Finding the Last Bits of Positional Information

Lauren McGough [1,2] Helena Casademunt,[3] Miloš Nikolić [1] Zoe Aridor,[1]
Mariela D. Petkova,[4] Thomas Gregor [1,5] and William Bialek [1]

[1] *Joseph Henry Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics,*
Princeton University, Princeton, New Jersey 08544, USA
[2] *Department of Ecology and Evolution, The University of Chicago, Chicago, Illinois 60637, USA*
[3] *Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[4] *Program in Biophysics, Harvard University, Cambridge, Massachusetts 02138, USA*
[5] *Department of Developmental and Stem Cell Biology, CNRS UMR3738 Paris Cité, Institut Pasteur, Paris, France*

In a developing embryo, information about the position of cells is encoded in the concentrations of morphogen molecules. In the fruit fly, the local concentrations of just a handful of proteins encoded by the gap genes are sufficient to specify position with a precision comparable to the spacing between cells along the anterior-posterior axis. This matches the precision of downstream events such as the striped patterns of expression in the pair-rule genes, but is not quite sufficient to define unique identities for individual cells. We demonstrate theoretically that this information gap can be bridged if positional errors are spatially correlated, with correlation lengths approximately 20% of the embryo length. We then show experimentally that these correlations are present, with the required strength, in the fluctuating positions of the pair-rule stripes, and this can be traced back to the gap genes. Taking account of these correlations, the available information matches the information needed for unique cellular specification, within error bars of approximately 2%. These observation support a precisionist view of information flow through the underlying genetic networks, in which accurate signals are available from the start and preserved as they are transformed into the final spatial patterns.

I. INTRODUCTION

During the development of an embryo, cell fates are determined in part by the concentrations of specific morphogen molecules that carry information about position [1–3]. For the early stages of fruit fly development, all of these molecules have been identified [4–6]. For patterning along the main body axis, spanning from anterior to posterior (AP), information flows from primary maternal morphogens to an interacting network of gap genes to the pair-rule genes [7,8], whose striped patterns of expression provide a precursor of the segmented body plan in the fully developed organism, visible within three hours after the egg is laid (Fig. 1). It has been known for some time that, at this stage in development, essentially every cell "knows" its fate [11], so it is natural to ask how this information is encoded, quantitatively, in the concentrations of the relevant morphogens.

Expression levels of the gap genes provide enough information to specify the positions of individual cells with an accuracy approximately 1% of the embryo's length [12]. This matches the precision with which the stripes of pair-rule expression are positioned and the precision of macroscopic developmental events such as the formation of the cephalic furrow [13]. Further, the algorithm that extracts optimal estimates of position from the expression levels of the gap genes also predicts, quantitatively, distortions of the striped pattern in mutant flies with deletions of the maternal inputs [10]. At the moment when pair-rule stripes are fully formed, just before gastrulation, there are fewer than 100 rows of cells along the length of the embryo, so it is tempting to think that positional signals with 1% accuracy define unique cellular identities. In fact, this is not quite correct [12]: If each cell makes independent positional errors drawn from a Gaussian distribution, then there is a small but significant probability that neighboring cells will get "crossed signals," driving errors in cell fate determination.

The small difference between 1% positional errors and unique cellular identities provides a test case in the search for a more quantitative understanding of living systems. In physics, we are used to the idea that small quantitative discrepancies can be signs of qualitatively new ideas or mechanisms. In complex biological systems, however, one might worry that small discrepancies reflect experimental errors or oversimplifications in interpretation. If correct, these concerns would limit our ambitions for quantitative theory in the physics tradition.

Here we revisit the small quantitative discrepancy in positional information in light of dramatic improvements in experimental precision [14–16]. On the theoretical side, we clarify the problem, defining an information gap, and show that this gap can be closed if errors in the positional signals are spatially correlated over relatively long distances. Early work by Lott *et al.* [17] detected such correlations in mRNA levels of gap and pair-rule genes; subsequent work found that noise
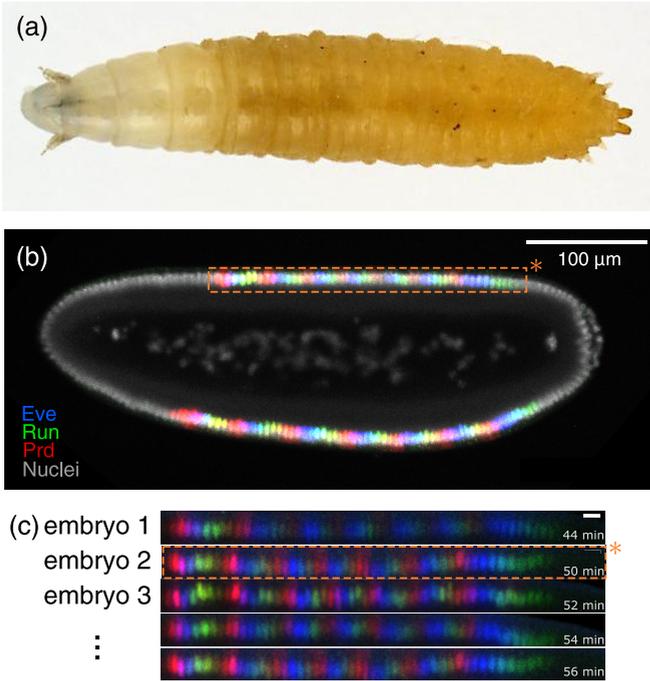
FIG. 1. Segmented *Drosophila* body plan. (a) Bright-field color image of a 5-mm-long third instar larva of the fruit fly *Drosophila melanogaster* [9] with clearly visible segments. (b) Optical section through an embryo stained for three of the pair-rule proteins, 50 min into nuclear cycle 14 (approximately 3 h after oviposition), showing striped patterns that align with the body segments (data from Ref. [10]). (c) Same as in (b), from multiple embryos, illustrating the pattern reproducibility. Time in nuclear cycle 14 is indicated at the bottom right of each profile. The asterisk marks the image in (b).

in different combinations of protein levels in the gap-gene network are correlated significantly over the entire length of the embryo [18]. On the experimental side, we reexamine these correlations, measuring the positions of stripes in the concentrations of pair-rule proteins. We find that the extent of these correlations is what is needed to close the information gap between positional errors and unique cellular identities, quantitatively.

II. DEFINING THE PROBLEM

In the early fly embryo, cells have access to the concentrations of morphogens, and these concentrations are continuously graded. From these concentrations, it is possible to decode an estimate of position, which we label as $\hat{x}_n$ in cell n [10]. We expect that these estimates are correct on average, so $\langle \hat{x}_n \rangle = nL/N$, where there are N cells along the length L of the embryo.[1] However the signals are noisy, so decoding in

---

[1]For simplicity, we imagine that the problem is one dimensional so that cells need to know their position along only one axis. In the early fly embryo, patterning signals along the two major axes are largely independent [19,20], justifying this simplification.

one cell will have errors,

$$\hat{x}_n = nL/N + \delta x_n, \tag{1}$$

$$\langle (\delta x_n)^2 \rangle = \sigma_x^2. \tag{2}$$

For simplicity, but guided by the experimental observations [10,12,21], we assume that σ_x is the same for all cells and that the distribution of δx_n is Gaussian (Appendix A). Here we are interested in the question of whether cells get signals that define the correct ordering along the axis so that $\hat{x}_{n+1} > \hat{x}_n$ for all cells or whether they can get crossed signals such that $\hat{x}_{n+1} < \hat{x}_n$.

