Stochastic multistability of clonal-like states in the Eigen model: A fidelity catastrophe

Emanuele Crosato [1,2,3,*] Richard E. Spinney [1,2,*] and Richard G. Morris [1,2,4,†]

[1]*School of Physics, UNSW, Sydney, New South Wales 2052, Australia*

[2]*EMBL Australia Node in Single Molecule Science, School of Biomedical Sciences, UNSW, Sydney, New South Wales 2052, Australia*

[3]*Living Systems Institute and Department of Mathematics and Statistics, University of Exeter, Exeter EX4 4QD, United Kingdom*

[4]*ARC Centre of Excellence for the Mathematical Analysis of Cellular Systems, UNSW Node, Sydney, New South Wales 2052, Australia*

(Received 20 August 2025; accepted 3 August 2026; published 16 September 2026)

The Eigen model is a prototypical toy model of evolution that is synonymous with the so-called error catastrophe: When mutation rates are sufficiently high, the genetic variant with the largest replication rate does not occupy the largest fraction of the total population because it acts as a source for the other variants. Here, we show that, in the stochastic version of the Eigen model, there is also a *fidelity* catastrophe. This arises due to the state-dependence of fluctuations and occurs when rates of mutation fall beneath a certain threshold, which we calculate. The result is a type of noise-induced multistability in which the system stochastically switches between short-lived regimes of effectively clonal behavior by different genetic variants. Most notably, when the number of possible variants—potentially $\sim 4^L$, with $L \gg 1$ the length of the genome—is significantly larger than the population size, there is only a vanishingly small interval of mutation rates for which the Eigen model is neither in the fidelity- nor error-catastrophe regimes.

DOI: 10.1103/dltj-jrj6

I. INTRODUCTION

Error-prone replication, or the copying of genetic sequences with occasional random mutations, is the primary form of genetic variation for asexual, single-celled organisms such as archaea and bacteria [1,2], a main thriving mechanism for viruses [3–5], and a source of inspiration for emerging engineering paradigms [6,7]. It is at the heart of Darwinian evolution and thought to have played a crucial role before the existence of higher functions such as recombination or horizontal gene transfer [8].

It was against this backdrop that Manfred Eigen, in the 1970s, introduced his eponymous model [9,10]. The Eigen model (EM) is a prototypical "toy" model of error-prone replication, invoking coupled ordinary differential equations (ODEs) to describe how the population fractions of different genetic variants change in time as they replicate, mutate, and deteriorate. The EM can be solved exactly [11] and has proven to be of conceptual importance; notions such as genotype clouds and quasispecies are widely attributed to the EM [12–14]. It is also synonymous with the so-called "error catastrophe": If mutation rates exceed a threshold in the EM, the fittest genetic variant—in the context of the model, the variant with the most rapid replication rate—does not occupy the largest fraction of the population.

Notably, the original formulation of the EM ignores the effects of fluctuations: It captures the deterministic $N \to \infty$ limit, where N is the total population size. As a result, several works have since explored the effects of stochasticity due to finite N [15–24]. These include (but are not limited to) the N-dependent scaling of the error threshold [15,17,18,22] and the existence of a dynamics that appears similar to that of "punctuated equilibria" [16,23].

However, conspicuously absent from these works are studies of so-called noise-induced phenomena [25–30], where finite-N fluctuations are state-dependent and lead to (typically multimodal) probability distributions whose extrema do not coincide with the fixed points of the system's deterministic, $N \to \infty$ dynamics. In other words, where a particular form of genetic drift—multiplicative demographic noise—acts to alter and reshape the extrema of the stationary distribution.

This omission is surprising for two reasons.

First, because such behaviors have been seen in the wider context of evolutionary dynamics, leading to quasicycles [31], selection "reversal" [32], and nonmonotonic rates of demographic switching under population growth [33,34].

Second, because the stochastic EM and asexual Wright-Fisher (W-F) diffusion have been shown to correspond to different limits of the same upstream stochastic process [35]; W-F diffusions are a paradigmatic family of models in population genetics that are generically characterized by state-dependent fluctuations [36–38] and therefore noise-induced behavior. This has been well characterized in low dimension,

*These authors contributed equally to this work.
†Contact author: r.g.morris@unsw.edu.au

such as when binary mutations are restricted to a single locus [36] and correspond to the so-called weak mutation limit: $\varepsilon \ll 1/N$, with ε the likelihood that a birth gives rise to a (point) mutation. The resulting dynamics involves a stochastic switching between two populations that are otherwise effectively clonal—i.e., they behave as if the other variant did not exist [36].

Despite this, noise-induced behavior in the (stochastic) EM has not yet been studied. One probable reason is that the model is manifestly high dimensional; while many genetic variants may not be viable (i.e., they carry zero or near-zero fitness), the number of possible states is nevertheless 4^L, where $L \gg 1$ is the genome length measured in base pairs (this can range from $\sim 10^6$ bp for small viruses to $\sim 10^{11}$ bp for eukaryotes). In contrast, and almost without exception, the detailed study of noise-induced effects has seemingly been restricted to cases where the number of variants, or species (dependent on the model), is $\leqslant 3$. This is because the typical approaches of directly analyzing state-dependent fluctuations—either visually or algebraically—are neither illuminating nor straightforwardly tractable in dimension >3.

Here, we proceed by calculating the nondispersive contribution to the probability current of an approximate Fokker-Planck equation (derived via a system-size expansion of the underlying master equation). We then investigate the zeros of this "force" when projected onto a given dimension of phase space. This allows us to identify a critical value for the rate of mutation above which the stationary distribution of the EM is unimodal, and below which it is multimodal. The latter regime, where the fidelity of replication is high, corresponds to a type of stochastic switching between clonal-like states. We term this a "fidelity catastrophe" in juxtaposition with the well-known error catastrophe: In both regimes, the fittest variant does not reliably occupy the largest fraction of the population, albeit with entirely different underlying mechanisms.

By replacing the generic notion of a weak mutation limit with an explicit threshold criterion (in terms of variant fitness, death rate, system size N, and number of variants m), we are able to make a surprising observation: In the limit that $m/N \to \infty$, which we might reasonably expect to characterize a large number of real systems, the fidelity catastrophe threshold coincides with that of the well-known error catastrophe. That is, the system either undergoes clonal switching *or* the fittest variant—the one with the most rapid replication rate—acts as a source for all other variants (due to high levels of mutation).

II. A STOCHASTIC EIGEN MODEL

The stochastic EM comprises a population of N genomes distributed across m variants, with N_i the number of genomes of variant $i \in \{1, \ldots, m\}$. The total number of variants is determined as $m = |\mathcal{A}|^L$, where L is the genome length expressed as a number of bases, which each take a value found in the alphabet $\mathcal{A}$, consisting of $|\mathcal{A}|$ different values (e.g., for DNA, $\mathcal{A} = \{C, G, A, T\}$, and $|\mathcal{A}| = 4$). The genomes replicate and deteriorate at average, variant-specific rates α_i and ω_i, respectively. During replication, a fraction μ_{ij} of the new genomes of variant i mutate into variant j, on average, with $\hat{\mu} = \mu_{ii}$ the fraction of genomes that do not mutate, which we refer to as the fidelity. The following reactions capture the population dynamics:

$$N_i \xrightarrow{\sum_j N_j \alpha_j \mu_{ji}} N_i + 1, \tag{1a}$$

$$N_i \xrightarrow{N_i \omega_i} N_i - 1. \tag{1b}$$

The only restrictions we place on the system, at this stage, are symmetry in the mutation rates, $\mu_{ij} = \mu_{ji}$, and equal fidelity across the variants, $\mu_{ii} = \hat{\mu}$, $\forall\, i$.

In the deterministic, $N \to \infty$, limit Eq. (1) can be approximated by coupled ODEs. These describe how the population converges, over time, on the demographic mix that corresponds to the largest eigenvalue of the linear algebra problem $W\mathbf{n} = \lambda\mathbf{n}$ [11], where $W_{ij} := \alpha_j \mu_{ji} - \omega_i \delta_{ij}$, and $\mathbf{n} := \{n_1, \ldots, n_m\}$ is the demographic mix in concentration space, i.e., $n_i := N_i/N$. At finite N, however, the behavior of the system significantly deviates from this solution.

Due to the relationship between the stochastic EM and W-F diffusions [35], as well as the prevalence of such behavior in evolutionary dynamics [31–34], we look for noise-induced bi- or multistability. That is, where intrinsic state-dependent fluctuations dictate the locations of the extrema (i.e., the peaks) of steady-state probability distributions [25]. Such fluctuations, and their corresponding noise-induced drifts, have size $\sim 1/\sqrt{N}$. As N decreases, their influence is typically marked by a transition from a unimodal distribution, peaked on a deterministic fixed point, to a multimodal distribution, peaked elsewhere. Such behavior has been observed in a variety of systems, ranging from idealized spins [27] and molecular autocatalysis [26,28] to collective motion in animals [29,30].

In the case of a W-F model in one dimension—i.e., binary, single locus—the switch from uni to bimodal is known as the weak mutation limit, which is characterized by $\mu \ll 1/N$. (Note: Despite being broadly adopted elsewhere, the terminology "noise-induced" is not used widely in the population genetics literature.) More specifically, the peak of the unimodal case corresponds to a mixed variant fixed-point, whereas the peaks of the bi-modal case correspond to populations where all individuals are from a single variant. As a result, the population switches stochastically between periods of behavior that is otherwise clonal.

III. EFFECTIVE FORCES

To capture the role of noise-induced drifts in the stochastic EM and, in particular, to understand when they compete with deterministic drifts, we focus on the notion of "effective" force. That is, the advective, nondispersive contribution to the probability flux, or current. (In a physical system, such as a particle subject to thermal fluctuations, this has the dimensions of a force.) As we will show, it turns out that the effective force is a helpful object when considering higher-dimensional systems.

Our starting point is the Master equation specified by Eq. (1). We perform a system size expansion [39,40] (Appendix A), which results in a set of Itô stochastic differential equations (SDEs) in $\{\mathbf{n}, N\} := \{n_1, \ldots, n_m, N\}$:

$$\mathrm{d}n_i = (\Phi_i(\mathbf{n}) + \Psi_i(\mathbf{n}))\,\mathrm{d}t + \frac{1}{\sqrt{N}} \sum_{j=1}^{m} q_{ij}(\mathbf{n})\,\mathrm{d}W_j, \tag{2}$$

$$\mathrm{d}N = N(\langle\alpha\rangle_{\mathbf{n}} - \langle\omega\rangle_{\mathbf{n}})\,\mathrm{d}t + \sqrt{N(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}})}\,\mathrm{d}W, \tag{3}$$

where

$$\Phi_i(\mathbf{n}) := A_i(\mathbf{n}) - n_i \sum_{j=1}^{m} A_j(\mathbf{n}), \tag{4}$$

$$\Psi_i(\mathbf{n}) := -N^{-1}\left(B_i(\mathbf{n}) - n_i \sum_{j=1}^{m} B_j(\mathbf{n}) \right), \tag{5}$$

with

$$A_i(\mathbf{n}) := \sum_{j=1}^{m} n_j \alpha_j \mu_{ji} - n_i \omega_i, \tag{6}$$

$$B_i(\mathbf{n}) := \sum_{j=1}^{m} n_j \alpha_j \mu_{ji} + n_i \omega_i. \tag{7}$$

Of note, the multiplicative noise strengths that appear in Eq. (2) have the form

$$q_{ij}(\mathbf{n}) := \begin{cases} (1-n_i)\sqrt{B_i(\mathbf{n})}, & i = j, \\ -n_i\sqrt{B_j(\mathbf{n})}, & i \neq j. \end{cases} \tag{8}$$

In the above, angle brackets $\langle\cdot\rangle_{\mathbf{n}}$ represent the mean with respect to the population fractions, such that $\langle\alpha\rangle_{\mathbf{n}} = \sum_{i=1}^{m} n_i\alpha_i$ and $\langle\omega\rangle_{\mathbf{n}} = \sum_{i=1}^{m} n_i\omega_i$. Meanwhile, $\mathrm{d}W_i$ are increments of independent Wiener processes, such that $\mathbb{E}[\mathrm{d}W_i \mathrm{d}W_j] = \delta_{ij}\mathrm{d}t$, where δ_{ij} is the Kronecker delta, and $\mathbb{E}[\cdot]$ represents an expectation over the noise. We note that the total noise experienced by each n_i is thus highly correlated, originating from the constraint $\sum_{i=1}^{m} n_i = 1$. Finally, the Wiener increment $\mathrm{d}W$ (no subscript) has correlation $\mathbb{E}[\mathrm{d}W\,\mathrm{d}W_i] = m^{-1/2}\mathrm{d}t$, for all i.

While describing the totality of this system is highly nontrivial, we can nevertheless characterize the effect of population size (and hence stochasticity) by studying the instantaneous dynamics for the n_i at arbitrary, fixed N. Due to the constraint $\sum_{i=1}^{m} n_i = 1$, this restricts us to an $m-1$-dimensional simplex and, as a consequence, the system is overdetermined, i.e., $p(n_1, \ldots, n_m) = \tilde{p}_{\setminus i}(n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_m)\delta(n_i - (1 - \sum_{k\neq i} n_k))$, for all i. This is reflected by a noninvertible diffusion matrix $\mathsf{Q} := \mathsf{q}\mathsf{q}^T = [Q_{ij}]$, based upon noise strength matrix $\mathsf{q} = [q_{ij}]$. To circumvent this, we focus on the Fokker-Planck equation (FPE) that corresponds to Eq. (2), but, without loss of generality, with the mth dimension removed (Appendix B). This leads to a reduced FPE that can be written

$$\dot{\tilde{p}}(\tilde{\mathbf{n}}, t) = -\sum_{i=1}^{m-1} \partial_{n_i}\Bigg[(\Phi_i(\tilde{\mathbf{n}}) + \Psi_i(\tilde{\mathbf{n}}) + \Theta_i(\tilde{\mathbf{n}}))\tilde{p}(\tilde{\mathbf{n}}, t) - \frac{1}{2N}\sum_{j=1}^{m-1} \mathcal{Q}_{ij}(\tilde{\mathbf{n}})\partial_{n_j}\tilde{p}(\tilde{\mathbf{n}}, t)\Bigg]. \tag{9}$$

Here, $\tilde{p}(\tilde{\mathbf{n}}, t) \equiv \tilde{p}_{\setminus m}$ is the probability density over population densities $\tilde{\mathbf{n}} = \{n_1, \ldots, n_{m-1}\}$, while $\mathcal{Q}_{ij}$ are elements of the invertible submatrix $\mathcal{Q} := [\mathsf{Q}]_{1:m-1;1:m-1} = [\mathcal{Q}_{ij}]$. Meanwhile, the Θ_i term is given by

$$\Theta_i(\tilde{\mathbf{n}}) := -\frac{1}{2N}\sum_{j=1}^{m-1} \partial_{n_j}\mathcal{Q}_{ij}(\tilde{\mathbf{n}}), \tag{10}$$

which we refer to as the *noise-induced effective force*. Importantly, for Eq. (9) to be a closed equation in terms of $\tilde{\mathbf{n}}$ we must have every appearance of n_m replaced by $1 - \sum_{j=1}^{m-1} n_j$, for example, $\Phi_i(\tilde{\mathbf{n}}) = \Phi_i(\mathbf{n})|_{n_m = 1 - \sum_{j=1}^{m-1} n_j}$.

The presentation of Eq. (9) deliberately casts the total flux—i.e., the terms inside the square brackets—as being formed from both an advective term, or *effective force*, which we write $F_i(\tilde{\mathbf{n}}) = \Phi_i(\tilde{\mathbf{n}}) + \Psi_i(\tilde{\mathbf{n}}) + \Theta_i(\tilde{\mathbf{n}})$, and a remaining dispersive component. The meaning of such contributions becomes clear if we consider the limiting distribution ($\dot{\tilde{p}} = 0$), under the assumption of zero stationary flux, such that we can write, in vector form,

$$\frac{\nabla\tilde{p}(\tilde{\mathbf{n}})}{\tilde{p}(\tilde{\mathbf{n}})} = 2N\mathcal{Q}^{-1}(\tilde{\mathbf{n}})F(\tilde{\mathbf{n}}). \tag{11}$$

Here, $\nabla = (\partial_{n_1}, \ldots, \partial_{n_{m-1}})$, and $\Phi(\tilde{\mathbf{n}}) = (\Phi_1(\tilde{\mathbf{n}}), \ldots, \Phi_{m-1}(\tilde{\mathbf{n}}))$ etc., such that $F(\tilde{\mathbf{n}}) = \Phi(\tilde{\mathbf{n}}) + \Psi(\tilde{\mathbf{n}}) + \Theta(\tilde{\mathbf{n}})$ represents the vector of the *total effective force*, which is proportional to the advective part of the total flux. Such a quantity can be used as a proxy for the expected behavior, or drift, of the system in any given population state, $\tilde{\mathbf{n}}$. For example, in the above case, where the total flux vanishes in the stationary state, extrema in $\tilde{p}(\tilde{\mathbf{n}})$ coincide with the zeros of $F(\tilde{\mathbf{n}})$. These therefore act as effective dynamical fixed points, which may be stable or unstable if they correspond to maxima and minima in the distribution, respectively.

Generically, however, we cannot guarantee that the stationary flux vanishes in the (stochastic) EM. We nevertheless conjecture that transition to switching behavior—i.e., from unimodality to multimodality—will be captured by the appearance (or disappearance) of zeros in $F(\tilde{\mathbf{n}})$. If true, this weaker property is still helpful because, while solution of the FPE is, in general, not possible, we can readily derive $F(\tilde{\mathbf{n}})$ (given in full in Appendix C).

IV. DIMENSIONAL REDUCTION

Since $F(\tilde{\mathbf{n}})$ is a vector field in $m-1$ dimensions, with m potentially very large, its investigation is both visually and analytically complicated (Appendix C). To simplify the analysis, we choose to inspect $F(\tilde{\mathbf{n}})$ along lines that bisect the center of the simplex and its vertices, characterized by co-ordinates $\mathbf{n}^{(c)}(x) = (n_1^{(c)}(x), \ldots, n_m^{(c)}(x))$, $x \in [0, 1]$, with

$$n_i^{(c)}(x) = \begin{cases} x, & i = c, \\ \frac{1-x}{m-1}, & i \neq c, \end{cases} \tag{12}$$

where c is a specifically chosen coordinate determining the bisection line. Explicitly, $\mathbf{n}^{(c)}(1)$ corresponds to a system composed only of variant c, while $\mathbf{n}^{(c)}(0)$ corresponds to a system completely absent variant c, and with all other variants present in equal proportion. The construction is shown in Fig. 1. Of note, when constrained to a bisecting line, the value below which a concentration has to fall in order to be smaller than that of the other variants is $1/m$.

We not only inspect $F(\tilde{\mathbf{n}})$ on the line $\mathbf{n}^{(c)}(x)$, but also *in the direction* of this line, which is given by the unit vector $\mathbf{u}^{(c)}$

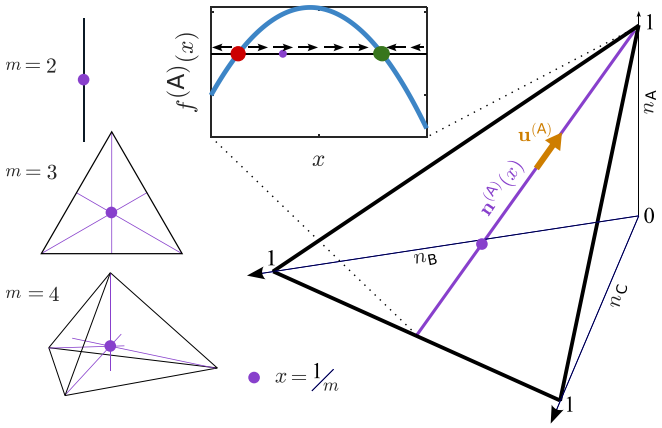

FIG. 1. Dimensional reduction. On the left, three simplexes are shown for the cases of $m = \{2, 3, 4\}$. The center of a simplex, where all bisecting lines (shown in purple) meet, corresponds to $n_i = n_j = 1/m \ \forall \ i, j$, where all variants have the same concentration. When inspecting the concentration of a "chosen" variant, $n_c^{(c)}$, along its bisecting line (i.e., the one that terminates in the corner of the simplex where $n_c^{(c)} = 1$) then, for that variant to have a smaller fraction of the population than the others, we must have $n_c^{(c)} < 1/m$. On the right, the case of $m = 3$ is shown in more detail. The three variants are labeled A, B, and C, with the purple line representing $\mathbf{n}^{(\mathsf{A})}(x)$, bisecting the A corner and the center of the simplex. Note that $\mathbf{n}^{(\mathsf{A})}(0)$ corresponds to a population with no A genomes, while $\mathbf{n}^{(\mathsf{A})}(1)$, i.e., the A corner itself, to a population of only A genomes. The orange arrow, arbitrarily positioned along the bisecting line, represents $\mathbf{u}^{(\mathsf{A})}$ [Eq. (13), not to scale]. Finally, the inset illustrates how projected forces over x are visualized.

pointing toward the corner $\mathbf{n}^{(c)}(1)$ of the simplex:

$$u_i^{(c)} = \begin{cases} \sqrt{\frac{m-1}{m}}, & i = c, \\ -\sqrt{\frac{m-1}{m}}\frac{1}{m-1}, & i \neq c. \end{cases} \tag{13}$$

To characterize the effective force along the simplex bisection line, we are ultimately interested in the scalar projection

$$f^{(c)}(x) = F(\mathbf{n}^{(c)}(x)) \cdot \mathbf{u}^{(c)} = \sum_i F_i(\mathbf{n}^{(c)}(x)) u_i^{(c)}, \tag{14}$$

and its partition into $f^{(c)}(x) = f_\Phi^{(c)}(x) + f_\Psi^{(c)}(x) + f_\Theta^{(c)}(x)$.

It is then instructive to consider the effective one-dimensional (1D) dynamics due to $f^{(c)}(x)$. To do so, we calculate exact expressions for each of the three contributions—$f_\Phi^{(c)}(x)$, $f_\Psi^{(c)}(x)$, and $f_\Theta^{(c)}(x)$—in the general case, in Appendix D. Crucially, all such components are quadratic functions of the parametrization variable, x. This means that there can be up to two zeros in the projected effective force within the relevant interval, [0, 1]. That is, up to two fixed points in the effective dynamics. In the case of stationary flux, such fixed points can be identified with the extrema of the stationary distribution: A stable fixed point corresponds to a maximum and an unstable fixed point to a minimum. While we cannot formally make this identification here, due to the absence of stationary flux, we assert (and later verify) that a *change* in the number of zeros within [0, 1] will still be indicative of a change in the number of peaks in the stationary distribution.

V. QUASISPECIES "TOY" MODEL

To illustrate these ideas, we consider the canonical "toy model" of quasispecies [15,18]: A single dominant "wild-type" (WT) variant, z, that replicates faster than all other variants, such that $\omega_i = \omega \ \forall i$, $\alpha_i = \alpha \ \forall i \neq z$, $\alpha_z = \alpha + \Delta\alpha$, with $\Delta\alpha > 0$. By assuming a fitness landscape that is flat for all variants other than the WT, this model grossly oversimplifies real fitness landscapes and likely overestimates the number of viable mutations. It is, nevertheless, useful for illustrating our arguments, and provides a convenient sandbox for comparison with the error catastrophe.

The salient properties of this toy system are captured by considering only the force projection onto the WT's bisection line, i.e., $c = z$ (an analysis of identically fit variants as well as non-WT variants is provided for completeness in Appendixes E and F). In this case, the effective force is a negative quadratic in x, illustrated qualitatively in Fig. 1. Generically, the upper fixed point is stable, while the lower fixed point is unstable. However, crucially, the position of these fixed points—including whether they lie inside or outside of the simplex—varies with system parameters. In particular, the variation with the fidelity ($\hat{\mu} = \mu_{ii}$) is illustrated in Fig. 2(a). Here, as the fidelity of the system is increased, the stable fixed point moves toward $x = 1$ (purely WT population) as might be expected. However, this also moves the unstable fixed point in the same direction. At high enough values of $\hat{\mu}$ (or conversely at low enough values of N), the unstable fixed point enters the interval [0, 1]. This can be attributed to the action of the noise-induced component of the effective projected force, $f_\Theta^{(z)}(x)$. When this component is removed, such as in the deterministic $N \to \infty$ limit, the unstable fixed point does not enter the simplex, irrespective of $\hat{\mu}$ [Figs. 2(b) and 2(c)].

Notably, when the unstable fixed point lies within the simplex, a qualitatively new dynamics can take place: Any state whose projection lies between that of the unstable fixed point and $x = 0$ will be driven toward $x = 0$ (and not $x = 1$). That is, if the system comprises only a small fraction of WTs, it will be driven toward states where there are none, despite the WT being the fittest variant. In terms of the probability distribution, this corresponds to the appearance of local minima, causing an otherwise unimodal distribution to become multimodal.

