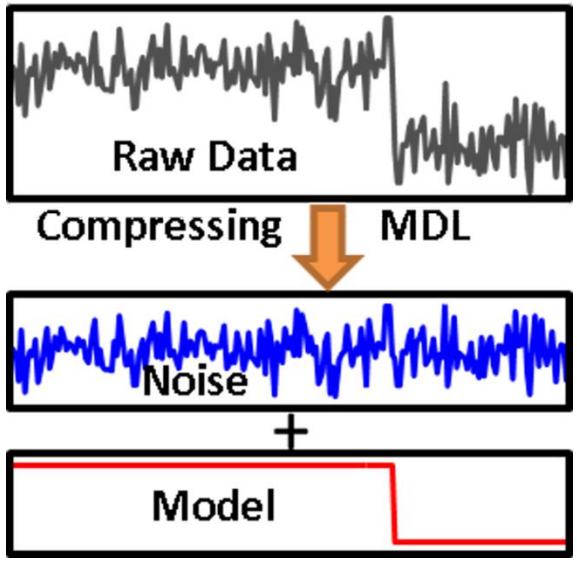
Fast Step Transition and State Identification (STaSI) for Discrete Single-Molecule Data Analysis

Bo Shuang

Landes Research Group

Rice University



QBI

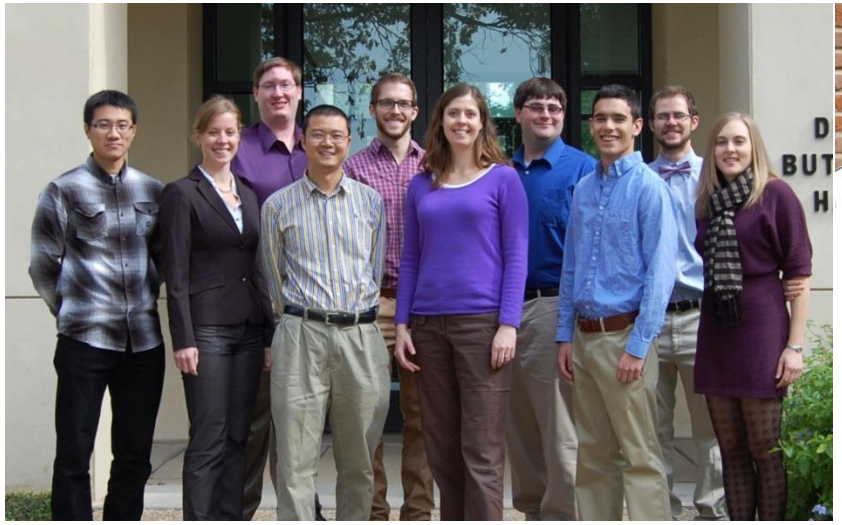
Jan. 08, 2015

Acknowledgements

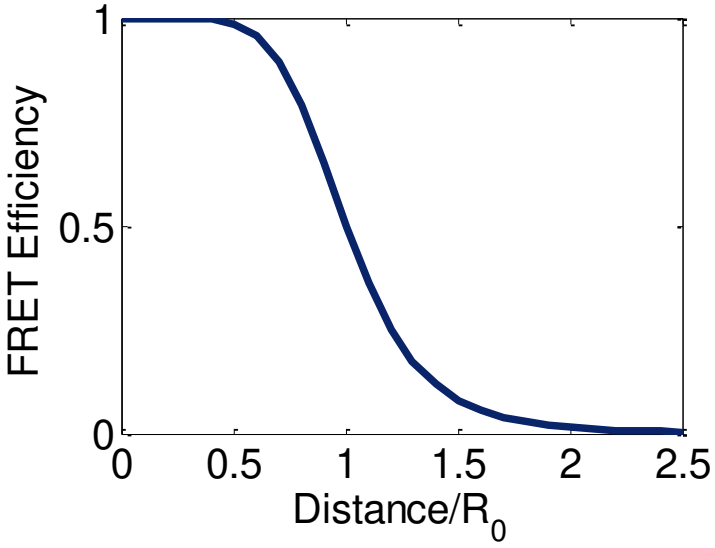
Landes Group
Dr. Christy Landes - PI
Dr. Jixin Chen
Lydia Kisley
David Cooper
Wenxiao Wang

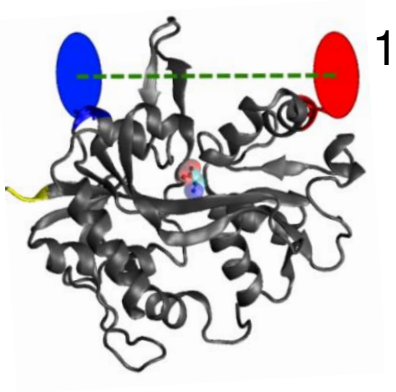
Dr. J. Nick Taylor
Dr. Chun-Biu Li
Dr. Tamiki Komatsuzaki