If we look at two neighboring cells, then the probability of incorrect ordering is

$$P_{\text{cross}} \equiv \Pr(\hat{x}_{n+1} < \hat{x}_n). \tag{3}$$

To find the probability of a wrong ordering we can take a look at the distribution of the distance to the next cell $y = \hat{x}_{n+1} - \hat{x}_n$. However, since both $\hat{x}_{n+1}$ and $\hat{x}_n$ are Gaussian, their difference y is also Gaussian, with mean equal to $\langle y \rangle = L/N$. If the noise is independent in each cell, then the variance of this difference signal will be $\langle (\delta y)^2 \rangle = 2\sigma_x^2$. Incorrect ordering happens when $y < 0$, which then has probability

$$P_{\text{cross}} = \int_{-\infty}^{0} \frac{dy}{\sqrt{4\pi \sigma_x^2}} e^{-(y-L/N)^2/4\sigma_x^2} \tag{4}$$

$$= \frac{1}{\sqrt{4\pi}} \int_{1/z}^{\infty} dx \, e^{-x^2/4}, \tag{5}$$

with $z = \sigma_x(N/L)$, as shown in Fig. 2. If positional errors are comparable to the spacing between cells, $\sigma_x \sim L/N$, the probability of an error is nearly 24%.

To make more quantitative statements we need a precise estimate of the number of cells N. Observations on the spacing between nuclei, or their areal density, reach back 40 years [22]. Recent measurements are broadly consistent, but with substantial variations [23,24]; it is not clear whether variations in density are correlated with variations in embryo length to result in more reproducible values of N. As explained in Appendix B, we have used images such as those in Fig. 1(b) to count the number of nuclei in the central 80% of the embryo, along the relatively straight dorsal side; the standard deviation across embryos is less than 5%. Assuming that the same density continues to the ends of the embryo, we have $N = 90 \pm 4$, which means that positional error is slightly less than the spacing between cells $\sigma_x \sim 0.9(L/N)$. Figure 2 then predicts that neighboring cells will cross signals with approximately 20% chance, and if the signals are independent the probability that all N come in the right order is vanishingly small.

This failure to specify unique cellular identities can be given a simple information-theoretic interpretation. To specify one cell uniquely out of N requires $I_{\text{unique}} = \log_2 N$ bits of information [25,26]. On the other hand, if we have signals that represent a continuous position x drawn uniformly from the range $0 < x \leqslant L$ and these signals have Gaussian noise with (small) standard deviation σ_x, as described above, then the amount of information the signal conveys about position is

$$I_{\text{position}} = \log_2 L - \log_2(\sqrt{2\pi e}\sigma_x), \tag{6}$$
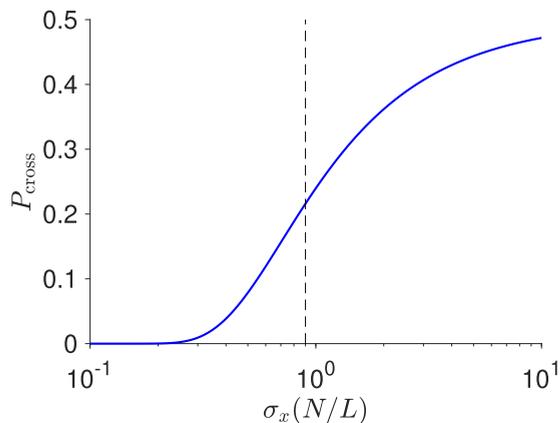
FIG. 2. Probability of crossed signals between two neighboring cells as a function of the positional error, assuming that noise is independent in each cell, from Eq. (5). Dashed vertical line marks the experimental value of positional noise $\sigma_x \sim 0.01L$, which corresponds to less than the mean distance between neighboring cells L/N [12].

where the first term is the entropy of the uniform distribution of positions and the second term is the entropy of the Gaussian noise distribution [26]. Combining these, we can define an information gap

$$I_{\text{gap}} \equiv I_{\text{unique}} - I_{\text{position}} = \log_2\left(\frac{N\sigma_x}{L}\sqrt{2\pi e}\right). \quad (7)$$

As discussed below, we obtain a more accurate estimate of the information gap by averaging over measurements of σ_x at multiple points along the embryo, defined by the pair-rule stripes, and we find $I_{\text{gap}} = 1.68 \pm 0.07$ bits (Appendix A). Importantly, this gap is measured per cell: It is not that the embryo is missing approximately 1.7 bits of information, but rather that *every cell* is missing this information.

III. EXTRA INFORMATION FROM CORRELATIONS

A. Theory

To address the information gap directly, we leverage the concept that correlated noise can facilitate enhanced information transmission. Correlated noise typically is viewed as challenging because it resists being averaged away. However, in the context of neighboring cells making errors in position, correlations mitigate the probability of receiving crossed signals as previously defined. Here we develop these considerations more formally.

Information is roughly the difference in entropy between the signal and the noise, where entropy measures the (logarithmic) volume in phase space that is occupied by a set of points. When random variables become correlated, the volume and hence the entropy is reduced, even if the variances of the individual variables are unchanged. In our example, with correlations, the full pattern of points $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\}$ fills a smaller volume in the space $[0, L]^N$ of possible positions for all the cells, and thus the embryo as a whole has access to more positional information.
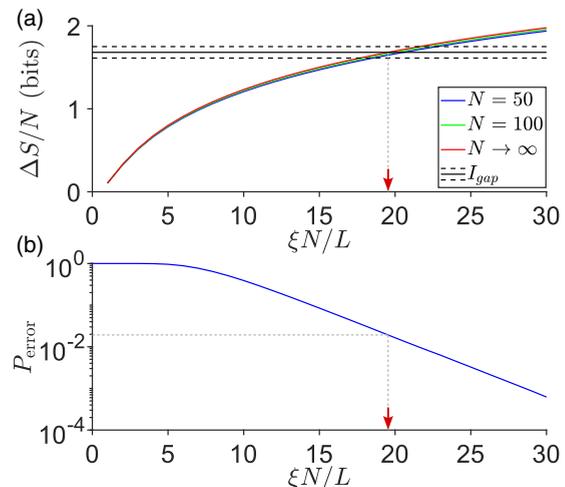

FIG. 3. Extra information from correlations, as a function of the correlation length. (a) Numerical results for $N = 50$ and 100 from Eq. (9) with the correlation matrix in Eq. (10); analytic results for $N \to \infty$ are from Eq. (15). Compare with the information gap from Appendix A (black solid line bracketed by dashed error bars). The intersection at $\xi = (19.5 \pm 1.9)(L/N)$ is marked by a vertical line and arrow. (b) Probability P_{error} of at least two signals being crossed, $\hat{x}_{n+1} < \hat{x}_n$, in a line of $N = 90$ cells, with $\sigma_x/L = 0.01$.

More formally, we can define the correlation matrix C,

$$\langle \delta x_n \delta x_m \rangle = \sigma_x^2 C_{nm}, \quad (8)$$

with diagonal elements $C_{nn} = 1$. Assuming again that the noise δx_n is Gaussian, the reduction in noise entropy for the entire set of variables $\{\delta x_n\}$ is given by the determinant of this matrix [26],

$$\Delta S = -\frac{1}{2}\log_2 \det C \text{ bits}, \quad (9)$$

and this reduction in entropy is the gain in information. Entropy is an extensive quantity, so when N is large the information gain per cell $\Delta S/N$ is finite. Can this be large enough to compensate for the information gap I_{gap}?

We expect that the correlation between fluctuations of positional signals in different cells depends on their spatial separation. Then C_{nm} is a function of the distance between cells n and m, $d_{nm} = |n - m|L/N$. A natural functional form is an exponential decay of correlations,

$$C_{nm} = e^{-d_{nm}/\xi}, \quad (10)$$

with correlation length ξ. This is what we would see if signals were encoded in the gradient of a single molecular species that has a lifetime τ and diffusion constant D, with $\xi = \sqrt{D\tau}$. Although this is oversimplified, it is useful for building intuition about how the range of correlations determines the additional information. Within this model it is straightforward to evaluate ΔS numerically, with results shown in Fig. 3(a).

