Trading bits in the readout from a genetic network

Marianne Bauer[a,b,c,d,1], Mariela D. Petkova[e], Thomas Gregor[a,b,f], Eric F. Wieschaus[b,c,d], and William Bialek[a,b,g,1]

[a]Joseph Henry Laboratories of Physics, Princeton University, Princeton, NJ 08544; [b]Lewis–Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544; [c]Department of Molecular Biology, Princeton University, Princeton, NJ 08544; [d]Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544; [e]Program in Biophysics, Harvard University, Cambridge, MA 02138; [f]Department of Developmental and Stem Cell Biology, UMR3738, Institut Pasteur, 75015 Paris, France; and [g]Initiative for the Theoretical Sciences, The Graduate Center, City University of New York, New York, NY 10016

Edited by Boris I. Shraiman, University of California, Santa Barbara, CA, and approved October 6, 2021 (received for review May 14, 2021)

In the regulation of gene expression, information of relevance to the organism is represented by the concentrations of transcription factor molecules. To extract this information the cell must effectively "measure" these concentrations, but there are physical limits to the precision of these measurements. We use the gap gene network in the early fly embryo as an example of the tradeoff between the precision of concentration measurements and the transmission of relevant information. For thresholded measurements we find that lower thresholds are more important, and fine tuning is not required for near-optimal information transmission. We then consider general sensors, constrained only by a limit on their information capacity, and find that thresholded sensors can approach true information theoretic optima. The information theoretic approach allows us to identify the optimal sensor for the entire gap gene network and to argue that the physical limitations of sensing necessitate the observed multiplicity of enhancer elements, with sensitivities to combinations rather than single transcription factors.

sensing | gene regulation | development | information bottleneck

Cells control the concentrations of proteins in part by controlling the transcription of corresponding genes into messenger RNA. This control is effected by the binding of transcription factor (TF) proteins to specific sites along the genome. Transcription factors can thus regulate the synthesis of other TFs, forming a genetic network. Regulatory mechanisms internal to the network must be precise enough to generate reliable relationships between the concentration of input signals and the levels of gene expression downstream. What must the cell do to extract and make efficient use of the information provided by variations in TF concentrations?

We usually think of transcription factors as controlling the level of gene expression, but we can also view the expression level as being the cell's measurement of the TF concentration (1, 2). As outside observers of the cell, we can measure the concentration of transcription factors with considerable accuracy (3). However, the cell's "measurement" of TF concentration is based on the arrival of these molecules at their binding sites, and this is a noisy process, because TF concentrations are low, in the nanomolar range (4–8). Physical limits to the measurement of such low concentrations were first explored in the context of bacterial chemotaxis (9), but have proved to be much more general (1, 10–12). What will be important for our discussion is not the precise values of these limits, but rather that the limits exist and are significant on the scale of biological function.

We focus on the example of the gap genes (more precisely, the transcription factor proteins expressed from them) that are crucial in the early events of embryonic development in fruit flies (13, 14). These four proteins form a network with inputs from primary maternal morphogen molecules and outputs in the striped patterns of pair-rule gene expression. These stripes are positioned with an accuracy of $\pm 1\%$ along the long (anterior–posterior) axis of the embryo, and this is the accuracy of subsequent developmental events such as the formation of the cephalic furrow (15, 16). The local concentrations of the gap proteins provide just enough information to support this level of precision (15). The algorithm that achieves optimal readout of this positional information predicts, quantitatively, the distortions of the pair-rule stripes in mutant flies where individual maternal inputs are deleted (17).

The gap gene network offers us the chance to ask how accurately the transcription factor concentrations need to be measured and to infer features of the regulatory architecture responsible for these measurements. The information that the gap genes convey about position along the anterior–posterior axis is what allows nuclei to make distinct cell fate decisions required for development; this ability to make distinct cell fate decisions needs to be possible even after the transcriptional apparatus has made noisy measurements of the TF concentrations (Fig. 1*A*): We investigate here how this can be seen as a sensing or signal processing problem. There is a tradeoff, such that less accurate sensing limits the complexity and reproducibility of the final decisions. We start with a more traditional view of how information is represented in the concentration of a single TF, through thresholds or expression domains, and then argue for a more abstract formulation of the problem as selective data compression—trading bits of accuracy in sensing for bits of information about position. In this abstract view, aspects of the transcriptional regulatory mechanisms can be seen as solutions to an information theoretic optimization problem. We apply this approach to analyze the information conveyed by the concentrations of all four gap proteins and find that some of the complexities in how these molecules function as transcription factors emerge naturally from solutions to the relevant optimization problem.

Significance

Many cellular processes depend on a quantitative response to the concentration of transcription factor molecules. A plethora of different mechanisms contribute to this concentration sensing: Multiple enhancers with a combination of binding sites regulate genes together based on spatially heterogeneous transcription factors. Using the early fly embryo as an example, we investigate abstract sensors with limited capacity due to noise and optimize so that the sensors capture as much information as possible about a cell's position in the embryo. The resulting optimal sensors have important features in common with the known mechanisms of enhancer function.

Author contributions: M.B., M.D.P., T.G., E.F.W., and W.B. designed research; M.B., M.D.P., T.G., E.F.W., and W.B. performed research; M.B., M.D.P., T.G., E.F.W., and W.B. contributed new reagents/analytic tools; M.B. and W.B. analyzed data; and M.B., M.D.P., T.G., E.F.W., and W.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: mb67@princeton.edu or wbialek@princeton.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2109011118/-/DCSupplemental.

Published November 12, 2021.

