Author for correspondence:
Mikhail Tikhonov
e-mail: tikhonov@fas.harvard.edu

Only accessible information is useful: insights from gradient-mediated patterning

Mikhail Tikhonov[1,3,4,5], Shawn C. Little[1,2] and Thomas Gregor[1,2]

[1]Joseph Henry Laboratories of Physics, and [2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
[3]Harvard Center of Mathematical Sciences and Applications, [4]Harvard John A. Paulson School of Engineering and Applied Sciences, and [5]Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, MA 02138, USA

(iD) MT, 0000-0002-9558-1121

Information theory is gaining popularity as a tool to characterize performance of biological systems. However, information is commonly quantified without reference to whether or how a system could extract and use it; as a result, information-theoretic quantities are easily misinterpreted. Here, we take the example of pattern-forming developmental systems which are commonly structured as cascades of sequential gene expression steps. Such a multi-tiered structure appears to constitute sub-optimal use of the positional information provided by the input morphogen because noise is added at each tier. However, one must distinguish between the total information in a morphogen and information that can be usefully extracted and interpreted by downstream elements. We demonstrate that quantifying the information that is *accessible* to the system naturally explains the prevalence of multi-tiered network architectures as a consequence of the noise inherent to the control of gene expression. We support our argument with empirical observations from patterning along the major body axis of the fruit fly embryo. We use this example to highlight the limitations of the standard information-theoretic characterization of biological signalling, which are frequently de-emphasized, and illustrate how they can be resolved.

1. Introduction

As an inspiring example of productive collaboration between computer science, physics and biology, information theory is

gaining popularity as a tool to characterize performance of biological systems. Although it may not have become the 'general calculus for biology', as predicted by Johnson in his 1970 review [1], the scope of its applications has been steadily expanding: from the earliest work measuring the information content in DNA, RNA and proteins to topics like neuroscience, collective behaviour, ecology, developmental biology, genetic regulation and signalling [2–5].

Specifically in the context of biochemical signalling, several recent reviews make compelling arguments that the mutual information between input and output of a signalling pathway is not just a useful quantity but is in fact the 'only natural framework' for characterizing the performance of such systems. However, implicit in these arguments is the assumption that the 'output' in question is the final target of signalling, the functionally relevant phenotypic trait. Unfortunately, in biological applications of information theory, the information content is usually assessed for signals that constitute intermediate steps, most commonly transcription factors, for example, NF-κB [6,7] or *Drosophila* patterning cues [8,9]. Such signals, however, still need to be interpreted by downstream processes. Therefore, the information they carry is useful only to the extent that it can be extracted and used by the system. As we will demonstrate, failure to recognize this can easily cause information-theoretic quantities to be misinterpreted.

To show this, we take the example of gradient-mediated patterning circuits in embryonic development. For a complex multicellular organism, the reliability of its developmental programme directly determines the probability of reaching reproductive age; therefore, low error rate and/or high error tolerance are likely to be key determinants of the structures of developmental circuits [10,11].

It thus seems surprising that, as we discuss below, many patterning circuits are structured as a cascade of several signalling steps, each of which is susceptible to loss of information due to noise inherent in biological control. We will see that treating information content of patterning cues as a one-size-fits-all method to characterize system performance erroneously predicts that a single-step readout strategy should be dominant in development. We will show that to understand the advantages of the multi-tiered architectures observed in real systems, it is essential to distinguish between the total information in a morphogen and the information that can be usefully extracted and interpreted. We support our reasoning with experiments on the well-studied segmentation gene network responsible for anterior–posterior (AP) patterning in the *Drosophila* embryo.

In many developing embryonic systems, cellular identities are conferred by graded input signals that induce dose-dependent gene expression programmes as outputs [12,13]. Such graded inputs, termed morphogens, often function as diffusible molecules produced by a localized expression source [14,15]. Localized expression generates concentration gradients in a field of otherwise naive and identical cells (presented in simplified form as a one-dimensional array in figure 1). Cells activate specific expression programmes in response to the local morphogen concentration $\hat{c}$. When $\hat{c}$ correlates closely with distance x from the source, such gradients carry a large amount of 'positional information' [16] quantified via the mutual information $I[\hat{c}, \hat{x}]$ (here and everywhere below, the 'hat' notation refers to random variables) [8,17]. In principle, a morphogen gradient carrying sufficient information could induce in each cell the gene expression programme appropriate for its position, thus generating the required spatial arrangement of cell fates [18] (figure 1a). In the most straightforward model, assuming the input morphogen is sufficiently reproducible [19], local morphogen concentration could be directly interpreted by each cell, i.e. the local input would activate all genes required at a given position, with no additional cycles of gene expression modulation. A central tenet of information theory, the information processing inequality, states that each transmission or processing step can only reduce the total information contained in a signal. Direct decoding might therefore be expected to dominate in early development as the optimal strategy for transmitting positional information. This expectation seems all the more valid given the widespread observation that the processes of transcription and translation exhibit considerable intrinsic variability, or noise [20,21]. Thus information loss in gene regulatory processes should be particularly notable.

Therefore, from the perspective of information theory, it is surprising that many gradient-based systems exhibit a multi-tiered architecture in which reiterated cycles of transcription and translation are required to attain patterning goals (illustrated in figure 1b). For example, in the vertebrate central nervous system, the unpatterned neuroectoderm exhibits a graded distribution of multiple diffusible signalling molecules. These signals subdivide the prospective brain into relatively large fore-, mid- and hindbrain territories, which are then segmented into smaller subunits by additional signalling activity [22–24]. Similar patterns of broad subdivision followed by short-range refinement are found during the specification the vertebrate neural crest by reiterated rounds of extracellular signalling [25]; in the formation of segmented muscle precursors (somites) by FGF and Notch followed by short-range
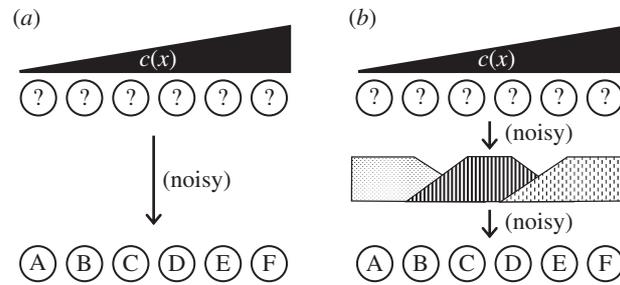
Figure 1. Direct versus multi-tiered decoding strategies for gradient-mediated patterning. (*a*) Direct decoding: to reduce noise introduced by intrinsically variable gene expression, patterning proceeds through a single cycle of transcription and translation. Differences in morphogen input $c(x)$ directly specify gene expression programmes A–F along axis x. (*b*) Multi-tiered decoding: morphogen first elicits expression of short-range diffusible factors in domains spanning several cells. These gene products then induce programmes A–F through a second cycle of transcription/translation. The added step introduces additional gene expression noise, reducing patterning information compared to direct decoding (*a*).

