

Noise can speed convergence in Markov chains

Brandon Franzke and Bart Kosko*

*Center for Quantum Information Science and Technology, Signal and Image Processing Institute, Department of Electrical Engineering,
University of Southern California, Los Angeles, California 90089, USA*

(Received 26 April 2011; revised manuscript received 29 June 2011; published 10 October 2011)

A new theorem shows that noise can speed convergence to equilibrium in discrete finite-state Markov chains. The noise applies to the state density and helps the Markov chain explore improbable regions of the state space. The theorem ensures that a stochastic-resonance noise benefit exists for states that obey a vector-norm inequality. Such noise leads to faster convergence because the noise reduces the norm components. A corollary shows that a noise benefit still occurs if the system states obey an alternate norm inequality. This leads to a noise-benefit algorithm that requires knowledge of the steady state. An alternative blind algorithm uses only past state information to achieve a weaker noise benefit. Simulations illustrate the predicted noise benefits in three well-known Markov models. The first model is a two-parameter Ehrenfest diffusion model that shows how noise benefits can occur in the class of birth-death processes. The second model is a Wright-Fisher model of genotype drift in population genetics. The third model is a chemical reaction network of zeolite crystallization. A fourth simulation shows a convergence rate increase of 64% for states that satisfy the theorem and an increase of 53% for states that satisfy the corollary. A final simulation shows that even suboptimal noise can speed convergence if the noise applies over successive time cycles. Noise benefits tend to be sharpest in Markov models that do not converge quickly and that do not have strong absorbing states.

DOI: [10.1103/PhysRevE.84.041112](https://doi.org/10.1103/PhysRevE.84.041112)

PACS number(s): 05.40.Ca, 05.10.Ln, 05.10.Gg

I. NOISE BENEFITS IN MARKOV CHAIN DENSITY ESTIMATION

We show that noise can speed convergence to equilibrium in discrete finite-state Markov chains. Judiciously adding noise directly to the state density speeds up the convergence time for the Markov chain simulation depending on the direction of an inequality. The noise appears to give the Markov chain system access to a statistically richer set of otherwise improbable states. Neural network researchers have observed related smoothing effects that arise when adding noise to training data [1,2]. Figures 1, 2, and 3 show a Markov chain noise benefit when a simulation adds noise to the current state density. Table II summarizes the 53%–64% performance improvement after adding noise in a simulation. Figure 10 shows that this noise benefit holds over successive cycles.

Many nonlinear signal systems benefit from adding small amounts of noise [3–16]. Too little noise produces little or no benefit while too much noise can swamp the system’s performance. This so-called “stochastic resonance” effect can take the form of an increased signal-to-noise ratio [17–19], entropy-based bit count [20–22], input-output correlation [23], or probability of detection [24,25]. The noise benefit for a simulated Markov chain is a shorter time to converge to the equilibrium probability density in the sense that the noise reduces the vector norm of the error.

Markov chains form a basis for powerful Markov chain Monte Carlo (MCMC) statistical simulations [26]. MCMC methods generate samples from a given posterior probability density function by constructing a Markov chain whose stationary density equals the posterior of interest [27,28]. The Metropolis-Hastings algorithm [29,30] and Gibbs samplers [31,32] are special and powerful MCMC frameworks that

compute Bayesian statistics. But MCMC methods suffer from problem-specific parameters that govern sample acceptance and convergence assessments [33,34]. A strong dependence on the initial conditions also biases the MCMC sampling unless the simulation allows a lengthy period of “burn-in” to allow the driving Markov chain to mix adequately [26,35].

The Markov chain noise benefit theorem in the next section shows how to construct a normalized state density at each time cycle for a finite time-homogeneous Markov chain with an irreducible and aperiodic state transition matrix. The theorem and corollary guarantee the existence of a component-wise noise benefit that decreases the time to convergence. They show that noise can perturb the current state of a Markov chain to explore novel regions in the state space and speed convergence to the steady-state distribution. The form of the noise depends on the direction of a state-related inequality. The theorem may ensure only minimal benefits for systems that exhibit fast convergence or that possess strong absorbing states.

Section III presents two algorithms that use the Markov chain noise benefit theorem to obtain a noise benefit. The first algorithm shows how the simulation can obtain an *optimal* noise benefit. The second algorithm describes how to obtain a noise benefit that uses only the current and past state of the Markov chain. A key limitation in applying this result to MCMC is that the system does not usually have direct access to the current state vector during the MCMC simulation. Table III shows that systems can still benefit from noise even without direct access to the state vector. Suitable guesses for the sign of the inequality should help further overcome this limitation in practice.

II. MARKOV CHAIN NOISE BENEFIT THEOREM

The Markov chain noise benefit theorem below shows that Markov chain simulations can benefit from noise through

*kosko@usc.edu

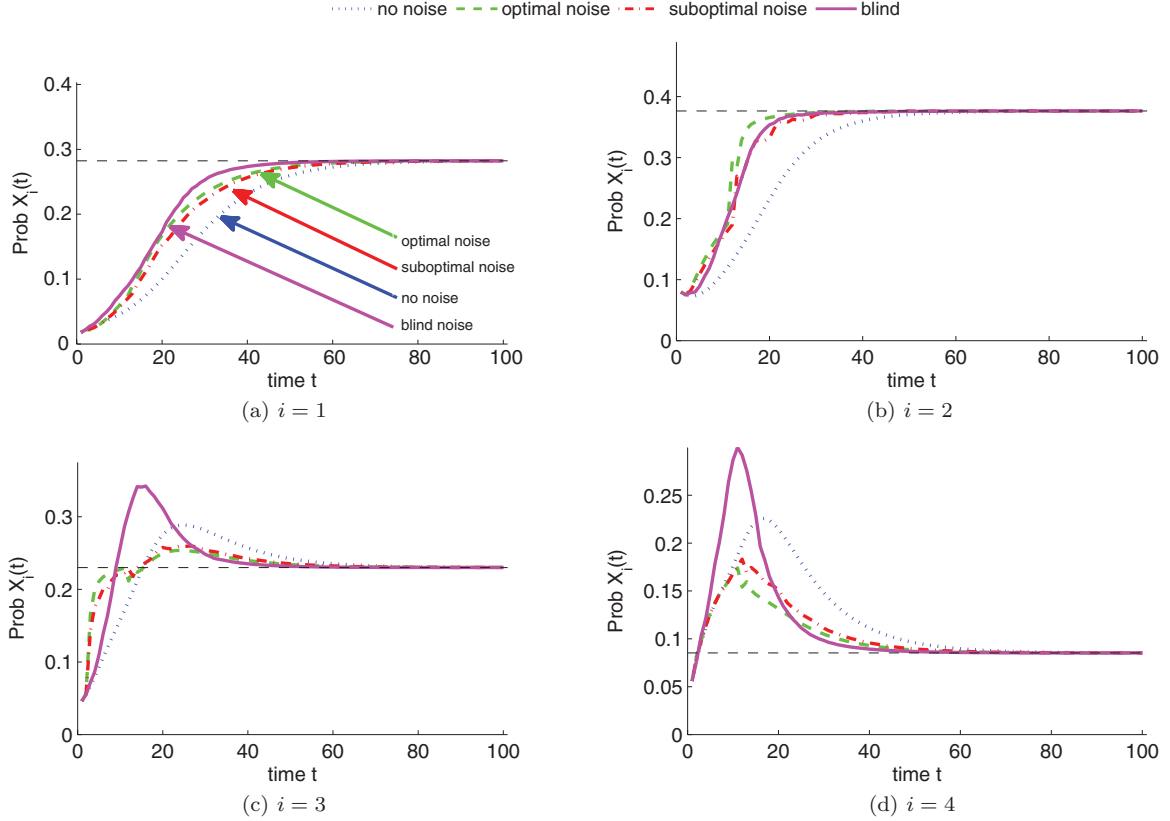


FIG. 1. (Color online) Noise benefits in the two-parameter (Krafft-Schaefer) Ehrenfest diffusion model. Noise enhances the two-parameter Ehrenfest diffusion model by decreasing the time to convergence. The figures summarize the results from a 12-molecule simulation with $s = 0.10$ and $t = 0.90$. These figures show the time evolution of the first four components of the 13-component state vector $X(t)$ (corresponding to $[X_\infty]_i > 0.002$). Each component of the state vector gives the probability of a particular distribution of the 12 molecules between compartments A and B . Case $i = 1$ corresponds to all 12 molecules in box A and $i = 2$ corresponds to 11 molecules in compartment A and 1 molecule in compartment B . The blue (dotted) curve plots the standard (no noise) Ehrenfest diffusion model. The green (dashed), red (dash-dot), and pink (solid) curves show noisy versions of the model. The noise benefit appears in the distinct shift to the left of the noise-enhanced simulations over the standard model. This shows that the simulations reach steady state sooner. The green (dashed) curve shows a simulation using the optimal noise N_{opt} according to the Markov chain noise benefit theorem and the red (dash-dot) curve shows the result by choosing suboptimal noise uniformly in $[0, N_{\text{opt}}]$. The pink (solid) curve shows the results of Algorithm 2. Algorithm 2 does not require prior knowledge or an estimate of the steady-state distribution. The figures show that this system nears steady state within 60 time steps.

faster convergence. Markov chain simulations employ a stochastic discrete time model to estimate the probability density over a system's state-space.

Suppose M is a time-homogeneous Markov chain over a finite state-space with $N < \infty$ states [26,36,37]. Let the $N \times 1$ column-vector $x(t)$ represent the state of the Markov chain at time t . Each component $x_i(t)$ represents the probability that the chain is in the corresponding state i at time t . Then

$$\sum_{i=1}^N x_i(t) = 1 \quad (1)$$

for all t because $x(t)$ is a probability density over the N states.

Let P represent the single-step state transition probability matrix where

$$P_{i,j} = P[x_{t+1} = j | x_t = i] \quad (2)$$

is the probability of the chain in state i at time t moving to state j at time $t + 1$. Then there exists a stationary vector x^∞

such that [37]

$$x^\infty = x^\infty P. \quad (3)$$

So x^∞ is always a left eigenvector of the transition probability matrix P that corresponds to the eigenvalue $\lambda = 1$.