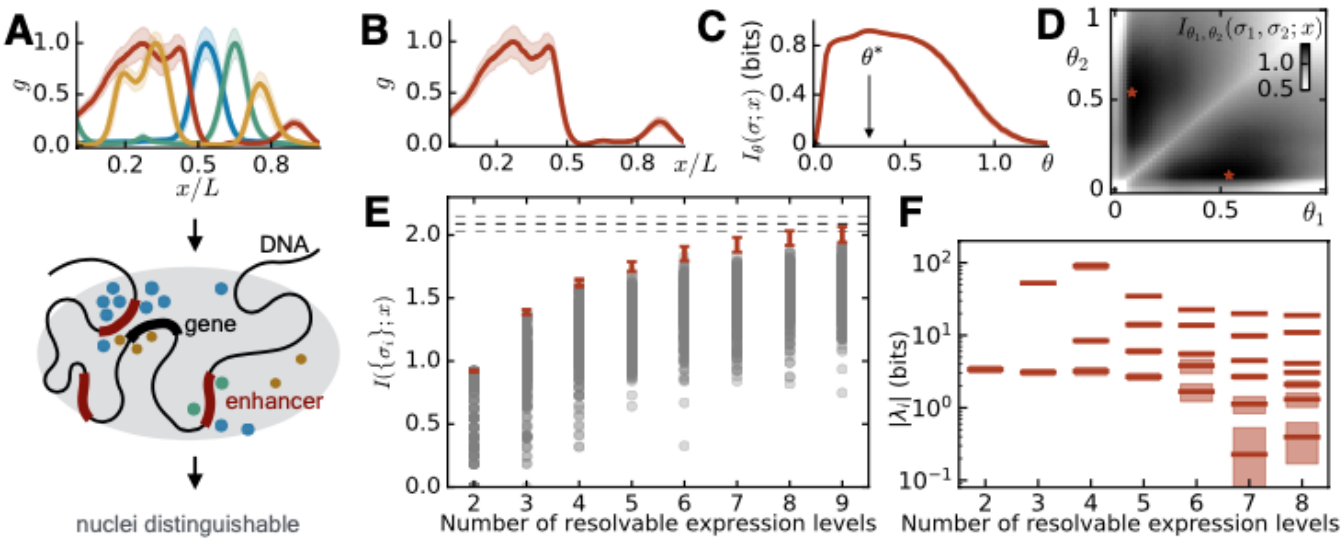

Fig. 1. Optimizing the flow of information provided through four transcription factors in the early fly embryo, here through thresholded sensing elements. (*A*) The four gap gene expression patterns (*kruppel, knirps, giant, and hunchback* in color, details in the text and later) provide information about distinguishable nuclear cell fates along the embryo's anterior–posterior axis (x), which needs to be identifiable after the fly's transcriptional apparatus measures or senses the TFs: Here we investigate an abstract sensor to learn what features of the gap expression profiles a sensor should concentrate on to optimize this information transfer. Biologically, this sensing is done by the regulatory elements. (*B*) Hb expression level vs. position along the anterior–posterior axis embryo. Shown are mean (line) $\pm$ 1 SD (shading) across $N_{\rm em} = 38$ embryos in a 5-min window (40 to 44 min) in nuclear cycle 14 (17). (*C*) Positional information vs. threshold, from Eq. **4**. (*D*) Positional information with two thresholds, $I_{\theta_1\theta_2}(\sigma_1, \sigma_2; x)$ (bits). (*E*) Positional information captured with $i = 1, \ldots, K$ thresholds, as a function of the number of resolvable levels $K + 1$. Error bars (red) are mean $\pm$ 1 SE of our estimate of the maximum. Circles (gray) are 300 values of $I(\{\sigma_i\}; x)$ at random settings of the K thresholds $\{\theta_i\}$. The black dashed line indicates the positional information $I(g; x)$, available from the expression levels if measured precisely, and gray dashed lines are $\pm$ 1 SE in our estimate of this information. (*F*) Eigenvalues $\{\lambda_i\}$ of the Hessian matrix χ, from Eq. **10**. The number of eigenvalues is the number of thresholds, one less than the number of resolvable expression levels. Shaded bands are $\pm$ 1 SE in our estimates.

Thresholds

The classical view of the gap genes is that they are expressed in domains (14). Implicitly this suggests that fine-scale variations in the concentration of these molecules are not important; rather all that matters is whether expression is on or off. The quantitative version of this idea is that subsequent events are sensitive to whether expression levels are above or below a threshold, corresponding to whether a cell is inside or outside an expression domain. We know that such simple thresholding loses a lot of the information that gap gene expression levels carry about position along the anterior–posterior axis (15). Still, we will look at this thresholding approach to gene regulation more precisely, using the expression profile for a single gap TF protein, Hb, as an example. While a single thresholding operation throws away more than half of the available information, we will see that this information could be recovered by multiple parallel thresholding mechanisms or equivalently by a single mechanism that could distinguish multiple "quantized" levels of expression. Importantly, in either case these thresholds do not need to be finely tuned, suggesting that there are plausible pathways for evolution to find mechanisms with close to optimal performance. This concrete discussion of thresholding also is meant to provide some foundation for the more abstract view of optimal sensing and compression that we introduce in the next section.

In Fig. 1 *B–E* we use the gap protein Hunchback (Hb) to illustrate the information loss associated with thresholding. At each point x there is an expression level g (Fig. 1*B*), drawn from a probability distribution $P(g|x)$; looking at many embryos we have samples out of this distribution. Experimental data are from ref. 17, where immunostaining of the proteins was used to obtain expression profiles of the genes. We focus on a time window of 40 to 44 min into nuclear cycle 14, the final cycle before blastoderm stage, during which the gap gene expression determines crucially the cell fates of nuclei along the embryo's anterior–posterior axis through pair-rule, segment polarity, and hox gene expression.

