitatively change the
 649 results. Furthermore, we assume that the tuning curves were circular (von Mises) tuning
 650 curves

$$651 f_i(s) = a_i \prod_{j=1}^D \exp \left(\frac{1}{w} \left(\cos \left(\frac{2\pi}{\lambda_i R} (s_j - s'_{i,j}) \right) - 1 \right) \right) + b = a_i \prod_{j=1}^D q_{i,j}(s) + b, \quad (A1.1)$$

652 where a_i is the peak amplitude of the stimulus-related tuning curve of neuron i , w is a width
 653 scaling parameter, λ_i defines the spatial period of the tuning, $s'_{i,j}$ determines the location of
 654 the firing field(s) in the j :th dimension, and b determines the amount of background activity.
 655 The amplitude parameters a_i were tuned such that all tuning curves had the same firing
 656 rate when averaged across all stimulus conditions (see Supplementary equation (A1.4)).

657 It is possible to reparametrize the stimuli into a phase variable, $\phi = \frac{s_j}{R}$. In the article, calcula-
 658 tions and numerical simulation are based on phase variables ϕ . This only changes the MSE
 659 and Fisher information by a constant scaling $\frac{1}{R^2}$. As we are interested in comparing the min-
 660 imal decoding time, not the MSE, we can drop the "unnormalized" stimulus s . The tuning
 661 curves in Supplementary equation (A1.1) can thus be rewritten using the phase variable ϕ
 662 as

$$663 f_i(\phi) = a_i \prod_{j=1}^D \exp \left(\frac{1}{w} \left(\cos \left(\frac{2\pi}{\lambda_i} (\phi_j - \phi'_{i,j}) \right) - 1 \right) \right) + b = a_i \prod_{j=1}^D q_{i,j}(\phi) + b. \quad (A1.2)$$

664 Given stimulus condition s (or ϕ) and decoding time T , the spike count of each neuron was
 665 independently sampled from a Poisson distribution with rate $T f_i(s)$. Thus, the probability
 666 of observing a particular spike count pattern $r = (r_1, \dots, r_N)$ given s is

$$667 p(r|s) = \prod_{i=1}^N p(r_i|s) = \prod_{i=1}^N \frac{(T f_i(s))^{r_i} \exp(-T f_i(s))}{r_i!}. \quad (A1.3)$$

677 **Adjusting amplitudes**

678 In order to make a fair comparison of decoding times across populations, we constrain each
 679 neuron to have the same average firing rate across the stimulus domain, \bar{f} . The average
 680 firing rate over the stimulus domain is

$$681 \bar{f}_i = b + a_i \frac{1}{R^D} \int_0^R \dots \int_0^R \prod_{j=1}^D q_{i,j}(\phi_j) d\phi_1 \dots d\phi_D.$$

682 Thus, given a desired stimulus-evoked firing rate, \bar{f}_{stim} , over a normalized stimulus range
 683 ($R = 1$), the amplitudes will be set to

687
688
689
690

$$a_i = \frac{\overline{f_{stim}}}{\int_0^1 \dots \int_0^1 \prod_{j=1}^D q_{i,j}(\phi_j) d\phi_1 \dots d\phi_D}. \quad (\text{A1.4})$$

691 Note that the integrals in equation (A1.4) are analytically solvable whenever the relative
692 spatial frequency $\xi_i = 1/\lambda_i$ is a positive integer, in which case we have

693
694
695
696

$$\int_0^1 \dots \int_0^1 \prod_{j=1}^D \exp \left(\frac{1}{w} \left(\cos \left(\frac{2\pi}{\lambda_i} (\phi_j - \phi'_{i,j}) \right) - 1 \right) \right) d\phi_1 \dots d\phi_D = B_0 \left(\frac{1}{w} \right)^D \exp \left(- \frac{D}{w} \right) \quad (\text{A1.5})$$

697 regardless of $\phi'_{i,j}$, here $B_0(\cdot)$ is the modified Bessel function of the first kind. In simulations,
698 $\overline{f_{stim,i}} = \overline{f_{stim}}$ was set such that tuning curves with integer spatial frequencies ($1/\lambda$) have am-
699 plitudes of 20 sp/s, i.e.,
700

701
702
703

$$\overline{f_{stim}(D)} = 20 B_0 \left(\frac{1}{w} \right)^D \exp \left(- \frac{D}{w} \right).$$

706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751

Appendix 2

Fisher information and the Cramér-Rao bound

Assuming a one-dimensional variable, the Cramér-Rao bound gives a lower bound on the MSE of any estimator G

$$\mathbb{E}[(G(\mathbf{r}) - s)^2] \geq \frac{[1 + b'_G(s)]^2}{J(s)} + b_G(s)^2,$$

where $b_G(s) = \mathbb{E}[G(\mathbf{r}) - s]$ is the bias of the estimator G and $J(s)$ is Fisher information, defined as

$$J(s) = \mathbb{E}\left[\frac{\partial}{\partial s} \log p(\mathbf{r}|s)\right]^2 = -\mathbb{E}\left[\frac{\partial^2}{\partial s^2} \log p(\mathbf{r}|s)\right] \quad (\text{A2.1})$$

where the last equality holds if $p(\mathbf{r}|s)$ is twice differentiable and the neural responses are conditionally independent (?). Assuming an unbiased estimator, the bound can be simplified to

$$\mathbb{E}[(G(s) - s)^2] = \text{Var}(G(\mathbf{r})) \geq \frac{1}{J(s)}.$$

For multi-parameter estimation, let $J(\mathbf{s})$ denote the Fisher information matrix, with elements defined analogously to Supplementary equation (A2.1)

$$J_{k,l}(\mathbf{s}) = -\mathbb{E}\left[\frac{\partial^2}{\partial s_k \partial s_l} \log p(\mathbf{r}|\mathbf{s})\right], \quad (\text{A2.2})$$

then (for unbiased estimators) the Cramér-Rao bound is instead stated as the following matrix inequality (?)

$$\text{Cov}(G) = \Sigma \geq J^{-1}(\mathbf{s})$$

in the sense that the difference $\Sigma - J^{-1}(\mathbf{s})$ is a positive semi-definite matrix. Thus, this implies the following lower bound for MSE of the k :th term

$$\text{Var}(G^{(s_k)}) = \Sigma_{k,k} \geq \{J^{-1}(\mathbf{s})\}_{k,k} \quad (\text{A2.3})$$

where $G^{(s_k)} = \hat{s}_k$, i.e., the estimation of s_k using estimator G . Note that if $J(\mathbf{s})$ is a diagonal matrix, i.e., $\{J(\mathbf{s})\}_{j,k} = 0$ for all $j \neq k$, then the following also holds

$$\{J^{-1}(\mathbf{s})\}_{k,k} = \{J_{k,k}(\mathbf{s})\}^{-1}.$$

For the tuning curves defined in Supplementary equation (A1.1), the diagonal elements of the Fisher information matrix can be analytically solved assuming $a_i \sim a$ within each module (see Supplementary equation (A2.8))

$$J_{k,k}(s) \approx (2\pi)^2 TN \frac{a}{R^2 w} B_0 \left(\frac{1}{w}\right)^{D-1} \exp\left(-\frac{D}{w}\right) B_1 \left(\frac{1}{w}\right) \bar{\lambda}^{-2} \quad (\text{A2.4})$$

where the bar indicates the sample average across modules. The off-diagonal elements, on the other hand, can be shown to be 0 (see below). Thus we have equality in the last inequality of Supplementary equation (A2.3), and the MSE for each stimulus dimension is lower bounded by the inverse of Supplementary equation (A2.4).

