

TEMPORAL PRECISION IN GENE EXPRESSION AND CELL MIGRATION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Shivam Gupta

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2021

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Andrew Mugler, Chair

Department of Physics and Astronomy

Dr. Srividya Iyer-Biswas

Department of Physics and Astronomy

Dr. David D. Nolte

Department of Physics and Astronomy

Dr. Martin Kruczenski

Department of Physics and Astronomy

Approved by:

Dr. John P. Finley

Head of the Graduate Program

ACKNOWLEDGMENTS

We acknowledge our collaborators Hendrik Korswagen (Hubrecht Institute, Netherlands) and Marie-Anne Félix (Institut de Biologie de l'Ecole Normale Supérieure, CNRS) and their Ph.D. students Erik Schild, Euclides E. Fernandes Póvoa and Clément Dubois for providing experimental data and insight described in chapter 4 and chapter 5 respectively. We also thank Ilya Nemenman for insight in chapter 6.

This work was supported by Human Frontier Science Program grant RGP0030/2016 and Simons Foundation grant 376198.

TABLE OF CONTENTS

	Page
ABBREVIATIONS	vi
ABSTRACT	vii
1. INTRODUCTION	1
2. THESIS PLAN	3
2.1 Temporal precision of regulated gene expression	3
2.2 Temporal precision of molecular events with regulation and feedback . .	4
2.3 Termination of QR.p cell migration	7
2.4 Detection of nutrient wave by single-celled organisms	8
3. TEMPORAL PRECISION OF REGULATED GENE EXPRESSION	9
3.1 Results	10
3.1.1 Regulation increases temporal precision	12
3.1.2 Optimal regulation balances regulator and target noise	13
3.1.3 Model predictions are consistent with neuroblast migration data	17
3.1.4 Results are robust to additional complexities including cell division	19
3.2 Discussion	22
3.3 Materials and methods	24
3.3.1 Computation of the first-passage time statistics	24
3.3.2 Deterministic dynamics	26
3.3.3 Details of the analytic approximations	27
3.3.4 Analysis of the experimental data	28
4. TEMPORAL PRECISION OF MOLECULAR EVENTS WITH REGULA- TION AND FEEDBACK	30
4.1 Introduction	30
4.2 Results	31

	Page
4.3 Experimental observation in QR.p cells in <i>C. elegans</i>	41
4.4 Discussion	44
4.5 Materials and methods	47
4.5.1 Calculation of the moments of the first passage time	47
4.5.2 Analytic minimization of timing variance using Lagrange multipliers	49
4.5.3 Derivation of the lower bound on timing variance	52
5. TERMINATION OF QR.P CELL MIGRATION	54
5.1 Results	56
5.1.1 Mathematical model for no compensation hypothesis	56
5.1.2 Mathematical model for partial compensation hypothesis	58
5.2 Discussion	61
6. TRAVELING WAVE DETECTION BY SINGLE-CELLED ORGANISMS	63
6.1 Mathematical model for traveling wave sensing	64
6.2 Discussion	69
6.3 Materials and methods	69
6.3.1 Calculation of the probability distribution for the next reaction	69
REFERENCES	72

ABBREVIATIONS

WT wild type

C. elegans *Caenorhabditis elegans*

ABSTRACT

Gupta, Shivam Ph.D., Purdue University, May 2021. Temporal Precision in Gene Expression and Cell Migration. Major Professor: Andrew Mugler Professor.

Important cellular processes such as migration, differentiation, and development often rely on precise timing. Yet, the molecular machinery that regulates timing is inherently noisy. How do cells achieve precise timing with noisy components? We investigate this question using a first-passage-time approach, for an event triggered by a molecule that crosses an abundance threshold. We investigate regulatory strategies that decrease the timing noise of molecular events. We look at several strategies which decrease the noise: i) Regulation performed by an accumulating activator, ii) Regulation due to a degrading repressor, iii) Auto-regulation and the presence of feedback. We find that either activation or repression outperforms an unregulated strategy. The optimal regulation corresponds to a nonlinear increase in the amount of the target molecule over time, arises from a tradeoff between minimizing the timing noise of the regulator and that of the target molecule itself, and is robust to additional effects such as bursts and cell division. Our results are in quantitative agreement with the nonlinear increase and low noise of *mig-1* gene expression in migrating neuroblast cells during *Caenorhabditis elegans* development. These findings suggest that dynamic regulation may be a simple and powerful strategy for precise cellular timing.

Autoregulatory feedback increases noise. Yet, we find that in the presence of regulation by a second species, autoregulatory feedback decreases noise. To explain this finding, we develop a method to calculate the optimal regulation function that minimizes the timing noise. Our method reveals that the combination of feedback and regulation minimizes noise by maximizing the number of molecular events that must happen in sequence before a threshold is crossed. We compute the optimal timing

precision for all two-node networks with regulation and feedback, derive a generic lower bound on timing noise, and compare our results with the neuroblast migration during *C. elegans* development, as well as two mutants. We find that indeed our model is aligned with the experimental findings.

Furthermore, we apply our framework of temporal regulation to explain how the stopping point of the migrating cells in *C. elegans* depends on the body size. Considering temporal regulation, we find the termination point of the cell for various larval sizes. We discuss three possible mechanisms: i) No compensation; here the migration velocity is constant across the mutants of *C. elegans*, and this results in the migration distance to be constant but the relative position to be different across various sizes; ii) Total compensation; here the velocity is compensated with body size, hence resulting in the same relative position of cells across mutants; and iii) Partial compensation; here the velocity of migration is correlated with body size to some degree, resulting in a non-linear relationship between termination point and body size. We find that our partial compensation model is consistent with experimental observations of cell termination.

Finally, we look at the detection of traveling waves by single-celled organisms. Cells must use temporal and spatial information to sense the direction of traveling waves, as seen in cAMP detection by the amoeba *Dictyostelium*. If a cell only uses spatial information to sense the direction of the wave then the cell will move forward when the wave hits the front of the cell, and move backward when the wave hits the back of the cell, resulting in neutral movement. Cells must use temporal information along with spatial information to effectively move towards the source. Here we develop a mechanism by which cells are able to integrate the spatial and temporal information through a system of inhibitors. We find the optimal time to release the inhibitors for maximizing the precision of directional sensing.

1. INTRODUCTION

Proper timing is crucial for biological processes, including cell division [1–3], cell differentiation [4], cell migration [5], viral infection [6], embryonic development [7, 8], and cell death [9]. These processes are governed by molecular events inside cells, i.e., production, degradation, and interaction of molecules. Molecular events are subject to unavoidable fluctuations, because molecule numbers are small and reactions occur at random times [10, 11]. Cells combat these fluctuations using networks of regulatory interactions among molecular species. This raises the fundamental question of whether there exist regulatory strategies that maximize the temporal precision of molecular events and, in turn, cellular behaviors.

A canonical mechanism by which a molecular event triggers a cellular behavior is accumulation to a threshold [3, 4, 12–14]: molecules are steadily produced by the cell, and once the molecule number crosses a particular threshold, the behavior is initiated. The temporal precision of the behavior is therefore bounded by the temporal precision of the threshold crossing. Threshold crossing has been shown to underlie cell cycle progression [3] and sporulation [4], although alternative strategies, such as derivative [9] or integral thresholding [15], oscillation [16], and dynamical transitions in the regulatory network [8], have also been investigated.

Recent work has investigated the impact of auto-regulation (i.e., feedback) on the temporal precision of threshold crossing [12, 13]. Interestingly, it was found that auto-regulation generically decreases the temporal precision of threshold crossing, meaning that the optimal strategy is a linear increase of the molecule number over time with no auto-regulation [12] (although auto-regulation can help if there is a large timescale separation and the threshold itself is also subject to optimization [13]). Indeed, even when the molecule also degrades, the optimal precision is achieved when positive auto-regulation counteracts the effect of degradation, preserving the linear increase

over time [12]. However, in many biological processes, such as the temporal control of neuroblast migration in *Caenorhabditis elegans* [5], the molecular species governing the behavior increases nonlinearly over time. This suggests that other regulatory interactions beyond auto-regulation may play an important role in determining temporal precision. In particular, the impact of activation and repression on temporal precision, where the activator or repressor has its own stochastic dynamics, remains unclear.

We apply the framework of first passage time statistics to describe the crossing of a threshold number in gene production, in order to understand the cell-to-cell variability in migration time of the neuroblast in *C. elegans*. We calculate the timing noise when regulation by a second gene and/or auto-regulation are incorporated. We use master equation framework for solving a discrete probabilistic gene system. We use analytic and numerical methods to determine the timing noise of gene networks. We find that regulation by a second gene increases the precision in time measurement by a cell. Interestingly we show that when auto-regulation is present by itself it increases the timing noise; however, if auto-regulation is present along with regulation by a second gene, the timing noise is further decreased than by regulation alone. Our collaborators at Hubrecht University and Institut de Biologie de l'Ecole Normale Supérieure, CNRS have performed experiments to test our predictions. In one experiment performed, we have shown that the observed timing noise in migrating neuroblasts can be attributed to a regulator mechanism as we describe [17]. We extend our method to explain the observed stopping point of a cell during its migration inside *C. elegans* [18].

The tools and insights we gathered from our previous studies can be applied to a vast variety of biological problems. For example, many single-celled organisms must sense the direction of chemical waves. We develop a theoretical mechanism by which cells can sense the direction of incoming nutrient waves, and which explains cell movement towards the nutrient source.

2. THESIS PLAN

Here, we briefly describe the basic components of the thesis. Further details on each component will follow in Chapters 3-6.

Cells keep track of time through molecular events such as production, degradation and interaction of molecules. Molecular events are inherently noisy, which generates noise in the timing of biological processes. Cells combat this noise using gene regulatory networks. In collaboration with two experimental groups, we study the regulatory networks that govern timing precision in migrating neuroblast cells called QR.p (Figure 2.1A, pink dot) in the roundworm *C.elegans*. The QR.p cells are present at the tail position when *C. elegans* is in early stage of development. These cells migrate from the tail position to the head position during development. Our collaborators have shown that in mutants of *C. elegans*, migration speed of QR.p is positively related to *elg-20* gene expression. Over expressing of *elg-20* leads to over migration of QR.p and under-expression of *elg-20* leads to under-migration of QR.p. Their experimental evidence suggests that the migration of QR.p cells happens only for fixed amount of time. This raises the fundamental question of how, and with what precision, the cells keep track of time. For my PhD thesis work, I have worked on four projects explaining the timing precision of cell migration and the final position of QR.p cell's descendant in *C. elegans*.

2.1 Temporal precision of regulated gene expression

Experiments on the migration of QR.p cells suggest that the migration is controlled by the number of *mig-1* mRNAs produced by QR.p cells. Cells start migrating after the *mig-1* production starts and they stop migrating when the *mig-1* crosses

threshold molecule number y^* (Figure 2.1B). *mig-1* production is regulated by a number of genes. We study activator and repressor gene networks and their effect on timing precision. In the activator mechanism, the production rate of *mig-1* increases as the number of activator molecules increase, whereas in the repressor mechanism, the production rate of *mig-1* decreases as the repressor molecules increase. For these two models we calculate the trajectory of *mig-1* expression and timing variance in crossing the threshold molecule number. To compare our analytical and numerical results with the experiments we define a linearity parameter (ρ) which measures the non-linear rise in the number of *mig-1* molecules. We calculate the timing variance as a function of linearity (Figure 2.1C) and compare with the experimental values. The increase in precision due to regulation is a consequence of a tradeoff between reducing the extrinsic noise of the regulator, and reducing the intrinsic noise of the target molecule itself. Our minimal model is sufficient to explain both the high degree of nonlinearity and the low degree of noise in the dynamics of *mig-1* in *C. elegans* neuroblasts, suggesting that these cells use regulated expression to terminate their migration with increased temporal precision [17].

2.2 Temporal precision of molecular events with regulation and feedback

Recent studies on gene regulatory networks suggest that feedback is disadvantageous for timing precision [12]. However, *mig-1* is thought to be regulated by feedback in QR.p cells. Having feedback on *mig-1* means that the production rate of *mig-1* depends on how many *mig-1* molecules are present. This raises the fundamental question, if the feedback is disadvantageous then why its present in gene regulatory networks? We hypothesized that there is an interplay between feedback and regulation which can be beneficial to the timing precision. Investigating this hypothesis with our stochastic framework, we found that indeed, feedback alone decreases timing precision, but, surprisingly, feedback can be beneficial in the presence of regulation. In this project, we sought to find the global regulation function of *mig-1* which min-

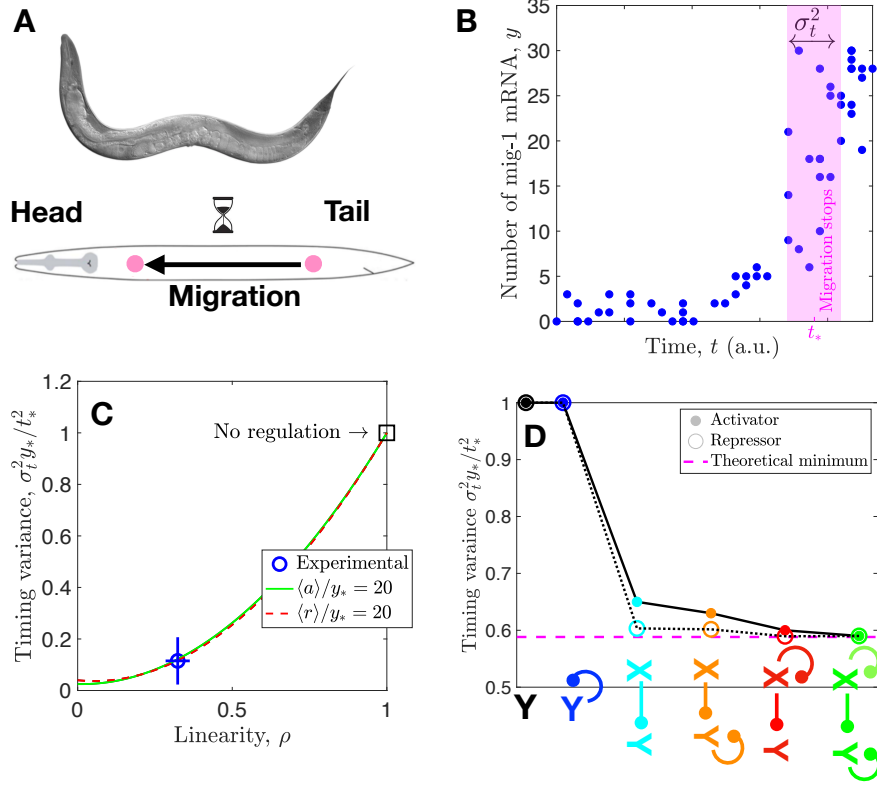


Fig. 2.1. A) In *C. elegans*, QR.p cell (pink dot) is present at the tail position at an early stage of development, and then it migrates to the head position. B) Experimentally measured rise in *mig-1* number during migration. C) Model prediction of timing variance and non-linear rise in *mig-1* agrees with experimental data. D) Timing variance comparison for various gene regulatory networks.

imizes the timing variance. We use analytic and gradient descent methods to find the global regulation function, and the theoretical global minimum timing variance achievable by a two-species regulatory network. Figure 2.1D is a comparison of the lowest possible noise achieved by various types of regulatory networks. Our findings suggest that a cellular process where timing precision is important should be governed by a molecular network with both multistep regulation and feedback, particularly one in which every species is subject to regulation. We find that adding freedom to the

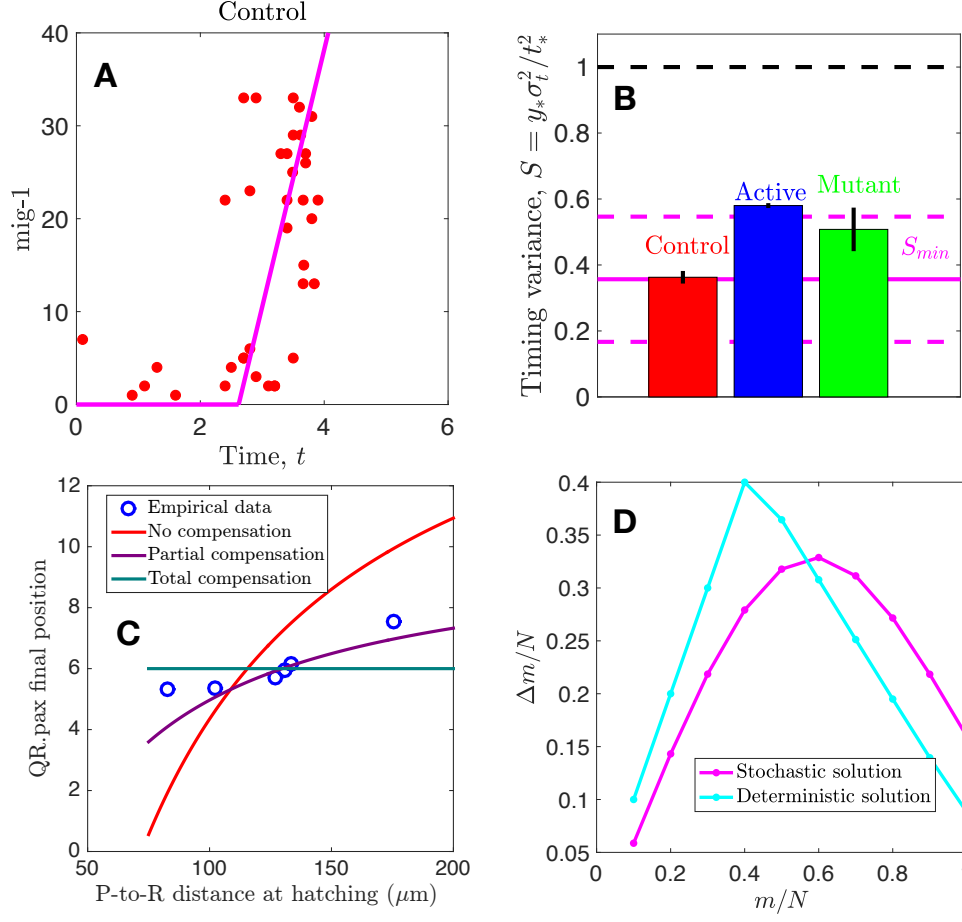


Fig. 2.2. A) Piece-wise linear fit to *mig-1* expression in control, B) Timing variance across three cases, magenta line is threshold minimum variance, C) In *C. elegans*, QR.p final position across the mutants with larva size ℓ_0 . No compensation assumes migration velocity is constant in all mutants, partial compensation model assumes migration velocity to be dependent on larva size ℓ_0 and total compensation is assuming cell stops at fixed position. D) Optimum receptors m needed to maximize the difference Δm between activated receptors on left side and activated receptors on the right side.

system can be utilized to obtain higher timing precision [19]. To test this mechanism, our collaborator performed experiments through CRISPR gene knockout. Through experimental studies we found that the *mig-1* expression is regulated by the *bar-1*

gene (Figure 2.2A). Furthermore, we looked at *mig-1* in three types of populations. i) Control (wild type); ii) Active, where *C. elegans* were mutated so that *bar-1* production was active, hence saturating the effect of *bar-1* on *mig-1*; and iii) Mutant, where individuals were mutated such that *bar-1* was not produced. We built a mathematical model based on empirical data to identify the *mig-1* regulation networks. We calculated the timing variance in crossing a threshold for each group (Figure 2.2B). We discovered that *mig-1* expression is controlled through positive feedback via *bar-1*. Furthermore, we found that the timing noise for the control group is the lowest and agrees with our analytical result for threshold global minimum variance (Figure 2.2B).

2.3 Termination of QR.p cell migration

In the previous project we discussed temporal regulation by QR.p cells (neuroblast) and quantified the precision of migration time. Here, we develop a model to explain the final position of the QR cell after migration is terminated. Recent experimental findings by our collaborators suggest that the larva size of *C. elegans* affects the final position of the QR.p cell. The empirical measurements reveal that the QR.p cell final position is correlated with larva size at hatching (Figure 2.2C). To explain these observations, we combined our model of timed cell migration in the previous projects, with a model of larval growth. Larval growth is similar to an expanding universe model, where every point is moving apart so that migration velocity is affected with respect to its background. We test whether migration velocity across the mutants is compensated by larval size. We find that the relative stopping position varies inversely with the larval size, and that larval growth softens this dependence into a roughly linear decrease. Our minimal model explains this effect in mutants with various larval size (Figure 2.2C).

2.4 Detection of nutrient wave by single-celled organisms

Our previous projects either investigated or were inspired by the timing precision of a migrating cell within *C. elegans*. However, the problem of precise timing is very common in biology. One common problem that many single-celled organisms face is the ability to sense nutrient waves. Signal sensing is used by many biological processes such as wound healing, neural patterning and tumor migration [20–22]. We study the mechanism by which cells are able to sense the direction of nutrient waves.

We hypothesized that cells sense the direction of nutrient waves through receptors present on the cell surface. As a case study, we assume the wave comes from the left direction and receptors are activated on upon interaction with the nutrient wave. Once m receptors are activated on the left side of the cell, they release inhibitors inside the cell. These inhibitors are free to diffuse inside the cell. When the inhibitors reach the right side, they deactivate the right-side receptors, preventing them from further activation by the nutrient wave. We assume that the nutrient wave can be described as a traveling wave. This mechanism creates a polarization due to the gradient of activated receptors on the left and receptors activated on the right side of the cell surface. The cell can control the release of inhibitors in time via m . We calculate the gradient of activated receptors at left and right side of cell. This gradient estimates the polarization direction for the cell. A higher gradient represents that the cell is able to sense direction of the wave precisely. Here we find a trade off between the diffusion time and the wave velocity for maximizing the gradient. If the diffusion is faster than the nutrient wave it will create a polarization of activated receptors towards the nutrient wave direction. We have developed a model to quantify the polarization towards the direction of the nutrient wave. We calculated the optimum number of receptors m needed out of total receptor number N for the release of inhibitors, hence maximizing the polarization (Figure 2.2D).

