



**HAL**  
open science

## Trading bits in the readout from a genetic network

Marianne Bauer, Mariela D Petkova, Thomas Gregor, Eric F Wieschaus,  
William Bialek

► **To cite this version:**

Marianne Bauer, Mariela D Petkova, Thomas Gregor, Eric F Wieschaus, William Bialek. Trading bits in the readout from a genetic network. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118 (46), pp.e2109011118. 10.1073/pnas.2109011118 . pasteur-03216308v2

**HAL Id: pasteur-03216308**

**<https://pasteur.hal.science/pasteur-03216308v2>**

Submitted on 13 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Trading bits in the readout from a genetic network

Marianne Bauer,<sup>1–4,\*</sup> Mariela D. Petkova,<sup>5</sup> Thomas Gregor,<sup>1,2,6</sup> Eric F. Wieschaus,<sup>2–4</sup> and William Bialek<sup>1,2,7,\*</sup>

<sup>1</sup>Joseph Henry Laboratories of Physics, <sup>2</sup>Lewis–Sigler Institute for Integrative Genomics, <sup>3</sup>Department of Molecular Biology, and <sup>4</sup>Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544 USA; <sup>5</sup>Program in Biophysics, Harvard University, Cambridge, MA 02138 USA; <sup>6</sup>Department of Developmental and Stem Cell Biology, UMR3738, Institut Pasteur, 75015 Paris, France; <sup>7</sup>Initiative for the Theoretical Sciences, The Graduate Center, City University of New York, 365 Fifth Ave, New York, NY 10016 USA; \*Corresponding authors: mb67@princeton.edu, wbialek@princeton.edu

**In the regulation of gene expression, information of relevance to the organism is represented by the concentrations of transcription factor molecules. In order 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 in order 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 proven 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 (16, 17). The local concentrations of the gap proteins provide just enough information to support this level of precision (16). 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 (18).

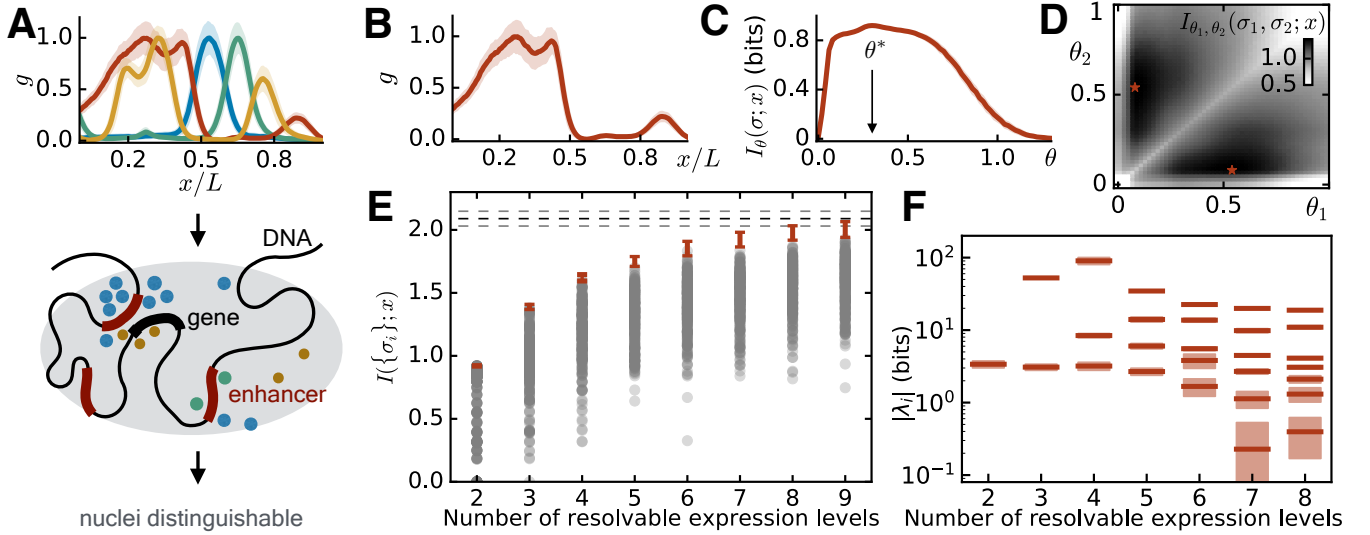
The gap gene network offers us the chance to ask how accurately the transcription factor concentrations need to be measured and to infer features of the regulatory architecture responsible for these measurements. The information that the gap genes convey about position along the anterior–posterior axis is what allows nuclei to make distinct cell fate decisions required for development; we investigate here how this can be seen as a sensing or signal processing problem (Fig 1A). 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. 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.

## 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 whether expression is on or off. The

### Significance Statement

Many cellular processes depend on a quantitative response to the concentration of transcription factor molecules. A plethora of different mechanisms that 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.



