Inference in microscopy and biological physics

Phil Nelson
University of Pennsylvania

For these slides see:
www.physics.upenn.edu/~pcn

Cartoon by Larry Gonick

Focus

"If your experiment requires statistics, then you ought to have done a better experiment." -- Ernest Rutherford

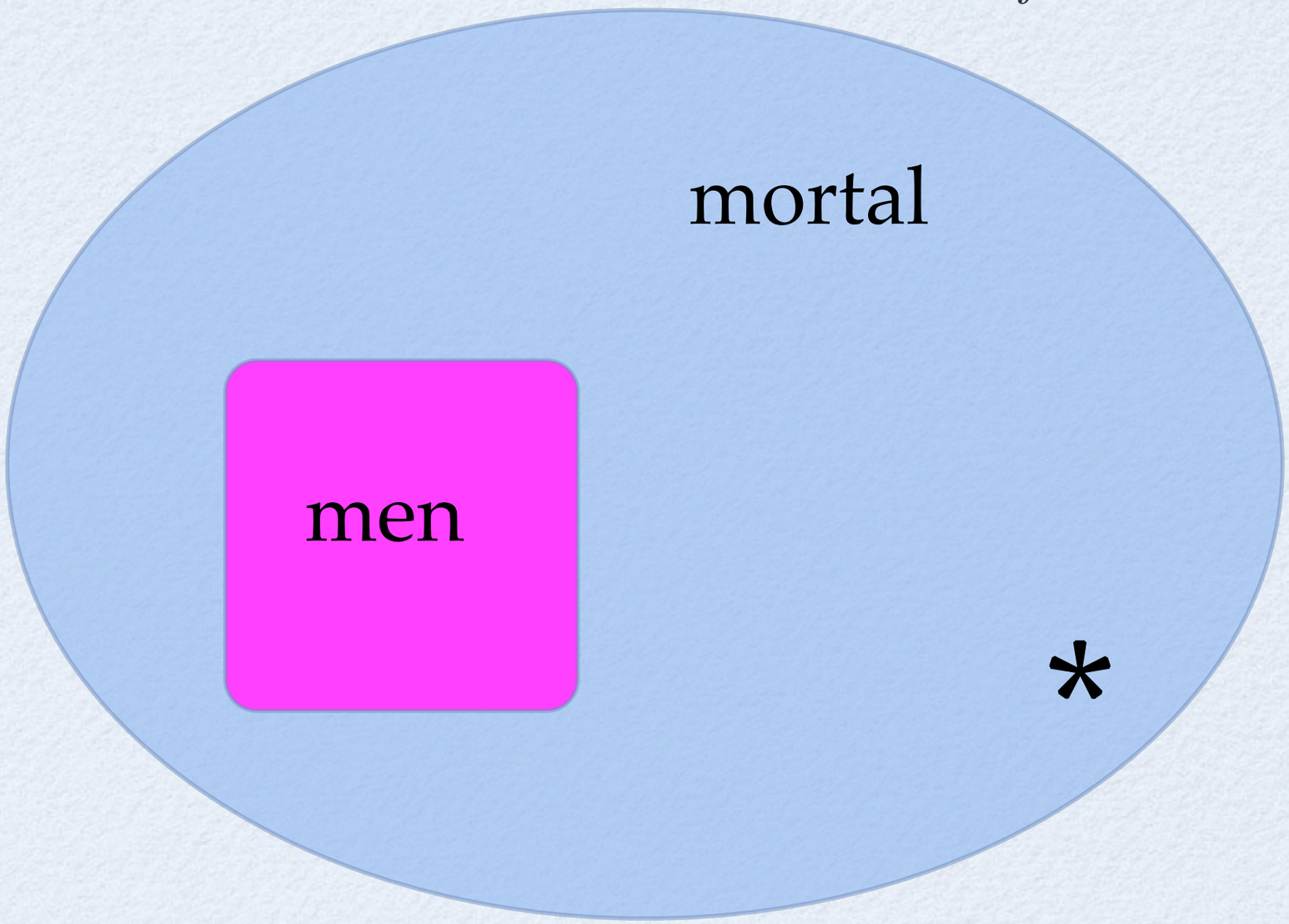
Well... Statistical inference may sound "too theoretical," but it is often needed to extract *information* from *data:*

- Sometimes suggests a new kind of measurement that tests a model more stringently, or distinguishes two different models more completely, than previous measurements.
- Sometimes our model is not obviously connected with what we can actually measure experimentally, and we need to make a connection.
- Sometimes the model that interests us involves the behavior of actors that we can only see indirectly in our data; theory may be needed to separate them out from each other, and from noise.

There's more, of course, but that's enough to get started.

Part I:

Suppose I stood here and said "*all men are mortal; Socrates is mortal; therefore Socrates is a man*."



In classical logic it's fairly easy to spot errors of inference.

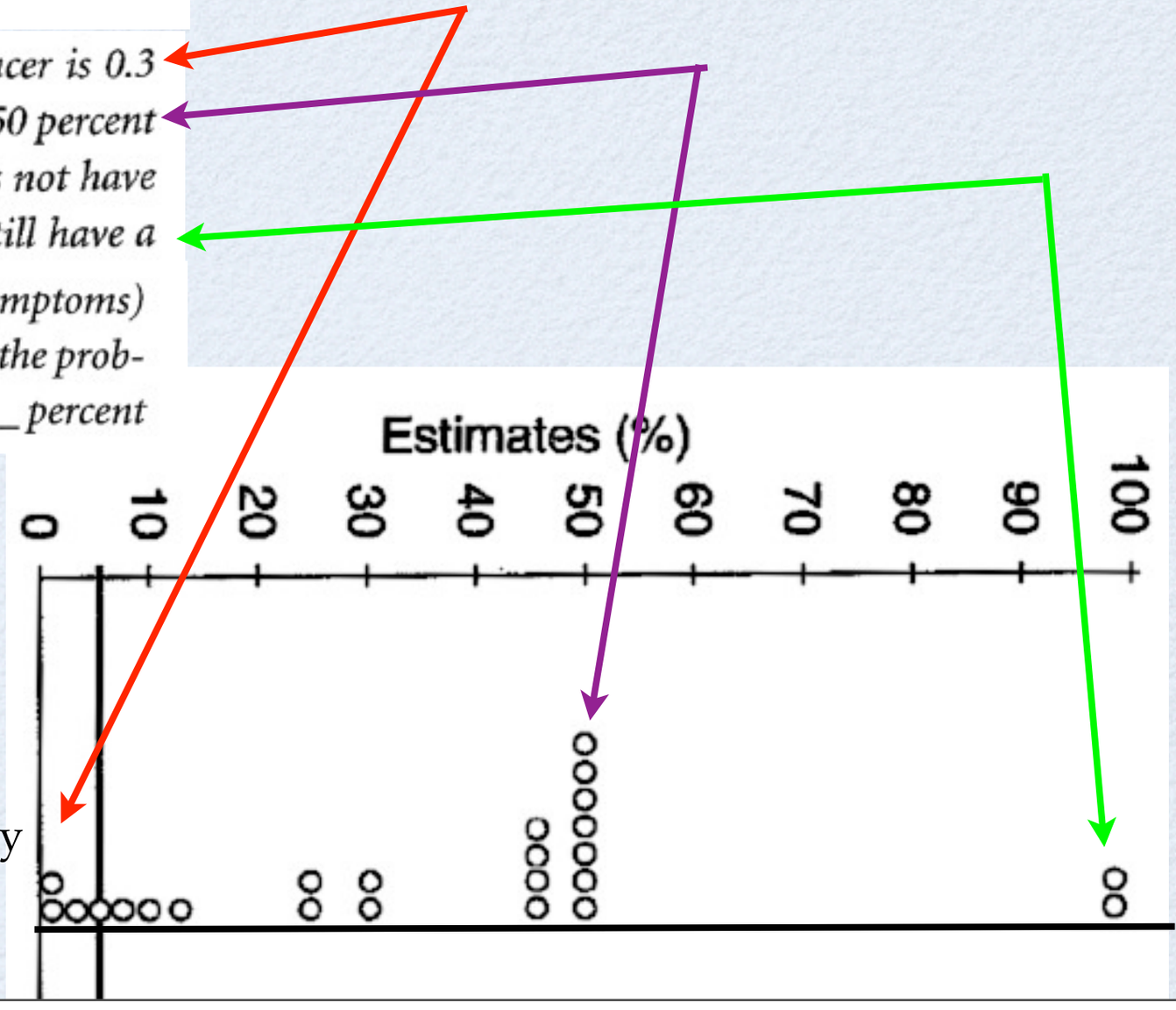
But what if I said "*92.7% of all men are mortal*..." Suddenly we find such questions tricky.

P. Nelson, Physical models of living systems (WH Freeman and Co 2015)

An everyday question in clinical practice

To diagnose colorectal cancer, the hemoccult test—among others—is conducted to detect occult blood in the stool. This test is used from a particular age on, but also in routine screening for early detection of colorectal cancer. Imagine you conduct a screening using the hemoccult test in a certain region. For symptom-free people over 50 years old who participate in screening using the hemoccult test, the following information is available for this region:

The probability that one of these people has colorectal cancer is 0.3 percent. If a person has colorectal cancer, the probability is 50 percent that he will have a positive hemoccult test. If a person does not have colorectal cancer, the probability is 3 percent that he will still have a positive hemoccult test. Imagine a person (over age 50, no symptoms) who has a positive hemoccult test in your screening. What is the probability that this person actually has colorectal cancer? _____ percent



Here are the replies of 24 practicing physicians, who had an average of 14 years of professional experience:

G. Gigerenzer, Calculated risks

Work it out

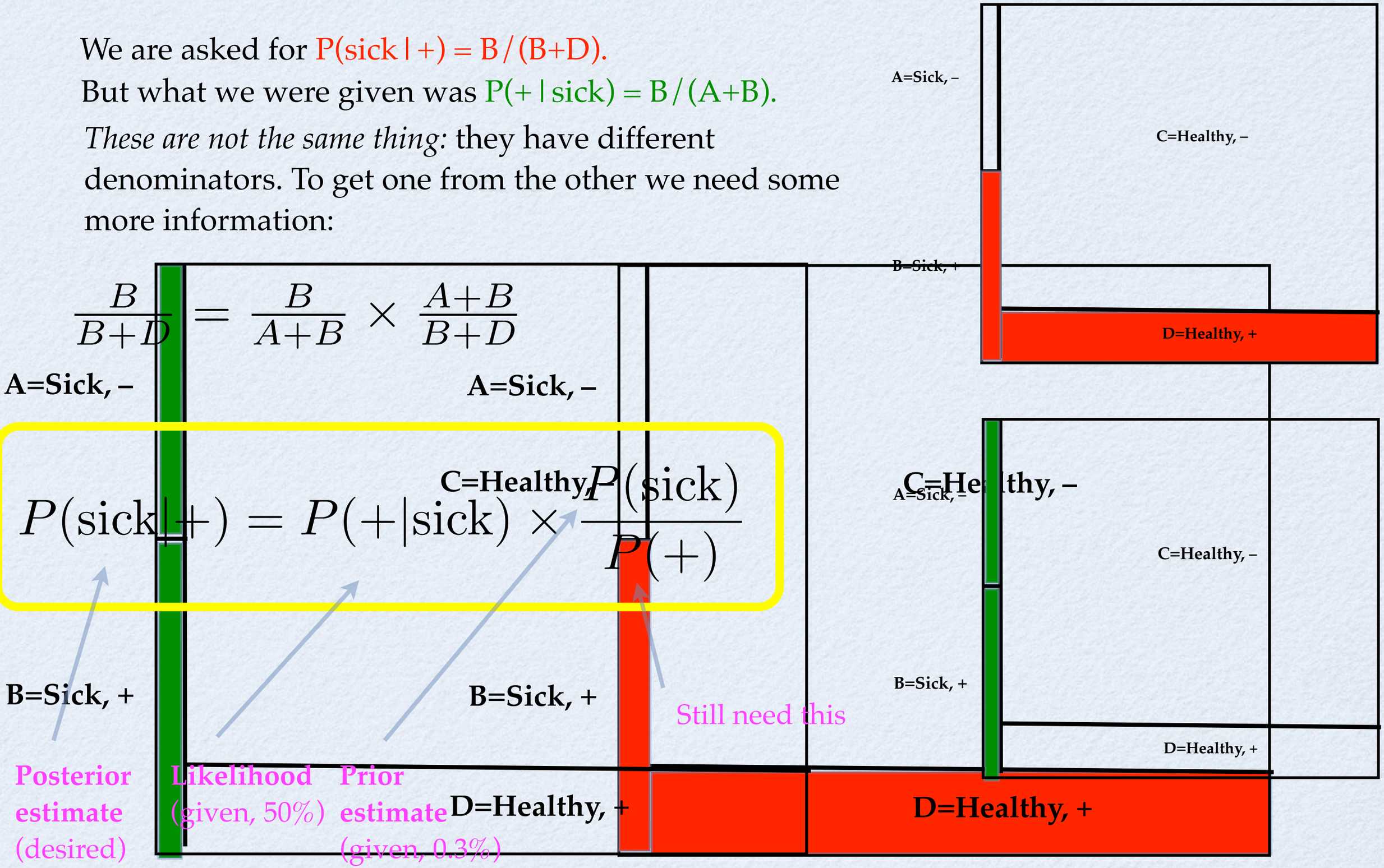
We are asked for P(sick | +) = B/(B+D).

But what we were given was P(+ | sick) = B/(A+B).

These are not the same thing: they have different denominators. To get one from the other we need some more information:

$$\frac{B}{B+D} = \frac{B}{A+B} \times \frac{A+B}{B+D}$$

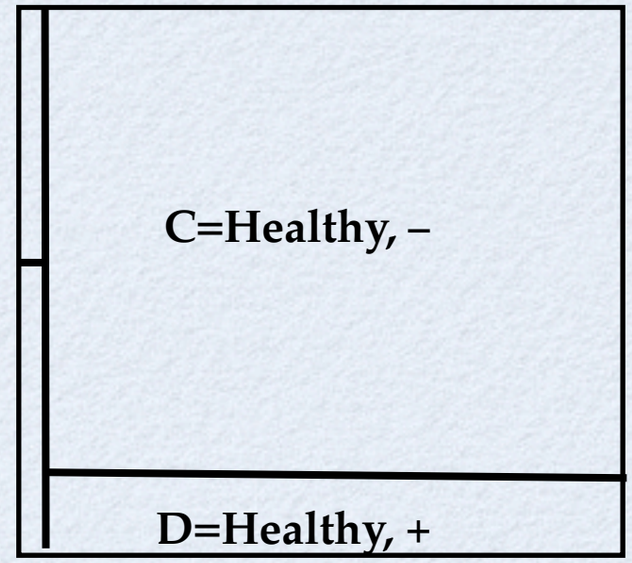
$$P(\text{sick}|+) = P(+|\text{sick}) \times \frac{P(\text{sick})}{P(+)}$$



Finish working it out

Bayes Formula:

$$P(\text{sick}|+) = P(+|\text{sick}) \times \frac{P(\text{sick})}{P(+)}$$



Is that last factor really important?

