Weak Transcription Factor Clustering at Binding Sites can Facilitate Information Transfer from Molecular Signals

Tamara Mijatović, Aimée R. Kok, Merlijn Brüggen, Jos W. Zwanikken, and Marianne Bauer*
Bionanoscience Department, Kavli Institute of Nanoscience, Delft University of Technology, van der Maasweg 9, 2629 Delft, The Netherlands

(Received 2 June 2025; accepted 3 December 2025; published 8 January 2026; corrected 26 January 2026)

Transcription factor concentrations provide signals to cells that allow them to regulate gene expression to make correct cell fate decisions. Calculations for noise bounds in gene regulation suggest that clustering or cooperative binding of transcription factors decreases signal-to-noise ratios at binding sites. However, clustering of transcription factor molecules around binding sites is frequently observed. We develop two complementary models for clustering transcription factors at binding site sensors that allow us to study information transfer from a signal, the morphogen Bicoid, to a variable relevant to development, namely, future cell fates. We find that weak cooperativity or clustering can allow for maximal information transfer, especially about the relevant variable. The timescale of measurement is crucial for predicting the optimal clustering strength: for short measurements, finite clustering is optimal because it allows for the implementation of a switch, while for long measurements, a range of weak clustering strengths allow binding site sensors to access near-maximal developmental information. Weak transcription factor clustering also helps binding site sensors achieve optimality consistent with the information bottleneck bound, which encodes an optimal trade-off between conveying relevant information and making costly measurements: changes in clustering in conjunction with changes in the binding energy can shift the binding site sensor along the optimal bound, and towards an optimal trade-off between obtaining information about the signal and obtaining relevant information.

DOI: 10.1103/cx3r-f2g2

I. INTRODUCTION

Cells differentiate and develop in response to chemical signals which convey information. It has been proposed that these cellular responses follow principles of information maximization: in development and gene regulation, but also in neuronal signaling or chemotaxis [1–10]. Early fly development is one of the canonical examples for precise information flow [8,11–15]. There, a few maternal morphogens, including the transcription factor Bicoid, regulate the development of cells along the head-to-tail body axis of the embryo [11,16,17]. This regulation occurs via a small network of genes, including the gap genes [16,18], with an expression profile precise to ca. 0.1% of the embryonal length [12]. Many parts of cell fate development, such as gap gene patterns and their regulation of cell fates, have been shown to be consistent with principles of information optimization with respect to cell fates [12,14,15,19,20].

On the molecular level, this information transfer is facilitated by binding of transcription factors to binding sites in the regulatory regions of the genome, including promoter or enhancer regions [21]. While the information optimality of binding site sensors is difficult to assess due to the requirement of knowledge of multiple parameters, precise experiments on promoter statistics and binding site occupancies are increasingly pushing this question to an empirical forefront [22–25].

Yet recent experiments on transcription factor behavior around binding sites might pose a challenge to these information-maximization principles. Transcription factors show clustered distributions inside cells or around these binding sites [24–27]. This clustering is associated with reduced information flow, since earlier theoretical calculations based on the Berg-Purcell bound [28–31] found that the signal-to-noise ratio at binding sites decreases when binding occurs cooperatively [32] or via local clustering [33,34]. We use the phrase "clustering" here to remain agnostic of the precise mechanism that leads to a positive affinity between proteins [35], as local clusters can indeed occur from a variety of mechanisms: transcription factors can cluster, for example, by binding cooperatively at binding sites [36], or by forming larger clusters around the binding site region, potentially as a consequence of binding or biomolecular condensation [35,37–41].

Here we address this apparent conflict of information-maximization principles in early fly development by investigating in detail how transcription factor clustering at binding sites affects the information transfer. We shift the focus from sensing errors and signal-to-noise ratios [32–34,42,43]

*Contact author: m.s.bauer@tudelft.nl

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

to the information that the binding site captures about the signal and conveys about possible downstream cell fates, the variable of functional relevance. We observe that this shift in focus resolves the apparent conflict. With two complementary models, we find that weak clustering is consistent with optimizing the information about at least this variable of relevance. We find further that the timescale in the biological system, i.e., the time during which transcription factor occupancy is measured by the binding site, is crucial for predicting the optimal amount of clustering.

To strengthen these findings, we investigate how transcription factor clustering affects information optimization with a constraint on the binding site sensor's capacity to measure the signal [14,44], corresponding to the information bottleneck (IB) goal [45]. We find that transcription factor clustering can help reach a particularly advantageous region along the optimal IB bound, and that tuning both clustering and binding site energy at the same time can move the sensor along the IB bound.

Taken together, our work gives new insights into the role of transcription factor clustering—complementing previous findings that clustering can buffer noise [46,47]—specifically in the context of gene regulation, where transcription factor concentrations encode signals. Our work adds to the ongoing discussion of optimality, by emphasizing a focus on information that is developmentally relevant, and by showing that at least in principle, mechanistic models for binding regions can reach optimal information transfer. It also adds to the discussion of achievability of these optima in high dimensions: we show that multiple mechanistic parameters generate binding sites regions that occupy near-identical points in the IB plane, which is in line with ideas that optima are practically accessible because of "sloppy" directions in parameter space [48,49]. Specifically for our problem, a concomitant modification of two molecular mechanisms (clustering and binding) allows cells to move on a smooth path from having binding sites sensors that may be optimal in one condition to sensors that are optimal in another.

Our paper is structured as follows: In Sec. II we introduce the problem of information transfer, here based on only one of the maternal signals in fly embryo development through a single binding site sensor, and describe our timescale estimates and the calculation procedure. In Sec. III we guide expectations with a model established in the context of cooperative binding, where the occupancy follows a Hill function. This model suggests that weak cooperativity improves information transfer. In Sec. IV we develop a more realistic model of molecules clustering at and around a single binding site sensor, inspired by Ising models [33,50,51], that goes beyond cooperative interactions. With this model, we find that weak clustering is beneficial for short timescales and barely affects relevant information for long timescales. Finally, we show in Sec. V that the IB bound can be achieved with both models. In Sec. VI we summarize and discuss future improvements of our work, including new experimental measurements to assess information optimality of binding site sensors, for example, through mRNA production [22] or binding site occupancy measurements [23,24,52].

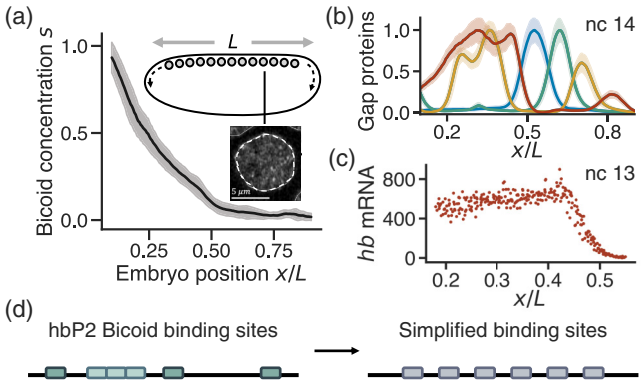

FIG. 1. Signal processing in the fly embryo. (a) The normalized maternal Bicoid gradient [11] along the embryonal head-to-tail axis, which regulates the four gap genes; inset shows hetereogeneous Bicoid concentration inside a single nucleus (airy-scan image with experimental design described in Ref. [24]). (b) The normalized concentration of the four gap gene proteins (*hunchback*, *giant*, *Krüppel*, *knirps* in red, yellow, blue, green respectively) in nc 14 as a function of embryonal position x, scaled to embryo length L, exemplifying the developing body segments [19]. (c) *hb* RNA expression in nc 13 as a function of embryonal position [53], when *hb* expression is mostly regulated through the proximal hbP2 enhancer region. (d) Sketch of real the hbP2 region [54], containing ca. six binding sites, and our simplification with equidistant binding sites of equivalent binding strength.

II. INFORMATION FLOW THROUGH BINDING SITE SENSORS

A. Clustering and information in early fly embryo development

The Bicoid morphogen gradient is the most easily experimentally accessible of the maternal signals that are required for correct body-part segmentation in the fly embryo [Fig. 1(a)] [11,17,55]. The information provided in these maternal signals is transferred downstream via the gap genes, whose expression begins consistently after twelve rounds of nuclear divisions, in nuclear cycle (nc) 13. In nc 14, these gap gene patterns become fully refined [Fig. 1(b)] and provide enough information for nuclei to develop into distinguishable cells [12,19]. This information is then transferred towards a larger group of genes, including the segmentation genes, which determine cell fates and scale with embryo size [12,20]. Although gap proteins act as transcription factors with mutually activating and repressive functions [18] that contribute towards refining their pattern, in nc 13 the gap gene *hunchback* (*hb*) can be considered as being regulated predominantly by Bicoid.

Bicoid has disordered or low-complexity domains, which are associated with condensation [56]. These condensation phenomena are most relevant for high concentrations, and transcription factors are often expressed at low (nM) concentrations: Concretely, Bicoid concentrations are on the order of 10 000 molecules at the anterior (head) and 0–10 molecules at the posterior (tail) of the embryo [11,26,57]. Yet, also at these lower concentrations, clustering can occur through affinity to different, highly expressed and clustering proteins,

or in the form of prewetted, local clusters only around binding sites [24,25,27,50]. Indeed, hubs or clusters involving Bicoid have been observed experimentally [24–27].

The expression of *hb* in nc 13 [Fig. 1(c)] is driven predominantly by the P2 (proximal) promoter or enhancer region (hbP2), with approximately six core binding sites [Fig. 1(d)]. This region has frequently served as a test region for gene expression [8,13,54,58–60]. Recently, mRNA expression statistics for all gap genes, including from this region, have been measured on the level of single polymerases [22]. Here we consider a simplified version of this enhancer as a bottleneck variable for information transfer in the context of transcription factor clustering. We study a single regulatory region with approximately equidistant binding sites with equal binding energies, and take the occupancy of this binding site sensor as the variable that determines cell fates downstream. We investigate only the local environment around this binding site sensor along the DNA, ignoring the three-dimensional microenvironment, potential competition effects between sensors or refining effects from shadow enhancers [54,61,62].

The occupancy C of the binding site sensor corresponds to the nucleus's measurement of the signal s and can be mathematically expressed as the averaged fraction of occupied sites during a time interval τ: $C = \frac{1}{\tau}\int_0^\tau c(t)dt$, with c denoting the occupancy at time t. We consider downstream gene expression to be approximately proportional to C, consistent with simple models of gene regulation, in which C can determine the number of polymerases recruited to the promoter. This approach neglects promoter noise, which is acceptable for weak promoters [63], and is consistent with equilibrium thermodynamic models of gene regulation [64]. Recent work points towards a necessity for nonequilibrium models in gene regulation [65–71]. While including polymerases in ways that match experimental promoter statistics for this region [22,71] presents an exciting direction for future work, the absence of nonequilibrium effects in our model affects our estimates for the time interval τ during which binding site sensors can measure and respond.

Estimating a realistic value for measurement time τ is difficult, since our model does not include all processes in the nucleus. We therefore consider two limiting measurement times: a comparatively short time, corresponding to one or a few measurements that determine gene expression for the rest of the nuclear cycle, and a long measurement time, which represents the maximum possible realistic timescale at this stage of development, corresponding to the entire lengthscale of nc 13 with generous parameter estimates. Nonequilibrium effects, such as decoupling between occupancy and transcription, chromatin accessibility or transcription factor rebinding would shorten this timescale estimate. Therefore, we expect that the realistic measurement time interval τ lies between our two limiting timescales.

B. Information flow calculations

To assess how much information about the external signal of concentration s the regulatory region (or sensor) can capture, we use the mutual information

$$I(C;s) = \int ds \int dC P(C,s)\log_2 \frac{P(C,s)}{P(C)P(s)}, \qquad (1)$$

which depends on probability distribution $P(C|s)$, describing the nucleus's measurement of s.

We use the Bicoid expression profile [11] [Fig. 1(a)] to obtain the prior distribution of signaling molecules $P(s)$, by taking $P(s|x)$ Gaussian and $P(x)$ uniform. For stochastic models of binding at a binding site, we can obtain $P(C|s)$ from simulations (Sec. IV). For sufficiently long time intervals τ, longer than the correlation time, the probability distribution $P(C|s)$ approaches a Gaussian distribution,

$$P(C|s) = \frac{1}{\sqrt{2\pi\sigma_C(s)^2}}\exp\left[-\frac{[C(s)-\bar{C}(s)]^2}{2\sigma_C(s)^2}\right], \qquad (2)$$

where $\sigma_C(s)^2$ is the variance and $\bar{C}$ the mean occupancy that occurs during the interval τ. For such long τ, $\bar{C}$ corresponds to the mean occupancy in steady state, $\bar{C} := \langle C\rangle = \frac{1}{\tau}\int_0^\tau \langle c\rangle dt = \langle c\rangle$, where $\langle\cdots\rangle$ denotes the ensemble average. We can calculate $\sigma_C(s)^2$ and $\bar{C}$ for a specific stochastic process [72], and use this Gaussian approximation to assess information transfer through a specific binding site.

The information provided by the signal could be different from the information that the embryo needs to develop correctly. For example, a signal may provide more information than needed in specific instances. In the fly embryo, correct development requires that nuclei along the embryonal axis can develop into distinct fates, determined by their position [12,73]. We thus use the information that the regulatory region can convey about nuclear position x, $I(C;x)$ [12,14], as a variable of relevance to the embryo. We calculate $I(C;x)$ using Eq. (1) with $P(C|x) = \int P(C|s)P(s|x)ds$, consistent with the information bottleneck framework [45] and the simplified problem we address here, where only Bicoid provides information about cell fates to the enhancer. We can therefore assess precision in the transfer of molecular signals through gene regulation with both the information captured about the signal, $I(C;s)$ as well as the information of relevance, $I(C,x)$, allowing for a more nuanced assessment than possible with signal-to-noise ratios.

