Biophysical Journal, Volume 99

Supporting Material

Title: Improved hidden Markov models for molecular motors. 1. Basic theory’
Authors: Fiona E Müllner, Sheyum Syed, Paul R Selvin, and Fred J. Sigworth

Supplement to:
Improved hidden Markov models for molecular motors. 1. Basic theory

Appendix. Obtaining rate constants from the transition probabilities

Equation (1) provides a simple way to obtain the rate constants between the molecular states, based on the optimized parameters of the hidden Markov model. This expression is only approximate, and ignores the problems of missed events and multiple steps occurring within the sample interval δt. A more rigorous approach to derive the transition probability matrix from the matrix of rate constants has been suggested by Milescu et al. (20); here we present a similar solution for the inverse problem.

Let the molecular state transition probabilities be given by the matrix **A** having elements

$$a_{ij} = P\left(s_{t+1} = j \,|\, s_t = i\right)$$

giving the probability of a molecular transition from *i* to *j* during the interval (*t*, *t* +1],

and the composite state transition probabilities be given by the matrix **B** having elements

$$b_{iujv} = P\left[(s_{t+1}, x_{t+1}) = (i,u) \,|\, (s_t, x_t) == (j,v)\right]$$

giving the probability of a molecular transition from the composite state (*i*,*u*) to (*j*,*v*) during the interval [*t*, *t* +1). The matrix **B** is related to the matrix of rate constants **Q** for the underlying chemical kinetics according to

$$\mathbf{B} = e^{\mathbf{Q}\delta t}.$$

Meanwhile we obtain from optimizing the HMM the probability of a transition from the composite state (i,u) to $(j,u+w)$,

$$c_{ij}(w) = P(s_{t+1} = j,\ x_{t+1} = u + w \,|\, s_t = i,\ x_t = u),$$

resulting in the step transition probability matrices **C**(*w*). The position change *w* of any molecular transition will be geometrically limited ($w \le W$), letting the **Q** matrix being banded with all entries equal zero for position changes $v\text{-}u > W$. Furthermore, the **Q** matrix is periodic with period *n* (the number of molecular states). Likewise, the transition probability matrix **B** is periodic with period *n* and banded, but with a larger band size. Within one sample interval, the motor can take any number of steps; however, the

probability that the molecule takes s steps within one sample interval decreases exponentially with s. If the sample interval is small relative to the smallest dwell time, the number of steps s is practically limited ($s \leq S$) and the transition probabilities approach zero for transitions involving position changes of size $s \cdot w > S \cdot W$.

A way to calculate the matrix of rate constants **Q** given **B** with minimal errors is to use truncated matrices as in Milescu et al. (20). First, a truncated transition probability matrix $\mathbf{B}^r$ of size greater than $(n \cdot W) \times (n \cdot W)$ is constructed such that it contains the full band of **B**; to minimize errors, B^r should have at least size $(2 \cdot n \cdot W) \times (2 \cdot n \cdot W)$. For example, B^r can be built blockwise from the $\mathbf{C}(w)$ matrices of size $n \times n$. Second, the matrix logarithm of B^r is numerically solved (we use the *logm* function in Matlab). The resulting matrix $Q^r \delta t = \log(B^r)$ contains the full band of the **Q** matrix. Third, the complete **Q** matrix can be reconstructed from one band of the $\mathbf{Q}^r$ matrix which should be taken from the center of the matrix to avoid edge effects. By using this approach, the full matrix of rate constants **Q** can be calculated by truncated matrices of comparably small size.

Since the probability of s steps occurring within a certain time interval δt is correlated with the interval length, a good strategy to keep the bands of **Q** and **A** small is to shorten the sample interval. A reasonable value for δt would be 1/10 of the shortest mean dwell time, which limits the number of steps within δt effectively to $s \leq 3$.

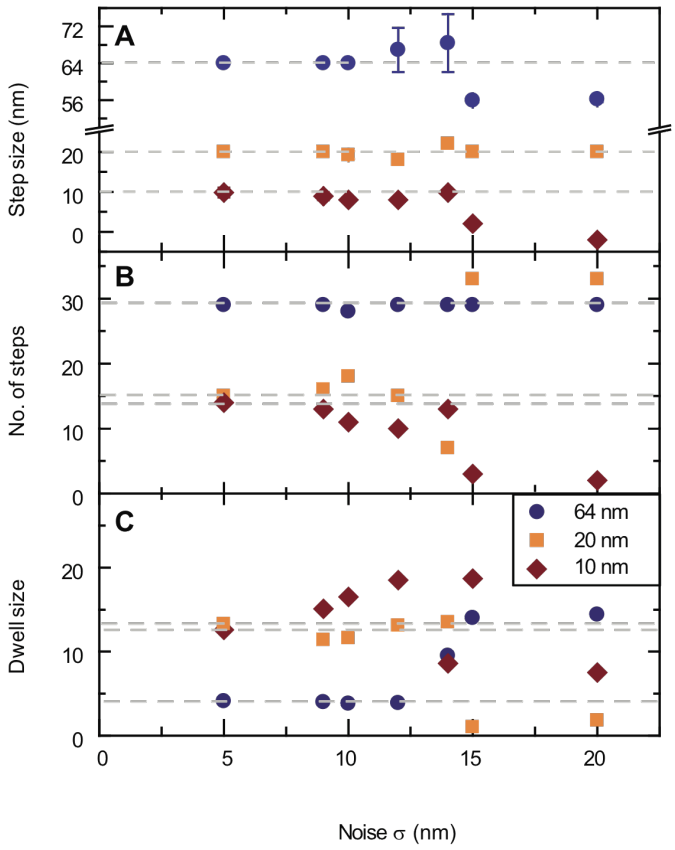

Figure S1

A demonstration of VS-HMM's performance at various noise levels. We generated a number of recordings each similar to that discussed in Fig. 1D (64-20/10 nm steps, with mean dwells of 5 and 10 data) but with the rms noise σ set to different values. For analysis of each trace, an initial HMM with broad c_{12} and c_{21} distributions was created and the algorithm allowed to converge on the most likely model parameters. In this figure we examine the three step sizes (A), frequencies of their occurrence (B), and the associated dwell times (C) as reported by the algorithm for input σ up to 20 nm. The results show that characteristics of the 64 and 20 nm steps are recovered unaltered, with < 20% excursions from their actual values, for σ<14 nm. For larger noise levels, the dwell periods are affected first, followed by deviations in the number of times the HMM algorithm detects these stepping events. Interestingly, the mean step sizes of the two populations remain very close to their true values even when σ = 20 nm. As expected, detection of the smallest steps in the simulations is most susceptible to noise. When the rms noise is about 12 nm, mean size of the 10 nm steps vary by ~20% but the estimated size of their dwell periods exceeds the original by ~50%. As σ surpasses the 14 nm mark, the algorithm is no longer able to recover the 10 nm steps (panel B).