If cells are only sensitive to whether expression levels are above or below a threshold θ, then the variable that matters is

$$\sigma = H(g - \theta), \qquad [1]$$

where H is the Heaviside step function, $H(y > 0) = 1$, and $H(y < 0) = 0$. Then we can estimate the θ (threshold)-dependent distribution $P_\theta(\sigma|x)$,

$$P_\theta(\sigma = 1|x) = \int dg\, H(g - \theta) P(g|x) \qquad [2]$$

$$P_\theta(\sigma = 0|x) = 1 - P_\theta(\sigma = 1|x). \qquad [3]$$

Finally, we compute the amount of (mutual) information that the discrete variable σ provides about different possible nuclear cell fates, quantified by the cell's position along the anterior–posterior axis,

$$I_\theta(\sigma; x) = \sum_\sigma \int dx\, P(x) P_\theta(\sigma|x) \log_2 \left[\frac{P_\theta(\sigma|x)}{P_\theta(\sigma)} \right] \text{ bits}, \qquad [4]$$

where $P(x) = 1/L$, as a priori all positions along the length of the embryo are equally likely, and

$$P_\theta(\sigma) = \int dx\, P(x) P_\theta(\sigma|x). \qquad [5]$$

It is important that in exploring the impact of thresholding we allow for the best possible choice of the threshold θ, which in this example proves to be at $\theta^* \sim 1/3$ of the maximal mean expression level (Fig. 1*C*).

If the expression level is represented only by the on/off or binary variable σ, then it can provide at most one bit of information (about anything). We see that the mutual information about

position obtained by a thresholded measurement comes close to this bound, with $I_{\rm max}(\sigma; x) = 0.92 \pm 0.01$ bits. But this is less than one-half of the information that is carried by the Hb expression levels,*

$$I(g;x) \equiv \int dg \int dx P(x)P(g|x) \log_2 \left[\frac{P(g|x)}{P(g)}\right] \quad [6]$$

$$= 2.09 \pm 0.06 \text{ bits}. \quad [7]$$

Following appendix A8 of ref. 18 we analyze subsets of the data to correct for effects of finite sample size and estimate errors.

One path to recovering the information that was lost by the thresholded measurement is to imagine that the cell can resolve more details, perhaps distinguishing reliably among three or four different expression levels rather than just two. This is equivalent to the cell having multiple readout mechanisms, each of which can distinguish only on/off, but with different on/off switches having different thresholds, in the spirit of the "French flag" model (19). Because we can always put the thresholds in order, having K binary switches is the same as distinguishing $K + 1$ different expression levels. It can be useful to think of thresholding as being implemented at individual binding sites for the TFs, or perhaps at cooperative arrays of binding sites in enhancers, but our arguments are independent of these microscopic details.

If we have two different elements, each of which reports on whether the expression level is above or below a threshold, then the relevant variables are

$$\sigma_1 = H(g - \theta_1) \quad [8]$$

$$\sigma_2 = H(g - \theta_2). \quad [9]$$

We see in Fig. 1*D* that there is a broad optimum in the positional information that these variables capture, $I_{\theta_1\theta_2}(\{\sigma_1, \sigma_2\}; x)$, when the two thresholds are quite different from one another, $\theta_1^* = 0.1$ and $\theta_2^* = 0.58$; these bracket the optimal single threshold $\theta = 0.34$. The maximum information now is $I_{\rm max}(\{\sigma_1, \sigma_2\}; x) = 1.4 \pm 0.015$ bits, noticeably more than in the case with one threshold but still far from capturing all the available information.

We can generalize this idea to multiple thresholding elements, which are described by a set of variables $\{\sigma_i\}$, with each $\sigma_i = H(g - \theta_i)$, for $i = 1, 2, \cdots, K$; the relevant quantity now is $I(\{\sigma_i\}; x)$. This positional information depends on all the thresholds $\{\theta_i\}$, and we perform a multidimensional optimization to find the maximum of $I(\{\sigma_i\}; x)$. Fig. 1*E* shows that for cells to extract all the positional information available from the Hb concentration, they must distinguish eight or nine different expression levels, representing g with $\sim \log_2 8 = 3$ bits of precision.

Distinguishing eight levels in this simple threshold picture requires the cell to set seven thresholds. It might seem as though this necessitates setting each threshold to its optimal value, a form of fine tuning. To explore this we choose thresholds at random, uniformly in the relevant interval $0 < \theta < 1$. As shown in Fig. 1*E*, typical random choices are far below the optimum, as expected. But Fig. 1 *C* and *D* shows that there is a broad plateau in information vs. one or two thresholds, which suggests that multiple threshold choices could yield good results. Indeed, even with eight thresholds we find that more than 1 in 1,000 of our random choices in in Fig. 1*E* come within error bars of the optimum.

Another way of looking at the issue of fine tuning is to examine the behavior of the information in the neighborhood of the optimum,

$$I(\{\theta_i\}) = I_{\rm max}(K) + \frac{1}{2}\sum_{i,j=1}^{K}(\theta_i - \theta_i^*)\chi_{ij}(\theta_j - \theta_j^*) + \cdots, \quad [10]$$

estimating the Hessian matrix χ numerically from the data. The matrix χ has units of bits, as we chose the thresholds to be dimensionless. The eigenvectors of χ determine the combinations of thresholds that have independent effects on the information, and the eigenvalues $\{\lambda_i\}$ of χ (also in bits) determine the sensitivity along these independent directions. As the number of thresholds increases, we find a broad spread of eigenvalues (Fig. 1*F*), as in a wide class of "sloppy models" (20, 21). This means that some combinations of thresholds are two orders of magnitude more important than others.

