

# Renewal Reward Process Formulation of Motor Protein Dynamics

Arjun Krishnan  
Bogdan I. Epureanu  
Department of Mechanical Engineering  
University of Michigan  
Ann Arbor, Michigan 48109-2125  
Email: arjunkc@umich.edu, epureanu@umich.edu

September 4, 2009

## Abstract

Renewal-reward processes are used to provide a framework for the mathematical description of single-molecule bead-motor assays of motor proteins. The main advantage of using such a formulation is that it makes available a set of formulae for the slopes of the asymptotes to the cumulants of the bead's displacement. The formulae relate the chemical reaction rates in the enzymatic cycle and the discrete mechanical steps of the motor to the slopes of the cumulants. The cumulant's slopes are commonly measured in bead-motor assays, the simplest examples being the bead's steady-state velocity and variance. To establish the usefulness of these cumulants and their robustness under experimental conditions, two other results are shown as simple consequences of the renewal-reward formulation. Namely, that mechanical substeps before the completion of a full-step either forwards or backwards, and initial state of the enzyme do not have any effect on the long-time slopes of the cumulants. To illustrate the uses of the formulation and the insight that it bestows, the Elston model and the Peskin-Oster model for kinesin are discussed in some detail. Then, the shortcomings of the randomness parameter - a function of the cumulants - when there is a possibility of forward and backward steps is discussed. To illustrate the use of the cumulant formulae, an alternative approach that gives more information than the randomness parameter is discussed about the motor's mechano-chemical cycle. To further encourage the use of this approach in bead-motor assays, the robustness of the approach under experimental error is also tested numerically and shown to be highly satisfactory.

# 1 Introduction

Motor proteins are enzymes that use energy from ATP hydrolysis to produce mechanical work. Different classes of motor proteins are involved in a variety of biological processes that range from cellular transport to DNA transcription. Our focus here is on processive enzymes; i.e., those in which ATP hydrolysis is accompanied by a mechanical displacement. Examples of such enzymes include kinesin, myosin and the rotary molecular motor F1-ATPase. Note that the actual physiological function of the enzyme is beyond the scope of this paper.

Motors like kinesin, myosin or RNA-polymerase are usually described as *linear*, in that they typically move along linear 'tracks' made of tubulin dimers, actin filaments or DNA strands (Howard and Clark, 2002) respectively. These tracks consist of a set of regularly or periodically spaced chemical binding sites. The motors walk by moving from one binding site to another in a preferred along these tracks. Once ATP binds to the motor protein, hydrolysis takes place in a sequence of chemical and mechanical processes (Cross, 2004). In one of the processes, the motor protein biases its diffusion in one direction by undergoing a conformational change (Rice et al., 1999). With the help of thermal fluctuations in the fluid medium, the motor completes a step by chemically binding to the next binding site on its track. Rotary motor proteins have similar mechanisms, but hydrolysis is accompanied by rotation of some radial appendage by a certain angle. For example, the rotation of F1-ATPase is composed of three separate rotations - as described by the binding change mechanism (Boyer, 1997) - of 120 degrees, each accompanied by the hydrolysis of one ATP molecule.

Single-molecule experiments on motor proteins focus on several different aspects of the chemical and mechanical phenomena. Some experiments attempt to determine chemical details of the enzymatic cycle such as rates, intermediate reactions, etc. (Rosenfeld et al., 2003). Other studies focus on structural aspects such as the different conformations of the protein and how these help it function (Rice et al., 1999). The experiments most relevant to the work herein are called bead-motor assays. These focus on the dynamic aspects of the motor protein's motion. In such an assay, a sphere of micrometer size is coupled to the motor in an ATP containing fluid medium. A force is applied to the bead using optical or magnetic tweezers (Gilbert et al., 1995; Svoboda et al., 1993) and this force, in turn, is transmitted to the motor. Such a force of several piconewtons is intended to simulate the effect of a cargo on the motor. The bead's position as the motor walks is optically measured. A great deal of insight into the enzyme's chemical cycle has been obtained through the analysis of the statistics of the bead's displacement (Block, 2003; Gilbert et al., 1995; Svoboda et al., 1993; Vale et al., 1996; Visscher et al., 1999).

Motor proteins have lengths of the order of several hundred nanometers and operate in the water-like cytoplasm. Due to the small size of the motor, the flow around it has a very low Reynolds number. Hence, the fluid's inertia may be ignored in the Navier Stokes' equations, and the familiar Stokes law for the viscous forces on the motor may be written in terms of a coefficient of friction  $\gamma$ . This coefficient of friction in Newton's law for the motor modeled as a point mass  $m$  corresponds to a (fast) time-scale of  $m/\gamma$ . In a bead-motor assay, this time-scale is of the order of fractions of a microsecond even for the micrometer sized beads. Hence, the effect of the inertia on the dynamics of the bead-motor system is negligible (Krishnan, 2008; Purcell, 1977). Such dynamics are generally called *overdamped*.

Their small size also makes them particularly susceptible to thermal noise; i.e., the influence of collisions with the fluid molecules. To put the magnitude of these fluctuations into perspective, consider the fact that kinesin, for example, consumes about 10 – 100 molecules of ATP per second (Astumian and Haenggi, 2002). ATP releases about 30 kJ of useful energy per hydrolysis - the reaction's Gibb's free energy - and this corresponds to an input power of about  $30 \cdot 10^6 / (6.023 \cdot 10^{23}) \approx 10^{-16}$  W. In contrast, the order of magnitude of thermal energy is  $k_B T$ , and the relaxation time of the collisions is about  $10^{-13}$  seconds, which gives a thermal power of  $10^{-8}$  W. Thus, the thermal fluctuations are nearly 10 times larger than the energy available to drive the motion. Thus, it is commonly accepted that thermal fluctuations play a significant role in the functioning of these motor proteins, much like Feynman's Brownian ratchets.

The motion of motor proteins is modeled on many length and time scales. Continuous, diffusion approximations with no internal details of the motor's chemistry or mechanics have been very useful in understanding

the origin of processivity or directed motion in a random thermal environment (Astumian and Haenggi, 2002; Bier, 1997). These are referred to as thermal-ratchet models. Models with more detailed chemistry model the enzyme's cycle as a Markov process through the chemical states (Qian and L. Elson, 2002) and often use Elston's method (Elston, 2000) to analyze the motor's walk superimposed over the chemistry (Fisher and Kolomeisky, 2001). Some models incorporate more detailed information from experimental work on the internal structure of the motor, and combine the diffusive motion and chemical cycle of the enzyme; the Peskin-Oster model for kinesin is a particularly elegant example (Peskin and Oster, 1995). While thermal-ratchet models have been successful in illustrating the general physical principles involved in this diffusion driven transport phenomenon, models that incorporate enzyme chemistry and protein physical structure have proved more useful in understanding experimental data.

A common assumption in these more fine-grained models is that diffusion is nearly instantaneous compared to the time required for ATP hydrolysis on the time-scale of the experiment. For example, the kinesin's chemical turnover time is of the order of hundreds of microseconds even when moving at its fastest velocity (Visscher et al., 1999). Estimates of the diffusion time-scale for even the large(r) micrometer sized bead using the low Reynold's number Stokes Law approximation, is at least three orders of magnitude smaller at about  $0.1 \mu s$ . Hence this assumption appears to be easily justifiable.

Under this separation of diffusion and chemical time-scales, the problem gains regenerative structure in a stochastic sense. Each time the enzymatic cycle completes, the motor returns to its original state (barring a step forwards or backwards); i.e., the process undergoes a *renewal*. Not surprisingly, results from renewal theory have been used very successfully to extract useful information about the chemistry and mechanics from bead-motor assays. The experimentally measured randomness parameter (or Peclet number), for example, has been used to gain more information about the number of rate determining substeps in the underlying enzymatic cycle (Svoboda et al., 1993). Time-correlation functions of the number of renewals have been used to obtain higher-order moments of the enzyme's chemical turnover time (Santos et al., 2005).

However, when backward steps or wasted ATP hydrolysis are present, renewal theory alone is not sufficient. This is because renewal theory is equipped to count only the number of enzymatic cycles that take place, and cannot take into account the mechanical stepping of the motor<sup>1</sup>. Backward steps, forward steps and wasted ATP hydrolyses are essentially mechanical phenomena, even though they may be closely coupled to the enzyme's chemical changes. Renewal-reward or cumulative processes are a useful extension of the renewal process that allows the association of an additional random variable to account for the stepping of the motor. Importantly, this formulation lets one derive a set of formulae that relate the long-time slopes of the bead's cumulants to the moments of the turnover time of the enzymatic cycle, *and* the superimposed mechanical steps of the motor. These formulae help one extract a host of statistical information from measurements the bead's displacement alone. The thesis of this paper is that these cumulants are robust statistical measurements that can be used to gain valuable insight into the physics of motor proteins.

Section 2 contains a short introduction to renewal theory, renewal-reward processes and a discussion about Markov chain models for enzymatic cycles. Section 3 describes the renewal reward formulation of bead-motor assays. Formulae for the cumulants of the bead's displacement are derived therein. Two simple propositions that help establish the slopes of the cumulants as robust statistical measures follow. The methods are then applied to analyze two models - Elston's model and the Peskin-Oster model - to illustrate their use. Due to the ubiquitous use of the randomness parameter in bead-motor assays, Section 4 describes its use and identifies several of its shortcomings: renewal-reward processes are very well-suited to analyze the statistical measures like the randomness parameter. An alternative to the randomness parameter that gives more information about the chemical cycle is suggested: specifically, it is possible to obtain estimates for the average chemical rate constants, the number of steps in the chemical reaction, the probability of backward steps and wasted hydrolyses solely by measuring the cumulants of the bead's displacement, and mapping these to a simple model.

---

<sup>1</sup>This may be taken into account using special formulations, as noted in Section 5

## 2 Mathematical Background

### 2.1 Renewal Theory

#### 2.1.1 Main Results in Renewal Theory

In this section, we review some of the important results in renewal theory for the reader's convenience. The following material follows closely the books of (Cinlar, 1975), (Karlin and Taylor, 1975), (Ross, 1983) and (Grimmett and Stirzaker, 2001).

A renewal process  $\{N(t), t \geq 0\}$  is a nonnegative, integer-valued stochastic process that registers successive occurrences of an event during the time interval  $(0, t]$ . The time intervals between successive events are given by a sequence of positive, independent, identically distributed random variables,  $\{X_k\}_{k=1}^{\infty}$  where  $X_k$  is the time interval between the  $(k-1)^{\text{th}}$  and  $k^{\text{th}}$  events. Let the distribution function of  $X_i$  be  $F(t)$ . This distribution is well-behaved and the density exists in most physical problems. It is also usually assumed that  $F(0) = 0$  and  $F(\infty) = 1$ , meaning that the renewal takes place in a finite time  $t > 0$  with probability 1 (Karlin and Taylor, 1975).

**Definition** (Waiting Time). The random variables  $\{S_n\}_1^{\infty}$  are defined as,

$$S_n = \sum_{i=1}^n X_i, \quad (1)$$

and  $S_n$  is called the waiting time until the occurrence of the  $n^{\text{th}}$  event. There is a natural equivalence between the sequence of waiting times and the counting process  $N(t)$ :

$$\mathbf{P}\{S_i \leq t\} \Leftrightarrow \mathbf{P}\{N(t) \geq i\}. \quad (2)$$

Two associated quantities, the *renewal function*  $M(t)$  and the *current life*  $\delta(t)$  are variables of interest, and their definitions are stated below.

**Definition** (Renewal Function and Excess Time). The renewal function  $M(t)$  is just the expected number of renewals  $N(t)$ . It can be written in terms of the  $k$ -fold convolution  $F_k(t)$  of the distribution function  $F$  as

$$E[N(t)] = M(t) = \sum_{k=1}^{\infty} F_k(t). \quad (3)$$

The current life  $\delta(t)$  is defined as

$$\delta(t) = t - S_{N(t)}, \quad (4)$$

and represents the time elapsed since the last renewal.

Let  $\mu$  and  $\sigma$  be the mean and variance of  $X_k$ , both finite. A version of the strong law applies to  $S_{N(t)}$  (Karlin and Taylor, 1975):

$$\lim_{t \rightarrow \infty} \frac{1}{t} S_{N(t)} \rightarrow \frac{1}{\mu}. \quad (5)$$

Indeed, since the sequence  $\{X_k\}_{k=1}^{\infty}$  contains identical, independently distributed random variables, a version of the central limit theorem holds:  $N(t)$  is asymptotically normal with mean  $t/\mu$  and variance  $\sigma^2 t/\mu^3$ .

An important result in renewal theory concerns certain integral equations called *renewal equations*. Their definition and solution are summarized below.

**Theorem 1** (Renewal Equations (Feller, 1968)). *Suppose  $a$  is a bounded function and  $F$  is a probability distribution function. Then there is a unique solution satisfying the renewal equation,*

$$A(t) = a(t) + \int_0^t A(t-x)dF(x). \quad (6)$$

The solution is

$$A(t) = a(t) + \int_0^t A(t-x)dM(x), \quad (7)$$

where  $M$  is the renewal function associated with  $F$ .