P(sick) was given, but we also need:

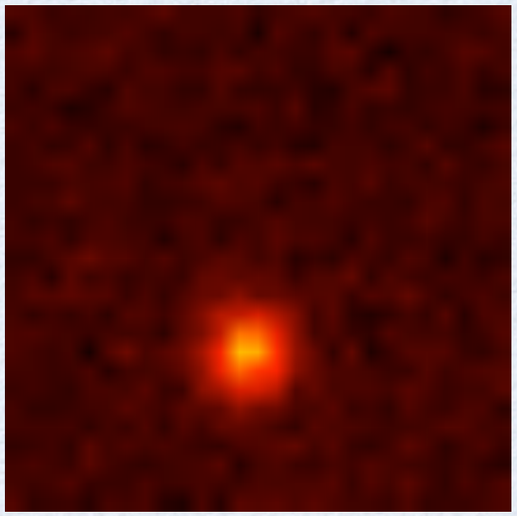
$$\begin{aligned} P(+) &= B + D \\ &= \frac{B}{A+B}(A+B) + \frac{D}{C+D}(C+D) \\ &= P(+|\text{sick})P(\text{sick}) + P(+|\text{healthy})P(\text{healthy}) \\ &= (0.5)(0.003) + (0.03)(0.997) \approx 0.03 \end{aligned}$$

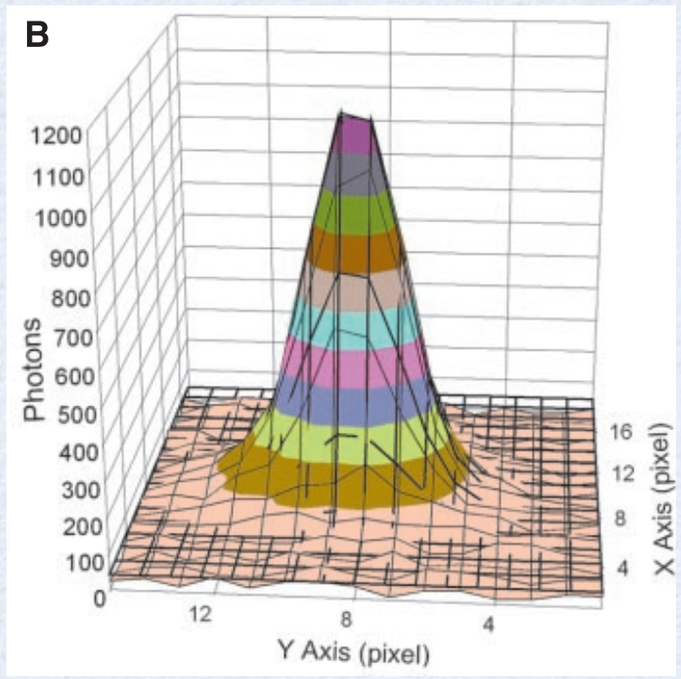
$$\frac{P(\text{sick})}{P(+)} \approx \frac{0.003}{0.03} \approx 0.1$$

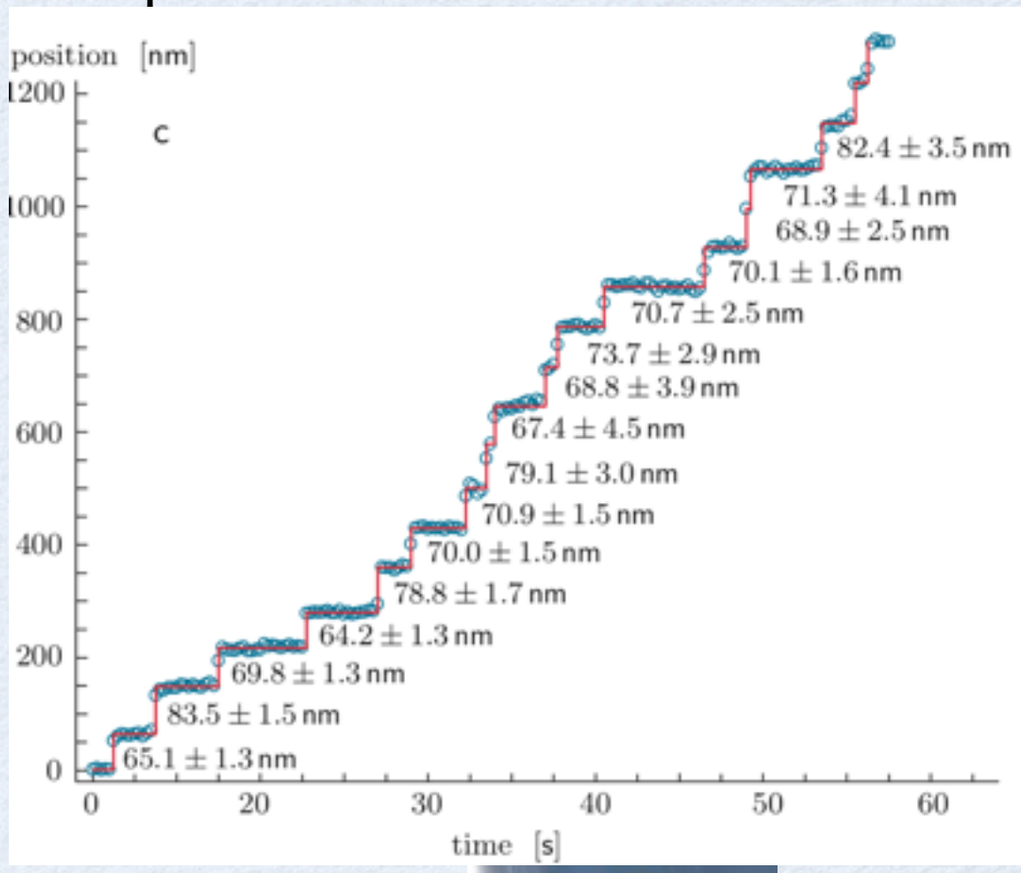
Yes, it's important: a positive test result means only a 5% chance you're sick. *Not 97%.*

Part II: Superresolution microscopy

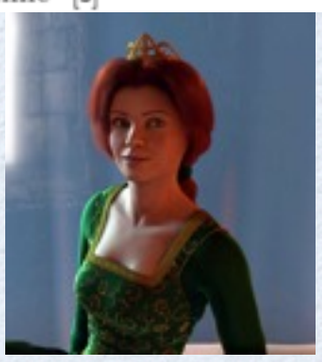
How does one measure myosin steps to within a few nm accuracy using visible light? The diffraction-limited spot is at least 200 nm wide!
The key point is to realize that although we cannot resolve *two* spots closer than this, sometimes all we want is to detect *motion* of *one* spot.







Fluorescence Imaging at One Nanometer Accuracy...



A. Yildiz, et al. Science 2003

F.I.O.N.A.

The posterior probability is

$$\mathcal{P}(x_*|x_1,\ldots,x_M) = \text{const.}\times \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-(x_1-x_*)^2/(2\sigma^2)}\times\cdots\times\frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-(x_M-x_*)^2/(2\sigma^2)}$$

want this...

know these...

(uniform prior)

likelihood is the product of independent terms

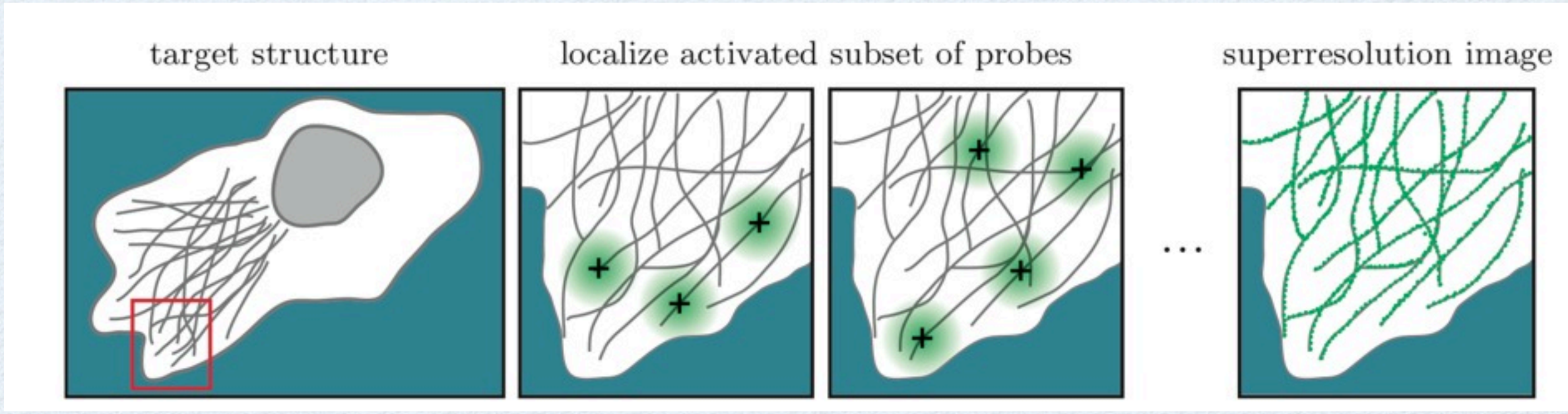
Its log is simple:

$$\ln\mathcal{P}(x_* \mid x_1,\ldots,x_M) = \sum_{i=1}^{M}\left[-\tfrac{1}{2}\ln(2\pi\sigma^2) - (x_i-x_*)^2/(2\sigma^2)\right].$$

We wish to maximize this function over x_*, holding σ and all the data $\{x_1,\ldots,x_M\}$ fixed. The beauty of this approach is that it can be generalized to include more a accurate point-spread function, background, etc.

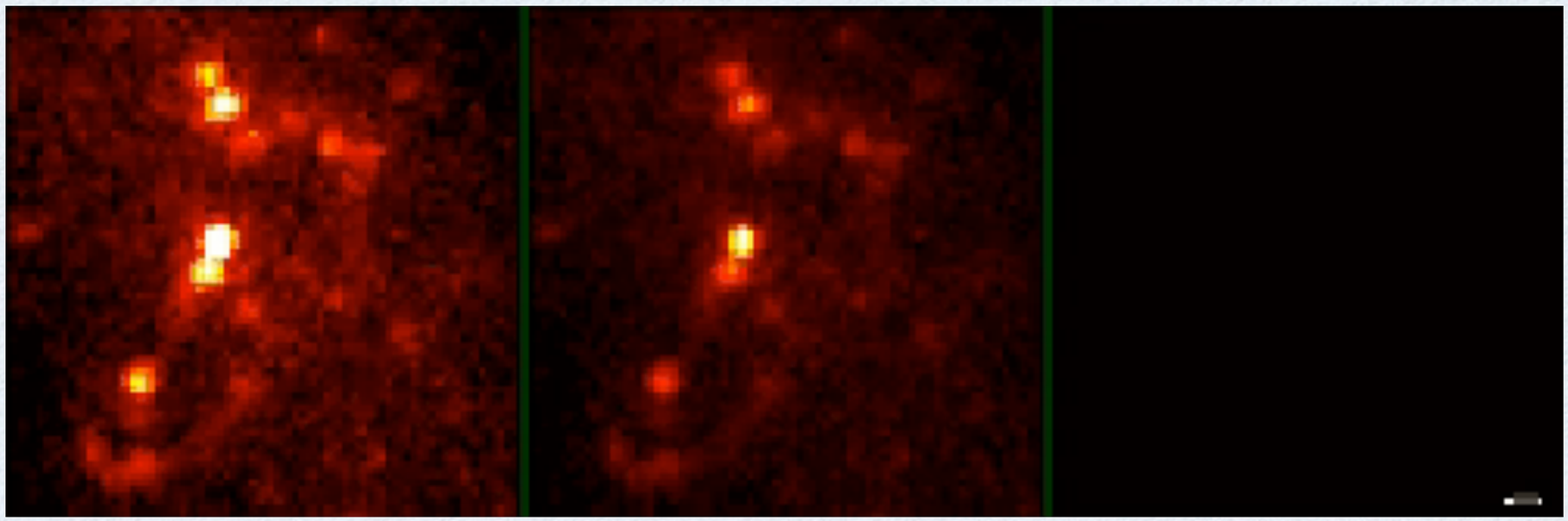
But usually we want an *image,* something a lot more structured than one point of light.

https://www.youtube.com/watch?v=RE70GuMCzww



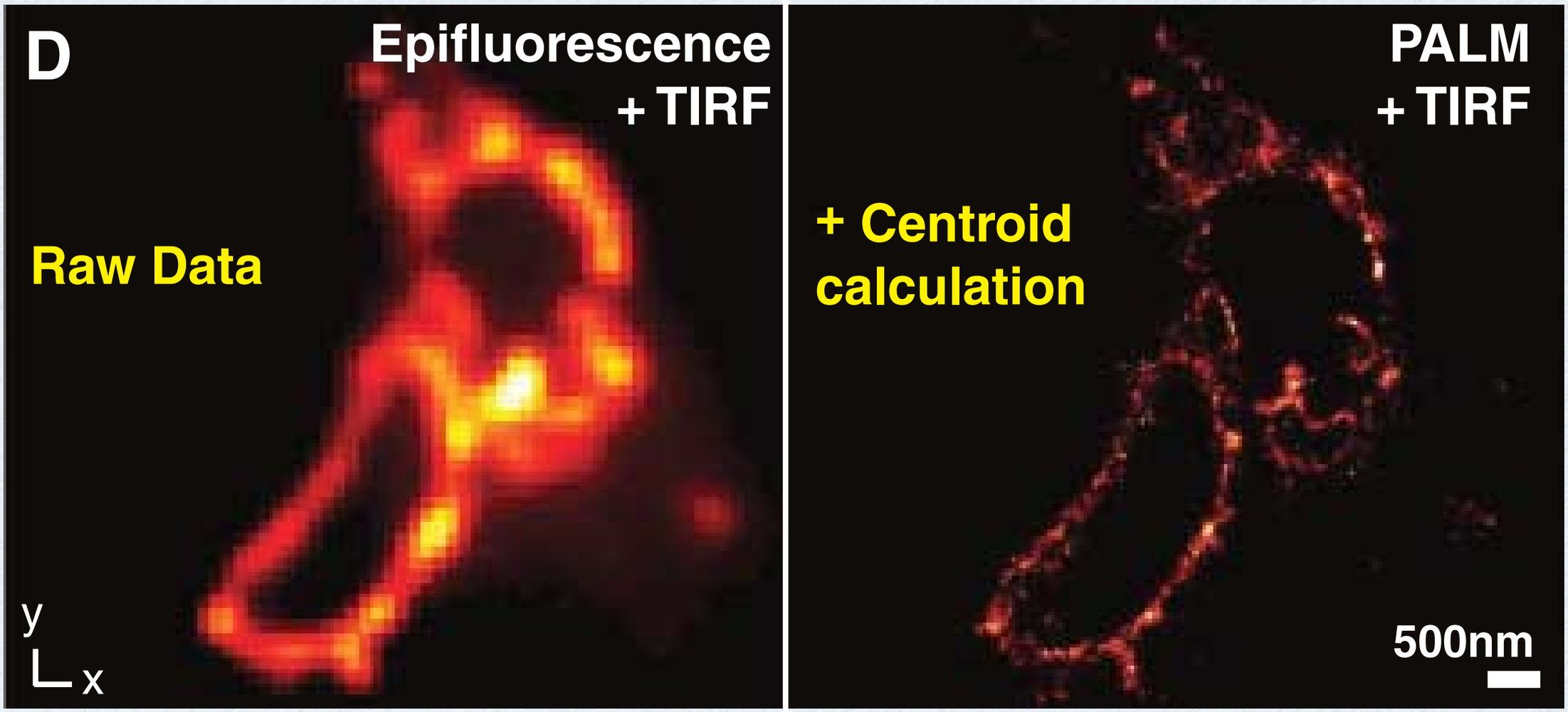
P. Nelson, Physical models of living systems (WH Freeman and Co 2015)

Can we obtain the nice features of FIONA, but for an image with more than just one object?