In more detail, we find that the eigenvector with the largest eigenvalue is concentrated on the lowest thresholds. For example, with three thresholds, the eigenvector associated with the largest eigenvalue is (–0.99, 0.08, 0.06). As more thresholds are added, the eigenvectors of the largest two eigenvalues are combinations of the lowest two thresholds or correspond to one of them directly, while the smaller eigenvalues more loosely correspond to linear combinations of higher thresholds.

Although we should be cautious about overly detailed molecular interpretations, it is natural to think of the mapping $g \to \{\sigma_i\}$ as being implemented by binding of the transcription factor to specific sites along the genome, so that thresholds are set by the binding constants or affinities of the TF for these sites. The spectrum of χ and the fact that the lowest threshold corresponds to the largest eigenvalue tell us that the affinity at the strongest binding site (for low concentrations) must be set carefully, but the weaker binding sites can be scattered more freely across the available dynamic range of concentrations. A near-optimal array of thresholds thus could evolve by duplication of a strong binding site, followed by sequence drift to weaker binding, and then selection for the more complex and reproducible patterns that result from capturing more positional information (22).

Beyond Thresholds

The idea that cells are sensitive only to whether the concentration of a transcription factor is above or below a threshold is used quite widely, if informally (23–27). This picture embodies the intuition that arbitrarily small changes in TF concentration cannot generate reliable responses. But if we take thresholding seriously, it involves a perfect, noise-free distinction between concentrations that are just above and just below threshold. We want to have a more realistic description while avoiding an explosion of parameters.

Transcription factors are thought to influence transcription only through their binding to target sites. These targets are defined by the presence of specific DNA sequences, termed regulatory elements or enhancers. In this broad class of molecular mechanisms, the cell does not have direct access to the TF concentration g, but only to the occupancy of the binding sites, perhaps averaged over time (28–31). A detailed model would include many components: There can be multiple interacting binding sites; these sites and the bound TF molecules can interact with a host of other molecules, perhaps condensed into a phase-separated droplet surrounding the site of active transcription (32, 33); and there can be many molecular steps through which TF binding actually influences the initiation of transcription. A full model including all these complexities would have many parameters and would lose much of its predictive power.

What is essential is that binding of TF molecules to their target sites is a noisy process, for fundamental physical reasons (1, 9–12). If we abstract away from the details, transcription is controlled not by the TF concentration directly, but by some intermediate variable, such as the occupancy of the relevant binding sites. We can think of this intermediate variable as a sensor of the TF concentration. Because the sensing mechanisms are noisy, this sensor can provide only a limited amount of information about the actual concentration.

*Integrals are evaluated with a bin size of $\Delta g \sim 0.03$ and $\Delta x = 0.005$.

Rather than trying to make a detailed model within which we can calculate the levels of noise and the resulting limits to information, we want to understand the consequences of these limits. We assume, generally, that the TF concentration g is being mapped into some other variable by the sensor, and we can call this variable C. This (noisy) mapping $g \to C$ can be expressed in a probability distribution $P(C|g)$, which describes the sensor. Since we do not know which of the molecular mechanisms the cell uses to measure, and thus how precision is limited, we want to assume the most general or unbiased version of limited precision. Thus, we describe limited precision by limiting the mutual information,

$$I(C;g) = \sum_C \int dg\, P(C,g) \log_2 \left[\frac{P(C|g)}{P(C)} \right], \qquad [11]$$

which is transmitted from the TF concentration variable g to the sensor's encoding C. Different molecular mechanisms generate different mappings $g \to C$, but in all mechanisms the low concentrations of the relevant molecules limit the information that is transmitted. Thus, a biological sensor, corresponding to a regulatory element or enhancer with biologically reasonable arrival statistics of TF molecules, necessarily experiences a limitation on its information capacity $I(C;g)$; this is a more general as well as realistic constraint than thresholding.

Information Bottleneck and the Optimal Sensor

We now want to find the mapping $g \to C$ that conveys the highest biologically relevant positional information, $I(C;x)$, for a range of limited capacities $I(C;g)$. We refer to these mappings as optimal sensors. For comparison, the thresholded sensors discussed in the previous section correspond to deterministic mappings [all $P(C|g) \in \{0,1\}$] with a small number of discrete states or levels $||C||$ in the variable C.

Instead of restricting to thresholds, we want to search over all mappings $g \to C$ with a fixed $I(C;g)$ and maximize $I(C;x)$. This can be expressed as an optimization problem,

$$\max_{P(C|g)} \left[I(C;x) - TI(C;g) \right], \qquad [12]$$

where T is a Lagrange multiplier that allows us to modulate the constraint on sensor capacity $I(C;g)$. This problem of optimizing $P(C|g)$ is known as the "information bottleneck" problem (34). Its solution gives an iterative algorithm that finds $P(C|g)$. This problem and the algorithm have implications for machine learning (35, 36) or finding efficient encodings in neuronal systems (37); in these fields, the optimal $P(C|g)$ is often described as a compression of g. Qualitatively, the algorithm identifies (potentially noisy) sets of values of g that are most informative about x and focuses $P(C|g)$ to make maximum use of those values.