III. SIMPLE MODEL FOR GENE REGULATION: WEAK COOPERATIVITY IS OPTIMAL

A. Hill function models: Obtaining probability distributions

In models for gene regulation, the mean occupancy of binding sites $\bar{C}$ for cooperative binding is frequently described by a Hill function [1,74–79],

$$\bar{C} = \frac{s^h}{k+s^h}. \qquad (3)$$

where s is the normalized signal concentration, k the dissociation equilibrium constant, and h the Hill coefficient corresponding to the number of binding sites or cooperatively binding molecules [Figs. 2(a) and 2(b)]. Hill functions originate from a heuristic deterministic description that compares the average number of free and bound binding sites over long time intervals, and accurately describe gene expression outputs in population-averaged experiments. Using Hill function models for gene expression in single cells is more complicated due to the lack of an entirely realistic mechanism, which we address with a stochastic description in Sec. IV. Nevertheless, Hill functions capture gene expression phenomenologically: sigmoidal activation profiles for gene regulation that can be

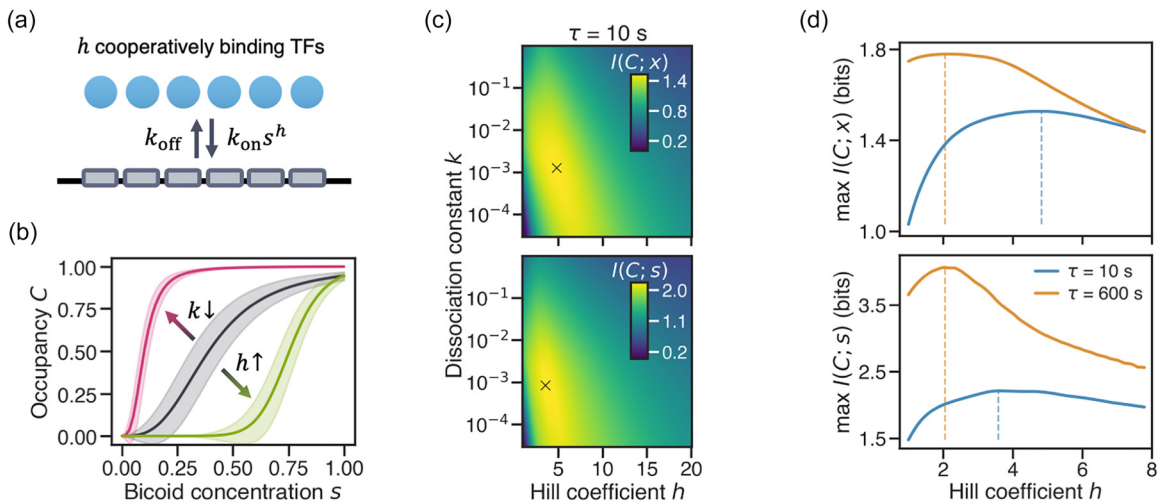

FIG. 2. Information transfer through a sensor with cooperative binding sites. (a) Sketch of cooperative binding of transcription factors (TFs) to six binding sites. (b) Mean occupancy according to the Hill function for different parameters h and k. (c) The information $I(C;x)$ (and $I(C;s)$) as a function of h and k shows a well-defined maximum (cross) in parameter space. We use $k_{\rm off} = 1/\rm s$. (d) The maximum possible information $I(C;x)$ and $I(C;s)$ over all values of k for different h (maxima indicated by vertical lines). For the longer measurement time $\tau = 600\,\rm s$ (orange), the maximal information value per h is higher than for $\tau = 10\,\rm s$ (blue), and maximal values are reached for lower h.

matched to Hill functions are frequently observed [64,66,68]; Hill functions with $h = 1$ correspond to MWC models, also canonically used for gene regulation [78]; and finally, Hill functions provide an interesting limiting case for a stochastic model as the steepest possible binding site occupancy for a regulatory region with h binding sites in single cells [80]. Therefore, we explore Hill function models here to guide expectations.

First, we design a stochastic process for which the mean occupancy can be represented with a Hill function, which will allow us calculate the variance in occupancy. We assume that binding only occurs when h molecules bind at the same time to h binding sites [Fig. 2(a)]. The probability for the binding site to be unoccupied at time t, $P(c = 0, t)$, can be described by the master equation for the two states of being occupied or not occupied, $c = 1$ and $c = 0$, $\frac{dP(c=0,t)}{dt} = k_{\rm off} - (k_{\rm off} + k_{\rm on}s^h)P(c = 0, t)$, where $k_{\rm on}$ and $k_{\rm off}$ are the rate constants for binding and unbinding, and we used $P(c = 1, t) = 1 - P(c = 0, t)$ The mean occupancy in steady state for this process corresponds indeed to a Hill function [Eq. (3)] with equilibrium constant $k = k_{\rm off}/k_{\rm on}$, if the time-averaged normalized mean occupancy $\bar{C}$ is approximately the ensemble-averaged, steady-state mean, $\bar{C} = \langle c \rangle$.

We assume that the probability with which a binding site sensor is occupied during a time interval at specific signal concentration s can be approximated with a Gaussian distribution [Eq. (2)] with this occupancy mean and the variance $\sigma_C(s)^2 = \langle (C - \bar{C})^2 \rangle$. An estimate of this variance can be obtained from the observation that the regulatory region is either occupied or unoccupied during each measurement; in time τ, there will be of order $N \approx \tau/\tau_c$ independent measurements, where the correlation time

$$\tau_c = \frac{1}{k_{\rm off} + k_{\rm on}s^h} \tag{4}$$

can be obtained from the master equation. Therefore, we expect a binomial variance of order $\sigma_C(s)^2 \approx \frac{1}{N}\bar{C}(1 - \bar{C})$, where in steady state $\bar{C}(1 - \bar{C})$ corresponds to the instantaneous variance [1,30,31]. We follow the calculation in Ref. [31] to obtain the exact result (Appendix A 1),

$$\sigma_{\bar{C}}(s)^2 = \frac{2\tau_c}{\tau}\bar{C}(1 - \bar{C}), \tag{5}$$

in agreement with Refs. [1,30,31]. This differs from the ad-hoc expectation by a factor of two; this factor is required when the concentration s is inferred from the occupancy during the time interval τ, whereas inferring the concentration from the discrete number of (independent) binding events allows for a lower variance [81].

The variance in occupancy [Eq. (5)] increases with the correlation time of the occupancy, which in turn increases with Hill coefficient h [Eq. (4)]. Therefore, one might expect that cooperativity in the form of $h > 1$ reduces the information transfer, consistent with Refs. [32–34].

With these expressions, we can calculate how precisely a binding site sensor can measure signal concentration s, provided the sensor is determined by parameters h and k and has access to a time-averaged occupancy during a particular interval τ. We calculate the distribution for the occupancy $P(C|s)$ using Eqs. (2), (3), and (5). We investigate two measurement times representative of a short and long measurement τ. The long measurement corresponds approximately to the duration of nc 13 (10 min), provided that transcription factor residence times of order $\approx$1s can set an effective timescale for $k_{\rm off} \sim 1/\rm s$, consistent with measurements from Ref. [26]. The short measurement is still longer than the correlation time, to ensure that the Gaussian approximation is valid; we chose a time of approximately $\tau = 10\,\rm s$.

B. The information about the signal and information of relevance increase for weak clustering

Both the information about the signal $I(C;s)$ as well as the information of relevance to the embryo, $I(C;x)$ show a clear maximum at a specific parameter combination of h and k per measurement time [Figs. 2(c) and 2(d), Appendix A 1]. This optimum occurs at finite h for all measurement times we studied, and for the longer of our two limiting τ, values close to this optimum are achieved by a range of cooperativities h [exact optimum at $h \approx 2.1$, Fig. 2(d) top]. The presence of this optimum at finite h is in agreement with earlier work that also incorporated noise in RNA production [82].

Both $I(C;x)$ and $I(C;s)$ increase with the measurement time τ, both when optimizing over h and k (Appendix A 1), and when just optimizing over k as a function of cooperativity h [Fig. 2(d)]. This is expected, since the variance decreases with measurement time. For long τ, the information of relevance $I(C;x)$ saturates at approximately 1.8 bits (Appendix A 1), as the binding site cannot extract more information about cell fates than the signal, i.e., the Bicoid concentration, provides (additional signals would allow for more information [12]). In this long measurement time limit and even for the unrealistic limit of $\tau \to \infty$, the optimal cooperativity saturates at $h \approx 1.38$ (Appendix A 1).

Even if we compare the cooperative measurement to one where h binding sites are treated independently, cooperative measurements with $h > 1$ are optimal (not shown). We find therefore that for realistic timescales, weak cooperativity is optimal independent of whether information about signal or cell fates are considered.

From an information-theoretic perspective, it makes sense that steep thresholds corresponding to high h are optimal for shorter measurement times, associated with high noise or variance in $P(C|s)$: binary, threshold-like measurements are known to provide the most information when noise is high [14,83,84]. The optimum occurs at finite cooperativity since the variance is nevertheless low enough for $\tau = 10\,\mathrm{s}$ to allow for a steep but not strict thresholded mean. For longer τ and more precise measurements, we find that weak cooperativity is optimal or does not reduce information up until $h \approx 3$ [Fig. 2(d) orange]. This may seem surprising, since one might expect a linear activation mean to become optimal as noise increases. However, all activation means we study here, even for $h = 1$, imply a nonlinear sigmoidal $\bar{C} = s^h/(k + s^h)$, and such nonlinear functions make it difficult to think about information-theoretic optima intuitively. In addition, for considering $I(C;x)$, we should note that the Bicoid gradient itself varies nonlinearly as a function of x; rate-distortion theory predicts that an optimal map $\bar{C}(s)$ that compresses a signal to convey information about x should be a nonlinear function of the signal if x is conveyed by the signal with a nonlinear map $\bar{s}(x)$ [concretely $\bar{C}(s) = \bar{x}(s)$ [85]]. Since optimizing $I(C;x)$ is additionally a convex problem, different nonlinear maps can achieve the optimum.

If the longer measurement time of 10 minutes is biologically realistic for the fly embryo, an optimally adjusted binding site sensor would have approximately two or three cooperatively binding molecules. Interestingly, this does indeed correspond to the number of strongly binding Bcd molecules in the hbP2 enhancer [54]. We do not expect an exact match, as our estimates for timescales are approximate, but are encouraged to find a reasonable estimate for h. We conclude that our heuristic Hill-function model suggests that some cooperativity is optimal, consistent with information-theoretical expectations considering the functional form of the signal. Yet the Hill-function model suffers from the significant shortcoming that we are forced to consider longer measurement times and cannot capture the binding and unbinding of individual molecules. To check the role of clustering in more detail, we next investigate a more realistic model that proposes a regulatory mechanism based on clustering transcription factors.

IV. MECHANISTIC MODEL FOR CLUSTERING TRANSCRIPTION FACTORS

A. Adapted Ising model for binding sites

To find a more realistic description of the occupancy of a binding site sensor that incorporates binding and unbinding of individual transcription factors, we adapt an Ising model for a small, one-dimensional region of the DNA around the binding sites [Fig. 3(a)]. We fix the concentration of the signal transcription factors with chemical potential μ, parametrize the binding energy for a binding site with ε_b, and describe clustering between neighboring transcription factors with clustering strength J. We calculate the energy of a specific configuration of transcription factors using

$$E(\{c_1, \ldots, c_N\}) = -J \sum_{i=1}^{N-1} c_i c_{i+1} - \mu \sum_{i=1}^{N} c_i - \varepsilon_b \sum_{i=1}^{N'} c_i, \quad (6)$$

where $c_i = \{0, 1\}$ denotes occupancy at site i and N (N') denotes the number of (binding) sites in the system, and where we refer to J, μ and ε_b in units of $k_B T$. The energy of configurations determines the rates for individual events that can occur, which we implement consistent with detailed balance [86]. We simulate two types of events, corresponding to transcription factors appearing from and leaving to a bath (the nucleus), with rates k_{on} and k_{off}. Since arrival statistics should not depend on local effects or the identity of the site (binding site or DNA site), we use the same k_{on} for all sites. This rate $k_{\mathrm{on}} = e^{\beta\mu} k_{\mathrm{off,free}}$ (with $\beta = 1$) then only depends on signal concentration s (through μ) and on a global timescale in the system (Appendix A 2), which is set by $k_{\mathrm{off,free}}$, the rate with which transcription factors leave an isolated or free DNA site [Fig. 3(a)]. We incorporate the local configuration, including site identity and clustering effects with a neighboring transcription factor, into the rate k_{off} (Appendix A 3).