3. TEMPORAL PRECISION OF REGULATED GENE EXPRESSION

This chapter has been published as S. Gupta, J. Varennes, HC. Korswagen, A. Mugler (2018), “Temporal precision of regulated gene expression,” PLoS Comput Biol 14(6): e1006201.

Here we investigate the temporal precision of threshold crossing for a molecule that is regulated by either an accumulating activator or a degrading repressor. Using a first-passage-time approach [12, 23–25] and a combination of computational and analytic methods, we find that, unlike in the case of auto-regulation, the presence of either an activator or a repressor increases the temporal precision beyond that of the unregulated case. Furthermore, the optimal regulatory strategy for either an activator or a repressor corresponds to a nonlinear increase in the regulated molecule number over time. We elucidate the mechanism behind these optimal strategies, which stems from a tradeoff between reducing the noise of the regulator and that of the target molecule, and is similar to the fact that a sequence of time-ordered stochastic events becomes more precisely timed with more events. These findings are robust to more complex features of the regulation process, including bursts of molecule production, more complex regulator dynamics, and cell division. Our results are quantitatively consistent with both the temporal precision and nonlinearity of the *mig-1* mRNA dynamics of the migrating neuroblast cells in *C. elegans* larvae [5]. The agreement of our simple model with these data suggests that many molecular timing processes may benefit from the generic regulatory strategies we identify here.

3.1 Results

We consider a molecular species X whose production is regulated by a second species, either an activator A or a repressor R (Fig. 3.1A). The regulator undergoes its own dynamics: the activator undergoes pure production at a zeroth-order rate k whereas the repressor undergoes pure degradation at a first-order rate μ , such that in either case the production rate of X increases over time. The activator does not degrade and the repressor is not produced, although we later relax this assumption. For the regulation function we take a Hill function, which is a generic model of cooperative regulation [12, 13, 26],

$$f_+(a) = \frac{\alpha a^H}{a^H + K^H} \quad (\text{activator}), \quad (3.1)$$

$$f_-(r) = \frac{\alpha K^H}{r^H + K^H} \quad (\text{repressor}). \quad (3.2)$$

Here a and r are the molecule numbers of A and R , respectively, α is the maximal production rate of X , K is the half-maximal regulator number, and H is the cooperativity. First we neglect additional complexities such as bursts of production, more complex regulator dynamics, cell division, auto-regulation, longer regulatory cascades, or transcriptional delay. Later we check the robustness of our results to bursts, more complex regulator dynamics, and cell division, and we speculate upon the effects of auto-regulation, longer regulatory cascades, and delay in the Discussion.

We suppose that a behavior is initiated when the molecule number x crosses a threshold x_* (Fig. 3.1B). Because the production of X and the dynamics of the regulator are stochastic, the time at which x first reaches x_* is a random variable. We characterize the precision of this event by the mean \bar{t} and variance σ_t^2 of this first-passage time, which we compute numerically from the master equation corresponding to the reactions in Fig. 3.1A (see Materials and Methods). The maximal production rate α is set to ensure that \bar{t} is equal to a target time t_* , which we assume is set by functional constraints on the initiated behavior. This leaves k , K , and H as free

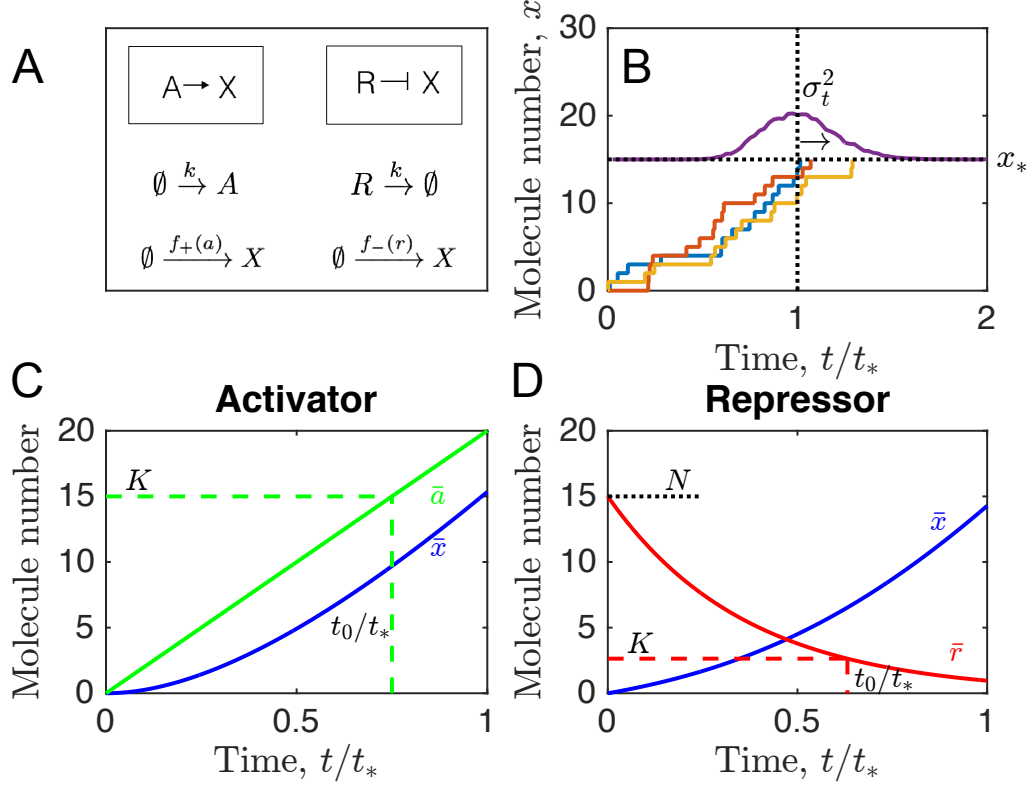


Fig. 3.1. **Threshold crossing of a regulated molecular species.** (A) A target species X is regulated by either an accumulating activator A or a degrading repressor R . (B) Temporal precision is quantified by the variance σ_t^2 of the first-passage time, at which the stochastic molecule number x first crosses the threshold x_* . (C, D) Deterministic dynamics illustrate the effects of regulation. Parameters are $kt_* = 20$ and $K = 15$ in C; $\mu t_* = 2.75$, $K = 2.6$, and $N = 15$ in B and D; and $x_* = 15$ and $H = 1$ throughout. t_0 is defined by $\bar{a}(t_0) = K$ in C and $\bar{r}(t_0) = K$ in D.

parameters of the regulation (with α a function of these parameters). In principle, these parameters can be optimized to minimize the timing variance σ_t^2 .

The deterministic dynamics, illustrated in Fig. 3.1C and D, neglect fluctuations but give an intuitive picture of the regulation. Whereas the amount of activator

increases linearly over time, the amount of repressor decays exponentially from an initial molecule number N :

$$\bar{a}(t) = kt, \quad (3.3)$$

$$\bar{r}(t) = Ne^{-\mu t}. \quad (3.4)$$

In either case, the production rate f_{\pm} of X increases over time, such that \bar{x} increases nonlinearly. N is an additional free parameter in the repressor case.

3.1.1 Regulation increases temporal precision

To investigate the effects of regulation on temporal precision, we consider the timing variance σ_t^2 as a function of the parameters k and K , or μ and K . The special case of no regulation corresponds to the limits $k \rightarrow \infty$ and $K \rightarrow 0$ in the case of activation, or $\mu \rightarrow \infty$ and $K \rightarrow \infty$ in the case of repression. In this case, the production of X occurs at the constant rate α . Reaching the threshold requires x_* sequential events, each of which occurs in a time that is exponentially distributed with mean $1/\alpha$. The total completion time for such a process is given by a gamma distribution with mean $\bar{t} = x_*/\alpha$ and variance $\sigma_t^2 = x_*/\alpha^2$ [25]. Ensuring that $\bar{t} = t_*$ requires $\alpha = x_*/t_*$, for which the variance satisfies $\sigma_t^2 x_*/t_*^2 = 1$. This expression gives the timing variance for the unregulated process.

In Fig. 3.2 we plot the scaled variance $\sigma_t^2 x_*/t_*^2$ as a function of the parameters k and K , or μ and K , for cooperativity $H = 3$ (color maps). In the case of activation (Fig. 3.2A), the variance decreases with increasing k and K . This means that the temporal precision is highest for an activator that accumulates quickly and requires a high abundance to produce X . In the case of repression (Fig. 3.2B), the variance has a global minimum as a function of μ and K . This means that the temporal precision is highest for a repressor with a particular well-defined degradation rate and abundance threshold. Importantly, we see that for both activation and repression, the scaled variance can be less than one, meaning that regulation allows improvement of the

temporal precision beyond that of the unregulated process. We have checked that this result holds for $H \geq 1$.

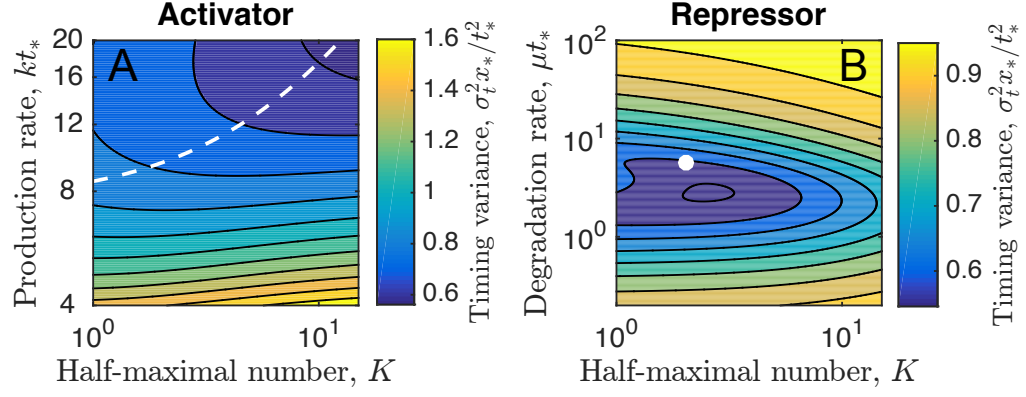


Fig. 3.2. **Optimal regulatory strategies.** Timing variance as a function of the regulatory parameters reveals (A) a trajectory along which the variance decreases in the case of the activator and (B) a global minimum in the case of the repressor. White dashed line in A and white dot in B show the analytic approximations in Eqs. 3.9 and 3.11, respectively. Parameters are $N = 15$ in B, and $x_* = 15$ and $H = 3$ in both.

3.1.2 Optimal regulation balances regulator and target noise

To understand the dependencies in Fig. 3.2, we develop analytic approximations. First, we assume that $H \rightarrow \infty$, such that the regulation functions in Eqs. 3.1 and 3.2 become threshold functions. In this limit, the production rate of X is zero if $a < K$ or $r > K$, and α otherwise. The deterministic dynamics of X become piecewise-linear,

$$\bar{x}(t) = \begin{cases} 0 & t < t_0 \\ \alpha(t - t_0) & t \geq t_0, \end{cases} \quad (3.5)$$

where t_0 is determined by either $\bar{a}(t_0) = K$ or $\bar{r}(t_0) = K$ according to Eqs. 3.3 and 3.4. Then, to set α , we use the condition $\bar{x}(t_*) = x_*$, which results in $\alpha = x_*/(t_* - t_0)$.

Lastly, we approximate the variance in the first-passage time using the variance in the molecule number and the time derivative of the mean dynamics [13]. Specifically, the timing variance of X arises from two sources: (i) uncertainty in the time when the regulator crosses its threshold K , which determines when the production of the target X begins, and (ii) uncertainty in the time when x crosses its threshold x_* , given that production begins at a particular time. The first source is regulator noise, and the second source is target noise. We estimate these timing variances from the associated molecule number variances, propagated via the time derivatives,

$$\sigma_t^2 \approx \underbrace{\sigma_y^2 \left(\frac{d\bar{y}}{dt} \right)^{-2} \Big|_{t_0}}_{\text{regulator}} + \underbrace{\sigma_x^2 \left(\frac{d\bar{x}}{dt} \right)^{-2} \Big|_{t_*}}_{\text{target}}, \quad (3.6)$$

where $y \in \{a, r\}$ denotes the regulator molecule number.

For the activator, which undergoes a pure production process with rate k , the molecule number obeys a Poisson distribution with mean kt . Therefore, the molecule number variance at time t_0 is $\sigma_a^2 = kt_0$. For the repressor, which undergoes a pure degradation process with rate μ starting from N molecules, the molecule number obeys a binomial distribution with number of trials N and success probability $e^{-\mu t}$. Therefore, the molecule number variance at time t_0 is $\sigma_r^2 = Ne^{-\mu t_0}(1 - e^{-\mu t_0})$. For the target molecule, which undergoes a pure production process with rate α starting at time t_0 , the molecule number obeys a Poisson distribution with mean $\alpha(t - t_0)$. Therefore, the molecule number variance at time t_* is $\sigma_x^2 = \alpha(t_* - t_0)$. Inserting these expressions into Eq. 3.6, along with the derivatives calculated from Eqs. 3.3-3.5 and the appropriate expressions for α and t_0 , we obtain

$$\frac{\sigma_t^2 x_*}{t_*^2} \approx \frac{Kx_*}{(kt_*)^2} + \left(1 - \frac{K}{kt_*} \right)^2 \quad (\text{activator}), \quad (3.7)$$

$$\frac{\sigma_t^2 x_*}{t_*^2} \approx \frac{(N - K)x_*}{NK(\mu t_*)^2} + \left[1 - \frac{\log(N/K)}{\mu t_*} \right]^2 \quad (\text{repressor}). \quad (3.8)$$

As a function of kt_* and K , the global minimum of Eq. 3.7 occurs as $kt_* \rightarrow \infty$ and $K \rightarrow \infty$. The path of descent toward this minimum is given by differentiating with respect to K at fixed kt_* and setting the result to zero, which yields the curve

$$K = \begin{cases} 0 & kt_* < \frac{x_*}{2} \\ kt_* - \frac{x_*}{2} & kt_* \geq \frac{x_*}{2}, \end{cases} \quad (3.9)$$

along which the variance satisfies

$$\frac{\sigma_t^2 x_*}{t_*^2} = \begin{cases} 1 & kt_* < \frac{x_*}{2} \\ \frac{x_*}{kt_*} \left(1 - \frac{x_*}{4kt_*}\right) & kt_* \geq \frac{x_*}{2}, \end{cases} \quad (3.10)$$

where the first case comes from the fact that K must be nonnegative. In contrast, the global minimum of Eq. 3.8 occurs at finite μt_* and K : differentiating with respect to each and setting the results to zero gives the values

$$K = e^{-2} N, \quad (3.11a)$$

$$\mu t_* = \frac{e^2 x_*}{2N} + 2, \quad (3.11b)$$

$$\frac{\sigma_t^2 x_*}{t_*^2} = \frac{x_*}{x_* + 4e^{-2} N}, \quad (3.12)$$

where we have assumed that $K/N \ll 1$ (see Materials and Methods), which is justified post-hoc by Eq. 3.11a.

These analytic approximations are compared with the numerical results for the activator in Fig. 3.2A (white dashed line, Eq. 3.9) and for the repressor in Fig. 3.2B (white circle, Eq. 3.11). In Fig. 3.2A we see that the global minimum indeed occurs as $kt_* \rightarrow \infty$ and $K \rightarrow \infty$, and the predicted curve agrees well with the observed descent. In Fig. 3.2B we see that the predicted global minimum lies very close to the observed global minimum. We have also checked along specific slices in the (K, kt_*) or $(K, \mu t_*)$ plane and found that the analytic approximations generally differ from

the numerical results by about 10% or less, despite the fact that the approximations take $H \rightarrow \infty$ whereas the numerics in Fig. 3.2 use $H = 3$.

The success of the approximations means that Eq. 3.6 describes the key mechanism leading to the optimal temporal precision. Eq. 3.6 demonstrates that the optimal regulatory strategy arises from a tradeoff between minimizing regulator and target noise. On the one hand, minimizing only the regulator noise would require that the regulator cross its threshold K with a steep slope $d\bar{y}/dt$ and therefore at an early time, meaning that the target molecule would be effectively unregulated and would increase linearly over time. On the other hand, minimizing only the target noise would require that the regulator cross its threshold only shortly before the target time t_* , such that the target molecule would cross its threshold x_* with a steep slope $d\bar{x}/dt$, leading to a highly nonlinear increase of the target molecule over time. In actuality, the optimal strategy is somewhere in between, with the regulator crossing its threshold at some intermediate time t_0 , and the target molecule exhibiting moderately nonlinear dynamics as in Fig. 3.1C and D.

Eqs. 3.10 and 3.12 demonstrate that the timing variance is small for large kt_*/x_* in the case of activation, and small for large N/x_* in the case of repression. This makes intuitive sense because each of these quantities scales with the number of regulator molecules: k is the production rate of activator molecules, while N is the initial number of repressor molecules. To make this intuition quantitative, we define a cost as the time-averaged number of regulator molecules,

$$\langle a \rangle = \frac{1}{t_*} \int_0^{t_*} dt \bar{a}(t) = \frac{1}{2} kt_*, \quad (3.13)$$

$$\langle r \rangle = \frac{1}{t_*} \int_0^{t_*} dt \bar{r}(t) = \frac{N}{\mu t_*} (1 - e^{-\mu t_*}), \quad (3.14)$$

where the second steps follow from Eqs. 3.3 and 3.4. We see that, indeed, $\langle a \rangle$ scales with k , and $\langle r \rangle$ scales with N . Thus, Eqs. 3.10 and 3.12 demonstrate that increased temporal precision comes at a cost, in terms of the number of regulator molecules that must be produced.

3.1.3 Model predictions are consistent with neuroblast migration data

We test our model predictions using data from neuroblast cells in *C. elegans* larvae [5]. During *C. elegans* development, particular neuroblast cells migrate from the posterior to the anterior of the larva. It has been shown that the migration terminates not at a particular position, but rather after a particular amount of time, and that the termination time is controlled by a temporal increase in the expression of the *mig-1* gene [5]. Since *mig-1* is known to be subject to regulation [27], we investigate the extent to which the dynamics of *mig-1* can be explained by the predictions of our model.

Figure 3.3A shows the number x of *mig-1* mRNA molecules per cell as a function of time t , obtained by single-molecule fluorescent in situ hybridization (from [5]). We analyze these data in the following way (see Materials and Methods for details). First, noting that the dynamics are nonlinear, we quantify the linearity using the area under the curve, normalized by that for a perfectly linear trajectory $x_*t_*/2$,

$$\rho = \frac{2}{x_*t_*} \int_0^{t_*} dt x(t). \quad (3.15)$$

By this definition, $\rho = 1$ for perfectly linear dynamics, and $\rho \rightarrow 0$ for maximally nonlinear dynamics (a sharp rise at t_*). Then, we estimate x_* , t_* , and the timing variance σ_t^2 from the data. Specifically, migration is known to terminate between particular reference cells in the larva [5], which gives an estimated range for the termination time t_* . This range is shown in magenta in Fig. 3.3A and corresponds to a threshold within the approximate range $10 \leq x_* \leq 25$. Therefore, we divide the x axis into bins of size Δx , choose bin midpoints x_* within this range, and for each choice compute the mean t_* and the variance σ_t^2 of the data in that bin. Fig. 3.3B shows the average and standard deviation of results for different values of x_* and Δx (blue circle).

The experimental data point in Fig. 3.3B exhibits two clear features: (i) the dynamics are nonlinear (ρ is significantly below 1), and (ii) the timing variance is

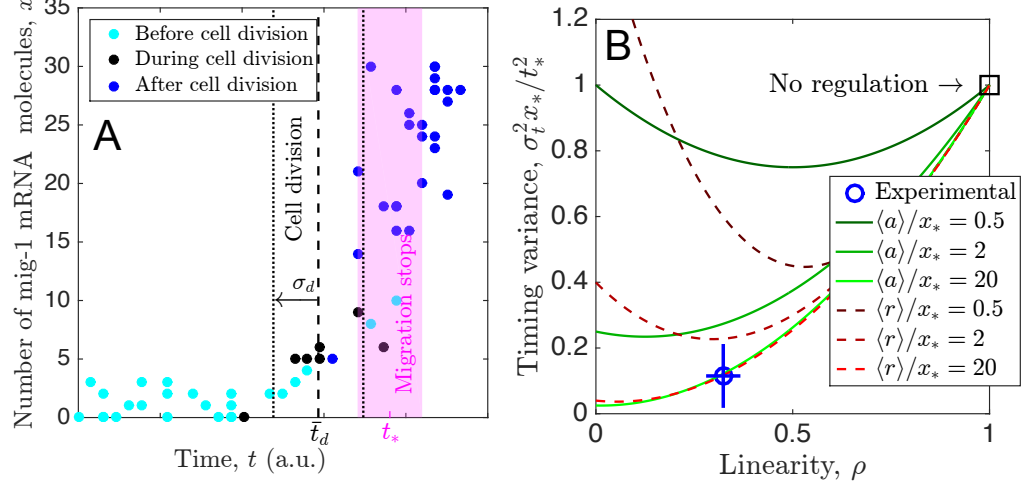


Fig. 3.3. Model predictions agree with neuroblast migration data.

(A) Number of *mig-1* mRNA molecules per cell as a function of time t , obtained by single-molecule fluorescent in situ hybridization, from [5]. Magenta shows approximate range of times when cell migration terminates. Black lines show mean \bar{t}_d (dashed) and standard deviation σ_d of cell division times (black points). (B) Timing variance vs. linearity of $x(t)$, both for experimental data in A (blue circle) and our model (curves, Eqs. 3.16 and 3.17). Data analyzed using ranges of threshold $10 \leq x_* \leq 25$ and bin size $3 \leq \Delta x \leq 12$; error bars show standard deviations of these results. We see that for sufficiently large cost $\langle a \rangle/x_*$ or $\langle r \rangle/x_*$, model predictions agree with experimental data point.

low ($\sigma_t^2 x_*/t_*^2$ is significantly below 1). Neither feature can be explained by a model in which the production of x is unregulated, since that would correspond to a linear increase of molecule number over time ($\rho = 1$) and a timing variance that satisfies $\sigma_t^2 x_*/t_*^2 = 1$ (square in Fig. 3.3B). Furthermore, since auto-regulation has been shown to generically increase timing variance beyond the unregulated case [12], it is unlikely that feature (ii) can be accounted for by a model with auto-regulation alone. Can these data be accounted for by our model with regulation?