Mechanistically, such behavior is the result of a positive feedback process. To see this, consider a random fluctuation that leads to an arbitrary non-WT variant possessing a larger population fraction than all the others. When N is large, there is an overwhelming likelihood that this difference is small, and the system will respond by reverting to the mean. However, when N is small, such a fluctuation can lead to substantive differences in population fractions. Here, since there is a fixed birth rate *per individual*, more individuals of that variant are likely to be produced, relative to the others, causing it to possess a yet larger fraction of the population. This feedback leads to the temporary fixation of the random variant—or so-called clonal behavior—and therefore depletion of the WT. Subsequent fluctuations, however, can then start the process all over again, leading to intermittent stochastic switching between periods of clonal-like behavior of random single variants.

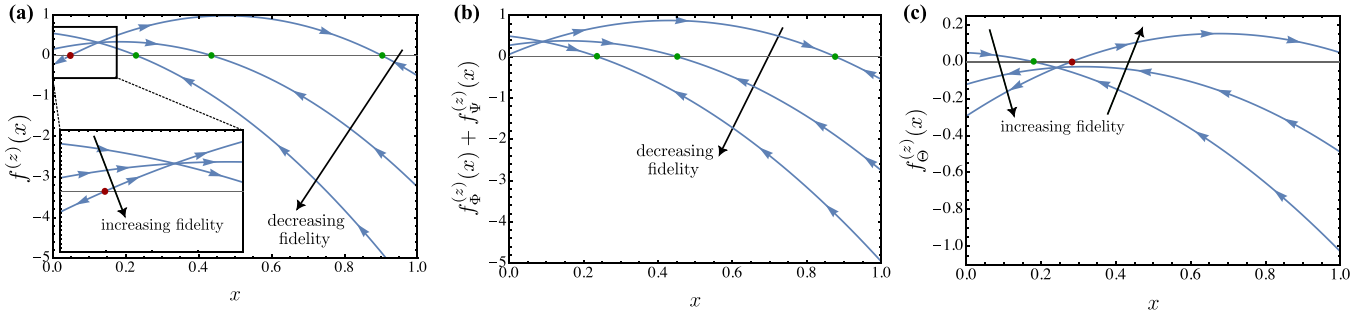

FIG. 2. Effective force and its fixed points: the role of noise-induced drifts. Projection of effective forces on the simplex bisection line of the WT variant. Across all panels, $m = 3$ and $N = 10$, $\alpha = 1$, $\Delta\alpha = 4$, and $\omega = 0.5$. (a) The total effective force, $f^{(z)}(x)$, at different fidelities, $\hat{\mu}$, taking values 0.1, 0.5, and 0.9. As the fidelity is increased, the lower fixed point (red dot) enters the simplex such that the force at $x = 0$ becomes negative. On the other hand, as the fidelity is reduced (error is increased), the upper, stable, fixed point (green dots) moves to lower values of x in the simplex. (b) The deterministic component of the effective force for the same values of $\hat{\mu}$: There is no unstable fixed point in any of the three cases. (c) The noise-induced component of the effective force for the same values of $\hat{\mu}$: For high fidelity, an unstable fixed point arises. For suitably low N, this can contribute sufficiently to the total force that such a fixed point appears in the total force [cf. panel (a)].

The way that this manifests, in a large-but-finite-N stochastic description of population fractions, is via fluctuations that are state-dependent. To see this, consider a state where a single variant is dominant—e.g., $\mathbf{n} = \{1, 0, \ldots, 0\}$. Here, the stochastic effects of birth and death do not lead to fluctuations. In contrast, in a fully mixed state—e.g., $\mathbf{n} = \{1/m, 1/m, \ldots, 1/m\}$—each birth and death event changes the relative composition of the system and hence generates fluctuations. This gradient in fluctuations, from mixed states (i.e., the center of the simplex) to single-variant states (i.e., the corners of the simplex), results in noise-induced forces that codify the above feedback mechanism in terms of a multimodal distribution peaked on single-variant states.

VI. A FIDELITY CATASTROPHE

We term the transition to a clonal switching regime a "fidelity catastrophe" in juxtaposition with the well-known error catastrophe; despite relying on different mechanisms, the salient feature of both regimes is that the WT variant does not reliably occupy the largest fraction of the population.

By identifying the parameter values at which the unstable fixed point enters the simplex, we can characterize a threshold for this behavior. Specifically, the unstable fixed point enters the simplex when $f^{(c)}(x = 0) = 0$, such that it lies exactly on the edge of the simplex. By solving for this condition in terms of $\hat{\mu}$, we thus determine the critical fidelity above which the distribution becomes multimodal, and below which the distribution remains unimodal (see Appendix F)

$$\hat{\mu}^{\text{fidel}}_{\text{crit}} = \frac{\alpha(m + 2N - 1) - m\omega + \omega}{\Delta\alpha(m - 1) + 2\alpha(m + N - 1)}. \tag{15}$$

Equivalently, we may fix $\hat{\mu}$ and calculate the critical N above which the distribution is unimodal, and below which it is multimodal, yielding

$$N^{\text{fidel}}_{\text{crit}} = \frac{(m - 1)(\alpha(2\hat{\mu} - 1) + \Delta\alpha\hat{\mu} + \omega)}{2\alpha(1 - \hat{\mu})}. \tag{16}$$

Here, we note that $N^{\text{fidel}}_{\text{crit}} \sim m$. In other words, the sense in which N can be considered small, such that there are noise-induced effects, grows with the dimensionality of the system.

To test our predictions, we simulate the system described by Eq. (1) using the Gillespie algorithm [41]. We adopt the standard setup, such that the mutation matrix, $\mu = [\mu_{ij}]$, has elements

$$\mu_{ij} = \left(\frac{\varepsilon}{|\mathcal{A}| - 1}\right)^{H_{ij}} (1 - \varepsilon)^{L - H_{ij}}, \tag{17}$$

where $\varepsilon = 1 - \hat{\mu}^{1/L}$ is the average replication error rate per base and H_{ij} is the Hamming distance—i.e., the number of loci where the genomes of i and j have different bases. In addition, in order to characterize the system at a given N, the population size is kept stable around its initial value $N(0)$ by adding a saturation limit (see Appendix H). Further, we make the selective advantage, $\Delta\alpha$, a function of ε, so that the deterministic solution (i.e., the composition of dominant quasispecies in the $N \to \infty$ limit) is approximately equal across all ε (Appendix H). This has several benefits. It shows that equivalent deterministic solutions to the Eigen model can differ drastically at finite N, and ensures that statistics of the distribution (i.e., β; see Fig. 3) can be meaningfully compared across the parameter space, since lines of β at fixed ε manifest as translations of the same approximate functional form. Finally, it also ensures that waiting times in clonal states are kept within a range such that sampling the limiting solution is computationally tractable.

For the simulations to remain numerically viable, we focus on the case of $m = 16$ variants ($L = 4$ and $|\mathcal{A}| = 2$), exploring behavior at various combinations of population size and error rate in a regime where the error catastrophe is avoided (Fig. 3). Two drastically different behaviors are observed: Either the concentrations fluctuate over time around the deterministic solution [Fig. 3(a)], or individual variants randomly take turns at dominating the entire population for long stretches of time, regardless of their Hamming distance to the WT [Fig. 3(b)]. The distribution of the concentration of the WT over time, $\rho(n_z)$, identifies the two regimes corresponding to observed behaviors [Fig. 3(a)]: At higher error rate (low fidelity) and larger population size, $\rho(n_z)$ is unimodal and peaked on the value predicted by the deterministic model, while at lower error rate (high fidelity) and smaller population size, $\rho(n_z)$ is

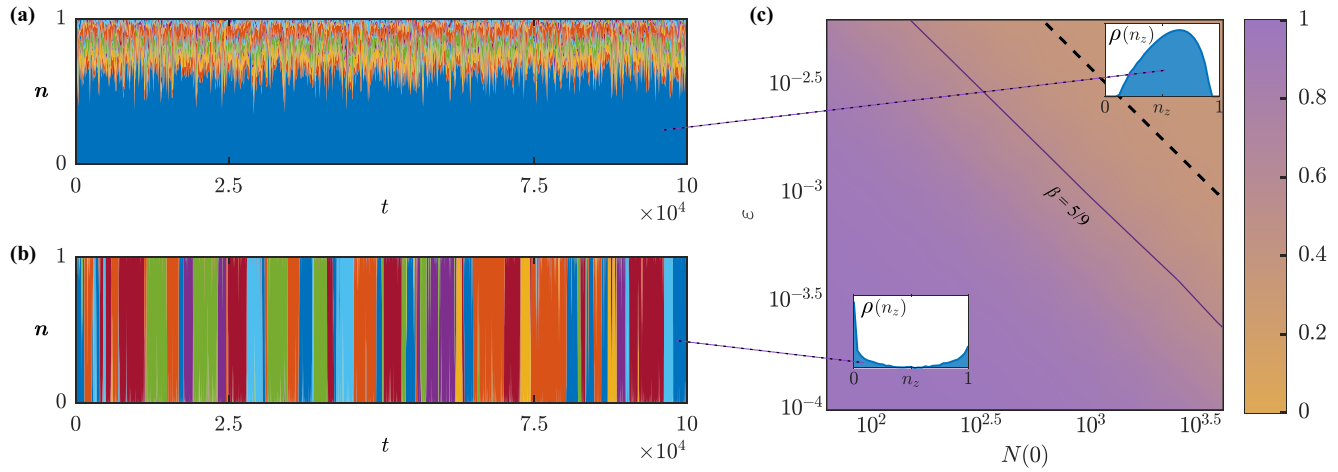

FIG. 3. Numerical evidence for the fidelity catastrophe. Gillespie simulations of Eq. (1), coupled with a saturation constraint (see Appendix H), for different combinations of ε and N. Two distinct behaviors can be observed. Either the WT reliably corresponds to the largest fraction of the population, subject to small fluctuations in time [panel (a)], or, the population stochastically switches between different clonal-like periods, where the entire population comprises only a single variant [panel (b)]. The former arises from a unimodal probability distribution over the WT population fraction, $\rho(n_z)$, while the latter arises from a bimodal distribution [panel (c), insets, logarithmic y axis]. Quantifying bimodality in terms of the coefficient $\beta = (\text{skew}[\rho(n_z)]^2 + 1)/\text{kurt}[\rho(n_z)]$ allows us to characterize the full phase space in terms of the per-base error ϵ and the $t = 0$ system size $N(0)$ [panel (c)]. This empirical approach agrees semiquantitatively with our analytical predictions. Evaluating Eq. (15) results in the black dashed line, marking the onset to the fidelity catastrophe and clonal-like switching (purple region). This has the same slope as the solid purple line, which marks the points where $\beta = 5/9$ (i.e., the value of β associated with a uniform distribution: a proxy for the transition between uni- and multimodality) [panel (c)]. Parameter values used: $|\mathcal{A}| = 2$, $L = 4$, $\omega_i = 1\ \forall i$, $\alpha_i = 1\ \forall i \neq z$, and μ_{ij} as per Eq. (17). We assume $\alpha_z = \alpha(1 + 10\varepsilon)$ so that $\lim_{N\to\infty} \rho(n_z)$ is independent of ε, and to ensure that waiting times in clonal states are computationally tractable (see Appendix H). Note that plots of $\rho(n_z)$ that appear in panel (c) are presented with a logarithmic axis.

bimodal with peaks at $n_z = 0$ and $n_z = 1$. While the former characterizes expected behavior of a quasispecies, the latter captures the fidelity catastrophe regime.

To compare with our analytics, we substitute $\hat{\mu}$ with ε in either Eq. (15) or Eq. (16) to obtain a line that bounds the fidelity catastrophe regime. The result yields the dashed black line in Fig. 3(c), while the solid purple line marks a nominal demarcation of unimodal and multimodal empirical probability distributions. We observe that our prediction well captures the trend of the transition between unimodality and bimodality, with weak mutation switching dynamics occurring at large $\hat{\mu}$ (low ε) where the unstable fixed point resides in the simplex. However, quantitatively, its position is overestimated compared to the nominal unimodal/multimodal line. Such a discrepancy could be attributed to several factors including limitations of the system size expansion that underlies Eq. (15), or the necessity for the unstable fixed point to be farther into the simplex for its effects to be noticeable.

VII. REVISITING THE ERROR CATASTROPHE

In light of the fidelity catastrophe, it is instructive to revisit the *error* catastrophe. This involves generalizing the classical, deterministic criterion for its onset to the stochastic case.

To do this, we focus on the *stable* fixed point of the generalized force. Generically, this has two types of response to changes in $\hat{\mu}$ [cf. Fig. 4(a)]: Either the response is approximately linear, or it is relatively independent of $\hat{\mu}$ and lies near $x = 1/m$ (see Appendix F). Notably, as m is taken larger, the transition between these two regimes becomes sharp. Consequently, we choose to characterize a threshold as the point of maximal curvature in the position of the fixed point $x_*^{(z)}$ with respect to $\hat{\mu}$, i.e., $\mathrm{d}^3 x_*^{(z)}(\hat{\mu})/\mathrm{d}\hat{\mu}^3 = 0$. Since the location of the fixed point is the solution of a quadratic equation, this has an analytical solution $\hat{\mu}_{\text{crit}}^{\text{error}}$, reported in Appendix F. Such an expression amounts to a powerful generalization of the original heuristic formula for the error threshold, containing both population and phase space/genome size dependence through N and m, respectively.

The error threshold places the opposite bound on the system in terms of the fidelity, $\hat{\mu}$, as that of the fidelity catastrophe, in that for the error catastrophe to be avoided we must have $\hat{\mu} \geqslant \hat{\mu}_{\text{crit}}^{\text{error}}$, while for the fidelity catastrophe, multimodal regime to be avoided we must obey $\hat{\mu} \leqslant \hat{\mu}_{\text{crit}}^{\text{fidel}}$. The result of this is that there is a characteristic "window," where $\hat{\mu}_{\text{crit}}^{\text{error}} \leqslant \hat{\mu} \leqslant \hat{\mu}_{\text{crit}}^{\text{fidel}}$, in which the system avoids both scenarios. This is illustrated in Figs. 4(a)–4(c) where the behavior of the two fixed points, and their relationship to the two thresholds, is clearly demonstrated.

Despite the dependence of our $\hat{\mu}_{\text{crit}}^{\text{error}}$ on m and N, the error threshold originally postulated by Eigen is easily recovered by progressively considering the deterministic ($N \to \infty$), and then large state space ($m \to \infty$), limits, obtaining $\alpha/(\alpha + \Delta\alpha)$ for our quasispecies example [9,10].

Interestingly, however, the location of this threshold changes if we consider the limits to be taken in the opposite order, such that we consider $m \gg N$. In such a regime, the generalized error threshold converges on $(\alpha - \omega)/(2\alpha + \Delta\alpha)$. Moreover, this is precisely the value that Eq. (15) takes in the same limit. Explicitly, in the limit $m \gg N$ the thresholds

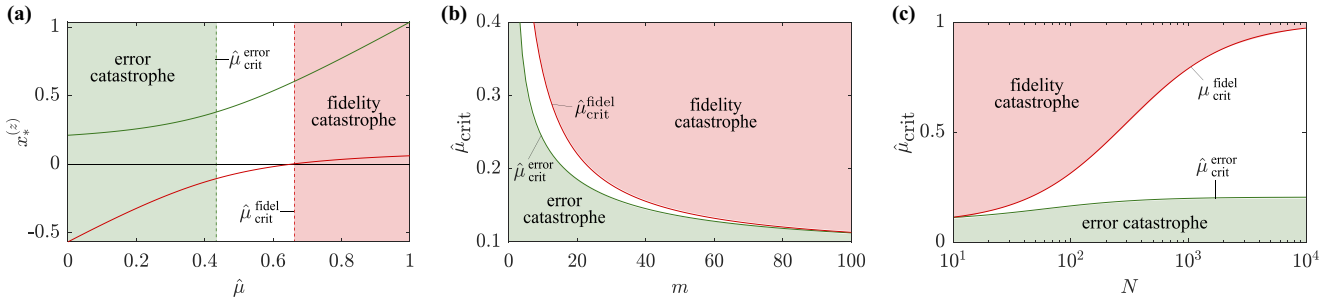

FIG. 4. The fidelity catastrophe vs the error catastrophe. Fixed points in the total effective force along the simplex bisection line of the WT variant for $\alpha = 1$, $\Delta\alpha = 4$, and $\omega = 0.5$. (a) With $m = 3$ and $N = 10$, the green and red lines show the position of the stable and unstable fixed point, respectively, as a function of $\hat{\mu}$. The critical values of $\hat{\mu}$ marking the onset of the error catastrophe and fidelity catastrophe are highlighted by dashed lines. The onset of the fidelity catastrophe is defined as the point where the unstable fixed point enters the simplex at $x = 0$, while the onset of the error catastrophe is defined by the point of maximum curvature in the location of the stable fixed point with $\hat{\mu}$. (b) Setting $N = 10$, the positions of the thresholds are shown as a function of m. (c) Setting $m = 100$, the positions of the thresholds are shown as a function of N.

to the error and fidelity catastrophes converge to the same value, with their ratio obeying

$$\frac{\hat{\mu}^{\text{fidel}}_{\text{crit}}}{\hat{\mu}^{\text{error}}_{\text{crit}}} = 1 + \frac{2\Delta\alpha}{2\alpha + \Delta\alpha}\frac{N}{m} + \mathcal{O}(m^{-1}). \tag{18}$$

This suggests that the "window" of possible $\hat{\mu}$ for which both the error catastrophe and fidelity catastrophe are avoided vanishes in this limit. In other words, in our quasispecies example, should the number of variants far exceed the number of replicators, in order to avoid the error catastrophe, one must be in the fidelity catastrophe regime such that the distribution is multimodal.

To understand the importance of the ratio N/m, we consider the position of the fidelity thresholds when varying m at fixed N [Fig. 4(b)]. As m increases, both thresholds move to lower values of $\hat{\mu}$, and we can clearly see the prediction of Eq. (18): The two thresholds tend to each other, reducing the fidelity interval where neither the error catastrophe nor fidelity catastrophe behavior is present. In contrast, the effect of varying N at fixed m is shown in Fig. 4(c). Here, as N increases, the position of the two thresholds moves to higher values of $\hat{\mu}$, as they also tend away from each other, thus increasing the fidelity interval where neither behavior is present. Here, we can also appreciate that the fidelity catastrophe is clearly a stochastic effect, with larger and larger fidelities (tending to 1) required to observe the effect as N is taken larger.

In totality, the central observation that underlies the behavior is the importance of the ratio N/m. In particular, we argue that it refines the idea of what constitutes a "low N" system, wherein stochastic effects, such as noise-induced multistability, become non-negligible. Crucially, when this ratio is large [low-m side of Fig. 4(b) and high-N side of Fig. 4(c)] there are a large number of replicators *per variant* and thus stochastic effects are small. In this limit, the fidelity catastrophe behavior only occurs for extremely large values of $\hat{\mu}$ ($\sim$1) and generally does not occur in the vicinity of the error catastrophe. In contrast, when N/m is small [high-m side of Fig. 4(b) and low-N side of Fig. 4(c)] there are, on average, very few replicators *per variant* (i.e., significantly less than one) and thus stochastic effects are highly significant. Indeed, we can directly quantify the critical ratio, $\xi := N/m$, below which we can expect noise-induced multistability by inspection of Eq. (16), giving

$$\xi^{\text{fidel}}_{\text{crit}} = \frac{m-1}{m}\frac{(\alpha(2\hat{\mu}-1) + \Delta\alpha\hat{\mu} + \omega)}{2\alpha(1-\hat{\mu})}, \tag{19}$$

which becomes independent of N and m, as $m \to \infty$, such that it is solely the ratio, ξ, which determines whether noise-induced effects arise. That is, we have

$$\lim_{m\to\infty} \hat{\mu}^{\text{fidel}}_{\text{crit}} = \frac{2\alpha\xi + \alpha - \omega}{2\alpha(\xi+1) + \Delta\alpha}. \tag{20}$$

It is then not irrelevant that for realistic models of genomes and populations, the number of variants m will be extremely large (much larger than N) such that $\xi \ll 1$ and fluctuations become important. When interpreted at face value, this implies that realistic systems cannot avoid both the error catastrophe and the fidelity catastrophe simultaneously.

The heuristic here is that, for large m, the fittest state occupies a vanishingly small fraction of phase space. Moreover, if $m \gg N$, then the majority of variants have zero population fraction. In this case, backmutation is highly unlikely (see discussion for a comparison with Muller's ratchet). Therefore, with sufficient mutation rates (low fidelity), the population becomes sparsely distributed among a sea of equally "unfit" variants away from the fittest state (error catastrophe), while at low mutation rates (high fidelity), any rare mutation effectively traps a lineage in a new clonal state until a new fluctuation occurs (fidelity catastrophe).

VIII. DISCUSSION AND CONCLUSIONS

We reveal the critical role of fluctuations in a stochastic variant of the EM: When replication occurs with high fidelity (i.e., mutation is weak), noise-induced drifts lead to a stochastic switching between clonal-like states that we term the fidelity catastrophe.

We do this by considering the projection, in phase space, of the effective force—i.e., the advective component of the probability current. This permits a qualitative characterization of the system through the description of two emergent fixed

points. One of these fixed points is stable and associated with the deterministic behavior of the system. It allows us to characterize an error threshold, by analogy with the error catastrophe that is synonymous with the deterministic version of the EM. The other fixed point is unstable and associated with noise-induced effects. It influences behavior at finite N only, with its entry into the simplex defining another threshold, this time at high fidelity, marking the entry to a regime where solutions become multimodal.

By calculating these thresholds precisely, we are able to determine how they vary with system parameters. The dependence on the dimension of the system, m, is particularly striking. Specifically, we observe that m plays an antagonistic role with system size, whereby high-dimensional systems need a commensurately large number of replicators, N, in order to avoid noise-induced behaviors from manifesting. Consequently, at fixed $\hat{\mu}$, it is the ratio N/m, i.e., the number of replicators *per variant*, which controls the onset of noise-induced multistability in systems with arbitrary dimension.

In light of this, it is highly relevant that many realistic scenarios will have very large dimension, such that $m \gg N$. In this regime, we have shown that the threshold to the fidelity catastrophe and a generalized error threshold all but coincide. The implication is that, assuming that the error threshold is to be avoided, these systems will generically be multistable and therefore exhibit noise-induced behaviors.

This result is surprising and, at first take, contrary to all standard expectations of evolutionary systems. However, while we have demonstrated a threshold for the existence of multistability, we have stopped short of precisely quantifying its dynamical aspects; a detailed analysis of the effective transition rates between clonal-like states is required to properly capture the relative likelihood of visiting, say, the master variant, as well as how long the system remains in that clonal state. For example, in cases such as those with large selection advantages, it may be that while a distribution is technically multimodal, one must wait for an extremely long time to observe a stochastic fluctuation away from the WT. Generically, precise expressions for the transition rates are likely to rely on the magnitude by which the replication rate of the fittest variant exceeds that of the other variants (i.e., $\Delta\alpha$ in our quasispecies example), the number of variants (i.e., the dimension, m) and the underlying structure of the mutation rates (i.e., a chosen Hamming or other model). Recently, piecewise deterministic Markov processes have been used to approximate transitions in stochastically multistable systems and this technique may also prove useful here [42–44].

The context of rare, but significant, fluctuations is particularly relevant for drawing out the parallels and differences between the phenomenon presented here and Muller's ratchet [22,45–48], which is a distinct finite N effect, characterized in terms of fluctuations, associated with replicating asexual populations. Generically, Muller's ratchet (MR) arises in the absence of backmutation, due to extremely large state spaces ($L \gg 1$, $m = |\mathcal{A}|^L \gg 1$), and in the context of a fitness landscape ordered by the loading of deleterious mutations such that a variant's fitness is inversely proportional to the number of mutations it carries. The absence of backmutation amounts to every mutation being effectively irreversible since the undoing of any given mutation will only occur in proportion to $1/L \to 0$. This results in a directed cascade of (effectively) irreversible loss of fitness in the population since, after a stochastic fluctuation that drives the population of the fittest variant to zero, no recovery of that variant is possible. This starts with the fittest variant and moves sequentially through states of increased mutational loading. Such (possibly rare) fluctuations, where $N_i \to 0$, are very similar in nature to the $n_j \to 1, n_{i\neq j} \to 0$ fluctuations which underpin the stochastic switching in this work. Consequently, there is value in carefully disambiguating the two phenomena.