The n -step transition probability matrix $P^{(n)}$ has entries

$$P_{i,j}^{(n)} = P[X_{t+n} = j | X_t = i] \quad (4)$$

$$= \sum_{k=0}^N P[X_{t+n} = j | X_t = i, X_{t+1} = k] \cdot P[X_{t+1} = k | X_t = i] \quad (5)$$

$$= \sum_{k=0}^N P[X_{t+n} = j | X_{t+1} = k] P[X_{t+1} = k | X_t = i] \quad (6)$$

$$= \sum_{k=0}^N P_{k,j}^{(n-1)} P_{i,k} \quad (7)$$

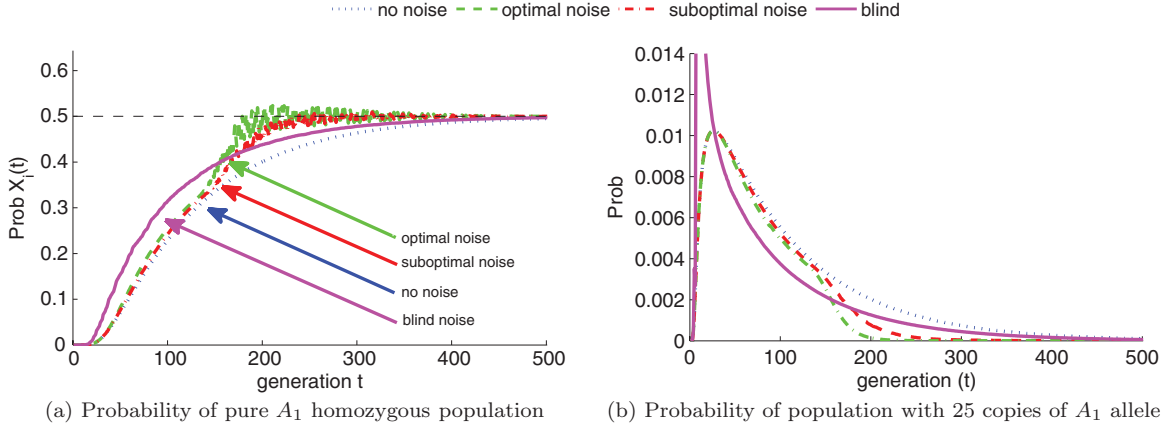


FIG. 2. (Color online) Noise benefits in Wright-Fisher population genetics. The figures show that the Wright-Fisher model benefits from noise in two ways: (1) noise moves the simulation toward the steady-state value faster than simulations without noise and (2) noise eliminates the asymptotic crawl toward the steady-state value. The figures represent allele distributions over 500 reproductive generations. The simulation models the distribution of a diallelic gene (A_1 and A_2) in a population with $N = 50$ diploid individuals ($2N = 100$ gene copies). The simulation initialized with $X_t \sim N(50, 1)$ is a normal distribution centered at a population with equal numbers of A_1 and A_2 . The symmetry of X_0 implies that the steady-state population will move toward either homozygous coalescent state ($A_1 A_1$ or $A_2 A_2$) with equal probability. The blue (dotted) curve plots the standard (no noise) Wright-Fisher model. The green (dashed), red (dash-dot), and pink (solid) curves show noisy versions of the model. The green (dashed) curve shows a simulation using the optimal noise N_{opt} prescribed by the noise-benefit theorem and the red (dash-dot) curve shows the result by choosing suboptimal noise uniformly in $[0, N_{opt}]$. The pink (solid) curve shows that noisy algorithm can benefit a Markov chain even if it cannot determine the steady-state values that the theorem assumes. (a) The probability for a homozygous A_1 ($A_1 A_1$) population. The initial distribution implies a near zero probability for a pure $A_1 A_1$ population but rapidly increases to its steady-state value of $1/2$. The noise-enhanced simulations (red and green) approach the steady-state value faster than the standard model (blue) and also reach the asymptotic value before settling. (b) The probability for a population with 25 copies of allele A_1 and 75 copies of allele A_2 . The noise-enhanced simulations (red and green) approach the steady-state value faster than the standard model (blue) and also reach the asymptotic value before settling.

where $P_{i,j}^{(n)}$ is the probability that the chain transitions from state i to state j in exactly n time steps. State j is *accessible* from state i if there is some nonzero probability of transitioning from state i to state j in any number of steps:

$$P_{i,j}^{(n)} > 0 \tag{8}$$

for some $n > 0$.

A Markov chain is *irreducible* if every state is accessible from every other state [26,37]. Irreducibility implies that for all states i and j there exists $m > 0$ such that

$$P[X_{n+m} = j | X_n = i] = P_{i,j}^{(n)} > 0. \tag{9}$$

This is equivalent to P is a *regular stochastic matrix* if M is a finite Markov chain.

The period d_i of state i is

$$d_i = \text{gcd}\{n \geq 1 : P_{i,i}^{(n)} > 0\} \tag{10}$$

or $d_i = \infty$ if $P_{i,i}^{(n)} = 0$ for all $n \geq 1$ where gcd denotes the greatest common divisor. State i is *aperiodic* if $d_i = 1$. A Markov chain with transition matrix P is *aperiodic* if $d_i = 1$ for all states i .

Suppose a Markov chain M is irreducible and aperiodic. Then the fixed point x^∞ is unique and

$$\lim_{k \rightarrow \infty} P^{(k)} = \mathbf{1} \otimes x^\infty \tag{11}$$

where $\mathbf{1}$ is the column vector with all entries equal to 1 [36,38]. The outer product generates a rank-one $N \times N$ matrix with each column equal to the stationary state density.

The Markov chain noise benefit theorem below shows that there is a component-wise noise benefit for any component that has not yet converged to its stationary value. The theorem assumes that the sign of a state-related inequality is in one of two directions. The corollary assumes it is in the other direction. The proof of the theorem is lengthy and appears in the Appendix.

Markov Chain Noise Benefit Theorem. Suppose M is a finite time-homogeneous Markov chain with N states and transition matrix P . Suppose further that M is irreducible and aperiodic. Then for all nonstationary state density vectors x there exists a noise benefit in the sense that there exists some $A > 0$ so that for all $a \in (0, A)$

$$|[\tilde{x}P - x^\infty]_i| < |[xP - x^\infty]_i| \tag{12}$$

for all states i with

$$\Delta_i = (x - x^\infty)P_i > 0 \tag{13}$$

where

$$\tilde{x} = \frac{1}{1+a}(x + n) \tag{14}$$

is the normalized state vector after adding a noise vector n with only one nonzero component

$$n_j = \begin{cases} a & j = k \\ 0 & j \neq k \end{cases} \quad (15)$$

for any k that satisfies

$$\Delta_k = (x - x^\infty)P_k > 0. \quad (16)$$

The following corollary provides a complementary result when the converse of inequality (13) holds ($\Delta_i < 0$) in the Markov chain noise benefit theorem.

Corollary. Suppose the hypotheses of the Markov chain noise benefit theorem hold. Then there exists a noise benefit for each nonstationary state density vector x in the sense that there exists some $A > 0$ so that for all $a \in (0, A)$

$$|[\tilde{x}P - x^\infty]_i| < |[xP - x^\infty]_i| \quad (17)$$

for all states i with

$$\Delta_i = (x - x^\infty)P_i < 0. \quad (18)$$

Proof. The Δ_i sign change does not affect the expansion in the proof of the theorem. So

$$|[\tilde{x}P - x^\infty]_i| = \left| \Delta_i - \frac{a_i}{1 + a_i}(xP_i - P_{k,i}) \right| \quad (19)$$

holds.

Now $\Delta_i < 0$ by hypothesis. Thus

$$\left| \Delta_i - \frac{a_i}{1 + a_i}(xP_i - P_{k,i}) \right| < |\Delta_i| \quad (20)$$

if and only if

$$\frac{a_i}{1 + a_i}(xP_i - P_{k,i}) > 2\Delta_i \quad (21)$$

$$\text{and } \frac{a_i}{1 + a_i}(xP_i - P_{k,i}) < 0 \quad (22)$$

since $|\Delta_i| > |\Delta_i - b|$ if and only if $2\Delta_i < b < 0$. The negativity constraint (22) holds if and only if $xP_i < P_{k,i}$. The lower bound (21) holds if and only if

$$a_i(xP_i - P_{k,i}) > 2\Delta_i(1 + a_i). \quad (23)$$

Therefore Eq. (21) holds if and only if

$$a_i(xP_i - P_{k,i} - 2\Delta_i) > 2\Delta_i. \quad (24)$$

If $2\Delta_i < xP_i - P_{k,i}$ then

$$a_i > \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} \quad (25)$$

and if $2\Delta_i > xP_i - P_{k,i}$ then

$$a_i < \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i}. \quad (26)$$

But if $2\Delta_i < xP_i - P_{k,i}$ then $\frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} < 0$. So any $a_i < 0$ suffices. Thus either

$$a_i < 0 \quad \text{if } 2\Delta_i > xP_i - P_{k,i} \quad (27)$$

or

$$a_i > \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} \quad \text{if } 2\Delta_i < xP_i - P_{k,i}. \quad (28)$$

Therefore if $a_i \in (-A_i, 0)$ with $A_i = -\frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} > 0$ then Eqs. (27) and (28) hold. So if $A = \min_i\{A_i\} > 0$ then the theorem holds for all states i that satisfy the inequality $\Delta_i = (x - x^\infty)P_i < 0$. ■

III. MARKOV CHAIN NOISE BENEFIT ALGORITHMS

This section presents two versions of the Markov chain noise benefit algorithm. The first algorithm shows how a Markov chain simulation can apply the Markov chain noise benefit theorem directly to realize an optimal noise benefit. The second algorithm shows a practical implementation that uses only the current and past states of the simulation.

Algorithm 1 shows a naive application of the Markov chain noise benefit theorem. The green lines on Figs. 1, 2, and 3 show simulation results from this algorithm. This algorithm has the practical limitation that it requires prior knowledge of the steady-state distribution. The algorithm finds the component with the smallest state error at each step. It then adds signed noise to compensate for the error.

