Exact fluctuations of nonequilibrium steady states from approximate auxiliary dynamics: Supplementary Information

Ushnish Ray and Garnet Kin-Lic Chan
Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125

David T. Limmer
Department of Chemistry, University of California, Berkeley, CA 94609
Kavli Energy NanoScience Institute, Berkeley, CA 94609 and
Materials Science Division, Lawerence Berkeley National Laboratory, Berkeley, CA 94609
(Dated: April 14, 2018)

GUIDING DISTRIBUTION FUNCTION IN THE CONTINUUM

Here we derive the tilted propagator used for importance sampling the driven brownian walker in the main text and discuss the adjoint operator for generating the Langevin dynamics. The Langevin equation that governs the overdamped motion of a particle in the presence of an external force, $F(\theta)$, is given by,

$$\partial_t \theta = F(\theta) + \eta \tag{S1}$$

As in the main text, the associated Fokker-Planck equation corresponding to this Langevin equation is

$$\partial_t p_t(\theta) = \mathcal{W} p_t(\theta) \tag{S2}$$

where $p_t(\theta)$ is the probability of observing the particle at θ at a time t and

$$\mathcal{W} = -\partial_\theta(F - \partial_\theta) \tag{S3}$$

is the Fokker-Planck operator. We are interested in the large deviation function of the entropy production, $s(t_N) = \sigma(t_N)t_N$, which is proportional to the current around the ring, unwrapped so that winding numbers are included,

$$s(t_N) = \int_0^{t_N} f\dot{\theta}(\tau)d\tau = f\Delta\theta(t_N) \equiv fx \tag{S4}$$

The generating function is related to the probability $p(\theta, s, t_N)$ of finding the particle at θ with entropy produced s at a time t_N by the Laplace transform,

$$\rho(\theta, \lambda, t) = \int ds\, e^{-\lambda s} p(\theta, s, t_N). \tag{S5}$$

Since we are interested in the behavior of probability distribution conditioned on x and θ and since both of them share the same noise source η, we expand the Fokker-Planck operator via $\partial_\theta \to \partial_\theta + \partial_x$ to obtain:

$$\tilde{\mathcal{W}} = \mathcal{W} + (2\partial_\theta - F(\theta))\partial_x + \partial_x^2 \tag{S6}$$

and the corresponding modified Fokker-Planck equation is

$$\partial_t p(\theta, s, t) = \tilde{\mathcal{W}} p(\theta, s, t). \tag{S7}$$

By differentiating Eq. S5 with respect to t and inserting Eq. S7 we get

$$\partial_t \rho(\theta, \lambda, t) = \int ds e^{-\lambda s} \tilde{\mathcal{W}} p(\theta, s, t). \tag{S8}$$

Now,

$$\begin{aligned}\int ds e^{-\lambda s} \tilde{\mathcal{W}} p(\theta, s, t) = &\int ds e^{-\lambda s} \mathcal{W} p(\theta, s, t) \\ &+ \int ds e^{-\lambda s}(2\partial_\theta - F(\theta))\partial_x p(\theta, s, t) \\ &+ \int ds e^{-\lambda s}\partial_x^2 p(\theta, s, t)\end{aligned} \tag{S9}$$

Performing integration by parts we get:

$$\begin{aligned}\partial_t \rho(\theta, \lambda, t) =& \mathcal{W} \int ds e^{-\lambda s} p(\theta, s, t) \\ &+ (2\partial_\theta - F(\theta))(f\lambda) \int ds e^{-\lambda s} p(\theta, s, t) \\ &+ (f\lambda)^2 \int ds e^{-\lambda s} p(\theta, s, t) \\ =& \mathbb{W}_\lambda \int ds e^{-\lambda s} p(\theta, s, t) = \mathbb{W}_\lambda \rho(\theta, \lambda, t),\end{aligned} \tag{S10}$$

where,

$$\mathbb{W}_\lambda = \mathcal{W} + 2f\lambda\partial_\theta + f\lambda(f\lambda - F(\theta)), \tag{S11}$$

is the tilted operator given by Eq. 8 in the main text used to obtain the modified or tilted dynamics [1]. The mapping of the second-order differential operator onto a stochastic diffusion process then follows the well-known Feynman-Kac theorem [2].

Following the general derivation of importance sampling in the main text, we use the specific form of $\mathbb{W}_\lambda$ to get

$$\tilde{\Xi}\mathbb{W}_\lambda(\tilde{\Xi}^{-1}\rho) = \rho[\tilde{\Xi}\mathbb{W}_\lambda\tilde{\Xi}^{-1} + \partial_\theta\zeta] - \partial_\theta(\zeta - \partial_\theta)\rho, \tag{S12}$$

with $\zeta = F - 2f\lambda + 2\partial_\theta \ln\tilde{\Xi}$. The last term corresponds to a Fokker-Planck type of operator with a force term equal to ζ and can be used to generate trajectories. The first term changes normalization and is, therefore, used for branching. The equation of motion for θ is obtained from the adjoint operator $\mathcal{W}_\lambda^\dagger$, where,

$$\tilde{\Xi}^{-1}\mathbb{W}_\lambda^\dagger\tilde{\Xi} = \tilde{\Xi}\mathbb{W}_\lambda\tilde{\Xi}^{-1} + \partial_\theta\zeta \tag{S13}$$

which follows from integration by parts. The specific form of the branching term following the operators in the main text is given by

$$\tilde{\Xi}^{-1}\mathbb{W}_\lambda^\dagger\tilde{\Xi} = \frac{1}{\tilde{\Xi}}\frac{d^2\tilde{\Xi}}{d\theta^2} + (F - 2\lambda f)\frac{d\ln\tilde{\Xi}}{d\theta} + f\lambda(f\lambda - F(\theta)), \tag{S14}$$

where $F(\theta) = f - \partial_\theta V(\theta)$, with f a constant, nonconservative force, and $V(\theta) = v_o\cos(\theta)$ is a periodic potential.