We first clarify that MR requires both an (effective) absence of back mutation, achieved through $L \gg 1$, and an ordered fitness landscape to achieve the progressive loss of fitness. In contrast, the fidelity catastrophe, while presented here using the example of a two-level, or "spiked," fitness landscape, does not strongly constrain the fitness landscape. Nor does it prohibit back mutation. The latter can be seen from Eq. (16), which describes the generic requirements to observe the fidelity catastrophe. Of note, the only constraint on L is that $m = |\mathcal{A}|^L > 1$, meaning that it can potentially be small. Indeed, explicit simulations at $L = 4$ demonstrate a fidelity catastrophe with repeated fluctuations back to the wild type [Fig. 3(b)]. In other words, the absence of MR does not imply that the system is either in, or is not in, the fidelity catastrophe regime. Similarly, we may ask if the presence of MR materially implies, or excludes, the fidelity catastrophe. On first inspection, insisting upon an absence of back mutation ($L \gg 1$) at finite N would seem to automatically satisfy the bounds $\hat{\mu} > \hat{\mu}_{\text{crit}}^{\text{fidel}}$ ($N < N_{\text{crit}}^{\text{fidel}}$), such that, given an ordered fitness landscape, both phenomena would be expected simultaneously. This is indeed possible and in such cases each $n_j \to 1$, $n_{i\neq j} \to 0$ stochastic switch associated with the fidelity catastrophe would be irreversible and represent a "click" of the ratchet, since the overwhelming likelihood is always that genotype j would be less fit than genotype i. In this case, once the master variant is lost, it is never returned to. However, the presence of MR does not imply that the system is in the fidelity catastrophe regime; as can be observed in Fig. 4(b) and Eq. (15), for any given m and N (including $m \gg N$) there is always a value of $\hat{\mu}$ below which the fidelity catastrophe does not arise and where the MR operates in its more traditionally conceived form without the concomitant $n_j \to 1$ fixation events. We note that our results do, however, predict that only a very small subset of this region operates outside the error catastrophe regime in this limit, due to Eq. (18). Interpreting the relative size of these regions ($\hat{\mu} < \hat{\mu}_{\text{crit}}^{\text{fidel}}$ and $\hat{\mu}_{\text{crit}}^{\text{error}} < \hat{\mu} < \hat{\mu}_{\text{crit}}^{\text{fidel}}$), so as to determine whether physical systems are likely to be found in either, is a delicate question. By writing our threshold in terms of fixed per-base error rate, $\varepsilon = 1 - \hat{\mu}^{1/L}$, and assuming $\alpha > \omega$, we obtain the requirement for the fidelity catastrophe $\varepsilon \lesssim 1 - e^{-1/L} \sim 1/L$. This mirrors the error threshold [Eq. (18) and Appendix F], suggesting that real systems that avoid the error catastrophe in nature (e.g., $L \sim 10^6$–10^7, $\varepsilon \sim 10^{-8}$–10^{-10} in many bacteria) fall into the fidelity catastrophe regime. The validity of such predictions, however, is unclear and demands further work. In particular, the effect of the various simplifying assumptions needs to be understood more fully.

One such assumption concerns the structure of the fitness landscape: While we have presented effective force

calculations for arbitrary fitness landscapes (Appendix C and D), we have restricted ourselves to the specific case of a single-peaked (and otherwise flat) landscape for the purpose of calculating explicit thresholds. This assumes that every one of the m possible genetic permutations generates a viable variant, and all but one of them share the same fitness. In practice, fitness landscapes can be expected to be far more complex. Many variants are likely not viable and have "zero fitness." For others, there is significant redundancy, such that only combinations of point mutations confer changes to fitness. As a result, the effective dimension, in terms of viable (nonzero fitness) phenotypes, is significantly smaller than m, challenging the notion that m/N is a crucial parameter. An interesting question therefore appears to be: How much of the (stochastic) Eigen model's phenomenology survives under a phenotypic coarse graining? We therefore welcome all further work in the area.

ACKNOWLEDGMENTS

E.C., R.E.S., and R.G.M. acknowledge support from the EMBL Australia program. R.G.M. acknowledges funding from the Australian Research Council Centre of Excellence for Mathematical Analysis of Cellular Systems (CE230100001). The authors thank K. Husain (UCL, London, U.K.) for critical feedback on initial versions of the manuscript.

DATA AVAILABILITY

The data used in this study is available in Ref. [49], together with the code that was used to generate them.

APPENDIX A: FORMULATION OF THE STOCHASTIC DYNAMICS

Let us denote by $p(\mathbf{N},t)$ the probability of a specific demographic mix $\mathbf{N} := \{N_1,\ldots,N_m\}$ at time t. The stochastic dynamics of the system are captured by the master equation

$$\dot{p}(\mathbf{N},t) = \left[\sum_{i,j=1}^{m}\left(E_{N_i}^{-1}-1\right)N_j\alpha_j\mu_{ji} + \sum_{i=1}^{m}\left(E_{N_i}^{+1}-1\right)N_i\omega_i\right]p(\mathbf{N},t), \tag{A1}$$

where the step operator E_x^r has action $E_x^r f(x) = f(x+r)$. We recall that α_i and ω_i are the replication and deterioration rates of variant i, respectively, while μ_{ji} is the probability that variant i is produced upon the replication of variant j, such that $\sum_{i=1}^{m}\mu_{ji}=1$. As such, we consider $1-\mu_{ii}$ to be the total replication error of variant i, while its complement, $\mu_{ii}\equiv\hat{\mu}$, is its fidelity.

A set of SDEs that approximate the dynamics can be obtained from the master equation through a system size expansion [39]. When $r \ll x$, the step operator E can be approximated using a Taylor series up to the second order, i.e.,

$$E_x^r - 1 \approx r\frac{\partial}{\partial x} + \frac{r^2}{2}\frac{\partial^2}{\partial x^2}. \tag{A2}$$

Hence, for large values of $N := \sum_{i=1}^{m}N_i$, we can substitute Eq. (A2) into Eq. (A1) to obtain the FPE

$$\frac{\partial p(\mathbf{N},t)}{\partial t} = \sum_{i,j=1}^{m}\left(-\frac{\partial}{\partial N_i}[N_j\alpha_j\mu_{ji}p(\mathbf{N},t)] + \frac{1}{2}\frac{\partial^2}{\partial N_i^2}[N_j\alpha_j\mu_{ji}p(\mathbf{N},t)]\right) + \sum_{i=1}^{m}\left(\frac{\partial}{\partial N_i}[N_i\omega_i p(\mathbf{N},t)] + \frac{1}{2}\frac{\partial^2}{\partial N_i^2}[N_i\omega_i p(\mathbf{N},t)]\right). \tag{A3}$$

Collecting the $\partial/\partial N_i$ and $\partial^2/\partial^2 N_i$ terms in Eq. (A3), respectively, yields

$$A_i(\mathbf{N}) := \sum_{j=1}^{m}N_j\alpha_j\mu_{ji} - N_i\omega_i, \tag{A4}$$

$$B_i(\mathbf{N}) := \sum_{j=1}^{m}N_j\alpha_j\mu_{ji} + N_i\omega_i, \tag{A5}$$

which can be used to rewrite Eq. (A3) in its common form

$$\frac{\partial p(\mathbf{N},t)}{\partial t} = -\sum_{i=1}^{m}\frac{\partial}{\partial N_i}[A_i(\mathbf{N})p(\mathbf{N},t)] + \frac{1}{2}\sum_{i=1}^{m}\frac{\partial^2}{\partial N_i^2}[B_i(\mathbf{N})p(\mathbf{N},t)]. \tag{A6}$$

Following Ref. [40] and Eq. (A6) implies the existence of the Itô SDEs

$$\mathrm{d}N_i = A_i(\mathbf{N})\mathrm{d}t + \sqrt{B_i(\mathbf{N})}\mathrm{d}W_i, \tag{A7}$$

where W_i are uncorrelated Wiener processes, i.e., $\mathbb{E}[\mathrm{d}W_i]=0$ and $\mathbb{E}[\mathrm{d}W_i(t)\mathrm{d}W_j(t')] = \delta_{ij}\delta(t-t')\mathrm{d}t$.

We aim at recasting the dynamics in terms of concentrations $n_i := N_i/N = N_i/\sum_{i=1}^{m}N_i$. To do this, we use the multivariate form of Itô's lemma, which, considering that the diffusion matrix is diagonal, yields the following:

$$\mathrm{d}n_i = \left(\sum_{j=1}^{m}A_j(\mathbf{N})\frac{\partial}{\partial N_j}n_i + \frac{1}{2}\sum_{j=1}^{m}B_j(\mathbf{N})\frac{\partial^2}{\partial N_j^2}n_i\right)\mathrm{d}t + \sum_{j=1}^{m}\sqrt{B_j(\mathbf{N})}\left(\frac{\partial}{\partial N_j}n_i\right)\mathrm{d}W_j. \tag{A8}$$

The first term can be simplified by noting that

$$\frac{\partial n_i}{\partial N_j} = \begin{cases} \frac{1-n_i}{N}, & i = j, \\ -\frac{n_i}{N}, & i \neq j, \end{cases} \tag{A9}$$

giving

$$\sum_{j=1}^{m} A_j(\mathbf{N}) \frac{\partial}{\partial N_j} n_i = A_i(\mathbf{n}) - n_i \sum_{j=1}^{m} A_j(\mathbf{n}) =: \Phi_i(\mathbf{n}), \tag{A10}$$

and, similarly, the second term can be simplified noting that

$$\frac{\partial^2 n_i}{\partial N_j^2} = \begin{cases} \frac{2(n_i-1)}{N^2}, & i = j, \\ \frac{2n_i}{N^2}, & i \neq j, \end{cases} \tag{A11}$$

giving

$$\frac{1}{2} \sum_{j=1}^{m} B_j(\mathbf{N}) \frac{\partial^2}{\partial N_j^2} n_i = -\frac{1}{N} \left(B_i(\mathbf{n}) - n_i \sum_{j=1}^{m} B_j(\mathbf{n}) \right) =: \Psi_i(\mathbf{n}). \tag{A12}$$

We note the use of $A_i(\mathbf{n}) := A_i(\mathbf{N})/N$ and $B_i(\mathbf{n}) := B_i(\mathbf{N})/N$ in the above. To deal with the third term of Eq. (A8) (i.e., the stochastic part), we again use Eq. (A9) to show that

$$\sqrt{B_j(\mathbf{N})} \frac{\partial n_i}{\partial N_j} = \frac{1}{\sqrt{N}} \times \begin{cases} (1-n_i)\sqrt{B_i(\mathbf{n})}, & i = j \\ -n_i \sqrt{B_j(\mathbf{n})}, & i \neq j \end{cases} =: \frac{1}{\sqrt{N}} q_{ij}(\mathbf{n}). \tag{A13}$$

Thus, Eq. (A8) can equally be expressed as

$$\mathrm{d}n_i = (\Phi_i(\mathbf{n}) + \Psi_i(\mathbf{n}))\,\mathrm{d}t + \frac{1}{\sqrt{N}} \sum_{j=1}^{m} q_{ij}(\mathbf{n})\,\mathrm{d}W_j, \tag{A14}$$

as appears in the main text.

To recover an SDE for the total number of individuals, N, we can merely sum Eq. (A7) yielding

$$\begin{aligned} \mathrm{d}N &= \sum_{i=1}^{m} \mathrm{d}N_i = \sum_{i=1}^{m} A_i(\mathbf{N})\mathrm{d}t + \sum_{i=1}^{m} \sqrt{B_i(\mathbf{N})}\mathrm{d}W_i \\ &= N \sum_{i=1}^{m} n_i(\alpha_i - \omega_i)\mathrm{d}t + \sqrt{N} \sum_{i=1}^{m} \sqrt{n_i(\alpha_i + \omega_i)}\mathrm{d}W_i \\ &= N(\langle \alpha \rangle_{\mathbf{n}} - \langle \omega \rangle_{\mathbf{n}})\mathrm{d}t + \sqrt{N(\langle \alpha \rangle_{\mathbf{n}} + \langle \omega \rangle_{\mathbf{n}})}\mathrm{d}W, \end{aligned} \tag{A15}$$

where $\langle \alpha \rangle_{\mathbf{n}} := \sum_{i=1}^{m} n_i \alpha_i$ and $\langle \omega \rangle_{\mathbf{n}} := \sum_{i=1}^{m} n_i \omega_i$ are the average replication and deterioration rates of the system in configuration $\mathbf{n}$, and where the final expression is in terms of increments of a newly introduced Wiener process $\mathrm{d}W$, which is statistically equivalent to the composite Wiener process $m^{-1/2} \sum_{i=1}^{m} \mathrm{d}W_i$ and is thus correlated with the $\mathrm{d}W_i$ such that $\mathbb{E}[\mathrm{d}W\,\mathrm{d}W_i] = m^{-1/2}\mathrm{d}t$.

APPENDIX B: THE FOKKER-PLANCK EQUATION IN CONCENTRATION SPACE

The SDEs introduced in Appendix A imply the Fokker-Planck equation

$$\frac{\partial p(\mathbf{n},t)}{\partial t} = -\sum_{i=1}^{m} \frac{\partial}{\partial n_i} [(\Phi_i(\mathbf{n}) + \Psi_i(\mathbf{n}))p(\mathbf{n},t)] + \frac{1}{2N} \sum_{i,j=1}^{m} \frac{\partial^2}{\partial n_i \partial n_j} [Q_{ij}(\mathbf{n})p(\mathbf{n},t)], \tag{B1}$$

where $\mathsf{Q} = [Q_{ij}] := \mathsf{q}\mathsf{q}^T$ and q is the matrix $[q_{ij}]$. Our central strategy then relies upon being able to rewrite this as

$$\begin{aligned} \frac{\partial p(\mathbf{n},t)}{\partial t} &= -\sum_{i=1}^{m} \frac{\partial}{\partial n_i} [(\Phi_i(\mathbf{n}) + \Psi_i(\mathbf{n}))p(\mathbf{n},t)] + \frac{1}{2N} \sum_{i,j=1}^{m} \left[\frac{\partial}{\partial n_i} p(\mathbf{n},t) \frac{\partial}{\partial n_j} Q_{ij}(\mathbf{n}) + \frac{\partial}{\partial n_i} Q_{ij}(\mathbf{n}) \frac{\partial}{\partial n_j} p(\mathbf{n},t) \right] \\ &= -\sum_{i=1}^{m} \frac{\partial}{\partial n_i} \left[\left(\Phi_i(\mathbf{n}) + \Psi_i(\mathbf{n}) - \frac{1}{2N} \sum_{j=1}^{m} \frac{\partial}{\partial n_j} Q_{ij}(\mathbf{n}) \right) p(\mathbf{n},t) - \frac{1}{2N} \sum_{j=1}^{m} Q_{ij}(\mathbf{n}) \frac{\partial}{\partial n_j} p(\mathbf{n},t) \right]. \end{aligned} \tag{B2}$$

Equation (B2) is more conveniently written in vector notation as

$$\dot{p}(\mathbf{n},t) = -\nabla_{\mathbf{n}} \cdot \mathcal{J}, \tag{B3}$$

with

$$\mathcal{J} := (\Phi(\mathbf{n}) + \Psi(\mathbf{n}) + \hat{\Theta}(\mathbf{n}))p(\mathbf{n},t) - \frac{1}{2N} \mathsf{Q}(\mathbf{n}) \nabla_{\mathbf{n}} p(\mathbf{n},t), \tag{B4}$$

where $\nabla_{\mathbf{n}} := (\partial/\partial n_1, \ldots, \partial/\partial n_m)$, $\Phi(\mathbf{n}) = (\Phi_1(\mathbf{n}), \ldots, \Phi_m(\mathbf{n}))$, $\Psi(\mathbf{n}) = (\Psi_1(\mathbf{n}), \ldots, \Psi_m(\mathbf{n}))$, and $\hat{\Theta}(\mathbf{n}) = (\hat{\Theta}_1(\mathbf{n}), \ldots, \hat{\Theta}_m(\mathbf{n}))$, and where

$$\hat{\Theta}_i(\mathbf{n}) = -\frac{1}{2N} \sum_{j=1}^{m} \frac{\partial}{\partial n_j} Q_{ij}(\mathbf{n}). \tag{B5}$$

It is then natural to attempt to solve for $\dot{p} = 0$, assuming absence of stationary flux, implying

$$\frac{1}{2N} \mathsf{Q}(\mathbf{n}) \nabla_{\mathbf{n}} p(\mathbf{n},t) = (\Phi(\mathbf{n}) + \Psi(\mathbf{n}) + \hat{\Theta}(\mathbf{n}))p(\mathbf{n},t), \tag{B6}$$

and therefore

$$\frac{\nabla_{\mathbf{n}} p(\mathbf{n},t)}{p(\mathbf{n},t)} = 2N\mathsf{Q}^{-1}(\mathbf{n})(\Phi(\mathbf{n}) + \Psi(\mathbf{n}) + \hat{\Theta}(\mathbf{n})), \tag{B7}$$

thus associating stationary points in the distribution with zeros of the right-hand side.

However, this reasoning is faulty since it relies on the inversion of the matrix Q, which is singular due to the simplex constraint that $\sum_i n_i = 1$. To proceed, as per the main text, we exploit the property that solution is

symmetrically singular in the sense that $p(n_1, \ldots, n_m) = p_{\setminus i}(n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_m)\delta(n_i - (1 - \sum_{k\neq i} n_k))$, for all i. As such, we can eliminate an arbitrary coordinate (we choose m without loss of generality) producing an FPE for $\tilde{p}(\tilde{\mathbf{n}}) \equiv p_{\setminus m}(\{n_1, \ldots, n_{m-1}\})$. This distribution is governed by an FP operator exactly as the above, but with all instances of n_m replaced as $n_m \to 1 - \sum_{k=1}^{m-1} n_k$ such that all quantities become functions of $\tilde{\mathbf{n}} = \{n_1, \ldots, n_{m-1}\}$, and with the appropriate mth element of the vectors, and mth row/column of the matrices, removed. We write the resulting reduced matrix as $\mathcal{Q}(\tilde{\mathbf{n}}) = [\mathsf{Q}(\mathbf{n})]_{1:m-1,1:m-1}$, comprised of elements $\mathcal{Q}_{ij}$. For vanishing stationary flux, the above reasoning then becomes valid and we identify

$$\frac{\nabla_{\tilde{\mathbf{n}}}\tilde{p}(\tilde{\mathbf{n}}, t)}{\tilde{p}(\tilde{\mathbf{n}}, t)} = 2N\mathcal{Q}^{-1}(\tilde{\mathbf{n}})(\Phi(\tilde{\mathbf{n}}) + \Psi(\tilde{\mathbf{n}}) + \Theta(\tilde{\mathbf{n}})), \tag{B8}$$

but now with $\Phi(\tilde{\mathbf{n}}) = (\Phi_1(\tilde{\mathbf{n}}), \ldots, \Phi_{m-1}(\tilde{\mathbf{n}}))$, etc., and with the noise-induced term now given by

$$\Theta_i(\tilde{\mathbf{n}}) := -\frac{1}{2N}\sum_{j=1}^{m-1}\frac{\partial}{\partial n_j}\mathcal{Q}_{ij}(\tilde{\mathbf{n}}), \tag{B9}$$

constructed from the reduced matrix $\mathcal{Q}(\tilde{\mathbf{n}})$, and with its constituent sum ranging from 1 to $m-1$. This distinction is crucial and we recognize that $\Theta(\tilde{\mathbf{n}}) \neq \hat{\Theta}(\mathbf{n})$ [i.e., Eqs. (B9) and (B5) are not equal]. This then allows us to identify the total effective force

$$F(\tilde{\mathbf{n}}) := \Phi(\tilde{\mathbf{n}}) + \Psi(\tilde{\mathbf{n}}) + \Theta(\tilde{\mathbf{n}}), \tag{B10}$$

which vanishes at stationary points of the limiting distribution in the case of zero stationary current, and is proportional to the total nondispersive flux in the general case.

1. Example solution, equivalent variants, $m = 2$

It is instructive to consider a simple, exactly solvable, example wherein we can illustrate the important qualitative features that the noise-induced forces can exert on the system. In particular, we aim to show (1) that knowing just the solution at the level of the mean is not sufficient to describe the overall behavior exhibited by the system, and (2) that at low population size N and/or high fidelity $\hat{\mu} := \mu_{ii}$, the probability distribution undergoes a qualitative change we associate with the fidelity catastrophe.

To this end, let us consider a simple example of equivalent species (equal replication and deterioration rates) in the case of just two variants, $m = 2$. In this case, we can enumerate all required objects as

$$\mu_{11} = \mu_{22} = \hat{\mu}, \tag{B11a}$$
$$\mu_{12} = \mu_{21} = 1 - \hat{\mu}, \tag{B11b}$$
$$A_1(\mathbf{n}) = \alpha\hat{\mu}n_1 + \alpha(1-\hat{\mu})n_2 - n_1\omega, \tag{B11c}$$
$$A_2(\mathbf{n}) = \alpha\hat{\mu}n_2 + \alpha(1-\hat{\mu})n_1 - n_2\omega, \tag{B11d}$$
$$B_1(\mathbf{n}) = \alpha\hat{\mu}n_1 + \alpha(1-\hat{\mu})n_2 + n_1\omega, \tag{B11e}$$
$$B_2(\mathbf{n}) = \alpha\hat{\mu}n_2 + \alpha(1-\hat{\mu})n_1 + n_2\omega, \tag{B11f}$$
$$\Phi_1(\mathbf{n}) = \alpha\hat{\mu}n_1 + \alpha(1-\hat{\mu})n_2 - n_1\alpha, \tag{B11g}$$
$$\Psi_1(\mathbf{n}) = -\Phi_1(\mathbf{n})/N, \tag{B11h}$$
$$\Phi_2(\mathbf{n}) = \alpha\hat{\mu}n_2 + \alpha(1-\hat{\mu})n_1 - n_2\alpha, \tag{B11i}$$
$$\Psi_2(\mathbf{n}) = -\Phi_2(\mathbf{n})/N, \tag{B11j}$$
$$q_{11}(\mathbf{n}) = (1-n_1)\sqrt{B_1}, \tag{B11k}$$
$$q_{12}(\mathbf{n}) = -n_1\sqrt{B_2}, \tag{B11l}$$
$$q_{21}(\mathbf{n}) = -n_2\sqrt{B_1}, \tag{B11m}$$
$$q_{22}(\mathbf{n}) = (1-n_2)\sqrt{B_2}, \tag{B11n}$$
$$\mathcal{Q}_{ij}(\mathbf{n}) = \sum_{k=1}^{2} q_{ik}(\mathbf{n})q_{jk}(\mathbf{n}). \tag{B11o}$$

As per Appendix B, constructing the Fokker-Planck equation in terms of $\mathbf{n} = \{n_1, n_2\}$ is problematic owing to the constraint $n_1 + n_2 = 1$ leading to a singular diffusion matrix, Q. To solve for this system, we need to consider the underlying nonsingular subsystem, which we can do by eliminating n_2 by writing $n_2 \to 1 - n_1$, such that we deal with $\tilde{\mathbf{n}} = n_1$. This then also requires the consideration of the sub-matrix, $\mathcal{Q}(n_1)$, which is now simply the single element, $\mathcal{Q}(n_1) = \mathcal{Q}_{11}(n_1) = Q_{11}(n_1, 1-n_1)$. As such, we have the following 1D system (for which the limiting state must have vanishing flux):

$$\mathcal{Q}(n_1) = -(n_1-1)n_1\omega - \alpha(\hat{\mu} + (4\hat{\mu}-3)(n_1-1)n_1 - 1), \tag{B12}$$

$$\begin{aligned} F_1(n_1) &= \Phi_1(n_1) + \Psi_1(n_1) + \Theta_1(n_1) \\ &= \Phi_1(n_1) + \Psi_1(n_1) - (2N)^{-1}\partial_{n_1}\mathcal{Q}(n_1) \\ &= \frac{(2n_1-1)(\alpha(2\hat{\mu} + 2(\hat{\mu}-1)N - 1) + \omega)}{2N}, \end{aligned} \tag{B13}$$

$$\frac{\partial_{n_1}\tilde{p}(n_1)}{\tilde{p}(n_1)} = 2N\frac{F_1(n_1)}{\mathcal{Q}(n_1)} = -\frac{(2n_1-1)(\alpha(2\hat{\mu} + 2(\hat{\mu}-1)N - 1) + \omega)}{(\alpha(\hat{\mu} + (4\hat{\mu}-3)(n_1-1)n_1 - 1) + (n_1-1)n_1\omega)}, \tag{B14}$$

with solution, up to a normalizing constant,

$$\tilde{p}(n_1) \propto \exp\left[\int_0^{n_1} dn_1'\, 2N\frac{F_1(n_1')}{\mathcal{Q}(n_1')}\right] = (\alpha(\hat{\mu}-1) + (n_1-1)n_1(\alpha(4\hat{\mu}-3)+\omega))^{-\frac{\alpha(2\hat{\mu}+2(\hat{\mu}-1)N-1)+\omega}{(\alpha(4\hat{\mu}-3)+\omega)}}. \tag{B15}$$

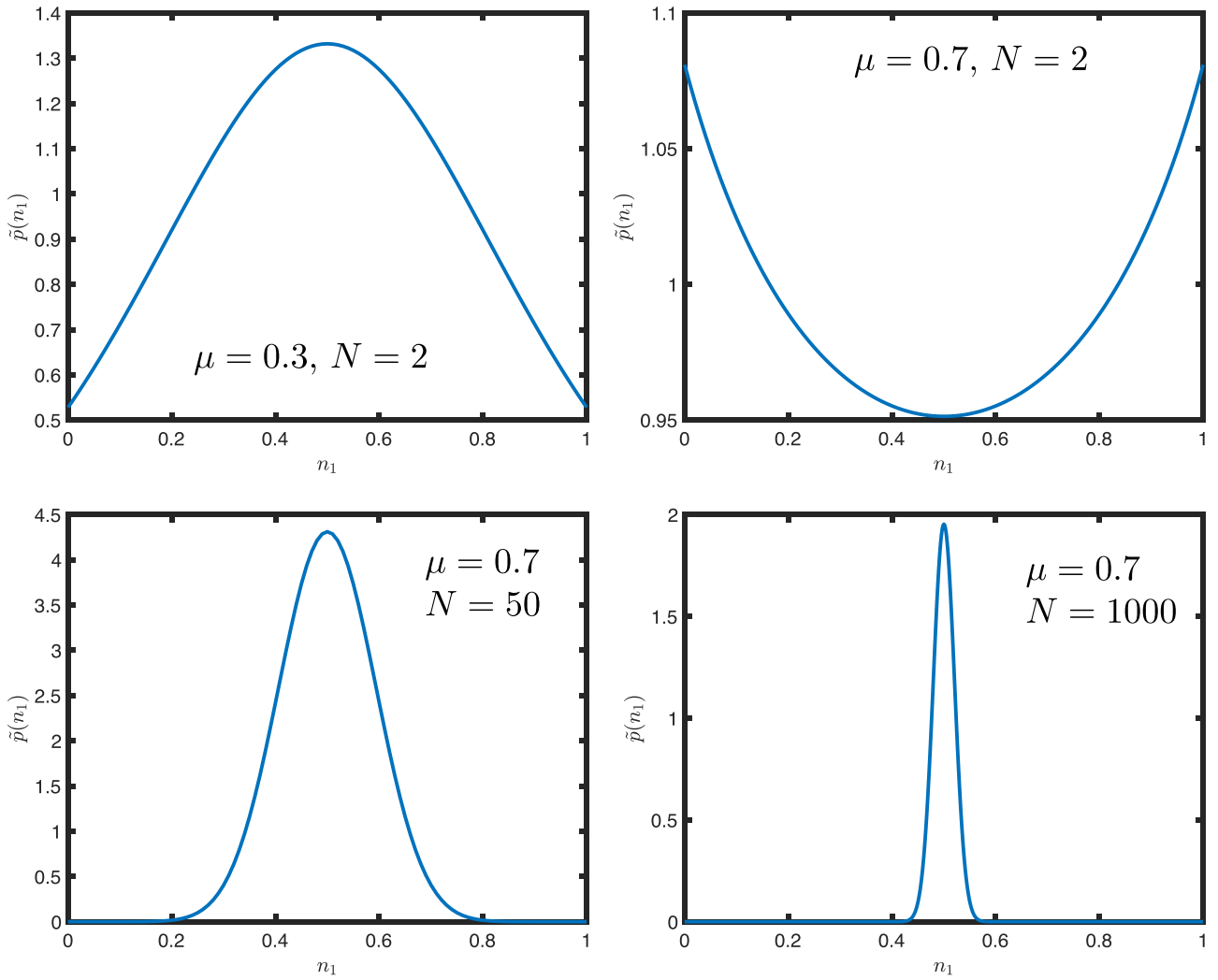

FIG. 5. Exact probability density functions for the $m = 2$ equivalent variant system for $\alpha = \omega = 1$. The behavior of the system depends strongly on N and $\hat{\mu}$, but all solutions share the mean value of $\langle n_1 \rangle = 1/2$.