The central theorem in renewal theory has two forms: one is a differentiated form of the asymptotic relationship (5), and the other is concerned with the renewal equation. They are equivalent.

**Theorem 2** (Renewal Theorem (Karlin and Taylor, 1975)). *Let  $F$  be a non-arithmetic distribution of a positive random variable  $X$  with mean  $\mu$ . Suppose  $a$  is Riemann integrable and  $A$  is the solution of the renewal equation (6). Then,*

$$\lim_{t \rightarrow \infty} A(t) = \begin{cases} \frac{1}{\mu} \int_0^{\infty} a(x)dx & \text{if } \mu < \infty \\ 0 & \text{if } \mu = \infty \end{cases} \quad (8)$$

The second equivalent statement is,

$$\lim_{t \rightarrow \infty} \frac{M(t) - M(t-h)}{h} = \frac{1}{\mu} \quad (9)$$

Suppose a second sequence of independent, identically distributed random variable  $\{H_i\}_{i=1}^{\infty}$  is associated with the corresponding sequence of renewal increments  $\{X_i\}_{i=1}^{\infty}$ .  $H_i$  is allowed to be dependent on  $X_i$ , but the tuples  $(X_i, H_i), (X_j, H_j)$  are independent for  $i \neq j$ . For the rest of the section, it will be assumed the rewards accumulate at the end of each renewal interval. For the statistical parameters that are of relevance here, however, it does not matter at what point the rewards accumulate. This will be elaborated on when substeps are discussed in Section 3. Define the *cumulative process*  $R(t)$  as (Karlin and Taylor, 1975)

$$R(t) = \sum_{k=1}^{N(t)} H_k. \quad (10)$$

The expectation of  $R(t)$  satisfies the asymptotic relationship

$$\lim_{t \rightarrow \infty} \frac{E[R(t)]}{t} = \frac{E[H_k]}{E[X]}. \quad (11)$$

### 2.1.2 Cumulants of the Cumulative Process $R(t)$

The moments and cumulants of a renewal-reward process can be computed from the moment and cumulant generating function. As usual, define the moment generating  $\phi_X(s)$  function and cumulant generating function  $g_X(s)$  of the random variable  $X$  taking a countable number of discrete values  $\{\lambda_k\}_0^{\infty}$  as

$$\phi_X(s) = \sum_{k=0}^{\infty} e^{s\lambda_k} \mathbf{P}\{X = \lambda_k\}. \quad (12)$$

$$g(s) = \log(\phi_X(s)).$$

The  $n^{\text{th}}$  cumulant  $\kappa_{X,n}$  is just the  $n^{\text{th}}$  derivative of  $g_X(s)$  at  $s = 0$ .

To find the cumulants of  $R(t)$ , fix  $t$  at some value and drop ignore it for the following. Let  $N$  be a  $\mathbb{N}$  valued random variable, let the sequence  $\{H_i\}_{i=1}^{\infty}$  be as in Section 3 and let  $R$  be as in (10). Assume that  $N$  is

independent of each  $H_i$ . Then, a standard derivation gives

$$\begin{aligned}
\phi_R(s) &= \sum_{k=0}^{\infty} e^{s\lambda_k} \mathbf{P}\{R = \lambda_k\} \\
&= \sum_{k=0}^{\infty} \sum_{n=1}^{\infty} e^{s\lambda_k} \mathbf{P}\{R = \lambda_k | N = n\} \mathbf{P}\{N = n\} \\
&= \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} e^{s\lambda_k} \mathbf{P}\{H_1 + \dots + H_n = \lambda_k | N = n\} \mathbf{P}\{N = n\} \\
&= \sum_{n=1}^{\infty} \phi_H^n(s) \mathbf{P}\{N = n\} \\
&= \phi_N(\log(\phi_H(s))),
\end{aligned} \tag{13}$$

where the independence of  $N$  and the  $\{H_i\}$  has been used.

The  $n^{\text{th}}$  cumulant of  $R(t)$  can be found by differentiating the logarithm of (13), and expanding the composition of functions using Faà di Bruno's formula (Johnson, 2002) to obtain

$$\frac{d^m g_R(s)}{ds^m} = \sum \frac{m!}{b_1! \dots b_m!} g_N^{(k)}(g_H(s)) \prod_{i=1}^m \binom{g_H^{(i)}(s)}{i!}, \tag{14}$$

where the sum is over all  $m$ -tuples  $(b_1, b_2, \dots, b_m)$  that satisfy the constraint  $\sum ib_i = m$ . Since  $g_X(0) = 0$  for every random variable  $X$ , it is clear that  $\kappa_{R,n}$  is a polynomial function of the cumulants of  $N$  and  $H_i$ .

At this point, the parameter  $t$  may be reintroduced into  $N$  and  $R$  to emphasize their dependence on it. Smith (1958) showed that the cumulants of  $N(t)$  satisfy

$$\kappa_{N,n}(t) = a_n t + b_n + \frac{\lambda(t)}{(1+t)^p}, \tag{15}$$

when the moments upto  $n + p + 1^{\text{th}}$  order of the renewal increment  $X_i$  are finite. It is noted that  $a_n, b_n$  are constants dependent on the first  $n + 1$  moments of  $X_i$ ,  $p \geq 0$ , and  $\lambda(t)$  is a function of bounded variation going to zero as  $t \rightarrow \infty$ . When all the moments of  $X_i$  are finite,  $\lambda(t) = 0$ . Formulae for the constants  $a_n, b_n$  are available for the first eight cumulants (see Section 4 and Section A.2). Equations (14) and (15) imply that the cumulative process has the asymptotic form

$$\lim_{t \rightarrow \infty} \kappa_{R,n}(t) = c_n t + d_n \tag{16}$$

when the appropriate moments of  $H_i$  and  $X_i$  exist.

It was assumed in (13) that the  $H_i$  and  $N$  are independent to obtain the cumulant generating function of  $R$ . This, in essence, means that even the  $(X_i, H_i)$  pairs are not allowed to be dependent. Formulas are harder to derive when the dependence of the  $(X_i, H_i)$  pairs is allowed for. A formula for the variance when  $(X_i, H_i)$  depend on each other is available. Let  $\mu_X$ ,  $\mu_H$ ,  $\sigma_X^2$  and  $\sigma_H^2$  be the variances and means of  $X_i$  and  $H_i$ . Let  $\rho$  represent their correlation coefficient. Then, the long-time variance of  $R(t)$  takes the form (Smith, 1958)

$$\lim_{t \rightarrow \infty} \frac{\text{Var}[R]}{t} = \frac{1}{\mu_X} \left( \sigma_H^2 - 2\rho\sigma_X\sigma_H \frac{\mu_H}{\mu_X} + \frac{\sigma_X^2}{\mu_X^2} \mu_H \right). \tag{17}$$

However, this assumption of independence will not be overly restrictive when formulating motor protein dynamics as a renewal-reward process. This, as will be seen in Section 2.2, Section 3 and Section A.1, is because of the usual Markov process formulation of the enzyme's chemical cycle.

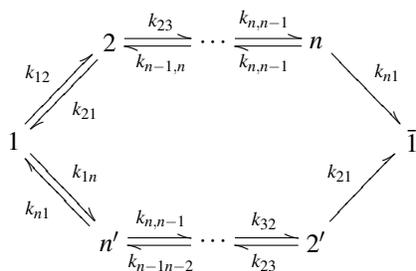
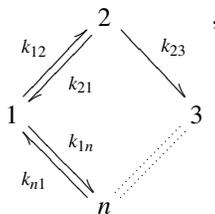


Figure 1: General first passage scheme based on Elston (2000)

## 2.2 Chemical Cycles and Markov Processes

Qian and L. Elson (2002) were the first to observe that the theory of Markov processes lends itself easily to the study of chemical cycles of single-molecule enzymes. Their model is inspired mainly by the fact that the times at which a single enzyme molecule encounters reactant molecules (e.g., ATP) in the medium are independent of each other and hence are Markovian events. Subsequent reactions like hydrolysis must, intuitively, depend on when they started, and a Markovian assumption may no-more be valid. To make such problems tractable, the most common approach is to approximate the distribution by a sum of exponentially distributed random variables; i.e., to model the single non-Markovian step as a set of artificial Markovian stages (Cox and Miller, 1977).

Following Elston (2000), for the enzymatic cycle of a motor protein that consumes ATP and goes through  $n - 1$  distinct intermediate states (labeled with the integers), a general *sequential* scheme can be written as



where  $k_{ij}$  denotes the rate of the reaction from state  $i$  to  $j$  and dotted lines represent the intermediate reactions between states 3 and  $n$ . Concentrations of ATP and by-products like ADP and  $P_i$  may be incorporated into the rate constants as multiplicative factors by assuming an appropriate reaction order.

The time required for the enzyme to cycle through all its states - the turnover or cycle time - is of primary interest. Although the directionality of the cycle is usually fixed by setting one or more backward rate constants to zero, a cycle may be considered to be complete if the the enzyme starts from state 1 and returns there after going through all the forward reactions or backward reactions at least once. That is, it takes one of the two paths from 1 to  $\bar{1}$  in Fig. 1, where the primed states have been artificially created to distinguish between forward and backward cycles. This cycle time for a single enzyme may be modeled as a *first-passage time* in a Markov chain/process. The standard approach to finding a first-passage time from state  $i$  to state  $j$  in a Markov chain is to make state  $j$  absorbing by setting all outward rate constants from state  $j$  to zero (Cox and Miller, 1977). In Fig. 1 state  $\bar{1}$  is absorbing.

Let  $X(t)$  be the Markov process taking values in  $\{1, \dots, n, 2', \dots, n, \bar{1}\}$ , as in Fig. 1. Let  $p_{ij} = \mathbf{P}\{X(t) = i | X(0) = j\}$  and let  $T$  be the first-passage time from state 1 to  $\bar{1}$ . Since  $\mathbf{P}\{T \leq t\} = p_{1\bar{1}}(t)$ , the objective is to find  $p_{1\bar{1}}$ . Let  $Q$  be the transition matrix of the Markov chain. Collect the  $p_{i\bar{1}}$  into a vector  $P$ , order the elements in  $P$  by adjacency starting with either  $p_{2'\bar{1}}$  or  $p_{n\bar{1}}$ , and notice that  $p_{\bar{1}\bar{1}}(t) = 1$ . Then,  $p_{i\bar{1}}(t)$  can be dropped from the equations. The transition rates into state  $\bar{1}$  from states  $n$  and  $2'$  are included as the first and last elements in vector  $B$ . Suppose also that we have some initial probability distribution on these states

assembled into a vector  $C$ . To be consistent, the initial probability distribution at  $t = 0$  will have to be on either the states in the forward half of the cycle or on the backward half, not both. Then the Kolmogorov backward equations for the Markov process can be written as

$$\begin{aligned}\frac{dP}{dt} &= QP + B, \\ y &= C^T P,\end{aligned}\tag{18}$$

where the superscript  $T$  represents the transpose and  $y$  gives the output; i.e., the distribution function of the first-passage time  $T$ .  $P, B$  and  $C$  are  $2n - 1$  dimensional vectors,  $y$  is a scalar and  $Q$  is a  $2n - 1 \times 2n - 1$  matrix.

One will note that this looks suspiciously like a state-space formulation of a control system - it is tempting to suggest that such a formulation makes available the wide range of tools in control theory<sup>2</sup>. Take the Laplace transform of (18) and denote the transformed quantities with an asterisk in the superscript. Then, the moment generating function of  $T$  is the *transfer function*  $G(s)$  of the system, and can be written as

$$\begin{aligned}E[e^{-sT}] &= \int_0^\infty e^{-st} \frac{dp_{1\bar{1}}}{dt} dt = sy^*(s) \\ &= sG(s) = C^T (sI - Q)B.\end{aligned}\tag{19}$$

In general, it is desirable to distinguish between these two (or more) pathways. One way to do this in the system defined above, would be to simply split the absorbing state into  $\bar{1}_b$  and  $\bar{1}_f$ . Then, the probabilities of absorption  $p_{1\bar{1}_f}$  and  $p_{1\bar{1}_b}$  can be found separately by dropping the either the first or last element from  $B$  in (18). Indeed, in this case  $P\{T \leq t\} = p_{1\bar{1}_b} + p_{1\bar{1}_f}$ . When formulating the bead-motor assay as a renewal-reward process, it will sometimes be necessary to distinguish between the backward and forward cycles (see Section 3.4.1).

There remains the question: is the first passage time identical if one assumes that the process starts in state  $i$  instead of state 1? While this seems intuitively obvious, it requires proof - a short one is presented in Section A.1. The proof hinges on conditioning on the path the system takes to the absorbing state. A well-known consequence is that absorption into a particular absorbing state is independent of the time of absorption. This will help justify the independence of  $H_i$  and  $X_i$  in Section 3.

---

<sup>2</sup>Note, for example, that one may also form observability and controllability matrices. In this particular problem, if one lets  $\mathcal{O} = \{C, CA, CA^2, CA^3, CA^4\}$  be the observability matrix and fix  $n = 3$ , one can show (by direct calculation) that the rank of  $\mathcal{O}$  is 3, although the transition matrix  $Q$  is  $5 \times 5$ . This is intuitive, since the extra states  $2'$  and  $3'$  were artificially constructed.

