

Spatial partitioning improves the reliability of biochemical signaling

Andrew Mugler, Filipe Tostevin, and Pieter Rein ten Wolde¹

Foundation for Fundamental Research on Matter Institute for Atomic and Molecular Physics, 1098 XG, Amsterdam, The Netherlands

Edited by Curtis G. Callan, Jr., Princeton University, Princeton, NJ, and approved February 22, 2013 (received for review October 19, 2012)

Spatial heterogeneity is a hallmark of living systems, even at the molecular scale in individual cells. A key example is the partitioning of membrane-bound proteins via lipid domain formation or cytoskeleton-induced corraling. However, the impact of this spatial heterogeneity on biochemical signaling processes is poorly understood. Here, we demonstrate that partitioning improves the reliability of biochemical signaling. We exactly solve a stochastic model describing a ubiquitous motif in membrane signaling. The solution reveals that partitioning improves signaling reliability via two effects: it moderates the nonlinearity of the switching response, and it reduces noise in the response by suppressing correlations between molecules. An optimal partition size arises from a trade-off between minimizing the number of proteins per partition to improve signaling reliability and ensuring sufficient proteins per partition to maintain signal propagation. The predicted optimal partition size agrees quantitatively with experimentally observed systems. These results persist in spatial simulations with explicit diffusion barriers. Our findings suggest that molecular partitioning is not merely a consequence of the complexity of cellular substructures, but also plays an important functional role in cell signaling.

information transmission | reversible modification | stochastic modeling

The cell membrane is a nexus of information processing. Once regarded as a simple barrier between a cell and its surroundings, it is now clear that the membrane is a hot spot of molecular activity, where signals are integrated and modulated even before being relayed to the inside of the cell (1). Moreover, the membrane itself is structurally complex. Regions enriched in glycosphingolipids, cholesterol, and other membrane components, often called lipid rafts, transiently assemble and float within the surrounding bilayer (2), providing platforms for molecular interaction (3). Additionally, interaction of the membrane with the underlying actin cytoskeleton forms compartments in which molecules are transiently trapped (4, 5). These membrane subdomains create a highly heterogeneous environment in which molecules are far from well mixed, and it is currently unclear what effect this heterogeneity has on cell signaling.

Membrane subdomains are thought to play a dominant role in the observed aggregation of signaling molecules into clusters (6). Interestingly, these clusters have a characteristic size of only a few molecules. For example, the GPI-anchored receptor CD59 is observed to form clusters of three to nine molecules upon interaction with the cytoskeleton and lipid rafts (7, 8). Similarly, the well-studied membrane-bound GTPase Ras forms clusters of six to eight molecules, which also depend on interactions with the cytoskeleton and rafts (9, 10). Despite the important findings that aggregation of proteins induced by subdomains can affect reaction kinetics (11), enhance oligomerization (1), modulate downstream responses (12, 13), and enhance signal fidelity (13, 14), the origin of this characteristic size remains unknown. Although it is quite possible that these domains owe their size to a thermodynamic or structural origin, we here address the question of whether this size can be optimized for signaling performance. We find that the partitioning imposed by subdomains gives rise to a trade-off in cell signaling, from which an optimal size of a few molecules emerges

naturally, suggesting that reliable signaling is intimately tied to the spatial structure of the membrane.

We study via stochastic analysis and spatial simulation a model that is directly motivated by both CD59 and Ras signaling at the membrane. Stimulated CD59 receptors induce the switching of several Src family kinases from an unphosphorylated to a phosphorylated state (7, 8). Similarly, stimulated EGF receptors induce the switching of Ras proteins from an inactive GDP-loaded state to an active GTP-loaded state (13). We therefore study the simple and ubiquitous motif of coupled switching reactions, in which the activation of one species (the receptor) triggers the activation of a second species (the downstream effector).

We exactly solve this stochastic model of coupled switching reactions, and we use the solution to compare signaling reliability in a spatially partitioned system to that in a well-mixed system. We demonstrate that partitioning can improve signaling performance by generating a more graded input–output relationship and by reducing the noise in the signaling response. This latter effect comes about because partitioning reduces the correlations between the states of the different output molecules. However, the stochastic exchange of proteins between partitions can generate configurations that isolate molecules and exclude them from the signaling process, thereby reducing the dynamic range of the response and increasing the output noise. The trade-off between these two effects results in an optimal partition size that agrees well with cluster sizes of signaling proteins that are observed experimentally (7–10), suggesting that cluster sizes are tuned so as to maximize information transmission.

Results

We model two coupled molecular species at the membrane, as depicted in Fig. 1A. A membrane-bound receptor (e.g., CD59 or EGF receptor) is activated via ligand stimulation, and the active receptor in turn activates a membrane-bound effector (e.g., a Src family kinase or Ras). A reaction scheme representing these processes is shown in Fig. 1B and consists of two protein species: the receptor \mathcal{X} and the downstream effector \mathcal{Y} . The switching of \mathcal{X} molecules from the X to the X^* state is driven by an external signal of strength α . Active X^* molecules act on inactive \mathcal{Y} molecules and promote switching to the Y^* state. Deactivation of both active protein species occurs spontaneously and independently.

We will be concerned with how the network response, the number of active Y^* molecules as a function of the input signal α , is affected by the spatial structure of the system. In particular, we ask how partitioning of the reaction system into noninteracting subdomains affects the reliability of signal transmission, which is

Author contributions: A.M., F.T., and P.R.t.W. designed research; A.M. performed research; A.M. contributed new reagents/analytic tools; A.M., F.T., and P.R.t.W. analyzed data; and A.M., F.T., and P.R.t.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: tenwolde@amolf.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1218301110/-DCSupplemental.

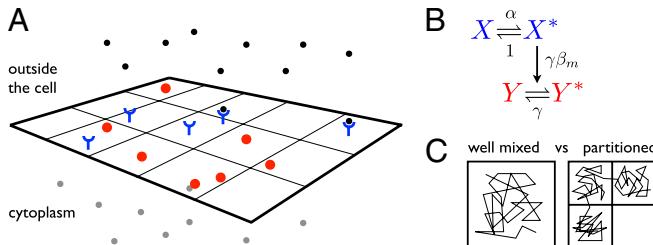


Fig. 1. Schematic depiction of the model system. (A) We consider a model representative of signal detection by receptors and signal transmission at the cell membrane. (B) The model consists of two molecular species (\mathcal{X} and \mathcal{Y}), which can each exist in active (X^* , Y^*) or inactive (X , Y) states. Molecules in the X state are activated by the external signal of strength α , and active X^* molecules subsequently activate Y molecules. (C) We consider these reactions taking place in a single domain with all components well mixed, or in a domain consisting of smaller compartments, which are each individually well mixed but between which no interaction is possible. The total system volumes in the two scenarios are equal and assumed to scale with the number of \mathcal{X} molecules.

determined by two principal factors: the input–output response and the output noise; together these properties determine to what extent different input signals can be reliably resolved from the network response. We focus on two system configurations, shown in Fig. 1C. In the first case, we assume that all molecules are present in a single well-mixed reaction compartment. In the second case, we consider a system partitioned into π compartments between which no interactions are permitted; here, we take the output of the system to be the total number of active Y^* molecules in all compartments. This choice of output corresponds to a readout of the Y^* signal by, e.g., a cytosolic component whose diffusion is much faster than the diffusion and signaling of \mathcal{X} and \mathcal{Y} on the membrane. In the partitioned system, we will for simplicity first assume that the molecules are uniformly and statically distributed among compartments. However, recognizing that this scenario will not generally be realized inside cells, we will later relax this assumption and consider exchange of molecules among partitions.

We model the dynamics of the well-mixed system, as well as each compartment within the partitioned system, using a stochastic equation of the same form. We denote the total numbers of \mathcal{X} and \mathcal{Y} molecules by M and N , respectively, and the numbers of active X^* and Y^* molecules by m and n , respectively. To parameterize the system, we scale units of time by the deactivation rate of X^* , such that the effective deactivation rate is 1. Then α denotes the rescaled activation rate of X ; γ is the rate of deactivation of Y^* relative to that of X^* ; and $\gamma\beta_m$ is the activation rate of a given Y molecule for a particular concentration of X^* molecules. The parameter α incorporates the effective strength of the input signal and determines the mean X^* activity via the occupancy $q \equiv \langle m \rangle / M = \alpha / (\alpha + 1)$. The precise m -dependence of the coupling function β_m will depend on the exact nature of the interactions between X^* and Y molecules. We take $\beta_m \propto m/v$, with v as the volume of the compartment in which the reactions are taking place. However, our conclusions are unaffected if we instead take a Michaelis–Menten form $\beta_m \propto m/(m + vK)$ (SI Appendix, Fig. S1). The total system volume V is assumed to scale with the total number of \mathcal{X} molecules, such that M/V is constant. The coupling function in partition $i \in \{1, \dots, \pi\}$ is then determined by m_i , the number of X^* molecules in partition i , according to $\beta_m^{(i)} \propto m_i / (V/\pi) = \beta \pi m_i / M$ for constant β .

The probability of having m proteins in the X^* state and n proteins in the Y^* state evolves according to the chemical master equation (CME),

$$\dot{p}_{mn} = -[\mathcal{L}_m(\alpha, M) + \gamma\mathcal{L}_n(\beta_m, N)]p_{mn}, \quad [1]$$

subject to suitable boundary conditions. The nature of the particular set of reactions in our model (Fig. 1B) means that the operators \mathcal{L}_m and \mathcal{L}_n have the same form,

$$\mathcal{L}_m(\alpha, M) = \alpha[1 - \mathbb{E}_m^{-1}](M - m) + [1 - \mathbb{E}_m^{+1}]m, \quad [2]$$

where $\mathbb{E}_m^i f(m) = f(m + i)$ defines the step operator. Despite the appearance of terms containing the product mn in the operator $\mathcal{L}_n(\beta_m, N)$, which make the direct calculation of moments of p_{mn} from the CME impossible, an exact solution to Eq. 1 can be found for arbitrary β_m using the method of spectral expansion (15, 16) as described in SI Appendix.

Partitioning Leads to a More Graded Response. We begin by analyzing the behavior of a minimal system with $M = N = 2$. In the well-mixed system, all molecules are contained within $\pi = 1$ domain of volume V . In the partitioned system, $\pi = 2$ subdomains with volume $V/2$ each contain one \mathcal{X} and one \mathcal{Y} molecule.

We first focus on the mean response $\langle n \rangle$. In the limits of small or large α , the mean response is the same in both the partitioned and mixed systems, $\langle n \rangle / N \rightarrow 0$ and $\langle n \rangle / N \rightarrow \beta / (\beta + 1)$, respectively. However, at all intermediate values of α , the mean response of the well-mixed system is larger than that of the partitioned system; equivalently, the partitioned system exhibits a more graded response than the well-mixed system to changes in the input signal (Fig. 2A, thick solid and dashed curves). The more graded response is due to higher fluctuations in X^* activity. When $\alpha \rightarrow 0$ or $\alpha \rightarrow \infty$, all \mathcal{X} molecules are inactive or active, respectively; however, at intermediate values of α , the number of active X^* molecules fluctuates. Partitioning reduces the number of \mathcal{X} molecules per reaction compartment, increasing the relative size of these fluctuations according to $\sigma_m^2 / (M/\pi)^2 = \pi q(1-q)/M$. These fluctuations are passed through the concave dependence of n on m , resulting in a smaller mean [via Jensen's inequality (17)], and therefore a more linear response curve (SI Appendix, Fig. S2A).

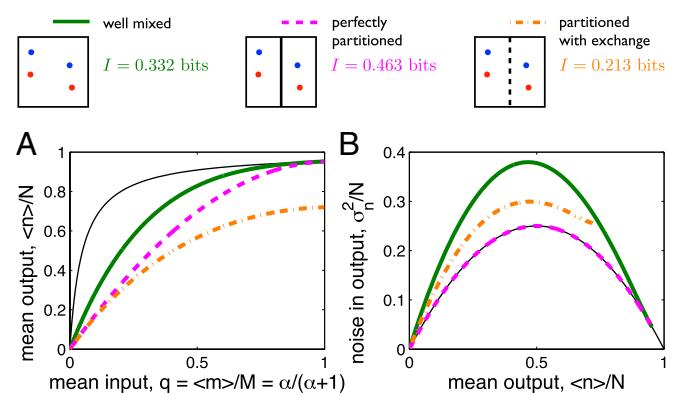


Fig. 2. Spatial partitioning improves signaling performance. (A) The mean response $\langle n \rangle / N$ as a function of the mean X^* activity $q = \langle m \rangle / M = \alpha / (\alpha + 1)$, and (B) the output variance σ_n^2 as a function of the mean response, plotted for a well-mixed system with $M = N = 2$ (thick solid) and a partitioned system of $\pi = 2$ compartments, each containing one \mathcal{X} and one \mathcal{Y} molecule (thick dashed). Partitioning linearizes the output response and reduces noise across the full range of responses, leading to a higher transmitted information. The thin solid curves show the mean field response $\langle n \rangle / N = \beta q / (\beta q + 1)$ in A and the binomial noise limit (3) in B. Allowing exchange of molecules between compartments (thick dot-dashed) compresses the output response and increases the noise compared with the perfectly partitioned system, dramatically reducing information transmission. Here, $\beta = 20$ and $\gamma = 1$.

A more graded input–output relationship can potentially enhance signaling by expanding the range of input signals which the network is able to transmit without saturating the response. However, to determine whether this larger input range can be resolved in the network it is crucial to examine how the noise in the response is affected.