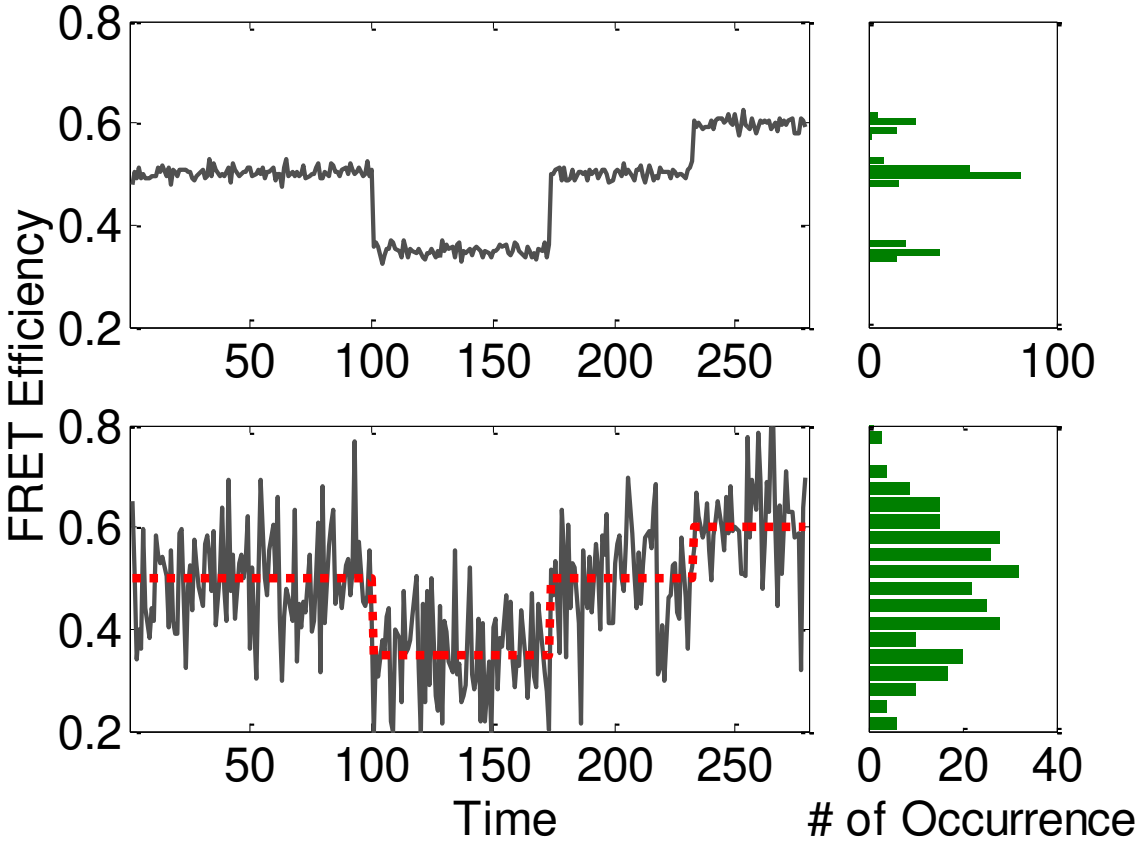
Link Group



Motivation: When Histogram is not Enough



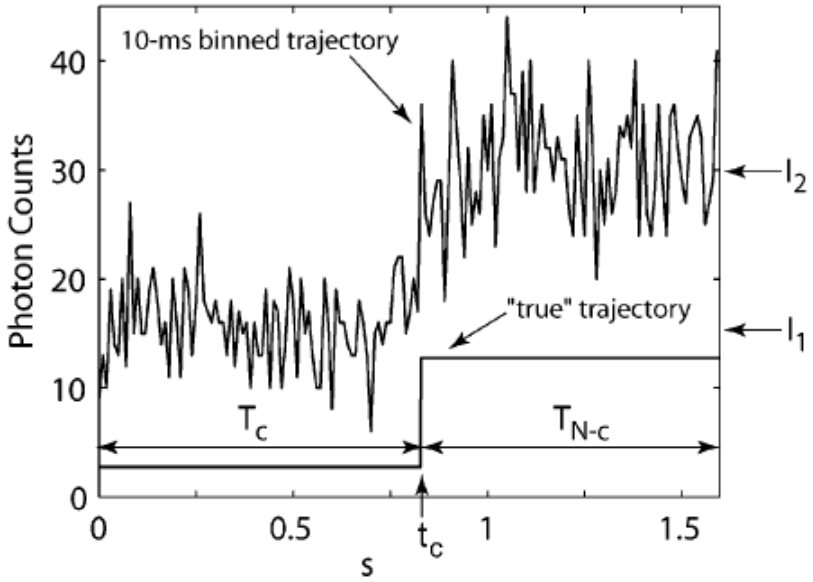




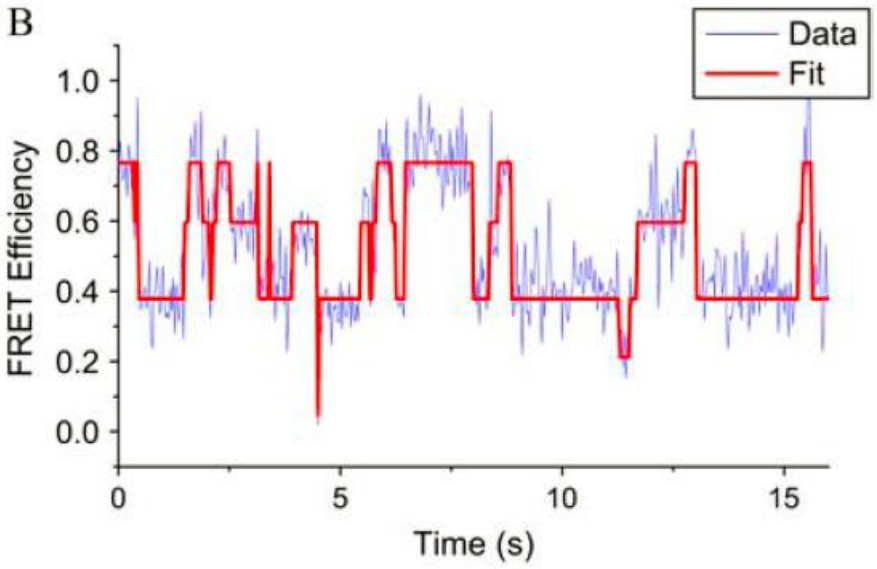
FRET: Förster(fluorescence) resonance energy transfer