A core difference between cooperativity and clustering is that a clustered hub of proteins can extend beyond the binding site region. Thus, we study two types of binding site sensors: first, a binding site sensor with $N' = N = 6$ binding sites [Fig. 3(a), binding site region], and second, an extended binding site sensor with $N' = 6$ binding sites surrounded by two sites on each side that represent the surrounding DNA region, such that $N = 10$ [Fig. 3(a), DNA region]. We will

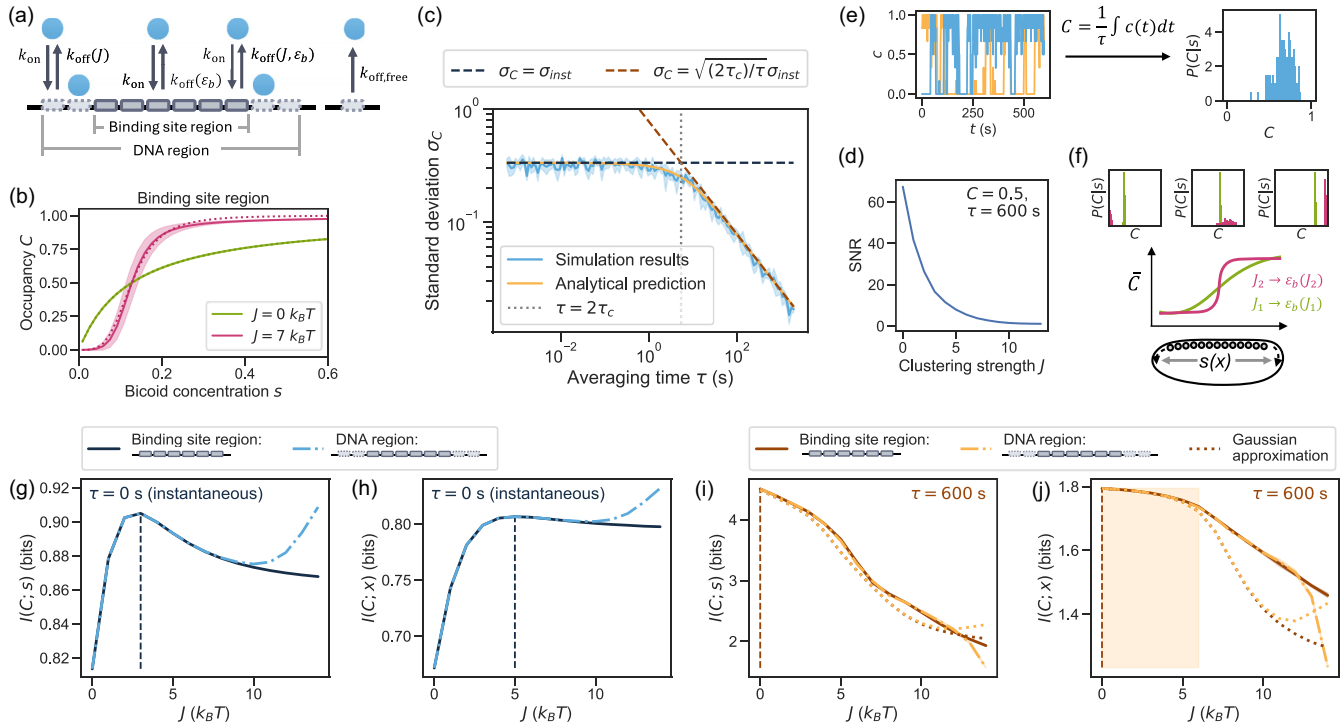

FIG. 3. Information transfer through a binding site or DNA region with individual clustering signal molecules. (a) Sketch of model binding sites (solid) and surrounding sites on DNA (dashed; DNA region), with on- and off-rates for transcription factors. (b) Steady-state occupancy $\bar{C}$ of the binding site region (no surrounding sites) for $J = 0$ and $\varepsilon_b \approx 14.1$, and $J = 7$ and $\varepsilon_b \approx 8.2$ as a function of signal concentration, s(shaded region denotes standard deviation for $\tau = 10$ min), and best-fit Hill function (dotted line), with $h \approx 1$ and $h \approx 4$, respectively. (c) Standard deviation σ_C from simulations of the binding site region matches the analytical solution from the master equation, decreasing from its instantaneous value to a $1/\sqrt{\tau}$-scaling for longer τ. Parameter values are $\mu = -14.0$, $J = 3.5$ and $\varepsilon_b \approx 11.1$, with error bars over 5 independent repeats of the calculation, each based on 20 simulations. (d) The signal-to-noise ratio (SNR) of a long measurement ($\tau = 600$ s) from the binding site region decays sharply with J for exemplary position $x = 0.47$. (e), (f) Sketch of the simulation procedure: for all μ corresponding to signal concentrations $s(x)$ along the embryonal axis, we perform Gillespie simulations and average occupancy over time τ to obtain C. For each s, we estimate $P(C|s)$ based on histograms of C from 100 repeats of the simulation. We calculate $I(C;s)$ and $I(C;x)$ from this distribution for each J, using a modified parameter ε_b to retain half maximum occupancy at $x \sim 0.47$. (g), (h) Mutual information in an instantaneous measurement: Clustering improves both $I(C;s)$ and $I(C;x)$, with a maximum for clustering strengths $J \approx 3$ and $J \approx 5$, respectively. For high J, the occupancy of additional sites along the DNA can further increase the information. (i), (j) Mutual information in a long measurement ($\tau = 600$ s): Clustering decreases $I(C;s)$. The information $I(C;x)$ also decreases with clustering, but with minor information loss (5% of the maximum) up until intermediate clustering strengths of $J = 6$ (shaded area).

see that an interesting effect of the DNA region sensor is that steeper thresholds are achievable for strong clustering strengths J, which can be beneficial for short measurement times.

We obtain probability distributions for the normalized time-averaged occupancy C of the binding site sensor (referred to as "occupancy" from now on) with Gillespie simulations [87,88] (Appendix A 4). To calculate the information, we again focus on limiting measurement durations that represent a short and long measurement time. We use an instantaneous measurement and a long timescale, corresponding again to an estimate of the maximal possible timescale in nc 13, ca. 10 min, which we estimate using diffusion constants and by generously matching lengthscales in our simulation to those in the fly embryo (Appendix A 2 [57,89–91]).

However, to guide expectations, we first explore the mean $\bar{C}$ and variance σ_C^2 of the occupancy C as a function of input parameters. We refer to Appendix A 3 for the stochastic master equation and detailed calculations.

B. Clustering causes steeper occupancy mean and increase in occupancy variance

The mean occupancy $\bar{C}(s)$ increases more steeply with signal concentration s for higher clustering strength J [Fig. 3(b)]. Clustering also shifts the signal concentration at which the mean occupancy of the binding site sensor reaches its half-maximal occupancy; varying the binding energy can compensate for this shift [Fig. 3(b)]. A Hill function can approximate the mean occupancy, justifying our approach in the first section.

The variance around the mean occupancy $\bar{C}$ in steady state can be obtained using the correlation function (see Appendix A 3, Refs. [33,92]). We find

$$\sigma_C^2 \approx 2\sigma_{\text{inst}}^2 \frac{e^{-\tau/\tau_c} + \frac{\tau}{\tau_c} - 1}{(\tau/\tau_c)^2}, \tag{7}$$

where σ_{inst}^2 is the instantaneous variance in steady state, which describes the fluctuations in (instantaneous, not

time-averaged over τ) occupancy around the steady-state mean at stationarity, $\sigma^2_{\text{inst}} = \langle (c - \langle c \rangle)^2 \rangle$. This expression reduces to $\sigma^2_C(\tau) \approx \sigma^2_{\text{inst}}$ for $\tau \ll 2\tau_c$ and to $\sigma^2_C(\tau) \approx \frac{2\tau_c}{\tau}\sigma_{\text{inst}}$ for $\tau \gg 2\tau_c$, as expected [31]. We validate this approximation by showing excellent agreement of the variance from Eq. (7) with a semianalytical expression using the master equation [Appendix A 3, Figs. 7(a) and 7(b)], which matches the stochastic simulations [Fig. 3(c)]. Clustering increases the correlation time of a measurement, which in turn increases the variance (Appendix A 3 [33]).

We thus conclude that the strength with which transcription factors cluster, J, affects both the mean and variance of the occupancy at the binding site sensor. Stronger clustering leads to strong decrease in signal-to-noise ratio, visualized in Fig. 3(d) at the middle of the embryo, for $x = 0.47$; this decrease may generate the expectation that transcription factor clustering is at odds with the idea of maximal information flow. However, while stronger clustering leads to increased variance, the concomitant effect on the mean occupancy may foil the conclusion that clustering always decreases information, just as in Sec. III.

C. In the fly embryo, weak clustering can lead to an increase in mutual information about cell fates

Before presenting how clustering affects information transfer in the fly embryo in this model, we explain one modification that we make to reduce the parameter space which we sample. In our model, increasing the clustering strength J shifts the mean occupancy towards anterior parts of the embryo. This behavior is consistent with the idea that clustering increases the local concentration of transcription factors [39,93]. Yet, as a consequence, many combinations of J and the binding energy ε_b lead to sensors that capture barely any information. Such sensors are not biologically realistic: If a particular clustering strength is set by properties of transcription factor proteins, and if the information transfer from Bicoid is important, the embryo would likely have faced evolutionary pressures that ensure that the binding energy is adjusted through mutations in binding sequences, to make information transfer feasible. Therefore, as we vary J in the following section, we choose to vary ε_b at the same time, setting ε_b such that the binding site sensor can produce half maximal expression at $x \approx 0.47$, as expected for *hb* expression in the fly embryo [Fig. 1(c), Appendix A 3, Fig. 7(e)]. This reduces the parameter combinations we study. While we do not prove that this choice is information-theoretically optimal, we see in Sec. V that it is close to the information bottleneck bound. For each J, we obtain mutual informations by calculating all $P(C|s)$ for values of chemical potential $\mu(x)$ that yield a particular signal (Bicoid) concentration s [Figs. 3(e) and 3(f)].

For instantaneous measurements, weak clustering increases information transfer: both $I(C;s)$ and $I(C;x)$ increase with J with local maxima at $J \approx 3$ and $J \approx 5$, respectively [Figs. 3(g) and 3(h)]. This is consistent with our results from Hill-function models, and the intuition that steeper, more thresholded measurements are optimal when the measurement is noisy. This robust maximum at finite clustering strengths, consistent across models, is noteworthy. It occurs at finite clustering strength, not at a sharp threshold or infinite clustering, since the instantaneous variance is finite and we are not in large-noise limit where strictly binary measurements are optimal (Appendix A 3).

For the DNA region sensor that includes a larger region around the binding site, stronger clustering of $J \approx 10$ improves information transfer again. This increase in information is related to the mean occupancy: the effective Hill coefficient for the binding site region reaches $h = 6$ from below for high J [Appendix A 3, Fig. 7(e)], since the Hill coefficient is bounded by the number of binding sites [65,80]; in contrast, the steepness for the DNA region can increase further, as clusters extend beyond just the binding sites. The information will eventually saturate at higher J, also for longer DNA sensors, beyond the valid regime of our simulations: For $J \gg 15$, the binding energy ε_b required to keep half-maximal occupancy at the center of the embryo becomes unrealistically small. We used an artificial sensor with periodic boundary conditions to estimate that information will saturate at $I(C;s) \approx 0.93$ and $I(C;x) \approx 0.85$ in our model.

The fact that clustering can lead to occupancy of surrounding DNA sites could therefore be an avenue by which clustering can increase information transfer in the embryo (see also Refs. [94,95]). However, for signals expressed at low concentrations, such strong clustering is likely also associated with a source of error that our model does not capture: for small numbers of signal molecules and multiple binding site regions, large clusters around one region may deplete signal molecules from other binding site regions, and therefore contribute to a more noisy measurement in realistic settings. Since we do not take into account these depletion effects, we cannot evaluate this effect here.

For long measurement times, clustering reduces the information about the signal [Fig. 3(i)]. This information loss occurs because of a combination of two effects: since measurements are more precise (low variance) for long τ, the steeper mean activation of occupancy (which is facilitated by increasing J) is less beneficial. In addition, clustering increases the variance, as the increasing correlation time implies that less independent measurements can be performed during interval τ, consistent with earlier work [33]. An important contrast is the information of functional relevance to the embryo: $I(C;x)$ exhibits a plateau for weak clustering strengths and does not decrease significantly (within five percent) for clustering up to $J \approx 6$ [Fig. 3(j)]. This information is barely affected by weak clustering, since the effect from the increasing variance can be partially compensated by the fact that for conveying relevant information, a steeper nonlinear mean can presen an optimal map if the map between signal and relevant variable is also nonlinear. This suggests that weak clustering is not detrimental for information transfer, and this result is even more striking when we investigate a model for gene regulation in which gene expression is only induced when the binding site region is fully occupied [13,96]: here we observe less than 1% information loss up to $J \approx 5$, corresponding to Hill coefficients up until $h = 3$ (Appendix A 5, Fig. 9).

For long τ, we can verify the validity of the Gaussian approximation, which we used in Sec. III, by comparison to simulations: this approximation works well until medium clustering strengths, $J \approx 6$, above which $P(C|s)$ becomes

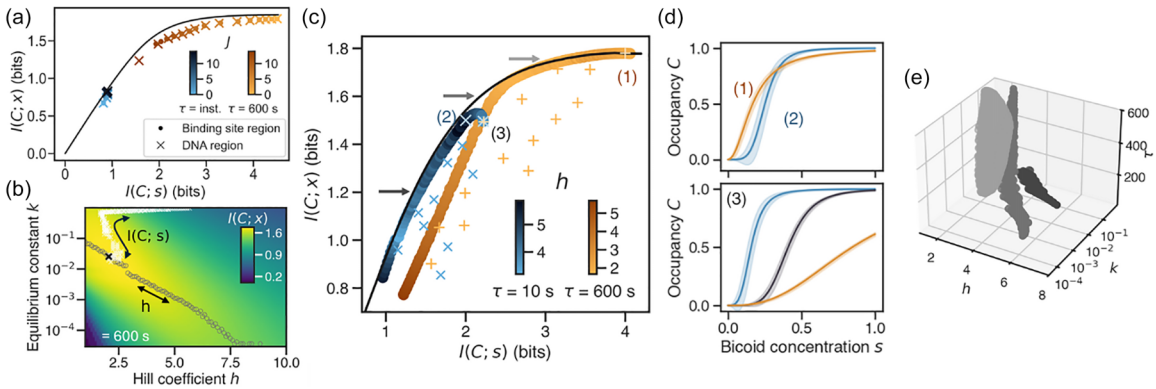

FIG. 4. Binding site sensors that incorporate cooperativity and clustering can theoretically achieve the information bottleneck optimization goal. (a) Sensors from the mechanistic model for instantaneous (blue) and long measurement times (orange) for changing clustering strength J (color shade) on the information plane, with optimal IB bound from data in black. (b) Different binding site sensors from the Hill function model, parameterized by h and k, are optimal for different constraints: grey circles and white crosses mark paths with a (variable) constraint on h and a (variable) constraint on $I(C;s)$ (IB goal), respectively. (c) Information plane with IB bound in black and a selection of exemplary binding site sensors (blue crosses and orange plusses for $\tau = 10$ s and $\tau = 600$ s, respectively) show that randomly chosen binding sites are well below the bound. The sensors that optimize $I(C;x)$ for a given $I(C;s)$ in our parameter ranges of h and k are shown with larger circles; color shade indicates the value of h. (d) Selected binding site sensors (1,2) from panel C representing the sensors along the IB bound (top), and sensors that represent point (3) in panel (c) (bottom, for $\tau = 10$ s in blue, 60 s in gray, and 600 s in orange). (e) Optimal values of h, k, and τ for binding site sensors fitted to $P(C|s)$ from IB optimization, for three values on the bound in panel (c) (gray arrows) show that a variety of binding site sensors are consistent with the IB bound.

too strongly peaked at empty or full occupancy [Figs. 3(i) and 3(j)]. We also observe that the difference between the extended DNA sensor and binding site region is marginal for the long measurement time, as the information is not limited by the steepness of the mean occupancy.