To address this question we calculate ρ and $\sigma_t^2 x_*/t_*^2$ from our model. For simplicity, we focus on the analytic approximations in Eqs. 3.7 and 3.8, since they have been

validated in Fig. 3.2. In these approximations, since $\bar{x}(t)$ is piecewise-linear (Eq. 3.5), calculating ρ via Eq. 3.15 is straightforward: $\rho = 1 - t_0/t_*$, where t_0 is once again determined by either $\bar{a}(t_0) = K$ or $\bar{r}(t_0) = K$ according to Eqs. 3.3 and 3.4. For a given ρ and cost $\langle a \rangle/x_*$ or $\langle r \rangle/x_*$, we calculate the minimum timing variance $\sigma_t^2 x_*/t_*^2$. For the activator, we use the expression for ρ along with Eq. 3.13 to write Eq. 3.7 in terms of ρ and $\langle a \rangle/x_*$,

$$\frac{\sigma_t^2 x_*}{t_*^2} = \frac{x_*}{2\langle a \rangle} (1 - \rho) + \rho^2. \quad (3.16)$$

For the repressor, we use the expression for ρ along with Eq. 3.14 to write Eq. 3.8 in terms of ρ and $\langle r \rangle/x_*$, and then minimize over N (see Materials and Methods) to obtain

$$\frac{\sigma_t^2 x_*}{t_*^2} = \frac{e^3}{27} \frac{x_*}{\langle r \rangle} (1 - \rho)^3 + \rho^2. \quad (3.17)$$

Eqs. 3.16 and 3.17 are shown in Fig. 3.3B (green solid and red dashed curves, respectively), and we see the same qualitative features for both cases: all curves satisfy $\sigma_t^2 x_*/t_*^2 = 1$ at $\rho = 1$, as expected; and as ρ decreases, each curve exhibits a minimum whose depth and location depend on cost. Specifically, as cost increases (lighter shades of green or red), the variance decreases, as expected. Importantly, we see that at a cost on the order of $\langle a \rangle/x_* = \langle r \rangle/x_* \sim 10$, the model becomes consistent with the experimental data: both the low timing variance and the low degree of linearity predicted by either the activator or repressor case agree quantitatively with the experiment. This suggests that either an accumulating activator or a degrading repressor is sufficient to account for the temporal precision observed in *mig-1*-controlled neuroblast migration.

3.1.4 Results are robust to additional complexities including cell division

Our minimal model neglects common features of gene expression such as bursts in molecule production [28] and additional sources of noise. Therefore we test the robustness of our findings to these effects here. First, we test the robustness of

the results to the presence of bursts by replacing the Poisson process governing the activator production with a bursty production process. Specifically, we assume that each production event increases the activator molecule count by an integer in $[1, \infty)$ drawn from a geometric distribution with mean b [29, 30]. The limiting case of $b = 1$ recovers the original Poisson process. The results are shown in Fig. 3.4A for $b = 1$, 3, and 5 (green solid, cyan dashed, and magenta dashed curves). We see that bursts in the activator increase the timing variance of the target molecule, as expected, but that there remain parameters for which the variance is less than that for the unregulated case, $\sigma_t^2 x_*/t_*^2 = 1$ (dashed black line). This result shows that even with bursts, regulation by an accumulating activator is beneficial for timing precision.

We also recognize that whereas the activator can be assumed to start with exactly zero molecules, it is more realistic for the repressor to start with an initial number of molecules that has its own variability. We incorporate this additional variability into the model by performing stochastic simulations [31] of the reactions in Fig. 3.1A and drawing the initial repressor molecule number from a Poisson distribution across simulations. The result is shown by the green dashed curve in Fig. 3.4B. We see that the additional variability gives rise to an increase in the timing variance of the target molecule, as expected (compare with the green solid curve). However, for most of the range of degradation rates, including the optimal degradation rate, the variance remains less than that of the unregulated case, $\sigma_t^2 x_*/t_*^2 = 1$ (dashed black line). This result indicates that the benefit of repression is robust to this additional source of noise.

Then, we test the robustness of the results to our assumptions that the activator undergoes pure production and the repressor undergoes pure degradation. Specifically, we introduce a degradation rate μ for the activator, and a production rate k for the repressor, such that either regulator reaches a steady state of k/μ . The blue curves in Fig. 3.4A and B show the case where the increasing activator's steady state k/μ is twice its regulation threshold K , or the decreasing repressor's steady state k/μ is half its regulation threshold K , respectively. In both cases, we see that the timing

variance of the target molecule increases because the regulator slows down on the approach to its regulation threshold. Nonetheless, we see that it is still possible for the variance to be lower than that of the unregulated case. The red curves show the case where the regulator's steady state is equal to its regulation threshold. Here we are approaching the regime in which threshold crossing is an exponentially rare event. As a result, the variance further increases, to the point where it is above that of the unregulated case for the full range of parameters shown. These results demonstrate that the benefit of regulation is robust to more complex regulator dynamics, but only if the regulator still crosses its regulation threshold at a reasonable mean velocity.

Finally, we test the robustness of the results to a feature exhibited by the experimental *mig-1* data: near the end of migration, cell division occurs (Fig. 3.3A, black data). One daughter cell continues migrating (Fig. 3.3A, dark blue data), while the other undergoes programmed cell death [5]. To investigate the effects of cell division, we perform stochastic simulations, and at a given time t_d we assume that the cell volume V is reduced by a factor of two. For each simulation, t_d is drawn from a Gaussian distribution with mean \bar{t}_d and variance σ_d^2 determined by the data (Fig. 3.3A, black). At t_d , we reduce the molecule numbers of both the regulator and the target molecule assuming symmetric partitioning, such that the molecule number after division is drawn from a binomial distribution with total number of trials equal to the molecule number before division and success probability equal to one half. We also reduce the molecule number threshold K by a factor of two because it is proportional to the cell volume via $K = K_d V$, where K_d is the dissociation constant.

Fig. 3.4C shows the dynamics of the mean molecule numbers of the activator (green solid) and its target (blue solid), or the repressor (red dashed) and its target (blue dashed). We see that the activator, repressor, and target drop in molecule number at division but that the abruptness of the drop is smoothed by the variability in the division time. The smoothing is more pronounced in the cases of the repressor and the target because the molecule numbers of these species are smaller at division. Thus, for either the activator or repressor mechanism, we see that the experimentally

observed variability in division time is sufficient to smooth out the dynamics of the target molecule number, consistent with the experimental data in Fig. 3.3A.

Additionally, we see in Fig. 3.4D that the timing variance of the target molecule in both the activator and repressor cases is similar to that without division in the region of the experimental division time. This further indicates that either model remains sufficient to account for the low experimental timing variance, even with the experimentally observed cell division. Taken together, the results in Fig. 3.4C and D show that the key results of the model are robust to the effects of cell division.

3.2 Discussion

We have demonstrated that regulation by an accumulating activator or a diminishing repressor increases the precision of threshold crossing by a target molecule, beyond the precision achievable with constitutive expression alone. The increase in precision results from a tradeoff between reducing the noise of the regulator and reducing the noise of the target molecule itself. Our minimal model is sufficient to account for both the high degree of nonlinearity and the low degree of noise in the dynamics of *mig-1* in *C. elegans* neuroblasts, providing evidence for the hypothesis that these cells use regulated expression to terminate their migration with increased temporal precision. These results suggest that regulation by a dynamic upstream species is a simple and generic method of increasing the temporal precision of cellular behaviors governed by threshold-crossing events.

Why does regulation increase temporal precision, whereas it has been shown that auto-regulation (feedback) does not [12]? After all, either regulation or positive feedback can produce an acceleration in molecule number over time, leading to a steeper threshold crossing. The reason is likely that positive feedback relies on self-amplification. In addition to amplifying the mean, positive feedback also amplifies the noise. In contrast, regulation by an external species does not involve self-amplification. Therefore, the noise increase is not as strong. The target molecule

certainly inherits noise from the regulator (Eq. 3.6), but the increase in noise does not outweigh the benefit of the acceleration, as it does for feedback. Future work could investigate the interplay of regulation and feedback, as well as active degradation of the target molecule, especially as *mig-1* is thought to be subject to feedback and degradation in addition to external regulation [5, 27].

Our finding that regulation increases temporal precision is related to the more basic phenomenon that a sequence of ordered events has a lower relative timing error than a single event [25, 32]. Specifically, if a single event occurs in a time that is exponentially distributed with a mean τ , then the relative timing error is $\sigma/\tau = 1$. For n such events that must occur in sequence, the total completion time follows a gamma distribution with relative timing error $\sigma/\tau = 1/\sqrt{n}$, which decreases with increasing n . Thus, at a coarse-grained level, the addition of a regulator can be viewed as increasing the length of the sequence of threshold-crossing events from one to two, and thus decreasing the timing error. This perspective suggests that the timing error could be decreased even further via a cascade of regulators.

Although we have demonstrated that our findings are robust to complexities such as bursts and cell division (Fig. 3.4), our model neglects additional features of regulated gene expression such as transcriptional delay. Transcriptional delay has been shown to play an important role in regulation [33, 34] and to have consequences for the mean and variance of threshold-crossing times [35]. If a delay were to arise due to a sequence of stochastic but irreversible steps, then we conjecture that the relative timing error would decrease with the number of these steps, due to the same cascading mechanism mentioned in the previous paragraph. However, it has been shown that if there are reversible steps or cycles within a multistep process, then the first passage time distribution can approach an exponential as the number of steps becomes large [32]. In this case the timing statistics would be captured by our simple model, which assumes single exponentially distributed waiting times. Future work could explore the effects of transcriptional delay in more detail.

Finally, we have shown that the *mig-1* data from migrating neuroblasts in *C. elegans* are quantitatively consistent with either the accumulating activator or diminishing repressor model, but the data do not distinguish between the two models. A direct approach to search for a distinction would be to use genetic knockout techniques to screen directly for regulators of *mig-1* and their effects on its abundance. A less direct approach would be to more closely investigate the effects of the cell division that occurs during migration. For example, we assumed in this study that the volume fraction is equal to the average molecule number fraction in the surviving cell after division. However, if they were found to be unequal for either *mig-1* or its regulator(s), then the concentrations of these species could undergo an abrupt change after division, which may have opposing consequences for the activator vs. the repressor mechanism. Future studies could use these or related approaches to more concretely identify the role of gene regulation in achieving precise timing during cellular processes.

3.3 Materials and methods

3.3.1 Computation of the first-passage time statistics

We compute the first-passage time statistics \bar{t} and σ_t^2 numerically from the master equation following [12], generalized to a two-species system. Specifically, the probability $F(t)$ that the molecule number crosses the threshold x_* at time t is equal to the probability $P_{y,x_*-1}(t)$ that there are y regulator molecules (where $y \in \{a, r\}$) and $x_* - 1$ target molecules, and that a production reaction occurs with rate $f_{\pm}(y)$ to bring the target molecule number up to x_* . Since this event can occur for any regulator molecule number y , we sum over all y ,

$$F(t) = \sum_{y=0}^Y f_{\pm}(y) P_{y,x_*-1}(t), \quad (3.18)$$

where $Y = \{a_{\max}, N\}$. The repressor has a maximum number of molecules N , whereas the activator number is unbounded, and therefore we introduce the numerical cutoff $a_{\max} = kt_* + \sqrt{10kt_*}$. The probability P_{yx} evolves in time according to the master equation corresponding to the reactions in Fig. 3.1A,

$$\dot{P}_{ax} = kP_{a-1,x} + f_+(a)P_{a,x-1} - [k + f_+(a)]P_{ax}, \quad (3.19a)$$

$$\dot{P}_{rx} = k(r+1)P_{r+1,x} + f_-(r)P_{r,x-1} - [kr + f_-(r)]P_{rx}. \quad (3.19b)$$

The moments of Eq. 3.18 are

$$\langle t^m \rangle = \sum_{y=0}^Y f_{\pm}(y) \int_0^{\infty} dt t^m P_{y,x_*-1}(t), \quad (3.20)$$

where $\bar{t} = \langle t \rangle$ and $\sigma_t^2 = \langle t^2 \rangle - \langle t \rangle^2$. Therefore computing \bar{t} and σ_t^2 requires solving for the dynamics of P_{yx} .

Because Eq. 3.19 is linear in P_{yx} , it is straightforward to solve by matrix inversion. We rewrite P_{yx} as a vector by concatenating its columns, $\vec{P}^T = [[P_{00}, \dots, P_{Y0}], \dots, [P_{0,x_*-1}, \dots, P_{Y,x_*-1}]]$ such that Eq. 3.19 becomes $\dot{\vec{P}} = \mathbf{M}\vec{P}$, where

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}^{(1)} & & & & \\ \mathbf{M}^{(2)} & \mathbf{M}^{(1)} & & & \\ & \mathbf{M}^{(2)} & \mathbf{M}^{(1)} & & \\ & & \ddots & \ddots & \\ & & & \mathbf{M}^{(2)} & \mathbf{M}^{(1)} \end{bmatrix}. \quad (3.21)$$

Here, for $i, j \in \{0, \dots, Y\}$, the $x_* - 1$ subdiagonal blocks are the diagonal matrix $\mathbf{M}_{ij}^{(2)} = f_{\pm}(i)\delta_{ij}$, and the x_* diagonal blocks are the subdiagonal matrix $\mathbf{M}_{ij}^{(1)} = -[k(1 - \delta_{ia_{\max}}) + f_+(i)]\delta_{ij} + k\delta_{i-1,j}$ or the superdiagonal matrix $\mathbf{M}_{ij}^{(1)} = -[ki + f_-(i)]\delta_{ij} + k(i+1)\delta_{i+1,j}$ for the activator or repressor case, respectively. The $\delta_{ia_{\max}}$ term prevents activator production beyond a_{\max} molecules. The final $\mathbf{M}^{(1)}$ matrix in Eq. 3.21 contains f_{\pm} production terms that are not balanced by equal and opposite terms any-

where in their columns. These terms correspond to the transition from $x_* - 1$ to x_* target molecules, for which there is no reverse transition. Therefore, the state with x_* target molecules (and any number of regulator molecules) is an absorbing state that is outside the state space of \vec{P} [12]. Consequently, probability leaks over time, and $\vec{P}(t \rightarrow \infty) = \vec{0}$. Equivalently, the eigenvalues of \mathbf{M} are negative.

The solution of the dynamics above Eq. 3.21 is $\vec{P}(t) = \exp(\mathbf{M}t)\vec{P}(0)$. Therefore, Eq. 3.20 becomes $\langle t^m \rangle = \vec{V}^\top [\int_0^\infty dt t^m \exp(\mathbf{M}t)] \vec{P}(0)$, where \vec{V}^\top is a length- $x_*(Y+1)$ row vector consisting of $[f_\pm(0), \dots, f_\pm(Y)]$ preceded by zeros. We solve this equation via integration by parts [12], noting that the boundary terms vanish because the eigenvalues of \mathbf{M} are negative, to obtain

$$\langle t^m \rangle = (-1)^{m+1} m! \vec{V}^\top (\mathbf{M}^{-1})^{m+1} \vec{P}(0). \quad (3.22)$$

We see that computing $\bar{t} = \langle t \rangle$ and $\sigma_t^2 = \langle t^2 \rangle - \langle t \rangle^2$ requires inverting \mathbf{M} , which we do numerically in Matlab. We initialize \vec{P} as $P_{ax}(0) = \delta_{a0}\delta_{x0}$ or $P_{rx}(0) = \delta_{rN}\delta_{x0}$ for the activator or repressor case, respectively.

When including cell division, we compute \bar{t} and σ_t^2 from 50,000 stochastic simulations [31].

3.3.2 Deterministic dynamics

The dynamics of the mean regulator and target molecule numbers are obtained by calculating the first moments of Eq. 3.19, $\partial_t \bar{y} = \sum_{yx} y \dot{P}_{yx}$ and $\partial_t \bar{x} = \sum_{yx} x \dot{P}_{yx}$, where $y \in \{a, r\}$. For the regulator we obtain $\partial_t \bar{a} = k$ or $\partial_t \bar{r} = -\mu \bar{r}$ in the activator or repressor case, respectively, which are solved by Eqs. 3.3 and 3.4. For the target molecule we obtain $\partial_t \bar{x} = \langle f_\pm(y) \rangle$, which is not solvable because f_\pm is nonlinear (i.e., the moments do not close). A deterministic analysis conventionally assumes

$\langle f_{\pm}(y) \rangle \approx f_{\pm}(\bar{y})$, for which the equation becomes solvable by separation of variables. For example, in the case of $H = 1$, using Eqs. 3.1-3.4, we obtain

$$\bar{x}(t) = \begin{cases} \alpha t - (\alpha K/k) \log[(kt + K)/K] & \text{(activator)} \\ (\alpha/\mu) \log[(N + Ke^{\mu t})/(N + K)] & \text{(repressor)}. \end{cases} \quad (3.23)$$

Equation 3.23 is plotted in Fig. 3.1C and D.

When including cell division, we compute the mean dynamics from the simulation trajectories (Fig. 3.4C).

3.3.3 Details of the analytic approximations

To find the global minimum of Eq. 3.8, we differentiate with respect to kt_* and K and set the results to zero, giving two equations. kt_* can be eliminated, leaving one equation for K ,

$$\frac{1}{2} \log \frac{N}{K} = 1 - \frac{K}{N} \quad (3.24)$$

This equation is transcendental. However, in the limit $K \ll N$, we neglect the last term, which gives Eq. 3.11.

To derive Eq. 3.17, we use

$$\rho = 1 - \frac{t_0}{t_*} = 1 - \frac{\log N/K}{\mu t_*} \quad (3.25)$$

where the second step follows from $\bar{r}(t_0) = K$ according to Eq. 3.4; and, from Eq. 3.14,

$$\langle r \rangle = \frac{N}{\mu t_*} (1 - e^{-\mu t_*}) \approx \frac{N}{\mu t_*} \quad (3.26)$$

where the second step assumes that the repressor is fast-decaying, $\mu t_* \gg 1$. We use Eqs. 3.26 and 3.25 to eliminate μt_* and K from Eq. 3.8 in favor of ρ and $\langle r \rangle$,

$$\frac{\sigma_t^2 x_*}{t_*^2} \approx \frac{x_*}{N} \left(e^{N(1-\rho)/\langle r \rangle} - 1 \right) \frac{\langle r \rangle^2}{N^3} + \rho^2. \quad (3.27)$$

For nonlinear dynamics ($\rho < 1$) we may safely neglect the -1 in Eq. 3.27. Then, differentiating Eq. 3.27 with respect to N and setting the result to zero, we obtain $N = 3\langle r \rangle / (1 - \rho)$. Inserting this result into Eq. 3.27 produces Eq. 3.17.

3.3.4 Analysis of the experimental data

To estimate the time at which migration terminates in Fig. 3.3A, we refer to [5]. There, the position at which neuroblast migration terminates is measured with respect to seam cells V1 to V6 in the larva (see Fig. 4D in [5]). In particular, in wild type larvae, migration terminates between V2 and the midpoint of V2 and V1. This range corresponds to the magenta region in Fig. 3.3A (see Fig. 4B, upper left panel, in [5]). Under the assumptions of constant migration speed and equal distance between seam cells, the horizontal axis in Fig. 3.3A represents time.

To compute ρ for the experimental data in Fig. 3.3A according to Eq. 3.15 we use a trapezoidal sum. For the choices of x_* and t_* described in the text, this produces the ρ values in Fig. 3.3B.

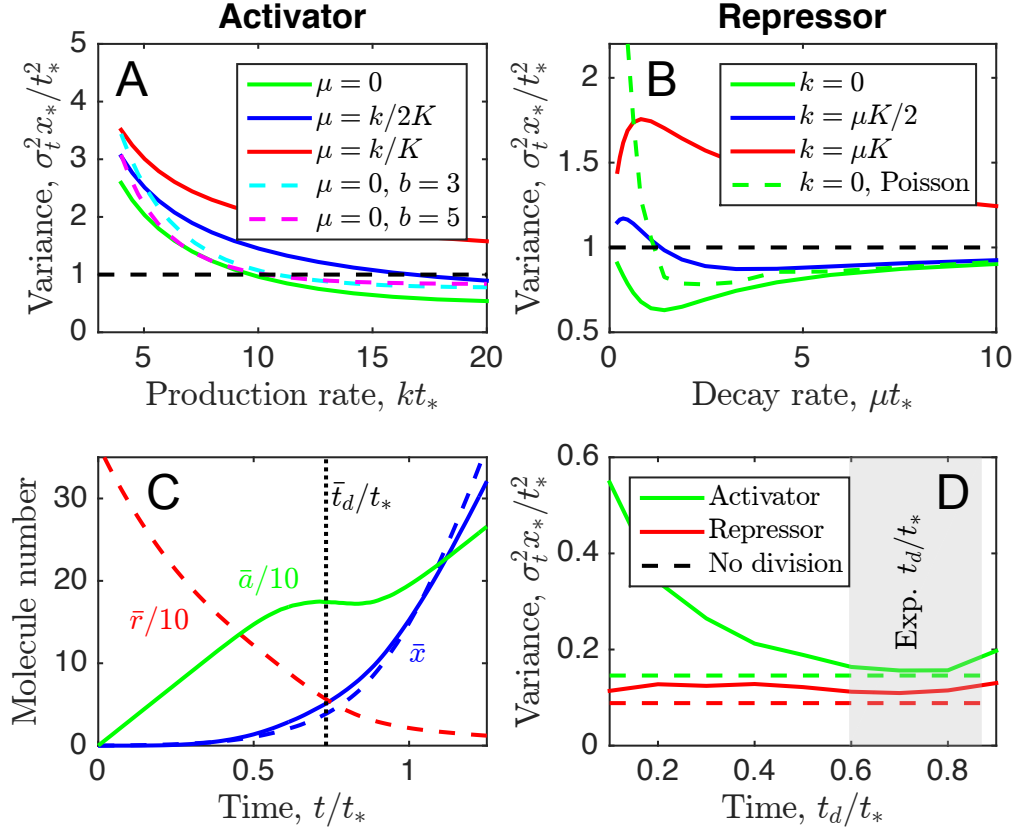


Fig. 3.4. Results are robust to additional complexities including cell division. (A, B) Green solid curves show slices from Fig. 3.2 with $K = 10$ while black dashed line shows unregulated limit $\sigma_t^2 x_*/t_*^2 = 1$. We see that regulation can reduce timing variance even with bursts in activator production of mean size b (A, cyan and magenta dashed), initial Poisson noise in repressor number (B, green dashed), or steady state k/μ in regulator dynamics (blue) unless it approaches regulation threshold K (red). (C) Mean dynamics of activator model (solid) and repressor model (dashed) in which cell division occurs at time \bar{t}_d on average. Abrupt reductions in molecule numbers are smoothed by noise in t_d and by binomial partitioning of molecules. (D) Timing variance approaches that with no division (dashed) within experimental division region (gray). In A and B, parameters are as in Fig. 3.2. In C and D, parameters are $x^* = 15$, $\langle a \rangle/x_* = \langle r \rangle/x_* = 10$, and $H = 3$, with kt_* , μt_* , and K set to optimal values (Fig. 3.2) and \bar{t}_d and σ_d set to experimental values. In all cases, α is set to ensure that mean threshold crossing time \bar{t} equals t_* .