As discussed in the text, the exact GDF can be computed by expanding $\tilde{\Xi}$ is plane waves. Specifically, we express

$$\tilde{\Xi}(\theta) = \sum_{n=-m}^{m} c_n e^{in\theta} \tag{S15}$$

and inserting this expansion into eigenvalue relation, and using the tilter operator in Eq. S11, we can derive a recursion relationship for the coefficients c_n. This yields,

$$c_n a_n - c_{n-1}b_- - c_{n+1}b_+ = c_n\psi(\lambda) \tag{S16}$$

with elements

$$a_n = in(f + f\lambda) - n^2/2 + \lambda f^2 + \lambda^2 f^2/2$$

$$b_\pm = \pm v_o(-i\lambda f + n \pm 1)/2$$

which specifies a tridiagonal matrix that can be diagonalized using standard techniques suitable for non-Hermitian matrices. In practice, we use $2M + 1$ basis functions with $M = 50$ for the exact calculations. The instantonic solution is generated using only 1 basis function. Note that Ref. [3] contains an error in the typesetting of the equations defining the expansion coefficients.

GUIDING DISTRIBUTION FUNCTIONS FROM DISCRETE MEAN-FIELD SOLUTIONS

In this section we outline the procedure needed to generate mean-field (MF) and cluster solutions that form the guiding distribution functions in our discrete models. We will illustrate the procedure for the SEP but it can be easily generalized to other models.

For the open boundary SEP the tilted matrix may be written as:

$$\mathbb{W}_\lambda = w_L + \sum_{i=1}^{L} w_i + w_R \tag{S17}$$

where the single particle transition matrices $\{w_L, w_i, w_R\}$ and their operator representations are given by

$$w_L = \begin{bmatrix} -\alpha & \gamma e^{\lambda} \\ \alpha e^{-\lambda} & -\gamma \end{bmatrix} \qquad w_R = \begin{bmatrix} -\nu & \beta e^{-\lambda} \\ \nu e^{\lambda} & -\beta \end{bmatrix} \tag{S18}$$

and

$$w_i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -q & pe^{-\lambda} & 0 \\ 0 & qe^{\lambda} & -p & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{S19}$$

These matrices can be combined to form a many-body matrix that in operator notation is written as

$$\begin{aligned}\mathbb{W}_\lambda = &-\alpha(\mathbb{1} - \hat{n}_1) + \alpha e^{-\lambda}\hat{c}_1^\dagger + \gamma e^{\lambda}\hat{c}_1 - \gamma\hat{n}_1 + pe^{-\lambda}\hat{c}_2^\dagger\hat{c}_1 - p\hat{n}_1(\mathbb{1} - \hat{n}_2) \\ &+ \sum_{i=2}^{L-1} pe^{-\lambda}\hat{c}_{i+1}^\dagger\hat{c}_i - p\hat{n}_i(\mathbb{1} - \hat{n}_{i+1}) + qe^{\lambda}\hat{c}_{i-1}^\dagger\hat{c}_i - q\hat{n}_i(\mathbb{1} - \hat{n}_{i-1}) \\ &- \nu(\mathbb{1} - \hat{n}_L) + \nu e^{\lambda}\hat{c}_L^\dagger + \beta e^{-\lambda}\hat{c}_L - \beta\hat{n}_L + qe^{\lambda}\hat{c}_{L-1}^\dagger\hat{c}_L - q\hat{n}_L(\mathbb{1} - \hat{n}_{L-1}).\end{aligned} \tag{S20}$$

Here $\hat{c}_i^\dagger$ creates a particle at site i, $\hat{c}_j$ destroys a particle at site j, and $\hat{n}_i$ counts the number of particles at site i. The combined operators $\hat{c}_i^\dagger\hat{c}_j$ correspond to kinetic-like terms that move a particle from site j to i, and $\hat{n}_i(\mathbb{1} - \hat{n}_j)$ represents a hard-core interaction that prevents double occupation of sites. The exact solution is the usual eigenvalue problem $\mathbb{W}_\lambda|\Xi\rangle = \varepsilon|\Xi\rangle$ where $(\varepsilon, |\Xi\rangle)$ is a particular eigenpair (the inverse eigenvectors $\{\langle\Xi|\}$ can be constructed from $\{|\Xi\rangle\}$).

An obvious route to explore in constructing approximate GDF is to use a mean-field (MF) solution. The MF approach is to approximate the many-body state as a product state. Starting with a product of single site states, we can systematically improve our results by mov-

ing to products of cluster states, where interactions in the cluster are treated explicitly, while the links between clusters are treated at the MF level. We first illustrate the procedure for single site clusters and then show how to generalize to multi-site clusters.

Site-Decoupled Mean-Field and Generalized Variational Approximation

For the single site clusters we approximate the solution by the form $|\tilde{\Xi}\rangle = \prod_{i=1}^{L} |\xi_i\rangle$, where the single site state $|\xi_i\rangle = \sum_{n=0}^{1} \xi_i(n)|n_i\rangle$ is written in a basis of occupation numbers $\{|n\rangle\}$. Here $\xi_i(n)$ is a scalar function dependent on the occupation number n. One subtlety is that because the tilted matrix is not Hermitian we must distinguish between approximating its left and its right eigenvector. In this work, we will always approximate the right eigenvector. The corresponding left eigenvector is then made of the product of left states $\langle\xi_i| = \sum_{n=0}^{1} \xi_i'(n)\langle n_i|$, where $\langle\xi_i|$ (and the coefficients ξ_i') can be obtained from the right states by the biorthogonality condition, $\langle\xi_i|\xi_j\rangle = \delta_{ij}$. For the following discussion $\{|\tilde{\Xi}\rangle, \langle\tilde{\Xi}|\}$ represents the biorthogonal pair.