### 3 Renewal Reward Formulation of Bead-Motor Assays

Suppose  $X_i$  is identified with the turnover time for the enzymatic reaction and the associated rewards  $H_i$  with the physical steps taken by the motor. The cumulative reward  $R(t)$  represents the net distance travelled by the motor. Then, for example, identification of motor's steady-state velocity with the limiting of  $E[R(t)]/t$  in (11) follows immediately.

To allow for mechanical substeps during the renewal interval, a *terminal reward*  $\tilde{H}(t)$  is included in the cumulative process  $R(t)$ , and the corresponding cumulative reward process  $\tilde{R}(t)$  can be written as

$$\tilde{R}(t) = R(t) + \tilde{H}(t). \quad (20)$$

For instance, the terminal reward may be defined in terms of the current life  $\delta(t)$ . The current life represents the time elapsed since the previous renewal, and whether a substep takes place or not is intuitively dependent on how long the motor is waiting in its current state. To make ideas concrete, consider a linear motor with a step size of length  $L$  and uniform substep size of  $L_s$  (forwards or backwards). A substep may be defined to have taken place if a certain fixed time  $t_s$  has elapsed since the last renewal, where  $p$  and  $q$  represent the probabilities of stepping forwards or backwards. Then,

$$\tilde{H} = \begin{cases} \tilde{L}_s & \text{with probability } p \cdot \mathbf{P}\{\delta(t) > t_s\} \\ -\tilde{L}_s & \text{with probability } q \cdot \mathbf{P}\{\delta(t) > t_s\}. \end{cases} \quad (21)$$

A commonly found definition in the literature is restated for use in subsequent sections.

**Definition** (Additive Function of a Random Variable). Let  $\mu_i$  be the moments of a random variable  $X$ . A function  $V(X) = f(\mu_1, \dots, \mu_n)$  from  $\mathbb{R}^n \rightarrow \mathbb{R}$  of a finite subset of the moments of  $X$  is said to be an additive function if, for two independent random variables  $X$  and  $Y$ ,

$$V(X + Y) = V(X) + V(Y). \quad (22)$$

The cumulants are such additive functions. As is well known, for  $n \leq 3$ , the cumulants are just the central moments. The rates of increase of the central moments (like the velocity and variance of the bead's position) and functions of these (like the randomness parameter) are commonly measured parameters in bead-motor assays (Guydosh and Block, 2006; Svoboda et al., 1993; ?). As will be seen in the next few sections, the appropriate generalization is to consider the all the cumulants of the bead's position.

#### 3.1 Substeps

The existence of substeps in kinesin's walk has been a subject of debate for the past decade and many different substep sizes and durations have been reported in the literature (Block, 2007). It has also been argued based on experimental data that substeps of duration greater than  $30 \mu s$  do not exist (Carter and Cross, 2005). It will be shown that that substeps (specifically their size), whether they exist or not, do not have any effect on the bead's velocity and in general, the rates of increase of the bead displacement's cumulants.

**Proposition 1** (Substep information is lost in the cumulants). *Let  $V(R(t))$  be some function of the moments of the cumulative process  $R(t)$  satisfying the additivity property. Let  $\tilde{R}(t)$  include a terminal substep  $\tilde{H}(t)$  as in (20) and let  $\lim_{t \rightarrow \infty} V(R(t))/t$  exist. In addition, let  $V(\tilde{H}(t))$  be bounded, and let  $\tilde{H}$  be independent of every  $H_i$ . Then,*

$$\lim_{t \rightarrow \infty} \frac{V(R(t))}{t} = \lim_{t \rightarrow \infty} \frac{V(\tilde{R}(t))}{t} \quad (23)$$

*Proof.* Trivially,

$$\lim_{t \rightarrow \infty} \left| \frac{V(R(t)) - V(\tilde{R}(t))}{t} \right| = \left| \frac{V(\tilde{H}(t))}{t} \right| = 0, \quad (24)$$

using independence, additivity and the finiteness of  $V(\tilde{H})$ . □

By setting  $\tilde{H} = H_{N(t)+1}$ , this result implies that it does not matter at what point in the renewal interval the reward accumulates - this is a standard result (Cox and Miller, 1977; Smith, 1958). The more important observation is that *in the rates of increase of all the cumulants of  $R(t)$ , substep information is absent*. This result has been seen in particular scenarios. For example, Tsygankov et al. (2006) present a set of methods to gain insight about the underlying chemical cycle by assuming that individual (sub)steps can be resolved in the assay. They observe that the formula for the velocity is unaffected by how the substeps are counted. Our simple result extends this observation to the slopes of all the cumulants.

### 3.2 Initial Conditions Fade

Another consequence of the renewal reward formulation of the motor protein's walk is the intuitive result that the initial state of the enzyme does not affect its steady-state velocity (and in general, the rates of increase of the cumulants). There has been some debate about the necessity of including the equilibrium probabilities as initial conditions in the Markov chain describing the enzymatic cycle (Elston, 2000; Fisher and Kolomeisky, 2001). The standard result on the *delayed renewal process* (Karlin and Taylor, 1975) is used here to show that this assumption is unnecessary, and that regardless of initial conditions, the slopes of the cumulants are the same. This result has been seen in particular applications like Wang's derivation of the randomness parameter (Wang, 2007).

**Proposition 2** (Delayed Renewal-Reward Process). *Let the renewal-reward pairs  $(X_k, H_k)$  continue to be independent, but let only  $\{X_2, X_3, \dots\}$  and  $\{H_2, H_3, \dots\}$  be identically distributed. Let  $X_1$  and  $X_2$  have distribution functions  $G$  and  $F$  respectively. Let  $M_D$  be the expectation of the number of renewals; i.e., the delayed renewal function which includes all the renewal and reward increments, and let  $M$  be the renewal function associated with the distribution  $F$  alone. Similarly, define  $R_D$  and  $R$  by excluding the first renewal-reward pair for the latter. Let  $V$  be an additive function of the moments, and let  $\lim_{t \rightarrow \infty} V(R)/t$  exist and equal some finite constant  $L$ . Then,  $M_D(t)$  satisfies the renewal theorem (9) and*

$$\lim_{t \rightarrow \infty} \frac{V(R)}{t} = \lim_{t \rightarrow \infty} \frac{V(R_D)}{t} = L. \quad (25)$$

*Proof.* The proof uses a standard tool called the renewal argument. This begins by conditioning on the time of the first renewal  $X_1$ . Note that,

$$E[N(t)|X_1 = x] = \begin{cases} 0 & \text{if } x > t \\ 1 + M(t-x), & \text{if } x \leq t \end{cases}, \quad (26)$$

and use the fact that  $M_D(t) = E[E[N(t)|X_1 = x]]$  to find

$$M_D(t) = G(t) + \int_0^t G(t-x)dM(x). \quad (27)$$

This is a standard result. Then, the self-same renewal argument can be applied to  $V(R_D)$  and  $V(R)$  to obtain

$$\begin{aligned} V(R_D(t)) &= \int_0^\infty V(R_D(t)|X_1 = x)dG(x) \\ &= \int_0^t V(H_1 + R(t-x))dG(x) \\ &= V(H_1) + \int_0^t V(R(t-x))dG(x), \end{aligned} \quad (28)$$

where the additivity property of  $V$  and the independence of  $H_1$  and  $R(t)$  has been used. Again, following a standard procedure, divide (28) by  $t$ , take the limit and split the integral on the right hand side into two parts to obtain

$$\frac{1}{t} \int_0^t V(R(t-x))dG(x) = \int_0^{t/2} \frac{V(R(t-x))}{t} dG(x) + \int_{t/2}^t \frac{V(R(t-x))}{t} dG(x). \quad (29)$$

Now, note first that both  $V(R(t-x))/t$  and  $G$  are bounded and converge to limits  $L$  and 1 as  $t \rightarrow \infty$ , and second that  $G$  is positive and nondecreasing. Then, we may use the mean value theorem of integration to state that there exists  $\tau_1$  in  $(0, t)$  for the first integral and  $\tau_2$  in  $(t/2, t)$  for the second integral such that,

$$\frac{1}{t} \int_0^t V(R(t-x)) dG(x) = \frac{V(R(t-\tau_1))}{\tau_1} G(t/2) + \frac{V(R(t-\tau_2))}{\tau_2} (G(t) - G(t/2)). \quad (30)$$

Taking limits, it is clear that the first term in (30) converges to  $L$ , and the second goes to 0.  $\square$

### 3.3 Cumulants Revisited

To derive the useful cumulant formulae in Section 2 it was assumed that the  $(H_i, X_i)$  pairs were independent. This assumption is easily justified if the  $H_i$  and  $X_i$  are generated from an underlying Markov chain. That is, suppose the Markov chain has  $n$  states of which  $r < n$  are absorbing. Then,  $X_i$  is modeled as a first-passage time problem to one of these  $r$  absorbing states as in Section 2.2. The reward increment  $H_i$  may be then said to take values in  $\{\lambda_k\}_{k=1}^r$ , depending on which state the process is absorbed into. Then,  $H_i$  is well-defined if and only if the renewal takes place at some finite time. The probability of being absorbed into a particular state *given* that renewal has taken place, however, is independent of the actual time of absorption. In an irreducible Markov chain with absorption states, any process that begins in a transient state is absorbed in a finite time with probability 1 (Karlin and Taylor, 1975). Then, we can write

$$\begin{aligned} \mathbf{P}\{H_i = \lambda_k \cap X_i = t\} &= \mathbf{P}\{H_i = \lambda_k | X_i = t < \infty\} \mathbf{P}\{X_i = t\} + \mathbf{P}\{H_i = \lambda_k | X_i = \infty\} \mathbf{P}\{X_i = \infty\} \\ &= \mathbf{P}\{H_i = \lambda_k\} \mathbf{P}\{X_i = t\}, \end{aligned} \quad (31)$$

and the assumption that  $H_i$  and  $X_i$  are independent is justified.

It is not necessary, of course, that the renewal and reward increments always have to be generated by an underlying Markov process as indicated above. It just appears to the authors that this will be the most commonly encountered situation, and their independence can be justified using the arguments above.  $H_i$  and  $X_i$  may very well arbitrarily be constructed to be independent, like in Section 4.

Given this independence assumption, the cumulants of  $R(t)$  as  $t \rightarrow \infty$  take the form in (16). Since formulas for the intercepts  $d_n$  are known and can be written in terms of  $a_n, b_n$  in (15) and the cumulants of  $H_i$ , it appears that it can be used to gain additional information about the underlying mechano-chemical cycle. However, this intercept is corrupted with information from the initial chemical state of the enzyme and substeps. To demonstrate this, let  $\tilde{R}(t)$  be as in (20). Using (13), the cumulant generating function  $g_{\tilde{R}}(s)$  can be related to the cumulant generating function  $g_R(s)$  as

$$g_{\tilde{R}}(s) = g_R(s) + g_{\tilde{H}}(s). \quad (32)$$

Since the moments of  $H_k$  (as we have defined it) are constants independent of time, it is clear from Prop. 1 that the asymptote to  $R(t)$  can differ from  $\tilde{R}$  only by a constant; i.e., they have different intercepts.

Initial conditions have a similar effect. Since the experiment may capture the motor in the middle of a substep, the initial reward will be different from the subsequent ones; i.e.,  $H_1$  is not identical to  $H_k$  for  $k > 1$ . If it is assumed for simplicity that the initial renewal increment is not different, it follows that

$$g_{\tilde{R}}(s) = g_R(s) + g_{H_1}(s) - g_H(s) \quad (33)$$

There seems to be no simple way of finding the initial condition of the enzyme or the size of the substep. Hence, it may be concluded that the intercept is not as reliable a parameter as the slope of the asymptote. However, if individual (sub)steps can be resolved in the experiment (like in Carter and Cross (2005)), the intercept might become more useful.

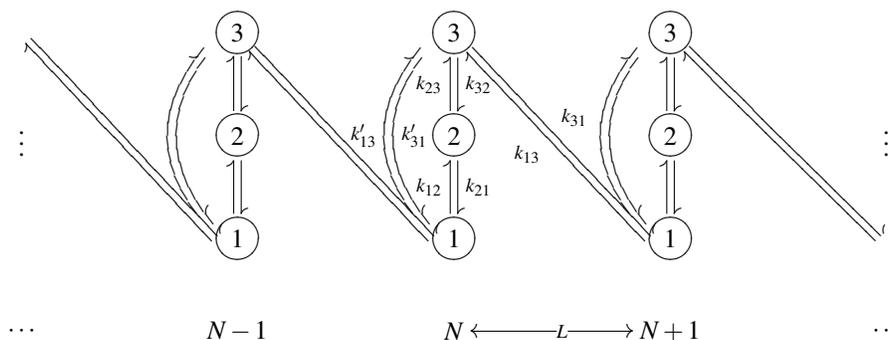


Figure 2: State diagram for Elston's model for  $n = 3$  from (Elston, 2000).