4. TEMPORAL PRECISION OF MOLECULAR EVENTS WITH REGULATION AND FEEDBACK

Parts of this chapter have been published as S. Gupta, S. Fancher, HC. Korswagen, A. Mugler (2020), “Temporal precision of molecular events with regulation and feedback,” Phys. Rev. E 101, 062420.

4.1 Introduction

Precise timing is crucial for many biological processes including cell division [1–3], cell differentiation [4], cell migration [5], embryonic development [7, 8], and cell death [9]. Ultimately the timing of these processes is governed by the timing of molecular events inside the cell. However, these events are inherently stochastic. Cells use regulatory networks to reduce this stochasticity, but the effects of particular regulatory features on timing precision remain poorly understood.

In chapter 3 we demonstrated that the time at which an accumulating molecular species crosses an abundance threshold is more precise if that species is regulated by a second species with its own stochastic dynamics [17]. In contrast, it was recently demonstrated that if the species is instead regulated by itself (feedback), then the crossing time is less precise [12]. It is difficult to predict in advance the effect of a particular regulatory feature on timing precision. Moreover, even if the effect of a regulatory feature is known for specific parameters, a systematic method of determining whether the effect is generic, or of determining the optimal regulatory function, is currently lacking.

Feedback cannot be generically harmful to timing precision, as feedback is common in many important timing processes. In yeast, the cyclin proteins that cross an abundance threshold to initiate the cell cycle [3] are subject to positive feedback

[1, 36, 37]. In *Caenorhabditis elegans*, the mig-1 protein that crosses an abundance threshold to terminate migration in QR neuroblasts [5] has been found in experiments on the sister QL lineage to be subject to feedback via Wnt signaling [27]. This raises the question of why feedback is observed in key timing processes if it has previously been shown to decrease timing precision.

Here we investigate the combined effect of regulation and feedback on timing precision. We develop a gradient-descent approach to find the globally optimal regulation function for a given network topology that minimizes the timing noise. We find that, despite the fact that feedback generically increases timing noise when it acts alone, feedback decreases timing noise when it acts in combination with regulation by an external species. We explain the mechanisms behind this counterintuitive result, derive a generic lower bound on the timing noise, and compare of our results to the timing of neuroblast migration in *C. elegans*.

4.2 Results

Consider a molecular species Y that is produced over time and first reaches a molecule-number threshold y_* at a particular time t_* on average (Figure 4.1A). In what follows we assume that Y can only be produced, not degraded, and that when regulated, the regulator of Y can only be produced or degraded, not both. This simplifying assumption limits the possible paths that the system can take through molecule number space, but the methods we outline below are readily generalized to species that are both produced and degraded, and we comment on this case in the Discussion.

Stochasticity in the accumulation process leads to variability in the crossing time t . The timing noise is given by the variance σ_t^2 . For unregulated production of Y , the time between each production event is exponentially distributed with mean t_*/y_* and variance $(t_*/y_*)^2$. Because the production events are independent, the variances

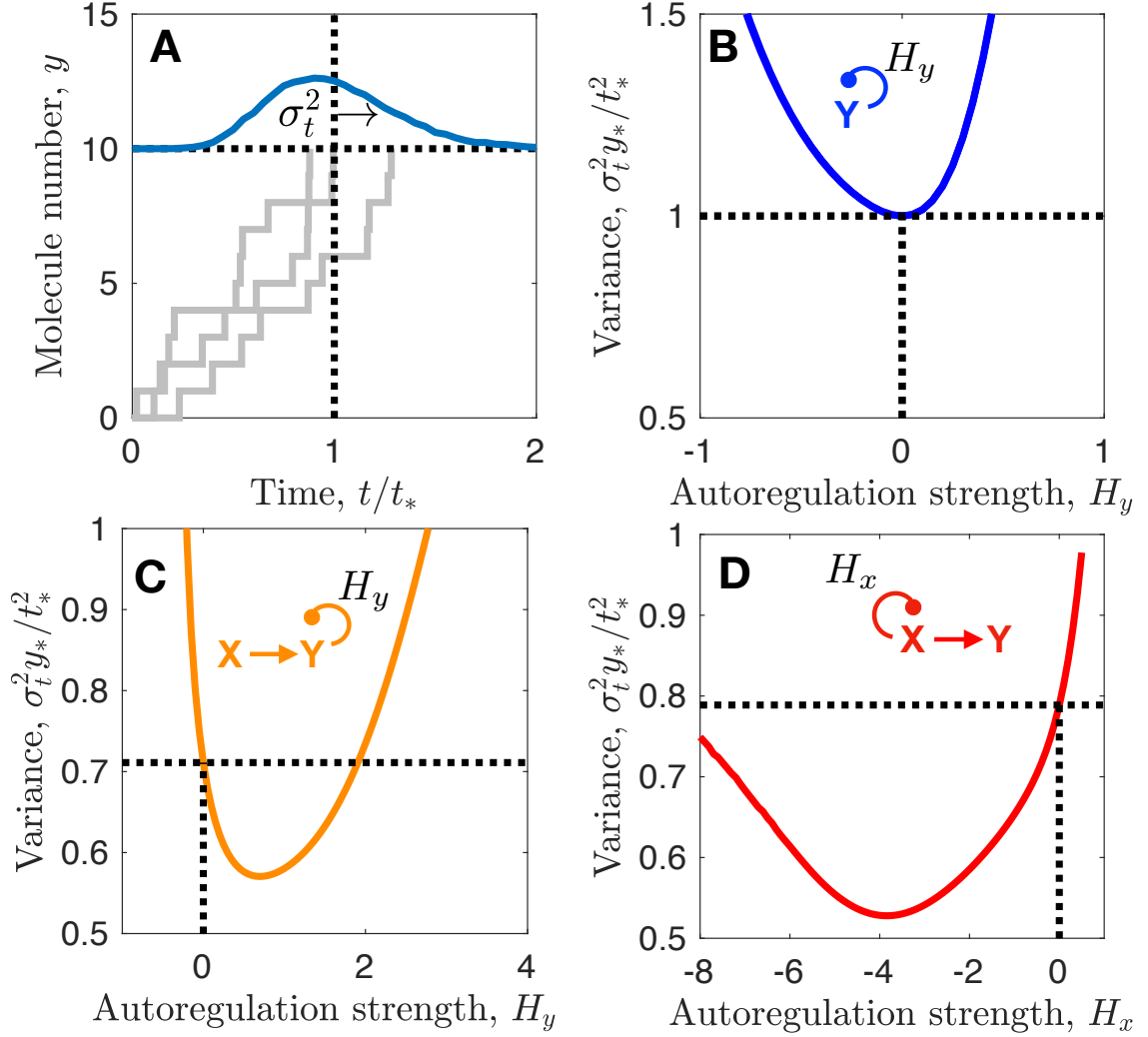


Fig. 4.1. Feedback increases timing precision in the presence but not absence of regulation. (A) A species Y crosses a molecule-number threshold y_* at mean time t_* with timing variance σ_t^2 . (B) Feedback increases the variance. However, in the presence of regulation by a second species X , feedback on either (C) Y or (D) X can decrease the variance. Parameters are $K_y = 2.5$ in B; $\alpha_0 t_* = 10$, $H_x = -0.5$, $H_{xy} = -H_y$, $K_x = 15$, $K_y = 5$, and $K_{xy} = 6$ in C; $\alpha_0 t_* = 10$, $H_y = 4$, $K_x = 10$, and $K_y = 7.5$ in D; and $y_* = 10$ throughout. A is generated by Gillespie simulation [31]; B-D are generated by solving the master equation numerically [17].

add, giving a total variance of $\sigma_t^2 = y_*(t_*/y_*)^2 = t_*^2/y_*$. Therefore we focus on the scaled variance $\sigma_t^2 y_*/t_*^2$, whose value is 1 for unregulated production.

First we investigate the effect of feedback on timing precision using a simple example: we suppose that the production rate of Y is not a constant but rather is a simple sigmoidal function of the current number of molecules y ,

$$\beta(y) = \beta_0 \{1 + \tanh[H_y(y/K_y - 1)]\}, \quad (4.1)$$

where positive (negative) H_y corresponds to positive (negative) feedback, $|H_y|$ is the maximum steepness, K_y is the molecule number at which β is half-maximal, and β_0 is set to ensure that the average time at which y first reaches y_* is t_* . We calculate the variance σ_t^2 from the master equation by matrix inversion [17]. In Fig. 4.1B we see that when there is no feedback ($H_y = 0$), the variance satisfies $\sigma_t^2 y_*/t_*^2 = 1$, and that either positive or negative feedback increases the variance. This result is consistent with previous findings for a species that does not degrade [12], and it has an intuitive explanation: a sequence of time-ordered stochastic events is most precisely timed if the mean time for each event to occur is equal, but feedback makes these times unequal.

Next we investigate the interplay of feedback and regulation by introducing a second species X that is produced at a constant rate α_0 . The Y production rate $\beta(x, y)$ is now a function of both molecule numbers x and y . We find that if it is a simple sum $\beta(x, y) = f_1(x) + f_2(y)$ or product $\beta(x, y) = f_1(x)f_2(y)$ then feedback continues to generically increase the timing variance, but if we include a coupling term $\beta(x, y) = f_1(x)f_2(y)f_3(xy)$ the situation is different. Specifically, Fig. 4.1C shows the case where

$$\begin{aligned} \beta(x, y) = & \beta_0 \{1 + \tanh[H_x(x/K_x - 1)]\} \\ & \times \{1 + \tanh[H_y(y/K_y - 1)]\} \\ & \times \{1 + \tanh[H_{xy}(xy/K_{xy}^2 - 1)]\}. \end{aligned} \quad (4.2)$$

We see that with no feedback ($H_y = 0$) we have $\sigma_t^2 y_*/t_*^2 < 1$, which demonstrates that regulation by a second species increases the timing precision as found previously [17]. However, now we also see that with positive feedback ($H_y > 0$), the variance can be even lower. Together with Fig. 4.1B, this result implies that although feedback increases timing noise in the absence of regulation, it can decrease timing noise in the presence of regulation.

Similarly we investigate the case where the feedback occurs on X , not Y . We take the production rates of x and y to be

$$\alpha(x) = \alpha_0 \{1 + \tanh[H_x(x/K_x - 1)]\}, \quad (4.3)$$

$$\beta(x) = \beta_0 \{1 + \tanh[H_y(x/K_y - 1)]\}, \quad (4.4)$$

respectively. We see in Fig. 4.1D that with negative feedback ($H_x < 0$) the variance is lower than with no feedback ($H_x = 0$), again implying that feedback can reduce timing noise when coupled to regulation.

To understand this effect, we develop a gradient-descent method to find the optimal regulation that minimizes the timing variance. The regulation is specified by the X and Y production rates $\alpha(x, y)$ and $\beta(x, y)$, respectively, which each depend on the molecule numbers x and y in general, but whose dependencies will later be restricted to consider particular feedback topologies. The probability of first reaching $y = y_*$ at time t is $P(t) = \sum_{\{\vec{s}\}} P(t|\vec{s})P(\vec{s})$, where

$$P(\vec{s}) = \prod_{i=0}^{S-1} \frac{r_i}{k_i}, \quad (4.5)$$

$$P(t|\vec{s}) = \left(\prod_{i=0}^{S-1} \int_0^\infty dt_i k_i e^{-k_i t_i} \right) \delta \left(t - \sum_{j=0}^{S-1} t_j \right). \quad (4.6)$$

In Eq. 4.5, $P(\vec{s})$ is the probability of taking a path \vec{s} from $(x_0, y_0) = (0, 0)$ to $(x_S, y_S) = (x_S, y_*)$ for any nonnegative x_S , where S is the length of the path. Each step i takes the system out of state (x_i, y_i) with rate $k_i = \alpha(x_i, y_i) + \beta(x_i, y_i)$ and into a new state

with probability r_i/k_i , where the new state is either $(x_i + 1, y_i)$ with $r_i = \alpha(x_i, y_i)$ or $(x_i, y_i + 1)$ with $r_i = \beta(x_i, y_i)$. In Eq. 4.6, $P(t|\vec{s})$ is the probability that traversing the given path \vec{s} takes a time t . The first term integrates over all values of each step's transition time t_i , which is exponentially distributed with rate k_i , and the second term ensures that the sum of these transition times is t . From $P(t)$ we calculate the moments in section 6.3, of which the first two are

$$\langle t \rangle = \sum_{\{\vec{s}\}} P(\vec{s}) \sum_{i=0}^{S-1} \frac{1}{k_i}, \quad (4.7)$$

$$\langle t^2 \rangle = \sum_{\{\vec{s}\}} P(\vec{s}) \left[\left(\sum_{i=0}^{S-1} \frac{1}{k_i^2} \right) + \left(\sum_{j=0}^{S-1} \frac{1}{k_j} \right)^2 \right]. \quad (4.8)$$

The optimal regulation function minimizes $\langle t^2 \rangle$ at fixed $\langle t \rangle = t_*$. Therefore, defining a vector $\vec{\gamma}$ whose components are all components of both the $\alpha(x, y)$ and $\beta(x, y)$ matrices, we initialize $\vec{\gamma}$ to satisfy $\langle t \rangle = t_*$ and update it as

$$\vec{\gamma}^{(n+1)} = \vec{\gamma}^{(n)} - \epsilon \vec{u}. \quad (4.9)$$

Here $\epsilon \ll 1$, and \vec{u} is such that $\vec{u} \cdot \nabla_{\vec{\gamma}} \langle t^2 \rangle$ is maximized with respect to the constraints $\vec{u} \cdot \nabla_{\vec{\gamma}} \langle t \rangle = 0$ and $|\vec{u}|^2 = 1$. To summarize this numerical method, we initialize the rates $\vec{\gamma}$, calculate the gradient of the timing variance $\sigma_t^2 = \langle t^2 \rangle - \langle t \rangle^2$ (Eqs. 4.7 and 4.8) with respect to $\vec{\gamma}$ at each point in state space, and update the rates such that $\nabla_{\vec{\gamma}} \sigma_t^2$ is maximized while keeping the target condition fixed by $\nabla_{\vec{\gamma}} \langle t \rangle = 0$. After many iterations this method converges to the minimum variance.

First we apply this method to the case where X regulates Y with no feedback. Thus, we fix $\alpha = \alpha_0$ and optimize $\beta(x)$. Figure 4.2A shows the result, and we see that the optimal $\beta(x)$ is an increasing function of x (i.e., X activates Y). The reason, clear from the mean dynamics in 4.2B, is that as x increases over time, $\beta(x)$ increases over time, which causes y to accelerate. The acceleration allows \bar{y} to cross y_* with a large slope, reducing the uncertainty of the crossing time. We observed this effect

previously with Hill-function activation [17], but the optimal regulation function was unknown.

Next we keep $\alpha = \alpha_0$, but we allow feedback on Y and find the optimal $\beta(x, y)$. Figure 4.2C shows the result, and we see that the optimal $\beta(x, y)$ depends on y , confirming that feedback is beneficial in the presence of regulation. Specifically, we see that $\beta(x, y)$ decreases with y (negative feedback) when x is small, and increases with y (positive feedback) when x is large. These two properties are also exhibited by Eq. 4.2 with $H_x < 0$, $H_y > 0$, and $H_{xy} < 0$ as in Fig. 4.1C. The first property ensures that Y is not prematurely activated at early times when x is small. The second property provides an additional acceleration of y at late times when x is large. Thus, X acts as a “timer” for Y , allowing Y to apply self-amplification only at late times. This has two advantages, as seen in Fig. 4.2D: (i) it increases the slope of \bar{y} at crossing, beyond that without feedback; and (ii) it allows the acceleration to begin at a \bar{y} value that is already close to y_* , thus reducing trajectory-to-trajectory variability caused by prolonged self-amplification [12].

Finally we consider the case where feedback acts on X instead of Y . Here, to provide a reasonable constraint on $x(t)$, we introduce a bound x_* and restrict $\alpha(x)$ such that $\bar{x}(t) \leq x_*$ over the range $0 \leq t \leq t_*$. The optimal regulation functions $\alpha(x)$ and $\beta(x)$ are shown in Fig. 4.2E. We see that X represses itself and activates Y , and that both regulation functions have a sharp transition when $x = x_*$. We see in Fig. 4.2F that the resulting dynamics are sharply kinked.

To understand the sharp nature of the optimal solution in Fig. 4.2E and F, we investigate our optimization scheme (Eqs. 4.5-4.9) analytically. The analytic version of Eq. 4.9 is $0 = \gamma_i \partial_{\gamma_i} (\langle t^2 \rangle - \lambda \langle t \rangle)$, where the Lagrange multiplier λ enforces $\langle t \rangle = t_*$, and the factor of γ_i in front enforces $\gamma_i > 0$ (see section 6.3). By inserting Eqs. 4.7 and 4.8 into this condition, we show in section 6.3 that it is satisfied when (i) α and β are such that all possible paths \vec{s} to reach $y = y_*$ have the same length S , and (ii) all transition rates along each of these paths are equal. Each such set of equal-length, constant-velocity paths is a local optimum, and the global optimum that minimizes

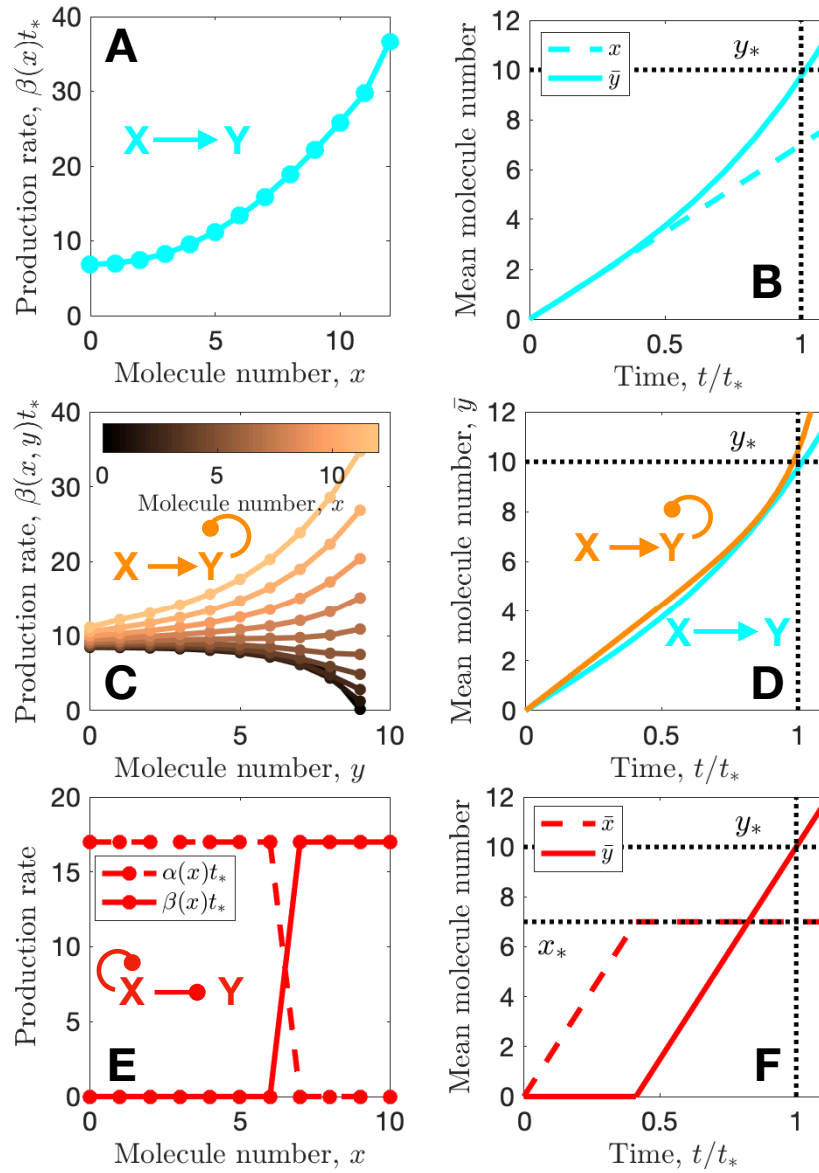


Fig. 4.2. Optimal regulation functions that minimize timing variance. (A) Without feedback, X activates Y , (B) allowing \bar{y} to accelerate before crossing y_* . (C) With feedback on Y , X acts as a “timer” for Y , allowing Y to self-repress at early times and self-activate at late times, and (D) providing further, late-phase acceleration of \bar{y} . (E) With feedback on X , it represses itself and activates Y sharply, (F) resulting in kinked dynamics where \bar{x} and \bar{y} growth are separated in time. Parameters are $\alpha_0 t_* = 7$ (A-D), $x_* = 7$ (E, F), and $y_* = 10$ throughout.

the timing variance is the set for which (iii) the path length S is as large as possible. More generally, if only property (ii) is satisfied, we show section that the timing variance satisfies

$$\frac{\sigma_t^2}{t_*^2} = \frac{\sigma_S^2}{\langle S \rangle^2} + \frac{1}{\langle S \rangle}, \quad (4.10)$$

where $\langle S \rangle$ and σ_S^2 are the mean and variance of the path lengths, weighted by the path probabilities $P(\vec{s})$. Clearly the variance is minimized when $\sigma_S^2 = 0$ and $\langle S \rangle$ is as large as possible, consistent with properties (i) and (iii) above, respectively.