We expect realistic timescales for the embryo to lie between the instantaneous and the long measurement time, i.e. between regimes with a clear optimum at weak clustering or negligible information loss for weak clustering. Therefore, our results show that, independently of the exact mechanistic model, weak clustering is consistent with optimizing the functionally relevant information transfer.

V. THE INFORMATION BOTTLENECK BOUND IS ACHIEVABLE WITH REALISTIC BINDING SITE SENSORS

In the previous sections, we discussed that the information of relevance $I(C;x)$ is a better quantity to use in maximizing information flow than the information about the signal $I(C;s)$. In this section, we investigate how clustering affects information maximization of $I(C;x)$ with a constraint. We focus on a constraint on the binding site capacity $I(C;s)$. Since measurements of the signal and therefore of each bit of $I(C;s)$ are costly for the organism, an organism may want to extract the maximal amount of relevant information $I(C;x)$ given a specific $I(C;s)$. An optimization with this constraint is an instance of an information bottleneck (IB) problem [45]. In the context of information flow in early fly development, this IB framework has predicted several features of enhancer networks that appear in the fly embryo, such as a need for multiple enhancers or a combination of binding sites for different signals [14]. Therefore, we want to investigate how clustering of transcription factors affects the ability of binding site regions or sensors to operate close to this bound.

We first find the probability distribution that describes the best binding site sensor C abstractly: The distribution $P(C|s)$ can be optimized numerically for the goal

$$\max_{P(C|s)} I(C;x) - \lambda I(C;s). \tag{8}$$

Here λ is a Lagrangian multiplier that constrains $I(C;s)$; we can find the optimal sensor and the information $I(C;x)$ it can convey for each value of λ (and thereby for each value of $I(C;s)$), without making assumptions about the sensing mechanism (Appendix A 6). This optimal IB bound of maximal $I(C;x)$ for each $I(C;s)$ is obtained directly from the distribution that maps the relevant variable to the signal, $P(s|x)$. It divides the information plane [Figs. 4(a) and 4(c)] into an allowed region on or below the optimal IB bound and a forbidden region above.

We compare realistic binding site sensors to this bound by showing pairs of $I(C;s)$ and $I(C;x)$ from binding site sensors with specific parameter combinations and models on the information plane. We begin with binding site sensors from the mechanistic simulations, for both measurement times and varying J [Fig. 4(a)]. Strikingly, all binding site sensors are close to the information bound, for all values of J. Sensors with longer measurement times τ measure more precisely and occupy regions higher in the information plane than sensors with instantaneous measurements, which also cover a smaller region of phase space [orange vs blue in Fig. 4(a)].

In general, we see sensors close to the IB bound as consistent with IB optimality, while sensors that are not optimal fall far below the bound; this is because we do not know *a priori* which $I(C;s)$ is physically possible and therefore what specific value point or range of values in the information plane are reasonable. Some binding site sensors are within 10% of the maximally achievable $I(C;x)$ (light orange); these

sensors involve weakly clustering transcription factors, up to clustering strengths of $J \approx 3$ [Fig. 4(a)].

More specifically, the elbow of the IB plot marks a particular "sweet spot" in optimization. We find that for Bicoid, an abstract IB-optimal sensor can operate within 20% of the bound of 1.8 bits when sensing with a capacity of 1.8 bits. We further find that sensors with finite clustering, $J \approx 5$, are closest to this elbow of the plot. This suggests that transcription factor clustering is beneficial especially from an IB perspective, since clustering helps the sensor optimize a trade-off by measuring as much relevant information as possible with the noisiest possible measurement.

It may be surprising that all binding site sensors from the mechanistic model for all values of clustering J are close to the IB bound. Our choice of constraining ε_b for each J to fix the half-maximal occupancy to the middle of the embryo thus implicitly presents an effective optimization consistent with the information bottleneck. While variation of ε_b may have allowed us to push even closer to the bound, it is interesting that our biologically motivated choice satisfies the bound well, as it suggests that this joint tuning of clustering strength and binding site energy can provide a smooth, potentially evolutionarily accessible path towards optimizing the relevant information.

To explore more binding site sensors, we return to the Hill-function models from Sec. III. Having access to a variety of parameters allows us to visualize the effect of changing the constraint from fixed h to fixed $I(C;s)$ [Fig. 4(b)]. The optimal sensor parameters differ between constraints, intersecting at the optimum. Along the IB constraint, h varies slightly while the equilibrium constant k increases untilinitially increases for decreasing $I(C;s)$, followed by a decrease until the equilibrium constant k is so large that the occupancy of an individual binding site sensor barely reaches 50% for the highest s (Appendix A 6). This suggests that forcing sensors for a single measurement time to occupy all parts of the IB plot is not sensible. Therefore, we do not consider binding site sensors with maximal occupancy below 0.5.

Binding site sensors with varying h and k can fall on a wide region below the IB bound [symbols in Fig. 4(c)], for both τ. Sensors that measure longer achieve more information, especially for more optimized sensors (orange crosses extending to higher $I(C;x)$). To investigate whether some binding site sensors match the IB bound, we numerically select the binding site sensor that maximizes $I(C;x)$ for each $I(C;s)$ (as in Fig. 4(b)), and show the corresponding information pairs in the information plane [blue and orange lines, with color shade indicating h, Fig. 4(c)]. The optimal sensors are on the IB bound for different regions of phase space: the sensors that measure for longer τ are unable to achieve the IB bound for low $I(C;s)$, due to our constraint on a sensor's mean occupancy. The good performance of a single sensor achieving high $I(C;x)$ may be surprising, but this likely reflects our maximum timescale estimate being too generous, or our rate constants not being realistic. Sensors for short τ achieve the bound for low $I(C;s)$ and are in general steeper than sensors for long τ close to the bound [Fig. 4(d), top].

Different sensors can yield the same $I(C;s)$ and $I(C;x)$ and therefore end up at the same region of the information plane [Fig. 4(d), bottom]. Indeed, multiple parameter combinations h, k and τ yield sensors with the same $I(C;s)$ and $I(C;x)$ along the bound [Fig. 4(e)]. This shows that multiple implementations of sensors can be consistent with information-theoretic optimization goals [48]; nevertheless, randomly selected sensors are not optimal.

In this section, we therefore showed that weak transcription factor clustering is more beneficial when the information of relevance is optimized with a constraint on binding site capacity. This suggests that it will be interesting to identify biologically realistic constraints, which could include molecule numbers or energy constraints [6,15,97,98], and investigate their impact on information optimizations.

VI. DISCUSSION AND CONCLUSION

A. Considering functionally relevant information resolves the paradox of clustering in info-max principles

We developed two models that allow us to calculate the precision with which binding site sensors measure transcription factor signals, and the information about cell fates that the binding site sensor can capture from the signal. Our first model based on Hill functions relies on a series of approximations, and does not describe the binding of individual molecules. To go beyond this, we developed a second more realistic model that also allows us to study clustering around a binding site region. We chose the occupancy of a binding site region, for which a decrease in the signal-to-noise ratio due to clustering had previously been predicted, as the variable that needs to convey information downstream. In both models, the presence of cooperativity or clustering can increase the relevant information that the binding site region can convey for short measurements, as steep enhancer activations are optimal for noisy sensing. Even for the longest measurement duration that could possibly be realistic according to our estimates, clustering does not significantly reduce this relevant information.

We therefore find that when drawing conclusions about the optimality of information transfer in a biological system, it is important to consider the functionally relevant information, rather than an estimate about the signal, which led to an apparent conflict of info-maximization principles based on previous estimates.

We also showed that binding site regions with clustering transcription factors are able to achieve information optimality as predicted by the information bottleneck bound, with clustering pivoting the sensor towards a particularly beneficial sensing trade-off. Here we showed that a variety of different binding site sensors are IB optimal, and that varying binding energy and clustering at the same time can tune the sensor along the IB bound. This makes IB optimal binding site sensors not abstract solutions in a parameter space that is impossible to search, but mechanistically and potentially evolutionarily achievable.

B. Connection to current experiments

Our result that clustering can be beneficial for information transfer connects with current work on the importance of clustering around binding site regions in the context of transcription: clustering can buffer signal fluctuations,

which improves accuracy for a constant signal that fluctuates [46,47]. Here we work towards understanding how clustering affects the processing of signals whose concentration conveys information. Our findings provide an information-theoretical understanding of the usefulness of clusters that stretch between and around binding sites [94,95], and of the importance of weak interactions in the context of transcription [99]. While current work for larger transcriptional hubs has discussed a Goldilock's principle of a narrow regime of optimal clustering strength [23,100], with our model we find that multiple clustering strengths can be close to optimal information transfer if binding energies are adjusted.

In order to allow for a more direct assessment of enhancer and promoter regions with respect to information optimality, three improvements to our work should be considered. First, our estimates of timescales suffer from difficulties with matching to new data for residence times and occupancy statistics of transcription factors [23,24,26]. A model improvement that would make this easier involves extending the size of our system to a larger 3D geometry with different binding site sensors. This geometry could then allow for transcription factors diffusion, and could capture competition effects between these sensors: If clustering occurs at multiple regulatory regions at the same time, transcription factors could be depleted [101] and these competition effects between these binding sites regions will reduce the accuracy of information transfer. Conversely, some regulatory regions regulate expression combinatorially [21].

Second, the promoter statistics should be included more explicitly in our model. Recently, mRNA production of the gap genes has been measured on the level of individual polymerases [22]. Models that reproduce these polymerase statistics suggest that an out-of-equilibrium drive is required in a model that includes polymerases [71]. An exciting direction is to add polymerases with nonequilibrium rates, to obtain promoter regions that match experimental statistics, and investigate their optimality following our pipeline. We expect that adding polymerases would significantly reduce information capacities in our system, but that it would provide clearer insights into how tightly squeezed biological systems are for achieving optimal information transfer.

Third, a more complete picture of information transfer would require incorporating additional proteins and combinatorial inputs, and potentially their dynamics. Multiple maternal signals provide information to multiple outputs, including the four gap genes, assisted by pioneer factors [102,103]. This combinatorial regulation can allow for modular optimal solutions that may fulfill biological functions and depend on steeper activation profiles, which clustering could help facilitate. In addition, polymerases and pioneer factors can cluster themselves [26,104], and it has recently been suggested that variables other than binding site occupancy assist in carrying relevant information, such as cluster shape or size [24,25,105]. Finally, the temporally changing dynamics of these proteins could present an exciting avenue for considering information in the full trajectory [106,107], including temporal smoothening of output profiles through diffusion [108], or the opening of enhancer regions due to pioneer activity [103,109,110].

These improvements would allow us to assess whether real regulatory regions with clustering transcription factors optimally processes information. For the single binding site region that we study here, a test for the connection between clustered occupancy and downstream processing would involve measurements of transcription factor occupancy at this region (ideally labeled [111,112]) in combination with concomitant output (RNA and protein). Experimental possibilities to tune the clustering strengths between molecules could further allow us to understand the landscape of optimal parameters that are still consistent with healthy development [48].

ACKNOWLEDGMENTS

We are grateful to Ned Wingreen, Eric Wieschaus, and William Bialek for many helpful discussions of questions, models, and calculations relevant to this project. We thank Thomas Gregor and Mikhail Tikhonov for providing data for *hb* mRNA for Fig. 1(c), and Rahul Munshi for providing an image of Bcd in Fig. 1(a). We thank Riccardo Rao, Linda Dierikx, Max Metlitski, Caroline Holmes, Michal Levo, Thomas Gregor and his laboratory, Trudi Schüpbach, and members of the Bauer group for comments and discussions. We acknowledge funding from the NWO Vidi Talent programme (NWO/VI.Vidi.223.169, M.B.), the NWO Science-XL Grant No. OCENW.XL21.XL21.115 (A.K. and M.B.), and a TUDelft start-up grant (J.Z.). A part of this work was performed at the Aspen Center for Physics, which is supported by National Science Foundation Grant No. PHY-2210452.

DATA AVAILABILITY

The data that support the findings of this article are openly available [72].

APPENDIX

1. Effective Hill function model: Variance and numerics

For the Hill function model, we assume that binding only occurs when h molecules bind at the same time to h binding sites. The chemical master equation for the two states of occupancy at the binding site, $c = 0$ and $c = 1$, reads

$$\frac{dP(c=0,t)}{dt} = k_{\text{off}} - (k_{\text{off}} + k_{\text{on}} s^h) P(c=0,t), \tag{A1}$$

where k_{on} and k_{off} are the rate constants for binding and unbinding. We assume that the signal concentration s is not affected by binding, since the number of binding sites is typically much less than the number of signal molecules in the nucleus. We can solve the master equation,

$$P(c=0,t) = \frac{k_{\text{off}}}{k_{\text{off}} + k_{\text{on}} s^h} + \left(P(c=0,t=0) - \frac{k_{\text{off}}}{k_{\text{off}} + k_{\text{on}} s^h} \right) \times e^{-(k_{\text{off}} + k_{\text{on}} s^h)t},$$

to find the mean at steady state. This mean indeed corresponds to the Hill function,

$$\bar{C} = \langle c \rangle = \frac{s^h}{k + s^h}, \tag{A2}$$

as in the main text. Although the effective ergodic, long τ approximation that we need for the first equality ($\bar{C} = \langle c \rangle$) may not hold for single nuclei and shorter timescales, we proceed with this approach, as Hill functions are established in gene regulation, and present an interesting and accurate limit for the true stochastic process, as described in the main text.