We can also give an analytic theory for ΔS in the large-N limit, leading to Eq. (15), below. If we define eigenvalues and eigenvectors of the matrix C_{nm},

$$\sum_m C_{nm}\phi_m^\mu = \lambda_\mu \phi_n^\mu, \quad (11)$$

then we have

$$\Delta S = -\frac{1}{2} \sum_\mu \log_2 \lambda_\mu \text{ bits.} \tag{12}$$

In the limit of large N at fixed N/L, the ends of the embryo are far away and there is an effective translation invariance. This means that the eigenvectors ϕ_n^μ are complex exponentials, $\phi_n^\mu \propto \exp(iq_\mu n)$, or equivalently that the matrix C_{nm} is diagonalized by a discrete Fourier transform;[2] allowed values of q_μ are in the interval $-\pi \leqslant q < \pi$. Then as $N \to \infty$ we find the eigenvalues

$$\lambda(q) \to \sum_{n=-\infty}^{\infty} e^{-|n|L/N\xi} e^{iqn} = \frac{\sinh(L/N\xi)}{\cosh(L/N\xi) - \cos(q)} \tag{13}$$

and the change in entropy

$$\frac{\Delta S}{N} \to -\frac{1}{2} \int_{-\pi}^{\pi} \frac{dq}{2\pi} \log_2 \lambda(q) \tag{14}$$

$$= -\frac{1}{2} \log_2(1 - e^{-2L/N\xi}). \tag{15}$$

In Fig. 3(a) we see that this analytic result agrees with numerical results at $N = 50$ and 100, which agree with one another, confirming that the fly embryo is large enough for the entropy to be extensive. We conclude that an information gap of approximately 1.7 bits can be closed if correlations extend over distances $\xi = (19.5 \pm 1.9)(L/N) \sim 0.2L$. Lott *et al.* saw significant correlations across this range of distances for all the genes that they probed [17], and combinations of gap-gene protein levels have even longer correlation lengths [18].

Beyond the perhaps abstract information-theoretic measures, we can evaluate the probability that all cells receive signals that are in the correct order, that is, $\hat{x}_{n+1} > \hat{x}_n$ for all $n = 1, 2, \ldots, N$. If correlations extend over a distance $\xi \sim 19.5(L/N)$, then all signals will have the correct ordering in approximately 98% of embryos, as illustrated in Fig. 3(b).

We emphasize that correlations extending over $\xi \sim 0.2L$ do not require special mechanisms to connect these long distances. As noted above, if the relevant signals are carried by a single molecule with diffusion constant D and lifetime τ, we expect that fluctuations in concentration will have a correlation length $\xi \sim \sqrt{D\tau}$. In a network of interacting molecules, as with the gap genes in the fly embryo, the role of τ is played by relaxation times for the network as a whole, and these emergent timescales can be much longer than the lifetime of the individual species because of feedback [18,27].

B. Experiment

To close the positional information gap, we *predict* that the noise in positional signals should be correlated over distances $\xi \sim 0.2L$. These distances are long compared to the separation between neighboring cells. The first indication that such correlations exist came from experiments marking the

boundaries of gene-expression domains as seen through measurements of mRNA for selected gap genes and the pair-rule gene *eve* [17]. At the same time, it was reported that fluctuations in the concentration of a single-gap-gene product protein are correlated only over short distances [28]. Analyzing simultaneous measurements on protein concentrations of four gap genes demonstrated that different combinations or modes of the network have different correlation lengths [18]; the longest correlation lengths are a significant fraction of the length of the embryo. Finally, errors in the position inferred from gap-gene-expression levels are reduced if we allow for alignment by translation along the x/L axis, indicating that errors in relative position are smaller than errors in absolute position [21]. All of this suggests that the noise in positional signals is spatially correlated. Can we make this more quantitative?

We analyze the experiments in Ref. [10], which used immunofluorescence stainings to measure spatial profiles of protein concentration for three of the pair-rule genes *eve*, *prd*, and *rnt* (Fig. 1). The data include $N_{\text{em}} = 109$ embryos, fixed and stained in the time window from 35 to 60 min after the start of nuclear cycle 14. This is the period of cellularization, and as in previous work the progress of the cellularization membrane provides a time marker with an accuracy of 1 min [16]. For each of the three genes, the seven peaks in the striped concentration profile can be found automatically, and their locations vary linearly with time throughout this period [29]. If we do not correct for this systematic dynamical behavior, the variance of stripe positions will be large and their fluctuations will be correlated, artificially. We consider the noise in position to be the deviation from the best-fit linear relation for each individual stripe marker. The standard deviations then are consistently slightly below $\sigma_x \sim 0.01L$ and the distribution of fluctuations is well approximated by a Gaussian. These results agree with previous work [10,12,29] and are summarized in Appendix A.

Before analyzing correlations, we can use these data to make a more precise estimate of the information gap. If each cell has access to a positional signal with errors $\sigma_x(n)$, which might vary with n, the average positional information available to a single cell is

$$I_{\text{position}} = \log_2 L - \langle \log_2[\sqrt{2\pi e}\sigma_x(n)] \rangle_n, \tag{16}$$

where $\langle \cdots \rangle_n$ denotes an average over cells, generalizing Eq. (6). Rather than making inferences about single cells, we have direct access to the signals that mark the locations of the stripes in the expression of three pair-rule genes, for a total of 21 features spread across half the AP axis. The mean separation between the nearest stripes is $\Delta\bar{x} = 0.023L$, just a few times larger than the spacing between cells. Rather than introducing a model that would interpolate, we take the stripe positions themselves as the signals x_n, now with $n = 1, 2, \ldots, 21$, and the average in Eq. (16) becomes an average over stripes.

The challenge in evaluating the positional information is that random errors in our estimates of the errors $\sigma_x(n)$ become systematic errors in estimates of information. This problem of systematic errors was appreciated in the very first efforts to use information-theoretic concepts to analyze biological experiments [30]. The analysis of neural codes has been an

---

[2]The discreteness is important. If we take a continuum limit so that the sum in Eq. (13) becomes an integral, the calculation is a bit simpler but leads to a significant overestimate of ΔS, even at large values of $\xi N/L$.
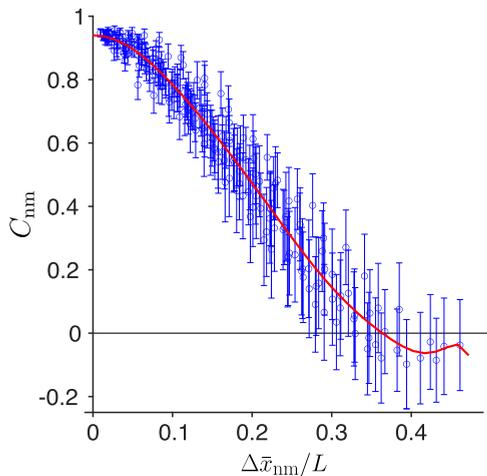
FIG. 4. Correlations between noise in peak positions of the *eve*, *run*, and *prd* stripe patterns, from Eq. (17), as a function of the mean separation between stripes. Error bars are estimated from the standard deviation across random halves of the data. With three genes, each having seven stripes, we observe $(21 \times 20)/2 = 210$ distinct elements of the correlation matrix C_{nm}. The red solid line is a smooth curve to guide the eye.

important testing ground for methods to address these errors [31–33]; for a review see Appendix A.8 of Ref. [26]. The approach we take here uses the fact that naive entropy estimates depend systematically on the size of the sample; if we can detect this systematic dependence then we can extrapolate to infinite data, as described in Appendix A. The result is that $I_{\text{gap}} = 1.68 \pm 0.07$ bits/cell.

The idea of positional information is that cells have access to a signal that represents position along the axis of the embryo [2,21]. In the discussion above we have taken this idea at face value, identifying the signal in each cell as $\hat{x}_n$. However, the signals we observe are the positions of stripes in three different pair-rule genes, and the different stripes for each gene are controlled by different enhancers responding to distinct combinations of transcription factors. We need to test the hypothesis that these multidimensional molecular concentrations encode a single positional variable.

We are looking at fluctuations in the positions of the stripes, δx_n. Figure 4 shows the elements of the correlation matrix

$$C_{nm} \equiv \frac{\langle \delta x_n \delta x_m \rangle}{[\langle (\delta x_n)^2 \rangle \langle (\delta x_m)^2 \rangle]^{1/2}} \qquad (17)$$

as a function of the mean separation $\Delta \bar{x}_{nm}$ between stripes n and m measured along the dorsal side of the embryo, starting with images as in Fig. 1. We see that, within experimental error, the correlations really are a function of distance. There is no obvious pattern linked to the identity of the enhancers that control these different features or to the identity of the transcription factors to which the enhancers respond: Nearby stripes are highly correlated, the decay of correlations with distance is the same whether we are looking at correlations between the same or different genes, and different pairs of stripes with the same mean separation have the same

correlation.[3] This suggests that, as in the theoretical discussion above, we can think about an abstract positional signal that is transmitted to each cell and controls the placement of the pair-rule stripes. Correspondingly, there are strong indications that the correlations are inherited from the structure of the noise in gap-gene expression (Appendix D). We see the same results along the ventral side of the embryo, though with larger errors because of difficulties associated with curvature of the contour.