### 3.4 Examples

To illustrate the use of the renewal-reward formulation, it is applied to two existing models: namely, the Elston model (Elston, 2000) and the Peskin-Oster (PO) model (Peskin and Oster, 1995). Elston's model is a Markov process formulation for any motor protein undergoing a cyclic chemical reaction. The PO model was developed specifically for kinesin. The essential difference between the two in terms of renewal-reward processes is mainly in their approach to the reward function. Elston considers the rewards to be related only to the chemical cycle. In contrast, the rewards in the PO model depend on mechanical diffusion and the chemistry. The PO model is especially interesting since it helps illustrate at what time- and length-scales the physical approximations are made to model the bead-motor assay as a renewal-reward process. Some of their results are recaptured herein and insights gained using the renewal-reward formulation are discussed.

#### 3.4.1 Elston's Kinetic Model

It is common to embed the discrete steps of a motor protein in a continuous "envelope" dynamics, and analyze the motion of the motor as a Brownian ratchet (Astumian and Haenggi, 2002; Bier, 1997; Prager et al., 2005). That is, the probability density of the motor's position  $p(x, t)$  is described by an equation of the form,

$$D_{eff} \frac{\partial^2 p}{\partial x^2} - v \frac{\partial p}{\partial x} - \frac{\partial p}{\partial t} = 0, \quad (34)$$

where  $D_{eff}$  is the effective diffusion coefficient and  $v$  is the average velocity of the motor. These transport properties tell us how effective the motor is in conducting unidirectional transport.

Elston's model shows how these transport properties can be related to an underlying chemical cycle. In the model, there are  $n$  chemical states, and an integer valued random variable  $N(t)$  which keeps track of which binding site the motor is at. The state diagram from (Elston, 2000) is reproduced for the readers' convenience. The motor steps forwards or backwards when the corresponding sequence of chemical reactions take place. Let  $p_{n,i}$  represent the probability of being in chemical state  $n$  and binding site  $i$ . The (infinite) set of evolution equations for the  $p_{n,i}(t)$ , the Kolmogorov forward equations or Fokker-Planck equations, can be conveniently represented in matrix form (Van Kampen, 2007). Then, expressions for the first two cumulants of  $N(t)$  can be found in terms of the rate constants of the underlying chemical cycle and related to  $v$  and  $D_{eff}$ .

The special case of the model for  $n = 3$  is considered here, and cast as a renewal-reward process. The procedure for general  $n$  is identical. There is an intuitive regenerative structure. Suppose the enzyme starts in state 1. Whenever it returns to 1 after going through all the other states  $(2, 3, \dots, k)$ , a renewal takes place. There are three ways in which this can happen, and each of these ways is associated with a backward step, a forward step, or a wasted step. Construct the appropriate absorbing states as shown in Fig. 3 to calculate the first passage times  $X_i$  and reward increments  $H_i$ .

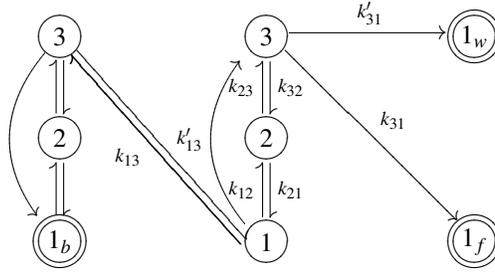


Figure 3: To find the first passage time and reward function. The absorbing states have been indicated with a double circle. Absorption into states with the subscripts  $b, f$  and  $w$  determines whether a backward step, forward step or wasted hydrolysis takes place.

Define the vectors  $P$ ,  $y$  and the matrices  $Q, B, C$  as in Section 2.2 to find the first passage time - with of course, the appropriate modifications to include the reaction rates  $k'_{13}$  and  $k'_{31}$ . Then, as before, the moment generating function of the first passage time is just the transfer function of the Laplace transformed set as in (19). The definition of the reward function  $H_i$  is,

$$H_i = \begin{cases} L & \text{if absorbed into } 1_b \\ 0 & \text{“ } 1_w \\ -L & \text{“ } 1_f \end{cases} \quad (35)$$

To find the  $k^{\text{th}}$  moment of the reward function  $E[H^k]$ , condition on absorption occurring in the infinitesimal time interval  $(\tau, \tau + d\tau)^3$ . Since  $p_{11_b}(t)$  is the distribution function of the first passage time given absorption into state  $1_b$  takes place, the probability that absorption into state  $1_b$  takes place in  $(\tau, \tau + d\tau)$  is just  $p'_{11_b}(\tau)d\tau$ . Then, one obtains

$$\begin{aligned} E[H^k] &= L^k \left( \int_0^\infty p_{11_f}(\tau)d\tau + (-1)^k p_{11_b}(\tau)d\tau \right) \\ &= L^k (p_{11_f}(\infty) + (-1)^k p_{11_b}(\infty)). \end{aligned} \quad (36)$$

Finding  $p_{11_f}(\infty)$  and  $p_{11_b}(\infty)$ , as noted in Section 2.2, is just a matter of writing  $B$  in (18) as  $B_{1_f} = k_{21} \delta_{1_i}$  or  $B_{1_b} = k_{31} \delta_{7_i}$ , where  $\delta_{ij}$  is the Kronecker delta function. Then, find the limit as  $s \rightarrow 0$  of  $sG(s)$  in (19) (with the appropriate  $B$ ) to obtain  $p_{11_f}(\infty)$  and  $p_{11_b}(\infty)$ .

Elston states solutions for the velocity and diffusion coefficient - the slopes of the mean and half the variance of  $R(t)$  - only for the case where the backward rates  $k_{21}$ ,  $k_{32}$ , and  $k'_{13}$  are zero. This simplifies the algebra considerably since backward steps are no longer possible. The same is done here. Since the chain has no backward rates, the first passage time is just a sum of exponentials, and the mean and variance ( $\mu$  and  $\sigma$ ) take the form

$$\begin{aligned} \mu &= \frac{1}{k_{12}} + \frac{1}{k_{23}} + \frac{1}{k_{31} + k'_{31}}, \\ \sigma^2 &= \frac{1}{k_{12}^2} + \frac{1}{k_{23}^2} + \frac{1}{(k_{31} + k'_{31})^2}. \end{aligned} \quad (37)$$

In this special case, it is easier to find  $p_{11_f}(\infty)$  and  $p_{11_b}(\infty)$  using the path decomposition described in Section A.1. First notice that the cycle completes only through state 3; i.e., 3 is the last state prior to absorption

<sup>3</sup>With backward chemical reactions present, it is hard to find the probabilities of absorption into particular states using the path decomposition in Section A.1 since the possible paths to absorption are too numerous.

into either  $1_b$  or  $1_f$ . Then it is clear from (60) that

$$p_{11_f}(\infty) = \frac{k_{31}}{k_{31} + k'_{31}}. \quad (38)$$

The variance of  $R(t)$  can be found using its generating function in (13) since  $H_i$  is independent of  $X_i$  (see Section A.1 and Section 2.2. Nevertheless, it is illuminating to consider the more complicated expression in (17) and show that the correlation coefficient it is zero. The numerator of the correlation coefficient can be written as

$$E[HX] - E[X]E[H] = L \left( \int_0^\infty t p'_{11_f}(t) dt - p_{11_f}(\infty) \int_0^\infty t (p'_{11_f}(t) dt + p'_{11_w}(t) dt) \right). \quad (39)$$

But the first term on the right hand side of (39) is

$$\int_0^\infty t p'_{11_f}(t) dt = \int_0^\infty \frac{p'_{11_f}(t)}{p'_{11_f}(t) + p'_{11_w}(t)} t (p'_{11_f}(t) + p'_{11_w}(t)) dt. \quad (40)$$

From the structure of (19), it is clear that the Laplace transform of  $p_{ix}$ , where  $x$  is either  $1_b$  or  $1_f$  can be written as

$$\begin{aligned} p_{i1_f}^*(s) &= f(s)k_{31}, \\ p_{i1_w}^*(s) &= f(s)k'_{31}. \end{aligned} \quad (41)$$

Due to the linearity of the Laplace transform, it follows that

$$\frac{p'_{11_f}(t)}{p'_{11_f}(t) + p'_{11_w}(t)} = \frac{k_{31}}{k_{31} + k'_{31}}, \quad (42)$$

and from (39) and (38) it follows that the correlation coefficient is zero.

It is just a matter of substitution and algebraic manipulation to verify using (11) and (17) that the expressions for velocity and the effective diffusion coefficient are identical to those obtained by Elston. The velocity is

$$v = \frac{k_{12}k_{23}k_{31}}{k_{23}(k_{31} + k'_{31}) + k_{12}(k_{23} + k_{31} + k'_{31})}, \quad (43)$$

and using the variance of reward process from (17) the diffusion coefficient is

$$\begin{aligned} D_{eff} &= \lim_{t \rightarrow \infty} \frac{E[R(t)^2] - E[R(t)]^2}{2t} \\ &= \frac{1}{2} \left( p_{11_f}^2(\infty) \frac{\sigma^2}{\mu^3} + \frac{(p_{11_f}^2(\infty) - p_{11_f}(\infty))}{\mu} \right) \end{aligned} \quad (44)$$

### 3.4.2 The Peskin-Oster Model

The Peskin-Oster model is a one-dimensional model developed specifically for kinesin. It describes experimental data obtained in bead-motor assays (Gilbert et al., 1995). The model is of interest because the reward function here depends on the chemistry and mechanical diffusion, in contrast to Elston's formulation. This makes it interesting to recapitulate the physical assumptions that have to be made to analyze the model as a renewal-reward process. To this end, a more-than-brief description of kinesin and the PO model is provided, and it is then cast as a renewal-reward process.

Kinesin is a motor protein that walks along track-like structures called microtubules (MTs) in cells. It has two heads made of globular proteins; most of its mass is concentrated here. These heads are connected

to each other through string-like structures called neck-linkers. A long tether connects the point of joining of the neck-linkers to the cargo. MTs have regularly spaced chemical binding sites. Kinesin's heads have chemical binding sites for both MTs and (ATP). Kinesin walks by alternately detaching and reattaching its heads from the MT. It derives the energy required for this process from ATP hydrolysis.

The model describes kinesin's chemical cycle as follows: kinesin starts with both heads bound to the MT. ATP binds to one of these heads, following which either the forward head or backward head detaches. Once one of the two heads detaches, it undergoes a conformational change that is modeled as a "power-stroke" (Rice et al., 1999); i.e., it undergoes a diffusion in the fluid medium along with the bead in a potential biased towards the forward MT binding site. ATP hydrolysis takes place simultaneously with the power-stroke. Once the hydrolysis completes, the diffusing head regains its affinity for the MT and binds to the nearest site nearly instantaneously. The process of ATP binding and either leading or trailing head detachment is quantified by the rate constants  $\beta_b$  and  $\beta_f$  respectively. ATP hydrolysis takes place with a rate  $\alpha$ .

To model the mechanical diffusion, the tether attached to the bead and the two neck linkers are modelled as linear, elastic springs and the heads and bead are modelled as point masses. There are two distinct states of the kinesin molecule: one in which only one head is diffusing while the other is bound, and a second in which both heads are bound to the MT (waiting for an ATP molecule to bind). Let  $x_b$  represent the bead position, let  $x$  be the free head location, let  $x_{bnd}$  be the position of the bound head, and let  $f$  be the external force on the bead. Each time the motor steps, the position variables are translated by  $\pm L$ , the distance between the binding sites. The potential energy of the system in the two states can then be written as

$$\begin{aligned}\phi_1(f, x_b) &= f(x_b - x_h) + \frac{1}{2}K_{th}(x_b - x_h)^2, \\ \phi_2(f, x_b, x) &= f(x_b - x_h) + \frac{K_{th}}{2}(x_h - x_b)^2 + W(x_h - x_{bnd}),\end{aligned}\tag{45}$$

where the  $x_h = (x_{bnd} + x)/2$  and  $W(x)$  denotes an interaction potential that biases the diffusion of the head towards the forward binding-site (the power-stroke). Following Atzberger and Peskin (2006), the interaction potential is modeled as  $W(x) = 1/2K_{bias}(x - x_0)^2$ .

To solve for the moments of the bead's displacement, two critical time-scale separations in Peskin and Oster (1995). First, since the bead is nearly 1000 times larger than each of the heads, its diffusion coefficient is proportionally smaller and the heads diffuse that many times faster than the bead. Hence one may assume that in the over-damped, low Reynolds number environment of the cell that the diffusing head is in mechanical equilibrium with the bead at all times. Second, they argue that the diffusion of the bead itself may be considered to be many times faster than the chemical rates, and diffusing times are insignificant on the time-scale of the experiment. To be more precise, the bead relaxes to its equilibrium density nearly instantaneously. These two time-scale separations allow them to calculate the probabilities of binding forwards  $p(f)$ , once the hydrolysis completes. When formulating this as a renewal-reward process, the diffusion of the bead during the hydrolysis phase may be viewed as a substep, and may be ignored (see Prop. 1). But it is necessary to account for bead diffusion when both heads are bound to MT. Hence this assumption is elaborated on.

