Reconstruction of complex bond potentials from multiple rupturing time distributions

Pak-Wing Fok\textsuperscript{a,1} and Tom Chou\textsuperscript{a,b}

\textsuperscript{a} Dept. of Biomathematics, UCLA, Los Angeles, CA 90095-1766
\textsuperscript{b} Dept. of Mathematics, UCLA, Los Angeles, CA 90095-1555

November 9, 2009

Abstract: We explore the numerical reconstruction of the potential energy profile of a molecular bond from distributions of bond rupture times. For a single measured first passage time (FPT) distribution the inverse problem is highly ill-conditioned and only a few attributes (such as the height and width of an energy barrier) can usually be reconstructed. Upon more careful analysis, we find optimal temperatures (described by the effective diffusivity of the stochastic process) and initial bond configurations that yield the most efficient numerical reconstruction of simple potentials. Moreover, we show that reconstruction of more complex potentials containing multiple minima can be achieved by simultaneously using two or more measured FPT distributions, obtained under different physical conditions. For example, by changing the effective potential energy surface by known amounts, the additional FPT distributions render the inverse problem less ill-posed. We demonstrate the feasibility of reconstructing potentials with multiple minima, motivate heuristic rules-of-thumb for optimizing the reconstruction, and discuss further applications and extensions.

1 Introduction

In many applications, one wishes to infer properties of a material or a process in an interior region of a sample not readily accessible to experimental probes. Examples of such inverse problems involving boundary data include radiological imaging, where radiation passing through tissues is detected outside the sample, electrical impedance tomography, where potentials are measured on the exterior of a body, and seismology, where reflected waves are measured at the earth’s surface.

In addition to boundary data in macroscopic, deterministic processes, many stochastic processes can also only be probed at the boundary. In these problems, one type of “boundary” data is a first passage time distribution (FPTD), describing the probability of a random variable first reaching a particular value within a certain time window. Here, the boundary data is the probability flux out of the domain. Figure 1(a) shows individual trajectories of a one-dimensional stochastic process and their corresponding first passage times. The FPTD is shown in Fig. 1(b) along with its Laplace transform in the inset. These types of first passage problems arise in many biophysical contexts.

\textsuperscript{1}Present Address: Dept. of Mathematical Sciences, University of Delaware, Newark, DE 19716-2553
For example, one realization of a stochastic first passage time problem is the voltage dynamics within neurons. The voltage across a nerve cell membrane fluctuates due to noisy inputs from other neurons, and can be described by a biased random walk determined by a constitutive voltage-current relationship intrinsic to the cell. When the fluctuating voltage exceeds a certain threshold, the potential spikes before resetting the stochastic process. The interspike times define the first passage times [1, 2] from which the neuron’s current-voltage relationship might be reconstructed.

Stochastic inverse problems are typically highly ill-posed: there may be several different interior structures that would yield nearly the same measured boundary data. Nonetheless, for many physical systems, reconstruction of constitutive relations from measured data can be cast in Sturm-Liouville form with an unknown spatially dependent coefficient [3, 4]. Given the eigenvalues of the problem and that the coefficient function is symmetric, the reconstruction is unique [5]. However, one eigenvalue spectrum is insufficient to determine a general (nonsymmetric) coefficient [5]. In certain cases, the boundary data directly yields the eigenvalues. For example, in the impulse response of a string with non-uniform density that is pinned at one end and free at the other, the eigenvalues can be found from the poles of the Fourier-transformed displacement of the free end [4]. Often, the spectrum of the Sturm-Liouville representation of the problem is not known or only partially known. For example, the diffusive nature of our stochastic problem renders it fundamentally different to the hyperbolic string problem in [4] and the FPT data no longer directly yields the eigenvalues. Therefore, many of the algorithms developed specifically for reconstruction through the eigenvalues [6, 7, 8] are of limited use. This motivates the development of new algorithms and techniques that deal directly with the boundary data.

Figure 1: (a) Three realizations of a representative random walk and their first passage times \( t_i \), \( i = 1, 2, 3 \). The random variable \( y(t) \) could represent the transmembrane voltage of a neuron or the bond coordinate of an unfolding macromolecule. (b) Histogram of the first passage times of a stochastic process started at position \( x = y(0) = 0, W(x = 0, t) \), obtained from 2000 realizations of the process shown in (a). The inset shows \( \tilde{W}(x, s) = \int_0^\infty W(x, t)e^{-st}dt \), which is used extensively in this paper. An arbitrary potential was used to generate the data.

