Signal processing for molecular and cellular biological physics: an emerging field

Max A. Little, MIT (maxl@mit.edu)
with Nick S. Jones, Imperial College London
www.maxlittle.net

MIT MEDIA LAB

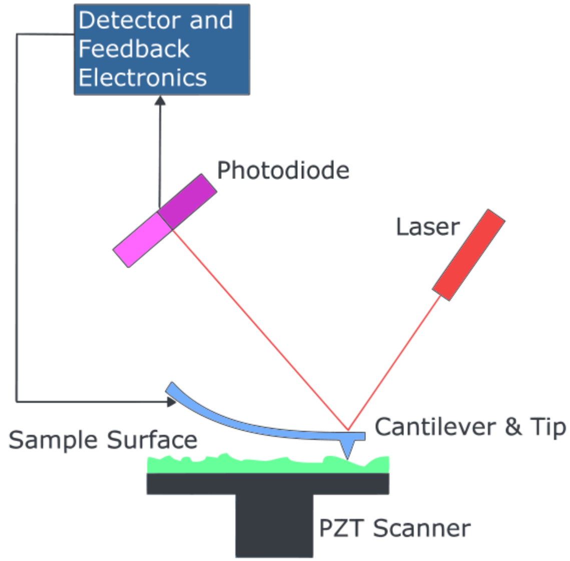
UNIVERSITY OF OXFORD

wellcome trust Fellow

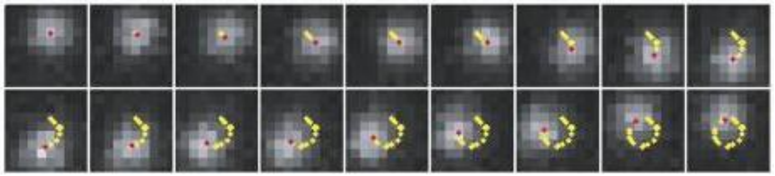
Common experimental assays, data pre-processing

- Single-molecule force spectroscopy (a)
- Forster resonance energy transfer
- Laser illuminated beads (b)
- CCD image pre-processing (c)