On the level of the mean, this distribution has identical behavior for all parameters, with the mean simply equal to $\langle n_1 \rangle = 1/2$, by symmetry. However, it exhibits very different behavior depending on the choice of parameters. These are illustrated in Fig. 5 for the case $\alpha = \omega = 1$. Note, however, that as $N \to \infty$ the distribution tends to an increasingly narrow Gaussian centered on $n_1 = 1/2$, as expected from the deterministic version of the Eigen model.

At low N, however, we particularly highlight that for various distinct parameter values the probability can be either concentrated on the mean, i.e., around $n_1 = n_2 = 1/2$, or bimodally, far from the mean, in the regions corresponding to homogeneous, one species dominance, i.e., $n_1 = 1$, $n_2 = 0$, and $n_1 = 0, n_2 = 1$. Specifically, for small N, where we expect the noise-induced behavior to dominate, the value of $\hat{\mu}$ controls the shape of the distribution between these two extremes. Where the distribution is bimodal in this way—at low N and high $\hat{\mu}$—we say that the system has suffered from the fidelity catastrophe.

This simple case is presented since it is exactly solvable; however, we wish to understand systems with much higher dimensionality. To do so, we implement the methodology expressed in the main text, whereby, in lieu of full solutions of the Fokker-Planck equation, we seek effective forces that can be associated with the nondispersive component of the limiting flux, even if we cannot state the full distribution. The expression for these forces, first in general, and then along illustrative lines in phase space, is the subject of the following sections.

APPENDIX C: FORCES IN THE SIMPLEX

As per the main text, we have identified a total effective force vector, $F(\tilde{\mathbf{n}})$, comprised of m components, $F_i(\tilde{\mathbf{n}})$, $i \in \{1, \ldots, m\}$, one for each variant (we note, only $m - 1$ are required in the FPE due to the simplex constraint). Each such component, in turn, is formed from three contributions

$$F_i(\tilde{\mathbf{n}}) := \Phi_i(\tilde{\mathbf{n}}) + \Psi_i(\tilde{\mathbf{n}}) + \Theta_i(\tilde{\mathbf{n}}), \tag{C1}$$

with the latter contribution identified as originating from the multiplicative nature of the noise, derived from the reduced diffusion matrix.

Here, we aim to derive expressions for these forces at any point within the simplex, making only the assumptions $\mu_{ii} = \mu_{jj}$ and $\mu_{ij} = \mu_{ji}$ for all $i, j \in \{1, \ldots, m\}$. We then note the essential corollary $\sum_{j=1}^{m} \mu_{ij} = \sum_{j=1}^{m} \mu_{ji} = 1$.

To do so, and for the sake of readability, we first define the following quantities:

$$\bar{\alpha} := \frac{1}{m}\sum_{j=1}^{m} \alpha_j \quad \text{and} \quad \langle \alpha \rangle_{\mathbf{n}} := \sum_{j=1}^{m} n_j \alpha_j, \tag{C2}$$

$$\bar{\omega} := \frac{1}{m}\sum_{j=1}^{m} \omega_j \quad \text{and} \quad \langle \omega \rangle_{\mathbf{n}} := \sum_{j=1}^{m} n_j \omega_j, \tag{C3}$$

$$\mathcal{T}_{ji} := \alpha_j \mu_{ji} \quad \text{and} \quad \langle \mathcal{T}_{\cdot a} \rangle_{\mathbf{n}} := \sum_{j=1}^{m} n_j \mathcal{T}_{ja} = \sum_{j=1}^{m} \alpha_j n_j \mu_{ja} = B_a(\mathbf{n}) - n_a \omega_a, \tag{C4}$$

$$\gamma_i := \sum_{j=1}^{m} \mathcal{T}_{ji} \quad \text{and} \quad \bar{\gamma} := \frac{1}{m}\sum_{j=1}^{m} \gamma_j = \frac{1}{m}\sum_{j=1}^{m}\sum_{k=1}^{m} \alpha_k \mu_{kj} = \bar{\alpha}. \tag{C5}$$

We also recognize that

$$\sum_{i=1}^{m} A_i(\mathbf{n}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} \alpha_j n_j \mu_{ji} - n_i \omega_i \right) = \langle \alpha \rangle_{\mathbf{n}} - \langle \omega \rangle_{\mathbf{n}}, \tag{C6}$$

$$\sum_{i=1}^{m} B_i(\mathbf{n}) = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} \alpha_j n_j \mu_{ji} + n_i \omega_i \right) = \langle \alpha \rangle_{\mathbf{n}} + \langle \omega \rangle_{\mathbf{n}}, \tag{C7}$$

and

$$\sum_{i=1}^{m} \gamma_i = \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_j \mu_{ji} = \sum_{j=1}^{m} \alpha_j = m\bar{\alpha}. \tag{C8}$$

Recalling $\Phi_i(\mathbf{n}) := A_i(\mathbf{n}) - n_i \sum_{j=1}^{m} A_j(\mathbf{n})$, with $A_i(\mathbf{n}) := \sum_{j=1}^{m} n_j \alpha_j \mu_{ji} - n_i \omega_i$, and $\Psi_i(\mathbf{n}) := -N^{-1}(B_i(\mathbf{n}) - n_i \sum_{j=1}^{m} B_j(\mathbf{n}))$, with $B_i(\mathbf{n}) := \sum_{j=1}^{m} n_j \alpha_j \mu_{ji} + n_i \omega_i$, we can immediately derive expressions for the two contributions deriving from the deterministic component of the dynamics

$$\begin{aligned} \Phi_i(\tilde{\mathbf{n}}) &\equiv \Phi_i(\mathbf{n}) \\ &= A_i - n_i(\langle \alpha \rangle_{\mathbf{n}} - \langle \omega \rangle_{\mathbf{n}}) \\ &= \langle \mathcal{T}_{\cdot i} \rangle_{\mathbf{n}} - n_i(\langle \alpha \rangle_{\mathbf{n}} + (\omega_i - \langle \omega \rangle_{\mathbf{n}})), \end{aligned} \tag{C9}$$

$$\begin{aligned} \Psi_i(\tilde{\mathbf{n}}) &\equiv \Psi_i(\mathbf{n}), \\ &= -N^{-1} B_i + N^{-1} n_i(\langle \alpha \rangle_{\mathbf{n}} + \langle \omega \rangle_{\mathbf{n}}) \\ &= -N^{-1} \langle \mathcal{T}_{\cdot i} \rangle_{\mathbf{n}} + N^{-1} n_i(\langle \alpha \rangle_{\mathbf{n}} - (\omega_i - \langle \omega \rangle_{\mathbf{n}})), \end{aligned} \tag{C10}$$

since these terms are invariant under either $\mathbf{n}$ or $\tilde{\mathbf{n}}$ (subject to the substitution $n_m \to 1 - \sum_{i=1}^{m-1} n_i$, assuming removal of the mth component).

Calculating the noise-induced component, however, is slightly more involved, since it is defined in terms of derivatives of the reduced matrix $\mathcal{Q}$. Explicitly, the form of this force component (assuming the reduction in state space by removing n_m) is given by

$$\Theta_i(\tilde{\mathbf{n}}) := -\frac{1}{2N}\sum_{j=1}^{m-1} \partial_{n_j} \mathcal{Q}_{ij}(\tilde{\mathbf{n}}). \tag{C11}$$

To start, we derive the expressions for $Q_{ii}(\mathbf{n})$ and $Q_{ij}(\mathbf{n})$, noting that these are in terms of the full matrix, Q, and vector, $\mathbf{n}$:

$$\begin{aligned} Q_{ii}(\mathbf{n}) &= \sum_{j=1}^{m} q_{ij}(\mathbf{n}) q_{ij}(\mathbf{n}) \\ &= (1 - n_i)^2 B_i(\mathbf{n}) + n_i^2 \sum_{j=1, j\neq i}^{m} B_j(\mathbf{n}), \end{aligned} \tag{C12}$$

$$\begin{aligned} Q_{ij}(\mathbf{n}) &= \sum_{k=1}^{m} q_{ik}(\mathbf{n}) q_{jk}(\mathbf{n}) \\ &= -(1 - n_i) n_j B_i(\mathbf{n}) - (1 - n_j) n_i B_j(\mathbf{n}) \\ &\quad + n_i n_j \sum_{k=1, k\notin\{i,j\}}^{m} B_k(\mathbf{n}). \end{aligned} \tag{C13}$$

Next, we consider the $Q_{ij}(\mathbf{n})$ as part of the reduced matrix, $\mathcal{Q}_{ij}(\tilde{\mathbf{n}})$, by replacing all instances of n_m as per

$$n_m \to 1 - \sum_{j=1}^{m-1} n_j. \tag{C14}$$

This change manifests itself in the forms of the derivatives that appear in Eq. (C11), noting that the use of the above substitution leads to special cases where the mth coordinate is involved such that

$$\partial_{n_i} n_i = 1, \quad \forall\, i \in \{1, \ldots, m-1\}, \tag{C15}$$

$$\partial_{n_i} n_m = -1, \quad \forall\, i \in \{1, \ldots, m-1\}, \tag{C16}$$

and

$$\partial_{n_i} B_i(\mathbf{n}) = \alpha_i \mu_{ii} + \omega_i - \alpha_m \mu_{mi}, \tag{C17}$$

$$\partial_{n_j} B_i(\mathbf{n}) = \begin{cases} \alpha_j \mu_{ji} - \alpha_i \mu_{ii} - \omega_i, & i = m,\ j \neq i, \\ \alpha_j \mu_{ji} - \alpha_m \mu_{mi}, & i \neq m,\ j \neq i. \end{cases} \tag{C18}$$

We can then use these to consider the derivatives of $\mathcal{Q}_{ij}(\tilde{\mathbf{n}})$, noting that the constituent sums for each element of $\mathcal{Q}$ are still over $\{1, \ldots, m\}$ since the individual $\mathcal{Q}_{ij}$ terms are still formed from m noise terms.

First, we consider $\partial_i \mathcal{Q}_{ii}(\tilde{\mathbf{n}})$, noting that the $i = m$ case is not needed since it does not appear in Eq. (C11):

$$\begin{aligned}
\partial_{n_i}\mathcal{Q}_{ii}(\tilde{\mathbf{n}}) &= \partial_{n_i}\left[(1-n_i)^2 B_i(\mathbf{n}) + n_i^2 \sum_{j=1,j\neq i}^{m} B_j(\mathbf{n})\right] \\
&= \partial_{n_i}\left[(1-n_i)^2 B_i(\mathbf{n}) + n_i^2 B_m(\mathbf{n}) + n_i^2 \sum_{j=1,j\notin\{i,m\}}^{m} B_j(\mathbf{n})\right] \\
&= -2(1-n_i)B_i(\mathbf{n}) + (1-n_i)^2\partial_{n_i}B_i(\mathbf{n}) + 2n_i B_m(\mathbf{n}) + n_i^2\partial_{n_i}B_m(\mathbf{n}) + 2n_i\sum_{j=1,j\notin\{i,m\}}^{m} B_j(\mathbf{n}) + n_i^2\sum_{j=1,j\notin\{i,m\}}^{m}\partial_i B_j(\mathbf{n}) \\
&= -2(1-n_i)B_i(\mathbf{n}) + (1-n_i)^2(\alpha_i\mu_{ii} + \omega_i - \alpha_m\mu_{mi}) + n_i^2(\alpha_i\mu_{im} - \alpha_m\mu_{mm} - \omega_m) \\
&\quad + 2n_i\sum_{j=1,j\neq i}^{m} B_j(\mathbf{n}) + n_i^2\sum_{j=1,j\notin\{i,m\}}^{m}(\alpha_i\mu_{ij} - \alpha_m\mu_{mj}) \\
&= -2B_i(\mathbf{n}) + 2n_i(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}} - \alpha_i\mu_{ii} - \omega_i + \alpha_m\mu_{mi}) + \alpha_i\mu_{ii} + n_i^2(\alpha_i + \omega_i - \alpha_m - \omega_m) + \omega_i - \alpha_m\mu_{mi}. \qquad \text{(C19)}
\end{aligned}$$

Next, we consider $\partial_{n_j}\mathcal{Q}_{ij}(\tilde{\mathbf{n}})$, noting that neither the $i = m$ nor $j = m$ cases are needed, but the constituent sum over k does include the mth term:

$$\begin{aligned}
\partial_{n_j}\mathcal{Q}_{ij}(\tilde{\mathbf{n}}) &= \partial_{n_j}\left[-(1-n_i)n_j B_i(\mathbf{n}) - (1-n_j)n_i B_j(\mathbf{n}) + n_i n_j \sum_{k=1,k\notin\{i,j\}}^{m} B_k(\mathbf{n})\right] \\
&= -(1-n_i)B_i(\mathbf{n}) - (1-n_i)n_j\partial_{n_j}B_i(\mathbf{n}) + n_i B_j(\mathbf{n}) - (1-n_j)n_i\partial_{n_j}B_j(\mathbf{n}) \\
&\quad + n_i\sum_{k=1,k\notin\{i,j\}}^{m} B_k(\mathbf{n}) + n_i n_j\partial_{n_j}B_m(\mathbf{n}) + n_i n_j\sum_{k=1,k\notin\{i,j,m\}}^{m}\partial_{n_j}B_k(\mathbf{n}) \\
&= -(1-n_i)B_i(\mathbf{n}) - (1-n_i)n_j(\alpha_j\mu_{ji} - \alpha_m\mu_{mi}) + n_i B_j(\mathbf{n}) - (1-n_j)n_i(\alpha_j\mu_{jj} + \omega_i - \alpha_m\mu_{mj}) \\
&\quad + n_i(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}} - B_i(\mathbf{n}) - B_j(\mathbf{n})) + n_i n_j(\alpha_j\mu_{jm} - \alpha_m\mu_{mm} - \omega_m) + n_i n_j\sum_{k=1,k\notin\{i,j,m\}}^{m}(\alpha_j\mu_{jk} - \alpha_m\mu_{mk}) \\
&= -B_i(\mathbf{n}) + \alpha_j(n_i(n_j - \mu_{ii}) - n_j\mu_{ji}) + \alpha_m(n_i(\mu_{mj} - n_j) + n_j\mu_{mi}) + n_i(-n_j\omega_m + (n_j-1)\omega_j + \langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}}). \qquad \text{(C20)}
\end{aligned}$$

We now consider the off-diagonal component of the sum in Eq. (C11), noting this excludes m:

$$\begin{aligned}
\sum_{j=1,j\notin\{i,m\}}^{m}\partial_{n_j}\mathcal{Q}_{ij}(\tilde{\mathbf{n}}) &= \sum_{j=1,j\notin\{i,m\}}^{m}[-B_i(\mathbf{n}) + \alpha_j(n_i(n_j - \mu_{ii}) - n_j\mu_{ji}) + \alpha_m(n_i(\mu_{mj} - n_j) + n_j\mu_{mi}) \\
&\quad + n_i(-n_j\omega_m + (n_j-1)\omega_j + \langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}})] \qquad \text{(C21)} \\
&= -(m-2)B_i(\mathbf{n}) + n_i(\langle\alpha\rangle_{\mathbf{n}} - n_i\alpha_i - n_m\alpha_m) - n_i\mu_{ii}(m\bar{\alpha} - \alpha_i - \alpha_m) - \sum_{j=1,j\notin\{i,m\}}^{m}\alpha_j n_j\mu_{ji} \\
&\quad + \alpha_m n_i(1 - \mu_{mi} - \mu_{mm}) - \alpha_m n_i(1 - n_i - n_m) + \alpha_m\mu_{mi}(1 - n_i - n_m) - \omega_m n_i(1 - n_i - n_m) \\
&\quad + n_i(\langle\omega\rangle_{\mathbf{n}} - n_i\omega_i - n_m\omega_m) - n_i(m\bar{\omega} - \omega_i - \omega_m) + (m-2)n_i(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}}) \qquad \text{(C22)} \\
&= n_i(\mu_{ii}(2\alpha_i - m\bar{\alpha}) + (m-1)(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}}) + 2\omega_i - 2\alpha_m\mu_{mi} - m\bar{\omega}) - (m-1)B_i(\mathbf{n}) \\
&\quad + n_i^2(-\alpha_i - \omega_i + \alpha_m + \omega_m) + \alpha_m\mu_{mi}. \qquad \text{(C23)}
\end{aligned}$$

Finally, we add the $\partial_i\mathcal{Q}_{ii}(\tilde{\mathbf{n}})$ component to get

$$\begin{aligned}
-2N\Theta_i(\tilde{\mathbf{n}}) &= \sum_{j=1}^{m-1}\partial_{n_j}\mathcal{Q}_{ij}(\tilde{\mathbf{n}}) \\
&= -(m+1)B_i(\mathbf{n}) + (m+1)n_i(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}}) + \mu_{ii}(\alpha_i - n_i m\bar{\alpha}) + (\omega_i - n_i m\bar{\omega}) \\
&= -(m+1)(\langle\mathcal{T}_{\cdot i}\rangle_{\mathbf{n}} + n_i\omega_i) + (m+1)n_i(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}}) + \mu_{ii}(\alpha_i - n_i m\bar{\alpha}) + (\omega_i - n_i m\bar{\omega}), \qquad \text{(C24)}
\end{aligned}$$

which is independent of any parameter or variable to do with the eliminated coordinate, but defines the noise-induced force in any Cartesian component of the system (i.e., running 1 to m).

Finally, we can now express all components of the total force, which can be evaluated at any point within the simplex:

$$F_i(\tilde{\mathbf{n}}) = \Phi_i(\tilde{\mathbf{n}}) + \Psi_i(\tilde{\mathbf{n}}) + \Theta_i(\tilde{\mathbf{n}}), \tag{C25}$$

$$\Phi_i(\tilde{\mathbf{n}}) = \langle \mathcal{T}_{\cdot i}\rangle_{\mathbf{n}} - n_i(\langle\alpha\rangle_{\mathbf{n}} + (\omega_i - \langle\omega\rangle_{\mathbf{n}})), \tag{C26}$$

$$\Psi_i(\tilde{\mathbf{n}}) = -N^{-1}[\langle \mathcal{T}_{\cdot i}\rangle_{\mathbf{n}} - n_i(\langle\alpha\rangle_{\mathbf{n}} - (\omega_i - \langle\omega\rangle_{\mathbf{n}}))], \tag{C27}$$

$$\Theta_i(\tilde{\mathbf{n}}) = (2N)^{-1}[(m+1)(\langle \mathcal{T}_{\cdot i}\rangle_{\mathbf{n}} + n_i\omega_i) - (m+1)n_i(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}}) - \mu_{ii}(\alpha_i - n_i m\bar{\alpha}) - (\omega_i - n_i m\bar{\omega})]. \tag{C28}$$

We stress, despite its expression as a function of $\tilde{\mathbf{n}} := \{n_1, \ldots, n_{m-1}\}$, there are still m component terms to each force—but they are constrained such that they lie within the $m-1$-dimensional simplex, such that only $m-1$ terms are required to uniquely determine it.