Since  $\phi_1(f, x_b, x_h)$  is quadratic in  $x_b$ , the stochastic differential equation describing the diffusion is in the form of the Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein, 1930). That is,

$$dX_b(t) = \frac{K_{th}}{\gamma}(X_b - x_{eq})dt + \sigma dW_t,\tag{46}$$

where  $\sigma = 2\gamma k_B T$  is the noise intensity,  $\gamma = 6\pi\mu R$  is the friction coefficient given by Stokes' law for a sphere in a low Reynolds number flow,  $x_{eq}$  is the equilibrium position of the bead in the potential given in , and  $W_t$  is the standard Wiener process. It can be shown with the help of the Maxwell-Boltzmann distribution that  $\sigma = 2\gamma k_B T$ , where  $k_B$  is the Boltzmann constant and  $T$  is the absolute temperature. Given  $X_b(0) = x_0$ , the Fokker-Plank equations may be solved to find the mean and variance of the bead as a function of time (Cox

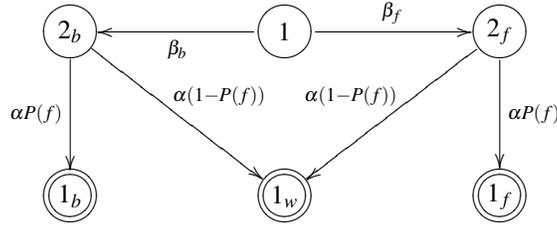


Figure 4: The state transition matrix to find the moments of the first passage time and reward function in the PO model. As in Fig. 3 the states absorption into  $1_f$ ,  $1_b$  or  $1_w$  indicates that a forward step, backward step or wasted hydrolysis has taken place. Notice that if two successive transitions take place, renewal is complete; i.e., the number of possible paths is finite.

and Miller, 1977) as

$$\begin{aligned} E[X_b(t)] &= x_{eq} + x_0 e^{-K_{th}t/\gamma}, \\ \text{Var}[X_b(t)] &= \frac{k_B T}{K_{th}} (1 - e^{-2K_{th}t/\gamma}). \end{aligned} \quad (47)$$

It is also known that  $X_b$  is Gaussian. Then,  $X_b(t)$  relaxes to a normal distribution with mean and variance given by the long term limits of (47), in a time-scale of  $\gamma/K_{th}$ .  $K_{th}$  may be obtained from physical estimates or even from a careful analysis of bead-motor assays (Atzberger and Peskin, 2006), and this turns out to be around  $1.5 \cdot 10^{-4} N/m$ . Using an average bead size of  $1 \mu m$  and a friction coefficient  $\gamma \approx 10^{-8} Ns/m$ , gives a time scale of around  $10^{-5} s$ . Since kinesin's velocities are around the  $100 nm/s$ , the chemical time is at least two orders of magnitude greater. Hence it may be assumed that  $X_b$  is normally distributed in the state when both heads are bound with mean and variance given by the long time limits (47).

Kinesin's chemical cycle in the PO model; i.e., the renewal increments  $X_i$ , is a sum of two exponentially distributed random variables. It consists of the waiting time for ATP binding and the time for ATP hydrolysis.<sup>4</sup> The reward increment can be written as,

$$H_i = \left\{ \begin{array}{ll} L & \text{with probability } \frac{\beta_b}{\beta_b + \beta_f} p(f) \\ 0 & \text{“ } \frac{\beta_b}{\beta_b + \beta_f} (1 - p(f)) + \frac{\beta_f}{\beta_b + \beta_f} p(f) \\ -L & \text{“ } \frac{\beta_f}{\beta_b + \beta_f} (1 - p(f)) \end{array} \right\} + Y \quad (48)$$

where  $Y$  represents the displacement of the bead diffusing in the potential given by (45). The equilibrium position of the bead depends only on which binding site the motor is at. Thus, the mean of  $Y$  may be assumed to be 0 in (47); i.e., it amounts to a constant translation of the coordinates and will not affect the slope of the asymptote.  $Y$  is independent of  $X_i$  under the time-scale separation, and its higher cumulants are identically zero since it is normally distributed.

The PO model displays substeps. The size of the substeps is dependent on the external force, as can be seen from (45). For the Markov process calculation in Peskin and Oster (1995), the substep is set to be uniformly  $L/2$ . Although it is not immediately noticeable in the equations for the moments, it is reassuring to note that this substep size does not end up appearing in the expressions for the bead's velocity and the variance in Peskin and Oster (1995). Figure 4 shows a Markov chain that can be used to calculate the first-passage time and probabilities for the reward function. In the PO model, the total number of paths to absorption is finite and the chemical process is cyclic. As in Section 3.4.1, the first passage time is the same whether the cycle is considered to start from state 1 or state 2, and  $H_i$  is independent of  $X_i$ . With the reward and renewal

<sup>4</sup>One can argue that the hydrolysis may have a more complicated distribution, and may be better approximated by the “method of stages” Section 2.2

increments defined, it is just a matter of applying (11) and (16) again to find expressions for the asymptotic velocity and rate of increase of the variance of the bead. The expressions obtained are identical, and are omitted for the sake of brevity.

## 4 The Randomness Parameter and an Alternative Approach

### 4.1 The Drawbacks of Using the Randomness Parameter

The randomness parameter or the inverse of the Peclet number has seen ubiquitous use in the analysis of bead-motor assays since its description in Svoboda et al. (1993). It is commonly defined as,

$$r = \lim_{t \rightarrow \infty} \frac{\text{Var}[x_b(t)]}{E[x_b(t)]L}, \quad (49)$$

where  $L$  is the uniform step length of the motor. Recalling the discussion (in Section 3.3) on the slopes and intercepts of the cumulants of a renewal-reward process used to represent a bead-motor assay, it seems clear that it is beneficial to exclude the effect of the unreliable intercept. Hence, the randomness parameter is better defined as,

$$r = \lim_{t \rightarrow \infty} \frac{\text{Var}[x_b(t)]}{t} \frac{t}{E[x_b(t)]L}. \quad (50)$$

When wasted chemical cycles and backward steps are not allowed and the motor steps uniformly forward, the randomness parameter is very useful. Suppose the chemical cycle of the enzyme can be approximated by a set of  $n$  Markovian steps with rate constants  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ . Then the cycle time or the renewal increment  $X_i$  is just a sum of  $n$  independent exponential random variables. Let  $\lambda_{\min}$ ,  $\lambda_{\max}$  be the minimum and maximum rate constants. Then the following inequality holds (Santos et al., 2005):

$$\frac{1}{n} \frac{\lambda_{\min}}{\lambda_{\max}} \leq r_N = \frac{\sum_{i=1}^n \frac{1}{\lambda_i^2}}{\sum_{i=1}^n \left(\frac{1}{\lambda_i}\right)^2} \leq 1, \quad (51)$$

where we have used  $r_N$  to highlight that the randomness here is of just the renewal process alone. The cumulative process does not need to be considered since the motor steps inexorably forwards.

If, however, even wasted hydrolyses are present - i.e., the reward  $H_k$  can take the value 0 with non-zero probability - (51) does not hold. This was anticipated by Svoboda et al. (1993) when the parameter's use was originally suggested. If backward steps are included, its failure is dramatic. To see this, let  $X_i$  be as above and define  $H_i$  as in (48), but with the probabilities of stepping forward and backward replaced by  $p$  and  $q$ . Here, it just needs to be asserted that  $H_i$  is independent of  $X_i$ . Then the moments of  $H_i$  are just  $E[H_i^k] = L(p + (-1)^k q)$ . Since the rewards are independent of the  $X_i$  by assertion, (17) is easy to apply and the randomness parameter of the cumulative process  $r_R$  can be written as

$$r_R = \frac{p+q}{p-q} + (r_N - 1)(p - q) \quad (52)$$

Using the bounds for  $r_N$  in (51), the bounds of  $r_R$  can be obtained as

$$0 \leq \frac{(p+q) - (p-q)^2}{p-q} \leq r_R \leq \frac{p+q}{p-q}. \quad (53)$$

When  $p$  and  $q$  are close to each other - for e.g., when the motor is close to stall -  $r_R$  may become really large. Another special case to consider is when there are no backward steps  $q = 0$ , but  $p < 1$ . In this case,  $1 - p \leq r_R \leq 1$ .

In Guydosh and Block (2006) and Visscher et al. (1999), the randomness parameter is greater than one (and rising fast, as in Fig. 5a) close to stall and at low ATP concentrations. Like in the PO model, ATP concentration and the force applied to the bead both affect  $p$  and  $q$ . This fact might, for instance, be used to argue for the existence of backsteps. However, this argument is no longer valid when backward *chemical reactions* become significant; i.e., when the ATP concentration is low, backward reactions may have rate constants similar to the forward reactions, and inequality (51) no longer applies. At saturating ATP, both of

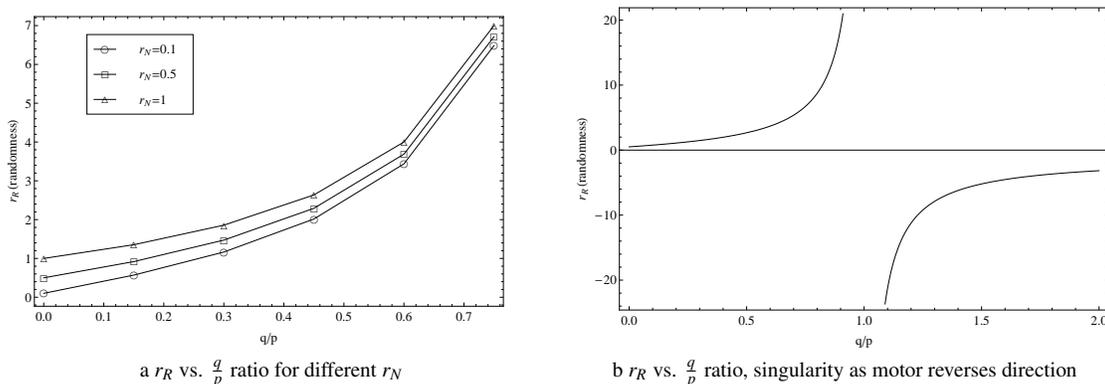


Figure 5: The variation of  $r_R$  with the  $\frac{q}{p}$ . Figure (a) shows variation with different  $r_N$  values. Note how  $r_R$  values become nearly identical as the values of  $p$  and  $q$  approach each other. Figure (b) shows the singularity at  $p = q$ . Clearly, this occurs when the velocity of the motor is zero.

the remaining two possibilities remain equally likely: close to stall, the external force may either affect the probability of binding forwards during the diffusing phase or it may directly affect the chemical cycle. Indeed, it is conceivable that both may be occurring simultaneously. Hence it is desirable to be able to quantify the effects of ATP concentration and load on *both* the rate of the chemical cycle and the probabilities of stepping backwards and forwards.

While the randomness parameter might, conceivably, help indicate the existence of backsteps, it loses its original purpose in their presence. Figure 5 shows that the measured randomness deviates significantly from the randomness of the underlying chemical cycle. It has been seen in experiments (in which individual steps are resolvable) that the ratio of backward to forward steps varies from  $10^{-4}$  at high ATP concentrations and low loads to  $10(!)$  close to stall.

## 4.2 An Alternative Approach

It is desirable to obtain an estimate for the number of rate determining steps in the enzyme's chemical cycle, the approximate average rate of each of these steps, and the probabilities of binding forward and backward. Given measurements of bead position - and its cumulants - the randomness parameter, in essence, fits it to a model. This model assumes that the motor steps uniformly forwards and that the chemical cycle is Erlang distributed with parameters  $n$  and  $\lambda$ . The model to which the bead experimental data will be mapped to will henceforth be referred to as the *test model*.

Clearly, this model is insufficient if backward steps and wasted chemical cycles are considered. Instead, if more information is obtained by measuring the slopes of the higher moments of  $R(t)$ , more complex models may be considered. To account for this, a reward function  $H_k$  which accounts for forward, backward and wasted steps may be considered along with the sequential chemical cycle. Now there are four unknowns in the model:  $\lambda$ ,  $n$ , and the probabilities  $p$  and  $q$ . Then, it is reasonable to assume that the first four slopes of the cumulants of the bead's position are enough to determine these parameters uniquely.

