

Optimizing information transmission in optogenetic Wnt signaling

Olivier Witteveen¹, Samuel J. Rosen², Ryan S. Lach³, Maxwell Z. Wilson^{4,5,6,7,*} and Marianne Bauer^{1,†}

¹*Department of Bionanoscience, Kavli Institute of Nanoscience Delft, Technische Universiteit Delft, Van der Maasweg 9, 2629 HZ Delft, The Netherlands*

²*Interdisciplinary Program in Quantitative Biosciences, University of California, Santa Barbara, California 93106, USA*

³*Integrated Biosciences, Inc., Redwood, California 94065, USA*

⁴*Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, California 93106, USA*

⁵*Neuroscience Research Institute, University of California, Santa Barbara, California 93106, USA*

⁶*Biomolecular Science and Engineering, University of California, Santa Barbara, California 93106, USA*

⁷*Center for BioEngineering, University of California, Santa Barbara, California 93106, USA*



(Received 6 July 2025; accepted 9 February 2026; published 18 March 2026)

Populations of cells regulate gene expression in response to external signals, but their ability to make reliable collective decisions is limited by both intrinsic noise in molecular signaling and variability between individual cells. In this work, we use optogenetic control of the canonical Wnt pathway as an example to study how reliably information about an external signal is transmitted to a population of cells, and determine an optimal encoding strategy to maximize information transmission from Wnt signals to gene expression. We find that it is possible to reach an information capacity beyond 1 bit only through an appropriate, discrete encoding of signals: using no Wnt, a short Wnt pulse, or a sustained Wnt signal. By averaging over an increasing number of outputs, we systematically vary the effective noise in the pathway. As the effective noise decreases, the optimal encoding comprises more discrete input signals. These signals do not need to be fine-tuned to achieve near-optimal information transmission. The optimal code transitions into a continuous code in the small-noise limit, which can be shown to be consistent with the Jeffreys prior. We visualize the performance of different signal encodings using decoding maps. Our results suggest that optogenetic Wnt signaling allows for regulatory control beyond a simple binary switch and provide a framework to apply ideas from information processing to single-cell *in vitro* experiments.

DOI: [10.1103/PhysRevResearch.8.013296](https://doi.org/10.1103/PhysRevResearch.8.013296)

I. INTRODUCTION

Cells respond to external signals by adapting their gene expression [1]. However, gene regulatory responses can fluctuate [2–5]. Especially in the context of development, precise responses are important for coordinated cell-fate decisions that lead to the healthy development of an organism [6–8]; therefore, the considerable cell-to-cell variability that has been observed in the downstream targets of signaling pathways crucial for development may seem surprising [9–11].

The mutual information between a signal and an output quantifies the precision in information transfer [12,13]. Experiments on mammalian signaling pathways often report values barely exceeding one bit [14,15], which is the minimum amount of information required for a reliable decision between only two states, for example, a differentiated and

an undifferentiated state. To understand whether cells actually have more information available, a variety of approaches have been proposed in different biological contexts, including specific computational strategies, signaling architectures, cell-to-cell contact adjustments, or incorporating the information contained in temporal dynamics [9,10,16–21]. Here, we turn to an optimization over the space of possible input signals to investigate how information transmission in an important signaling pathway can be made more precise.

We focus on the canonical Wnt signaling pathway, a key regulator of cell-fate decisions during development and maintenance of adult tissues. Wnt signaling is crucial for the differentiation of stem cells into lineages such as skin, bone, and other tissues [22,23]. We study cellular responses using an established reporter of Wnt transcriptional activity, TopFlash, following optogenetic activation of the canonical Wnt signaling pathway [24,25]. We vary the duration of the optogenetic Wnt signals and observe long-tailed distributions of gene expression; we describe them using gamma distributions. Similar distributions have been widely observed for protein expression among cells in a population using fluorescent reporters [5], and are consistent with stochastic gene expression models [26,27].

We optimize information transmission between optogenetic inputs and gene expression outputs by determining the

*Contact author: mzw@ucsb.edu

†Contact author: M.S.Bauer@tudelft.nl

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

optimal prior distribution over input signals. We find that if signals are chosen uniformly from all possible signal durations, the mutual information is approximately 0.7 bits, whereas with a prior that involves only two discrete signals—no Wnt, or a long sustained Wnt signal—we can reach almost one bit. We can obtain more than one bit when we fully optimize the distribution of input signals, and find that the optimal prior consists of three different input signals: a Wnt off state, a short Wnt pulse, and a long sustained Wnt signal. Since different maps between input and output can have the same mutual information, we visualize which input signals can be distinguished based on the output using decoding maps, for which we provide analytic expressions whenever possible.

It is possible that cells have more information available than observed with our fluorescent reporter. This increased precision could be a consequence of different biological scenarios: for example, if cells average expressed proteins among neighbors [28–30], or if they derive differentiation outcomes in response to not one but multiple relevant Wnt targets [16,31–33]. Therefore, we explore how these optimal signals change if gene expression responses are represented by those we observe, but more precise; mathematically, we describe this increased precision by a narrowing gamma distribution. We use the Blahut-Arimoto algorithm to optimize the input distribution. A central result is that the optimal input signals are discrete for the noisy gene expression we observe and smoothly transition to a continuous encoding as the noise decreases. We show that the continuous distribution in the small-noise limit is equivalent to the Jeffreys prior. Finally, we present calculations to show that the discrete optimal priors yield “sloppy” optima [34,35]: This means that they do not need to be fine-tuned.

This signal-level optimization has potential application in engineering contexts: Our optimized signals correspond to those that external users should supply such that the output can be decoded with high precision. It may also offer insight into the signals that cells might encounter in natural contexts: Ideas from efficient coding [36] suggest that signal transmission is optimized, and therefore would predict that the signals we calculate should be those typically encountered by cells (see Refs. [37–39] for work on neurons).

While it is unclear if cells need to decode Wnt signal durations, how exactly Wnt provides information is also not yet established: hypotheses in different contexts include Wnt timing or duration [40–43], fold-changes [44,45], or absolute concentrations, either in gradients or dynamics [42,46–49]. Timing and duration of signals can play a key role in guiding differentiation [25,50,51], and we chose to investigate the duration of Wnt here because it is easily accessible in the opto-Wnt experiments. Since our work depends only on the parametrization of the conditional output distribution, our results will also apply to other gene regulatory outputs and input signals. We discuss how cells may be able to interpret signals with discrete priors inspired by recent work that optimizes an understanding or model of the world with finite samples [52].

More broadly, we present a systematic framework for investigating how heterogeneity in gene expression across a population impacts information transmission, including cases where the realistic biological noise is difficult to characterize.

II. CELLULAR RESPONSES TO WNT SIGNALING

We explore the expression of genes that respond to the canonical Wnt signaling pathway in a clonal established human embryonic kidney cell line (HEK293T) engineered to respond to optogenetic Wnt signals [24,25]. The duration of the Wnt signal can be varied experimentally, and we use this duration t as an input signal. We measure cellular responses to Wnt signaling using a synthetic fluorescent TCF/LEF (TopFlash-type) iRFP reporter that reflects the activation of Wnt/ β -catenin target genes [24]. This reporter is established for Wnt targets [24,53–55] and we refer to it as TopFlash from now on. At the molecular level, TopFlash and many canonical Wnt/ β -catenin target genes are activated as a result of β -catenin accumulation in the cytoplasm and nucleus, following the binding of extracellular Wnt ligands to membrane receptors [56].

We collect the output expression levels (fluorescent intensity per cell) of TopFlash, denoted by g , of ca. 1500 ± 800 cells to optogenetic Wnt input signals of varying durations t ranging from 0 to 20 h [Fig. 1(a)]. The experiment is conducted using a high-throughput light stimulation device, the LITOS plate, which enables optogenetic activation across multiple experimental conditions simultaneously (Appendix A) [57]. To ensure that the measured fluorescence has stabilized and to remove effects from residual signaling dynamics, we include a 4-h cool-down period after signal termination before measuring g . This allows Wnt pathway effectors, such as stabilized β -catenin, to return to baseline levels [25]. Since g is not degraded and cell division in the cool-down period is negligible, the value of g represents a robust measure for gene expression as a consequence of the Wnt pulse.

Histograms of g for a given signal duration t are left-skewed, long-tailed, and unimodal [see Fig. 1(b) for signal durations $t = 5, 10, 15$, and 20 h]. We observe that the histograms are well described by gamma distributions, and that the mean response $\mu_g(t)$ is directly proportional to the standard deviation $\sigma_g(t)$ [Fig. 1(c)]. The latter implies that the gamma distribution is parametrized by a constant *shape* parameter k and a time-dependent *scale* parameter $\theta(t)$:

$$p(g|t) = \frac{1}{\Gamma(k)\theta(t)^k} g^{k-1} e^{-g/\theta(t)}, \quad (1)$$

where Γ is a gamma function. The mean and variance of g are given by

$$\mu_g(t) = k\theta(t), \quad (2)$$

$$\sigma_g^2(t) = k\theta(t)^2, \quad (3)$$

respectively. The distribution of Eq. (1) predicts that one can collapse all histograms by normalizing each by their standard deviation [Fig. 1(d)]: Indeed, after rescaling, all data collapse onto a gamma distribution with shape parameter $\hat{k} = 2.88 \pm 0.01$ and unit variance.

To identify how the scale parameter $\theta(t)$ depends on the Wnt signal duration t , we plot the mean $\mu_g(t)$ for all experimental conditions [Fig. 1(e)]. For Wnt signals longer than ~ 1 h, we find that the mean gene expression grows linearly with the signal duration. Therefore, the scale parameter

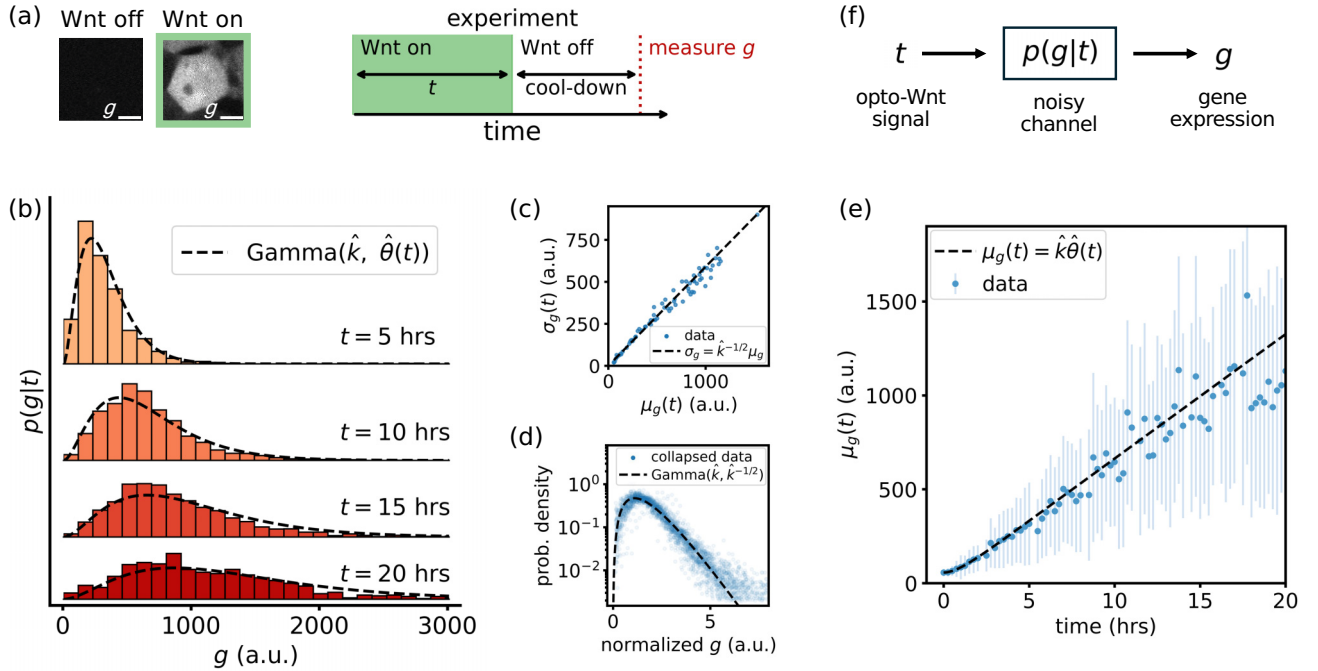


FIG. 1. (a) Optogenetic control of Wnt signaling. In the absence of light, there is no Wnt signal and no expression of TopFlash g . When the light is activated, Wnt target genes are expressed. We vary the duration t of the Wnt signal and measure the resulting gene expression g . (b) The histograms over g are long-tailed, left-skewed, and unimodal; shown here for Wnt signal durations $t = 5, 10, 15$, and 20 h. Black lines show the gamma distribution from Eq. (1), evaluated at the appropriate t . (c) The mean $\mu_g(t)$ of each histogram scales linearly with the standard deviation $\sigma_g(t)$. The black line shows the linear dependence predicted by Eq. (1). (d) Rescaling the histograms (dividing by the standard deviation) shows a collapse of gene expression data. The collapsed data are described well by a gamma distribution with shape parameter $\hat{k} \approx 2.88 \pm 0.01$ and unit variance (black line). (e) The mean gene expression $\mu_g(t)$ grows linearly with time, as captured by Eq. (1) (black line). Error bars show the standard deviation. (f) We can view our system analogously to a communication channel where input t is mapped to output g via the noisy transmission probability $p(g|t)$.