Partitioning Reduces Noise. Fig. 2B shows the variance of the output σ_n^2 as a function of the mean response $\langle n \rangle$ for the system with $M = N = 2$, as the input signal strength α is varied. We see that the output noise is reduced in the partitioned system relative to the well-mixed system across the full range of response levels. The noise reduction is surprising: one might expect that the increased fluctuations in X^* activity that come with partitioning would propagate to fluctuations in Y^* activity. Indeed, this is the case: in a single compartment, as the number of X molecules is reduced, the noise in the output increases (SI Appendix, Fig. S2B). However, this effect is overcome by a second effect: partitioning reduces correlations among output molecules.

To see the effect of partitioning on correlations, we consider the expressions for the variance. In the partitioned case, because the two Y molecules switch independently, the variance of n is simply that of a pair of independent binomial switches with activation probability $\langle n \rangle/N$,

$$\frac{\sigma_n^2}{N} = \frac{\langle n \rangle}{N} \left(1 - \frac{\langle n \rangle}{N} \right). \quad [3]$$

In contrast, in the well-mixed case, the two Y molecules are not independent. Because both are driven by the same set of X molecules, fluctuations in β_m lead to correlations between the states of the two Y molecules as their switching becomes more synchronized (Fig. 3). This in turn leads to an increase in the variance, which can be written as follows:

$$\frac{\sigma_n^2}{N} = \frac{\langle n \rangle}{N} \left(1 - \frac{\langle n \rangle}{N} \right) + \frac{\Delta}{N}, \quad [4]$$

where Δ is a correction term accounting for the correlation between Y molecules, which is due to “extrinsic” fluctuations in the input $m(t)$. The functional form of Δ for any M and N follows directly from the spectral solution of the CME (SI Appendix, Eq. 68); for $M = N = 2$, one finds by inspection that Δ is manifestly positive, meaning that correlations increase the noise across all values of the mean. Importantly, this effect is independent of the parameters of the switching reactions.

The reduction of noise upon partitioning extends beyond the case of one Y molecule per partition. Indeed, the same

phenomenon is observed if we consider larger molecule numbers $M > \pi$ and $N > \pi$, and compare the well-mixed system to a system with uniform partitioning of the X and Y molecules into the π compartments. In the well-mixed case, all Y molecules respond to the same signal $m(t)$ and hence are correlated with all other Y molecules in the system. By contrast, in the partitioned case, the $N/\pi > 1$ Y molecules within each partition are correlated, and indeed because the fluctuations in $m_i(t)$ will be larger than $m(t)$ for the mixed system, such correlations will be stronger; yet the Y molecules in different partitions are uncorrelated. This latter effect is sufficient to overcome the increase in correlations within each partition, such that the total noise is reduced.

To see the noise reduction explicitly, we again consider the expression for the variance. Because the dynamics of different partitions is independent, assuming that both M and N are multiples of π , the variance can be written as follows:

$$\frac{\sigma_n^2}{N} = \frac{\langle n \rangle}{N} \left(1 - \frac{\langle n \rangle}{N} \right) + \pi \frac{\Delta(\tilde{M}, \tilde{N})}{N}, \quad [5]$$

where $\tilde{M} \equiv M/\pi$ and $\tilde{N} \equiv N/\pi$ are the numbers of X and Y molecules per compartment, respectively. Here, as before, $\Delta(\tilde{M}, \tilde{N})$ represents the additional fluctuations due to correlations between the states of Y molecules within each compartment. The N -dependence of $\Delta(\tilde{M}, \tilde{N})$, which reflects the number of correlated pairs of Y molecules, can be straightforwardly factored out as $\Delta(\tilde{M}, \tilde{N}) = \tilde{N}(\tilde{N} - 1)\Delta(\tilde{M})$, where $\Delta(\tilde{M})$ describes how strongly correlated are Y molecules within each compartment. The exact form for $\Delta(\tilde{M})$, although straightforward to calculate for a given \tilde{M} , is difficult to generalize for all \tilde{M} ; nonetheless, inspection of numerical and analytic results for specific combinations of \tilde{M} and \tilde{N} reveals in all cases that increasing π leads to an overall reduction in σ_n^2 . Additionally, if the switching of Y molecules is much slower than that of X molecules, $\gamma \ll 1$, then $\Delta(\tilde{M})$ takes the following form:

$$\Delta(\tilde{M}) \approx \frac{\alpha\beta^2\gamma}{\tilde{M}(1 + \alpha + \alpha\beta)^3}. \quad [6]$$

Inserting this expression into Eq. 5 with $\tilde{M} = M/\pi$ and $\tilde{N} = N/\pi$, one can straightforwardly see that the variance is a decreasing function of π for $\pi < N$, indicating that the noise is reduced as the system is more finely partitioned.

Partitioning Increases Information Transmission. We have seen that partitioning has two beneficial effects on signal propagation: the

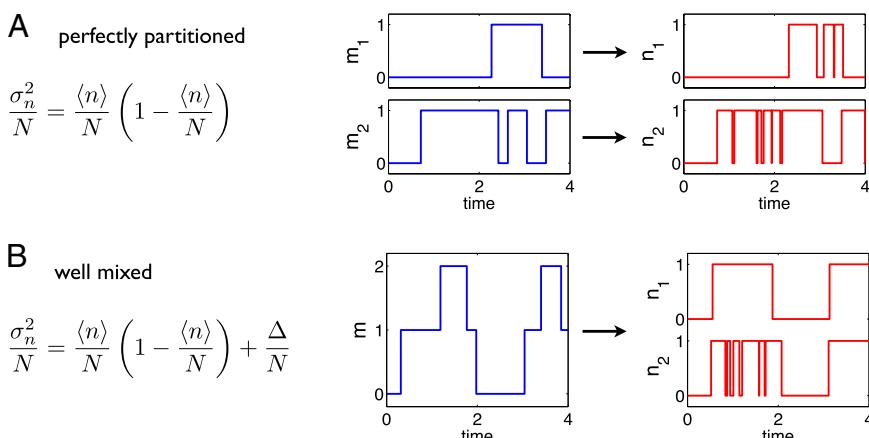


Fig. 3. Partitioning reduces correlations between output modules. (A) In the partitioned system, each Y molecule receives an independent signal $m_i(t)$. The variance is simply that of independent two-state switches. (B) In the well-mixed system, each Y molecule reacts to the same $m(t)$, which leads to correlations between the states of different Y molecules and an increase in the variance σ_n^2 . Sample trajectories are generated using parameters as in Fig. 2, with $\alpha = 1$.

input–output response becomes more graded, and the output noise at a given response level is reduced. Together, these effects mean that a larger number of distinct input signals can be encoded in the network response. To quantify the ability of the network to transmit signals, we calculate the mutual information $I[\alpha, n]$ (18) between the input and the number of active Y^* molecules, as described in *SI Appendix, Methods*. The mutual information quantifies the number of resolvable output signals given a particular distribution of input signals. We find that, indeed, in the case of $M = N = 2$ (Fig. 2), $I[\alpha, n]$ is significantly larger for the partitioned system ($I = 0.463$ bits) than for the well-mixed system ($I = 0.332$ bits), confirming that signal transmission is dramatically improved by partitioning.

Exchange Between Partitions Compromises Signaling Reliability. Thus far, we have considered only the perfectly uniform and stationary partitioning of molecules. In reality, physical transport processes such as diffusion will also give rise to a variety of configurations with different numbers of proteins in each compartment, as depicted in Fig. 4. Each of these configurations will have different properties for the transmission of the signal from α to n . It is therefore important to consider whether the benefits of partitioning described above persist once these additional configurations are taken into account.

Single-molecule tracking experiments have revealed that the timescale of diffusive mixing within a compartment ($\sim 100 \mu\text{s}$) is two orders of magnitude faster than the timescale of molecular exchange between compartments ($\sim 10 \text{ ms}$) (19). This observation allows us to treat each configuration as static on the timescale of mixing, and then compute the total response by averaging over all configurations. Inherent in this treatment is the assumption that the timescale of signaling is also faster than that of exchange between compartments. We later relax this assumption using spatially resolved simulations and nonetheless find similar results.

The total response is computed by first enumerating the possible configurations of M X molecules and N Y molecules distributed among π partitions. For each such configuration c , we then solve for the distribution $p_{mn|c}$ and combine these distributions, weighted by the probability p_c of each configuration occurring if molecules are randomly assigned to different partitions with uniform and independent probability, to give the overall distribution $p_{mn} = \sum_c p_{mn|c} p_c$.

Fig. 2B (dot-dashed curve) shows that the exchange of molecules between compartments increases the noise relative to the perfectly partitioned system considered previously when $M = N = 2$. This is because many of the alternative configurations generated by exchange lead to significant correlations between the states of the different Y molecules. Nevertheless, we see that the noise remains lower than that of the well-mixed system, because of the existence of some configurations in which the Y molecules are independent. However, the appearance of alternate configurations also affects the mean response (Fig. 2A); in particular, the appearance of configurations in which X and Y molecules do not

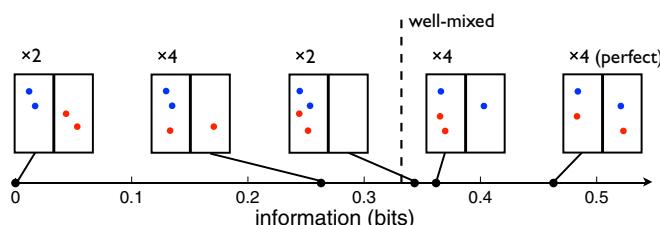


Fig. 4. Exchange between partitions leads to different configurations of the system with a range of signaling performance. Multiplicities listed above each configuration are due to symmetry. Parameters are as in Fig. 2.

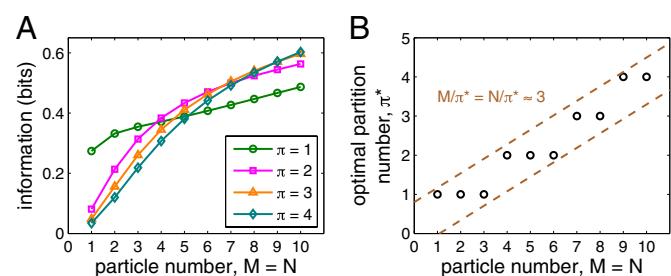


Fig. 5. An optimal partition size. (A) For $M = N > 3$ molecules, a system with $\pi > 1$ partitions achieves higher information transmission than a well-mixed system ($\pi = 1$). (B) As $M = N$ is increased, the optimal partition number also increases such that the optimal number of proteins per partition $M/\pi^* = N/\pi^* \sim 3$ is roughly constant. Parameters are as in Fig. 2.

occupy the same partitions, and hence no signal can be propagated, means that the maximal output level is reduced. Given this simultaneous change in both the input–output function and the noise, it is not immediately clear whether signaling reliability is improved relative to the well-mixed system. Computing the mutual information, we see that the information transmitted by the system with exchange ($I = 0.213$ bits) is significantly lower than that for the well-mixed system ($I = 0.332$ bits), showing that the reduction of the output range compromises signal transmission to an extent that cannot be overcome by the corresponding reduction in noise.

The decrease in information transmission upon incorporating molecule exchange in the system with $M = N = 2$ is the result of the appearance of suboptimal protein configurations, for which signal propagation is compromised (or even impossible). However, the number and performance of such configurations will in general depend on the relative values of M , N , and π (which need not equal M or N). Although molecule exchange may make partitioning unfavorable in the extreme case of $M = N = 2$, for systems with higher protein numbers it can be beneficial to partition the system into $\pi > 1$ compartments, as we will see next.

Optimal Partition Size. To study the performance of systems with higher protein numbers and different partition sizes, we compare the information transmission, including molecule exchange, for different partition numbers π as the number of proteins in the system is varied while holding $M = N$. Fig. 5A shows that for $M = N > 3$ protein copies, systems with $\pi > 1$ partition do indeed outperform the well-mixed system. Furthermore, as $M = N$ is increased, the optimal partition number also increases such that the optimal number of proteins per partition $M/\pi^* = N/\pi^* \sim 3$ is roughly constant (Fig. 5B). This result is robust to variations in β and γ : changing each over several orders of magnitude results in optimal partition sizes in the range $M/\pi^* = N/\pi^* \sim 1-10$ (*SI Appendix, Fig. S3 A and B*). The assumption of $M = N$ is also not crucial for this result. In fact, we find that the value of M/π^* has only a weak dependence on N (*SI Appendix, Fig. S4*).

The optimal partition size arises from a trade-off between the reliability and efficiency of signaling. Increasing the number of partitions decreases the typical number of proteins per partition, which leads to the beneficial effects of a more graded response and reduced noise, increasing signaling reliability. However, due to molecule exchange, reducing the number of molecules per partition also increases the probability that any partition contains proteins of only one species that are therefore excluded from the signaling process, which leads to a reduced maximal response, reducing signaling efficiency.

The optimal size revealed by our study of $\sim 1-10$ molecules per species per partition shows good quantitative agreement with the observed aggregation of CD59 receptors [3–9 molecules (7, 8)]

and Ras proteins [6–8 molecules (9, 10)], which each signal via the present motif and are known to interact with rafts and the cytoskeleton. It is of further interest that a recent experiment in which T-cell receptors were artificially partitioned on supported membranes found that the minimum number of agonist-bound receptors per partition necessary for downstream signaling is approximately four (20).

Explicitly Spatial Model. Last, we confirm that the effects observed in these minimal model systems, where the contents of each compartment are well mixed and exchange can occur between any pair of compartments, persist in a more realistic model in which the diffusion of molecules in space is included explicitly. We simulate the diffusion and reaction of X and Y molecules on a 2D lattice, as described in *SI Appendix, Methods*. The system is partitioned into a number of subdomains by the introduction of diffusion barriers, which are crossed with a reduced probability p_{hop} relative to regular diffusion steps on the lattice. Results of such simulations are shown in Fig. 6.