Algorithm 1 The optimal Markov chain noise benefit algorithm

```

1: procedure MARKOVCHAIN( $x_0, P, x^\infty$ )
2:    $x_t \leftarrow x_0$ 
3:   repeat
4:      $x_t \leftarrow x_t P$ 
5:      $x_t \leftarrow \text{NOISYSTEP}(x_t, P, x^\infty)$ 
6:   until ISCONVERGED( $x_t$ )
7:   return  $x_t$ 

8: procedure NOISYSTEP( $x_t, P, x^\infty$ )
9:    $n_t \leftarrow \text{CALCNOISE}(x_t, P, x^\infty)$ 
10:   $\tilde{x}_t \leftarrow \frac{1}{1 + \sum n_t}(x_t + n_t)$ 
11:  return  $\tilde{x}_t$ 

12: procedure CALCNOISE( $x_t, P, x^\infty$ )
13:   $\Delta \leftarrow (x_t - x^\infty)P$ 
14:   $L \leftarrow \text{LENGTH}(\Delta)$ 
15:   $A \leftarrow \Delta[0]$ 
16:   $k \leftarrow 0$ 
17:  for  $j \leftarrow 1, L$  do
18:    if  $|\Delta[j]| < A$  then
19:       $A \leftarrow -\Delta[j]$ 
20:       $k \leftarrow j$ 
21:   $n \leftarrow \text{ZEROVECTOR}(L)$ 
22:   $n[k] \leftarrow A$ 
23:  return  $n$ 

```

Algorithm 2 overcomes the limitation of Algorithm 1 because it does not require knowledge of the steady-state values. It uses only the past state probabilities to determine the noise at each time step. Algorithm 2 picks the state that changes the most at each time step and then adds noise to drive that state further in its current direction. The pink lines on Figs. 1, 2, and 3 show that Algorithm 2 speeds convergence in the three Markov chain simulations on average.

Algorithm 2 The blind Markov chain noise benefit algorithm

```

1: procedure MARKOVCHAIN( $x_0, P$ )
2:    $x_t \leftarrow x_0$ 
3:   repeat
4:      $x_t \leftarrow x_t P$ 
5:      $x_t \leftarrow \text{NOISYSTEP}(x_t, P, x_{t-1})$ 
6:   until ISCONVERGED( $x_t$ )
7:   return  $x_t$ 

8: procedure NOISYSTEP( $x_t, P, x_{t-1}$ )
9:    $n_t \leftarrow \text{CALCNOISE}(x_t, n_{t-1}, x_{t-1})$ 
10:   $\tilde{x}_t \leftarrow \frac{1}{1 + \sum n_t} (x_t + n_t)$ 
11:  return  $\tilde{x}_t$ 

12: procedure CALCNOISE( $x_t, n_{t-1}, x_{t-1}$ )
13:   $\Delta \leftarrow x_t - n_{t-1} - x_{t-1}$ 
14:   $L \leftarrow \text{LENGTH}(\Delta)$ 
15:   $A \leftarrow \Delta[0]$ 
16:   $k \leftarrow 0$ 
17:  for  $j \leftarrow 1, L$  do
18:    if  $|\Delta[j]| > A$  then
19:       $A \leftarrow -\Delta[j]$ 
20:       $k \leftarrow j$ 
21:   $n \leftarrow \text{ZEROVECTOR}(L)$ 
22:   $n[k] \leftarrow \text{Uniform}(0, A)$ 
23:  return  $n$ 

```

IV. MARKOV CHAIN EXPERIMENTAL RESULTS

The simulations below show that the proposed noise benefit applies to a wide range of Markov chain models. The three simulations show the evolution of the state density by direct computation of $x_{t+1} = x_t P$. Figures 1, 2, and 3 show the probability of several states over time. The first simulation applies noise to the two-parameter Ehrenfest diffusion model. The simulation reaches a steady state about 24% faster than the simulation without noise and provides evidence that the

Markov chain noise benefit theorem can apply to birth-death processes. The second simulation demonstrates that the Wright-Fisher population genetics model benefits from noise by decreasing the time to convergence. The third simulation shows that noise can speed simulations of a proposed chemical reaction whose state transition matrix derives from empirical measurement data.

A. Noise benefits in the Ehrenfest diffusion model

The first simulation shows a noise benefit in the Ehrenfest diffusion model. Ehrenfest proposed a diffusion model in the early 1900s as a statistical interpretation of the second law of thermodynamics [39,40]. The model demonstrates the increase in entropy of a closed system over time [41].

The simulation shows that the noise benefit theorem applies to a class of Markov models called *birth-death* processes. A birth-death process has the constraint $P_{i,j} = 0$ if $|i - j| > 1$ [40,42–45]. The simulation also demonstrates a noise benefit in a model that converges only in distribution. Figure 1 illustrates the noise benefit in an $N = 12$ molecule Ehrenfest diffusion simulation. Table I shows how each state i corresponds to a distribution of 12 molecules divided between two compartments A and B .

The simulation employed a two-parameter generalized model with $s = 0.10$ and $t = 0.90$. The figure shows that the components approach their steady-state values 24.2% faster on average with added noise (error $< 0.5\%$ of steady state).

The simplest Ehrenfest diffusion model uses a rectangular container with a permeable membrane separating two equally sized compartments called compartment A and compartment B [46–48]. The container holds N gas molecules that the membrane allows to pass between compartments (Fig. 4).

The model randomly selects a molecule at each time step t and then moves that molecule to the other compartment. $x(t)$ denotes the number of molecules in compartment A at each time step. So $x(t) \in \{0, 1, 2, \dots, N\}$. The simulation tends toward a steady-state distribution with maximal entropy as

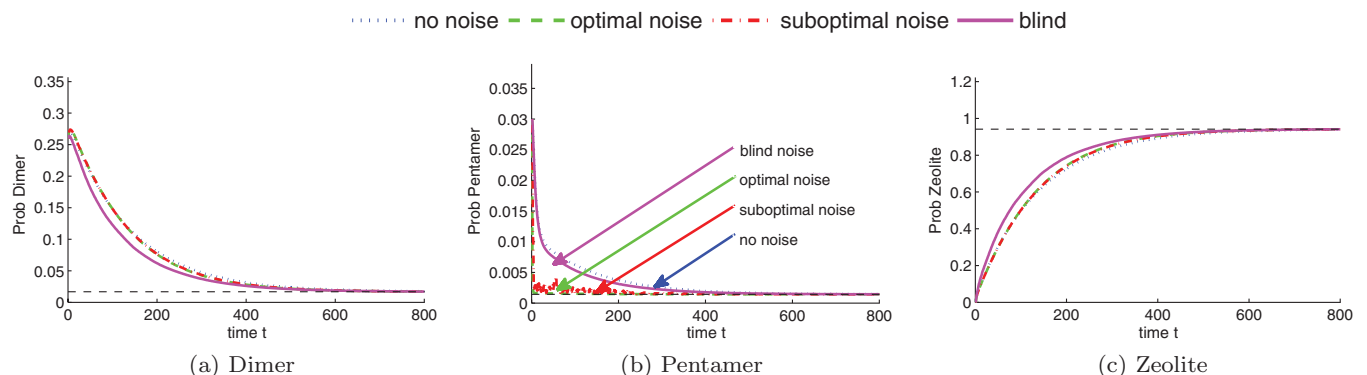


FIG. 3. (Color online) Noise benefits in an empirical chemical network Markov model. The figures show that noise enhances a crystallization model for zeolite resulting from a complex chemical process. The benefit appears in the shift to the left of the noise-enhanced simulations over the standard model. This indicates that the simulations reach steady state sooner. This simulation shows that the predicted noise benefit may be small but the benefit will exist for any simulation that has not converged. These curves represent concentrations of three species involved in a hypothetical zeolite synthesis over time. The blue (dotted) curve plots the standard (no noise) Markov chemical reaction model. The green (dashed), red (dash-dot), and pink (solid) curves show noisy versions of the model. The green (dashed) curve shows a simulation using the optimal noise N_{opt} in accord with the Markov chain noise benefit theorem. The red (dash-dot) curve shows the result when the simulated added suboptimal noise drawn uniformly from $[0, N_{\text{opt}}]$. The pink (solid) curve shows the results of Algorithm 2. Over time the zeolite concentrations dominate the other species.

TABLE I. Number of molecules ($N = 12$) per compartment in simulation state i .

State i	No. molecules in A	No. molecules in B
1	12	0
2	11	1
...
12	1	11
13	0	12

$t \rightarrow \infty$ [49]. The Ehrenfest model is a birth-death process because $x(t)$ either increases or decreases by one at each time step [50]. Suppose the container contains N molecules and has $0 < M < N$ molecules in compartment A at time t . Then

$$x(t) = M$$

and

$$P[x(t+1) = M-1] = \frac{M}{N}$$

$$P[x(t+1) = M+1] = 1 - \frac{M}{N}.$$

The Markov chain $x(t)$ evolves according to the state transition matrix P where

$$P_{ij} = \begin{cases} \frac{N-i}{N} & j = i+1 \\ \frac{i}{N} & j = i-1 \\ 0 & \text{else} \end{cases}$$

for $0 \leq i, j \leq N$ [51]. This model converges in distribution since $x(t) \neq x(t+1)$ for all t .

The Krafft-Schaefer extension adds two new parameters to the Ehrenfest diffusion model to describe asymmetry between transitions from $A \rightarrow B$ and $B \rightarrow A$ [52]. The two

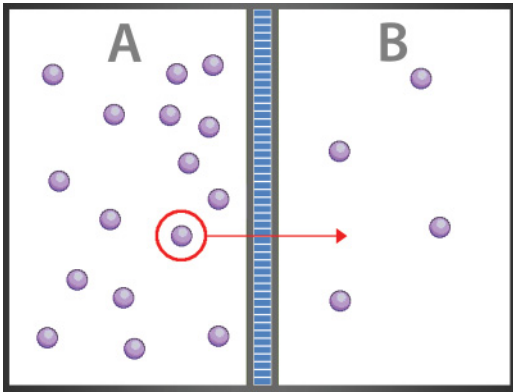


FIG. 4. (Color online) Two compartment Ehrenfest diffusion model. The figure illustrates the diffusion experiment of Ehrenfest. The box contains $N = 20$ molecules. The compartments A and B partition the box. $x(t)$ represents the number of molecules in box A at time t . This example assumes $x(t) = 16$. The simulation randomly selects a molecule at each time step (red circle) and moves the selected molecule to the other compartment (red arrow). Here $x(t+1) = 16 - 1 = 15$ since one molecule moves from A to B. The model exhibits a dynamic equilibrium because molecules continue to shuttle across the membrane for all t so $x(t) \neq x(t+1)$. So the occupancy $x(t)$ converges in distribution.