To calculate the variance, we consider the temporal correlation function at stationarity: $\kappa(\delta t) = \langle (C(0) - \bar{C})(C(\delta t) - \bar{C}) \rangle$. This correlation function depends only on the time difference δt [113]. This allows us to write for the variance

$$\begin{aligned} \sigma_C(s)^2 &= \langle (C - \bar{C})^2 \rangle, \\ &= \frac{1}{(\tau)^2} \int_0^\tau \int_0^\tau \langle [C(t_1) - \bar{C})(C(t_2) - \bar{C}] \rangle dt_1 dt_2, \\ &= \frac{1}{(\tau)^2} \int_0^\tau \int_{-t_1}^{\tau - t_1} \kappa(t) dt dt_1, \end{aligned} \tag{A3}$$

following the calculation in Ref. [31]. We swap the order of integration in Eq. (A3) and exploit symmetry of $C(t)$ to perform the integration over dt_1:

$$\langle (C - \bar{C})^2 \rangle = \frac{2}{\tau^2} \int_0^\tau dt \kappa(t)(\tau - t). \tag{A4}$$

This can be approximated further if the correlation function $\kappa(t)$ decays on a timescale shorter than the measurement time $\tau_c \ll \tau$,

$$\sigma_C(s)^2 = \langle (C - \bar{C})^2 \rangle \approx \frac{2}{\tau} \int_0^\tau dt \kappa(t).$$

We can obtain $\kappa(t)$ from the solution from the master equation (A1),

$$\kappa(t) = \bar{C}(1 - \bar{C}) e^{-t/\tau_c} = k_{\text{off}} k_{\text{on}} s^h \tau_c^2 e^{-t/\tau_c}. \tag{A5}$$

where τ_c is the correlation time from Eq. (4), and $\bar{C}(1 - \bar{C})$ is the instantaneous variance as in the main text. Using $k = \frac{k_{\text{off}}}{k_{\text{on}}}$, the variance can be expressed as

$$\begin{aligned} \sigma_C(s)^2 &= \frac{2}{\tau k_{\text{on}}} \frac{k s^h}{(k + s^h)^3}, \\ &= \frac{2}{\tau k_{\text{off}}} \frac{k^2 s^h}{(k + s^h)^3}, \\ &= \frac{2\tau_c}{\tau} \bar{C}(1 - \bar{C}), \end{aligned} \tag{A6}$$

for large τ.

We use Eq. (A6) to calculate the variance for our binding site sensors depending on parameters $k = k_{\text{off}}/k_{\text{on}}$, h and τ. The system depends on three free parameters, which can equivalently be chosen as τk_{on}, k_{off} and h, or as τk_{off}, k_{on} and h [see Eq. (A6)]. We opt for the second choice, as the first options errs towards a fallacy where the information can always increase as k_{off} increases [Fig. 5(a) with the maximum at the boundary of our range for k]: in our system, the information to the binding site is provided only through the binding, and therefore a fast unbinding rate (and instantaneous disappearing of the molecule in our diffusive-free system) allows for more independent measurements. We therefore make the practical choice to keep $k_{\text{off}} = 1$ fixed and vary k_{on}.

For comparison, Fig. 5(b) shows the information provided through a binding site with varying h and k (for constant k_{off}, as in the main text) for both measurement times side by side. The maximum information, indicated by a cross, occurs for finite h and k. We calculate probability distributions using the Gaussian assumption numerically, with 100 bins in the variable C in the main text. To ensure that the finite number of bins does not affect our result, we recalculated the information as a function of binding site parameters h and k for a different numerical parametrization, with 1000 discrete bins for the variable C. We observe minor shifts for the parameters h and k that make up the maximum (Fig. 6), but the results do not change otherwise; therefore, we decided to use 100 bins for the main text.

We use 160 positions between 0.1 and 0.9 embryo length (for details see Sec. A 4 for information about x-position bins).

2. Time and length scales for the mechanistic model

The length scale of the model is set by the shortest distance between the centers of two Bicoid binding sites in the *hb* enhancer: 12 bp or approximately 4 nm [89]. The shortest distance between the centers of two Bicoid binding sites in the *hb* enhancer is 12 bp or approximately 4 nm [89]. We thus use a grid cell of 4 nm. With this grid size, the concentrations used in our simulations are all below 10^{-12} molecules/grid cell, and thus satisfy the condition that the occupancy at each binding site is much less than 1.

We assume that on the surrounding sites in the DNA region the transcription factors do not interact with DNA. Therefore, the off-rate in the absence of binding and clustering depends only on the timescale in which a transcription factor diffuses away. To simplify the calculations, we rescale time such that $k_{\text{off,free}}$ is equal to 1 per time unit. One time unit is then equal to the mean time it takes a transcription factor to diffuse through the grid cell. We roughly estimate this timescale using the equation for the mean squared displacement (MSD) of a diffusing particle:

$$\text{MSD} = 2KDt \rightarrow t = d_x^2 / (2KD), \tag{A7}$$

where D is the diffusion constant, d_x is the grid cell size and K is the number of dimensions in which the transcription factor can leave the grid, with $K = 3 - 1$ for a one-dimensional grid. Values for the diffusion constant of Bicoid in literature range form $0.3\,\mu\text{m}^2/\text{s}$ [90] to $7.7\,\mu\text{m}^2/\text{s}$ [91]. We use the more likely and recent value of $D \approx 7\,\mu\text{m}^2/\text{s}$ as found by [57]. With these values we obtain a time unit of $t = 1/k_{\text{off,free}} \approx 5.7 \times 10^{-7}$ s or $k_{\text{off,free}} = 1.8 \times 10^6/\text{s}$. With typical binding energies that we use for weak clustering, these timescales typically correspond to $k_{\text{off,free}} \sim 1/\text{s}$, which is consistent with the timescale estimate we use in the Hill function model. Residence times of order 1s are consistent with recent experiments for Bicoid [26], as well as previous experiments in

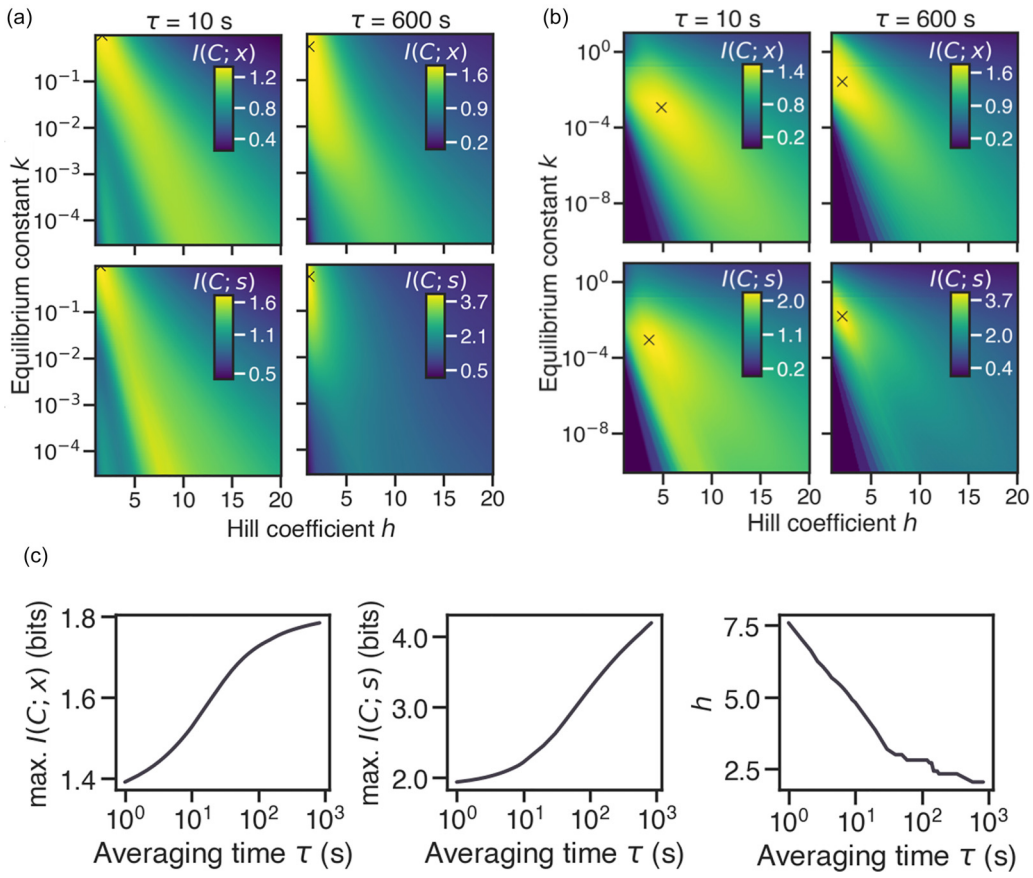

FIG. 5. $I(C;x)$ and $I(C;s)$ as a function of h and k for two representative measurement times. (a) $I(C;x)$ and $I(C;s)$ for $k_{\rm on} = 10\,{\rm s}^{-1}\,{\rm nM}^{-h}$ constant and varying $k_{\rm off}$; the maxima (cross) are always at the boundaries of parametrization space. (b) Longer measurements with $\tau = 600\,{\rm s}$ show a similar behavior to $\tau = 10\,{\rm s}$, with a single optimum (cross). The left panels are replotted from the main with an increased range of k values. ((C) left and middle) The optimized information $I(C;x)$ and $I(C;s)$ over all k and h as a function of measurement time τ, with $I(C;x)$ reaching its maximal bound $I(s,x)$ at approximately $\tau \sim 1000s$. ((C) right) The value of h at the optimum of $I(C;x)$ decreases with measurement time τ.

the fly embryo but also in other systems [91,114–117]. We note that we assume that the estimate for our longer measurement is likely longer than what is realistically possible for enhancers: this is partially because we chose generous estimates for diffusion and grid size, and because the enhancer is likely not actively measuring for the entire duration of the cycle.

3. Semianalytical calculation of standard deviation for the mechanistic model

In the absence of clustering, N states with different discrete occupancy values n are enough to describe a system with N sites. For a system with $N = 6$ binding sites, the probability of the state with n transcription factors bound at a timepoint t, $P_n(t)$ is given by

$$\begin{aligned}\frac{dP_n(t)}{dt} &= \{[N-(n-1)]k_{\rm on}\}P_{n-1}(t) + [(n+1)k_{\rm off}]P_{n+1}(t)\\ &\quad - [(N-n)k_{\rm on} + nk_{\rm off}]P_n(t), \quad \text{for} \quad 0 < n < N,\\ \frac{dP_0(t)}{dt} &= k_{\rm off}P_1(t) - Nk_{\rm on}P_0(t),\\ \frac{dP_N(t)}{dt} &= k_{\rm on}P_{N-1}(t) - Nk_{\rm off}P_N(t), \qquad \text{(A8)}\end{aligned}$$

where P_0 and P_N describe probabilities for empty and full states. The variable n is related to c in the main manuscript by a normalization constant, i.e., $c = n/N$. Probabilities need to satisfy the constraint $\sum_{n=0}^{N} P_n(t) = 1$. The expression for the energy of a configuration in the main text can be used to determine the energies of a particular state and thereby determine the rates $k_{\rm on}$ and $k_{\rm off}$, where $k_{\rm on}$ only depends on signal concentration s and through it on μ, and where $k_{\rm off}$ depends on the site identity and the local neighborhood,

$$k_{\rm on} = e^{\beta\mu}k_{\rm off,free} \quad \text{and} \quad k_{\rm off} = e^{-\beta(\varepsilon_b + Jn)}k_{\rm off,free}. \qquad \text{(A9)}$$

In the limit of a single binding site, we recover Eq. (A1) with $h = 1$ and, using $\mu = k_B T \ln s$, the steady-state mean corresponding to the Hill function [Eq. (3), $\frac{s}{s+e^{-\beta\varepsilon_b}}$].

To incorporate clustering, we need to take into account all possible states of occupancy in what could be seen as a graph-theoretical approach [67]: different states with the same occupancy are connected with different rates. Taking into account all states requires a $2^N \times 2^N$ transition matrix with elements M_{ij}, where i and j run over all 2^N different states, described by their occupancy for each binding site $\{n_1, \ldots, n_N\}$, where each $n_i \in \{0, 1\}$. For example, at a particular time, a binding site sensor might be in state $\{1, 1, 1, 0, 0, 0\}$. We

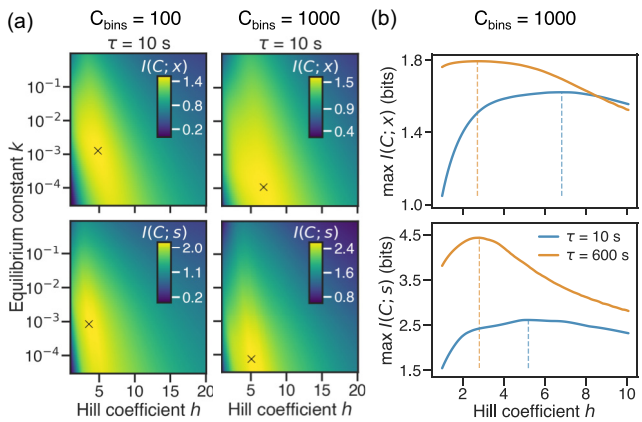

FIG. 6. $I(C;x)$ and $I(C;s)$ using a changed number of bins for the numerical calculation of $P(C|s)$. (a) The information $I(C;x)$ and $I(C;s)$ as a function of h and k for $\tau = 10\,\mathrm{s}$ shows maximum for binding site parameters (cross) at finite h, independent of numerical details. These details can affect the exact parametrization of the optimum, especially the dissociation equilibrium constant k. (b) The maximum possible information $I(C;x)$ and $I(C;s)$ over all values of k for different h. Finite h is optimal for information transfer and the h at the optimal mutual information $I(C;x)$ decreases with τ, as in the main text [Fig. 2(d)].

denote the occupancy of each such state by $\tilde{c}_i = \sum_{k=1}^{N}(\{n_k\})/N$, so that $\tilde{c}$ is also a 2^N vector.

The probability to occupy state i, $\tilde{P}_i$, can be obtained from the solution of the master equation, via

$$\tilde{P}_i(t) = \sum_{j=1}^{2^N}(e^{Mt})_{ij}\tilde{P}_j(0), \tag{A10}$$

where $\tilde{P}_j(0)$ denotes the initial probability of state j, which we take to be the steady-state probability. The transition matrix M has one zero eigenvalue, corresponding to steady state. Since all transition matrices at thermodynamic equilibrium can be transformed into a symmetric transition matrix $\tilde{W}$ with the same eigenvalues [92], we know that this matrix is diagonalizable and will continue below with the representation using orthonormal eigenvectors, following references [33,92].