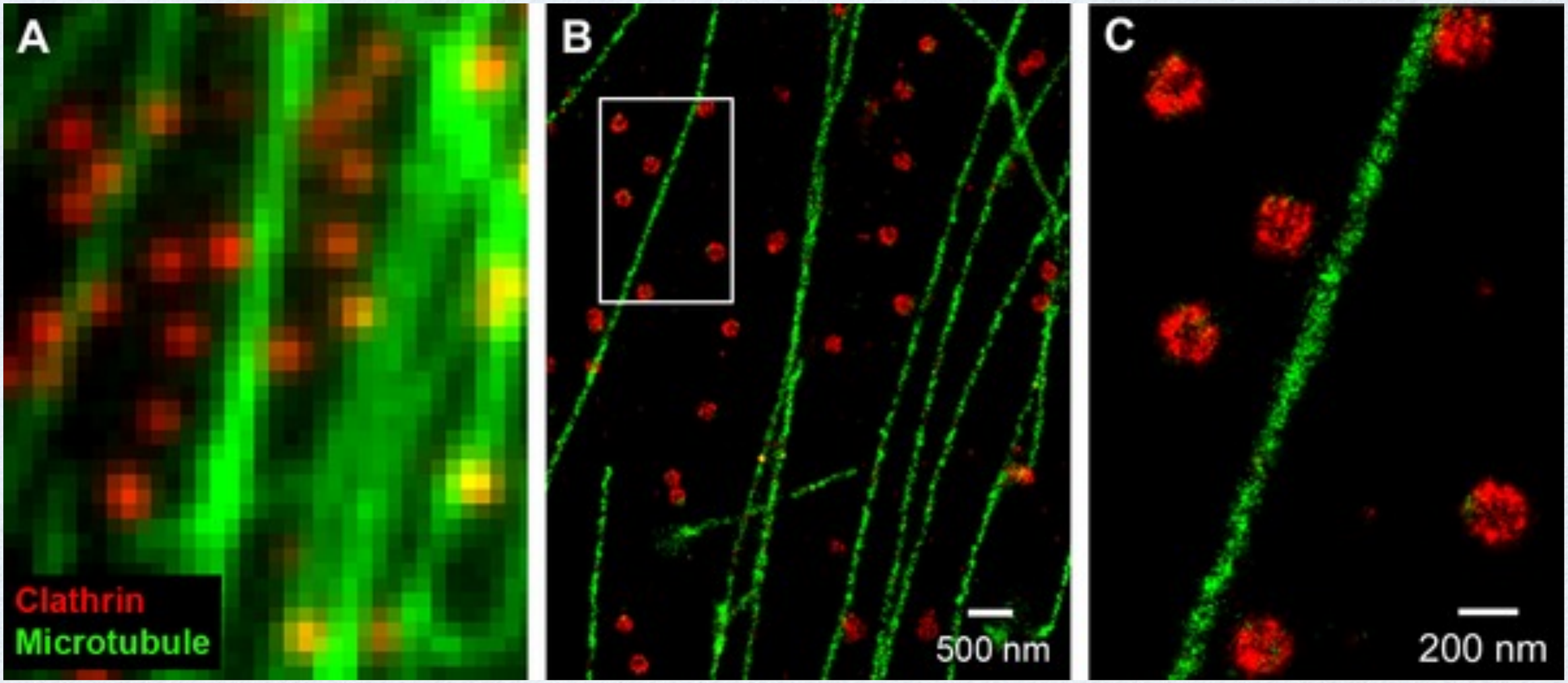


See movie: http://www.sciencemag.org/content/suppl/2006/08/08/1127344.DC1/1127344s1.mov. (Betzig et al 2006)

Partial summed molecule TIRF image (center) and PALM image (right) constructed during the acquisition of 300 single molecule frames (left) out of the 20,000 frames used to construct the images in Fig. 2. Scale bar is 0.5 μm.

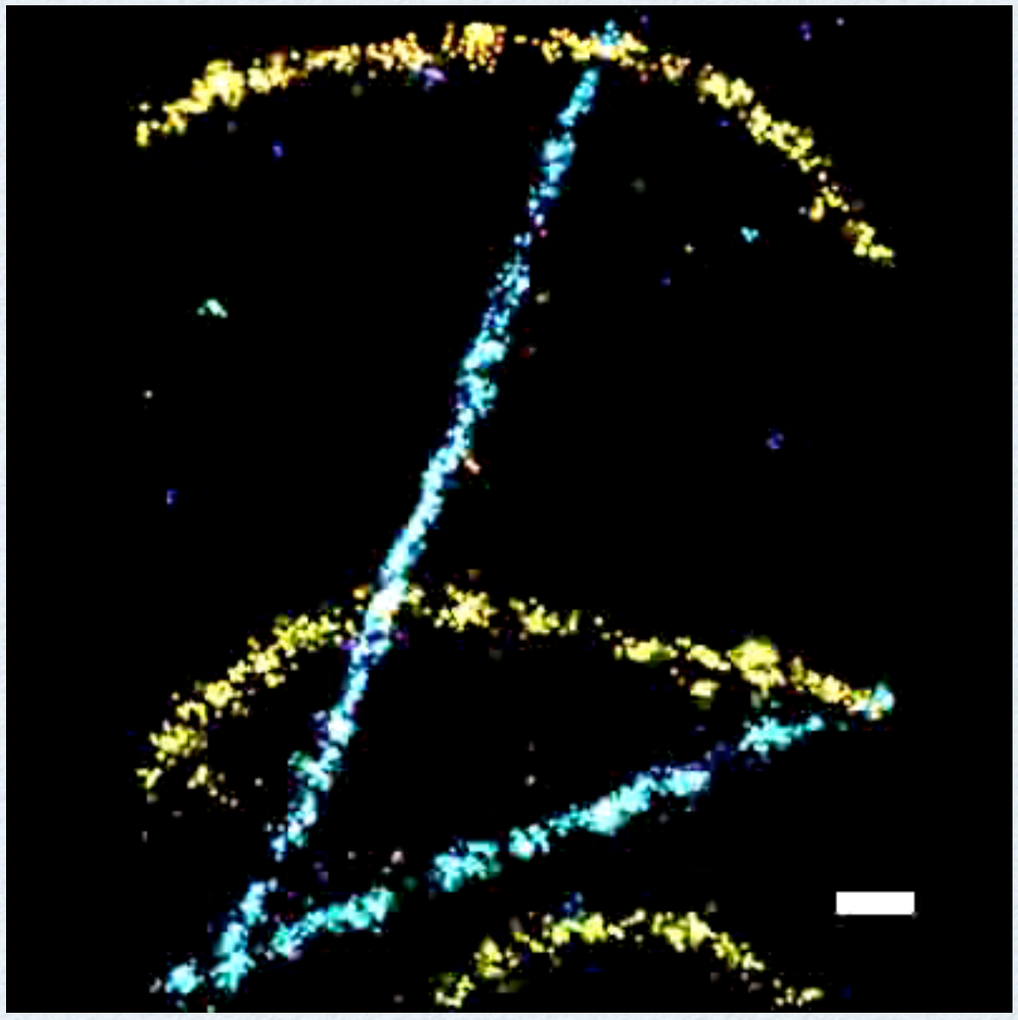


Epifluorescence versus PALM recording of a cryoprepared section from a mammalian cell expressing a lysosomal transmembrane protein tagged with a photoswitchable protein; both images were recorded with a TIRF setup. PALM resolution ranges between 20 and 60 nm, whereas individual protein localizations can be 2 nm (12).



Two-color STORM distinguishes two kinds of objects. You might not have realized from the conventional image on left that the green filaments are narrower than the red objects.

Bates et al.. Science (2007) vol. 317 (5845) pp. 1749-1753



(See movie http://zhuang.harvard.edu/movies/STORM_Movie1.mpg)

Part III: Changepoint analysis in single-molecule TIRF

JF Beausang, Yale Goldman, PN

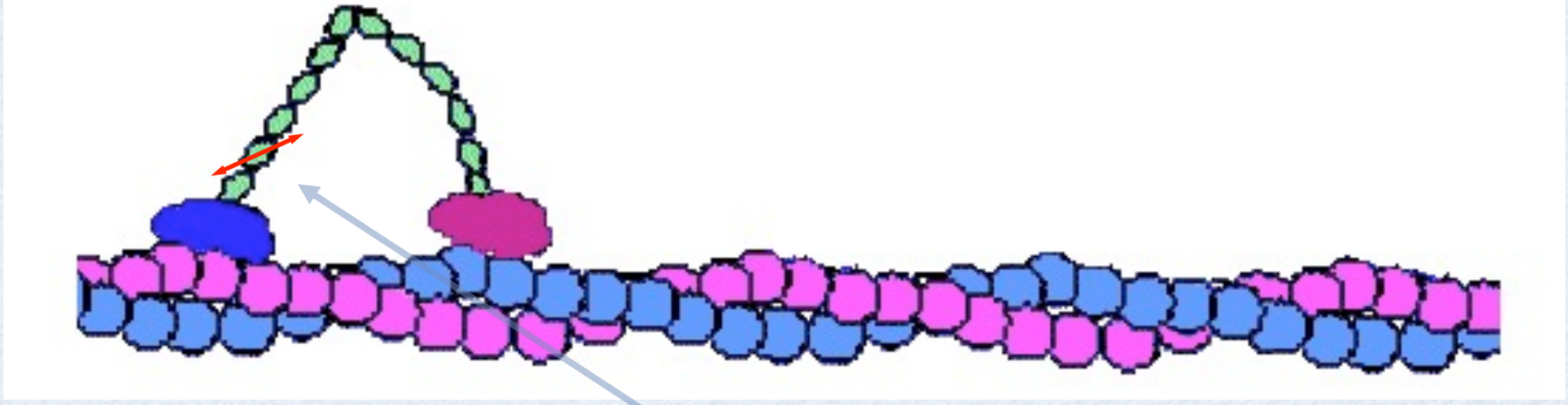
- **Sometimes our model is not obviously connected with what we can actually measure experimentally, and we need to makes a connection.**
- Sometimes the model that interests us involves the behavior of actors that we can only see indirectly in our data; theory may be needed to separate them out from each other, and from noise.

Many thanks to Haw Yang. See also Lucas P. Watkins and Haw Yang J. Phys. Chem. B 2005

Myosin V stepping

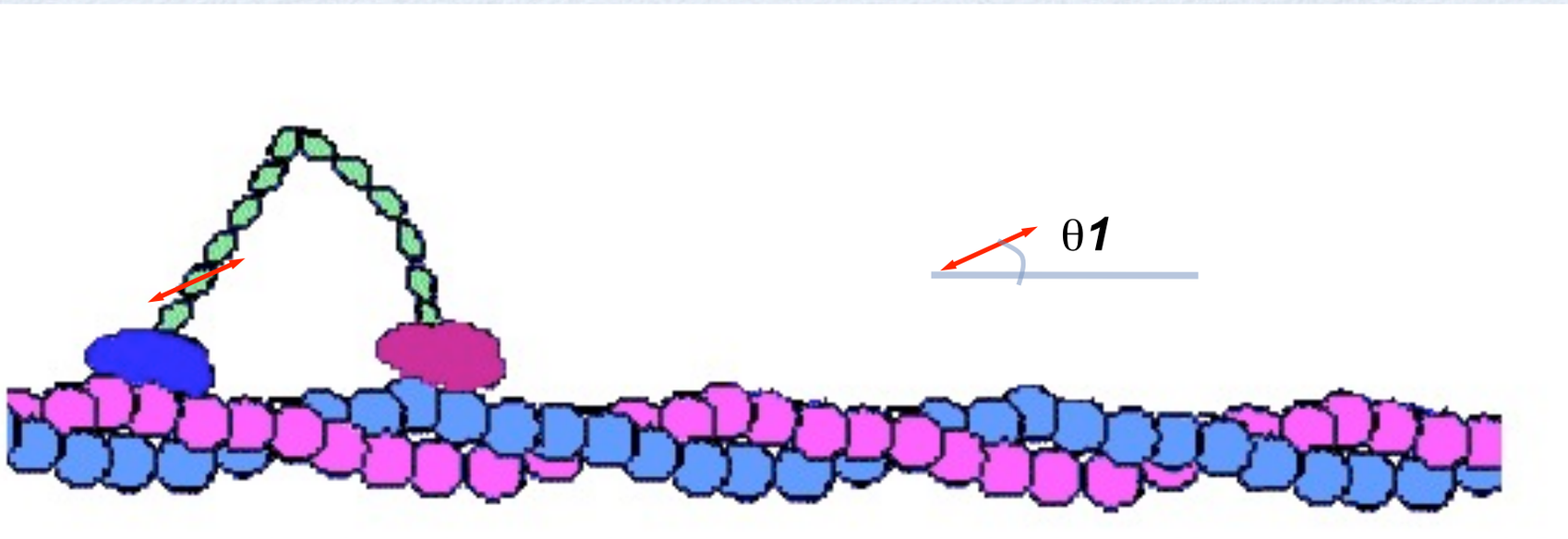
Defects in myosin V are associated with human immunological and neurological disorders.

We'd like to know things like: How does it walk? What are the steps in the kinetic pathway? What is the geometry of each state?
One classic approach is to monitor the position in space of a marker (e.g. a bead) attached to the motor. But this does not address the geometry of each state.