Now we can understand why the the optimal solution in Fig. 4.2E and F looks the way it does. The sharp nature of the regulation functions ensures that at early times only x changes, and at late times only y changes, confining the stochastic dynamics to only one possible path in (x, y) space [property (i)]. The values of α and β , when they are nonzero, are constant and equal to each other, ensuring that the velocity along this path is constant [property (ii)]. Finally, both x and y attain their maximal values x_* and y_* , ensuring that the path is as long as possible [property (iii)].

Indeed, Fig. 4.3 shows the optimal solutions for all of the networks considered thus far in terms of these three properties. Specifically, Fig. 4.3A shows the mean dynamics in (x, y) space; Fig. 4.3B shows the velocity $v(t) = \sqrt{(d\bar{x}/dt)^2 + (d\bar{y}/dt)^2}$ along this path, normalized by its time average $\bar{v} = t_*^{-1} \int_0^{t_*} dt v(t)$; and Fig. 4.3C shows the variance σ_S^2 in the path length across all paths. With only Y and no X (blue), there is only one possible path (Fig. 4.3A), and therefore $\sigma_S^2 = 0$ (Fig. 4.3C). The optimal solution has constant velocity along the path (Fig. 4.3B), which is achieved with no feedback. When X regulates Y (cyan, orange), the mean path extends into the (x, y) plane (Fig. 4.3A), which increases its length and thus lowers the timing variance. However, it also makes the velocity non-constant (Fig. 4.3B) and allows for many possible paths such that $\sigma_S^2 > 0$ (Fig. 4.3C). Only upon allowing X to also regulate itself (red) does the path become as long as possible (Fig. 4.3A), constant-velocity (Fig. 4.3B), and unique (Fig. 4.3C).

The minimal values of the timing variance for the networks are shown by the filled circles in Fig. 4.4A. We see that the single species Y achieves the standard $\sigma_t^2 y_*/t_*^2 = 1$

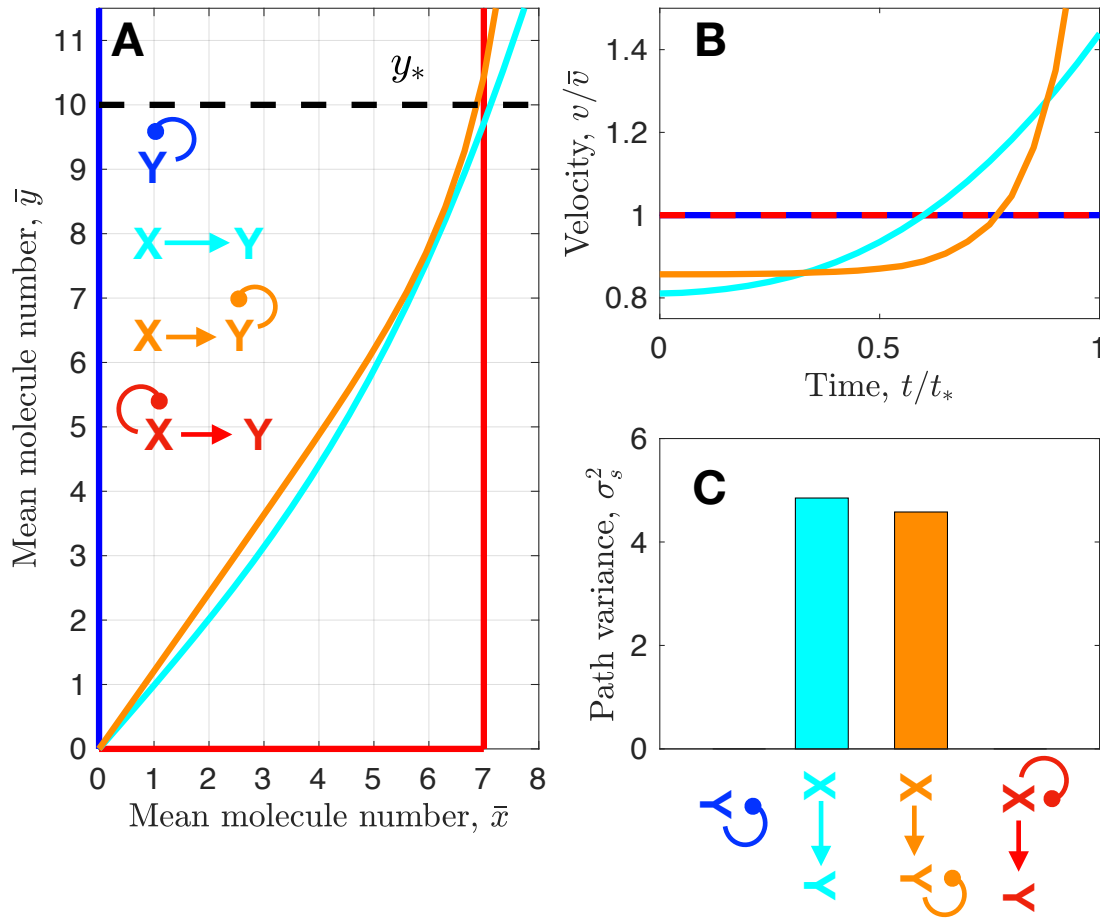


Fig. 4.3. Properties that minimize timing variance: (A) large path length S , (B) constant velocity $v(t)$ along path, and (C) small path length variance σ_S^2 . Parameters as in Fig. 4.2.

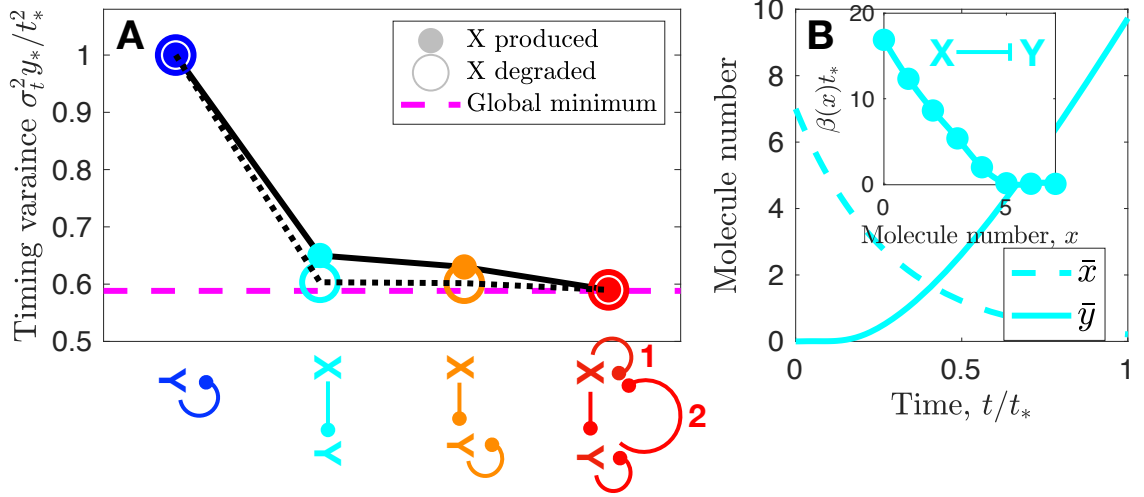


Fig. 4.4. (A) Ranking of timing variance for all one- and two-node networks. Global minimum is $\sigma_t^2 y_*/t_*^2 = y_*/(x_* + y_*)$. In red network, link 1, 2, or both is required. Parameters as in Fig. 4.2. (B) Mean dynamics and regulation function (inset) for case when X is degraded. Here $\alpha_0 t_* = 3.5$.

(blue), regulation by X lowers the variance (cyan), feedback on Y lowers it further (orange), and regulation of X lowers it to the global minimum given by Eq. 4.10 with $\sigma_S^2 = 0$ and $\langle S \rangle = x_* + y_*$, namely $\sigma_t^2 y_*/t_*^2 = y_*/(x_* + y_*)$. Because the results in Fig. 4.4A are minima, it does not matter in the last case whether the regulation of X is by X itself (red link 1), by Y (red link 2), or both; the optimal regulation functions will produce the red path in Fig. 4.3 regardless.

Thus far we have only considered the scenario where X is produced over time. However, X could alternatively be degraded over time [17]. In the cases where X is unregulated (cyan, orange), this corresponds to replacing its production propensity α_0 (for $x \rightarrow x + 1$) with a degradation propensity $\alpha_0 x$ (for $x \rightarrow x - 1$). The resulting minimal values of the timing variance are shown by the open circles in Fig. 4.4A, and we see that they are lower than the corresponding values when X is produced over time (filled circles). The reason, illustrated for the case where X regulates Y in Fig. 4.4B, is that when X is produced over time it increases linearly (Fig. 4.2B dashed),

whereas when X is degraded over time it decreases exponentially (Fig. 4.4B dashed). The curvature of the exponential begins to approximate the kinked dynamics of the globally optimal solution (Fig. 4.2F dashed). Specifically, X is most dynamic at early times (Fig. 4.4B dashed), and Y is only produced once x drops below a particular value (Fig. 4.4B inset) allowing it to be most dynamic at late times (Fig. 4.4B solid). Thus, even without feedback, the nonlinear dynamics of a degraded regulator allow its target to more closely approach the globally optimal timing precision.

4.3 Experimental observation in QR.p cells in *C. elegans*

Our collaborators performed experiments on mutants of *C. elegans* to test our hypothesis that timing variance decreases in the presence of regulation and feedback. During *C. elegans* development a cell called QR.p migrates for specific amount of time [5]. It has been shown that temporal regulation of QR.p is controlled by expression of *mig-1* through Wnt signaling [5]. The experiments were performed using CRISPR gene knockout. Through this method we found that the *bar-1* gene regulates the production of *mig-1*, and *mig-1* in turn feeds back on *bar-1*. We measured the number of *mig-1* mRNA produced by a QR.p cell during migration. Experiments were performed on three types of populations (Figure 4.5): i) Wild type (WT); ii) Active, where *bar-1* was over-expressed such that the effect of *bar-1* on *mig-1* was saturated; and iii) Mutant, where *bar-1* was removed. Each data point gathered, corresponds to a particular *C. elegans* individual; thus experiments were performed on the population to acquire data at various time points (Figure 4.5). We calculate the variance of stopping times in all cases. We considered that the migration stops when a threshold of *mig-1* expression (y_*) is crossed. We assumed that this threshold can be anywhere higher than $y_* = 10$. We chose the minimum threshold to be 10 because the data show a clear distinction between low copy number of *mig-1* and high copy number (Figure 4.5A). We calculate variance in stopping times ($\sigma_{t_*}^2 y_*/t_*^2$) for each population (4.5A, B, C). The graphical representation of regulation for each popula-

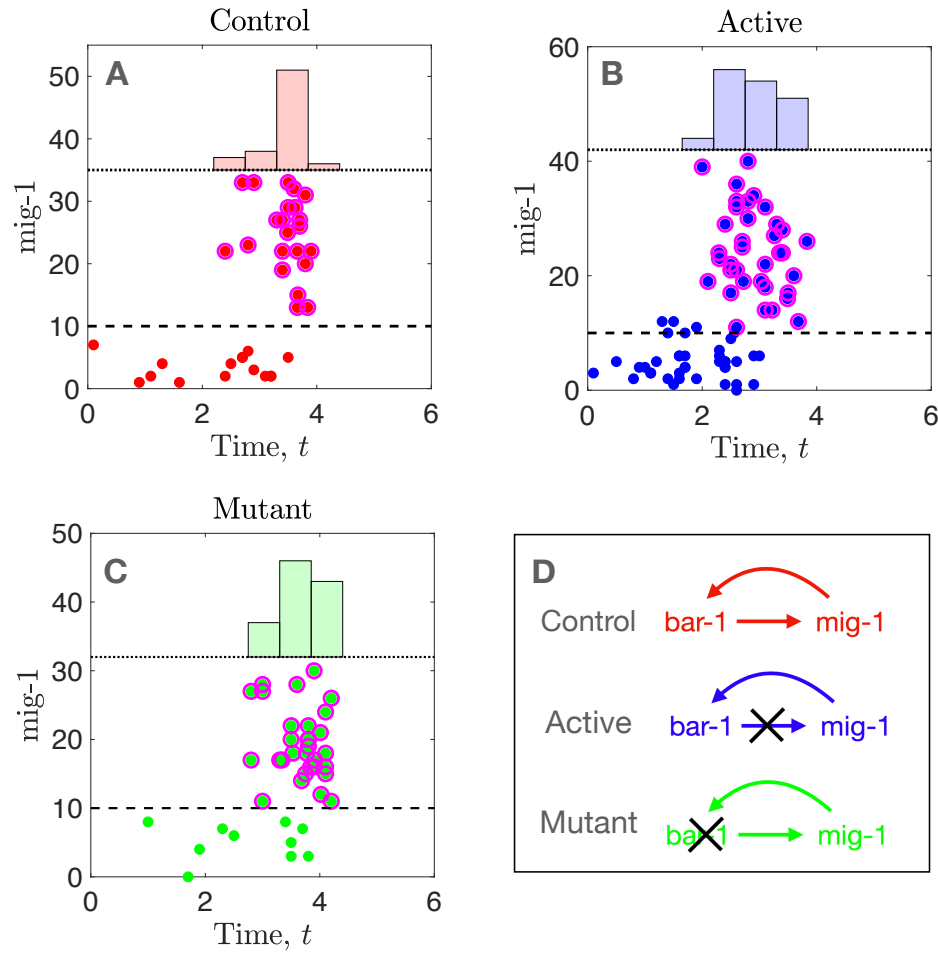


Fig. 4.5. *mig-1* expression in QR.p cell. The bar graph represent the distribution of points above the black dashed line. We choose black dashed line to be at protein number to be 10. (A) *mig-1* expression for wild type population. , (B) *mig-1* expression for active population, (C) *mig-1* expression for mutant population, and (D) systematic gene regulation network of control, active, and mutant.

tion is shown in Figure 4.5D. For the control group *mig-1* expression is regulated by *bar-1*, and *bar-1* is regulated by *mig-1*. For the active group *bar-1* is overly expressed, such that it's effect is saturated on *mig-1*; hence, no regulation on *mig-1*. Similarly for the mutant group *bar-1* is absent; therefore no effect on *mig-1*.

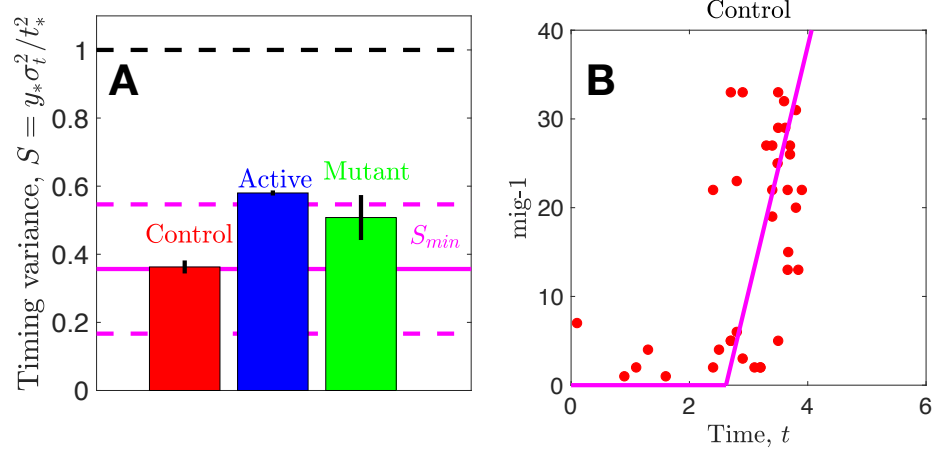


Fig. 4.6. A) The bar graph represent the distribution of points above the black dashed line. Black dashed line represent the minimum threshold for y_* . B) *mig-1* expression for wild type population and magenta line is piece-wise linear fit to WT.

The timing variance for each case is plotted in figure 4.6A. We find that the control population has the lowest timing variance, lower than the active and mutant groups. This suggests that the *bar-1* regulation on *mig-1* reduces the timing variance. We expect that without any regulation the timing variance should be 1 in these units, but we see the variance is still less than one in the active and mutant groups. This suggest that *mig-1* production is regulated by other genes in addition to the *bar-1*.

We further tested the degree to which *mig-1* is regulated in the control group. Our model predicts that for the two gene regulatory network the minimum global timing variance is when dynamics are sharp (Figure 4.3). We compare this regulation strat-

egy to control group. We fit the data as a piece-wise linear function. We considered error in x and y to fit the line. The error function is,

$$E = (1 - \gamma) \sum_{i=1}^n (\bar{x} - x_i)^2 + \gamma \sum_{i=1}^n (\bar{y} - y_i)^2, \quad (4.11)$$

where γ is a weight factor for y error, and n is the number of data points. For $\gamma = 1$, Eq. 4.11 reduces to a standard linear regression. We chose to include error in both terms because the x and y are both measured variables and subjected to error. We varied γ from 0 to 1 and calculated the fit line, and used the mean of all fitted lines. We calculated the global minimum timing variance obtained through the optimal sharp regulation strategy. That is, *bar-1* production is constant, and when a threshold of *bar-1* (x_*) is crossed, then *mig-1* production starts with a same rate as the *bar-1* production rate. The threshold x_* can be estimated as the intercept of best fitted line (Figure 4.6B). Using x_* and y_* we calculated the global minimum variance, $y_*/(x_* + y_*)$ as discussed below Eq. 4.10. The global minimum variance is plotted in Figure 4.6A as a magenta line. We see that the control variance is close to the threshold minimum and within its error bound, propagated from the errors in x_* and y_* .

4.4 Discussion

We have developed a gradient-descent approach that provides the optimal regulation functions for a given network topology that minimize the timing noise of a threshold-crossing event. The approach has revealed that feedback reduces timing noise in the presence but not absence of regulation because the combination of the two increases the number of transitions that must happen sequentially in molecular state space. More generally, our work suggests a perspective where noise is not minimized by finding the right network topology, but rather by finding the right combination of regulation functions that produce a path through state space that is as long, steady, and unique as possible. We anticipate that this perspective applies

broadly to biological processes where timing is crucial. Although our work is limited to two-species networks, whereas most biochemical networks are much larger, our approach is straightforward to generalize to larger and more complex networks given sufficient computational power.

Our findings generalize previously known results. It was previously found that Hill-function autoregulation increases noise [12]. Here, we have shown that in fact, any autoregulation function increases noise, as it results in unequal production rates at different points in state space, which is suboptimal. We previously found that Hill-function regulation by an upstream species reduces noise [17], but we lacked a procedure to determine whether Hill-function optimization is optimal. Here, we have computed the optimal regulation function that results in the minimal noise, and we see that it is not necessarily Hill-like but rather concave-up (Fig. 4.2A).

We have considered networks in which species can be produced or degraded, but not both. However, species that can be both produced and degraded are ubiquitous in biochemical networks. We have checked for small system sizes in the case where X regulates Y without feedback, that adding degradation to either X or Y does not significantly change the optimal regulation function from that in Fig. 4.2A, but it does increase the timing variance. The increase in the timing variance makes sense because degradation introduces a much larger set of possible paths for the molecule numbers to take through state space, as now one of the species can go down in molecule number as well as up. Consistent with our results herein, a larger path number should lead to higher path stochasticity and therefore larger threshold-crossing noise. This finding also suggests that if the degradation rate is also optimized over, the optimal degradation rate would be zero, and therefore that our results on theoretical optimality apply also when degradation is included. We leave a more comprehensive investigation of simultaneous production and degradation to future work.

We find that timing noise is minimized by following a single deterministic path through state space, which is likely unrealistic for biochemical reactions. However, this result is nonetheless useful because it demonstrates that under ideal conditions

any network will have a fundamental minimum timing variance. Here we provide the value of that variance for the simple cases of one- and two-species networks. This is important because it provides a bound to which an actual network may (or may not) come close, even if that network does not take a single deterministic path through state space. Moreover, the more realistic examples in Fig. 1 demonstrate that significant reductions in timing noise are possible due to regulation and feedback even without the globally optimal regulation functions. We anticipate that our general insights will serve as guiding predictions for experimental investigations, such as the findings that in order to increase timing precision species should be only produced or degraded but not both, and that different species should change molecule number at separate times.

Our findings suggest that a cellular process where timing precision is important should be governed by a molecular network with both multistep regulation and feedback, particularly one in which every species is subject to regulation as in Fig. 4.4A (red). An experimental example in which timing precision is particularly well studied is neuroblast migration in developing *C. elegans* larvae. Here, the QR neuroblast produces a protein called *mig-1* that crosses an abundance threshold to terminate migration; overproduction causes undermigration and vice versa [5]. Our collaborators have discovered that *mig-1* is regulated by *bar-1*. We compared the timing variance of wild type with active (where *bar-1* was over expressed) and mutant (where *bar-1* was absent) groups. We discovered that the timing variance is lower for the WT case compared to other two. The over-expression and no expression of *bar-1* reduces the path length to reach the target, hence increasing the timing variance. Furthermore, we compared the timing variance of wild type with the globally minimum timing noise. The noise of the wild type was found to be close to the global minimum.