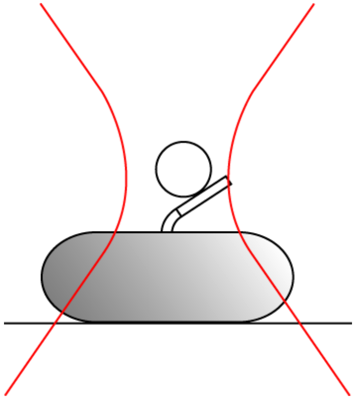
a



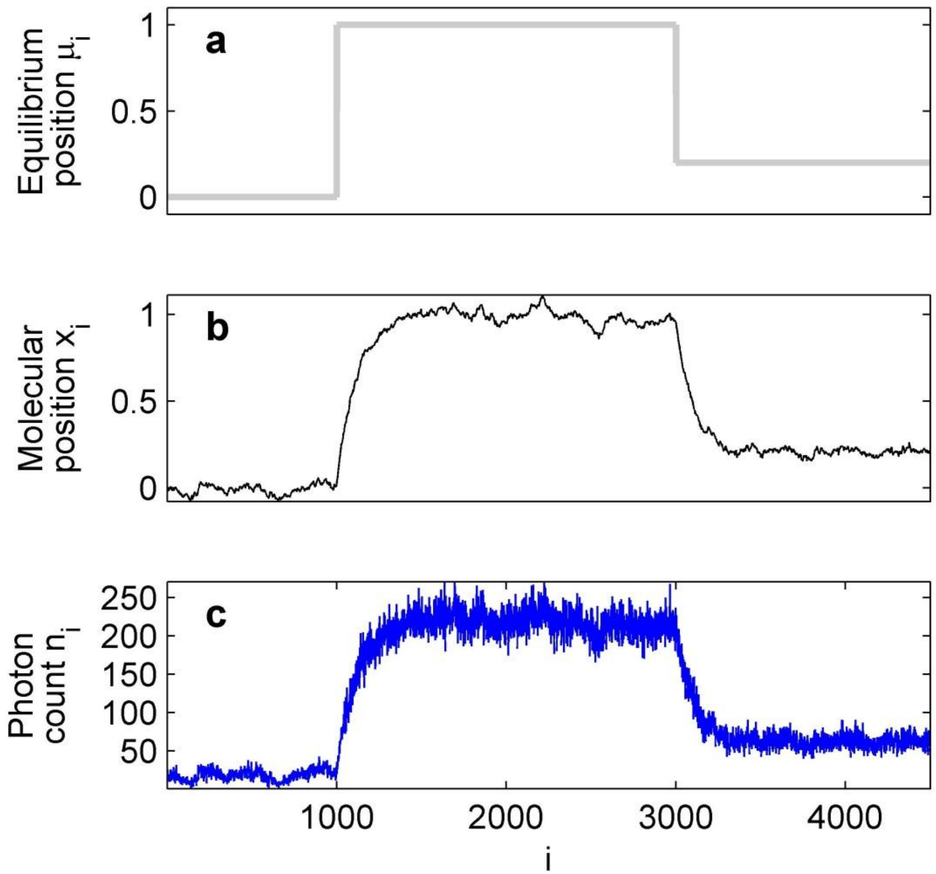
c



b



The distinctive character of biophysical time series



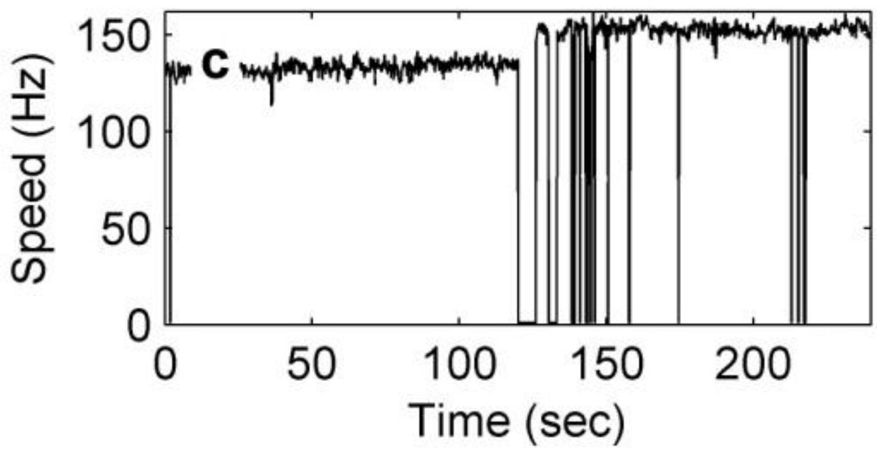
a. Step-like behaviour
b. Langevin dynamics: autocorrelation