Fig. 6A and B reveal that as the strength of the diffusion barriers is increased, the mean response becomes more graded, and the variance of Y^* activity is reduced, analogous to the two effects observed in the minimal model system (Fig. 2). When $p_{\text{hop}} = 0$, one molecule of each species is permanently confined to a compartment, producing the graded response predicted for the perfectly partitioned system (Fig. 6A) and the associated minimal, binomial noise (Fig. 6B). Low but finite p_{hop} allows exchange of molecules between neighboring compartments but preserves a separation of timescales between intracompartimental and intercompartmental mixing. This results in a graded mean response whose maximal level is reduced (Fig. 6A) and reduced noise (Fig. 6B), precisely the features observed in the minimal model of partitioning with exchange (Fig. 2). When $p_{\text{hop}} = 1$, there are no barriers, and the system approaches the well-mixed limit (CME). Interestingly, however, the response remains more graded and the noise remains lower than the predictions of the CME due to the finite speed of diffusion (Fig. 6A and B), with agreement only reached when the ratio of diffusion to reaction propensities is much greater than 1. This observation reveals that finite diffusion imposes an effective

partitioning even when no actual partitions exist: molecules remain correlated with reaction partners within a typical distance set by diffusion, but uncorrelated with partners beyond this distance. As such, in the context of coupled reversible modification, we find that slower diffusion can linearize the response and reduce the noise, thereby improving information transmission. Interestingly, this result is in marked contrast to the case of boundary establishment in embryonic development, where faster diffusion reduces noise within each nucleus by washing out bursts of gene expression in the input signal (21). Although in the present system faster diffusion will similarly reduce any super-Poissonian component of the noise within each partition individually, this averaging does not reduce the noise in the total output across all partitions. In fact, the latter noise is enhanced with faster diffusion by virtue of increased correlations between partitions.

While slower diffusion can lead to effective partitioning in a system without explicit diffusion barriers, the noise reduction in such a system is much smaller than that in a system with actual partitions. Fig. 6B shows that finite diffusion reduces the maximal noise by $(1.25 - 1)/1.25 = 20\%$, whereas strong partitioning ($p_{\text{hop}} = 0.001$) reduces the maximal noise by $(1.25 - 0.4)/1.25 \sim 70\%$. Therefore, partitioning, which introduces not only a slower effective “hop” diffusion but also a separation of timescales between intracompartimental and intercompartmental mixing, is far more effective at conveying an information enhancement.

Fig. 6C confirms that the transmitted information varies non-monotonically with the number of barriers in a fixed area, indicating that an optimal partition size also appears in systems where space is modeled explicitly. Like in the minimal model, this optimum persists with changes in β and γ , spanning the range of ~ 1 – 10 molecules per partition (*SI Appendix, Fig. S3 C and D*). Fig. 6C also provides a measure of the scale of information transmitted by this motif. In absolute terms, the optimal information (1.35 bits) is consistent with values recently measured for signaling via the TNF–NF- κ B pathway (~ 0.5 – 1.5 bits) (22) and for patterning in the *Drosophila* embryo (1.5 ± 0.15 bits) (23). In relative terms, we see that partitioning increases information over the unpartitioned system by $(1.35 - 1.04)/1.04 \sim 30\%$ (Fig.

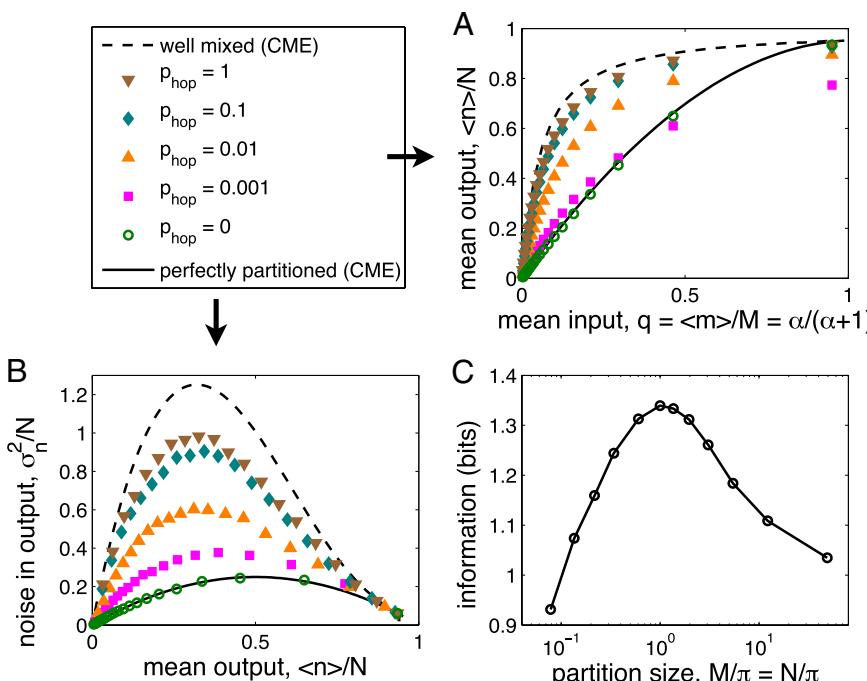


Fig. 6. The effects of partitioning persist in simulations with explicit diffusion. As the probability of crossing a diffusion barrier p_{hop} is decreased, (A) the mean response becomes more graded, and (B) the output noise decreases. (C) The information transmission has a maximum as a function of the partition size. Here $M = N = 49$, $\beta = 20$, $\gamma = 1$, the system is $\lambda = 70$ lattice spacings squared, and the ratio of diffusion to reaction propensities is $p_D/p_r = 1$. In A and B, $\pi = 49$; in C, $p_{\text{hop}} = 0.001$, and the partition size is varied by taking $\sqrt{\pi}$ from 25 to 1.

6C) and decreases the maximal noise by $(1 - 0.4)/1 = 60\%$ (Fig. 6B). Thus, in both absolute and relative terms, we see that partitioning plays a critical role in producing informative and reliable membrane signaling.

As a final test, we use simulation to confirm that the effects of partitioning persist in the presence of features that are more realistic for signaling systems at the membrane, including extrinsic noise in the input (*SI Appendix*, Fig. S5) and receptor dimerization (*SI Appendix*, Fig. S6). The fact that the effects of partitioning, including the emergence of an optimal partition size, are robust to these details further underscores the generality of our findings.

Discussion

We have seen that the partitioning of a biochemical signaling system into a number of noninteracting subsystems improves the reliability of signaling via two effects. First, the nonlinear response of the network means that a reduction in the number of input molecules translates into a more graded input–output response. Second, partitioning significantly reduces the noise in the response by eliminating correlations between the states of the different output molecules, an effect that, remarkably, overcomes the increase in noise associated with fewer input molecules in each subsystem. However, we have seen that the introduction of diffusion or exchange of molecules between partitions enhances the variance and reduces the range of the response, thereby reducing signaling performance. This result is due to the presence of configurations in which the two species are isolated from one another, compromising or even arresting signal transmission in certain partitions. The interplay between these two effects leads to a partition size that optimizes information transmission, corresponding to a few molecules per partition on average, in quantitative agreement with experiments. These effects are generic, and hence the emergence of an optimal partition size is robust to the specific parameters of the model. Notably, the underlying mechanism revealed here, namely the removal of correlations, differs fundamentally from that based on cooperativity in protein activation, which has been argued to underlie optimal cluster size in sensory systems (24, 25).

- Grecco HE, Schmick M, Bastiaens PIH (2011) Signaling from the living plasma membrane. *Cell* 144(6):897–909.
- Eggeling C, et al. (2009) Direct observation of the nanoscale dynamics of membrane lipids in a living cell. *Nature* 457(7233):1159–1162.
- Lingwood D, Simons K (2010) Lipid rafts as a membrane-organizing principle. *Science* 327(5961):46–50.
- Kusumi A, Sako Y (1996) Cell surface organization by the membrane skeleton. *Curr Opin Cell Biol* 8(4):566–574.
- Kusumi A, Shirai YM, Koyama-Honda I, Suzuki KGN, Fujiwara TK (2010) Hierarchical organization of the plasma membrane: Investigations by single-molecule tracking vs. fluorescence correlation spectroscopy. *FEBS Lett* 584(9):1814–1823.
- Kholodenko BN, Hancock JF, Kolch W (2010) Signalling ballet in space and time. *Nat Rev Mol Cell Biol* 11(6):414–426.
- Suzuki KGN, et al. (2007) GPI-anchored receptor clusters transiently recruit Lyn and G alpha for temporary cluster immobilization and Lyn activation: Single-molecule tracking study 1. *J Cell Biol* 177(4):717–730.
- Suzuki KGN, Fujiwara TK, Edidin M, Kusumi A (2007) Dynamic recruitment of phospholipase C gamma at transiently immobilized GPI-anchored receptor clusters induces IP3-Ca²⁺ signaling: Single-molecule tracking study 2. *J Cell Biol* 177(4):731–742.
- Prior IA, Muncke C, Parton RG, Hancock JF (2003) Direct visualization of Ras proteins in spatially distinct cell surface microdomains. *J Cell Biol* 160(2):165–170.
- Plowman SJ, Muncke C, Parton RG, Hancock JF (2005) H-ras, K-ras, and inner plasma membrane raft proteins operate in nanoclusters with differential dependence on the actin cytoskeleton. *Proc Natl Acad Sci USA* 102(43):15500–15505.
- Kalay Z, Fujiwara TK, Kusumi A (2012) Confining domains lead to reaction bursts: Reaction kinetics in the plasma membrane. *PLoS One* 7(3):e32948.
- Mugler A, Bailey AG, Takahashi K, ten Wolde PR (2012) Membrane clustering and the role of rebinding in biochemical signaling. *Biophys J* 102(5):1069–1078.
- Tian T, et al. (2007) Plasma membrane nanoswitches generate high-fidelity Ras signal transduction. *Nat Cell Biol* 9(8):905–914.
- Gurry T, Kahramanogullari O, Endres RG (2009) Biophysical mechanism for ras-nanocluster formation and signaling in plasma membrane. *PLoS One* 4(7):e6148.
- Walczak AM, Mugler A, Wiggins CH (2009) A stochastic spectral analysis of transcriptional regulatory cascades. *Proc Natl Acad Sci USA* 106(16):6529–6534.
- Mugler A, Walczak AM, Wiggins CH (2009) Spectral solutions to stochastic models of gene expression with bursts and regulation. *Phys Rev E Stat Nonlin Soft Matter Phys* 80(4 Pt 1):041921.
- Jensen JLWV (1906) Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math* 30(1):175–193.
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27: 379–423, 623–656.
- Murase K, et al. (2004) Ultrafine membrane compartments for molecular diffusion as revealed by single-molecule techniques. *Biophys J* 86(6):4075–4093.
- Manz BN, Jackson BL, Petit RS, Dustin ML, Groves J (2011) T-cell triggering thresholds are modulated by the number of antigen within individual T-cell receptor clusters. *Proc Natl Acad Sci USA* 108(22):9089–9094.
- Erdmann T, Howard M, ten Wolde PR (2009) Role of spatial averaging in the precision of gene expression patterns. *Phys Rev Lett* 103(25):258101.
- Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A (2011) Information transduction capacity of noisy biochemical signaling networks. *Science* 334(6054):354–358.
- Tkacik G, Callan CG, Jr., Bialek W (2008) Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci USA* 105(34):12265–12270.
- Aquino G, Clausznitzer D, Tollis S, Endres RG (2011) Optimal receptor-cluster size determined by intrinsic and extrinsic noise. *Phys Rev E Stat Nonlin Soft Matter Phys* 83(2 Pt 1):021914.
- Skoge M, Meir Y, Wingreen NS (2011) Dynamics of cooperativity in chemical sensing among cell-surface receptors. *Phys Rev Lett* 107(17):178101.
- Takahashi K, Tanase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proc Natl Acad Sci USA* 107(6): 2473–2478.
- van Zon JS, Lubensky DK, Altena PRH, ten Wolde PR (2007) An allosteric model of circadian KaiC phosphorylation. *Proc Natl Acad Sci USA* 104(18):7420–7425.
- Miller P, Zhabotinsky AM, Lisman JE, Wang XJ (2005) The stability of a stochastic CaMKII switch: Dependence on the number of enzyme molecules and protein turnover. *PLoS Biol* 3(4):e107.
- Nash P, et al. (2001) Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* 414(6863):514–521.
- Zeke A, Lukács M, Lim WA, Reményi A (2009) Scaffolds: Interaction platforms for cellular signalling circuits. *Trends Cell Biol* 19(8):364–374.

Reversible modification reactions are ubiquitous in cell signaling, and interactions with the cytoskeleton and lipids provide general mechanisms for the formation of subdomains. We therefore expect the results revealed by our study to be applicable to a wide class of signaling systems at the membrane. We have focused in this paper on coupled single-site modification reactions because this motif governs pathways specifically known to be affected by the formation of membrane subdomains. However, the effects we uncover also pertain to multisite modification reactions, which are very common in cell signaling (26–29). Moreover, we have focused on systems where the reactant species are confined by a boundary that limits diffusion. However, similar effects could be observed in systems where proteins are localized to raft domains, or even scaffolds or large macromolecular complexes. In the latter case, each complex would effectively provide an independent reaction “compartment,” and the exchange between compartments would be the result of rare dissociation events, after which proteins could diffuse rapidly through the cytoplasm to a different complex. Even if the signal within each complex was not mediated via diffusive encounters, but rather via cooperative or allosteric interactions, the fundamental mechanism that we reveal here—that partitioning into subsystems removes correlations between subsystems—remains at play. The presence of scaffolds and macromolecular complexes at early stages of signaling pathways is extremely common (30), suggesting that the effects discussed here are of wide biological relevance.