Finally, our work has connections to other active areas of research and other biological systems. In principle, confining a stochastic system to a single deterministic path in state space should come at a large thermodynamic cost. Although we do not consider this cost here, the connection between noise reduction and energy ex-

penditure is a rich and active field [38–40], beginning with the seminal example of kinetic proofreading [41]. Furthermore, our finding that timing precision is maximized by systems that only move in one direction (i.e., production or degradation but not both), and whose species progress one at a time, are reminiscent of molecular motors on filaments. Systems of motors may exhibit these properties mechanically, via ratcheted motion and steric hinderance, and therefore may be promising examples of precise biological timers [25, 42–44].

4.5 Materials and methods

4.5.1 Calculation of the moments of the first passage time

Using Eqs. 4.5 and 4.6, we write the first passage time distribution as

$$\begin{aligned}
 P(t) &= \sum_{\{\vec{s}\}} P(\vec{s}) P(t|\vec{s}) \\
 &= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \frac{r_i}{k_i} \right) \left(\prod_{j=0}^{S-1} \int_0^\infty dt_j k_j e^{-k_j t_j} \right) \\
 &\quad \times \delta \left(t - \sum_{\ell=0}^{S-1} t_\ell \right) \\
 &= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \int_0^\infty dt_i r_i e^{-k_i t_i} \right) \delta \left(t - \sum_{j=0}^{S-1} t_j \right). \tag{4.12}
 \end{aligned}$$

The n th moment is

$$\begin{aligned}
\langle t^n \rangle &= \int_0^\infty dt \, t^n P(t) \\
&= \int_0^\infty dt \, t^n \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \int_0^\infty dt_i r_i e^{-k_i t_i} \right) \\
&\quad \times \delta \left(t - \sum_{j=0}^{S-1} t_j \right) \\
&= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \int_0^\infty dt_i r_i e^{-k_i t_i} \right) \left(\sum_{j=0}^{S-1} t_j \right)^n.
\end{aligned} \tag{4.13}$$

Specifically, the first and second moments are

$$\begin{aligned}
\langle t \rangle &= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \int_0^\infty dt_i r_i e^{-k_i t_i} \right) \left(\sum_{j=0}^{S-1} t_j \right) \\
&= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \frac{r_i}{k_i} \right) \sum_{j=0}^{S-1} \frac{1}{k_j} \\
&= \sum_{\{\vec{s}\}} P(\vec{s}) \sum_{j=0}^{S-1} \frac{1}{k_j}
\end{aligned} \tag{4.14}$$

and

$$\begin{aligned}
\langle t^2 \rangle &= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \int_0^\infty dt_i r_i e^{-k_i t_i} \right) \left(\sum_{j=0}^{S-1} t_j \right)^2 \\
&= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \int_0^\infty dt_i r_i e^{-k_i t_i} \right) \\
&\quad \times \left(\sum_{j=0}^{S-1} t_j^2 + \sum_{j=0}^{S-2} \sum_{\ell=j+1}^{S-1} 2t_j t_\ell \right) \\
&= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \frac{r_i}{k_i} \right) \sum_{j=0}^{S-1} \sum_{\ell=j}^{S-1} \frac{2}{k_j k_\ell} \\
&= \sum_{\{\vec{s}\}} \left(\prod_{i=0}^{S-1} \frac{r_i}{k_i} \right) \left(\left(\sum_{j=0}^{S-1} \frac{1}{k_j^2} \right) + \left(\sum_{j=0}^{S-1} \frac{1}{k_j} \right)^2 \right), \\
&= \sum_{\{\vec{s}\}} P(\vec{s}) \left(\left(\sum_{j=0}^{S-1} \frac{1}{k_j^2} \right) + \left(\sum_{j=0}^{S-1} \frac{1}{k_j} \right)^2 \right), \tag{4.15}
\end{aligned}$$

as in Eqs. 4.7 and 4.8, where the last line in each case recalls Eq. 4.5.

4.5.2 Analytic minimization of timing variance using Lagrange multipliers

To find the minimum variance when the mean is fixed to be t^* , we utilize Lagrange multipliers. Because the variance is a function of only the first and second moments and is monotonically increasing with the second moment, finding the minimum of the variance with a fixed mean is equivalent to finding the minimum of the second moment with a fixed mean. Thus, the set of r_ℓ values which produces the minimum variance is the set which solves

$$0 = \frac{\partial}{\partial r_\ell} (\langle t^2 \rangle - \lambda \langle t \rangle) \tag{4.16}$$

for Lagrange multiplier λ .

However, Eq. 4.16 raises an issue. Assume that $x_* = y_* = 1$. In this case, there are only three possible rates α_{xy} and β_{xy} , namely α_{00} , β_{00} , and β_{10} . There are also only two possible paths: $\vec{s}_1 = [\{0, 0\}, \{0, 1\}]$ and $\vec{s}_2 = [\{0, 0\}, \{1, 0\}, \{1, 1\}]$. Putting these rates and paths into Eqs. 4.14 and 4.15 yields

$$\begin{aligned} \langle t \rangle &= \frac{\beta_{00}}{\alpha_{00} + \beta_{00}} \frac{1}{\alpha_{00} + \beta_{00}} \\ &\quad + \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} \frac{\beta_{10}}{\beta_{10}} \left(\frac{1}{\alpha_{00} + \beta_{00}} + \frac{1}{\beta_{10}} \right) \\ &= \frac{1}{\alpha_{00} + \beta_{00}} \left(1 + \frac{\alpha_{00}}{\beta_{10}} \right) \end{aligned} \quad (4.17)$$

and

$$\begin{aligned} \langle t^2 \rangle &= \frac{\beta_{00}}{\alpha_{00} + \beta_{00}} \frac{2}{(\alpha_{00} + \beta_{00})^2} \\ &\quad + \frac{\alpha_{00}}{\alpha_{00} + \beta_{00}} \frac{\beta_{10}}{\beta_{10}} \left(\frac{2}{(\alpha_{00} + \beta_{00})^2} \right. \\ &\quad \left. + \frac{2}{(\alpha_{00} + \beta_{00}) \beta_{10}} + \frac{2}{\beta_{10}^2} \right) \\ &= \frac{2}{(\alpha_{00} + \beta_{00})^2} \left(1 + \frac{\alpha_{00}}{\beta_{10}} + \frac{\alpha_{00} (\alpha_{00} + \beta_{00})}{\beta_{10}^2} \right). \end{aligned} \quad (4.18)$$

By putting Eqs. 4.17 and 4.18 into Eq. 4.16 and solving the resulting system of equations, one obtains that some rates must be negative or even undefined depending on the order in which they are solved. Since negative rates are unphysical, we can enforce positivity by making the substitutions $\alpha_{xy} = \exp(a_{xy})/t^*$ and $\beta_{xy} = \exp(b_{xy})/t^*$ and finding the minimum variance in (a_{xy}, b_{xy}) space rather than $(\alpha_{xy}, \beta_{xy})$ space. This procedure can be done without ever leaving $(\alpha_{xy}, \beta_{xy})$ space by noting that $\partial/\partial a = (\partial\alpha/\partial a) \partial/\partial\alpha = \alpha (\partial/\partial\alpha)$ and similarly that $\partial/\partial b = \beta (\partial/\partial\beta)$. This allows Eq. 4.16 to be rewritten as

$$0 = r_\ell \frac{\partial}{\partial r_\ell} (\langle t^2 \rangle - \lambda \langle t \rangle). \quad (4.19)$$

Putting Eqs. 4.17 and 4.18 into Eq. 4.19 yields two possible solutions to the resulting equations: $[\beta_{00}, \alpha_{00}, \beta_{10}] = [1/t_*, 0, \beta_{10}]$ with $\sigma^2 = t_*^2$ for any value of β_{10} or $[\beta_{00}, \alpha_{00}, \beta_{10}] = [0, 2/t_*, 2/t_*]$ with $\sigma^2 = t_*^2/2$. Of important note is the fact that when $\alpha_{00} = 0$ only the \vec{s}_1 path is available, while when $\beta_{00} = 0$ only the \vec{s}_2 path is available. Thus, the variance is seen to be extremized when only one possible path is available and all rates along that path are equal. Additionally, the longer path yields a smaller variance.

This can be seen to be a simple case of a larger trend. For any possible values of x_* and y_* it is possible to choose a set of reaction rates such that there is only one possible path through (x, y) space. When this is done, the product terms in Eqs. 4.14 and 4.15 become identically 1 since $r_i = k_i$ must be true along the one possible path. All other paths will have $r_i = 0$ for some i and will thus not contribute. This allows Eq. 4.19 to be easily calculated for any r_ℓ that is in the single possible path,

$$\begin{aligned} 0 &= r_\ell \frac{\partial}{\partial r_\ell} \left(\left(\sum_{i=0}^{S-1} \frac{1}{r_i^2} \right) + \left(\sum_{i=0}^{S-1} \frac{1}{r_i} \right)^2 - \lambda \left(\sum_{i=0}^{S-1} \frac{1}{r_i} \right) \right) \\ &= \frac{\lambda}{r_\ell} - \frac{2}{r_\ell^2} - \frac{2}{r_\ell} \left(\sum_{i=0}^{S-1} \frac{1}{r_i} \right) \end{aligned} \quad (4.20)$$

Eq. 4.20 is true for all r_ℓ along the single path if and only if all r_ℓ along that path have the same value, which, from the restriction that the mean first passage time must be t^* and Eq. 4.14, means $r_\ell = S/t^*$. Putting these values back into Eq. 4.15 then allows the variance to be simply calculated to be $\sigma^2 = t^{*2}/S$.

Eq. 4.19 must hold for all off-path reactions as well. This can be seen to be true by noting that for all other paths at least one r_i must be 0 in the product term. If $\ell \neq i$ this fact is not changed and that path will still have 0 contribution. If $\ell = i$ then the r_ℓ in front of the derivative operator will still force that path to have 0 contribution since no k_i can be 0. Similarly, if r_ℓ is not a reaction that occurs at any state along the one possible path then the derivative will cause it to vanish since the contribution from the one possible path does not depend on rates that exist in other states, while

if r_ℓ is a 0 rate that exists at a state in the one possible path then the factor of r_ℓ in front of the derivative will cause the whole expression to vanish. Thus, choosing a set of reaction rates such that there is a single possible path and all rates along that path are equal is a solution to Eq. 4.19 for all r_ℓ . Additionally, since $\sigma^2 = t^{*2}/S$, the longer that path is the smaller the variance will be.

4.5.3 Derivation of the lower bound on timing variance

If all rates are the same, $k_i = k$, then Eqs. 4.14 and 4.15 become

$$\langle t \rangle = \sum_{\{\vec{s}\}} P(\vec{s}) \frac{S}{k} = \frac{\langle S \rangle}{k} \quad (4.21)$$

and

$$\langle t^2 \rangle = \sum_{\{\vec{s}\}} P(\vec{s}) \left(\frac{S}{k^2} + \frac{S^2}{k^2} \right) = \frac{\langle S \rangle}{k^2} + \frac{\langle S^2 \rangle}{k^2}. \quad (4.22)$$

We then have

$$\begin{aligned} \frac{\sigma_t^2}{\langle t \rangle^2} &= \frac{\langle t^2 \rangle - \langle t \rangle^2}{\langle t \rangle^2} \\ &= \frac{k^2}{\langle S \rangle^2} \left(\frac{\langle S \rangle}{k^2} + \frac{\langle S^2 \rangle}{k^2} - \frac{\langle S \rangle^2}{k^2} \right) \\ &= \frac{1}{\langle S \rangle} + \frac{\sigma_S^2}{\langle S \rangle^2}, \end{aligned} \quad (4.23)$$

as in Eq. 4.10 of the main text. Eq. 4.23 implies that the system does not need to be restricted to only one path, but rather to any set of paths of the same length. Then, $\sigma_S^2 = 0$ still, and a local minimum in the variance is still obtained.

We summarize the results of this and the preceding appendix by establishing three rules which state that the variance in first passage time is minimized when:

1. All possible paths have the same length,

2. The rate at which the system moves through state space is as constant as possible, and
3. The path length through state space is maximized.

5. TERMINATION OF QR.P CELL MIGRATION

This chapter has been adapted from C. Dubois, S. Gupta*, M.A. Felix, A. Mugler, “Sensitivity to perturbations of a cell migration under temporal regulation,” bioRxiv 2020.07.21.213710; under review in Development (* equal contributors)*

Cell migration is a key process during the development of many animal tissues. The direction of migration, the signal transduction and the cytoskeletal dynamics behind the movement has been widely studied [45–52]. However, little is known about the termination of migration, although it is obviously of key importance for final cell and organ position [53–55]. Termination of migration has been thought to be spatially triggered by homogeneous concentration of guiding cues [56], adhesion to a specific target [57] and/or physical barriers [55, 58]. However, recent studies have shown that the termination of migration is due to temporal cues in some cell types [5]. The occurrence of cell migration begs the question of the degree of precision in the final cell position. Most studies have been focused on the precision of the final position due to genetic effects or experimental perturbations [46, 55]. Here, we study the final position of QR.p cells subjected to natural genetic variation. We explain the termination point of QR.p cells by considering temporal precision in the migration of QR.p cells.

The QR neuroblast is a cell that migrates a long distance from the posterior to a more anterior position during first larval stage of the *C. elegans*. Three rounds of cell division take place during the migration [59]. The progeny are named according to their anterior or posterior position at each successive cytokinesis: QR.p is the posterior daughter of QR, and QR.pa the anterior daughter of QR.p. Finally, the daughter cells of QR.pa, QR.paa and QR.pap (hence after called QR.pax), acquire a neuronal identity [60]. The posterior-to-anterior migration of the QR.pa lineage stops upon expression of the Wnt receptor *mig-1* [5]. The expression of *mig-1* in

QR.pa is not induced by the cell reaching a certain position in the body; rather, it responds to temporal regulation. Indeed, preventing QR migration or increasing its speed does not alter the timing of *mig-1* expression [5]. After QR.pa stops migrating, the daughter cells of QR.pa, called QR.paa and QR.pap (hence after called QR.pax) separate in a dorsoventral direction while crossing each other in an antero-posterior direction [61,62]; they then differentiate without further change in cell body position (Figure 5.1).

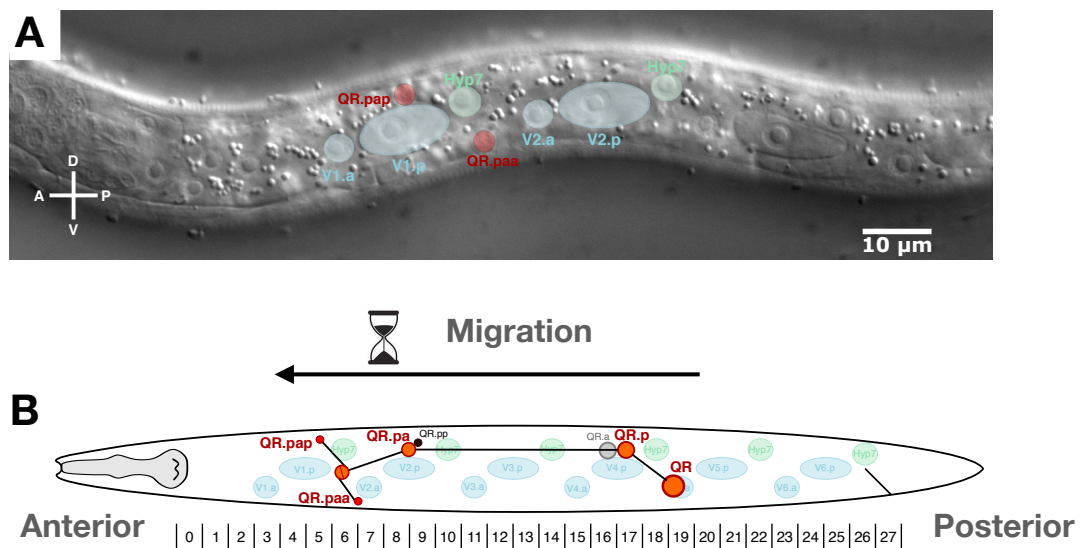


Fig. 5.1. QR.p neuroblast migration during the L1 stage. A) Nomarski micrograph of a L1 larva, where QR.pap and QR.paa have reached their final position. The relevant cells are outlined. B) Illustration of QR.p neuroblast migration during the L1 stage showing the final position of QR.paa and QR.pap and the relative scale used in scoring their position. This relative scale from 0 to 27 was constructed relative to the seam cells, as shown here.

Given the temporal regulation of *mig-1* expression and of the migration endpoint, a change in L1 larva body size should lead to a shift in the final cell position relative to other landmarks along the antero-posterior axis, provided that cell migration occurs

at the same speed in *C. elegans* of different body sizes. The expectation is that QR.pa stops in a more posterior position relatively to body landmarks in longer *C. elegans*. To assess the effect of body size on the relative position of QR.pax, we developed a mathematical model of the expected relationship between body size and QR.pax final position, while accounting for larval growth during cell migration. Our collaborator used different body size mutants in the reference genetic background N2 to confirm the hypothesis.

5.1 Results

The cessation of QR.pa migration relies on the timing of expression of *mig-1* and not the position of the cell in *C. elegans* [5]. From this observation, we predicted that the body size affects QR.pax final position: in a longer body, the cell should stop at the same time, i.e. at a more posterior position. Conversely, the cell should stop at a more anterior position in a shorter body (Figure 5.2A, left). To make this intuition quantitative, we develop a mathematical model for the no compensation hypothesis.

5.1.1 Mathematical model for no compensation hypothesis

For the no compensation model, we assume that the migration velocity is constant across all mutants. We treat the dynamics of the QR cell lineage as one dimensional within a growing larva. Taking the pharynx to be stationary at the origin $x = 0$ and rectum to moving away with constant velocity u due to growth, the length of the pharynx-to-rectum region evolves in time according to

$$\ell(t) = \ell_0 + ut, \quad (5.1)$$

where ℓ_0 is the length at hatching ($t = 0$). The dynamics of the cell position x are

$$\frac{dx}{dt} = -v(t) + \frac{x}{\ell(t)}u, \quad (5.2)$$

where the first term is the leftward velocity of the cell, and the second term is an effective rightward velocity due to the growth. In the second term we enforce uniform expansion of the larva during growth, such that intermediate points move according to the fraction of the distance to the origin. We assume that the cell begins migrating at a constant velocity v_0 at a time $t = \tau$ after hatching and stops migrating at time $t = T$,

$$v(t) = \begin{cases} 0, & \text{if } t < \tau \\ T, & \text{if } t \leq \tau \leq T \\ 0, & \text{if } t > T \end{cases} \quad (5.3)$$

Integrating Eq. 5.2 gives

$$x(t) = \begin{cases} \ell(t) \frac{x_0}{l_0}, & \text{if } t < \tau \\ \ell(t) \left[\frac{x_0}{l_0} - \frac{v_0}{u} \log \left(\frac{\ell(t)}{\ell(\tau)} \right) \right], & \text{if } t \leq \tau \leq T \\ \ell(t) \left[\frac{x_0}{l_0} - \frac{v_0}{u} \log \left(\frac{\ell(T)}{\ell(\tau)} \right) \right], & \text{if } t > T \end{cases} \quad (5.4)$$

where x_0 is the initial position of the cell. The relative position on the semi-discrete scale is $p = Nx/\ell$, where $N = 27$ (Figure 5.1B). Inserting Eq. 5.4, we have

$$p(t) = \begin{cases} p_0, & \text{if } t < \tau \\ p_0 - \frac{Nv_0}{u} \log \left(\frac{\ell(t)}{\ell(\tau)} \right), & \text{if } t \leq \tau \leq T \\ p_0 - \frac{Nv_0}{u} \log \left(\frac{\ell(T)}{\ell(\tau)} \right), & \text{if } t > T \end{cases} \quad (5.5)$$

where $p_0 = 19$ is the initial position of the cell (Figure 5.1B). We take $u = (35 \mu\text{m})/(6 \text{ hours}) = 5.6 \mu\text{m/hr}$ (Figure 5.3A) (every mutant grows approximately $35 \mu\text{m}$ in 6h, from the experiments), and $\tau = 2 \text{ hr}$ and $T = 8 \text{ hr}$ from the literature [59, 63, 64], leaving only v_0 as a fit parameter. Figure 5.2C (red curve) shows p versus ℓ_0 from Eq. 5.5 ($t > T$) with best-fit value $v_0 = 11.6 \mu\text{m/hr}$. Our collaborators measured the final position of QR.p, and we compare with our hypothesis. We see that the red curve even with best fit does not agree with the data. Without body size compensation, the empirical

measurements show a smaller effect of body size, indicating the presence of a partial compensation mechanism.

We investigated two different scenarios to decipher the relationship between body size and QR.pax final position in the context of partial compensation: an adaptive QR.p migration or an adaptive QR.p cell velocity which is positively related to the body size. In the first case, given that the temporal regulation of *mig-1* expression occurs at the end of the migration the relative position of QR.p division should be similar between long and short animals. Our experimental measurements on final position of QR.p does not support this mechanism (Figure 5.3B); the QR.p position is more anterior in small animals. Hence, we use adaptive migration velocity. Therefore, we adjust our model for partial compensation.