With the eigenvalues of M, λ_k, we can write $\frac{1}{\tau}\int_0^\tau e^{Mt}dt = \frac{1}{\tau}UAU^{-1}$, where A is a diagonal matrix, the k diagonal elements of which correspond to the integrals $\int_0^\tau e^{\lambda_k t}dt$ over all eigenvalues of M and where U is the matrix of eigenvectors of M. Since the first eigenvalue is zero, corresponding to the steady-state eigenvalue, we obtain $A_{11} = \tau$ and $A_{ii} = \frac{1}{\lambda_i}(e^{\lambda_i\tau} - 1)$ for $i > 1$. This allows us to find the expectation value of the time-averaged occupancy C,

$$\bar{C}(\tau) = \frac{1}{\tau}\sum_{i=1}^{2^N}\tilde{c}_i\left(\sum_{k=1}^{2^N}U_{ik}A_{kk}\sum_{j=1}^{2^N}(U^{-1})_{kj}\tilde{P}_j(0)\right). \tag{A11}$$

For small systems of $N = 10$ and $N = 6$ sites, Eq. (A11) can be solved semianalytically: we can obtain an exact analytic expression that is too complicated to write down. We plot this solution for the mean occupancy from this semianalytic treatment in Fig. 3(b). We note that the mean occupancies for two different J in Fig. 3(b) are calculated with a different value for ε_b, so that the mean occupancy shows half its maximal value, i.e., a value of 0.5, at $x = 0.47$. We choose the value of ε_b such that this is the case for every J [Fig. 7(d)]. Fitting Hill functions to the means obtained with Eq. (A11) for these parameter combinations shows that for increasing J the fitted Hill coefficient approaches the number of binding sites from below ($h \to N = 6$) for the binding site region [see Fig. 7(e)]; for the extended DNA sensor, we see that for high clustering strengths, higher Hill coefficients can be achieved.

We calculate the variance using Eq. (A4). As for the mean, we can obtain a semianalytic solution by solving the master equation, analogously to Eq. (A11). Reference [33] obtained an analytic expression for the variance for the system with $N = 6$ binding sites under periodic boundary conditions. We show the agreement of this limit and the semianalytic solution for the variance from our master equation treatment in Figs. 7(a) and 7(b) and discuss below how to obtain this limit for our system.

We follow Refs. [33] and [92], noting that a symmetric transition matrix, $\tilde{W}$, which shares eigenvalues with M, can be obtained from M via

$$\tilde{W}_{ij} = (\tilde{P}_j(0)/\tilde{P}_i(0))^{1/2}M_{ij}, \tag{A12}$$

where $\tilde{P}_i(0)$ corresponds to the probability of state i at equilibrium. The eigenvectors $\tilde{v}^{(k)}$ of $\tilde{W}$ are orthonormal, and can be related to the eigenvectors $v^{(k)}$ of M: $v_i^{(k)} = \sqrt{\tilde{P}_i(0)}\tilde{v}_i^{(k)}$ [92]. This implies that the elements of U^{-1} can be expressed in terms of the elements of U via $(U^{-1})_{ki}\tilde{P}_i(0) = U_{ik}$. With that, the probabilities $\tilde{P}_i(t)$ can be fully expressed in terms of the eigenvectors and eigenvalues of M. Using $\delta\tilde{c} = \tilde{c} - \langle\tilde{c}\rangle$ to denote deviations from the occupancy at steady state, we can then write the variance starting from Eq. (A4),

$$\begin{aligned}
\sigma_C^2 &= \frac{2}{\tau^2}\int_0^\tau dt\kappa(t)(\tau - t),\\
&= \frac{2}{\tau^2}\int_0^\tau dt\sum_{i=1}^{2^N}\delta\tilde{c}_i\left(\sum_{k=1}^{2^N}v_i^{(k)}e^{\lambda_k t}\sum_{j=1}^{2^N}v_j^{(k)}\delta\tilde{c}_j\right)(\tau - t),\\
&= \frac{2}{\tau^2}\sum_{k=1}^{2^N}\left(\sum_{i=1}^{2^N}\delta\tilde{c}_i v_i^{(k)}\right)\left(\sum_{j=1}^{2^N}\delta\tilde{c}_j v_j^{(k)}\right)\int_0^\tau dt e^{\lambda_k t}(\tau - t),\\
&= 2\sum_{k=2}^{2^N}\left(\sum_{i=1}^{2^N}\delta\tilde{c}_i v_i^{(k)}\right)^2\left(\frac{e^{\lambda_k\tau} - \lambda_k\tau - 1}{\tau^2\lambda_k^2}\right),
\end{aligned} \tag{A13}$$

where the terms $\sum_{i=1}^{2^N}\delta\tilde{c}_i v_i^{(k)}$ correspond to a projection of the fluctuations in occupancy along the kth eigenvector and the sum in the last line starts with the second eigenvector, as $\sum_{i=1}^{2^N}\delta\tilde{c}_i v_i^{(0)} = 0$. Finally, the instantaneous variance at equilibrium corresponds to the correlation function $\kappa(0)$ and can be expressed as

$$\sigma_{\mathrm{inst}}^2 = \sum_{k=2}^{2^N}\left(\sum_{i=1}^{2^N}\delta\tilde{c}_i v_i^{(k)}\right)^2. \tag{A14}$$

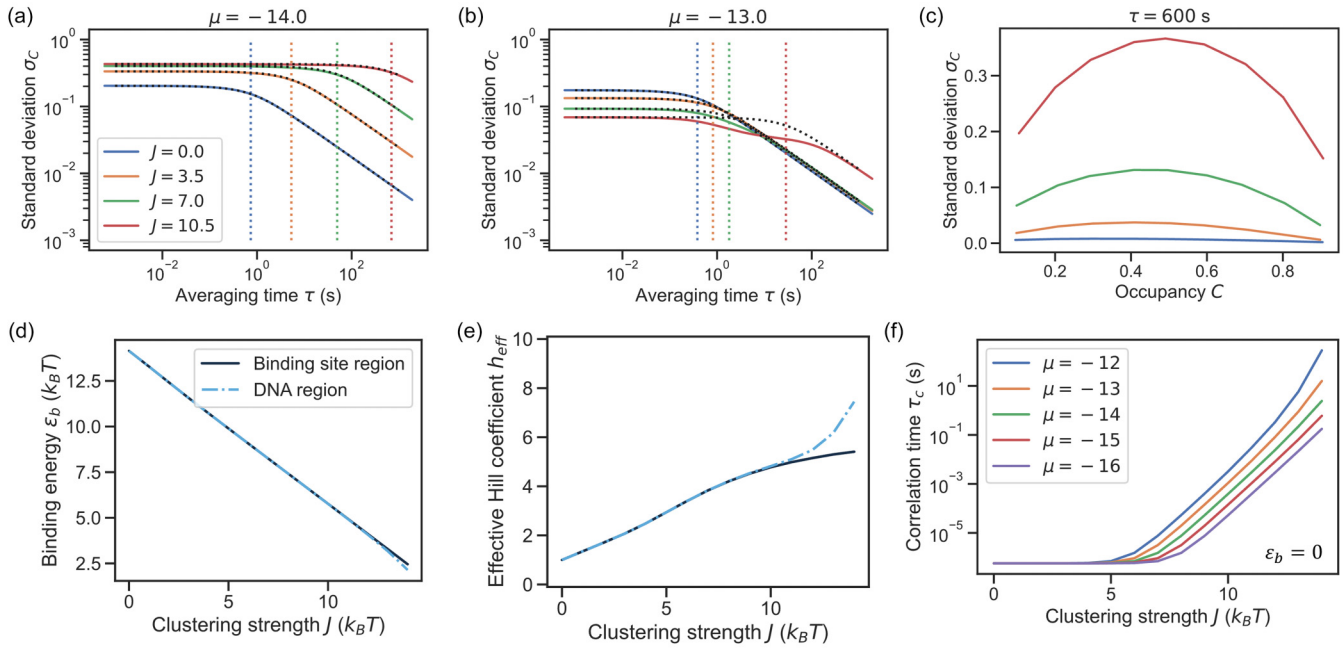

FIG. 7. Mean and standard deviation of the mechanistic model from the master equation and analytic approximations. (a), (b) Comparison of standard deviation from diagonalization of Eqs. (A10) and (A16) a value chemical potential μ corresponding to the anterior of the embryo, for different J; vertical bars indicate $2\tau_c$. We note that we shifted the value of ε_b for these simulations, as shown in panel *D*. (c) Standard deviation as a function of occupancy $\bar{C}$ shows characteristic $\bar{C}(1-\bar{C})$ shape and increases with clustering strength. (d) The binding energy ε_b that guarantees half occupancy at position $x \approx 0.47$ decreases almost linearly as a function of clustering strength J. (e) Hill function fits to the mean occupancy $\bar{C}$ show that the effective Hill coefficient h approaches the six binding sites from below but can exceed this number for an extended sensor with 10 sites. (f) The correlation time increases with clustering strengths throughout different values of chemical potential along the embryo.

While in principle all eigenvalues contribute to the sum, for long averaging times, we can only consider the contribution from the largest negative eigenvalue. Similarly, while the correlation time

$$\tau_c = \frac{1}{\sigma_{\text{inst}}^2}\int_0^\infty \kappa(t)dt \tag{A15}$$

in principle involves multiple timescales, for many parameter values it is a valid approximation to consider only the largest of the nonzero (negative) eigenvalues, meaning that $\tau_c \approx 1/\lambda_2$. Therefore, following Ref. [33], we find that

$$\sigma_C^2 \approx \sigma_{\text{inst}}^2 \mathcal{T}(\tau/\tau_c), \tag{A16}$$

with

$$\mathcal{T}(x) = 2\frac{e^{-x}+x-1}{x^2}. \tag{A17}$$

We show the match of this expression to our semianalytical calculation from master equation in Figs. 7(a) and 7(b), as a function of measurement time τ for different J and for a constant value of μ. The variance follows the expected behavior from Eq. (A16) for all J, with a clear switch from the instantaneous variance to the scale where it decays as approximately $1/\sqrt{\tau}$ at approximately $\tau \approx 2\tau_c$. We show the match of this semianalytical calculation and simulations in the main manuscript, for a single value of μ.

In Fig. 7(c) we show the variance for several values of J as a function of occupancy $\bar{C}$, which illustrates clearly the increase of the variance with J, due to the increase in the correlation time τ_c with J, in agreement with [33].

Finally, we show the correlation time from Eq. (A15) from the numerical diagonalization of the master equation for different values of chemical potential as a function of clustering strength J in Fig. 7(f). In agreement with previous calculations, the correlation time increases with clustering strength.

In the main text, we discuss the robust optimum at finite clustering strengths for short τ. In Fig. 8 we show three mean activation profiles to at values of J below, at, and beyond the optimum value of $I(C;x)$ at $J \approx 5$, to illustrate that the instantaneous variance is clearly in a regime where we are not yet in a large-noise limit, and where therefore an infinitely steep threshold is not optimal.

4. Concentration and occupancy sampling for estimating $P(C|s)$ using the mechanistic model

We model presence and absence of transcription factors with 1s and 0s to avoid symmetry between the presence and absence of transcription factors.

To estimate the conditional probabilties $P(C|s)$ from simulations of the mechanistic model for clustering transcription factors, both the Bicoid concentration and the occupancy must be discretized. To this end, we divide the concentration and occupancy ranges into bins, and represent each bin using the value from the center of that bin.

The range of mean Bicoid concentrations (0–147 nM) obtained from experimental data [Fig. 1(a)] was divided into

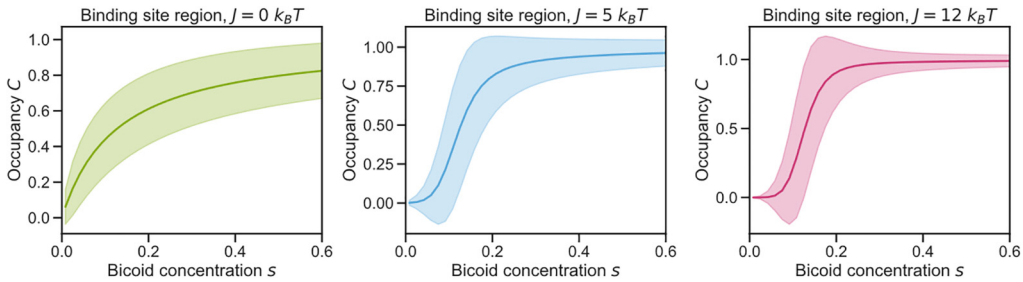

FIG. 8. Mean and standard deviation for three values of J for instantaneous measurement: both change with J in a way that presents a tradeoff for optimizing $I(C;x)$ (optimum at $J \approx 5$).

60 bins. Five additional bins were added, covering higher concentrations, to capture the fluctuations around the maximum mean concentration, resulting in a total of 65 bins covering a range of concentrations from 0 to 159 nM. This number of bins was chosen such that the mutual information between the position and concentration differed by less than 0.01 from it's limit as the number of bis goes to infinity (as estimated by the mutual information obtained with 10^5 bins). The analytical calculations were repeated with a higher concentration sampling using ten times as many bins, to confirm that the choice of concentration sampling does not affect our conclusions (data not shown).

The number of occupancy bins chosen for the histogram used to estimate $P(C|s)$ [Figs. 3(e) and 3(f)] influences the results obtained for the mutual information. When the number of bins is too low, the true distribution will not be accurately represented by the histogram, but if the number of bins is too high, the discrete nature of the simulations results will lead to artifacts in the final distribution. For this reason it is important to choose the right number of bins.

To this end, we compared the mutual information obtained using the histogram with that obtained using the Gaussian approximation for different numbers of bins. Since the Gaussian approximation is continuous, the mutual information results obtained using this approximation will converge to the true values as the number of bins increases. By comparing this to the mutual information calculated using the histogram, we can distinguish changes resulting from the mutual information converging to the true value from artifacts caused by the discrete nature of the simulation results.

Using this method we chose to use 400 bins for the histogram. For the Gaussian approximation, which does not suffer from the same type of artifacts we used 1000 bins. In the case of all-or-nothing expression described in Appendix A 5 100 bins were used.