**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 (*Krupel*, *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. Mean (line)  $\pm$  one standard deviation (shading) across  $N_{em} = 38$  embryos in a five minute window (40–44 min) in nuclear cycle 14 (18). (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  $i = 1, \dots, K$  thresholds, as a function of the number of resolvable levels  $K + 1$ . Error bars (red) are mean  $\pm$  one standard error of our estimate of the maximum. Circles (grey) 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  $\pm$  one standard error in our estimate of this information. (F) Eigenvalues  $\{\lambda_i\}$  of the Hessian matrix  $\chi$ , from Eq (10). The number of eigenvalues is the number of thresholds, one less than the number of resolvable expression levels. Shaded bands are  $\pm$  one standard error in our estimates.

67 quantitative version of this idea is that subsequent events are  
 68 sensitive to whether expression levels are above or below a  
 69 threshold, corresponding to whether a cell is inside or outside  
 70 an expression domain. We know that such simple thresholding  
 71 loses a lot of the information that gap gene expression  
 72 levels carry about position along the anterior-posterior axis  
 73 (16). Still, we will look at this thresholding approach to gene  
 74 regulation more precisely, using the expression profile for a  
 75 single gap TF protein, Hb, as an example. While a single  
 76 thresholding operation throws away more than half of the  
 77 available information, we will see that this information could  
 78 be recovered by multiple parallel thresholding mechanisms,  
 79 or equivalently by a single mechanism that could distinguish  
 80 multiple “quantized” levels of expression. Importantly, in  
 81 either case these thresholds do not need to be finely tuned,  
 82 suggesting that there are plausible pathways for evolution to  
 83 find mechanisms with close to optimal performance. This  
 84 concrete discussion of thresholding also is meant to provide  
 85 some foundation for the more abstract view of optimal sensing  
 86 and compression that we introduce in the next section.

87 In Figure 1B-E we use the gap protein *hunchback* (Hb) to  
 88 illustrate the information loss associated with thresholding.  
 89 At each point  $x$  there is an expression level  $g$  (Fig. 1B), drawn  
 90 from a probability distribution  $P(g|x)$ ; looking at many embryos  
 91 we have samples out of this distribution. Experimental  
 92 data are from Ref (18), where immunostaining was used to  
 93 obtain expression profiles of the gap proteins. We focus on a  
 94 time window of 40 – 44 min into nuclear cycle 14, the final  
 95 cycle before blastoderm stage, during which the gap gene  
 96 expression determine crucially the cell fates of nuclei along the  
 97 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  $\theta$ , then the variable which matters  
 is

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

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

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

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

Finally we compute the amount of (mutual) information that  
 the discrete variable  $\sigma$  provides about different possible nuclear  
 cell fates, quantified by the cell’s position along the anterior-  
 posterior axis,

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

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

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

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

120 If the expression level is represented only by the on/off or  
 121 binary variable  $\sigma$ , then it can provide at most one bit of infor-  
 122 mation (about anything). We see that the mutual information  
 123 about position obtained by a thresholded measurement comes  
 124 close to this bound, with  $I_{\max}(\sigma; x) = 0.92 \pm 0.01$  bits. But  
 125 this is less than one half of the information that is carried by  
 126 the Hb expression levels, \*

$$127 \quad I(g; x) \equiv \int dg \int dx P(x) P(g|x) \log_2 \left[ \frac{P(g|x)}{P(g)} \right] \quad [6]$$

$$128 \quad = 2.09 \pm 0.06 \text{ bits.} \quad [7]$$

129 Following Appendix A8 of Ref (19) we analyze subsets of the  
 130 data to correct for effects of finite sample size and estimate  
 131 errors.

132 One path to recovering the information that was lost by the  
 133 thresholded measurement is to imagine that the cell can resolve  
 134 more details, perhaps distinguishing reliably among three or  
 135 four different expression levels rather than just two. This is  
 136 equivalent to the cell having multiple readout mechanisms,  
 137 each of which can only distinguish on/off, but with different  
 138 on/off switches having different thresholds, in the spirit of  
 139 the ‘‘French flag’’ model (15). Because we can always put the  
 140 thresholds in order, having  $K$  binary switches is the same  
 141 as distinguishing  $K + 1$  different expression levels. It can  
 142 be useful to think of thresholding as being implemented at  
 143 individual binding sites for the TFs, or perhaps at cooperative  
 144 arrays of binding sites in enhancers, but our arguments are  
 145 independent of these microscopic details.