5.1.2 Mathematical model for partial compensation hypothesis

For the partial compensation model, we used Eq. 5.2 to infer the cell velocity from measurements of x and ℓ at two time points $t_1 = 3$ hr and $t_2 = 6$ hr. Calling $x(t_1) = x_1$, $x(t_2) = x_2$, $\ell(t_1) = \ell_1$, and $\ell(t_2) = \ell_2$ for short, and recognizing that both t_1 and t_2 fall between τ and T , we have from Eq. 5.4,

$$\frac{x_1}{\ell_1} = \frac{x_0}{\ell_0} - \frac{v_0}{u} \log \left(\frac{\ell_1}{\ell(\tau)} \right), \quad (5.6)$$

$$\frac{x_2}{\ell_2} = \frac{x_0}{\ell_0} - \frac{v_0}{u} \log \left(\frac{\ell_2}{\ell(\tau)} \right), \quad (5.7)$$

at each point. Subtracting Eq. 5.7 from Eq. 5.6 gives

$$\frac{x_1}{\ell_1} - \frac{x_2}{\ell_2} = \frac{v_0}{u} \log \left(\frac{\ell_2}{\ell_1} \right). \quad (5.8)$$

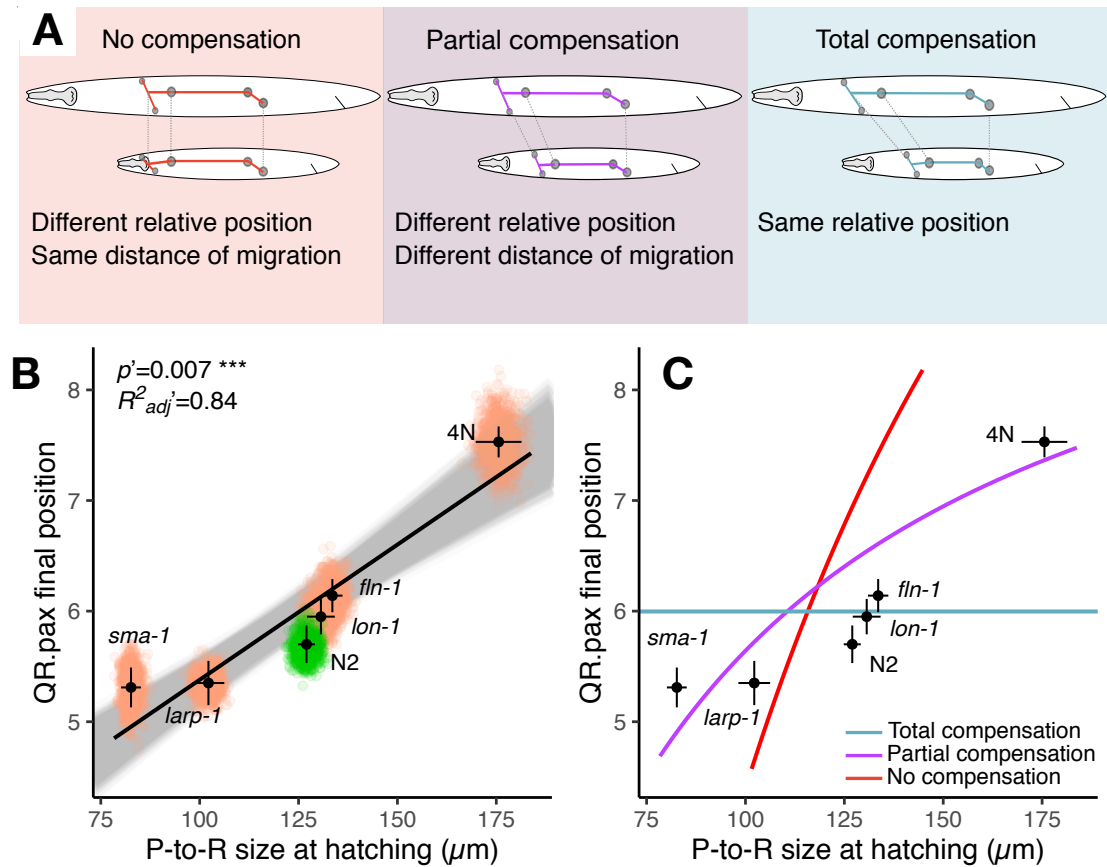


Fig. 5.2. Sensitivity of QR.pax position to body size. A) Schematics of the relative position of QR.pax in a long versus a short animal in the absence of body size compensation (left), with full compensation (right) or with a partial compensation (middle). B) Relationship between QR.pax final position and Pharynx-to-Rectum distance at hatching relationship in a subset of the panel. Orange and green dots represent one round of subsampling of the data. Black dots and error bars represent the mean and confidence intervals (CI 95%) from the data for each genotype. The black line represents the regression line from the data. The grey area represents regression lines after each iteration of subsampling. C) Mathematical model of the relationship between QR.pax final position and Pharynx-to-Rectum distance assuming full body size compensation, no compensation (with one fit parameter, cell velocity), or partial compensation (with no fit parameters).

Because growth is linear in time (Eq. 5.1), we have $u = (\ell_2 - \ell_1)/(t_2 - t_1)$. Inserting this expression into Eq. 5.8 and solving for v_0 gives

$$v_0 = \left(\frac{x_1}{\ell_1} - \frac{x_2}{\ell_2} \right) \frac{\ell_2 - \ell_1}{(t_2 - t_1) \log(\ell_2/\ell_1)} \quad (5.9)$$

Eq. 5.9 was used to calculate the cell velocities in the *lon-1* ($x_1 = 97.9 \mu\text{m}$, $x_2 = 73.5 \mu\text{m}$, $\ell_1 = 160.1 \mu\text{m}$, $\ell_2 = 195.4 \mu\text{m}$) and *sma-1* ($x_1 = 69.8 \mu\text{m}$, $x_2 = 46.1 \mu\text{m}$, $\ell_1 = 103.2 \mu\text{m}$, $\ell_2 = 110.9 \mu\text{m}$) mutants, giving $v_0 = 13.9 \mu\text{m/hr}$ and $v_0 = 9.3 \mu\text{m/hr}$, respectively (Figure 5.3B). Assuming a linear relationship between velocity and body size,

$$v_0 = m\ell_2 + b, \quad (5.10)$$

the two values of v_0 and ℓ_2 imply $m = 0.054 \text{ hr}^{-1}$ and $b = 3.3 \mu\text{m/hr}$. Recognize that $\ell_2 = \ell_0 + ut_2$, Eq. 5.10 becomes

$$v_0 = m(\ell_0 + ut_2) + b, \quad (5.11)$$

with $ut_2 = 35 \mu\text{m}$ (Figure 5.3A). Figure 5.2C (purple curve) shows p versus ℓ_0 from Eq. 5.5 ($t < T$) with Eq. 5.11 inserted for v_0 .

In this scenario, we assumed that the cell migrates faster in long animals. Based on measurements of cell position at two time points during migration ($t = 3\text{h}$ and $t = 6\text{h}$ after hatching) and the inference of the cell velocity through Eq. 5.9. We found that the migration speed of QR lineage from 3h to 6h after hatching was higher in *lon-1* ($13.9 \mu\text{m/hr}$) compared to small mutant *sma-1* ($9.3 \mu\text{m/hr}$) (Figure 5.3B). These results are consistent with the partial compensation mechanism of body size acting on cell velocity. Incorporating these data into the model gives good agreement with the measurement using no free parameters (Figure 5.2C, purple curve).

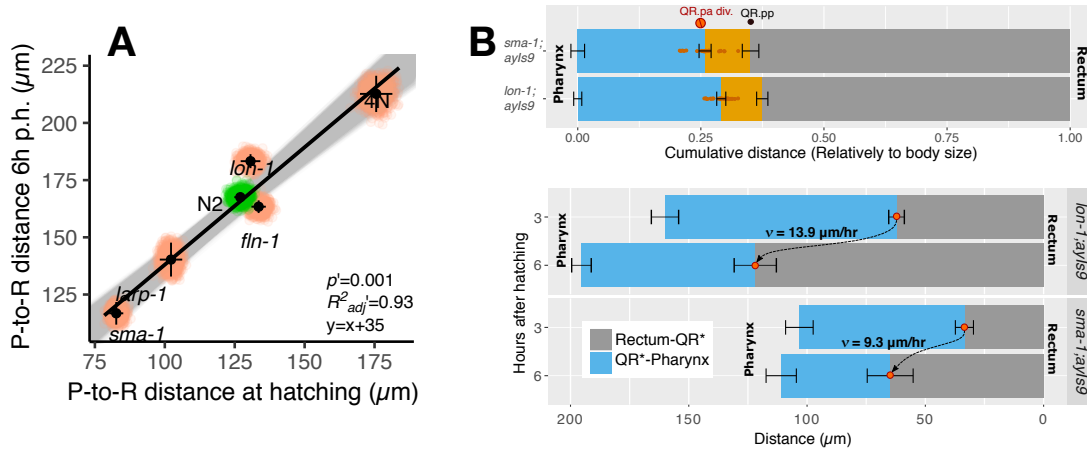


Fig. 5.3. (A) Relationships between pharynx-to-rectum (P-to-R) distance at 0 and 6h after hatching in N2 (green), size mutants (pink). The black line represents the regression line from the data. (B) Absolute distances between the rectum, QR cell* and the pharynx 3h and 6h after hatching in *lon-1*; *ayIs9* and *ayIs9 sma-1* animals. At 3h QR* is either QR or QR.p in some animals, at 6h QR* is either QR.p division or QR.pa in *lon-1*; *ayIs9* and mostly QR.p in *ayIs9 sma-1*.

5.2 Discussion

We observed that cells migrate farther, relative to the body size, when the body size is small, and vice versa. This observation is qualitatively consistent with the fact that these cells stop after a certain amount of time: if the speed is constant, they will migrate a constant distance, which will be larger relative to a smaller body. However, quantitatively, the degree of this effect was observed to be less than predicted by the model, even for the best-fit value of the constant speed. We therefore hypothesized that partial compensation may be acting through a change in cell speed as a function of body size, i.e. if QR and its progeny migrate faster at larger body size. To test this hypothesis, we measured the speed in *sma-1* versus *lon-1* mutants and using the model, we found that partial compensation of body size did operate through a change

in cell speed. The improved model provided a quantitatively accurate description of the measurements. Moreover, given the known inputs (cell start position, cell start and stop time, cell velocity in two mutants, and larval growth speed), the improved model succeeded with no fit parameters.

Our mathematical model is minimal in its construction yet quantitatively accounts for the observation. Some simplifying assumption are supported by the data, for example the observation that all larva, independent of mutant strain, grow a constant amount in 6 hours. We have checked other assumption explicitly; for example, we find that the results are negligibly changed if the cell accelerates and decelerates instead of instantaneously starting and stopping it's migration. This result also suggests that our result would be robust to details such as temporal variation in cell speed or pause due to cell division.

A possible mechanism of cell speed dependence on body size could be the following. First, migration speed could be increased at larger cell size. Nonetheless, evidence from literature tends to associate a negative correlation between cell size and cell velocity in vitro [65,66]. Second, body size could affect extracellular matrix density, such that larger cells secrete a less tight matrix, resulting in faster net migration speed, as in the *emb-9* matrix collagen mutant [5,67]. Finally, body size could affect the properties of the Wnt gradient influencing QR migration; this could operate if for instance a larger body size resulted in stronger Wnt concentration at a given relative position, resulting in faster cell speed [5].

6. TRAVELING WAVE DETECTION BY SINGLE-CELLED ORGANISMS

Previous projects investigated cell migration due to temporal regulation. However, the problem of cell migration is very common in biology. Cells also migrate in response to spatially heterogeneous cues from the surrounding environment. Chemotaxis is the migration of cells due to a chemical gradient of a chemoattractant. Cell migration due to spatial cues plays an important role in many biological processes including embryogenesis, neural patterning, and tumor dissemination [20–22, 68]. Cell migration due to a static gradient of chemoattractant is well understood [69]. However, natural chemical gradients to which cells respond are often dynamic, with both spatial and temporal components [70, 71]. Therefore chemotaxis in natural environments often requires integration of temporal and spatial information. One such example is the self-organized chemoattractant field arising during the development of the social amoeba *Dictyostelium discoideum*. Here, nondissipating waves of chemoattractant travel outward from an aggregation center and provide long-range cues to direct the migration of cells toward the wave source. If cells only sense the spatial cues when the traveling wave hits the front of the cell, the cell will move forward, and when the wave hits the back of the cell, it will reverse the direction, resulting neutral movement. Cells must integrate temporal information to sense the direction of the wave and thereby overcome neutral movement. In one experiment using spatially uniform concentrations of chemoattractant, neutrophils cells were observed to maintain their polarity when the concentration was increasing, but reversed polarity when the concentration was decreasing [72]. In another study, *Dictyostelium* exhibited a chemokinetic response to temporal oscillation of chemoattractant [73, 74]. It has also been shown that *Dictyostelium* moves towards the direction of cAMP wave source, and reverses its direction for longer wave period [75]. Despite substantial evidence for

this pattern of behavior, we do not understand the mechanism for encoding temporal and spatial information.

6.1 Mathematical model for traveling wave sensing

We hypothesize that cells sense the direction of incoming traveling waves using spatial and temporal information through a system of inhibitors. We assume the signaling wave is traveling from left to right in one dimension (Figure 6.1). When the wave encounters receptors on the surface of cell, the receptors activate. Upon activation of a sufficient number of receptors, we assume the receptors release inhibitors. Inhibitors are free to diffuse inside the cell. Once inhibitors reach the right side of the cell, inhibitors deactivate the receptors on the right hence no activation of the right side receptors through the wave. This mechanism encodes the direction of the traveling wave.

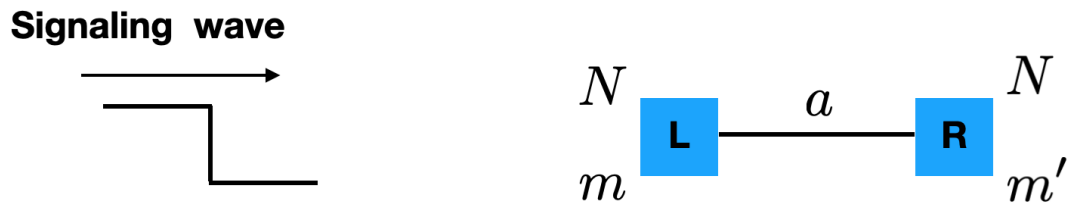


Fig. 6.1. The signaling wave traveling towards the cell. Cell has diameter a and it is divided in two parts left (L) and right (R), each parts of cell surface has N receptors. m Receptors are needed to activate to release the inhibitors on left side of cell. m' is the receptors activated by signaling wave on the right side before inhibitor diffuses to right side.

First we consider a simplified model by using a deterministic approach to demonstrate wave detection by the cell. Later we consider a general class of traveling wave and use a stochastic approach. We limit our study to a one dimensional system for simplicity; however, the model can be extended to three dimensions.

We consider a step wave moving to the right with a velocity v . The wave can be described as,

$$\phi(x, t) = \begin{cases} 1, & \text{if } x \leq vt \\ 0, & \text{otherwise} \end{cases} \quad (6.1)$$

We assume that there are N receptors on each side of the cell. Receptors release inhibitors when m receptors are activated. The probability distribution of the time for any one receptor to activate is exponentially distributed,

$$P_1 = ke^{-kt}, \quad (6.2)$$

where $k = k_0\phi(x, t)$ is the activation rate for each receptors. We can write the probability distribution for m receptors getting activated out of N by considering there are $\binom{N}{m}$ possibilities [76],

$$P_{m|N}(t) = \binom{N}{m} m P_1(t) \left(\int_0^t dt' P_1(t') \right)^{m-1} \left(1 - \int_0^t dt' P_1(t') \right)^{N-m} \quad (6.3)$$

Thus the mean time $t_{m|N}$ for m receptors to be activated out of N total receptors can be calculated as $\int t P_{m|N}(t)$. Integrating over all time we get,

$$\bar{t}_{m|N} = \binom{N}{m} \frac{m}{k_0} \sum_{l=0}^{m-1} \frac{(-1)^l}{(N - m + l + 1)^2} \quad (6.4)$$

We assume at $t = 0$ the wave is at the left end of the cell and that it takes $t_{m|N}$ time to activate m receptors. Inhibitors are released in the cell once m receptors are activated. Inhibitors are free to diffuse inside the cell. The distribution of times for diffusing molecules to diffuse to the right side with a hard boundary can be described as follows [77],

$$P_D(t|a) = \sum_{n=0}^{\infty} (-1)^n \frac{(2n+1)\pi D}{a^2} e^{-(n+1/2)^2 \pi^2 D t / a^2}, \quad (6.5)$$

where a is the diameter of the cell and D is the diffusion constant. The mean time for inhibitors to diffuse to the right end of the cell can be calculated using Eq. 6.5 and is found to be $t_D = a^2/2D$. After inhibitors reach the right side, they deactivate receptors on the right side from further activation from wave. We calculate the number of right-side receptors activated by the wave m' before inhibitors reach the right side. The time taken by the wave to reach right side is $t_w = a/v$. The wave has $\bar{t} = t_{m|N} + t_D - t_w$ time to activate the receptors. Each receptor has ke^{-kt} probability to activate. Therefore, the number of receptors, m' activated in time \bar{t} can be calculated as,

$$m' = N \int_0^{\bar{t}} dt k e^{-kt} \quad (6.6)$$

If the wave reaches the right end after the inhibitors arrive, inhibitors will deactivate all receptors on right side in this case, hence $m' = 0$. By solving Eq. 6.6 we get,

$$m' = \begin{cases} N \left(1 - e^{-k\bar{t} - (\frac{\tau_1}{2} - \tau_2)} \right), & \text{if } k\bar{t} \geq -(\frac{\tau_1}{2} - \tau_2) \\ 0, & \text{otherwise} \end{cases} \quad (6.7)$$

where $\tau_1 = ka^2/D$, and $\tau_2 = ka/v$ are dimensionless parameters. We calculate the gradient of activated receptors as $(m - m')/N$. This gradient estimates the wave direction memory. For a large gradient, the cell is able to detect the wave direction precisely because of a large difference in activated receptors between the left and right side of the cell. Figure 6.2A (cyan line) plots the gradient for varying m for $\tau_1 = 1$ and $\tau_2 = 1$. We see that the gradient has an optimum. This optimum occurs at the point when the wave and inhibitors reach at right side at the same time. If the wave reaches earlier than the inhibitors then it activates receptors on right side hence reducing the gradient. However, if the wave reaches after the inhibitors then having a large m increases the gradient. We calculated the location of optimum m^*/N and its value $\Delta m^*/N$ for the range of τ_1 and τ_2 (figure 6.2B, C). We see a boundary where m^* is 1. This boundary represents the trade-off line for the time at which inhibitors and the wave reach the right side.

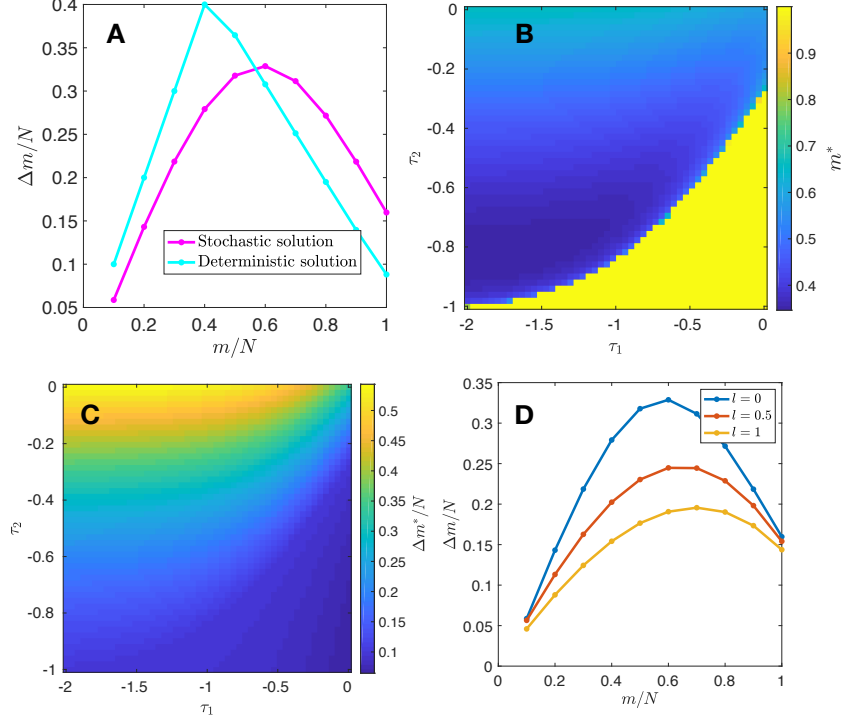


Fig. 6.2. A) Stochastic and deterministic solution to Δm for $\tau_1 = 1$ and $\tau_2 = 1$, B, C) Optimum m^* and $\Delta m/N$ value at optimum for various τ_1 and τ_2 , D) Stochastic solution for varying width $l = w/a$. $N = 10$

We further study more a generalized system by considering the signaling wave to be a smooth traveling wave. A traveling wave can be described by analytic solution of the Fisher-Kolmogorov equation [78]. We assume that the wave has a finite width w (as opposed to the step function case where $w = 0$). The analytic solution for Fisher-Kolmogorov equation is [78],

$$\phi(x, t) = \frac{1}{(1 + \exp((x - vt)/w))^2}, \quad (6.8)$$

hence the activation rate for any receptors is time dependent $k(x, t) = k_0 \phi(x, t)$. We derive the time distribution for the next reaction happening in section 6.3. The probability for any one receptor getting activated is given by $P_1(x, t) = k(x, t) e^{-k_0 \int_{-\infty}^t dt' \phi(x, t')}$.