To obtain a value of signal concentration s from the chemical potential μ, we calculate the average concentration on a free, nonbinding site. This sets the values of chemical potential that we use along the embryonal axis, x, allowing us to simulate the binding site sensor essentially in a grand-canonical ensemble. We use 160 positions between 0.1 and 0.9 embryo length. At each position x, the distribution of concentrations is assumed to be Gaussian; the mean concentration at each x is found by assuming that the maximum mean concentration $s = 1$ corresponds to 147 nM [57].

5. All-or-nothing expression

We assumed for almost all of the manuscript that gene expression regulated by the binding site sensor is proportional to the averaged occupancy at the sensor. This seems biologically relevant, as polymerases could be recruited proportional to this average occupancy, though of course this recruitment would add further noise into the process. Nevertheless, recent work on the hbP2 enhancer also discusses a scenario where gene expression (or polymerase recruitment) can only occur when the binding site is fully occupied [13].

We study this scenario numerically. We analyze our simulated trajectories for the long measurement time and instantaneous measurement, to obtain mean and variance for such a measurement C'. These can be used to calculate probability distributions $P(C'|s)$ as before. The mutual information $I(C';s)$ and $I(C';x)$ is shown in Fig. 9. For all measure-

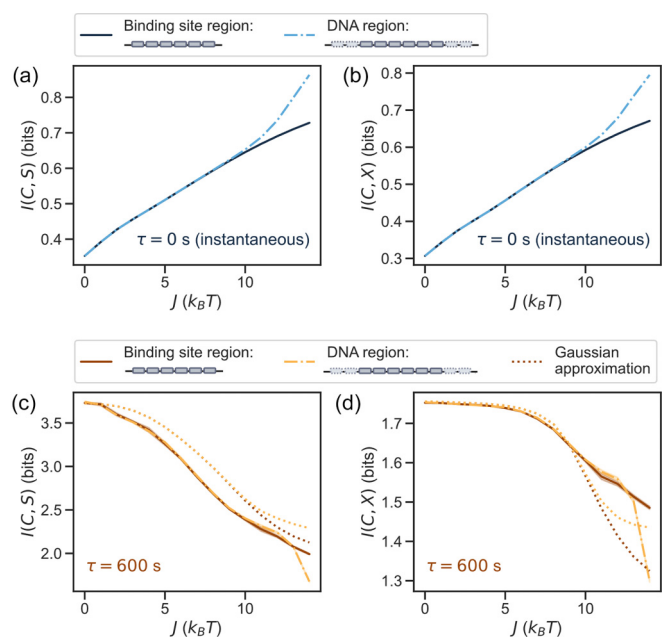

FIG. 9. Mutual information for a transcription factor measurement which is processed in "all-or-nothing" fashion, i.e., where occupancy is read out only in terms of two states: one state corresponding to a fully occupied binding site, that can activate gene expression downstream, and another state representing all other occupancy states.

ments, the values of information are slightly lower than for the measurement that reads out the time-averaged occupancy, since the all-or-nothing readout presents a course graining that reduces information, consistent with the data-processing inequality. For the instantaneous measurement, clustering increases both relevant information quantities for all clustering strengths studied here; for the longer measurement, both informations decrease, with a strong plateau for $I(C';x)$ for weak clustering strengths as in the main text. Therefore, we conclude that our result that weak clustering is information-theoretically acceptable applies also to this different readout protocol.

6. IB optimization

We find the optimal information bound following [45]: we iteratively optimize $P(C|s)$ from a random starting distribution, using

$$P(C|s) = \frac{P(C)}{Z(s,\lambda)} \exp\left[-\frac{1}{\lambda}\int dx P(x|s) \ln\left(\frac{P(x|s)}{P(x|C)}\right)\right], \tag{A18}$$

where Z is a normalization constant that ensures that $\int P(C|s)dC = 1$. For each value of the Lagrangian multiplier λ, we iterate until $P(C|s)$ converges; from this optimal $P(C|s)$ we can then calculate a particular point along the bound in the information plane. The optimal $P(C|s)$ is thus determined from the data distribution $P(s,x)$, without any physical or mechanistic constraints on $P(C|s)$. $I(C;s)$ and $I(C;x)$ can be calculated from the optimal $P(C|s)$. Practically, we perform a discrete sum instead of an integral. We also used an approximate expression for the information bound [44], in which both $I(C;x)$ and $I(C;s)$ can be expressed using only a changing Lagrangian multiplier λ and $I(s;x)$ (not shown). While it captures the bound approximately, it is slightly below the optimal bound and therefore also slightly below some binding site sensors, making it confusing for the purpose of this manuscript. To find the binding site sensors that also lie on the optimal bound, we search our sensors numerically: we first find all sensors with a particular $I(C;s)$ and then choose the sensor with the highest $I(C;x)$. This yields the white crosses in Figs. 4(b) and 10 (for $\tau = 600$ s and $\tau = 10$ s, respectively). In Fig. 10 we see the white path begin at smaller h than the optimum, which corresponds to the part of the optimal bound in Fig. 4(c) where $I(C;x)$ decreases for $I(C;s) > 2$. As $I(C;x)$ decreases from its largest possible value, the path in the binding site parameter plane moves away from the optimum, towards larger values of k and towards the boundary of the parameters we sample. When $I(C;s)$ just decreases away from the maximum, h increases, but then decreases as the $I(C;s)$ decreases further. Eventually, the path arrives at the maximal value for k we consider; if we increase k further, the path continues along its current trajectory. However, in this case the binding site sensors along the optimal bound yield a mean occupancy that does not reach $\bar{C} = 1$ even at high s, i.e., these binding sites are never fully occupied. Since we believe that such binding sites are no longer sensible, we typically only consider binding sites with a maximal normalized occupancy $\bar{C} = 0.6$; if we do so, the path continues towards significantly higher h for almost constant k. This behavior is again in principle consistent with the idea that lower $I(C;s)$ implies higher h, though we do note that we have effectively imposed an additional constraint on the binding site occupancy; without this constraint, h would continue to decrease slightly and k would increase significantly, which implies an unphysically small k_{on}.

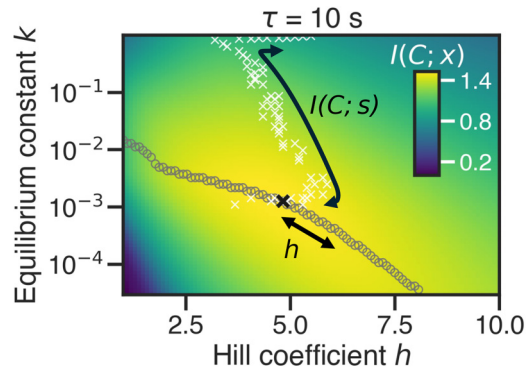

FIG. 10. The information $I(C;x)$ as a function of binding site parameters h and k with symbols marking trajectories for a given constraint, for shorter $\tau = 10$ s compared to Fig. 4(b). If h are constrained to varying values of h, the optimal $I(C;x)$ lies along the circles, with lower k for higher h. If $I(C;s)$ is constrained, the optimal $I(C;x)$ is marked with white crosses. The path of white crossed begins beyond the optimum [see blue line in Fig. 4(c), which decreases as $I(C;s)$ is pushed beyond two bits]. The maximum $I(C;x)$ is marked with a black cross.

We hope to show with this trajectory in Figs. 4(b) and 10 that the constraints imposed matter significantly for finding the optimal binding site parameters.

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

[2] H. B. Barlow, *et al.*, Possible principles underlying the transformation of sensory messages, Sensory Commun. **1**, 217 (1961).

[3] S. E. Palmer, O. Marre, M. J. Berry, and W. Bialek, Predictive information in a sensory population, Proc. Natl. Acad. Sci. USA **112**, 6908 (2015).

[4] D. E. Koshland Jr., Bacterial chemotaxis in relation to neurobiology, Annu. Rev. Neurosci. **3**, 43 (1980).

[5] H. H. Mattingly, K. Kamino, B. B. Machta, and T. Emonet, *Escherichia coli* chemotaxis is information limited, Nat. Phys. **17**, 1426 (2021).

[6] A. J. Tjalma, V. Galstyan, J. Goedhart, L. Slim, N. B. Becker, and P. R. ten Wolde, Trade-offs between cost and information in cellular prediction, Proc. Natl. Acad. Sci. USA **120**, e2303078120 (2023).

[7] G. Reddy, A. Celani, and M. Vergassola, Infomax strategies for an optimal balance between exploration and exploitation, J. Stat. Phys. **163**, 1454 (2016).

[8] G. Tkačik, C. G. Callan Jr., and W. Bialek, Information flow and optimization in transcriptional regulation, Proc. Natl. Acad. Sci. USA **105**, 12265 (2008).

[9] J. Selimkhanov, B. Taylor, J. Yao, A. Pilko, J. Albeck, A. Hoffmann, L. Tsimring, and R. Wollman, Accurate information transmission through dynamic biochemical signaling networks, Science **346**, 1370 (2014).

[10] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, Information transduction capacity of noisy biochemical signaling networks, Science **334**, 354 (2011).

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

[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] J. Desponds, M. Vergassola, and A. M. Walczak, A mechanism for hunchback promoters to readout morphogenetic positional information in less than a minute, eLife **9**, e49758 (2020).

[14] M. Bauer, M. Petkova, T. Gregor, E. F. Wieschaus, and W. Bialek, Trading bits in the readout from a genetic network, Proc. Natl. Acad. Sci. USA **118**, e2109011118 (2021).

[15] T. R. Sokolowski, T. Gregor, W. Bialek, and G. Tkačik, Deriving a genetic regulatory network from an optimization principle, Proc. Natl. Acad. Sci. USA **122**, e2402925121 (2025).

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

[17] W. Driever and C. Nüsslein-Volhard, A gradient of bicoid protein in *drosophila* embryos, Cell **54**, 83 (1988).

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

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

[20] 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, Proc. Natl. Acad. Sci. USA **121**, e2403265121 (2024).

[21] E. E. Furlong and M. Levine, Developmental enhancers and chromosome topology, Science **361**, 1341 (2018).

[22] P.-T. Chen, M. Levo, B. Zoller, and T. Gregor, A conserved coupling of transcriptional on and off periods underlies bursting dynamics, Nat. Struct. Mol. Biol. **32**, 1959 (2025).

[23] S. Fallacaro, A. Mukherjee, M. A. Turner, H. G. Garcia, and M. Mir, Transcription factor hubs exhibit gene-specific properties that tune expression, bioRxiv (2025).

[24] R. Munshi, J. Ling, S. Ryabichko, E. F. Wieschaus, and T. Gregor, Transcription factor clusters as information transfer agents, Sci. Adv. **11**, eadp3251 (2025).

[25] M. Mir, M. R. Stadler, S. A. Ortiz, C. E. Hannon, M. M. Harrison, X. Darzacq, and M. B. Eisen, Dynamic multifactor hubs interact transiently with sites of active transcription in *Drosophila* embryos, eLife **7**, e40497 (2018).

[26] M. Mir, A. Reimer, J. E. Haines, X.-Y. Li, M. Stadler, H. Garcia, M. B. Eisen, and X. Darzacq, Dense bicoid hubs accentuate binding along the morphogen gradient, Genes Dev. **31**, 1784 (2017).

[27] J. Dufourt, A. Trullo, J. Hunter, C. Fernandez, J. Lazaro, M. Dejean, L. Morales, S. Nait-Amer, K. N. Schulz, M. M. Harrison, C. Favard, O. Radulescu, and M. Lagha, Temporal control of gene expression by the pioneer factor zelda through transient interactions in hubs, Nat. Commun. **9**, 5194 (2018).

[28] H. C. Berg and E. M. Purcell, Physics of chemoreception, Biophys. J. **20**, 193 (1977).

[29] W. Bialek and S. Setayeshgar, Physical limits to biochemical signaling, Proc. Natl. Acad. Sci. USA **102**, 10040 (2005).

[30] P. R. ten Wolde, N. B. Becker, T. E. Ouldridge, and A. Mugler, Fundamental limits to cellular sensing, J. Stat. Phys. **162**, 1395 (2016).

[31] K. Kaizu, W. H. de Ronde, J. Paijmans, K. Takahashi, F. Tostevin, and P. R. ten Wolde, The Berg-Purcell limit revisited, Biophys. J. **106**, 976 (2014).

[32] W. Bialek and S. Setayeshgar, Cooperativity, sensitivity, and noise in biochemical signaling., Phys. Rev. Lett. **100**, 258101 (2008).

[33] M. Skoge, Y. Meir, and N. S. Wingreen, Dynamics of cooperativity in chemical sensing among cell-surface receptors, Phys. Rev. Lett. **107**, 178101 (2011).

[34] M. Skoge, S. Naqvi, Y. Meir, and N. S. Wingreen, Chemical sensing by nonequilibrium cooperative receptors, Phys. Rev. Lett. **110**, 248102 (2013).

[35] D. T. McSwiggen, M. Mir, X. Darzacq, and R. Tjian, Evaluating phase separation in live cells: Diagnosis, caveats, and functional consequences, Genes Dev. **33**, 1619 (2019).

[36] D. Lebrecht, M. Foehr, E. Smith, F. J. P. Lopes, C. E. Vanario-Alonso, J. Reinitz, D. S. Burz, and S. D. Hanes, Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*, Proc. Natl. Acad. Sci. USA **102**, 13176 (2005).

[37] A. A. Hyman, C. A. Weber, and F. Jülicher, Liquid-liquid phase separation in biology, Annu. Rev. Cell Dev. Biol. **30**, 39 (2014).

[38] Y. Shin and C. P. Brangwynne, Liquid phase condensation in cell physiology and disease, Science **357**, eaaf4382 (2017).

[39] D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp, A phase separation model for transcriptional control, Cell **169**, 13 (2017).

[40] B. R. Sabari, A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, I. I. Cisse, R. G. Roeder *et al.*, Coactivator condensation at super-enhancers links phase separation and gene control, Science **361**, eaar3958 (2018).

[41] C. E. Hannon and M. B. Eisen, Intrinsic protein disorder is insufficient to drive subnuclear clustering in embryonic transcription factors, eLife **12**, RP88221 (2024).