$$\mathrm{d}x = \nu(\mu - x)\mathrm{d}t + \sigma\,\mathrm{d}W$$

c. Poisson observation noise

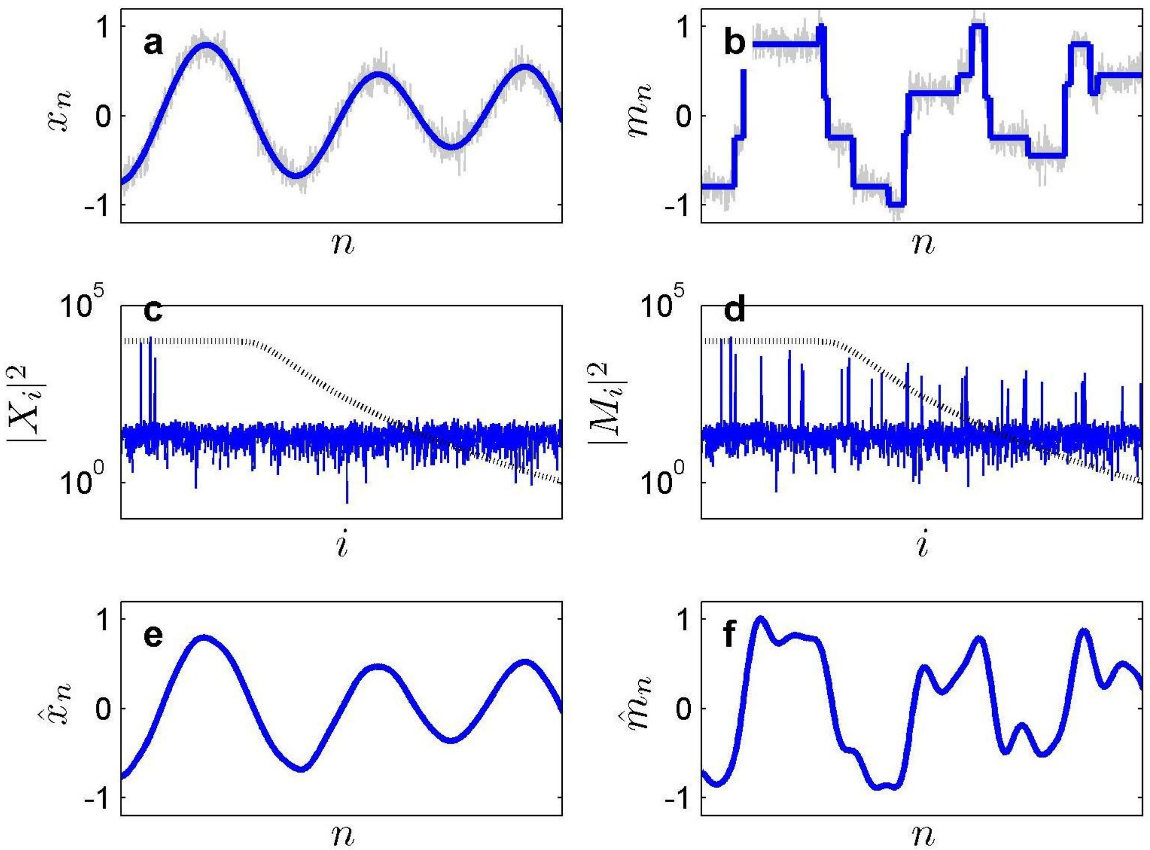
$$n \sim \mathrm{Poisson}(ax + b)$$

Noise removal from step-like signals



- Remove noise whilst retaining the edges

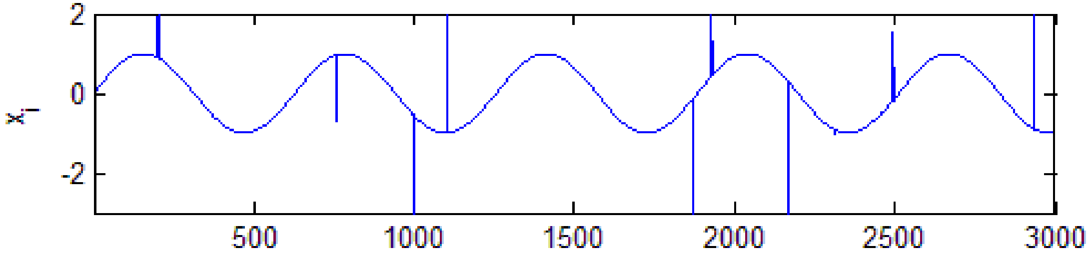
Classical LTI DSP fails in this important circumstance



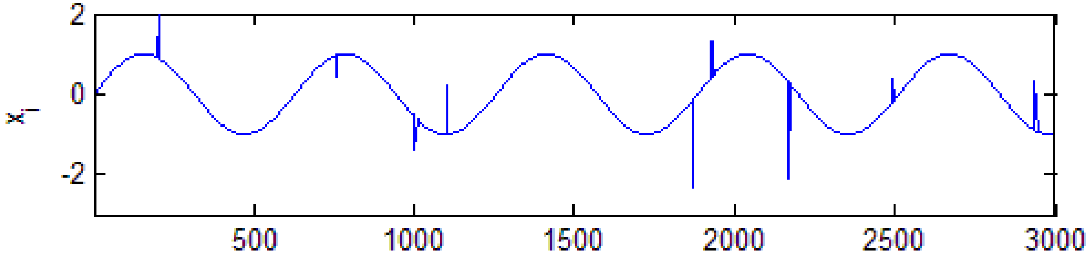
- Edges and noise *overlap in the Fourier domain*

Nonlinear running filters

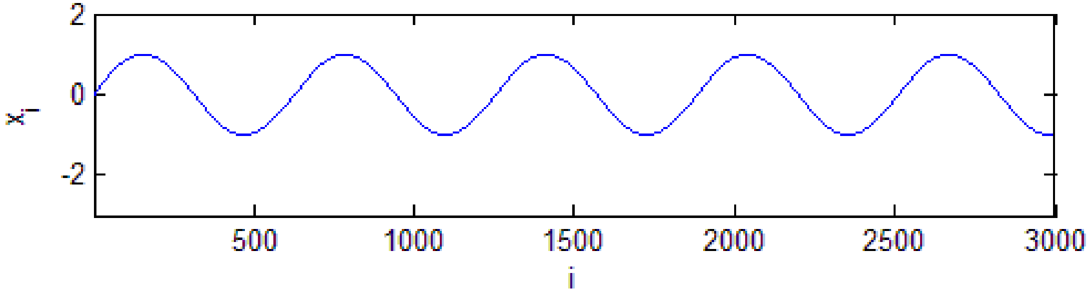
Original signal corrupted by spike noise