Qualitatively, the correlations that we see in Fig. 4 decay over distances $\xi \sim 0.2L$, consistent with the scale needed to close the information gap and with early measurements [17]. Quantitatively, the decay of correlations is not well described by a single exponential function of distance, so we cannot simply transcribe the predictions of the theory. Instead, we would like to make a direct estimate of the positional information from the data. Conceptually this is simple: We estimate the correlation matrix from the data and then compute the (logarithm of the) determinant of this matrix following Eq. (9). As with the information gap itself (above), the problem is that random errors in our estimates of individual matrix elements become systematic errors in the entropy. We follow the same strategy of identifying the dependence of this error on the number of embryos that we include in our analysis and extrapolating to large data sets (Appendix C).

We emphasize that our estimates of $\Delta S/N$ are based on the joint distribution of the stripe positions, ultimately including all 21 measured stripes. We are able to make progress because the distribution of positional noise is well approximated as Gaussian (Appendix A), so the 21-dimensional distribution is summarized by the 21×21 matrix C_{nm} and we can thus estimate the information per stripe contained in the entire pattern. These information measures are independent of the molecular mechanisms that give rise to the underlying correlations.

By definition, to see the extra information hidden in correlations we have to look at the positions of multiple stripes. We start with two neighboring stripes and gradually work toward all $N = 21$ stripes; results are shown in Fig. 5. Note that at each N we are measuring an information gain per stripe, and small differences among the stripes are included in our error bars. The added information grows to $\Delta S/N = 1.65 \pm 0.08$ bits/stripe, and we see that this is relatively constant for $N \geqslant 19$ stripes. This suggests that our analysis includes distances long enough to capture all of the relevant correlations, so ΔS becomes truly proportional to N, as in the discussion of Fig. 3(a). Strikingly, this plateau matches the information gap $I_{\text{gap}} = 1.68 \pm 0.07$ bits/cell within errors.

---

[3] We see hints of weak negative correlations at long distances also in our analysis of the gap genes (Appendix D), although the error bars make these measurements consistent with zero. Negative correlations between individual gene-expression levels emerge naturally in networks with mutually repressive interactions, as with the gap genes, but it is unclear how these would project into errors in position. Small negative correlations at long distances could also be spurious, resulting from imperfect azimuthal alignment of the embryos (Appendix A).
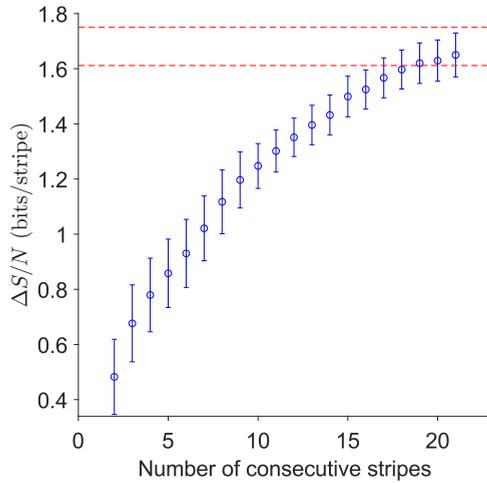
FIG. 5. Extra information from correlations, $\Delta S/N$, computed from the observed correlations in pair-rule stripe fluctuations C_{nm} through Eq. (9), including different numbers of contiguous stripes. Circles and error bars (blue) are the extrapolated estimates from Appendix C. Red dashed lines are plus or minus one standard error of the mean around the best estimate of the information gap $I_{\text{gap}} = 1.68 \pm 0.07$ bits/cell from Appendix A.

IV. DISCUSSION

There is strong evidence that, early in embryonic development, each cell acquires a distinct identity [11]; it is less clear how this information is encoded. In the fruit fly embryo, positional information along the anterior-posterior axis is orchestrated through a sequential cascade involving three primary maternal inputs, a select number of gap genes, and the pair-rule genes. The conventional perspective suggests that the information flow through this cascade entails a gradual refinement, with noisy inputs ultimately generating a precise and reproducible pattern [34,35], in the spirit of the Waddington landscape [36].

In contrast to the picture of noisy inputs and precise outputs, at least one maternal input itself exhibits a high level of precision, consistently reproducible across embryos [28,37]. Moreover, the expression levels of gap genes within a single cell prove sufficient to determine positions with an error smaller than the distance between neighboring cells [10,12]. Notably, this precision agrees with that observed in downstream events such as the pair-rule stripes. In parallel, crucial developmental events exhibit highly reproducible temporal trajectories [38]. These quantitative observations challenge the conventional view of refinement and error correction, supporting instead a precisionist perspective in which locally available information is processed and preserved with near optimal efficiency. Given that all relevant molecules are present at low copy numbers, this places significant constraints on the architecture of the underlying networks [38–41].

Despite their precision, local signals in the fly embryo do not quite provide enough information to uniquely specify all $N = 90 \pm 4$ cellular identities along the AP axis, $I_{\text{unique}} = \log_2 N$: Errors in the position that a cell can infer from molecular concentrations come from a distribution and distributions have tails [12]. The result is that there is

a substantial (approximately 25%) gap between the information provided by the gap genes, or the pair-rule stripes, and I_{unique}.

Previous measurements have characterized the noise in local estimates of position for each cell individually; however, there are many hints from previous work that this noise is correlated [12,17,18]. Extra information can be hiding in these correlations, and we have seen in Sec. III A that if correlations extend over distances $\xi \sim 0.2L$ then this would be enough to close the information gap. This prompts a more detailed examination of the noise correlations, which really do seem to be a function of distance independent of gene identity (Fig. 4).

The perhaps surprising conclusion of Sec. III B is that the extra information contained in the correlations, $\Delta S/N$, matches the information gap I_{gap} almost perfectly, with an error bar of less than 2%:

$$I_{\text{gap}} - \Delta S/N = (0.0048 \pm 0.0162)I_{\text{unique}}. \qquad (18)$$

This agreement strongly supports the precisionist view of information flow in this system.

Historically, the lack of precise data on gene-expression levels, with uncertainties extending to factors of 2, led to skepticism regarding the relevance of more refined measurements to general mechanisms of genetic control. These expectations stood in contrast, for example, to our understanding of signaling in rod photoreceptors, where the quantitative reproducibility of responses to single molecular events provides important constraints on the underlying biochemical mechanisms [42].

The fly embryo has provided a laboratory within which to explore precision vs noisiness in the function of an intact living system. We have seen reproducible protein and mRNA concentrations across embryos with an accuracy of 10% [16,28,37], and these concentrations encode position with an accuracy of approximately 1% of the embryo's length [10,12,21]. The present study adds a layer to this understanding, demonstrating that the available positional information, including the subtle effects of correlated noise, matches the threshold for specifying unique cellular identities, and this match itself has an accuracy of better than 2%. Beyond the fly embryo, these results suggest a more general conclusion: Quantitative measurements in living systems merit serious consideration, even at high precision, as in other areas of physics.

APPENDIX A: STATISTICS OF INDIVIDUAL STRIPES

The raw data for our analyses are the profiles of fluorescence intensity vs position along the length of the embryo, as in Fig. 1. These embryos have been fixed and stained with antibodies against the proteins encoded by the pair-rule genes *eve*, *prd*, and *rnt* and fluorescently tagged antibodies against those antibodies [10]. Independent experiments demonstrate that these classical staining methods, used carefully, yield fluorescence intensities that are linear in protein concentrations [16]. The data set used here, which contains a large number of wild-type embryos, comes from Ref. [10].

