

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,^{1-4,*} Mariela D. Petkova,⁵ Thomas Gregor,^{1,2,6} Eric F. Wieschaus,²⁻⁴ and William Bialek^{1,2,7,*}

¹ Joseph Henry Laboratories of Physics, ²Lewis–Sigler Institute for Integrative Genomics, ³Department of Molecular Biology, and ⁴Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544 USA; ⁵Program in Biophysics, Harvard University, Cambridge, MA 02138 USA; ⁶Department of Developmental and Stem Cell Biology, UMR3738, Institut Pasteur, 75015 Paris, France; ⁷Initiative 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 2 molecules. In order to extract this information the cell must effectively 3 "measure" these concentrations, but there are physical limits to the precision of these measurements. We use the gap gene network 5 in the early fly embryo as an example of the tradeoff between the 6 precision of concentration measurements and the transmission of 7 relevant information. For thresholded measurements we find that 8 lower thresholds are more important, and fine tuning is not required 9 for near-optimal information transmission. We then consider general 10 sensors, constrained only by a limit on their information capacity, and 11 find that thresholded sensors can approach true information theoretic 12 optima. The information theoretic approach allows us to identify the 13 optimal sensor for the entire gap gene network, and to argue that the 14 physical limitations of sensing necessitate the observed multiplicity 15 of enhancer elements, with sensitivities to combinations rather than 16 single transcription factors. 17

Sensing | gene regulation | development | information bottleneck

ells control the concentrations of proteins in part by • controlling the transcription of corresponding genes into 2 messenger RNA. This control is effected by the binding of 3 transcription factor (TF) proteins to specific sites along the 4 genome. Transcription factors can thus regulate the synthesis 5 of other TFs, forming a genetic network. Regulatory mech-6 anisms internal to the network must be precise enough to generate reliable relationships between the concentration of 8 input signals and the levels of gene expression downstream. What must the cell do in order to extract and make efficient 10 use of the information provided by variations in TF concen-11 trations? 12

We usually think of transcription factors as controlling the 13 level of gene expression, but we can also view the expression 14 level as being the cell's measurement of the TF concentration 15 (1, 2). As outside observers of the cell, we can measure the con-16 centration of transcription factors with considerable accuracy 17 (3). However the cell's "measurement" of TF concentration 18 is based on the arrival of these molecules at their binding 19 sites, and this is a noisy process, because TF concentrations 20 are low, in the nanoMolar range (4-8). Physical limits to the 21 measurement of such low concentrations were first explored 22 in the context of bacterial chemotaxis (9), but have proven 23 to be much more general (1, 10-12). What will be important 24 for our discussion is not the precise values of these limits, but 25 rather that the limits exist and are significant on the scale of 26 biological function. 27

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 33 stripes are positioned with an accuracy of $\pm 1\%$ along the long 34 (anterior-posterior) axis of the embryo, and this is the accuracy 35 of subsequent developmental events such as the formation of 36 the cephalic furrow (16, 17). The local concentrations of the 37 gap proteins provide just enough information to support this 38 level of precision (16). The algorithm that achieves optimal 39 readout of this positional information predicts, quantitatively, 40 the distortions of the pair-rule stripes in mutant flies where 41 individual maternal inputs are deleted (18). 42

The gap gene network offers us the chance to ask how ac-43 curately the transcription factor concentrations need to be 44 measured and to infer features of the regulatory architecture 45 responsible for these measurements. The information that the 46 gap genes convey about position along the anterior-posterior 47 axis is what allows nuclei to make distinct cell fate decisions 48 required for development; we investigate here how this can be 49 seen as a sensing or signal processing problem (Fig 1A). We 50 start with a more traditional view of how information is repre-51 sented in the concentration of a single TF, through thresholds 52 or expression domains, and then argue for a more abstract 53 formulation of the problem as selective data compression. In 54 this abstract view, aspects of the transcriptional regulatory 55 mechanisms can be seen as solutions to an information theo-56 retic optimization problem. We apply this approach to analyze 57 the information conveyed by the concentrations of all four gap 58 proteins, and find that some of the complexities in how these 59 molecules function as transcription factors emerge naturally 60 from solutions to the relevant optimization problem. 61

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

62

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 $i = 1, \ldots, K$ 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 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 one standard error in our estimates.