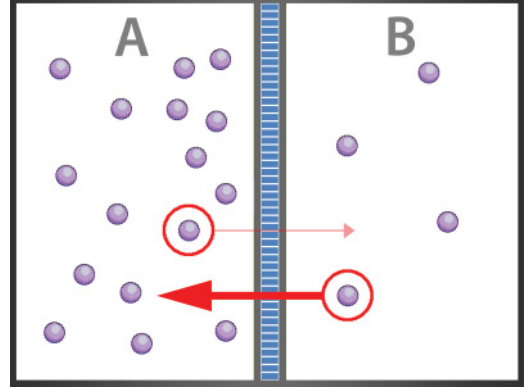


FIG. 5. (Color online) Two-compartment Krafft-Schaefer asymmetric diffusion model. The figure illustrates the membrane “preference” in the asymmetric Krafft-Schaefer diffusion model. Here $s \gg t$. So $P[B \rightarrow A] \gg P[A \rightarrow B]$ for a particular molecule (indicated by the relative size of the arrows). The asymmetry shifts the equilibrium to the left so that more molecules tend to accumulate in A at steady state.

parameters s and t characterize the transition asymmetry and scale the respective conditional transition probabilities from $A \rightarrow B$ and $B \rightarrow A$ [51,53]. This corresponds physically to the membrane “preferring” diffusion in one direction over the other (Fig. 5).

The generalized diffusion model evolves as a birth-death process with state transition matrix P where

$$P_{ij} = \begin{cases} \frac{N-i}{N}s & j = i+1 \\ \frac{i}{N}t & j = i-1 \\ \frac{1}{N}[(1-s)N + i(s-t)] & j = i \\ 0 & \text{else} \end{cases}$$

for $s, t \in [0, 1]$ and integers $0 \leq i, j \leq N$ [53]. The Krafft-Schaefer model converges with probability one only for the trivial case where one of the compartments is a perfect sink (when $s = 0$ or $t = 0$). The model weakens to convergence in distribution for all other s and t such that $0 < s, t < 1$.

Figure 1 shows a simulation that initialized x_0 as a normalized random state vector. This represents starting the diffusion simulation with uncertainty in the system’s configuration. The simulation used $s = 0.10$ and $t = 0.90$ to slow convergence and highlight the noise benefit. The asymmetry due to $s = 0.10$ and $t = 0.90$ collapses the dominant eigengap $|\lambda_1| - |\lambda_2|$ where λ_i is the i th largest magnitude eigenvalue. This increases the time for the simulation to reach steady state. A similar benefit exists for all s and t in $(0, 1)$. A wider eigengap $|\lambda_1| - |\lambda_2|$ ensures that the chain quickly converges toward a steady state. This results in a smaller noise benefit.

B. Noise benefits in a population genetics model

The second simulation shows a noise benefit in the Wright-Fisher population genetics model. The Wright-Fisher model uses a Markov chain to simulate stochastic genotypic drift during successive generations [54–56]. Figure 2 illustrates the noise benefit in a simulation with 2 alleles and $N = 50$ diploid individuals.

The Wright-Fisher model applies generally to populations under the following assumptions [57]:

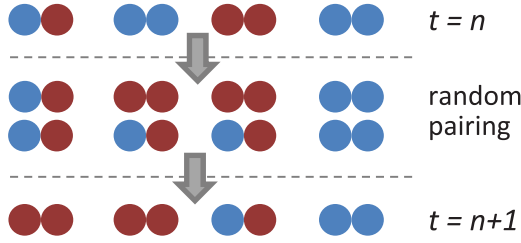


FIG. 6. (Color online) Wright-Fisher mating. The figure illustrates how the Wright-Fisher model produces successive generations. Each doublet in the first row represents the genotype of a diploid individual from a population $N = 4$. Each organism possesses a pair of alleles [blue (gray) = A_1 and red (black) = A_2]. The two middle rows show how the model randomly pairs individuals with replacement to form mates. Each doublet in the last row represents an offspring. The offspring inherit one allele (A_1 or A_2) randomly from each parent. Then the simulation “kills” the $t = n$ population and the offspring become the new $t = n + 1$ population.

- (1) the population size N remains constant between generations,
- (2) no selective difference between alleles,
- (3) nonoverlapping generations.

Consider a gene with 2 alleles (A_1 and A_2) in a population with N diploid individuals. The population contains $2N$ copies of the gene since each diploid individual has two copies of the gene. Let the state vector $x(t)$ represent the allele distribution at time t [58]. Then at time t

$$\begin{aligned} x_0(t) &= P[0 \text{ copies } A_1, 2N \text{ copies } A_2], \\ x_1(t) &= P[1 \text{ copies } A_1, 2N - 1 \text{ copies } A_2], \\ x_2(t) &= P[2 \text{ copies } A_1, 2N - 2 \text{ copies } A_2], \\ &\dots \\ x_{2N}(t) &= P[2N \text{ copies } A_1, 0 \text{ copies } A_2]. \end{aligned}$$

The Wright-Fisher model produces successive generations with a two-step process (Fig. 6). The model first creates N pairs of parents selected randomly and with replacement from the population. Then each pair produces a single offspring with its genotype inherited by selecting one gene from each parent. All parents die after mating.

The allele distribution $x(t)$ is a Markov chain that advances by random sampling with replacement from the pool of parent genes (Fig. 6) [59,60]. The density of alleles evolves according to a binomial probability density with

$$P[x(t+1) = j | x(t) = i] \sim \text{Bin}\left(j; 2N, \frac{i}{2N}\right). \quad (29)$$

Thus the Markov chain transition matrix has elements [58]

$$P_{i,j} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}.$$

Figure 7 also demonstrates how the allele distribution $x(t)$ converges to the steady state. $x(t)$ converges with probability one to either of the homozygous populations: either (A_1, A_1) or (A_2, A_2) [61]. This convergence is much stronger than the convergence in distribution found in the Ehrenfest diffusion model.

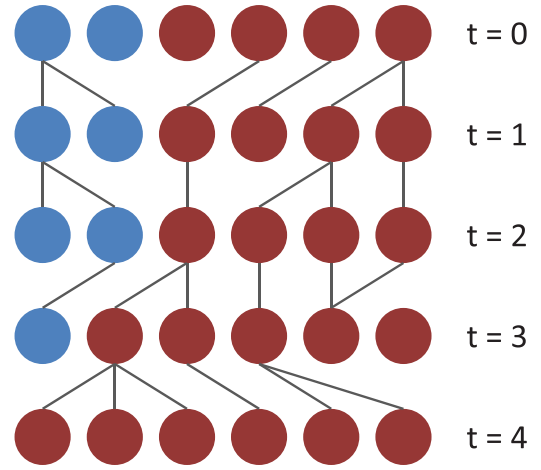


FIG. 7. (Color online) Markov dynamics of a Wright-Fisher genotype. Each of the six circles for $t = 0$ represents an allele for a particular gene [blue (gray) = A_1 and red (black) = A_2]. The Wright-Fisher model generates the $t = 1$ offspring by randomly sampling the $t = 0$ population with replacement. The connections indicate the surviving genes and their offspring. The A_1 allele becomes extinct by the fourth generation in this example. The steady state for this example is homozygous (A_2, A_2) because future generations can no longer inherit the extinct A_1 gene.

The Wright-Fisher simulation used a population $N = 50$ diploid individuals. The simulation tracked the allele distribution of a diallelic gene: A_1 and A_2 . It initialized the allele distribution $x(0)$ according to a normal distribution with a mean of 50 copies of A_1 and 50 copies of A_2 . This initial distribution represents imperfect information about the population’s initial genotypic makeup. The simulation evolved four separate copies of the initial population following the Fisher-Wright procedure: (1) standard (no noise), (2) applying Algorithm 1 by adding optimal noise N_{opt} at each iteration as prescribed by the theorem, (3) adding suboptimal noise uniformly chosen from $[0, N_{\text{opt}}]$, (4) applying Algorithm 2. Each copy ran for 500 generations.

Figure 2 shows two modes of noise benefit in the Wright-Fisher simulation: (1) noise shifts the over-damped system (damping ratio $\zeta > 1$) into a near critically damped regime ($\zeta \approx 1$) and (2) noise speeds the asymptotic approach toward the steady-state distribution. Each plot in the figure represents the estimate of the probability for a single genotypic distribution: Fig. 2(a) shows $P[100 \text{ copies } A_1, 0 \text{ copies } A_2]$ and Fig. 2(b) shows $P[50 \text{ copies } A_1, 50 \text{ copies } A_2]$ during the 500-step simulation. The population will reach one of two homozygous steady states: (A_1, A_1) or (A_2, A_2) . Stochastic dynamics during the simulation control the probabilities of the possible steady states. This particular simulation shows that $P[\text{steady state} = (A_1, A_1)] = 0.5 = P[\text{steady state} = (A_2, A_2)]$. This is the expected result because of the symmetric initial uncertainty for A_1 and A_2 .

Figure 2(a) shows that Algorithm 2 can introduce oscillations in the density estimate. The oscillations have a short-lived effect in this simulation. The *ringing* quickly dies down and the estimate settles to the theoretical limit = 0.5. The simulations in the other sections do not show this ringing artifact. We do not know if this ringing artifact arises from some relation between

the state transition probabilities, the number of states, or some other condition unique to this model.

Figure 2 also shows that even nonoptimal noise can benefit the simulation dynamics. The probability of the homozygous state in Fig. 2(a) is one of the two distributions with nonzero steady-state probability: $P[\text{steady state} = (A_1, A_1)] \neq 0$. The suboptimal noise simulation (red curve) shows similar benefits to the optimal noise simulation (green curve) since the traces of the two simulations resemble each other. This also appears to be an artifact of some special condition in this model.