APPENDIX D: FORCES ALONG THE BISECTING SIMPLEX LINE

Here, we demonstrate how to evaluate the forces in a particular direction within the simplex, and in particular to evaluate it along a line that bisects the center of the simplex and one of its vertices. Let c be the coordinate we are investigating (i.e., the vertex of the simplex), and $b = \{1, \ldots, m\} \setminus \{c\} \equiv \{1, \ldots, c-1, c+1, \ldots, m\}$, the set of all other coordinates such that we define the parametric line, $\mathbf{n}^{(c)}(x)$, as

$$n_i^{(c)}(x) := \begin{cases} n_c = x, & i = c, \\ n_b = \frac{1-x}{m-1}, & i \in b. \end{cases} \tag{D1}$$

We use this to find simplified parametrized forms of key variables, starting with

$$B_i(x) = \alpha_c n_c(x)\mu_{ci} + \sum_{j\neq c} \alpha_j n_b(x)\mu_{ji} + n_i(x)\omega_i \tag{D2}$$

$$= \alpha_c n_c(x)\mu_{ci} + n_b(x)(\gamma_i - \alpha_c\mu_{ci}) + n_i(x)\omega_i \tag{D3}$$

$$= \alpha_c\mu_{ci}(n_c(x) - n_b(x)) + n_b(x)\gamma_i + n_i(x)\omega_i \tag{D4}$$

$$= \alpha_c\mu_{ci}\frac{mx-1}{m-1} + \frac{1-x}{m-1}\gamma_i + n_i(x)\omega_i \tag{D5}$$

$$= \begin{cases} \frac{\alpha_c\mu_{cc}(mx-1)+\gamma_c(1-x)}{m-1} + x\omega_c, & i = c, \\ \frac{\alpha_c\mu_{ci}(mx-1)+(\gamma_i+\omega_i)(1-x)}{m-1}, & i \in b. \end{cases} \tag{D6}$$

We also consider the parametrized form of the instantaneous averages of the replication and deterioration rates

$$\langle\alpha\rangle_{\mathbf{n}} = x\alpha_c + \frac{1-x}{m-1}(m\bar{\alpha} - \alpha_c), \tag{D7}$$

$$\langle\omega\rangle_{\mathbf{n}} = x\omega_c + \frac{1-x}{m-1}(m\bar{\omega} - \omega_c). \tag{D8}$$

This allows us to evaluate the functions on the line in terms of the parametrization variable x, but we also need to project them along this line. We do this by taking the inner product with the direction vector, $\mathbf{u}^{(c)}$, into the c vertex of the simplex:

$$u_i^{(c)} := \begin{cases} u_c^{(c)} = \sqrt{\frac{m-1}{m}}, & i = c, \\ u_b^{(c)} = -\sqrt{\frac{m-1}{m}}\frac{1}{m-1}, & i \in b. \end{cases} \tag{D9}$$

Using such a construction, we first compute the noise-induced force in the $\mathbf{u}^{(c)}$ direction along the line as

$$
\begin{aligned}
f_{\Theta}^{(c)}(x) &:= \Theta_c(\mathbf{n}^{(c)}(x))u_c^{(c)} + \sum_{i\in b}\Theta_i(\mathbf{n}^{(c)}(x))u_b^{(c)} \\
&= \sqrt{\frac{m-1}{4mN^2}}\Bigg[(m+1)B_c(x) - \frac{m+1}{m-1}\sum_{i\in b}B_b(x) - (m+1)(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}})\left(n_c(x) - \frac{1}{m-1}\sum_{i\in b}n_b(x)\right) \\
&\quad + (\mu_{cc}m\bar{\alpha} + m\bar{\omega})\left(n_c(x) - \frac{1}{m-1}\sum_{j\in b}n_b(x)\right) - (\mu_{cc}\alpha_c + \omega_c) + \frac{1}{m-1}\sum_{i\in b}(\mu_{cc}\alpha_i + \omega_i)\Bigg] \\
&= \sqrt{\frac{m-1}{4mN^2}}\Bigg[(m+1)\left(\frac{\alpha_c\mu_{cc}(mx-1)+\gamma_c(1-x)}{m-1} + x\omega_c\right) - \frac{m+1}{m-1}\sum_{i\in b}\left(\frac{\alpha_i\mu_{ci}(mx-1)+(\gamma_i+\omega_i)(1-x)}{m-1}\right) \\
&\quad - (m+1)(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}})\left(\frac{mx-1}{m-1}\right) + (\mu_{cc}m\bar{\alpha} + m\bar{\omega})\left(\frac{mx-1}{m-1}\right) \\
&\quad - (\mu_{cc}\alpha_c + \omega_c) + \frac{1}{m-1}(\mu_{cc}m\bar{\alpha} - \mu_{cc}\alpha_c + m\bar{\omega} - \omega_c)\Bigg] \\
&= \sqrt{\frac{m-1}{4mN^2}}\Bigg[(m+1)\left(\frac{\alpha_c\mu_{cc}(mx-1)+\gamma_c(1-x)}{m-1} + t\omega_c\right) \\
&\quad - \frac{m+1}{m-1}\left(\frac{\alpha_c(1-\mu_{cc})(mx-1)+(m\bar{\alpha}-\gamma_c+m\bar{\omega}-\omega_c)(1-x)}{m-1}\right) - (m+1)(\langle\alpha\rangle_{\mathbf{n}} + \langle\omega\rangle_{\mathbf{n}})\left(\frac{mx-1}{m-1}\right) \\
&\quad + (\mu_{cc}m\bar{\alpha} + m\bar{\omega})\left(\frac{mx-1}{m-1}\right) - (\mu_{cc}\alpha_c + \omega_c) + \frac{1}{m-1}(\mu_{cc}m\bar{\alpha} - \mu_{cc}\alpha_c + m\bar{\omega} - \omega_c)\Bigg] \\
&= \sqrt{\frac{m}{4N^2(m-1)^3}}[\alpha_c(m\mu_{cc}(mx+x-2) + (m+1)x(1-mx)) - \omega_c((m+1)m(x-1)x + m - 1) \\
&\quad + \bar{\alpha}(m-1)mx\mu_{cc} + (m+1)(x-1)(\bar{\alpha}mx - \gamma_c) + mx\bar{\omega}(mx+x-2)]. \qquad \text{(D10)}
\end{aligned}
$$

Similarly, we can compute the other force terms. We first compute the requisite terms in terms of the parameter x:

$$
\begin{aligned}
A_i(x) &= \alpha_c n_c(x)\mu_{ci} + \sum_{j\neq c}\alpha_j n_b(x)\mu_{ji} - n_i(x)\omega_i \\
&= \begin{cases} \frac{\gamma_c(1-x)+\alpha_c\mu_{cc}(mx-1)}{m-1} - x\omega_c, & i = c, \\ \frac{(\gamma_i-\omega_i)(1-x)+\alpha_c\mu_{ci}(mx-1)}{m-1}, & i \in b, \end{cases} \qquad \text{(D11)} \\
\Phi_i(x) &= A_i(x) - n_i(x)(\langle\alpha\rangle_{\mathbf{n}}(x) - \langle\omega\rangle_{\mathbf{n}}(x)) \\
&= \begin{cases} \frac{\gamma_c(1-x)+\alpha_c\mu_{cc}(mx-1)}{m-1} - x\omega_c - x(\langle\alpha\rangle_{\mathbf{n}}(x) - \langle\omega\rangle_{\mathbf{n}}(x)), & i = c, \\ \frac{(\gamma_i-\omega_i)(1-x)+\alpha_c\mu_{ci}(mx-1)}{m-1} - \frac{1-x}{m-1}(\langle\alpha\rangle_{\mathbf{n}}(x) - \langle\omega\rangle_{\mathbf{n}}(x)), & i \in b, \end{cases} \\
&= \begin{cases} \frac{m(x-1)x\bar{\alpha}-m(x-1)t\bar{\omega}-\alpha_c\mu_{cc}+mx\alpha_c\mu_{cc}-mx^2\alpha_c+mx^2\omega_c-mx\omega_c+x\alpha_c-(x-1)\gamma_c}{m-1}, & i = c, \\ -\frac{(x-1)(m(x-1)\bar{\alpha}+\bar{\omega}(m-mx)-(mx-1)(\alpha_c-\omega_c))}{(m-1)^2} + \frac{\alpha_c\mu_{ci}(mx-1)+(x-1)(\omega_i-\gamma_i)}{m-1}, & i \in b, \end{cases} \qquad \text{(D12)} \\
N\Psi_i(x) &= -B_i + n_i(x)(\langle\alpha\rangle_{\mathbf{n}}(x) + \langle\omega\rangle_{\mathbf{n}}(x)) \qquad \text{(D13)} \\
&= \begin{cases} -\frac{\alpha_c\mu_{cc}(mx-1)+\gamma_c(1-x)}{M-1} - x\omega_c + x(\langle\alpha\rangle_{\mathbf{n}}(x) + \langle\omega\rangle_{\mathbf{n}}(x)), & i = c, \\ -\frac{\alpha_c\mu_{ci}(mx-1)+(\gamma_i+\omega_i)(1-x)}{m-1} + \frac{(1-x)}{m-1}(\langle\alpha\rangle_{\mathbf{n}}(x) + \langle\omega\rangle_{\mathbf{n}}(x)), & i \in b \end{cases} \qquad \text{(D14)} \\
&= \begin{cases} \frac{-m(x-1)x\bar{\alpha}-m(x-1)x\bar{\omega}+\alpha_c\mu_{cc}-mx\alpha_c\mu_{cc}+mx^2\alpha_c+mx^2\omega_c-mx\omega_c-x\alpha_c+(x-1)\gamma_c}{m-1}, & i = c, \\ \frac{(x-1)(m(x-1)\bar{\alpha}+m(x-1)\bar{\omega}-(mx-1)(\alpha_c+\omega_c))}{(m-1)^2} + \frac{\alpha_c\mu_{ci}(1-mx)+(x-1)(\omega_i+\gamma_i)}{m-1}, & i \in b. \end{cases} \qquad \text{(D15)}
\end{aligned}
$$

To find the resultant forces along the path, we then take the inner product with $\mathbf{u}^{(c)}$. Doing so gives

$$
\begin{aligned}
f_\Phi^{(c)}(x) &= \Phi_c(\mathbf{n}^{(c)}(x))u_c^{(c)} + \sum_{i\in b}\Phi_i(\mathbf{n}^{(c)}(x))u_b^{(c)} \\
&= \sqrt{\frac{m-1}{m}}\Bigg[\frac{m(x-1)x\bar{\alpha} - m(x-1)x\bar{\omega} - \alpha_c\mu_{cc} + mx\alpha_c\mu_{cc} - mx^2\alpha_c + mx^2\omega_c - mx\omega_c + x\alpha_c - (x-1)\gamma_c}{m-1} \\
&\quad - \frac{1}{m-1}\sum_{i\in b}\left(-\frac{(x-1)(m(x-1)\bar{\alpha} + \bar{\omega}(m-mx) - (mx-1)(\alpha_c-\omega_c))}{(m-1)^2} + \frac{\alpha_c\mu_{ci}(mx-1) + (x-1)(\omega_c-\gamma_i)}{m-1}\right)\Bigg] \\
&= \sqrt{\frac{m-1}{m}}\Bigg[\frac{m(x-1)x\bar{\alpha} - m(x-1)x\bar{\omega} - \alpha_c\mu_{cc} + mx\alpha_c\mu_{cc} - mx^2\alpha_c + mx^2\omega_c - mx\omega_c + x\alpha_c - (x-1)\gamma_c}{m-1} \\
&\quad - \Bigg(-\frac{(x-1)(m(x-1)\bar{\alpha} + \bar{\omega}(m-mx) - (mx-1)(\alpha_c-\omega_c))}{(m-1)^2} \\
&\quad + \frac{\alpha_c(1-\mu_{cc})(mx-1) + (x-1)(\omega_c(m-1) - (m\bar{\alpha}-\gamma_c))}{(m-1)^2}\Bigg)\Bigg] \\
&= \sqrt{\frac{m-1}{m}}\left[\frac{m((x-1)(mx(\bar{\alpha}-\bar{\omega}+\omega_c)+\bar{\omega}-\omega_c) - \alpha_c(mx-1)(x-\mu_{cc}) + \gamma_c - x\gamma_c)}{(m-1)^2}\right] \\
&= \sqrt{\frac{m}{(m-1)^3}}[(x-1)(mx(\bar{\alpha}-\bar{\omega}+\omega_c)+\bar{\omega}-\omega_c) - \alpha_c(mx-1)(x-\mu_{cc}) + \gamma_c(1-x)].
\end{aligned}
\tag{D16}
$$

Similarly, we find

$$
N f_\Psi^{(c)}(x) = \sqrt{\frac{m}{(m-1)^3}}[-(x-1)((mx-1)(\bar{\omega}-\omega_c) + mx\bar{\alpha}) + \alpha_c(mx-1)(x-\mu_{cc}) + (x-1)\gamma_c]. \tag{D17}
$$

This then allows us to state the general form of the forces along, and in the direction of, the simplex bisection line as

$$
f^{(c)}(x) = f_\Phi^{(c)}(x) + f_\Psi^{(c)}(x) + f_\Theta^{(c)}(x), \tag{D18}
$$

$$
f_\Phi^{(c)}(x) = \sqrt{\frac{m}{(m-1)^3}}[(x-1)(mx(\bar{\alpha}-\bar{\omega}+\omega_c)+\bar{\omega}-\omega_c) - \alpha_c(mx-1)(x-\mu_{cc}) + \gamma_c(1-x)], \tag{D19}
$$

$$
N f_\Psi^{(c)}(x) = \sqrt{\frac{m}{(m-1)^3}}[-(x-1)((mx-1)(\bar{\omega}-\omega_c) + mx\bar{\alpha}) + \alpha_c(mx-1)(x-\mu_{cc}) + (x-1)\gamma_c], \tag{D20}
$$

$$
\begin{aligned}
f_\Theta^{(c)}(x) = \sqrt{\frac{m}{4N^2(m-1)^3}}[&\alpha_c(m\mu_{cc}(mx+x-2) + (m+1)x(1-mx)) - \omega_c((m+1)m(x-1)x + m - 1) \\
&+ \bar{\alpha}(m-1)mx\mu_{cc} + (m+1)(x-1)(\bar{\alpha}mx - \gamma_c) + mx\bar{\omega}(mx+x-2)].
\end{aligned}
\tag{D21}
$$

The total force along the bisecting line is then revealed as a quadratic function of the parametrization variable x. Consequently, there are up to two fixed points, one stable and one unstable, which will control the qualitative behavior of the system. We will explore this behavior in the next sections.

APPENDIX E: FIDELITY CATASTROPHE IN THE CASE OF IDENTICALLY FIT VARIANTS

In the first instance, a simple, but instructive, fitness landscape to consider for the purpose of illustrating the fidelity catastrophe is that of equally fit variants. In this case, we have $\alpha_i = \gamma_i = \bar{\alpha} = \alpha$ and $\omega_i = \bar{\omega} = \omega$, which results in a special case where the forms of the forces reduce from a quadratic to linear. As in the preceding development, we assume only $\mu_{ij} = \mu_{ji}$ and $\mu_{ii} = \mu_{jj} = \hat{\mu}$ for all i and j. Explicitly, this yields total force and components

$$
f_\Phi^{(c)}(x) = \sqrt{\frac{m}{(m-1)^3}}\alpha(1-mx)(1-\hat{\mu}), \tag{E1}
$$

$$
f_\Psi^{(c)}(x) = \sqrt{\frac{m}{N^2(m-1)^3}}\alpha(mx-1)(1-\hat{\mu}), \tag{E2}
$$

$$
\begin{aligned}
f_\Theta^{(c)}(x) = &\sqrt{\frac{m}{4N^2(m-1)^3}}(mx-1) \\
&\times [\alpha(m(2\hat{\mu}-1)-1) + (m-1)\omega],
\end{aligned}
\tag{E3}
$$

$$
\begin{aligned}
f^{(c)}(x) = &\sqrt{\frac{m}{4N^2(m-1)^3}}[(mx-1)[2\alpha(\hat{\mu}-1)N \\
&+ \alpha(2\hat{\mu}-1)(m-1) + (m-1)\omega]].
\end{aligned}
\tag{E4}
$$

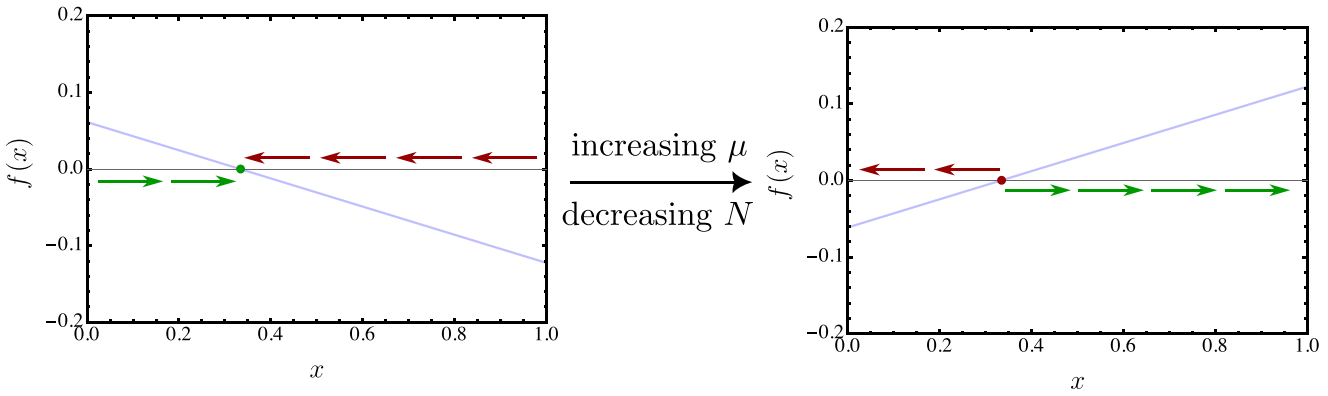

FIG. 6. Cartoon of the effective force experienced along a simplex bisection line in the case of equally fit variants. The force is a simple linear function with a fixed point at $x_* = 1/m$ (shown here for $m = 3$). As $\hat{\mu}$ is increased or N is decreased, the stability of this fixed point can change from stable (green point) to unstable (red point). When this happens, the system is forced to the points $x = 0$ and $x = 1$ and the system's distribution changes from a unimodal one centered at $x = 1/m$ to a bimodal one centered at $x = 0$ and $x = 1$. This is the fidelity catastrophe. When all such bisection lines are considered, we understand that the total distribution is m-modal across the whole simplex.

Here, there is only a single fixed point, trivially at $x = 1/m$ (i.e., the center of the simplex), and its location remains unchanged for all values of $\hat{\mu}$, N, α, and ω. The stability of this fixed point, however, can change.

As there is no master variant in this fitness landscape, there is also no concept of an error catastrophe; however, the fidelity catastrophe can still be observed, and this is related to the stability of the sole fixed point. Since the forces are simple linear functions, the stability of the fixed point is simply related to the sign of the multiplying factor to the term in $(mx - 1)$.

Explicitly, when the prefactor to the term in $(mx - 1)$ is negative, the fixed point at $x = 1/m$ is stable. As $\hat{\mu}$ is increased, or N is decreased, the prefactor can change sign, becoming positive, thus leaving the fixed point unstable. This is illustrated in Fig. 6.

When the fixed point is unstable, the distribution along the bisecting line becomes bimodal, with the edges of the simplex, namely, $x = 0$ and $x = 1$ along all bisecting lines, become effective attractors. This is precisely what we observe in Fig. 5 when we see the distribution become bimodal. In this case, when the prefactor is precisely zero, the force vanishes everywhere along the bisecting line. The expected behavior of the system when not experiencing the fidelity catastrophe, such that fixed point is stable, is that of a well-mixed system with approximately equal distribution between the variants. When the system experiences the fidelity catastrophe, such that the fixed point is unstable, we instead expect stochastic switching between temporary dominance of different single variants, with none being preferred, on average.

We can easily characterize the onset of the fidelity catastrophe in this case by simply solving for when the total force vanishes (i.e., when the fixed point changes stability). Solving in terms of a threshold N, we have

$$N_{\text{crit}}^{\text{fidel}} = \frac{(m-1)(\alpha(2\hat{\mu}-1)+\omega)}{2\alpha(1-\hat{\mu})}. \tag{E5}$$

In general terms, changing the value of N changes the relative weighting of the noise-induced force as compared to the deterministic contribution. The deterministic contribution acts to push the system into the center of the simplex causing the fixed point at $x = 1/m$ to act more like a stable fixed point, while, for sufficiently high $\hat{\mu}$, the noise-induced component acts to push the system to the edges of the simplex.

Alternatively, solving in terms of a threshold fidelity, $\hat{\mu}$, yields

$$1 - \hat{\mu}_{\text{crit}}^{\text{fidel}} = \frac{(m-1)(\alpha+\omega)}{2\alpha(m+N-1)}, \tag{E6}$$

revealing that, for suitably low N, a minimum fidelity is required to observe the switching phenomenon.

Finally, we note that the onset of the switching behavior arises due to the varied behavior of the noise-induced force, which changes as a function of $\hat{\mu}$. Specifically, the noise-induced component itself [Eq. (E3)] can act to either force the system into or away from the fixed point at $x = 1/m$ depending on the value of $\hat{\mu}$—at low $\hat{\mu}$, the noise-induced force stabilises the system pushing the state toward the fixed point; however, at high $\hat{\mu}$ the noise-induced force serves to push the system away from the fixed point.

As such, we can consider the point where the noise-induced force vanishes as a function of $\hat{\mu}$, characterizing when the noise-induced force changes its character from one that pushes inward toward the center of the simplex, to one where it pushes outward towards its edges. This is given by

$$1 - \hat{\mu}_{\text{crit}}^{\text{noise}} = \frac{m-1}{m} \times \frac{\alpha+\omega}{2\alpha}, \tag{E7}$$

which is conveniently a threshold independent of the other main control parameter, N. We note that in the case of approximately stable populations ($\alpha \sim \omega$), this threshold fidelity is a simple function of the size of the state space $\hat{\mu}_{\text{crit}}^{\text{noise}} \sim 1/m$ such that for almost any practical system (i.e., where $m \gg 1$) the threshold for the noise-induced force to direct the system

away from the center of the simplex becomes very low. We also note that we can use this threshold to empirically test our predictions of the nature of the noise-induced force, which we perform in Appendix G.

We can then consider a more realistic genetic model, as per the main text, where genomes are made of L bases in the set $\mathcal{A}$ (e.g., $\mathcal{A} = \{\mathrm{A}, \mathrm{T}, \mathrm{G}, \mathrm{C}\}$), leading to a total of $m = |\mathcal{A}|^L$ variants connected by the mutation rate matrix

$$\mu_{ij} = \left(\frac{\varepsilon}{|\mathcal{A}| - 1}\right)^{H_{ij}} (1 - \varepsilon)^{L - H_{ij}}, \tag{E8}$$

where ε is the average error rate per base during replication and H_{ij} is the Hamming distance between i and j—i.e., the number of loci where the genomes of i and j have different bases. In this case, we can identify $\hat{\mu} = (1 - \varepsilon)^L$ and $m = |\mathcal{A}|^L$. This allows us to express the per-base error rate that causes the noise-induced force to invert, viz.

$$\begin{aligned}\varepsilon_{\mathrm{crit}}^{\mathrm{noise}} &= 1 - \left(\frac{\alpha(|\mathcal{A}|^L + 1) - \omega(|\mathcal{A}|^L - 1)}{2\alpha|\mathcal{A}|^L}\right)^{\frac{1}{L}} \\ &\simeq \frac{1}{L}\ln\left(\frac{2\alpha}{\alpha - \omega}\right) + \mathcal{O}(L^{-2}) + \mathcal{O}(m^{-1}).\end{aligned} \tag{E9}$$

APPENDIX F: SINGLE DOMINANT VARIANT QUASISPECIES CASE

Next, we consider the case of a single dominant variant as often used in models of quasispecies. This is conveniently the simplest system parametrization that retains the full quadratic character of the forces along the bisecting simplex line. We characterize such a situation by considering $m - 1$ variants with replication and deterioration rate, α and ω, respectively. Then, we consider an additional single master variant, having index z, which has the same deterioration rate, ω, but an elevated replication rate, $\alpha + \Delta\alpha$. The mutation rate μ_{ij} is arbitrary except for symmetry, $\mu_{ij} = \mu_{ji}$, and equal fidelity between the variants, i.e., $\mu_{ii} = \mu_{jj} = \hat{\mu}$ for all i and j.

In this case, we have the summation variables

$$\bar{\omega} = \omega, \quad \bar{\alpha} = \alpha + \frac{\Delta\alpha}{m}, \quad \gamma_i = \alpha + \Delta\alpha\mu_{zi}, \tag{F1}$$

in turn allowing us to calculate the forces along the bisection line. We note that we must be careful to distinguish the form of the forces depending on which bisection line of the simplex we choose.

First, we consider the bisection line related to the vertex of the dominant species (such that $\alpha_c = \alpha_z = \alpha + \Delta\alpha$). This gives us forces

$$f_{\Phi}^{(z)}(x) = \sqrt{\frac{m}{(m-1)^3}}[\alpha(1 - mx)(1 - \hat{\mu}) - \Delta\alpha(m - 1)x(x - \hat{\mu})], \tag{F2}$$

$$f_{\Psi}^{(z)}(x) = -N^{-1} f_{\Phi}^{(z)}(x), \tag{F3}$$

$$f_{\Theta}^{(z)}(x) = \sqrt{\frac{m}{4N^2(m-1)^3}}\Big[(mx - 1)[\alpha(m(2\hat{\mu} - 1) - 1) + (m - 1)\omega] + \Delta\alpha(m - 1)[((m + 2)x - 1)\hat{\mu} - (m + 1)x^2]\Big]. \tag{F4}$$

On the other hand, forces along all other bisection lines (where $\alpha_c = \alpha$) are given by

$$f_{\Phi}^{(c\neq z)}(x) = \sqrt{\frac{m}{(m-1)^3}}[\alpha(1 - mx)(1 - \hat{\mu}) - \Delta\alpha(1 - x)(x - \mu_{zc})], \tag{F5}$$

$$f_{\Psi}^{(c\neq z)}(x) = -N^{-1} f_{\Phi}^{(c\neq z)}(x), \tag{F6}$$

$$\begin{aligned}f_{\Theta}^{(c\neq z)}(x) = \sqrt{\frac{m}{4N^2(m-1)^3}}&[(mx - 1)[\alpha(m(2\hat{\mu} - 1) - 1) + (m - 1)\omega] \\ &+ \Delta\alpha[(1 + m)(x - 1)x + (m - 1)x\hat{\mu} - (1 + m)(x - 1)\mu_{zc}]].\end{aligned} \tag{F7}$$

Combining the deterministic and noise-induced forces along the bisecting line for the master and other variants gives parametrized total forces

$$\begin{aligned}f^{(z)}(x) = \sqrt{\frac{m}{4N^2(m-1)^3}}&[(mx - 1)[2\alpha(\hat{\mu} - 1)N + \alpha(2\hat{\mu} - 1)(m - 1) + (m - 1)\omega] \\ &- \Delta\alpha(m - 1)(\hat{\mu} + x^2(m + 2N - 1) - \hat{\mu}x(m + 2N))],\end{aligned} \tag{F8}$$

$$\begin{aligned}f^{(c\neq z)}(x) = \sqrt{\frac{m}{4N^2(m-1)^3}}&[(mx - 1)[2\alpha(\hat{\mu} - 1)N + \alpha(2\hat{\mu} - 1)(m - 1) + (m - 1)\omega] \\ &+ \Delta\alpha(m - 1)(x(\hat{\mu} + x - \mu_{zc} - 1) + \mu_{zc}) + 2\Delta\alpha N(x - 1)(x - \mu_{zc})].\end{aligned} \tag{F9}$$

We note, setting $\Delta\alpha = 0$, such that all species become equivalent, results in both expressions being equal (as they must), but also reduces the form of the forces to the linear expressions in x found in Appendix E, with a single fixed point at $x = 1/m$.