Materials and Methods

The CME (1) is solved using the method of spectral expansion (15, 16). Details of this method, the computation of mutual information, and the spatial simulations are described in *SI Appendix*. Source code, written in MATLAB, C++, and Mathematica, used to generate all results and figures in the main text and *SI Appendix*, is freely available at <http://partitioning.sourceforge.net>.

ACKNOWLEDGMENTS. We thank Philippe Nghe for a critical reading of the manuscript. This work is part of the research program of Stichting voor Fundamenteel Onderzoek der Materie, which is financially supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

Supporting Information for “Spatial partitioning improves the reliability of biochemical signaling”

Andrew Mugler¹, Filipe Tostevin¹, and Pieter Rein ten Wolde¹

¹FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

This document contains the methods, supporting analytic results, and supplementary Figures S1-S8.

1 Methods

1.1 Spectral solution of the master equation

The chemical master equation (CME) is solved using the method of spectral expansion [1, 2], described in detail in the next section. Briefly, the structure of the CME, in which the dynamics can be separated into two operators that act only on m or n but not both, allows for its solution to be written in the form $p_{mn}(t) = \sum_{j=0}^M \sum_{k=0}^N G_{jk}(t; \bar{\beta}) \phi_m^j(\alpha) \phi_n^k(\bar{\beta})$, where $\phi_m^j(\alpha)$ is the j^{th} eigenvector of the operator $\mathcal{L}_m(\alpha)$ and similarly for $\phi_n^k(\bar{\beta})$, and $\bar{\beta}$ is an expansion parameter on which p_{mn} does not ultimately depend. The expansion coefficients $G_{jk}(t; \bar{\beta})$ can be calculated straightforwardly, as shown in the next section. Importantly, this spectral expansion dramatically decreases the computational complexity of calculating p_{mn} : rather than solving the $(M + 1)(N + 1) \times (M + 1)(N + 1)$ system of the original CME, it is only necessary to solve N linear systems of size $(M + 1) \times (M + 1)$ for the vectors of coefficients \vec{G}_k . We emphasize that since the system has a finite state-space, no approximations are made in using the spectral expansion, and the solution remains exact. Furthermore, the moments of the steady-state distribution p_{mn} can be conveniently expressed in terms of the expansion coefficients G_{jk} ; in particular, $\langle n \rangle = G_{01}$ and $\langle n^2 \rangle = 2G_{02} + G_{01}$.

1.2 Mutual information

The mutual information between network input and response is given by the standard expression [3] $I[\alpha, m] = \langle \log\{p(\alpha, m)/[p(\alpha)p(m)]\} \rangle$, where the average is taken over the joint distribution $p(\alpha, m) = p(m|\alpha)p(\alpha)$, and $p(m|\alpha) = \sum_{n=0}^M p(m, n|\alpha)$ is given by the steady state of the CME. The calculation of the mutual information requires specification of the distribution of input signals $p(\alpha)$. We choose N_α values of α such that $q = \alpha/(\alpha + 1) = \langle m \rangle / M$ is uniformly-spaced over the range $0 \leq q \leq 1$; then $p(n) = \sum_{i=1}^{N_\alpha} p(n|\alpha_i)p(\alpha_i)$ and $p(\alpha_i) = N_\alpha^{-1}$. However, our conclusions are unaffected if we instead take a input distribution that is unimodal or bimodal (Fig. S7). We take $N_\alpha > 30$, for which $I[\alpha, m]$ converges to within 1% of its large- N_α limit (Fig. S8).

1.3 Spatial simulations

The diffusion and reactions of M \mathcal{X} molecules and N \mathcal{Y} molecules are simulated on a two-dimensional square lattice of side length λ using a fixed-time-step integration scheme. During each step of duration δt , each particle is moved to a random neighboring lattice site with probability $p_D = (D/\ell^2)\delta t$, where D is the diffusion constant, and ℓ is the lattice spacing. Molecules have steric interactions on the lattice, such that only one molecule can be present at each lattice site at any time. Attempted moves to an occupied site are rejected, with the particle remaining at its original position. If a molecule in the X^* state is adjacent to a molecule in the Y state, the latter is converted to the Y^* state with probability $p_r = \gamma(\beta\lambda^2/M)\delta t$. To make π partitions, linear diffusion barriers are placed at $i\lambda/\sqrt{\pi}$ in each direction, where $i \in \{0, 1, \dots, \sqrt{\pi} - 1\}$. A diffusion step which crosses such a barrier is accepted with probability reduced by a factor p_{hop} . The time step δt is chosen sufficiently small that no probability exceeds one.

2 Solution of the master equation by spectral expansion

This section describes the solution via the method of spectral expansion, or the ‘spectral method’, of the CME introduced in the main text. The spectral method has been used fruitfully in the context of gene regulation to solve CMEs describing cascades [1], bursts [2], and oscillations [4], and a pedagogical treatment is available in [5]. Here we apply the spectral method to coupled reversible switching.

From Eqns. 1-2 of the main text, the stochastic dynamics of the system under study are given by the CME

$$\dot{p}_{mn} = -[\mathcal{L}_m(\alpha, M) + \gamma\mathcal{L}_n(\beta_m, N)] p_{mn}, \quad (1)$$

where both operators \mathcal{L}_m and \mathcal{L}_n have the form

$$\mathcal{L}_m(\alpha, M) = \alpha [1 - \mathbb{E}_m^{-1}] (M - m) + [1 - \mathbb{E}_m^{+1}] m, \quad (2)$$

with $\mathbb{E}_m^i f(m) = f(m+i)$ defining the step operator. The CME describes the evolution of the probability of having m \mathcal{X} proteins in the active state and n \mathcal{Y} proteins in the active state, with β_m the coupling function by which \mathcal{X} drives the activation of \mathcal{Y} .

2.1 The moments do not close

We first demonstrate that direct computation of the moments from the CME is not possible because the moments do not close. The reason is that a nonlinearity is present in the first term of Eqn. 2 in the form of the factor $\beta_m n$. As a result, the first moment depends on a higher moment, which in turn depends on an even higher moment, and so on.

To see explicitly that the moments do not close, we consider computing the dynamics of the first moment of the driven species, the mean $\langle n \rangle$, by summing the CME over m and n against n . We obtain

$$\frac{1}{\gamma} \partial_t \langle n \rangle = -\langle n \rangle + N \langle \beta_m \rangle - \langle \beta_m n \rangle, \quad (3)$$

where averages are taken over p_{mn} . We see that indeed the final term carries the nonlinearity. Even for the simplest coupling function, i.e. linear coupling $\beta_m = cm$, one finds a hierarchy of moment dependencies that

does not close:

$$\partial_t \langle n \rangle = -\gamma \langle n \rangle + \gamma c N \langle m \rangle - \gamma c \langle mn \rangle, \quad (4)$$

$$\partial_t \langle mn \rangle = \alpha M \langle n \rangle - (\alpha + \gamma + 1) \langle mn \rangle + \gamma c N \langle m^2 \rangle - \gamma c \langle m^2 n \rangle, \quad (5)$$

$$\partial_t \langle m^2 n \rangle = \dots \quad (6)$$

That is, the dynamics of $\langle n \rangle$ depend on $\langle mn \rangle$, whose dynamics depend on $\langle m^2 n \rangle$, and so on.

The fact that the moments cannot be computed—indeed, not even the mean output $\langle n \rangle$ —makes it particularly important to actually solve the CME in order to learn about the statistical properties of this system.

2.2 The spectrum of the switch operator

The CME is a linear equation. Even when the rates are nonlinear functions of the molecule numbers, the CME is still linear in its degree of freedom, the joint probability. The most straightforward way to solve a linear equation is to write its solution as an expansion in the eigenfunctions of the linear operator. Although it is difficult to derive the eigenfunctions of the coupled operator $\mathcal{L}_m(\alpha, M) + \gamma \mathcal{L}_n(\beta_m, N)$, it is straightforward to derive the eigenfunctions of the uncoupled operator $\mathcal{L}_m(\alpha, M)$, which we call the switch operator. Indeed, we will see that expanding the joint probability in eigenfunctions of the uncoupled operator greatly simplifies the form of the CME, yielding an exact solution in terms of matrix algebra.

The switch operator governs the CME for the first species \mathcal{X} ; explicitly,

$$\dot{p}_m = -\mathcal{L}p_m = \alpha[M - (m - 1)]p_{m-1} + (m + 1)p_{m+1} - [\alpha(M - m) + m]p_m, \quad (7)$$

where for notational simplicity we have taken $\mathcal{L}_m(\alpha, M) \rightarrow \mathcal{L}$. Its eigenvalue relation is written

$$\mathcal{L}\phi_m^j = \lambda_j \phi_m^j, \quad (8)$$

for eigenvalues λ_j and eigenvectors ϕ_m^j .

2.2.1 Eigenvalues

The matrix form of the operator \mathcal{L} can be read directly from Eqn. 7:

$$\mathbf{L} = \begin{pmatrix} M\alpha & -1 & & & \\ -M\alpha & (M-1)\alpha + 1 & -2 & & \\ & -(M-1)\alpha & (M-2)\alpha + 2 & -3 & \\ & & \ddots & \ddots & \ddots \\ & & & -3\alpha & 2\alpha + (M-2) & -(M-1) \\ & & & & -2\alpha & \alpha + (M-1) & -M \\ & & & & & -\alpha & M \end{pmatrix}. \quad (9)$$

The tridiagonal structure follows from the fact that molecule numbers only increase or decrease by one at a time. Practically speaking, the eigenvalues can be obtained using the fact that the determinant of a tridiagonal matrix can be computed recursively. Performing the computation for $M = 0, 1, 2, \dots$ reveals the pattern

$$\lambda_j = (\alpha + 1)j, \quad j \in \{0, 1, 2, \dots, M\}. \quad (10)$$

However, Eqn. 10 can be derived more rigorously by making use of a generating function. We present this derivation next, since the generating function formalism will also prove quite useful in deriving the eigenvectors and solving the CME.

The generating function is an expansion in any complete basis for which the probability distribution provides the expansion coefficients [6]. Choosing as our basis the set of polynomials in some continuous variable x , the generating function is defined

$$G(x) = \sum_{m=0}^M p_m x^m. \quad (11)$$

The probability distribution is recovered via the inverse transform

$$p_m = \frac{1}{m!} \partial_x^m [G(x)]_{x=0}. \quad (12)$$

A key utility of the generating function is turning the CME, which is a set of ordinary differential equations (ODEs), into a single partial differential equation. Indeed, summing Eqn. 7 against x^m yields

$$\dot{G} = -(x - 1)[(\alpha x + 1)\partial_x - \alpha M]G, \quad (13)$$

where the appearances of x and ∂_x arise from the shifts $m - 1$ and $m + 1$, respectively. Eqn. 13 directly gives the form of the operator in x space: $\mathcal{L} = (x - 1)[(\alpha x + 1)\partial_x - \alpha M]$. The eigenfunctions are then obtained from the relation $\mathcal{L}\phi^j(x) = \lambda_j\phi^j(x)$ by separating variables and integrating:

$$\phi^j(x) = (\alpha + 1)^{-M}(x - 1)^{\lambda_j/(\alpha+1)}(\alpha x + 1)^{M-\lambda_j/(\alpha+1)}. \quad (14)$$

Here, the constant factor $(\alpha + 1)^{-M}$ is determined by application of the normalization condition $G(1) = 1$ to the steady state solution, which is obtained by setting $\lambda_j = 0$:

$$G(x) = \left(\frac{\alpha x + 1}{\alpha + 1} \right)^M. \quad (15)$$

We will solve Eqn. 13 in two ways: by the method of characteristics and by expansion in the eigenfunctions; together these solutions will reveal the eigenvalues.

First, the method of characteristics [7] posits that the dependence of G on x and t occurs via some parametric variable s , i.e. $G(x, t) = G[x(s), t(s)]$. The chain rule then gives $dG/ds = (\partial G/\partial x)(dx/ds) + (\partial G/\partial t)(dt/ds)$, which when compared term by term with Eqn. 13 yields three ordinary differential equations:

$$\frac{dt}{ds} = 1, \quad \frac{dx}{ds} = (x - 1)(\alpha x + 1), \quad \frac{dG}{ds} = \alpha M(x - 1). \quad (16)$$

The first identifies $s = t$, with which the second is solved by

$$z = \frac{x - 1}{\alpha x + 1} e^{-(\alpha+1)t}, \quad (17)$$

where z is a constant of integration. The crux of the method is that Eqn. 17 defines a characteristic curve on which G must depend, i.e. $G(x, t) = f[z(x, t)]g(x, t)$, where f and g are unknown functions, and z has been promoted to a characteristic function of x and t . The function g is identified by realizing that steady state is reached as $t \rightarrow \infty$, for which $f(z) \rightarrow f(0)$ no longer depends on x or t . Therefore, g must be the steady state function given in Eqn. 15:

$$G(x, t) = \left(\frac{\alpha x + 1}{\alpha + 1} \right)^M f(z). \quad (18)$$

Although we still do not know f , we may Taylor expand it around the point $z = 0$, yielding

$$G(x, t) = \left(\frac{\alpha x + 1}{\alpha + 1} \right)^M \sum_{j=0}^{\infty} c_j z^j = \left(\frac{\alpha x + 1}{\alpha + 1} \right)^M \sum_{j=0}^{\infty} c_j \left(\frac{x - 1}{\alpha x + 1} \right)^j e^{-(\alpha+1)jt}, \quad (19)$$

where $c_j \equiv \partial_z^j[f(z)]_{z=0}/j!$.