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

$$149 \quad \sigma_1 = H(g - \theta_1) \quad [8]$$

$$150 \quad \sigma_2 = H(g - \theta_2). \quad [9]$$

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

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

169 Distinguishing eight levels in this simple threshold picture  
 170 requires the cell to set seven thresholds. It might seem as  
 171 though this necessitates setting each threshold to its optimal  
 172 value, a form of fine tuning. To explore this we choose thresh-  
 173 olds at random, uniformly in the relevant interval  $0 < \theta < 1$ .  
 174 As shown in Fig 1E, typical random choices are far below the  
 175 optimum, as expected. But Figures 1C and D show that there

176 is a broad plateau in information vs one or two thresholds,  
 177 which suggests that multiple threshold choices could yield  
 178 good results. Indeed, even with eight thresholds we find that  
 179 more than 1 in 1000 of our random choices in in Fig 1E come  
 180 within error bars of the optimum.

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

$$184 \quad I(\{\theta_i\}) = I_{\max}(K) + \frac{1}{2} \sum_{i,j=1}^K (\theta_i - \theta_i^*) \chi_{ij} (\theta_j - \theta_j^*) + \dots, \quad [10]$$

185 estimating the Hessian matrix  $\chi$  numerically from the data.  
 186 The matrix  $\chi$  has units of bits, as we chose the thresholds  
 187 to be dimensionless. The eigenvectors of  $\chi$  determine the  
 188 combinations of thresholds that have independent effects on  
 189 the information, and the eigenvalues  $\{\lambda_i\}$  of  $\chi$  (also in bits)  
 190 determine the sensitivity along these independent directions.  
 191 As the number of thresholds increases we find a broad spread  
 192 of eigenvalues (see Fig 1F), as in a wide class of ‘‘sloppy models’’  
 193 (20, 21). This means that some combinations of thresholds  
 194 are two orders of magnitude more important than others.

195 In more detail, we find that the eigenvector with the largest  
 196 eigenvalue is concentrated on the lowest thresholds. For exam-  
 197 ple, with three thresholds, the eigenvector associated with the  
 198 largest eigenvalue is  $(-0.99, 0.08, 0.06)$ . As more thresholds  
 199 are added, the eigenvectors of the largest two eigenvalues are  
 200 combinations of the lowest two thresholds or correspond to one  
 201 of them directly, while the smaller eigenvalues more loosely  
 202 correspond to linear combinations of higher thresholds.

203 Although we should be cautious about overly detailed molec-  
 204 ular interpretations, it is natural to think of the mapping  
 205  $g \rightarrow \{\sigma_i\}$  as being implemented by binding of the transcrip-  
 206 tion factor to specific sites along the genome, so that thresholds  
 207 are set by the binding constants or affinities of the TF for  
 208 these sites. The spectrum of  $\chi$  and the fact that the lowest  
 209 threshold corresponds to the largest eigenvalue tells us that the  
 210 affinity at the strongest binding site (for low concentrations)  
 211 must be set carefully, but the weaker binding sites can be  
 212 scattered more freely across the available dynamic range of  
 213 concentrations. A near optimal array of thresholds thus could  
 214 evolve by duplication of a strong binding site, followed by  
 215 sequence drift to weaker binding, and then selection for  
 216 the more complex and reproducible patterns that result from  
 217 capturing more positional information (22).

## 218 Beyond thresholds

219 The idea that cells are sensitive only to whether the concentra-  
 220 tion of a transcription factor is above or below a threshold is  
 221 used quite widely, if informally (23–27). This picture embodies  
 222 the intuition that arbitrarily small changes in TF concentration  
 223 can’t generate reliable responses. But if we take thresholding  
 224 seriously, it involves a perfect, noise-free distinction between  
 225 concentrations that are just above and just below threshold.  
 226 We would like to have a more realistic description while avoid-  
 227 ing an explosion of parameters.

228 Transcription factors are thought to influence transcription  
 229 only through their binding to target sites. These targets are  
 230 defined by the presence of specific DNA sequences, termed reg-  
 231 ulatory elements or enhancers. In this broad class of molecular  
 232 mechanisms, the cell does not have direct access to the TF

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

233 concentration  $g$ , but only to the occupancy of the binding sites,  
 234 perhaps averaged over time (28–31). A detailed model would  
 235 include many components: there can be multiple interacting  
 236 binding sites; these sites and the bound TF molecules can  
 237 interact with a host of other molecules, perhaps condensed  
 238 into a phase-separated droplet surrounding the site of active  
 239 transcription (32, 33); and there can be many molecular steps  
 240 through which TF binding actually influences the initiation  
 241 of transcription. A full model including all these complexities  
 242 would have many parameters, and would lose much of its  
 243 predictive power.

244 What is essential is that binding of TF molecules to their  
 245 target sites is a noisy process, for fundamental physical reasons  
 246 (1, 9–12). If we abstract away from the details, transcription is  
 247 controlled not by the TF concentration directly, but by some  
 248 intermediate variable, such as the occupancy of the relevant  
 249 binding sites. We can think of this intermediate variable as  
 250 a sensor of the TF concentration, and because the sensing  
 251 mechanisms are noisy it can provide only a limited amount of  
 252 information about the actual concentration.