1 Cooper, D. et. al. unpublished image

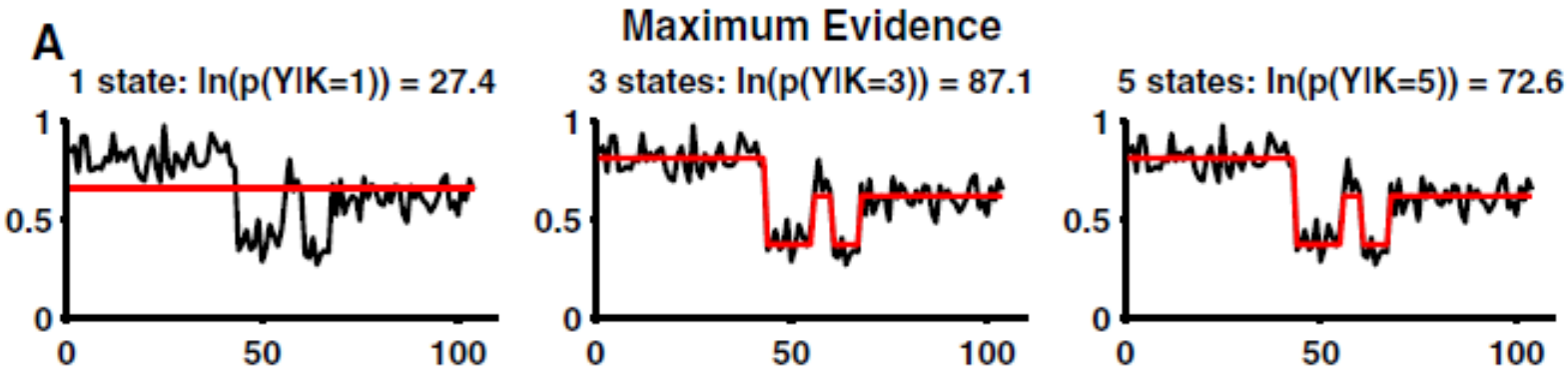
Cutting Edge Methods



Watkins and Yang, *J. Phys. Chem. B* **2005**



McKinney, Joo and Ha, *Biophys. J.* **2006**



Bronson, et. al. *Biophys. J.* **2009**

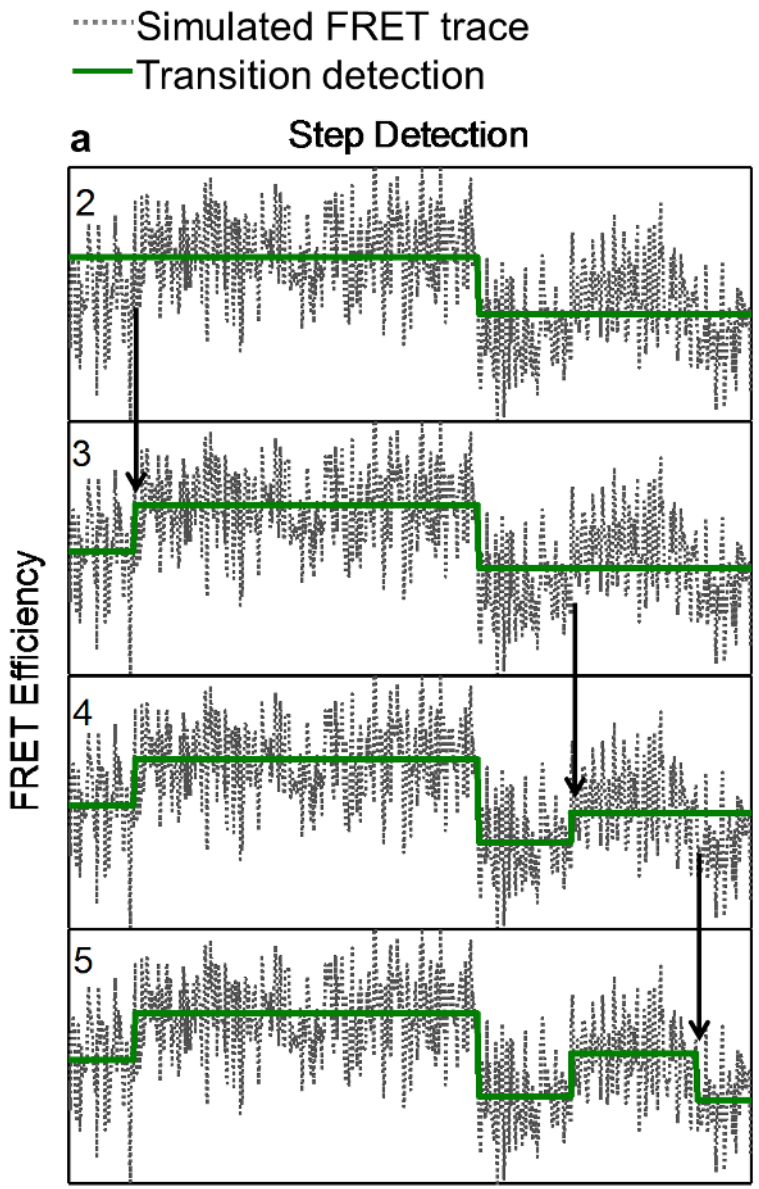
Our Approach

- Targeting binned data
- 1) Fitting of different number of states
  - Student's t test
- 2) Determine the optimum number of states
  - Minimum Description Length Principle (MDL)

Pre-selection: Step Detection and State Identification for Different Number of States

Step 1: Step Detection using Student's *t*-test:

$$R(t_i) = \frac{|I_2(t_{i+1}, t_N) - I_1(t_1, t_i)|}{\sigma\sqrt{\frac{1}{i} + \frac{1}{N-i}}}$$