Second, because Eqn. 13 is linear, we may also write down its solution as an expansion in the eigenfunctions of its linear operator:

$$G(x, t) = \sum_j C_j(t) \phi^j(x). \quad (20)$$

Under the assumption that the eigenfunctions are orthogonal (which will be shown in the next section), inserting Eqn. 20 into Eqn. 13 yields an independent ODE for each C_j , $\dot{C}_j = -\lambda_j C_j$, which is solved by $C_j(t) = c_j e^{-\lambda_j t}$ for initial conditions c_j . Inserting this functional form and that for $\phi_j(t)$ (Eqn. 14) into Eqn. 20 yields

$$G(x, t) = \left(\frac{\alpha x + 1}{\alpha + 1} \right)^M \sum_j c_j \left(\frac{x - 1}{\alpha x + 1} \right)^{\lambda_j / (\alpha + 1)} e^{-\lambda_j t}. \quad (21)$$

Comparison of Eqns. 19 and 21 reveals both the expression for the eigenvalues, $\lambda_j = (\alpha + 1)j$, and a limit on their domain, the nonnegative integers $j \in \{0, 1, 2, \dots, \infty\}$. Of course, the domain can be a subset of the nonnegative integers; then some c_j in Eqn. 21 would be zero. Indeed, since \mathbf{L} is a finite matrix of size $M + 1$ by $M + 1$ (Eqn. 9), it is spanned by $M + 1$ linearly independent eigenvectors, meaning we expect only $M + 1$ eigenvalues. In fact, the only set of $M + 1$ nonnegative integers that satisfies the requirement that the trace of \mathbf{L} , $\sum_{m=0}^M [(M - m)\alpha + m] = (\alpha + 1)M(M + 1)/2$, equals the sum of the eigenvalues, $\sum_j (\alpha + 1)j$, is $j \in \{0, 1, 2, \dots, M\}$. Thus, we arrive at the result

$$\lambda_j = (\alpha + 1)j, \quad j \in \{0, 1, 2, \dots, M\}, \quad (22)$$

as proposed by inspection in Eqn. 10.

2.2.2 State space notation

The linear algebraic manipulations we have done thus far can be cast in the more abstract notation of state spaces, commonly used in quantum mechanics [8]. We will find this notation useful in later sections, for example in transforming between the molecule number basis and the eigenbasis. Specifically, we introduce a state $|p\rangle$ that can be projected into $\langle m|$ space to give the probability distribution, or into $\langle x|$ space to give the generating function:

$$\langle m|p\rangle = p_m, \quad \langle x|p\rangle = G(x). \quad (23)$$

In the same way, the j th eigenstate $|j\rangle$ is projected into $\langle m|$ space to give the j th eigenvector, or into $\langle x|$ space to give the j th eigenfunction:

$$\langle m|j\rangle = \phi_m^j, \quad \langle x|j\rangle = \phi^j(x). \quad (24)$$

This notation offers new insight into our definition of the generating function. For example, Eqn. 11 can now be written

$$\langle x|p\rangle = \sum_{m=0}^M \langle x|m\rangle \langle m|p\rangle, \quad (25)$$

where we have recognized

$$\langle x|m\rangle = x^m \quad (26)$$

as the projection of the state $|m\rangle$ into $\langle x|$ space. Eqn. 25 has a clear interpretation: we have inserted a complete set of $|m\rangle$ states. Similarly, Eqn. 12 can now be written

$$\langle m|p\rangle = \oint dx \frac{G(x)}{x^{m+1}} = \oint dx \langle m|x\rangle \langle x|p\rangle. \quad (27)$$

In the first step, we have rewritten Eqn. 12 using Cauchy's theorem, where $dx \equiv dx/2\pi i$, and the contour surrounds the pole at $x = 0$. In the second step, we have recognized

$$\langle m|x\rangle = \frac{1}{x^{m+1}} \quad (28)$$

as the conjugate to $\langle x|m\rangle$. Eqn. 27 has the clear interpretation of inserting a complete set of $|x\rangle$ states, under an inner product defined by the complex integration. The choice of inner product and of conjugate state are made such that orthonormality is preserved, a fact which we may confirm by again employing Cauchy's theorem:

$$\langle m|m'\rangle = \oint dx \langle m|x\rangle \langle x|m'\rangle = \oint dx \frac{x^{m'}}{x^{m+1}} = \frac{1}{m!} \partial_x^m \left[x^{m'} \right]_{x=0} \theta(m > 0) = \delta_{mm'}. \quad (29)$$

Finally, the dynamics in Eqn. 13 can be written in state space as

$$|\dot{p}\rangle = -\hat{\mathcal{L}}|p\rangle = -(\hat{a}^+ - 1)[(\alpha\hat{a}^+ + 1)\hat{a}^- - \alpha M]|p\rangle, \quad (30)$$

where we have defined the operators \hat{a}^+ and \hat{a}^- whose projections in x space are $\langle x|\hat{a}^+ = x$ and $\langle x|\hat{a}^- = \partial_x$. These are analogous to the raising and lowering operators in the well known treatment of the quantum harmonic oscillator. This operator formalism for the generating function was first developed in the 1970s; for a review see [9].

2.2.3 Eigenvectors

The state space notation facilitates a derivation of the functional form of the eigenvectors:

$$\phi_m^j = \langle m|j\rangle = \oint dx \langle m|x\rangle \langle x|j\rangle = \oint dx \frac{1}{x^{m+1}} \frac{(x-1)^j (\alpha x + 1)^{M-j}}{(\alpha + 1)^M}. \quad (31)$$

Here we have inserted the eigenfunctions

$$\phi_j(x) = \langle x|j\rangle = \frac{(x-1)^j (\alpha x + 1)^{M-j}}{(\alpha + 1)^M} \quad (32)$$

from Eqn. 14, with eigenvalues given by Eqn. 22. We use Cauchy's theorem to perform the integration and recognize that derivatives of a product follow a binomial expansion:

$$\phi_m^j = \frac{1}{(\alpha + 1)^M} \frac{1}{m!} \partial_x^m \left[(\alpha x + 1)^{M-j} (x-1)^j \right]_{x=0} \quad (33)$$

$$= \frac{1}{(\alpha + 1)^M} \frac{1}{m!} \sum_{\ell=0}^m \binom{m}{\ell} \partial_x^\ell \left[(\alpha x + 1)^{M-j} \right]_{x=0} \partial_x^{m-\ell} \left[(x-1)^j \right]_{x=0} \quad (34)$$

$$= \frac{1}{(\alpha + 1)^M} \frac{1}{m!} \sum_{\ell=0}^m \frac{m!}{(m-\ell)!\ell!} \left[\frac{(M-j)!\alpha^\ell}{(M-j-\ell)!} \theta(\ell \leq M-j) \right] \left[\frac{j!(-1)^{j-m+\ell}}{(j-m+\ell)!} \theta(m-\ell \leq j) \right] \quad (35)$$

$$= \frac{(-1)^{j-m}}{(\alpha + 1)^M} \sum_{\ell \in \Omega} \binom{M-j}{\ell} \binom{j}{m-\ell} (-\alpha)^\ell. \quad (36)$$

Here the domain Ω results from the derivatives and is defined by $\max(0, m-j) \leq \ell \leq \min(m, M-j)$. Eqn. 36 gives the expression for the eigenvectors. For $j = 0$ the expression reduces to the binomial distribution in terms of the occupancy $q = \alpha/(\alpha + 1)$, as it must, since this is the steady state of the uncoupled process:

$$\phi_m^0 = \binom{M}{m} \frac{\alpha^m}{(\alpha + 1)^M} = \binom{M}{m} q^m (1-q)^{M-m}. \quad (37)$$

This function has one maximum, and in general the j th eigenvector has $j+1$ extrema, making the eigenvectors qualitatively similar to Fourier modes or eigenfunctions of the quantum harmonic oscillator.

The switch operator $\hat{\mathcal{L}}$ is not Hermitian. A consequence is that its conjugate eigenvectors $\psi_m^j = \langle j|m\rangle$ (row vectors) are not complex conjugates of its eigenvectors $\phi_m^j = \langle m|j\rangle$ (column vectors). Rather, they

are distinct functions that must be constructed to obey an orthonormality relation in order to constitute a complete basis. The orthonormality relation can be used to derive their form in x space, $\psi^j(x) = \langle j|x\rangle$:

$$\delta_{jj'} = \langle j|j'\rangle = \oint dx \langle j|x\rangle \langle x|j'\rangle = \oint dx \psi^j(x) \frac{(x-1)^{j'}(\alpha x+1)^{M-j'}}{(\alpha+1)^M} = \oint dz_0 z_0^{j'} f_j(z_0). \quad (38)$$

Here we have defined $z_0 \equiv (x-1)/(\alpha x+1)$ and $f_j(z_0) \equiv \psi^j(x)(\alpha x+1)^{M+2}/(\alpha+1)^{M+1}$ in order to draw an equivalence between Eqn. 38 and Eqn. 29, which then implies $f_j(z_0) = 1/z_0^{j+1} = (\alpha x+1)^{j+1}/(x-1)^{j+1}$, or

$$\psi^j(x) = \frac{(\alpha+1)^{M+1}}{(\alpha x+1)^{M-j+1}(x-1)^{j+1}}. \quad (39)$$

Eqn. 39 gives the form of the conjugate eigenfunctions in x space, which can be used to derive the expression for the conjugate eigenvectors as in Eqns. 31-36:

$$\psi_m^j = \langle j|m\rangle = \oint dx \langle j|x\rangle \langle x|m\rangle = \oint dx \frac{(\alpha+1)^{M+1}}{(\alpha x+1)^{M-j+1}(x-1)^{j+1}} x^m \quad (40)$$

$$= \sum_{\ell \in \Omega} \binom{M-j+\ell}{\ell} \binom{m}{j-\ell} (-\alpha)^\ell (\alpha+1)^{j-\ell}. \quad (41)$$

Here Ω is defined by $\max(0, j-m) \leq \ell \leq j$. Eqn. 41 gives the expression for the conjugate eigenvectors. They are j th order polynomials in m .

2.3 Expanding the coupled problem in uncoupled eigenfunctions

We now solve the CME by expanding the solution in the eigenfunctions of the uncoupled operator. This procedure is most easily done in state space, in which the CME reads

$$|\dot{p}\rangle = -[\hat{\mathcal{L}}_x(\alpha) + \gamma \hat{\mathcal{L}}_{xy}]|p\rangle \quad (42)$$

where

$$\hat{\mathcal{L}}_x(\alpha) = (\hat{a}_x^+ - 1)[(\alpha \hat{a}_x^+ + 1)\hat{a}_x^- - \alpha M], \quad (43)$$

$$\hat{\mathcal{L}}_{xy} = (\hat{a}_y^+ - 1)[(\hat{\beta}_x \hat{a}_y^+ + 1)\hat{a}_y^- - \hat{\beta}_x N], \quad (44)$$

as in Eqn. 30, and we have introduced the operator $\hat{\beta}_x$ whose action on the state $|m\rangle$ yields the coupling function, $\hat{\beta}_x|m\rangle = \beta_m|m\rangle$. The first step is to write the full operator as two uncoupled operators plus a correction term. Introducing the constant $\bar{\beta}$ to parameterize the second uncoupled operator, the CME becomes

$$|\dot{p}\rangle = -[\hat{\mathcal{L}}_x(\alpha) + \gamma \hat{\mathcal{L}}_y(\bar{\beta}) + \gamma \hat{\Gamma}_x \hat{\Delta}_y]|p\rangle \quad (45)$$

where we have explicitly denoted the fact that the correction term $\hat{\mathcal{L}}_{xy} - \hat{\mathcal{L}}_y(\bar{\beta})$ factorizes into two operators that act on each of the x and y sectors alone:

$$\hat{\Gamma}_x \equiv \hat{\beta}_x - \bar{\beta}, \quad (46)$$

$$\hat{\Delta}_y \equiv (\hat{a}_y^+ - 1)(\hat{a}_y^+ \hat{a}_y^- - N). \quad (47)$$

The second step is to expand the solution in the eigenfunctions of the two uncoupled operators. Introducing k as the mode index for the eigenstates of $\hat{\mathcal{L}}_y(\bar{\beta})$, we write

$$|p\rangle = \sum_{j=0}^M \sum_{k=0}^N G_{jk}|j,k\rangle. \quad (48)$$

Inserting this form into the CME, projecting with the conjugate state $\langle j', k' |$, and summing over j and k yields the dynamics for the expansion coefficients G_{jk} :

$$\dot{G}_{jk} = -[(\alpha + 1)j + \gamma(\bar{\beta} + 1)k]G_{jk} - \gamma \sum_{j'=0}^M \Gamma_{jj'} \sum_{k'=0}^N \Delta_{kk'} G_{j'k'}. \quad (49)$$

Here the first term is diagonal and reflects the actions of the uncoupled operators on their eigenstates. The second term contains the corrections $\Gamma_{jj'} = \langle j|\hat{\Gamma}_x|j'\rangle$ and $\Delta_{kk'} = \langle k|\hat{\Delta}_y|k'\rangle$. The first correction is directly evaluated by inserting a complete set of m states:

$$\Gamma_{jj'} = \sum_{m=0}^M \langle j|(\hat{\beta}_x - \bar{\beta})|m\rangle \langle m|j'\rangle = \sum_{m=0}^M \langle j|m\rangle (\beta_m - \bar{\beta}) \langle m|j'\rangle \quad (50)$$

$$= \sum_{m=0}^M \psi_m^j (\beta_m - \bar{\beta}) \phi_m^{j'}. \quad (51)$$

We see that $\Gamma_{jj'}$ is the simply the difference between the coupling function and the constant parameter, rotated into eigenspace. Notably, for linear coupling, $\Gamma_{jj'}$ is tridiagonal (see Sec. 2.5). The second correction is most easily evaluated by inserting a complete set of y states; the result, derived in Sec. 2.5, is

$$\Delta_{kk'} = k\delta_{kk'} - (N - k + 1)\delta_{k-1,k'}. \quad (52)$$

We see that $\Delta_{kk'}$ is subdiagonal in k , which simplifies the dynamics of G_{jk} to

$$\dot{G}_{jk} = - \sum_{j=0}^M \Lambda_{jj'}^k G_{j'k} + \gamma(N - k + 1) \sum_{j=0}^M \Gamma_{jj'} G_{j',k-1}, \quad (53)$$

where we define the matrix acting on the diagonal part as

$$\Lambda_{jj'}^k \equiv [(\alpha + 1)j + \gamma(\bar{\beta} + 1)k]\delta_{jj'} + \gamma k \Gamma_{jj'}. \quad (54)$$

The subdiagonality allows one to write the steady state of Eqn. 53 as an iterative scheme, by which the k th column of G_{jk} is computed from the $(k - 1)$ th column:

$$\vec{G}_k = \gamma(N - k + 1) \boldsymbol{\Lambda}_k^{-1} \mathbf{T} \vec{G}_{k-1}. \quad (55)$$

The scheme is initialized with

$$\vec{G}_0 = \delta_{j0} \quad (56)$$

(see Sec. 2.5), and the joint distribution is recovered via

$$p_{mn} = \sum_{j=0}^M \sum_{k=0}^N G_{jk} \phi_m^j \phi_n^k, \quad (57)$$

which is the projection of Eqn. 48 into $\langle m, n |$ space.