In this paper, we investigate the physical and mathematical features of a stochastic inverse problem in the context of yet another system commonly encountered in biophysics: macromolecule unfolding and molecular adhesion bond rupturing. Typically, macromolecular bond displacements are modeled using a single bond coordinate, represented by a fluctuating Brownian “particle” in a one-dimensional energy landscape. The metastable bond is considered to be broken the instant the
bond coordinate reaches a critical extension. This problem is of great interest in single-molecule biophysics and has been recently studied by a number of researchers, particularly in the context of dynamic force spectroscopy (DFS) experiments [9]. In DFS, a pulling force protocol is applied to the bond and the force at the instant of rupture is recorded. The mean rupture force by itself would give very little information about the molecular potential [10] (many different potentials would yield the same mean rupture force). However, can one reconstruct the bond potential from the measured rupture force distribution? All of the recent theoretical treatments of this problem have either analyzed the forward problem [11], used physical approximations to derive simple force and time-dependent dissociation rates [12, 13], and/or considered simple 2-3 parameter single minimum potentials [14, 15, 16, 17]. In addition to the force at rupture, the rate of force increase as a function of displacement (the rupture stiffness) has also been incorporated into a procedure to fit basic parameters of simple potentials [18]. However, by imposing such simple two or three parameter forms for the reconstructed potential, one loses details such as multiple minima.

Here, we will approach the inverse problem more generally by allowing a wider class of potentials, including those with multiple minima. Within a class of potentials, we numerically determine the ones that best fit the entire measured FPTD. Although the ill-posedness of obtaining eigenvalues from FPTD data is avoided, our stochastic bond-breaking problem retains an intrinsic ill-posedness from its diffusive nature. Therefore, it is not surprising that almost all of the theoretical work to date has focused on reconstructing only two or three attributes of the stochastic process, typically, the energy barrier height and width. In the next Section, we formulate the problem mathematically through the backward equation of a Brownian process with an energy potential-derived drift. After decomposing the target drift function into basis functions, we develop an iterative optimization procedure to find the coefficients of the basis functions. Experimental protocols that best improve the convergence of the reconstruction are analyzed and discussed in the Results section. We also discuss different ways for obtaining multiple FPTDs, and how these strategies can facilitate the reconstruction. In the Summary and Conclusions, we discuss limitations and refinements.

2 Stochastic Theory

Consider a one-dimensional Brownian motion described by a diffusivity $D(x)$ and a convective drift $-D(x)(k_B T)^{-1}(d\Phi(x)/dx)$ derived from a time-independent molecular bond potential $\Phi(x)$. Although we restrict ourselves to time-independent potentials, corresponding to static forces, the rupture force distribution can be transformed into a first rupture time distribution (FPTD) in the quasi-adiabatic limit [16]. The continuous Brownian process can be described by the probability $P(y, t|x)dy$ for the bond coordinate to be between positions $y$ and $y + dy$ at time $t$, given that it started at position $x$ at initial time $t = 0$. This probability density obeys the backward equation [19]

$$\frac{\partial P(y, t|x)}{\partial t} + \frac{D(x)}{k_B T} \frac{d\Phi(x)}{dx} \left( \frac{\partial P(y, t|x)}{\partial x} \right) = D(x) \frac{\partial^2 P(y, t|x)}{\partial x^2}. \tag{1}$$

Since the bond is irreversibly ruptured when stretched to a known position $y = L$, we impose the absorbing boundary condition $P(y = L, t|x) = 0$. The bond survival probability at time $t$, given that it started initially at position $x$ is found from integrating the probability density over all the final coordinates that an unruptured bond can take, i.e., $S(x, t) = \int_0^L P(y, t|x)dy$. From $S(x, t)$, we
can define the FPTD $w(x,t) = -\partial_t S(x,t)$. Upon integrating Eq. (1) over allowable final positions $0 < y < L$, and differentiating with respect to the final time, we find an equation for the FPTD

$$\frac{\partial w(x,t)}{\partial t} + \frac{D(x) d\Phi(x)}{k_B T} \frac{\partial w(x,t)}{\partial x} = D(x) \frac{\partial^2 w(x,t)}{\partial x^2},$$

subject to initial condition $w(x,0) = 0$, and boundary conditions $\partial_x w(x,t)|_{x=0} = 0$ and $w(x = L, t) = \delta(t)$.