752 Approximating Fisher information

753 To analytically approximate the Fisher information for a given neural population, we will
 754 neglect the impact of spontaneous activity b . Then, the tuning curves in Supplementary
 755 equation (A1.1) factorize as $f_i(s) = aq_{1,i}(s_1)\dots q_{D,i}(s_D)$ and the log-likelihood for N neurons
 756 with conditionally independent spike counts becomes

$$755 \quad \log p(\mathbf{r}|\mathbf{s}) = \sum_{i=1}^N r_i \log(T f_i(\mathbf{s})) - T f_i(\mathbf{s}) - \log r_i!$$

757 By taking the second derivatives w.r.t. stimulus dimension, we get for $k = l$:

$$759 \quad \frac{\partial^2}{\partial s_k^2} \log p(\mathbf{r}|\mathbf{s}) = \dots = \sum_i^N r_i \left(\frac{q''_{k,i}}{q'_{k,i}} - \left(\frac{q'_{k,i}}{q_{k,i}} \right)^2 \right) - T f_i \frac{q''_{k,i}}{q'_{k,i}}$$

761 and for $k \neq l$

$$763 \quad \frac{\partial^2}{\partial s_k \partial s_l} \log p(\mathbf{r}|\mathbf{s}) = \dots = \sum_i^N -T f_i \frac{q'_{k,i} q'_{l,i}}{q_{k,i} q_{l,i}}.$$

765 Consequently, the elements of the Fisher information matrix are given by

$$767 \quad J_{k,k}(\mathbf{s}) = -\mathbb{E} \left[\frac{\partial^2}{\partial s_k^2} \log p(\mathbf{r}|\mathbf{s}) \right] = \sum_{i=1}^N T f_i(\mathbf{s}) \left(\frac{q'_{k,i}(s_k)}{q_{k,i}(s_k)} \right)^2 \quad (\text{A2.5})$$

770 and for $k \neq l$

$$771 \quad J_{k,l}(\mathbf{s}) = -\mathbb{E} \left[\frac{\partial^2}{\partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right] = \sum_{i=1}^N T f_i(\mathbf{s}) \frac{q'_{k,i}(s_k) q'_{l,i}(s_l)}{q_{k,i}(s_k) q_{l,i}(s_l)}. \quad (\text{A2.6})$$

774 To simplify calculations, it is possible to reparametrize the stimulus as in Supplementary
 775 equation (A1.2) using the formula for Fisher information under reparametrization (?)

$$776 \quad J_{k,l}(\phi) = \sum_m \sum_n \frac{\partial s_m}{\partial \phi_k} \frac{\partial s_n}{\partial \phi_l} J_{k,l}(\mathbf{s}) = R^2 J_{k,l}(\mathbf{s})$$

779 to obtain

$$780 \quad J(\mathbf{s}) = \frac{1}{R^2} J_{k,l}(\phi).$$

782 We can approximate the elements of the Fisher information matrix $J(\phi)$ in the limit of large
 783 N by replacing the sums with integrals, e.g.,

$$784 \quad J_{k,k}(\phi) = \sum_{i=1}^N T f_i(\phi) \left(\frac{q'_{k,i}(\phi_k)}{q_{k,i}(\phi_k)} \right)^2 \approx \\ 785 \quad \approx T \sum_{j=1}^L \frac{1}{\lambda_j^D} a_j \int_{\phi_1 - \frac{1}{2}\lambda_j}^{\phi_1 + \frac{1}{2}\lambda_j} \dots \int_{\phi_D - \frac{1}{2}\lambda_j}^{\phi_D + \frac{1}{2}\lambda_j} \left[\prod_{p=1}^D \exp \left(\frac{1}{w} \left(\cos \left(\frac{2\pi}{\lambda_j} (\phi_p - \phi'_p) \right) - 1 \right) \right) \right] \frac{(2\pi)^2 \sin^2 \left(\frac{2\pi}{\lambda_j} (\phi_k - \phi'_k) \right)}{\lambda_j^2 w^2} d\phi'$$

790 where L is the number of distinct modules, M is the number of neurons in each module,
 791 $d\phi' = d\phi'_1 \dots d\phi'_D$, and the D-dimensional integral is taken over the interval $[\phi_p - \frac{1}{2}\lambda_j, \phi_p + \frac{1}{2}\lambda_j]$
 792 along each dimension. Making the variable substitution $\theta_p = \frac{2\pi}{\lambda_j} (\phi_p - \phi'_p)$ for $p = \{1, \dots, D\}$ we
 793 have

$$\begin{aligned}
J_{k,k}(\phi) &\approx M T a \sum_{j=1}^L \frac{1}{\lambda_j^D} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \left[\prod_{p=1}^D \exp \left(\frac{1}{w} (\cos(\theta_p) - 1) \right) \right] \frac{(2\pi)^2 \sin^2(\theta_k)}{\lambda_j^2 w^2} \frac{\lambda_j^D}{(-1)^D (2\pi)^D} d\theta = \\
&= \dots = \frac{(2\pi)^2 N T a}{w} B_0 \left(\frac{1}{w} \right)^{D-1} B_1 \left(\frac{1}{w} \right) \exp \left(-\frac{D}{w} \right) \overline{\lambda^{-2}}
\end{aligned}$$

where the sample average is taken over the population's distribution of spatial frequencies
 and $B_a(\cdot)$ is the modified Bessel function of the first kind. Similar calculations for the case
 $k \neq l$ yield

$$\begin{aligned}
J_{k,l}(\phi) &= \dots \approx M T \sum_{j=1}^L \frac{a}{w^2} \exp \left(-\frac{D}{w} \right) B_0 \left(\frac{1}{w} \right)^{D-2} \int_{-\pi}^{\pi} \frac{1}{\lambda_j} \sin(\theta_k) \exp \left(\frac{1}{w} \cos(\theta_k) \right) d\theta_k \\
&\quad \int_{-\pi}^{\pi} \frac{1}{\lambda_j} \sin(\theta_l) \exp \left(\frac{1}{w} \cos(\theta_l) \right) d\theta_l = 0.
\end{aligned} \tag{A2.7}$$