To determine the coefficients $\xi_i(n)$, we will use a generalized variational procedure to make the functional

$$\langle\tilde{\Xi}|(\mathbb{W}_\lambda - \varepsilon)|\tilde{\Xi}\rangle \tag{S21}$$

stationary with respect to independent linear variations in $\langle\tilde{\Xi}|$ and $|\tilde{\Xi}\rangle$. Although stationarity of Eq. S21 does not give a minimum principle, it leads to a mean-field approximation that approximates the right (or left) eigenvector equation when projected into an appropriate space, similar to the use of collocation to solve a partial differential equation. For example, considering stationarity with respect to small variations in the left state, we obtain

$$\langle\delta\tilde{\Xi}|\mathbb{W}_\lambda - \varepsilon|\tilde{\Xi}\rangle = 0 \tag{S22}$$

The small variations in the left state take the explicit form

$$\begin{aligned}\langle\delta\tilde{\Xi}| &= \sum_{i=1}^{L} \langle\delta\xi_i| \prod_{j\neq i}^{L} \langle\xi_j| \\ &= \sum_{i=1}^{L} \left[\sum_{n_i=0}^{1} \delta\xi_i'(n)\langle n_i|\right] \prod_{j\neq i}^{L} \left[\sum_{n_j=0}^{1} \xi_i'(n)\langle n_j|\right]\end{aligned} \tag{S23}$$

We then evaluate Eq. S22 with respect to the specific variations in Eq. S23 to obtain mean-field equations. Writing $\mathbb{W}_\lambda = (\mathcal{W}_\lambda - V_i) + V_i$ where V_i represents all terms in $\mathcal{W}_\lambda$ that contains terms involving site i, we obtain the following equation for each i:

$$\langle\delta\tilde{\Xi}|V_i|\tilde{\Xi}\rangle - \epsilon_i\langle\delta\xi_i|\xi_i\rangle = 0, \tag{S24}$$

which is equivalent to an eigenvalue equation with eigenvalue ϵ_i, where we have defined

$$\epsilon_i = \varepsilon - \left[\prod_{j\neq i}\langle\xi_j|\right](\mathbb{W}_\lambda - V_i)\left[\prod_{k\neq i}|\xi_k\rangle\right] \tag{S25}$$

As we are interested in the maximal eigenvalue/eigenvector pair of $\mathbb{W}_\lambda$, we are interested in the maximal eigenvalue/eigenvector pair for each site i.

For the SEP $\mathbb{W}_\lambda$, Eq. S24 implies solving the following decoupled eigenvalue equation for $i = 1$

$$\begin{aligned}&[(-\alpha - q\langle\hat{n}_2\rangle)(\mathbb{1} - \hat{n}_1) + (pe^{-\lambda}\langle\hat{c}_2^\dagger\rangle + \gamma e^{\lambda})\hat{c}_1 \\ &+ (\alpha e^{-\lambda} + qe^{\lambda}\langle\hat{c}_2\rangle))\hat{c}_1^\dagger + (-\gamma - p(1 - \langle\hat{n}_2\rangle))\hat{n}_1]|\xi_1\rangle \\ &= \varepsilon_1|\xi_1\rangle\end{aligned} \tag{S26}$$

where we have used the shorthand $\langle\ldots\rangle$ to denote the expectation value $\langle\tilde{\Xi}|\ldots|\tilde{\Xi}\rangle$. For $i = L$, the eigenvalue equation is

$$\begin{aligned}&[(-\nu - p\langle\hat{n}_{L-1}\rangle)(\mathbb{1} - \hat{n}_L) + (qe^{\lambda}\langle\hat{c}_{L-1}^\dagger\rangle + \beta e^{-\lambda})\hat{c}_{L-1} \\ &+ (\nu e^{\lambda} + pe^{-\lambda}\langle\hat{c}_{L-1}\rangle))\hat{c}_L^\dagger + (-\beta - q(1 - \langle\hat{n}_{L-1}\rangle))\hat{n}_L]|\xi_L\rangle \\ &= \varepsilon_L|\xi_L\rangle\end{aligned} \tag{S27}$$

Finally for all other i, the eigenvalue equation is

$$\begin{aligned}&[(-p\langle\hat{n}_{i-1}\rangle - q\langle\hat{n}_{i+1}\rangle)(\mathbb{1} - \hat{n}_i) \\ &+ (pe^{-\lambda}\langle\hat{c}_{i+1}^\dagger\rangle + qe^{\lambda}\langle\hat{c}_{i-1}^\dagger\rangle)\hat{c}_i \\ &+ (qe^{\lambda}\langle\hat{c}_{i+1}\rangle + pe^{-\lambda}\langle\hat{c}_{i-1}\rangle)\hat{c}_i^\dagger \\ &+ (-p(1 - \langle\hat{n}_{i+1}\rangle) - q(1 - \langle\hat{n}_{i-1}\rangle))\hat{n}_i]|\xi_i\rangle \\ &= \varepsilon_i|\xi_i\rangle\,.\end{aligned} \tag{S28}$$

The maximal eigenvalue/eigenvector pair approximation can be obtained by solving the above equations for each site i and choosing the set of states $\{|\xi_i\rangle\}$ corresponding to the largest eigenvalues $\{\epsilon_i\}$. Notice that even when collecting terms involving site i, there will inevitably be terms involving other sites due to the two-body interaction present in the matrix (S20). The natural way to solve this system, thus, requires the use of a self-consistent procedure. We start with an initial set of guess values for $\langle\hat{n}_i\rangle$, $\langle\hat{c}_i\rangle$ and $\langle\hat{c}_i^\dagger\rangle$ and proceed to solve the individual eigenvalue problems for each site. At the end of each iteration we use the states $\{|\xi_i\rangle\}$ to recompute $\langle\hat{n}_i\rangle$, $\langle\hat{c}_i\rangle$, and $\langle\hat{c}_i^\dagger\rangle$ and use them for the next iteration. This is continued until $\{\langle\hat{n}_i\rangle, \langle\hat{c}_i\rangle, \langle\hat{c}_i^\dagger\rangle\}$ do not change.