Eqn. 57 constitutes an exact steady state solution to the CME, with G_{jk} computed iteratively via Eqns. 55 and 56, auxiliary matrices defined in Eqns. 51 and 54, and the eigenvectors given by Eqns. 36 and 41. Importantly, the computational complexity of the solution has been dramatically reduced: rather than solving the original CME (Eqn. 1), which requires inverting its operator of size $(M + 1)(N + 1) \times (M + 1)(N + 1)$, Eqn. 55 makes clear that it is only necessary to invert N smaller matrices of size $(M + 1) \times (M + 1)$, i.e. the matrices $\boldsymbol{\Lambda}_k$ for $k \in \{1, 2, \dots, N\}$.

2.4 Exact expressions for moments

Now that we have an exact solution to the CME in terms of a spectral expansion, moments take an exact form in terms of the expansion coefficients. We thus circumvent the problem of moment closure, instead arriving at compact expressions that require only the inversion and multiplication of finite matrices via Eqn. 55.

Moments are most easily computed from the generating function, $G(x, y)$. For example, the ν th moment of the output is

$$\langle n^\nu \rangle = [(y\partial_y)^\nu G(x=1, y)]_{y=1}. \quad (58)$$

In terms of the expansion, the generating function is $G(x, y) = \langle x, y | p \rangle = \sum_{j=0}^M \sum_{k=0}^N G_{jk} \langle x | j \rangle \langle y | k \rangle$, and using the fact that $\langle x = 1 | j \rangle = \delta_{j0}$ (Eqn. 32), we have

$$\langle n^\nu \rangle = \sum_{k=0}^N G_{0k} [(y\partial_y)^\nu \langle y | k \rangle]_{y=1}. \quad (59)$$

Inserting the expression for $\langle y | k \rangle$ (Eqn. 32) and defining $w \equiv \log y$, we obtain

$$\langle n^\nu \rangle = \sum_{k=0}^N G_{0k} \partial_w^\nu \left[\frac{(e^w - 1)^k (\bar{\beta} e^w + 1)^{N-k}}{(\bar{\beta} + 1)^N} \right]_{w=0}. \quad (60)$$

At this point we recall that $\bar{\beta}$ is a constant we introduce to parameterize the expansion. The expression for the moments therefore cannot depend on $\bar{\beta}$: if we change $\bar{\beta}$, the expression in brackets changes, but the expansion coefficients G_{0k} also change, such that Eqn. 60 evaluates to the same $\bar{\beta}$ -independent form. We are therefore free to set $\bar{\beta}$ to any value, and the choice $\bar{\beta} = 0$ makes the derivative easiest to evaluate. Thus we have

$$\langle n^\nu \rangle = \sum_{k=0}^N G_{0k} \partial_w^\nu [(e^w - 1)^k]_{w=0}, \quad (61)$$

where it is now understood that G_{0k} is computed with $\bar{\beta} = 0$. Evaluating the derivative yields

$$\langle n^\nu \rangle = \sum_{k=0}^N G_{0k} \left[\sum_{\ell=1}^{\min(k, \nu)} \begin{Bmatrix} \nu \\ \ell \end{Bmatrix} \frac{k!}{(k-\ell)!} e^{\ell w} (e^w - 1)^{k-\ell} \right]_{w=0} \quad (62)$$

$$= \sum_{k=0}^N G_{0k} \sum_{\ell=1}^{\min(k, \nu)} \begin{Bmatrix} \nu \\ \ell \end{Bmatrix} \frac{k!}{(k-\ell)!} \delta_{k\ell} \quad (63)$$

$$= \sum_{k=1}^{\min(\nu, N)} G_{0k} \begin{Bmatrix} \nu \\ k \end{Bmatrix} k! \quad (64)$$

in terms of the Stirling numbers of the second kind,

$$\begin{Bmatrix} \nu \\ k \end{Bmatrix} = \frac{1}{k!} \sum_{\ell=0}^k (-1)^{k-\ell} \binom{k}{\ell} \ell^\nu. \quad (65)$$

For example, the first moment, second moment, and variance are

$$\langle n \rangle = G_{01}, \quad (66)$$

$$\langle n^2 \rangle = G_{01} + 2G_{02}, \quad (67)$$

$$\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 = G_{01} + 2G_{02} - G_{01}^2. \quad (68)$$

These are exact expressions for the moments in terms of the expansion coefficients G_{0k} , which are obtained by matrix inversion and multiplication via Eqn. 55, e.g. in Mathematica.

An informative special case is immediately revealed when $N = 1$, for which G_{02} does not exist, i.e. $\langle n \rangle = G_{01}$ and $\sigma_n^2 = G_{01} - G_{01}^2$, or

$$\sigma_n^2 = \langle n \rangle (1 - \langle n \rangle) \quad (N = 1). \quad (69)$$

Here there is only one output molecule. The relationship between its mean activation and the associated noise must therefore obey the known result for a single binary switch, Eqn. 69.

2.5 Auxiliary calculations

Here we show that $\Gamma_{jj'}$ is tridiagonal for linear $\beta_m = cm$:

$$\Gamma_{jj'} = \langle j | \hat{\Gamma}_x | j' \rangle \quad (70)$$

$$= \langle j | (c\hat{a}_x^+ \hat{a}_x^- - \bar{\beta}) | j' \rangle \quad (71)$$

$$= -\bar{\beta}\delta_{jj'} + c \oint dx \langle j | x \rangle \langle x | \hat{a}_x^+ \hat{a}_x^- | j' \rangle \quad (72)$$

$$= -\bar{\beta}\delta_{jj'} + c \oint dx \langle j | x \rangle x \partial_x \langle x | j' \rangle \quad (73)$$

$$= -\bar{\beta}\delta_{jj'} + c \oint dx \langle j | x \rangle x \partial_x \frac{(x-1)^{j'}(\alpha x+1)^{M-j'}}{(\alpha+1)^M} \quad (74)$$

$$= -\bar{\beta}\delta_{jj'} + c \oint dx \langle j | x \rangle \frac{x}{(\alpha+1)^M} \left[j'(x-1)^{j'-1}(\alpha x+1)^{M-j'} \right. \\ \left. + (x-1)^{j'}(M-j')(\alpha x+1)^{M-j'-1}\alpha \right] \quad (75)$$

$$= -\bar{\beta}\delta_{jj'} + c \oint dx \langle j | x \rangle \frac{x(x-1)^{j'-1}(\alpha x+1)^{M-j'-1}}{(\alpha+1)^M} [j'(\alpha x+1) + (x-1)(M-j')\alpha] \quad (76)$$

$$= -\bar{\beta}\delta_{jj'} + \frac{c}{\alpha+1} \oint dx \langle j | x \rangle \frac{x(x-1)^{j'-1}(\alpha x+1)^{M-j'-1}}{(\alpha+1)^M} \left\{ j'(\alpha x+1)^2 \right. \\ \left. + [\alpha(M-j') + j'](\alpha x+1)(x-1) \right. \\ \left. + \alpha(M-j')(x-1)^2 \right\} \quad (77)$$

$$= -\bar{\beta}\delta_{jj'} + \frac{c}{\alpha+1} \oint dx \langle j | x \rangle \left\{ \frac{(x-1)^{j'-1}(\alpha x+1)^{M-j'+1}}{(\alpha+1)^M} [j'] \right. \\ \left. + \frac{(x-1)^{j'}(\alpha x+1)^{M-j'}}{(\alpha+1)^M} [\alpha(M-j') + j'] \right. \\ \left. + \frac{(x-1)^{j'-1}(\alpha x+1)^{M-j'+1}}{(\alpha+1)^M} [\alpha(M-j')] \right\} \quad (78)$$

$$= -\bar{\beta}\delta_{jj'} + \frac{c}{\alpha+1} \oint dx \langle j | x \rangle \{ \langle x | j' - 1 \rangle j' + \langle x | j' \rangle [\alpha(M-j') + j'] + \langle x | j' + 1 \rangle \alpha(M-j') \} \quad (79)$$

$$= -\bar{\beta}\delta_{jj'} + \frac{c}{\alpha+1} \{ \langle j | j' - 1 \rangle j' + \langle j | j' \rangle [\alpha(M-j') + j'] + \langle j | j' + 1 \rangle \alpha(M-j') \} \quad (80)$$

$$= \frac{cj'}{\alpha+1} \delta_{j,j'-1} + \left\{ \frac{c[\alpha(M-j') + j']}{\alpha+1} - \bar{\beta} \right\} \delta_{jj'} + \frac{c\alpha(M-j')}{\alpha+1} \delta_{j,j'+1}. \quad (81)$$

Eqn. 71 recognizes that $\hat{\beta}_x = c\hat{a}_x^+ \hat{a}_x^-$ is the operator representation of β_m (since $\hat{a}^+ \hat{a}^-$ is the number operator, i.e. $\hat{a}_x^+ \hat{a}_x^- |m\rangle = m|m\rangle$), and Eqn. 77 uses the algebraic fact that $x[j'(\alpha x+1) + (x-1)(M-j')\alpha](\alpha+1) = j'(\alpha x+1)^2 + [\alpha(M-j') + j'](\alpha x+1)(x-1) + \alpha(M-j')(x-1)^2$, which is straightforward to verify.

Here we derive Eqn. 52:

$$\Delta_{kk'} = \langle k | \hat{\Delta}_y | k' \rangle \quad (82)$$

$$= \langle k | (\hat{a}_y^+ - 1)(\hat{a}_y^+ \hat{a}_y^- - N) | k' \rangle \quad (83)$$

$$= \oint dy \langle k | y \rangle \langle y | (\hat{a}_y^+ - 1)(\hat{a}_y^+ \hat{a}_y^- - N) | k' \rangle \quad (84)$$

$$= \oint dy \langle k | y \rangle (y - 1)(y \partial_y - N) \langle y | k' \rangle \quad (85)$$

$$= \oint dy \langle k | y \rangle (y - 1)(y \partial_y - N) \frac{(y - 1)^{k'} (\bar{\beta}y + 1)^{N-k'}}{(\bar{\beta} + 1)^N} \quad (86)$$

$$= \oint dy \langle k | y \rangle \frac{(y - 1)}{(\bar{\beta} + 1)^N} \left[yk'(y - 1)^{k'-1} (\bar{\beta}y + 1)^{N-k'} + y(y - 1)^{k'}(N - k')(\bar{\beta}y + 1)^{N-k'-1} \bar{\beta} - N(y - 1)^{k'} (\bar{\beta}y + 1)^{N-k'} \right] \quad (87)$$

$$= \oint dy \langle k | y \rangle \frac{(y - 1)^{k'} (\bar{\beta}y + 1)^{N-k'-1}}{(\bar{\beta} + 1)^N} \left[yk'(\bar{\beta}y + 1) + y(y - 1)(N - k')\bar{\beta} - N(y - 1)(\bar{\beta}y + 1) \right] \quad (88)$$

$$= \oint dy \langle k | y \rangle \frac{(y - 1)^{k'} (\bar{\beta}y + 1)^{N-k'-1}}{(\bar{\beta} + 1)^N} [k'(\bar{\beta}y + 1) - (y - 1)(N - k')] \quad (89)$$

$$= \oint dy \langle k | y \rangle \left[k' \frac{(y - 1)^{k'} (\bar{\beta}y + 1)^{N-k'}}{(\bar{\beta} + 1)^N} - (N - k') \frac{(y - 1)^{k'+1} (\bar{\beta}y + 1)^{N-(k'+1)}}{(\bar{\beta} + 1)^N} \right] \quad (90)$$

$$= \oint dy \langle k | y \rangle [k' \langle y | k' \rangle - (N - k') \langle y | k' + 1 \rangle] \quad (91)$$

$$= k' \langle k | k' \rangle - (N - k') \langle k | k' + 1 \rangle \quad (92)$$

$$= k' \delta_{kk'} - (N - k') \delta_{k,k'+1} \quad (93)$$

$$= k \delta_{kk'} - (N - k + 1) \delta_{k-1,k'}. \quad (94)$$

Here we derive Eqn. 56:

$$\vec{G}_0 = G_{j0} \quad (95)$$

$$= \langle j, k = 0 | p \rangle \quad (96)$$

$$= \sum_{m=0}^M \sum_{n=0}^N \langle j | m \rangle \langle k = 0 | n \rangle \langle m, n | p \rangle \quad (97)$$

$$= \sum_{m=0}^M \sum_{n=0}^N \langle j | m \rangle p_{mn} \quad (98)$$

$$= \sum_{m=0}^M \langle j | m \rangle p_m \quad (99)$$

$$= \sum_{m=0}^M \langle j | m \rangle \langle m | j = 0 \rangle \quad (100)$$

$$= \langle j | j = 0 \rangle \quad (101)$$

$$= \delta_{j0}. \quad (102)$$

Eqn. 98 uses Eqn. 41 to obtain $\langle k = 0 | n \rangle = 1$, and Eqn. 100 recognizes that p_m is the steady state of the uncoupled operator, $p_m = \phi_m^0 = \langle m | j = 0 \rangle$.