Thus, the stimulus parameters will be asymptotically orthogonal for all of the populations
 considered in this paper. That is, the covariance matrix will be diagonal. The per-neuron
 average contribution to each diagonal element of the Fisher information matrix, as reported
 in the main text, is, therefore

$$\bar{J}_{k,k}(s) \approx T \frac{(2\pi)^2 a}{R^2 w} B_0 \left(\frac{1}{w} \right)^{D-1} B_1 \left(\frac{1}{w} \right) \exp \left(-\frac{D}{w} \right) \overline{\lambda^{-2}}. \tag{A2.8}$$

812 Maximum of the joint likelihood function (2 module case)

813 Assuming that the responses of the two modules are independent, the joint likelihood function
 814 $p(\mathbf{r}|s)$ can be decomposed into a product of the likelihood functions of the two modules.
 815 Using the approximation of each $Q_1(\mathbf{r}|s)$ and $Q_2(\mathbf{r}|s)$ as Gaussian sums (see Methods and
 816 Materials), we have the following

$$817 p(\mathbf{r}|s) = Q_1(s)Q_2(s) \approx \hat{Q}_1(s)\hat{Q}_2(s) = \\ 818 = A_1 \sum_{n_1=-K_1}^{K_1} \exp\left(-\frac{J_1}{2}(s - (s_{ML}^{(1)} + n_1 \lambda_1 R))^2\right) A_2 \sum_{n_2=-K_2}^{K_2} \exp\left(-\frac{J_2}{2}(s - (s_{ML}^{(2)} + n_2 \lambda_2 R))^2\right)$$

821 Thus, the contribution of the p :th and q :th mode of \hat{Q}_1 and \hat{Q}_2 to the joint likelihood function
 822 is

$$823 \hat{Q}_1^p \hat{Q}_2^q = A_1 A_2 \exp\left(-\frac{\Sigma_1}{2}(s - (s_{ML}^{(1)} + p \lambda_1 R))^2\right) \exp\left(-\frac{\Sigma_2}{2}(s - (s_{ML}^{(2)} + q \lambda_2 R))^2\right) \\ 824 = A_1 A_2 \exp\left(-\frac{\Sigma_1}{2}(s - s_p)^2 - \frac{\Sigma_2}{2}(s - s_q)^2\right)$$

827 where we in the last step renamed $s_{ML}^{(1)} + p \lambda_1 R$ and $s_{ML}^{(2)} + q \lambda_2 R$ to s_p and s_q , respectively.

828 Unless the width w of the tuning curves or the range R is very large, all the modes of \hat{Q}_1
 829 and \hat{Q}_2 , respectively, are well separated (see the end of the section). Thus, it is a reasonable
 830 approximation that the maximum of $p(\mathbf{r}|s) = Q_1(s)Q_2(s)$ is defined by the maximum
 831 of $\hat{Q}_1^p(s)\hat{Q}_2^q(s)$ across all combinations of p and q . Each combination $\hat{Q}_1^p(s)\hat{Q}_2^q(s)$ reaches its
 832 maximum for some stimulus $s_{p,q}^*$:

$$835 s_{p,q}^* = \arg \max_s \hat{Q}_1^p(s)\hat{Q}_2^q(s) = \arg \min_s \frac{\Sigma_1}{2}(s - s_p)^2 + \frac{\Sigma_2}{2}(s - s_q)^2$$

838 Taking the derivative w.r.t. s on the rightmost terms and solving gives

$$840 s_{p,q}^* = \frac{\Sigma_1 s_p + \Sigma_2 s_q}{\Sigma_1 + \Sigma_2}.$$

843 Thus, using $\delta_{p,q} = (s_p - s_q)$, the maximal value of each pair $\hat{Q}_1^p(s)\hat{Q}_2^q(s)$ is

$$844 \hat{Q}_1^p(s_{p,q}^*)\hat{Q}_2^q(s_{p,q}^*) = A_1 A_2 \exp\left(-\frac{J_1}{2}(s_{p,q}^* - s_p)^2 - \frac{J_2}{2}(s_{p,q}^* - s_q)^2\right) = \dots = \\ 845 = A_1 A_2 \exp\left(-\frac{J_1}{2}\left(\frac{-J_2 \delta_{p,q}}{J_1 + J_2}\right)^2 - \frac{J_2}{2}\left(\frac{J_1 \delta_{p,q}}{J_1 + J_2}\right)^2\right) = \dots = \quad (\text{A3.1}) \\ 846 = A_1 A_2 \exp\left(-\frac{1}{2} \frac{J_1 J_2}{J_1 + J_2} \delta_{p,q}^2\right)$$

847 Thus, the maximum likelihood choice will approximately be $s_{p,q}^*$ for the p :th and q :th mode
 848 with the smallest $\delta_{p,q}^2$, i.e., the smallest distance between the modes.

849 Lastly, all modes of \hat{Q}_1 and \hat{Q}_2 , respectively, need to be sufficiently separated such that
 850 no two pairs of p and q reinforce each other. However, it is well known the full width at half

maximum for a Gaussian function is FWHM = $2\sqrt{2 \ln 2}\sigma_i$, where for our functions \hat{Q}_1^p and \hat{Q}_2^q , $\sigma_1 = 1/\sqrt{J_1}$ and $\sigma_2 = 1/\sqrt{J_2}$. Thus, given the expression for Fisher information in equation (A2.4), the FWHM can be expressed as

$$FWHM = \frac{2}{\pi} \lambda_i \sqrt{\frac{\ln(2)wR^2}{2aTM B_0(1/w)^{D-1} B_1(1/w) \exp(-D/w)}}$$

Thus, for the modes to be separated, it is reasonable to require that the FWHM is no longer than one period length of the module, i.e., λ_i . Hence, we have that

$$\frac{2}{\pi} \lambda_i \sqrt{\frac{\ln(2)wR^2}{2aTM B_0(1/w)^{D-1} B_1(1/w) \exp(-D/w)}} < \lambda_i$$

Rewriting this into a bound on the time T needed for the assumption of separation, we get

$$T > \left(\frac{\pi}{2}\right)^2 \frac{\ln(2)wR^2}{2aMB_0(1/w)^{D-1}B_1(1/w)\exp(-D/w)}$$

For the parameters used in our simulations, this is satisfied very fast, on the order of tens of microseconds. However, note that the assumption of each module providing efficient estimates, which is a prerequisite for these approximations, requires significantly longer time scales. Thus, the assumption that the individual modes of \hat{Q}_1 and \hat{Q}_2 are well-separated is, in our case, not likely to be a restrictive assumption.