Much of the behavior of the quasispecies system can be understood through an appreciation of the forces that appear in Eqs. (F8) and (F9). First, we consider the force along the bisecting line of the master variant, which, on its own, is enough to characterize the nature of both the error and fidelity catastrophes.

The force along the master bisecting line is a simple quadratic force with a negative term in x^2 and as such we can understand the qualitative behavior of the system in terms of the resultant, up to two, fixed points in the simplex where the force vanishes. Solving for vanishing force reveals fixed points

$$x_*^{(z)} = \frac{1}{2\Delta\alpha(m-1)(2N+m-1)}(\nu \pm \sqrt{\zeta}), \tag{F10}$$

$$\nu = \Delta\alpha\hat{\mu}(m-1)(m+2N) + \alpha m(2(\hat{\mu}-1)N + (2\hat{\mu}-1)(m-1)) + (m-1)m\omega, \tag{F11}$$

$$\begin{aligned}\zeta = {} & (-2\Delta\alpha\hat{\mu}N + m^2(\alpha(2\hat{\mu}-1) + \Delta\alpha\hat{\mu} + \omega) + m(2\alpha(\hat{\mu}-1)N - 2\alpha\hat{\mu} + \alpha + \Delta\alpha\hat{\mu}(2N-1) - \omega))^2 \\ & - 4\Delta\alpha(m-1)(m+2N-1)(2\alpha(\hat{\mu}-1)N + (m-1)(\Delta\alpha\hat{\mu}+\omega) + \alpha(2\hat{\mu}-1)(m-1)).\end{aligned} \tag{F12}$$

Similarly, we may consider the force along the bisecting line of a variant, which is not the master. This too is a quadratic but has a positive x^2 term. Solving for its vanishing fixed points yields

$$x_*^{(c\neq z)} = \frac{1}{2\Delta\alpha(2N+m-1)}(\nu \pm \sqrt{\zeta}), \tag{F13}$$

$$\nu = m^2(-2\alpha\hat{\mu}+\alpha-\omega) + m(\alpha(-2\hat{\mu}N + 2\hat{\mu} + 2N - 1) + \Delta\alpha(-\hat{\mu}) + \Delta\alpha + \omega + \Delta\alpha\mu_{zc}) + \Delta\alpha(\hat{\mu} + 2N + (2N-1)\mu_{zc} - 1), \tag{F14}$$

$$\begin{aligned}\zeta = {} & [\Delta\alpha(\hat{\mu}+2N-1) + m^2(-2\alpha\hat{\mu}+\alpha-\omega) + m(\alpha(-2\hat{\mu}N + 2\hat{\mu} + 2N - 1) + \Delta\alpha(-\hat{\mu}) + \Delta\alpha + \omega) + \Delta\alpha(m+2N-1)\mu_{zc}]^2 \\ & - 4\Delta\alpha(m+2N-1)(\alpha(-2\hat{\mu}N + 2\hat{\mu} - 2\hat{\mu}m + m + 2N - 1) - m\omega + \Delta\alpha(m+2N-1)\mu_{zc} + \omega).\end{aligned} \tag{F15}$$

First, we turn our attention to the behavior of the force along the bisection line of the master variant.

In this case, because the quadratic prefactor is negative, the upper of these fixed points is stable and associated with the concentration of probability around the dominant variant in the sense of the deterministic Eigen model. The lower fixed point is unstable and is broadly associated with noise-induced forces. The overall behavior induced by the quadratic force, and the relevance of the fixed points, is shown in Fig. 7.

The existence of the unstable fixed point marks a significant deviation from the behavior of the deterministic Eigen model. When it exists within the simplex, a large enough fluctuation (i.e., to a low enough value of x) will cause the parametrized state of the system to run away to $x = 0$ characterizing relative depletion of the master variant, such that the system is temporarily dominated by a variant that is not the master. This behavior occurs only in cases when the fidelity, $\hat{\mu}$, is above a certain threshold, as broadly observed in the equal two variant case in Fig. 5 when the distribution becomes bimodal. Consequently, as with the equal variant case in the previous section, we call this behavior the fidelity catastrophe, and we consider the threshold of this behavior to occur at the value of $\hat{\mu}$ where the unstable fixed point enters the simplex, $\hat{\mu}_{\text{crit}}^{\text{fidel}}$.

In contrast, the upper, stable, fixed point corresponds to the expected behavior of the deterministic model, serving as an attractor of the system associated with dominance of the master variant. Here, decreases in fidelity (increased error) disrupt the dominance of the master species, causing the stable fixed point to reside at progressively lower values of x. When this reduction reaches a threshold level, $\hat{\mu}_{\text{crit}}^{\text{error}}$, we say that the system has suffered from the error catastrophe.

We characterize these thresholds in the subsequent subsections.

1. Fidelity catastrophe

We first consider the fidelity catastrophe. This occurs when the fidelity, $\hat{\mu}$, is too high, or equivalently, when the ratio N/m

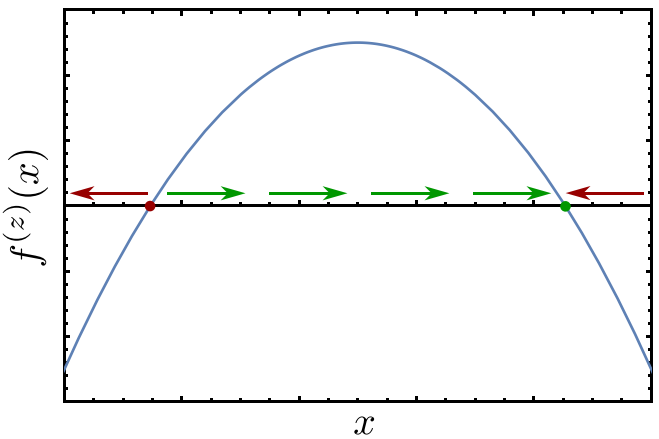

FIG. 7. Cartoon of the effective force experienced along the simplex bisection line of the master variant. The form is a negative quadratic; thus, there are two potential fixed points. A stable one (green point) that acts as an attractor, and an unstable one (red point), which repels the system. In general, the stable and unstable fixed points are bounded by $x > 0$ and $x < 1$ (the lower and upper physical limits of the system), respectively, but each point need not necessarily lie inside the simplex—i.e., the unstable fixed point can lie below $x = 0$ and the stable fixed point can lie above $x = 1$. If the unstable fixed point lies within the simplex (i.e., for $x > 0$), then the master variant extinction state ($x = 0$) effectively acts as an attractor, destabilizing unimodal quasideterministic solutions, whereas the system is repelled from the master extinction state if the fixed point lies below $x = 0$, stabilizing them. In contrast, the location of the stable fixed point determines the behavior of the system, which comports with the deterministic Eigen model.

is too low and causes edges of the simplex to become attractors. The signature of this behavior experienced in the force along the master species bisecting line is the entry into the simplex of the unstable fixed point from outside the simplex, at $x = 0$. This then causes the system in configurations near $x = 0$ to experience a force that drives them toward extinction. We can characterize this very simply by solving for vanishing force along the bisecting line at $x = 0$, i.e., $f^{(c)}(0) = 0$. Solving for this condition in terms of $\hat{\mu}$ reveals a threshold fidelity

$$\hat{\mu}_{\text{crit}}^{\text{fidel}} = \frac{\alpha(m + 2N - 1) - m\omega + \omega}{\Delta\alpha(m - 1) + 2\alpha(m + N - 1)}, \tag{F16}$$

or alternatively a threshold system size

$$N_{\text{crit}}^{\text{fidel}} = \frac{(m - 1)(\alpha(2\hat{\mu} - 1) + \Delta\alpha\hat{\mu} + \omega)}{2\alpha(1 - \hat{\mu})}. \tag{F17}$$

Note this definition subsumes the definitions introduced in the equal variants case since setting $\Delta\alpha = 0$ recovers Eqs. (E5) and (E6).

We note that we may express the threshold fidelity in terms of an upper bound threshold per-base error rate so that the fidelity catastrophe is observed for $\varepsilon < \varepsilon_{\text{crit}}^{\text{fidel}} = 1 - (\hat{\mu}_{\text{crit}}^{\text{fidel}})^{1/L}$, where

$$\begin{aligned}\varepsilon_{\text{crit}}^{\text{fidel}} &= 1 - \left(\frac{\alpha(m + 2N - 1) - m\omega + \omega}{2\alpha(m + N - 1) + \Delta\alpha(m - 1)}\right)^{\frac{1}{L}} \\ &= 1 - \left(\frac{\alpha(|\mathcal{A}|^L + 2N - 1) - |\mathcal{A}|^L\omega + \omega}{2\alpha(|\mathcal{A}|^L + N - 1) + \Delta\alpha(|\mathcal{A}|^L - 1)}\right)^{\frac{1}{L}} \\ &\simeq \frac{1}{L}\ln\left(\frac{2\alpha + \Delta\alpha}{\alpha - \omega}\right) + \mathcal{O}(L^{-2}) + \mathcal{O}(m^{-1}).\end{aligned} \tag{F18}$$

Note, in this limit (and with $\Delta\alpha = 0$), this is the same threshold as for the noise-induced becoming repulsive away from the center of the simplex in Eq. (E9). In other words, in this limit the system is dominated by noise/drift and so the behavior, and thus the presence of the fidelity catastrophe, depends on f_Θ alone. This relation $\varepsilon_{\text{crit}}^{\text{fidel}} \sim L^{-1}$ is significant as it mirrors the expected behavior of the error catastrophe (see below).

For completeness, we note that the calculation of $\hat{\mu}_{\text{crit}}^{\text{fidel}}$ through $f^{(c)}(0) = 0$ can be trivially extended to the general fitness landscape case since $f^{(c)}(x)$ remains quadratic. Specifically, assuming $c = z$ is a master sequence [i.e., so that $f^{(c)}(x)$ is a negative quadratic, in the deterministic limit, generically achieved with $\lim_{N\to\infty} \gamma_c/\hat{\mu} \lesssim \alpha_c$], the condition becomes

$$\hat{\mu}_{\text{crit}}^{\text{fidel,gen}} = \frac{\left(\sum_{i\neq c}\mu_{ic}\alpha_i\right)(m + 2N - 1) + (3 - m + 2N)\omega_c - 2(N + 1)\bar{\omega}}{\alpha_c(m - 1)}, \tag{F19}$$

where we have written $\gamma_c = \alpha_c\hat{\mu} + \sum_{i\neq c}\mu_{ic}\alpha_i$. Note, generally, μ_{ic} in the above is itself a function of $\hat{\mu}_{\text{crit}}^{\text{fidel,gen}}$ rendering the above a condition rather than a formula. For instance, this constraint implies the specific cases above under substitution $\sum_{i\neq c}\mu_{ic}\alpha_i \to (1 - \hat{\mu}_{\text{crit}}^{\text{fidel,gen}})\alpha$. Finally, we note that in the general case the system cannot be assumed to possess the symmetries around the projection line found in the flat and single dominant variant fitness landscapes considered in this contribution, possibly altering the effectiveness of our force projection technique. As per the discussion in the main text, we acknowledge the importance of such questions for future work; however, we consider such questions out of scope here.

2. Error catastrophe

Since this single dominant variant quasispecies fitness landscape possesses a well-defined master variant, it is instructive to consider the manifestation of the classic error catastrophe or error threshold in this system.

This fixed point does not have as simply defined a threshold as with the fidelity catastrophe. Instead, we note that for appropriate system parameters, as $\hat{\mu}$ is reduced (error is increased), the system experiences a rapid change between a regime of approximately linear response in the position of $x_*^{(z)}$ to a regime where the position of the fixed point is relatively independent of $\hat{\mu}$ near $x = 1/m$. We choose to characterize this crossover behavior in terms of a threshold through the point of maximal curvature in the position of the fixed point with respect to $\hat{\mu}$, i.e., $d^3x_*^{(z)}(\hat{\mu})/d\hat{\mu}^3 = 0$. Since the fixed point is the solution of a quadratic, this has an analytical solution and is given by

$$\hat{\mu}_{\text{crit}}^{\text{error}} = \frac{A}{B}, \tag{F20}$$

$$\begin{aligned}A = {} & \Delta\alpha(m - 1)(\alpha(m + 2N - 1)(m^2 + 2(m + 2)N + 4m - 4) \\ & - (m - 1)m\omega(m + 2N)) + 2\alpha m^2(m + N - 1) \\ & \times (\alpha(m + 2N - 1) - m\omega + \omega) \\ & + 2(\Delta\alpha)^2(m - 1)^2(m + 2N - 1),\end{aligned} \tag{F21}$$

$$B = (2\alpha m(m + N - 1) + \Delta\alpha(m - 1)(m + 2N))^2. \tag{F22}$$

Such an expression amounts to a powerful generalization of the original heuristic formula for the error threshold, containing both population and phase space/genome size dependence through N and m, respectively. As such, the location of the threshold can deviate significantly from that originally postulated by Eigen if N is not large enough compared to m to effect a deterministic limit or generally if m is independently not large. However, we can easily obtain the original result by progressively considering the deterministic, and large m, limits. First, we expand in orders of N^{-1} to obtain

$$\hat{\mu}_{\text{crit}}^{\text{error}} = \frac{\alpha(m(\Delta\alpha + m(\alpha + \Delta\alpha)) - 2\Delta\alpha)}{(\Delta\alpha - m(\alpha + \Delta\alpha))^2} + \mathcal{O}(N^{-1}), \tag{F23}$$

thus revealing the location of the error threshold in the deterministic, but small state space, limit. By then expanding in

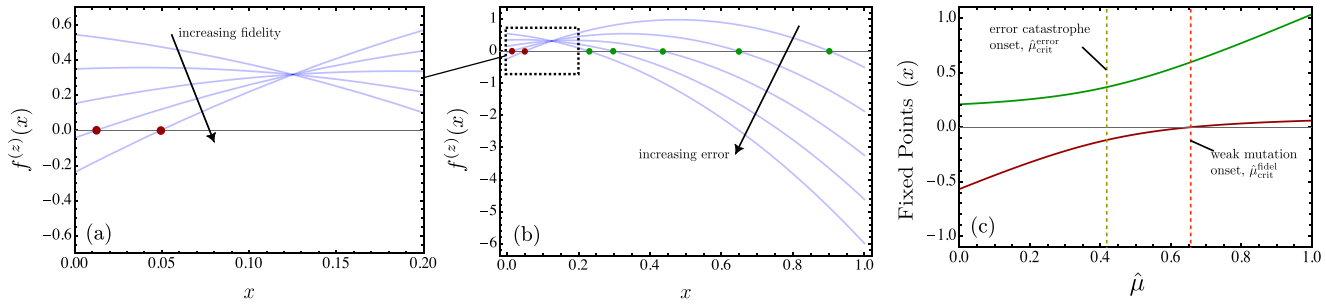

FIG. 8. Effective force experienced along the simplex bisection line of the master variant for a system with $m = 3, N = 10, \alpha = 1, \Delta\alpha = 4$, and $\omega = 1/2$. Blue lines are the effective force along the simplex bisecting line and vary between fidelity parameter, $\hat{\mu}$, taking values from 0.1 to 0.9 in steps of 0.2. As the fidelity is increased, the lower fixed point (red dots) enters the simplex such that the force at $x = 0$ becomes negative. The behavior of this fixed point is shown by the red line in panel (c) with the onset of the fidelity catastrophe defined as the point where it enters the simplex at $x = 0$. On the other hand, as the fidelity is reduced (error is increased), the upper, stable, fixed point (green dots) moves to lower values of x in the simplex. The behavior of this fixed point is shown by the green line in panel (c) with the onset of the error catastrophe defined by the point of maximum curvature in the fixed point position with $\hat{\mu}$.

powers of m^{-1}, we find

$$\hat{\mu}_{\rm crit}^{\rm error} = \frac{\alpha}{\alpha + \Delta\alpha} + \frac{3\alpha\Delta\alpha}{(\alpha + \Delta\alpha)^2}\frac{1}{m} + \mathcal{O}(m^{-2}) + \mathcal{O}(N^{-1}), \tag{F24}$$

agreeing with the classical result [9,10] as $m \to \infty$.

Note, however, that this agreement holds in the limit $N \gg m \gg 1$ only. If we exchange the order of the limits such that we consider $m \gg N$, we instead obtain

$$\hat{\mu}_{\rm crit}^{\rm error} = \frac{\alpha - \omega}{2\alpha + \Delta\alpha} + \frac{2(\alpha^2 N + N\omega(\alpha + \Delta\alpha) + 2\alpha\Delta\alpha + (\Delta\alpha)^2)}{(2\alpha + \Delta\alpha)^2}\frac{1}{m} + \mathcal{O}(m^{-2}). \tag{F25}$$

In terms of a threshold per-base error rate, the error catastrophe occurs when $\varepsilon > \varepsilon_{\rm crit}^{\rm error}$, where

$$\varepsilon_{\rm crit}^{\rm error} \simeq \begin{cases} \frac{1}{L}\ln\left(\frac{2\alpha+\Delta\alpha}{\alpha-\omega}\right) + \mathcal{O}(L^{-2}) + \mathcal{O}(m^{-1}), & m \gg N, \\ \frac{1}{L}\ln\left(\frac{\alpha+\Delta\alpha}{\alpha}\right) + \mathcal{O}(L^{-2}) + \mathcal{O}(N^{-1}), & N \gg m \gg 1. \end{cases} \tag{F26}$$

3. Relationship between the error and fidelity catastrophes

As can be observed from the limiting forms of the error and fidelity catastrophe thresholds above, in the limit $m \gg N$, these two thresholds converge on each other. Specifically, we find

$$\frac{\hat{\mu}_{\rm crit}^{\rm fidel}}{\hat{\mu}_{\rm crit}^{\rm error}} = 1 + \frac{2\Delta\alpha}{2\alpha + \Delta\alpha}\frac{N}{m} + \mathcal{O}(m^{-1}). \tag{F27}$$

Consequently, in this highly relevant limit, $m \gg N$, there is a fundamental trade-off between these two (potentially) deleterious behaviors. As such, unless the number of replicators can be taken large enough that it exceeds the number of variants, assuming that the error catastrophe must be avoided in nature, we expect to see noise-induced effects causing the master species to be unstable over time.

4. Illustration of key behaviors

We now illustrate the main behavior of the master variant fixed points and their relation to the error and fidelity catastrophe thresholds.

In Fig. 8, we illustrate the position of the fixed points with varying fidelity, $\hat{\mu}$, for an illustrative set of system parameters. In the left two subpanels are the effective forces along the bisecting line of the simplex for the master variant, with fixed points marked as red and green dots. The behavior of the position of these fixed points is then plotted in the rightmost subpanel. As $\hat{\mu}$ is increased (or equivalently, error is reduced), we observe the lower fixed point enter the simplex (red dots), causing total depletion of the master variant to become a stable fixed point associated with the fidelity catastrophe. Conversely, as we reduce the fidelity (increase the error) the upper fixed point moves to lower and lower positions (values of x) within the simplex, causing relative depletion of the master variant, eventually associated with the error catastrophe.

In Fig. 9, we illustrate the effect of varying the state space of the system (number of variants, m) on the positions of the fixed point within the simplex. Shown are positions of the fixed points varying with $\hat{\mu}$ for different values of m. Here, we see the resolution of the error catastrophe become much clearer, with the vertical dashed green lines increasingly separating distinct behavior as m increases. Moreover, we can also clearly see the prediction of Eq. (F27) with the two thresholds tending to each other.

Finally, in Fig. 10 we illustrate the effect of varying the number of replicators in the system, N, on the positions of the fixed point within the simplex. Shown are positions of the fixed points varying with $\hat{\mu}$ for different values of N at fixed m. Here, we see that the onset of the fidelity catastrophe is clearly a stochastic effect, with larger and larger fidelities (tending to 1) required to observe the effect as N is taken larger. We also note that in this figure (specifically in the top right subpanel), we consider the model in the approach to the joint limit $N \gg m$ and $m \gg 1$ required for it to replicate the deterministic limit of the large state space regime. As such, the error threshold here can be seen to reside at $\sim$0.2, exactly at the value predicted by Eq. (F24) and which relates to the classic error threshold.

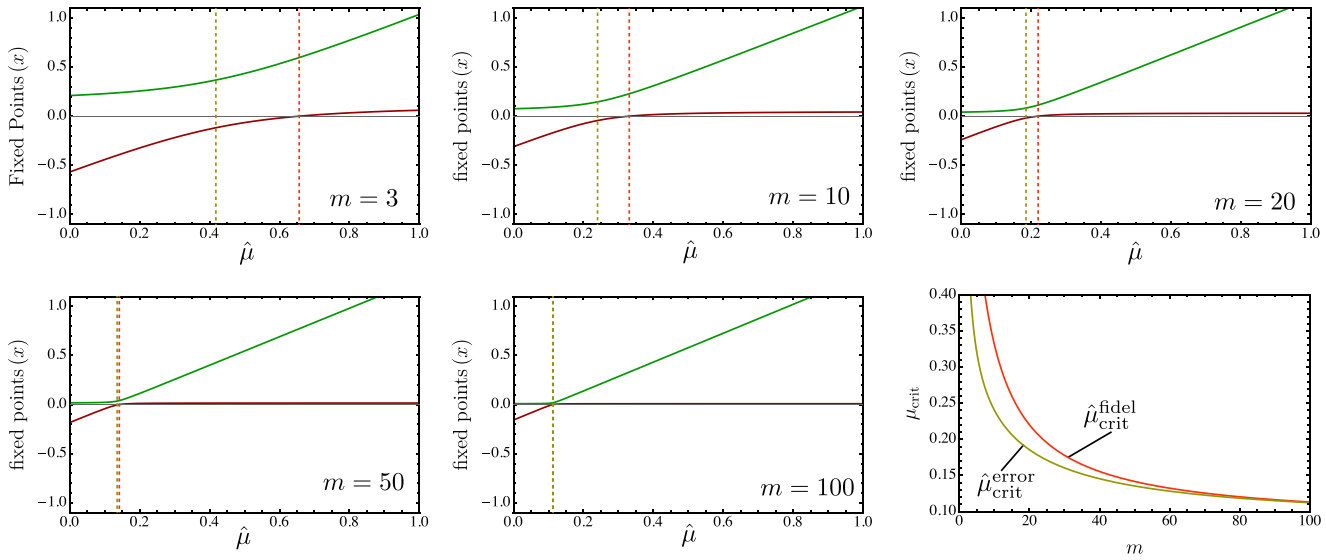

FIG. 9. Behavior of the fixed points with $\hat{\mu}$, and related catastrophe thresholds, for different values of m. Keeping N fixed for a system with parameters $N = 10$, $\alpha = 1$, $\Delta\alpha = 4$, and $\omega = 1/2$, increasing m increases the ratio m/N reducing the number of replicators per variant. From top left, through top right, to lower center the location of the fixed points is shown for increasing m from $m = 3$ to $m = 100$. Green lines indicate the position of the upper, stable, fixed point, while the red lines indicate the position of the lower, unstable, fixed point. The subpanel in the lower right shows the position of the onset of the error catastrophe (green) and fidelity catastrophe (red) as a function of m. As m increases the position of the thresholds moves to lower values of $\hat{\mu}$. Moreover, the locations of the thresholds tend closer to each other, reducing the region of error/fidelity space, $\hat{\mu}$, where one of the phenomena is not present. Further, as m increases, the behavior around the thresholds can be seen to get sharper, with the error catastrophe, in particular, becoming a more clearly resolved phenomenon. We note that at large m the center of the simplex is at lower values of x (i.e., at $x \sim 1/m$) such that it becomes harder to see the fidelity catastrophe on the linear scales shown in the above plots.