We briefly summarize the imaging protocol and describe the procedure for localizing the stripe positions. Images are taken in the midsagittal plane showing a row of nuclei along the dorsal and ventral side of the embryo. For consistency and to avoid geometric distortion, we focus on the dorsal profiles, as was done previously (but see Fig. 11 below). In order to include the entire embryos in a single image, large field-of-view images, with a pixel size of 445 nm, are acquired with a 20×0.7 numerical aperture (NA) objective on a Leica SP5 confocal microscope. Fluorescence intensity is averaged inside a sliding window of the size of a nucleus and the position of the window center is recorded. In a given embryo, positions of the seven stripes are first roughly identified by finding local maxima in the profile of an individual embryo. To make this quantitative, we tried several methods. First, we used an iterative procedure in which the mean peak shape is used as a template [29]. Second, we fitted a model of seven Gaussians with variable amplitudes and widths to the entire profile. Finally, we fitted individual Gaussians to each stripe, using a window centered on the local maximum with a width of 5% the embryo length. These methods give consistent results, and importantly global fits do not generate larger correlations than local fits. In the end, we use the local Gaussian fits, as in Fig. 6(a).

The age of embryos is estimated to 1-min precision in nuclear cycle 14 by measuring the length of the cellularization membrane [12]. At 30 min into this cycle, the stripes of *prd* first start to become visible and the other two genes have well-defined stripes by that time, so we confine our attention to $t > 30$ min.

Stripe patterns are dynamic, with positions that depend on time. If we do not take account of this systematic variation, then across an ensemble of embryos with different ages we would see artificial correlations among fluctuations in stripe position. For example, we would see an artificial negative correlation between the first and the last stripe position because they move in the opposite direction (towards the middle of the embryo) during the course of the nuclear cycle 14. Stripe movement is small, however, and we can use a linear fit to remove the effect of the temporal shifts, separately for each of the 21 stripes, across the population of embryos:

$$x_n(t) = x_n(t_0) + s_n(t - t_0). \tag{A1}$$

Results are shown in Figs. 6(b) and 6(c). For each embryo we find an equivalent position of all the stripes at a reference time $t_0 = 45$ min [29]. Attempts to fit the shifts with more complex functions of time do not reduce the variance $\sigma_x^2(n)$ nor do they change the correlations C_{nm}.
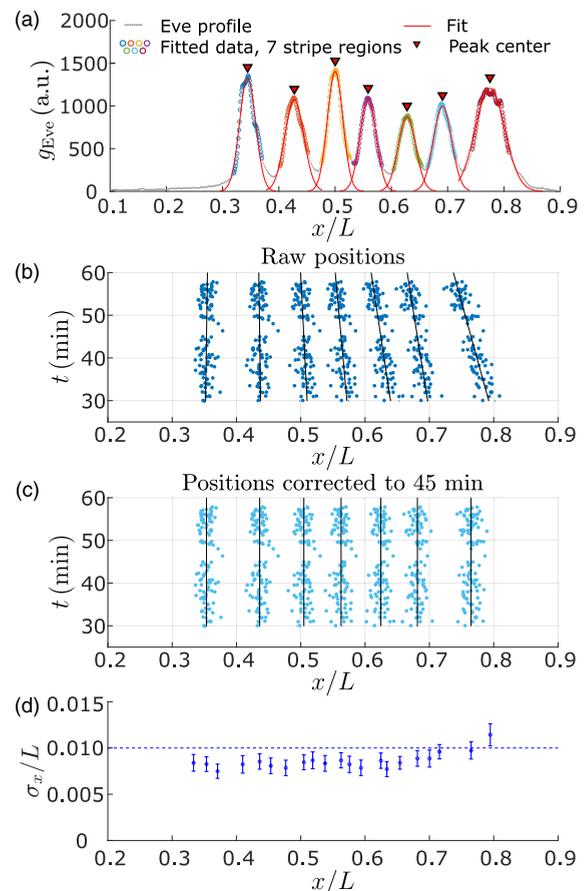

FIG. 6. Pair-rule stripe positions. (a) Concentration of Eve protein in a single embryo. Colored circles indicate regions which were fitted with a Gaussian function to calculate the stripe position. Each stripe is fitted individually, with fits shown in red. Red triangles indicate the centers of each fitted peak. (b) Stripe positions as a function of time in the nuclear cycle 14. Linear fits from Eq. (A1) are shown as black lines. (c) Peak positions $x_n(t_0)$ corrected to $t_0 = 45$ min. (d) Positional error of the pair-rule stripes. The magnitude of the error $\sigma_x(n)$ is plotted against the mean position $\bar{x}_n$ for each of the *eve*, *prd*, and *rnt* stripes. Errors in $\bar{x}_n$ are standard errors of the mean; errors in σ_x are standard deviations across random halves of the data. Dashed line marks the rough estimate $\sigma_x/L \sim 0.01$.

Another possible experimental source of artificial correlations is the azimuthal orientation of the embryo. Pair-rule stripes splay outward from the dorsal to the ventral side, and errors in azimuthal orientation would generate correlated errors in position. The errors are small for nearby stripes and negative for the most separated stripes. If we make azimuthal errors of approximately $10°$, then for the first and seventh stripes there would be positional errors approximately $\pm 0.003L$, which would generate a correlation coefficient approximately equal to $-(0.003/0.01)^2 \sim -0.09$; importantly this is much smaller for stripes that are closer together. Even though long-distance correlations in Figs. 4 and 10(c) are both zero within the error bar, the experimental uncertainty in azimuthal orientation might explain why the mean correlation coefficient dips below zero when $\Delta x/L > 0.35$.

With x_n the position of each pair-rule stripe, we have the mean and variance

$$\bar{x}_n = \langle x_n \rangle, \tag{A2}$$

$$\sigma_x^2(n) = \langle (x_n - \bar{x}_n)^2 \rangle, \tag{A3}$$

where $\langle \cdots \rangle$ denotes an average over our complete experimental ensemble of $N_{\text{em}} = 109$ embryos. Results are shown in Fig. 6(d), where we confirm that positional errors are almost all smaller than 1% of the embryo length.

Beyond measuring the variance, we can estimate the distribution of positional errors. Since the different stripes have slightly different σ_x, we normalize the positional errors for each stripe individually,

$$z_n = (x_n - \bar{x}_n)/\sigma_x(n). \tag{A4}$$

With this normalization we can pool across all 21 stripes, and we estimate the distribution of z as usual by making bins and counting the number of examples in each bin, with results shown in Fig. 7(a). Qualitatively, the distribution is close to being Gaussian, but what matters for our analysis is the entropy of this distribution.

When we estimate a probability distribution and use this estimate to compute the entropy, the random errors in the distribution that arise from the finiteness of our sample become systematic errors in the entropy. The general version of this problem goes back to the very first efforts to use information-theoretic concepts to analyze biological experiments [30]; for a review see Appendix A.8 of Ref. [26]. Briefly, naive entropy estimates depend systematically on the size of the sample, and if we can detect this systematic dependence we can extrapolate to infinite data, thus providing an unbiased estimate of the entropy. In Fig. 7(b) we show the difference between the entropy of the estimated distribution $P(z)$ and the entropy of a Gaussian. We see that when we base our estimates on N_{em} embryos there is a (small) term approximately equal to $1/N_{\text{em}}$, as expected. Extrapolating $N_{\text{em}} \to \infty$, we see that the entropy difference goes to zero within the small (less than 0.01 bit) error bars. We conclude, for the purposes of our discussion, that it is safe to approximate the positional errors as being Gaussian.

Finally, we can use the same extrapolation methods to provide a better estimate of the information gap defined in the main text. Equation (16) defines the positional information contained in the local signals, I_{position}, and the information gap is the difference between this and $I_{\text{unique}} = \log_2 N$. Figure 7(c) shows the values of

$$I_{\text{gap}} = I_{\text{unique}} - I_{\text{position}} = \left\langle \log_2 \left(\sqrt{2\pi e} \frac{N\sigma_x(n)}{L} \right) \right\rangle_n \tag{A5}$$

estimated from fractions of our data set and then extrapolated. The result is $I_{\text{gap}} = 1.68 \pm 0.07$ bits/cell.