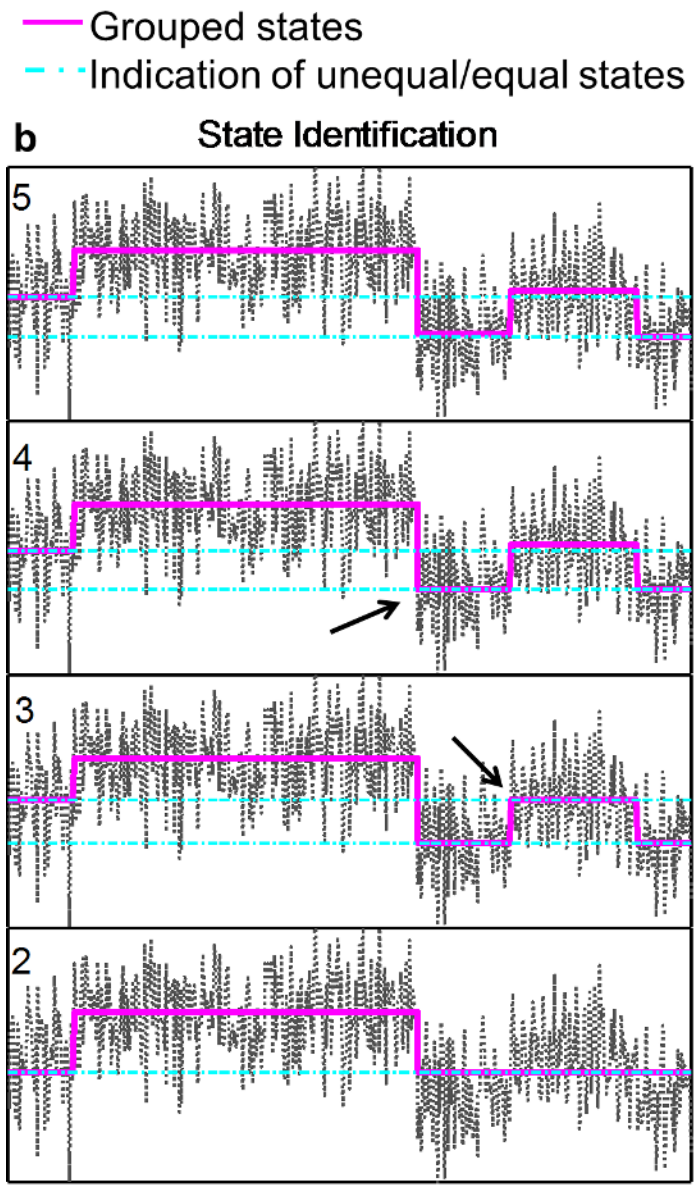


Shuang, B., et. Al., *J. Phys. Chem. Lett.*, **2014**, 5, 3157-3161

Pre-selection: Step Detection and State Identification for Different Number of States

Step 2: Group similar states using a greedy algorithm

$$M(i,j) \propto \left(m_i + m_j\right) * {I_{i,j}}^2 - \left(m_i * I_i^2 + m_j * I_j^2\right)$$



Shuang, B., et. Al., *J. Phys. Chem. Lett.,* **2014**, 5, 3157-3161

Step 3: Determine the Optimum Number of States

$$\mathrm{MDL} = F + G$$

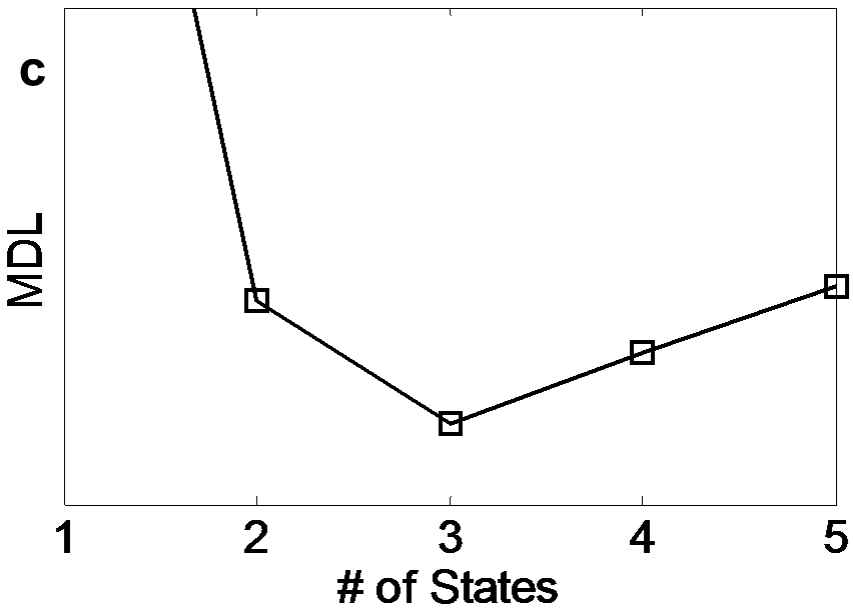
$$F = \frac{\sum_{i=1}^{N} |y(t_i) - y_{fit}(t_i)|}{2\sigma}$$

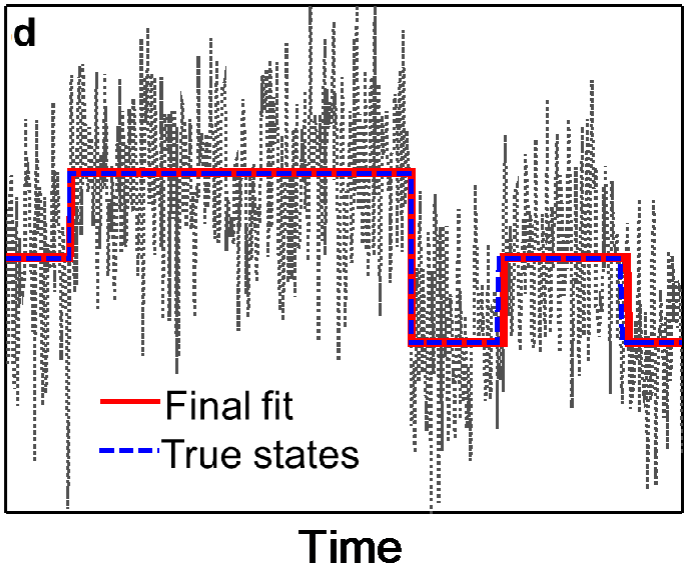
$$G = \frac{k}{2}\ln\frac{1}{2\pi} + k\ln\frac{V}{\sigma} + \frac{N_{\mathrm{tp}}}{2}\ln N + \frac{1}{2}\left(\sum_{i=1}^{k}\ln n_i + \sum_{j=1}^{N_{\mathrm{tp}}}\ln\frac{T_j^2}{\sigma^2}\right)$$