In the forward problem, $\Phi(x)$ and $D(x)$ are given and one solves Eq. 2 for $0 < x < L$ and $t > 0$ to find the function $w(x,t)$, as shown in Fig. 2(a). Each slice of the surface $w(x,t)$ (Fig. 2(b)) represents the FPTD for a particle that started the random walk at position $x$. In the inverse problem, the functions $\Phi(x)$ and $D(x)$ are unknown (except for possible prior information) and need to be determined from an experimentally measured or simulated FPTDs. In general, the unique pointwise reconstruction of both $D(x)$ and $\Phi(x)$ from FPT data is impossible [20]. At best, only half of either $D(x)$ or $\Phi(x)$ can be uniquely determined from a single FPTD [21, 20]. It has been shown that if we assume $D$ is a constant, then $\Phi(x)$ is uniquely identifiable from a single FPTD provided it is known over a certain interval within $(0,1]$ [20]. In the approach presented here, we do not know $\Phi(x)$ over the requisite interval, so we shall express $\Phi(x)$ in terms of basis functions to make the reconstruction of the expansion coefficients computationally feasible. Henceforth, we nondimensionalize the problem by measuring distance in units of $L$, time in units of $L^2/D$, and the potential $\Phi(x)$ in units of the thermal energy $k_BT$. Finally, to avoid numerically representing the $\delta$-function in the boundary condition $w(L,t) = \delta(t)$, we work with the Laplace transform $\tilde{w}(x,s) = \int_0^\infty w(x,t) e^{-st} dt$, which obeys the infinite set (for each $s \in \mathbb{R}_{\geq 0}$) of uncoupled ODEs

$$\frac{\partial^2 \tilde{w}(x,s)}{\partial x^2} + U(x) \frac{\partial \tilde{w}(x,s)}{\partial x} = s\tilde{w}(x,s),$$

subject to the Laplace transformed boundary conditions

$$\frac{\partial \tilde{w}(x,s)}{\partial x} \bigg|_{x=0} = 0, \quad \tilde{w}(1,s) = 1.$$
In Eq. 3, \( U(x) \equiv -\left(\frac{d\Phi(x)}{dx}\right) \) is the dimensionless convection which we wish to reconstruct. The molecular potential can be found, modulo an irrelevant constant, from integrating \( U(x) \). The first boundary condition of Eqs. 4 represent a reflecting wall at \( x = 0 \) and ensures a nonnegative bond coordinate. Ultimately, our numerical optimization suffers little in the Laplace domain; nonetheless, the diffusive nature of our stochastic inverse problem renders it fundamentally different from the hyperbolic string problem [4] mentioned in the Introduction. Since Eq. 2 is parabolic, it cannot support traveling waves and the eigenvalues of Eq. 3 are negative. In contrast to the string problem [4], the boundary data \( \tilde{w}(x, s \geq 0) \) does not yield the eigenvalues, nor can we extract the normalizing constants that, together with the eigenvalues, guarantee the unique reconstruction of \( U(x) \).

Equation 3 is a differential equation in the initial bond position for the Laplace-transformed rupture time distribution \( \tilde{w}(x, s) \). Mathematically, our objective is to find the function \( \Phi(x) \), within a reduced representation, that renders, for a chosen value of \( x \) and all \( s \in \mathbb{R}_{\geq 0} \), \( \tilde{w}(x, s) \) as close as possible to the measured or simulated, Laplace-transformed FPTD \( \tilde{W}(x, s) \).

3 Algorithm to find \( U(x) \) by Newton Iteration

Our method for reconstructing \( U(x) \) consists of using spectral methods [22] to repeatedly solving the forward problem Eqs. 3-4 to refine our estimate for the potential. Since it is known that approximating a function using a basis of monomials leads to a very ill-conditioned problem [23], we represent the convection in terms of orthonormal polynomials:

\[
U(x) = \sum_{i=0}^{n} a_i u_i(x),
\]

where \( \{a_i\} \equiv a \) are expansion coefficients and the first few orthonormal basis functions are

\[
u_0(x) = 1, \quad u_1(x) = \sqrt{3}(1 - 2x), \quad u_2(x) = \sqrt{5}(1 - 6x + 6x^2), \ldots
\]