Running mean filter (size 3)



Running median filter (size 3)



Constant spline and level-set filtering of piecewise constant signals

- Existence of generalized functional

$$E = \sum_t \sum_s \Lambda\left(x_t - \hat{x}_s, \hat{x}_t - \hat{x}_s, x_t - x_s, t - s\right)$$

- Total variation regularization

$$\Lambda = \frac{1}{2}\left(x_t - \hat{x}_s\right)^2 I\left(t = s\right) + \gamma\left|\hat{x}_t - \hat{x}_s\right| I\left(t - s = 1\right)$$

- Mean shift clustering

$$\Lambda = \min\left(\frac{1}{2}\left(\hat{x}_t - \hat{x}_s\right)^2, W\right)$$

- Robust total variation regularization

$$\Lambda = \frac{1}{2}\left|x_t - \hat{x}_s\right| I\left(t = s\right) + \gamma\left|\hat{x}_t - \hat{x}_s\right| I\left(t - s = 1\right)$$

Markov chain analysis

- HMM with Gaussian observation noise
- Number of states using AIC/BIC
- E-M inference
- Independent states collapses to GMM

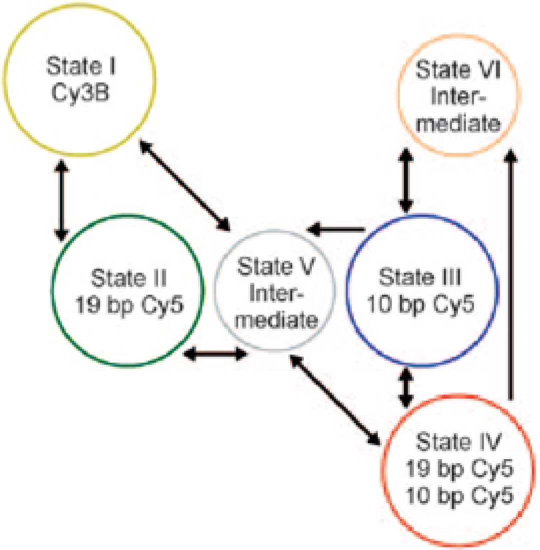
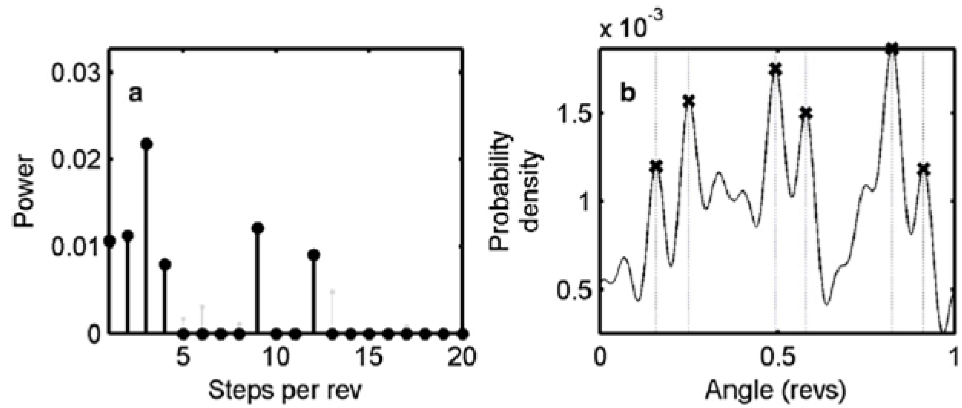


Figure from Uphoff et al. ChemPhysChem (2011), DOI: 10.1002/cphc.201000834

Periodic distribution estimates

- Multi-modal (>20 modes) distributions
- Mixture modeling intractable
- Histogram and kernel density estimates
- *Empirical characteristic function and hard shrinkage estimation*