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

$$265 \quad I(C; g) = \sum_C \int dg P(C, g) \log_2 \left[ \frac{P(C|g)}{P(C)} \right], \quad [11]$$

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

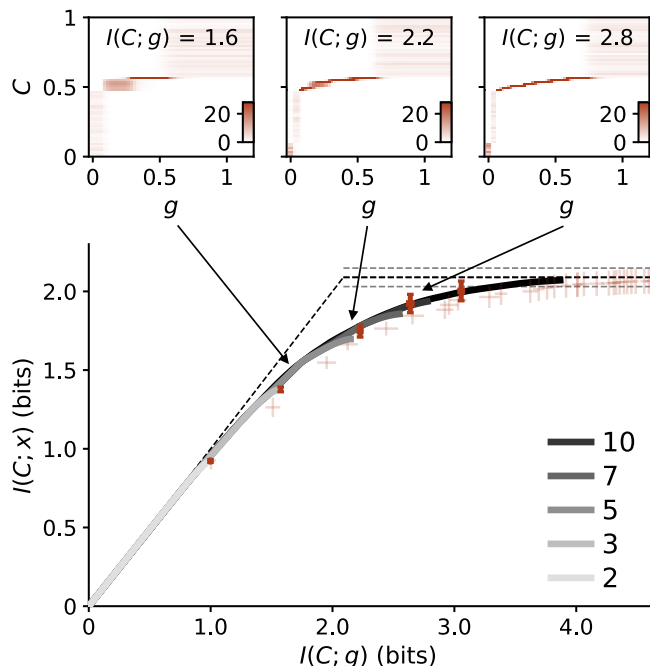
## 275 Information bottleneck and the optimal sensor

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

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

$$286 \quad \max_{P(C|g)} [I(C; x) - TI(C; g)], \quad [12]$$

287 where  $T$  is a Lagrange multiplier that allows us to modulate  
 288 the constraint on sensor capacity  $I(C; g)$ . This problem of



289 **Fig. 2.** The information bottleneck for positional information carried by Hb expression  
 290 levels. We map expression into some compressed description,  $g \rightarrow C$ , and find  
 291 the maximum  $I(C; x)$  at fixed  $I(C; g)$ , from Eq (12), shown as the solid line with  
 292 different greyshades indicating different numbers of states  $\|C\|$ . Solid red points with  
 293 error bars are the  $I(\theta_i; g) - I(\theta_i; x)$  pairs from the optimal discretization by multiple  
 294 thresholds in Fig 1E, and match with the  $T \rightarrow 0$  limit of the bottleneck solutions with  
 295 fixed  $\|C\|$ . The light points are from an explicitly deterministic formulation of the bot-  
 296 tleneck problem (39). Upper panel shows snapshots probability distributions  $P(C|g)$   
 297 at different information capacities  $I(C; g)$  along the bottleneck curve; intermediate  
 298 levels of  $g \in [0.05, 0.8]$  are progressively better resolved as the capacity increases.

299 optimizing  $P(C|g)$  is known as the “information bottleneck”  
 300 problem (34). Its solution gives an iterative algorithm which  
 301 finds  $P(C|g)$ . This problem and algorithm have implications  
 302 for machine learning (35, 36) or finding efficient encodings  
 303 in neuronal systems (37); in these fields, the optimal  $P(C|g)$   
 304 is often described as a compression of  $g$ . Qualitatively, the  
 305 algorithm identifies (potentially noisy) sets of values of  $g$  that  
 306 are most informative about  $x$ , and focuses  $P(C|g)$  to make  
 307 maximum use of those values.

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

317 plateaus at the available positional information  $I(g; x)$ .  
 318 The optimal thresholding sensors from Fig 1 correspond to  
 319 the endpoints of the bottleneck solutions with  $\|C\|$  equal to  
 320 the number of resolvable expression levels. We see that these  
 321 thresholded sensors, or deterministic endpoints of bottleneck  
 322 solutions with finite  $\|C\|$ , are almost on the optimal bounding  
 323 curve. This is unusual for general compression problems, where  
 324 the optimal thresholded sensor falls below the optimal curve.  
 325 Thus, although the picture of multiple noiseless thresholds is  
 326 physically wrong, it does correspond, almost quantitatively,  
 327 to an information theoretic optimization of positional infor-  
 328 mation with the constraint of limited information capacity  
 329  $I(C; g)$  in the sensor. This is important, because it suggests  
 330 that the intuition behind the French flag model or the biologi-  
 331 cal importance of the gap expression boundaries corresponds  
 332 more closely than expected to a true information theoretic  
 333 optimization.