877 **Calculate $M_{k,l,m}$**

878 To approximate the minimal decoding time, we need to calculate (see equations (17 - 18),
 879 main text)

$$880 \quad 881 \quad M_{k,l,m}(\mathbf{s}) = \frac{\partial^3}{\partial s_m \partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \approx \mathbb{E} \left[\frac{\partial^3}{\partial s_m \partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) \right]. \\ 882$$

883 For $k \neq l \neq m$, using Supplementary equations (A1.1-A1.3), we have

$$884 \quad 885 \quad \frac{\partial^3}{\partial s_m \partial s_l \partial s_k} \log p(\mathbf{r}|\mathbf{s}) = - \sum_{i=1}^N T f_i(\mathbf{s}) \frac{q'_{k,i}(s_k) q'_{l,i}(s_l) q'_{m,i}(s_m)}{q_{k,i}(s_k) q_{l,i}(s_l) q_{m,i}(s_m)}$$

886 Thus, $M_{k,l,m}$ for $k \neq l \neq m$ becomes

$$887 \quad 888 \quad M_{k,l,m}(\mathbf{s}) = - \sum_{i=1}^N T f_i(\mathbf{s}) \frac{q'_{k,i}(s_k) q'_{l,i}(s_l) q'_{m,i}(s_m)}{q_{k,i}(s_k) q_{l,i}(s_l) q_{m,i}(s_m)} \approx \\ 889 \quad 890 \quad \approx \sum_{j=1}^L \frac{M}{(\lambda_j R)^D} \int_{s_1 - \frac{R}{2}\lambda_j}^{s_1 + \frac{R}{2}\lambda_j} \dots \int_{s_D - \frac{R}{2}\lambda_j}^{s_D + \frac{R}{2}\lambda_j} T a \left[\prod_{j=1}^D \exp \left(\frac{1}{w} \left(\cos \left(\frac{2\pi}{\lambda_j R} (s_j - s'_j) \right) - 1 \right) \right) \right] (A4.1) \\ 891 \quad 892 \quad \frac{(2\pi)^3}{\lambda_j^3 w^3} \sin \left(\frac{2\pi}{\lambda_j R} (s_k - s'_k) \right) \sin \left(\frac{2\pi}{\lambda_j R} (s_l - s'_l) \right) \sin \left(\frac{2\pi}{\lambda_j R} (s_m - s'_m) \right) d\mathbf{s}' = 0$$

893 as odd functions over even intervals integrate to zero. For $k \neq l = m$ (note that $k = l \neq m$ and
 894 $k = m \neq l$ follows by symmetry) we have

$$895 \quad 896 \quad \frac{\partial^3}{\partial s_l^2 \partial s_k} \log p(\mathbf{r}|\mathbf{s}) = - \sum_{i=1}^N T f_i(\mathbf{s}) \frac{q'_{k,i}(s_k) q''_{l,i}(s_l)}{q_{k,i}(s_k) q_{l,i}(s_l)}$$

897 and hence,

$$898 \quad 899 \quad M_{k,l,l} = - \sum_{i=1}^N T f_i(\mathbf{s}) \frac{q'_{k,i}(s_k) q''_{l,i}(s_l)}{q_{k,i}(s_k) q_{l,i}(s_l)} \approx \\ 900 \quad 901 \quad \approx - \sum_{j=1}^L \frac{M}{(\lambda_j R)^D} \int_{s_1 - \frac{R}{2}\lambda_j}^{s_1 + \frac{R}{2}\lambda_j} \dots \int_{s_D - \frac{R}{2}\lambda_j}^{s_D + \frac{R}{2}\lambda_j} T a \left[\prod_{j=1}^D \exp \left(\frac{1}{w} \left(\cos \left(\frac{2\pi}{\lambda_j R} (s_j - s'_j) \right) - 1 \right) \right) \right] (A4.2) \\ 902 \quad 903 \quad \frac{(2\pi)^3}{\lambda_j^3 w^3} \sin \left(\frac{2\pi}{\lambda_j R} (s_k - s'_k) \right) \left(w \cos \left(\frac{2\pi}{\lambda_j R} (s_l - s'_l) \right) - \sin^2 \left(\frac{2\pi}{\lambda_j R} (s_l - s'_l) \right) \right) d\mathbf{s}' = 0.$$

904 Lastly, for $k = l = m$ we have,

$$905 \quad 906 \quad \frac{d^3}{ds_k^3} \log p(\mathbf{r}|\mathbf{s}) = \sum_{i=1}^N (r_i - T f_i(\mathbf{s})) \frac{q'''_{k,i}(s_k)}{q_{k,i}(s_k)} - 3r_i \frac{q'_{k,i}(s_k) q''_{k,i}(s_k)}{q_{k,i}(s_k)^2} + 2r_i \left(\frac{q'_{k,i}(s_k)}{q_{k,i}(s_k)} \right)^3.$$

907 Thus, $M_{k,k,k}(\mathbf{s}^*)$ becomes

$$\begin{aligned}
M_{k,k,k}(\mathbf{s}^*) &= \mathbb{E}_{\mathbf{r} \sim \text{Pois}(f(\mathbf{s}^\circ))} \left[\frac{d^3}{ds_k^3} \log p(\mathbf{r}|\mathbf{s}) \Big|_{\mathbf{s}=\mathbf{s}^*} \right] = \\
&= \sum_{i=1}^N (T f_i(\mathbf{s}^\circ) - T f_i(\mathbf{s}^*)) \frac{q'''_{k,i}(s_k^*)}{q_{k,i}(s_k^*)} - 3 T f_i(\mathbf{s}^\circ) \frac{q'_{k,i}(s_k^*) q''_{k,i}(s_k)}{q_{k,i}(s_k)^2} + 2 T f_i(\mathbf{s}^\circ) \left(\frac{q'_{k,i}(s_k)}{q_{k,i}(s_k)} \right)^3.
\end{aligned}$$

914 Each term above have a dependence on $\sin(\frac{2\pi}{\lambda_j R}(s_k^* - s'_{k,i}))$, with an odd power. Therefore,
 915 when multiplying with $f(\mathbf{s}^*)$ and integrating as above, these terms vanish. Hence, we can
 916 focus only on the terms including $f(\mathbf{s}^\circ)$. After some calculus manipulation, it is possible to
 917 reduce the expression to include only $T \sin(\frac{2\pi}{\lambda_j R}(s_k^* - s'_{k,i})) f_i(\mathbf{s}^\circ)$ (for all i).