C. Noise benefits in a chemical reaction model

The third simulation shows a noise benefit in a zeolite crystallization model. Figure 3 shows a benefit in a six-state chemical network simulation. The simulation extended an earlier study [62,63] that investigated a proposed crystallization process for natural zeolite [64]. The figure shows that the components approach their steady-state values (within 0.5% of steady state) 18.1% faster on average with added noise.

Thus the noise benefit can extend generally to a large domain of problems that employ *observed* transition matrices. We rarely deal with a pure Markov process in practice. We are even less likely to have complete knowledge of the state transition matrix. Researchers that model complex processes often estimate the transition matrix with approximate conditional transition probabilities calculated from a series of observations [65–72].

Zeolites are a class of aluminosilicates that form naturally under geologic conditions [73,74]. Geologists have identified 40 naturally occurring zeolite frameworks. Chemists have synthesized over 175 unique varieties [75,76]. Zeolites find uses in many industries. These include water purification [77–79], detergents [80,81], catalysis [82–85], and nuclear reprocessing [86–88].

The exact natural hydrothermal synthesis of many zeolites is not known [89–92]. Researchers have employed Markov models to predict properties from the complex chemistry involved in their formation [93–95]. Geologists constructed an observed transition matrix in one such model based on ²⁹Si concentration profiles during a formation experiment [63].

They then determined rate constants, equilibrium constants, and free energies for elementary zeolite-forming reactions for a hypothetical zeolite-formation process using Markov simulations with the estimated transition matrix and initial species concentrations. Our simulations show that noise benefits such a Markov model.

Hawkins [62] empirically found the following state transition probability matrix for six silica oligomers from aggregate NMR data using weighted least squares:

$$P = \begin{bmatrix} 0.9274 & 0.0700 & 0.0025 & 8 \times 10^{-5} & 10^{-5} & 10^{-5} \\ 0.0500 & 0.8395 & 0.1000 & 0.0100 & 0.0004 & 0.0001 \\ 0.0600 & 0.0600 & 0.8495 & 0.0300 & 0.0004 & 0.0001 \\ 0.0500 & 0.0100 & 0.0400 & 0.5400 & 0.0600 & 0.3000 \\ 0.0500 & 0.0200 & 0.0200 & 0.0500 & 0.8595 & 0.0005 \\ 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.0001 & 0.99953 \end{bmatrix} \quad (30)$$

corresponding to steady-state probability density x^∞

$$x^\infty = [0.026 \quad 0.017 \quad 0.013 \quad 0.002 \quad 0.002 \quad 0.942]. \quad (31)$$

Figure 8 summarizes the principal reaction pathways.

We used the experimental ²⁹Si NMR data reported earlier [62] to initialize the species concentrations to

$$x_0 = [0.430 \quad 0.260 \quad 0.220 \quad 0.060 \quad 0.030 \quad 0.000],$$

and advanced the Markov chain to simulate the crystallization of zeolite.

Our simulations show that noise benefits the empirical estimation but the observed benefit was small. The performance metric showed a strong benefit of 18.1% despite some states experiencing only minimal noise benefits [Figs. 3(a) and 3(c)]. This is because the noise quickly moved a few components to their steady-state value [Fig. 3(b)]. The Markov chain noise benefit theorem could not provide an additional benefit to the system after this initial boost because the theorem relies on the magnitude of the component closest to its steady-state value. Several components converged within a few time steps.

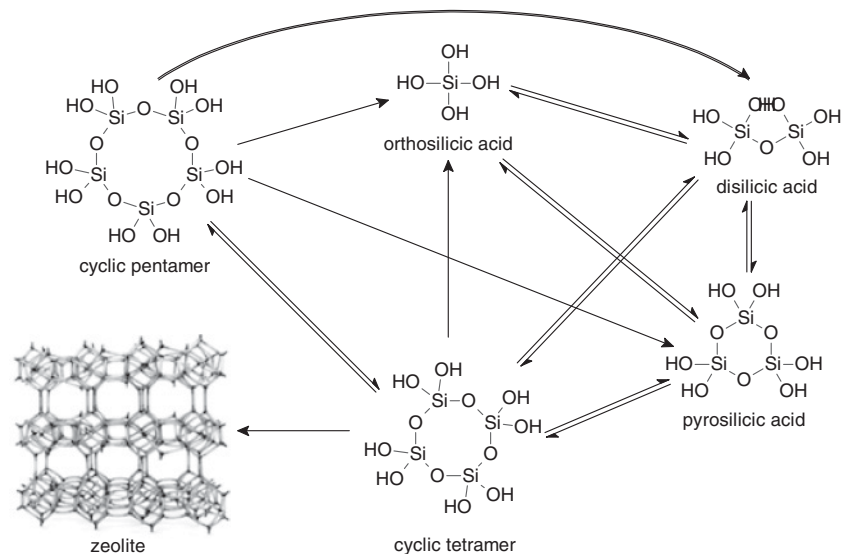


FIG. 8. Zeolite reaction scheme of Hawkins [62]. Simulations show that noise speeds the convergence of this model to its steady-state concentrations. The model synthesizes zeolite from five silicate oligomers. The reaction arrows show the dominant model pathways. The state transition matrix (30) defines each pathway as a Markov transition probability from one species to another during one time step. The vector (31) lists the steady-state concentrations of the six reactants. The system saturates with zeolite because the model lacks strong pathways that consume zeolite.

So the theorem-based noise added only small corrections to the states for the rest of the simulation. This shows that the theorem confers a larger benefit to systems with states that converge at approximately the same rate. But other Markov systems still receive some noise benefit.

V. MARKOV CHAIN NOISE BENEFIT THEOREM SIMULATION

Two final simulations show the noise benefit that exists for Markov chain simulations. The simulations show how the Markov chain noise benefit theorem might speed convergence in modern algorithms such as the Google PageRankTM link analysis algorithm [96–98]. The PageRank algorithm constructs a probability density that represents the likelihood that a person randomly clicking on links will arrive at a particular page over all indexed pages on the Internet [99]. The algorithm operates on a dataset called the *Google matrix*. This matrix is equivalent to a Markov state transition matrix spanning tens of billions of dimensions [100,101]. The noise benefit theorem shows that the algorithm should benefit from noise.

A. One-step Markov chain simulation

The first simulation shows that a Markov chain can benefit from additive noise (Fig. 9). The simulation shows the benefit after one time step as a decrease of the absolute error between the posterior state density and the stationary state density. Tables II and III show a large decrease in the absolute error in the noisy simulations compared to the no-noise simulations.

Table II summarizes the one-step experiment with and without noise. The simulation classified the states as satisfying either the conditions of the Markov chain noise benefit theorem $\Delta_i = (x - x^\infty)P_i > 0$ or the Corollary $\Delta_i = (x - x^\infty)P_i < 0$. It set the noise strength to $A = \min\{a_i\}$ for each class in accord with the theorem. This gave $A = 0.0682$ for the states with $\Delta_i > 0$ and $A = -0.1594$ for the states with $\Delta_i < 0$. The simulation calculated the total absolute error for each class using the respective values for the noise strength A .

Table III summarizes a simulation with and without noise that does not have access to signs of the inequality. It shows that a noise benefit exists even if the simulation cannot classify individual states according to $\Delta_i > 0$ or $\Delta_i < 0$. The table summarizes the relative improvement over all $N = 6$ states when setting the noise strength to $A = \text{sign}(a_i) \cdot \min(|a_i|)$. This gave $A = \min(0.0682, 0.1594) = 0.0682$. The simulation calculated the total absolute error for the posterior state density using this value of A .

The Markov chain noise benefit theorem ensures that there exists a noise distribution that reduces the state error. Figure 9 illustrates this because it shows that the error decreases as the

TABLE II. Noise benefits in one-step Markov chain simulation.

	States satisfying MC Theorem	States satisfying corollary
No noise	0.1547	0.1547
With noise	0.0547	0.0724
Error decrease (%)	64.64%	53.20%

TABLE III. Noise benefits in one-step Markov chain simulation with unknown sign error.

No noise	0.3093
With noise	0.2370
Error decrease (%)	23.38%

noise strength increases from zero. The theorem and corollary also establish that past some noise strength ($A > 0$) the error will increase. Thus properly signed noise with magnitude less than A guarantees that the absolute error will be lower in the noisy simulation than in the no-noise simulation.

Figure 9 shows an example where three of the $N = 6$ states obey the inequality (13) in the main theorem and the remaining three states obey the inequality (18) in the corollary. Not all transition matrices P have this even splitting. But any given matrix will have at least one state that satisfies each case since the sum of the signed errors must equal 0.