We solve the optimization problem in Eq. **12** numerically, considering C to be a variable with a discrete set of values or states and varying the number of these states, $||C||$. At fixed $||C||$, decreasing T allows $I(C;g)$ to be larger and pushes the noisy mapping $P(C|g)$ toward being deterministic. Results of the bottleneck analysis for Hb are shown in Fig. 2 as trajectories (solid gray and black lines) in the plane $I(C;x)$ vs. $I(C;g)$. Only the region below the dashed diagonal and horizontal lines is theoretically accessible due to the data-processing inequality [$I(C;x) \le I(C;g)$ and $I(C;x) \le I(g;x)$], which implies that even an optimal sensor cannot know more about positional information or nuclear cell fates than is provided by the protein expression itself. Often, for example for neuronal systems (37), the bounding curve for the optimal sensor at large $||C||$ (solid black line) is farther away from the data-processing bound than here. This optimal bounding curve that emerges from the information bottleneck analysis separates the plane into a physically possible region (below the curve) and an impossible region (above the curve). As $I(C;g)$ becomes large, the curve plateaus at the available positional information $I(g;x)$.

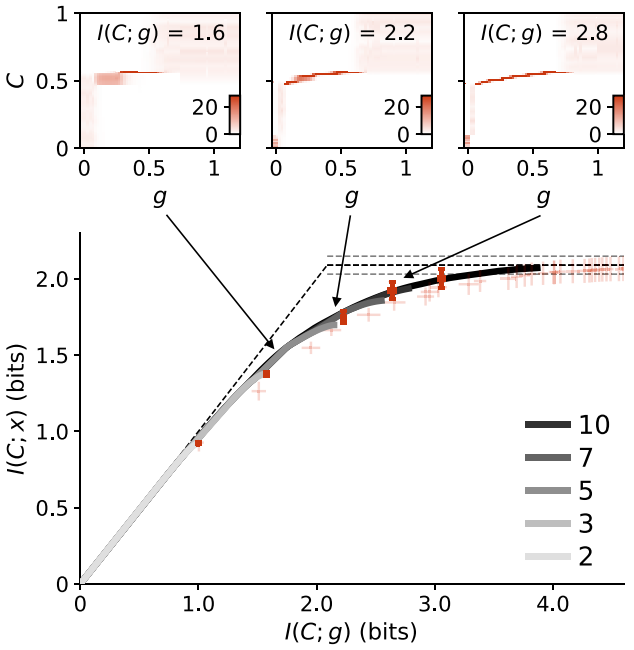

Fig. 2. The information bottleneck for positional information carried by Hb expression levels. We map expression into some compressed description, $g \to C$, and find the maximum $I(C;x)$ at fixed $I(C;g)$, from Eq. **12**, shown as the solid line with different grayshades indicating different numbers of states $||C||$. Solid red points with error bars are the $I(\theta_i;g)$–$I(\theta_i;x)$ pairs from the optimal discretization by multiple thresholds in Fig. 1*E* and match with the $T \to 0$ limit of the bottleneck solutions with fixed $||C||$. The light red crosses are from an explicitly deterministic formulation of the bottleneck problem (39). *Top* shows probability distributions $P(C|g)$ at different information capacities $I(C;g)$ along the bottleneck curve, here for $||C|| = 70$ states; intermediate levels of $g \in [0.05, 0.8]$ are progressively better resolved as the capacity increases.

The optimal thresholding sensors from Fig. 1 correspond to the endpoints of the bottleneck solutions with $||C||$ equal to the number of resolvable expression levels. We see that these thresholded sensors, or deterministic endpoints of bottleneck solutions with finite $||C||$, are almost on the optimal bounding curve. This is unusual for general compression problems, where the optimal thresholded sensor falls below the optimal curve. Thus, although the picture of multiple noiseless thresholds is physically wrong, it does correspond, almost quantitatively, to an information theoretic optimization of positional information with the constraint of limited information capacity $I(C;g)$ in the sensor. This is important, because it suggests that the intuition behind the French flag model or the biological importance of the gap expression boundaries corresponds more closely than expected to a true information theoretic optimization.

We can understand more about the structure of the optimal mappings $g \to C$ by looking at the distributions $P(C|g)$, shown in Fig. 2, *Top*. These $P(C|g)$ correspond to the three $I(C;g)$, marked by the arrows, of the black information bottleneck curve, where we have used $||C|| = 70$ numerically but normalized to 1 to emphasize the almost continuous character of C. At small $I(C;g)$ whole ranges of g are mapped uniformly into ranges of C, while at larger $I(C;g)$ we see the emergence of a reliably graded mapping, especially in the range bracketing half-maximal expression. In all panels in Fig. 2, the optimal sensor focuses on the low expression levels of Hb, which are biologically the most precise expression levels (Fig. 1*A*). That the optimal sensor resolves these levels more than noisily expressed levels to receive the most information about the system is expected from intuition

for optimal sensor arrangements in neurons, in the spirit of ref. 38.

The light red crosses in Fig. 2 correspond to a greedy, deterministic approximation to the full optimization problem in the way of refs. 39 and 40; we provide more details on this calculation in *SI Appendix*. This approximation generates thresholded sensors, but as we add more thresholds one cannot go back to readjust the existing thresholds. Despite this restriction, the results are very close to the true optimum, so that there is a hierarchical evolutionary path to nearly optimal performance.

A detailed discussion of how the optimal sensor corresponds to models of sensing that involve binding-site occupation (41, 42) would go beyond the scope of this paper. Qualitatively, however, we note that Fig. 2, *Top* could be compared to such sensors, with the steep change in C vs. g corresponding to highly cooperative binding; interestingly the predicted degree of cooperativity depends on the sensor capacity $I(C;g)$.

Multiple Regulatory Elements for Hb

We see from Fig. 2 that capturing all the positional information encoded by Hb requires measuring the expression level with a sensor capacity of $I(C;g)\sim 3$ bits of precision. This is consistent with our conclusions from the analysis of thresholded sensors, where the optimal sensors with seven to nine thresholds also have a capacity of $I(C;g)\sim 3$ bits. We have done the same analysis for the other gap TF proteins (Kruppel, Kr; Giant, Gt; and Knirps, Kni) and also find that ~ 3 bits of capacity are required in each case.