334 We can understand more about the structure of the opti-  
 335 mal mappings  $g \rightarrow C$  by looking at the distributions  $P(C|g)$ ,  
 336 shown in the top panels of Fig 2. These  $P(C|g)$  correspond to  
 337 the three  $I(C; g)$ , marked by the arrows, of the black IB curve,  
 338 where we have used  $\|C\| = 70$  numerically but normalized  
 339 to 1 to emphasize the almost continuous character of  $C$ . At  
 340 small  $I(C; g)$  whole ranges of  $g$  are mapped uniformly into  
 341 ranges of  $C$ , while at larger  $I(C; g)$  we see the emergence of  
 342 a reliably graded mapping, especially in the range bracketing  
 343 half-maximal expression. In all panels, the optimal sensor  
 344 focuses on the low expression levels of Hb, which are biologi-  
 345 cally the most precise expression levels (see Fig 1 A). That the  
 346 optimal sensor resolves these levels more than noisily expressed  
 347 levels in order to receive the most information about the sys-  
 348 tem is expected from intuition for optimal sensor arrangements  
 349 in neurons, in the spirit of Ref. (38).

350 The light crosses in Fig 2 correspond to a greedy, determin-  
 351 istic approximation to the full optimization problem in the way  
 352 of Ref. (39, 40); we provide more details on this calculation in  
 353 the SI. This approximation generates thresholded sensors, but  
 354 as we add more thresholds one cannot go back to readjust the  
 355 existing thresholds. Despite this restriction, the results are  
 356 very close to the true optimum, so that there is a hierarchical  
 357 evolutionary path to nearly optimal performance.

358 A detailed discussion of how the optimal sensor corresponds  
 359 to models of sensing that involve binding site occupation  
 360 (41, 42) would go beyond the scope of this paper. Qualita-  
 361 tively, however, we note that the top panels in Fig 2 could  
 362 be compared such sensors, with the steep change in  $C$  vs  
 363  $g$  corresponding to highly cooperative binding; interestingly  
 364 the predicted degree of cooperativity depends on the sensor  
 365 capacity  $I(C; g)$ .

### 366 Multiple regulatory elements for Hb

367 We see from Fig 2 that capturing all the positional information  
 368 encoded by Hb requires measuring the expression level with  
 369 a sensor capacity of  $I(C; g) \sim 3$  bits of precision. This is  
 370 consistent with our conclusions from the analysis of thresholded  
 371 sensors, where the optimal sensors with 7-9 thresholds also  
 372 have a capacity of  $I(C; g) \sim 3$  bits. We have done the same  
 373 analysis for the other gap TF proteins (*krupel* Kr, *giant* Gt,  
 374 and *knirps* Kni), and also find that  $\sim 3$  bits of capacity is  
 375 required in each case.

376 How does this information capacity compare with the in-

377 formation capacity of biological regulatory elements, such as  
 378 enhancers? Estimates based both on direct measurements  
 379 and on more detailed models indicate that the capacity of a  
 380 regulatory element is in the range of 1–3 bits (2, 43). These es-  
 381 timates depends on the absolute concentrations of the relevant  
 382 molecules, on the time available for reading out the infor-  
 383 mation, the length of the regulatory elements, and on other  
 384 details of the different noise sources in the system (2, 43). At  
 385 one extreme, if the capacity of a biological regulatory element  
 386 is three bits, then a single regulatory element is sufficient to  
 387 capture the full positional information; in this case, it should  
 388 have been possible for the fly’s transcriptional apparatus to  
 389 extract all the available positional information using only one  
 390 regulatory element or enhancer, but this requires that this  
 391 element operates close to the physical limits to information  
 392 capacity. But if the capacity of a single element is only one bit,  
 393 then we need multiple regulatory elements even in response to  
 394 a single transcription factor. It is clear from Fig 2 that there  
 395 is a very big difference between a capacity of 1 bit and 3 bits.

### 396 Optimal sensor for all the gap proteins

397 One might argue that the fly does not need to extract this much  
 398 positional information about cell fates from Hb, as the other  
 399 gap proteins provide information as well. Indeed, we know that  
 400 biologically all four gap TF proteins (Kr, Kni, Gt and Hb), are  
 401 important for nuclei to take their correct cell fates. Practically,  
 402 the temporal changes in the expression patterns could also be  
 403 important (44, 45), but in the first instance we again focus  
 404 on a sensor that measure the expression profiles 40 – 44 min  
 405 into cycle 14, as it has been shown that these are sufficient to  
 406 predict the positions of pair-rule stripes (18). Thus, we need to  
 407 find the optimal sensor for the joint gap expression profiles in  
 408 order to draw biologically relevant lessons from our approach.  
 409 Rather than considering, as above, the mapping  $g_{Hb} \rightarrow C$ , we  
 410 can consider mappings from combinations of expression levels  
 411 of multiple gap TFs (Fig 3A) into  $C$ , corresponding to a single  
 412 optimal sensing element; i.e.  $\{g_i\} \equiv \{g_{Kr}, g_{Kni}, g_{Gt}, g_{Hb}\} \rightarrow C$ .  
 413 The analog of Eq (12) is the optimization problem

