When you see a title like that, you need to worry: “Uh-oh, sounds philosophical.”
Well, I just wanted to tell you two concrete stories about cases when my colleagues and I managed to do something useful by virtue of knowing something about inference. The ideas we needed were things I didn’t know a few years ago, so I thought you might be interested too.

Inference in biological physics

Phil Nelson
University of Pennsylvania

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

we need to make sure we don't forget to tell our undergraduates why theory exists. (Are there any students here today?) Remember, to them "theory" is a strongly negative word. ("How's it going in Physics N? Too much theory.") To them it means "stuff that won't get me a job because it's not relevant to anything practical." To them it's at best decorative, like music or art. If we think there's more to it than that, we'd better not forget to tell them

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

"Of course that's what we do -- everybody knows that." -- No. They don't.

we need to make sure we don't forget to tell our undergraduates why theory exists. (Are there any students here today?) Remember, to them "theory" is a strongly negative word. ("How's it going in Physics N? Too much theory.") To them it means "stuff that won't get me a job because it's not relevant to anything practical." To them it's at best decorative, like music or art. If we think there's more to it than that, we'd better not forget to tell them

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

"Of course that's what we do -- everybody knows that." -- No. They don't.

we need to make sure we don't forget to tell our undergraduates why theory exists. (Are there any students here today?) Remember, to them "theory" is a strongly negative word. ("How's it going in Physics N? Too much theory.") To them it means "stuff that won't get me a job because it's not relevant to anything practical." To them it's at best decorative, like music or art. If we think there's more to it than that, we'd better not forget to tell them

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

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

"Of course that's what we do -- everybody knows that." -- No. They don't.

we need to make sure we don't forget to tell our undergraduates why theory exists. (Are there any students here today?) Remember, to them "theory" is a strongly negative word. ("How's it going in Physics N? Too much theory.") To them it means "stuff that won't get me a job because it's not relevant to anything practical." To them it's at best decorative, like music or art. If we think there's more to it than that, we'd better not forget to tell them

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

Well... Statistical inference sounds "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.

"Of course that's what we do -- everybody knows that." -- No. They don't.

we need to make sure we don't forget to tell our undergraduates why theory exists. (Are there any students here today?) Remember, to them "theory" is a strongly negative word. ("How's it going in Physics N? Too much theory.") To them it means "stuff that won't get me a job because it's not relevant to anything practical." To them it's at best decorative, like music or art. If we think there's more to it than that, we'd better not forget to tell them

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

Well... Statistical inference sounds "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.

"Of course that's what we do -- everybody knows that." -- No. They don't.

we need to make sure we don't forget to tell our undergraduates why theory exists. (Are there any students here today?) Remember, to them "theory" is a strongly negative word. ("How's it going in Physics N? Too much theory.") To them it means "stuff that won't get me a job because it's not relevant to anything practical." To them it's at best decorative, like music or art. If we think there's more to it than that, we'd better not forget to tell them

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

Well... Statistical inference sounds "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.

"Of course that's what we do -- everybody knows that." -- No. They don't.

we need to make sure we don't forget to tell our undergraduates why theory exists. (Are there any students here today?) Remember, to them "theory" is a strongly negative word. ("How's it going in Physics N? Too much theory.") To them it means "stuff that won't get me a job because it's not relevant to anything practical." To them it's at best decorative, like music or art. If we think there's more to it than that, we'd better not forget to tell them

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

Well... Statistical inference sounds "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 t

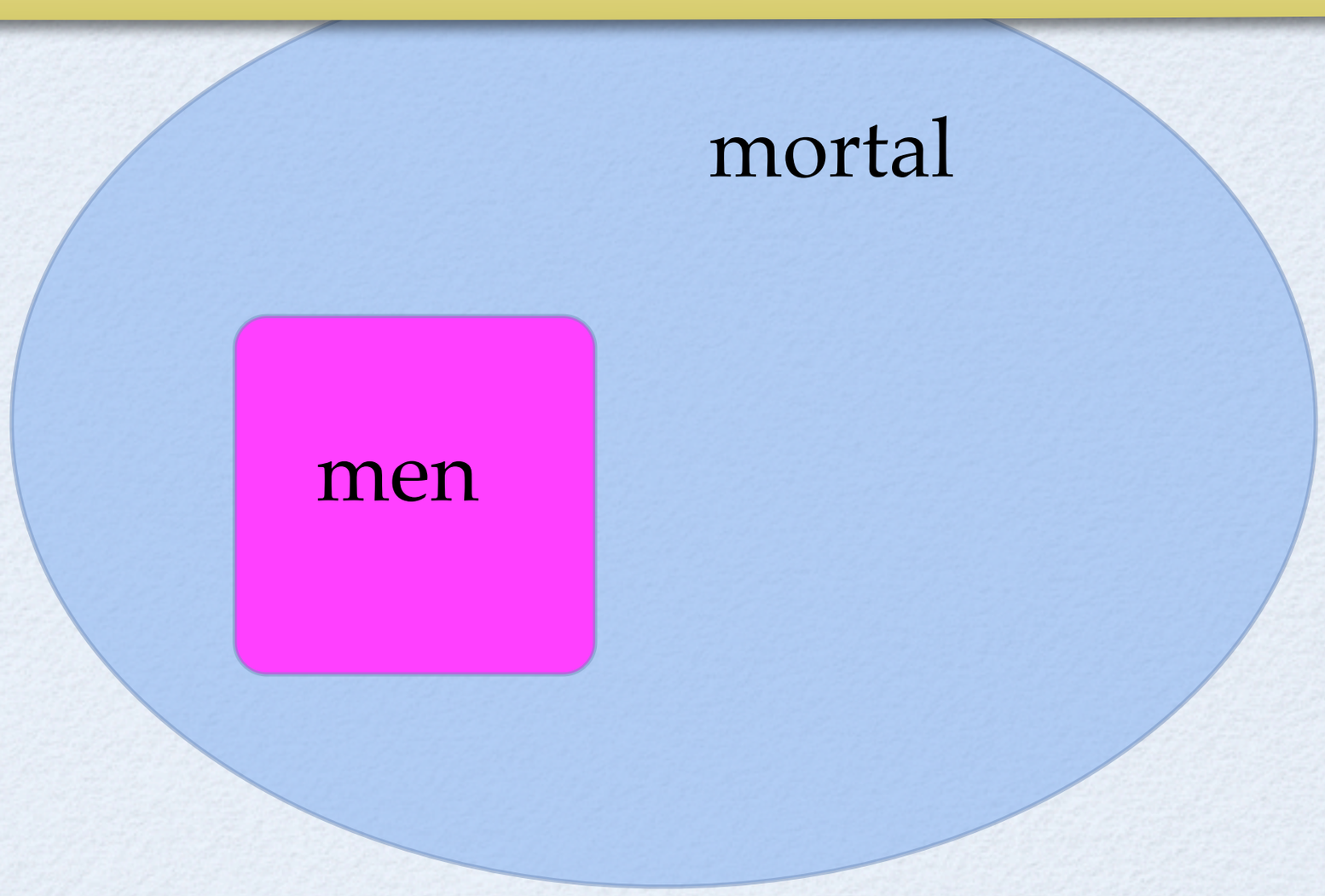
"Of course that's what we do -- everybody knows that." -- No. They don't.

Part I:

start with a topic that may not be obviously biophysical in character.
Suppose I stood here and said "all men are mortal; Socrates is mortal; therefore Socrates is a man"

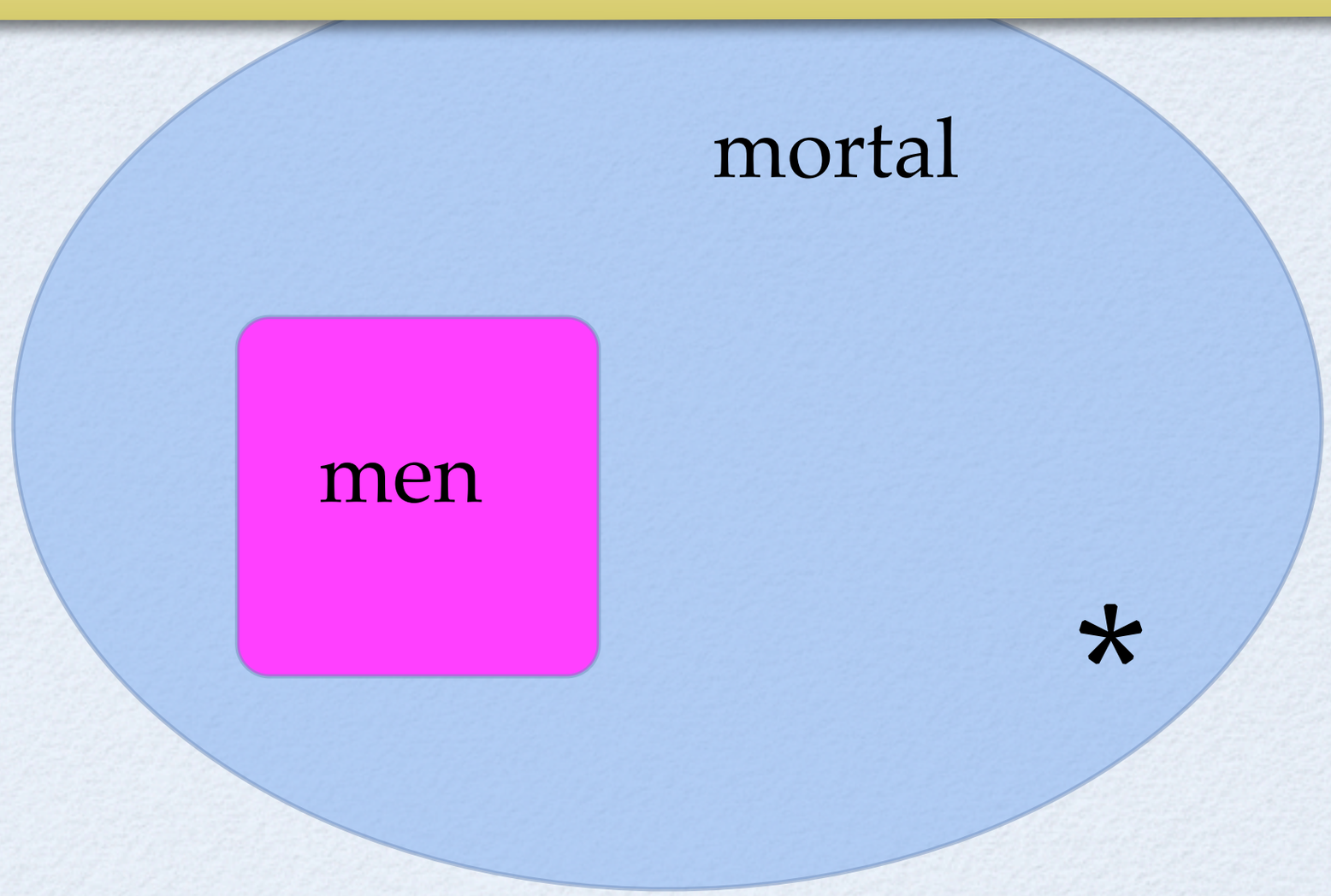
Part I:

start with a topic that may not be obviously biophysical in character.
Suppose I stood here and said "all men are mortal; Socrates is mortal; therefore Socrates is a man"



Part I:

start with a topic that may not be obviously biophysical in character.
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.

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

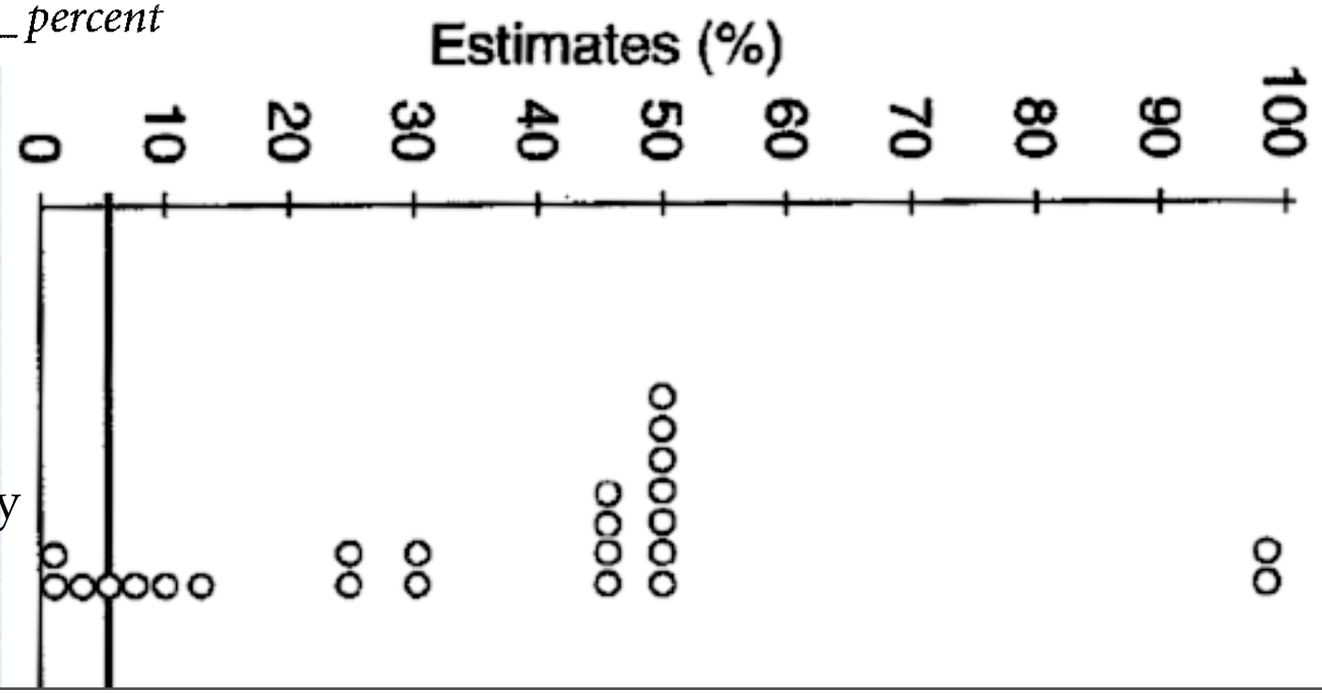
G. Gigerenzer, *Calculated risks*

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

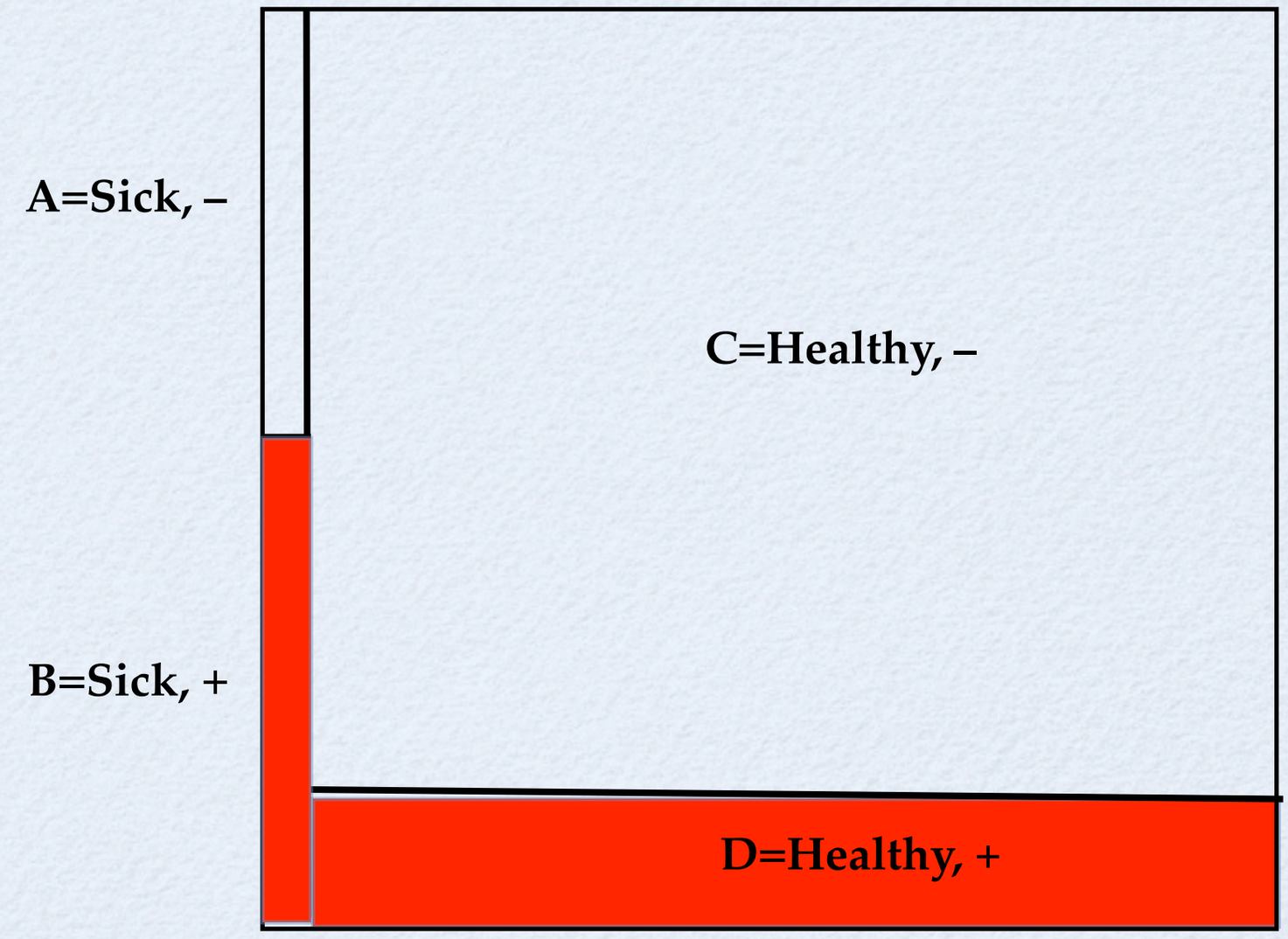
Here are the replies of 24 practicing physicians, who had an average of 14 years of professional experience:



G. Gigerenzer, *Calculated risks*

Work it out

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



Work it out

We are asked for $P(\text{sick} \mid +) = B/(B+D)$.

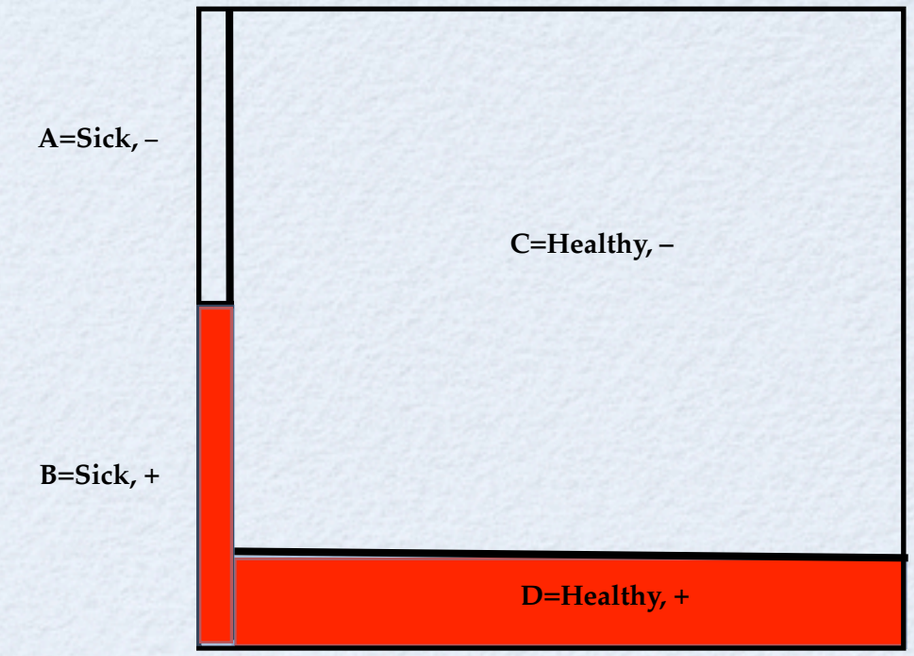
Work it out

We are asked for $P(\text{sick} \mid +) = B/(B+D)$.

Work it out

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

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



Work it out

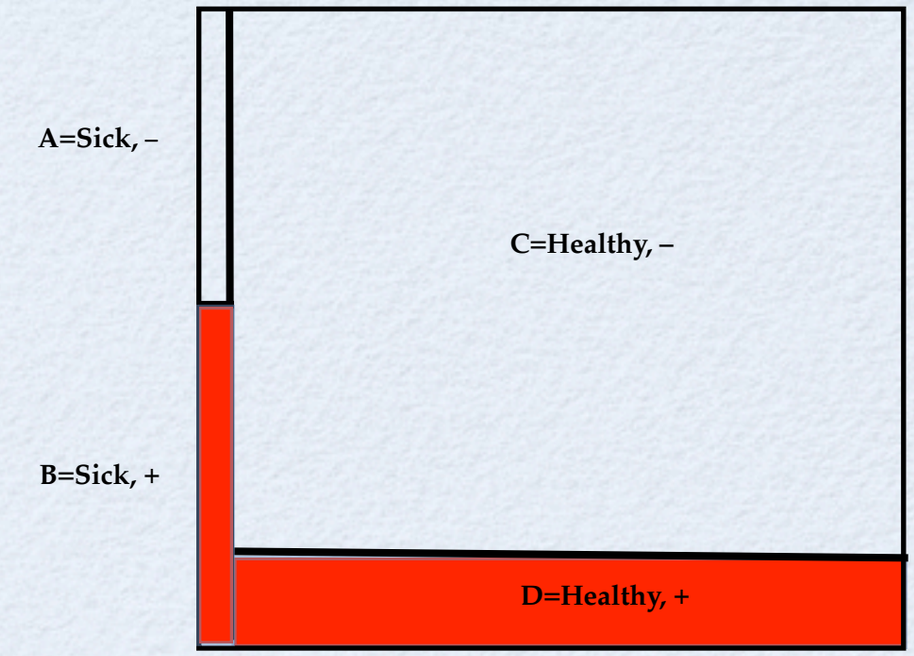
We are asked for $P(\text{sick} \mid +) = B/(B+D)$.

But what we were given was $P(+ \mid \text{sick}) = B/(A+B)$.