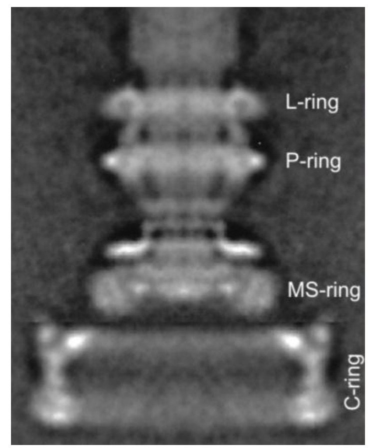
$$P(f) \approx \frac{1}{N}\sum_{i=1}^{N} \exp(jfx_i)$$

$$Q(f) = \begin{cases} P(f) & |P(f)| \geq \sqrt{2\log F}\,\zeta \\ 0 & \text{otherwise} \end{cases}$$

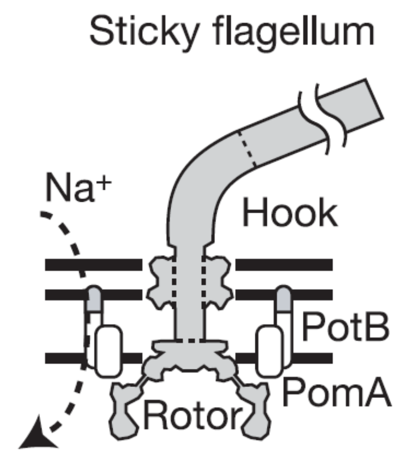
$$p(x) \approx \sum_{f=-F}^{F} \exp(-jfx)\,\bar{Q}(f)$$

Example application: the periodicity of the bacterial flagellar motor

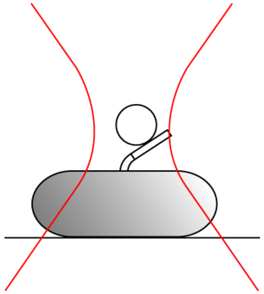
- Motor rotates flagellum: tail-like structure, propels bacteria
- Sodium ion electrochemical gradient drives rotor
- Direction change controlled by chemotaxis pathway



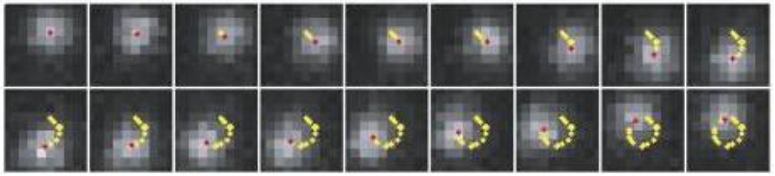
Electron micrograph



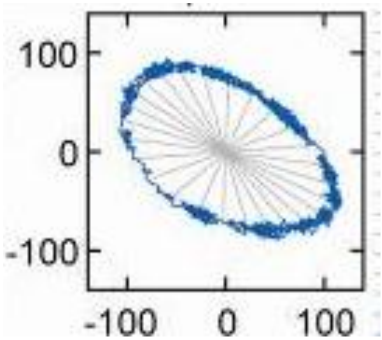
Example application: the periodicity of the bacterial flagellar motor