The constants  $a_n$  in (15) are needed. The first eight values are available in Smith (1959), and the first four are given in Section A.2 (with a small extension). For the test model with the Erlang distributed  $X_i$ ,  $a_i = \lambda/n^i$  in general. While this result is hard to see from (72), and (70), there is an intuitive way of deriving this using the properties of the Poisson process and Proposition 1 (see Section A.2). The cumulants of the test model can be calculated using (13). Let the experimentally measured slopes of the first four cumulants be  $s_i$ . The

resulting set of equations are stated in full for completeness:

$$\begin{aligned}
s_1 &= \lambda \left( \frac{p-q}{n} \right), \\
s_2 &= \lambda \left( \frac{(p+q) - (p-q)^2}{n^2} + \frac{(p-q)^2}{n} \right), \\
s_3 &= \lambda \left( \left( \frac{p-q}{n} \right)^3 + \frac{3(q(1-q) + p(1+2q) - p^2)}{n^2} + \frac{1-3q+3p+4pq+2p^2+2q^2}{n} \right), \\
s_4 &= \lambda \left( \left( \frac{p-q}{n} \right)^4 + 6 \frac{(p-q)^2((p-q)^2 - (p+q))}{n^3} + \right. \\
&\quad \left. \frac{7(p+q)^2 - 18(p+q)(p-q)^2 + 11(p-q)^4 - 16pq}{n^2} + \right. \\
&\quad \left. \frac{-6(p-q)^4 + 12(p-q)^2(p+q) - 7(p+q)^2 + 16pq + (p+q)}{n} \right).
\end{aligned} \tag{54}$$

The four polynomial equations in (54) may be solved numerically in general. To illustrate this fitting procedure, it may be applied to the PO and Elston models. That is, the cumulants for both of these models may be obtained separately and plugged into (54).

As a check, the cumulants for the PO model were obtained using the methods of renewal-reward processes and compared with those obtained using the Markov process formulation in Peskin and Oster (1995). As noted in Section 3.4.2, the renewal increment of the PO-model is a sum of two exponentials with rates  $\alpha$  and  $\beta_b + \beta_f$ . So, it is reasonable to expect that  $n$  should be in the vicinity of 2, and that  $\lambda$  should be close to the average of the two rates. It turns out that for this special case,  $\lambda$  and  $n$  take exactly those values. The analytical result was guessed at by examining numerical solutions for particular values of the constants. The full solution is

$$\begin{aligned}
n &= 2, \\
\lambda &= \frac{\alpha + \beta_b + \beta_f}{2}, \\
p &= \frac{4p(f)\alpha\beta_b}{(\alpha + \beta_b + \beta_f)^2}, \\
q &= \frac{4(1-p(f))\alpha\beta_f}{(\alpha + \beta_b + \beta_f)^2},
\end{aligned} \tag{55}$$

where, as before,  $p(f)$  is the force-dependent probability of stepping forwards.

The results are different for special case of the Elston model considered in Section 3.4.1. The set of four equations may be reduced to two, but the analytical solution of these simultaneous cubic equations in  $n$  and  $\lambda$  is not worth examining, as we believe no further insight can be gained from doing this. However, it may be tested in a certain restricted sense to reassure ourselves that the method works satisfactorily: the forward rate constants  $k_{12}$ ,  $k_{23}$  and  $k_{31}$  were fixed to certain arbitrary values, and the reverse rate  $k'_{31}$  was varied over a range. Equations (54) were solved numerically to obtain  $n$ ,  $\lambda$ ,  $p$  and  $q$ . The results are as expected:  $n$  is close to 3 over the entire range,  $\lambda$  is the close to the average of the rate of chemical turnover  $(k_{12} + k_{23} + k_{31} + k'_{31})/3$ , and importantly  $q$  is 0, which indicates that there are no backward steps in the cycle. The randomness parameter, in contrast, tells us that there are about 1.3 steps in the cycle and nothing more. The results are shown in Fig. 6.

To demonstrate the robustness of the procedure in the presence of experimental error, artificial errors are introduced into the slopes of the cumulants.  $k'_{31}$  was first fixed at 9.2, and 1000 separate sets of uniformly distributed errors ( $\pm 10\%$ ) were added to each of the four slopes. Of course, it's not reasonable to expect an exact solution will make sense under an arbitrary perturbation such as this. Instead, the Euclidean norm of

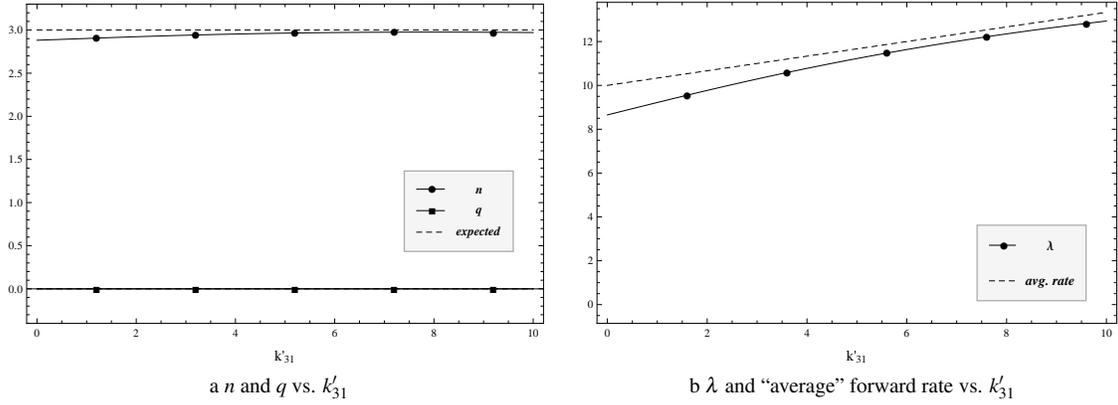


Figure 6: The values of the rate constants in Elston’s model for  $n = 3$  were fixed. The backward rates were set to 0,  $k_{12} = 10$ ,  $k_{21} = 15$  and  $k_{31} = 5$ .  $k'_{31}$  was varied in a range and  $\lambda$ ,  $n$ ,  $p$  and  $q$  were solved for.

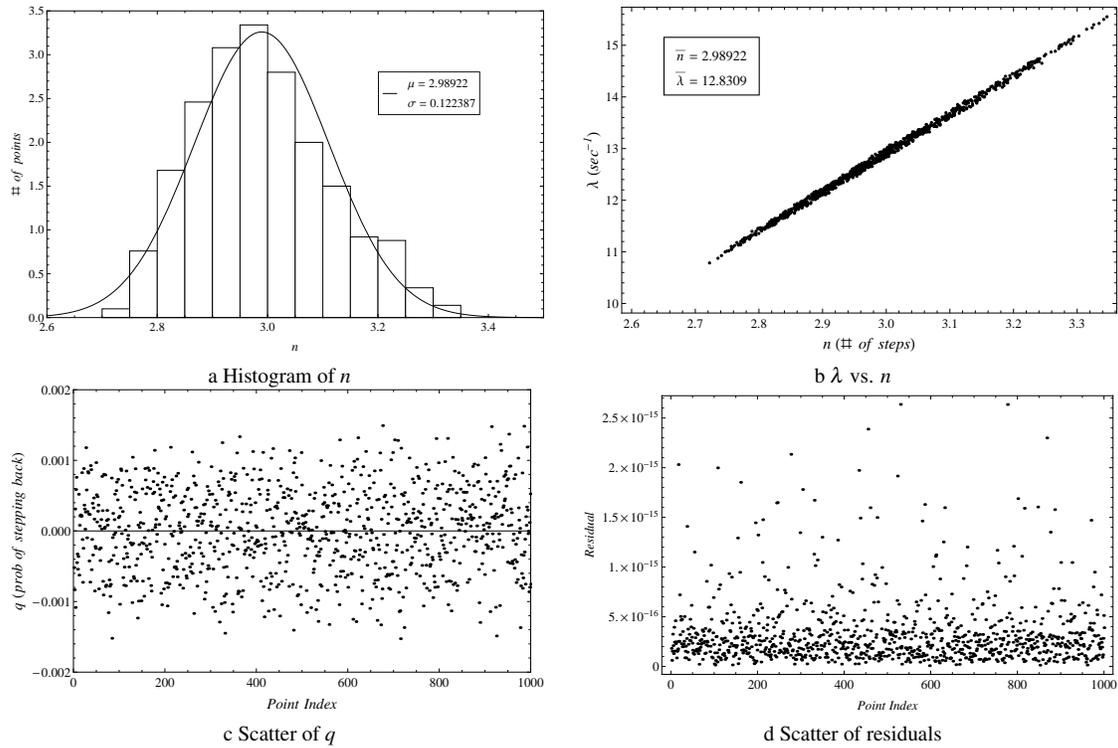


Figure 7: The values of the rate constants in Elston’s model for  $n = 3$  were fixed, as in Fig. 6 and  $k'_{31} = 9.2$ . Errors in a  $\pm 10\%$  range were added to the four slopes, and  $\lambda$ ,  $n$ ,  $p$  and  $q$  were solved for. Figure 7a shows a histogram of the  $n$  values obtained. All the values fall in the  $(2.6, 3.4)$  interval. Figure 7b shows the variation of  $\lambda$  with  $n$ ;  $\lambda$  appears to vary linearly with  $n$ . Figure 7c shows that most of the  $q$  fall in a  $(-10^{-3}, 10^3)$  interval. Although the negative values are inadmissible, this a strong indicator that there are no backward steps.

the set of four residuals of Eqs. (54) was minimized. The  $n$  values had a mean of 2.988 and a sample variance of 0.12. The residual had an upper bound of about  $10^{-14}$ . A histogram for  $n$ , the residuals and other results are shown in Fig. 7. These results show that the proposed approach works very well.

In practise, such a simple chemical scheme might not work very well. Instead, one might consider a serial scheme with forward and reverse reactions; i.e., one that has two additional constants  $\mu$  and  $m$  representing the overall rate and order of the reverse reactions. Such simple serial schemes have been considered before (Fisher and Kolomeisky, 2001; Tsygankov et al., 2006). With two new parameters, six equations have to be considered. Rather than applying it to an existing model, this procedure may more fruitfully be tested in an actual experiment. The authors believe that quantifying the effects of ATP and external force on  $\lambda$ ,  $\mu$ ,  $n$ ,  $m$ ,  $p$  and  $q$  will prove particularly insightful.

## 5 Outlook and Conclusions

Renewal reward processes are particularly well-suited to extracting cumulants from underlying Markov models. It is easy to build simple models like those in Section 4.2 fit these to experimental data. However, since the details of the underlying chemistry and physics are embedded in a cycle-time variable  $X_i$  and associated reward increments  $H_i$ , it has its drawbacks. In detail, these are,

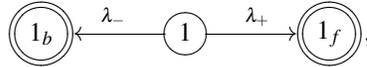
- One of the major advantages of using cumulants of the bead’s displacement is that it is not really necessary to be able to resolve individual steps. However, when individual steps can be resolved, the methods herein must be supplemented by other approaches to gain more information and insight (Fisher and Kolomeisky, 2001; Santos et al., 2005; Tsygankov et al., 2006).
- Detachment and reattachment phenomena result in the loss of regenerative structure; it appears that their inclusion into the renewal-reward analysis is not trivial. It is conceivable that the possibility of detachment may be included by assuming that there is a nonzero probability that renewal does not complete in finite time; i.e., it is a “terminating” renewal process (Karlin and Taylor, 1975). Of course, by considering the entire underlying Markov model, methods in Elston (2000); Mogilner et al. (2001) may be used to understand these phenomena.
- Independence between the reward increments and the renewal increments  $X_i$  cannot be assumed in general. However, if  $X_i$  and  $H_i$  are appropriately generated from an underlying Markov chain, the simplifying assumption of independence can be made. Consider the the situation where there are two possible chemical reactions with rates  $\lambda_+$  and  $\lambda_-$  respectively. Suppose a forward step takes place if the  $\lambda_+$  reaction takes place and a backward step if  $\lambda_-$  takes place. Then one may suppose that the turnover time  $X_i$  may be represented as

$$X_i = \begin{cases} \text{Exp}[\lambda_+] & \text{with probability } p \\ \text{Exp}[\lambda_-] & \text{with probability } q \end{cases} \quad (56)$$

where  $\text{Exp}[\lambda]$  is the exponential distribution with rate  $\lambda$ . Associate with this a reward increment  $H_i$  which takes the values 1 and  $-1$  according as whether a forward or backward step has taken place. Then,

$$\begin{aligned} & \left( \mathbb{P}\{X_i \leq t \cap H_i = 1\} = p(1 - e^{-\lambda_+ t}) \right) \\ & \neq \left( \mathbb{P}\{H = 1\} \mathbb{P}\{X \leq t\} = p(p(1 - e^{-\lambda_+ t}) + q(1 - e^{-\lambda_- t})) \right) \end{aligned} \quad (57)$$

except for a few special values of  $p, q$  and  $t$ . If however, a single step enzymatic cycle based on a Markov chain is constructed,



the arguments in Section A.1 may be used to show  $H_i$  and  $X_i$  are independent. Indeed, one may directly observe that

$$\mathbb{P}\{X_i \leq t \cap H_i = \pm 1\} = \frac{\lambda_{\pm}}{\lambda_- + \lambda_+} (1 - e^{-(\lambda_+ + \lambda_-)t}) = \mathbb{P}\{X_i \leq t\} \mathbb{P}\{H_i = \pm 1\}. \quad (58)$$

If this independence does not hold, (13) cannot be used to derive formulas for the cumulants; the more complicated expression in (17) must be used to take the correlation of  $H_i$  and  $X_i$  into account. Formulas for the higher cumulants do not appear to be available.

- Renewal reward process are applicable when motion takes place in one dimension, where the regenerative structure is more obvious. Multi-dimensional effects like those in (Block, 2003) are harder to capture.