Work it out

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

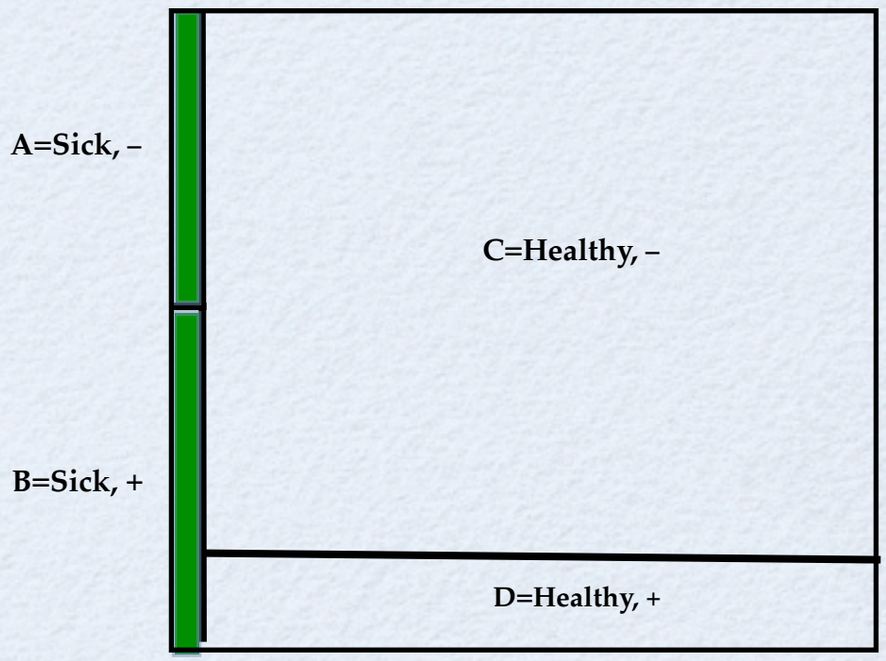
But what we were given was P(+ | sick) = B/(A+B).



Work it out

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

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



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:

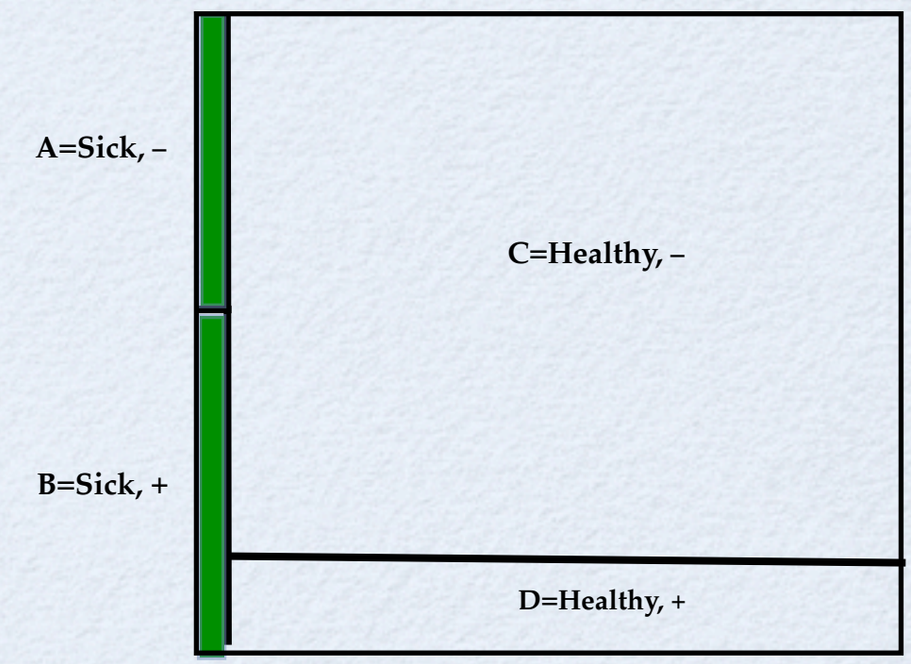
Work it out

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:

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



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:

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

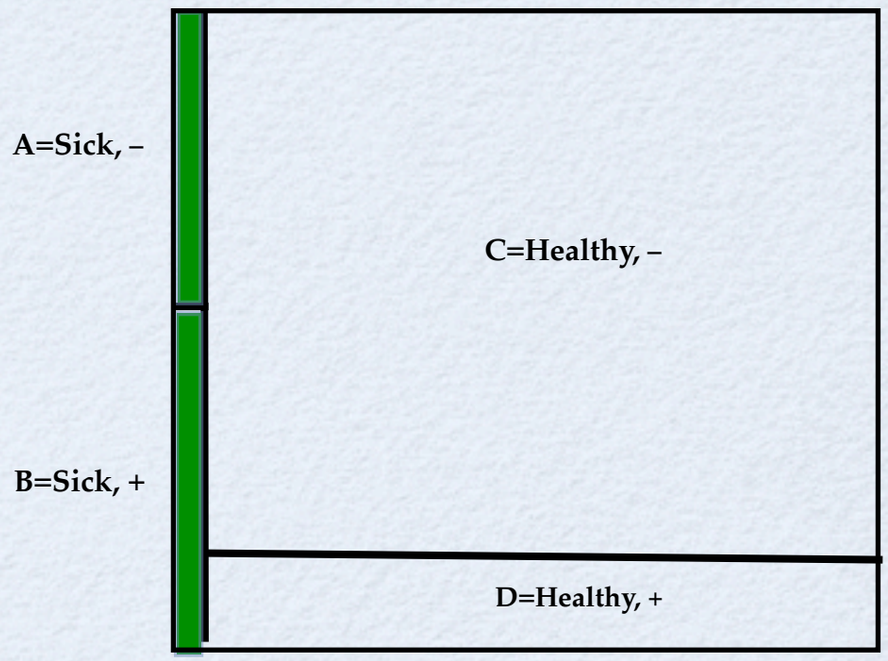
$$P(\text{sick}|+) = P(+|\text{sick}) \times \frac{P(\text{sick})}{P(+)}$$

Posterior estimate (desired)

Likelihood (given)

Prior estimate (given)

Still need this



Finish working it out

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

Is that last factor a big deal?

P(sick) was given, but we need:

Finish working it out

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

Is that last factor a big deal?

P(sick) was given, but we need:

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

Finish working it out

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

Is that last factor a big deal?

P(sick) was given, but we need:

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

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

It's huge: a positive test result means only a 5% chance you're sick. *Not* 97%.

Part II: Changepoint analysis in single-molecule TIRF

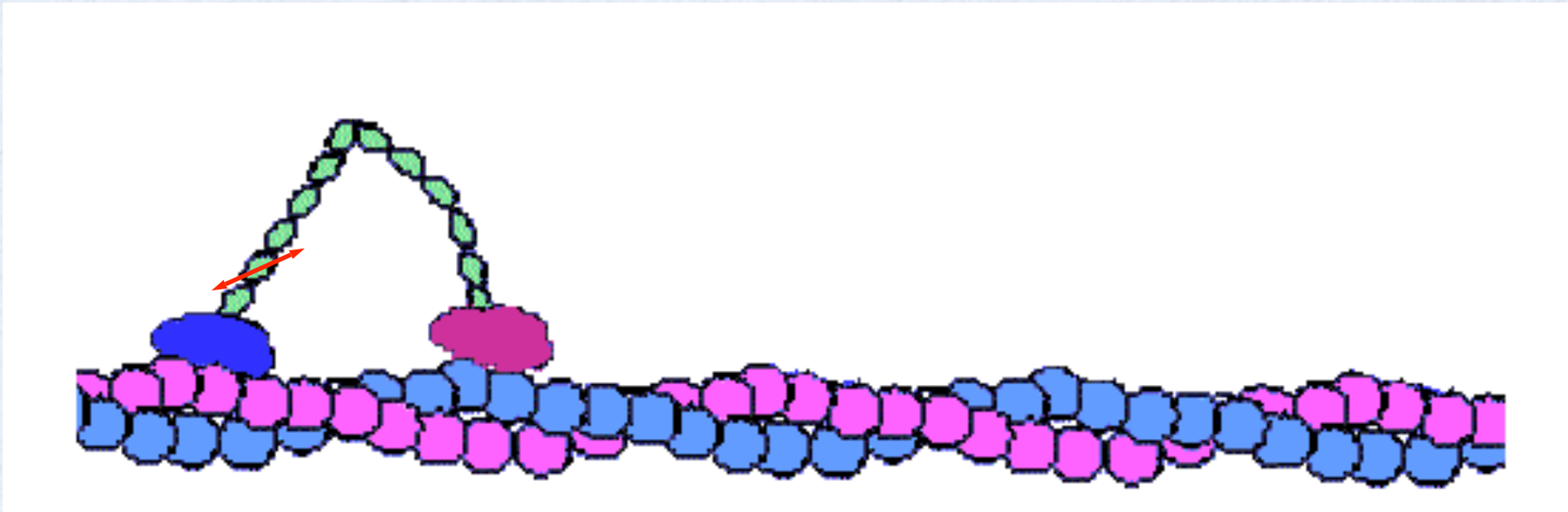
JF Beausang, Yale Goldman, PN

- **Sometimes our model is not obviously connected with what we can actually measure experimentally, but 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**

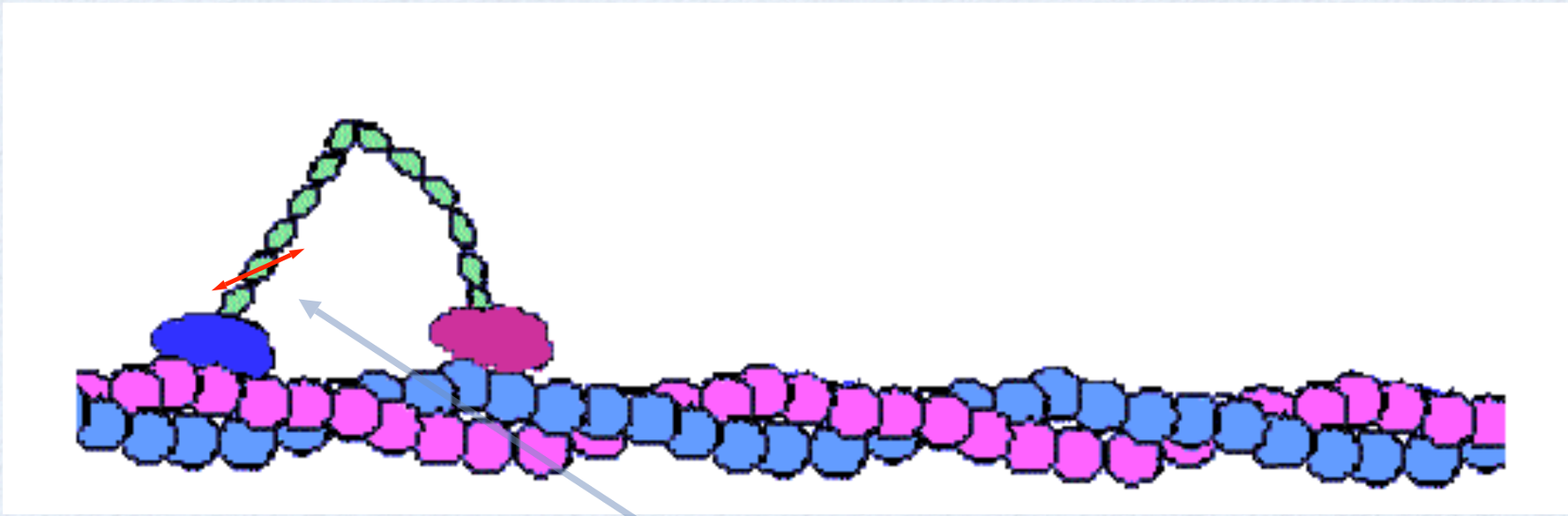
Myosin V Processivity

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.



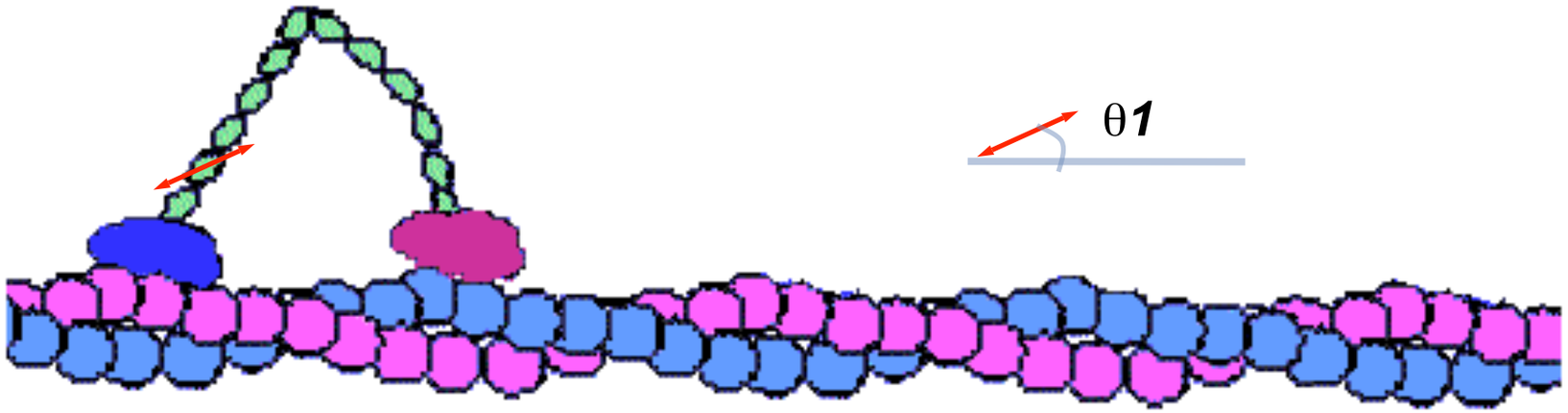
Myosin V Processivity

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.

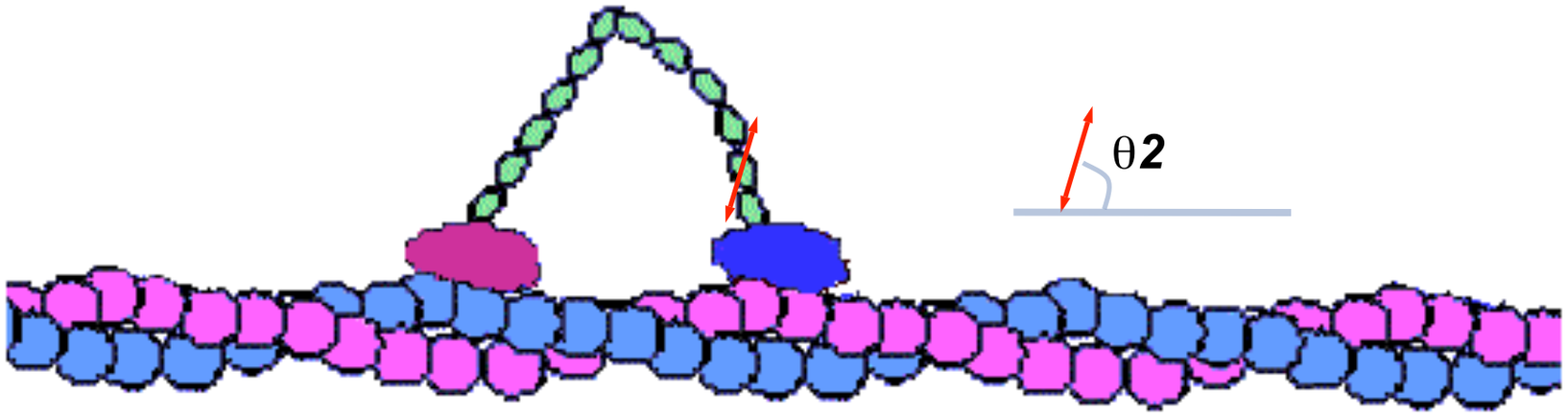


The approach I'll discuss involves attaching a bifunctional fluorescent label to one lever arm. The label has a dipole moment whose *orientation* in space reflects that of the arm.