$\theta(t) \approx at$ must also grow linearly with time, where we estimate $\hat{a} = 23.0 \pm 0.1$ a.u. h^{-1} . We add a small correction term $\epsilon e^{-t/\tau}$ to the scale parameter $\theta(t)$ to fit the data in the regime $t \lesssim \tau \approx 1$ h (Appendix B).

We note that the gamma distribution was fit empirically and is thus a phenomenological description of the data. We choose it here due to its convenient parametrization. Other long-tailed distributions, such as log-normal, may also fit the data well. These long-tailed distributions, as well as the linear relationship between the standard deviation and the mean, have been suggested to hold universally for protein fluctuations among cells in a population [5] and emerge in bottom-up stochastic models of gene expression [26,27,58].

Next, we quantify how precisely we can reconstruct the Wnt signal from the gene expression. Given the broad, long-tailed distributions with substantial overlap between experimental conditions, we anticipate that the information transmission in the pathway will appear limited.

III. INFERRING THE WNT SIGNAL FROM GENE EXPRESSION IN SINGLE CELLS

We can view our system analogously to a communication channel $t \rightarrow g$ with transmission probability $p(g|t)$ [Fig. 1(f)]. As such, we quantify how much information about the input t is captured by the output g using the *mutual*

information [12,59]:

$$I(g;t) = \int_0^\infty dt \int_0^\infty dg p(g|t) p(t) \log_2 \left(\frac{p(g|t)}{p(g)} \right). \quad (4)$$

This mutual information $I(g;t)$ captures (in *bits*) how much we expect to learn about the Wnt signal by observing the gene expression. For the rest of this manuscript, we use the phrase “input signal” or “Wnt signal” to refer to the Wnt signal duration t .

The mutual information requires knowledge of the distribution of input signals $p(t)$, also referred to as the *prior distribution* [12,13,59]. A sensible prior distribution that favors no particular signal condition, like the experiment, is one that is uniform over all available signals $t \in [0, \infty)$. For this uniform prior, we obtain $I(g;t) \approx 0.67$ bits. Since 1 bit is the minimum required to reliably distinguish two states (e.g., an “on-off” switch), this result suggests that the gene expression carries less than the information required to support even a binary regulatory decision.

The numerical value of the mutual information can be difficult to assess abstractly. It is bounded from above by the entropy of the input distribution, which in turn depends on the size of the state space of possible input signals. For example, if the input distribution includes several discrete states, a mutual information of 1 bit does not necessarily imply that any two particular states are neatly distinguishable. Therefore, it can be useful to employ quantities other than the mutual

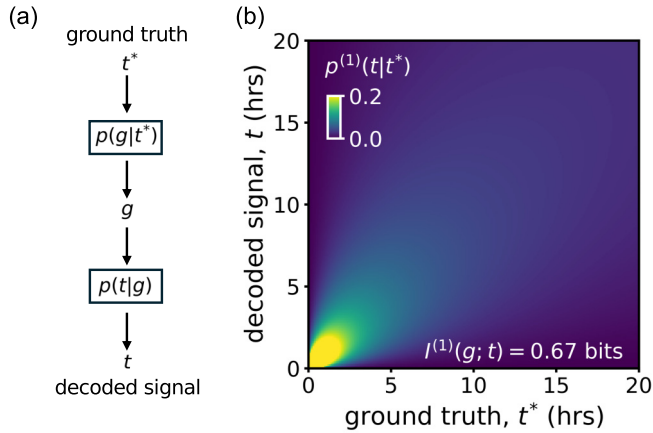


FIG. 2. (a) The input signal t^* leads to a gene output g drawn from the transmission probability $p(g|t^*)$. Based on a measurement of g , one can use the posterior distribution $p(t|g)$ to infer the input signal. (b) Decoding map $p^{(1)}(t|t^*)$ from Eq. (7), showing the average probability assigned to t by the posterior $p(t|g)$ given that the true signal is t^* . Here, the input distribution $p(t)$ is uniform over all possible signals $t \in [0, \infty)$ h.

information that allow us to more clearly identify which signals become confused in the information transmission from input to output.

To do so, we ask how well one can infer the optogenetic Wnt signal t from a measurement of the gene expression g . This is captured by the posterior distribution $p(t|g)$, which one obtains from Bayes' theorem, $p(t|g) = p(g|t)p(t)/p(g)$, where $p(g) = \int_0^\infty dt p(g|t)p(t)$. With a uniform distribution $p(t)$ over the interval $t \in [0, \infty)$, we find

$$p(t|g) = \frac{p(g|t)}{\int_0^\infty dt' p(g|t')} \approx a(k-1) p(g|t), \quad (5)$$

where the final expression is valid in the regime $t \gtrsim 1$ h. In principle, features of the posterior $p(t|g)$ can be used to quantify the precision in this inference problem. Decoding errors, such as the variance of inferred t around its true value, are often used to quantify inference precision [60]. However, such metrics rely on selecting a decoding rule, such as the posterior mean or MAP estimate, which may be misleading when the posterior $p(t|g)$ is skewed, heavy-tailed, or multimodal [61]. Here, we use a *decoding map* to quantify our ability to decode without subscribing to an estimator. Decoding maps have been used to quantify positional precision from gap gene expression patterns in the early fly embryo [61,62].

The decoding map quantifies the average posterior $p(t|g)$ generated from a true input t^* . To construct the decoding map, we consider a Markov chain in Fig. 2(a) and integrate out the regulatory output through which we intend to infer

$$p^{(1)}(t|t^*) = \int_0^\infty dg p(t|g) p(g|t^*). \quad (6)$$

The superscript “(1)” refers to the fact that we are considering gene expression from a single ($N = 1$) cell. While the benefit of decoding maps is most obvious for multidimensional g , where they provide a means to visualize the precision in the inference in a two-dimensional object, they can also be useful

for scalar g : We will use them later to visualize the performance of different signal encodings. If the gene expression provides enough information to reconstruct the Wnt signal accurately, the density $p^{(1)}(t|t^*)$ will be sharply peaked around the diagonal $t = t^*$.

We can compute the distribution $p^{(1)}(t|t^*)$ analytically in the regime $t, t^* \gtrsim 1$ h, by inserting the posterior from Eq. (5) into Eq. (6) and performing the change of variables $g' = g(1/\theta(t) + 1/\theta(t^*))$, to obtain

$$p^{(1)}(t|t^*) \approx \frac{\Gamma(2k-1)}{\Gamma(k)\Gamma(k-1)} \frac{(tt^*)^{k-1}}{(t+t^*)^{2k-1}}. \quad (7)$$

The distribution in Eq. (7) is a beta-prime distribution, and the normalizing constant can be identified as a beta-function $B(k, k-1) = \Gamma(k)\Gamma(k-1)/\Gamma(2k-1)$ [63]. We plot the decoding map in Fig. 2(b) and observe that the width of the decoding map broadens linearly with the Wnt signal duration t^* . This suggests that absolute decoding errors also grow proportional to t^* , while relative errors remain constant.

Next, we ask how reliable information transfer from Wnt to a single target gene could be possible. To do so, we note that the mutual information between the Wnt signal and gene expression depends not only on the channel $p(g|t)$, which we take as given from the experimental data, but also on how one chooses the input signals $p(t)$.

IV. OPTIMAL ENCODING OF WNT SIGNALS USES A DISCRETE DISTRIBUTION

As a first step toward optimizing the signal distribution, we focus on a binary input distribution consisting of a Wnt “off” state ($t = 0$ h) and a single Wnt “on” state of duration $t = \Delta t$ h [Fig. 3(a)]. We explore this setup, since for noisy channels with limited capacity (on the order of 1 bit or less), an efficient coding strategy is to use two maximally distinguishable signal states [12,64–67]. We find that an optogenetic Wnt signal of approximately $\Delta t \approx 10$ h is necessary to reliably distinguish the “on” state from the “off” state, approaching the information-theoretic upper bound $I(g; t) \leq 1$ bit.

This finding is of biological interest: cells may use Wnt signaling to make binary cell-fate decisions, for example, between remaining undifferentiated or committing to mesoderm [33]. In this context, our results suggest that such a binary decision is only reliable if the Wnt “on” signal persists for durations longer than ~ 10 h. This timescale is biologically realistic and lies well within the doubling time of the cells (ca. 20–30 h [68–70]). In addition, recent work in the intestinal crypt has shown that stem cells commit to differentiation after Wnt signaling is lost for ca. 10 h [50]. This observation suggests that a binary encoding scheme, based on the sustained presence or absence of Wnt signaling, could have biological relevance.

Next, we optimize the signal distribution $p(t)$ to obtain the maximally achievable mutual information or *channel capacity*:

$$I_* = \max_{p(t)} I(g; t). \quad (8)$$

The capacity-achieving distribution $p_*(t)$ tells us how to encode Wnt signals to create maximally distinguishable gene

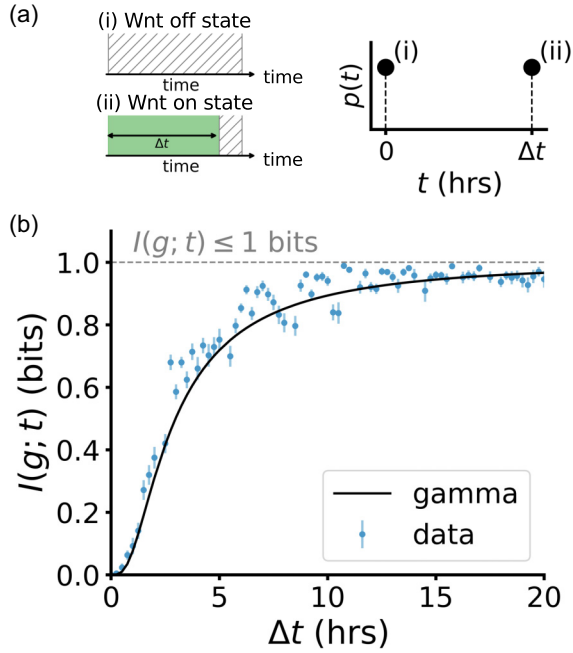


FIG. 3. (a) Binary input distribution of optogenetic Wnt signals, containing (i) an “off” state of $t = 0$ h and (ii) an “on” state of $t = \Delta t$ h. (b) Mutual information $I(g; t)$ as a function of the duration Δt of the on state. Results from the data (blue) and predictions from the gamma distribution in Eq. (1) (black) are shown. Error bars obtained via subsampling. The upper bound $I \leq 1$ (gray line) corresponds to perfect distinguishability of the on and off states.

expression outcomes within the noisy constraints. In most cases, this optimization is analytically intractable. Instead, we optimize numerically using the Blahut-Arimoto (BA) algorithm [71,72]. The algorithm converges to a discrete solution [Fig. 4(a)]: The optimal encoding of optogenetic Wnt signals selects a set of three discrete signals (or “symbols”) at t_1 , t_2 , and t_3 . We obtain a capacity of

$$I_{\star}^{(1)} \approx 1.12 \text{ bits}, \quad (9)$$

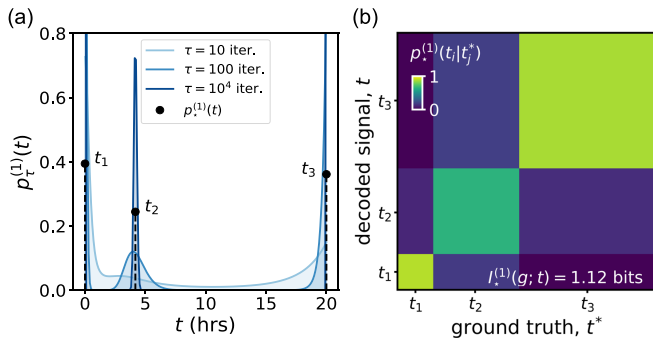


FIG. 4. (a) Optimal encoding of optogenetic Wnt signals for single cells: the Blahut-Arimoto algorithm converges to a discrete solution $p_{\star}^{(1)}(t)$, consisting of three optimally distinguishable Wnt signal durations or “symbols.” (b) Decoding map $p_{\star}^{(1)}(t|t^*)$ obtained using the optimal prior: at the cost of discretizing the space of input signals, we gain distinguishability [cf. Fig. 2(b)].

which is a significant improvement over the naive uniform encoding.

Convergence of the BA algorithm to the discrete solution $p_{\star}(t)$ is slow compared to the convergence to the information capacity $I_{\star}$, especially if the density of symbols is high. We can exploit the knowledge that $p_{\star}(t)$ is discrete to significantly accelerate convergence to the optimal solution [52,73,74]. To initialize the distribution, we use a weighted sum of K delta functions, representing K discrete symbols:

$$p_{\star}(t) = \sum_{i=1}^K w_i \delta(t - t_i). \quad (10)$$

We iteratively optimize their locations t_i using gradient descent, while updating the weights w_i using a BA-type update rule. To find the optimal K , we use lower and upper bounds to the information capacity to either add or remove a delta function after convergence (Appendix C).