# of states

# of data points in ith state

transitions

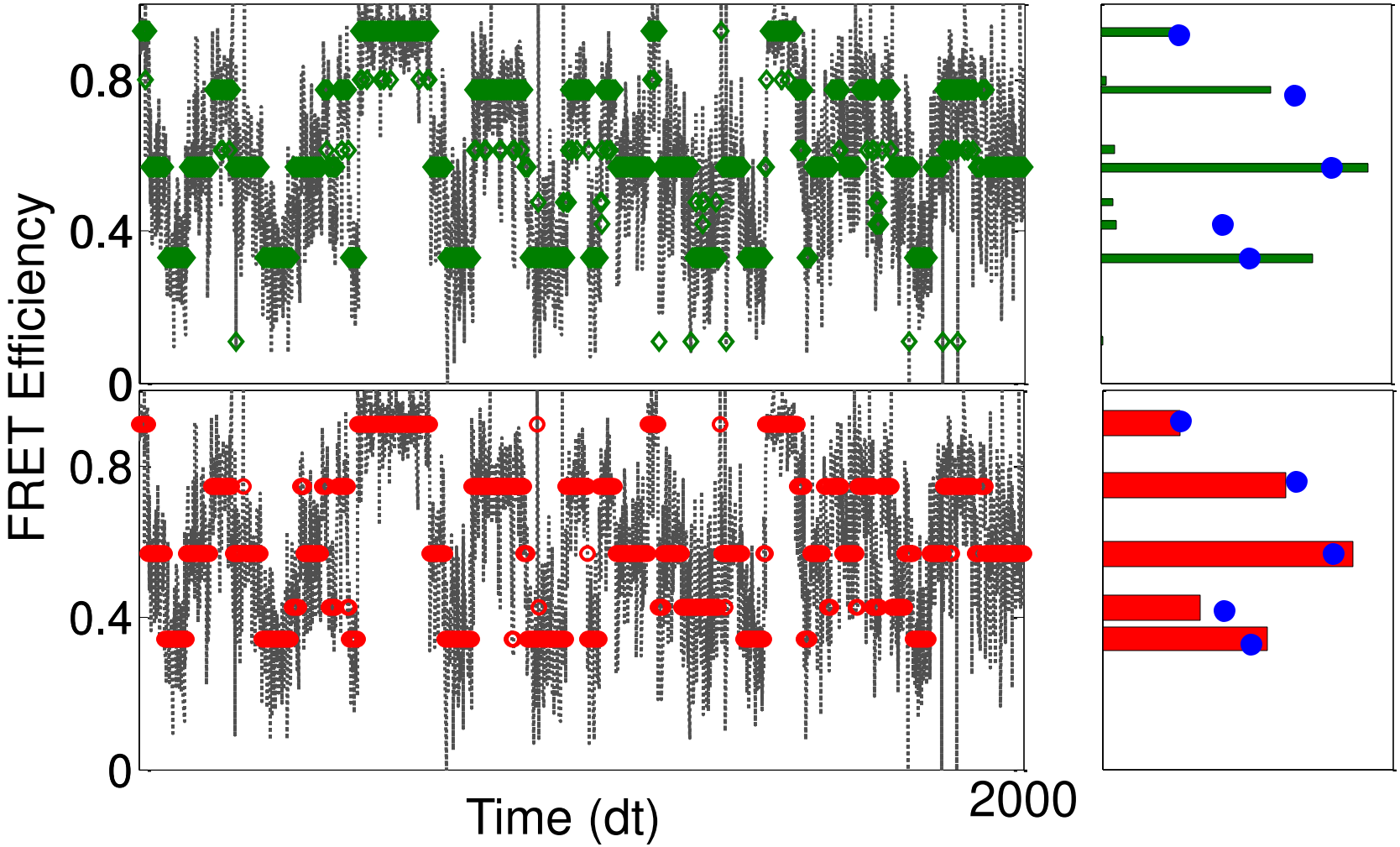




Shuang, B., et. Al., *J. Phys. Chem. Lett.*, **2014**, 5, 3157-3161

Performance: More Accurate

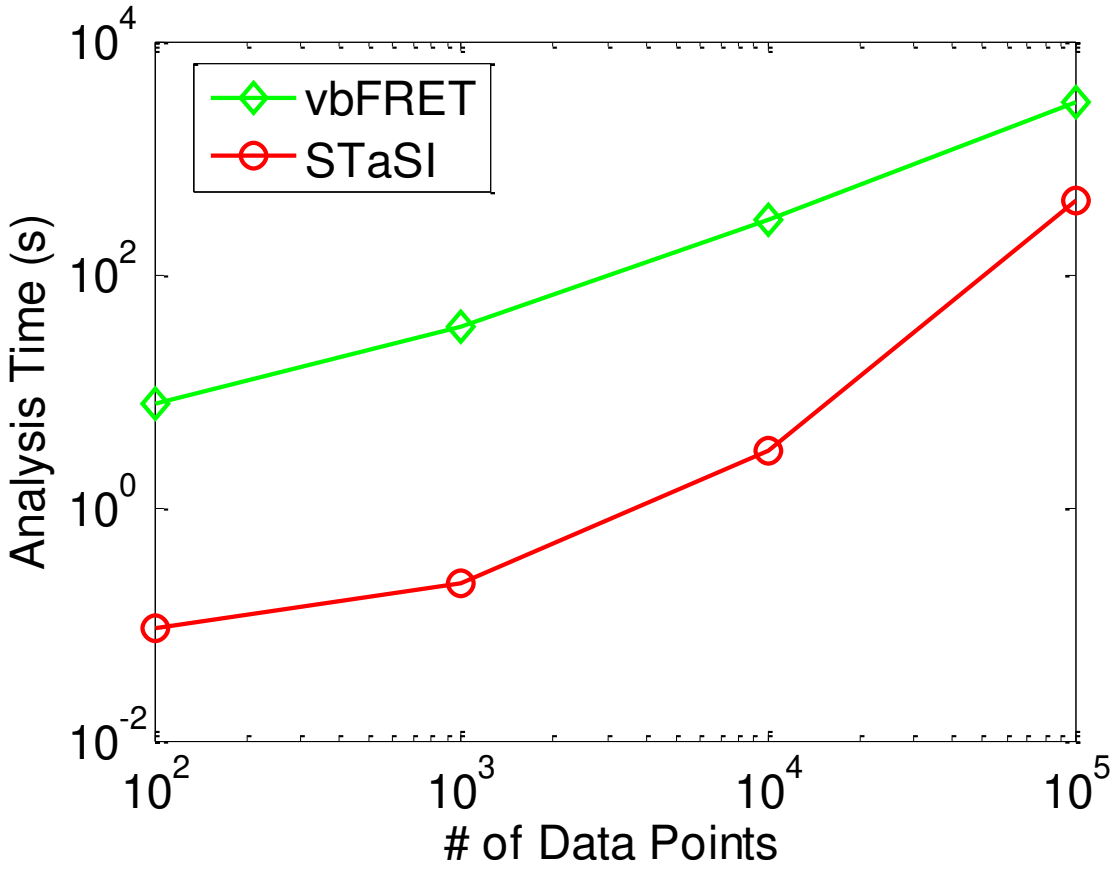
Gray: Simulated FRET trace Red: StaSI fit
Green: vbFRET fit[1] Blue: True states