Ephrin activity [26,27]; in the dorsal–ventral patterning of the *Drosophila* body axis, first by a gradient of NF-κB activity (also called Dorsal) and then by members of the BMP family of secreted signalling molecules [28,29]; and also in the fruit fly, in the patterning of the AP axis by gradients of diffusible transcription factors within the shared cytoplasm of the nuclear syncytium [18,30,31].

These examples and others illustrate a common theme where long-range signalling gradients subdivide a large field into smaller domains, within which the patterned expression of secondary factors establishes elaborated patterns (figure 1*b*). Since each cycle of transcription and translation introduces more noise, the widespread use of the multi-tiered architecture appears to conflict with the expectation that development should favour circuits exhibiting efficient information utilization.

This apparent conflict arises because Shannon's information content of a signal [17] has two important limitations. First, the information content of a patterning cue or other biological signal is defined locally in space and time, whereas its interpretation is non-local, and instead occurs over time and frequently involves diffusive signals. For this reason, the naive application of information processing inequality in these systems is incorrect, and the local, instantaneous information content in a signal does not in fact provide an upper bound for the performance of downstream processes interpreting this signal [7,32,33]. Second, the same amount of information can be encoded in formats that are more or less easy for the system to access, since the interpreting circuit is itself subject to noise. Thus, the local information content of a signal is neither an upper bound nor a fair estimate of the amount of information this signal can 'transmit' to the downstream circuit. This is well illustrated by the recent experimental work on ERK, calcium and NF-κB pathways [7]. If the output of any of these pathways is reduced to a single scalar, it is found to transmit very little information about the input. If the output is treated as a dynamical variable, its apparent information content increases considerably [32]. Neither of these quantities, however, can be interpreted before it is established what fraction of that information can actually be extracted and used by the system. Here, we use a simplified model to illustrate these limitations of what we call 'raw' information content, contrasting it with 'accessible information' that we introduce. We argue that assessing the usefulness of a signal must always take into account the so-called input noise [34] of the downstream circuit interpreting this signal.

2. Results

2.1. Responding to a graded input signal

Consider a one-dimensional array of cells i located at positions $(0 < x_i < L)$ and exposed to a noisy linear gradient of an input morphogen $\hat{c}$ spanning the range $[0, c_{\max}]$. To build intuition, we will assume the noise of input $\hat{c}$ to be Gaussian, of constant magnitude σ_0, and uncorrelated between cells[1]:

[1]The assumption of uncorrelated noise is intentionally strong. In a real system, correlated noise can be introduced, for example, by variations in the total amount of morphogen deposited maternally. These fluctuations, which cannot be reduced by averaging, lead to imperfect *reproducibility* of morphogen activity at a given location across multiple embryos. Much work has focused on investigating the limitations imposed on patterning by this type of fluctuations [8,35,36]. In contrast, our model is applicable for understanding the effects of imperfect *precision* of gene expression (at a given location within the same embryo). The distinction between 'raw' and 'accessible' information does not rely on the assumption of uncorrelated noise.
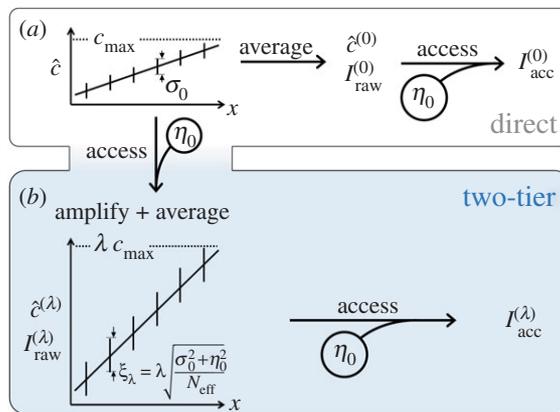
Figure 2. The two patterning strategies. (*a*) In the direct strategy, target genes are controlled directly by averaged morphogen $\hat{c}$ with only one 'access' operation. (*b*) The two-tier strategy involves a second patterning factor $\hat{c}^{(\lambda)}$; target genes are separated from the input by two tiers of 'access' operations. I_{raw}, raw information content of the controlling morphogen; I_{acc}, accessible information.

$\hat{c} = (x/L)\,c_{\mathrm{max}} + \hat{\sigma}$, where σ_i is a Gaussian random variable of variance σ_0^2 (figure 2*a*). Cells respond to morphogen $\hat{c}$ by modulating gene expression through intrinsically noise-prone signal transduction and regulation processes. We will model this response as a composition of three steps, three elementary operations that constitute the 'toolkit' with which cells can access and process information contained in patterning cues: *access*, *amplify* and *average*.

Let g^{out} be a gene product whose expression is controlled by $\hat{c}$. The simplest readout is achieved by placing gene g^{out} under the control of a promoter that is responsive to $\hat{c}$ and by accumulating the output protein for some time τ. In our model, we express the amount of g^{out} produced during this time by a cell as $\hat{g}^{\mathrm{out}} = F(\hat{c}^{\mathrm{est}})$, where $\hat{c}^{\mathrm{est}}$ is a noisy estimate of the true concentration $\hat{c}$ that the system could obtain in time τ ('access'), and F is some deterministic input–output function ('amplify'); for simplicity, we first consider F to be pure linear amplification with coefficient λ, denoted F_λ. The 'access' operation (a noisy estimate of the input) is the key element of our framework. Specifically, we write

$$\hat{c}^{\mathrm{est}} = \hat{c} + \hat{\eta},$$

where $\hat{\eta}$ reflects the intrinsic stochasticity of transcriptional regulation due to promoter switching, random arrival of molecules and, in principle, many other noise sources. In our simplified framework, this 'input noise' [34] is the only type we consider, and we will model $\hat{\eta}$ as a Gaussian random variable of variance η_0^2. In other words, we postulate that each 'access' operation takes time τ and comes at the price of corrupting the signal with extra noise of magnitude η_0.