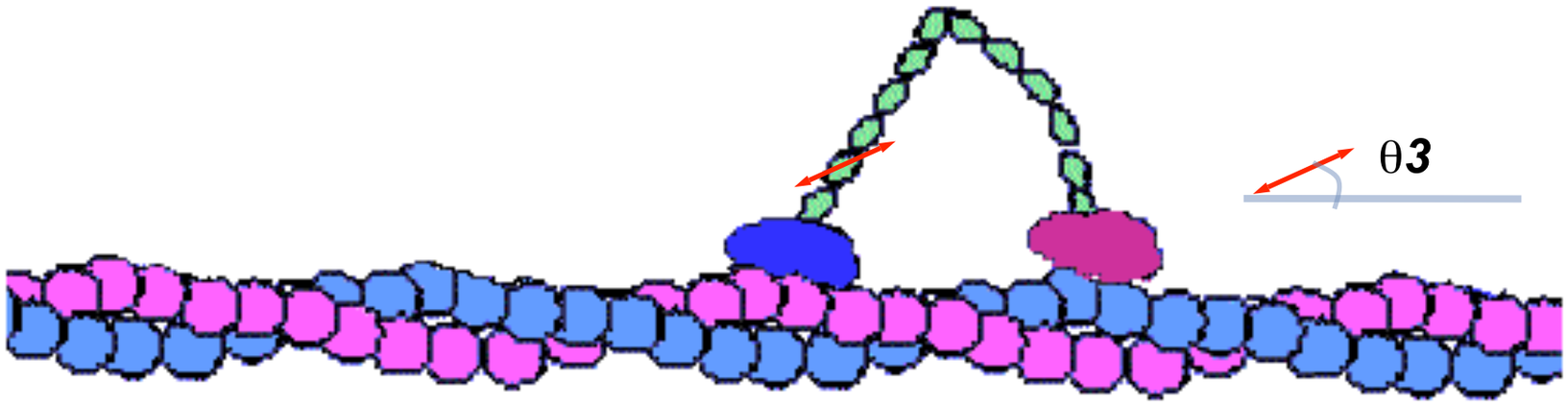
Myosin V Processivity



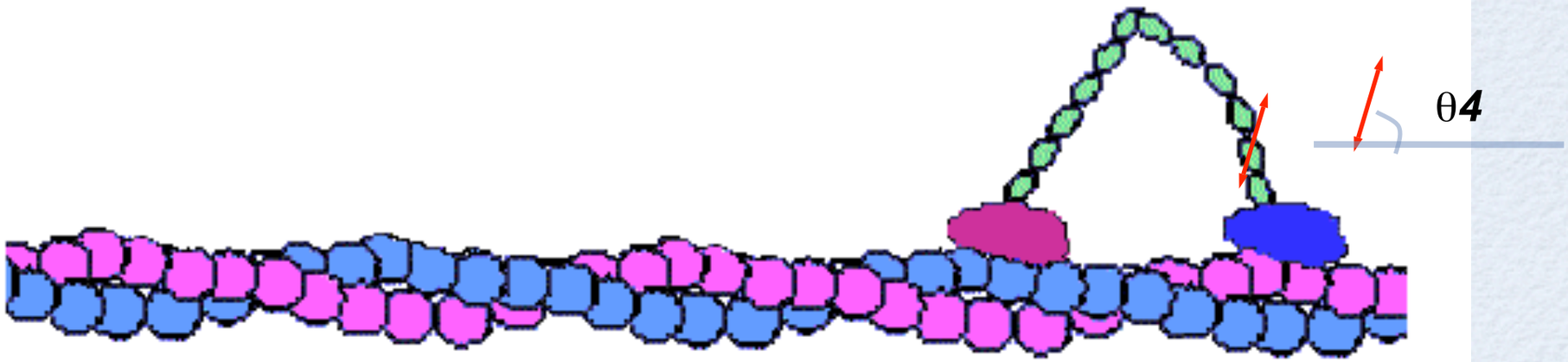
Myosin V Processivity



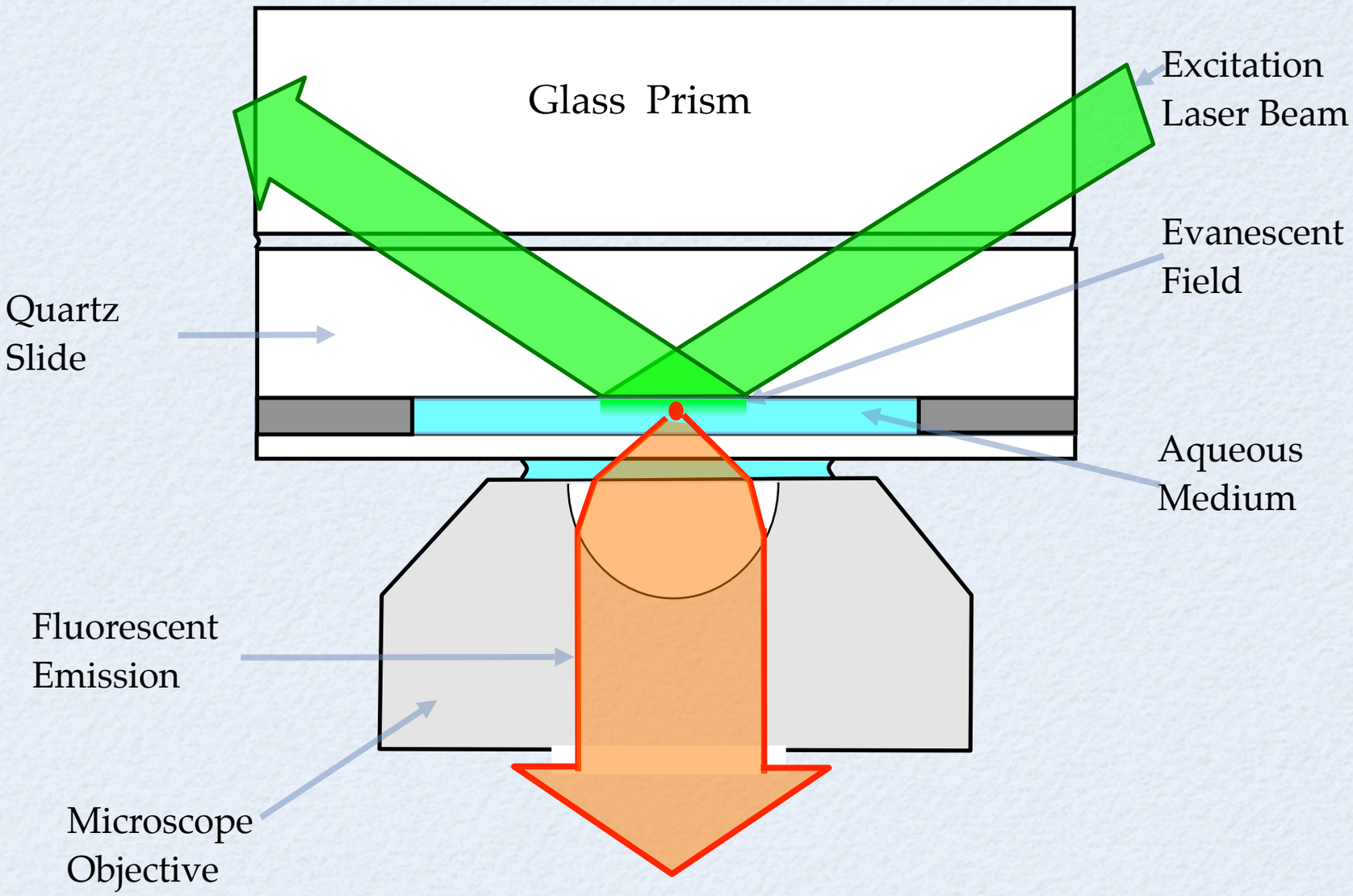
Myosin V Processivity



Myosin V Processivity

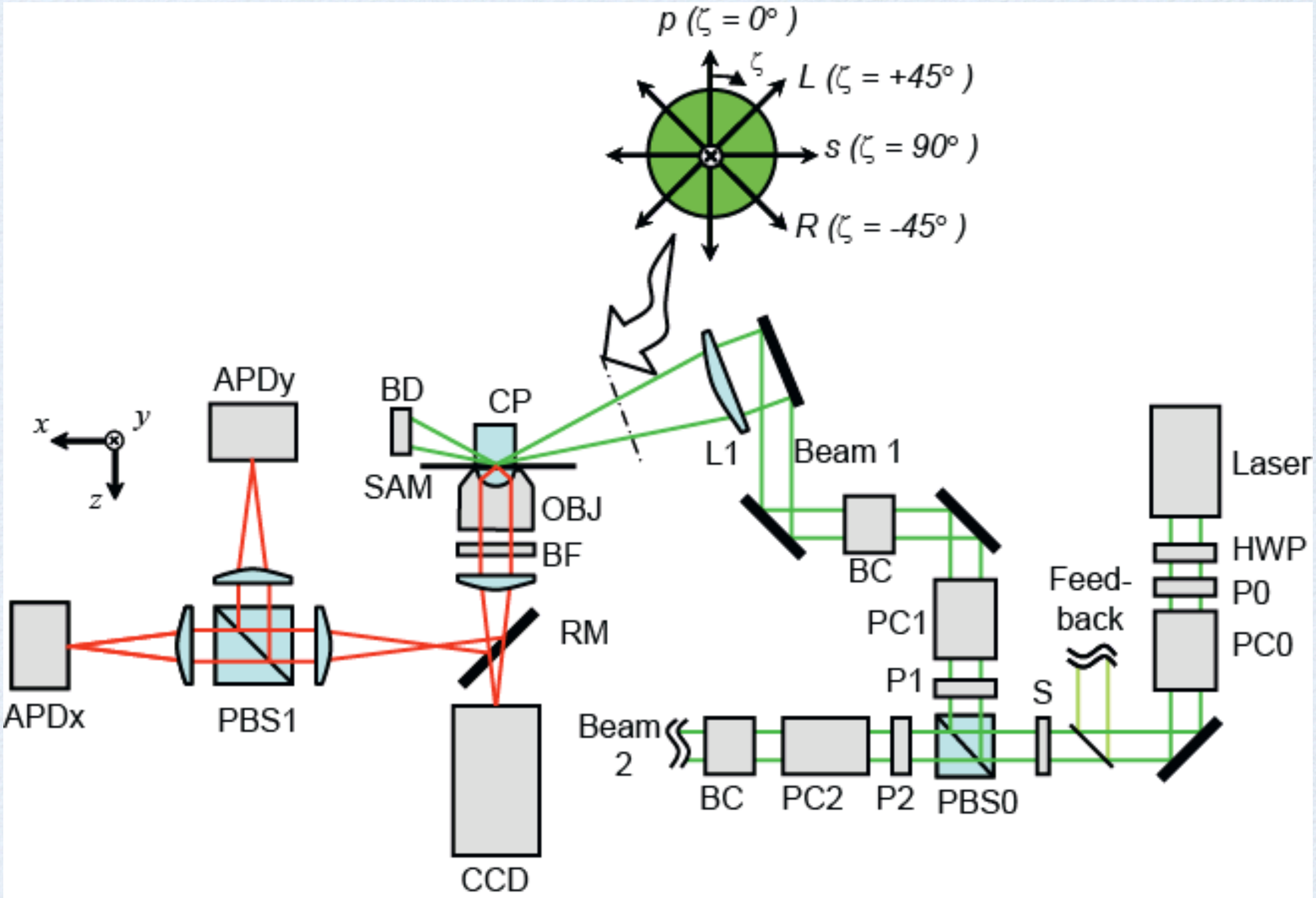


Polarized total internal reflection fluorescence microscopy

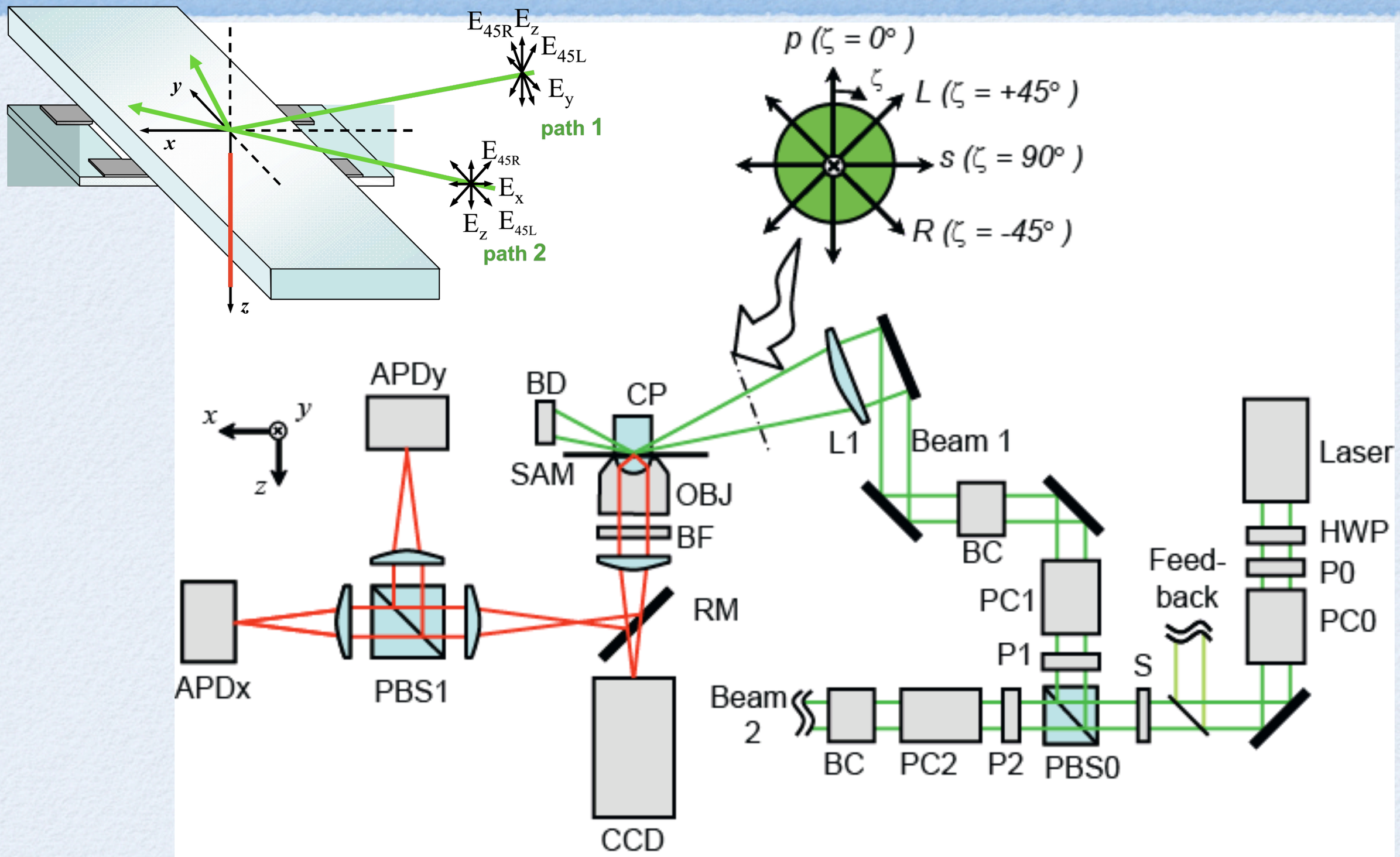


Fluorescence illumination by the evanescent wave eliminates a lot of noise, and importantly, maintains the polarization of the incident light.
To tickle the fluorophore with every possible polarization, we need the incoming light to have at least two different beam directions.

pol-TIRF setup

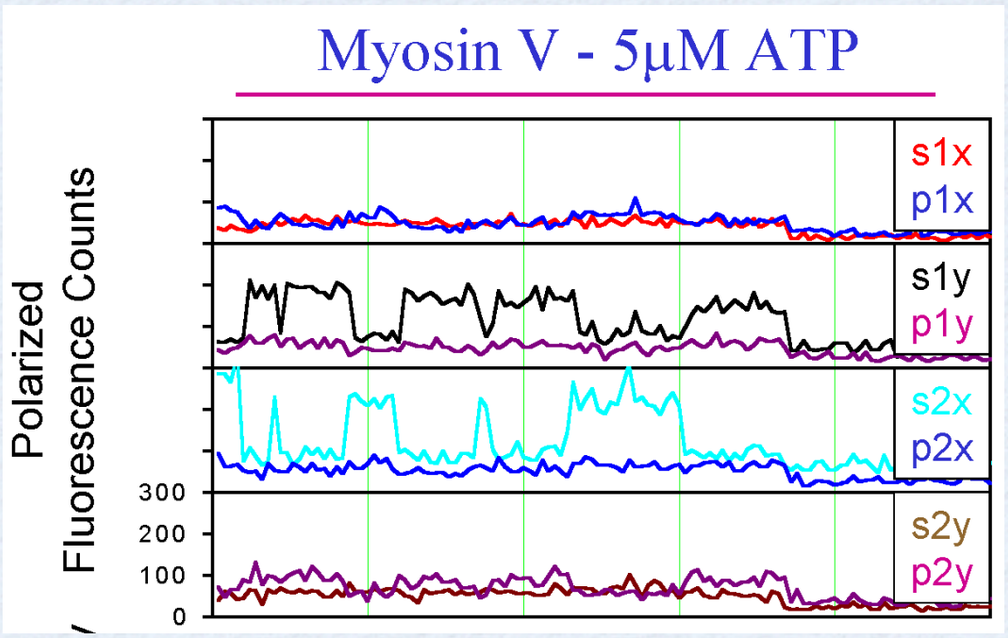


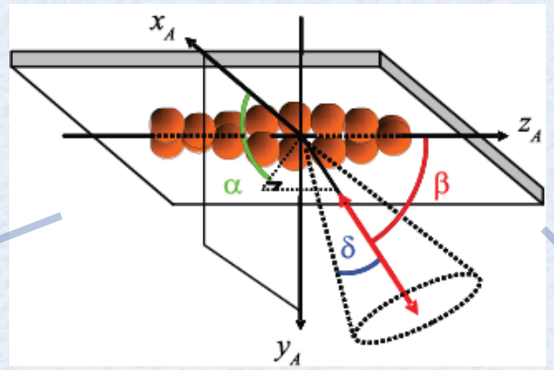
pol-TIRF setup

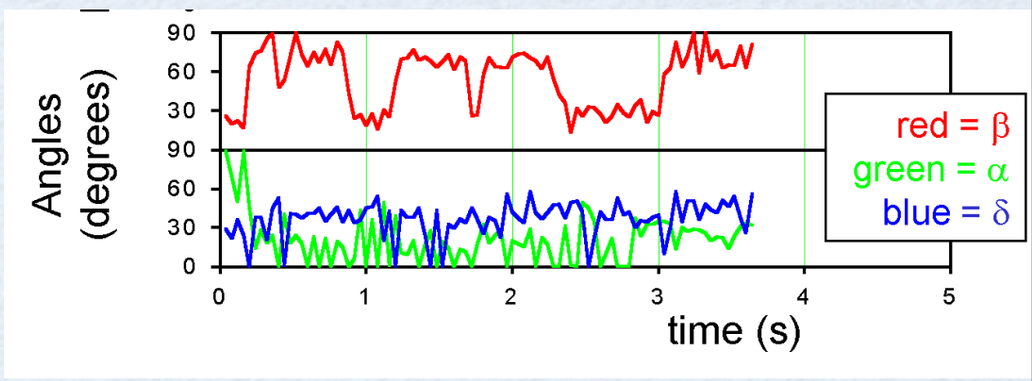


8 polarized illuminations x 2 detectors = 16 fluorescent intensities per cycle

Current state of the art







It's a bit more meaningful to convert from lab-frame angles θ,ϕ to actin-frame angles α,β. Even then, however, state of the art calculations give pretty noisy determinations, with pretty poor time resolution.

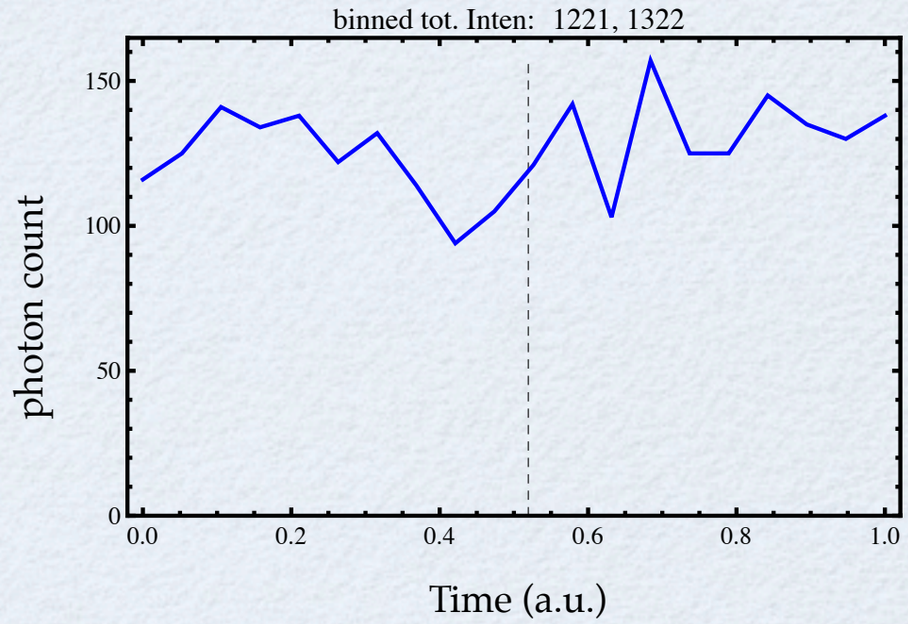
You could easily miss a short-lived state -- e.g. the elusive diffusive-search step (if it exists).

Can't we do better?

JN Forkey et al. Nature 2003

Unfortunately, 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 Poiss on counts, of course.
([ATP]=10uM)

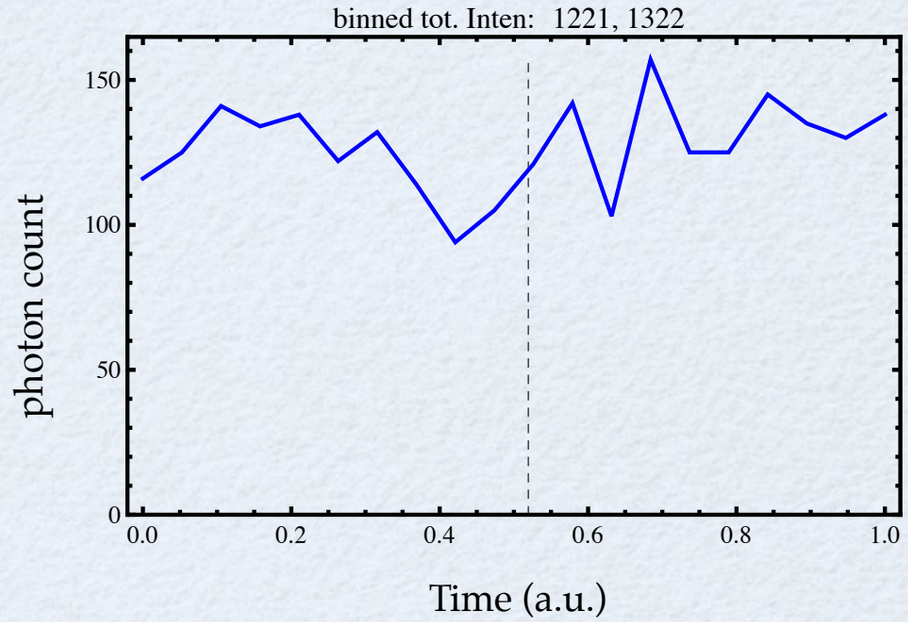
Horizontal axis is time. Vertical axis is binned photon count, PFI =polarized fluorescence intensity

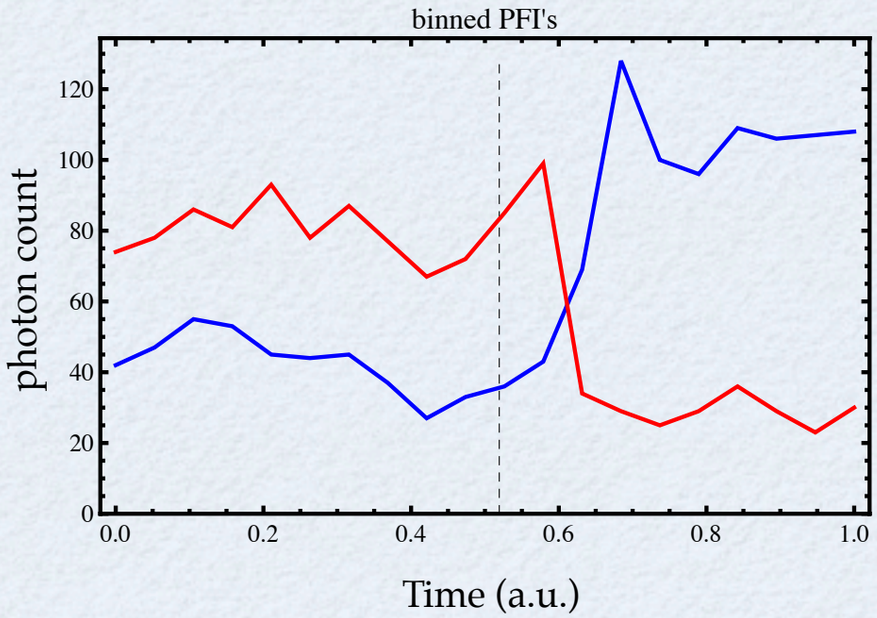


JF Beausang, YE Goldman, and PCN, Meth. Enzymol. 487:431 (2011).

Unfortunately, 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 Poiss…on counts, of course.
([ATP]=10uM)

Horizontal axis is time. Vertical axis is binned photon count, PFI =polarized fluorescence intensity



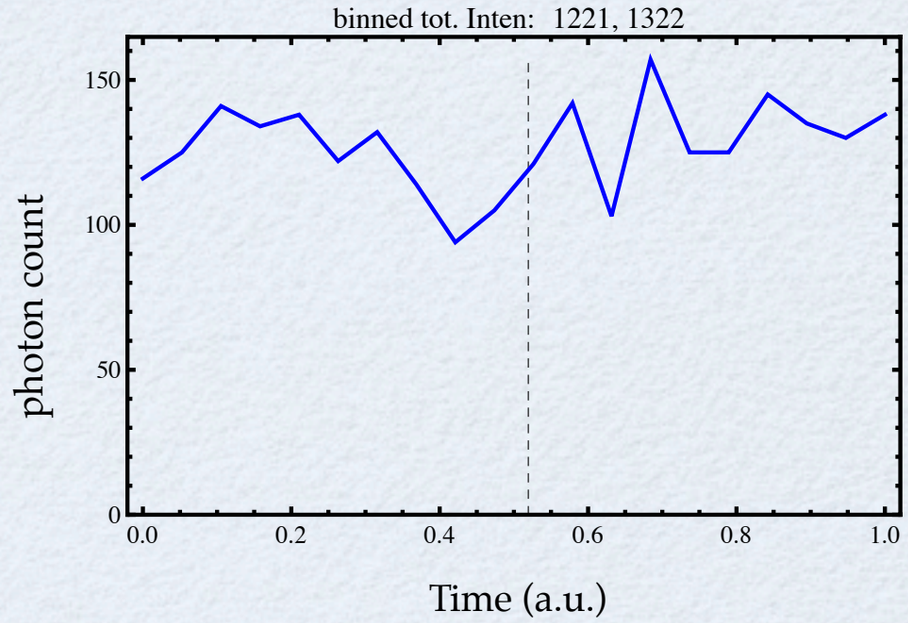


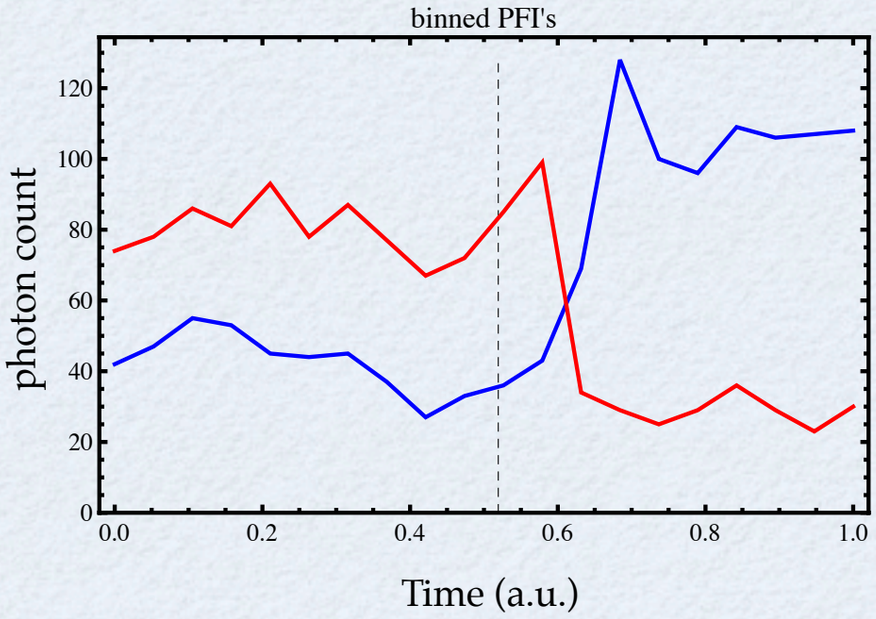
If we classify the photons by polarization and bin them separately, that reveals a definite changepoint. But when exactly did it occur? Probably not at the dashed line shown, but how can we be more precise?