The analysis based on renewal-reward processes must be supplemented with usual methods of Markov chain analysis to overcome these drawbacks. On the other hand, renewal reward processes are easily adapted to handle other periodic situations (time and space) where regenerative structure in the problem can be exploited. To elaborate,

- If some variable in a bead-motor assay fluctuates *spatially* with a period equal to some rational multiple of the binding-site separation, the regenerative structure of the problem can still be exploited (in theory). For example, consider the PO model with an external force of the form  $f(x) = A\sin(x\pi/2L) + f_0$ , where  $f_0$  is a constant and  $L$  is the distance between binding sites. This force affects only the diffusion and the probability of binding forwards; it doesn't change the cycle time. By the force's periodicity, there must be two probabilities of binding forwards  $p_0, p_1$ . These can be calculated using the methods in Peskin and Oster (1995).

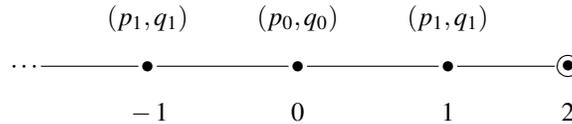


Figure 8: Simple periodic random walk. State 2 is an absorbing barrier.

Although a renewal-reward process formulation can be constructed, it is easier to just consider the simple random walk with an absorbing barrier at state 2 as shown in Fig. 8. Then, the first passage time to the absorbing barrier gives us the renewal increments  $X_i$ . It needs to be ensured that the renewal takes place in a finite number of steps with probability 1. Absorption into state 2, given the process starts in state 1, is certain if

$$p_0 p_1 \geq q_0 q_1.$$

The calculation is based on a simple extension of the “single absorbing barrier” problem in a simple random walk Cox and Miller (1977). Some details are given in Section A.4.

- This involves the definition of an initial-time dependent renewal increment and renewal function. The methods in Prager et al. (2005) are well suited to this purpose.

To conclude, renewal-reward processes are used in this paper to analyze the bead-displacement time-trace in single molecule bead-motor assays of motor proteins. The approximations involved in reducing this to such a process are justified by appealing to the physics at the length and time-scales in question. General insight about the influence of substeps and initial conditions is given. When the renewal increments (the cycle time of the enzyme) and the reward increment (the associated mechanical step) are generated by a single Markov model, formulas for calculating the cumulants to arbitrary order is given. Then, a method of fitting simple models to experimental data using the cumulants of the bead-displacement time-trace is presented. This approach is shown to yield more information about the underlying chemistry than previously used statistical measures like like the randomness parameter. Its robustness in the presence of error is demonstrated.

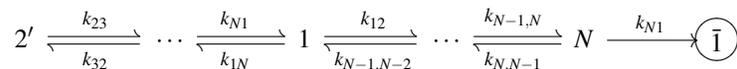


Figure 9: Section of Fig. 1 showing the forward path alone. The rate  $k_{21}$  has been set to zero to prevent absorption from state  $2'$  into state  $\bar{1}$ ; this would result in the completion of a backward cycle. Solving a first passage problem on this Markov chain, given that the process starts in state 1, helps one find  $F_{T_f}(t)$ , the distribution function of  $T_f$ .

## A Appendix

### A.1 Path Formulation of Chemical Cycles

The result discussed in this subsection is not essential to the thesis of this paper and appears to be a special case of a more general result in Markov chains. The proof is worth repeating due to its simplicity in the case of cyclic chemical schemes. For the purposes of this paper, it is not necessary to consider the problem in all its generality.

Consider the chemical scheme described in Section 2.2. The first-passage problem formulated in Fig. 1 is called a  $1 \rightarrow \bar{1}$  form of the problem, for obvious reasons. The objective is to show that the first passage time is identical, whether the chemical cycle is formulated as  $1 \rightarrow \bar{1}$  problem, or as a  $n \rightarrow \bar{n}$  problem. To show this, it is first shown that the first-passage time can be found by conditioning on the particular path a system takes. This is a rearrangement of the Chapman-Kolmogorov equations for Markov chains; nevertheless, it's interesting to proceed purely from first-principles.

Absorption through the forward path and absorption through the backward path are mutually exclusive events. Then, if  $T$  is the first-passage time in the  $1 \rightarrow \bar{1}$  problem, we can write  $\mathbf{P}\{T \leq t\} = \mathbf{P}\{T_f \leq t\} + \mathbf{P}\{T_b \leq t\}$ , where  $T_x$  is the first-passage time through the forward or backward path for  $x = f, b$  respectively. Then, restrict attention to the forward half of Fig. 1 with  $N$  states as shown in Fig. 1. Let  $F_{T_f}(t)$  be the distribution function of the first passage time through the forward path alone. Let  $r_{i,n}$  represent some path of length  $n$ , where the index  $i$  is used to distinguish between paths of the same length. It is clear that the path can be represented by a sequence of rate constants  $u_i = k_{mn}$  as

$$r_{i,n} \equiv \{u_1, u_2, \dots, u_n\}, \quad (59)$$

and the probability that path  $r_{i,n}$  is chosen can be written as

$$\mathbf{P}\{r_{i,n}\} = \prod_{p=1}^n \frac{k_{i_p j_p}}{\sum_{q=1}^n k_{i_p q}}. \quad (60)$$

Let  $F_{T_{i,n}}$  represent the density of the first passage time, given that absorption took place through path  $r_{i,n}$ . Note that once a particular sequence of reactions; i.e., a path is specified in this reaction, the conditional first passage time is just a sum of exponentials. It seems intuitive that  $F_{T_f}(t)$  can be conditioned on the path taken to absorption, and written in terms of the  $F_{T_{i,n}}$ .

**Lemma 1.** *The series,*

$$F_{T_f}(t) = \sum_{n=N}^{\infty} \sum_i \mathbf{P}\{r_{i,n}\} F_{T_{i,n}}(t) \quad (61)$$

*is convergent (pointwise in  $[0, \infty)$ ).*

*Proof.* The proof has two parts: the first estimates the number of possible paths to absorption, and the second bounds the CDF of the first-passage time given it takes a path of a certain size. Given these two observations, the inner sum in (61) can be bounded and the outer sum can be shown to be convergent.

Notice first that when all the reverse rates in Fig. 9 are zero, there is only one path to absorption starting from state 1 and that path has length  $N$ . If one reverse reaction is allowed, additional paths of length  $N + 2r$ ,  $r \in \mathbb{N}$  become available. For example, the process can step back once along the reverse reaction, forward again and continue inexorably forwards to absorption to form a path of length  $N + 2$ . Each time the process “doubles back” on itself, it chooses one of the  $k$  reverse reactions. Combinatorially, there are  $k^r$  paths of length  $N + 2r$ . Then, the number of paths of length  $M$  (denoted by  $l_M$ )

$$l_M = \begin{cases} k^r & M = N + 2r \\ 0 & \text{otherwise} \end{cases} \quad \text{for } r = 0, 1, \dots \quad (62)$$

Second, consider some path of length  $M$ , with rate constants  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ . Then the probability density,  $F'_{T_i, n}$ , as mentioned earlier, is the convolution of  $M$  exponential distributions. Let these  $M$  exponentially distributed random variables have densities and distributions  $\rho(\lambda_i, t)$  and  $p(\lambda_i, t)$  respectively, and let their sum have density and distribution  $\rho(t)$  and  $p(t)$ . Since the convolution is a product under a Laplace transform, it follows that

$$\begin{aligned} p^*(s) &= \frac{\rho^*(s)}{s} \\ &= s^{M-1} \prod_{i=1}^M p^*(\lambda_i, s) \\ &\leq s^{M-1} p^*(\lambda_0, s)^M, \end{aligned} \quad (63)$$

where  $\lambda_0$  represents the maximum of the set  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ . The last inequality in (63) follows from the fact that if  $\lambda_1 \leq \lambda_2$ ,  $p^*(\lambda_1, t) \leq p^*(\lambda_2, t)$ . The distribution function of a sum of  $M$  independent, identically distributed, exponential random variables is usually written in terms of the incomplete gamma function as  $\gamma(M, \lambda t)/(M-1)!$ . Then,

$$p(t) \leq \int_0^t x^{M-1} e^{-\lambda_0 t} dt \leq \frac{(\lambda_0 t)^{M-1}}{(M-1)!}. \quad (64)$$

Let  $a$  be the maximum probability of taking a particular reaction given the process is in a state from which multiple paths can be taken. That is

$$a = \sup_{i,j} \left\{ \frac{k_{i,j}}{\sum_{j=1}^N k_{i,j}} \left\| \frac{k_{i,j}}{\sum_{j=1}^N k_{i,j}} \neq 1, i, j \in \{1, \dots, N\} \right\} \right\}. \quad (65)$$

Then using (64), (62) and (65), the series in (61) can be bounded by a geometric series

$$0 \leq \sum_{r=0}^{\infty} \sum_i^{k^r} \mathbf{P}\{r_{i, N+2r}\} F_{T_i, N+2r}(t) \leq \sum_{r=0}^{\infty} k^r a^r \frac{(\lambda_0 t)^{N+2r-1}}{(N+2r-1)!} \quad (66)$$

Using the usual ratio test, the geometric series is pointwise convergent for any  $t \in [0, \infty)$ , and it follows by comparison that the series for  $F_{T_j}(t)$  converges.  $\square$

By conditioning on paths, it is easy to show that the first-passage time is invariant under cyclic permutations of the states and rate constants. Consider now the first-passage problems  $m \rightarrow \bar{m}$  and  $n \rightarrow \bar{n}$ , where  $m \neq n$ . Restrict attention to the forward paths to absorption alone; an identical argument can be applied to the backward paths. Let  $r_{i,l} \equiv u_1, u_2, \dots, u_l$  be a path in the  $m \rightarrow \bar{m}$  problem. Clearly,  $u_1 = k_{mj}$ , and  $u_l = k_{m-1m}$ , where  $j$  is arbitrary. Let  $u_k$  be the first rate constant in  $r_{i,l}$  of the form  $k_{n-1n}$ . Since it is the forward path, this  $k$  must exist, and  $k \neq l$ . Then, construct the path  $\bar{r}_{i,n} \equiv \{u_{k+1}, \dots, u_N, u_1, \dots, u_k\}$ . This is a valid path in the  $n \rightarrow \bar{n}$ . This transformation is clearly one-one and implies that there is a one-to-one correspondence between the terms in the series in (61) for the  $m \rightarrow \bar{m}$  and  $n \rightarrow \bar{n}$ . It's clear that the two first-passage times are identical. It has recently been brought to our attention that these and other issues are addressed in greater generality and detail by Wang and Qian (2007).

Another critical observation that follows from this path decomposition is that the probability of absorption into a particular absorbing state, given that absorption has taken place, just the probability of taking one of the possible paths to that absorbing state. This probability is just the sum of the probabilities of taking each individual path, since these are mutually exclusive. From (60), it is clear that this probability is a constant independent of the time of absorption. Hence any reward increment  $H_i$  whose value depends only on which state the process was absorbed into is dependent only on that fact that absorption has occurred at some finite time, and not on the time itself. Since absorption takes place with probability 1 in a finite time, this justifies the assumption that the  $(H_i, X_i)$  pairs are independent (also see Section 3.3).

## A.2 Formulas for Cumulants of a General Renewal Process

To find formulas for the constants  $a_i$  and  $b_i$  in (15) Smith (1959) finds the cumulants  $\psi_n(t)$  of a function of  $N(t)$  which has a moment generating function given by

$$\Phi(\eta) = E \left[ \frac{1}{(1-\eta)^{N(t)+1}} \right].$$

Then,  $\psi_n(t)$  has the form

$$\psi_i(t) = \alpha_i t + \beta_i + \lambda(t). \quad (67)$$

where  $\lambda(t) \rightarrow 0$  as  $t \rightarrow \infty$  as described in Section 2.1.2.