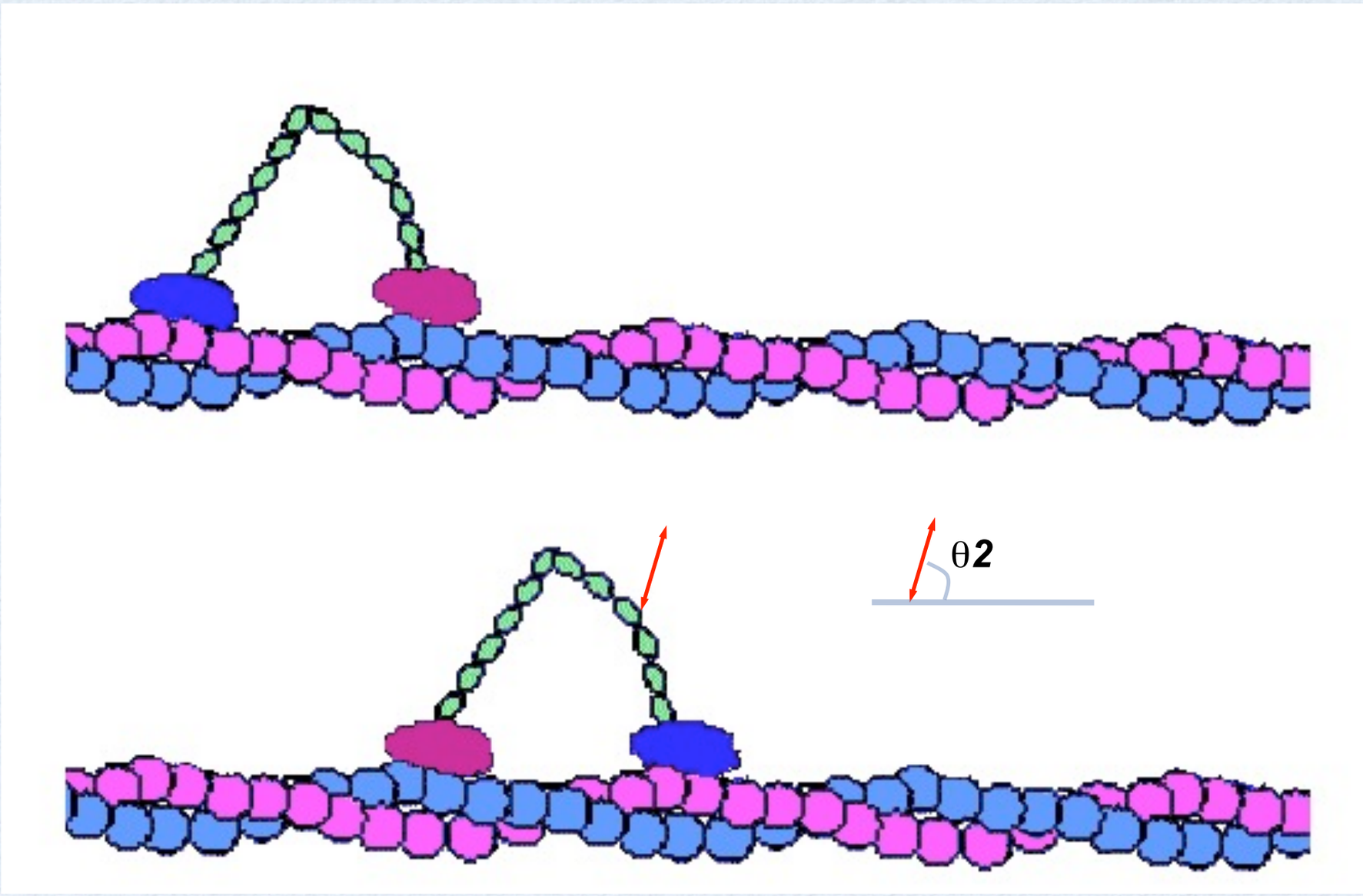


The approach I'll discuss involves attaching a bifunctional fluorescent label to one lever arm. The label has a dipole moment whose *orientation* in space reflects that of the arm.