The final toolkit operation is averaging. Because patterning systems typically act over durations that are long (hours) compared to the time required to synthesize mRNA and protein (minutes), cells can perform temporal averaging by allowing stable gene products to accumulate [37]: if T is the time available for patterning, the system can effectively perform T/τ access operations. In addition, the production of soluble factors that can be shared between cells gives rise to spatial averaging [37,38]. Conveniently, for linear expression profiles, spatial averaging affects only the fluctuations, leaving the mean shape intact except at the boundary of the patterned region; for weak averaging, boundary effects can be neglected for our purposes. Both types of averaging offer the system some capacity to perform multiple measurements of the input, which we capture formally by an averaging operator $G_{N_{\mathrm{eff}}}$. Here, N_{eff} indicates the effective number of independent measurements, so that application of $G_{N_{\mathrm{eff}}}$ to a morphogen, by definition, reduces expression fluctuations by a factor $1/N_{\mathrm{eff}}$.

We distinguish between two patterning strategies. In the first ('direct strategy'; figure 2*a*), cell-fate-specific target genes are controlled directly by $\hat{c}$ and no other patterning factors are involved. Any available averaging mechanisms are applied to $\hat{c}$ itself. In the second ('two-tier') strategy, cells perform an amplifying readout of $\hat{c}$ with input–output function F_λ to establish a spatial profile of a second factor $\hat{c}^{(\lambda)}$ (figure 2*b*). The pattering time T is spent on accumulating and averaging $\hat{c}^{(\lambda)}$. Mathematically,

in the two scenarios, the cell-fate-specific target genes are controlled by

$$\hat{c}^{(0)} = G_{N_{\text{eff}}}[\hat{c}] = \frac{x}{L}c_{\text{max}} + \frac{\hat{\sigma}}{\sqrt{N_{\text{eff}}}} \qquad \text{(direct strategy)} \tag{2.1}$$

and
$$\hat{c}^{(\lambda)} = G_{N_{\text{eff}}}[F_\lambda(\hat{c} + \hat{\eta})] = \frac{x}{L}\lambda c_{\text{max}} + \lambda\frac{\hat{\sigma} + \hat{\eta}}{\sqrt{N_{\text{eff}}}} \quad \text{(two-tier strategy)} \tag{2.2}$$

We now ask: when, if ever, does the noisy amplification step of the two-tier strategy provide a benefit to the system?

2.2. Standard information-theoretic considerations do not explain the benefits of amplification

For a linear morphogen $\hat{c}$ with dynamic range c_{max} and noise σ_0, the standard positional information $I[\hat{c}, \hat{x}]$, which we will call its 'raw information content', is given by

$$I_{\text{raw}}[\hat{c}, \hat{x}] = \ln\left(\frac{c_{\text{max}}}{\sigma_0\sqrt{2\pi e}}\right)$$

(this expression assumes that noise is small and the number of cells is large; see the electronic supplementary material for details). It depends only on the ratio $\phi = c_{\text{max}}/\sigma_0$; for convenience, we define $\mathcal{I}(\phi) \equiv \ln(\phi/\sqrt{2\pi e})$, which is an increasing function of ϕ.

Let us compare the two patterning strategies from the point of view of the raw information content carried by the controlling signal. In the direct strategy (2.1), the application of $G_{N_{\text{eff}}}$ reduces the noise to $\sigma_0/\sqrt{N_{\text{eff}}}$ and so the controlling signal $\hat{c}^{(0)}$ carries $I_{\text{raw}}^{(0)} = \mathcal{I}(c_{\text{max}}/(\sigma_0/\sqrt{N_{\text{eff}}}))$ bits of raw information. In the two-tier strategy (2.2), the amplified profile $\hat{c}^{(\lambda)}$ is characterized by noise $\xi_\lambda = \lambda\sqrt{(\sigma_0^2 + \eta_0^2)/N_{\text{eff}}}$, and its raw information content is therefore

$$I_{\text{raw}}^{(\lambda)} = \mathcal{I}\left(\frac{\lambda c_{\text{max}}}{\xi_\lambda}\right) = \mathcal{I}\left(c_{\text{max}}\sqrt{\frac{N_{\text{eff}}}{\sigma_0^2 + \eta_0^2}}\right) < I_{\text{raw}}^{(0)}. \tag{2.3}$$

Averaging mitigates the loss of positional information caused by noisy readout [38]. If N_{eff} is sufficiently large, the amplified-and-averaged profile carries even more information than the original input. (Note that the information-processing inequality is not violated, as it states only that the output cannot carry more information than N_{eff} independent copies of the input.) Nevertheless, applying averaging directly to the input (the direct strategy) always yields more raw information; thus, the multi-step scenario appears inferior to a direct readout.

In real systems, the three operations we treat as independent may be mechanistically linked. For example, if $\hat{c}$ is an intracellular factor while spatial averaging requires a small diffusible molecule, then performing an extra readout can provide access to an otherwise unavailable averaging mechanism. By assuming that the two strategies (2.1) and (2.2) can benefit from equal amounts of averaging, which in our model simply reduces expression noise and is obviously beneficial, we can focus specifically on the effect of signal amplification. Multi-tier patterning proceeds through rounds of amplification: small differences in input result in large differences in gene expression so as to establish increasingly sharp boundaries delimiting expression domains [39], yet in our expression (2.3) for the information content of the amplified profile $\hat{c}^{(\lambda)}$, the amplification factor λ cancels out. Thus, considerations based on raw information content fail to explain the prevalence of signal amplification.