APPENDIX B: COUNTING NUCLEI

We quantify the number of nuclei in a single row (one dimension) along the AP axis using living *Drosophila* embryos that express a transgene with fluorescently labeled histone on the second chromosome (his-RFP/CyO). Embryos are imaged in a Zeiss LSM880 confocal microscope with a
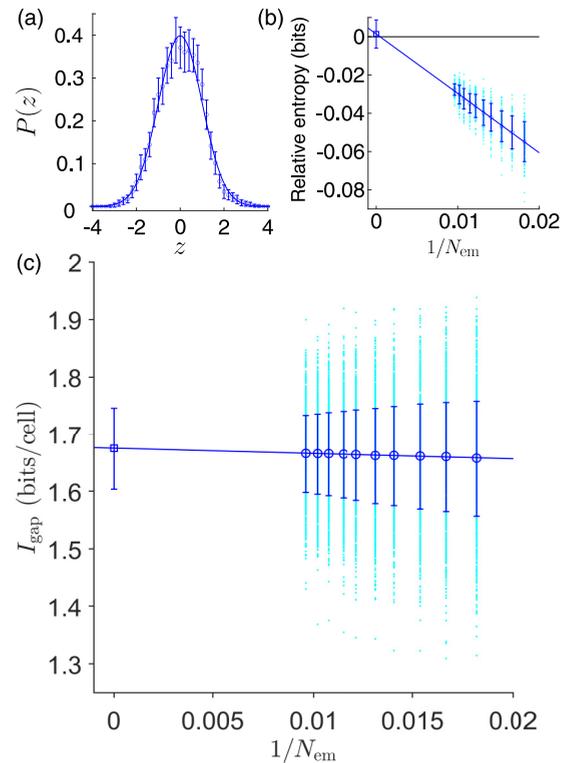

FIG. 7. (a) Estimate of the distribution of normalized errors (A4). Positional errors are well approximated as Gaussian. Open circles are means pooled across all stripes and embryos, error bars are standard deviations across random halves of the embryos, and the line is the Gaussian with zero mean and unit variance. (b) Entropy difference between this estimated distribution and the Gaussian, as a function of the (inverse) number of embryos we include in our analysis. Cyan points are examples from random choices out of the full ensemble of embryos, open circles with error bars are the mean and standard deviations of these points, and the line is a linear extrapolation [26,30–33]. (c) Estimates of the information gap (A5). Cyan points are examples from random choices out of the full ensemble of embryos, blue open circles with error bars are the mean and standard deviations of these points, and the line is a linear extrapolation to $I_{\text{gap}} = 1.68 \pm 0.07$ bits/cell.

20×0.8 NA objective and pinhole equivalent to 1 Airy unit. Pixel size is 0.35 μm, corresponding to about 7% of the size of the nucleus. We acquire a z-stack that includes the midsagittal plane during the middle of nuclear cycle 14. Both fluorescent and bright-field image stacks are collected using a z-step of 1 μm. From this stack, the midsagittal plane is identified by inspection of the largest extent of the embryo, where the embryo edge is in focus.

A mask of the embryo is created from the bright-field image using edge detection to separate the embryo from the uniform background (custom code in PYTHON). The embryo length L is measured as the length of the straight line from the most anterior to the most posterior points of the mask. This line defines the AP axis, from which we determine the central 80% [Fig. 8(a)]. We manually count the number of nuclei along the middle 80% (from 10% to 90% of L) of the dorsal side [Fig. 8(b)]. We count 72 ± 3 nuclei ($n = 26$ embryos) in the middle 80% [Fig. 8(c)], which corresponds to 90 ± 4
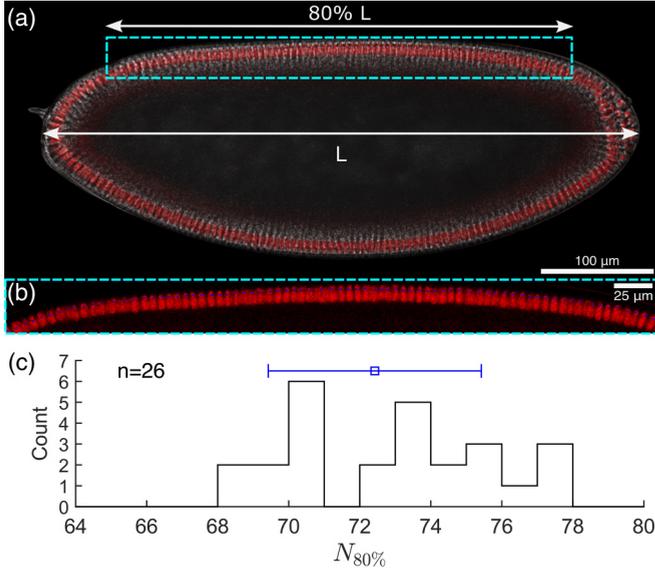
FIG. 8. Counting nuclei in nuclear cycle 14. (a) Fluorescence image of an embryo with labeled histones highlighting the nuclei underlaid with a bright-field image of the same embryo. Focus is on the midsagittal plane. (b) Close-up of the central 80% on the dorsal side, showing that we can count nuclei by hand. (c) Results from $n = 26$ embryos. Histogram has a mean plus or minus standard deviation of 72 ± 3.

nuclei for the entire length of the AP axis assuming uniform nuclear density along the one-dimensional line. The error bar of the nuclear count includes both embryo-to-embryo variability and ambiguities in nuclear identification during counting.

APPENDIX C: ENTROPY ESTIMATES

Figure 9 shows estimates of the extra information $\Delta S/N$ [Eq. (9)] based on measurements in different numbers of embryos, for $N = 10$ and 20 contiguous pair-rule stripes. We see the expected dependence on $1/N_{\text{em}}$, and the steepness of this dependence is twice as large at $N = 20$ than at $N = 10$,
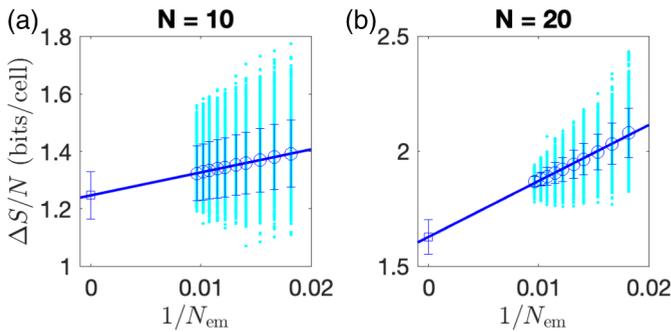

FIG. 9. Entropy reduction by correlations among the pair-rule stripe fluctuations, estimated from different numbers of embryos N_{em}, for (a) $N = 10$ stripes and (b) $N = 20$ stripes. Cyan points are examples from random choices out of the full ensemble of embryos, open blue circles with error bars are the mean and standard deviations of these points, and the line is a linear extrapolation to the square.

as expected [26]. This gives us confidence in the extrapolation $N_{\text{em}} \to \infty$ [26,30–33].

APPENDIX D: ORIGIN OF THE CORRELATIONS

The precision of pair-rule stripe placement matches, quantitatively, the noise in optimal estimates of position based on the local expression levels of the gap genes [10,12]. To be consistent with this result, the correlations should also be visible in the gap genes. As noted above, Lott *et al.* saw correlations in expression boundaries for selected gap genes [17], and later measurements showed that combinations of gap-gene-expression levels have correlations extending over a significant fraction of the embryo [18]. Here we revisit these measurements and connect fluctuations in gap-gene expression to positional noise. Note that for the pair-rule genes we can work directly with the positions of the stripes, but for the gap genes we have to think more carefully about how positions are encoded in expression levels.