Potentials and drift functions that diverge at \( x = 0 \) lead to highly singular differential equations, but remain easily solved using spectral methods [22]. Note that the coefficients \( a_i \) encode both the shape and relative magnitude of the potential and associated drift. Since the magnitude of the potential sets the scale of drift relative to diffusion, we can define an effective Peclet number for our problem

\[
\alpha = \sqrt{\frac{N}{\sum_{i=1}^{N} a_i^2}}.
\]

While the relative magnitudes of the \( a_i \) determine the shape of the potential \( \Phi(x) \) and drift \( U(x) \), the Peclet number \( \alpha \) determines the scale of the potential. Experimentally, the shapes of potentials are fixed by molecular details; however, the Peclet number \( \alpha \) is inversely proportional to temperature and can in principle be tuned experimentally.

Starting with an initial guess for \( U(x) \) (say, \( U(x) = 0 \)), we use spectral methods to solve Eqs. 3-4 for many positive values of \( s \) to obtain \( \tilde{w}(x, s; a) \). We then compute the “distance” between the \( \tilde{w}(x, s; a) \) and the given data \( \tilde{W}(x, s) \) using the objective function
\[ \Pi(a) = \int_0^\infty A(s)|\tilde{w}(x, s; a) - \tilde{W}(x, s)|^2 ds, \]

where \( A(s) \) is a weighting function chosen so that the integral remains convergent.\(^1\) By adjusting \( U(x) \) through the coefficients \( a \), we seek to decrease \( \Pi(a) \). The adjustments in \( a \) are repeated until \( \Pi \) is minimized. Although many different algorithms can be used to minimize \( \Pi \), we first consider \( A(s) = 1 \) and choose a safe-guarded Newton strategy that relies on computing the Hessian of \( \Pi \), and describe details in the Supporting Information (SI).

## 4 Results and Discussion

We test our algorithm and discuss salient aspects of reconstructing the drift function \( U(x) \) from (i) a single, perfectly measured distribution of rupturing times, (ii) multiple perfectly measured distributions of rupturing times, realized under different experimental conditions. We first generate “data” by solving the forward problem using a hypothetical target potential function \( \Phi^*(x) \) (and corresponding \( U^*(x) = \sum_{i=0}^n a_i^* u_i(x) \)). This initializing solution will serve as data \( \tilde{W}(x, s) \) from which we wish to reconstruct \( U^*(x) \) by comparing with \( \tilde{w}(x, s; a) \) and minimizing \( \Pi(a) \). We are interested in if and how \( a_i \) approaches \( a_i^* \) during the iteration process and the number of coefficients \( a_i \) that can be reliably reconstructed. In addition to analyzing perfect (up to numerical precision) data, we also show in the Supporting Information (SI) that reasonable reconstruction is possible from two or more noisy rupture time distribution data derived from Monte Carlo simulations. Using a single FPTD data set, we find that reconstruction of \( \Phi^*(x) \) is badly conditioned, but using two sets of data allows us to find up to 6 coefficients of \( \Phi^*(x) \) in many cases.

### 4.1 Single measurement

We first assume a target potential \( \Phi^*(x) = \left( \frac{1}{3} + \frac{\sqrt{3}}{4} + \frac{\sqrt{505}}{12} \right) x - \left( \frac{\sqrt{3}}{4} + \frac{\sqrt{505}}{4} \right) x^2 + \frac{\sqrt{505}}{6} x^3 \) corresponding to a simple three-parameter potential well. The associated target drift function \( U^*(x) \) is described by \( (a_0^*, a_1^*, a_2^*) = (1/5, 9/10, \sqrt{3}/20) \), which correspond to a Peclet number \( \alpha = 1 \). Figure 3(a) shows that starting with the initial guess \( U(x) = 0 \), minimizing the objective function Eq. 8 leads to accurate convergence to the target drift \( U^*(x) \) occurs within about 10 iterations. We find that a five-parameter potential is typically a marginal case in that it can only be occasionally reconstructed, and only after a large number of iterations as shown in Fig. 3(b). However, we are typically not able to accurately reconstruct a potential with six parameters (see Fig. 3(c)), regardless of the number of iterations.