It is interesting to note that under noisy conditions, nontrivial inputs beyond binary “on/off” encoding may be optimal for cells using Wnt signaling. The decoding map in Fig. 4(b) visualizes the way the optimal encoding improves information transmission: Even though the input signals are not perfectly distinguishable based on their outputs, information transmission is improved compared to a binary input distribution containing two (almost) completely distinguishable states.

V. DECODING FROM N OUTPUTS

We pursued the above analysis under the assumption that the probability with which a particular cell expresses a target gene g is set by the observed experimental response of TopFlash. However, the distribution of TopFlash observed in our cell culture may be noisier than a biologically relevant output of cells in realistic tissue settings [75–77]. Therefore, we ask in this section how our analysis changes if the biologically relevant output is systematically more precise than the TopFlash distributions we observe. For example, cells might communicate output with neighbors via surface signaling or molecular exchange, or have access to multiple regulatory targets. If the observed TopFlash distribution is representative also of other such outputs, we can think of these extra outputs as additional samples from the same distribution.

We consider N gene expression outputs $\mathbf{g} = (g_1, g_2, \dots, g_N)$, where each output g_i is sampled from a gamma distribution, and use $\mathbf{g}$ to decode the Wnt signal t . This can correspond to, for example, cells producing N regulatory outputs; alternatively, cells may have access to the mean output of N neighbors, $\bar{g} = \sum_{i=1}^N g_i / N$, through cell-cell communication. Mathematically, it turns out that both options are equivalent. If the responses g_i are independent, one can show that the sample mean $\bar{g}$ is a *sufficient statistic*, implying that all information contained in $\mathbf{g}$ about the input signal t is preserved in $\bar{g}$ (Appendix D). Indeed, for the HEK293T cells in our experiment, we find negligible correlation between TopFlash expression of neighboring cells (Appendix E). Hence, the mutual information satisfies $I(\bar{g}; t) = I(\mathbf{g}; t)$ and decoding is identical $p(t|\bar{g}) = p(t|\mathbf{g})$. Therefore, we can consider the sample mean $\bar{g}$ of N cells in what follows.

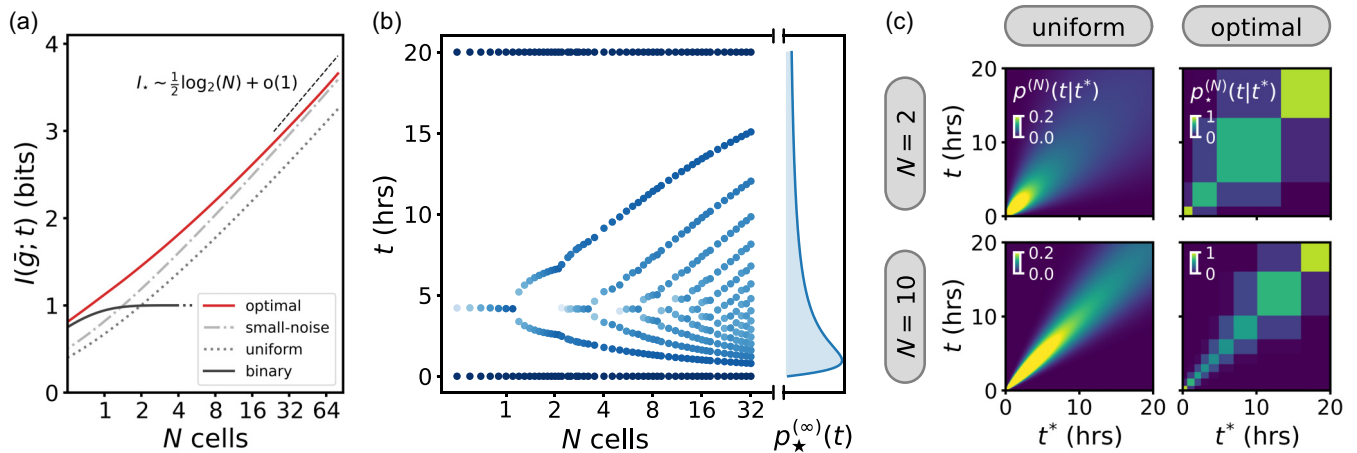


FIG. 5. (a) We show the information capacity achieved by the optimal encoding of Wnt signals (red line), and show convergence to analytical results in the small- and large-noise regimes. (b) The optimal code for Wnt signaling consists of a discrete number of symbols (blue dots). As the effective noise decreases, the optimal number of symbols increases, and approaches a continuous optimal code $p_\star^{(\infty)}(t)$ with a heavy tail that decays as $\sim 1/t$. The color of the markers (blue shade) indicates the relative probability mass of the symbols: darker blue indicates a higher probability mass. (c) Decoding maps visualize how encoding strategies affect signal inference. Shown are decoding maps $p^{(N)}(t|t^*)$ for ensembles of $N = 2$ (top row) and $N = 10$ (bottom row) cells. A uniform prior over opto-Wnt signals (left column) leads to broad posterior distributions, while the optimized discrete choice of signals (right column) yields more distinguishable responses and higher mutual information $I_\star^{(N)}(\bar{g}; t)$, at the cost of discretizing the space of input signals.

Since $\bar{g}$ is the mean of N identically gamma-distributed random variables with shape parameter k and scale parameter $\theta(t)$, its distribution is also gamma, with shape parameter Nk and scale parameter $\theta(t)/N$. The distribution of $\bar{g}$ given a Wnt signal t is thus

$$p(\bar{g}|t) = \frac{1}{\Gamma(Nk)(\theta(t)/N)^{Nk}} \bar{g}^{Nk-1} e^{-N\bar{g}/\theta(t)}. \quad (11)$$

The mean and variance are given by $\mu_{\bar{g}}(t) = k\theta(t)$ and $\sigma_{\bar{g}}^2(t) = k\theta(t)^2/N$, respectively; as such, N in Eq. (11) not only represents an integer number of samples, but can also be seen as a continuous parameter that changes the effective noise by a factor of $1/\sqrt{N}$.

We now ask how the mutual information and decoding change as we increase N . We can expect the mutual information to scale asymptotically as $I \sim (1/2) \log_2 N$ (Appendix F). We confirm this scaling for both a prior distribution $p(t)$ that is uniform and one that is optimized [Fig. 5(a)]. The mutual information using the binary prior, where our input is restricted to two signal durations, is limited by definition to 1 bit and is not a good choice for maximizing the mutual information when the effective map from input to output becomes more precise. However, in regimes where the noise is large ($N \lesssim 1$), a binary encoding comes close to achieving the information capacity.

In the limit of large N , or small effective noise, one can derive an analytic expression for the optimal $p(t)$ [6,13,62,78–82]. In this small-noise approximation, we assume that $p(\bar{g}|t)$ is a narrow Gaussian distribution, and that we can calculate $p(t|\bar{g})$ by performing an expansion (Appendix F); then, taking a variational derivative of $I(\bar{g}; t)$ with respect to $p(t)$, one finds

$$p_\star^{(\infty)}(t) \propto \frac{1}{\sigma_{\bar{g}}(t)} \left| \frac{d\mu_{\bar{g}}(t)}{dt} \right|. \quad (12)$$

Since both $\mu_{\bar{g}}(t)$ and $\sigma_{\bar{g}}(t)$ grow linearly with t for longer Wnt signals, it follows from Eq. (12) that the tail of $p_\star^{(\infty)}(t)$ decays as $\sim 1/t$.

As expected, the small-noise approximation approaches the information capacity from below [Fig. 5(a)] and is a good approximation for $N \gtrsim 20$. Notably, we can show that this optimal continuous encoding from the small-noise limit in Eq. (12) is equivalent to the *Jeffreys prior* (Appendix F). The Jeffreys prior is a noninformative prior that is invariant to changes in parametrization, defined as $p_J(t) \propto |\mathcal{I}(t)|^{1/2}$, where $\mathcal{I}(t)$ is the *Fisher information* [83]:

$$\mathcal{I}(t) = \int_0^\infty d\bar{g} p(\bar{g}|t) \left(\frac{\partial \ln p(\bar{g}|t)}{\partial t} \right)^2. \quad (13)$$

Indeed, it is known that in the limit of an infinite number of identical, independent trials of the same experiment (i.e., $N \rightarrow \infty$), the prior that maximizes the mutual information between input and output converges weakly to the Jeffreys prior [84].

Next, we investigate how the numerically optimized prior $p_\star^{(N)}(t)$ changes as N increases. Since we know that the optimal prior consists of three discrete symbols for $N = 1$ and should approach the continuous distribution in Eq. (12) for large N , we expect that it will admit an increasing number of symbols as N increases. We find that this is indeed the case: Figure 5(b) shows a bifurcationlike diagram of the positions and weights of the optimal prior distribution, where symbols split into two and additional symbols are added as N increases. For high N , the density of the symbols starts approaching the optimal distribution $p_\star^{(\infty)}(t)$ from the small-noise approximation.

As before, we visualize how the optimization improves decoding performance and the mutual information with decoding maps [Fig. 5(c)]. Unlike the uniform prior, which leads

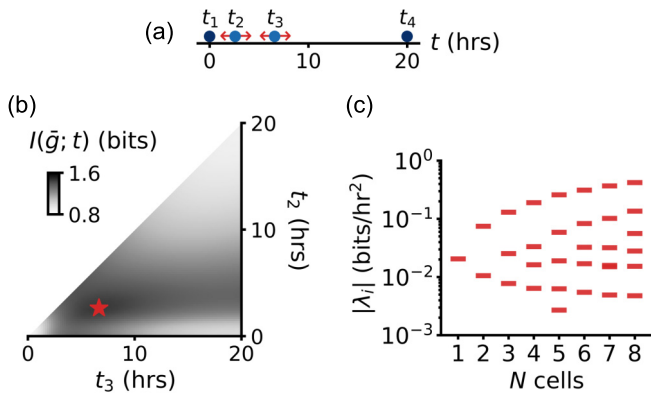


FIG. 6. The exact duration of individual symbols does not need to be fine-tuned. (a) For $N = 2$, the optimal prior consists of $K = 4$ symbols [cf. Fig. 5(b)]. We vary the positions of two symbols t_2 and t_3 , while keeping their weights fixed. (b) Information $I(\bar{g}; t)$ (color-shade) shows a broad optimum (red star) as a function of the position of peaks t_2 and t_3 . (c) The Hessian matrix [Eq. (14)] has a sloppy spectrum that widens as N increases: Symbols at longer durations become more sloppy and symbols at shorter durations become more stiff.

to smoothly narrowing posteriors as N increases (approaching a Gaussian for large N), the optimized prior increases the mutual information by admitting more discrete symbols. The optimal number of symbols K in the optimal prior $p_\star^{(N)}(t)$ follows an asymptotic scaling law $I_\star \sim (3/4) \log_2 K$, consistent with recent literature (Appendix G) [52,74]. The discretization enables better distinguishability between inputs, as illustrated by increased activity along the diagonal of the decoding map. At the same time, the optimal prior does not achieve perfect distinguishability of symbols, reflected by the remaining off-diagonal activity. The optimal input distribution, like in the $N = 1$ case, finds a balance between distinguishability and adding additional symbols. An alternative strategy is to encode fewer symbols $K' < K$ than optimal, and approach the bound $\log_2 K'$ bits: This leads to better distinguishability but does not achieve the maximum mutual information.

The fact that the optimal input distribution is discrete is interesting: there may be biological situations in which cells want to distinguish between a discrete number of cell fates, such as the different germ layers. In practice, cell fates are regulated by complex networks of multiple input signals and genetic targets, a more intricate setting than the one we study here. It is therefore not clear if our optimal input distribution carries direct biological meaning. Yet, it is conceivable that also for these more complex input spaces, discrete inputs are optimal and it is therefore interesting to investigate if cells attempt to map signals they receive to discrete input states. We also note that the optimal input distribution is one that permits slight errors, which suggests that some uncertainty in inference of signals is inherently part of optimal information processing; as such, some observed cellular noise in differentiation could be part of an information-maximization strategy.

We observe that the numerical optimization for the optimal prior distribution converges more quickly to the correct value of the mutual information than to the correct number of delta

functions K and their positions t_i , especially as N becomes larger [52]. This implies that the information landscape at the optimum is smooth, and has some directions where parameters for the prior distribution still change while the optimum is almost attained. These directions are typically referred to as “sloppy” directions [34,35] and their presence has important implications for the ability of biological systems to show variability in parameter space, even at the optimum [85]. Indeed, in Fig. 6(a) we show the mutual information for $N = 2$ as a function of the positions of two out of four delta functions and observe a broad optimum with different sensitivities depending on the direction one moves away from the optimum.