The simulation generated a Markov chain from a fixed random transition matrix where $\hat{P}_{i,j} = P[X_{k+1} = j | X_k = i] \sim U(0,1)$. The simulation used $N = 6$ states. The theorem and corollary guarantee the benefit for transition matrices with any finite dimension. But uniformly chosen transition matrices tend to generate a uniform stationary density

$$x_j^\infty = \frac{1}{N}. \quad (32)$$

We transformed each transition probability by $U(0,1) + \epsilon$ to construct a network of states with nonuniform importance. We chose $\epsilon > 0.04$ to avoid numerical instability. This gives a transition matrix

$$\hat{P}_{i,j} = \frac{U_{i,j}}{V_{i,j} + \epsilon} \quad (33)$$

where $U_{i,j} \sim U(0,1)$ and $V_{i,j} \sim U(0,1)$. We normalized the rows of \hat{P} to form a proper stochastic matrix:

$$P_{i,j} = \frac{\hat{P}_{i,j}}{\sum_{k=1}^N \hat{P}_{i,k}}. \quad (34)$$

We chose the initial state density x as the uninformed prior [36] (uniform distribution) over the six states so that

$$x_j = \frac{1}{N}. \quad (35)$$

We used MATLAB R2009b to perform the simulations with transition matrix P

$$P = \begin{bmatrix} 0.038 & 0.040 & 0.077 & 0.070 & 0.065 & 0.710 \\ 0.017 & 0.109 & 0.140 & 0.128 & 0.234 & 0.372 \\ 0.014 & 0.022 & 0.062 & 0.174 & 0.005 & 0.723 \\ 0.027 & 0.053 & 0.068 & 0.184 & 0.058 & 0.611 \\ 0.071 & 0.075 & 0.015 & 0.132 & 0.011 & 0.696 \\ 0.181 & 0.177 & 0.484 & 0.017 & 0.068 & 0.073 \end{bmatrix}$$

corresponding to steady-state probability density x^∞

$$x^\infty = [0.089 \quad 0.102 \quad 0.241 \quad 0.094 \quad 0.065 \quad 0.408].$$

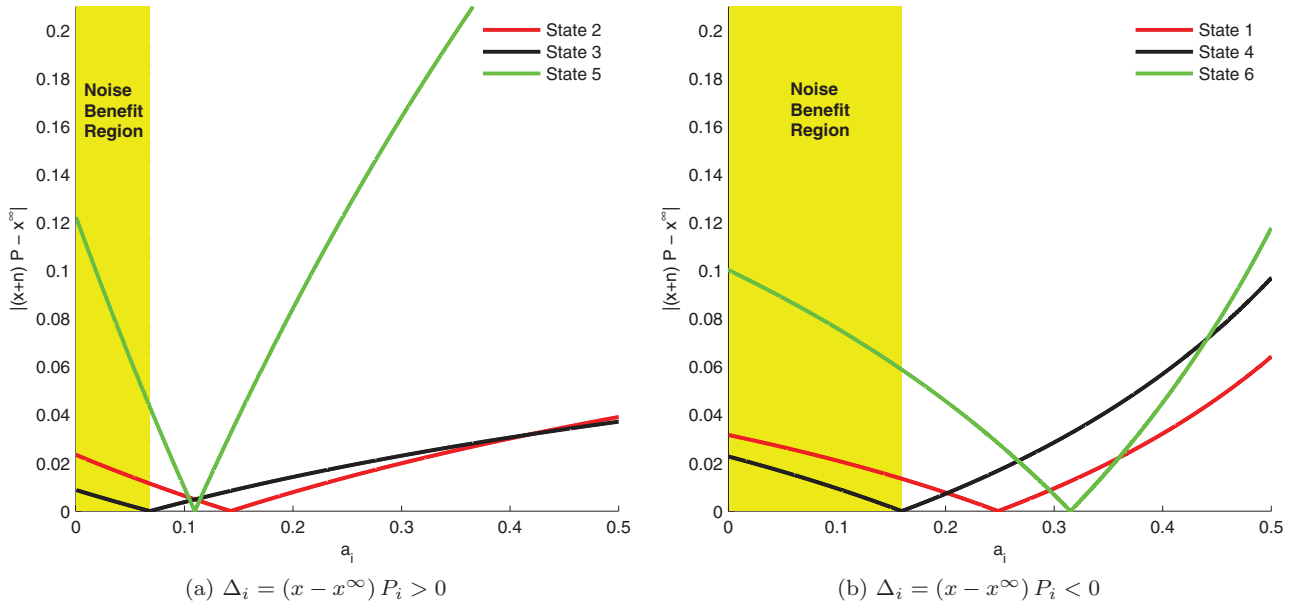


FIG. 9. (Color online) Noise benefits in Markov chain density estimation. These figures show the relation between the error magnitude of each Markov state and the noise strength a_i . The simulation used a six-state Markov chain and the figure shows the single-step absolute errors by state. Each of the six states satisfied either (a) the Markov chain noise benefit theorem: $(x - x^\infty)P_i > 0$ or (b) the Corollary: $(x - x^\infty)P_i < 0$. (a) Three states satisfy the inequality $(x - x^\infty)P_i > 0$ in this simulation. Each curve represents the absolute error $|\tilde{x}P - x^\infty|_i$ of the i th state as a_i increases. The standard zero-noise condition corresponds to $a_i = 0$. Each state has an optimal noise level A_i indicated by the point where the curve meets the a_i axis. The optimal noise A_i will exactly drive the state to its stationary value. The Markov chain noise benefit theorem first shows that the benefit exists for all $a_i < A_i$. The theorem also guarantees the existence of a global $A = \min\{A_i\} > 0$ such that any noise $a < A$ benefits every state that satisfies the inequality. All curves decrease (strictly) monotonically until they reach A_i . Thus any point between the no-noise condition and A_i shows some benefit and $A = \min\{A_i\}$ satisfies this constraint for each such state. (b) Three states satisfy the alternative inequality $(x - x^\infty)P_i < 0$. These correspond to the states that satisfy the corollary. The corollary ensures a point A so that any noise-strength less than A benefits every such state.

B. Two-step Markov chain simulation

The second simulation shows that the noise benefits in the one-step simulation extend over successive time steps (Fig. 10). We measure the benefit as a decrease in the absolute error between the posterior state density and the stationary state density. The simulation also shows that even suboptimal noise in one time step can still benefit successive steps. The proof guarantees that there exists a noise density that will reduce the error over multiple time steps.

We generated a transition matrix P using the same procedure as in the one-step simulation (34):

$$P = \begin{bmatrix} 0.147 & 0.013 & 0.051 & 0.667 & 0.062 & 0.061 \\ 0.158 & 0.030 & 0.088 & 0.622 & 0.012 & 0.090 \\ 0.078 & 0.061 & 0.095 & 0.582 & 0.077 & 0.108 \\ 0.138 & 0.106 & 0.055 & 0.565 & 0.039 & 0.098 \\ 0.171 & 0.085 & 0.213 & 0.085 & 0.170 & 0.276 \\ 0.048 & 0.028 & 0.070 & 0.804 & 0.030 & 0.020 \end{bmatrix}.$$

This corresponds to the steady-state probability density

$$x^\infty = [0.129 \quad 0.077 \quad 0.068 \quad 0.582 \quad 0.048 \quad 0.094].$$

VI. CONCLUSION

We have shown that noise can benefit Markov chain estimation by speeding up the convergence time if the algorithm can calculate the sign of the state error. We have also shown how a simulation can use estimates of the error magnitude to update its current estimate of the underlying state density. Simulations confirm that noise can benefit a single-step or multistep system even if the system has insufficient information to determine the optimal noise.

Versions of the Markov chain noise benefit theorem may well hold for weaker assumptions and other Markov chain models. An open question is whether the results hold for noise-perturbed Markov transition matrices instead of noise-perturbed state densities. This may apply to simulations with noisy estimates of the transition matrix or to simulations with transition matrix estimates based on only a few observations. This might also apply to specific MCMC algorithms under suitable assumptions. Adaptive algorithms may be able to find optimal noise amounts in many of these cases.

APPENDIX: PROOF OF THE MARKOV CHAIN NOISE BENEFIT THEOREM

Markov Chain Noise Benefit Theorem. Suppose M is a finite time-homogeneous Markov chain with N states and transition matrix P . Suppose further that M is irreducible and aperiodic.

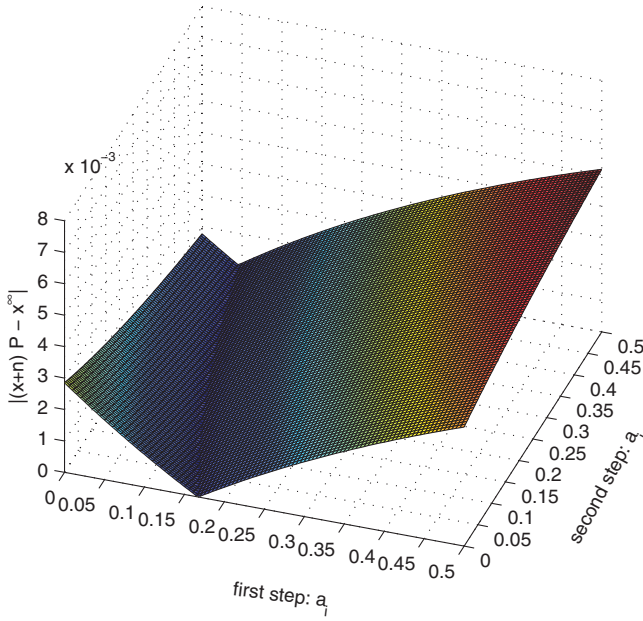


FIG. 10. (Color online) Multicycle noise benefits in Markov chain density estimation. This figure shows that the noise benefits apply for successive Markov steps. It further shows that even suboptimal noise in one iteration can still benefit successive steps. The simulation evaluated the deciding inequalities (13) and (18) for a single state at two successive time steps and used only the sign (+ or -) to determine the direction of beneficial noise for the state. The plot shows the relation between the state's absolute error and the noise magnitude during the “first step” and “second step” (with the appropriate sign). The origin of the $a_1^{(\text{step1})}$ and $a_2^{(\text{step2})}$ axes corresponds to a zero-noise two-step Markov chain. The optimal noise corresponds to $a_1^{(\text{step1})} \approx 0.165$ during the first step. Then there is a strictly positive value for $a_2^{(\text{step2})}$ that yields a lower error than the zero-noise case even if the system applies suboptimal noise during the first step such as $a_1 = 0.10$.

Then for all nonstationary state density vectors x there exists a noise benefit in the sense that there exists some $A > 0$ so that for all $a \in (0, A)$

$$|[\tilde{x}P - x^\infty]_i| < |[xP - x^\infty]_i| \quad (\text{A1})$$

for all states i with

$$\Delta_i = (x - x^\infty)P_i > 0 \quad (\text{A2})$$

where

$$\tilde{x} = \frac{1}{1+a}(x+n) \quad (\text{A3})$$

is the normalized state vector after adding a noise vector n with only one nonzero component

$$n_j = \begin{cases} a & j = k \\ 0 & j \neq k \end{cases} \quad (\text{A4})$$

for any k that satisfies

$$\Delta_k = (x - x^\infty)P_k > 0. \quad (\text{A5})$$

Proof. Fix x as a state vector of the Markov chain M . Note first that \tilde{x} is a probability density function over the states of M because of (a) and (b) as follows:

(a) \tilde{x} is a N -vector with $\tilde{x}_i \geq 0$ since

$$[\tilde{x}]_i = \left[\frac{1}{1+a}(x+n) \right]_i \quad (\text{A6})$$

$$= \begin{cases} \frac{1}{1+a}x_i & i \neq k \\ \frac{1}{1+a}(x_i + a) & i = k \end{cases} \quad (\text{A7})$$

$$\geq \frac{1}{\max(1, A)}x_i \quad (\text{A8})$$

$$\geq 0 \quad (\text{A9})$$

since $a > 0$ and $A > 0$.