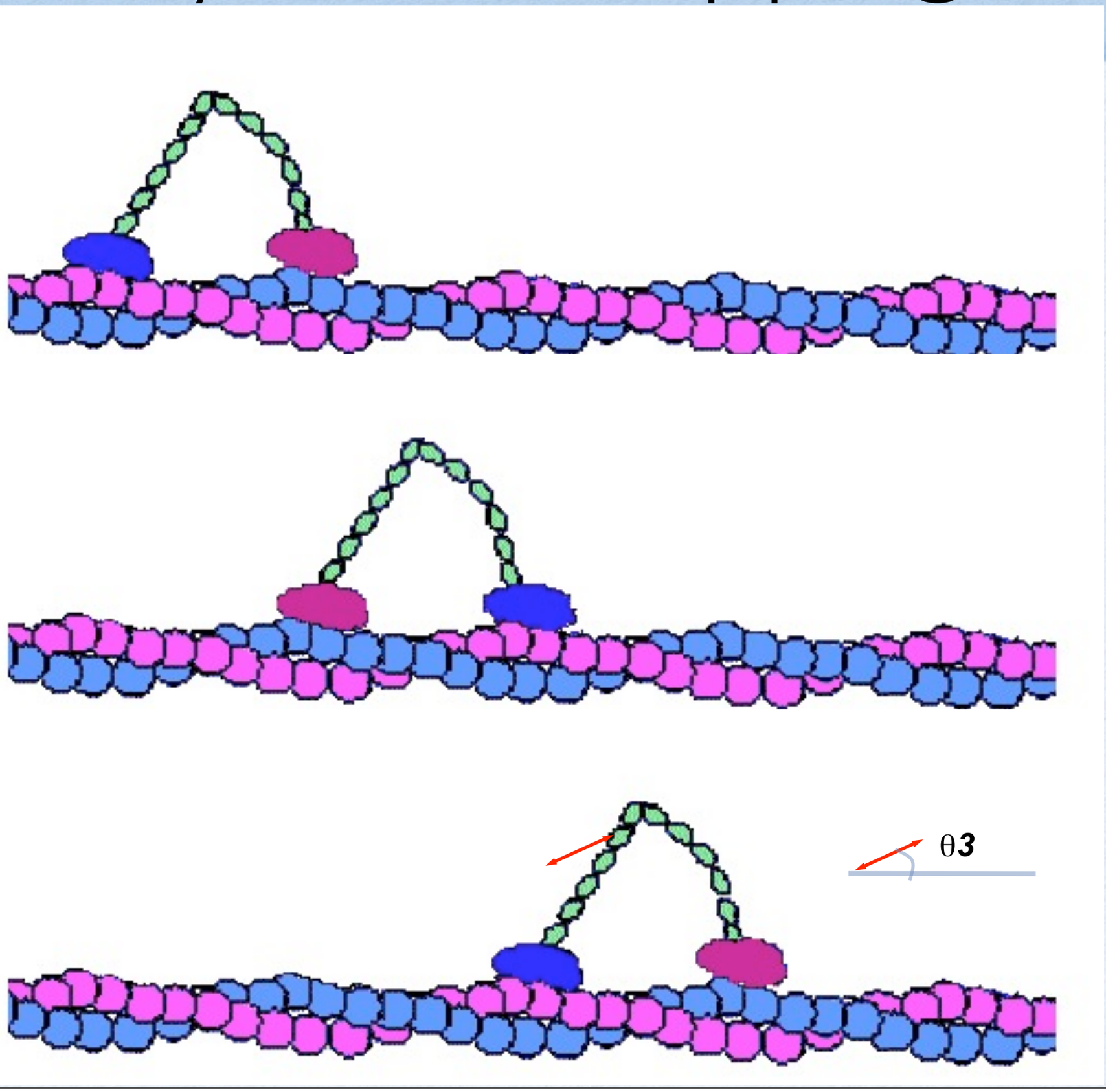
Myosin V stepping



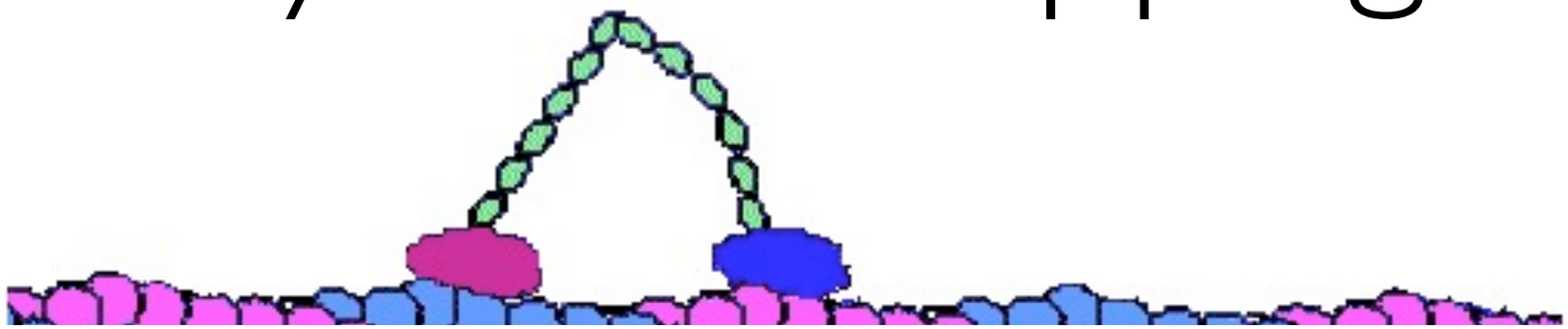
Myosin V stepping

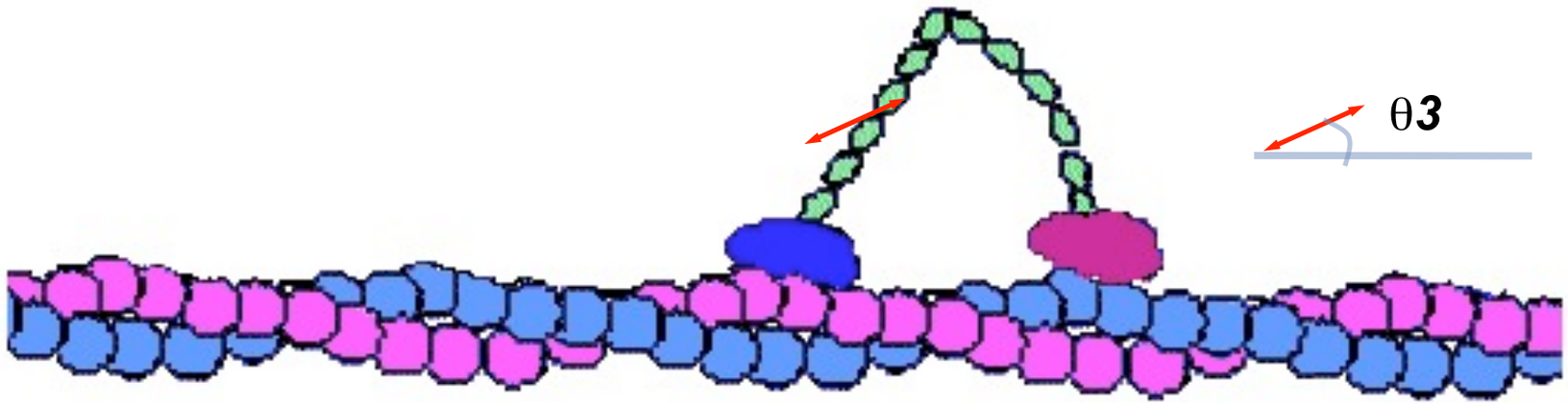


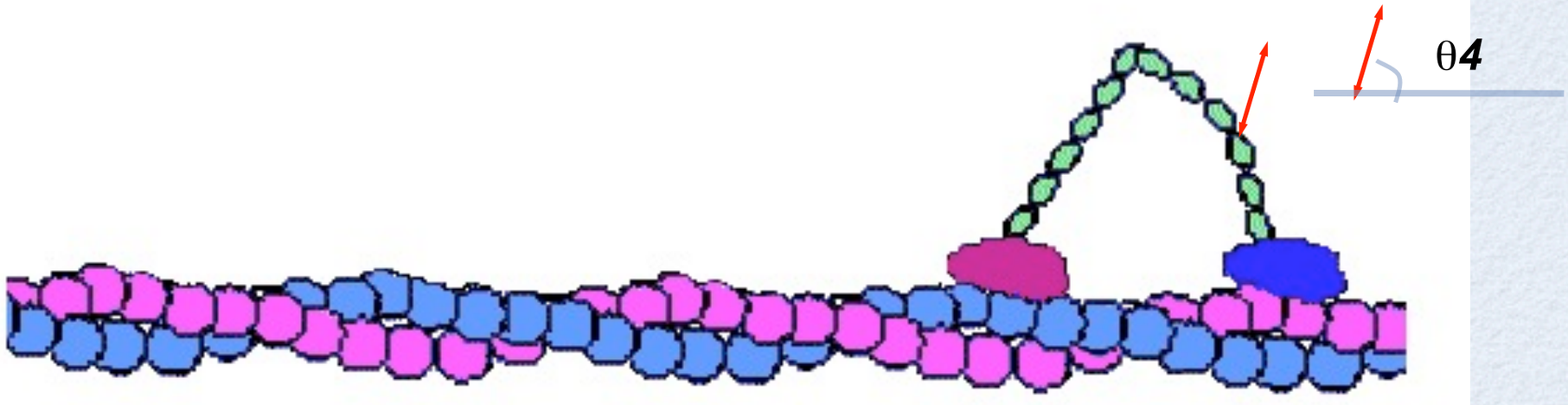
Myosin V stepping



Myosin V stepping

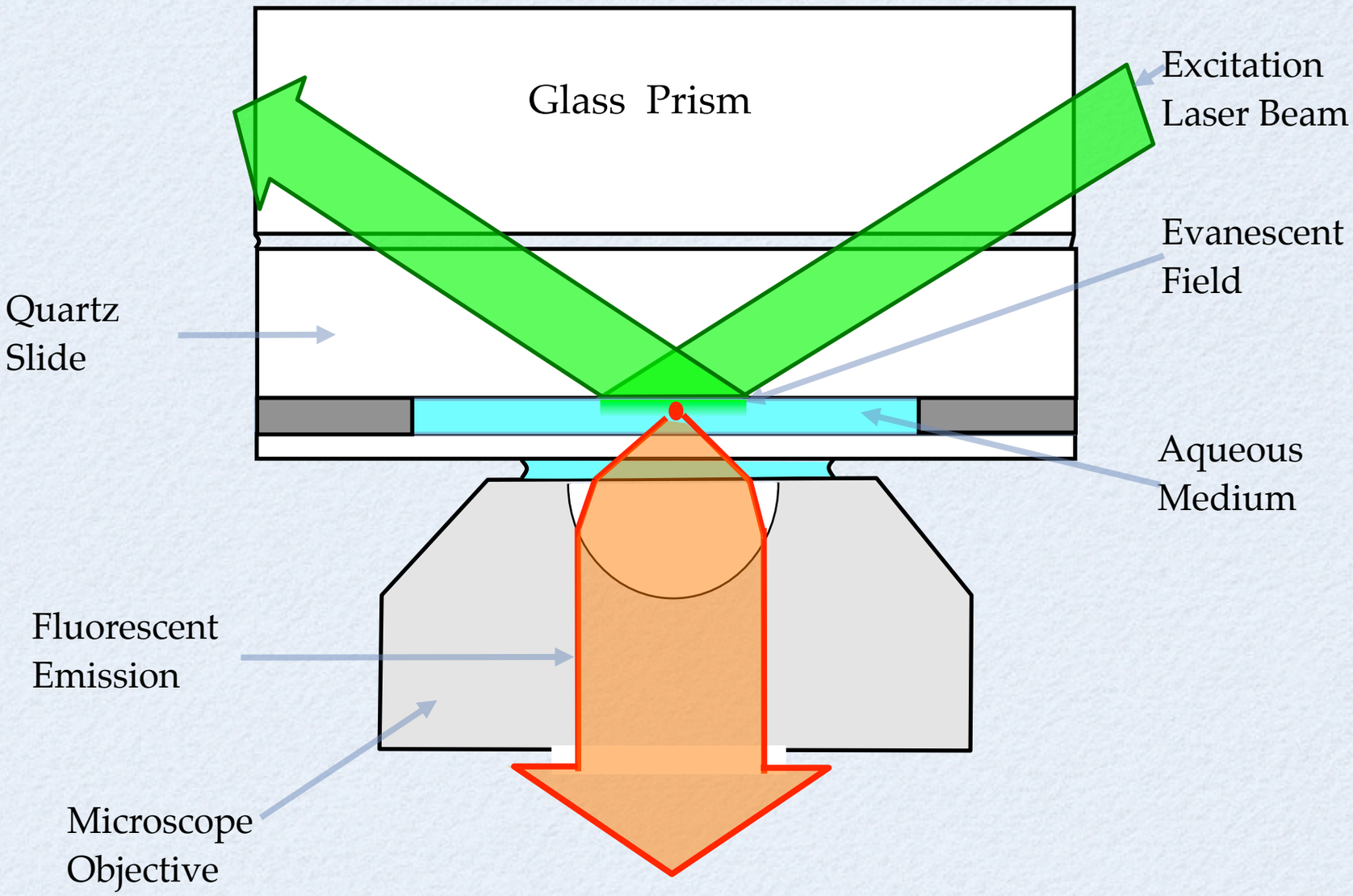






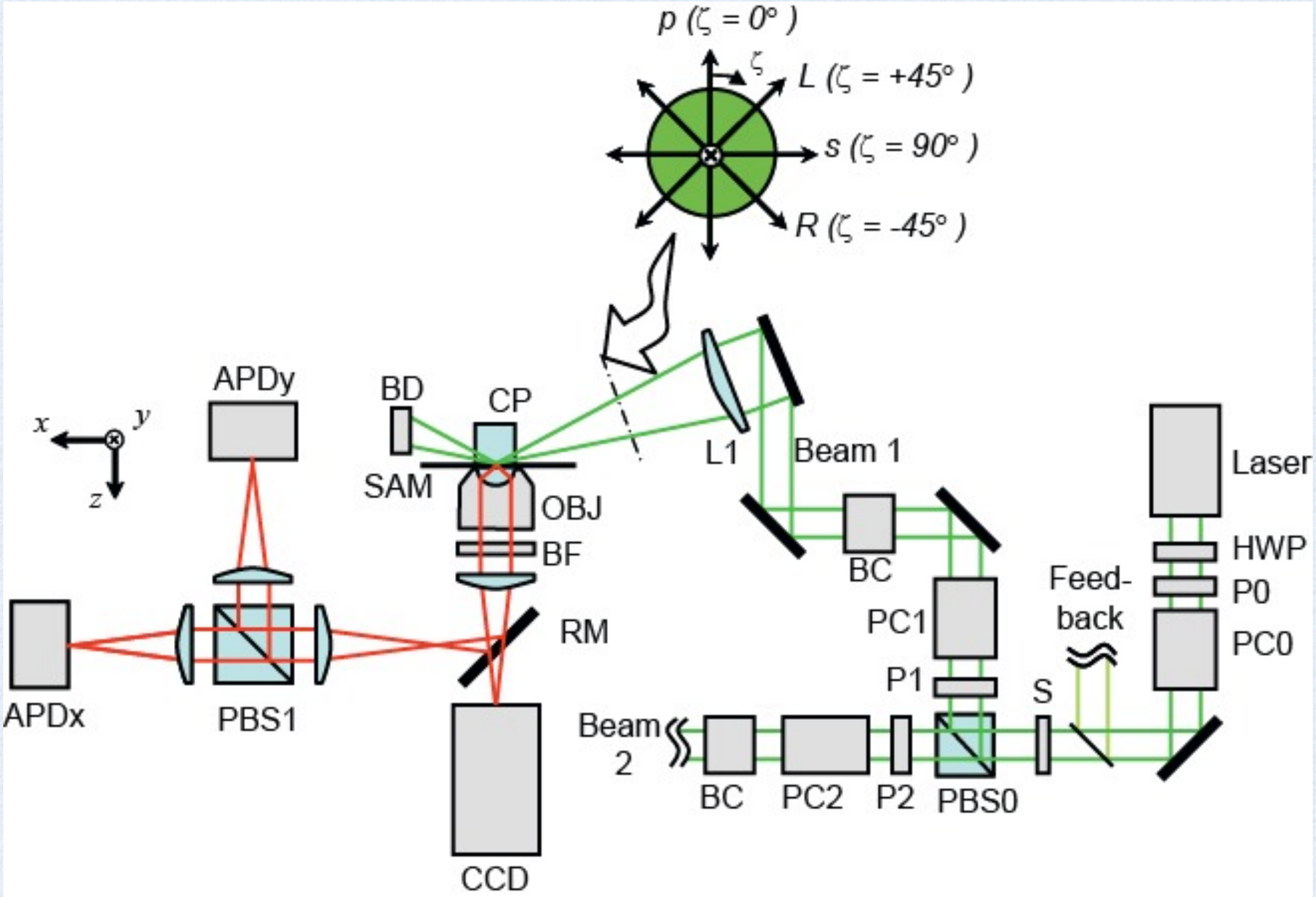
To read out the orientation, we send in polarized light and see how many fluorescence photons, in each polarization, emerge.

Polarized total internal reflection fluorescence microscopy

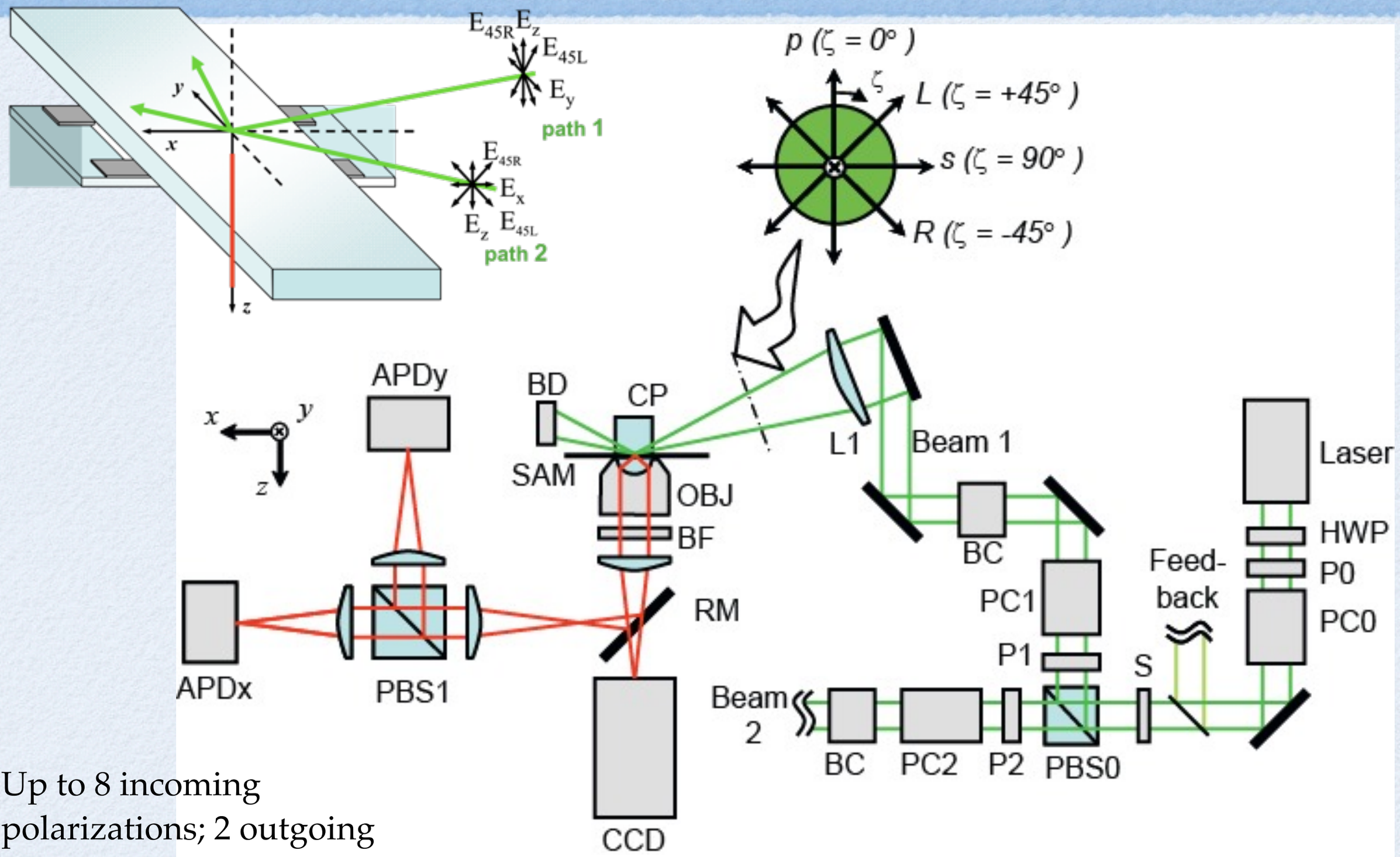


Fluorescence illumination by the evanescent wave eliminates a lot of noise, and importantly, maintains the polarization of the incident light.
To tickle the fluorophore with every possible polarization, we need the incoming light to have at least two different beam directions.

pol-TIRF setup



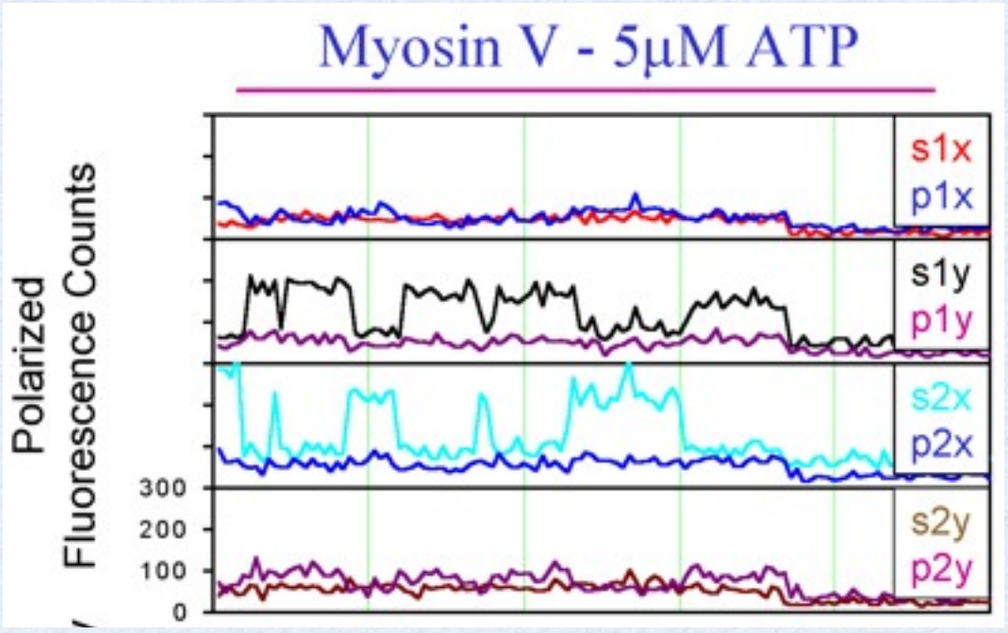
pol-TIRF setup

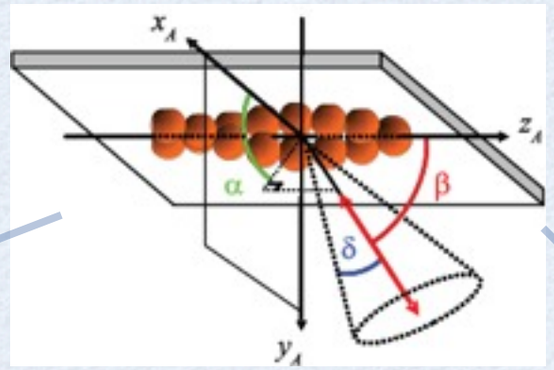


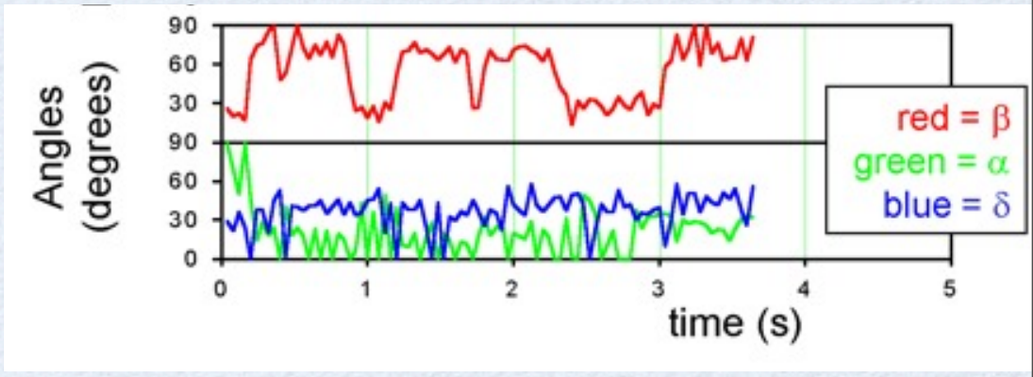
Up to 8 incoming polarizations; 2 outgoing polarizations

Previous state of the art

For our purposes, the upshot is that: ***We need to know the arrival rates of photons in each of several channels.*** Unfortunately, existing analyses gave pretty noisy determinations, with pretty poor time resolution.