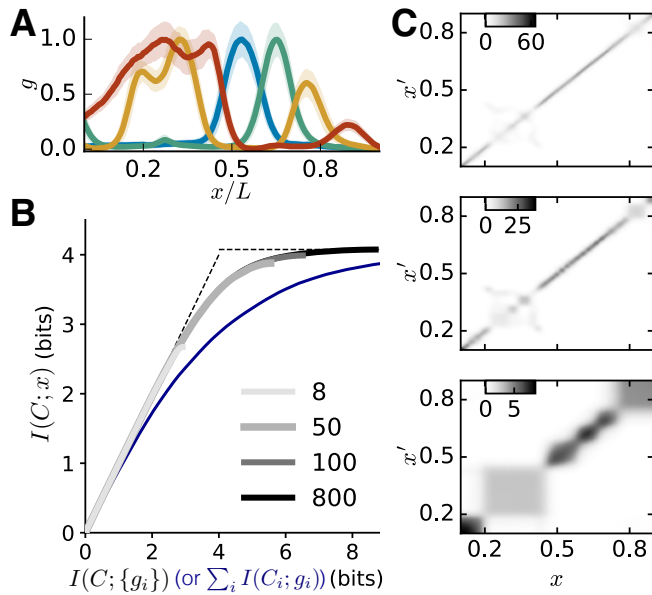
$$314 \max_{P(C|\{g_i\})} [I(C; x) - TI(C; \{g_i\})]. \quad [13]$$

415 We apply the information bottleneck scheme to find the opti-  
 416 mal sensor, and see that we can capture a significant fraction  
 417 of the information provided by all gap TFs by keeping only  
 418 four bits of information about their expression levels, or just  
 419 one bit per gene (Fig 3B), but four bits still captures less than  
 420 90% of the available information.

421 We can visualize what is being gained as the sensor capacity  
 422  $I(C; \{g_i\})$  increases using the decoding maps introduced in  
 423 Ref (18). The decoding map at the top of Fig 3C is the best  
 424 possible decoding map given the expression levels that we  
 425 observe experimentally. The map show the distribution of  
 426 positions  $x'$  consistent with the gap gene expression levels seen  
 427 in nuclei at the true position  $x$ ,

$$428 P(x'|x) = P(x'|\{g_i\}) \Big|_{\{g_i=g_i(x)\}}; \quad [14]$$

429 for simplicity we show this averaged over all the expression  
 430 levels found at  $x$ . Using all the available information,  $P(x'|x)$   
 431 forms a narrow band around  $x' = x$ , with width  $\sigma_x/L \sim 0.01$   
 432 (18). In the lower panels we imagine that inference is based



**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). Mean (solid) and standard deviation (shading) across  $N_{\text{emb}} = 38$  embryos in a five minute window (40–44 min) in nuclear cycle 14 (18). (B) Information bottleneck results, as in Fig 2. Optimal solutions with  $||C|| = 8, 100$  and  $800$  (shades of grey), 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).

expression levels, so that  $g_{\text{Hb}} \rightarrow C_1, g_{\text{Kr}} \rightarrow C_2$ , 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  $P_i(C_i|g_i)$ , and the objective function is

$$\mathcal{F} = I(\{C_i\}; x) - T \sum_{i=1}^4 I(C_i; g_i). \quad [16]$$

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). This 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, 49, 50). 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. In order 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.

**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 National Science Foundation, through the Center for the Physics of Biological Function (PHY-1734030) and the Center for the Science of Information (CCF-0939370); by National Institutes of Health Grant R01GM097275; by the Alexander von Humboldt Stiftung; and by the Howard Hughes Medical Institute.

1. W Bialek and S Setayeshgar, Physical limits to biochemical signaling. *Proc Natl Acad Sci (USA)* **102**, 10040–10045 (2005).
2. G Tkačik, CG Callan Jr, and W Bialek, Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci (USA)* **105**, 12265–12270 (2008).
3. JO Dubuis, R Samanta and T Gregor, Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol Sys Biol* **9**, 639 (2013).

not on the actual expression levels but on the compressed version  $C$ ,

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

as explained in more detail in the SI; 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  $\sigma_x$ ) 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 (18).