We start with a brief review of ideas about decoding positional information [10]. Measurements of gap-gene expression in multiple embryos provide samples from the conditional distribution $P(\{g_i\}|x)$, at all values of the position x along the anterior-posterior axis; we focus on the $d = 4$ gap genes expressed in the middle approximately 80% of the embryo, *hunchback*, *giant*, *krüppel*, and *knirps*. To a good approximation this distribution is Gaussian,

$$P(\{g_i\}|x) = \frac{1}{Z(x)} \exp\left(-\frac{1}{2}\chi^2(\{g_i\};x)\right), \quad \text{(D1)}$$

$$Z(x) = [(2\pi)^d \det \hat{C}(x)]^{1/2}, \quad \text{(D2)}$$

$$\chi^2(\{g_i\};x) = \sum_{i,j=1}^{d} [g_i - \bar{g}_i(x)][\hat{C}^{-1}(x)]_{ij}[g_j - \bar{g}_j(x)], \quad \text{(D3)}$$

where $\bar{g}_i(x)$ is the mean expression level of gene i at position x and

$$[\hat{C}(x)]_{ij} = \langle \delta g_i \delta g_j \rangle_x \quad \text{(D4)}$$

is the covariance matrix of fluctuations around these means. To decode the position of a cell from the local expression levels we need to construct

$$P(x|\{g_i\}) = \frac{P(\{g_i\}|x)P(x)}{P(\{g_i\})}. \quad \text{(D5)}$$

However, because nuclei are arrayed uniformly along the length of the embryo, $P(x)$ is uniform and hence the dependence on x is captured in Eq. (D1).

A cell at the actual position x_{true} has expression levels

$$g_i = \bar{g}_i(x_{\text{true}}) + \delta g_i, \quad \text{(D6)}$$

and if the positional noise is small we can write

$$\bar{g}_i(x) = \bar{g}_i(x_{\text{true}}) + (x - x_{\text{true}})\frac{d\bar{g}_i(x)}{dx}\bigg|_{x=x_{\text{true}}} + \cdots, \quad \text{(D7)}$$

which we substitute into $P(x|\{g_i\})$. With uniform prior $P(x) = 1/L$, the best estimate of x maximizes $P(\{g_i\}|x)$. In principle, there is a contribution from the normalization $Z(x)$,

or more generally from derivatives of the covariance matrix $\hat{C}(x)$. However, if the noise level is small these contributions to maximizing $P(\{g_i\}|x)$ are suppressed by a factor of the noise variance itself. Unless $\hat{C}(x)$ varies very rapidly with x—and we have checked that it does not—this is sufficient to make minimizing χ^2 a good approximation to maximizing $P(\{g_i\}|x)$. This estimate can be written as

$$\hat{x} = x_{\text{true}} + \delta x, \quad \text{(D8)}$$

$$\delta x(x_{\text{true}}) = \left(\sigma_x^2(x) \sum_{i,j=1}^{d} \delta g_i [\hat{C}^{-1}(x)]_{ij} \frac{d\bar{g}_j(x)}{dx} \right)_{x=x_{\text{true}}}, \quad \text{(D9)}$$
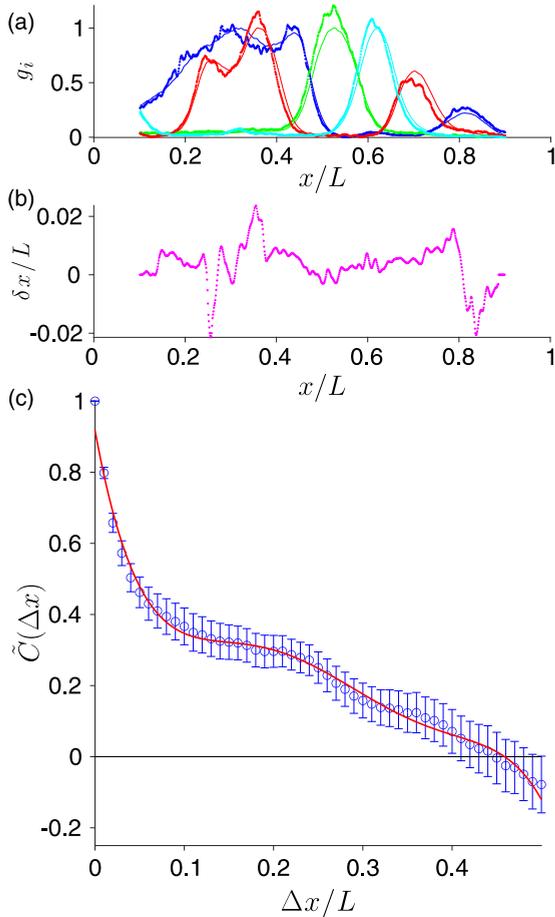

(a)

(b)

(c)

FIG. 10. Decoding gap-gene-expression levels in a single embryo and correlations in the resulting pattern of positional errors. (a) Expression of Hb (blue), Kr (green), Gt (red), and Kni (cyan). Thin solid lines are means across $N_{\text{em}} = 38$ embryos in a small window $40 \leqslant t \leqslant 44$ min in nuclear cycle 14; dense points are data from a single embryo [10]. (b) Positional errors computed from Eq. (D9). (c) Correlations in the positional noise inferred from gap-gene expression. For each embryo α we compute the correlation function in Eq. (D12) and then normalize to give $\tilde{C}(\Delta x) = C(\Delta x)/C(0)$. Blue circles with error bars are mean and standard error across $N_{\text{em}} = 38$ embryos; the red solid line is a smooth curve to guide the eye.

where the variance of positional noise is defined by

$$\frac{1}{\sigma_x^2(x)} = \sum_{i,j=1}^{d} \frac{d\bar{g}_i(x)}{dx} [\hat{C}^{-1}(x)]_{ij} \frac{d\bar{g}_j(x)}{dx}; \quad \text{(D10)}$$

for consistency we have

$$\langle [\delta x(x)]^2 \rangle = \sigma_x^2(x). \quad \text{(D11)}$$

Note the connection to Eqs. (1) and (2) in Sec. II.

Previous work has emphasized the scale of positional errors σ_x [10,12,21]. However, the optimal decoding of gap-gene-expression levels [10] maps the deviation of expression levels from the mean into a decoding error for each embryo individually, as in Eq. (D9). An example is in Fig. 10, where the small fluctuations of expression levels around the mean [Fig. 10(a)] translate into proportionally small errors δx [Fig. 10(b)].

For each embryo α we can take the positional errors $\delta x_\alpha(x)$ and compute the correlation function

$$C_\alpha(\Delta x) = \frac{1}{L - \Delta x} \int dx\, \delta x_\alpha(x) \delta x_\alpha(x + \Delta x). \quad \text{(D12)}$$

Figure 10(c) shows the mean and standard error of the normalized correlation function across all $N_{\text{em}} = 38$ embryos in our experimental ensemble. Qualitatively, correlations in the positional noise encoded by the gap genes extend over distances similar to the correlation in positional noise of the pair-rule stripes (Fig. 4). Quantitatively, the gap-gene correlations include an additional component with a short correlation length. One possibility is that this component is averaged away by interactions among neighboring cells during expression of the pair-rule stripes. Another possibility is that a modest fraction of the noise in gap-gene expression reflects local noise in the measurements, as discussed previously [16]; this measurement noise has only a small impact on our estimates of the effective noise σ_x but a larger impact on the shape of the correlation function. It seems likely that both effects contribute.
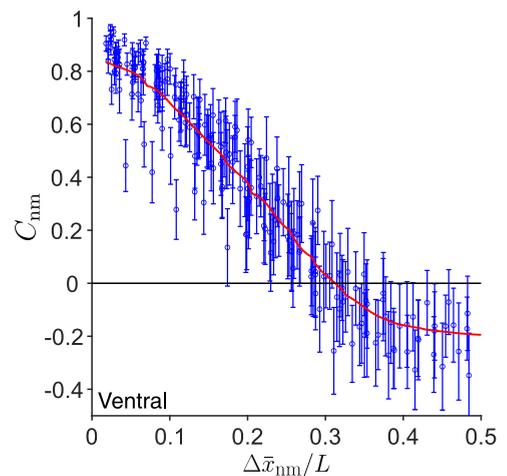

FIG. 11. Correlations between noise in peak positions of the *eve*, *run*, and *prd* stripe patterns, as in Fig. 4, but with stripe positions measured along the ventral side of the embryo. Error bars are estimated from the standard deviation across random halves of the data. With three genes, each having seven stripes, we observe $(21 \times 20)/2 = 210$ distinct elements of the correlation matrix C_{nm}. The red solid line is a smooth curve to guide the eye.