How does this information capacity compare with the information capacity of biological regulatory elements, such as enhancers? Estimates based both on direct measurements and on more detailed models indicate that the capacity of a regulatory element is in the range of 1 to 3 bits (2, 43). These estimates depend on the absolute concentrations of the relevant molecules, on the time available for reading out the information, on the length of the regulatory elements, and on other details of the different noise sources in the system (2, 43). At one extreme, if the capacity of a biological regulatory element is 3 bits, then a single regulatory element is sufficient to capture the full positional information; in this case, it should have been possible for the fly's transcriptional apparatus to extract all the available positional information using only one regulatory element or enhancer, but this requires that this element operates close to the physical limits to information capacity. But if the capacity of a single element is only 1 bit, then we need multiple regulatory elements even in response to a single transcription factor. It is clear from Fig. 2 that there is a very big difference between a capacity of 1 bit and that of 3 bits.

Optimal Sensor for All the Gap Proteins

One might argue that the fly does not need to extract this much positional information about cell fates from Hb, as the other gap proteins provide information as well. Indeed, we know that biologically all four gap TF proteins (Kr, Kni, Gt, and Hb) are important for nuclei to take their correct cell fates. Practically, the temporal changes in the expression patterns could also be important (44, 45), but in the first instance we again focus on a sensor that measures the expression profiles 40 to 44 min into cycle 14, as it has been shown that these are sufficient to predict the positions of pair-rule stripes (17). Thus, we need to find the optimal sensor for the joint gap expression profiles to draw biologically relevant lessons from our approach. Rather than considering, as above, the mapping $g_{\rm Hb}\to C$, we can consider mappings from combinations of expression levels of multiple-gap TFs (Fig. 3*A*) into C, corresponding to a single optimal sensing

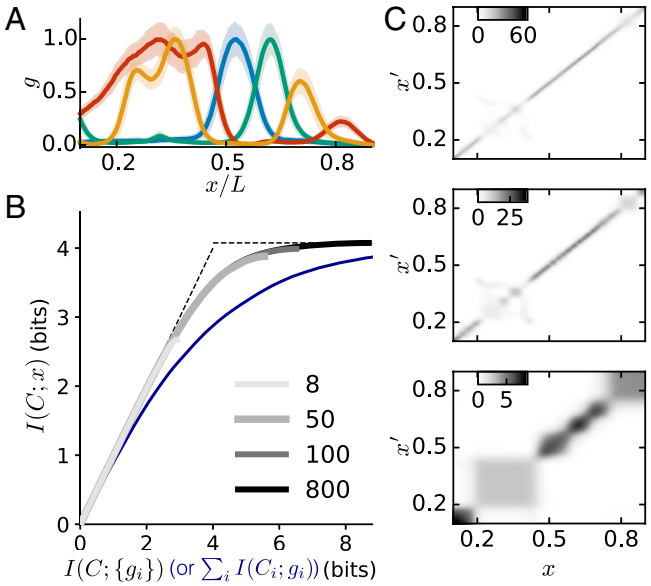

Fig. 3. The information bottleneck for positional information carried by all four gap gene expression levels. (*A*) Expression vs. position along the anterior–posterior axis for Hb (red), Kr (blue), Kni (green), and Gt (mustard). Shown are mean (solid) and SD (shading) across $N_{\rm em}=38$ embryos in a 5-min window (40 to 44 min) in nuclear cycle 14 (17). (*B*) Information bottleneck results, as in Fig 2. Shown are optimal solutions with $||C||=8$, 100, and 800 (shades of gray) and solutions with independent compression of each gene expression level (blue). (*C*) Decoding maps $P(x'|x)$ based on compressed representations of the expression levels: no compression (*Top*), $I(C;\{g_i\})=4$ bits (middle), and $I(C;\{g_i\})=2$ bits (*Bottom*).

element; i.e., $\{g_i\}\equiv\{g_{\rm Kr},g_{\rm Kni},g_{\rm Gt},g_{\rm Hb}\}\to C$. The analog of Eq. **12** is the optimization problem

$$\max_{P(C|\{g_i\})}\left[I(C;x)-TI(C;\{g_i\})\right]. \qquad [13]$$

We apply the information bottleneck scheme to find the optimal sensor and see that we can capture a significant fraction of the information provided by all gap TFs by keeping only 4 bits of information about their expression levels or just 1 bit per gene (Fig. 3*B*), but 4 bits still capture less than 90% of the available information.

We can visualize what is being gained as the sensor capacity $I(C;\{g_i\})$ increases, using the decoding maps introduced in ref. 17. The decoding map in Fig. 3 *C*, *Top* is the best possible decoding map given the expression levels that we observe experimentally. The maps show the distribution of positions x' consistent with the gap gene expression levels seen in nuclei at the true position x,

$$P(x'|x)=P(x'|\{g_i\})\Big|_{\{g_i=g_i(x)\}}; \qquad [14]$$

for simplicity we show this averaged over all the expression levels found at x. Using all the available information, $P(x'|x)$ forms a narrow band around $x'=x$, with width $\sigma_x/L\sim 0.01$ (17). In Fig. 3 *C*, *Middle* and *Bottom* we imagine that inference is based not on the actual expression levels but on the compressed version C,

$$P(x'|x)=\sum_C P(x'|C)P(C|\{g_i\})\Big|_{\{g_i(x)\}}, \qquad [15]$$

as explained in more detail in *SI Appendix*; we do this for the optimal compressions with $I(C;\{g_i\})=2$, and 4 bits. We see that as the compression becomes more severe, the inference becomes more uncertain (larger σ_x) and genuinely ambiguous.