67 quantitative version of this idea is that subsequent events are

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

In Figure 1B-E we use the gap protein hunchback (Hb) to 87 illustrate the information loss associated with thresholding. 88 89 At each point x there is an expression level q (Fig. 1B), drawn from a probability distribution P(q|x); looking at many em-90 bryos we have samples out of this distribution. Experimental 91 data are from Ref (18), where immunostaining was used to 92 obtain expression profiles of the gap proteins. We focus on a 93 time window of 40 - 44 min into nuclear cycle 14, the final 94 cycle before blastoderm stage, during which the gap gene ex-95 pression determine crucially the cell fates of nuclei along the 96 embryo's anterior-posterior axis through pair-rule, segment 97

polarity, and hox gene expression.

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

$$\sigma = H(g - \theta), \qquad [1] \quad {}_{102}$$

98

99

100

101

where H is the Heaviside step function, H(y > 0) = 1 and H(y < 0) = 0. Then we can estimate the θ (threshold)- 104 dependent distribution $P_{\theta}(\sigma|x)$, 105

$$P_{ heta}(\sigma = 1|x) = \int dg \, H(g - \theta) P(g|x)$$
 [2] 100

$$P_{\theta}(\sigma = 0|x) = 1 - P_{\theta}(\sigma = 1|x).$$
 [3] 107

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, 111

$$I_{\theta}(\sigma; x) = \sum_{\sigma} \int dx P(x) P_{\theta}(\sigma|x) \log_2 \left[\frac{P_{\theta}(\sigma|x)}{P_{\theta}(\sigma)} \right] \text{ bits, } [4] \quad \text{112}$$

where P(x) = 1/L, as a priori all positions along the length of the embryo are equally likely, and 114

$$P_{\theta}(\sigma) = \int dx P(x) P_{\theta}(\sigma|x).$$
 [5] 115

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 (see Fig. 1 C).

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

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

[7]

 2.09 ± 0.06 bits.

128

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

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

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

149
$$\sigma_1 = H(g -$$

150

$$\sigma_1 = H(g - \theta_1)$$

$$\sigma_2 = H(g - \theta_2).$$
[8]
[9]

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

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

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

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

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

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

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

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

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

Beyond thresholds

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

218

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

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

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

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

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

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

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

275 Information bottleneck and the optimal sensor

We now want to find the mapping $g \to C$ which 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,

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

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



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 greyshades 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 1E, and match with the $T \to 0$ limit of the bottleneck solutions with fixed ||C||. The light points are from an explicitly deterministic formulation of the bottleneck problem (39). Upper panel shows snapshots probability distributions P(C|g) at different information capacities I(C; g) along the bottleneck curve; intermediate levels of $g \in [0.05, 0.8]$ are progressively better resolved as the capacity increases.

optimizing P(C|q) is known as the "information bottleneck" 289 problem (34). Its solution gives an iterative algorithm which 290 finds P(C|g). This problem and algorithm have implications 291 for machine learning (35, 36) or finding efficient encodings 292 in neuronal systems (37); in these fields, the optimal P(C|g)293 is often described as a compression of g. Qualitatively, the 294 algorithm identifies (potentially noisy) sets of values of q that 295 are most informative about x, and focuses P(C|q) to make 296 maximum use of those values. 297

We solve the optimization problem in Eq (12) numerically, 298 considering C to be a variable with a discrete set of values or 299 states and varying the number of these states, ||C||. At fixed 300 ||C||, decreasing T allows I(C; g) to be larger, and pushes the 301 noisy mapping P(C|g) toward being deterministic. Results of 302 the bottleneck analysis for Hb are shown in Fig 2 as trajectories 303 (solid grey and black lines) in the plane I(C; x) vs I(C; g). Only 304 the region below the dashed diagonal and horizontal lines is 305 theoretically accessible due to the data processing inequality 306 $[I(C;x) \leq I(C;g)$ and $I(C;x) \leq I(g;x)$, which implies that 307 even an optimal sensor cannot know more about positional 308 information or nuclear cell fates than is provided by the protein 309 expression itself. Often, for example for neuronal systems (37), 310 the bounding curve for the optimal sensor at large ||C|| (solid 311 black line) is further away from the data processing bound 312 than here. This optimal bounding curve that emerges from 313 the information bottleneck analysis separates the plane into a 314 physically possible region (below the curve) and an impossible 315 region (above the curve). As I(C; q) becomes large, the curve 316

265

³¹⁷ plateaus at the available positional information I(g; x).