We use Eq. 6.3 to calculate $P_{m|N}(x, t)$. To include the stochastic behavior of receptor activation and diffusion of inhibitors we perform a convolution of the probability distribution instead of using the mean time. We write the probability distribution of time for inhibitors reaching at the right side of the cell as,

$$P_I(t) = \int_{-\infty}^t dt' P_{m|N}(t') P_D(t - t'), \quad (6.9)$$

where t' is a instance of time when m receptors activate and $t - t'$ is the time left for inhibitors to reach at right side. On the right side of the cell, the probability of any one receptor getting activated in time t can be calculated as $\int_{-\infty}^t dt' P_1(t'|x = a)$. Hence total number of receptors activated before the wave reaches is,

$$m' = N \int_{-\infty}^{\infty} dt (1 - e^{-k_0 f(t-a/v)}) P_I(t) \quad (6.10)$$

where $f(x, t) = \int_{-\infty}^t dt' \phi(x, t')$. Integrating $\phi(x, t)$ we get,

$$k_0 f(x, t) = \ell \tau_2 \left[\frac{1}{1 + e^{-\frac{x}{w} + \frac{u}{\ell \tau_2}}} + \log \left(1 + e^{-\frac{x}{w} + \frac{u}{\ell \tau_2}} \right) - 1 \right] \quad (6.11)$$

where $u = k_0 t$ and $\ell = w/a$ are dimensionless time and length respectively. We solve Eq. 6.10 numerically and calculate the gradient for activated receptors. Figure 6.2A (magenta) plots the gradient against m/N for dimensionless parameter $l = 0$ (step wave). We see that the optimum persists for stochastic solution. The reason for this optimum is due to competition between the wave reaching the right end and inhibitors reaching the right end. By tuning when to release inhibitors the cell can maximize the precision by which it can sense the direction. However as the width of the wave increases, the optimum starts to disappear (Figure 6.2D). For higher width of wave the optimal m is increasing. This suggests a cell can sense better if it releases inhibitors at a later time. The reason is due to flat nature of wave. If the wave is smoothed, both ends of the cell are sensing the wave approximately at the same time

and with the same intensity. Hence the directional information of the wave is being lost.

6.2 Discussion

We have hypothesized a mechanism for a cell to detect the direction of a nutrient traveling wave. Our model is minimal but presents a strong hypothesis for wave sensing. The mechanism by which cells sense the direction requires the integration of spatial and temporal information. Cells use a system of receptors and inhibitors to encode the direction of the wave. Cells can maximize the direction sensing by controlling the time to release the inhibitors. Optimal time for the cell to release inhibitors is such that inhibitors and the wave reach the right side at the same time. Future works will be required to further validate our mechanism. We plan to extend our study by including noise in the traveling wave, as traveling waves are subjected to stochastic fluctuations.

6.3 Materials and methods

6.3.1 Calculation of the probability distribution for the next reaction

The probability of a next reaction happening at time t for a given reaction rate k can be derived as follows:

Assuming we divide time t in small sub-intervals of length Δt , we write the probability of a reaction at time t as having no reaction in first n sub-intervals and the reaction happening in next sub-interval,

$$P(t)\Delta t = (1 - k\Delta t)^n k\Delta t \quad (6.12)$$

where $n = \frac{\Delta t}{\Delta t}$. The first part of the right side represents the probability that the reaction does not happen in n sub-intervals, and second part represents that the

reaction happens in next sub-interval. Using the exponential approximation we can write equation 6.12 as,

$$P(t) = ke^{-kt} \quad (6.13)$$

For a traveling wave, the reaction rate k is time dependent. We assume the time dependent reaction rate can be described by,

$$k(t) = k_0 \frac{1}{(1 + \exp((x - vt)/w))^2} \quad (6.14)$$

Writing the probability of the reaction happening at time t by dividing the time in small sub-intervals Δt ,

$$P(t)\Delta t = (1 - k_1\Delta t)(1 - k_2\Delta t)....(1 - k_n\Delta t)k_{(n+1)}\Delta t \quad (6.15)$$

Taking the log of both sides,

$$\log P(t)\Delta t = \log(1 - k_1\Delta t) + \log(1 - k_2\Delta t).... + \log(1 - k_n\Delta t) + \log k_{(n+1)}\Delta t \quad (6.16)$$

Using Taylor's approximation,

$$\log P(t)\Delta t = -k_1\Delta t - k_2\Delta t.... - k_n\Delta t + \log k_{(n+1)}\Delta t \quad (6.17)$$

$$\log P(t)\Delta t = -\Delta t \sum_{i=1}^n k_i + \log k_{(n+1)}\Delta t \quad (6.18)$$

$$P(t) = k_{n+1}e^{-\Delta t \sum_{i=1}^n k_i} \quad (6.19)$$

Note that for $k_i = k$ we get the exponential time distribution. Solving for the exponential part,

$$f = -\Delta t \sum_{i=1}^n k_i \quad (6.20)$$

$$f = -\Delta t \sum_{i=1}^n k_0 \frac{1}{(1 + \exp((x - vi\Delta t)/w))^2} \quad (6.21)$$

For limit the $\Delta t \rightarrow 0$, $n \rightarrow \infty$. Hence we convert $\sum_i \rightarrow \int di$

$$f = -k_0 \Delta t \int_{i=0}^{i=t/\Delta t} di \frac{1}{(1 + \exp((x - vi\Delta t)/w))^2} \quad (6.22)$$

Substituting $i\Delta t \rightarrow t'$, we obtain the final result.

$$f = -k_0 \Delta t \int_{t'=0}^t dt' \frac{1}{(1 + \exp((x - vt')/w))^2}. \quad (6.23)$$

REFERENCES

REFERENCES

- [1] J. M. Bean, E. D. Siggia, and F. R. Cross, “Coherence and timing of cell cycle start examined at single-cell resolution,” *Molecular cell*, vol. 21, no. 1, pp. 3–14, 2006.
- [2] I. Nachman, A. Regev, and S. Ramanathan, “Dissecting timing variability in yeast meiosis,” *Cell*, vol. 131, no. 3, pp. 544–556, 2007.
- [3] B. L. Schneider, J. Zhang, J. Markwardt, G. Tokiwa, T. Volpe, S. Honey, and B. Futcher, “Growth rate and cell size modulate the synthesis of, and requirement for, g1-phase cyclins at start,” *Molecular and cellular biology*, vol. 24, no. 24, pp. 10 802–10 813, 2004.
- [4] K. Carniol, P. Eichenberger, and R. Losick, “A threshold mechanism governing activation of the developmental regulatory protein σ in bacillus subtilis,” *Journal of Biological Chemistry*, vol. 279, no. 15, pp. 14 860–14 870, 2004.
- [5] R. A. Mentink, T. C. Middelkoop, L. Rella, N. Ji, C. Y. Tang, M. C. Betist, A. van Oudenaarden, and H. C. Korswagen, “Cell intrinsic modulation of wnt signaling controls neuroblast migration in c. elegans,” *Developmental cell*, vol. 31, no. 2, pp. 188–201, 2014.
- [6] A. Amir, O. Kobiler, A. Rokney, A. B. Oppenheim, and J. Stavans, “Noise in timing and precision of gene activities in a genetic cascade,” *Molecular Systems Biology*, vol. 3, no. 1, p. 71, 2007.
- [7] H. Meinhardt, *Models of biological pattern formation*. Academic Press Inc, 1982.
- [8] D. E. Tufcea and P. François, “Critical timing without a timer for embryonic development,” *Biophysical journal*, vol. 109, no. 8, pp. 1724–1734, 2015.
- [9] J. Roux, M. Hafner, S. Bandara, J. J. Sims, H. Hudson, D. Chai, and P. K. Sorger, “Fractional killing arises from cell-to-cell variability in overcoming a caspase activity threshold,” *Molecular systems biology*, vol. 11, no. 5, p. 803, 2015.
- [10] N. Van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier, 1992, vol. 1.
- [11] H. H. McAdams and A. Arkin, “Stochastic mechanisms in gene expression,” *Proceedings of the National Academy of Sciences*, vol. 94, no. 3, pp. 814–819, 1997.
- [12] K. R. Ghusinga, J. J. Dennehy, and A. Singh, “First-passage time approach to controlling noise in the timing of intracellular events,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 4, pp. 693–698, 2017.

- [13] A. D. Co, M. C. Lagomarsino, M. Caselle, and M. Osella, “Stochastic timing in gene expression for simple regulatory strategies,” *Nucleic acids research*, vol. 45, no. 3, pp. 1069–1078, 2017.
- [14] E. Yurkovsky and I. Nachman, “Event timing at the single-cell level,” *Briefings in functional genomics*, p. els057, 2012.
- [15] F. Jafarpour, M. Vennettilli, and S. Iyer-Biswas, “Biological timekeeping in the presence of stochasticity,” *arXiv preprint arXiv:1703.10058*, 2017.
- [16] M. Monti and P. R. ten Wolde, “The accuracy of telling time via oscillatory signals,” *Physical biology*, vol. 13, no. 3, p. 035005, 2016.
- [17] S. Gupta, J. Varennes, H. C. Korswagen, and A. Mugler, “Temporal precision of regulated gene expression,” *PLoS computational biology*, vol. 14, no. 6, p. e1006201, 2018.
- [18] C. Dubois, S. Gupta, A. Mugler, and M.-A. Félix, “Sensitivity to perturbations of a cell migration under temporal regulation,” *Development*.
- [19] S. Gupta, H. C. K. Sean Fancher, and A. Mugler, “Temporal precision of molecular events with regulation and feedback,” *Phys. Rev. E*, vol. 101, no. 6, p. 062420, 2020.
- [20] A. J. Ridley, M. A. Schwartz, K. Burridge, R. A. Firtel, M. H. Ginsberg, G. Borisy, J. T. Parsons, and A. R. Horwitz, “Cell migration: Integrating signals from front to back,” *Science*, vol. 302, no. 5651, pp. 1704–1709, 2003.
- [21] D. Dormann and C. J. Weijer, “Chemotactic cell movement during development,” *Current Opinion in Genetics and Development*, vol. 13, no. 4, pp. 358 – 364, 2003.
- [22] M. Tessier-Lavigne and C. S. Goodman, “The molecular biology of axon guidance,” *Science*, vol. 274, no. 5290, pp. 1123–1133, 1996. [Online]. Available: <https://science.sciencemag.org/content/274/5290/1123>
- [23] S. Redner, *A guide to first-passage processes*. Cambridge University Press, 2001.
- [24] T. Chou and M. D’orsogna, “First passage problems in biology,” *First-Passage Phenomena and Their Applications*, vol. 35, p. 306, 2014.
- [25] S. Iyer-Biswas and A. Zilman, “First-passage processes in cellular biology,” *Advances in Chemical Physics, Volume 160*, pp. 261–306, 2016.
- [26] A. M. Walczak, A. Mugler, and C. H. Wiggins, “Analytic methods for modeling stochastic regulatory networks,” *Computational Modeling of Signaling Networks*, pp. 273–322, 2012.
- [27] N. Ji, T. C. Middelkoop, R. A. Mentink, M. C. Betist, S. Tonegawa, D. Mooijman, H. C. Korswagen, and A. van Oudenaarden, “Feedback control of gene expression variability in the *caenorhabditis elegans* wnt pathway,” *Cell*, vol. 155, no. 4, pp. 869–880, 2013.
- [28] A. Raj and A. van Oudenaarden, “Nature, nurture, or chance: stochastic gene expression and its consequences,” *Cell*, vol. 135, no. 2, pp. 216–226, 2008.

- [29] N. Friedman, L. Cai, and X. S. Xie, "Linking stochastic dynamics to population distribution: an analytical framework of gene expression," *Physical review letters*, vol. 97, no. 16, p. 168302, 2006.
- [30] V. Shahrezaei and P. S. Swain, "Analytical distributions for stochastic gene expression," *Proceedings of the National Academy of Sciences*, vol. 105, no. 45, pp. 17 256–17 261, 2008.
- [31] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The journal of physical chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.
- [32] G. Bel, B. Munsky, and I. Nemenman, "The simplicity of completion time distributions for common complex biochemical processes," *Physical biology*, vol. 7, no. 1, p. 016003, 2009.
- [33] H. H. McAdams and L. Shapiro, "Circuit simulation of genetic networks," *Science*, vol. 269, no. 5224, pp. 650–656, 1995.
- [34] J. Mahaffy and C. Pao, "Models of genetic control by repression with time delays and spatial effects," *Journal of mathematical biology*, vol. 20, no. 1, pp. 39–57, 1984.
- [35] K. Josić, J. M. López, W. Ott, L. Shiau, and M. R. Bennett, "Stochastic delay accelerates signaling in gene networks," *PLoS computational biology*, vol. 7, no. 11, p. e1002264, 2011.
- [36] L. Dirick and K. Nasmyth, "Positive feedback in the activation of gl cyclins in yeast," *Nature*, vol. 351, no. 6329, p. 754, 1991.
- [37] F. R. Cross and A. H. Tinkelenberg, "A potential positive feedback loop controlling *cln1* and *cln2* gene expression at the start of the yeast cell cycle," *Cell*, vol. 65, no. 5, pp. 875–883, 1991.
- [38] U. Seifert, "Stochastic thermodynamics, fluctuation theorems and molecular machines," *Reports on progress in physics*, vol. 75, no. 12, p. 126001, 2012.
- [39] P. R. ten Wolde, N. B. Becker, T. E. Ouldridge, and A. Mugler, "Fundamental limits to cellular sensing," *Journal of Statistical Physics*, vol. 162, no. 5, pp. 1395–1424, 2016.
- [40] P. Mehta, A. H. Lang, and D. J. Schwab, "Landauer in the age of synthetic biology: energy consumption and information processing in biochemical networks," *Journal of Statistical Physics*, vol. 162, no. 5, pp. 1153–1166, 2016.
- [41] J. J. Hopfield, "Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity," *Proceedings of the National Academy of Sciences*, vol. 71, no. 10, pp. 4135–4139, 1974.
- [42] O. Campas, Y. Kafri, K. Zeldovich, J. Casademunt, and J.-F. Joanny, "Collective dynamics of interacting molecular motors," *Physical review letters*, vol. 97, no. 3, p. 038101, 2006.
- [43] P. C. Bressloff and J. M. Newby, "Stochastic models of intracellular transport," *Reviews of Modern Physics*, vol. 85, no. 1, p. 135, 2013.

- [44] P. J. Mlynarczyk and S. M. Abel, “First passage of molecular motors on networks of cytoskeletal filaments,” *Physical Review E*, vol. 99, no. 2, p. 022406, 2019.
- [45] J. Whangbo and C. Kenyon, “A wnt signaling system that specifies two patterns of cell migration in *c. elegans*,” *Mol Cell*, vol. 4, no. 5, pp. 851–8, 1999.
- [46] C. Branda and M. Stern, “Mechanisms controlling sex myoblast migration in *caenorhabditis elegans* hermaphrodites,” *Dev Biol*, vol. 226, no. 1, pp. 137–51, 2000.
- [47] P. Duchek, K. Somogyi, G. Jékely, S. Beccari, and P. Rørth, “Guidance of cell migration by the drosophila pdgf/vegf receptor,” *Cell*, vol. 107, no. 1, pp. 17 – 26, 2001. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0092867401005025>
- [48] A. M. Pani and B. Goldstein, “Direct visualization of a native wnt in vivo reveals that a long-range wnt gradient forms by extracellular dispersal,” *eLife*, vol. 7, p. e38325, 2018.
- [49] A. Szabó and R. Mayor, “Mechanisms of neural crest migration,” *Annual Review of Genetics*, vol. 52, no. 1, pp. 43–63, 2018.
- [50] M. Kim, B. Bjorke, and G. S. Mastick, “Motor neuron migration and positioning mechanisms: New roles for guidance cues,” *Seminars in Cell Developmental Biology*, vol. 85, pp. 78 – 83, 2019, axon guidance: Signaling pathways old and new across a variety of developmental contexts. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1084952117304652>
- [51] S. van Helvert, C. Storm, and P. Friedl, “Mechanoreciprocity in cell migration,” *Nat Cell Biol*, vol. 20, no. 1, pp. 8–20, 2018.
- [52] R. G. Endres and N. S. Wingreen, “Accuracy of direct gradient sensing by single cells,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 41, pp. 15 749–15 754, 2008.
- [53] C. M. Meighan and J. E. Schwarzbauer, “Mechanoreciprocity in cell migration,” *Nat Cell Biol*, vol. 20, no. 1, pp. 8–20, 2018.
- [54] A. Aman and T. Piotrowski, “Cell migration during morphogenesis,” *Developmental Biology*, vol. 341, no. 1, pp. 20 – 33, 2010, special Section: Morphogenesis. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0012160609013608>
- [55] A. Paksa, J. Bandemer, B. Hoeckendorf, N. Razin, K. Tarbashevich, S. Minina, D. Meyen, A. Biundo, S. A. Leidel, N. Peyrieras, N. S. Gov, P. J. Keller, and E. Raz, “Repulsive cues combined with physical barriers and cell-cell adhesion determine progenitor cell positioning during organogenesis,” *Nature communications*, vol. 7, p. 11288, 2016. [Online]. Available: <https://europepmc.org/articles/PMC4837475>
- [56] M. Doitsidou *et al.*, “Guidance of primordial germ cell migration by the chemokine sdf-1,” *Cell*, vol. 111, no. 5, pp. 647–659, 2002.

- [57] M. R. Rohrschneider and J. Nance, "The union of somatic gonad precursors and primordial germ cells during *caenorhabditis elegans* embryogenesis," *Developmental Biology*, vol. 379, no. 2, pp. 139 – 151, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0012160613001498>
- [58] W. Halfter, S. Dong, Y.-P. Yip, M. Willem, and U. Mayer, "A critical function of the pial basement membrane in cortical histogenesis," *Journal of Neuroscience*, vol. 22, no. 14, pp. 6029–6040, 2002.
- [59] J. Sulston and H. Horvitz, "Post-embryonic cell lineages of the nematode, *caenorhabditis elegans*," *Developmental Biology*, vol. 56, no. 1, pp. 110 – 156, 1977.
- [60] M. Chalfie and J. Sulston, "Developmental genetics of the mechanosensory neurons of *caenorhabditis elegans*," *Developmental Biology*, vol. 82, no. 2, pp. 358 – 370, 1981.
- [61] Z. Altun and D. Hall, *Handbook of C. elegans anatomy*. In WormAtlas, 2020.
- [62] L. Rella, E. E. Fernandes Póvoa, and H. C. Korswagen, "The *caenorhabditis elegans* q neuroblasts: A powerful system to study cell migration at single-cell resolution in vivo," *genesis*, vol. 54, no. 4, pp. 198–211, 2016.
- [63] A. Ebbing, T. C. Middelkoop, M. C. Betist, E. Bodewes, and H. C. Korswagen, "Partially overlapping guidance pathways focus the activity of *unc-40/dcc* along the anteroposterior axis of polarizing neuroblasts," *Development*, vol. 146, no. 18, 2019.
- [64] G. Ou and R. D. Vale, "Molecular signatures of cell migration in *C. elegans* Q neuroblasts," *Journal of Cell Biology*, vol. 185, no. 1, pp. 77–85, 04 2009.
- [65] K. Hennig, I. Wang, P. Moreau, L. Valon, S. DeBeco, M. Coppey, Y. A. Miroshnikova, C. Albiges-Rizo, C. Favard, R. Voituriez, and M. Balland, "Stick-slip dynamics of cell adhesion triggers spontaneous symmetry breaking and directional migration of mesenchymal cells on one-dimensional lines," *Science Advances*, vol. 6, no. 1, 2020.
- [66] A. Leal-Egaña, G. Letort, J.-L. Martiel, A. Christ, T. Vignaud, C. Roelants, O. Filhol, and M. Théry, "The size-speed-force relationship governs migratory cell response to tumorigenic factors," *Molecular Biology of the Cell*, vol. 28, no. 12, pp. 1612–1621, 2017, pMID: 28428257.
- [67] T. Kawano, H. Zheng, D. C. Merz, Y. Kohara, K. K. Tamai, K. Nishiwaki, and J. G. Culotti, "*C. elegans* mig-6 encodes papilin isoforms that affect distinct aspects of dtc migration, and interacts genetically with mig-17 and collagen iv," *Development*, vol. 136, no. 9, pp. 1433–1442, 2009.
- [68] E. Roussos, J. Condeelis, and A. Patsialou, "Chemotaxis in cancer," *Nat Rev Cancer*, vol. 11, no. 8, pp. 573–87, 2011.
- [69] P. N. Devreotes and S. H. Zigmond, "Chemotaxis in eukaryotic cells: A focus on leukocytes and dictyostelium," *Annual Review of Cell Biology*, vol. 4, no. 1, pp. 649–686, 1988.

- [70] T. Lämmermann, P. Afonso, B. Angermann *et al.*, “Neutrophil swarms require Itb4 and integrins at sites of cell death in vivo,” *Nature*, vol. 498, pp. 371–375, 2013.
- [71] P. Niethammer, C. Grabher, A. Look *et al.*, “A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish,” *Nature*, vol. 459, pp. 996–999, 2009.
- [72] E. Albrecht and H. R. Petty, “Cellular memory: Neutrophil orientation reverses during temporally decreasing chemoattractant concentrations,” *Proceedings of the National Academy of Sciences*, vol. 95, no. 9, pp. 5039–5044, 1998.
- [73] D. Wessels, J. Murray, and D. Soll, “Behavior of dictyostelium amoebae is regulated primarily by the temporal dynamic of the natural camp wave,” *Cell motility and the cytoskeleton*, vol. 23, no. 2, p. 145–156, 1992.
- [74] J. Geiger, D. Wessels, and D. Soll, “Human polymorphonuclear leukocytes respond to waves of chemoattractant, like dictyostelium,” *Cell Motil Cytoskeleton*, vol. 56, no. 1, pp. 27–44, 2003.
- [75] M. Skoge, H. Yue, M. Erickstad, A. Bae, H. Levine, A. Groisman, W. F. Loomis, and W.-J. Rappel, “Cellular memory in eukaryotic chemotaxis,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 40, pp. 14 448–14 453, 2014.
- [76] X. Cheng, L. Merchan, M. Tchernookov, and I. Nemenman, “A large number of receptors may reduce cellular response time variation,” *Physical Biology*, vol. 10, no. 3, p. 035008, jun 2013. [Online]. Available: <https://doi.org/10.1088/1478-3975/10/3/035008>
- [77] S. Redner, *A Guide to First-Passage Processes*. Cambridge University Press, 2001.
- [78] H. Breuer, W. Huber, and F. Petruccione, “Fluctuation effects on wave propagation in a reaction-diffusion process,” *Physica D: Nonlinear Phenomena*, vol. 73, no. 3, pp. 259 – 273, 1994. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0167278994901619>