[42] G. Malaguti and P. R. ten Wolde, Theory for the optimal detection of time-varying signals in cellular sensing systems, eLife **10**, e62574 (2021).

[43] T. Mora and I. Nemenman, Physical limit to concentration sensing in a changing environment, Phys. Rev. Lett. **123**, 198101 (2019).

[44] M. Bauer and W. Bialek, Information bottleneck in molecular sensing, Phys. Rev. X **1**, 023005 (2023).

[45] N. Tishby, F. C. Pereira, and W. Bialek, The information bottleneck method, in *Proceedings of the 37th Annual Allerton Conference on Communication Control and Computing*, edited by B. Hajek and R. S. Sreenivas (University of Illinois Press, Illinois, Urbana, 2000), pp. 368–377.

[46] A. Klosin, F. Oltsch, T. Harmon, A. Honigmann, F. Jülicher, A. A. Hyman, and C. Zechner, Phase separation provides a mechanism to reduce noise in cells, Science **367**, 464 (2020).

[47] D. Deviri and S. A. Safran, Physical theory of biological noise buffering by multicomponent phase separation, Proc. Natl. Acad. Sci. USA **118**, e2100099118 (2021).

[48] M. Bauer, W. Bialek, C. Goddard, C. M. Holmes, K. Krishnamurthy, S. E. Palmer, R. Pang, D. J. Schwab, and L. Susman, Optimization and variability can coexist, arXiv:2505.23398.

[49] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, Universally sloppy parameter sensitivities in systems biology models, PLoS Comput. Biol. **3**, e189 (2007).

[50] J. A. Morin, S. Wittmann, S. Choubey, A. Klosin, S. Golfier, A. A. Hyman, F. Jülicher, and S. W. Grill, Sequence-dependent surface condensation of a pioneer transcription factor on DNA, Nat. Phys. **18**, 271 (2022).

[51] P. Hillenbrand, U. Gerland, and G. Tkačik, Beyond the french flag model: Exploiting spatial and gene regulatory interactions for positional information, PLoS ONE **11**, e0163628 (2016).

[52] J. P. Bothma, M. R. Norstad, S. Alamos, and H. G. Garcia, Llamatags: A versatile tool to image transcription factor dynamics in live embryos, Cell **173**, 1810 (2018).

[53] S. C. Little, M. Tikhonov, and T. Gregor, Precise developmental gene expression arises from globally stochastic transcriptional activity, Cell **154**, 789 (2013).

[54] J. Park, J. Estrada, G. Johnson, B. J. Vincent, C. Ricci-Tam, M. D. Bragdon, Y. Shulgina, A. Cha, Z. Wunderlich, J. Gunawardena, and A. H. DePace, Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity, eLife **8**, e41266 (2019).

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

[56] M. Kato and S. L. McKnight, A solid-state conceptualization of information transfer from gene to message to protein, Annu. Rev. Biochem. **87**, 351 (2018).

[57] A. Abu-Arish, A. Porcher, A. Czerwonka, N. Dostatni, and C. Fradin, High mobility of bicoid captured by fluorescence correlation spectroscopy: Implication for the rapid establishment of its gradient, Biophys. J. **99**, L33 (2010).

[58] J. Ling, T. S. K. Y. Umezawa, and S. Small, Bicoid dependent activation of the target gene hunchback requires a two-motif sequence code in a specific basal promoter, Mol. Cell **75**, 1178 (2019).

[59] E. Eck, J. Liu, M. Kazemzadeh-Atoufi, S. Ghoreishi, S. A. Blythe, and H. G. Garcia, Quantitative dissection of transcription in development yields evidence for transcription-factor-driven chromatin accessibility, eLife **9**, e56429 (2020).

[60] I. Neri and S. Pigolotti, Amendable decisions in living systems, bioRxiv (2025).

[61] M. W. Perry, A. N. Boettiger, J. P. Bothma, and M. Levine, Shadow enhancers foster robustness of *drosophila* gastrulation, Curr. Biol. **20**, 1562 (2010).

[62] J. P. Bothma, H. G. Garcia, S. Ng, M. W. Perry, T. Gregor, and M. Levine, Enhancer additivity and non-additivity are determined by enhancer strength in the *drosophila* embryo, eLife **4**, e07956 (2015).

[63] F. M. Weinert, R. C. Brewster, M. Rydenfelt, R. Phillips, and W. K. Kegel, Scaling of gene expression with transcription-factor fugacity, Phys. Rev. Lett. **113**, 258101 (2014).

[64] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips, Transcriptional regulation by the numbers: Applications, Curr. Opin. Genet. Dev. **15**, 125 (2005).

[65] J. Estrada, F. Wong, A. DePace, and J. Gunawardena, Information integration and energy expenditure in gene regulation, Cell **166**, 234 (2016).

[66] C. Scholes, A. H. DePace, and Á. Sánchez, Combinatorial gene regulation through kinetic control of the transcription cycle, Cell Syst. **4**, 97 (2017).

[67] F. Wong and J. Gunawardena, Gene regulation in and out of equilibrium, Annu. Rev. Biophys. **49**, 199 (2020).

[68] B. Zoller, T. Gregor, and G. Tkačik, Eukaryotic gene regulation at equilibrium, or non? Current Opin. Syst. Biol. **31**, 100435 (2022).

[69] R. Shelansky, S. Abrahamsson, C. R. Brown, M. Doody, T. L. Lenstra, D. R. Larson, and H. Boeger, Single gene analysis in yeast suggests nonequilibrium regulatory dynamics for transcription, Nat. Commun. **15**, 6226 (2024).

[70] G. Tkačik and P. R. ten Wolde, Information processing in biochemical networks, Annu. Rev. Biophys. **54**, 249 (2025).

[71] B. Zoller, A. Bénichou, T. Gregor, and G. Tkačik, Invariant non-equilibrium dynamics of transcriptional regulation optimize information flow, arXiv:2507.12395.

[72] T. Mijatović, A. R. Kok, M. Brüggen, J. W. Zwanikken, and M. Bauer (2025), https://github.com/bauer-group-delft/binding.

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

[74] A. V. Hill, The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves, J. Physiol. **40**, 4 (1910).

[75] H. D. Jong, Modeling and simulation of genetic regulatory systems: A literature review., J. Comput. Biol. **9**, 67 (2002).

[76] J. N. Weiss, The hill equation revisited: Uses and misuses, FASEB J. **11**, 835 (1997).

[77] G. Tkačik and A. Walczak, Information transmission in genetic regulatory networks: A review, J. Phys.: Condens. Matter **23**, 153102 (2011).

[78] S. Marzen, H. G. Garcia, and R. Phillips, Statistical mechanics of Monod–Wyman–Changeux (MWC) models, J. Mol. Biol. **425**, 1433 (2013).

[79] M. Bauer, How does an organism extract relevant information from transcription factor concentrations?, Biochem. Soc. Trans. **50**, 1365 (2022).

[80] R. Martinez-Corral, K.-M. Nam, A. H. DePace, and J. Gunawardena, The hill function is the universal Hopfield barrier for sharpness of input–output responses, Proc. Natl. Acad. Sci. USA **121**, e2318329121 (2024).

[81] R. G. Endres and N. S. Wingreen, Maximum likelihood and the single receptor, Phys. Rev. Lett. **103**, 158101 (2009).

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

[83] H. H. Mattingly, M. K. Transtrum, M. C. Abbott, and B. B. Machta, Maximizing the information learned from finite data selects a simple model, Proc. Natl. Acad. Sci. USA **115**, 1760 (2018).

[84] O. Witteveen, S. J. Rosen, R. S. Lach, M. Z. Wilson, and M. Bauer, Optimizing information transmission in the canonical Wnt pathway, arXiv:2506.22633.

[85] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons, Hoboken, 2012).

[86] R. J. Glauber, Time-dependent statistics of the Ising model, J. Math. Phys. **4**, 294 (1963).

[87] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. **22**, 403 (1976).

[88] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. **81**, 2340 (1977).

[89] G. Fernandes, H. Tran, M. Andrieu, Y. Diaw, C. Perez Romero, C. Fradin, M. Coppey, A. M. Walczak, and N. Dostatni, Synthetic reconstruction of the hunchback promoter specifies the role of bicoid, zelda and hunchback in the dynamics of its transcription, eLife **11**, e74509 (2022).

[90] T. Gregor, E. F. Wieschaus, A. P. McGregor, W. Bialek, and D. W. Tank, Stability and nuclear dynamics of the bicoid morphogen gradient, Cell **130**, 141 (2007).

[91] A. Porcher, A. Abu-Arish, S. Huart, B. Roelens, C. Fradin, and N. Dostatni, The time to measure positional information: Maternal Hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription, Development **137**, 2795 (2010).

[92] C. Zhang and R. S. Berry, Time autocorrelation function analysis of master equation and its application to atomic clusters, J. Chem. Phys. **123**, 094103 (2005).

[93] J. Trojanowski, L. Frank, A. Rademacher, N. Mücke, P. Grigaitis, and K. Rippe, Transcription activation is enhanced by multivalent interactions independent of phase separation, Mol. Cell **82**, 1878 (2022).

[94] K. Rippe, Liquid–liquid phase separation in chromatin, Cold Spring Harbor Perspect. Biol. **14**, a040683 (2022).

[95] B. W. Pontius, Close encounters: Why unstructured, polymeric domains can increase rates of specific macromolecular association, Trends Biochem. Sci. **18**, 181 (1993).

[96] G. Ravanelli, K.-M. Nam, J. Gunawardena, and R. Martinez-Corral, Decoupling between activation time and steady-state level in input-output responses, bioRxiv (2025).

[97] A. J. Tjalma and P. R. ten Wolde, Predicting concentration changes via discrete receptor sampling, Phys. Rev. Res. **6**, 033049 (2024).

[98] G. Nicoletti and D. M. Busiello, Tuning transduction from hidden observables to optimize information harvesting, Phys. Rev. Lett. **133**, 158401 (2024).

[99] A. Gao, K. Shrinivas, P. Lepeudry, H. I. Suzuki, P. A. Sharp, and A. K. Chakraborty, Evolution of weak cooperative interactions for biological specificity, Proc. Natl. Acad. Sci. USA **115**, E11053 (2018).

[100] S. Chong, T. G. Graham, C. Dugast-Darzacq, G. M. Dailey, X. Darzacq, and R. Tjian, Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription, Mol. Cell **82**, 2084 (2022).

[101] L. Galbusera, G. Bellement-Theroue, T. Julou, and E. van Nimwegen, Rapid transcription factor fluctuations drive nonequilibrium gene regulatory dynamics in bacteria, Phys. Rev. X **3**, 033006 (2025).

[102] B. Lim, C. J. Dsilva, T. J. Levario, H. Lu, T. Schüpbach, I. G. Kevrekidis, and S. Y. Shvartsman, Dynamics of inductive erk signaling in the *drosophila* embryo, Curr. Biol. **25**, 1784 (2015).

[103] T. T. Harden, B. J. Vincent, and A. H. DePace, Transcriptional activators in the early *drosophila* embryo perform different kinetic roles, Cell Syst. **14**, 258 (2023).

[104] W.-K. Cho, J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, and I. I. Cisse, Mediator and RNA polymerase II clusters associate in transcription-dependent condensates, Science **361**, 412 (2018).

[105] T. Mittag and R. V. Pappu, A conceptual framework for understanding phase separation and addressing open questions and challenges, Mol. Cell **82**, 2201 (2022).

[106] M. Reinhardt, G. Tkačik, and P. R. ten Wolde, Path weight sampling: Exact Monte Carlo computation of the mutual information between stochastic trajectories, Phys. Rev. X **13**, 041017 (2023).

[107] A.-L. Moor and C. Zechner, Dynamic information transfer in stochastic biochemical networks, Phys. Rev. Res. **5**, 013032 (2023).

[108] T. R. Sokolowski and G. Tkačik, Optimizing information flow in small genetic networks. IV. Spatial coupling, Phys. Rev. E **91**, 062710 (2015).

[109] K. N. Schulz and M. M. Harrison, Mechanisms regulating zygotic genome activation, Nat. Rev. Genet. **20**, 221 (2019).

[110] A. Birnie, A. Plat, C. Korkmaz, and J. P. Bothma, Precisely timed regulation of enhancer activity defines the binary expression pattern of Fushi tarazu in the *drosophila* embryo, Curr. Biol. **33**, 2839 (2023).

[111] B. J. Beliveau, A. N. Boettiger, M. S. Avendaño, R. Jungmann, R. B. McCole, E. F. Joyce, C. Kim-Kiselak, F. Bantignies, C. Y. Fonseka, J. Erceg *et al.*, Single-molecule super-resolution imaging of chromosomes and *in situ* haplotype visualization using oligopaint fish probes, Nat. Commun. **6**, 7147 (2015).

[112] A. M. C. Gizzi, D. I. Cattoni, J.-B. Fiche, S. M. Espinola, J. Gurgo, O. Messina, C. Houbron, Y. Ogiyama, G. L. Papadopoulos, G. Cavalli *et al.*, Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms, Mol. Cell **74**, 212 (2019).

[113] C. W. Gardiner, *Elements of Stochastic Methods* (AIP Publishing, Melville, NY, 2021).

[114] A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, and X. Darzacq, Ctcf and cohesin regulate chromatin loop stability with distinct dynamics, eLife **6**, e25776 (2017).

[115] J. Chen, Z. Zhang, L. Li, B.-C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig *et al.*, Single-molecule dynamics of enhanceosome assembly in embryonic stem cells, Cell **156**, 1274 (2014).

[116] J. M. Kim, P. Visanpattanasin, V. Jou, S. Liu, X. Tang, Q. Zheng, K. Y. Li, J. Snedeker, L. D. Lavis, T. Lionnet *et al.*, Single-molecule imaging of chromatin remodelers reveals role of atpase in promoting fast kinetics of target search and dissociation from chromatin, eLife **10**, e69387 (2021).

[117] C. Fountas and T. L. Lenstra, Better together: How cooperativity influences transcriptional bursting, Curr. Opin. Genet. Dev. **89**, 102274 (2024).

Correction: Figure 1 formatting errors have been fixed.