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

Unfortunately, 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 Poiss…on counts, of course.
([ATP]=10uM)

Horizontal axis is time. Vertical axis is binned photon count, PFI =polarized fluorescence intensity





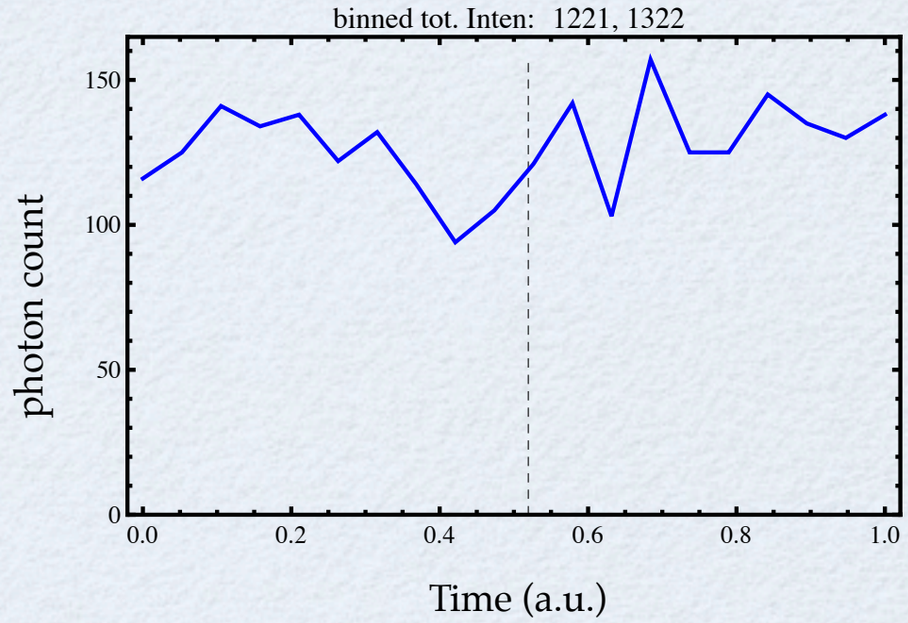
If we classify the photons by polarization and bin them separately, that reveals a definite changepoint. But when exactly did it occur? Probably not at the dashed line shown, but how can we be more precise?

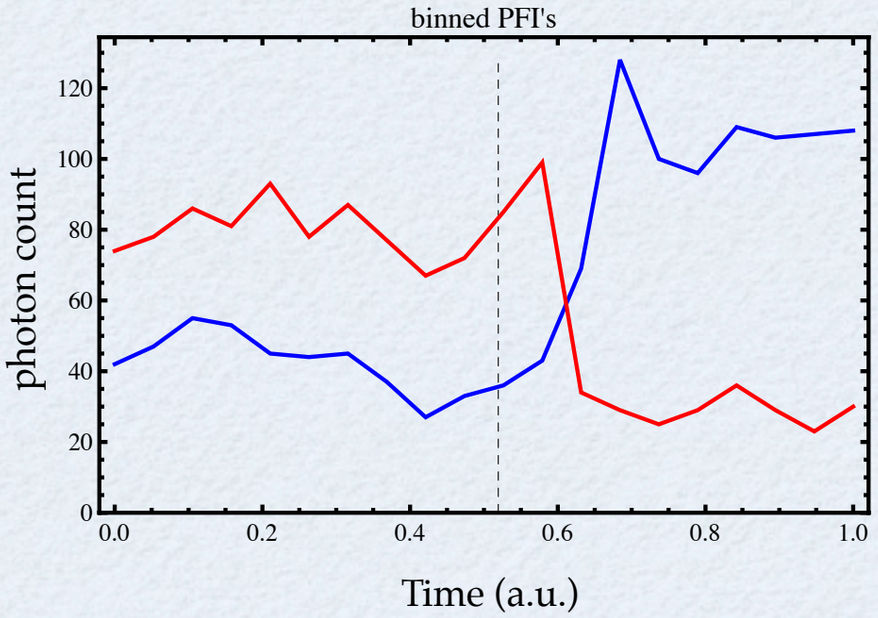
If we choose wider bins, we'll get worse time resolution; if we choose narrower bins, we'll get worse shot-noise errors.

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

Unfortunately, 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 Poiss on counts, of course.
([ATP]=10uM)

Horizontal axis is time. Vertical axis is binned photon count, PFI =polarized fluorescence intensity





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?

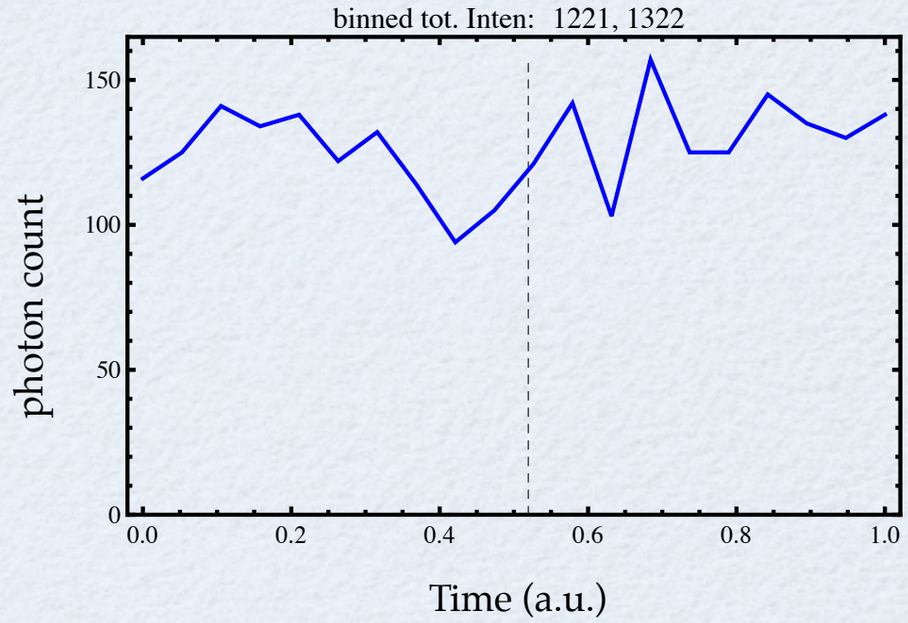
If we choose wider bins, we'll get worse time resolution; if we choose narrower bins, we'll get worse shot-noise errors.

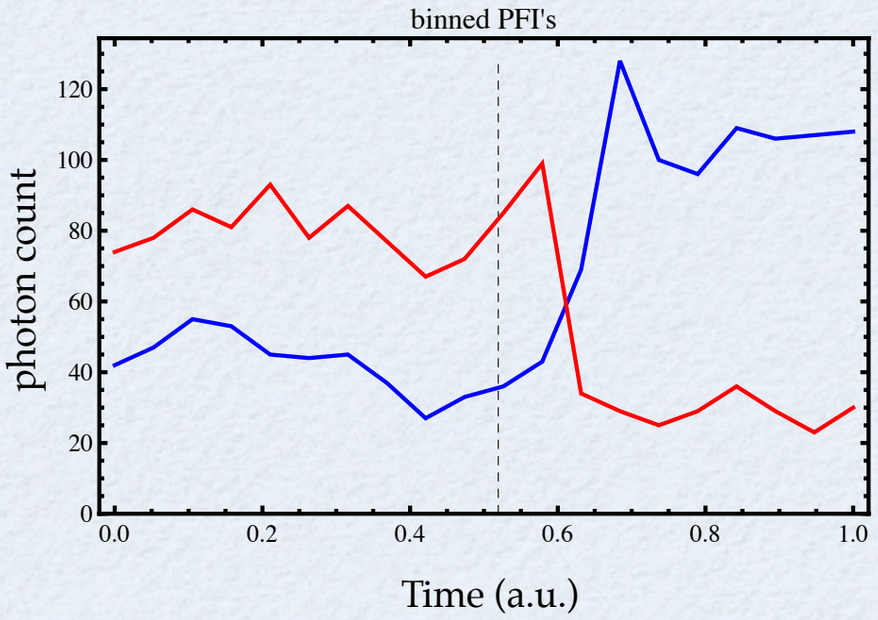
Can we evade the cruel logic of photon statistics?

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

Unfortunately, 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 Poiss…on counts, of course.
([ATP]=10uM)

Horizontal axis is time. Vertical axis is binned photon count, PFI =polarized fluorescence intensity



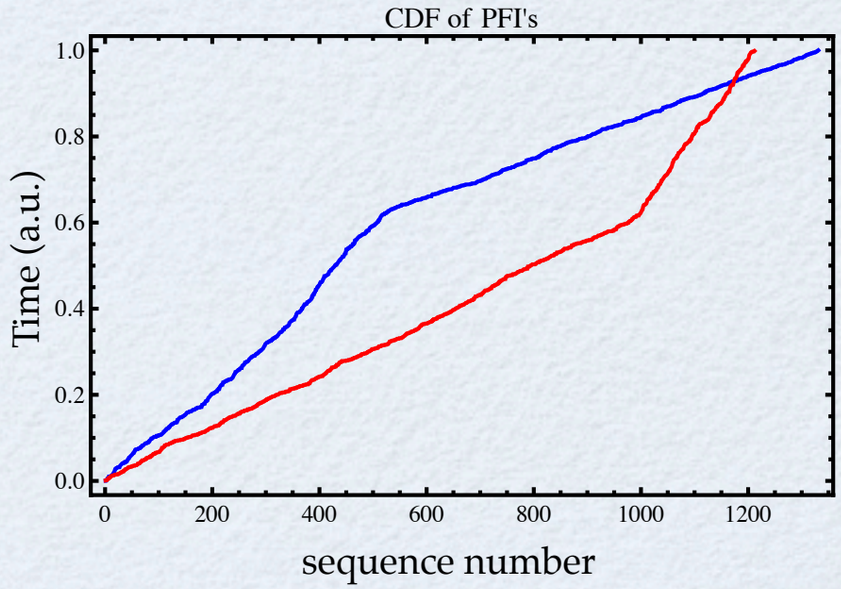


If we classify the photons by polarization and bin them separately, that reveals a definite changepoint. But when exactly did it occur? Probably not at the dashed line shown, but how can we be more precise?

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?

It turns out that ***binning the data destroyed some information.*** Something magical happens if instead of binning, we just we 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.

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.

The appropriate tool is maximum-likelihood analysis:

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.

The appropriate tool is maximum-likelihood analysis:
Focus on just one "flavor" of photons (e.g. one polarization).

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.

The appropriate tool is maximum-likelihood analysis:
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.

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.

The appropriate tool is maximum-likelihood analysis:
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'.

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.

The appropriate tool is maximum-likelihood analysis:
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', find confidence intervals for them, and compare the null hypothesis that there was no changepoint.

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.

The appropriate tool is maximum-likelihood analysis:
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', find confidence intervals for them, and compare the null hypothesis that there was no changepoint.

To do this, we ask for 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:

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.

The appropriate tool is maximum-likelihood analysis:
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', find confidence intervals for them, and compare the null hypothesis that there was no changepoint.

To do this, we ask for 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:

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

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.

The appropriate tool is maximum-likelihood analysis:
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', find confidence intervals for them, and compare the null hypothesis that there was no changepoint.

To do this, we ask for 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 $\Delta t \to 0$:

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 $\Delta t \to 0$:

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

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

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':

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

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.

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 $\Delta t \to 0$:

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

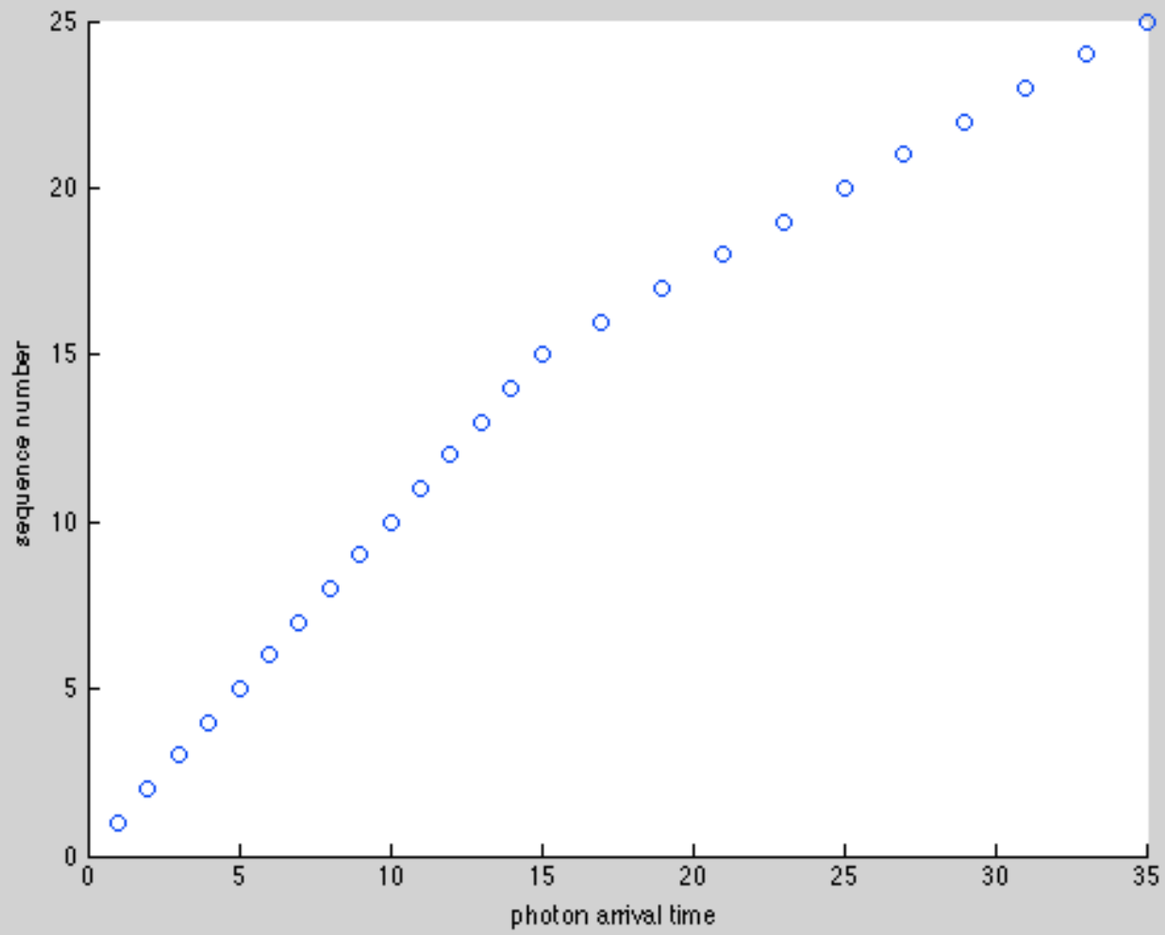
Maximize this first over R and R':

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

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:

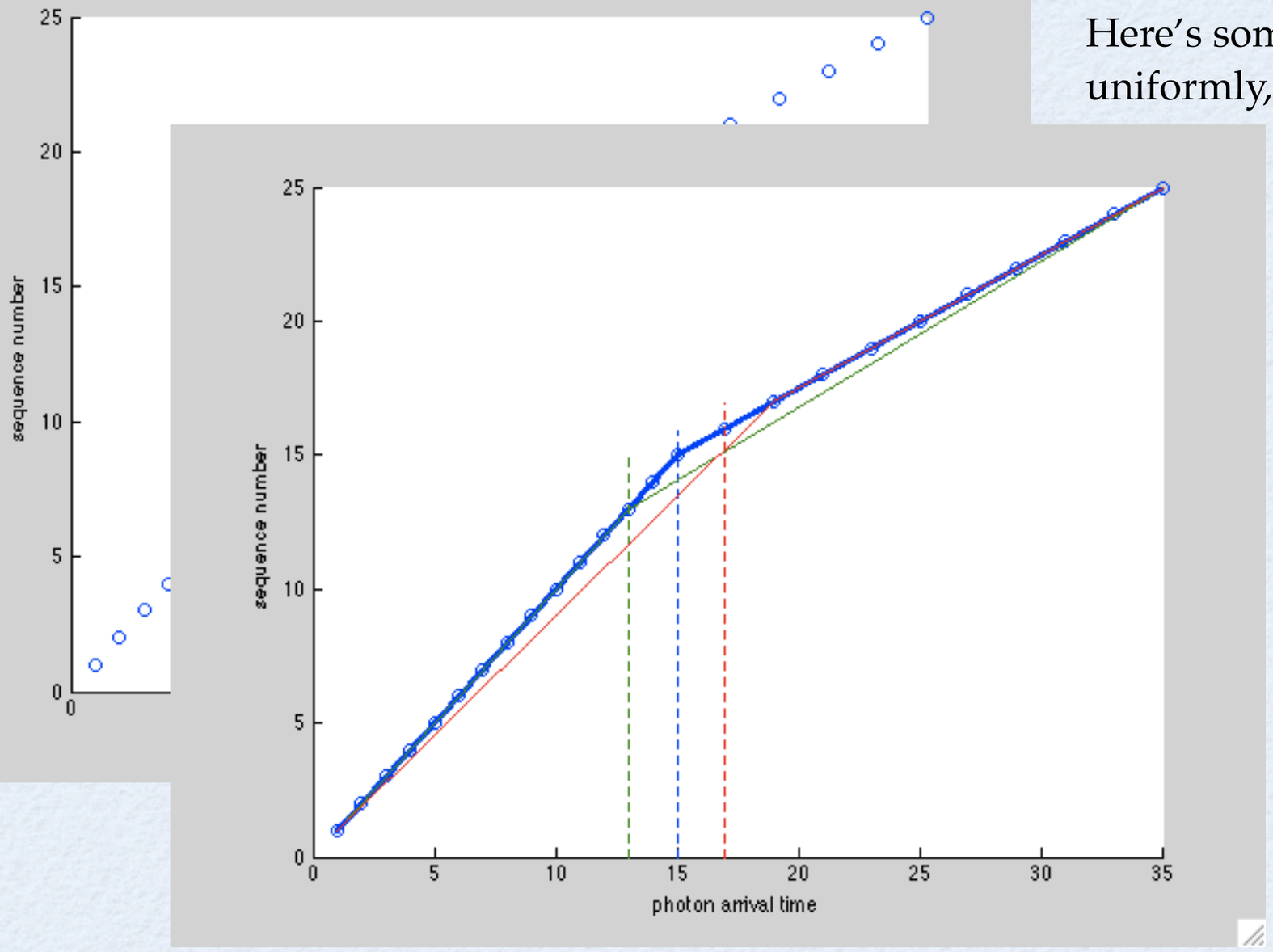
Application

Application



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

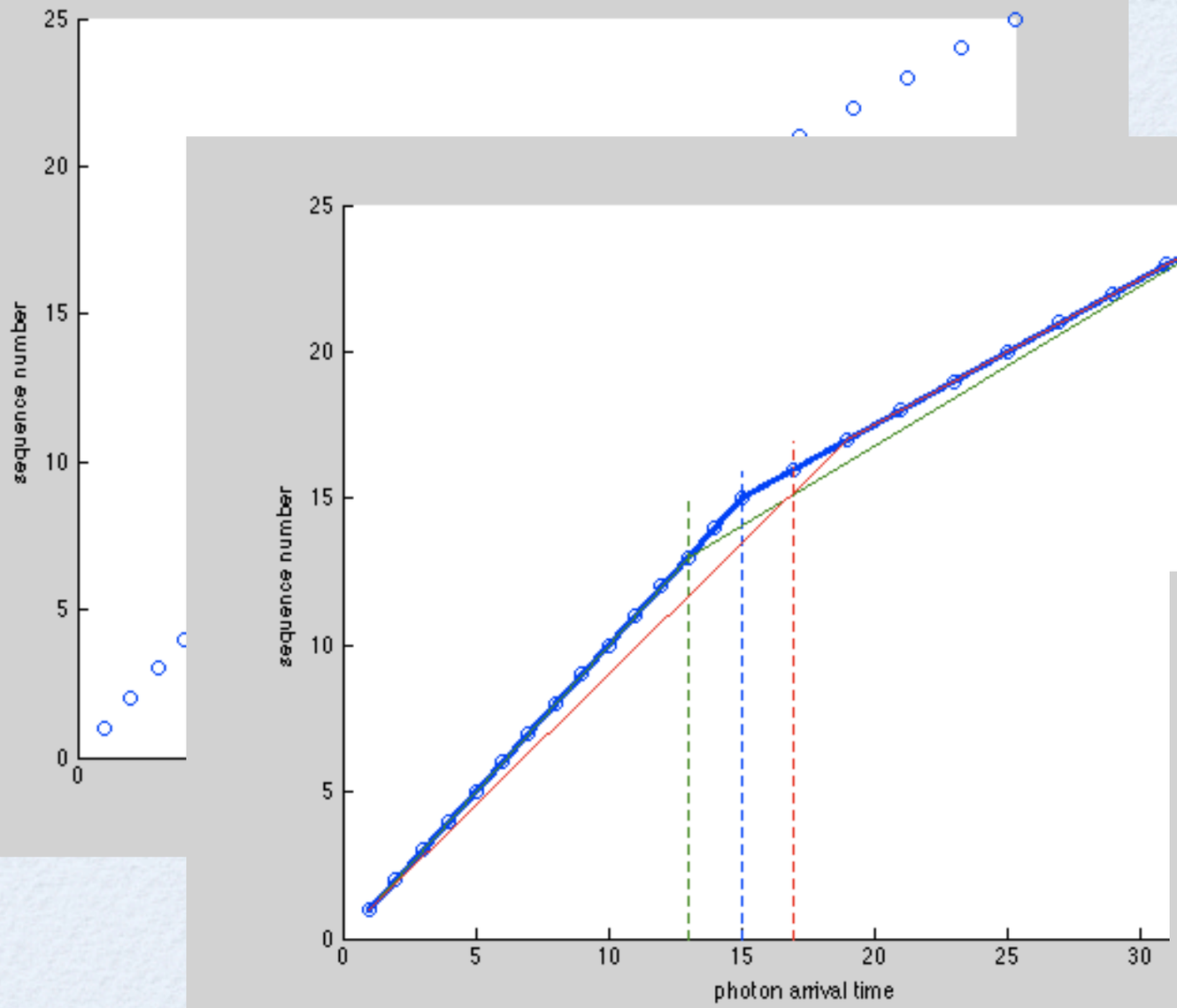
Application



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

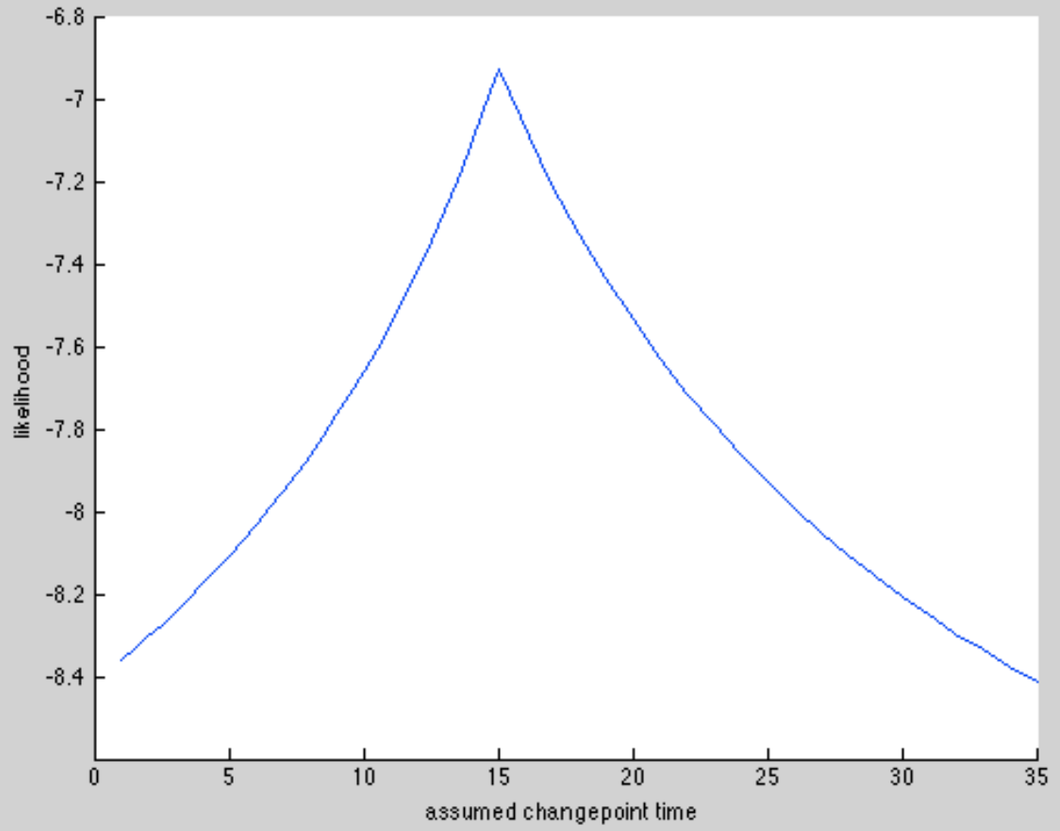
Here are two lines corresponding to non-optimal choices of the changepoint. We'd like to see the likelihood function and how it selects the "right" changepoint, which for fake data is known.