2.3. The benefits of the multi-tiered strategy lie in making the 'raw' information more accessible

The benefits of amplification and the advantages of the multi-tier strategy become clear when we observe that, due to the intrinsic noise in the regulatory readout, the raw information content is an inadequate measure of a morphogen's usefulness to the system. The purpose of a morphogen is to activate downstream processes; the relevant quantity is therefore not the amount of information a morphogen carries, but the amount of information it can transmit to its downstream targets. Since biological control is intrinsically noisy, the two quantities are distinct.

Our model was designed to make this particularly clear: since the system can never access the true concentration of any signal $\hat{c}_\star$, but only its noisy estimate $\hat{c}_\star^{\text{est}}$, $I_{\text{raw}}[\hat{c}_\star]$ is beyond the system's reach.
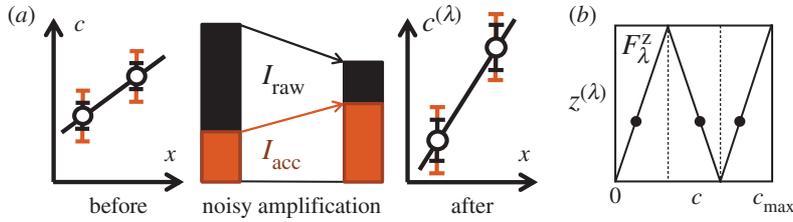
Figure 3. (*a*) Noisy amplification can increase accessible information even if raw information is reduced. Inner error bars are the signal variability and increase when amplification adds new noise, reducing I_{raw}. Outer error bars represent the signal observed by the noisy cell machinery (corrupted by noise η_0). After amplification, the relative importance of η_0 is reduced, increasing I_{acc}. (*b*) The 'segmentation' input–output function F_λ^z for integer λ (here $\lambda = 3$) preserves the dynamic range of morphogen concentration. Locations such as those indicated by dots now have identical expression levels of $z^{(\lambda)}$ (the *y*-axis), but can be distinguished using the input morphogen c (the *x*-axis on this plot).

We define *accessible* information I_{acc} in a morphogen as the amount of information the system can access in time τ:

$$I_{acc}[\hat{c}_\star] \equiv I_{raw}[\hat{c}_\star^{est}] \equiv I_{raw}[\hat{c}_\star + \hat{\eta}], \tag{2.4}$$

where $\hat{\eta}$ (the input noise of the downstream readout circuit), again, is a Gaussian noise of magnitude η_0 within our model.

The amount of accessible information provided by the direct strategy (equation 2.1) is given by

$$I_{acc}[\hat{c}^{(0)}] = I_{raw}\left[\frac{x}{L}c_{max} + \frac{\hat{\sigma}}{\sqrt{N_{eff}}} + \hat{\eta}\right] = \mathcal{I}\left(\frac{c_{max}}{\sqrt{\frac{\sigma_0^2}{N_{eff}} + \eta_0^2}}\right). \tag{2.5}$$

For the amplified profile $c^{(\lambda)}$ (equation 2.2), a similar calculation yields

$$I_{acc}[\hat{c}^{(\lambda)}] = \mathcal{I}\left(\frac{\lambda c_{max}}{\sqrt{\lambda^2 \frac{\sigma_0^2 + \eta_0^2}{N_{eff}} + \eta_0^2}}\right) = \mathcal{I}\left(\frac{c_{max}}{\sqrt{\frac{\sigma_0^2 + \eta_0^2}{N_{eff}} + \frac{\eta_0^2}{\lambda^2}}}\right). \tag{2.6}$$

In this expression, the intrinsic noise η_0 of regulatory readout enters twice, corresponding to the two 'access operations' required in the two-tier strategy (figure 2).

The amplification factor λ no longer cancels out in (2.6); amplifying dynamic range is beneficial, since it reduces the relative importance of the intrinsic readout noise (figure 3*a*). Comparing (2.5) and (2.6), we find that the extra tier of noisy amplification is beneficial if and only if

$$\eta_0^2 \left(1 - \frac{1}{N_{eff}} - \frac{1}{\lambda^2}\right) > 0 \tag{2.7}$$

Note that the condition (2.7) is never satisfied if $N_{eff} = 1$ (no averaging) or $\lambda = 1$ (no amplification). Intuitively, our argument demonstrates that the patterning system is a mechanism that invests some effort into making a careful measurement ($N_{eff} > 1$) and encodes this information in a more accessible format where steeper concentration changes ($\lambda > 1$) can be interpreted with a faster, and therefore noisier readout. This mechanism is useful precisely because regulatory readout is intrinsically noisy, otherwise direct readout would have been the better strategy. In other words, to understand the purpose of the patterning system, it is essential to distinguish between the total information in a morphogen and information that can be usefully extracted and interpreted.

2.4. Multiple tiers can improve gradient interpretation even when raw information decreases

So far, we considered the information content (raw or accessible) in each tier separately. However, in principle, downstream processes could access all patterning cues and not simply the final tier [40,41]. As a result, extra readout tiers can be beneficial even when they carry very little information on their own.

To see this, consider the input–output function F_λ^z depicted in figure 3*b*. In some respects, it is more realistic than the purely amplifying linear readout F_λ considered above, since real patterning systems must operate within a limited global dynamic range of morphogen concentrations. Let $\hat{z}^{(\lambda)}$ be the

morphogen profile established by the new F_λ^z-shaped readout of c:

$$\hat{z}^{(\lambda)} = G_{N_{\mathrm{eff}}}[F_\lambda^z(\hat{c} + \hat{\eta})] = \bar{z}^{(\lambda)}(x) + \lambda \frac{\hat{\sigma} + \hat{\eta}}{\sqrt{N_{\mathrm{eff}}}};$$

here $\bar{z}^{(\lambda)}(x)$ is the zigzag-shaped mean profile (cf. figure 3*b*; once again, we assume for simplicity that the spatial component of averaging operator is sufficiently weak that its smoothing effect on the mean profile shape can be neglected; for example, we can assume that averaging is predominantly temporal). This morphogen $\hat{z}^{(\lambda)}$ has noise magnitude ξ_λ (same as the noise in $\hat{c}^{(\lambda)}$), but is folded onto itself λ times, reminiscent of the spatially reiterated expression of genes involved in *Drosophila* axis segmentation. Repeatedly using the same output values at multiple positions naturally reduces mutual information between the output concentration and position:

$$I_{\mathrm{raw}}[\hat{z}^{(\lambda)}] = I_{\mathrm{raw}}[\hat{c}^{(\lambda)}] - \ln \lambda$$

and

$$I_{\mathrm{acc}}[\hat{z}^{(\lambda)}] = I_{\mathrm{acc}}[\hat{c}^{(\lambda)}] - \ln \lambda.$$