Once the solutions have converged it is straightforward to obtain the MF estimate of the LDF, $\epsilon(\lambda) = \langle\tilde{\Xi}|\mathbb{W}_\lambda|\tilde{\Xi}\rangle$. The MF approximation to the state $\langle\tilde{\Xi}|$ is precisely the GDF that we need to construct the auxiliary process that will importance sample the LDF. The effective matrices incorporating the importance sampling are:

$$\tilde{w}_L = \begin{bmatrix} -\alpha & \gamma e^{\lambda \frac{\xi_1(0)}{\xi_1(1)}} \\ \alpha e^{-\lambda \frac{\xi_1(1)}{\xi_1(0)}} & -\gamma \end{bmatrix} \tag{S29}$$

$$\tilde{w}_i = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -q & p e^{-\lambda \frac{\xi_i(0)\xi_{i+1}(1)}{\xi_i(1)\xi_{i+1}(0)}} & 0 \\ 0 & q e^{\lambda \frac{\xi_{i-1}(1)\xi_i(0)}{\xi_{i-1}(0)\xi_i(1)}} & -p & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{S30}$$

$$\tilde{w}_R = \begin{bmatrix} -\nu & \beta e^{-\lambda \frac{\xi_L(0)}{\xi_L(1)}} \\ \nu e^{\lambda \frac{\xi_L(1)}{\xi_L(0)}} & -\beta \end{bmatrix} \tag{S31}$$

Notice that these are not normalized and their renormalization determines the branching weights. In the text, we use a first order Trotterization on the short-time importance-sampled propagator $(\langle\tilde{\Xi}|e^{dt\mathbb{W}_\lambda}|\tilde{\Xi}\rangle)$ to obtain the Markovian transition probability matrix $\tilde{\mathbb{U}}_\lambda \equiv \mathbb{1} + dt\langle\tilde{\Xi}|\mathbb{W}_\lambda|\tilde{\Xi}\rangle$. Therefore, $\tilde{\mathbb{U}}_\lambda$ follows directly from the transformed components of $\mathbb{W}_\lambda$, and is given by

$$\tilde{\mathbb{U}}_\lambda = \tilde{u}_L + \sum_{i=1}^{L} \tilde{u}_i + \tilde{u}_R, \tag{S32}$$

where,

$$\tilde{u}_L = \begin{bmatrix} 1-\alpha dt & \gamma e^{\lambda \frac{\xi_1(0)}{\xi_1(1)}} \\ \alpha e^{-\lambda \frac{\xi_1(1)}{\xi_1(0)}} & 1-\gamma dt \end{bmatrix} \tag{S33}$$

$$\tilde{u}_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-qdt & p e^{-\lambda \frac{\xi_i(0)\xi_{i+1}(1)}{\xi_i(1)\xi_{i+1}(0)}} & 0 \\ 0 & q e^{\lambda \frac{\xi_{i-1}(1)\xi_i(0)}{\xi_{i-1}(0)\xi_i(1)}} & 1-pdt & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{S34}$$

$$\tilde{u}_R = \begin{bmatrix} 1-\nu dt & \beta e^{-\lambda \frac{\xi_L(0)}{\xi_L(1)}} \\ \nu e^{\lambda \frac{\xi_L(1)}{\xi_L(0)}} & 1-\beta dt \end{bmatrix} \tag{S35}$$

are the associated transition probabilities at the ends of the lattice, $\tilde{u}_L$ and $\tilde{u}_R$, or in its interior, $\tilde{u}_i$. At every time step or Monte Carlo sweep, the current state $|\mathcal{C}\rangle = |n_1 n_2 ... n_L\rangle$ that corresponds to a column of $\tilde{\mathbb{U}}_\lambda$ is used to propose moves such that the outgoing state $|\mathcal{C}'\rangle$ has the probability $\tilde{\mathbb{U}}_\lambda(\mathcal{C}', \mathcal{C})/\mathcal{N}(C)$ of being accepted, where $\mathcal{N}(C) \equiv \sum_{\mathcal{C}'} \tilde{\mathbb{U}}_\lambda(\mathcal{C}', \mathcal{C})$ is the normalization factor. Over the course of the short trajectory generated in between branching steps, a walker's weight is accumulated as a product of these normalization factors.