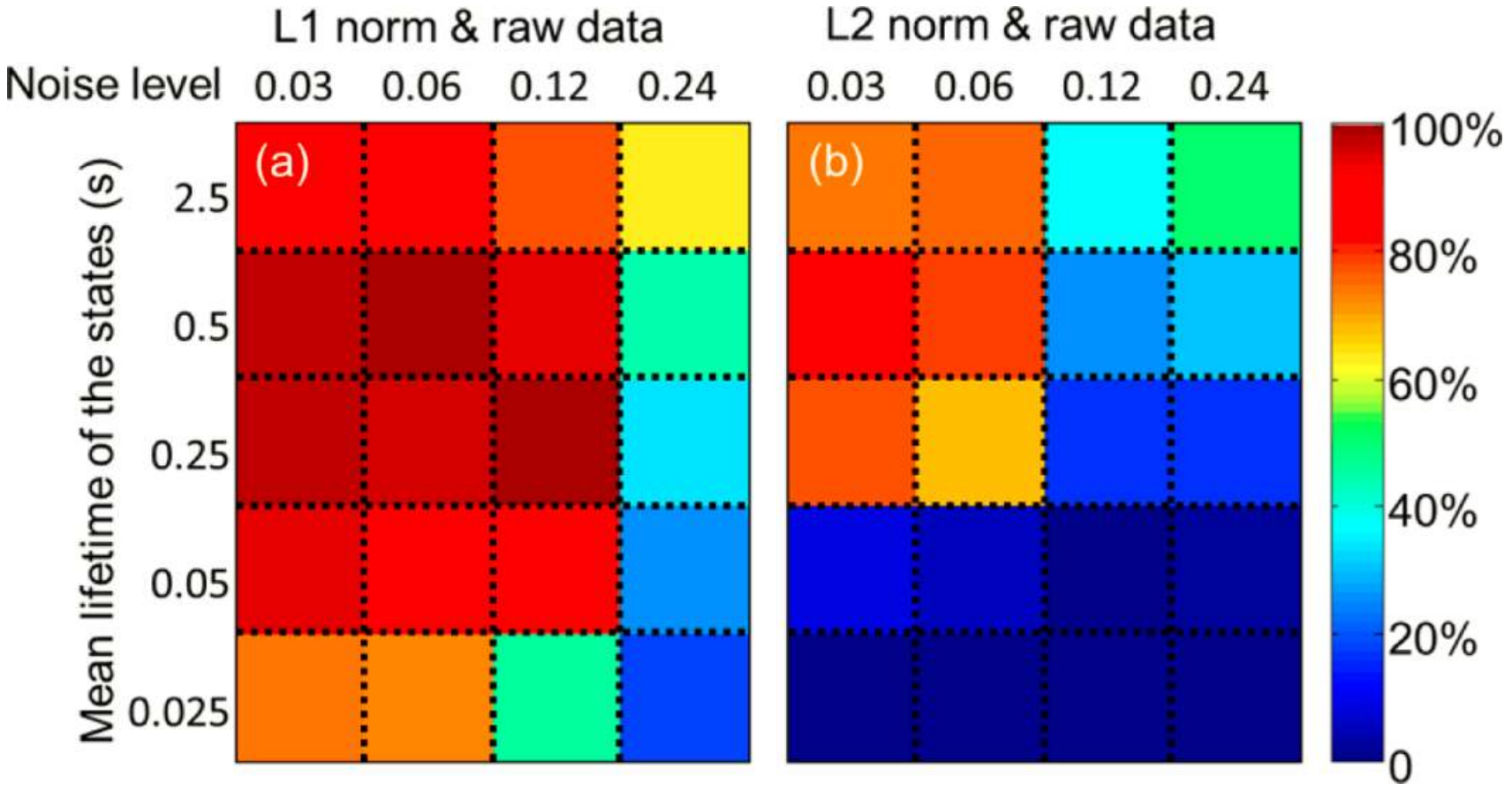
1 Bronson, J. E., et. al. *Biophys. J.* **2009**, *97*, 3196-3205.