$$\begin{aligned}
M_{k,k,k}(\mathbf{s}^*) &\approx T \sum_{j=1}^L a \frac{M}{(\lambda_j R)^D} \frac{(2\pi)^3}{\lambda_j^3 R^3 w} \int_{s_1^* - \frac{R}{2}\lambda_j}^{s_1^* + \frac{R}{2}\lambda_j} \dots \int_{s_D^* - \frac{R}{2}\lambda_j}^{s_D^* + \frac{R}{2}\lambda_j} \sin \left(\frac{2\pi(s_k^* - s'_k)}{\lambda_j R} \right) \times \\
&\quad \times \prod_{j=1}^D \exp \left(\frac{1}{w} (\cos \left(\frac{2\pi}{\lambda_j R} (s_j^\circ - s'_j) \right) - 1) \right) d\mathbf{s}' = \\
&= T \sum_{j=1}^L a \frac{M}{(\lambda_j R)} \frac{(2\pi)^3}{\lambda_j^3 R^3 w} \exp \left(-\frac{D}{w} \right) B_0 \left(\frac{1}{w} \right)^{(D-1)} \times \\
&\quad \times \int_{s_k^* - \frac{R}{2}\lambda_j}^{s_k^* + \frac{R}{2}\lambda_j} \sin \left(\frac{2\pi(s_k^* - s'_k)}{\lambda_j R} \right) \exp \left(\frac{1}{w} (\cos \left(\frac{2\pi}{\lambda_j R} (s_k^\circ - s'_k) \right) - 1) \right) ds'_k
\end{aligned}$$

922 Unfortunately, as this integral includes both s_k^* and s'_k , no simple expression can be obtained.
 923 Using the variable substitution $\theta_k^* = \frac{2\pi}{\lambda_j R}(s_k^* - s'_k)$, we can simplify it slightly to

$$\begin{aligned}
M_{k,k,k}(\mathbf{s}^*) &\approx T a M \frac{(2\pi)^2}{R^3 w} \exp \left(-\frac{D}{w} \right) B_0 \left(\frac{1}{w} \right)^{(D-1)} \sum_{j=1}^L \frac{1}{\lambda_j^3} \times \\
&\quad \times \int_{-\pi}^{\pi} \sin(\theta_k^*) \exp \left(\frac{1}{w} (\cos(\theta_k^* + \frac{2\pi}{\lambda_j R}(s_k^\circ - s_k^*)) - 1) \right) d\theta_k^*
\end{aligned}$$

928 Instead, we focus on the difference $\phi_j^* = \phi_k(\lambda_j) = s_k^\circ - s_k^*$, which maximizes the above integral
 929 for each module. Thus, all integrals can be upper bounded by a constant C^* , yielding the
 930 upper bound
 931

$$M_{k,k,k}(\mathbf{s}^*) \leq M^* = (2\pi)^2 C^* \frac{T a N}{R^3 w} \exp \left(-\frac{D}{w} \right) B_0 \left(\frac{1}{w} \right)^{(D-1)} \frac{1}{\lambda^{-3}} \quad (\text{A4.3})$$

935 Note that the constant C^* can be incorporated into the regression coefficient K_1 in equation
 936 (23).

938
940
941
942
Approximating minimal required spike count

Given the approximation of minimal decoding time in equation (8) (main text), we seek to reformulate the approximation in terms of the required total spike count, instead. The average total spike count for a given population and stimulus condition is

$$943 \quad 944 \quad 945 \quad \mu(\mathbf{s}) = \mathbb{E}_r \left[\sum_{i=1}^N r_i | \mathbf{s} \right] = \sum_{i=1}^N T f_i(\mathbf{s})$$

where T is the decoding time. Thus, the average spike count over both stimulus conditions (assuming uniformly distributed stimulus and integer spatial frequencies) and trials for the entire population is

$$946 \quad 947 \quad 948 \quad 949 \quad 950 \quad 951 \quad \mu = \mathbb{E}_{\mathbf{s}} \left[\mathbb{E}_r \left[\sum_{i=1}^N r_i | \mathbf{s} \right] \right] = \frac{1}{R^D} \int_0^R \cdots \int_0^R \sum_{i=1}^N T f_i(\mathbf{s}) d\mathbf{s} = NT(aB_0(1/w)^D \exp(-D/w) + b).$$

Consequently, the number of spikes evoked by the stimulus-related tuning of the population is

$$952 \quad 953 \quad 954 \quad 955 \quad \mu_{stim} = NT a B_0 (1/w)^D \exp(-D/w). \quad (\text{A5.1})$$

Inserting Supplementary equation (A5.1) into equation (22) (main text) reveals the number of stimulus-evoked spikes, μ_{stim}^* , the population must produce before reaching the predicted lower bound

$$956 \quad 957 \quad 958 \quad 959 \quad 960 \quad 961 \quad \mu_{stim}^* \gg A(w) B_0 (1/w) \frac{\overline{\lambda^{-3}}^2}{\overline{\lambda^{-3}}}.$$

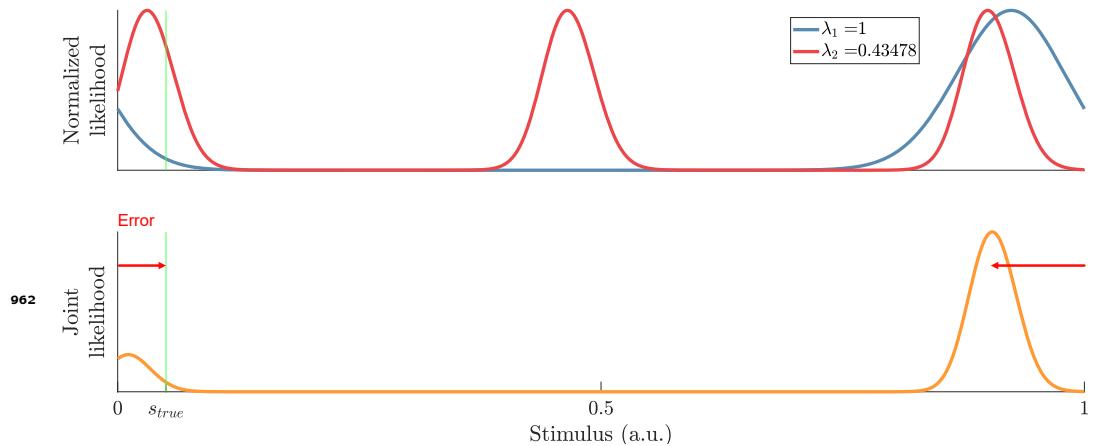


Figure 2—figure supplement 1. Top: Sampled likelihood functions of two modules with $\lambda_1 = 1$ and $\lambda_2 = 1/2.3$. Bottom: The joint likelihood function is shifted across the periodic boundary. Such shifts across the periodic boundary can become more pronounced when λ_2 is slightly below a multiple of λ_1 .