Cluster Approach

The idea of a site-decoupled mean-field can be extended to multiple sites collected into clusters. The MF ansatz is then $|\tilde{\Xi}\rangle = \prod_{c=1}^{L/c_L} |\xi_c\rangle$ where c_L is the number of sites that constitutes a cluster. $|\xi_c\rangle$ is expanded in the occupation basis of the cluster $|\xi_c\rangle = \sum_{\{n\}} \xi_c(n_{(c-1)c_L+1}, ..., n_{cc_L})|n_{(c-1)c_L+1}, ..., n_{cc_L}\rangle$. For each cluster (c), we can write down the analogous V_i operator which contains all terms in Eq. S20 that involve sites in the cluster (V_c). All terms involving only sites inside the given cluster are treated exactly (i.e. treated as operators) while terms that involve sites that are inside two different clusters are split up via the MF approximation $\hat{A}_i\hat{B}_j \sim \langle\hat{A}_i\rangle\hat{B}_j + \hat{A}_i\langle\hat{B}_j\rangle - \langle\hat{A}_i\rangle\langle\hat{B}_j\rangle$, where as before $\langle\hat{A}_i\rangle$ and $\langle\hat{B}_j\rangle$ are determined self-consistently.

More explicitly, the mean-field cluster V_c operator is

$$\begin{aligned} V_c = {} & \hat{L}(j = c_L(c-1)+1) + \hat{R}(j = c_L c - 1) \\ & + \sum_{j=c_L(c-1)+1}^{c_L c - 1} p e^{-\lambda}\hat{c}_{j+1}^\dagger \hat{c}_j - p\hat{n}_j(\mathbb{1} - \hat{n}_{j+1}) + q e^{\lambda}\hat{c}_j^\dagger \hat{c}_{j+1} - q\hat{n}_{j+1}(\mathbb{1} - \hat{n}_j) \end{aligned} \tag{S36}$$

where

$$\begin{aligned} \hat{L}(j) &= -p\langle\hat{n}_{j-1}\rangle(\mathbb{1} - \hat{n}_j) + q e^{\lambda}\langle\hat{c}_{j-1}^\dagger\rangle\hat{c}_j + p e^{-\lambda}\langle\hat{c}_{j-1}\rangle\hat{c}_j^\dagger - q\hat{n}_j(1 - \langle\hat{n}_{j-1}\rangle) && \text{for } c > 1 \\ \hat{L}(j) &= -\alpha(\mathbb{1} - \hat{n}_1) + \alpha e^{-\lambda}\hat{c}_1^\dagger + \gamma e^{\lambda}\hat{c}_1 - \gamma\hat{n}_1 && \text{for } c = 1 \\ \hat{R}(j) &= -q\langle\hat{n}_{j+1}\rangle(\mathbb{1} - \hat{n}_j) + p e^{-\lambda}\langle\hat{c}_{j+1}^\dagger\rangle\hat{c}_j + q e^{\lambda}\langle\hat{c}_{j+1}\rangle\hat{c}_j^\dagger - p\hat{n}_j(1 - \langle\hat{n}_{j+1}\rangle) && \text{for } c < L/c_L \\ \hat{R}(j) &= -\nu(\mathbb{1} - \hat{n}_L) + \nu e^{\lambda}\hat{c}_L^\dagger + \beta e^{-\lambda}\hat{c}_L - \beta\hat{n}_L && \text{for } c = L/c_L \end{aligned}$$

This treatment ensures that all cluster based eigenvalue problems can be solved separately using only expectation values to estimate the couplings between clusters. The latter couplings are calculated separately at the end of each self-consistent step. This is continued until the averages do not change. Once the MF calculations converge we will obtain the required GDF $\langle\tilde{\Xi}|$ needed to construct the generator of auxiliary dynamics, similar to the case for the single site MF. However, now the $\xi_c(n_{(c-1)c_L+1},...,n_{cc_L})$ involve multiple sites and thus the proposal matrix must be updated accordingly to distinguish between inter- and intra- cluster states. For instance, for a cluster of size $c_L = 2$ the intra-cluster hopping matrix for hops within a cluster is given by,

$$\tilde{u}_c = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-qdt & pe^{-\lambda}\frac{\xi_c(0,1)}{\xi_c(1,0)} & 0 \\ 0 & qe^{\lambda}\frac{\xi_c(1,0)}{\xi_c(0,1)} & 1-pdt & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{S37}$$

For inter-cluster hopping the two matrices $\tilde{u}_c$ and $\tilde{u}_{c+1}$ must be combined and the hopping between edge sites n_{cc_L} and $n_{(c)c_L+1}$ will depend on the state configuration of the two clusters. For e.g. the transition matrix elements given by $P(n_{cc_L-1}, n_{cc_L}; n_{cc_L+1}, n_{cc_L+2} \to n_{cc_L-1}, n_{cc_L}; n_{cc_L+1}, n_{cc_L+2})$ are:

$$\begin{aligned}
&P(i,0;0,j \to i,0;0,j) = 1 \\
&P(i,1;0,j \to i,0;1,j) = pe^{-\lambda}\frac{\xi_c(i,0)\xi_{c+1}(1,j)}{\xi_c(i,1)\xi_{c+1}(0,j)} \\
&P(i,1;0,j \to i,1;0,j) = 1-pdt \\
&P(i,0;1,j \to i,1;0,j) = qe^{\lambda}\frac{\xi_c(i,1)\xi_{c+1}(0,j)}{\xi_c(i,0)\xi_{c+1}(1,j)} \\
&P(i,0;1,j \to i,0;1,j) = 1-qdt \\
&P(i,1;1,j \to i,1;1,j) = 1
\end{aligned} \tag{S38}$$

This can be generalized to other cluster sizes.