Application

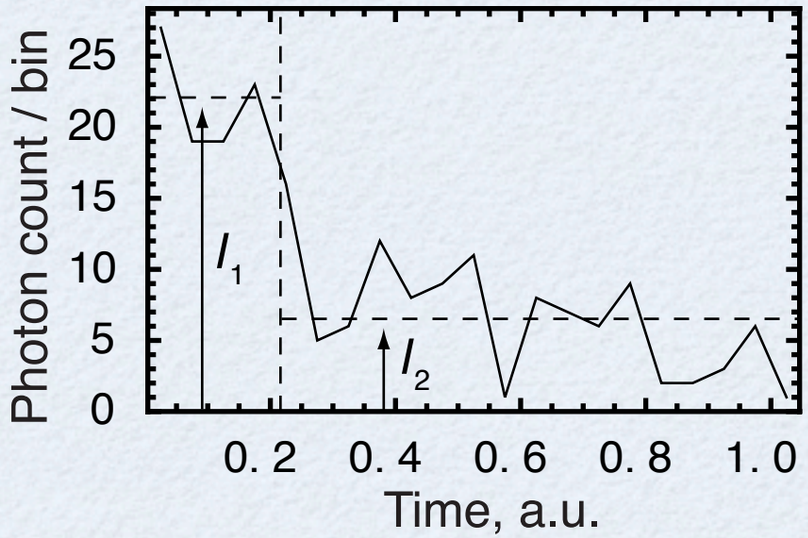


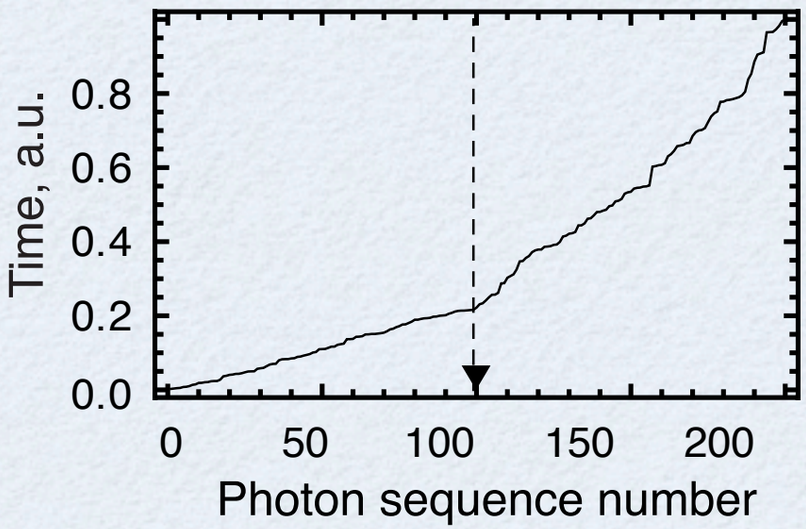
Here's some very fake data; the photons arrive uniformly, not at random.

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.



Here is our log-likelihood function as a function of putative changepoint time.

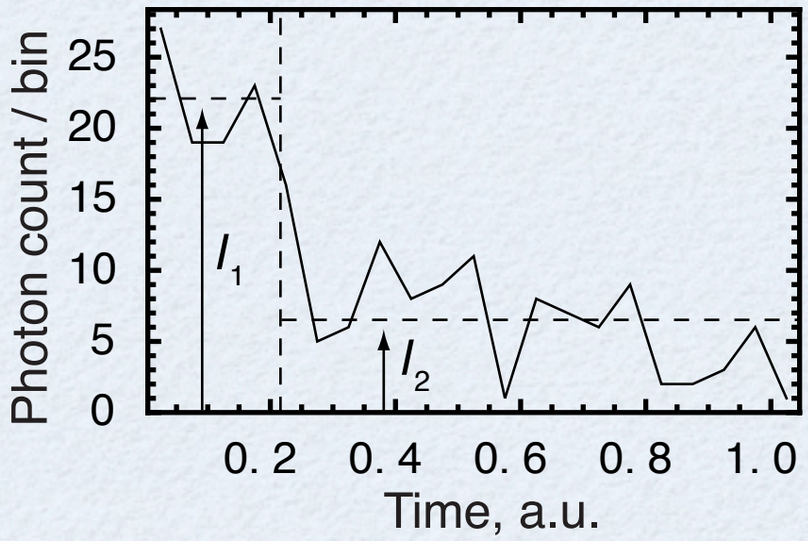


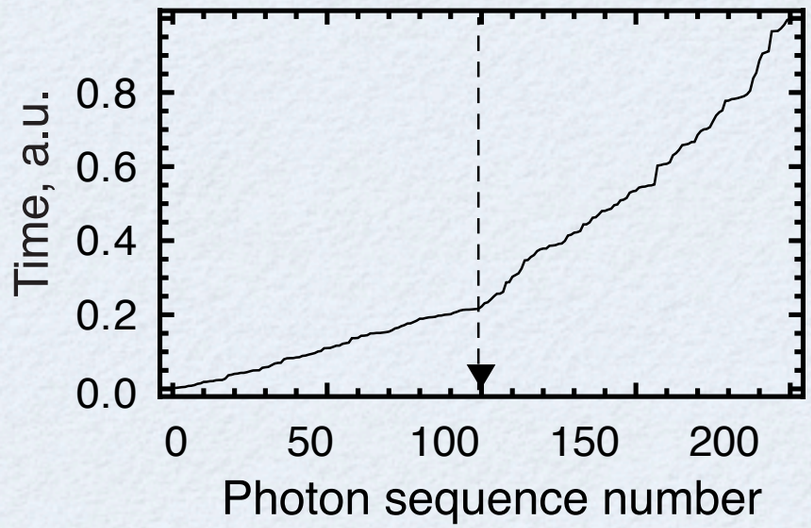


Left: Some more realistic (Poisson-arrival) simulated data, shown in traditional binned form and in the improved version.

Right: Likelihood function for placement of the changepoint. Dashed line, maximum-likelihood point. Black triangle: Actual changepoint used in the simulation.

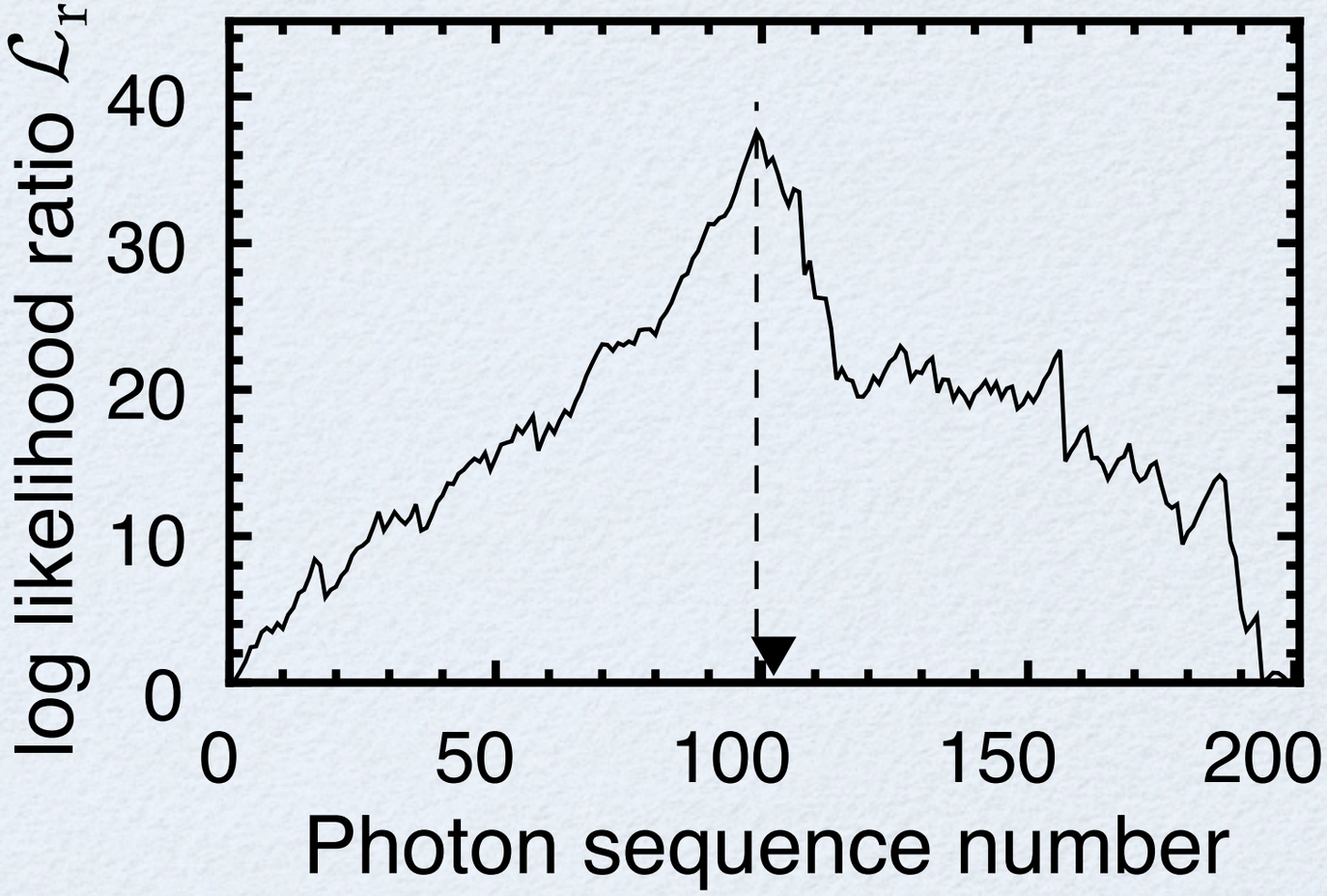
JF Beausang, YE Goldman, and PCN, Meth. Enzymol. 487:431 (2011).





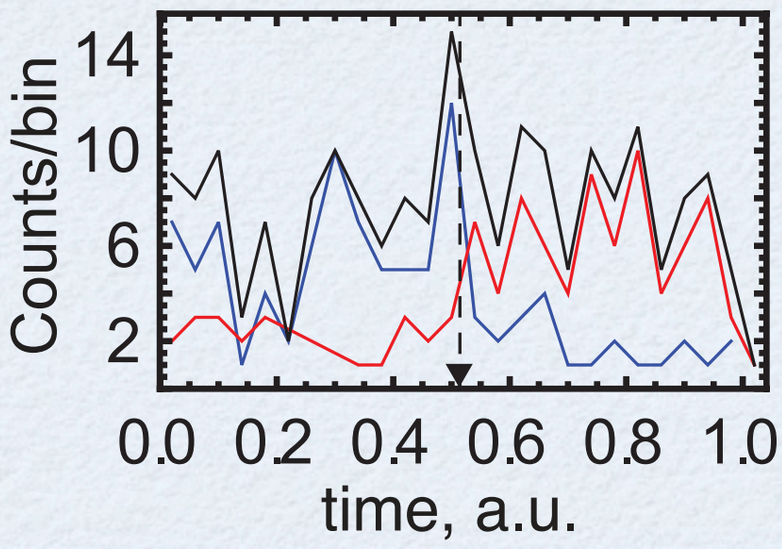
Left: Some more realistic (Poisson-arrival) simulated data, shown in traditional binned form and in the improved version.

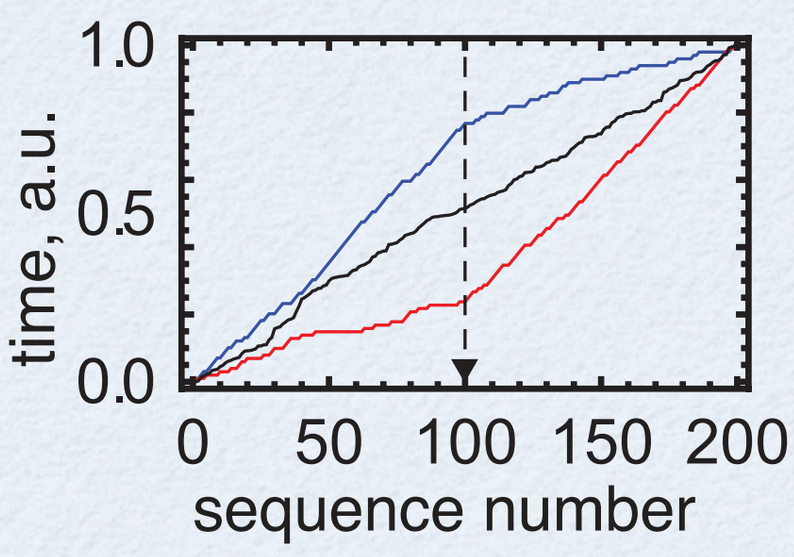
Right: Likelihood function for placement of the changepoint. Dashed line, maximum-likelihood point. Black triangle: Actual changepoint used in the simulation.

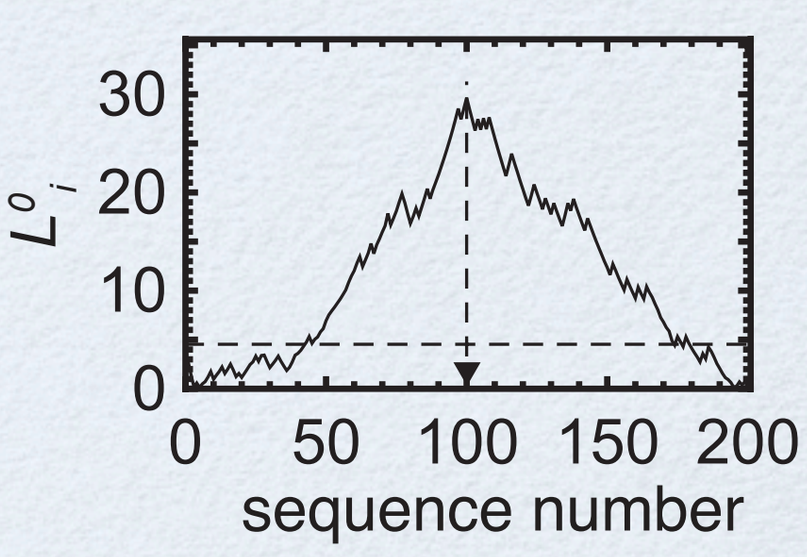


JF Beausang, YE Goldman, and PCN, Meth. Enzymol. 487:431 (2011).

Oh, yes -- the method also works on multiple-channel data. *Left:* one channel (red) starts with rare photons, then jumps to higher intensity. Another channel (blue) does the opposite. The sum of the intensities (black) doesn't change much at all.
Middle: "kink" representations of the same data. *Right:* both channels contribute to a likelihood function with a robust peak, even though there were only a total of just 200 photons in the entire dataset.

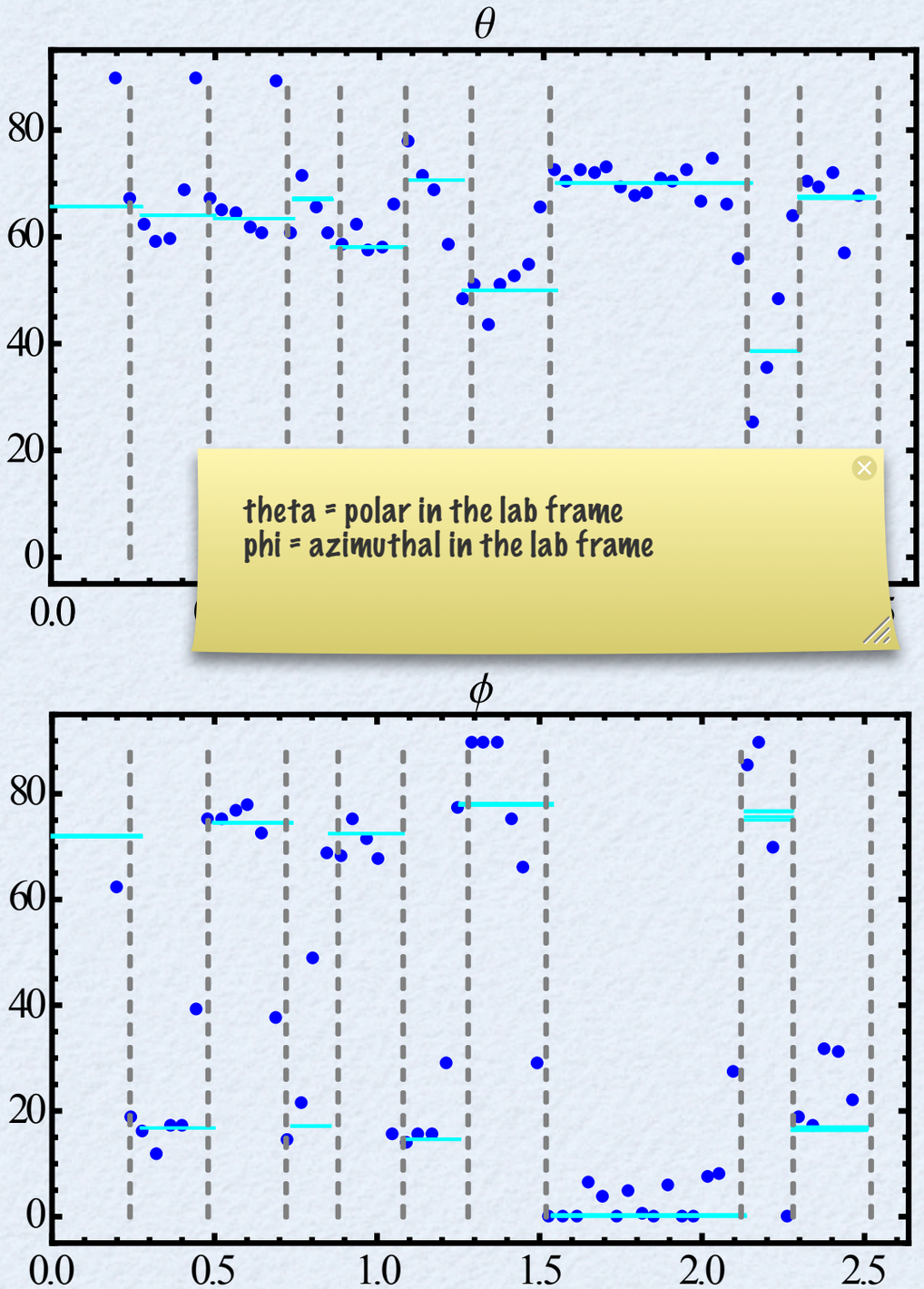


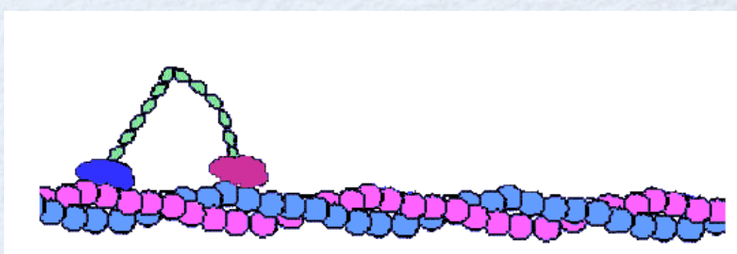




JF Beausang, YE Goldman, and PCN, Meth. Enzymol. 487:431 (2011).