However, the λ locations with identical concentrations of $\hat{z}^{(\lambda)}$ are made distinguishable by the original morphogen $\hat{c}$ (figure 3*b*; this statement invokes the small-noise assumption; see the electronic supplementary material). Therefore, under the small-noise assumption, the *joint* information that the original and the amplified profiles together provide about a cell's location is the same for F_λ^z as it was for F_λ:

$$I[\{\hat{c}, \hat{z}^{(\lambda)}\}, x] = I[\{\hat{c}, \hat{c}^{(\lambda)}\}, x].$$

Replacing information content of a single profile by this joint information, our argument demonstrating that amplification increases accessible information can now be repeated verbatim[2], and we again find that the extra readout is beneficial as long as (2.7) is satisfied (see electronic supplementary material for more detail). Note, however, that on its own, $\hat{z}^{(\lambda)}$ may carry *less* information than the original morphogen $\hat{c}$. The easiest way to see this is to compare their noise levels:

$$\left(\frac{\xi_\lambda}{\sigma_0}\right)^2 = \frac{\lambda^2}{N_{\mathrm{eff}}}\left(1 + \frac{\eta_0^2}{\sigma_0^2}\right)$$

If the effect of amplification is stronger than that of averaging, we find $\xi_\lambda/\sigma_0 > 1$. In this scenario, the amplified profile $\hat{z}^{(\lambda)}$ has the same dynamic range but lower precision than the original morphogen $\hat{c}$, and therefore, on its own, carries less information (whether raw or accessible). This shows that evaluating the usefulness of a particular cue from information-theoretic standpoint can lead to misleading results, unless all other relevant cues (which are often hard to establish) are taken into account simultaneously. Here, we demonstrated that systems can benefit from multi-tiered interpretation even in cases where intermediate amplification steps occur at a net loss of information, increasing noise.

2.5. The multi-tier structure of *Drosophila* segment patterning increases information accessibility

To illustrate our theoretical point in a real system, we consider the AP axis patterning of *Drosophila*. In this system, segmentation of the AP axis proceeds through four tiers of gene activity, termed maternal gradients, gap genes, pair-rule genes and segment polarity genes [31]. The sequential activity of each tier subdivides the naive blastoderm into smaller domains of gene expression with increasingly sharp boundaries, culminating in the designation of each row of cells with its own unique set of expressed genes (figure 4*a*). This process is subject to transcriptional noise with a large intrinsic component [37], as well as several other noise sources with different signatures [34,42–44]. No single value of η_0 adequately characterizes such readout noise. Nevertheless, we can gain important insight by computing accessible information $I_{\mathrm{acc}}^{\eta_0}[c_\star]$ as a function of η_0, treating it as a variable parameter: the decay of $I_{\mathrm{acc}}^{\eta_0}[c_\star]$ with η_0 characterizes the tolerance to added noise of the information encoded in a morphogen (or set of morphogens) $c_\star$. Applied to gene expression data from the early *Drosophila* segmentation gene network, this analysis will show how our simple model explains the use of multi-tier gradient interpretation in a real system (figure 4).

We focus on a particular node in this network whereby, in early embryos, two gap genes, *hb* and *Kr*, regulate a pair-rule gene *eve*. For $0.37 < x_{\mathrm{AP}} < 0.47$, where *Kr* and *hb* expression form opposing

---

[2]For multiple profiles $\{c^{(1)}, c^{(2)}, \ldots\}$, we define accessible information as the joint information content in the set of morphogen profiles, independently corrupted with noise of magnitude η_0 (compare with equation (2.4)): $I_{\mathrm{acc}}(\{c^{(1)}, c^{(2)}, \ldots\}) \equiv I_{\mathrm{raw}}(\{c^{(1)} + \eta^{(1)}, c^{(2)} + \eta^{(2)}, \ldots\})$.
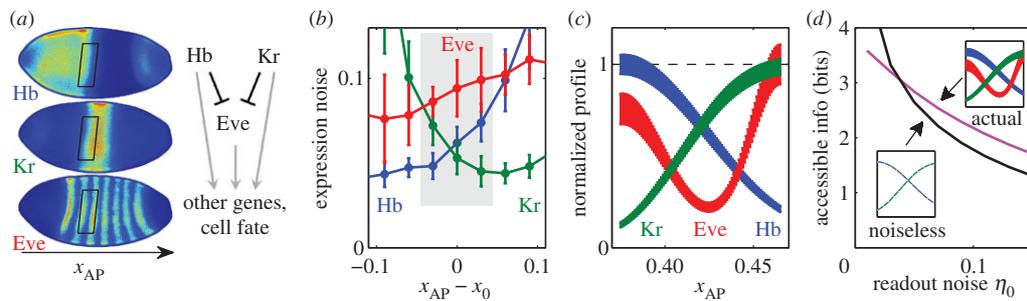
Figure 4. (a) Immunostaining of three AP axis patterning genes in the same embryo. Rather than specifying cell fate directly, the 'gap genes' such as *hunchback* (Hb; top) and *Krüppel* (Kr; middle) control 'pair-rule' genes such as *even-skipped* (Eve, bottom). Both tiers regulate other genes further downstream. Boxes indicate the selected ROI, where at this time, Hb and Kr are the only relevant inputs to Eve, as shown in the cartoon. (b) Within the ROI (shaded), Eve exhibits higher expression noise than either Hb or Kr. Expression noise computed as RMS difference between expression level of a nucleus and its immediate dorsal or ventral neighbour (see the electronic supplementary material), plotted against AP distance from the Hb/Kr boundary (denoted x_0). Error bars are standard deviation over $N = 8$ embryos. (c) Idealized morphogen profiles, restricted to the ROI. Profile shape obtained as smooth spline-fit to expression values and noise magnitudes calculated for the profiles of panel (a) after projection onto the AP axis. (d) For all but the lowest readout noise magnitude, joint accessible information content in the triplet (Hb, Kr, Eve) exceeds the accessible information provided by Hb and Kr alone, even in an extreme hypothetical case when they are rendered entirely noiseless.

boundaries, they are jointly responsible for creating the trough between *eve* stripes 2 and 3; other inputs to *eve* are negligible in this region at this time [45,46]. Protein levels are measured simultaneously in each nucleus by a triple immunostaining experiment (figure 4a) in $N = 8$ single embryos. We determine the expression noise of each gene by comparing levels in a given nucleus with those of its immediate dorsal and ventral neighbours (see the electronic supplementary material).