RICE

Performance: Faster in Computation



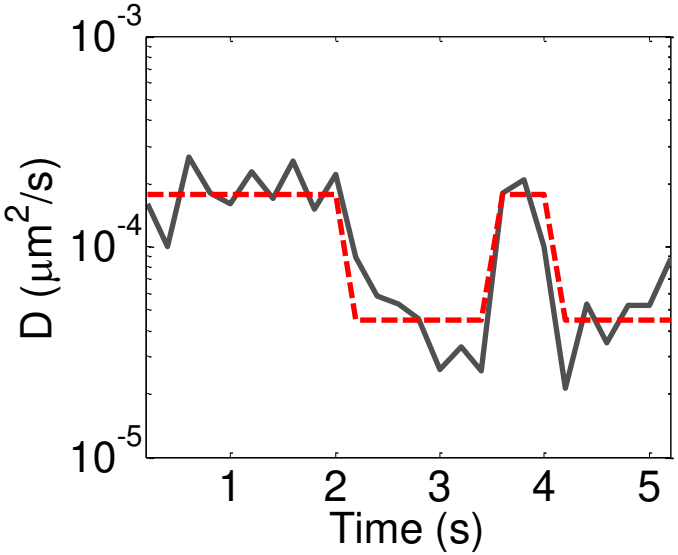
The Importance of L1 norm



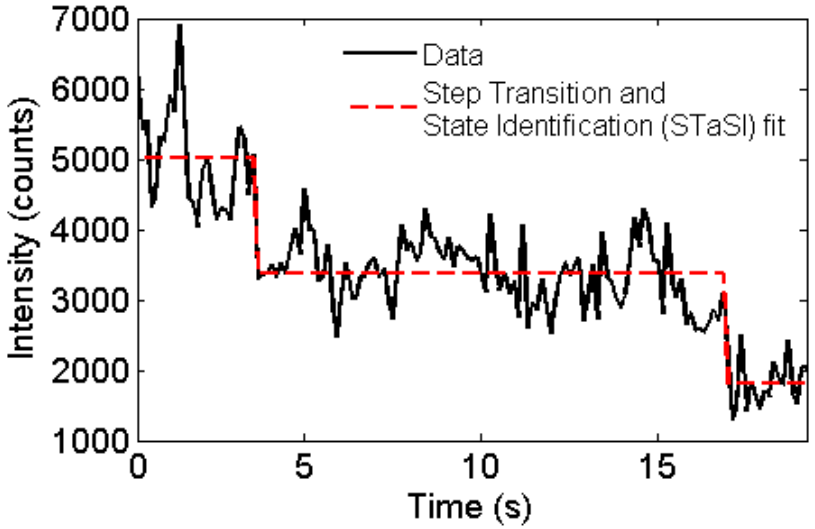
$$F = \frac{\sum_{i=1}^{N} |y(t_i) - y_{fit}(t_i)|}{2\sigma} \qquad F = \frac{\sum_{i=1}^{N} \left(y(t_i) - y_{fit}(t_i)\right)^2}{2\sigma^2}$$