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

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

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

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

366 Multiple regulatory elements for Hb

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

376 How does this information capacity compare with the in-

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

Optimal sensor for all the gap proteins

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

$$\max_{P(C|\{g_i\})} \left[I(C;x) - TI(C;\{g_i\}) \right].$$
 [13] 414

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 four bits of information about their expression levels, or just one bit per gene (Fig 3B), but four bits still captures 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 (18). The decoding map at the top of Fig 3C is the best possible decoding map given the expression levels that we observe experimentally. The map show the distribution of positions x' consistent with the gap gene expression levels seen in nuclei at the true position x, 421

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

1

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



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 $N_{\rm em} = 38$ 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).

⁴³³ not on the actual expression levels but on the compressed ⁴³⁴ version C,

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

as explained in more detail in the SI; we do this for the 436 optimal compressions with $I(C; \{g_i\}) = 2$, and 4 bits. We see 437 that as the compression becomes more severe, the inference 438 becomes more uncertain (larger σ_x) and genuinely ambiguous. 439 This more noisy inference has biological consequences: sensors 440 with capacity of much less than 4 bits do not capture enough 441 information to predict the patterns of pair-rule expression 442 stripes in mutants, following the analysis in Ref (18). 443

To extract all the available positional information requires 444 445 mechanisms that preserve eight or more bits of information 446 about the combined expression levels of the four gap genes. The greyscale indicates that a ||C|| of at least 30–50 levels 447 would be required. Again, if we think that a single sensor 448 can implement one threshold, this means that more than one 449 sensor would be required, even in the best possible case where 450 information-theoretically optimal sensing is possible. We know 451 that there are several dozen enhancer sites that respond to 452 453 the gap gene TFs, and we see that this degree of complexity may be required by information theoretic constraints, even if 454 these sensors make optimal use of the available information. 455

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 expression levels, so that $g_{\rm Hb} \rightarrow C_1$, $g_{\rm Kr} \rightarrow C_2$, 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

$$\mathcal{F} = I(\{C_i\}; x) - T \sum_{i=1}^{4} I(C_i; g_i).$$
[16] 467

482

We find that such a set of four sensors always is substantially 468 worse than a single optimal sensor, as indicated by the blue 469 line in Fig 3, even with same total information capacity (for 470 more details see SI). This indicates the importance of having 471 regulatory mechanisms that are sensitive to combinations 472 of transcription factors. In fact, the readout of positional 473 information encoded in the gap proteins is implemented by 474 the array of enhancers controlling pair rule gene expression, 475 and these enhancers are prototypical instances of regulatory 476 elements that respond to combinations of transcription factors 477 (29, 32, 49, 50). While there is some distance between our 478 abstract formulation and the molecular details, it is attractive 479 to see that this mechanistic complexity is required as a response 480 to basic physical and information theoretic limitations. 481

Conclusion

To summarize, individual regulatory mechanisms have limited 483 information capacity, and our central result is that this ca-484 pacity in turn sets strict limits on the amount of positional 485 information that can be extracted from the gap gene expres-486 sion levels. In this paper, we see the measurement of the 487 transcription factors as a problem of efficient sensing or com-488 pression, and use the information bottleneck algorithm to 489 identify an optimal sensor for this network. Precise compari-490 son with ideas about thresholded reading of the gap TF Hb 491 shows that the thresholds do not need to be fine tuned and 492 exhibit a hierarchy of sensitivities. Crucially, we find that it 493 almost certainly is not possible to read out enough positional 494 information with a single enhancer element. In order for the 495 nuclei to obtain at least 90% of the information provided by 496 the gap TF network, a large number of thresholds (30-50) or a 497 high capacity in the optimal sensor is required, and this must 498 be realized by multiple enhancers. Further, if each enhancer 499 responds to a single TF, there is a dramatic loss of efficiency. 500 The information theoretic optimization principle we have ex-501 plored here thus predicts that expression levels must be read 502 by multiple enhancers, each sensitive to combinations of the 503 gap TFs. This complex enhancer logic indeed is how gap gene 504 expression levels drive downstream events in the fly embryo. 505