In the defined region of interest (ROI), *eve* expression noise is higher than the respective noise in *hb* or *Kr* expression (figure 4b). The information content of *eve* must therefore be lower than that carried by either of its two inputs. Due to the curvature of the embryo (figure 4a), the positional information of a real morphogen is only approximately related to that derived from projection onto the imaginary AP axis. Therefore, to estimate the information content for each of the three genes, we consider 'idealized' Gaussian-noise profiles (figure 4c) with mean and noise obtained by smoothing the measured values in real embryos. This approach neglects correlations in the fluctuations of these genes which may be significant in practice [47]; this will not be an issue, as detailed below. The idealized profiles are normalized to the same maximum and are, by construction, functions of x_{AP} carrying positional information $I(c(x_{AP}), x_{AP})$. Restricted to the ROI, the information content of Hb and Kr is, respectively, 2.6 and 2.7 bits, whereas the larger noise of Eve reduces its information content to only 2.0 bits. Why, then, does the system use Eve to regulate downstream processes, rather than utilizing Kr and Hb directly?

The answer becomes clear when we consider the accessibility of information encoded in these morphogens, namely $I_{acc}^{\eta_0}$ as a function of η_0 (figure 4d). A patterning strategy lacking Eve can access only Hb and Kr. Even if some hypothetical filtering mechanism could reduce their expression noise to arbitrarily low level (in particular, making the correlation of noise between the two genes irrelevant), the readout noise magnitude $\eta_0 > 0$ imposes an upper bound that $I_{acc}^{\eta_0}[c_{Hb}, c_{Kr}]$ must satisfy. This corresponds to the information in a hypothetical pair of noiseless Hb and Kr and cannot be achieved in practice; it is a theoretical best-case scenario for any strategy lacking Eve.

When the readout noise η_0 is zero, $I_{acc}^{\eta_0}$ coincides with the raw information content, which for perfectly noiseless Hb and Kr would be infinite. However, as readout noise increases, the performance bound becomes finite and drops quickly (figure 4d, black curve). This behaviour contrasts with the joint accessible information of the triplet (Hb, Kr, Eve) (magenta curve) as calculated for the idealized profiles using their actual measured noise. The accessible information content in the triplet is, of course, always finite, but it is also more tolerant to readout noise: due to the steeper slopes of the Eve profile, as η_0 increases, the accessible information content of the triplet (Hb, Kr, Eve) decreases slowly: importantly, more slowly than the black curve. Therefore, a crossing point is observed, whose presence does not qualitatively depend on the specifics of the readout noise model (e.g. absolute noise magnitude can be replaced by fractional). Remarkably, although Eve is measurably noisier than either of its inputs, its presence enables the system to access more information than could have been extracted from Hb and Kr alone, even if these inputs could be rendered perfectly noiseless. In practice, the enhancers of the

pair-rule genes also contain binding sites for maternal transcription factors [40,41], which may lead to a further increase in the precision of gene expression. However, our framework demonstrates that even if Eve were regulated by Hb and Kr only, and so were fully redundant in the standard information-theoretic sense, the additional tier would still confer an advantage, because transcriptional regulation is intrinsically noisy.

3. Discussion

The *Drosophila* patterning network has been described as performing a 'transition from analogue to digital specification' of cell identity [39]. The 'digital' metaphor has its limitations: even for Eve, the graded distribution within gene expression domains contains information [8]; nevertheless, it expresses the correct intuition that the final pattern is more tolerant to noise. Importantly, the standard information-theoretic formalism does not capture this intuition: for instance, the profile depicted in figure 3*b* has the *same* information content for all λ. Noise tolerance—a critically important feature in biological systems—becomes manifest only when the readout process is considered explicitly, for example, as we have done in our definition of accessible information. This point is implicit in the theoretical work investigating the so-called 'input noise' [34], but has not been emphasized. This is because in a theoretical discussion of an abstract biochemical circuit, the quantities for which information is computed are easily postulated to be the complete input and the final output; in this manner, valid theoretical results can be derived without a concern for information accessibility (for some recent examples, see [48,49]). However, when information-theoretic arguments are applied to experimental data where the measured quantity is only an intermediate step, e.g. a transcription factor regulating downstream events, the question of information accessibility (the unavoidable input noise of the downstream circuit) can no longer be neglected.

For example, it has been suggested that certain signalling circuits may have evolved towards optimal information transmission [4,5]. Although the argument is plausible, applying it in practice requires two conditions. First, information transmission must be calculated from the input to the entire set of functional genes; in the case of developmental circuits, this means hundreds of cell-fate-specific targets. Second, its optimization must be performed under some 'bounded complexity' constraint: presumably, information transmission could be enhanced if all regulatory elements were as complex as gap genes enhancers of *Drosophila* with combinatorial, cooperative regulation.

Unfortunately, considering hundreds of genes is intractable, and the constraint on regulatory complexity is hard to formulate. The usual, more economical approach to patterning recognizes that the bulk of the patterning task is accomplished by only a small subset of dedicated, cross-regulated genes that establish the pattern that all other genes can then interpret simply. If we focus only on this core subset, the problem becomes tractable, and the 'economy of complexity' constraint is conveniently imposed by construction. We must, however, realize that maximizing information transmission to the target genes (downstream of the patterning core) imposes a different requirement than merely efficient information transfer within the core itself. Instead, the core circuit must function as a format converter, re-encoding information at its input into a format that can be accessed with a simpler and faster readout, that of a patterning cue by a functional gene.