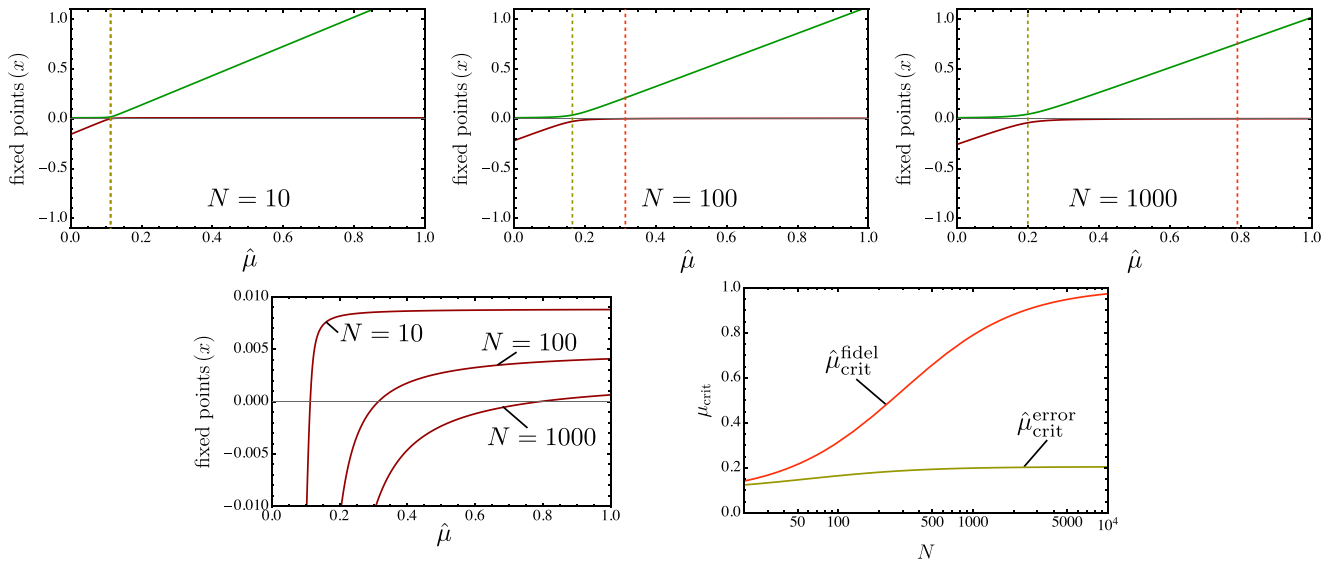

FIG. 10. Behavior of the fixed points with $\hat{\mu}$, and related catastrophe thresholds, for different values of N. Keeping m fixed for a system with parameters $m = 100$, $\alpha = 1$, $\Delta\alpha = 4$, and $\omega = 1/2$, increasing N decreases the ratio m/N increasing the number of replicators per variant. From top left to top right, the locations of the fixed points are shown for increasing N from $N = 10$ to $N = 1000$. Green lines indicate the position of the upper, stable, fixed point, while the red lines indicate the position of the lower, unstable, fixed point. The subpanel in the lower left shows an enlarged view of the position of the lower fixed point for all values of N. The subpanel in the lower right shows the position of the onset of the error catastrophe (green) and fidelity catastrophe (red) as a function of N on a logarithmic scale. As N increases the position of the thresholds moves to higher values of $\hat{\mu}$. Moreover, the locations of the thresholds tend away from each other, increasing the region of error/fidelity space, $\hat{\mu}$, where neither phenomena (error catastrophe and weak mutation regime) is present. In particular, as N increases far beyond the value of m, such that there are many replicators per variant, the fidelity required to achieve the fidelity catastrophe tends to 1. We also note that for $N \gg m$ and $m \gg 1$, most closely replicated by the top right panel ($N = 1000$) the error threshold tends to the classical result $\alpha/(\alpha + \Delta\alpha) = 0.2$.

In totality, the central observation that underlies the behavior is the importance of the ratio m/N. When m/N is small, there is a large number of replicators per variant and thus stochastic effects are small. In this limit, the fidelity catastrophe only occurs for extremely large values of $\hat{\mu}$ ($\sim$1) and generally does not occur in the vicinity of the error catastrophe. In addition, if the state space is independently large ($m \gg 1$), then the error threshold coincides with the conventional limit associated with the Eigen model, $\alpha/(\alpha + \Delta\alpha)$. In contrast, when m/N is large, there are, on average, very few replicators per variant (i.e., significantly less than one) and thus stochastic effects are highly significant. In this limit, the error and fidelity thresholds coincide such that it becomes practically impossible not to observe one of the catastrophes for a given set of system parameters. It is then not irrelevant that for realistic models of genomes, the number of variants m will be extremely large. It is instructive to then consider the role and relevance of the assumptions that underpin the traditional, deterministic, Eigen model, which, by ignoring stochastic effects, essentially assumes $N \gg m$.

We have established thresholds for the fidelity and error catastrophes from the total force experienced by the master variant; however, this has only established values of $\hat{\mu}$ and N where an extinction force arises for the master variant and where the position of the stable fixed point for the master variant qualitatively changes behavior, respectively. It is instructive, therefore, to also examine the behavior of the nonmaster variants. The nature of the force on the nonmaster variants is inherently more complicated since it depends on the mutation fraction μ_{zc}. These values, in concert with the other parameters, will control the overall qualitative behavior of the system in a nontrivial way; however, we can still find analogous thresholds in the behavior of these nonmaster variants.

First, we consider behavior in a nonmaster variant that is analogous to the fidelity catastrophe. Here, at large enough $\hat{\mu}$ and low enough N, this variant, which is not expected to be dominant in the deterministic Eigen model, can develop an attractor at $x = 1$ (i.e., population dominance) when the upper, unstable, fixed point enters the simplex at $x = 1$. This is because the form of the force along this bisection line is a quadratic with positive curvature. Moreover, by defining a threshold for this behavior in terms of when this fixed point enters the simplex (by solving for vanishing force at $x = 1$) we find a result that is independent of μ_{zc}. Solving in terms of a threshold $\hat{\mu}$ and N here yields

$$\hat{\mu}^{\text{fidel}}_{\text{crit},\neq z} = \frac{\alpha(m + 2N - 1) + \omega - m\omega}{\Delta\alpha + 2\alpha(m + N - 1)} \tag{F28}$$

and

$$N^{\text{fidel}}_{\text{crit},\neq z} = \frac{\Delta\alpha\hat{\mu} + \alpha(2\hat{\mu} - 1)(m - 1) + (m - 1)\omega}{2\alpha(1 - \hat{\mu})}. \tag{F29}$$

These expressions are very similar to, but not exactly equal to the earlier expressions for $\hat{\mu}^{\text{fidel}}_{\text{crit}}$ and $N^{\text{fidel}}_{\text{crit}}$, differing by a factor of $(m - 1)$ into terms in $\Delta\alpha$ in the denominator and numerator, respectively. As a consequence, these thresholds occur at larger and smaller values of $\hat{\mu}$ and N, respectively.

Similarly, behavior in the lower, stable, fixed point exhibits a response that acts like an analog to the error catastrophe. Specifically, as $\hat{\mu}$ is decreased (such that error is increased) the fixed point can also enter the simplex, such that extinction in the nonmaster variant can cease to locally be the most stable configuration of the system. We can solve for when this occurs yielding a threshold

$$\hat{\mu}^{\text{error}}_{\text{crit},\neq z} = \frac{\alpha(m + 2N - 1) - m\omega + \omega + \Delta\alpha(m + 2N - 1)\mu_{\text{zc}}}{2\alpha(m + N - 1)} \tag{F30}$$

and

$$N^{\text{error}}_{\text{crit},\neq z} = \frac{(m - 1)(\alpha(2\hat{\mu} - 1) + \omega - \Delta\alpha\mu_{\text{zc}})}{2(\alpha(1 - \hat{\mu}) + \Delta\alpha\mu_{\text{zc}})}. \tag{F31}$$

These thresholds, however, depend on the mutation fraction μ_{zc}.

The behavior of the force for a nonmaster variant, alongside the behavior of the fixed points and their relation to the above thresholds, is shown in Fig. 11.

APPENDIX G: EMPIRICAL EVIDENCE FOR THE NOISE-INDUCED FORCE

To confirm the qualitative behavior of the noise-induced forces that underpin the overall behavior, we can simulate the noise terms for the system in a trivial case and show that it possesses the same behavior predicted by Eq. (D10). Specifically, we simulate the SDEs in the case $m = 2$ and $\alpha_i = \omega_i = 1$, for all i, but remove the deterministic component to fully isolate the noise-induced behavior. In this case, the SDEs reduce to

$$dn_1 = \sum_{j=1}^{2} q_{1j} dW_j = (1 - n_1)\sqrt{B_1} dW_1 - n_1\sqrt{B_2} dW_2, \tag{G1}$$

$$dn_2 = \sum_{j=1}^{2} q_{2j} dW_j = (1 - n_2)\sqrt{B_2} dW_2 - n_2\sqrt{B_1} dW_1, \tag{G2}$$

where

$$B_1 = \sum_{j=1}^{2} n_j \alpha \mu_{j1} + n_1\omega = n_1\alpha\mu_{11} + n_2\alpha\mu_{21} + n_1\omega, \tag{G3}$$

$$B_2 = \sum_{j=1}^{2} n_j \alpha \mu_{j2} + n_2\omega = n_2\alpha\mu_{22} + n_1\alpha\mu_{12} + n_2\omega. \tag{G4}$$

As per the assumptions of the model and dimension of the current example, we then have $\mu_{11} = \mu_{22} = \hat{\mu}$ and $\mu_{12} = \mu_{21} = 1 - \hat{\mu}$. For these values, Eq. (D10) simplifies to

$$N f^{(c)}_{\Theta(x)}(x) = 2\sqrt{2}(1 - 2x)(1 - 2\hat{\mu}). \tag{G5}$$

Crucially, this predicts that the force has a zero at $x = 1/2$, i.e., the fully mixed state $n_1 = n_2 = 1/2$, and, moreover, the direction of this force inverts from one that pushes toward this center point, to one that pushes outward from this point, at a value of $\hat{\mu} = 1/2$, also in agreement with Eq. (E7). Simulating the above SDEs under the additional stipulation of reflecting boundaries exactly corroborates this behavior as shown in the probability density functions of system state shown in Fig. 12, indicating that it has been formulated correctly.

Finally, for completeness, we offer empirical evidence for the form of the noise-induced force in the slightly less trivial

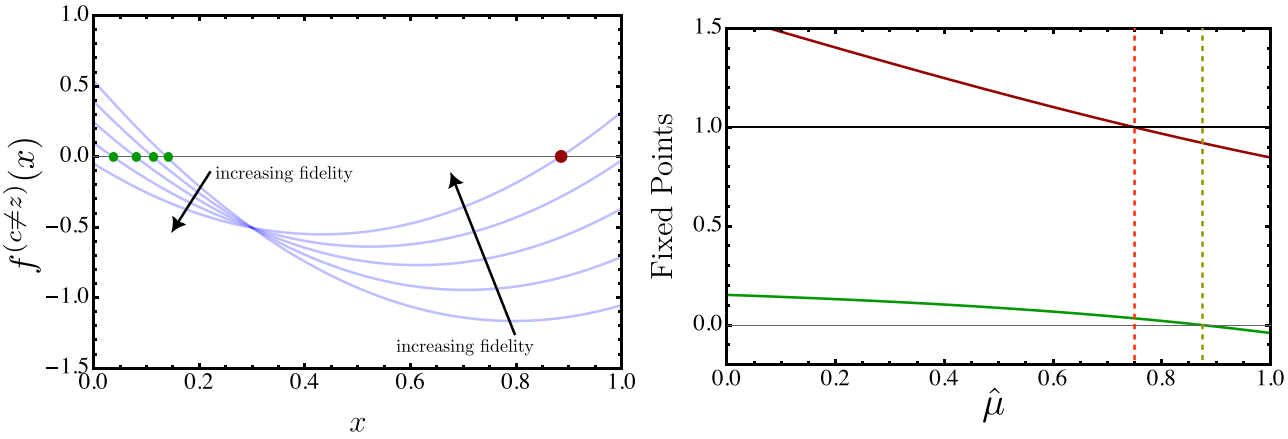

FIG. 11. Effective force experienced along the simplex bisection line of a nonmaster variant, c, for a system with $\mu_{zc} = 0$, $m = 3$, $N = 10$, $\alpha = 1$, $\Delta\alpha = 4$, and $\omega = 1/2$. Blue lines are the effective force along the simplex bisecting line and vary with the fidelity parameter, $\hat{\mu}$, taking values from 0.1 to 0.9 in steps of 0.2. As the fidelity is increased, the upper fixed point (red dots) enters the simplex such that the force at $x = 1$ becomes positive. The behavior of this fixed point is shown by the red line in the right subpanel with the threshold where it enters the simplex at $x = 1$ shown by the vertical red line. As fidelity is increased further, the lower, stable, fixed point (green dots) moves out of the simplex, such that extinction becomes a stable fixed point. The behavior of this fixed point is shown by the green line in the right subpanel with the threshold shown by the vertical green line.

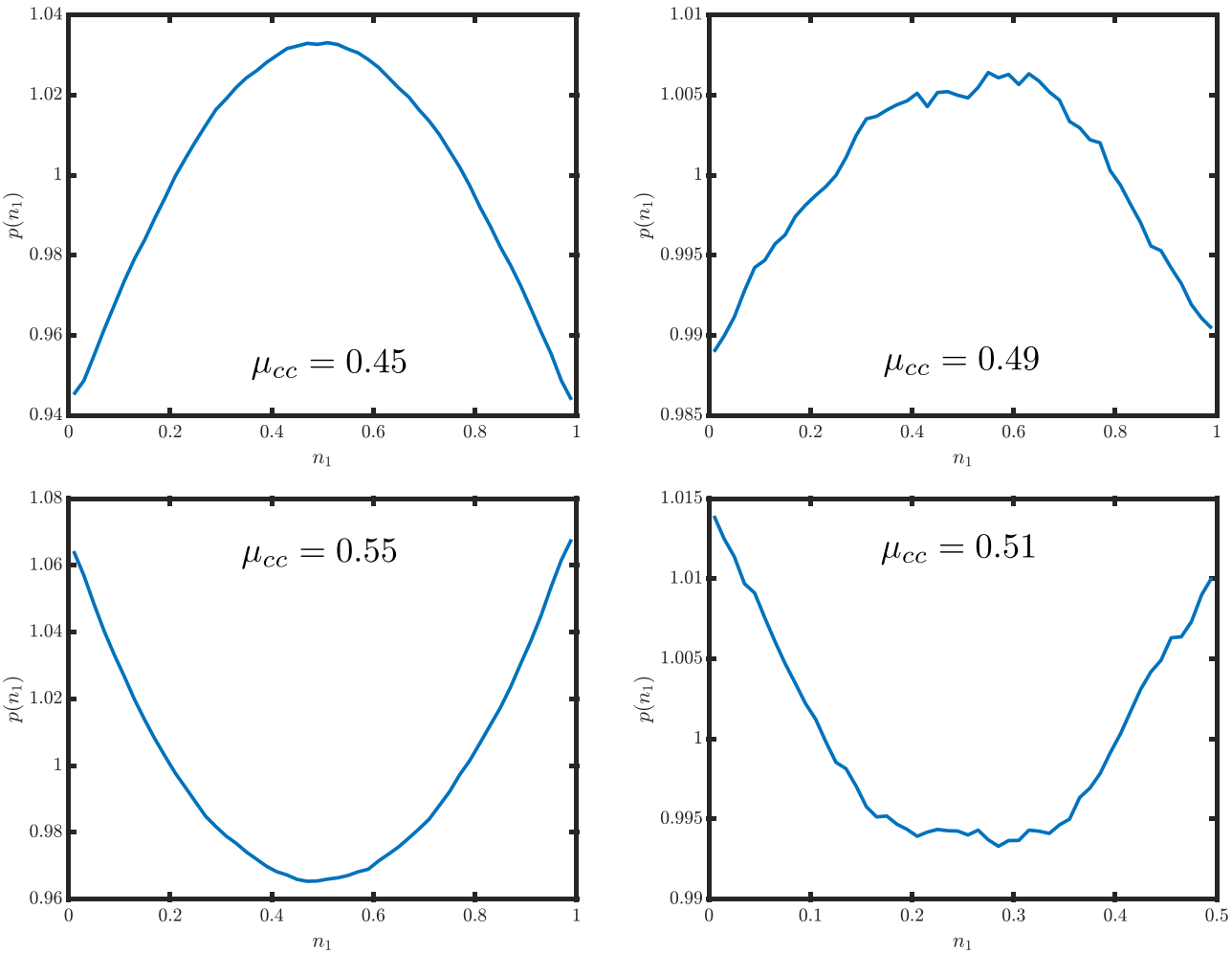

FIG. 12. Empirical probability density functions for the system comprised only of stochastic terms for $m = 2$, $\alpha_i = \omega_i = 1$, for all i, and $\mu_{11} = \mu_{22} = \hat{\mu}$. The behavior inverts at $\hat{\mu} = 1/2$, as predicted by Eq. (E7).

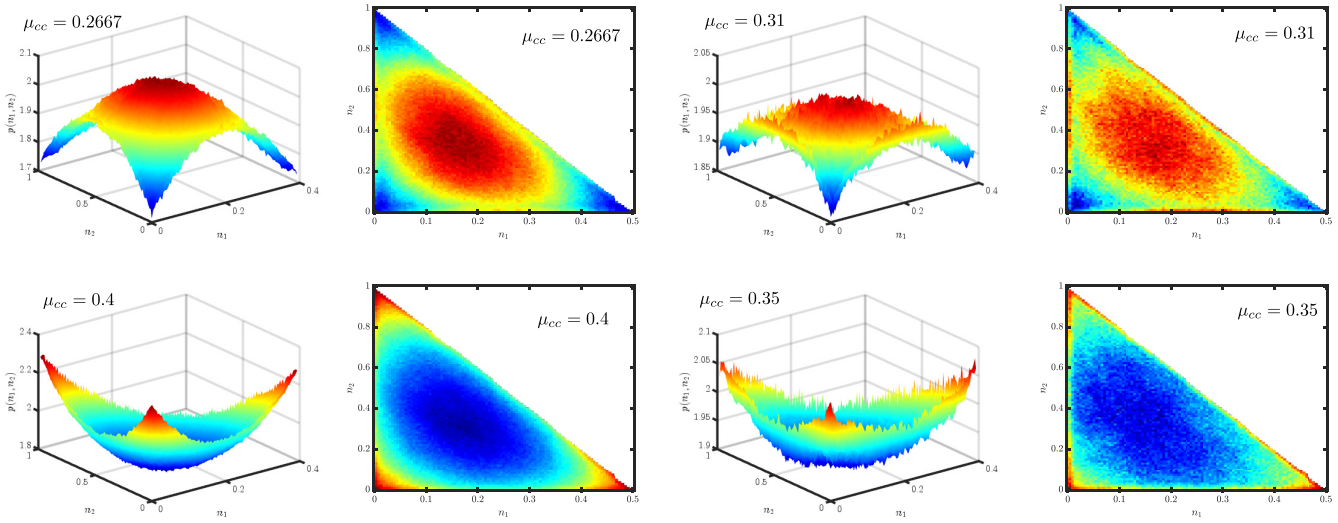

FIG. 13. Empirical probability density functions for the system comprised only of stochastic terms for $m = 3$, $\alpha_i = \omega_i = 1$, for all i, and $\mu_{11} = \mu_{22} = \mu_{33} = \hat{\mu}$ and $\mu_{ij} = (1 - \hat{\mu})/2$ for all $i \neq j$. The behavior inverts at $\hat{\mu} = 1/3$, as predicted by Eq. (E7).

case of $m = 3$. Here, Eq. (E7) predicts that such an inversion (for the same parameters) occurs at $\hat{\mu} = 1/3$. Again, we see exact corroboration of this behavior, as seen in the probability density functions, for the system comprising only the stochastic terms, shown in Fig. 13.

APPENDIX H: SIMULATION

1. Gillespie simulation and saturation limit

Here, we discuss numerical simulation of the underlying system

$$N_i \xrightarrow{\sum_j N_j \alpha_j \mu_{ji}} N_i + 1, \tag{H1a}$$

$$N_i \xrightarrow{N_i \omega_i} N_i - 1, \tag{H1b}$$

which underpins the master equation in Eq. (A1) that has formed the basis of our analysis. Moreover, such simulations also generate the data shown in Fig. 3.

We simulate the system starting with $N(t = 0) = N = \sum_{i=1}^{m} N_i(t = 0)$ individuals at time $t = 0$ through the use of the Gillespie algorithm, wherein events occur at intervals in time according to an exponential distribution with rate $\Lambda = \sum_{i=1}^{m} N_i(t)(\alpha_i + \omega_i)$, where α_i and ω_i are the replication and deterioration rates of variant i, respectively. The event that occurs may be a deterioration of a member of variant i, $(N_i \to N_i - 1)$, chosen with probability $N_i\omega_i/\Lambda$, or production of a member of variant j due to replication of variant i, $(N_j \to N_j + 1)$, chosen with probability $N_i\alpha_i\mu_{ij}/\Lambda$, such that the probabilities of all possible events sum to 1. Individual mutation rates, μ_{ij}, are determined as per

$$\mu_{ij} = \left(\frac{\varepsilon}{|\mathcal{A}| - 1}\right)^{H_{ij}} (1 - \varepsilon)^{L - H_{ij}}, \tag{H2}$$

where ε is the uniform replication error rate *per base*, $\mathcal{A}$ is the alphabet of values a base can take, L is the genome length, and H_{ij} is the Hamming distance between variants i and j—i.e., the number of loci where the genomes of i and j have distinct bases.

Application of such rules leads to a simulation of the system where $N(t)$ varies in time from its initial value $N(0) = N$. As such, in order to investigate the effects of a particular N, we then modify these dynamics such that they remain close to $N(t) \simeq N = N(0)$ throughout the simulation. This is achieved by dynamically modifying the replication rates, $\alpha_i \to \tilde{\alpha}_i(t)$, according to

$$\tilde{\alpha}_i(t) = \alpha_i \left[\lambda - \frac{1}{1 + e^{-\kappa(N(t) - N(0))}}\right]. \tag{H3}$$

Here, κ and λ are then parameters, which we set as $\lambda = 3/2$ and $\kappa = 10/N(0)$, such that $\tilde{\alpha}_i(t) = \alpha_i$ when $N(t) = N(0) = N$. This will then be a fixed point of the system in cases where we have $\langle \alpha \rangle_{\mathbf{n}} \sim \langle \omega \rangle_{\mathbf{n}}$, which is the case in our examples.

2. Matching of $N \to \infty$ quasispecies solutions and motivation

In Fig. 3, we present data from simulation across distinct combinations of ε (the uniform per-base replication error rate) and N, for the quasispecies model. The fidelity thus varies according to Eq. (H2) with $i = j$ such that

$$\hat{\mu} = (1 - \varepsilon)^L, \tag{H4}$$

where we recall L is the genome length.

In the case of the quasispecies model, once ε, $N(0)$, L, and $|\mathcal{A}|$ are set, the system is fully specified by choosing the remaining parameters, α, ω, and $\Delta\alpha = \alpha_z - \alpha$, where z is the index that relates to the wild type.