This more noisy inference has biological consequences: Sensors with capacity of much less than 4 bits do not capture enough information to predict the patterns of pair-rule expression stripes in mutants, following the analysis in ref. 17.

To extract all the available positional information requires mechanisms that preserve ≥ 8 bits of information about the combined expression levels of the four gap genes. The grayscale in Fig. 3 indicates that a $||C||$ of at least 30 to 50 levels would be required. Again, if we think that a single sensor can implement one threshold, this means that more than one sensor would be required, even in the best possible case where information-theoretically optimal sensing is possible. We know that there are several dozen enhancer sites that respond to the gap gene TFs, and we see that this degree of complexity may be required by information theoretic constraints, even if these sensors make optimal use of the available information.

We end with a final note regarding the splitting of the optimal sensor into multiple sensors. We investigate four sensors C_i ($\mathrm{i} \in \{1, 4\}$), where each sensor can respond only to a single one of the four gap transcription factors. Mathematically, this corresponds to demanding that the compressed variables be constructed from individual gene expression levels, so that $g_{\mathrm{Hb}} \rightarrow C_1$, $g_{\mathrm{Kr}} \rightarrow C_2$, etc., but all the states of the compressed variable $C = \{C_1, C_2, C_3, C_4\}$ can provide positional information. More precisely, we optimize all of the individual distributions $P_{\mathrm{i}}(C_{\mathrm{i}}|g_{\mathrm{i}})$, and the objective function is

$$\mathcal{F} = I(\{C_{\mathrm{i}}\}; x) - T\sum_{\mathrm{i}=1}^{4} I(C_{\mathrm{i}}; g_{\mathrm{i}}). \qquad [16]$$

We find that such a set of four sensors always is substantially worse than a single optimal sensor, as indicated by the blue line in Fig. 3, even with same total information capacity (for more details see *SI Appendix*). This loss of information indicates the importance of having regulatory mechanisms that are sensitive to combinations of transcription factors. In fact, the readout of positional information encoded in the gap proteins is implemented by the array of enhancers controlling pair-rule gene expression, and these enhancers are prototypical instances of regulatory elements that respond to combinations of transcription factors (29, 32, 46, 47). While there is some distance between our abstract formulation and the molecular details, it is attractive to see that this mechanistic complexity is required as a response to basic physical and information theoretic limitations.

Conclusion

To summarize, individual regulatory mechanisms have limited information capacity, and our central result is that this capacity in turn sets strict limits on the amount of positional information that can be extracted from the gap gene expression levels. In this paper, we see the measurement of the transcription factors as a problem of efficient sensing or compression and use the information bottleneck algorithm to identify an optimal sensor for this network. Precise comparison with ideas about thresholded reading of the gap TF Hb shows that the thresholds do not need to be fine-tuned and exhibit a hierarchy of sensitivities. Crucially, we find that it almost certainly is not possible to read out enough positional information with a single enhancer element. For the nuclei to obtain at least 90% of the information provided by the gap TF network, a large number of thresholds (30–50) or a high capacity in the optimal sensor is required, and this must be realized by multiple enhancers. Further, if each enhancer responds to a single TF, there is a dramatic loss of efficiency. The information theoretic optimization principle we have explored here thus predicts that expression levels must be read by multiple enhancers, each sensitive to combinations of the gap TFs. This complex enhancer logic indeed is how gap gene expression levels drive downstream events in the fly embryo.

Data Availability. Previously published data were used for this work (17).

ACKNOWLEDGMENTS. We thank P.-T. Chen, M. Levo, R. Munshi, B. van Opheusden, and R. Rao for helpful discussions. This work was supported in part by the US NSF, through the Center for the Physics of Biological Function (PHY–1734030) and the Center for the Science of Information (CCF–0939370); by NIH Grant R01GM097275; by the Alexander von Humboldt Stiftung; and by the Howard Hughes Medical Institute.