Laser illuminated bead attached to flagellar hook

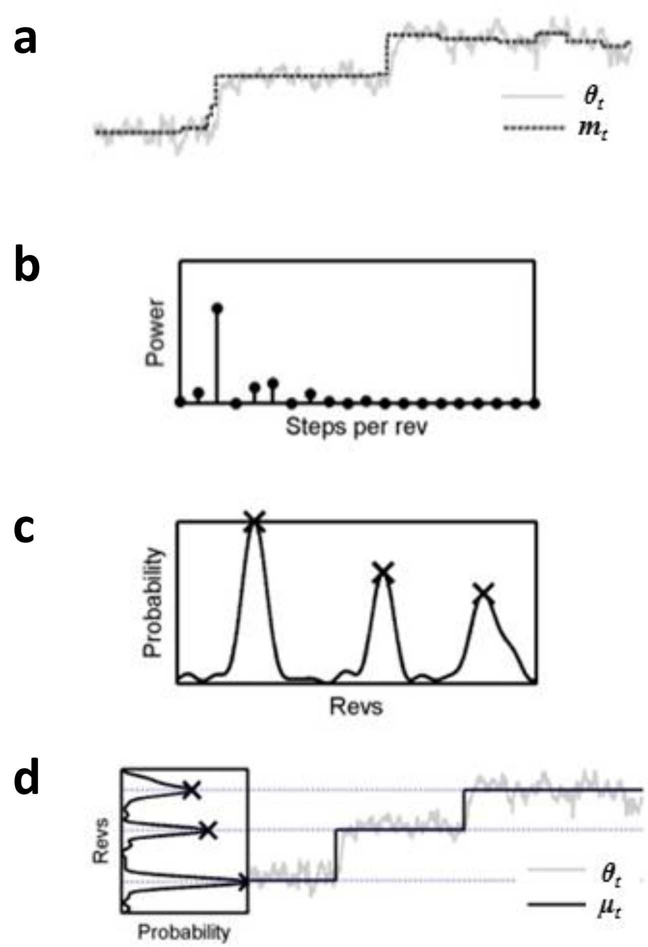


Track bead position by microscopic bead image fitting to CCD movies



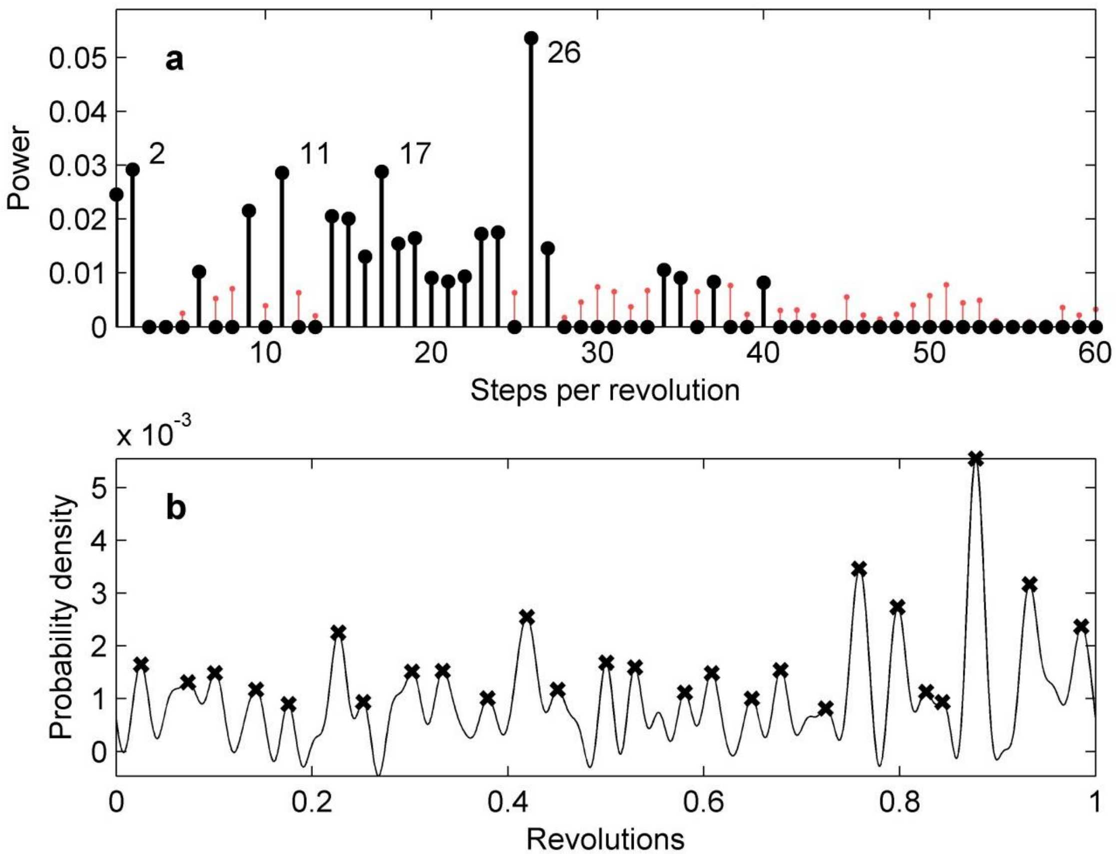
Angle of bead rotation extracted from bead position estimates

Example application: the periodicity of the bacterial flagellar motor



a. Langevin model-based step filtering
b. ECF, shrink
c. Inverse ECF, find peaks
d. Classify to nearest peaks