The sloppiness can be quantified using the Hessian matrix of the cost-function, in this case the mutual information $I(\bar{g}; t)$ for a given N [62,85]. Calculating this Hessian can be numerically difficult. Here, we have access to the functional form of the probability distribution $p(g|t)$, and can therefore calculate the Hessian with respect to the positions t_i of the discrete symbols in the optimal encoding $p_\star^{(N)}(t)$ (Appendix H). Writing $p_\star^{(N)}(t)$ as in Eq. (10), the Hessian matrix becomes

$$\chi_{ij} = \frac{\partial^2 I(\bar{g}; t)}{\partial t_i \partial t_j} = \frac{w_i}{\ln 2} \int_0^\infty d\bar{g} \left\{ \delta_{ij} \left[\frac{\partial^2 p(\bar{g}|t_i)}{\partial t_i^2} \ln \left(\frac{p(\bar{g}|t_i)}{p(\bar{g})} \right) + \frac{1}{p(\bar{g}|t_i)} \left(\frac{\partial p(\bar{g}|t_i)}{\partial t_i} \right)^2 \right] - \frac{w_j}{p(\bar{g})} \frac{\partial p(\bar{g}|t_i)}{\partial t_i} \frac{\partial p(\bar{g}|t_j)}{\partial t_j} \right\}. \quad (14)$$

The eigenvectors of χ determine directions in parameter space t_i that have independent effects on the mutual information, and the eigenvalues λ_i tell us the sensitivity along these directions. We evaluate the Hessian at the stationary point that maximizes $I(\bar{g}; t)$. After diagonalization, we indeed observe a sloppy spectrum [Fig. 6(b)], with eigenvalues spanning ca. two decades. The most stiff eigendirections correspond to the shorter durations, where the density of symbols is highest. As N increases, the spectrum broadens: symbols at longer durations becoming more sloppy, and those at shorter durations become more stiff. Practically, the fact that the optimal prior is sloppy implies that the optimal signal encoding does not need to be fine-tuned [62,85]; this could indeed be one advantage of information transmission using channels with similarly long-tailed distributions of gene expression outputs.

In deriving Eq. (14), we assumed weights w_i are fixed; alternatively, one can keep the weights optimized while varying the position of the symbols. In that case, we also obtain a sloppy spectrum (Appendix H). We emphasize that Eq. (14) (and the extension for variable weights in Appendix H) are general and can be used to obtain the sensitivity spectrum of a discrete prior for any choice of transmission probability.

VI. DISCUSSION

In this work, we optimized information transmission from optogenetic Wnt signals to (representative) Wnt target gene expression. We measured expression of a fluorescent reporter in a HEK cell line for different optogenetic signal durations, and found that gamma distributions with a constant shape parameter describe the output distributions well across different input conditions. We calculated the distribution of input

signals that maximizes information transfer and found that it consists of a discrete set of signals that allows for optimal decoding from the output. We also explored how this optimal set of signals would change if cells had access to multiple instances of the output we measure: for example, because they average multiple genetic outputs or exchange outputs with neighboring cells. As the effective noise decreases, the optimal input distribution evolves from three discrete symbols to a continuous distribution: We showed that the latter can be obtained from the small-noise approximation or, equivalently, the Jeffreys prior. Using decoding maps, we visualized how the optimal choice of optogenetic inputs improves signal inference, and finally, we showed that the optimal signals do not need to be fine-tuned. In the following, we discuss the biological and theoretical impact of our findings.

a. Discrete Wnt input signals. The canonical Wnt signaling pathway is active in different types of cells, often with the goal to direct mammalian cells toward differentiation. The effective features of Wnt signals that cells respond to depend on context and may include absolute or relative concentrations, thresholds, timers, and combinatorial interactions with other signaling pathways. Here, we focus on the Wnt signal duration as a potential input, since it can be precisely controlled in our optogenetic setup. Durations of Wnt signals may have developmental relevance, particularly in the context of organoid development [41–43,50]. We found that two Wnt input signals (a binary prior) provide more information than a uniform distribution when cells have to decide based on a single output: When the Wnt input is either no signal or a signal lasting at least ca. 10 h, the output expression is precise enough to reliably distinguish between the two. Such a binary response between two states, e.g., cell differentiation or remaining in the undifferentiated state, could be relevant for Wnt biologically [86,87]. This timescale of ca. 10 h could also be biologically reasonable: it fits well within a cell's division cycle, and connects to recent findings in gut cells, where differentiation is triggered only after Wnt signal loss lasting about 10 h, while shorter transient losses leave cells undifferentiated [50].

Our work to infer optimal priors would imply, in analogy with ideas with efficient coding, that cells either experience strongly different Wnt signals, or map the Wnt signals they experience onto signals that are strongly differentiable. This idea is similar to ideas from optimal parameter inference with limited data [52]. While it is unclear whether this happens biologically, it is interesting to observe that both recent work on differentiation timers [50], as well as clustering of cells with similar inputs [88], could suggest that cells indeed try to map the input they receive onto clear binary states (here, the presence or absence of Wnt). In our analysis, we find that cells could, in principle, resolve a third, intermediate state as well. We do not know whether this additional state is used as such *in vivo*, but it seems possible that multiple Wnt input states can be decoded in some contexts [89,90] and it is interesting that this possibility emerges purely from the distributions observed in the experiment and a model-free optimization framework.

In addition to suggestions for what signals cells may receive naturally, our work also has meaning from an engineering perspective: the optimal optogenetic signals we infer could be used with the goal to engineer signals that

populations of cells will be able to respond to precisely. This could be interesting also in the context of synthetic development.

b. Optimal priors for systematically lower noise, connections to population coding, and decoding maps. Our work connects to previous work on optimizing information transmission in gene regulation. That a binary prior (or switch) can optimize transmission for noisy systems has also previously been discussed in the context of genetic networks [62,66,67,80]. An interesting and perhaps more surprising finding of our work is the increasing number of discrete signals as the effective noise in the signaling pathway decreases: While the precise phase diagram of symbols depends on the underlying distribution, in general, additional symbols emerge as the noise decreases since the Shannon-optimal prior balances complexity with the ability to resolve different input signals. Reference [52] showed this effect in the context of Bayesian inference, to predict what models should be inferred for noisy data; in our work, it is either the cell or the engineer that would employ the optimal priors for reliable signal processing. The fact that there is sloppiness of the exact positions of these signals can make near-optimal information transmission easier than expected.

We investigated signaling via multiple cells or outputs as a way to improve information transmission; this is equivalent to using multiple communication channels and is related to population coding in neural systems [91–93]. An alternative approach to consider if information transmission from signals to outputs is higher than observed is to consider repeated stimuli on single cells: indeed, recent work shows that single cells responding to repeated optogenetic stimuli show a more reproducible output—higher mutual information—compared to population-level measurements [94].

To visualize how the encoding of input signals influences inference, we use decoding maps alongside the mutual information. They are reminiscent of recurrence plots for nonlinear dynamical systems, showing at what times a dynamical system reverts to a state it has visited before. Recurrence plots can be viewed as a measure of the “predictability” of the system [95], similar to how decoding maps can visualize ambiguity in inference from multimodal posterior distributions [61,62]. In our case, the decoding maps visualize how the optimal encoding discretizes the input space to improve distinguishability between signals, and how signal inference becomes more precise as the effective noise decreases.

c. Outlook and future work. Our work is directly applicable to gene expression responses that follow gamma distributions, and we expect qualitatively similar results for other long-tailed distributions that are commonly observed for gene expression [5,26,27,58]. This suggests that the implications of our work may go beyond Wnt signaling. Our systematic reduction of noise would also predict for those signaling pathways that discrete input distributions maximize information flow. While we do not know whether these discrete intermediate durations are a relevant input *in vivo*, Wnt/ β -catenin systems exhibit multistate dynamic decoding in other contexts, and developing embryoid systems pass through intermediate Wnt activity states during patterning [89,90]. In the context of general gene regulatory responses, it seems possible that smooth inputs are mapped onto a discrete but

not infinite number of outputs, such as a finite number of segments in the fly embryo body plan: In this context, it may be beneficial to map signal inputs onto different discrete states. A more detailed investigation of evidence of the use of discrete signal priors in gene regulatory or other biological contexts is an interesting direction for future work.

Our work relates closely to the problem of selecting effective models that maximize the information extracted from finite data, as discussed in Ref. [52]. In their framework, this corresponds to choosing a Bayesian prior that maximizes the mutual information between parameters and predictions. In our system, there are two concrete applications: First, the optimal prior can predict signal durations that should be used in synthetic, engineered experiments if cells are to respond distinguishably. Second, if cells operate consistent with the model selection work of Ref. [52], the cell itself should map the signal onto a discrete state space to optimally extract information. It will be interesting to explore the optimal input distributions in multi-input settings [33,96,97] in the future.

ACKNOWLEDGMENTS

We thank Florian Berger, William Bialek, Aneta Koseska, Pieter Rein ten Wolde, and the members of the Bauer group for useful discussions. We acknowledge funding from the NWO VIDI Talent Programme, NWO/VI.Vidi.223.169 (M.B. and O.W.). We also acknowledge funding from the U.S. Army Research Office and experiments were accomplished under contract W911NF-19-D-0001 and cooperative agreement W911NF-19-2-0026 for the Institute for Collaborative Biotechnologies (S.J.R., R.S.L., and M.Z.W.).

DATA AVAILABILITY

The data that support the findings of this article are openly available [98,99].

APPENDIX A: EXPERIMENTAL METHODS

Human 293T cells were cultured at 37°C and 5% CO₂ in Dulbecco's Modified Eagle Medium, high glucose GlutaMAX (Thermo Fisher Scientific, 10566016) medium supplemented with 10% fetal bovine serum (Atlas Biologicals, F-0500-D) and 1% penicillin-streptomycin (Thermo Fisher Cat. No. 15140-122). Clonal 293Ts containing CRISPR tdmRuby3- β -cat, oLRP6-Puro (AddGene ID: 249712), and pPig-8X-TOPFlash-tdIRFP-Puro (AddGene ID: 249713) were obtained from Dr. Ryan Lach's previous experiments [24,25].

We seeded these cells onto a 96-well glass-bottom plate (Cellvis Cat. No.: P96-1.5H-N) coated with fibronectin (Thermo Fisher Cat. No.: 33010018) in Dulbecco's Modified Eagle Medium, high glucose GlutaMAX (Thermo Fisher Scientific, 10566016) medium supplemented with 10% fetal bovine serum (Atlas Biologicals, F-0500-D) and 1% penicillin-streptomycin. Cells were allowed to adhere to the plate for 24 h. Afterward, cells were stimulated via a benchtop LED array purpose-built for light delivery to cells in standard tissue culture plates (LITOS) [57]. Light was patterned to cover the entire surface of intended wells of plates used, rather than a single microscope imaging field. Post-LITOS

stimulation, cells were moved into dark conditions for 4 h to ensure the Wnt pathway was properly deactivated prior to imaging. After 4 h, cells were then imaged using a Nikon W2 SoRa spinning-disk confocal microscope equipped with an incubation chamber maintaining cells at 37°C and 5% CO₂. We then used our previously published image analysis pipeline (methods, image analysis from Ref. [25]). We quantified the TopFlash-tdiRFP fluorescent intensity (measured in arbitrary units, a.u.) by taking the median pixel value for each individual segmented cell; we refer to this quantity as g . The data from the experiment are openly available in Ref. [98], along with the code used for its analysis [99].

We analyzed g for populations of ca. 1500 ± 800 cells that were optically stimulated for up to 24 h, which is of order of the doubling time of 293T cells (20–30 h [68–70]). We included data of up to until 20 h in this experiment, since 20 h is on the lower end of possible cell cycle estimates, and we observe a clear linear increase of mean TopFlash expression with time in this time frame. After 1–3 division cycles, cells grow beyond the volume of the plate, and tracking becomes more difficult as this time frame is approached. Therefore, a time frame of ca. 20 h marks both a natural end to our experiment and lies well within the regime of reliable experimental data acquisition. Finally, we note that cell division is negligible during ca. 4-h cooldown window after the Wnt signal, since only ca. 15% of cells are expected to divide in this time frame. We therefore expect no decay of TopFlash intensity in this regime, and TopFlash accumulation has been found to be stable in similar experiments even for over 10 h after the end of the Wnt “on” condition; for details on TopFlash dynamics, we refer to Ref. [25]. Analysis from the smaller dataset of Ref. [25] (not shown) suggests that TopFlash distributions at longer cool-down windows are also gamma-distributed with the same shape parameter k but different scale parameters; since most analytic results in our manuscript depend explicitly on k , our results do not sensitively depend on the cool-down window and could easily be extended to others.

APPENDIX B: FITTING THE GENE EXPRESSION DATA

Gene expression distributions observed in cell cultures are frequently long-tailed, and have been successfully modeled using, e.g., gamma, negative binomial, or log-normal distributions [5]. These choices are not only empirically well matched to the data, but also mechanistically plausible in systems where gene expression is shaped by multiple interacting timescales [26,27,58]. In our case, we find that our distributions are particularly well described by a gamma distribution:

$$p(g|k, \theta(t)) = \frac{1}{\Gamma(k)\theta(t)^k} g^{k-1} e^{-g/\theta(t)}, \quad (\text{B1})$$

parametrized by a constant shape parameter k and a time-dependent scale parameter $\theta(t)$. In fact, one can show that this parametrization is our only choice given that the mean $\mu_g(t)$ and standard deviation $\sigma_g(t)$ are directly proportional [Fig. 1(c)]. Our aim here is to estimate k and $\theta(t)$ from the data using a maximum-likelihood estimate.