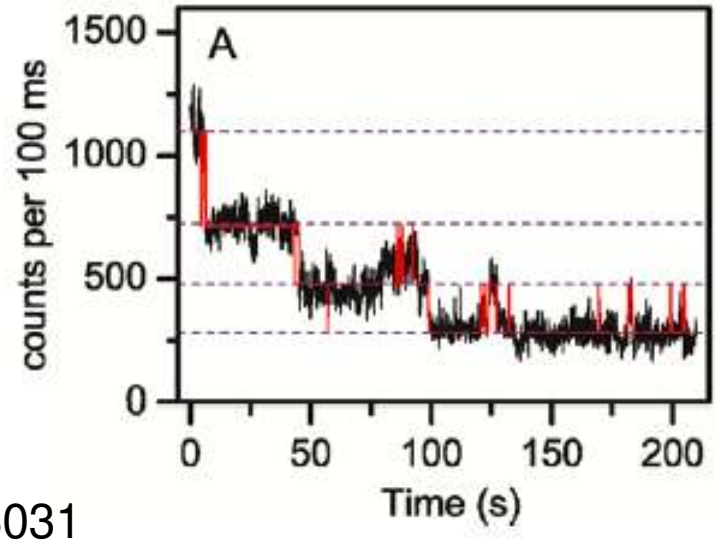
Other Applications

Instantaneous diffusion coefficient



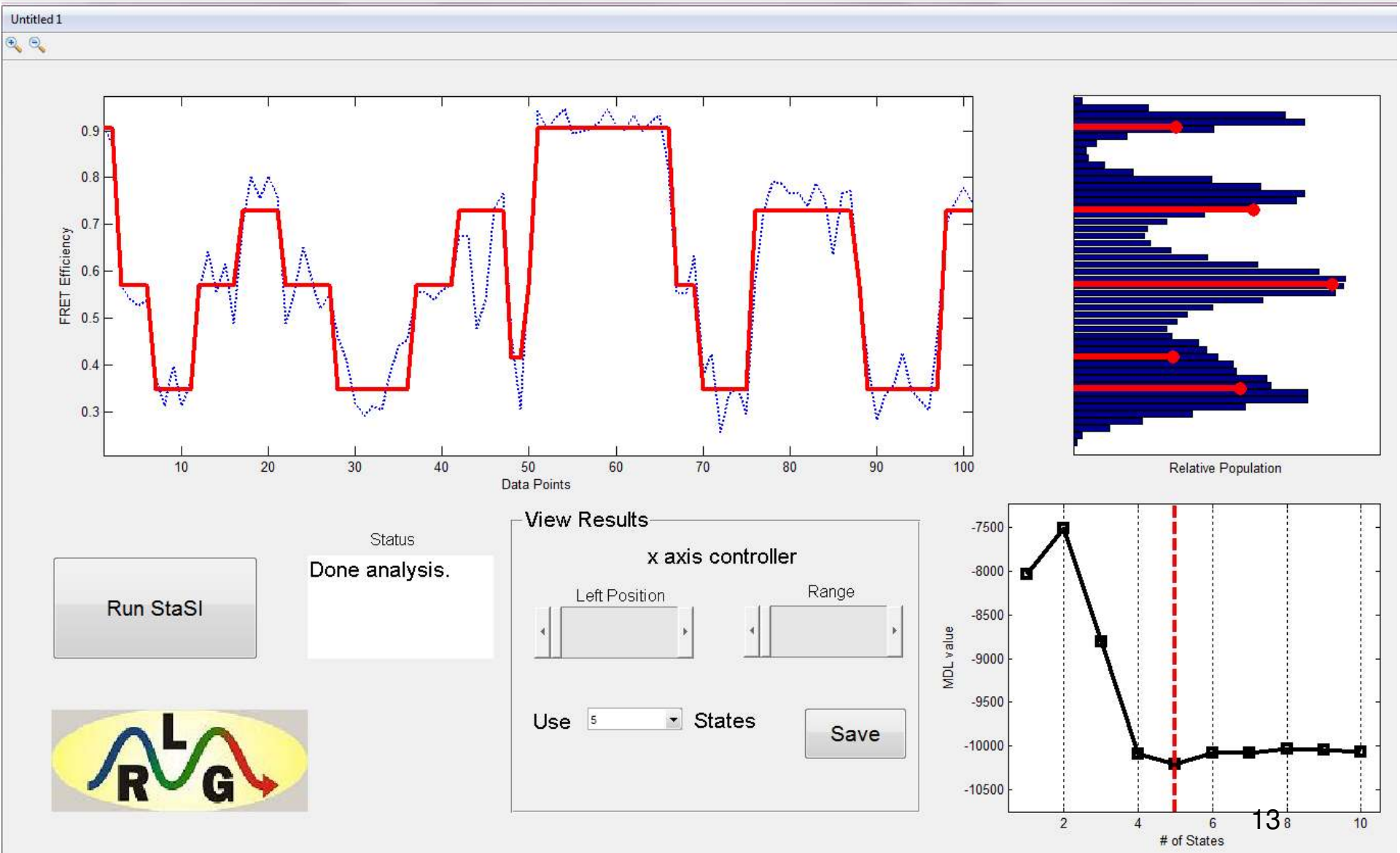
Photobleaching steps





Hu, Z. et. al. *JACS*, **2014**, 136 (45), 16023–16031

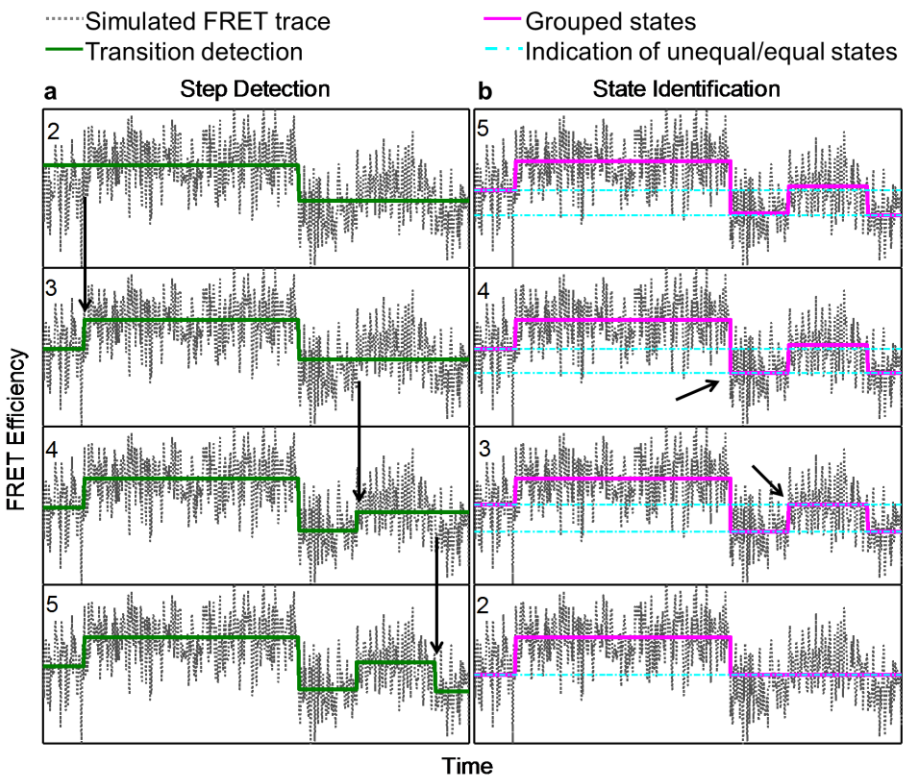
The Program with GUI

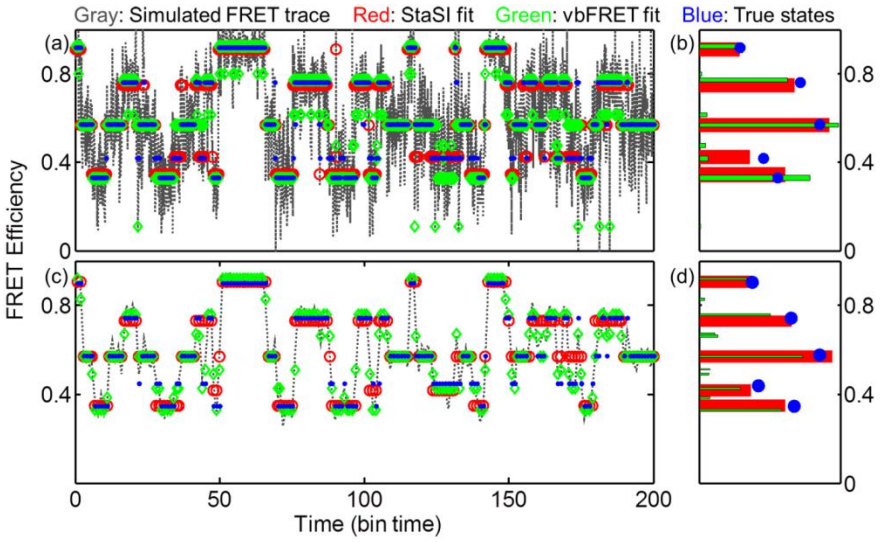


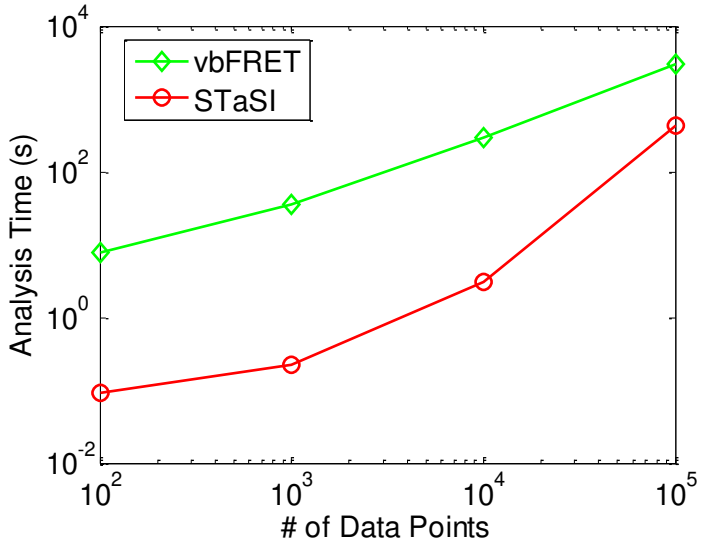
Conclusion

- STaSI provides comprehensive, objective analysis of multiple traces requiring few user inputs about the underlying physical models, and is faster and more precise in determining the number of states than established and cutting-edge methods for single-molecule data analysis.
- This method improves the state determination for noisy data or data with fast dynamics, which potentially increases the temporal resolution in single-molecule measurements.
- We plan to expend this method to data with variance change or frequency change.
- Our program (in MATLAB) can be downloaded at: http://lrg.rice.edu/Content.aspx?id=96

Questions


Simulated FRET trace
Transition detection
Grouped states
Indication of unequal/equal states
a
Step Detection
b
State Identification
FRET Efficiency
Time


Gray: Simulated FRET trace
Red: StaSI fit
Green: vbFRET fit
Blue: True states
FRET Efficiency
Time (bin time)


vbFRET
STaSI
Analysis Time (s)
# of Data Points