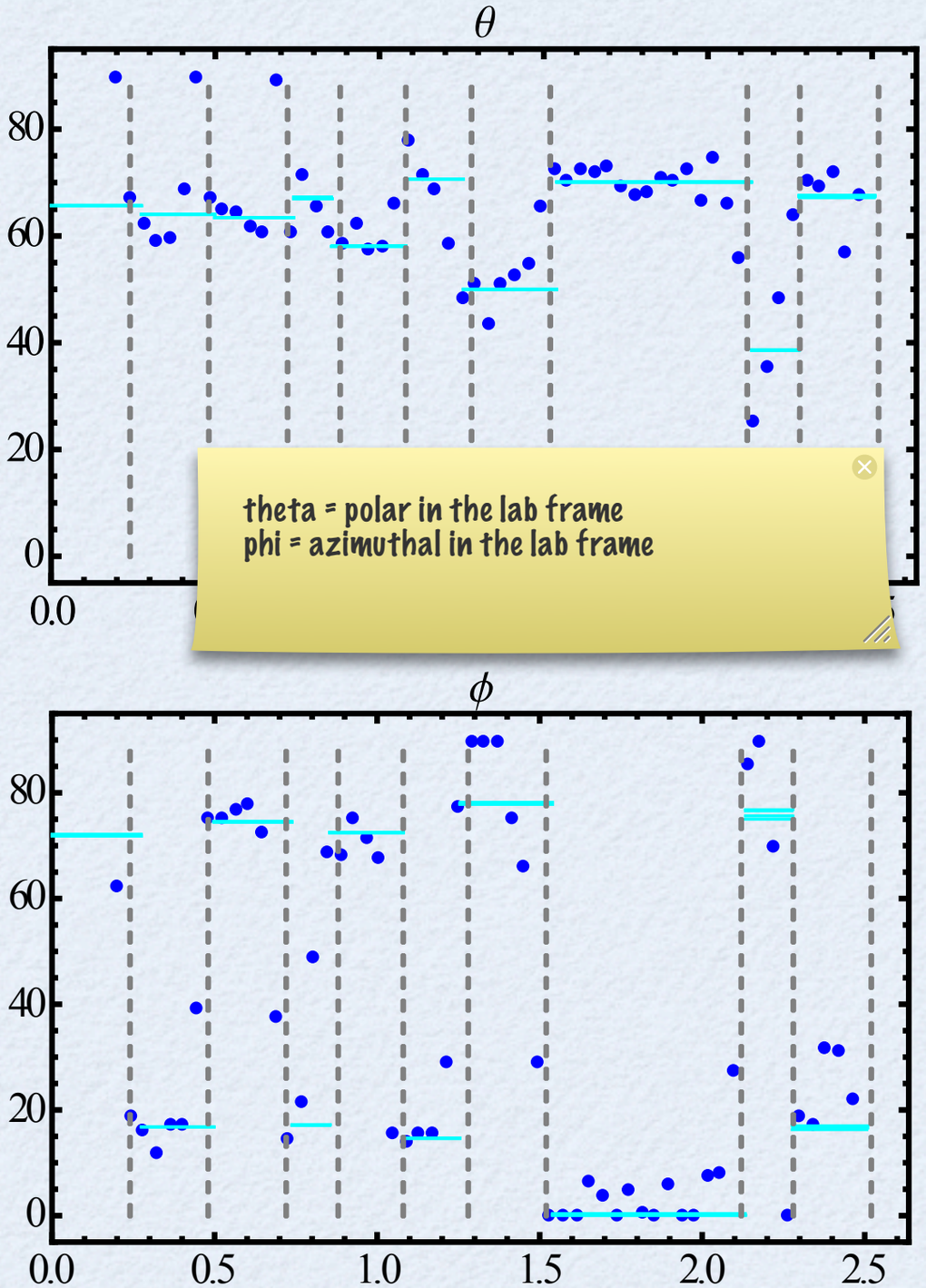
Payoff





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

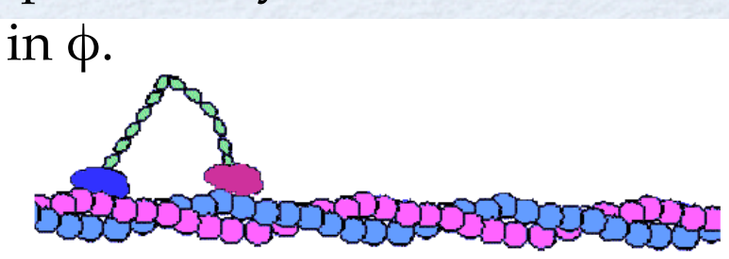
Payoff



Oh, yes--it also works on real experimental data.

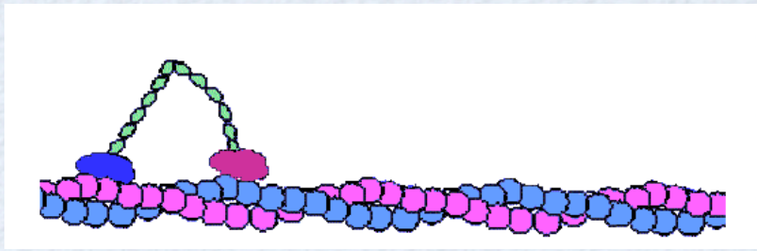
Now we can get back to the original motivation. Previously, people would take data from multiple polarizations, bin it, and pipe the inferred intensities into a maximum-likelihood estimator of the orientation of the fluorophore. That procedure leads to the rather noisy dots shown here.
One problem is 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.



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

Summary Part II



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

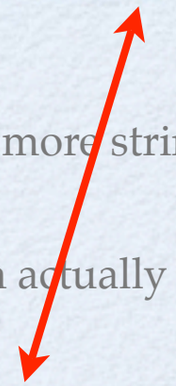
✴A simple maximum-likelihood analysis accomplishes this.

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

Part III: Parallel recordings from dozens of individual neurons

- 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.**

Part III: Parallel recordings from dozens of individual neurons



- 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.**

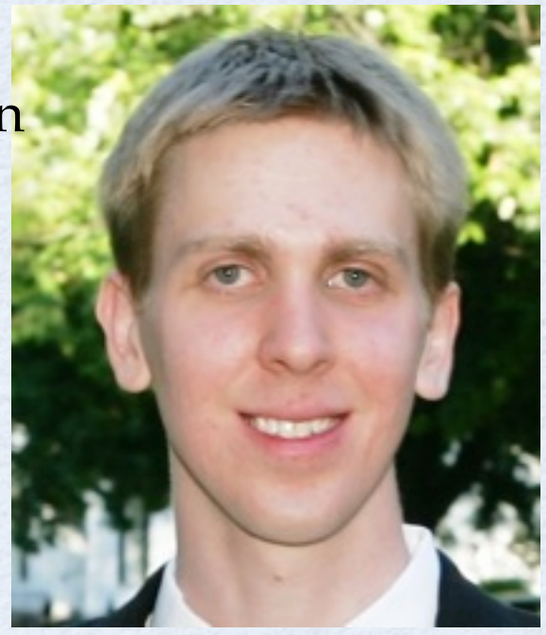
Sources of energy

Experiments done in the lab of Vijay Balasubramanian (Penn).

Sources of energy

Experiments done in the lab of Vijay Balasubramanian (Penn).

Jason Prentice, Penn Physics



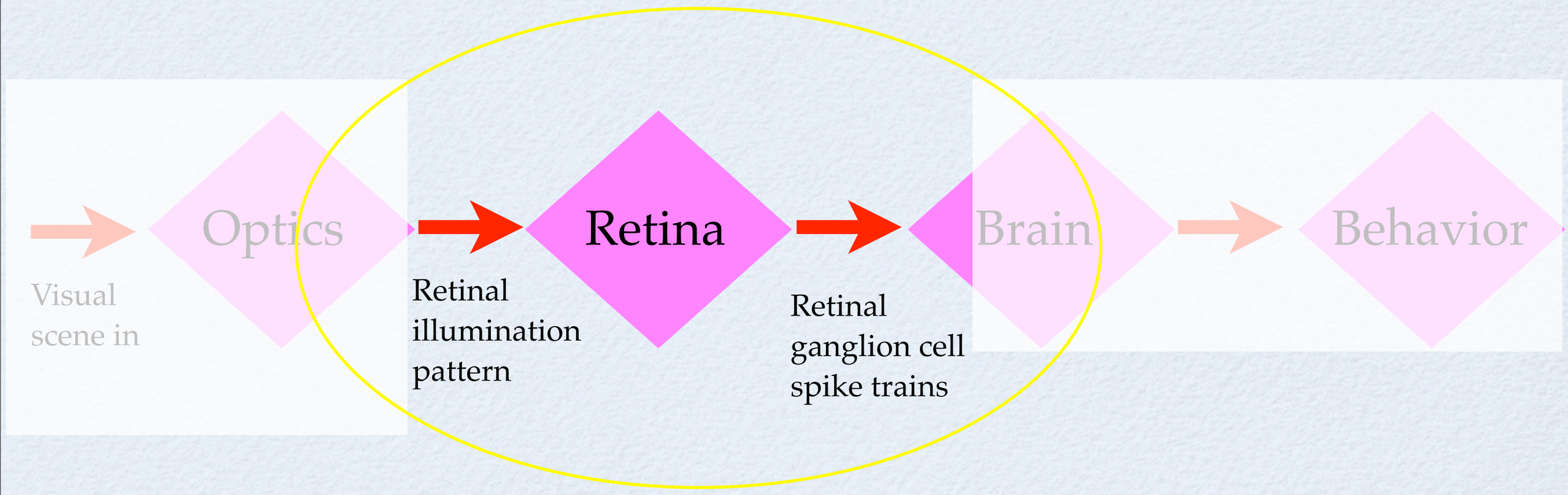
Kristy Simmons, Penn Neuroscience

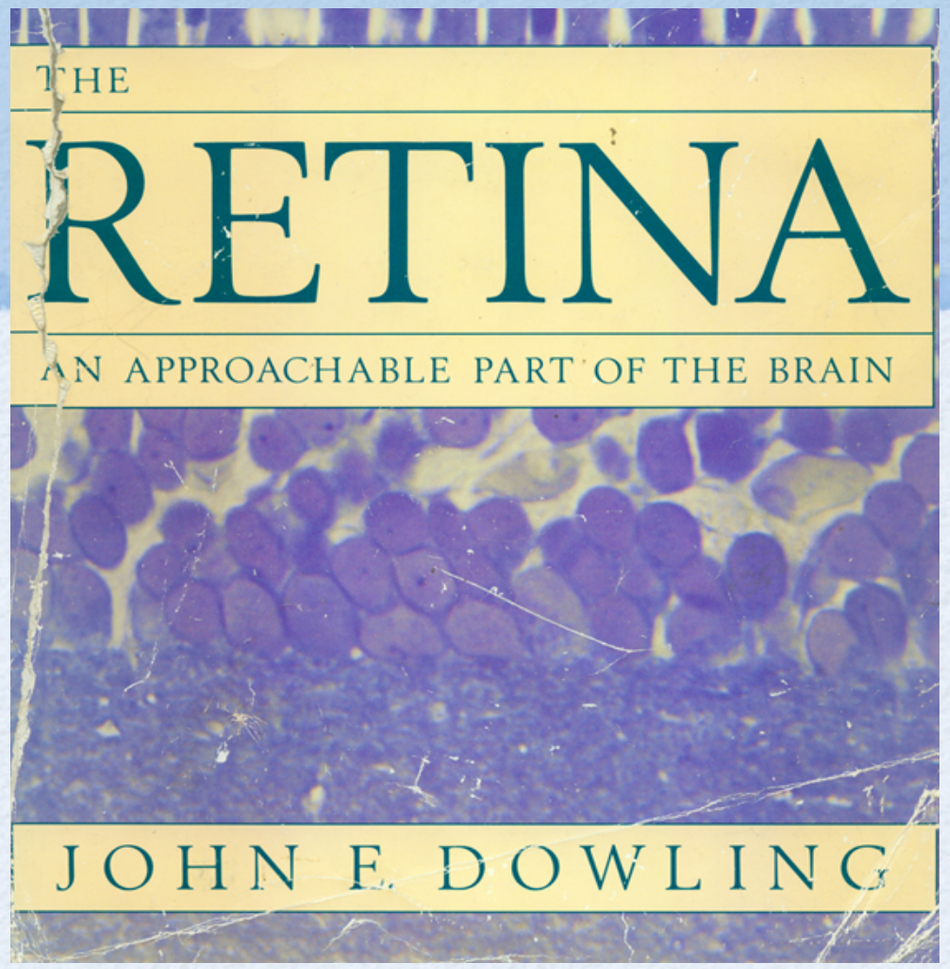
Jan Homann, Penn Physics

(plus Gasper Tkacik.)
(Many thanks to Michael Berry and Olivier Marre, Princeton; Bart Borghuis, Janelia Farms; Michael Freed and others at Penn Retina Lab; Joerg Sander, U Alberta; Ronen Segev, BGU, Chris Wiggins, Columbia.)

Really big picture

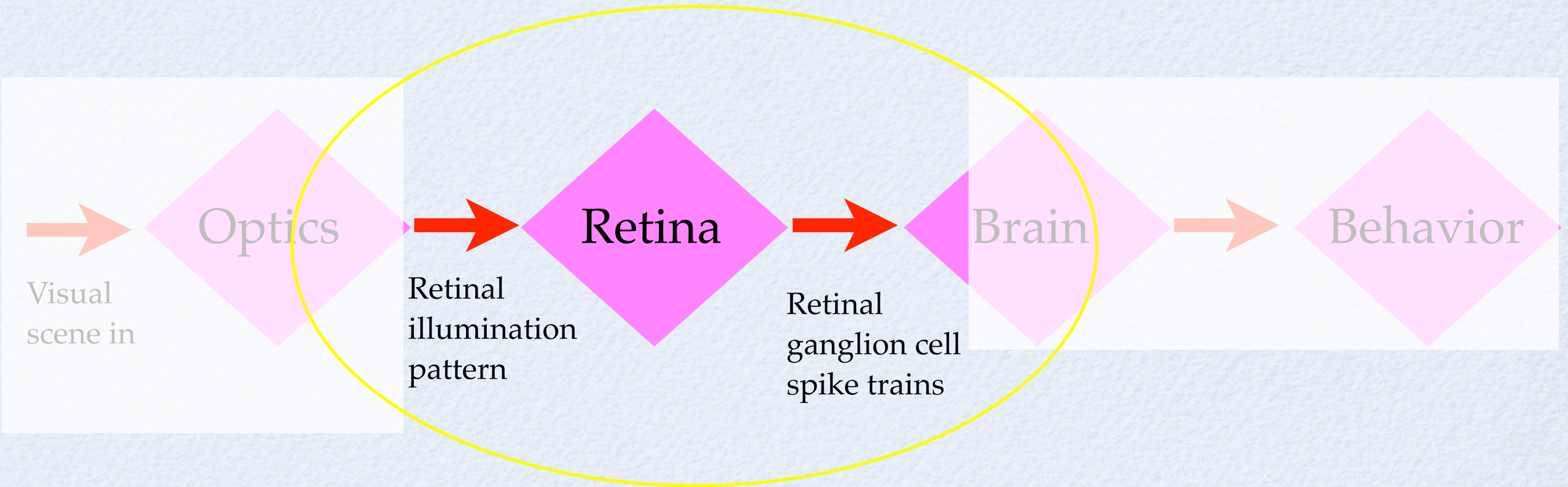
Retina is also an approachable, yet still complex, part of the brain. It's a 2D carpet consisting of "only" three layers of neurons.





Really big picture

Retina is also an approachable, yet still complex, part of the brain. It's a 2D carpet consisting of "only" three layers of neurons.



It matters

afari File Edit View History Bookmarks Window Help (100%) Thu 4:21

Artificial retina more capable of restoring normal vision; animal study shows including retina's neural 'code' improved prosthetic

http://www.sciencedaily.com/releases/2010/11/101116102647.htm

shiela nirenberg

511PA – Travel Info To Go | nelsonbook wiki | pennspikes | Dropbox | Secure Access SSL VPN | Umbrella Today? | Bb 280 | Bb 516 | Penn InTouch Advisor | Penn Library-Franklin | VerizonNetM

Artificial retina more capable ... | Now I See You – Technology ... | Retinal Implant Restores Visi... | Neuroscience meeting: Resear... | Blind mice can see where the... | Nirenberg

ScienceDaily®

Your source for the latest research news

News | Articles | Videos | Images | Books | Search

Health & Medicine | Mind & Brain | Plants & Animals | Earth & Climate | Space & Time | Matter & Energy | Computers & Math | Fossils & Ruins

Science News

... from universities, journals, and other research organizations

Save Email Print Share

Artificial Retina More Capable of Restoring Normal Vision; Animal Study Shows Including Retina's Neural 'Code' Improved Prosthetic

ScienceDaily (Nov. 16, 2010) — Researchers have developed an artificial retina that has the capacity to reproduce normal vision in mice. While other prosthetic strategies mainly increase the number of electrodes in an eye to capture more information, this study concentrated on incorporating the eye's neural "code" that converts pictures into signals the brain can understand.

The research was presented at

Just In:
Antarctic Blue Whales Shows High Diversity
more breaking science news

Social Networks

*Share this story on **Facebook**, **Twitter**, & **Google**:*

Like 99 Tweet 0 +1 0

Other social bookmarking and sharing tools:

Digg StumbleUpon LinkedIn Email
Reddit Slashdot Fark ShareThis

It matters

afari File Edit View History Bookmarks Window Help (100%) Thu 4:21

Artificial retina more capable of restoring normal vision; animal study shows including retina's neural 'code' improved prosthetic

http://www.sciencedaily.com/releases/2010/11/101116102647.htm

shiela nirenberg

511PA – Travel Info To Go | nelsonbook wiki | pennspikes | Dropbox | Secure Access SSL VPN | Umbrella Today? | Bb 280 | Bb 516 | Penn InTouch Advisor | Penn Library-Franklin | VerizonNetM

Artificial retina more capable ... | Now I See You – Technology ... | Retinal Implant Restores Visi... | Neuroscience meeting: Resear... | Blind mice can see where the... | Nirenberg

ScienceDaily®

Your source for the latest research news

News | Articles | Videos | Images | Books | Search

Health & Medicine | Mind & Brain | Plants & Animals | Earth & Climate | Space & Time | Matter & Energy | Computers & Math | Fossils & Ruins

Science News ... from universities, journals, and other research organizations

Save Email Print Share

Artificial Retina More Capable of Restoring Normal Vision; Animal Study Shows Including Retina's Neural 'Code' Improved Prosthetic

ScienceDaily (Nov. 16, 2010) — Researchers have developed an artificial retina that has the capacity to reproduce normal vision in mice. While other prosthetic strategies mainly increase the number of electrodes in an eye to capture more information, this study concentrated on incorporating the eye's neural "code" that converts pictures into signals the brain can understand.

The research was presented at

Just in:
Antarctic Blue Whales Shows High Diversity
more breaking science news

Social Networks

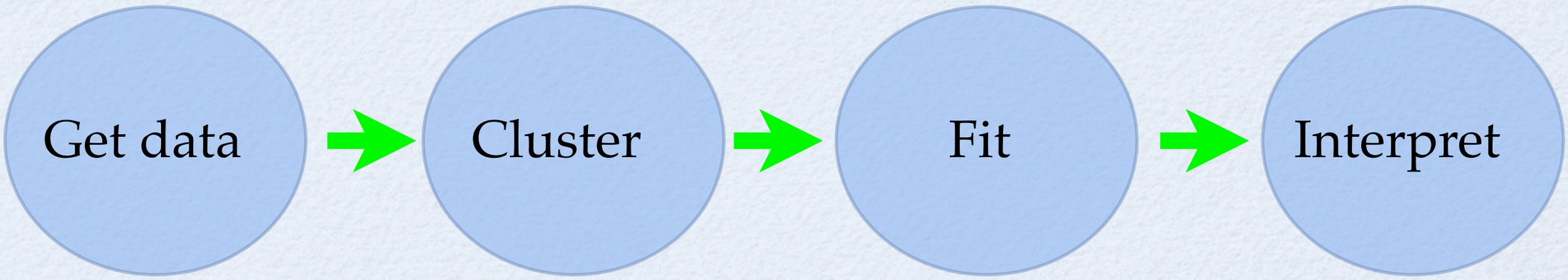
*Share this story on **Facebook**, **Twitter**, & **Google**:*

Like 99 | Tweet 0 | +1 0

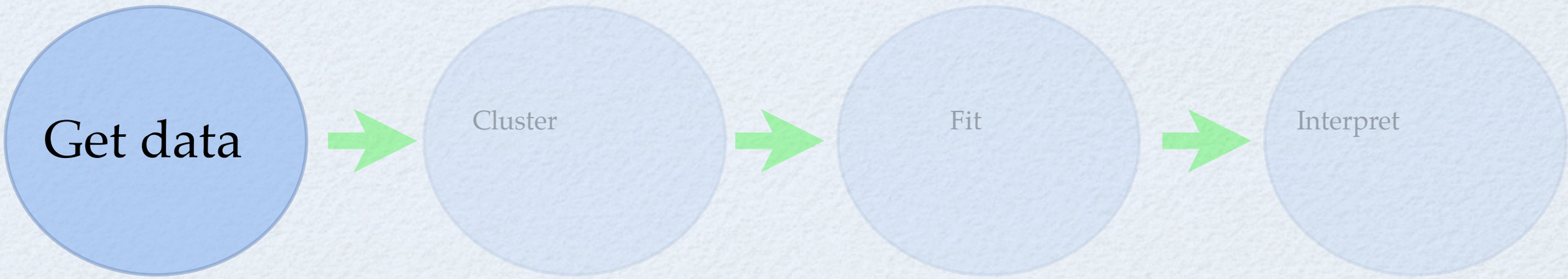
Other social bookmarking and sharing tools:

Digg | StumbleUpon | LinkedIn | Email | Reddit | Slashdot | Fark | ShareThis

Summary, Part III



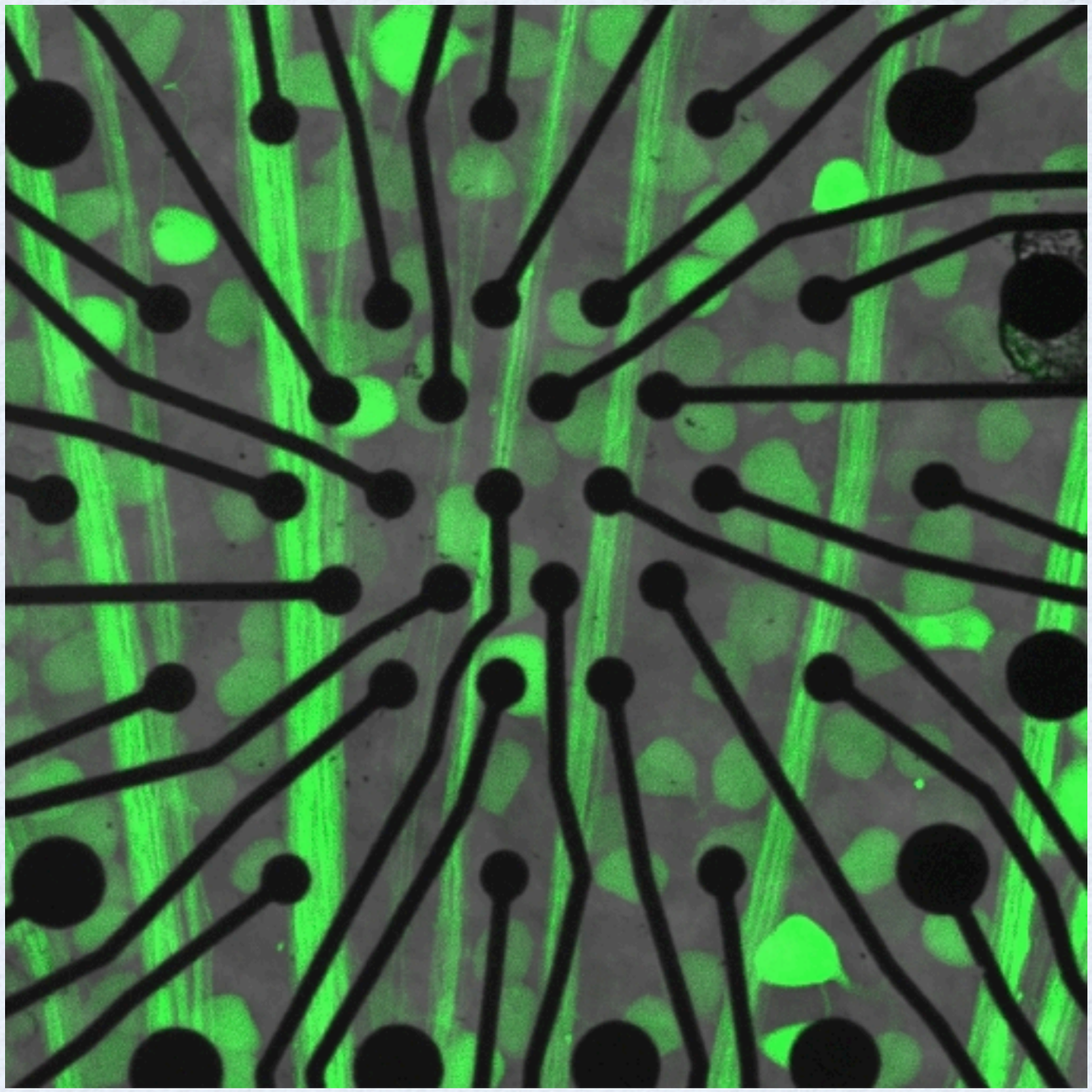
1. Experiment
2. Clustering
3. Fitting
4. Performance



Cf Meister, Pine, and Baylor 1994.
Incredibly, one can keep a
mammalian retina alive in a dish
for over 6 hours while presenting it
stimuli and recording its activity.