We denote the data for TopFlash as g_{ij} and the signal durations as t_i , where $i = 1, \dots, n$ and $j = 1, \dots, m_i$. The i th experimental condition is populated by m_i different cells at

the end of the experiment. We first consider the regime where $\theta(t) \approx at$. Thus, our task is to estimate k and a . The likelihood function is given by

$$\mathcal{L}(k, a) = \prod_{i=1}^n \prod_{j=1}^{m_i} p(g_{ij}|k, a, t_i), \quad (\text{B2})$$

and hence our log-likelihood is

$$\begin{aligned} \ln(\mathcal{L}(k, a)) &= \sum_{i=1}^n \sum_{j=1}^{m_i} \ln(p(g_{ij}|k, a, t_i)) \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} \left[-\ln(\Gamma(k)) - k \ln(a) - k \ln(t_i) \right. \\ &\quad \left. + (k-1) \ln(g_{ij}) - \frac{g_{ij}}{at_i} \right]. \end{aligned} \quad (\text{B3})$$

Setting

$$\left. \frac{\partial \ln(\mathcal{L})}{\partial a} \right|_{\hat{k}, \hat{a}} = 0, \quad (\text{B4})$$

we obtain

$$\hat{a} = \frac{1}{\hat{k}} \left(\frac{1}{\sum_{l=1}^n m_l} \sum_{i=1}^n \sum_{j=1}^{m_i} \frac{g_{ij}}{t_i} \right). \quad (\text{B5})$$

Further, setting

$$\left. \frac{\partial \ln(\mathcal{L})}{\partial k} \right|_{\hat{k}, \hat{a}} = 0, \quad (\text{B6})$$

we get

$$\begin{aligned} \psi^{(0)}(\hat{k}) &= \ln(\hat{k}) + \frac{1}{\sum_{l=1}^n m_l} \sum_{i=1}^n \sum_{j=1}^{m_i} \left[-\ln(g_{ij}) \right. \\ &\quad \left. + \ln \left(\frac{1}{\sum_{r=1}^n m_r} \sum_{p=1}^n \sum_{q=1}^{m_p} \frac{g_{pq}}{t_p} \right) + \ln(t_i) \right], \end{aligned} \quad (\text{B7})$$

where $\psi^{(0)}(k) = d \ln(\Gamma(k))/dk$ is the polygamma function of order 0. We can solve this numerically for $\hat{k}$ without too much trouble, though we can continue analytically to very good approximation using the asymptotic expansion:

$$\psi^{(0)}(k) \sim \ln(k) - \sum_{l=1}^{\infty} \frac{B_l}{lk^l}. \quad (\text{B8})$$

Here, B_l are Bernoulli numbers with the convention $B_1 = +\frac{1}{2}$. Keeping the first two terms in the sum, we obtain the following estimator for k :

$$\hat{k} = \frac{1 + \sqrt{1 + \frac{4}{3}C}}{4C}, \quad (\text{B9})$$

where

$$\begin{aligned} C &= \frac{1}{\sum_{l=1}^n m_l} \sum_{i=1}^n \sum_{j=1}^{m_i} \left[-\ln(g_{ij}) \right. \\ &\quad \left. + \ln \left(\frac{1}{\sum_{r=1}^n m_r} \sum_{p=1}^n \sum_{q=1}^{m_p} \frac{g_{pq}}{t_p} \right) + \ln(t_i) \right]. \end{aligned} \quad (\text{B10})$$

We obtain, using Eqs. (B5) and (B9),

$$\hat{k} = 2.88 \pm 0.01, \quad (\text{B11})$$

$$\hat{a} = 23.0 \pm 0.1 \text{ h}^{-1}, \quad (\text{B12})$$

where the error is obtained using the asymptotic normality of maximum-likelihood estimators.

If we take the scale parameter to be $\theta(t) = at$, the gamma distribution in Eq. (B1) has a singularity at $t = 0$. The singularity is not physical, as it implies that any signal $t > 0$ is perfectly distinguishable from $t = 0$, and more importantly does not match the experimental data in this regime. To regularize the behavior of $p(g|t)$ near $t = 0$, we add a small exponential term to the scale parameter:

$$\theta(t) = a(t + \epsilon e^{-t/\tau}), \quad (\text{B13})$$

where $\hat{\epsilon} = 0.86\text{h}$ and $\hat{\tau} = 0.87\text{h}$ are positive constants that we estimate from the data. The mean and variance are therefore given by

$$\mu_g(t) = ka(t + \epsilon e^{-t/\tau}), \quad (\text{B14})$$

$$\sigma_g^2(t) = ka^2(t + \epsilon e^{-t/\tau})^2, \quad (\text{B15})$$

respectively. At times $t \gtrsim \tau$, the exponential term becomes irrelevant and we recover $\theta(t) \approx at$.

APPENDIX C: ALGORITHM FOR COMPUTING THE CHANNEL CAPACITY

We want to maximize the mutual information $I(g; t)$ with respect to the input distribution $p(t)$:

$$I_* = \max_{p(t)} I(g; t), \quad (\text{C1})$$

where I_* is the channel capacity and $p(g|t)$ is fixed. In most cases, this optimization is analytically intractable and we must proceed numerically. The BA algorithm is the standard algorithm for solving this problem [71,72]. One starts with an initial guess $p^{(0)}(t)$ for the input distribution, and with each iteration it is updated as follows:

$$p^{(\tau)}(t) = \frac{1}{Z^{(\tau-1)}} p^{(\tau-1)}(t) e^{f_{\text{KL}}^{(\tau-1)}(t)}, \quad (\text{C2})$$

where

$$f_{\text{KL}}^{(\tau)}(t) = \int_0^\infty dg p(g|t) \ln \left(\frac{p(g|t)}{p^{(\tau)}(g)} \right) \quad (\text{C3})$$

and

$$p^{(\tau)}(g) = \int_0^T dt p(g|t) p^{(\tau)}(t). \quad (\text{C4})$$

Practically, we restrict ourselves to a finite domain $t \in [0, T]$, where T is the maximum signal duration.

After τ iterations, the lower bound to the channel capacity is given by

$$I_L^{(\tau)} = \frac{1}{\ln 2} \int_0^T dt p^{(\tau)}(t) f_{\text{KL}}^{(\tau)}(t), \quad (\text{C5})$$

and an upper bound is given by

$$I_U^{(\tau)} = \max_t \frac{f_{\text{KL}}^{(\tau)}(t)}{\ln 2}. \quad (\text{C6})$$

As such, we iterate until convergence

$$I_U^{(\tau)} - I_L^{(\tau)} < \epsilon \quad (\text{C7})$$

and use $I_* \approx I_L^{(\tau)}$ and $p_*(t) \approx p^{(\tau)}(t)$ as our estimates for the channel capacity and the optimal input distribution, respectively. The optimization problem in Eq. (C1) is convex and guaranteed to converge to the global maximum [12].

As shown in Fig. 4 and as noted by Mattingly *et al.* [52], convergence to a discrete solution in the interior of the domain $t \in [0, T]$ is rather slow compared to the boundaries, especially when there is high density of delta functions. To overcome this, we can exploit the knowledge that $p_*(t)$ is discrete by starting with a sum of K delta functions:

$$p(t) = \sum_{i=1}^K w_i \delta(t - t_i) \quad (\text{C8})$$

and adjusting the weights w_i and positions t_i iteratively. Starting with K equally spaced delta functions, we use the BA algorithm to adjust the weights w_i until convergence. After this, we use the gradient

$$\frac{\partial I}{\partial t_i} = \frac{w_i}{\ln 2} \int_0^\infty dg \frac{\partial p(g|t_i)}{\partial t_i} \ln \left(\frac{p(g|t_i)}{p(g)} \right) \quad (\text{C9})$$

to adjust the positions t_i . Iterating the adjustment of the weights via the BA algorithm and the positions by gradient ascent, we can converge to the optimal discrete solution.

In contrast to the BA algorithm, the optimization over w_i and t_i is not convex: in particular, it depends on the number of peaks K we define beforehand. After convergence, we can compute $f_{\text{KL}}(t)$ everywhere in the domain. If $\max_i f_{\text{KL}}(t)$ is greater than $f_{\text{KL}}(t)$ evaluated at any of the peaks t_i , we have to add another delta function [52]. This way, we ensure that (1) we have converged to the global maximum and (2) that we have used the optimal number of delta functions.

The accompanying code [99] includes a self-contained example for computing the channel capacity using the procedure above.

APPENDIX D: SUFFICIENT STATISTICS FOR INDEPENDENT, IDENTICAL, GAMMA-DISTRIBUTED VARIABLES

We find that there is negligible spatial correlation in the gene expression g (Appendix E). Hence, we can treat the cells as responding independently conditional on the Wnt signal t . Below, we show that when one considers a group of N cells, the arithmetic mean $\bar{g}$ is a sufficient statistic for the signal duration t . That is, $\bar{g}$ contains as much information about t as the whole dataset $\mathbf{g}$, as claimed in the main text.

A single cell i in a group of N cells exposed to a Wnt signal of duration t responds independently by expressing output $g_i \sim \text{Gamma}(k, \theta(t))$. Hence, the likelihood function for the

group of N cells is given by

$$p(\mathbf{g}|t) = \prod_{i=1}^N p(g_i|t) \quad (\text{D1})$$

$$= \frac{\prod_{i=1}^N g_i^{k-1}}{\Gamma(k)^N \theta(t)^{Nk}} e^{-\sum_{i=1}^N g_i / \theta(t)}. \quad (\text{D2})$$

In Sec. V, we claim that the arithmetic mean $\bar{g} \equiv \frac{1}{N} \sum_{i=1}^N g_i$ is a sufficient statistic for t . A quick way to see this is by observing that Eq. (D2) satisfies Fisher-Neyman factorization [100]. That is, it can be written in the form $p(\mathbf{g}|t) = h(\mathbf{g}) f(\bar{g}, t)$ for nonnegative functions h and f , from which the defining property $p(\mathbf{g}|\bar{g}, t) = p(\mathbf{g}|\bar{g})$ follows. Below, however, we derive this property explicitly.

The likelihood of $\bar{g}$ conditioned on t can be written as

$$p(\bar{g}|t) = \frac{1}{\Gamma(Nk)(\theta(t)/N)^{Nk}} \bar{g}^{Nk-1} e^{-N\bar{g}/\theta(t)}, \quad (\text{D3})$$

which follows from the addition of N independent and gamma-distributed random variables. Using Bayes' theorem, we obtain

$$p(\mathbf{g}|\bar{g}, t) = \frac{p(\mathbf{g}|t)}{p(\bar{g}|t)}. \quad (\text{D4})$$

Substituting Eqs. (D2) and (D3) into the above equation, we get

$$p(\mathbf{g}|\bar{g}, t) = \frac{\prod_{i=1}^N g_i^{k-1}}{\bar{g}^{Nk-1}} \frac{\Gamma(Nk)}{(N^k \Gamma(k))^N} = p(\mathbf{g}|\bar{g}) \quad (\text{D5})$$

as required. The last equality follows as all t dependence has been canceled out.

Sufficiency of $\bar{g}$ also implies the posterior distributions $p(t|\mathbf{g})$ and $p(t|\bar{g})$ are identical. This can be seen by applying Bayes' theorem $p(t|\mathbf{g}) = p(\mathbf{g}|t)p(t)/p(\mathbf{g})$. Substituting $p(\mathbf{g}|t) = p(\mathbf{g}|\bar{g})p(\bar{g}|t)$ and $p(\mathbf{g}) = p(\mathbf{g}|\bar{g})p(\bar{g})$, we obtain

$$p(t|\mathbf{g}) = \frac{p(\bar{g}|t)p(t)}{p(\bar{g})} = p(t|\bar{g}) \quad (\text{D6})$$

as required.

We can also show that the mutual information satisfies $I(t; \mathbf{g}) = I(t; \bar{g})$. This makes precise the statement that the statistic $\bar{g}$ contains as much information about the signal t as the whole dataset $\mathbf{g}$. To show this, we use the data-processing inequality for the mutual information in two ways. First, as $\bar{g}$ is a function of the dataset $\mathbf{g}$, we must have

$$I(t; \mathbf{g}) \geq I(t; \bar{g}). \quad (\text{D7})$$

Second, we have just shown that $p(\mathbf{g}|\bar{g}, t) = p(\mathbf{g}|\bar{g})$. This implies that we have a Markov chain $t \rightarrow \bar{g} \rightarrow \mathbf{g}$, to which we can also apply the data-processing inequality:

$$I(t; \mathbf{g}) \leq I(t; \bar{g}). \quad (\text{D8})$$

Together, these inequalities imply that $I(t; \mathbf{g}) = I(t; \bar{g})$, as required.