2D Cluster Approximation

For the 2D WASEP problem that we consider in the main text, the cluster approach outlined above for 1D systems can be generalized to 2D. The tilted operator for this problem is given by,

$$\begin{aligned}
\mathbb{W}_{\lambda_x,\lambda_y} = \sum_{i=1}^{L_x}\sum_{j=1}^{L_y} & p_x e^{-\lambda_x}\hat{c}^\dagger_{i+1,j}\hat{c}_{i,j} - p_x\hat{n}_{i,j}(\mathbb{1}-\hat{n}_{i+1,j}) + q_x e^{\lambda_x}\hat{c}^\dagger_{i-1,j}\hat{c}_{i,j} - q_x\hat{n}_{i,j}(\mathbb{1}-\hat{n}_{i-1,j}) \\
& p_y e^{-\lambda_y}\hat{c}^\dagger_{i,j+1}\hat{c}_{i,j} - p_y\hat{n}_{i,j}(\mathbb{1}-\hat{n}_{i,j+1}) + q_y e^{\lambda_y}\hat{c}^\dagger_{i,j-1}\hat{c}_{i,j} - q_y\hat{n}_{i,j}(\mathbb{1}-\hat{n}_{i,j-1}),
\end{aligned} \tag{S39}$$

where p_x, q_x, p_y, q_y are transition rates associated with hopping along $\pm x, \pm y$ directions. Additionally for this model we use periodic boundary conditions. Here, (L_x, L_y) are the number of sites in the respective directions. Notice that the biasing parameter is a vector $\boldsymbol{\lambda} = (\lambda_x, \lambda_y)$. Additionally to accommodate the extra dimension, operators are now indexed via the coordinates (i, j). Much like the 1D case the MF solutions can be constructed as before except now the coupling terms are defined for the perimeter of the cluster at (c_x, c_y). Suppose there are (c^x_L, c^y_L) sites per cluster; then in explicit terms the cluster MF tilted operator V_{c_x,c_y} is given by:

$$\begin{aligned}
V_{c_x,c_y} = \sum_{\substack{i=c^x_L(c_x-1)+1 \\ j=c^y_L(c_y-1)+1}}^{c^x_L c_x-1, c^y_L c_y-1} & p_x e^{-\lambda_x}\hat{c}^\dagger_{i+1,j}\hat{c}_{i,j} - p_x\hat{n}_{i,j}(\mathbb{1}-\hat{n}_{i+1,j}) + q_x e^{\lambda_x}\hat{c}^\dagger_{i,j}\hat{c}_{i+1,j} - q_x\hat{n}_{i+1,j}(\mathbb{1}-\hat{n}_{i,j}) \\
& + p_y e^{-\lambda_y}\hat{c}^\dagger_{i,j+1}\hat{c}_{i,j} - p_y\hat{n}_{i,j}(\mathbb{1}-\hat{n}_{i,j+1}) + q_y e^{\lambda_y}\hat{c}^\dagger_{i,j}\hat{c}_{i,j+1} - q_y\hat{n}_{i,j+1}(\mathbb{1}-\hat{n}_{i,j}) \\
& + \hat{L}(i=(c_x-1)c^x_L+1, c_y) + \hat{R}(i=c_x c^x_L, c_y) + \hat{T}(c_x, j=c_y c^y_L) + \hat{B}(c_x, j=(c_y-1)c^y_L+1)
\end{aligned} \tag{S40}$$

where

$$\hat{L}(i,c_y) = \sum_{j=c^y_L(c_y-1)+1}^{c^y_L c_y-1} p_x e^{-\lambda_x}\hat{c}^\dagger_{i,j}\langle\hat{c}_{i-1,j}\rangle - p_x\langle\hat{n}_{i-1,j}\rangle(\mathbb{1}-\hat{n}_{i,j}) + q_x e^{\lambda_x}\langle\hat{c}^\dagger_{i-1,j}\rangle\hat{c}_{i,j} - q_x\hat{n}_{i,j}(\mathbb{1}-\langle\hat{n}_{i-1,j}\rangle)$$

$$\hat{R}(i,c_y) = \sum_{j=c_L^y(c_y-1)+1}^{c_L^y c_y - 1} p_x e^{-\lambda_x}\langle \hat{c}^\dagger_{i+1,j}\rangle \hat{c}_{i,j} - p_x \hat{n}_{i,j}(\mathbb{1} - \langle \hat{n}_{i+1,j}\rangle) + q_x e^{\lambda_x}\hat{c}^\dagger_{i,j}\langle \hat{c}_{i+1,j}\rangle - q_x\langle \hat{n}_{i+1,j}\rangle(\mathbb{1}-\hat{n}_{i,j})$$

$$\hat{B}(c_x,j) = \sum_{i=c_L^x(c_x-1)+1}^{c_L^x c_x - 1} p_y e^{-\lambda_y}\hat{c}^\dagger_{i,j}\langle \hat{c}_{i,j-1}\rangle - p_y\langle \hat{n}_{i,j-1}\rangle(\mathbb{1}-\hat{n}_{i,j}) + q_y e^{\lambda_y}\langle \hat{c}^\dagger_{i,j-1}\rangle \hat{c}_{i,j} - q_y \hat{n}_{i,j}(\mathbb{1}-\langle \hat{n}_{i,j-1}\rangle)$$

$$\hat{T}(c_x,j) = \sum_{i=c_L^x(c_x-1)+1}^{c_L^y c_x - 1} p_y e^{-\lambda_y}\langle \hat{c}^\dagger_{i,j+1}\rangle \hat{c}_{i,j} - p_y\hat{n}_{i,j}(\mathbb{1}-\langle \hat{n}_{i,j+1}\rangle) + q_y e^{\lambda_y}\hat{c}^\dagger_{i,j}\langle \hat{c}_{i,j+1}\rangle - q_y\langle \hat{n}_{i,j+1}\rangle(\mathbb{1}-\hat{n}_{i,j})$$

The operators $\hat{L}$, $\hat{R}$, $\hat{B}$ and $\hat{T}$ are the border terms for the cluster at (c_x, c_y). They contain the mean-field coupling parameters that are enclosed with $\langle ... \rangle$. These are determined self-consistently much as in the 1D case. Finally the determined cluster states can then be used to construct the transformed tilted operator from which moves can be proposed.