When \( U^*(x) \) is more structurally complex, extremely slow or nonconvergence to \( a^* \) arises because the curvature of \( \Pi \) near the true minimum, in at least one direction, becomes extremely small. For \( n \) coefficients \( a_i^* \), the \( n \) eigenvalues \( \lambda \) of the Hessian matrix \( H_{ij} = \partial_{a_i} \partial_{a_j} \Pi|_{a=a^*} \) quantify the curvatures along the eigendirections. In addition to increasing the number of eigendirections, increasing \( n \) rapidly decreases the minimum eigenvalue \( \lambda_{\text{min}} \). These properties render the problem

\(^1\)If we had chosen to work in the time domain, a reasonable objective function that measures the difference between measured and computed FPTDs would be \( \Pi(a) = \int_0^\infty A(t)|w(x, t; a) - W(x, t)|^2 dt \).
Figure 3: Reconstruction of three, five and six parameter potentials where the Peclet number $\alpha = 1$. In row (a) the drift $U^*(x)$ is given by the coefficients $(a_0^*, a_1^*, a_2^*) = \left(\frac{1}{5}, \frac{9}{10}, \frac{\sqrt{3}}{20}\right)$ which defines three parameter potential $\Phi^*(x) = -\left(\frac{1+\sqrt{3}}{5}\right)x + \frac{12\sqrt{3}}{5}x^2 - \sqrt{3}x^3$. The reconstructed potential $\Phi(x)$ (circles) is in excellent agreement with the original function $\Phi^*(x)$ (solid curve). The starting position was $x = 0.433$. The algorithm described in Section 3 quickly converges to the drift coefficients after about 10 iterations. In row (b) we attempt to reconstruct the five parameter double-well potential $\Phi^*(x) \approx -4.5696x + 21.7325x^2 - 43.1192x^3 + 37.6357x^4 - 11.8794x^5$ obtained from $a^* = \left(\frac{1}{5}, \frac{3}{5}, \frac{2}{5}, \frac{3}{5}, \frac{\sqrt{2}}{5}\right)$. The Newton algorithm barely converges to the correct drift coefficients, and only after many iterations. The accurate reconstruction of double-well potentials is usually not possible with one FPTD. In row (c), we show that our algorithm fails to reconstruct a 6-parameter, triple-well potential when only one FPTD is available. The coefficients $a_i$ converge to the wrong values and one of the local minima is not resolved. In all plots, the coefficient values $a_0, a_1, a_2, a_3, a_4, a_5$ at each iteration are indicated by open circles, filled circles, open triangles, filled triangles, open squares, and filled squares, respectively.
Figure 4: The objective function $\Pi(a_0, a_1, a_2^*) \equiv \int_0^\infty ||w(x = 0.3, s; a_0, a_1, a_2^*) - W(x = 0.3, s; a_i^*)||^2 ds$ for the potential shown in Fig 3(a) as function of $a_0$ and $a_1$. The scale factors are (a) $\alpha = 1$, and (b) $\alpha = 1.5$. Projected onto $a_0 - a_1$ space, $\Pi$ exhibits a much smaller curvature in one direction compared to the orthogonal direction. Curvatures in both directions at the minimum are larger for larger $\alpha = 1.5$ as shown in (b). Although the smallest curvature is maximized at intermediate $\alpha = O(1)$, we find that the condition number $\kappa$ is decreases as $\alpha \to 0$, as shown in (c).

badly conditioned and is the underlying mathematical reason for the difficulty of extracting more than three parameters from a given potential landscape.

To explicitly illustrate the ill-conditioning of the problem, we plot in Fig. 4(a-b) the objective function $\Pi(a_0, a_1, a_2^*)$ as a function of the two parameter directions $a_0$ and $a_1$. Although $(a_1^*, a_2^*)$ is the global minimum, it is clear that the curvature along one direction is extremely small, making the minimum difficult to find numerically.

Fig. 4 suggests that for a single FPTD measurement, one can optimize the reconstruction by tuning the potential strength $\alpha$ by e.g., performing the experiment at an optimal temperature. As shown in Fig. 4(c) the smallest eigenvalue $\lambda_{\text{min}}$ of the Hessian matrix has a maximum at $\alpha \sim O(1)$, when the potential wells are on the order of the thermal energy. This can be understood by considering the small and large $\alpha$ limits. When $\alpha \to 0$, the thermal energy dominates and we expect that the FPTD will very nearly resemble that of pure diffusion. In the opposite limit of $\alpha \to \infty$, the system is purely deterministic and the FPTD is proportional to $\delta(t - t^*)$ where $t^*$ is finite only if $\Phi^*(x)$ is monotonically decreasing as $x$ increases. If only one FPTD can be measured, $\lambda_{\text{min}}$ may be maximized with respect to the effective Peclet number $\alpha$. Therefore, we might expect that an intermediate drift magnitude balancing Brownian motion would optimize the reconstruction.