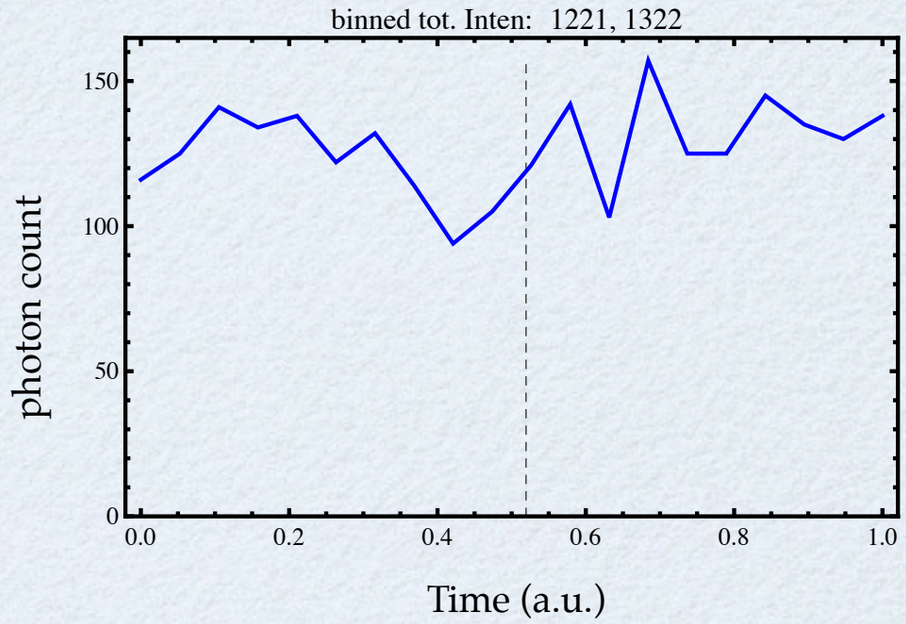


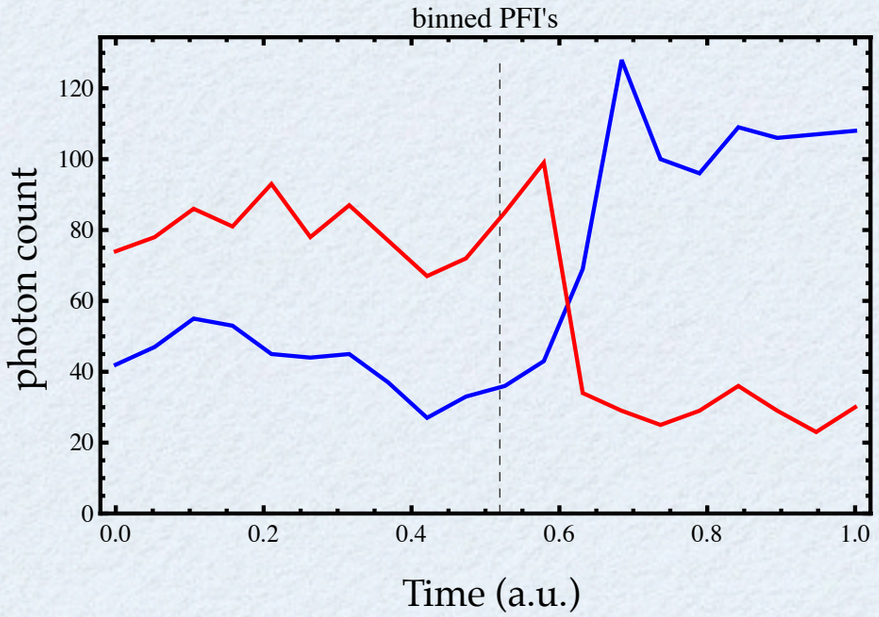


Left: binned photon counts in 8 channels.
Right: Polar and azimuthal angles of the fluorescent label, inferred from data on the left.
You could easily miss a short-lived state -- e.g. the elusive diffusive-search step (if it exists). *Can we do better?*

JN Forkey et al. Nature 2003

The problem is that the total photon counts from a fluorescent probe may not be very informative. Here we divided a time period of interest into 20 bins. There is some Poisson noise in the photon counts.





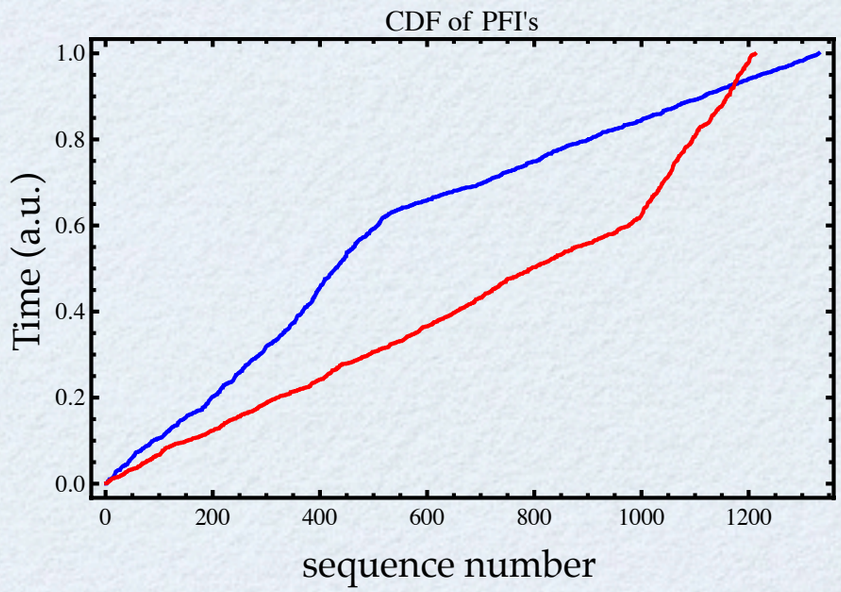
If we classify the photons by polarization and bin them separately, that reveals a definite changepoint. But when exactly did it occur? Probably not at the dashed line shown, but how can we be more precise?

If we choose wider bins, we'll get worse time resolution; if we choose narrower bins, we'll get worse shot-noise errors.

Can we evade the cruel logic of photon statistics?

Key point: ***binning the data destroyed some information.*** Something magical happens if instead of binning, we just plot photon arrival time versus photon sequence number. Despite some ripples from Poisson statistics, it's obvious that each trace has a sharp changepoint, and moreover that the two changepoints found independently in this way are simultaneous.

(A similar approach in the context of FRET was pioneered by Haw Yang.)



JF Beausang, YE Goldman, and PCN, Meth. Enzymol. 487:431 (2011).

Now that I have your attention

- ***Why did that trick work?*** How did we get such great time resolution from such cruddy data?
- ***How well does it work?*** If we have even fewer photons, for example because a state is short-lived, how can we quantify our confidence that any changepoint occurred at all?
- ***Could we generalize and automate this trick?*** Ultimately we'll want to handle data with multiple polarizations, and find lots of changepoints.

Focus on just one "flavor" of photons (e.g. one polarization).
Suppose that in total time T we catch N photons at times t_1,... t_N.
We wish to explore the hypothesis that photons are arriving in a Poisson process with rate R from time 0 to time t_* , and thereafter arrive in another Poisson process with rate R'.
We want to find our best estimates of the three parameters t_*, R, and R', and compare the null hypothesis that there was no change.

To do this, we need the Bayes formula, which involves the "likelihood," the probability that the data we actually observed *would have been observed* in a world described by our model with particular values of the unknown fit parameters:

$$\log P(t_1,\ldots,t_N|R,R',t_*) = \sum_{k=1}^{t_*/\Delta t} \log \begin{cases} R\,\Delta t & \text{if a photon in this slice} \\ (1-R\,\Delta t) & \text{otherwise} \end{cases}$$

$$+ \sum_{k'=t_*/\Delta t+1}^{T/\Delta t} \log \begin{cases} R'\,\Delta t & \text{if a photon in this slice} \\ (1-R'\,\Delta t) & \text{otherwise} \end{cases}$$

From previous slide: In total time T we catch N photons at times t_1,... t_N.
Hypothesis is that photons are arriving in a Poisson process with rate R from time 0 to time t_* , and thereafter arrive in another Poisson process with rate R'.

$$\log P(t_1,\ldots,t_N|R,R',t_*) = \sum_{k=1}^{t_*/\Delta t} \log \begin{cases} R\,\Delta t & \text{if a photon in this slice} \\ (1-R\,\Delta t) & \text{otherwise} \end{cases}$$

$$+ \sum_{k'=t_*/\Delta t+1}^{T/\Delta t} \log \begin{cases} R'\,\Delta t & \text{if a photon in this slice} \\ (1-R'\,\Delta t) & \text{otherwise} \end{cases}$$

Now: Divide the N photons into n that arrived before the putative changepoint, and n'=N-n that arrived after.
Take the limit $\Delta t \to 0$:

$$P \approx N\log(\Delta t) + n\log R + n'\log R' - \left(\frac{t_*}{\Delta t} - n\right)\left(R\Delta t\right) - \left(\frac{T-t_*}{\Delta T} - 1 - (N-n)\right)\left(R'\Delta t\right)$$

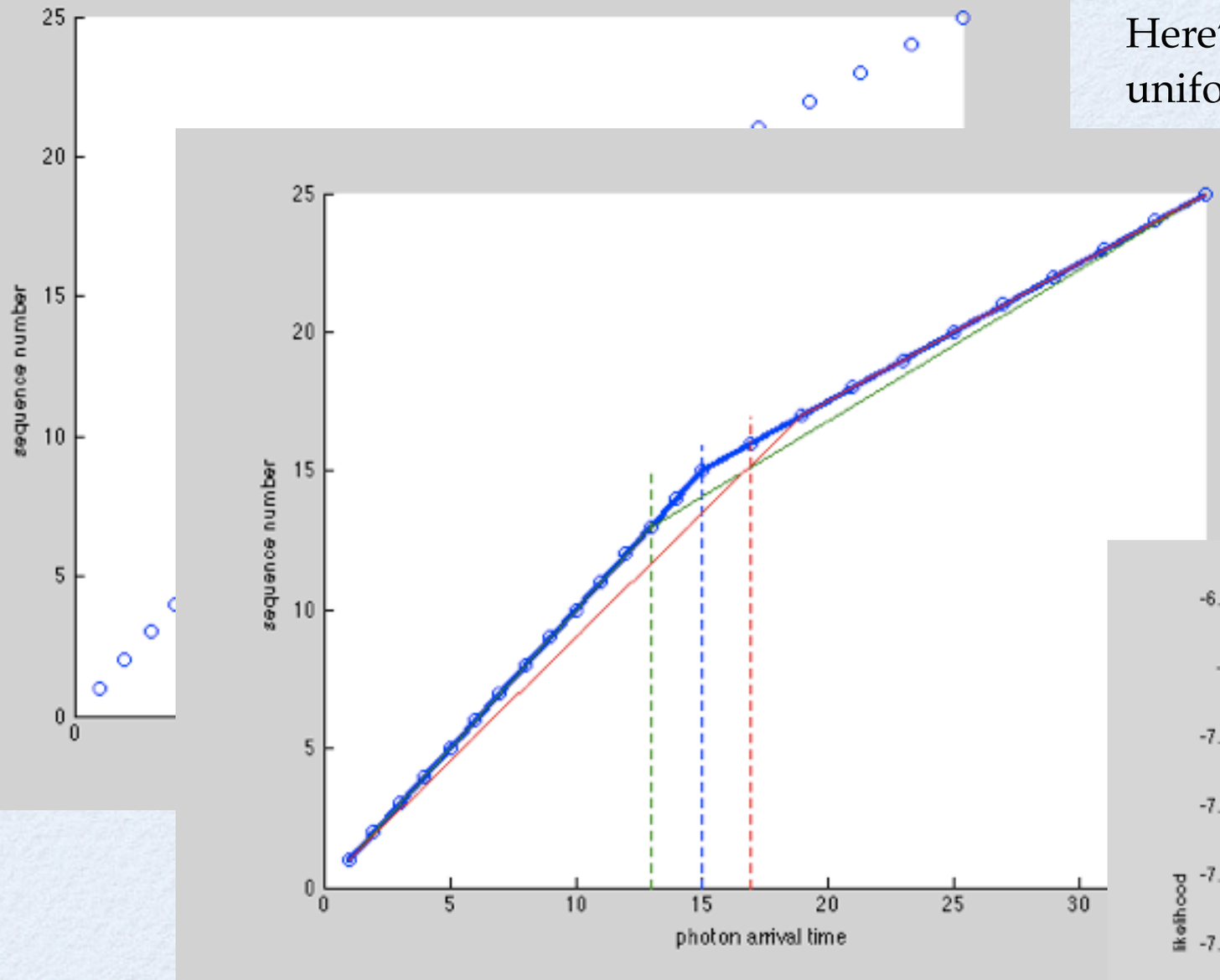
$$\approx \text{const} + n\log R + n'\log R' - Rt_* - R'(T-t_*)$$

Maximize this first over R and R':

$$R = n/t_* \ , \qquad R' = n'/(T-t_*)$$

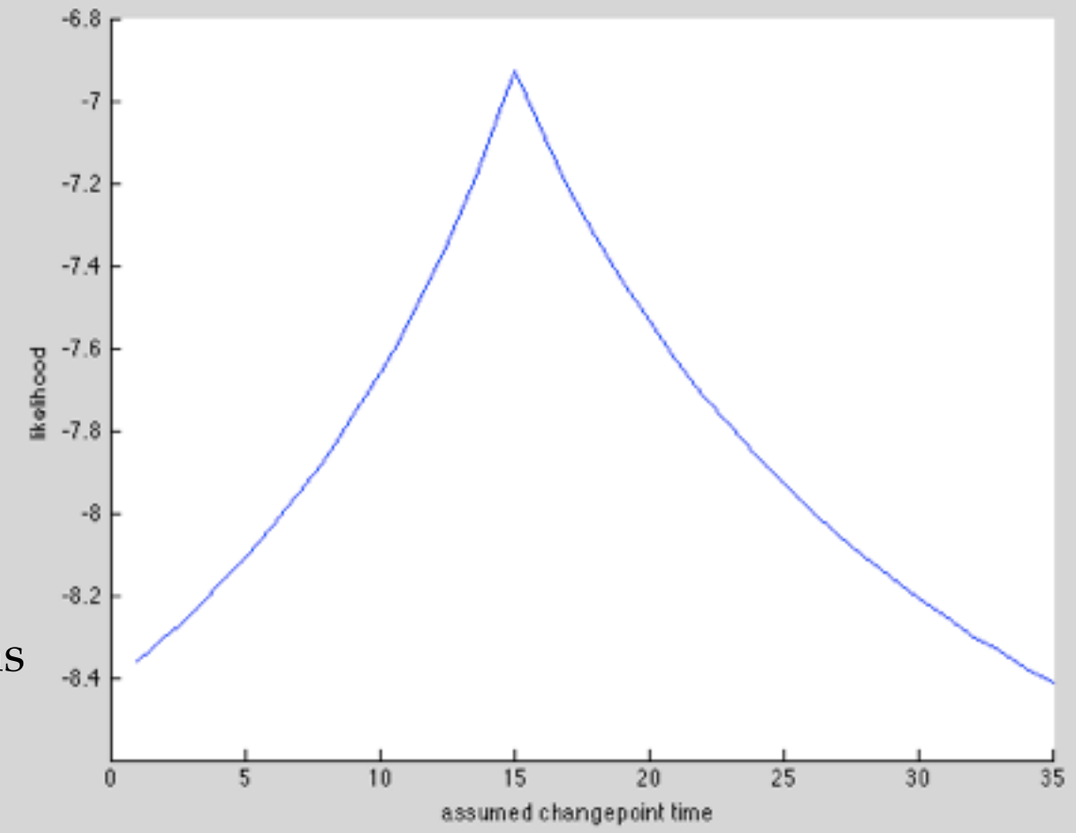
OK, **duh**, that was no surprise! But it does explain why we can just lay a ruler along the cumulative plot to get our best estimate of the before and after rates.
More interestingly, we can substitute these optimal rates into the formula for P to find the likelihood as a function of putative changepoint:

Illustration

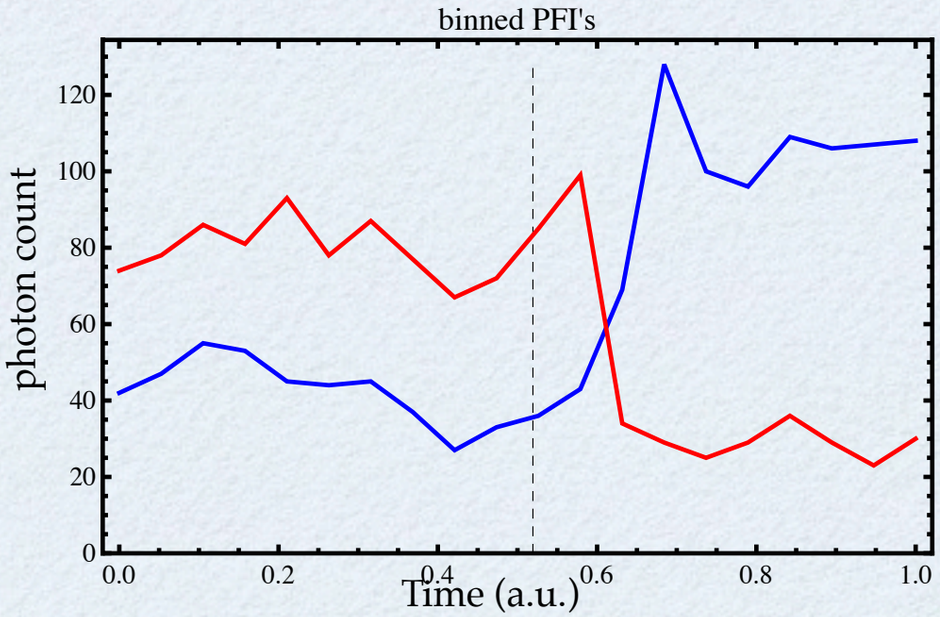


Here's some very fake data; the photons arrive uniformly, not at random.

Here are two lines corresponding to non-optimal choices of the changepoint. We'd like to see the likelihood function and how it selects the "right" changepoint, which for fake data is known.

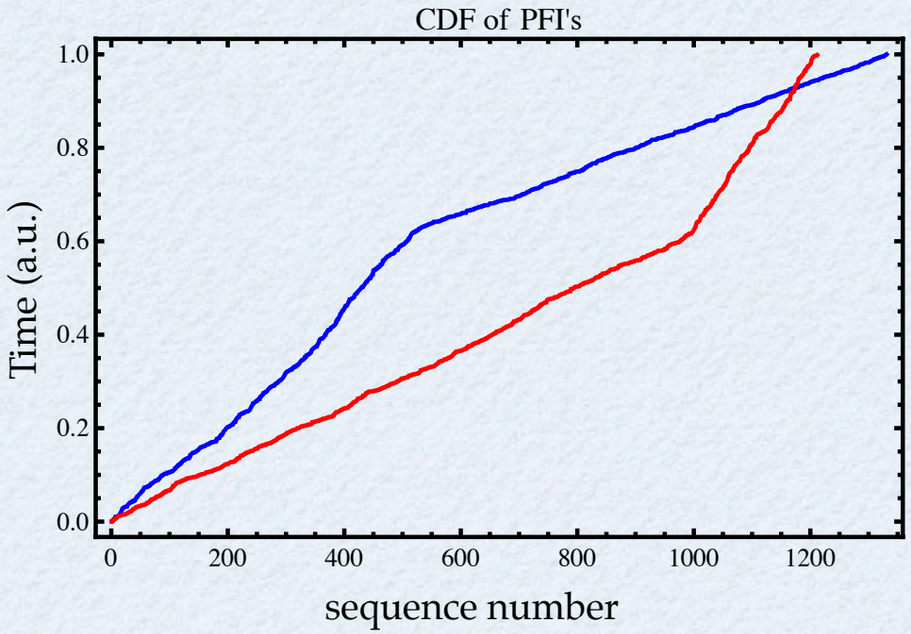


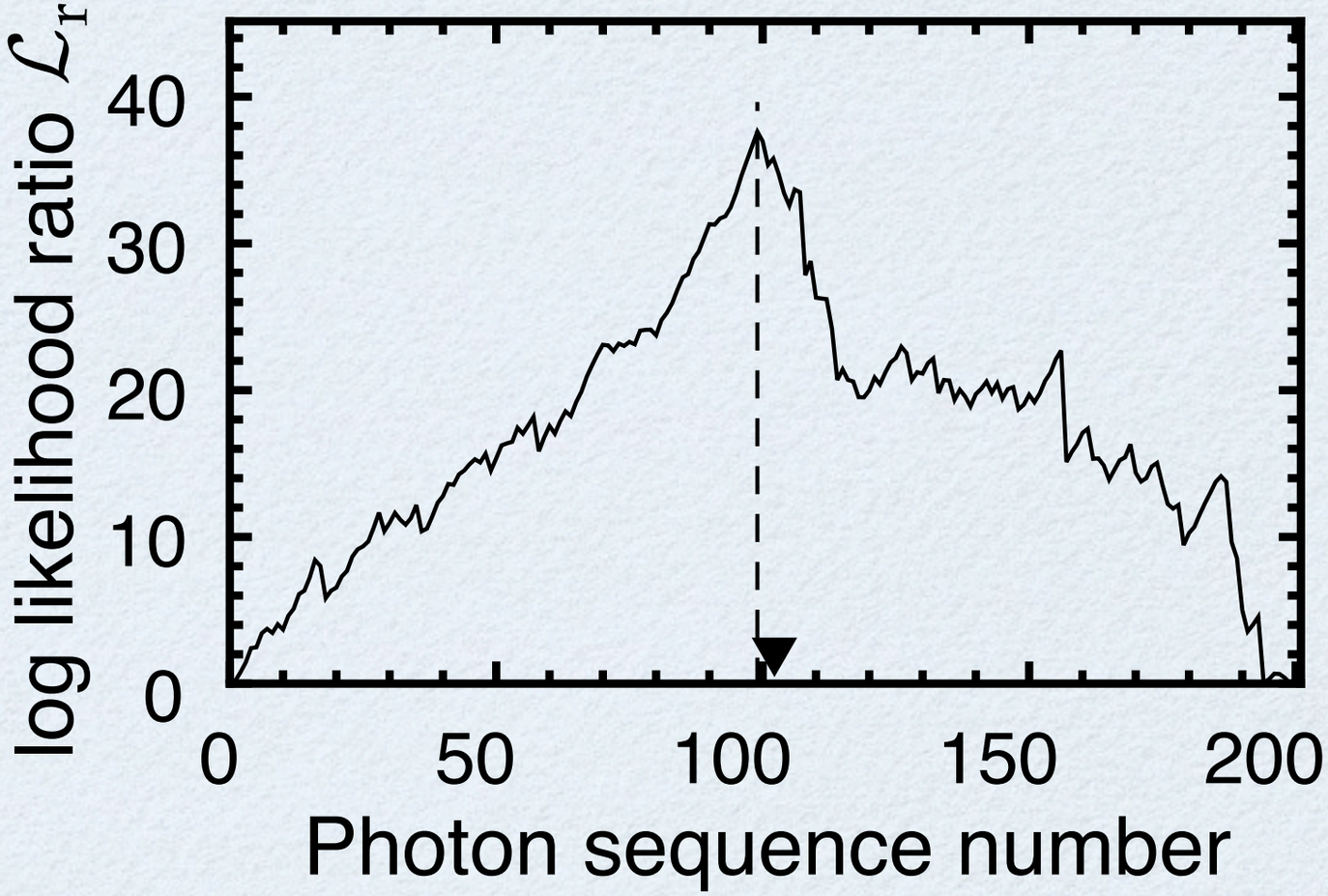
Our log-likelihood function has a huge peak as a function of putative changepoint time.



Left: Some data, shown in traditional binned form and in the improved version.

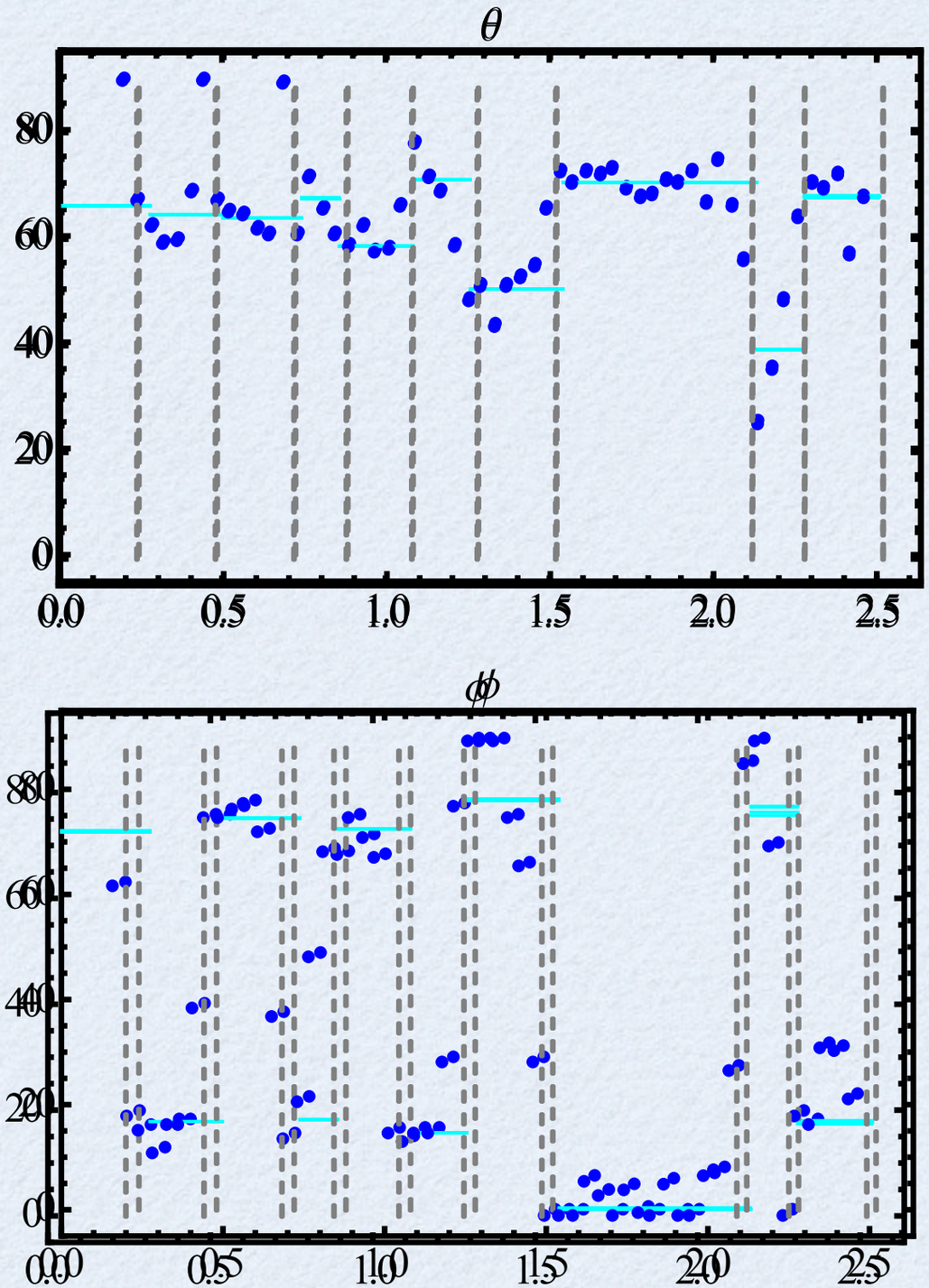
Below: Likelihood function for placement of the changepoint. Dashed line, maximum-likelihood point. Black triangle: Actual changepoint used to generate the simulated data. The analysis found a robust changepoint, even though there were a total of just 200 photons in the entire dataset.





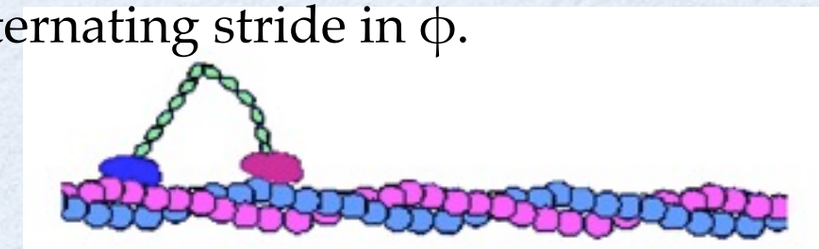
JF Beausang, YE Goldman, and PCN, Meth. Enzymol. 487:431 (2011).

Payoff



Oh, yes--it also works on real experimental data, multiple-channel data, data with many different changepoints...
Previously, people would take data from multiple polarizations, bin it, and pipe the inferred intensities into a maximum-likelihood estimator of the orientation of the fluorophore.
That procedure leads to the rather noisy dots shown here.
One problem is that if a transition happens in the middle of a time bin, then the inferred orientation in that time bin can be crazy.

Here the solid lines are the inferred orientations of the probe molecule during successive states defined by changepoint analysis. We see a nice alternating stride in ϕ.



We got a ***50-fold improvement*** in time resolution for finding changepoints, compared to the binning method, ***without changing the apparatus.***

JF Beausang, YE Goldman, and PCN, Meth. Enzymol. (2011); JF Beausang, DY Shroder, PCN, and YE Goldman, Biophys J (2013).