(b) $\sum \tilde{x}_i = 1$ since

$$\sum_{i=1}^N \tilde{x}_i = \frac{1}{1+a} \left(\sum_{i=1}^N x_i + \sum_{i=1}^N n_i \right) \quad (\text{A10})$$

$$= \frac{1}{1+a}(1+a) \quad (\text{A11})$$

$$= 1. \quad (\text{A12})$$

Note that

$$|[\tilde{x}P - x^\infty]_i| < |[xP - x^\infty]_i| \quad (\text{A13})$$

$$= |[xP - x^\infty P]_i| \quad (\text{A14})$$

$$= |(x - x^\infty)P|_i \quad (\text{A15})$$

$$= |\Delta_i|. \quad (\text{A16})$$

The proof proceeds by showing that such an A_i exists for each component i that satisfies $\Delta_i = (x - x^\infty)P_i > 0$. This will complete the proof because $(0, A) = \cap_{i=1}^N (0, A_i) \neq \emptyset$ since $N < \infty$ and $\mu[(0, A_i)] > 0$ for each A_i .

Let i in $1 \leq i \leq N$ be any state that satisfies the inequality $\Delta_i = (x - x^\infty)P_i > 0$. Choose k in $1 \leq k \leq N$ and define

$$\tilde{x} = \frac{1}{1+a_i}(x+n) \quad (\text{A17})$$

with

$$n_j = \begin{cases} a_i & j = k \\ 0 & j \neq k \end{cases} \quad (\text{A18})$$

and $a_i > 0$. Then

$$|[\tilde{x}P - x^\infty]_i| = |[\tilde{x}P]_i - [x^\infty]_i| \quad (\text{A19})$$

$$= |[\tilde{x}P]_i - [x^\infty P]_i| \quad (\text{A20})$$

since $x^\infty = x^\infty P$. Expand \tilde{x} :

$$|[\tilde{x}P - x^\infty]_i| = \left| \sum_{j=1}^n \tilde{x}_j P_{j,i} - \sum_{j=1}^n x_j^\infty P_{j,i} \right| \quad (\text{A21})$$

$$= \left| \sum_{j=1}^n \left[\frac{1}{1+a_i}(x+n) \right]_j P_{j,i} - \sum_{j=1}^n x_j^\infty P_{j,i} \right| \quad (\text{A22})$$

$$= \left| \sum_{j=1}^n \frac{1}{1+a_i}(x_j + n_j) P_{j,i} - \sum_{j=1}^n x_j^\infty P_{j,i} \right| \quad (\text{A23})$$

$$= \left| \frac{1}{1+a_i} \sum_{j=1}^n (x_j P_{j,i} + n_j P_{j,i}) - \sum_{j=1}^n x_j^\infty P_{j,i} \right|. \quad (\text{A24})$$

Then add $0 = (\frac{a_i}{1+a_i} \sum x_j P_{j,i} - \frac{a_i}{1+a_i} \sum x_j P_{j,i})$ and group:

$$|[\tilde{x}P - x^\infty]_i| = \left| \frac{1}{1+a_i} \sum_{j=1}^n x_j P_{j,i} + \frac{1}{1+a_i} \sum_{j=1}^n n_j P_{j,i} + \left(\frac{a_i}{1+a_i} \sum_{j=1}^n x_j P_{j,i} - \frac{a_i}{1+a_i} \sum_{j=1}^n x_j P_{j,i} \right) - \sum_{j=1}^n x_j^\infty P_{j,i} \right| \quad (\text{A25})$$

$$= \left| \sum_{j=1}^n x_j P_{j,i} + \frac{1}{1+a_i} \sum_{j=1}^n n_j P_{j,i} - \frac{a_i}{1+a_i} \sum_{j=1}^n x_j P_{j,i} - \sum_{j=1}^n x_j^\infty P_{j,i} \right| \quad (\text{A26})$$

$$= \left| \left(\sum_{j=1}^n x_j P_{j,i} + \sum_{j=1}^n x_j^\infty P_{j,i} \right) - \frac{1}{1+a_i} \left(a_i \sum_{j=1}^n x_j P_{j,i} - \sum_{j=1}^n n_j P_{j,i} \right) \right| \quad (\text{A27})$$

$$= \left| (xP_i - x^\infty P_i) - \frac{1}{1+a_i} \left(a_i \sum_{j=1}^n x_j P_{j,i} - \sum_{j=1}^n n_j P_{j,i} \right) \right| \quad (\text{A28})$$

$$= \left| \Delta_i - \frac{1}{1+a_i} \left(a_i \sum_{j=1}^n x_j P_{j,i} - a_i P_{k,i} \right) \right| \quad (\text{A29})$$

$$= \left| \Delta_i - \frac{a_i}{1+a_i} \left(\sum_{j=1}^n x_j P_{j,i} - P_{k,i} \right) \right|. \quad (\text{A30})$$

So

$$|[\tilde{x}P - x^\infty]_i| = \left| \Delta_i - \frac{a_i}{1+a_i} (xP_i - P_{k,i}) \right|. \quad (\text{A31})$$

Now $\Delta_i > 0$ by hypothesis. Thus

$$\left| \Delta_i - \frac{a_i}{1+a_i} (xP_i - P_{k,i}) \right| < |\Delta_i| \quad (\text{A32})$$

if and only if

$$\frac{a_i}{1+a_i} (xP_i - P_{k,i}) > 0 \quad (\text{A33})$$

$$\text{and } \frac{a_i}{1+a_i} (xP_i - P_{k,i}) < 2\Delta_i \quad (\text{A34})$$

since $|\Delta_i| > |\Delta_i - b|$ if and only if $0 < b < 2\Delta_i$. The positivity constraint (A33) holds if and only if $xP_i > P_{k,i}$. The upper bound (A34) holds if and only if

$$a_i(xP_i - P_{k,i}) < 2\Delta_i(1+a_i). \quad (\text{A35})$$

Therefore Eq. (A34) holds if and only if

$$a_i(xP_i - P_{k,i} - 2\Delta_i) < 2\Delta_i. \quad (\text{A36})$$

If $2\Delta_i < xP_i - P_{k,i}$ then

$$a_i < \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} \quad (\text{A37})$$

and if $2\Delta_i > xP_i - P_{k,i}$ then

$$a_i > \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i}. \quad (\text{A38})$$

But if $2\Delta_i > xP_i - P_{k,i}$ then $\frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} < 0$. So any $a_i > 0$ suffices. Thus either

$$a_i > 0 \quad \text{if } 2\Delta_i < xP_i - P_{k,i} \quad (\text{A39})$$

or

$$a_i < \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} \quad \text{if } 2\Delta_i < xP_i - P_{k,i}. \quad (\text{A40})$$

Therefore if $a_i \in (0, A_i)$ with $A_i = \frac{2\Delta_i}{xP_i - P_{k,i} - 2\Delta_i} > 0$ then Eqs. (A39) and (A40) hold. So if $A = \min_i \{A_i\} > 0$ then the theorem holds for all states i that satisfy the inequality $\Delta_i = (x - x^\infty)P_i > 0$. ■

- [1] C. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, New York, 1996).
 [2] W. J. Freeman, H. J. Chang, B. C. Burke, P. A. Rose, and J. Badler, *IEEE Trans. Circuits Syst. I* **44**, 989 (1997).
 [3] B. Kosko, *Noise* (Viking, New York, 2006).
 [4] P. Hänggi, *Chem. Phys. Chem.* **3**, 285 (2002).
 [5] M. D. McDonnell, N. G. Stocks, C. E. M. Pearce, and D. Abbott, *Stochastic Resonance: From Suprathreshold Stochastic*

Resonance to Stochastic Signal Quantization (Cambridge University Press, Cambridge, England, 2008).

- [6] M. M. Wilde and B. Kosko, *J. Phys. A: Math. Theor.* **42**, 465309 (2009).
 [7] I. Lee, X. Liu, C. Zhou, and B. Kosko, *IEEE Trans. Nanotechnol.* **5**, 613 (2006).
 [8] A. Patel and B. Kosko, *Neural Netw.* **22**, 697 (2009).
 [9] B. Kosko and S. Mitaim, *Phys. Rev. E* **70**, 031911 (2004).