Nonetheless, it is clear that relatively long-range correlations, which are crucial to closing the information gap, are present already in the gap-gene-expression levels, as suggested in earlier work [12,17,18].

While new experiments will be needed to estimate the information that is encoded in the gap-gene correlations, one can ask how the different gap genes are contributing to these correlations. In particular, it is interesting that the correlations at long distances depend on correlations among different combinations of genes. As an example, near $x/L = 0.4$ only Hb and Gt have strongly nonzero expression, so it is some combination of fluctuations in the expression levels of these two genes that determines the local positional error. Near $x/L = 0.6$, however, only Kr and Kni have significant expression, and so positional errors are determined by a combination of expression fluctuations in these two genes. However, the data show that positional noises at points separated by $\Delta x/L \sim 0.2$ are correlated. Thus, not only are fluctuations in gene-expression levels correlated over long distances, but the relevant correlations are among different genes, as emphasized previously by Krotov *et al.* [18]. It is plausible that these intergene correlations are a signature of interactions in the gap-gene network, which can propagate along the length of the embryo via diffusion.

Finally, as a check, we redo the analysis of Fig. 4 using measurements of stripe positions along the ventral edge of the embryo. We expect to see essentially the same pattern of correlations, although with larger errors since measurements along a curved contour are more challenging. This is what is shown in Fig. 11.

[1] A. M. Turing, The chemical basis of morphogenesis, Philos. Trans. R. Soc. London B **237**, 37 (1952).

[2] L. Wolpert, Positional information and the spatial pattern of cellular differentiation, J. Theor. Biol. **25**, 1 (1969).

[3] G. Tkačik and T. Gregor, The many bits of positional information, Development **148**, dev176065 (2021).

[4] C. Nüsslein-Volhard and E. Wieschaus, Mutations affecting segment number and polarity in *Drosophila*, Nature (London) **287**, 795 (1980).

[5] P. A. Lawrence, *The Making of a Fly: The Genetics of Animal Design* (Blackwell Scientific, Oxford, 1992).

[6] E. Wieschaus and C. Nüsslein-Volhard, The Heidelberg screen for pattern mutants of *Drosophila*: A personal account, Annu. Rev. Cell Dev. Biol. **32**, 1 (2016).

[7] R. Rivera-Pomar and H. Jackle, From gradients to stripes in *Drosophila* embryogenesis: Filling in the gaps, Trends Genet. **12**, 478 (1996).

[8] J. Jaeger, The gap gene network, Cell. Mol. Life Sci. **68**, 243 (2011).

[9] See https://bugguide.net/node/view/1308194/bgimage.

[10] M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, and T. Gregor, Optimal decoding of cellular identities in a genetic network, Cell **176**, 844 (2019).

[11] J. P. Gergen, D. Coulter, and E. F. Wieschaus, in *Gametogenesis and The Early Embryo*, edited by J. G. Gall (Liss, New York, 1986), pp. 195–220.

[12] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, Positional information, in bits, Proc. Natl. Acad. Sci. USA **110**, 16301 (2013).

[13] F. Liu, A. H. Morrsison, and T. Gregor, Dynamic interpretation of maternal inputs by the *Drosophila* segmentation gene network, Proc. Natl. Acad. Sci. USA **110**, 6724 (2013).

[14] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie, Probing gene expression in live cells, one protein molecule at a time, Science **311**, 1600 (2006).

[15] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells, Science **329**, 533 (2010).

[16] J. O. Dubuis, R. Samanta, and T. Gregor, Accurate measurements of dynamics and reproducibility in small genetic networks, Mol. Syst. Biol. **9**, 639 (2013).

[17] S. E. Lott, M. Kreitman, A. Palsson, E. Alekseeva, and M. Z. Ludwig, Canalization of segmentation and its evolution in *Drosophila*, Proc. Natl. Acad. Sci. USA **104**, 10926 (2007).

[18] D. Krotov, J. O. Dubuis, T. Gregor, and W. Bialek, Morphogenesis at criticality, Proc. Natl. Acad. Sci. USA **111**, 3683 (2014).

[19] C. Nüsslein-Volhard and S. Roth, Axis determination in insect embryos, Ciba Found Symp. **144**, 37 (1989).

[20] C. Nüsslein-Volhard, Determination of the embryonic axes of *Drosophila*, Development **113**(Suppl. 1), 1 (1991).

[21] G. Tkačik, J. O. Dubuis, M. D. Petkova, and T. Gregor, Positional information, positional error, and read-out precision in morphogenesis: A mathematical framework, Genetics **199**, 39 (2015).

[22] V. E. Foe and B. M. Alberts, Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis, J. Cell Sci. **61**, 31 (1983).

[23] S. Dutta, N. J.-V. Djabrayan, S. Torquato, S. Y. Shvartsmanm and M. Krajnc, Self-similar dynamics of nuclear packing in the early *Drosophila* embryo, Biophys. J. **117**, 743 (2019).

[24] X. C. Li, L. Gandara, M. Ekelöf, K. Richter, T. Alexandrov, and J. Crocker, Laboratory evolution of flies to morphogen dosage via rapid maternal changes reveals predictable outcomes, bioRxiv: 509860.

[25] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[26] W. Bialek, *Biophysics: Searching for Principles* (Princeton University Press, Princeton, 2012).

[27] M. Nikolić, V. Antonetti, F. Liu, G. Muhaxheri, M. D. Petkova, M. Scheeler, E. M. Smith, W. Bialek, and T. Gregor, Scale invariance in early embryonic development, arXiv:2312.17684.

[28] T. Gregor, D. W. Tank, E. F. Wieschaus, and W. Bialek, Probing the limits to positional information, Cell **130**, 153 (2007).

[29] V. Antonetti, W. Bialek, T. Gregor, G. Muhaxheri, M. Petkova, and M. Scheeler, Precise spatial scaling in the early fly embryo, arXiv:1812.11384.

[30] G. A. Miller, in *Information Theory in Psychology: Problems and Methods II-B*, edited by H. Quastler (Free Press, Glencoe, 1955), pp. 95–100.

[31] S. Panzeri and A. Treves, The upward bias in measures of information derived from limited data samples, Neural Comput. **7**, 399 (1995).

[32] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, Entropy and information in neural spike trains, Phys. Rev. Lett. **80**, 197 (1998).

[33] L. Paninski, Estimation of entropy and mutual information, Neural Comput. **15**, 1191 (2003).

[34] A. M. Arias and P. Hayward, Filtering transcriptional noise during development: Concepts and mechanisms, Nat. Rev. Genet. **7**, 34 (2006).

[35] T. C. Lacalli, Patterning, from conifers to consciousness: Turing's theory and order from fluctuations, Front. Cell Dev. Biol. **10**, 871950 (2022).

[36] C. H. Waddington, *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology* (Allen and Unwin, London, 1957).

[37] M. D. Petkova, S. C. Little, F. Liu, and T. Gregor, Maternal origins of developmental reproducibility, Curr. Biol. **24**, 1283 (2014).

[38] P. L. Ferree, V. E. Deneke, and S. D. Talia, Measuring time during early embryonic development, Semin. Cell. Dev. Biol. **55**, 80 (2016).

[39] G. Tkačik, A. M. Walczak, and W. Bialek, Optimizing information flow in small genetic networks, Phys. Rev. E **80**, 031920 (2009).

[40] A. M. Walczak, G. Tkačik, and W. Bialek, Optimizing information flow in small genetic networks. II. Feed-forward interaction, Phys. Rev. E **81**, 041905 (2010).

[41] T. R. Sokolowski, T. Gregor, W. Bialek, and G. Tkačik, Deriving a genetic regulatory network from an optimization principle, arXiv:2302.05680.

[42] T. Doan, A. Mendez, P. B. Detwiler, J. Chen, and F. Rieke, Multiple phosphorylation sites confer reproducibility of the rod's single-photon responses, Science **313**, 530 (2006).