Summary Part III

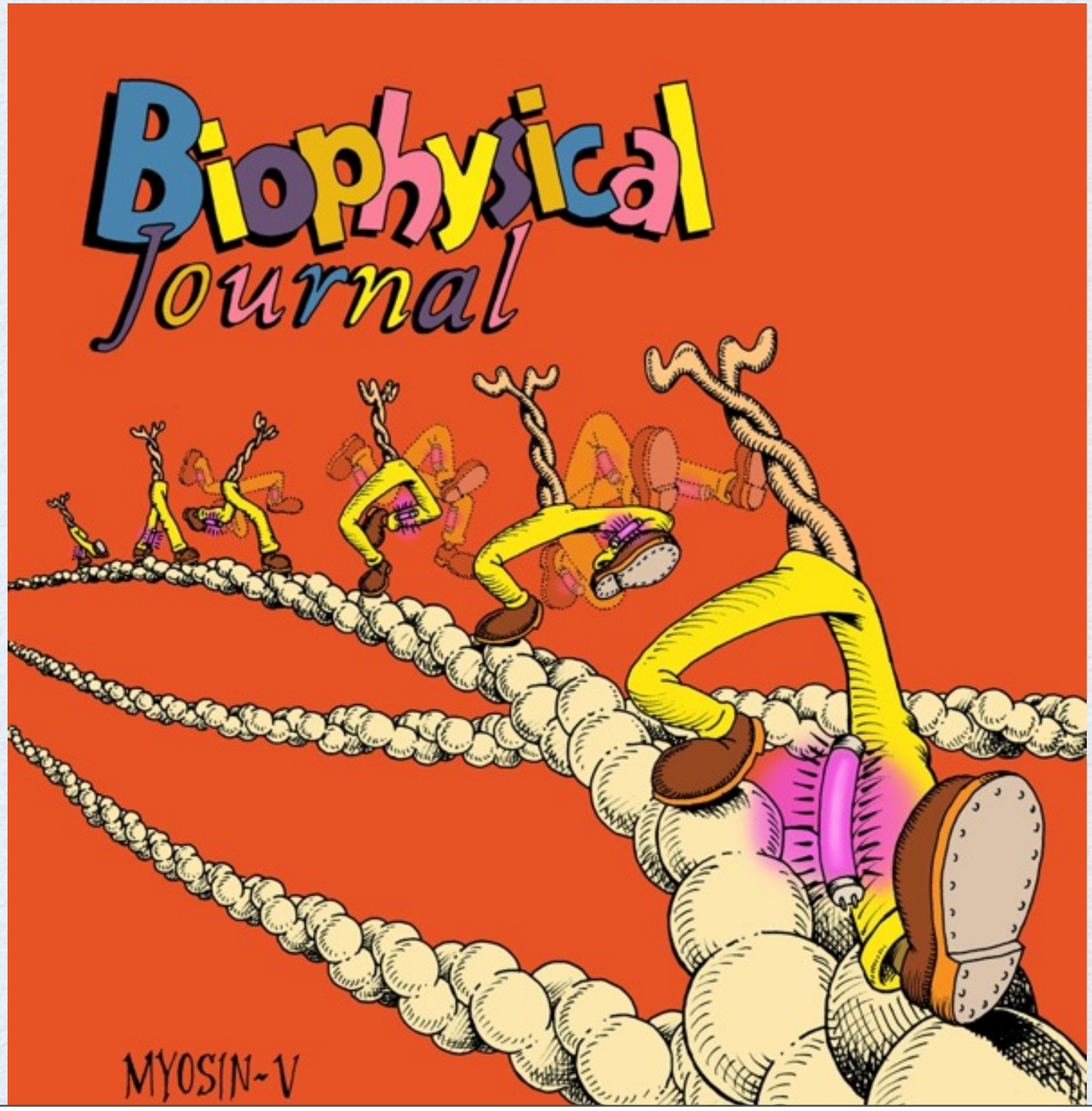
- When you only get a million photons, you'd better make every photon count.

- A simple maximum-likelihood analysis accomplishes this.

- In the context of TIRF it can dramatically improve the tradeoff between time resolution and accuracy.

- That can help you find substeps, like the diffusive-search step in myosin-V's kinetic scheme.

JF Beausang, DY Shroder, PCN, and YE Goldman, Biophys J (2013).

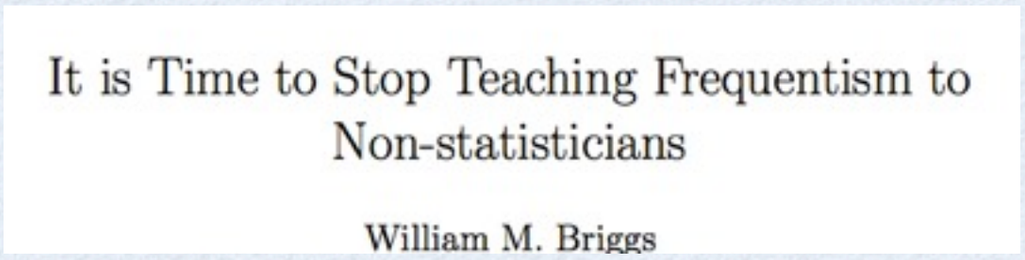


Go long

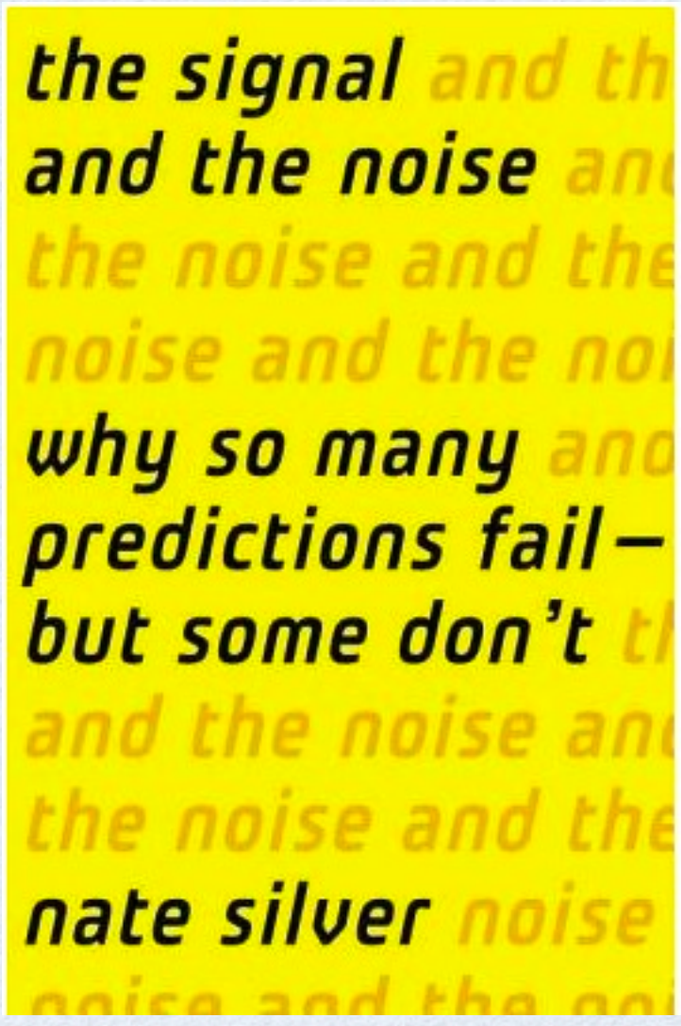

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

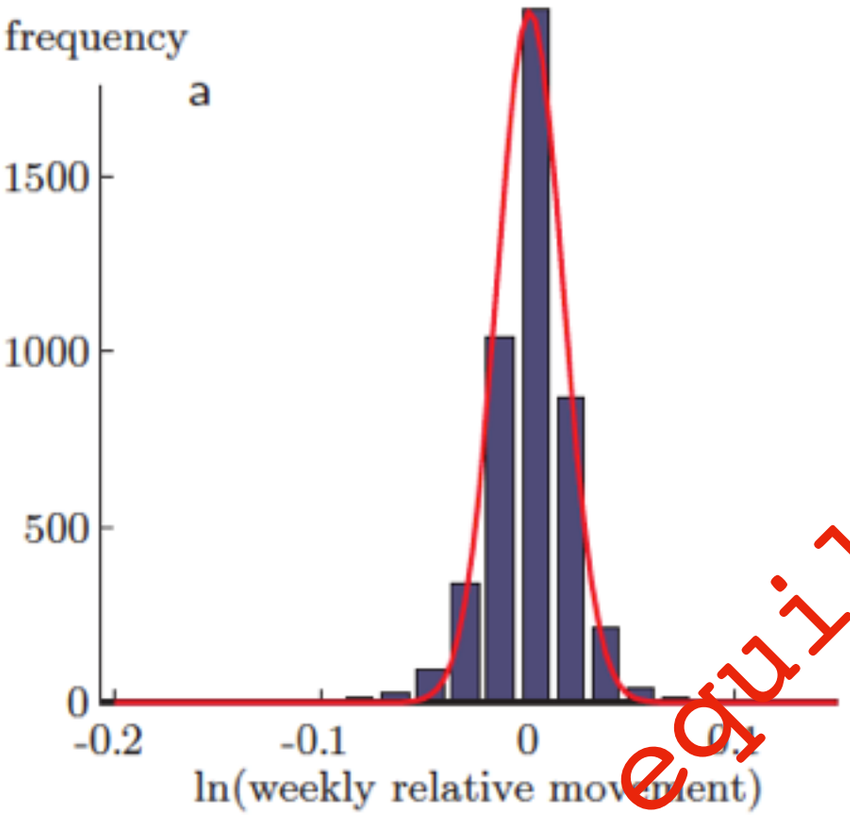

It is Time to Stop Teaching Frequentism to Non-statisticians

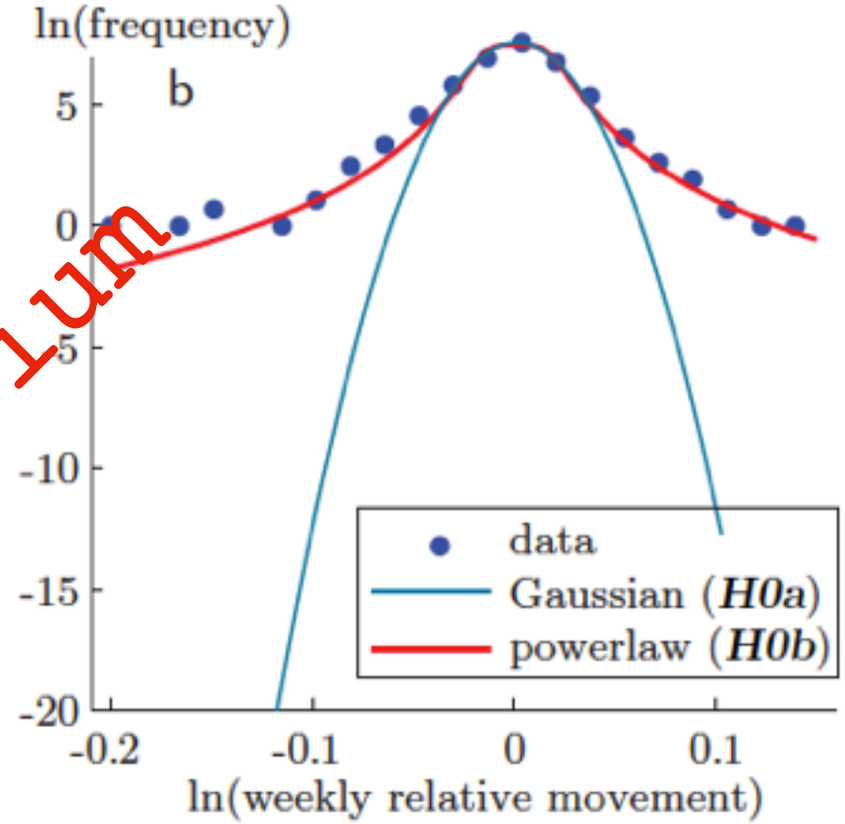
William M. Briggs



"Overfitting: The biggest scientific problem you've never heard of."

Power laws are ubiquitous





OK, this is not biophysics. But it would be the same figure if I were talking about enzyme turnovers.

How much weight should I give to data in the tail? Depends on how reliable they are, but also on what the model says there. Different points convey different amounts of information.

Likelihood takes care of all that.

Again... these two models have vastly different likelihood scores.

P. Nelson, Physical models of living systems (WH Freeman and Co 2015)

Go longer

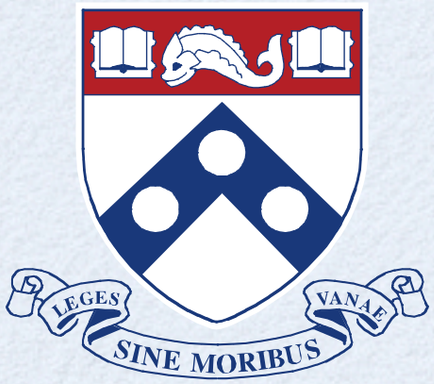
Often, when we want to justify physical modeling, we scratch our heads and say, "Well Hodgkin and Huxley was a big deal."
Indeed. But that sort of cherry-picking approach can leave the impression that this is something that happens every 50 years or so. It's also too reverent.

Another context in which theory enters laboratory discussions is, "We need some theory to get this thing published. Go do some theory, run some ANOVA, whatever."
I'd just like to suggest that this attitude, though common, misses out on some of what theory can do for you.

Getting back to the opening point, I'd say ***theory is needed every day***. It's our microscope; our Geiger counter; it helps us to see the invisible. It squeezes out the information from the data. To emphasize that, I didn't select famous examples; instead I have told you about the two things I happen to be working on right now (a random choice, you'll agree).

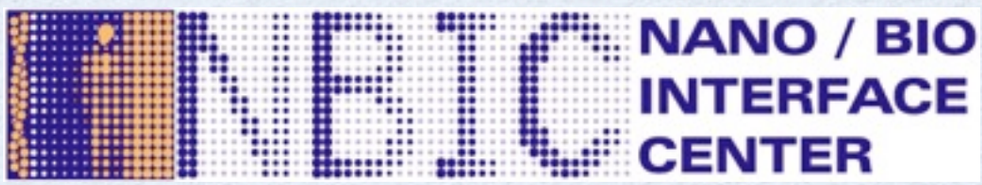
We like to teach famous success stories in science. When appropriate, let's remember to present them as showcases of the utility of physical modeling. Let's communicate to our students the "unreasonable effectiveness of physical models in biology."

Thanks



University of Pennsylvania

NSF DMR,
BIO, IBN



NSF NSEC

For these slides see:
www.physics.upenn.edu/~pcn

(Things I wish I'd read earlier)

M. Denny and S. Gaines
Chance in biology
Princeton Univ. Press 2000

D. S. Sivia and J. Skilling
Data analysis: A Bayesian tutorial
Oxford Univ. Press 2006

P. Nelson
Physical models of living systems
WH Freeman and Co 12/26/2014

H. J. C. Berendsen
A student's guide to data and error analysis
Cambridge Univ. Press 2011

Louis Lyons
Discovery or fluke: statistics in particle physics
Physics Today, 2012

P. Nelson and Tom Dodson
Student's guide to physical modeling in MATLAB
Free at
www.physics.upenn.edu/biphys/PMLS

Jesse Kinder and P. Nelson
Student's guide to physical modeling in Python
Coming soon to lulu.com

For these slides see:
www.physics.upenn.edu/~pcn