ACKNOWLEDGMENTS. We thank P-T Chen, M Levo, R Munshi, 506 B van Opheusden, and R Rao for helpful discussions. This work was 507 supported in part by the US National Science Foundation, through 508 the Center for the Physics of Biological Function (PHY-1734030) 509 and the Center for the Science of Information (CCF-0939370); 510 by National Institutes of Health Grant R01GM097275; by the 511 Alexander von Humboldt Stiftung; and by the Howard Hughes 512 Medical Institute. 513

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

435

- T Gregor, DW Tank, EF Wieschaus, and W Bialek, Probing the limits to positional information.
 Cell 130, 153–164 (2007).
- A Abu-Arish, A Porcher, A Czerwonka, N Dostatni and C Fradin, High mobility of Bicoid captured by fluorescence correlation spectroscopy: Implication for the rapid establishment of its gradient. *Biophys J* 99, L33–35 (2010).
 - JA Drocco, O Grimm, DW Tank and EF Wieschaus, Measurement and perturbation of morphogen lifetime: Effects on gradient shape. *Biophys J* 101, 1809–1815 (2011).
 - CE Hannon, SA Blythe and EF Wieschaus, Concentration dependent chromatin states induced by the bicoid morphogen gradient. *eLife* 6, e28275 (2017).
- SE Keenan, SA Blythe, RA Marmion, NJ-V Djabrayan, EF Wieschaus, and SY Shvartsman,
 Rapid dynamics of signal-dependent transcriptional repression by Capicua. *Dev Cell* 52,
 794–801 (2020).
- 9. HC Berg and EM Purcell, Physics of chemoreception. *Biophys J* 20, 193–219 (1977).
- K Kaizu, WH de Ronde, J Paijmans, K Takahashi, F Tostevin, and PR ten Wolde, The
 Berg–Purcell limit revisited. *Biophys J* 106, 976–985 (2014).
- T Friedlander, R Prizak, CC Guet, NH Barton, and G Tkačik, Intrinsic limits to gene regulation
 by global crosstalk. *Nat Commun* 7, 12307 (2016).
- T Mora and I Nemenman, Physical limit to concentration sensing in a changing environment.
 Phys Rev Lett 123, 198101 (2019).
- C Nüsslein–Vollhard and EF Wieschaus, Mutations affecting segment number and polarity in Drosophila. Nature 287, 795–801 (1980).
- 541 14. J Jaeger, The gap gene network. Cell Mol Life Sci 68, 243–274 (2011).