Curiously, it has been shown that in small networks with a realistic model of noise, maximizing raw information transmission leads to network structures exhibiting features such as tiling of patterned range with amplifying input/output readouts [50–52], i.e. features that tend to also make information more accessible, even though the optimization scheme employed in these studies did not specifically consider the encoding format. This remarkable coincidence, however, should not obscure the fact that ultimately the two tasks—maximizing information transmission and re-encoding it in a more accessible format—could be conflicting.

Information theory is a powerful tool; its formalism does not, however, aim to replace considerations of what constitutes useful information or how it might be used by the system. As it is gaining popularity in biological applications, it is important to remember that for a channel $X \mapsto Y$, the relation between mutual information $I(X, Y)$ and the ability to use Y to determine X is only asymptotic: Shannon [17] proved that it is the maximum rate of error-free communication via this channel, *in the limit of infinite uses* of the channel. Importantly, in development and biological signalling, the number of channel uses (e.g. integration time of the signal) is fundamentally finite [3]. Further, Shannon's results assumed an encoder/decoder of infinite computational power [17]. This asymptotic rate is never in fact achieved in practice [53], but in biological context, performance is constrained even further, since the 'encoding

scheme' is usually limited to measuring the same signal multiple times. In communication theory, this is called a 'repetition code' and is formally classified as a 'bad code', i.e. a code that does not attain Shannon's bound even asymptotically. This means that extracting all the 'raw' information from a signal is impossible even in principle. For example, a signalling pathway with capacity of 1 bit is never sufficient to make a reliable binary decision [3], and therefore should not be conceptualized as a binary switch.

Taken narrowly, the results presented here explain an architectural property shared by multiple patterning circuits responding to a graded signal. Perhaps more importantly, the example we construct highlights a general theoretical point: assessing the usefulness of a signal must always take into account the input noise of the downstream circuit interpreting this signal. Here, this type of noise manifests itself in the distinction that we draw between 'raw' and 'accessible' information. Our definition of the latter relied on a simplistic noise model; in general, encoding the effects of input noise in a re-defined notion of information content is not possible. In general, quantifying the usefulness of information-bearing signals in contexts where channel uses are limited will require reinstating considerations of rate/fidelity trade-off, which Shannon could eliminate by taking the limit of infinite-time communication. Nevertheless, information theory remains a most adequate framework to address these issues, provided its limitations are understood.

References

1. Johnson HA. 1970 Information theory in biology after 18 years. *Science* **168**, 1545–1550. (doi:10.1126/science.168.3939.1545)

2. Waltermann C, Klipp E. 2011 Information theory based approaches to cellular signaling. *Biochim. Biophys. Acta* **1810**, 924–932. (doi:10.1016/j.bbagen.2011.07.009)

3. Bowsher CG, Swain PS. 2014 Environmental sensing, information transfer, and cellular decision-making. *Curr. Opin. Biotechnol.* **28**, 149–155. (doi:10.1016/j.copbio.2014.04.010)

4. Levchenko A, Nemenman I. 2014 Cellular noise and information transmission. *Curr. Opin. Biotechnol.* **28**, 156–164. (doi:10.1016/j.copbio.2014.05.002)

5. Tkacik G, Bialek W. 2014 Information processing in living systems. arXiv:1412.8752v1 [q-bio.QM]. (http://arxiv.org/abs/1412.8752)

6. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A. 2011 Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358. (doi:10.1126/science.1204553)

7. Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, Tsimring L, Wollman R. 2014 Systems biology. Accurate information transmission through dynamic biochemical signaling networks. *Science* **346**, 1370–1373. (doi:10.1126/science.1254933)

8. Dubuis JO, Tkacik G, Wieschaus EF, Gregor T, Bialek W. 2013 Positional information, in bits. *Proc. Natl Acad. Sci. USA* **110**, 16 301–16 308. (doi:10.1073/pnas.1315642110)

9. Dubuis JO, Samanta R, Gregor T. 2013 Accurate measurements of dynamics and reproducibility in small genetic networks. *Mol. Syst. Biol.* **9**, 639. (doi:10.1038/msb.2012.72)

10. Hironaka K, Morishita Y. 2012 Encoding and decoding of positional information in morphogen-dependent patterning. *Curr. Opin. Genet. Dev.* **22**, 553–561. (doi:10.1016/j.gde.2012.10.002)

11. Lander A. 2013 How cells know where they are. *Science* **339**, 923–927. (doi:10.1126/science.1224186)

12. Rogers KW, Schier AF. 2011 Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* **27**, 377–407. (doi:10.1146/annurev-cellbio-092910-154148)

13. Nahmad M, Lander AD. 2011 Spatiotemporal mechanisms of morphogen gradient interpretation. *Curr. Opin. Genet. Dev.* **21**, 726–731. (doi:10.1016/j.gde.2011.10.002)

14. Wartlick O, Kicheva A, Gonzalez-Gaitan M. 2009 Morphogen gradient formation. *Cold Spring Harb. Persp. Biol.* **1**, a001255. (doi:10.1101/cshperspect.a001255)

15. Muller P, Rogers KW, Yu SR, Brand M, Schier AF. 2013 Morphogen transport. *Development* **140**, 1621–1638. (doi:10.1242/dev.083519)

16. Wolpert L. 1969 Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47. (doi:10.1016/S0022-5193(69)80016-0)

17. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656. (doi:10.1002/j.1538-7305.1948.tb01338.x)

18. Gergen JP, Coulter D, Wieschaus EF. 1986 Segmental pattern and blastoderm cell identities. In *Gametogenesis and the early embryo* (ed J. Gall), pp. 195–220. New York, NY: Alan R. Liss.