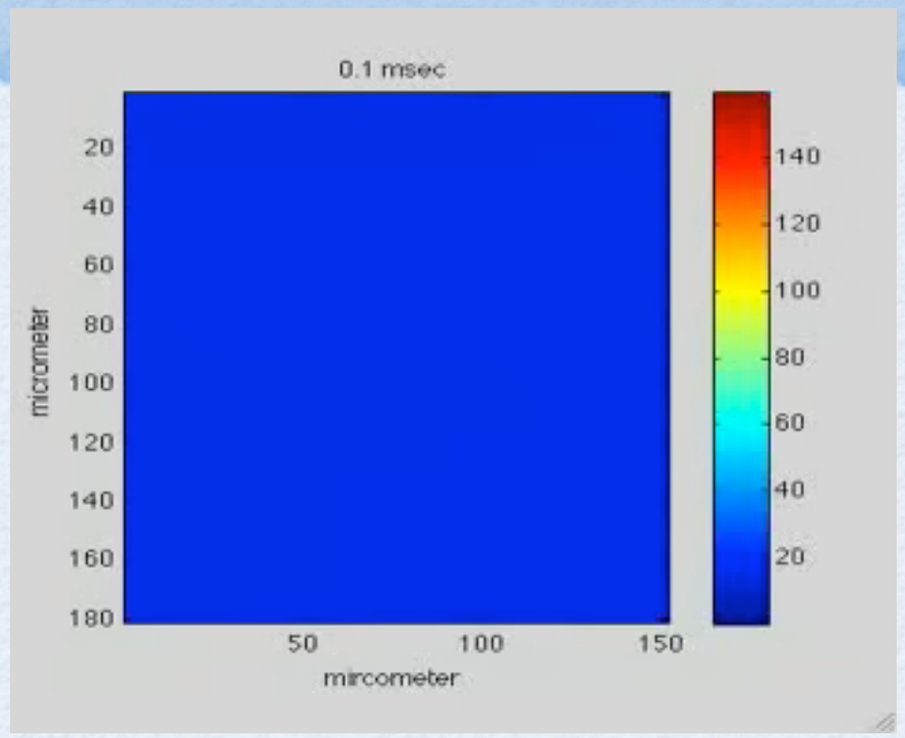
Cf Meister, Pine, and Baylor 1994.
Incredibly, one can keep a
mammalian retina alive in a dish
for over 6 hours while presenting it
stimuli and recording its activity.

What's in the dish



Michael Berry, Princeton

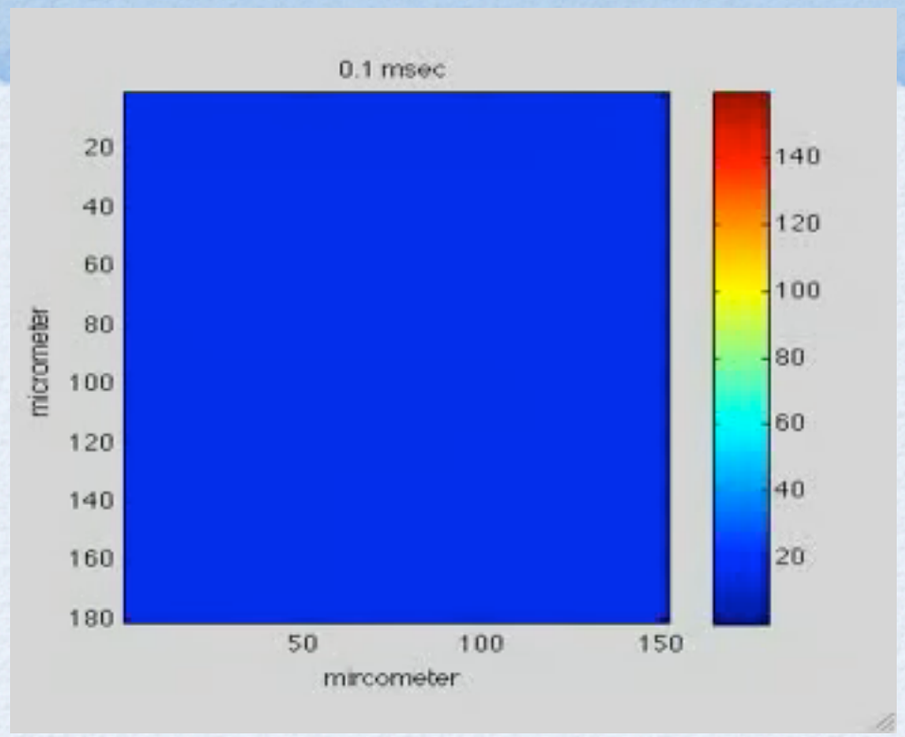
Simple events



67 ms of data,
viewed as a movie.
[data have been smoothed]

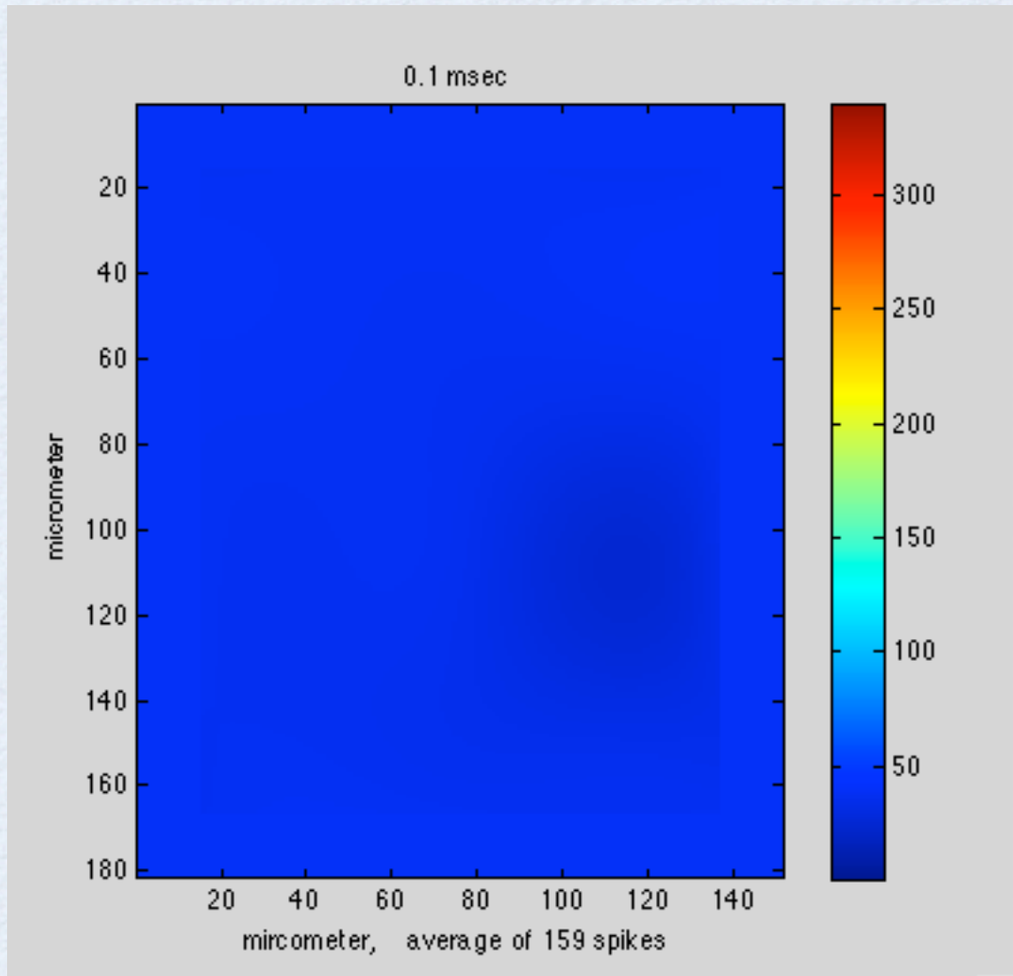
Classic: Gerstein+Clark 1964; Abeles+Goldstein 1977; Schmidt 1984.

Simple events



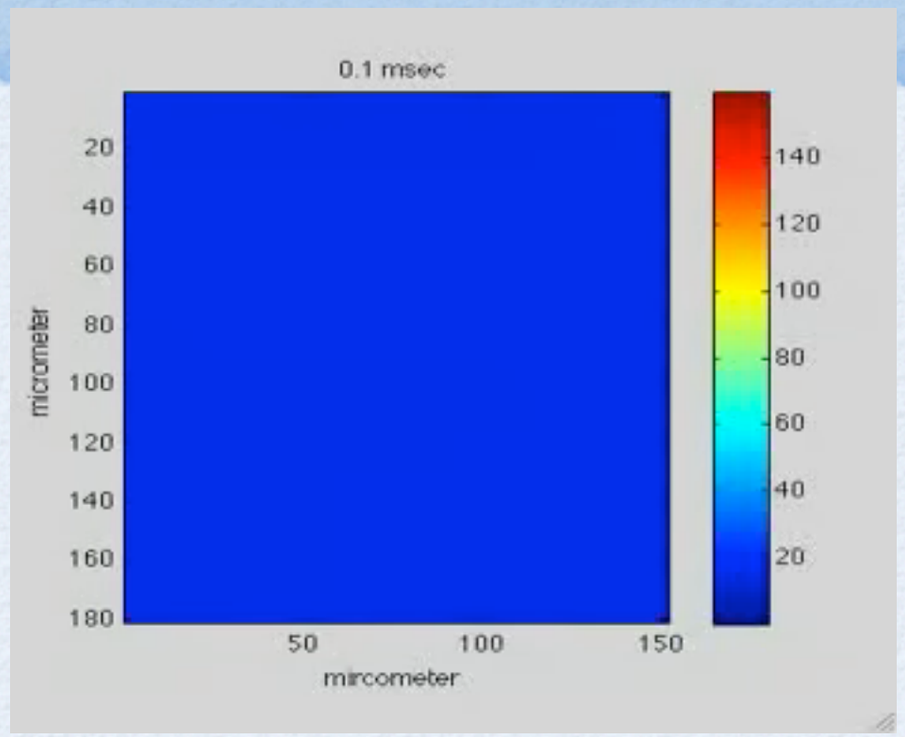
67 ms of data,
viewed as a movie.
[data have been smoothed]

Some spikes move across the array:



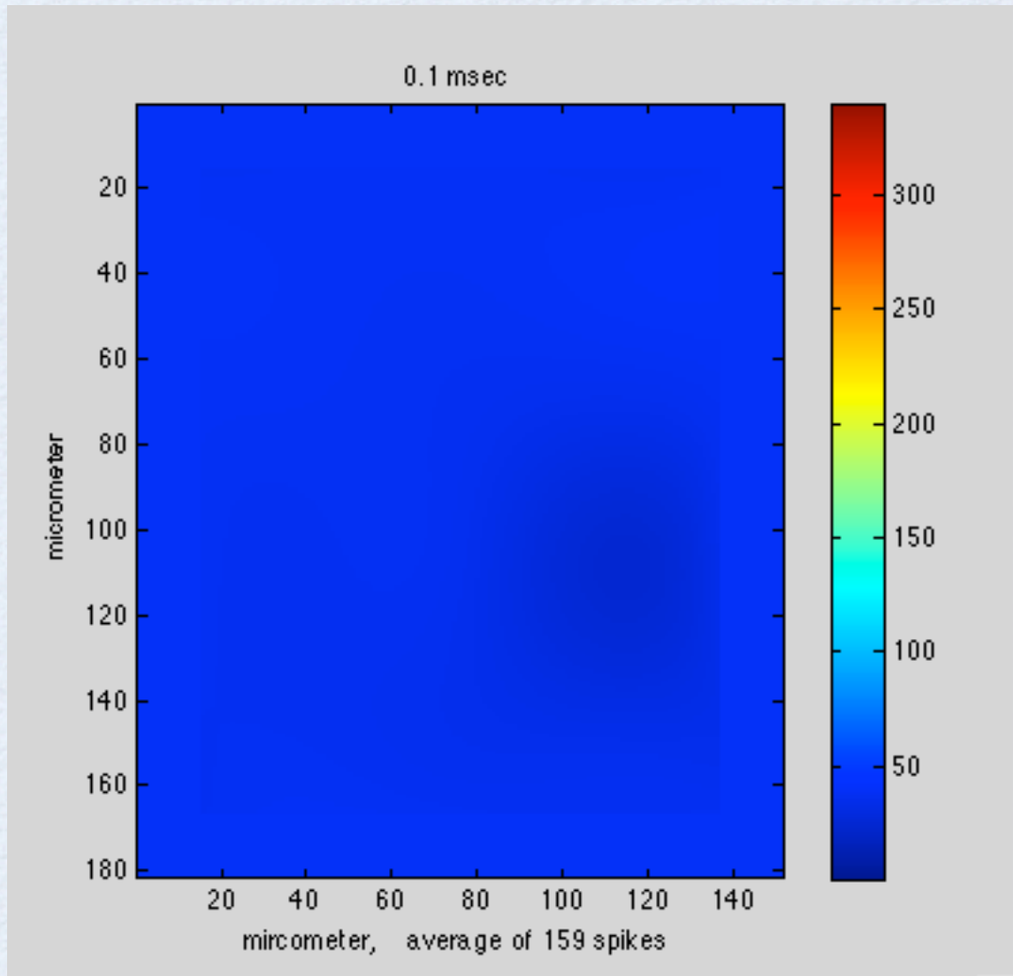
Classic: Gerstein+Clark 1964; Abeles+Goldstein 1977; Schmidt 1984.

Simple events



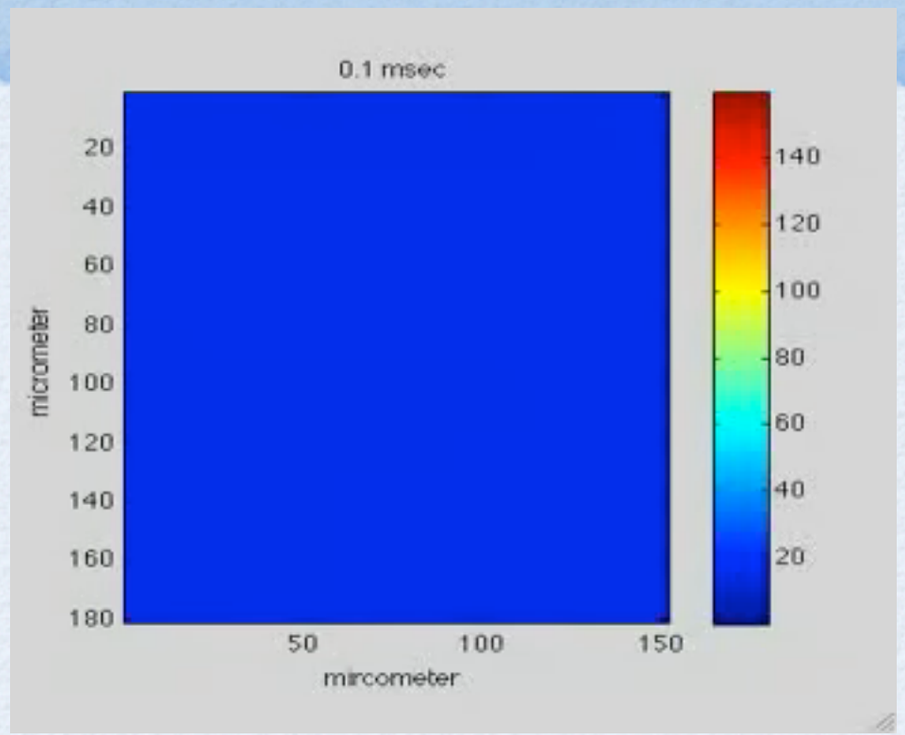
67 ms of data,
viewed as a movie.
[data have been smoothed]

Some spikes move across the array:



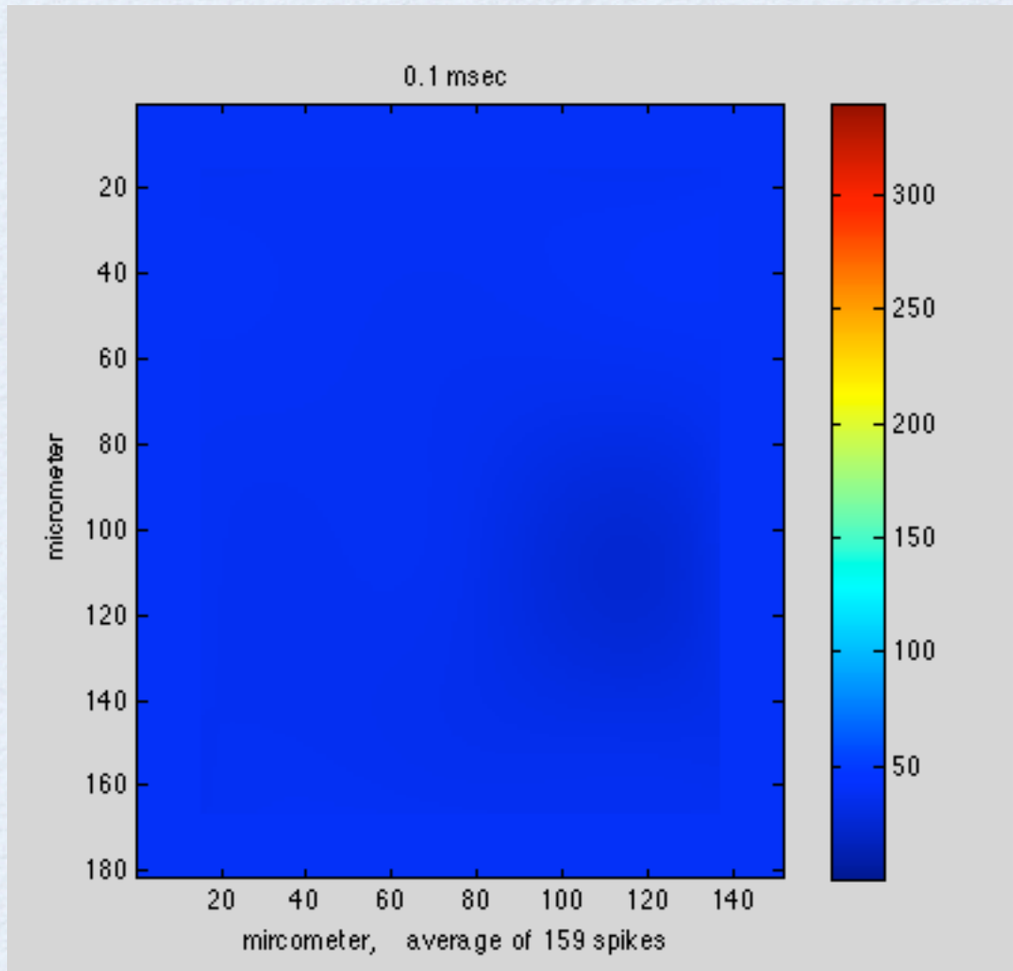
Classic: Gerstein+Clark 1964; Abeles+Goldstein 1977; Schmidt 1984.

Simple events



67 ms of data,
viewed as a movie.
[data have been smoothed]

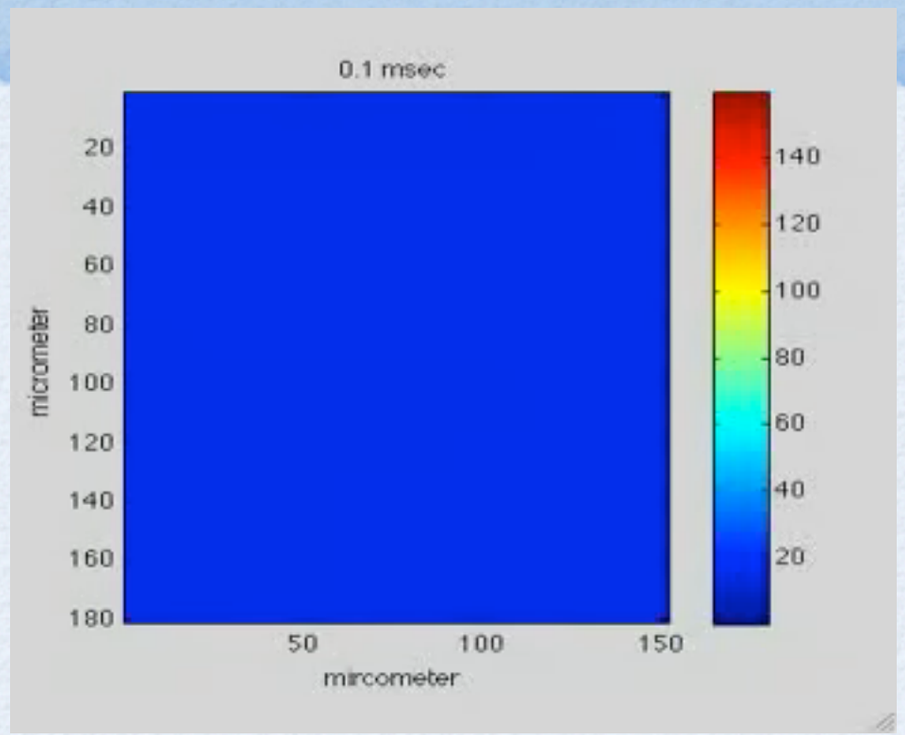
Some spikes move across the array:



Mostly we are hearing retinal ganglion cells, as desired, because they're the ones that spike.