3 Supplementary figures

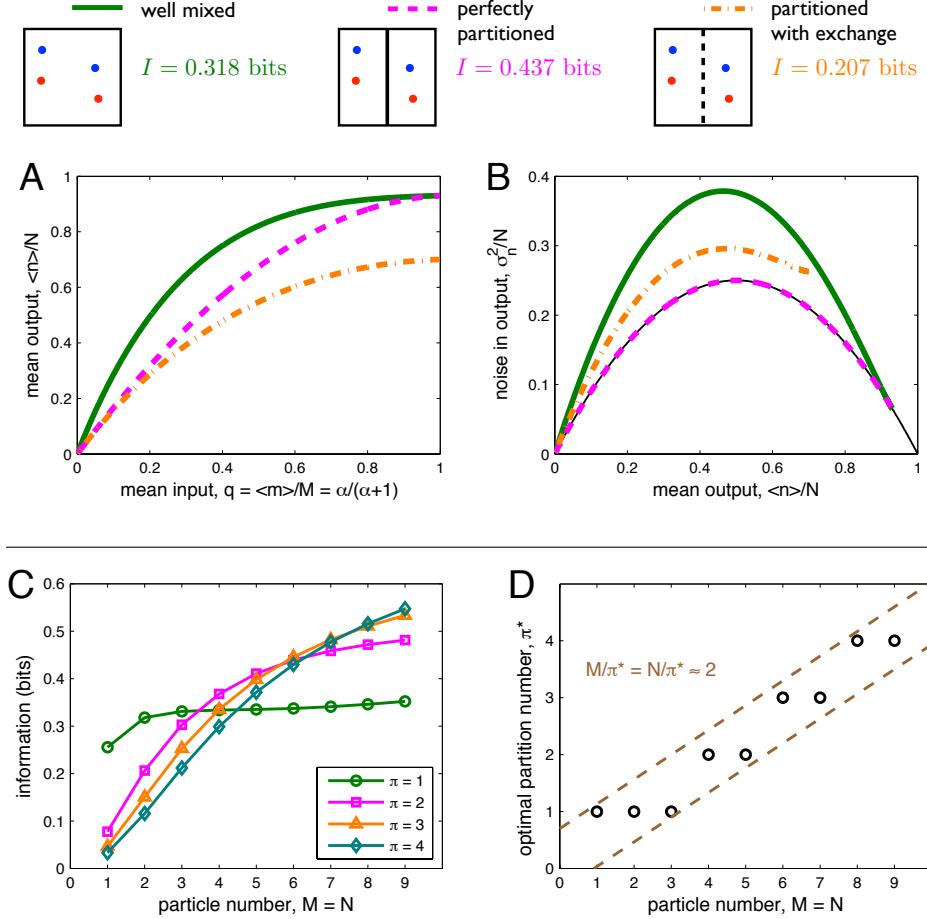


Figure S1: The effects of partitioning persist for Michaelis-Menten coupling. The coupling is described by $\beta_m^{(i)} = \beta m_i / [m_i + (V/\pi)K] = \beta m_i / (m_i + \phi M/\pi)$, where m_i is the number of X^* molecules in partition $i \in \{1, \dots, \pi\}$, and $\phi \equiv KV/M$ is a constant. Here $\beta = 20$, $\phi = 1/2$, and $\gamma = 1$.

A, B As in Fig. 2 of the main text, with $M = N = 2$, perfect partitioning linearizes the input-output relation and reduces the noise, transmitting more information than the well-mixed system; further, allowing exchange among partitions compresses the response and increases the noise compared to the perfectly partitioned system, transmitting less information than the well-mixed system.

C, D As in Fig. 5 of the main text, an information-optimal partition size, here $M/\pi^* = N/\pi^* \approx 2$, emerges due to the trade-off between optimizing signaling reliability and avoiding unfavorable configurations.

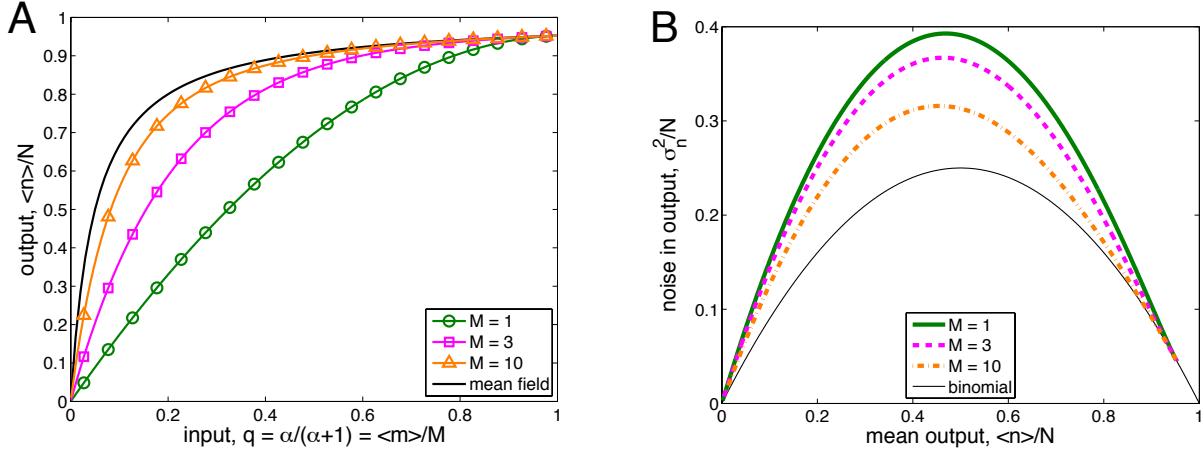


Figure S2: Reducing the number of input molecules linearizes the input-output response and increases the noise in the output. Here $\pi = 1$, $\beta = 20$, and $\gamma = 1$.

A The output (the mean activity of $N = 2$ \mathcal{Y} molecules) vs. the input (the mean activity of M \mathcal{X} molecules) for several values of M . As M is reduced the response becomes more linear, deviating more strongly from the mean-field response $\langle n \rangle / N = \beta q / (\beta q + 1)$. Symbols show 20 uniformly spaced values of q to highlight the effect of saturation on the state space.

B The noise vs. the mean for the output, shown for the same values of M . As M is reduced the noise increases for all values of the mean.

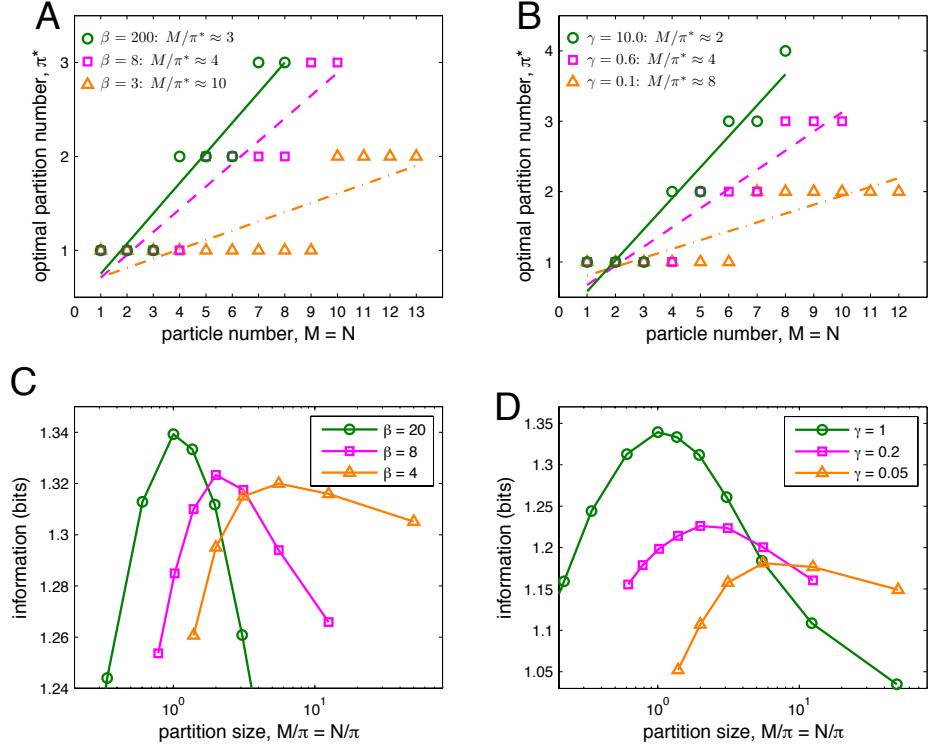


Figure S3: The emergence of an optimal partition size is robust to parameter variations.

A, B Results from the minimal system, described by the chemical master equation, as in Fig. 5B of the main text. The information-optimal partition number π^* is plotted as a function of molecule number $M = N$ for various values of β (A) and γ (B). Linear fits provide estimates of the optimal partition size M/π^* , as indicated in the legends. In A, $\gamma = 1$; in B, $\beta = 20$.

C, D Results from the lattice simulation, in which space is accounted for explicitly, as in Fig. 6C of the main text. The information is plotted as a function of the partition size, directly revealing an optimum, for various values of β (C) and γ (D). Parameters are as in Fig. 6C: $M = N = 49$, $p_{\text{hop}} = 0.001$, $\lambda = 70$, and $p_D/p_r = 1$. In C, $\gamma = 1$; in D, $\beta = 20$.

As discussed in the main text, the optimum arises due to a tradeoff between two key effects of partitioning: on the one hand, partitioning removes correlations in the states of \mathcal{Y} molecules, reducing noise; on the other hand, partitioning isolates molecules, reducing the maximal response. The first effect favors few molecules per partition, while the second effect favors many molecules per partition.

As seen here in both the minimal system (A, B) and the simulated system (C, D), lowering β or γ increases the optimal number of molecules per partition. This result has an intuitive explanation in terms of the above tradeoff: lowering either β or γ slows the rate of switching from the Y to the Y^* state, with respect to the timescale of \mathcal{X} switching. As a result, \mathcal{Y} molecules are less sensitive to individual fluctuations in the state of \mathcal{X} molecules. The states of the \mathcal{Y} molecules therefore exhibit weaker correlations, which in turn weakens the benefit that partitioning imparts in terms of the removal of these correlations. The opposing effect of molecular isolation thus begins to dominate, pushing the optimum toward a larger number of molecules per partition.

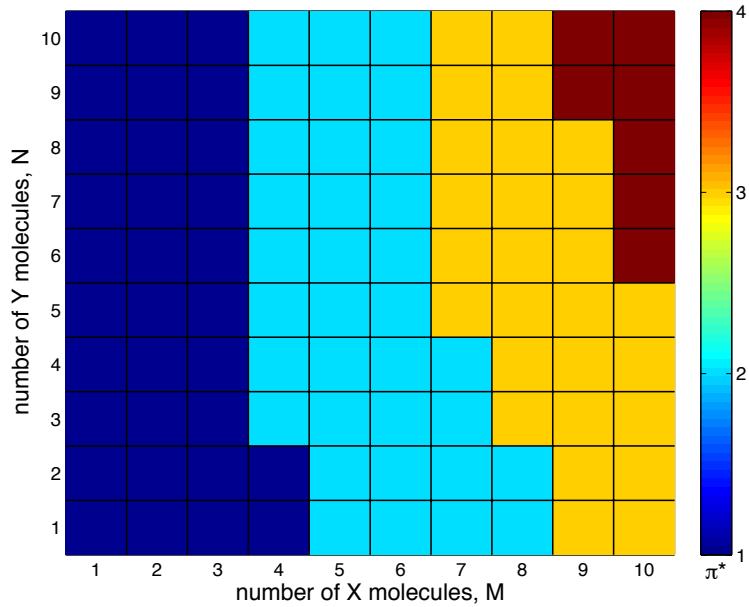


Figure S4: The optimal partition size has only a weak dependence on the number of output molecules. The information-optimal partition number π^* is plotted as a function of the number of \mathcal{X} molecules M and the number of \mathcal{Y} molecules N . The dependence of π^* on N is weak, such that the partition size $M/\pi^* \approx 3 - 4$ is roughly constant over the range of N values.

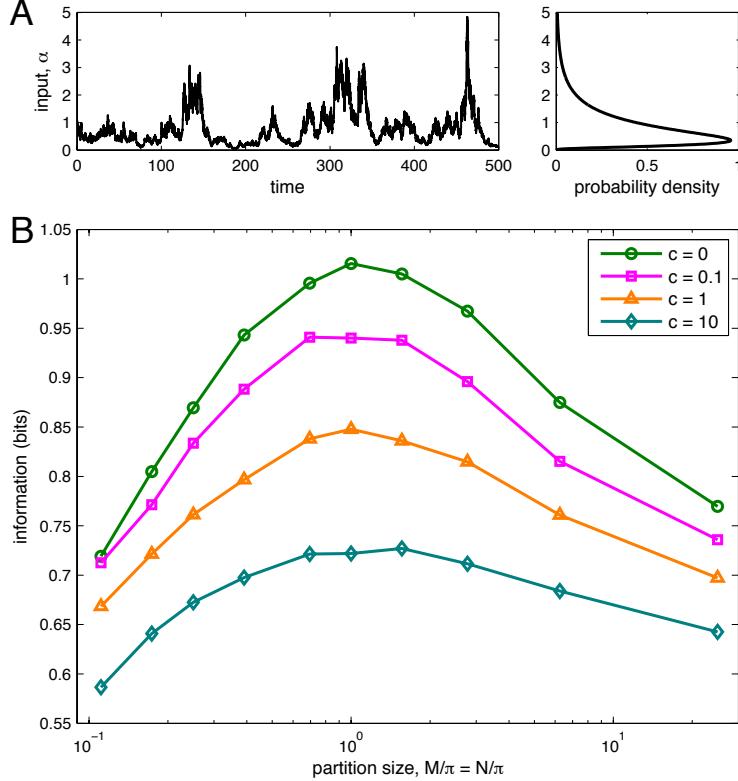


Figure S5: The effects of partitioning are robust to extrinsic noise.