However, the ease of reconstruction in our Newton algorithm relies on inversion of the Hessian matrix, and hence, its condition number. For reliable reconstruction, the experimental aim is to minimize the condition number $\kappa$ of the Hessian matrix $H_{ij}$. While $\lambda_{\text{min}}$ has an expected maximum at $\alpha = O(1)$, the maximum eigenvalue also increases with $\alpha$ such that $\kappa^{-1}$ is monotonically decreasing with increasing $\alpha$ (cf. Fig. 5). Therefore, from a computational point of view, the conditioning of the reconstruction problem is optimized as $\alpha \to 0$, or when the drift is negligible. This limit corresponds to a vanishingly small drift perturbing the purely diffusive problem. Although the magnitudes of $a_i$ are vanishingly small, their incremental effect on reducing the condition number $\kappa$ is nonetheless greatest in this limit. This optimal limit arises from computational considerations and is not a predicted on physical grounds. However, system and experimental constraints will preclude measurement of effective potentials at extremely high temperatures, suggesting that an optimal, intermediate temperature may still arise in practice.
Figure 5: The inverse condition number as a function of both starting position $x$ and potential strength $\alpha$ for three different potential shapes. (a) $\Phi^*(x) = -(1/2)x^2$ (b) $\Phi^*(x) = -\frac{2\sqrt{3}}{7}x + \frac{2\sqrt{3}}{7}x^2$ (c) $\Phi^*(x) = -(\frac{1+7\sqrt{3}}{5})x + \frac{12\sqrt{3}}{5}x^2 - \sqrt{3}x^3$. The ratio $\kappa^{-1} = \lambda_{\min}/\lambda_{\max}$ is typically maximal as $\alpha \to 0$ and at $x \sim 0.8$.

The behavior of $\kappa^{-1}$ as a function of starting position is more intriguing. For all potentials we tested, the optimal value of the starting position occurs near $x = 0.8$. This starting position is close to the rupture point at $x = 1$ and is somewhat insensitive to $\alpha$, except for very large $\alpha$ where the optimal starting position decreases. This is expected since for large drifts toward the rupture point, starting too near the rupture point prevents adequate sampling of the small $x$ region for its reconstruction. Nevertheless, for the three qualitatively different potentials used in Fig. 5, $x \sim 0.6 - 0.85$ for a wide range of values $\alpha$. This optimal starting position near $x = 0.8$ is a mathematical property of the problem of minimizing the condition number of the Hessian matrix associated with our objective function $\Pi$ when the weighting $A(s)$ is constant. For nonconstant $A(s)$, as would be expected when experimental measurements of the moments of the first passage times are weighted differently, the optimal starting position $x$ can change (cf. SI).

### 4.2 Multiple measurements

Since a single measurement $\tilde{W}(x, s)$ is typically insufficient to reconstruct the potential well beyond three coefficients, even after optimizations with respect to $\alpha$ and $x$, we consider how additional data can be used to refine the reconstruction of $U^*(x)$. As indirectly suggested by the analysis of varying $\alpha$ and $x$, the potential can be perturbed by specified amounts to yield modified potentials that can provide a different FPTD from the unperturbed potential. By imposing any number of perturbations, multiple FPTD data $\tilde{W}$ can be measured aiding in the reconstruction of the unperturbed potential.

We propose three protocols for modifying the potential to be reconstructed. Experimentally, these protocols correspond to changing the system temperature, removing an existing external force immediately at the start of the stochastic process, and adding an applied force at the start of the stochastic process. Mathematically, these perturbations correspond to specific changes in $\alpha$, $x$, and the form of the potential $\Phi^*$, respectively. The multiple FPT data for all conditions can then be combined into a multi-distribution objective function. We summarize the protocols below:
• **Changing system temperature** - The first way to obtain additional data is by changing the scale \( \alpha \) of the potential. This can be achieved for example by changing the system temperature. At each different value of \( \alpha \), a separate FPTD can be measured. For example, if \( \alpha \) is decreased, we expect the FPTD will be shifted to shorter times if the target potential is attractive and has at least one local minimum.