CALCULATION DETAILS FOR RESULTS IN THE MAIN TEXT

Driven brownian motion: All of the calculations on the driven brownian walker were accomplished with a second order stochastic Runge-Kutta integrator with a timestep of 0.01. Observation times of 20.0 were needed to converge the calculations and branching steps attempted every 0.05 unit of time.

1D SEP: For the 1D system consisting of $L = 8$ sites, clusters were made using $c_L = 1, 2, 4$ sites. All calculations were done using $N_\mathrm{w} = 2 \times 10^4$ walkers, with a time step $dt = 0.001$, observation time $t_N = 1$ with branching occurring at intervals of time $t_\mathrm{int} = 0.01$. In the main text, we have used calculations done with cluster size of 0 ("zeroth-order" MF) to indicate a sampling strategy where the exponential $e^{-\lambda\mathcal{O}}$ has been absorbed directly into the proposed Monte Carlo moves. This essentially means $\xi_i = 1$ for equations S33-S35. We note that this last way of sampling should always be used if no higher order MF solutions are available. For the 2D WASEP system we discuss next, the bare sampling strategy is to at least use a zeroth-order MF. Generating trajectories from the unbiased distribution, i.e., without directly incorporating the exponential in the proposed Monte Carlo moves makes it impossible to converge calculations for the range of λ we explore and the number of walkers we deploy.

2D WASEP: The 2D WASEP system discussed in the main text is a generalization of the 1D model to an $L \times L$ lattice with periodic boundary conditions. Along the principal direction, x, particles are subject to biased hopping rates that are scaled by the length of the system $p_x = e^{-E/L_x}/2$ along x and $q_x = e^{E/L_x}/2$ along $-x$ with $L_x = 12$ and $E = 10$. Along the y direction particles diffuse symmetrically, with rates $p_y = q_y = 0.5$ [4]. These

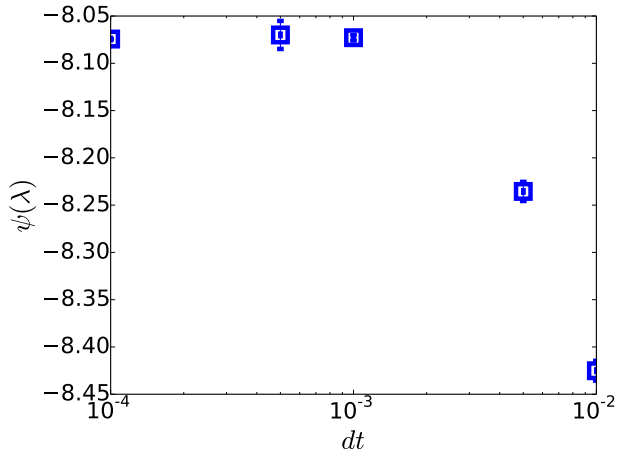

FIG. S1. Convergence of the 2D SEP DMC simulations with integration time step for $\lambda = -6$.

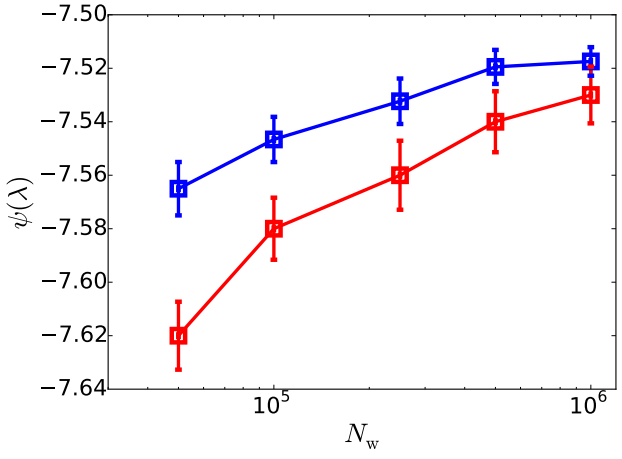

FIG. S2. Systematic error associated with walker population for 2D SEP DMC simulations for the large deviation function. Shown in red are the results for no importance sampling and in blue the results of the cluster GDF. Both were computed at $\lambda = -5.0$.

calculations are resource intensive so a careful study of convergence and statistical properties is needed to be able to compute the susceptibility $\chi(\lambda)$.

Towards this end we determine the time step error associated with Trotterization as shown in Fig. S1 for the largest λ value that we wish to access, since this sets the upper-bound on the error. We find that $dt = 0.001$ is sufficient to converge the error.

The second major source of error in these calculations is the systematic error due to finite walker population. Since the variance grows exponentially with $|\lambda|$ [5], it is sufficient to determine the largest number of walkers needed for maximal $|\lambda|$ we use in our simulations. Fig. S2 shows the convergence of $\psi(\lambda)$ with $N_{\rm w}$ for $\lambda = -5.0$ both using and not using importance sampling. It is evident that using auxiliary dynamics, even with $N_{\rm w} = 5 \times 10^5$ our results would have been sufficiently converged. Comparing against calculations without auxiliary dynamics we have $\sim$ 4-fold increase in efficiency mirroring the ratio of independent walkers of Fig. 3c in the main text. Despite the mean field GDF employed for this model not being particularly good at large $|\lambda|$, as the traveling wave state is not well described by a small single cluster, we still get a factor of 2-4 reduction in the required number of walkers to converge results (c.f. Fig. 3c in main text).