- [10] A. Patel and B. Kosko, *IEEE Trans. Signal Process.* **57**, 1655 (2009).
- [11] A. Patel and B. Kosko, *IEEE Trans. Neural Netw.* **19**, 1993 (2008).
- [12] D. Applebaum, *IEEE Trans. Neural Netw.* **20**, 1993 (2009).
- [13] A. R. Bulsara and A. Zador, *Phys. Rev. E* **54**, R2185 (1996).
- [14] M. Guerriero, P. Willett, S. Marano, and V. Matta, in *ICASSP* (IEEE, New York, 2008), pp. 3901–3904.
- [15] D. Rousseau and F. Chapeau-Blondeau, *Signal Process.* **85**, 571 (2005).
- [16] A. Patel and B. Kosko, *IEEE Signal Process. Lett.* **17**, 1005 (2010).
- [17] S. Mitaim and B. Kosko, in *Proceedings of the IEEE: Special Issue on Intelligent Signal Processing* (IEEE, New York, 1998), pp. 2152–2183.
- [18] A. A. Saha and G. V. Anand, *Signal Process.* **83**, 1193 (2003).
- [19] M. Guerriero, S. Marano, V. Matta, and P. Willett, *Trans. Sig. Proc.* **57**, 2 (2009).
- [20] M. D. McDonnell, N. G. Stocks, C. E. Pearce, and D. Abbott, *Phys. Lett. A* **352**, 183 (2006).
- [21] S. Mitaim and B. Kosko, *IEEE Trans. Neural Netw.* **15**, 1526 (2004).
- [22] A. Patel and B. Kosko, *Neural Netw.* **18**, 467 (2005).
- [23] P. Parmananda, G. J. Escalera Santos, M. Rivera, and K. Showalter, *Phys. Rev. E* **71**, 031110 (2005).
- [24] H. Chen, P. Varshney, S. Kay, and J. Michels, *IEEE Trans. Signal Process.* **55**, 3172 (2007).
- [25] S. Kay, *IEEE Signal Process. Lett.* **7**, 8 (2000).
- [26] C. R. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. (Springer, New York, 2004).
- [27] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis* (Cambridge University Press, Cambridge, England, 2005).
- [28] L. Tierney, *The Annals of Statistics* **22**, 1701 (1994).
- [29] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1954).
- [30] W. K. Hastings, *Biometrika* **57**, 97 (1970).
- [31] S. Geman and D. Geman, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**, 721 (1984).
- [32] A. E. Gelfand and A. F. M. Smith, *J. Am. Stat. Assoc.* **85**, 398 (1990).
- [33] Y. J. Sung and C. J. Geyer, *The Annals of Statistics* **35**, 990 (2007).
- [34] W. R. Gilks, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (CRC Press, London, 1996).
- [35] M. K. Cowles and B. P. Carlin, *J. Am. Stat. Assoc.* **91**, 883 (1996).
- [36] G. Grimmett and D. Stirzaker, *Probability and Random Processes*, 3rd ed. (Oxford University Press, New York, 2001).
- [37] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, 2nd ed. (Cambridge University Press, Cambridge, England, 2009).
- [38] R. W. Shonkwiler and F. Mendivil, *Explorations in Monte Carlo Methods* (Springer, New York, 2009).
- [39] F. P. Kelly, *Reversibility and Stochastic Networks* (John Wiley & Sons, New York, 1979).
- [40] R. Durrett, *Probability: Theory and Examples*, 3rd ed. (Thompson Brooks/Cole, Florence, KY, 2005).
- [41] M. J. Klein, *Physica* **22**, 569 (1956).
- [42] A. S. Novozhilov, G. P. Karev, and E. V. Koonin, *Briefings in Bioinformatics* **7**, 70 (2006).
- [43] S. C. Kou and S. G. Kou, *Adv. Appl. Probab.* **35**, 641 (2003).
- [44] H. Berthiaux and V. Mizonov, *The Canadian J. Chem. Eng.* **82**, 1143 (2004).
- [45] D. Diermeier and J. A. V. Mieghem, *Math. Comput. Modell.* **48**, 1497 (2008).
- [46] K. Schulten, Z. Schulten, and A. Szabo, *Physica A: Stat. Theor. Phys.* **100**, 599 (1980).
- [47] S. Balaji, H. M. Mahmoud, and O. Watanabe, *Statistics & Probability Letters* **76**, 666 (2006).
- [48] S. Prestipino, *Physica A: Stat. Mech. Appl.* **340**, 373 (2004).
- [49] K. Lindgren, *Phys. Rev. A* **38**, 4794 (1988).
- [50] J. Palacios and P. Tetali, *Statistics & Probability Letters* **30**, 119 (1996).
- [51] S. Karlin and J. McGregor, *J. Appl. Probab.* **2**, 352 (1965).
- [52] O. Krafft and M. Schaefer, *J. Appl. Probab.* **30**, 964 (1993).
- [53] H. Dette, *J. Appl. Probab.* **31**, 930 (1994).
- [54] R. A. Fisher, *The Genetical Theory of Natural Selection*, 2nd ed. (Dover, New York, 1958).
- [55] S. Wright, *Genetics* **16**, 97 (1931).
- [56] W. Feller, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman (University of California Press, Berkeley, CA, 1951), pp. 227–246.
- [57] G. Baxter, R. Blythe, and A. McKane, *Math. Biosci.* **209**, 124 (2007).
- [58] D. Waxman, *J. Theor. Biol.* **257**, 245 (2009).
- [59] T. Lenormand, D. Roze, and F. Rousset, *Trends in Ecology & Evolution* **24**, 157 (2009).
- [60] D. Waxman, *J. Theor. Biol.* **269**, 79 (2011).
- [61] P. Billingsley, *Probability and Measure*, 3rd ed. (Wiley Interscience, New York, 1995).
- [62] D. B. Hawkins, *Clays and Clay Minerals* **37**, 433 (1989).
- [63] R. J. Donahoe and J. Liou, *Geochim. Cosmochim. Acta* **49**, 2349 (1985).
- [64] C. S. Cundy and P. A. Cox, *Microporous Mesoporous Mater.* **82**, 1 (2005).
- [65] Q. Dai, X. Qing Liu, and T. Ming Wang, *J. Mol. Struct.* **803**, 115 (2007).
- [66] J. Yang, W. J. Bruno, W. S. Hlavacek, and J. E. Pearson, *Biophys. J.* **91**, 1136 (2006).
- [67] M. Lindén and M. Wallin, *Biophys. J.* **92**, 3804 (2007).
- [68] J. M. Conroy, T. G. Kolda, D. P. O’Leary, and T. J. O’Leary, *Laboratory Investigation* **80**, 1629 (2000).
- [69] Z. Song, X. Geng, A. Kusiak, and C. Xu, *Expert Systems with Applications* **38**, 10229 (2011).
- [70] E. Renshaw, *Computational Statistics & Data Analysis* **45**, 765 (2004).
- [71] G. Pulford, R. A. Kennedy, and S.-H. Chung, *Signal Processing* **43**, 207 (1995).
- [72] Y.-T. Hu, R. Kiesel, and W. Perraudin, *Journal of Banking & Finance* **26**, 1383 (2002).
- [73] C. Colella, in *Zeolites and Ordered Mesoporous Materials: Progress and Prospects*, edited by J. Cejka and H. van Bekkum, Vol. 157 of *Studies in Surface Science and Catalysis* (Elsevier, New York, 2005), pp. 13–40.

- [74] Y. Yukselen-Aksoy, *Applied Clay Science* **50**, 130 (2010).
- [75] C. Baerlocher, L. B. McCusker, and D. H. Olson, *Atlas of Zeolite Framework Types*, 6th ed. (Elsevier, New York, 2007).
- [76] R. V. Gaines, H. C. Skinner, E. E. Foord, B. Mason, and A. Rosenzweig, *DANA's New Mineralogy*, 8th ed. (John Wiley & Sons, New York, 1997).
- [77] C. Covarrubias, R. García, R. Arriagada, J. Yáñez, H. Ramanan, Z. Lai, and M. Tsapatsis, *J. Membr. Sci.* **312**, 163 (2008).
- [78] S. Wang and Y. Peng, *Chem. Eng. J.* **156**, 11 (2010).
- [79] E. Álvarez Ayuso, A. García-Sánchez, and X. Querol, *Water Res.* **37**, 4855 (2003).
- [80] C. J. Adams, A. Araya, S. W. Carr, A. P. Chapple, K. R. Franklin, P. Graham, A. R. Minihan, T. J. Osinga, and J. A. Stuart, in *Progress in Zeolite and Microporous Materials, Proceedings of the 11th International Zeolite Conference*, edited by S.-K. I. Hakze Chon and Y. S. Uh, Vol. 105 (Elsevier, New York, 1997), pp. 1667–1674.
- [81] I. Yamane and T. Nakazawa, in *New Developments in Zeolite Science and Technology, Proceedings of the 7th International Zeolite Conference*, edited by A. I. Y. Murakami and J. Ward, Vol. 28 of *Studies in Surface Science and Catalysis* (Elsevier, New York, 1986), pp. 991–1000.
- [82] J. Weitkamp, *Solid State Ionics* **131**, 175 (2000).
- [83] M. Stöcker, *Microporous Mesoporous Mater.* **82**, 257 (2005).
- [84] J. A. Rabo and M. W. Schoonover, *Appl. Catal., A* **222**, 261 (2001).
- [85] A. Bielanski and A. Malecka, *Zeolites* **6**, 249 (1986).
- [86] A. Dyer and D. Keir, *Zeolites* **4**, 215 (1984).
- [87] A. Dyer and H. Faghihian, *Microporous Mesoporous Mater.* **21**, 27 (1998).
- [88] M. Flytzani-Stephanopoulos, I.-S. Nam, and X. Verykios, edited by, *Applied Catalysis B: Environmental* **5**, N34 (1995).
- [89] H. Ghobarkar, O. Schäf, B. Paz, and P. Knauth, *J. Solid State Chem.* **173**, 27 (2003).
- [90] Z. A. Lethbridge, J. J. Williams, R. I. Walton, K. E. Evans, and C. W. Smith, *Microporous Mesoporous Mater.* **79**, 339 (2005).
- [91] H. Ghobarkar and O. Schäf, *Mater. Res. Bull.* **34**, 517 (1999).
- [92] H. Ghobarkar and O. Schäf, *Microporous Mesoporous Mater.* **23**, 55 (1998).
- [93] M. D. Hollingsworth, *Science* **295**, 2410 (2002).
- [94] M. Falcioni and M. W. Deem, *J. Chem. Phys.* **110**, 1754 (1999).
- [95] M. G. Wu, M. W. Deem, S. A. Elomari, R. C. Medrud, S. I. Zones, T. Maesen, C. Kibby, C.-Y. Chen, and I. Y. Chan, *J. Phys. Chem. B* **106**, 264 (2002).
- [96] L. Page, U.S. Patent no. 6285999: Method for node ranking in a linked database (Sep. 4, 2001).
- [97] Google, Inc. annual report (Form 10-K) (2010).
- [98] Google, Inc. annual report (Form 10-K) (2009).
- [99] S. Brin and L. Page, *Proceedings of the Seventh International World-Wide Web Conference (WWW 1998)* (International World-Wide Web Conference Committee IW3C2, Brisbane, Australia, 1998).
- [100] Y. Lin, X. Shi, and Y. Wei, *J. Comput. Appl. Math.* **224**, 702 (2009).
- [101] S. Kamvar, T. Haveliwala, C. Manning, and G. Golub, *Proceedings of the Twelfth International World-Wide Web Conference (WWW 2003)* (International World-Wide Web Conference Committee IW3C2, Budapest, Hungary, 2003).