Finally, we can show that the decoding maps $p_g(t|t^*)$ and $p_{\bar{g}}(t|t^*)$ are identical. Starting with $p_g(t|t^*)$, we can introduce

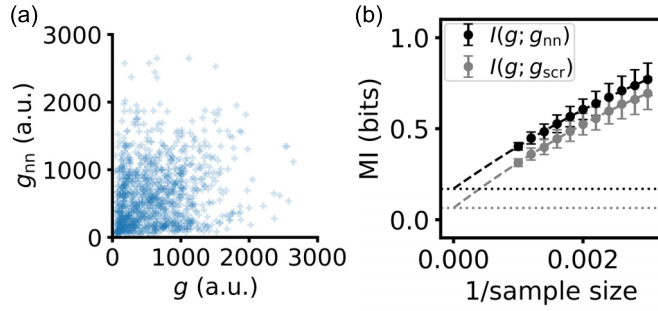


FIG. 7. There are negligible spatial correlations between responses g to the optogenetic Wnt signal. Here, we show data from a repeat experiment for $t = 12.5$ h. (a) Scatterplot showing TopFlash expression g vs the expression of the nearest neighboring cell, g_{nn} . (b) Mutual information $I(g; g_{nn})$ and $I(g; g_{scr})$, where g_{scr} is a scrambled permutation of g , as a function of inverse sample size. Extrapolating to infinite sample size we obtain $I(g; g_{nn}) \approx 0.2$ bits and $I(g; g_{scr}) \approx 0.1$ bits, which are negligible compared to the joint entropies $H(g, g_{nn}) \approx 21.0$ bits and $H(g, g_{scr}) \approx 21.1$ bits.

an integral over $\bar{g}$ using the law of total probability:

$$p_g(t|t^*) = \int_0^\infty dg p(t|g) p(g|t^*) \quad (D9)$$

$$= \int_0^\infty dg \int_0^\infty d\bar{g} p(t|g) p(g|\bar{g}) p(\bar{g}|t^*). \quad (D10)$$

By changing the order of integration and using the fact that $p(t|g) = p(t|\bar{g})$, we obtain

$$p_g(t|t^*) = \int_0^\infty d\bar{g} p(t|\bar{g}) p(\bar{g}|t^*) \int_0^\infty dg p(g|\bar{g}) \quad (D11)$$

$$= p_{\bar{g}}(t|t^*). \quad (D12)$$

As such, the decoding maps are identical whether we consider the gene expression in the whole group g or just the mean expression $\bar{g}$.

APPENDIX E: NEGLIGIBLE SPATIAL CORRELATIONS IN TOPFLASH EXPRESSION

In the main text, we consider the vector response $\mathbf{g}$ of a group of N cells to an optogenetic Wnt signal of duration t . We treated each response as independent. To justify this assumption, we investigate here whether there are spatial correlations; specifically, whether the TopFlash response g of a particular cell is correlated to that of its nearest neighbor g_{nn} . Indeed, we find no significant correlation between g and g_{nn} . The results from a repeat experiment for $t = 12.5$ h are shown in Fig. 7. The scatterplot between g and g_{nn} reveals an uncorrelated cloud [Fig. 7(a)], and the mutual information $I(g; g_{nn})$ is negligible [Fig. 7(b)].

APPENDIX F: SMALL-NOISE REGIME AND EQUIVALENCE TO THE JEFFREYS PRIOR

As before, we seek to find the distribution $p_*(t)$ that maximizes the mutual information $I(\bar{g}; t)$. In this Appendix, we consider regimes where the optimization is analytically tractable. In Fig. 5, we show that we can analytically compute

the channel capacity I_* in regimes where the effective noise is really small ($N \gg 1$). We derive these results here.

As the effective noise approaches zero, it is known that the optimal code for a communication channel becomes continuous [52,74]. In this regime, we can use the fact that the noise is small to derive $p_*(t)$ analytically. In literature, this is often referred to as the small-noise approximation and is widely used for studying information transmission in biological systems [6,62,78–82]. We will also illustrate that the prior obtained in this limit is formally the same as the Jeffreys prior, a noninformative prior that is often used in Bayesian statistics [83].

We explore the gene regulatory response for a particular signal duration, $t \rightarrow \bar{g}$, which occurs with probability

$$p(\bar{g}|t) = \frac{1}{\Gamma(Nk)(\theta(t)/N)^{Nk}} \bar{g}^{Nk-1} e^{-N\bar{g}/\theta(t)}. \quad (F1)$$

In the limit of small noise ($N \gg 1$), we can approximate the gamma distribution as a narrow Gaussian distribution with a mean and variance given by $\mu_{\bar{g}}(t) = k\theta(t)$ and $\sigma_{\bar{g}}^2(t) = k\theta(t)^2/N$, respectively.

We write the mutual information as a difference of entropies:

$$I(\bar{g}; t) = H(\bar{g}) - H(\bar{g}|t), \quad (F2)$$

where

$$H(\bar{g}) = - \int_0^\infty d\bar{g} p(\bar{g}) \log_2 p(\bar{g}) \quad (F3)$$

and

$$H(\bar{g}|t) = - \int_0^T dt p(t) \int_0^\infty d\bar{g} p(\bar{g}|t) \log_2 p(\bar{g}|t). \quad (F4)$$

When the noise $\sigma_{\bar{g}}(t)$ is small, the mapping $p(\bar{g}|t)$ is almost deterministic. As such, we can write

$$p(t) \approx p(\bar{g}) \left| \frac{d\mu_{\bar{g}}}{dt} \right|. \quad (F5)$$

This allows us to write entropy in Eq. (F3) as

$$H(\bar{g}) \approx H(t) + \int_0^T dt p(t) \log_2 \left| \frac{d\mu_{\bar{g}}}{dt} \right|. \quad (F6)$$

Further, using the Gaussian approximation for $p(\bar{g}|t)$, the conditional entropy in Eq. (F4) becomes

$$H(\bar{g}|t) \approx \frac{1}{2} \int_0^T dt p(t) \log_2 (2\pi e \sigma_{\bar{g}}^2(t)). \quad (F7)$$

Having written both $H(\bar{g})$ and $H(\bar{g}|t)$ in a way where $p(t)$ is the only “free” distribution that we can vary, we can now proceed with the optimization. We add a Lagrangian multiplier to ensure normalization of $p(t)$, and optimize

$$\mathcal{L}[p(t)] = I(\bar{g}; t) - \beta \int_0^T dt p(t). \quad (F8)$$

Taking the variational derivative with respect to $p(t)$, we get

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta p(t)} = & \log_2 \left| \frac{d\mu_{\bar{g}}}{dt} \right| - \log_2 p(t) - \frac{1}{\ln 2} \\ & - \frac{1}{2} \log_2 (2\pi e \sigma_{\bar{g}}^2(t)) - \beta. \end{aligned} \quad (F9)$$

Setting $\delta\mathcal{L}/\delta p(t) = 0$, we obtain

$$p_*(t) = \frac{1}{Z} \frac{1}{\sigma_{\bar{g}}(t)} \left| \frac{d\mu_{\bar{g}}}{dt} \right|, \quad (\text{F10})$$

where

$$Z = \int_0^T \frac{dt}{\sigma_{\bar{g}}(t)} \left| \frac{d\mu_{\bar{g}}}{dt} \right| \quad (\text{F11})$$

is a normalizing constant. Using the optimal input distribution in Eq. (F10) to evaluate $I(\bar{g}; t)$, we obtain the channel capacity given by [13]

$$I_* = \log_2 \left[\frac{Z}{\sqrt{2\pi e}} \right]. \quad (\text{F12})$$

Since $\theta(t) \propto t$ for times $t \gtrsim 1h$, both $\mu_{\bar{g}}(t)$ and $\sigma_{\bar{g}}(t)$ grow linearly with t in this regime. It follows from Eq. (F10) that the tail of the optimal prior decays functionally as $\sim 1/t$ [cf. Fig. 5(b)].

In Fig. 5, we show the mutual information obtained using the prior in Eq. (F10) derived from the small-noise approximation. Indeed, asymptotically with increasing N , we converge to the channel capacity I_* . To derive the functional behavior of the channel capacity I_* with large N we can proceed by inspection. By construction, the mean $\mu_{\bar{g}}(t)$ does not depend on N . The variance obeys $\sigma_{\bar{g}}^2(t) \propto 1/N$, and therefore the normalizing constant Z in Eq. (F11) satisfies $Z \propto N^{1/2}$. As such, the channel capacity in Eq. (F12) asymptotically scales as

$$I_* \sim \frac{1}{2} \log_2 N + o(1). \quad (\text{F13})$$

We also note that the prior obtained in the small-noise limit is formally identical to the Jeffreys prior [83]. It is known that in the limit of an infinite number of identically, independent trials of the same experiment (i.e., $N \rightarrow \infty$), the prior that is Shannon-optimal converges weakly to the Jeffreys prior [84]. Indeed, the same is true for a single precise experiment in the limit of small noise. We derive the Jeffreys prior below and verify that it is indeed identical to Eq. (F10).

The Jeffreys prior is designed to be invariant under reparametrization of the probability distribution, and is defined as

$$p_J(t) \propto |\mathcal{I}(t)|^{1/2}, \quad (\text{F14})$$

where $\mathcal{I}(t)$ is the Fisher information,

$$\mathcal{I}(t) = \int_0^\infty d\bar{g} p(\bar{g}|t) \left(\frac{\partial \ln p(\bar{g}|t)}{\partial t} \right)^2. \quad (\text{F15})$$

In our case, $p(\bar{g}|t)$ is gamma distributed with shape parameter Nk and a time-dependent scale parameter $\theta(t)/N$:

$$\begin{aligned} \ln p(\bar{g}|t) &= -\ln \Gamma(Nk) - Nk \ln(\theta(t)/N) \\ &\quad + (Nk - 1) \ln \bar{g} - \frac{N\bar{g}}{\theta(t)}. \end{aligned} \quad (\text{F16})$$

Taking the first derivative and substituting $\mu_{\bar{g}}(t) = k\theta(t)$ and $\sigma_{\bar{g}}^2(t) = k\theta(t)^2/N$, we obtain

$$\left(\frac{\partial \ln p(\bar{g}|t)}{\partial t} \right)^2 = \frac{1}{\sigma_{\bar{g}}(t)^4} \left(\frac{d\mu_{\bar{g}}(t)}{dt} \right)^2 (\bar{g} - \mu_{\bar{g}}(t))^2. \quad (\text{F17})$$

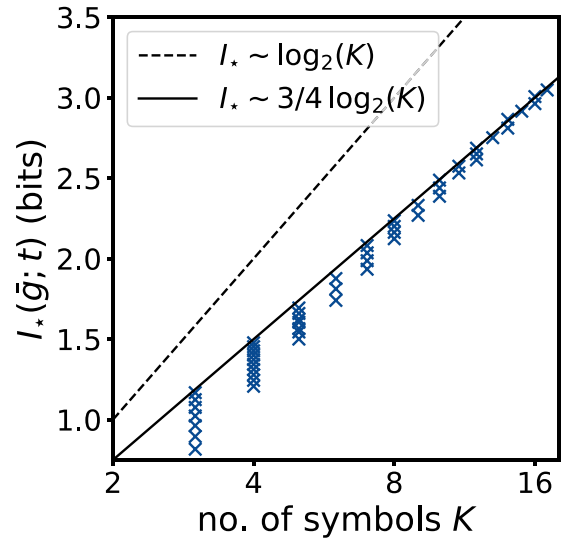


FIG. 8. Asymptotic scaling law between the number of symbols K in the optimized prior $p_*(t)$ and the mutual information $I(g; t)$, in agreement with literature [52,74]. The bound $I(\bar{g}; t) \leq \log_2 K$ is shown using a dotted line.

The expectation over $(\bar{g} - \mu_{\bar{g}}(t))^2$ is just the variance $\sigma_{\bar{g}}^2(t)$; hence, the Fisher information in Eq. (F15) becomes

$$\mathcal{I}(t) = \frac{1}{\sigma_{\bar{g}}(t)^2} \left(\frac{d\mu_{\bar{g}}(t)}{dt} \right)^2. \quad (\text{F18})$$

The Jeffreys prior can now be written as

$$p_J(t) \propto \frac{1}{\sigma_{\bar{g}}(t)} \left| \frac{d\mu_{\bar{g}}}{dt} \right|, \quad (\text{F19})$$

which is identical to the Shannon-optimal prior in the limit of small noise in Eq. (F10). We have thus verified that the Jeffreys prior and the prior that optimizes the mutual information are equivalent in the limit of small noise.

APPENDIX G: ASYMPTOTIC SCALING LAW FOR THE CHANNEL CAPACITY

In the main text, we explored what happens to our optimal prior $p_*^{(N)}(t)$ as we considered multiple cells N . As N becomes larger, the effective noise level decreases—analogueous to sending the same message multiple times in a communication channel. As the noise level approaches zero, it is known that the number of symbols K in the optimal code follows an asymptotic scaling law. In this limit, the channel capacity I_* scales with the logarithm of the number of symbols K as $I_* \sim (3/4) \log_2 K$ [74]. Indeed, we confirm the scaling law for our system in Fig. 8 as a nontrivial check of our optimization and to verify consistency with existing literature. The dotted line shows the fundamental limit $I_* \leq \log_2 K$, which would reach equality if the K symbols were perfectly distinguishable.

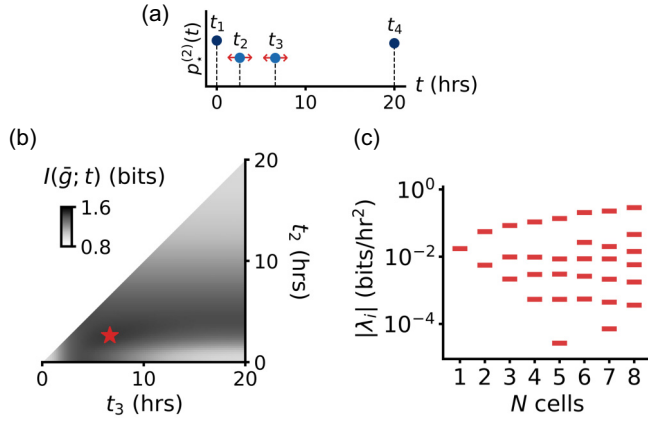


FIG. 9. Fine-tuning of symbols in the optimal code; here, weights $\{w_i\}$ are kept optimized as positions $\{t_i\}$ are varied. (a) For $N = 2$, the optimal prior consists of $K = 4$ symbols [cf. Fig. 5(b)]. We vary the positions of two symbols t_2 and t_3 , while keeping the weights w_i optimized. (b) Information $I(\bar{g}; t)$ (colorshade) shows a broad optimum (red star) as a function of the position of peaks t_2 and t_3 . (c) The Hessian matrix [cf. Eq. (H8)] has a sloppy spectrum that widens as N increases.