To extract all the available positional information requires mechanisms that preserve eight or more bits of information about the combined expression levels of the four gap genes. The greyscale indicates that a  $||C||$  of at least 30–50 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 the joint sensing of four sensors  $C_i$  ( $i \in \{1, 4\}$ ), where each sensor can only respond to a single one of the four gap transcription factors. Mathematically, this corresponds to demanding that the compressed variables be constructed from individual gene

- 520 4. T Gregor, DW Tank, EF Wieschaus, and W Bialek, Probing the limits to positional information. *Cell* **130**, 153–164 (2007). 604
- 521
- 522 5. A Abu-Arish, A Porcher, A Czerwonka, N Dostatni and C Fradin, High mobility of Bicoid 605
- 523 captured by fluorescence correlation spectroscopy: Implication for the rapid establishment of 606
- 524 its gradient. *Biophys J* **99**, L33–35 (2010). 607
- 525 6. JA Drocco, O Grimm, DW Tank and EF Wieschaus, Measurement and perturbation of morphogen 608
- 526 lifetime: Effects on gradient shape. *Biophys J* **101**, 1809–1815 (2011). 609
- 527 7. CE Hannon, SA Blythe and EF Wieschaus, Concentration dependent chromatin states induced 610
- 528 by the bicoid morphogen gradient. *eLife* **6**, e28275 (2017). 611
- 529 8. SE Keenan, SA Blythe, RA Marmion, NJ-V Djabrayan, EF Wieschaus, and SY Shvartsman, 612
- 530 Rapid dynamics of signal-dependent transcriptional repression by Capicua. *Dev Cell* **52**, 613
- 531 794–801 (2020). 614
- 532 9. HC Berg and EM Purcell, Physics of chemoreception. *Biophys J* **20**, 193–219 (1977). 615
- 533 10. K Kaizu, WH de Ronde, J Pajmans, K Takahashi, F Tostevin, and PR ten Wolde, The 616
- 534 Berg–Purcell limit revisited. *Biophys J* **106**, 976–985 (2014). 617
- 535 11. T Friedlander, R Prizak, CC Guet, NH Barton, and G Tkačik, Intrinsic limits to gene regulation 618
- 536 by global crosstalk. *Nat Commun* **7**, 12307 (2016). 619
- 537 12. T Mora and I Nemenman, Physical limit to concentration sensing in a changing environment. 620
- 538 *Phys Rev Lett* **123**, 198101 (2019). 621
- 539 13. C Nüsslein–Vollhard and EF Wieschaus, Mutations affecting segment number and polarity in 622
- 540 *Drosophila*. *Nature* **287**, 795–801 (1980). 623
- 541 14. J Jaeger, The gap gene network. *Cell Mol Life Sci* **68**, 243–274 (2011). 624
- 542 15. L Wolpert, Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* 625
- 543 **25**, 1–47 (1969). 626
- 544 16. JO Dubuis, G Tkačik, EF Wieschaus, T Gregor, and W Bialek, Positional information, in bits. 627
- 545 *Proc Natl Acad Sci (USA)* **110**, 16301–16308 (2013). 628
- 546 17. F Liu, AH Morrison, and T Gregor, Dynamic interpretation of maternal inputs by the *Drosophila* 629
- 547 segmentation gene network. *Proc Natl Acad Sci (USA)* **110**, 6724–6729 (2013). 630
- 548 18. MD Petkova, G Tkačik, W Bialek, EF Wieschaus, and T Gregor, Optimal decoding of cellular 631
- 549 identities in a genetic network. *Cell* **176**, 844–855 (2019). 632
- 550 19. W Bialek, *Biophysics: Searching for Principles*. (Princeton University Press, Princeton, 2012). 633
- 551 20. RN Gutenkunst, JJ Waterfall, FP Casey, KS Brown, CR Myers, and JP Sethna, Universally 634
- 552 sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* **3**, e189 (2007). 635
- 553 21. MK Transtrum, B Machta, K Brown, BC Daniels, CR Myers, and J Sethna, Sloppiness and 636
- 554 emergent theories in physics, biology, and beyond. *J Chem Phys* **143**, 010901 (2015). 637
- 555 22. P François and ED Siggia, Predicting embryonic patterning using mutual entropy fitness and 638
- 556 in silico evolution. *Development* **137**, 2385–2395 (2010). 639
- 557 23. JH Lewis, JMW Slack, and LJ Wolpert, Thresholds in development. *J. Theor. Biol.* **65**, 579–590 640
- 558 (1977). 641
- 559 24. JB Green and JC Smith, Growth factors as morphogens: do gradients and thresholds establish 642
- 560 body plan? *Trends Genet.* **7**, 245 (1991). 643
- 561 25. JB Green, HV New, and JC Smith, Responses of embryonic *Xenopus* cells to activin and FGF 644
- 562 are separated by multiple dose thresholds and correspond to distinct axes of the mesoderm. 645
- 563 *Cell* **71**, 731 (1992). 646
- 564 26. KW Rogers, and AF Schier, Morphogen gradients: from generation to interpretation. *Annu.* 647