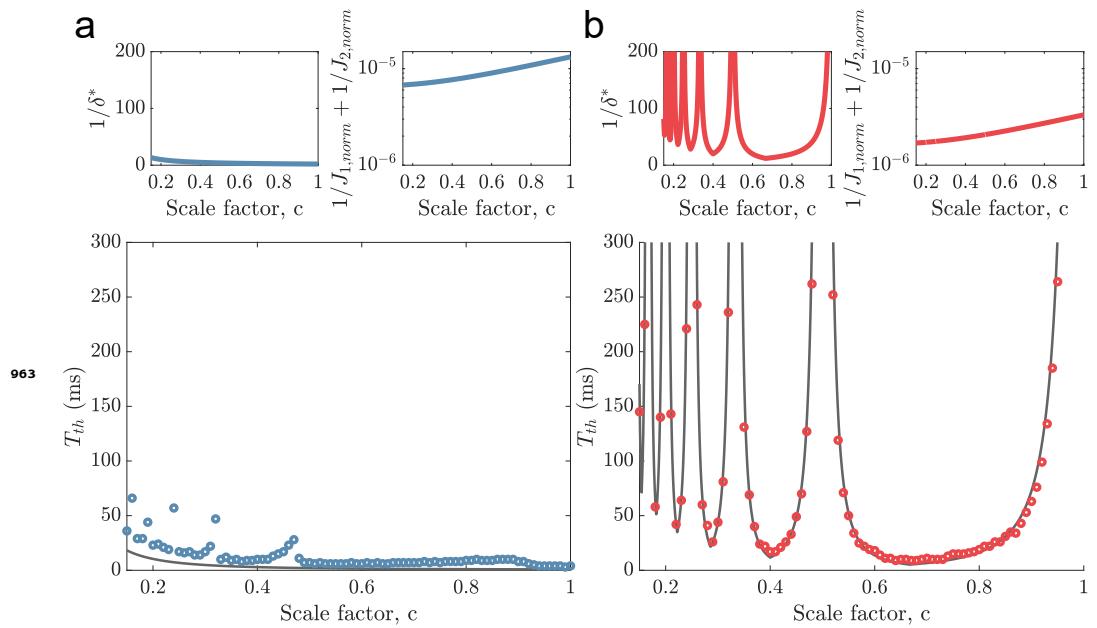


Figure 2—figure supplement 2. Same as Figure 2d-e) but using threshold factor $\alpha = 1.2$. Notice the deviations from the predicted minimal decoding time for $\lambda_1 = 1$ when c is slightly below $1/2, 1/3, 1/4, \dots$, etc.

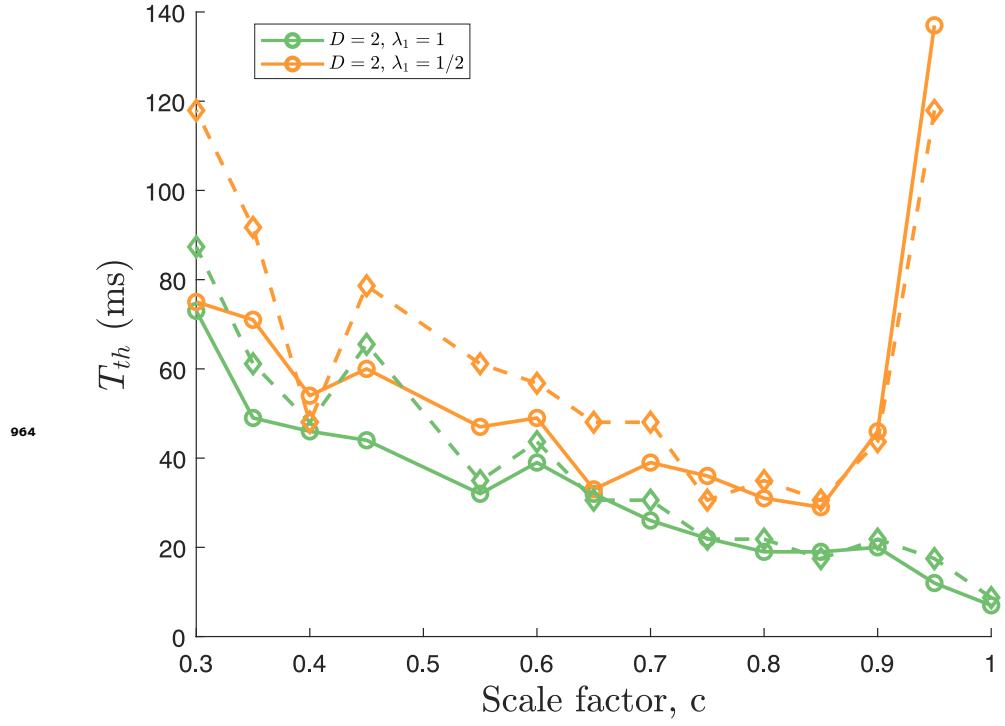


Figure 3—figure supplement 1. Replotting the minimal decoding time for $D = 2$ from Figure 3c (solid lines with circles, color code the same as in the main figure) and the predicted minimal decoding times for $D = 2$ using the data for $D = 1$ scaled by $\frac{\exp(1/w)}{B_0(1/w)}$ (dashed lines with diamond shapes, color code same as for $D = 2$). The predicted scaling of the minimal decoding time with stimulus dimensionality is in good agreement with the actual scaling.

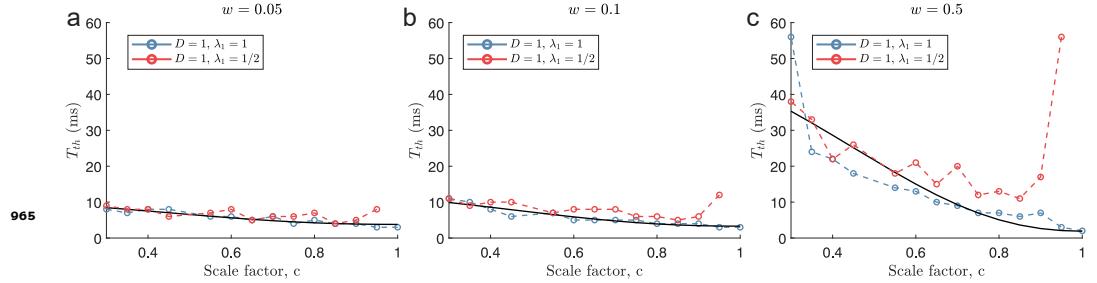


Figure 3—figure supplement 2. a-c) Minimal decoding times as a function of scale factor, c , for various width parameters w . All other parameters are as in Figure 3. As in Figure 3c, there is a trend of increasing minimal decoding time with decreasing scale factor c . However, the range of minimal decoding times decreases with decreasing widths (for $D = 1$).