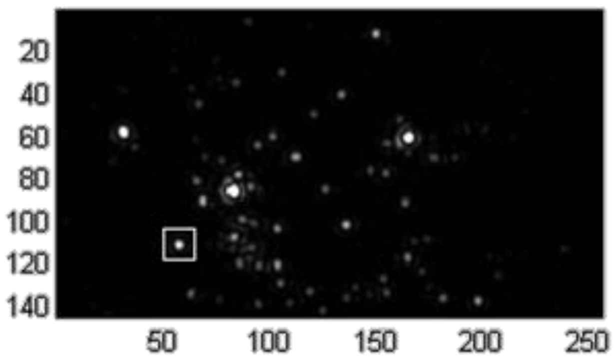
Example application: the periodicity of the bacterial flagellar motor



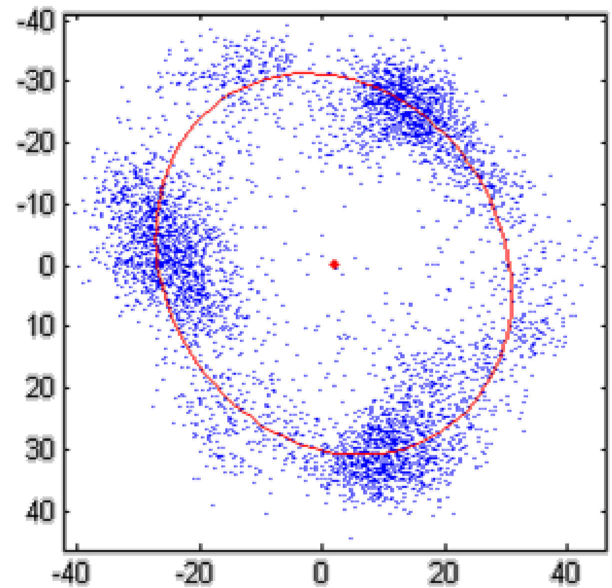
- 26- and 11-fold symmetry confirmed
- Non-Poisson dwell times

Example application: detecting cascaded rate-limiting steps in ATP synthesis

- ATP standard cellular “energy unit”
- F1-ATPase rotary motor synthesizes ATP using proton gradient