- 565 *Rev. Cell Dev. Biol.* **27**, 377–407 (2011). 648
- 566 27. J Briscoe and S Small, Morphogen rules: Design principles of gradient-mediated embryo 649
- 567 patterning. *Development* **142**, 3996–4009 (2015). 650
- 568 28. L Bintu, NE Buchler, HG Garcia, U Gerland, T Hwa, J Kondev, and R Phillips, Transcriptional 651
- 569 regulation by the numbers: models. *Curr Opin Genet Dev* **15**, 116–124 (2005). 652
- 570 29. E Segal, T Raveh-Sadka, M Schroeder, U Unnerstall, and U Gaul, Predicting expression 653
- 571 patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**, 535–540 (2008). 654
- 572 30. G Tkačik, AM Walczak, and W Bialek, Optimizing information flow in small genetic networks. 655
- 573 *Phys Rev E* **80**, 031920 (2009); arXiv:0903.4491 [q-bio.MN] (2009). 656
- 574 31. O Pulkkinen and R Metzler, Distance matters: The impact of gene proximity in bacterial gene 657
- 575 regulation. *Phys Rev Lett* **110**, 198101 (2013). 658
- 576 32. EE Furlong and M. Levine, Developmental enhancers and chromosome topology. *Science* 659
- 577 **361**, 1341–1345 (2018). 660
- 578 33. BR Sabari et al, Coactivator condensation at super-enhancers links phase separation and 661
- 579 gene control. *Science* **361** eaar3958 (2018). 662
- 580 34. N Tishby, FC Pereira, and W Bialek, The information bottleneck method. In *Proceedings of the* 663
- 581 *37th Annual Allerton Conference on Communication, Control and Computing*, B Hajek and RS 664
- 582 Sreenivas, eds, pp 368–377 (University of Illinois, 1999); arXiv:physics/0004057 (2000). 665
- 583 35. R Shwartz-Ziv and N Tishby, Opening the Black Box of Deep Neural Networks via Information, 666
- 584 arXiv:1703.00810 (2017). 667
- 585 36. AA Alemi, I Fischer, JV Dillon, and K Murphy, Deep variational information bottleneck. 668
- 586 arXiv:1612.00410 (2016). 669
- 587 37. S Palmer, O Marre, MJ Berry II, and W Bialek, Predictive information in a sensory population. 670
- 588 *Proc Natl Acad Sci (USA)* **112**, 6911 (2015). 671
- 589 38. S Laughlin, A simple coding procedure enhances a neuron's information capacity. *Z. Natur-* 672
- 590 *forsch.* **36**, 910–912 (1981). 673
- 591 39. N Slonim and N Tishby, Agglomerative information bottleneck. In *Advances in Neural Informa-* 674
- 592 *tion Processing 12*, S Solla, T Leen, and K Müller, eds, pp 617–623 (MIT Press, Cambridge, 675
- 593 2000). 676
- 594 40. DJ Strouse and DJ Schwab, The deterministic information bottleneck. *Neural Comp* **29**, 677
- 595 1611–1630 (2017). 678
- 596 41. J Estrada, F Wong, A DePace, and J Gunawardena, Information integration and energy 679
- 597 expenditure in gene regulation. *Cell* **166**, 234–244 (2016). 680
- 598 42. SE Marzen and JP Crutchfield, Prediction and Dissipation in Nonequilibrium Molecular Sen- 681
- 599 sors: Conditionally Markovian Channels Driven by Memoryful Environments, *Bull Math Biol* 682
- 600 **82**, 1 (2020). 683
- 601 43. G Tkačik, CG Callan Jr, and W Bialek, Information capacity of genetic regulatory elements. 684
- 602 *Phys Rev E* **78**, 011910 (2008). 685
- 603 44. O Wartlick, P Mumcu, A Kicheva, T Bittig, C Seum, F Jülicher, and M Gonzalez–Gaitan, 686
- Dynamics of Dpp signaling and proliferation control. *Science*, **331**, 1154 (2011). 687
45. NB Becker, A Mugler, PR ten Wolde, Prediction and dissipation in biochemical sensing. *Phys.* 688
- Rev. Lett.* **115**, 258103 (2015). 689
46. C Schulz and D Tautz, Autonomous concentration-dependent activation and repression of 690
- Krüppel* by *hunchback* in the *Drosophila* embryo. *Development* **120**, 3043–3049 (1994). 691
47. L Descheemaeker, E Peters, and S de Buyl, Non-monotonic auto-regulation in single gene 692
- circuits. *PLoS ONE* **14**, e0216089 (2019). 693
48. MV Staller, BJ Vincent, MDJ Bragdon, T Lydiard-Martin, Z Wunderlich, J Estrada, and AH 694
- DePace, Shadow enhancers enable *Hunchback* bifunctionality in the *Drosophila* embryo. *Proc* 695
- Natl Acad Sci (USA)* **112**, 785–790 (2015). 696
49. JO Yáñez-Cuna, EZ Kwon, and A Stark, Deciphering the transcriptional cis-regulatory code, 697
- Trends Genet.* **29**, 11–22 (2013). 698
50. J Crocker, GR Ilsley, DL Stern, Quantitatively predictable control of *Drosophila* transcriptional 699
- enhancers in vivo with engineered transcription factors *Nat Genet.* **48**, 292–298 (2016). 700