Following this analysis, we used $N_{\rm w} = 5 \times 10^5$ walkers for $\lambda \leq -3.0$ with $t_N = 100$ where branching was done every $t_{\rm int} = 0.01$. For $\lambda > -3.0$ simulations used $N_{\rm w} = 10^6$ walkers with $t_N = 72$ and $t_{\rm int} = 1.4$. In Fig. 3c of the main text the ratio of independent clones was determined using independent clone counts at time $t = 0.32 t_N$. Fig. S3 shows a comparison of the fraction of independent walkers, $f_{\rm I}(t)$, along the entire t_N trajectory for $\lambda = -5.0$. It highlights the importance of using GDF for sampling purposes to ensure that that there are enough uncorrelated contributions to the estimator.

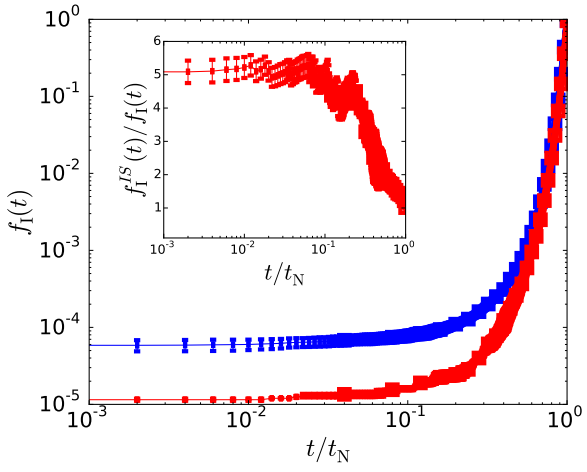

FIG. S3. Fraction of independent walkers $f_I(t)$ as a function of simulation time for $\lambda = -5.0$ using GDF (blue) and not using GDF (red). Inset shows the ratio of fraction of independent walkers with GDF and without as a function of time.

CONNECTION WITH THE ITERATIVE FEEDBACK METHOD OF NEMOTO ET AL

In their work, Nemoto et al. have outlined an iterative feedback control algorithm by which they attempt to construct a potential to affect an importance sampling of the large deviation function[6, 7]. These papers illustrate their method via two models: diffusion of a particle in a quartic potential and the FA model[8]. The basic idea in the context of a lattice model is to modify the transition rate $w(\mathcal{C} \to \mathcal{F}[\mathcal{C}])$ between configurations $\mathcal{C}$ and $\mathcal{F}[\mathcal{C}]$ via a auxilary potential $U(\mathcal{C})$ as

$$\tilde{w}(\mathcal{C} \to \mathcal{F}[\mathcal{C}]) = e^{-\lambda} w(\mathcal{C} \to \mathcal{F}[\mathcal{C}]) e^{\frac{1}{2}[U(\mathcal{C}) - U(\mathcal{F}[\mathcal{C}])]}. \quad \text{(S41)}$$

where they biased on the activity, or number of configurational changes. Here $\mathcal{F}[\mathcal{C}]$ is a function that maps the current configuration $\mathcal{C}$ to a new configuration and λ is a counting field conjugate to the activity. Typically Nemoto et al. parametrize $U(\mathcal{C})$ iteratively using the criteria that in the optimum Doob transformed dynamics the distribution of order parameter values computed over the full simulation time – the "average distribution" and those constructed from the final "end distribution" must be identical. Of the models they consider the more complicated many-particle-interacting simulations for the FA case is done by setting $U(\mathcal{C}) - U(\mathcal{F}_i[\mathcal{C}]) \equiv u_d(n_{i-d}, ..., n_i, ..., n_{i+d})$. Here $\mathcal{F}_i$ simply constitutes a spin-flip on site i and $\mathcal{C}$ represents a 1D lattice of spins.

Nemoto et al. alluded to the relationship between their potential and the dominant left eigenvector of the tilted operator, but did not give an explicit expression. From the formalism presented in the text it is evident that

$$e^{\frac{1}{2}[U(\mathcal{C}) - U(\mathcal{F}[\mathcal{C}])]} = \frac{\tilde{\Xi}(\mathcal{F}[\mathcal{C}])}{\tilde{\Xi}(\mathcal{C})}, \quad \text{(S42)}$$

where $\tilde{\Xi}(\mathcal{C})$ is the GDF obtained by approximating the left eigenvector of the tilted operator, as we have discussed in the main text. This connection simplifies the importance sampling in the sense that GDFs with larger overlap with the exact left eigenvector of the tilted operator immediately lead to better sampling. The iterative mechanism outlined by Nemoto et al. may be seen as an analog of the variation of parameters needed to determine a parametrized GDF. However, since the variational determination of the GDFs is carried out outside of the dynamics itself, we expect it to be numerically less intensive in most situations than the iterative feedback algorithm.

[1] J. Mehl, T. Speck, and U. Seifert, Physical Review E **78**, 011123 (2008).
[2] R. Chetrite and H. Touchette, in *Annales Henri Poincaré*, Vol. 16 (Springer, 2015) pp. 2005–2057.
[3] P. Tsobgni Nyawo and H. Touchette, Phys. Rev. E **94**, 032101 (2016).
[4] N. Tizón-Escamilla, C. Pérez-Espigares, P. L. Garrido, and P. I. Hurtado, Physical Review Letters **119**, 090602 (2017).
[5] U. Ray, G. K.-L. Chan, and D. T. Limmer, The Journal of Chemical Physics **148**, 124120 (2018).
[6] T. Nemoto, F. Bouchet, R. L. Jack, and V. Lecomte, Phys. Rev. E **93**, 062123 (2016).
[7] T. Nemoto, R. L. Jack, and V. Lecomte, Physical Review Letters **118**, 115702 (2017).
[8] G. H. Fredrickson and H. C. Andersen, Physical review letters **53**, 1244 (1984).