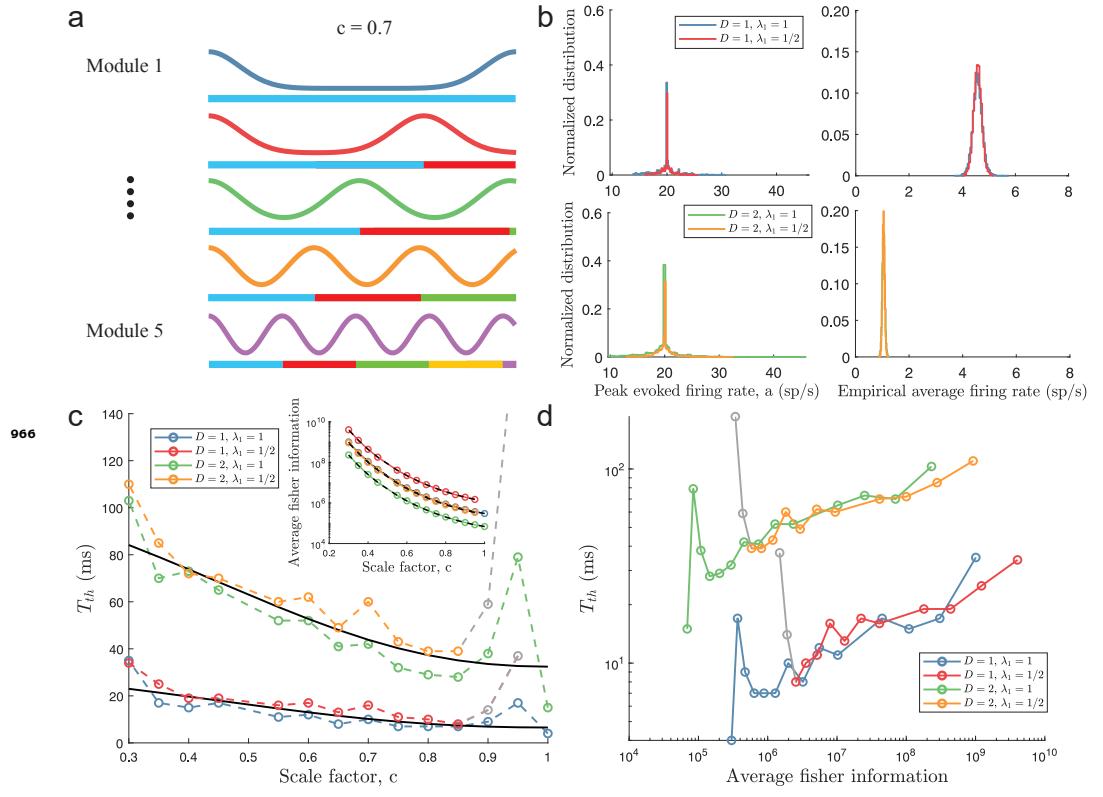


Figure 3—figure supplement 3. Same as Figure 3 in the main text, but simulated using threshold factor $\alpha = 1.2$. For each T , the MSE is evaluated based on 15000 random stimulus samples.

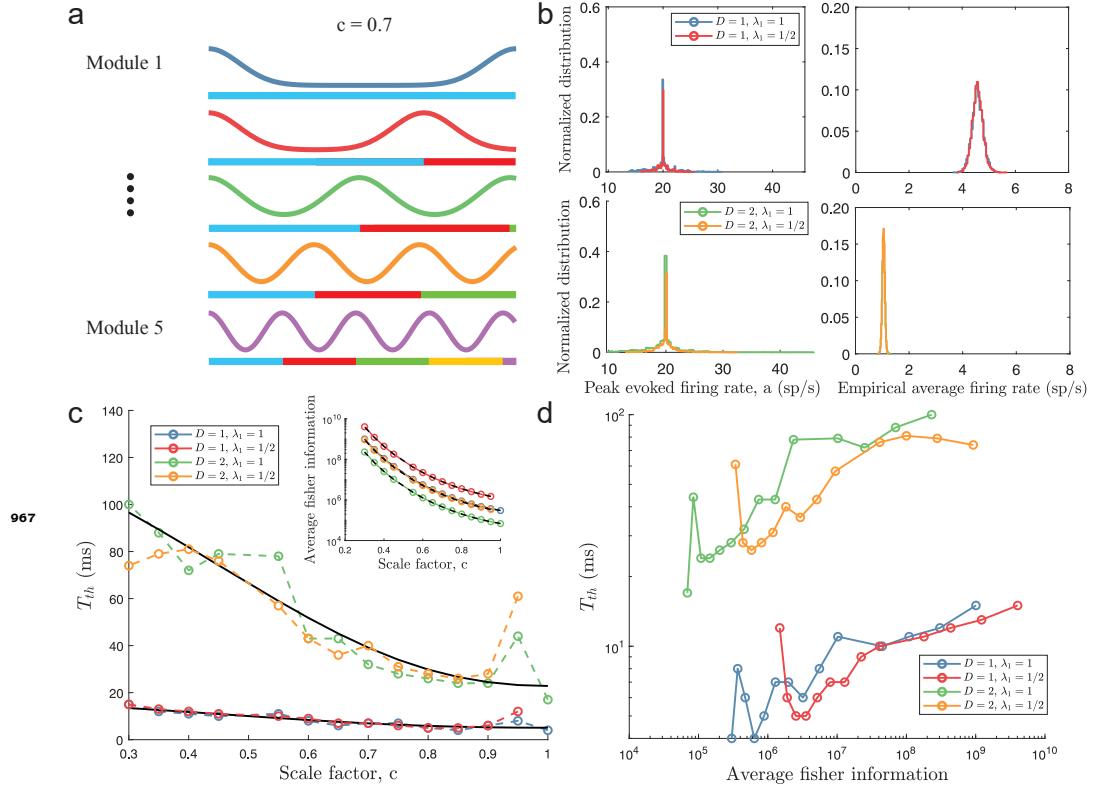


Figure 3—figure supplement 4. Same as Figure 3 in the main text, but simulated using another criterion on minimal decoding time. Instead of comparing the empirical MSE to the predicted lower bound, we here compared the full empirical distribution of errors to the predicted Gaussian distribution provided by the Cramér-Rao bound. The minimal decoding time is reached whenever the empirical distribution is indistinguishable from the predicted distribution using a one-sided KS test with significance level $\alpha = 0.05/j$, where j is the j :th time comparison. For each T , the MSE is evaluated based on 15000 random stimulus samples.