525

526

527

528

- L Wolpert, Positional information and the spatial pattern of cellular differentiation. J Theor Biol 25, 1–47 (1969).
- JO Dubuis, G Tkačik, EF Wieschaus, T Gregor, and W Bialek, Positional information, in bits.
 Proc Natl Acad Sci (USA) 110, 16301–16308 (2013).
- F Liu, AH Morrison, and T Gregor, Dynamic interpretation of maternal inputs by the Drosophila
 segmentation gene network. *Proc Natl Acad Sci (USA)* 110, 6724–6729 (2013).
- MD Petkova, G Tkačik, W Bialek, EF Wieschaus, and T Gregor, Optimal decoding of cellular identities in a genetic network. *Cell* **176**, 844–855 (2019).
- 550 19. W Bialek, Biophysics: Searching for Principles. (Princeton University Press, Princeton, 2012).
- RN Gutenkunst, JJ Waterfall, FP Casey, KS Brown, CR Myers, and JP Sethna, Universally
 sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* 3, e189 (2007).
- MK Transtrum, B Machta, K Brown, BC Daniels, CR Myers, and J Sethna, Sloppiness and emergent theories in physics, biology, and beyond. J Chem Phys 143, 010901 (2015).
- P François and ED Siggia, Predicting embryonic patterning using mutual entropy fitness and in silico evolution. *Development* 137, 2385–2395 (2010).
- JH Lewis, JMW Slack, and LJ Wolpert, Thresholds in development. J. Theor. Biol. 65, 579-590 (1977).
- JB Green and JC Smith, Growth factors as morphogens: do gradients and thresholds establish
 body plan? *Trends Genet.* 7, 245 (1991).
- JB Green, HV New, and JC 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 (1992).
- KW Rogers, and AF Schier, Morphogen gradients: from generation to interpretation. Annu. Rev. Cell Dev. Biol. 27, 377-407 (2011).
- J Briscoe and S Small, Morphogen rules: Design principles of gradient-mediated embryo patterning. *Development* 142, 3996–4009 (2015).
- L Bintu, NE Buchler, HG Garcia, U Gerland, T Hwa, J Kondev, and R Phillips, Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev* 15, 116–124 (2005).
- E Segal, T Raveh-Sadka, M Schroeder, U Unnerstall, and U Gaul, Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* 451, 535–540 (2008).
- G Tkačik, AM Walczak, and W Bialek, Optimizing information flow in small genetic networks.
 Phys Rev E 80, 031920 (2009); arXiv:0903.4491 [q–bio.MN] (2009).
- O Pulkkinen and R Metzler, Distance matters: The impact of gene proximity in bacterial gene regulation. *Phys Rev Lett* **110**, 198101 (2013).
- EE Furlong and M. Levine, Developmental enhancers and chromosome topology. Science
 361, 1341–1345 (2018).
- BR Sabari et al, Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361 eaar3958 (2018).
- N Tishby, FC Pereira, and W Bialek, The information bottleneck method. In *Proceedings of the* 37th Annual Allerton Conference on Communication, Control and Computing, B Hajek and RS Sreenivas, eds, pp 368–377 (University of Illinois, 1999); arXiv:physics/0004057 (2000).
- R Shwartz-Ziv and N Tishby, Opening the Black Box of Deep Neural Networks via Information, arXiv:1703.00810 (2017).
- AA Alemi, I Fischer, JV Dillon, and K Murphy. Deep variational information bottleneck.
 arXiv:1612.00410 (2016).
- S Palmer, O Marre, MJ Berry II, and W Bialek, Predictive information in a sensory population.
 Proc Natl Acad Sci (USA) 112, 6911 (2015).
- S Laughlin, A simple coding procedure enhances a neuron's information capacity. Z. Naturforsch. 36, 910-912 (1981).
- N Slonim and N Tishby, Agglomerative information bottleneck. In Advances in Neural Information Processing 12, S Solla, T Leen, and K Müller, eds, pp 617–623 (MIT Press, Cambridge, 2000).
- 594
 40. DJ Strouse and DJ Schwab, The deterministic information bottleneck. *Neural Comp* 29, 1611–1630 (2017).
- J Estrada, F Wong, A DePace, and J Gunawardena. Information integration and energy expenditure in gene regulation. *Cell* 166, 234-244 (2016).
- SE Marzen and JP Crutchfield, Prediction and Dissipation in Nonequilibrium Molecular Sensors: Conditionally Markovian Chan-nels Driven by Memoryful Environments, *Bull Math Biol* 82, 1 (2020).
- G Tkačik, CG Callan Jr, and W Bialek, Information capacity of genetic regulatory elements.
 Phys Rev E 78, 011910 (2008).
- 44. O Wartlick, P Mumcu, A Kicheva, T Bittig, C Seum, F Jülicher, and M Gonzalez-Gaitan,

- Dynamics of Dpp signaling and proliferation control. Science, 331, 1154 (2011).
 604

 NB Becker, A Mugler, PR ten Wolde, Prediction and dissipation in biochemical sensing. Phys.
 605

 Rev. Lett. 115,, 258103 (2015).
 606
- C Schulz and D Tautz, Autonomous concentration-dependent activation and repression of Krüppel by hunchback in the Drosophila embryo. Development 120, 3043–3049 (1994).

45.

- 47. L Descheemaeker, E Peters, and S de Buyl, Non-monotonic auto-regulation in single gene circuits. *PLoS ONE* 14, e0216089 (2019).
- MV Staller, BJ Vincent, MDJ Bragdon, T Lydiard-Martin, Z Wunderlich, J Estrada, and AH DePace, Shadow enhancers enable Hunchback bifunctionality in the *Drosophila* embryo. *Proc Natl Acad Sci (USA)* **112**, 785–790 (2015).
- JO Yáñez-Cuna, EZ Kvon, and A Stark, Deciphering the transcriptional cis-regulatory code, *Trends Genet*, 29, 11–22 (2013).
- J Crocker, GR Ilsley, DL Stern, Quantitatively predictable control of Drosophila transcriptional enhancers in vivo with engineered transcription factors Nat Genet, 48, 292-298 (2016).

607