In the data presented in the main text (and in Fig. 14), α and ω are fixed as the constants $\alpha = 1$ and $\omega = 1$. However, in contrast, we choose to vary $\Delta\alpha$ as a function of the per-base error rate, ε. Specifically, we set

$$\frac{\Delta\alpha}{\alpha\varepsilon} = \text{const}, \tag{H5}$$

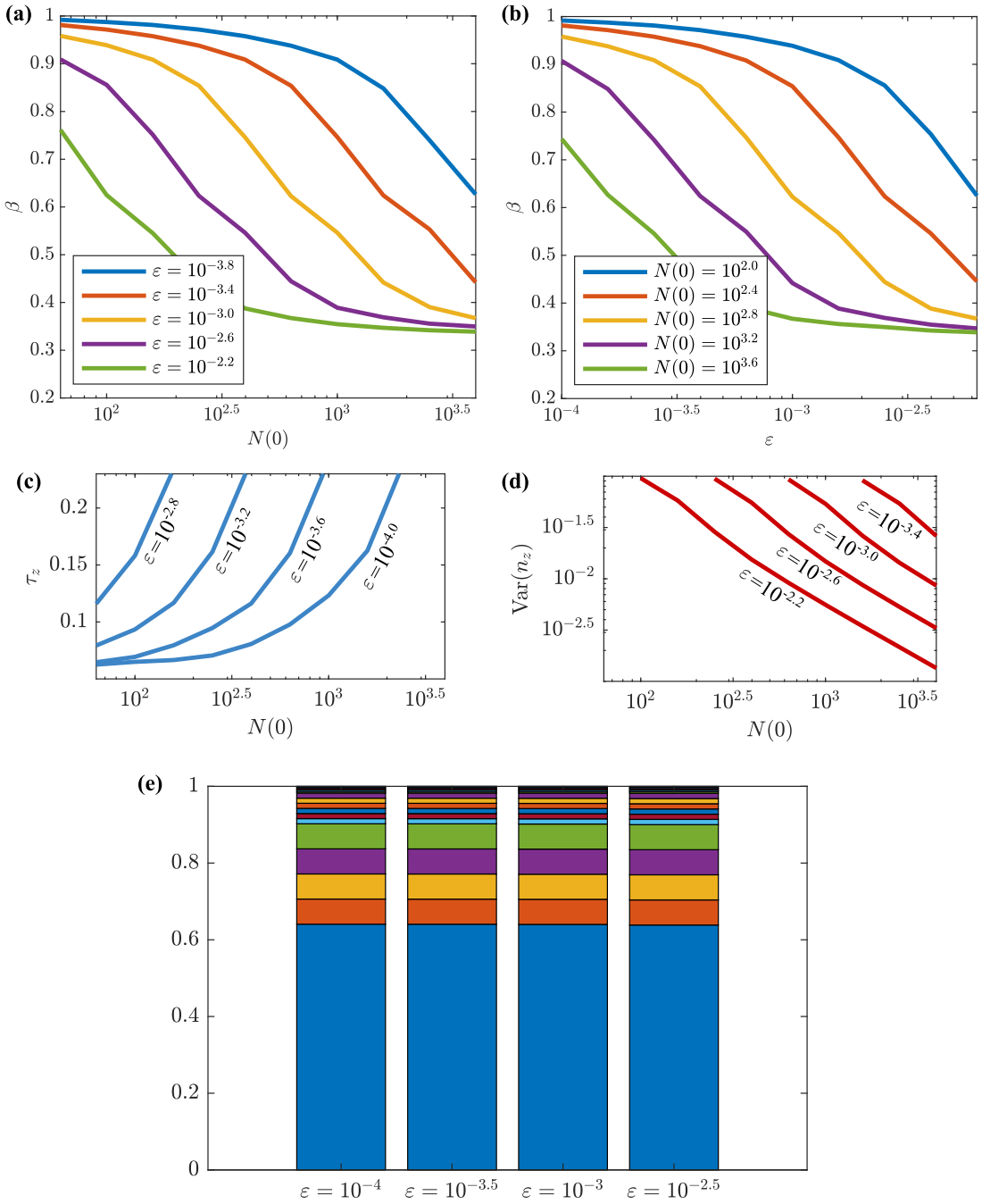

FIG. 14. Simulation data for the quasispecies example with $m = 16$ ($|\mathcal{A}| = 2$, $L = 4$), $\alpha = 1$, $\omega = 1$, $\hat{\mu} = (1-\varepsilon)^L$, and $\Delta\alpha = 10\varepsilon$. Panel (a) shows the response of the bimodality coefficient with $N(0)$ at fixed ε, while panel (b) shows its response with ε at fixed $N(0)$. Panel (c) shows the response to varying $N(0)$ of the empirical time fraction for which the most abundant variant in $\mathbf{n}$ is the wild type. Panel (d) shows the response to varying $N(0)$ of the variance of the empirical distribution for the fractional abundance of the wild type, $\rho(n_z)$. Panel (e) shows the population states, $\mathbf{n}' = (n'_1, \ldots, n'_m)$, associated with the solution of the deterministic ($N \to \infty$) Eigen model, $p^{\mathrm{det}}(\mathbf{n}) = \delta(\mathbf{n}' - \mathbf{n})$, for varying ε (and thus $\hat{\mu}$ and $\Delta\alpha$; cf. Appendix H 2), with the largest fraction (blue) corresponding to the wild type.

and, in particular, for Fig. 3, choose const = 10. This choice is not arbitrary; setting $\Delta\alpha/(\alpha\varepsilon) = \mathrm{const}$ ensures that the limiting deterministic solution is approximately independent of ε, for $\varepsilon \ll 1$.

In this Appendix, we first demonstrate this claim before giving a rationalization for why enforcing ε independent distributions in the deterministic $N \to \infty$ limit is desirable.

a. Equivalent deterministic solutions

To validate this claim, we first recall that the limiting, deterministic, distribution, at finite m, is $p^{\mathrm{det}}(\mathbf{n}) = \lim_{N\to\infty} p(\mathbf{n}) = \delta(\mathbf{n}' - \mathbf{n})$, i.e., with all probability concentrated on $\mathbf{n}'$, with $\mathbf{n}'$ the normalized eigenvector associated with the largest eigenvalue of $W = [W_{ij}]$, $W_{ij} = \alpha_j \mu_{ji} - \omega_j \delta_{ij}$.

Taking a binary alphabet, $|\mathcal{A}| = 2$, we demonstrate the approximate ε independence first, theoretically, in the $L \to \infty$ limit, and then empirically. The former is achieved by adapting the result found in Eqs. (20) and (21) in Ref. [11], applicable to our quasispecies example with a single wild-type variant with elevated fitness, $\alpha_i = \alpha + \delta_{iz}\Delta\alpha$, $\omega_i = \omega\ \forall i$ (note, where ω is variant independent, as here, the deterministic solution is independent of ω, so we may set $\omega = 0$, making our case applicable to Eqs. (20) and (21) in Ref. [11]). This yields

$$\varrho_i = \begin{cases} \frac{(1-\varepsilon)^L(\alpha+\Delta\alpha)-\alpha}{\Delta\alpha}, & i = 0, \\ \frac{\alpha}{\Delta\alpha}\sum_{j=0}^{i-1}\frac{i!\varrho_j\varepsilon^{i-j}}{(i-j)!j!}, & i \geqslant 1, \end{cases} \tag{H6}$$

where ϱ_i is a weight associated with all of the variants at Hamming distance i from the wild type. This then associates the population fraction, $\hat{n}^{(i)}$, with those individual variants at Hamming distance i from the wild type, as

$$\hat{n}^{(i)} = \frac{\varrho_i}{\sum_{j=0}^{L+1}\binom{L}{j}\varrho_j}, \tag{H7}$$

where $\binom{L}{j} = L!/(j!(L-j)!)$ is the binomial coefficient, accounting for the number of distinct variants at each Hamming distance, since $|\mathcal{A}| = 2$. This formula is valid for $L \to \infty$, assuming $\varepsilon \sim 1/L$, such that $\hat{\mu} = (1-\varepsilon)^L \to e^{-\varepsilon L} = \mathcal{O}(1)$. Indeed, in that limit, where the error catastrophe manifests as a sharp phase transition, this formula fails precisely at the transition where $\rho_1 = 0$, such that $\hat{\mu}_{\text{crit}}^{\text{error}} = (1-\varepsilon)^L = \alpha/(\alpha+\Delta\alpha)$, in agreement with Eq. (F24), derived under the same circumstances, i.e., $\lim_{L\to\infty} p^{\text{det}} = \lim_{m\to\infty}\lim_{N\to\infty} p$. Despite its asymptotic validity, we may still use Eq. (H7) at finite L as an approximation, provided $\varrho_0 > 0$. In our case, for $L = 4$ as utilized in Fig. 3, this yields the population fractions

$$\begin{aligned}
\hat{n}^{(0)} &= \frac{\Delta\alpha^4}{\Sigma}, \\
\hat{n}^{(1)} &= \frac{\alpha\Delta\alpha^3\varepsilon}{\Sigma}, \\
\hat{n}^{(2)} &= \frac{\alpha\Delta\alpha^2\varepsilon^2(2\alpha+\Delta\alpha)}{\Sigma}, \\
\hat{n}^{(3)} &= \frac{\alpha\Delta\alpha\varepsilon^3(6\alpha^2+6\alpha\Delta\alpha+\Delta\alpha^2)}{\Sigma}, \\
\hat{n}^{(4)} &= \frac{\alpha\varepsilon^4(2\alpha+\Delta\alpha)(12\alpha^2+12\alpha\Delta\alpha+\Delta\alpha^2)}{\Sigma}, \\
\Sigma &= 24\alpha^4\varepsilon^4 + 12\alpha^3\Delta\alpha\varepsilon^3(3\varepsilon+2) \\
&\quad + 2\alpha^2\Delta\alpha^2\varepsilon^2(\varepsilon(7\varepsilon+12)+6) \\
&\quad + \alpha\Delta\alpha^3\varepsilon(\varepsilon+2)(\varepsilon(\varepsilon+2)+2) + \Delta\alpha^4.
\end{aligned} \tag{H8}$$

Fixing $x := \Delta\alpha/(\alpha\varepsilon) = \text{const}$ allows us to expand around $\varepsilon = 0$ yielding

$$\begin{aligned}
\hat{n}^{(0)} &= \frac{x^4}{x^4+4x^3+12x^2+24x+24} + \mathcal{O}(\varepsilon), \\
\hat{n}^{(1)} &= \frac{x^3}{x^4+4x^3+12x^2+24x+24} + \mathcal{O}(\varepsilon), \\
\hat{n}^{(2)} &= \frac{2x^2}{x^4+4x^3+12x^2+24x+24} + \mathcal{O}(\varepsilon), \\
\hat{n}^{(3)} &= \frac{6x}{x^4+4x^3+12x^2+24x+24} + \mathcal{O}(\varepsilon), \\
\hat{n}^{(4)} &= \frac{24}{x^4+4x^3+12x^2+24x+24} + \mathcal{O}(\varepsilon),
\end{aligned} \tag{H9}$$

independent of ε, as claimed, with $\lim_{\varepsilon\to 0}\sum_{i=0}^{L}\binom{L}{i}\hat{n}^{(i)} = 1$. Note that this expansion becomes exact in the classical Eigen scaling limit $\varepsilon \sim 1/L$, with $L \to \infty$. For our choices, $L = 4$, $|\mathcal{A}| = 2$, and $x = 10$, the limiting $\varepsilon \to 0$ approximation for the deterministic quasispecies population state, denoted by $\mathbf{n}^*$, comprises

$$\mathbf{n}^* = \{n_1^*, \ldots, n_{16}^*\}, \tag{H10}$$

$$n_i^* = \hat{n}^{(H_{iz})}, \tag{H11}$$

$$\begin{aligned}
&\{\hat{n}^{(0)}, \hat{n}^{(1)}, \hat{n}^{(2)}, \hat{n}^{(3)}, \hat{n}^{(4)}\} \\
&\quad = \left\{\tfrac{1250}{1933}, \tfrac{125}{1933}, \tfrac{25}{1933}, \tfrac{15}{3866}, \tfrac{3}{1933}\right\} \\
&\quad \simeq \{0.647, 0.0647, 0.0129, 0.003\,88, 0.001\,55\},
\end{aligned} \tag{H12}$$

where $H_{iz} \in \{0, 1, 2, 3, 4\}$ is the Hamming distance of variant i from the wild type, z. Denoting the normalized dominant eigenvector of W for $\varepsilon = \varepsilon_x$ (and with $L = 4$, $\alpha = 1$, and $x = 10$), $\mathbf{n}^{\varepsilon_x} = \{n_1^{\varepsilon_x}, \ldots, n_m^{\varepsilon_x}\}$, $\mathbf{n}^*$ can be favorably compared with, for example, $\mathbf{n}^{\varepsilon_x} = \{0.641, \ldots, 0.0654, \ldots, 0.0133, \ldots, 0.00400, \ldots, 0.00154\}$ for $\varepsilon_x = 10^{-4}$.

More broadly, we can quantify the accuracy of the approximation with the total, L^1, distance between $\mathbf{n}^*$ and $\mathbf{n}^{\varepsilon_x}$, defined as

$$D(\mathbf{n}^*, \mathbf{n}^{\varepsilon_x}) = \sum_{i=1}^{m}\left|n_i^* - n_i^{\varepsilon_x}\right|, \tag{H13}$$

for the range $\varepsilon_x \in \boldsymbol{\varepsilon} = \{10^{-4}, 10^{-3.5}, 10^{-3}, 10^{-2.5}, 10^{-2}\}$ utilized in Fig. 3. For matching $\alpha = 1$ and $x = \Delta\alpha/(\alpha\epsilon) = 10$, the mean $\mathbb{E}_{\varepsilon_x\in\boldsymbol{\varepsilon}}[D(\mathbf{n}^*, \mathbf{n}^{\varepsilon_x})]$ was found to be $\sim$0.0161.

Of more importance, for our purposes, is that the exact limiting population, using our given parameters, is approximately independent of ε across our simulations in Fig. 3. This is confirmed empirically in Fig. 14(e), which shows the limiting, deterministic, population states, $\mathbf{n}^{\varepsilon_x}$ [such that $p^{\text{det}}(\mathbf{n}) = \delta(\mathbf{n}^{\varepsilon_x} - \mathbf{n})$], for the parameters corresponding to Fig. 3, across representative values of ε_x. We can then quantify their approximate equivalence in terms of the maximum L^1 variance across the range of ε, which was found to be $\max_{\varepsilon_x,\varepsilon_y\in\boldsymbol{\varepsilon}} D(\mathbf{n}^{\varepsilon_x}, \mathbf{n}^{\varepsilon_y}) \sim 0.0035$.

b. Motivation

The advantages conferred by our prescription that $\lim_{N\to\infty} p(\mathbf{n})$ is independent of ε (and thus $\hat{\mu}$) are threefold.

First, it illustrates that the deterministic, $\lim_{N\to\infty}$, quasispecies solution is well described by a single reduced parameter $\Delta\alpha/(\alpha\varepsilon)$, such that the original deterministic EM is effectively overdetermined, with changes in (inverse) error rate and fitness advantage being indistinguishable. However, this property is broken when realistic, finite-size effects are introduced, such that two identical populations, under the

classical EM, can behave very differently at finite N. This can be seen plainly in Fig. 3, where the threshold to the fidelity catastrophe occurs at differing N despite equivalence of the deterministic, $N \to \infty$, solutions.

Second, by constraining $\lim_{N\to\infty} p(\mathbf{n})$ to be ε ($\hat{\mu}$) independent, we ensure that differences in the variation of the distribution across ε ($\hat{\mu}$), due to changes in N, manifest merely as translations across our phase space. In broad terms, this ensures that behavior observed in the multimodal, fidelity catastrophe regime is approximately independent of $\hat{\mu}$, as is the behavior in the unimodal regime, with only the *onset* of the transition between them occurring at different values of N for varying $\hat{\mu}$. This is illustrated in panels (a)–(d) of Fig. 14 where we present descriptive statistics across vertical [varying ε (and thus $\hat{\mu}$ and $\Delta\alpha$), fixed N] and horizontal (varying N, fixed ε) lines of our phase space that appears in Fig. 3. Panels (a) and (b) of Fig. 14 plot the bimodality coefficient

$$\beta = (\text{skew}[\rho(n_z)]^2 + 1)/\text{kurt}[\rho(n_z)], \tag{H14}$$

associated with the empirical distribution for the fractional abundance of the wild type, $\rho(n_z') = T^{-1}\int_0^T \delta(n_z' - n_z(t))\mathrm{d}t$, resulting from simulation, for fixed ε and $N(0)$, respectively. Similarly, panel (c) plots the empirical time fraction for which the most abundant variant in $\mathbf{n}$ is the wild type, z (i.e., $\tau_z = T^{-1}\int_0^T \delta_{z,\text{argmax}_i n_i(t)}\mathrm{d}t$), while panel (d) plots the variance of the empirical distribution for the fractional abundance of the wild type, $\text{var}(\rho(n_z))$, as a function of $N(0)$ at distinct ε. Across all panels, we observe uniformly shifted responses, with the same sigmoidal variation of β trending from high to low as N is increased, the same growth in τ_z as the system exits the fidelity catastrophe region with increasing N as the system stabilises on the classic quasispecies solution, and with the variance falling with increasing N as the distribution tends from being multimodal with weight at the edges of the simplex, to being unimodal, concentrated toward the center of the simplex. In practice, this common variation allows us to more reasonably associate a line of fixed β (taken as $\beta = 5/9$, the value for the uniform distribution) across distinct ε and $N(0)$ with nominally equivalent behavior, thus allowing for a more meaningful comparison with the predicted variation associated with Eqs. (15) and (16) of the main text (black line, Fig. 3), which shares the same slope across ε and $N(0)$, lending confidence to our characterization.

Finally, the independence of $\lim_{N\to\infty} p(\mathbf{n}')$ across ε ensures that it is feasible to numerically sample the distributions across our whole phase space, such that the measured, empirical, distribution, $\rho(\mathbf{n}') = T^{-1}\int_0^T \delta(\mathbf{n}' - \mathbf{n}(t))\mathrm{d}t$ reasonably approximates $p(\mathbf{n}') = \lim_{T\to\infty} T^{-1}\int_0^T \delta(\mathbf{n}' - \mathbf{n}(t))\mathrm{d}t$, i.e., $\rho(\mathbf{n}) \simeq p(\mathbf{n})$ over reasonable computation timescales, $T < \infty$. More specifically, by ensuring that the distributions are comparable across the phase space, we observe similar residence times in the temporary clonal states that characterize the fidelity catastrophe. Thus, across values of ε ($\hat{\mu}$), a similar amount of computational time is required to observe a sufficient number of transitions between such clonal states to achieve good sampling. In contrast, if we, for instance, did not decrease $\Delta\alpha$ with reduced ε, the deterministic behavior would become increasingly singular [i.e., with $\mathbf{n}'$ in $p^{\text{det}}(\mathbf{n}) = \delta(\mathbf{n}' - \mathbf{n})$ tending to $\mathbf{n}' = \{1, 0, \ldots, 0\}$ as $\varepsilon \to 0$]. When such a configuration is in the fidelity catastrophe regime, the distribution becomes very narrowly distributed on the distinct clonal configurations, thus requiring very large (and thus rare) fluctuations to transition between them, leading to ever increasing required computational time.

[1] B. Alberts, R. Heald, A. Johnson, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 7th ed. (W. W. Norton, 2022).

[2] B. Tippin, P. Pham, and M. F. Goodman, Error-prone replication for better or worse, Trends Microbiol. **12**, 288 (2004).

[3] J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow, Rates of spontaneous mutation, Genetics **148**, 1667 (1998).

[4] P. D. Sniegowski, P. J. Gerrish, T. Johnson, and A. Shaver, The evolution of mutation rates: Separating causes from consequences, Bioessays **22**, 1057 (2000).

[5] S. Duffy, L. A. Shackelton, and E. C. Holmes, Rates of evolutionary change in viruses: Patterns and determinants, Nat. Rev. Genet. **9**, 267 (2008).

[6] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing* (Springer, Berlin, 2015).

[7] F. Silva, M. Duarte, L. Correia, S. M. Oliveira, and A. L. Christensen, Open issues in evolutionary robotics, Evol. Comput. **24**, 205 (2016).

[8] P. Adamski, M. Eleveld, A. Sood, Á. Kun, A. Szilágyi, T. Czárán, E. Szathmáry, and S. Otto, From self-replication to replicator systems en route to de novo life, Nat. Rev. Chem. **4**, 386 (2020).

[9] M. Eigen, Selforganization of matter and the evolution of biological macromolecules, Naturwissenschaften **58**, 465 (1971).

[10] M. Eigen and P. Schuster, A principle of natural self-organization: Part A: Emergence of the hypercycle, Naturwissenschaften **64**, 541 (1977).

[11] D. B. Saakian and C.-K. Hu, Exact solution of the Eigen model with general fitness functions and degradation rates, Proc. Natl. Acad. Sci. USA **103**, 4935 (2006).

[12] M. Eigen, Viral quasispecies, Sci. Am. **269**, 42 (1993).

[13] E. Domingo, J. Sheldon, and C. Perales, Viral quasispecies evolution, Microbiol. Mol. Biol. Rev. **76**, 159 (2012).

[14] E. Domingo, C. García-Crespo, and C. Perales, Historical perspective on the discovery of the quasispecies concept, Ann. Rev. Virol. **8**, 51 (2021).

[15] D. Alves and J. F. Fontanari, Error threshold in finite populations, Phys. Rev. E **57**, 7008 (1998).

[16] Y. Bakhtin, M. I. Katsnelson, Y. I. Wolf, and E. V. Koonin, Evolution in the weak-mutation limit: Stasis periods punctuated by fast transitions between saddle points on the fitness landscape, Proc. Natl. Acad. Sci. USA **118**, e2015665118 (2021).

[17] D. Bonnaz and A. J. Koch, Stochastic model of evolving populations, J. Phys. A: Math. Gen. **31**, 417 (1998).

[18] P. R. A. Campos and J. F. Fontanari, Finite-size scaling of the error threshold transition in finite populations, J. Phys. A: Math. Gen. **32**, L1 (1999).

[19] L. Demetrius, P. Schuster, and K. Sigmund, Polynucleotide evolution and branching processes, Bull. Math. Biol. **47**, 239 (1985).
[20] B. L. Jones and H. K. Leung, Stochastic analysis of a nonlinear model for selection of biological macromolecules, Bull. Math. Biol. **43**, 665 (1981).
[21] J. S. McCaskill, A stochastic theory of macromolecular evolution, Biol. Cybern. **50**, 63 (1984).
[22] M. Nowak and P. Schuster, Error thresholds of replication in finite populations mutation frequencies and the onset of Muller's ratchet, J. Theor. Biol. **137**, 375 (1989).
[23] Y.-C. Zhang, Quasispecies evolution of finite populations, Phys. Rev. E **55**, R3817 (1997).
[24] F. Musso, A stochastic version of the Eigen model, Bull. Math. Biol. **73**, 151 (2011).
[25] W. Horsthemke and R. Lefever, *Noise-Induced Transitions*, *Springer Series in Synergetics* (Springer, Berlin. 2006), Vol. 15.
[26] T. Biancalani, T. Rogers, and A. J. McKane, Noise-induced metastability in biochemical networks, Phys. Rev. E **86**, 010106(R) (2012).
[27] R. G. Morris and T. Rogers, Growth-induced breaking and unbreaking of ergodicity in fully-connected spin systems, J. Phys. A: Math. Theor. **47**, 342003 (2014).
[28] F. Jafarpour, T. Biancalani, and N. Goldenfeld, Noise-induced symmetry breaking far from equilibrium and the emergence of biological homochirality, Phys. Rev. E **95**, 032407 (2017).
[29] T. Biancalani, L. Dyson, and A. J. McKane, Noise-induced bistable states and their mean switching time in foraging colonies, Phys. Rev. Lett. **112**, 038101 (2014).
[30] J. Jhawar, R. G. Morris, U. R. Amith-Kumar, M. D. Raj, T. Rogers, H. Rajendran, and V. Guttal, Noise-induced schooling of fish, Nat. Phys. **16**, 488 (2020).
[31] A. J. Bladon, T. Galla, and A. J. McKane, Evolutionary dynamics, intrinsic noise, and cycles of cooperation, Phys. Rev. E **81**, 066122 (2010).
[32] G. W. A. Constable, T. Rogers, A. J. McKane, and C. E. Tarnita, Demographic noise can reverse the direction of deterministic selection, Proc. Natl. Acad. Sci. USA **113**, E4745 (2016).
[33] P. Ashcroft, C. E. R. Smith, M. Garrod, and T. Galla, Effects of population growth on the success of invading mutants, J. Theor. Biol. **420**, 232 (2017).
[34] E. Crosato, J. N. Philippson, S. Thutupalli, and R. G. Morris, Dynamical demographic phases explain how population growth and mutation control the evolutionary impact of bottlenecks, Phys. Rev. Res. **5**, 013093 (2023).
[35] F. Musso, On the relation between the Eigen model and the asexual Wright–Fisher model, Bull. Math. Biol. **74**, 103 (2012).
[36] M. Kimura and J. F. Crow, *An Introduction to Population Genetics Theory* (Scientific Publishers, 2017).
[37] E. Baake and W. Gabriel, Biological evolution through mutation, selection, and drift: An introductory review, Annu. Rev. Comput. Phys. **7**, 203 (2000).
[38] R. A. Blythe and A. J. McKane, Stochastic models of evolution in genetics, ecology and linguistics, J. Stat. Mech. (2007) P07018.
[39] N. Godfried and V. Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, North-Holland, 1992), Vol. 1.
[40] C. Gardiner, *Stochastic Methods* (Springer, Berlin, 2009), Vol. 4.
[41] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. **22**, 403 (1976).
[42] P. G. Hufton, Y. T. Lin, T. Galla, and A. J. McKane, Intrinsic noise in systems with switching environments, Phys. Rev. E **93**, 052119 (2016).
[43] P. G. Hufton, Y. T. Lin, and T. Galla, Model reduction methods for population dynamics with fast-switching environments: Reduced master equations, stochastic differential equations, and applications, Phys. Rev. E **99**, 032122 (2019).
[44] J. R. Worsfold and R. G. Morris, Stochastically bistable growth and decay in the Togashi-Kaneko model, Phys. Rev. E **112**, 014131 (2025).
[45] H. J. Muller, The relation of recombination to mutational advance, Mutat. Res. **1**, 2 (1964).
[46] J. Felsenstein, The evolutionary advantage of recombination, Genetics **78**, 737 (1974).
[47] J. Haigh, The accumulation of deleterious genes in a population—Muller's ratchet, Theor. Popul Biol. **14**, 251 (1978).
[48] J. J. Metzger and S. Eule, Distribution of the fittest individuals and the rate of Muller's ratchet in a model with overlapping generations, PLoS Comput. Biol. **9**, e1003303 (2013).
[49] E. Crosato, R. E. Spinney, and R. G. Morris, Fidelity Catastrophe - Simulation Code and Dataset, Version 1, Zenodo, 2025, https://doi.org/10.5281/zenodo.16910145.