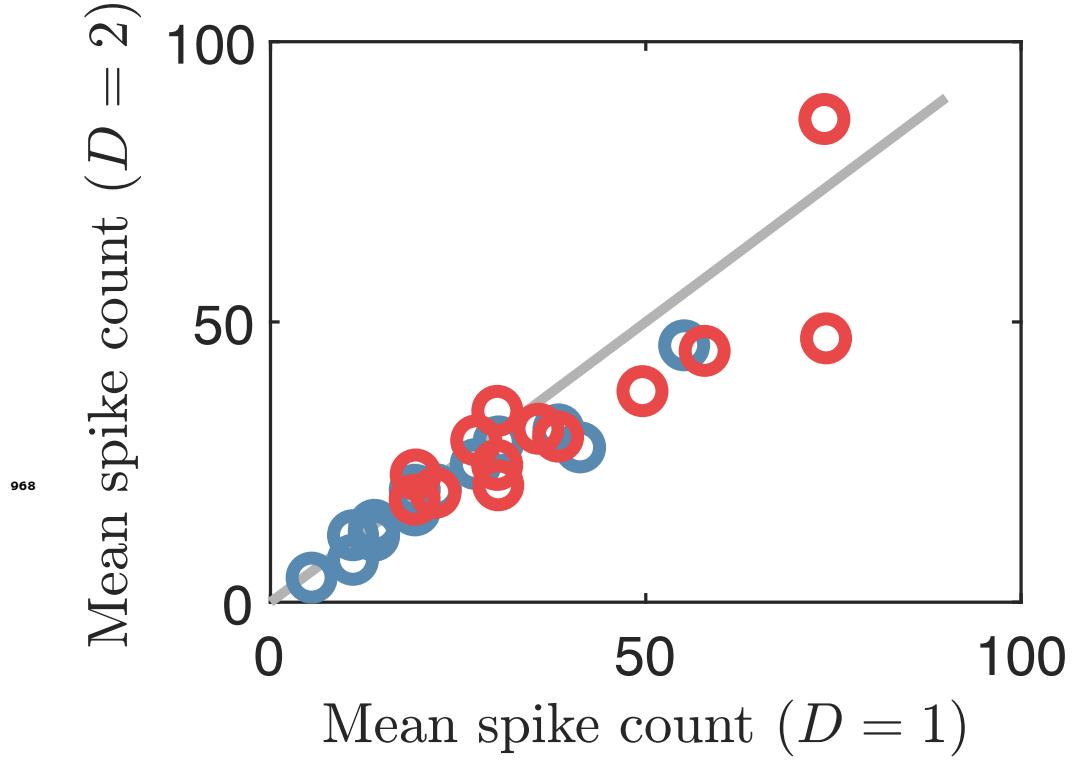


Figure 3—figure supplement 5. Plot of the mean spike counts (summed over the population) required to remove catastrophic errors for the populations in Figure 3. Each circle indicates the minimal spike count for a single population with a constant scale factor encoding either a 1-dimensional (x-axis) or a 2-dimensional stimulus (y-axis). Blue circles indicate $\lambda_1 = 1$ and red circles $\lambda_1 = 1/2$. Being on the grey line corresponds to having the same required spike count for both stimulus cases.

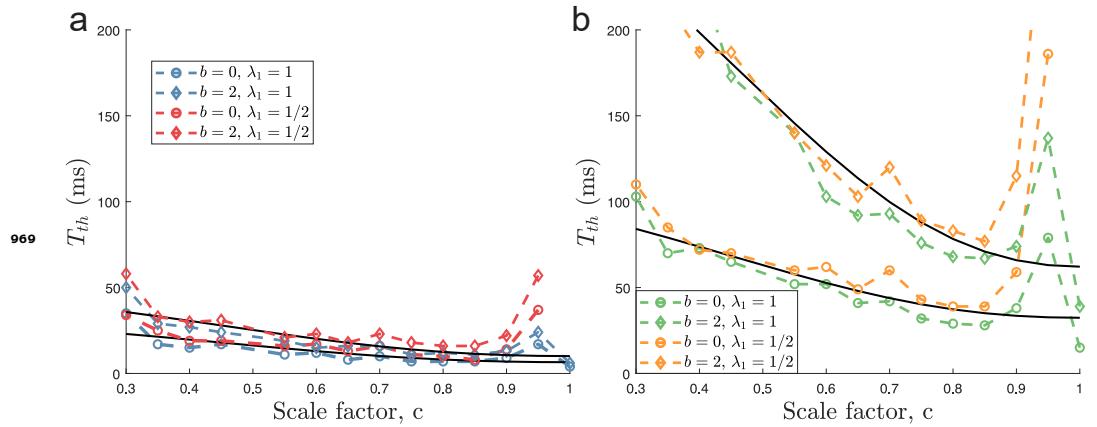


Figure 4—figure supplement 1. Same as Figure 4 in the main text, but using $\alpha = 1.2$.

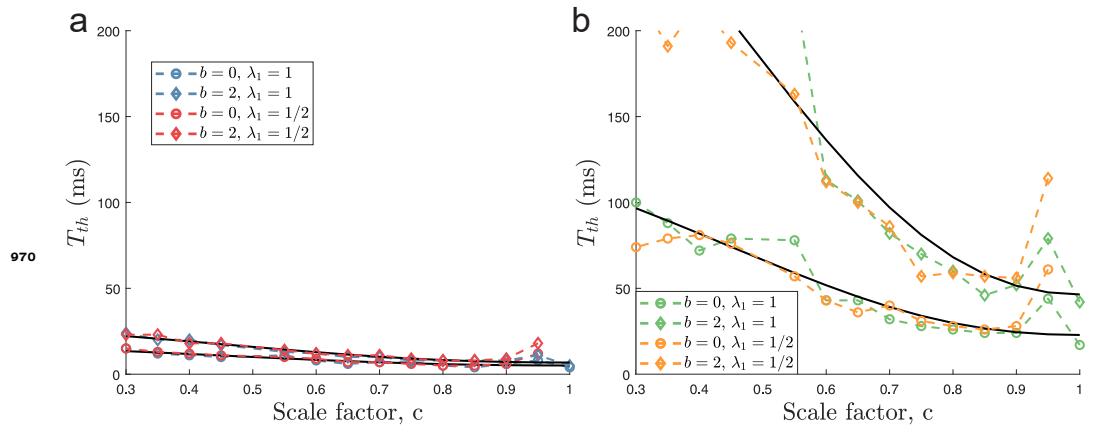


Figure 4—figure supplement 2. Same as Figure 4 in the main text, but using the one-sided KS-test criterion described before (see Figure 3 - Figure Supplement 3).