• **Tuning starting positions** - By adding a force to the system before the start of the process, one can adjust the initial position \( x \) of the bond. At \( t = 0 \), this force is released, and the stochastic process proceeds under the original target drift \( U^*(x) \), provided the potential relaxes quickly to \( \Phi^*(x) \). Stochastic bond dynamics starting at different positions \( x \) yield different measured FPT distributions.

• **Adding Probe Forces** - Finally, one can add known potentials to the original target potential immediately after the start of the stochastic process to obtain additional FPT distribution data. Here, \( \Phi^*(x) \rightarrow \Phi^*(x) + \Delta \Phi(x) \), where \( \Delta \Phi(x) \) is known. The associated drift then changes according to \( U^*(x) \rightarrow U^*(x) + \Delta U(x) \), where \( \Delta U(x) \) is implemented through a change in the expansion coefficients \( \Delta a \) and represents an externally applied force imposed by e.g., a pulling device such as an AFM tip or an optical tweezer. The associated external potential in such cases may be of the form \( \Delta \Phi(x) = -F_{\text{ext}} x - K x^2 / 2 \), where \( F_{\text{ext}} \) is the externally applied time-independent force and \( K \) is the elastic response of the pulling device. The added drift \( \Delta U(x) = F_{\text{ext}} + K x \) changes the potential (in this case, though a known change in the coefficients \( \Delta a_0 = F_{\text{ext}} \) and \( \Delta a_1 = K \)), in which the dynamics occur, and gives rise to a different measured FPTD.

An objective function that incorporates all \( M \) FPT distributions measured under different conditions described above can be defined as

\[
\Pi_M(a) = \sum_{m=1}^{M} \int_{0}^{\infty} A(s) \left[ \tilde{w}(x_m, s; a + \Delta a_m, \alpha_m) - \tilde{W}_m(a^*, s) \right]^2 ds, 
\]

(9)

where \( x_m, \Delta a_m, \) and \( \alpha_m \) denote the starting position, additional pulling force, and temperature of the \( m \)th measurement. At least one of the parameters \( x_m, \Delta a_m, \alpha_m \) should be different for each measurement \( m \). More data provides additional constraints, increasing the curvature of the objective function near \( a^* \). In general, the lowest eigenvalues \( \lambda \) of the Hessian matrix associated with the multi-dataset objective function \( \Pi_M \) increases with \( M \) (cf. SI).

To illustrate how additional data can improve potential reconstruction, we compare how including two FPT distributions \( (M = 2) \) in our objective function \( \Pi \) affects reconstruction relative to using each FPTD separately. The two distributions will arise from two ideal measurements taken at two different conditions within each of the proposed experimental protocols described above. In Figure 6(a), we attempt to reconstruct the five-parameter, double-well potential \( \Phi^*(x) = -28 x + \frac{1415}{10} x^2 - 307 x^3 + 290 x^4 - 100 x^5 \) using two “temperatures” corresponding to \( \alpha = 6.11755 \ldots \) and \( \alpha = 1.22351 \ldots \). Each single FPTD reconstruction fails, as does using both FPTD data sets. In Fig. 6(b), we see that while starting from either initial position \( x = 0.182 \) or \( x = 0.727 \) fails to reconstruct the potential, using both initial positions allows us to accurately determine \( \Phi^*(x) \). Similarly, adding the perturbing potential \( \Delta \Phi = -x^2/2 \) provides another FPTD which allows accurate reconstruction of a double-well potential (Fig. 6(c)).
Figure 6: (a) Single measurements with $\alpha = 6.11755\ldots$ (blue dotted) and $\alpha = 1.22351\ldots$ (red dashed) both fail to give the correct 5-parameter potential $\Phi^*(x) = -28x + \frac{1451}{16}x^2 - 307x^3 + 290x^4 - 100x^5$ (solid black). This potential arises from the parameters $a^* = \{-1/1011\sqrt{3}/5, 89/(14\sqrt{5}), 8/\sqrt{7}, 50/21\}$ and has an intrinsic $\alpha = 6.11755\ldots$. However, even fitting to two data sets (corresponding to both $\alpha = 6.11755\ldots$ and $\alpha = 1.22351\ldots$) simultaneously (circles) fails to yield the right potential. The starting position was $x = 0.170$. (b) Single measurements with starting position $x = 0.182$ (blue dotted) and $x = 0.727$ (red dashed) yield inaccurate potentials. However, fitting to both sets of data (circles) gives an excellent approximation to the original 5-parameter, double-well potential defined by $a = \{2/3, 2, 5/3, 2, 4/3\}/\sqrt{13}$, corresponding to $\Phi^* \approx -4.7563x + 23.9612x^2 - 50.0252x^3 + 46.21670x^4 - 15.5316x^5$ and $\alpha = 1$ (solid black). (c) Single measurements with no probe forces yields the incorrect potential (blue dotted). Adding the probe force $\Delta \Phi = -x^2/2$ produces a second set of data and the total potential gives the reconstruction depicted by the red dashed curve. Upon using data sets derived from both $\Delta \Phi = 0$ and $\Delta \Phi = -x^2/2$, the reconstruction (circles) yields the correct original double-well potential defined by $a = \{1/5, 3/5, 2/5, 1/2, \sqrt{19/100}\}$, corresponding to $\alpha = 1$ and $\Phi^*(x) = -4.7642x + 24.7365x^2 - 54.2477x^3 + 52.3828x^4 - 18.3074x^5$ (solid black). The starting position here was $x = 0.642$. 