APPENDIX H: FINE-TUNING OF THE OPTIMAL CODE

Here, we consider a prior distribution $p(t)$ composed of a set of K discrete symbols $\{t_i\}$ with respective weights $\{w_i\}$:

$$p(t) = \sum_{i=1}^K w_i \delta(t - t_i), \quad (\text{H1})$$

and ask to what extent the capacity-achieving distribution $p_*(t)$ needs to be fine-tuned. Since the BA algorithm converges much more quickly to the channel capacity I_* than to the optimal prior $p_*(t)$, we anticipate a sloppy information-landscape around the optimum.

To investigate fine-tuning of the prior, we write the mutual information $I(\bar{g}; t)$ as a function of the positions $\{t_i\}$ while keeping the weights $\{w_i\}$ fixed:

$$I(\bar{g}; t | \{t_i\}) = \sum_{i=1}^K w_i \int_0^\infty d\bar{g} p(\bar{g}|t_i) \log_2 \left(\frac{p(\bar{g}|t_i)}{p(\bar{g})} \right). \quad (\text{H2})$$

Denoting the optimized positions as $\{t_i^*\}$, we can expand the mutual information around its maximum, the capacity

$$I(\bar{g}; t | \{t_i\}) = I_* + \frac{1}{2} \sum_{i,j=1}^K (t_i - t_i^*) \chi_{ij} (t_j - t_j^*) + \dots, \quad (\text{H3})$$

where χ_{ij} is the Hessian matrix:

$$\chi_{ij} = \left. \frac{\partial^2 I}{\partial t_i \partial t_j} \right|_{\{t_i^*\}}. \quad (\text{H4})$$

The eigenvectors of χ determine directions in parameter space $\{t_i\}$ that have independent effects on the mutual information, whereas the eigenvalues $\{\lambda_i\}$ tell us the sensitivity along these directions [62]. Since we have fit the data with a functional form for $p(\bar{g}|t_i)$, we can proceed analytically to compute the Hessian.

To proceed, we express the partial derivatives of the marginal distribution $p(\bar{g})$ with respect to symbol t_i ,

$$\frac{\partial p(\bar{g})}{\partial t_i} = w_i \frac{\partial p(\bar{g}|t_i)}{\partial t_i}. \quad (\text{H5})$$

We can then compute the first derivative of the mutual information with respect to t_i :

$$\frac{\partial I}{\partial t_i} = \frac{w_i}{\ln 2} \int_0^\infty d\bar{g} \frac{\partial p(\bar{g}|t_i)}{\partial t_i} \ln \left(\frac{p(\bar{g}|t_i)}{p(\bar{g})} \right). \quad (\text{H6})$$

Note that this is the same gradient as the one used to iteratively adjust the positions in Eq. (C9). Taking a second derivative with respect to t_j , we obtain

$$\begin{aligned} \chi_{ij} &= \frac{\partial^2 I}{\partial t_i \partial t_j} \\ &= \frac{w_i}{\ln 2} \int_0^\infty d\bar{g} \left\{ \delta_{ij} \left[\frac{\partial^2 p(\bar{g}|t_i)}{\partial t_i^2} \ln \left(\frac{p(\bar{g}|t_i)}{p(\bar{g})} \right) \right. \right. \\ &\quad \left. \left. + \frac{1}{p(\bar{g}|t_i)} \left(\frac{\partial p(\bar{g}|t_i)}{\partial t_i} \right)^2 \right] - \frac{w_j}{p(\bar{g})} \frac{\partial p(\bar{g}|t_i)}{\partial t_i} \frac{\partial p(\bar{g}|t_j)}{\partial t_j} \right\}. \end{aligned} \quad (\text{H7})$$

To compute the integral, we insert the functional form $p(\bar{g}|t)$ and integrate numerically, after which we can find the spectrum $\{\lambda_i\}$ by diagonalizing χ [Fig. 5(d)]. We note that in deriving Eq. (H7), we have not assumed a functional form for $p(\bar{g}|t)$, and thus the equation holds for any choice of transmission probability and discrete prior.

In the above and the main text, we kept the weights $\{w_i\}$ fixed while adjusting the positions $\{t_i\}$ of the symbols. Alternatively, one can keep the weights optimized while varying the position of the symbols. In that case, the Hessian becomes a total derivative $\chi_{ij} = d^2 I / dt_i dt_j$ and receives a contribution from adjusting the weights as one departs from the optimum. Denoting the partial derivatives of the mutual information by $(I_{tt})_{ij} = \partial^2 I / \partial t_i \partial t_j$, $(I_{wt})_{ij} = \partial^2 I / \partial w_i \partial t_j$, and $(I_{ww})_{ij} = \partial^2 I / \partial w_i \partial w_j$, we can write the Hessian as

$$\chi = I_{tt} - I_{wt}^T M I_{wt}, \quad (\text{H8})$$

where M is a matrix given by

$$M = I_{ww}^{-1} - \frac{I_{ww}^{-1} \mathbf{1} \mathbf{1}^T I_{ww}^{-1}}{\mathbf{1}^T I_{ww}^{-1} \mathbf{1}}, \quad (\text{H9})$$

and $\mathbf{1} = (1, \dots, 1)$. Again, these expressions are general and hold for any choice of transmission probability and discrete prior. In Fig. 9(a), we plot the mutual information for $N = 2$ as a function of the positions of two out of four delta functions while keeping the weights optimized. We observe a broad optimum similar to Fig. 6(a). In Fig. 9(b), we use Eq. (H8) to plot the spectrum as a function of N ; spanning four decades, it is more sloppy than the case for fixed weights in the main text.

- [1] F. Jacob and J. Monod, Genetic regulatory mechanisms in the synthesis of proteins, *J. Mol. Biol.* **3**, 318 (1961).
- [2] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, *Science (New York, N.Y.)* **297**, 1183 (2002).
- [3] J. Elf, J. Paulsson, O. G. Berg, and M. Ehrenberg, Near-critical phenomena in intracellular metabolite pools, *Biophys. J.* **84**, 154 (2003).
- [4] J. Paulsson, Summing up the noise in gene networks, *Nature (London)* **427**, 415 (2004).
- [5] H. Salman, N. Brenner, C.-K. Tung, N. Elyahu, E. Stolovicki, L. Moore, A. Libchaber, and E. Braun, Universal protein fluctuations in populations of microorganisms, *Phys. Rev. Lett.* **108**, 238105 (2012).
- [6] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, Positional information, in bits, *Proc. Natl. Acad. Sci. USA* **110**, 16301 (2013).
- [7] M. Merle, L. Friedman, C. Chureau, A. Shoushtarizadeh, and T. Gregor, Precise and scalable self-organization in mammalian pseudo-embryos, *Nat. Struct. Mol. Biol.* **31**, 896 (2024).
- [8] A. Warmflash, B. Sorre, F. Etoc, E. D. Siggia, and A. H. Brivanlou, A method to recapitulate early embryonic spatial patterning in human embryonic stem cells, *Nat. Methods* **11**, 847 (2014).
- [9] A. Stanoev, C. Schröter, and A. Koseska, Robustness and timing of cellular differentiation through population-based symmetry breaking, *Development* **148**, dev197608 (2021).
- [10] R. Suderman, J. A. Bachman, A. Smith, P. K. Sorger, and E. J. Deeds, Fundamental trade-offs between information flow in single cells and cellular populations, *Proc. Natl. Acad. Sci. USA* **114**, 5755 (2017).
- [11] Y. el Azhar, P. Schulthess, M. J. van Oostrom, S. D. C. Weterings, W. H. M. Meijer, N. Tsuchida-Straeten, W. M. Thomas, M. Bauer, and K. F. Sonnen, Unravelling differential Hes1 dynamics during axis elongation of mouse embryos through single-cell tracking, *Development* **151**, dev202936 (2024).
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
- [13] W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton, 2012).
- [14] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, Information transduction capacity of noisy biochemical signaling networks, *Science (New York, N.Y.)* **334**, 354 (2011).
- [15] S. Uda, T. H. Saito, T. Kudo, T. Kokaji, T. Tsuchiya, H. Kubota, Y. Komori, Y.-i. Ozaki, and S. Kuroda, Robustness and compensation of information transmission of signaling pathways, *Science* **341**, 558 (2013).
- [16] K. S. Iyer, C. Prabhakara, S. Mayor, and M. Rao, Cellular compartmentalisation and receptor promiscuity as a strategy for accurate and robust inference of position during morphogenesis, *eLife* **12**, e79257 (2023).
- [17] R. Bettoni, G. Dupont, A. M. Walczak, and S. de Buyl, Optimizing information transmission in neural induction constrains cell surface contacts of ascidian embryos, *PRX Life* **3**, 033004 (2025).
- [18] J. Selimkhanov, B. Taylor, J. Yao, A. Pilko, J. Albeck, A. Hoffmann, L. Tsimring, and R. Wollman, Accurate information transmission through dynamic biochemical signaling networks, *Science* **346**, 1370 (2014).
- [19] F. Tostevin and P. R. ten Wolde, Mutual information between input and output trajectories of biochemical networks, *Phys. Rev. Lett.* **102**, 218101 (2009).
- [20] M. Reinhardt, G. Tkačik, and P. R. ten Wolde, Path weight sampling: Exact Monte Carlo computation of the mutual information between stochastic trajectories, *Phys. Rev. X* **13**, 041017 (2023).
- [21] A.-L. Moor and C. Zechner, Dynamic information transfer in stochastic biochemical networks, *Phys. Rev. Res.* **5**, 013032 (2023).
- [22] K. M. Cadigan and R. Nusse, Wnt signaling: a common theme in animal development, *Genes Dev.* **11**, 3286 (1997).
- [23] C. P. Arnold, B. W. Benham-Pyle, J. J. Lange, C. J. Wood, and A. Sánchez Alvarado, Wnt and TGF β coordinate growth and patterning to regulate size-dependent behaviour, *Nature (London)* **572**, 655 (2019).
- [24] R. S. Lach, C. Qiu, E. Z. Kajbaf, N. Baxter, D. Han, A. Wang, H. Lock, O. Chirikian, B. Pruitt, and M. Z. Wilson, Nucleation of the destruction complex on the centrosome accelerates degradation of β -catenin and regulates Wnt signal transmission, *Proc. Natl. Acad. Sci. USA* **119**, e2204688119 (2022).
- [25] S. J. Rosen, O. Witteveen, N. Baxter, R. S. Lach, E. Hopkins, M. Bauer, and M. Z. Wilson, Anti-resonance in developmental signaling regulates cell fate decisions, *eLife* **14**, RP107794 (2026).
- [26] N. Friedman, L. Cai, and X. S. Xie, Linking stochastic dynamics to population distribution: An analytical framework of gene expression, *Phys. Rev. Lett.* **97**, 168302 (2006).
- [27] L. Cai, N. Friedman, and X. S. Xie, Stochastic protein expression in individual cells at the single molecule level, *Nature (London)* **440**, 358 (2006).
- [28] A. J. Mikels and R. Nusse, Wnts as ligands: Processing, secretion and reception, *Oncogene* **25**, 7461 (2006).
- [29] T. Erdmann, M. Howard, and P. R. ten Wolde, Role of spatial averaging in the precision of gene expression patterns, *Phys. Rev. Lett.* **103**, 258101 (2009).
- [30] A. N. Gasparski and K. A. Beningo, Mechanoreception at the cell membrane: More than the integrins, *Arch. Biochem. Biophys.* **586**, 20 (2015).
- [31] N. S. Funa, K. A. Schachter, M. Lerdrup, J. Ekberg, K. Hess, N. Dietrich, C. Honoré, K. Hansen, and H. Semb, β -Catenin regulates primitive streak induction through collaborative interactions with SMAD2/SMAD3 and OCT4, *Cell Stem Cell* **16**, 639 (2015).
- [32] K. Kim, J. Cho, T. S. Hilzinger, H. Nunns, A. Liu, B. E. Ryba, and L. Goentoro, Two-element transcriptional regulation in the canonical Wnt pathway, *Curr. Biol.* **27**, 2357 (2017).
- [33] A. Kicheva and J. Briscoe, Control of tissue development by morphogens, *Annu. Rev. Cell Dev. Biol.* **39**, 91 (2023).
- [34] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, Universally sloppy parameter sensitivities in systems biology models, *PLoS Comput. Biol.* **3**, e189 (2007).
- [35] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, Perspective: Sloppiness and emergent theories in physics, biology, and beyond, *J. Chem. Phys.* **143**, 010901 (2015).