Let  $\mu_n$  be the  $n^{\text{th}}$  moment of the renewal increments  $X_i$ , and let its distribution function be  $F$ . For the first  $n$  moments of  $X_i$  to be finite, a necessary and sufficient condition, is that there exists another distribution  $F^{(k)}$  such that

$$F^*(s) = 1 - \mu_1 s + \frac{\mu_2}{2!} s^2 - \dots + \frac{\mu_{n-1}}{(n-1)!} (-s)^{(n-1)} + \frac{\mu_n}{n!} F^{(k)*}(s). \quad (68)$$

If  $z_1(\eta)$  is the solution of  $F^*(s) = 1 - \eta$  such that  $z_1(0) = 0$ ,  $\alpha_i$  can be written as

$$\alpha_i = \frac{n!}{2\pi i} \oint_{C'} \frac{z_1(\eta)}{\eta^{n+1}} d\eta, \quad (69)$$

and the constants  $a_n$  can be written in terms of  $\alpha_i$  as

$$\begin{aligned} a_1 &= \alpha_1, \\ a_2 &= \alpha_2 - \alpha_1, \\ a_3 &= \alpha_3 - 3\alpha_2 + \alpha_1, \\ a_4 &= \alpha_4 - 8i\alpha_3 + 7\alpha_2 + \alpha_1. \end{aligned} \quad (70)$$

The first eight values for  $\alpha_s$  are given in (Smith, 1959), but  $\mu_1$  is set to 1. For the formulas to hold, the appropriate moments of  $F$  have to be finite, and hence  $F$  has a representation as in (68). Then, setting  $\mu_1 = 1$  is equivalent to a coordinate transformation  $\bar{s} = s\mu_1$  in (68). Let  $\bar{F}(\bar{s}) = F(s)$ , and let the quantities associated with  $\bar{F}(\bar{s})$  be  $\bar{\alpha}_i$ ,  $\bar{z}_1$ , and  $\bar{\mu}_i$  for  $i > 1$ . Then, it follows that

$$\begin{aligned} \bar{z}_i(\eta) &= z_i(\eta) \\ \mu_i &= \frac{\bar{\mu}_i}{\mu_1^i} \text{ for } i > 1, \\ \bar{\alpha}_i &= \frac{\alpha_i}{\mu_1}. \end{aligned} \quad (71)$$

The first four formulas for the  $\alpha_i$  from Smith (1959) become

$$\begin{aligned}\alpha_1 &= \frac{1}{\mu_1}, \\ \alpha_2 &= \frac{\mu_2}{\mu_1^3}, \\ \alpha_3 &= \frac{-\mu^3}{\mu_1^4} + 3\frac{\mu_2^2}{\mu_1^5}, \\ \alpha_4 &= \frac{\mu_4}{\mu_1^5} - 10\frac{\mu_2\mu_3}{\mu_1^6} + 15\frac{\mu_2^3}{\mu_1^7}.\end{aligned}\tag{72}$$

### A.3 Cumulants of a Renewal Process with an Erlang Distributed Increment

Let  $N(t)$  be the renewal process with renewal increment  $X_i$  that corresponds to a single exponential variable with rate  $\lambda$  - this is the well understood Poisson process (Ross, 1983). The moment generating functions of the Poisson process  $\phi_N(z)$  is given by,

$$\phi_N(z) = \sum_{i=0}^{\infty} \exp(-\lambda t(e^z - 1)).\tag{73}$$

Since the cumulant generating function  $\psi_N(z) = \log(\phi_N(z))$ , the slopes of the cumulants  $N(t)$  are all  $\lambda$ . The objective is to find the cumulants of the renewal process  $\bar{N}(t)$  with Erlang distributed increments with parameters  $k$  and  $\lambda$ ; this is a sum of  $k$  exponential random variables with rate  $\lambda$ . Then

$$\bar{N}(t) = \left\lfloor \frac{N(t)}{k} \right\rfloor,\tag{74}$$

where  $\lfloor \cdot \rfloor$  denotes the greatest integer function. Associate a cumulative process  $R(t)$  with  $N(t)$  that has increments  $H_i = 1/k$  and a terminal reward  $\tilde{H}(t) = \lfloor N(t)/k \rfloor - N(t)/k$ . The terminal reward subtracts out the “excess” steps in  $N(t)$ , and the number of renewals in  $\bar{N}(t)$  is the same as the value  $R(t)$ . Prop. 1 states that the terminal reward has no effect on the cumulants of, and hence the terminal reward  $\tilde{H}(t)$  can be dropped from  $R(t)$ . It follows from (10) that  $R(t) = N(t)/k$ , and using a standard property of cumulants, one obtains

$$\kappa_{R,n}(t) = \frac{\kappa_{\bar{N},n}(t)}{k^n},\tag{75}$$

### A.4 Periodic Simple Random Walk

This is based on a standard method (Cox and Miller, 1977) for a simple random walk in one-dimension with one absorbing barrier at  $a > 0$ . For the periodic problem in Section 5, Fig. 8, it is noted that if the process begins in state 0, it will be absorbed into state 2 if it is first absorbed into state 1 and subsequently into state 2. Since the steps of the random walk are independent, the first passage time into state 2, given the process starts in state 0 is the sum of the first passage times from  $0 \rightarrow 1$  and  $1 \rightarrow 2$ . More formally, let  $X_n$  be an integer valued random variable that represents the position of the random walk after  $n$  steps, and suppose that  $X_0 = 0$ . Let,

$$f_n = \mathbb{P}\{X_m < 1 (m = 1, \dots, n-1), X_n = 1\},$$

and let

$$F(s) = \sum_{i=0}^{\infty} f_i s^i\tag{76}$$

be its probability generating function. Let the random variable  $N_f$  take values in  $\mathbb{N}$  and represent the probability of absorption in that many steps. If  $A_{01}$  represents the event that absorption into 1 takes place, given the process starts at 0,

$$F(s) = \mathbb{P}\{A_{01}\}E[N_f|A_{01}].$$

Similarly,  $g_n$ ,  $G(s)$ ,  $N_g$ , and  $A_{12}$  be the corresponding variables for absorption into state 2 given the process starts in state 1 ( $X_0 = 1$ ). It is clear that the probability generating functional for the first passage time into state 2, given the process starts in state 0 is just  $F(s)G(s)$ . Conditioning on the first step, the following simultaneous equations for  $F(s)$  and  $G(s)$  can be obtained:

$$F(s) = p_0s + (1 - p_0 - q_0)F(s) + q_0sF(s)G(s) \quad (77)$$

$$G(s) = p_1s + (1 - p_1 - q_1)G(s) + q_1sF(s)G(s). \quad (78)$$

These equations can be solved by eliminating (say)  $G(s)$  from the first equation using the second, and then solving a quadratic for  $F(s)$ . One root can be discarded by noticing that since  $F(s)$  represents functions of the form (76), they must be well behaved as  $s \rightarrow 0$ ; one of the roots goes to infinity and the other goes to zero. For the renewal increment defined as the first passage time to state 2 given  $X_0 = 1$  to be finite with probability 1, it is clear that  $F(1) = \mathbf{P}\{A_{01}\} = 1$  and  $G(1) = \mathbf{P}\{A_{12}\} = 1$  are needed. Setting  $s = 1$  in the well-behaved roots of  $F(s)$  and  $G(s)$ , we obtain

$$F(s) = \frac{p_0p_1 + q_0q_1 + 2p_1q_0 - \sqrt{(p_0p_1 - q_0q_1)^2}}{2p_1q_0 + 2q_0q_1} \quad (79)$$

$$G(s) = \frac{p_0p_1 + q_0q_1 + 2p_0q_1 - \sqrt{(p_0p_1 - q_0q_1)^2}}{2p_0q_1 + 2q_0q_1} \quad (80)$$

, which both result in the requirement that  $p_0p_1 \geq q_0q_1$ . If  $T$  is the random variable representing the cycle time of the enzyme, the renewal increment is

$$X_i = \sum_{i=1}^N T,$$

where  $N = N_f + N_g$ . The moments and cumulants of  $X_i$  can be determined using a generating function like in (13), and the cumulants of the renewal process constructed using the  $X_i$  can be obtained from the equations in Section A.2.

## References

- R.D. Astumian and P. Haenggi. Brownian motors. *Physics Today*, 55(11):33–39, 2002.
- P.J. Atzberger and C.S. Peskin. A Brownian Dynamics Model of Kinesin in Three Dimensions Incorporating the Force-Extension Profile of the Coiled-Coil Cargo Tether. *Bulletin of Mathematical Biology*, 68(1):131–160, 2006.
- M. Bier. Brownian ratchets in physics and biology. *Contemporary Physics*, 38(6):371–379, 1997.
- S.M. Block. Probing the kinesin reaction cycle with a 2 D optical force clamp. *Proceedings of the National Academy of Sciences*, 100(5):2351–2356, 2003.
- S.M. Block. Kinesin Motor Mechanics: Binding, Stepping, Tracking, Gating, and Limping. *Biophysical Journal*, 92(9):2986, 2007.
- P.D. Boyer. The ATP synthase—a splendid molecular machine. *Annual review of biochemistry*, 66(1):717–749, 1997.
- N.J. Carter and R.A. Cross. Mechanics of the kinesin step. *Nature*, 435(7040):308–312, 2005.
- E. Cinlar. Introduction to Stochastic Processes. *Englewood Cliffs*, 1975.
- D.R. Cox and H.D. Miller. *The Theory of Stochastic Processes*. Chapman & Hall/CRC, 1977.
- R.A. Cross. The kinetic mechanism of kinesin. *Trends in Biochemical Sciences*, 29(6):301–309, 2004.
- T.C. Elston. A macroscopic description of biomolecular transport. *Journal of Mathematical Biology*, 41(3):189–206, 2000.
- W. Feller. An Introduction to Probability Theory and Applications, Vol. 2. *New York: Wiley*, 4:153–170, 1968.
- M.E. Fisher and A.B. Kolomeisky. Simple mechanochemistry describes the dynamics of kinesin molecules. *Proceedings of the National Academy of Sciences*, page 141080498, 2001.
- S.P. Gilbert, M.R. Webb, M. Brune, and K.A. Johnson. Pathway of processive ATP hydrolysis by kinesin. *Nature*, 373(6516):671, 1995.
- G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, USA, 2001.
- N.R. Guydosh and S.M. Block. Backsteps induced by nucleotide analogs suggest the front head of kinesin is gated by strain. *Proceedings of the National Academy of Sciences*, 103(21):8054–8059, 2006.
- J. Howard and RL Clark. Mechanics of motor proteins and the cytoskeleton. *Applied Mechanics Reviews*, 55:B39, 2002.
- W.P. Johnson. The curious history of Faà di Bruno’s formula. *American Mathematical Monthly*, pages 217–234, 2002.
- S. Karlin and H.M. Taylor. *A First Course in Stochastic Processes*. Academic Press, 1975.
- A. Krishnan. The Random Walker: Stochastic Mechano-Chemical Models for Motor Proteins. *Masters Thesis, Mechanical Engineering, University of Michigan*, 2008.
- A. Mogilner, A.J. Fisher, and R.J. Baskin. Structural Changes in the Neck Linker of Kinesin Explain the Load Dependence of the Motor’s Mechanical Cycle. *Journal of Theoretical Biology*, 211(2):143–157, 2001.

- C.S. Peskin and G. Oster. Coordinated hydrolysis explains the mechanical behavior of kinesin. *Biophysical Journal*, 68(4 Suppl), 1995.
- T. Prager, L. Schimansky-Geier, and IM Sokolov. Periodic driving controls random motion of Brownian steppers. *Journal Of Physics: Condensed Matter*, 17(47):3661, 2005.
- E.M. Purcell. Life at low Reynolds number. *American Journal of Physics*, 45(1):3–11, 1977.
- H. Qian and E. L. Elson. Single-molecule enzymology: stochastic Michaelis–Menten kinetics. *Biophysical Chemistry*, 101:565–576, 2002.
- S. Rice, A.W. Lin, D. Safer, C.L. Hart, N. Naber, B.O. Carragher, S.M. Cain, E. Pechatnikova, E.M. Wilson-Kubalek, and M. Whittaker. A structural change in the kinesin motor protein that drives motility. *Nature*, 402:778–784, 1999.
- S.S. Rosenfeld, P.M. Fordyce, G.M. Jefferson, P.H. King, and S.M. Block. Stepping and Stretching: How Kinesin Uses Internal Strain to Walk Processively. *Journal of Biological Chemistry*, 278(20):18550–18556, 2003.
- S.M. Ross. *Stochastic Processes [M]*. John Willey & Sons, 1983.
- J.E. Santos, T. Franosch, A. Parmeggiani, and E. Frey. Renewal processes and fluctuation analysis of molecular motor stepping. *Phys. Biol*, 2:207–222, 2005.
- W.L. Smith. Renewal theory and its ramifications. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 243–302, 1958.
- W.L. Smith. On the cumulants of renewal processes. *Biometrika*, 46(1-2):1–29, 1959.
- K. Svoboda, C.F. Schmidt, B.J. Schnapp, and S.M. Block. Direct observation of kinesin stepping by optical trapping interferometry. *Nature*, 365(6448):721–727, 1993.
- D. Tsygankov, M. Lindén, and M.E. Fisher. Back-stepping, hidden substeps, and conditional dwell times in molecular motors. *Arxiv preprint q-bio.BM/0611051*, 2006.
- G.E. Uhlenbeck and L.S. Ornstein. On the Theory of the Brownian Motion. *Physical Review*, 36(5):823–841, 1930.
- R.D. Vale, T. Funatsu, D.W. Pierce, L. Romberg, Y. Harada, and T. Yanagida. Direct observation of single kinesin molecules moving along microtubules. *Nature*, 380(6573):451–453, 1996.
- NG Van Kampen. *Stochastic processes in physics and chemistry*. North-Holland, 2007.
- K. Visscher, MJ Schnitzer, and SM Block. Single kinesin molecules studied with a molecular force clamp. *Nature*, 400(6740):184–9, 1999.
- H. Wang. A new derivation of the randomness parameter. *Journal of Mathematical Physics*, 48:103301, 2007.
- H. Wang and H. Qian. On detailed balance and reversibility of semi-Markov processes and single-molecule enzyme kinetics. *Journal of Mathematical Physics*, 48:013303, 2007.