1. W. Bialek, S. Setayeshgar, Physical limits to biochemical signaling. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10040–10045 (2005).
2. G. Tkačik, C. G. Callan Jr., W. Bialek, Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12265–12270 (2008).
3. J. O. Dubuis, R. Samanta, T. Gregor, Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol. Syst. Biol.* **9**, 639 (2013).
4. T. Gregor, D. W. Tank, E. F. Wieschaus, W. Bialek, Probing the limits to positional information. *Cell* **130**, 153–164 (2007).
5. A. Abu-Arish, A. Porcher, A. Czerwonka, N. Dostatni, C. Fradin, High mobility of bicoid captured by fluorescence correlation spectroscopy: Implication for the rapid establishment of its gradient. *Biophys. J.* **99**, L33–L35 (2010).
6. J. A. Drocco, O. Grimm, D. W. Tank, E. Wieschaus, Measurement and perturbation of morphogen lifetime: Effects on gradient shape. *Biophys. J.* **101**, 1807–1815 (2011).
7. C. E. Hannon, S. A. Blythe, E. F. Wieschaus, Concentration dependent chromatin states induced by the bicoid morphogen gradient. *eLife* **6**, e28275 (2017).
8. S. E. Keenan *et al.*, Rapid dynamics of signal-dependent transcriptional repression by Capicua. *Dev. Cell* **52**, 794–801.e4 (2020).
9. H. C. Berg, E. M. Purcell, Physics of chemoreception. *Biophys. J.* **20**, 193–219 (1977).
10. K. Kaizu *et al.*, The Berg-Purcell limit revisited. *Biophys. J.* **106**, 976–985 (2014).
11. T. Friedlander, R. Prizak, C. C. Guet, N. H. Barton, G. Tkačik, Intrinsic limits to gene regulation by global crosstalk. *Nat. Commun.* **7**, 12307 (2016).
12. T. Mora, I. Nemenman, Physical limit to concentration sensing in a changing environment. *Phys. Rev. Lett.* **123**, 198101 (2019).
13. C. Nüsslein-Volhard, E. Wieschaus, Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
14. J. Jaeger, The gap gene network. *Cell. Mol. Life Sci.* **68**, 243–274 (2011).
15. J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, W. Bialek, Positional information, in bits. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16301–16308 (2013).
16. F. Liu, A. H. Morrison, T. Gregor, Dynamic interpretation of maternal inputs by the Drosophila segmentation gene network. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6724–6729 (2013).
17. M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, T. Gregor, Optimal decoding of cellular identities in a genetic network. *Cell* **176**, 844–855.e15 (2019).
18. W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton, NJ, 2012).
19. L. Wolpert, Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969).
20. R. N. Gutenkunst *et al.*, Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, 1871–1878 (2007).
21. M. K. Transtrum *et al.*, Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *J. Chem. Phys.* **143**, 010901 (2015).
22. P. François, E. D. Siggia, Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development* **137**, 2385–2395 (2010).
23. J. Lewis, J. M. W. Slack, L. Wolpert, Thresholds in development. *J. Theor. Biol.* **65**, 579–590 (1977).
24. J. B. Green, J. C. Smith, Growth factors as morphogens: Do gradients and thresholds establish body plan? *Trends Genet.* **7**, 245–250 (1991).
25. J. B. Green, H. V. New, J. C. Smith, Responses of embryonic Xenopus cells to activin and FGF are separated by multiple dose thresholds and correspond to distinct axes of the mesoderm. *Cell* **71**, 731–739 (1992).
26. K. W. Rogers, A. F. Schier, Morphogen gradients: From generation to interpretation. *Annu. Rev. Cell Dev. Biol.* **27**, 377–407 (2011).
27. J. Briscoe, S. Small, Morphogen rules: Design principles of gradient-mediated embryo patterning. *Development* **142**, 3996–4009 (2015).
28. L. Bintu *et al.*, Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005).
29. E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, U. Gaul, Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**, 535–540 (2008).
30. G. Tkačik, A. M. Walczak, W. Bialek, Optimizing information flow in small genetic networks. *Phys. Rev. E* **80**, 031920 (2009).
31. O. Pulkkinen, R. Metzler, Distance matters: The impact of gene proximity in bacterial gene regulation. *Phys. Rev. Lett.* **110**, 198101 (2013).
32. E. E. M. Furlong, M. Levine, Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
33. B. R. Sabari *et al.*, Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
34. N. Tishby, F. C. Pereira, W. Bialek, "The information bottleneck method" in *Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing*, B. Hajek, R. S. Sreenivas, Eds. (University of Illinois, Champaign, IL, 1999), pp 368–377.
35. R. Shwartz-Ziv, N. Tishby, Opening the black box of deep neural networks via information. arXiv [Preprint] (2017). https://arxiv.org/abs/1703.00810 (Accessed 3 November 2021).

36. A. A. Alemi, I. Fischer, J. V. Dillon, K. Murphy, Deep variational information bottleneck. arXiv [Preprint] (2016). https://arxiv.org/abs/1612.00410 (Accessed 3 November 2021).
37. S. E. Palmer, O. Marre, M. J. Berry 2nd, W. Bialek, Predictive information in a sensory population. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6908–6913 (2015).
38. S. Laughlin, A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch. C Biosci.* **36**, 910–912 (1981).
39. N. Slonim, N. Tishby, "Agglomerative information bottleneck" in *Advances in Neural Information Processing 12*, S. Solla, T. Leen, K. Müller, Eds. (MIT Press, Cambridge, MA, 2000), pp. 617–623.
40. D. J. Strouse, D. J. Schwab, The deterministic information bottleneck. *Neural Comput.* **29**, 1611–1630 (2017).
41. J. Estrada, F. Wong, A. DePace, J. Gunawardena, Information integration and energy expenditure in gene regulation. *Cell* **166**, 234–244 (2016).
42. S. E. Marzen, J. P. Crutchfield, Prediction and dissipation in nonequilibrium molecular sensors: Conditionally Markovian channels driven by memoryful environments. *Bull. Math. Biol.* **82**, 25 (2020).
43. G. Tkačik, C. G. Callan Jr., W. Bialek, Information capacity of genetic regulatory elements. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **78**, 011910 (2008).
44. O. Wartlick *et al.*, Dynamics of Dpp signaling and proliferation control. *Science* **331**, 1154–1159 (2011).
45. N. B. Becker, A. Mugler, P. R. ten Wolde, Prediction and dissipation in biochemical sensing. *Phys. Rev. Lett.* **115**, 258103 (2015).
46. J. O. Yáñez-Cuna, E. Z. Kvon, A. Stark, Deciphering the transcriptional cis-regulatory code. *Trends Genet.* **29**, 11–22 (2013).
47. J. Crocker, G. R. Ilsley, D. L. Stern, Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors. *Nat. Genet.* **48**, 292–298 (2016).
48. C. Schulz, D. Tautz, Autonomous concentration-dependent activation and repression of *Krüppel* by *hunchback* in the *Drosophila* embryo. *Development* **120**, 3043–3049 (1994).
49. L. Descheemaeker, E. Peeters, S. de Buyl, Non-monotonic auto-regulation in single gene circuits. *PLoS One* **14**, e0216089 (2019).
50. M. V. Staller *et al.*, Shadow enhancers enable Hunchback bifunctionality in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 785–790 (2015).