11
5 Summary and Conclusions

We have developed a procedure for reconstructing the drift of a stochastic process from first passage time distributions. For single distributions, only very coarse attributes (approximately three parameters) can be reconstructed. We demonstrate how to optimize the efficiency of the reconstruction through judicious choice of the potential strength through the effective inverse temperature $\alpha$ and starting position $x$ of the stochastic process. If only one FPTD can be measured, our analysis suggests that $\alpha \rightarrow 0$ and $x \sim 0.8$ are the most likely parameters to give the best chance for reconstructing relatively simple potentials. This estimate is based purely on mathematical analysis and assumes no physical constraints. However, our computations were performed assuming perfect data, uniform diffusivity, and precisely defined starting positions $x$. Moreover, we computed the inverse condition numbers $\kappa^{-1}$, shown in Figs. 4(c) and 5, using a constant weighting function $A(s) = 1$. In practice, finite time resolution, noisy data, and other experimental limitations might weight the measured FPTDs differently in time and one should use an objective function with the appropriate $A(s)$ (or $A(t)$ if $\Pi$ were a functional of $w(x, t)$ and the data $W(x, t)$). We show in the SI that the optimal parameters $\alpha$ and $x$ change when hypothetical physical constraints manifest themselves through a nonconstant $A(s)$. While, the mathematically-derived optimal values of $\alpha \rightarrow 0$ and $x \sim 0.8$ may be different from optimal values in practice, these parameter values might provide an experimental starting point.

Even more powerful conditioning can be achieved by imposing additional constraints provided by multiple FPTD data. We propose that additional measurements can be used to provide better conditioning of the problem, allowing finer details of the drift function to be extracted. These additional FPT distributions can be experimentally obtained by changing the temperature of the system or judiciously adding forces before or after the stochastic process is started. The total objective function including the constraints from all $M$ measurements has a sharper minimum, increasing the efficiency of standard optimization algorithms.

We proposed three ways of performing measurements under different experimental conditions: tuning the effective Peclet number $\alpha$ through the system temperature, adjusting the starting position $x$ via an initially applied force, and adding a known “probe” potential $\Delta \Phi$. This later potential can be realized in a number of ways, from directly mechanically pulling on the bond, to using mutagenesis to systematically change local properties of the bond energy profile. Such mutant bonds may provide additional FPTD data facilitating reconstruction of the original “wild-type” potential.

A number of refinements remain to be investigated, and some are discussed in the SI. For example, rather than Laplace-transforming the data, one can directly fit to $W(x, t)$. Although this approach is computationally more expensive, it would allow us to treat time-dependent potentials $U(x, t)$, and directly analyze dynamic force spectroscopy experiments [9, 11, 18], or scenarios in which the temperature is changed in a time-dependent way [24]. Other technical refinements and extensions should also be explored, including employing better basis functions for $U(x)$, using more sophisticated optimization methods, exploring the reconstruction efficiency for large $M$, defining the experimentally-imposed weighting function $A(s)$, generalization to discrete Markov processes, reconstruction of the diffusivity $D(x)$ itself, and systematically exploring the effects of noise in the data. Here, a Bayesian approach to estimate likelihood functions for the coefficients $a^*$ may be required [25].
The authors thank A. Furhmann and S. Getfert for discussions. This work was supported by NSF grant DMS-0349195 and NIH grant K25 AI058672

References