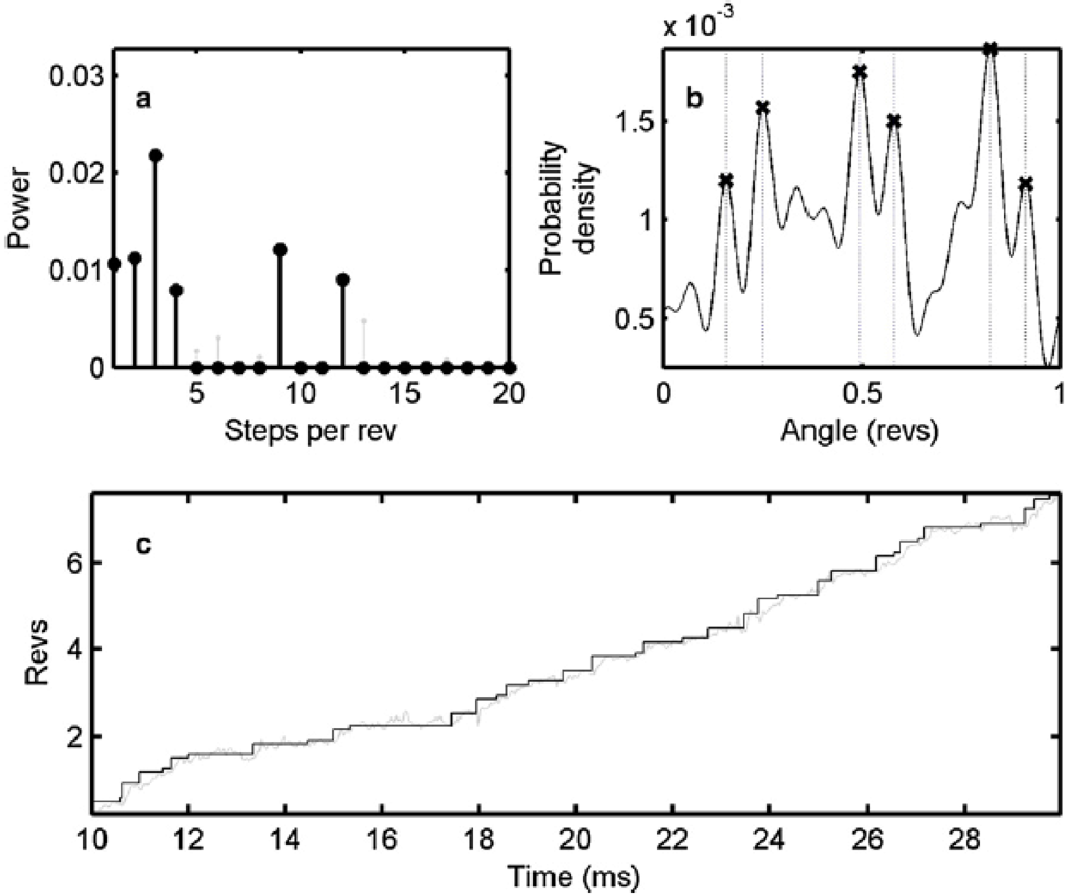


Laser illuminated bead imaged with high-speed CCD movie camera



Obtain motor angle at each movie frame

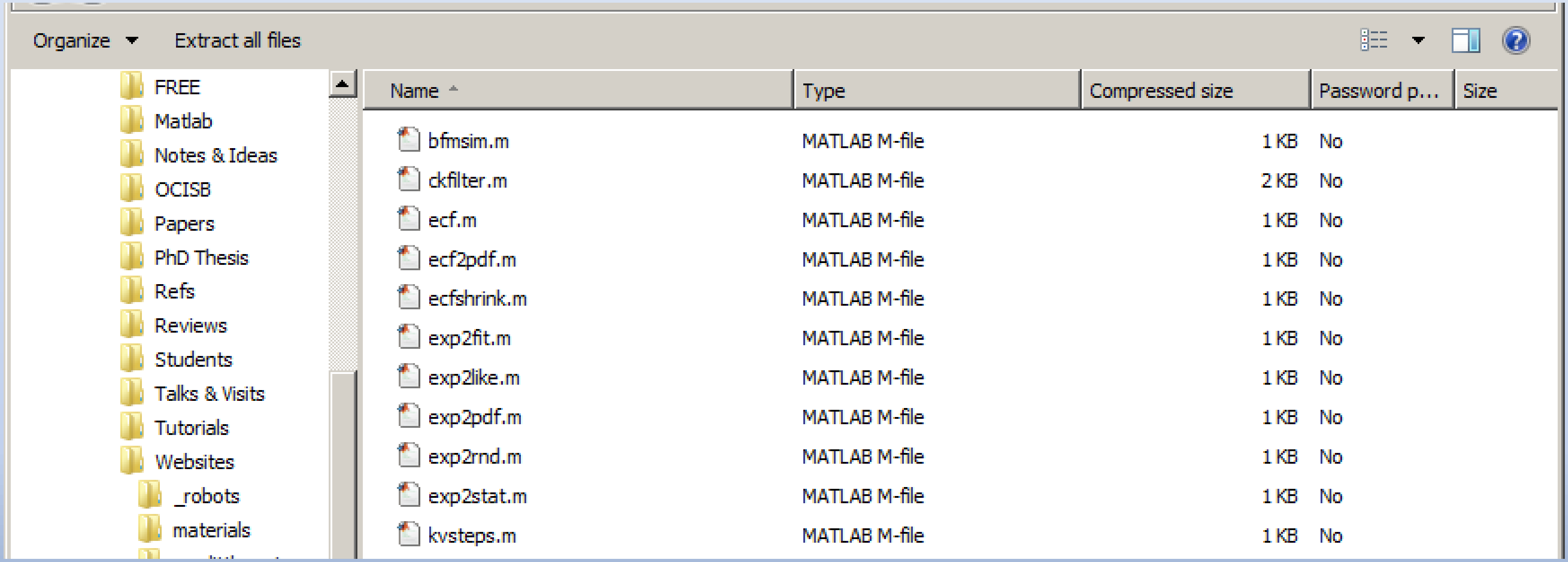
Example application: detecting cascaded rate-limiting steps in ATP synthesis



a. ECF magnitudes
b. Density estimate
c. Angle-time signal

- k=2 gamma-distributed dwell times indicates cascaded sub-steps

Biophysical signal processing code



- Open-source Matlab toolbox: *Langevin model-based step filtering, ECF and inverse ECF, shrinkage, classification etc.*
- **www.maxlittle.net/software/steps_bumps_toolkit.zip**
- Your contributions welcome! Get in touch: maxl@mit.edu