- [36] H. B. Barlow, Possible principles underlying the transformations of sensory messages, in *Sensory Communication* (The MIT Press, Cambridge, 1961).
- [37] S. Laughlin, A simple coding procedure enhances a neuron's information capacity, *Z. Naturforsch. Sect. C, Biosci.* **36**, 910 (1981).
- [38] W. Bialek, F. Rieke, R. R. de Ruyter van Steveninck, and D. Warland, Reading a neural code, *Science* **252**, 1854 (1991).
- [39] N. Brenner, W. Bialek, and R. de Ruyter van Steveninck, Adaptive rescaling maximizes information transmission, *Neuron* **26**, 695 (2000).
- [40] R. Anton, H. A. Kestler, and M. Kühl, β -Catenin signaling contributes to stemness and regulates early differentiation in murine embryonic stem cells, *FEBS Lett.* **581**, 5247 (2007).
- [41] J. R. Spence, C. N. Mayhew, S. A. Rankin, M. F. Kuhar, J. E. Vallance, K. Tolle, E. E. Hoskins, V. V. Kalinichenko, S. I. Wells, A. M. Zorn, N. F. Shroyer, and J. M. Wells, Directed differentiation of human pluripotent stem cells into intestinal tissue *in vitro*, *Nature (London)* **470**, 105 (2011).
- [42] T. Sato and H. Clevers, Growing self-organizing mini-guts from a single intestinal stem cell: Mechanism and applications, *Science* **340**, 1190 (2013).
- [43] S. C. van den Brink, A. Alemany, V. van Batenburg, N. Moris, M. Blotenburg, J. Vivić, P. Baillie-Johnson, J. Nichols, K. F. Sonnen, A. M. Arias, and A. van Oudenaarden, Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids, *Nature (London)* **582**, 405 (2020).
- [44] L. Goentoro and M. W. Kirschner, Evidence that fold-change, and not absolute level, of β -catenin dictates Wnt signaling, *Mol. Cell* **36**, 872 (2009).
- [45] L. Goentoro, O. Shoval, M. W. Kirschner, and U. Alon, The incoherent feedforward loop can provide fold-change detection in gene regulation, *Mol. Cell* **36**, 894 (2009).
- [46] A. Aulehla, W. Wiegand, V. Baubet, M. B. Wahl, C. Deng, M. Taketo, M. Lewandoski, and O. Pourquié, A β -catenin gradient links the clock and wavefront systems in mouse embryo segmentation, *Nat. Cell Biol.* **10**, 186 (2008).
- [47] E. Cimet, C. Cannizzaro, R. James, T. Biechele, R. T. Moon, N. Elvassore, and G. Vunjak-Novakovic, Microfluidic device generating stable concentration gradients for long term cell culture: application to Wnt3a regulation of β -catenin signaling, *Lab. Chip* **10**, 3277 (2010).
- [48] K. F. Sonnen, V. M. Lauschke, J. Uraji, H. J. Falk, Y. Petersen, M. C. Funk, M. Beaupex, P. François, C. A. Merten, and A. Aulehla, Modulation of phase shift between Wnt and Notch signaling oscillations controls mesoderm segmentation, *Cell* **172**, 1079 (2018).
- [49] E. J. Cooper and S. Scholpp, Transport and gradient formation of Wnt and FGF in the early zebrafish gastrula, *Curr. Top. Dev. Biol.* **157**, 125 (2024).
- [50] R. Kok, X. Zheng, W. Spoelstra, T. Clement, Y. van de Grift, H. Clevers, R. van Amerongen, S. Tans, and J. van Zon, Loss of Paneth cell contact starts a Wnt differentiation timer in intestinal crypts, *Res. Square* (2024).
- [51] M. Yadav, D. Koch, and A. Koseska, Homeorhetic regulation of cellular phenotype (2025), *bioRxiv*.
- [52] H. H. Mattingly, M. K. Transtrum, M. C. Abbott, and B. B. Machta, Maximizing the information learned from finite data selects a simple model, *Proc. Natl. Acad. Sci. USA* **115**, 1760 (2018).
- [53] A. Ferrer-Vaquer, A. Piliszek, G. Tian, R. J. Aho, D. Dufort, and A.-K. Hadjantonakis, A sensitive and bright single-cell resolution live imaging reporter of Wnt/ β -catenin signaling in the mouse, *BMC Dev. Biol.* **10**, 121 (2010).
- [54] S. Ogami, M. Yamamichi, K. Sato, and T. Ishitani, Dynamics of Wnt/ β -catenin reporter activity throughout whole life in a naturally short-lived vertebrate, *npj Aging* **10**, 23 (2024).
- [55] C. Fuerer and R. Nusse, Lentiviral vectors to probe and manipulate the Wnt signaling pathway, *PLoS One* **5**, e9370 (2010).
- [56] J. Liu, Q. Xiao, J. Xiao, C. Niu, Y. Li, X. Zhang, Z. Zhou, G. Shu, and G. Yin, Wnt/ β -catenin signalling: function, biological mechanisms, and therapeutic opportunities, *Sig. Transduct. Target. Ther.* **7**, 3 (2022).
- [57] T. C. Höhener, A. E. Landolt, C. Dessauges, L. Hinderling, P. A. Gagliardi, and O. Pertz, LITOS: A versatile LED illumination tool for optogenetic stimulation, *Sci. Rep.* **12**, 13139 (2022).
- [58] A. L. Koch, The logarithm in biology I. Mechanisms generating the log-normal distribution exactly, *J. Theor. Biol.* **12**, 276 (1966).
- [59] C. E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* **27**, 379 (1948).
- [60] L. Zdeborová and F. Krzakala, Statistical physics of inference: Thresholds and algorithms, *Adv. Phys.* **65**, 453 (2016).
- [61] M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, and T. Gregor, Optimal decoding of cellular identities in a genetic network, *Cell* **176**, 844 (2019).
- [62] M. Bauer, M. D. Petkova, T. Gregor, E. F. Wieschaus, and W. Bialek, Trading bits in the readout from a genetic network, *Proc. Natl. Acad. Sci. USA* **118**, e2109011118 (2021).
- [63] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, Applied Mathematics Series (Dover Publications, New York, 1965).
- [64] J. G. Smith, The information capacity of amplitude- and variance-constrained scalar Gaussian channels, *Inf. Control* **18**, 203 (1971).
- [65] A. Levchenko and I. Nemenman, Cellular noise and information transmission, *Curr. Opin. Biotechnol.* **28**, 156 (2014).
- [66] G. Tkačik, C. G. Callan, and W. Bialek, Information capacity of genetic regulatory elements, *Phys. Rev. E* **78**, 011910 (2008).
- [67] T. Mijatović, A. R. Kok, J. W. Zwanikken, and M. Bauer, Weak transcription factor clustering at binding sites can facilitate information transfer from molecular signals, *PRX Life* **4**, 013003 (2026).
- [68] A. Bairoch, The Cellosaurus, a cell-line knowledge resource, *J. Biomol. Tech.* **29**, 25 (2018).
- [69] J. Yang, P. Guertin, G. Jia, Z. Lv, H. Yang, and D. Ju, Large-scale microcarrier culture of HEK293T cells and Vero cells in single-use bioreactors, *AMB Express* **9**, 70 (2019).
- [70] J. Moosemiller, HEK Cell Splitting and Maintenance, <https://receptor.nsm.uh.edu/research/protocols/experimental/hekcells-split>, accessed 2 June 2025.
- [71] R. Blahut, Computation of channel capacity and rate-distortion functions, *IEEE Trans. Inf. Theory* **18**, 460 (1972).
- [72] S. Arimoto, An algorithm for computing the capacity of arbitrary discrete memoryless channels, *IEEE Trans. Inf. Theory* **18**, 14 (1972).

- [73] J. Dauwels, Numerical computation of the capacity of continuous memoryless channels, in *Proceedings of the 26th Symposium on Information Theory in the Benelux* (WIC, Brussels, Belgium, 2005).
- [74] M. C. Abbott and B. B. Machta, A scaling law from discrete to continuous solutions of channel capacity problems in the low-noise limit, *J. Stat. Phys.* **176**, 214 (2019).
- [75] S. Chakrabarti and F. Michor, Circadian clock effects on cellular proliferation: Insights from theory and experiments, *Curr. Opin. Cell Biol.* **67**, 17 (2020).
- [76] R. Sandberg and I. Ernberg, The molecular portrait of *in vitro* growth by meta-analysis of gene-expression profiles, *Genome Biol.* **6**, R65 (2005).
- [77] L. Liu and A. Warmflash, Self-organized signaling in stem cell models of embryos, *Stem Cell Rep.* **16**, 1065 (2021).
- [78] G. Tkačik, C. G. Callan, Jr., and W. Bialek, Information flow and optimization in transcriptional regulation, *Proc. Natl. Acad. Sci. USA* **105**, 12265 (2008).
- [79] G. Tkačik and A. M. Walczak, Information transmission in genetic regulatory networks: A review, *J. Phys.: Condens. Matter* **23**, 153102 (2011).
- [80] A. M. Walczak, G. Tkačik, and W. Bialek, Optimizing information flow in small genetic networks. II. Feed-forward interactions, *Phys. Rev. E* **81**, 041905 (2010).
- [81] M. Bauer, How does an organism extract relevant information from transcription factor concentrations? *Biochem. Soc. Trans.* **50**, 1365 (2022).
- [82] M. Bauer and W. Bialek, Information bottleneck in molecular sensing, *PRX Life* **1**, 023005 (2023).
- [83] H. Jeffreys, An invariant form for the prior probability in estimation problems, *Proc. R. Soc. London A* **186**, 453 (1946).
- [84] H. R. Scholl, Shannon optimal priors on independent identically distributed statistical experiments converge weakly to Jeffreys' prior, *Test* **7**, 75 (1998).
- [85] M. Bauer, W. Bialek, C. Goddard, C. M. Holmes, K. Krishnamurthy, S. E. Palmer, R. Pang, D. J. Schwab, and L. Susman, Optimization and variability can coexist, [arXiv:2505.23398](https://arxiv.org/abs/2505.23398).
- [86] A. Amel, A. Rabeling, S. Rossouw, and M. Goolam, Wnt and BMP signalling direct anterior–posterior differentiation in aggregates of mouse embryonic stem cells, *Biol. Open* **12**, bio059981 (2023).
- [87] S. C. van den Brink, P. Baillie-Johnson, T. Balayo, A.-K. Hadjantonakis, S. Nowotschin, D. A. Turner, and A. Martinez Arias, Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells, *Development* **141**, 4231 (2014).
- [88] N. A. Repina, H. J. Johnson, X. Bao, J. A. Zimmermann, D. A. Joy, S. Z. Bi, R. S. Kane, and D. V. Schaffer, Optogenetic control of Wnt signaling models cell-intrinsic embryonic patterning using 2D human pluripotent stem cell culture, *Development* **150**, dev201386 (2023).
- [89] A. B. Rosenbloom, M. Tarczyński, N. Lam, R. S. Kane, L. J. Bugaj, and D. V. Schaffer, β -catenin signaling dynamics regulate cell fate in differentiating neural stem cells, *Proc. Natl. Acad. Sci. USA* **117**, 28828 (2020).
- [90] H. M. McNamara, S. C. Solley, B. Adamson, M. M. Chan, and J. E. Toettcher, Recording morphogen signals reveals mechanisms underlying gastruloid symmetry breaking, *Nat. Cell Biol.* **26**, 1832 (2024).
- [91] B. B. Averbeck, P. E. Latham, and A. Pouget, Neural correlations, population coding and computation, *Nat. Rev. Neurosci.* **7**, 358 (2006).
- [92] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli, Spatio-temporal correlations and visual signalling in a complete neuronal population, *Nature (London)* **454**, 995 (2008).
- [93] E. Schneidman, W. Bialek, and M. J. Berry, II, Synergy, redundancy, and independence in population codes, *J. Neurosci.* **23**, 11539 (2003).
- [94] M. Kramar, L. Hahn, A. M. Walczak, T. Mora, and M. Coppey, Single cells can resolve graded stimuli, *PRX Life* **3**, 043016 (2025).
- [95] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, Recurrence plots for the analysis of complex systems, *Phys. Rep.* **438**, 237 (2007).
- [96] E. Camacho-Aguilar, S. T. Yoon, M. A. Ortiz-Salazar, S. Du, M. C. Guerra, and A. Warmflash, Combinatorial interpretation of BMP and Wnt controls the decision between primitive streak and extraembryonic fates, *Cell Syst.* **15**, 445 (2024).
- [97] S. Lehr, D. B. Brückner, T. G. Minchington, M. Greunz-Schindler, J. Merrin, E. Hannezo, and A. Kicheva, Self-organized pattern formation in the developing mouse neural tube by a temporal relay of BMP signaling, *Dev. Cell* **60**, 567 (2025).
- [98] S. J. Rosen, Experimental data for Witteveen *et al.*, Zenodo (2026), doi:[10.5281/zenodo.18601164](https://doi.org/10.5281/zenodo.18601164).
- [99] O. Witteveen, GitHub (2026), <https://github.com/olivierwitteveen/optimizing-info-in-opto-wnt>.
- [100] P. R. Halmos and L. J. Savage, Application of the Radon-Nikodym theorem to the theory of sufficient statistics, *Ann. Math. Stat.* **20**, 225 (1949).