19. Gregor T, Tank DW, Wieschaus EF, Bialek W. 2007 Probing the limits to positional information. *Cell* **130**, 153–164. (doi:10.1016/j.cell.2007.05.025)

20. Munsky B, Neuert G, van Oudenaarden A. 2012 Using gene expression noise to understand gene regulation. *Science* **336**, 183–187. (doi:10.1126/science.1216379)

21. Sanchez A, Golding I. 2013 Genetic determinants and cellular constraints in noisy gene expression. *Science* **342**, 1188–1193. (doi:10.1126/science.1242975)

22. Lumsden A, Krumlauf R. 1996 Patterning the vertebrate neuraxis. *Science* **274**, 1109–1115. (doi:10.1126/science.274.5290.1109)

23. Pera EM, Acosta H, Gouignard N, Climent M, Arregi I. 2014 Active signals, gradient formation and regional specificity in neural induction. *Exp. Cell Res.* **321**, 25–31. (doi:10.1016/j.yexcr.2013.11.018)

24. Raible F, Brand M. 2004 *Divide et Impera* – the midbrain–hindbrain boundary and its organizer. *Trends Neurosci.* **27**, 727–734. (doi:10.1016/j.tins.2004.10.003)

25. Patthey C, Gunhaga L. 2014 Signaling pathways regulating ectodermal cell fate choices. *Exp. Cell Res.* **321**, 11–16. (doi:10.1016/j.yexcr.2013.08.002)

26. Saga Y. 2012 The mechanism of somite formation in mice. *Curr. Opin. Genet. Dev.* **22**, 331–338. (doi:10.1016/j.gde.2012.05.004)

27. Watanabe T, Takahashi Y. 2010 Tissue morphogenesis coupled with cell shape changes.

Curr. Opin. Genet. Dev. **20**, 443–447. (doi:10.1016/j.gde.2010.05.004)

28. Little SC, Mullins MC. 2006 Extracellular modulation of BMP activity in patterning the dorsoventral axis. *Birth Defects Res. C Embryo Today* **78**, 224–242. (doi:10.1002/bdrc.20079)

29. Rushlow CA, Shvartsman SY. 2012 Temporal dynamics, spatial range, and transcriptional interpretation of the Dorsal morphogen gradient. *Curr. Opin. Genet. Dev.* **22**, 542–546. (doi:10.1016/j.gde.2012.08.005)

30. Driever W, Nüsslein-Volhard C. 1988 A gradient of bicoid protein in *Drosophila* embryos. *Cell* **54**, 83–93. (doi:10.1016/0092-8674(88)90182-1)

31. Kornberg TB, Tabata T. 1993 Segmentation of the *Drosophila* embryo. *Curr. Opin. Genet. Dev.* **3**, 585–593. (doi:10.1016/0959-437X(93)90094-6)

32. Tostevin F, ten Wolde PR. 2009 Mutual information between input and output trajectories of biochemical networks. *Phys. Rev. Lett.* **102**, 218101. (doi:10.1103/PhysRevLett.102.218101)

33. Sokolowski TR, Tkacik G. 2015 Optimizing information flow in small genetic networks. IV. Spatial coupling. arXiv:1501.04015v1 [q-bio.MN]. (http://arxiv.org/abs/1501.04015)

34. Tkacik G, Gregor T, Bialek W. 2008 The role of input noise in transcriptional regulation. *PLoS ONE* **3**, e2774. (doi:10.1371/journal.pone.0002774)

35. Tkacik G, Dubuis JO, Petkova MD, Gregor T. 2015 Positional information, positional error, and read-out precision in morphogenesis: a mathematical framework. *Genetics* **199**, 39–59. (doi:10.1534/genetics.114.171850)

36. Petkova MD, Little SC, Liu F, Gregor T. 2014 Maternal origins of developmental reproducibility. *Curr. Biol.* **24**, 1283–1288. (doi:10.1016/j.cub.2014.04.028)

37. Little SC, Tikhonov M, Gregor T. 2013 Precise developmental gene expression arises from globally stochastic transcriptional activity. *Cell* **154**, 789–800. (doi:10.1016/j.cell.2013.07.025)

38. Erdmann T, Howard M, ten Wolde PR. 2009 Role of spatial averaging in the precision of gene expression patterns. *Phys. Rev. Lett.* **103**, 258101. (doi:10.1103/PhysRevLett.103.258101)

39. Gilbert SF. 2013 *Developmental biology*, 10th edn. Sunderland, MA: Sinauer Associates, Inc.

40. Li XY et al. 2008 Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6**, e27. (doi:10.1371/journal.pbio.0060027)

41. MacArthur S et al. 2009 Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* **10**, R80. (doi:10.1186/gb-2009-10-7-r80)

42. Krivega I, Dean A. 2012 Enhancer and promoter interactions – long distance calls. *Curr. Opin. Genet. Dev.* **22**, 79–85. (doi:10.1016/j.gde.2011.11.001)

43. Kwak H, Lis JT. 2013 Control of transcriptional elongation. *Annu. Rev. Genet.* **47**, 483–508. (doi:10.1146/annurev-genet-110711-155440)

44. Maheshri N, O'Shea EK. 2007 Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys.*

Biomol. Struct. **36**, 413–434. (doi:10.1146/annurev.biophys.36.040306.132705)

45. Kraut R, Levine M. 1991 Spatial regulation of the gap gene giant during *Drosophila* development. *Development* **111**, 601–609.

46. Small S, Blair A, Levine M. 1996 Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo. *Dev. Biol.* **175**, 314–324.

47. Krotov D, Dubuis JO, Gregor T, Bialek W. 2014 Morphogenesis at criticality. *Proc. Natl Acad. Sci. USA* **111**, 3683–3688. (doi:10.1073/pnas.1324186111)

48. Bowsher CG, Voliotis M, Swain PS. 2013 The fidelity of dynamic signaling by noisy biomolecular networks. *PLoS Comput. Biol.* **9**, e1002965. (doi:10.1371/journal.pcbi.1002965)

49. de Ronde W, ten Wolde PR. 2014 Multiplexing oscillatory biochemical signals. *Phys. Biol.* **11**, 026004. (doi:10.1088/1478-3975/11/2/026004)

50. Tkacik G, Walczak AM, Bialek W. 2009 Optimizing information flow in small genetic networks. *Phys. Rev. E* **80**, 031920. (doi:10.1103/PhysRevE.80.031920)

51. Walczak AM, Tkacik G, Bialek W. 2010 Optimizing information flow in small genetic networks II. Feed-forward interactions. *Phys. Rev. E* **81**, 041905. (doi:10.1103/PhysRevE.81.041905)

52. Tkacik G, Walczak AM, Bialek W. 2012 Optimizing information flow in small genetic networks III. A self-interacting gene. *Phys. Rev. E* **85**, 041903. (doi:10.1103/PhysRevE.85.041903)

53. MacKay DJC 2012 *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.