A Simulations are performed with extrinsic noise introduced to the input parameter α . To keep $\alpha \geq 0$, the quantity $z \equiv \log \alpha$ is described by the simple mean-reverting Ornstein-Uhlenbeck process $dz = r(\mu - z)dt + \eta\sqrt{rdt}\xi$, where ξ is a Gaussian random variable with mean 0 and variance 1; this results in a log-normal distribution for α . The quantity $1/r$ is the autocorrelation time, and the choices $\mu = \log[\bar{\alpha}^3/(\bar{\alpha} + c)]/2$ and $\eta = \sqrt{2\log(1+c/\bar{\alpha})}$ ensure that the mean of α is $\bar{\alpha}$ and that the variance of α scales with the mean via $\sigma_\alpha^2 = c\bar{\alpha}$.

B As the magnitude of the extrinsic noise (set by c) increases, the information $I[\bar{\alpha}, n]$ decreases for all partition sizes, while the presence of an information-optimal partition size persists.

Here $M = N = 25$, $\beta = 20$, $\gamma = 1$, $p_{\text{hop}} = 0.001$, the system is $\lambda = 50$ lattice spacings squared, the ratio of diffusion to reaction propensities is $p_D/p_r = 1$, and $r = 1$ in units of the $X^* \rightarrow X$ reaction rate (which sets the timescale of switching). In A, $\bar{\alpha} = c = 1$ and time is scaled by $1/r$. In B, when $c = 0$, the information transmission is lower than that in Fig. 6C of the main text because $M = N$ is lower.

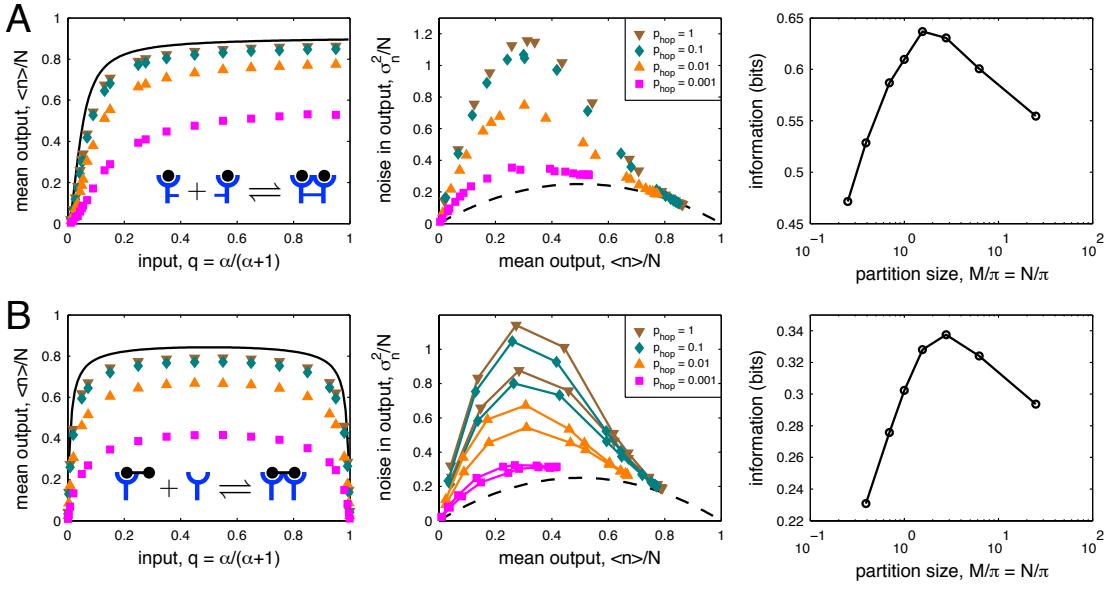


Figure S6: The effects of partitioning are robust to receptor dimerization. Two dimerization schemes are simulated, which are paradigmatic for receptor tyrosine kinases, including EGF receptor [10]: **A** Dimerization is receptor-mediated (left, inset), meaning two active receptors X^* form a complex C , or **B** dimerization is ligand-mediated (left, inset), meaning an active receptor X^* and an inactive receptor X form a complex C . The latter scheme admits a “dead-end” state at ligand saturation, when all receptors are ligand-bound and no complexes can form, leading to a non-monotonic response curve (B, left), as observed e.g. for the Ret receptor [11]. Both schemes are described by the reactions $X \xrightleftharpoons[1]{\alpha} X^*$, $C + Y \xrightarrow{\gamma\beta} C + Y^*$, and $Y^* \xrightarrow{\gamma} Y$, with dimer formation described by $X^* + X^* \xrightleftharpoons[\chi]{\epsilon} C$ in A, or $X^* + X \xrightleftharpoons[\chi]{\epsilon} C$ in B. Here $M = N = 25$, $\beta = 20$, $\chi = \gamma = 1$, the system is $\lambda = 50$ lattice spacings squared, and the ratio of diffusion to reaction propensities is $p_D/p_r = 1$. In A, $\epsilon = 20$; in B, $\epsilon = 5$.

Left As in Fig. 6A of the main text, as the probability of crossing a diffusion barrier p_{hop} is decreased, the maximal value of the mean response decreases. In A, the response also becomes more linear, but to less of a degree than in Fig. 6A of the main text. Note that due to both finite diffusion and finite molecule number, even the unpartitioned response ($p_{\text{hop}} = 1$) deviates from the mean-field response (black solid line), which is given by $\langle n \rangle/N = \beta f/(1 + \beta f)$, where f is the fraction of X molecules in the dimer state; in A, $f = \epsilon g^2$ with $g \equiv \langle m \rangle/M = (\sqrt{1 + 8\epsilon q^2} - 1)/(4\epsilon q)$, while in B, $f = \epsilon g(1 - g)/(2\epsilon g + 1)$ with $g \equiv \langle m \rangle/M = [\sqrt{1 + 8\epsilon q(1 - q)} - 1]/[4\epsilon(1 - q)]$. Here $\pi = 25$. Legends in middle panels apply to left panels as well.

Middle As in Fig. 6B of the main text, as the probability of crossing a diffusion barrier p_{hop} is decreased, the output noise decreases. Black dashed line shows the binomial noise limit $\sigma_n^2/N = (\langle n \rangle/N)(1 - \langle n \rangle/N)$. In B, lines connecting data points are provided to reveal that, as there are two values of q that give the same mean $\langle n \rangle/N$ (left), the noise is higher for the smaller value of q . Here $\pi = 25$.

Right As in Fig. 6C of the main text, the tradeoff between reliable signaling (reduced noise) and efficient signaling (maintaining a high maximal response) leads to an information-optimal partition size. Here $p_{\text{hop}} = 0.001$. Here, the information transmission is lower than that in Fig. 6C of the main text because $M = N$ is lower and additionally, in B, because of the non-monotonic mean response.

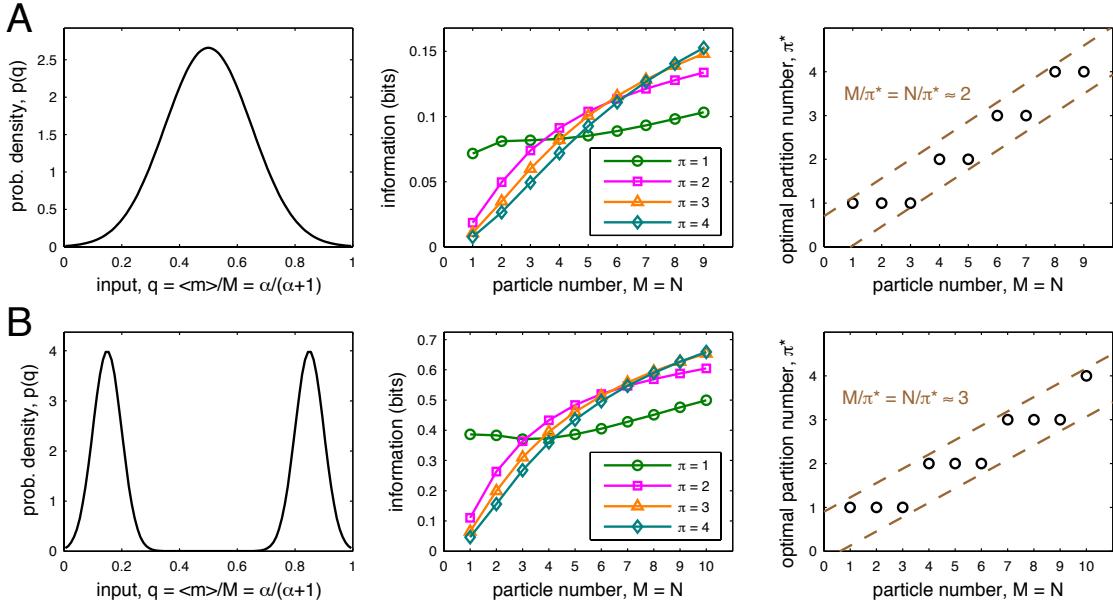


Figure S7: The effects of partitioning are robust to the shape of the input distribution. As in Fig. 5 of the main text, which takes a uniform input distribution $p(q)$, an information-optimal partition size $M/\pi^* = N/\pi^*$ persists with an input distribution that is **A** unimodal or **B** bimodal.

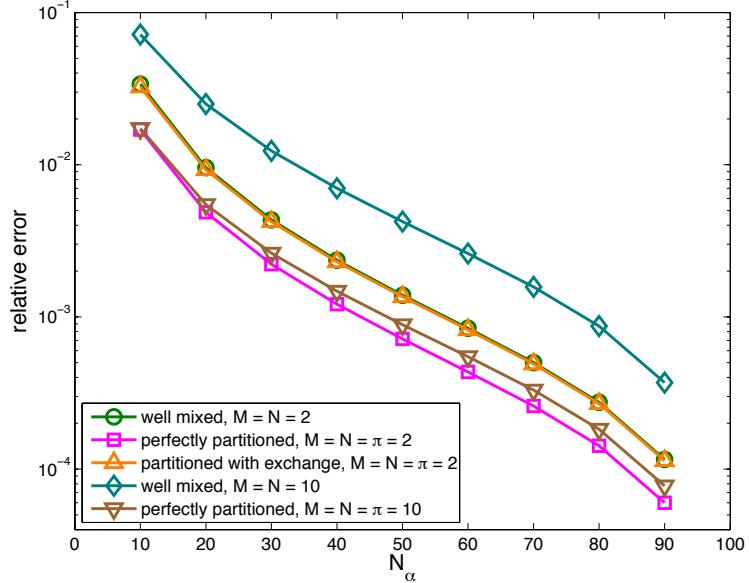


Figure S8: Computation of the mutual information converges as the input is more finely discretized. The relative error $|I - I_0| / I_0$, where I_0 is the information at $N_\alpha = 100$, is plotted against the number N_α of values of α [uniformly spaced in $q = \alpha/(\alpha + 1)$] used in the computation. Five conditions are tested, as indicated in the legend. It is seen that the relative error falls below $\sim 1\%$ in all conditions for $N_\alpha \gtrsim 30$.

References

- [1] A M Walczak, A Mugler, and C H Wiggins. A stochastic spectral analysis of transcriptional regulatory cascades. *Proc Natl Acad Sci USA*, 106:6529–6534, 2009.
- [2] A Mugler, A M Walczak, and C H Wiggins. Spectral solutions to stochastic models of gene expression with bursts and regulation. *Phys Rev E*, 80:041921, 2009.
- [3] C. E. Shannon. A Mathematical Theory of Communication. *Bell Syst Tech J*, 27:379–423, 623–656, 1948.
- [4] A Mugler, A M Walczak, and C H Wiggins. Information-optimal transcriptional response to oscillatory driving. *Phys Rev Lett*, 105:058101, 2010.
- [5] A M Walczak, A Mugler, and C H Wiggins. Analytic methods for modeling stochastic regulatory networks. In X Liu and M Betterton, editors, *Methods in Molecular Biology, Vol. 880: Computational Modeling of Signaling Networks*. Humana Press, 2012.
- [6] N G van Kampen. *Stochastic processes in physics and chemistry*. Elsevier Science, 2nd edition, 1992.
- [7] S J Farlow. *Partial differential equations for scientists and engineers*. Dover Publications, 1993.
- [8] J S Townsend. *A modern approach to quantum mechanics*. University Science Books, 2000.
- [9] D C Mattis and M L Glasser. The uses of quantum field theory in diffusion-limited reactions. *Rev Mod Phys*, 70:979–1001, 1998.
- [10] M A Lemmon and J Schlessinger. Cell signaling by receptor tyrosine kinases. *Cell*, 141:1117–34, 2010.
- [11] S Schlee, P Carmillo, and A Whitty. Quantitative analysis of the activation mechanism of the multi-component growth-factor receptor Ret. *Nat Chem Biol*, 2:636–44, 2006.