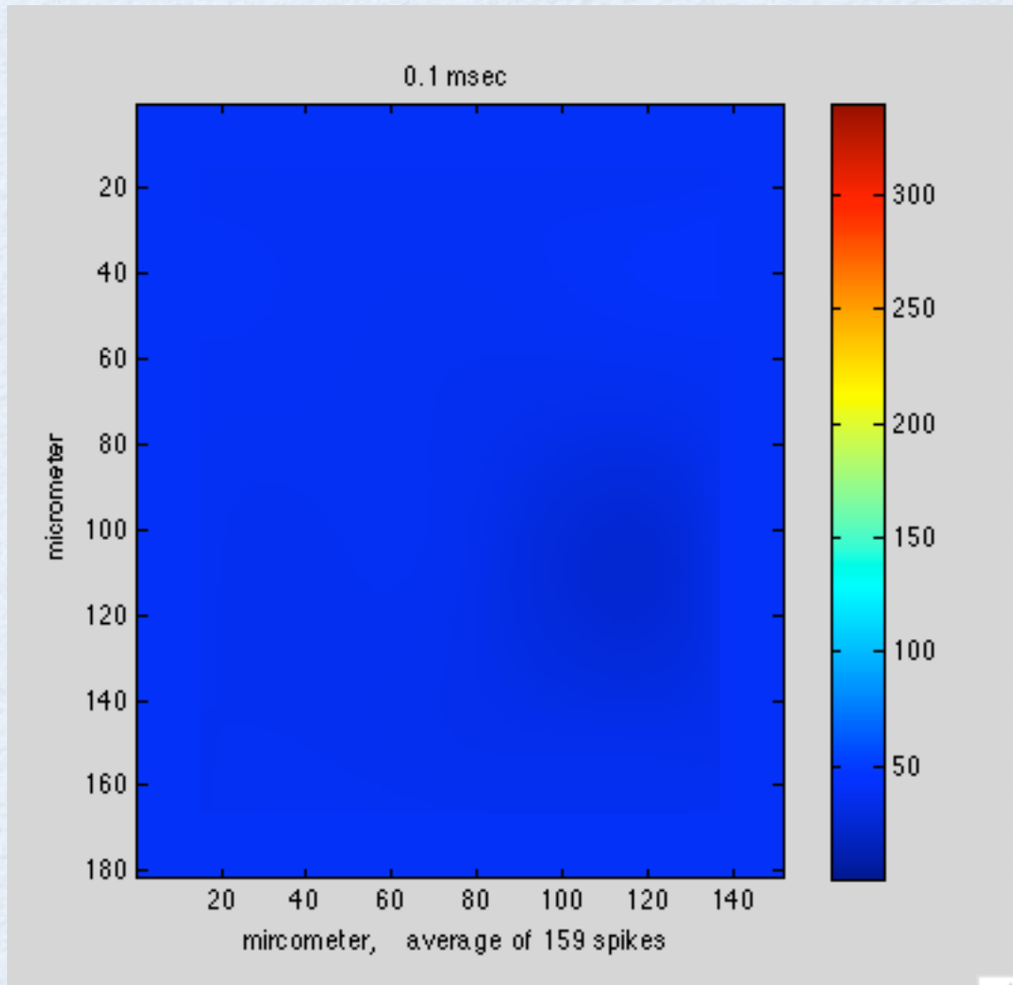
Classic: Gerstein+Clark 1964; Abeles+Goldstein 1977; Schmidt 1984.

Simple events



67 ms of data,
viewed as a movie.
[data have been smoothed]

Some spikes move across the array:



Mostly we are hearing retinal ganglion cells, as desired, because they're the ones that spike.

The spike-sorting problem is: Given raw data like these, convert to a list of discrete events (which cells fired at what times).

Classic: Gerstein+Clark 1964; Abeles+Goldstein 1977; Schmidt 1984.

Not-so-simple events

Unfortunately many events are complex, with multiple overlapping spikes in many locations. And of course these may be the most interesting ones!

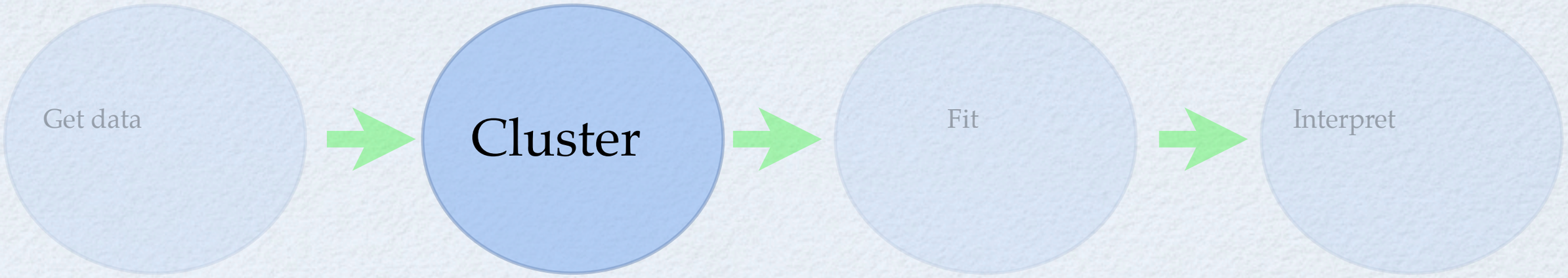
It really matters because "Failure in identification of overlapping spikes from multiple neuron activity causes artificial correlations" [Bar-Gad '01]. Moreover, when we graduate to bigger arrays, nearly all events will involve overlaps in time!!

Many authors say bursts are a big problem, but here is a nice fit that we obtained with no special effort. See later.

We even handle overlapping spikes, which some algorithms do not attempt. See later.

JS Prentice, J Homann, KD Simmons, G Tkacik, V Balasubramanian, PCN, PLoS ONE 6(7): e19884 (2011).

1. Experiment
2. Clustering
3. Fitting
4. Performance



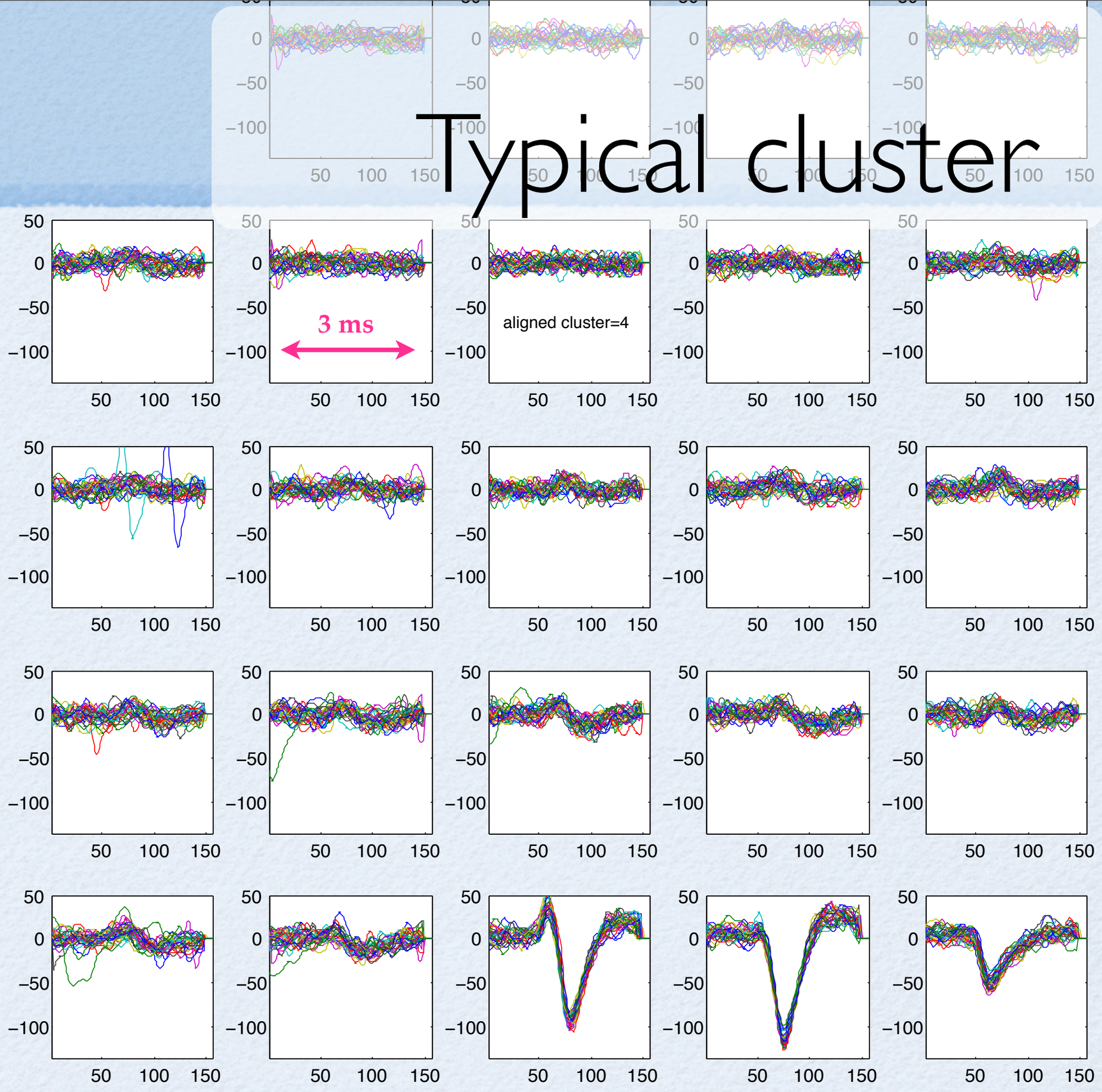
[Sorry, no time to discuss our method for this step.]

Typical cluster

Superposing 50 traces chosen from 284 in this cluster shows that they really do all resemble each other.

Occasional events in which this event collides with another don't affect the "**archetype waveform**" (template) (next slide).

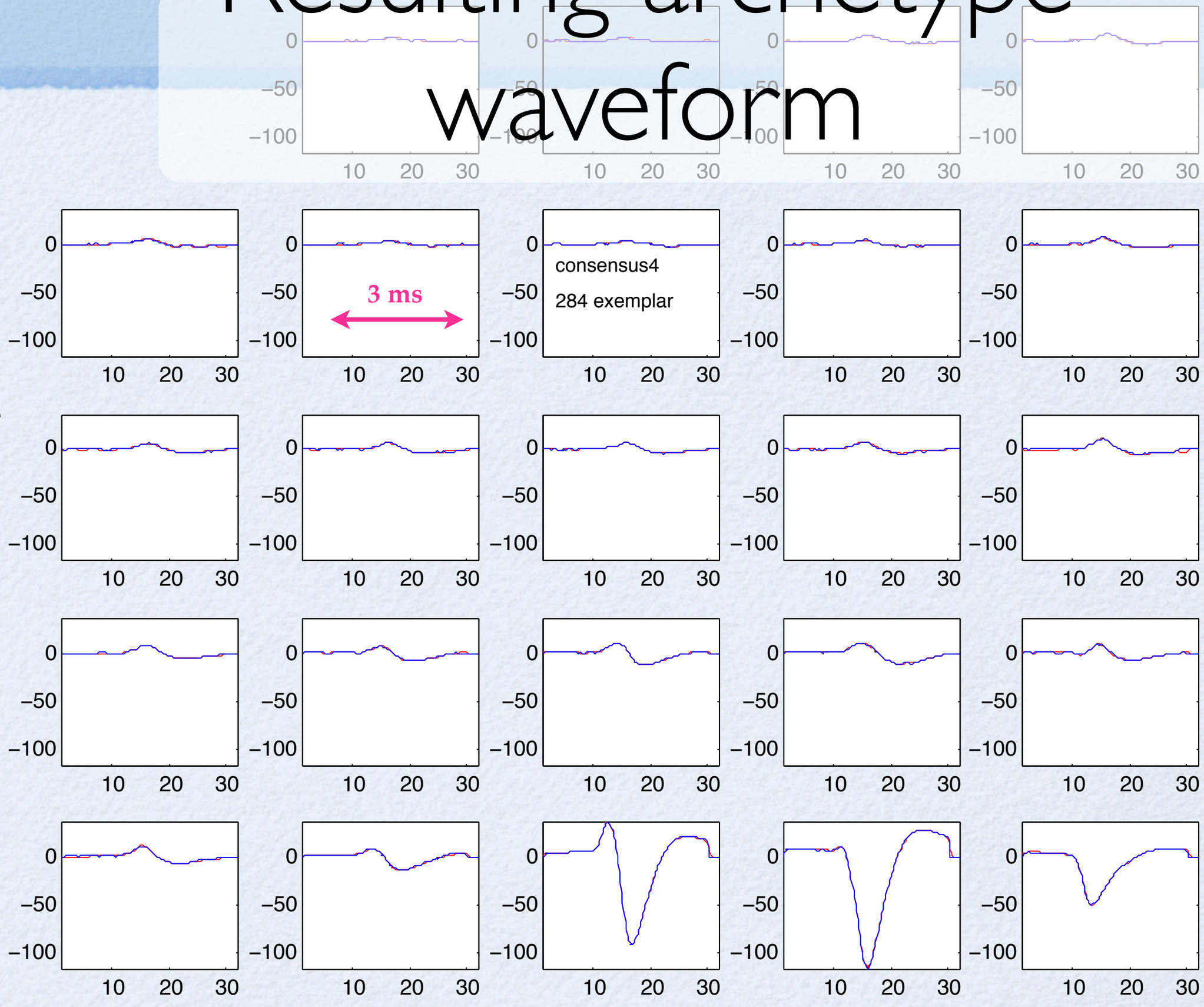
Although the *shape* of each instance of the archetype is quite constant, still its *amplitude* has significant variation.



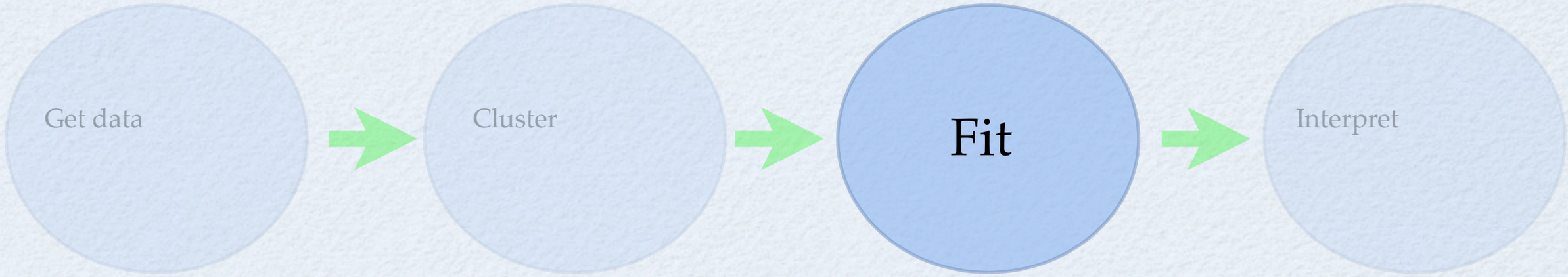
JS Prentice, J Homann, KD Simmons, G Tkacik, V Balasubramanian, PCN, PLoS ONE 6(7): e19884 (2011).

Resulting archetype waveform

We scaled each instance of each archetype to get best agreement with the others, then took the median at each time point to find our best estimate of the consensus waveform (blue). As a check, the pointwise mean waveform looks the same (red).



1. Experiment
2. Clustering
3. Fitting
4. Performance



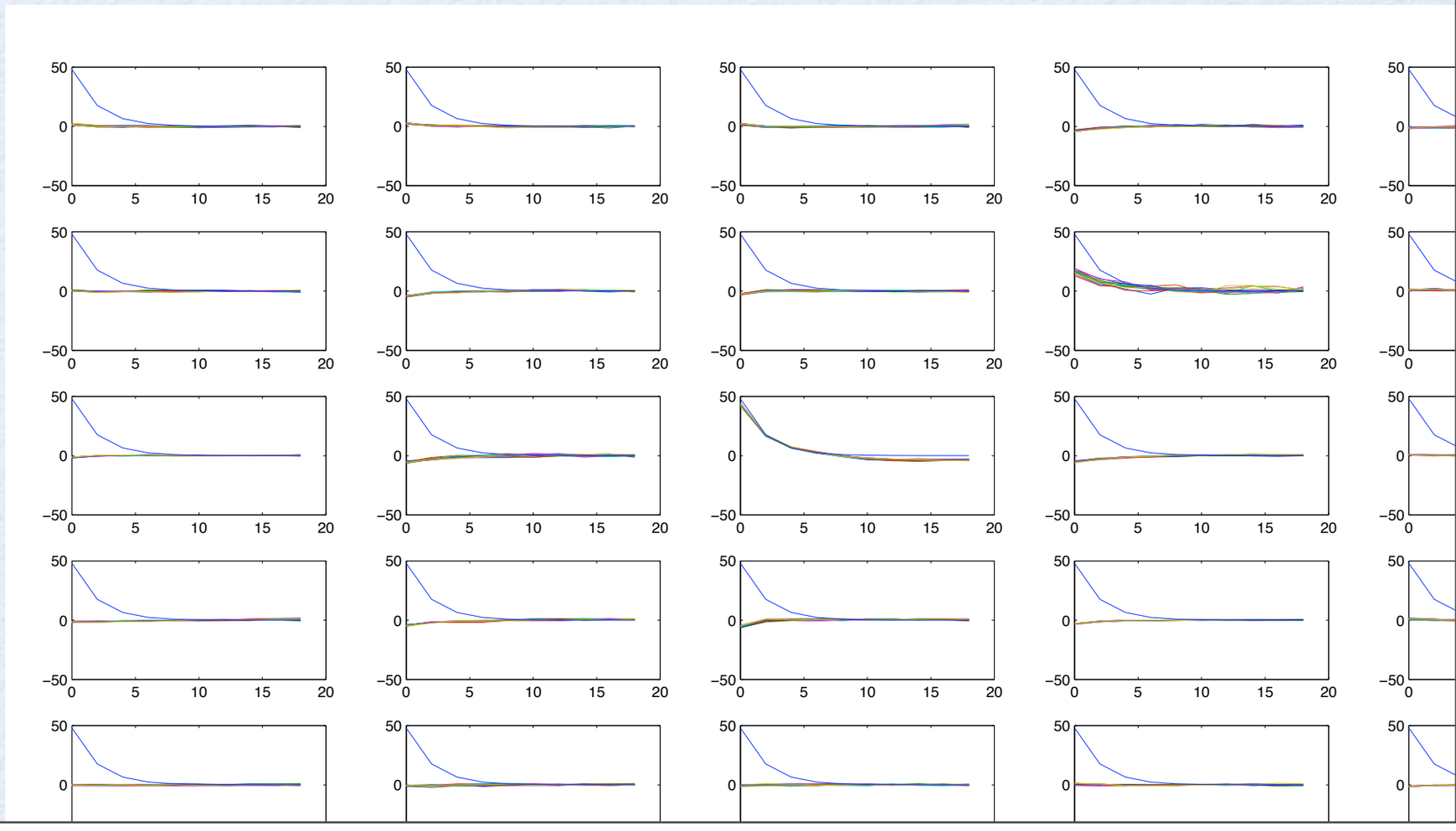
Noise covariance

Vanilla least-squares fitting is not appropriate for time series, because it assumes that every sample is independent of all others--whereas actually, successive samples are correlated.

Here is the covariance of channel #13 with all other channels (after an initial spatial filter, also obtained from data). For reference, each channel has a single blue curve showing an exponential function.

We see that #13 is correlated only with itself, and it has a simple covariance matrix that is easy to invert. The inverse covariance thus obtained defines our correlated Gaussian model of the noise.

[Again: The covariance is **not** a delta function, contrary to what is assumed in naive least-squares fitting.]



On inference

Suppose we measure some experimental data, and wish to make an inference about some situation that we cannot directly observe. That is, we imagine a variety of worlds with different values of X, and ask which is most probable given the observed data.

See M. Denny and S. Gaines, *Chance in Biology*.

On inference

Suppose we measure some experimental data, and wish to make an inference about some situation that we cannot directly observe. That is, we imagine a variety of worlds with different values of X, and ask which is most probable given the observed data.

If we know the probability that those data would have arisen in a world with a particular value of X, then Bayes's formula gives us what we actually want:

$$P(X|\text{observed data}) = P(\text{data}|X)\frac{P(X)}{P(\text{data})}$$

We can ignore the denominator, if all we want is to compare two hypotheses (e.g. maximize over X).

See M. Denny and S. Gaines, *Chance in Biology*.

On inference

Suppose we measure some experimental data, and wish to make an inference about some situation that we cannot directly observe. That is, we imagine a variety of worlds with different values of X, and ask which is most probable given the observed data.

If we know the probability that those data would have arisen in a world with a particular value of X, then Bayes's formula gives us what we actually want:

$$P(X|\text{observed data}) = P(\text{data}|X)\frac{P(X)}{P(\text{data})}$$

We can ignore the denominator, if all we want is to compare two hypotheses (e.g. maximize over X).

For our application, we'd like $\mathbf{P}(\text{spikes} \mid \text{data})$, where "data" is an observed waveform and "spikes" refers to a collection of spike archetypes $\mu_1, \ldots$ occurring at times $t_1, \ldots$ with amplitudes $A_1, \ldots$ relative to the amplitude of the corresponding archetype (neuron). Bayes's formula gives what we want as

$$\mathbf{K} \times (\text{likelihood}) \times (\text{prior}) = \mathbf{KP}\,(\text{data} \mid \text{spikes})\mathbf{P}\,(\text{spikes})$$

See M. Denny and S. Gaines, *Chance in Biology*.

Bayesian idea

Previous slide expressed $\mathbf{P}(\text{spikes} \mid \text{data})$ as:

$$\mathbf{K} \times (\text{likelihood}) \times (\text{prior}) = \mathbf{K}\ \mathbf{P}(\text{data} \mid \text{spikes})\ \mathbf{P}(\text{spikes})$$

Here "spikes" refers to a collection of spike archetypes $\mu_1, \ldots$ occurring at times $t_1, \ldots$ with amplitudes $A_1, \ldots$ relative to the amplitude of the corresponding archetype.

Priors co
itself--an
benefit o
and can b

Bayesian idea

Previous slide expressed $\mathbf{P}(\text{spikes} \mid \text{data})$ as:

$$\mathbf{K} \times (\text{likelihood}) \times (\text{prior}) = \mathbf{K}\ \mathbf{P}(\text{data} \mid \text{spikes})\ \mathbf{P}(\text{spikes})$$

Here "spikes" refers to a collection of spike archetypes $\mu_1, \ldots$ occurring at times $t_1, \ldots$ with amplitudes $A_1, \ldots$ relative to the amplitude of the corresponding archetype.

To get the **prior**, $\mathbf{P}(\text{spikes})$, assume that for a single spike it has the form

$$P^{\text{cell}}(\mu)\, P^{\text{time}}(t)\, P^{\text{ampl}}(A|\mu)$$

The three factors are respectively the popularity of this neuron, uniform in time, and a Gaussian reflecting its typical amplitude and amplitude variability. We get these priors from the data subset used in clustering.

Priors co
itself--an
benefit o
and can b

Bayesian idea

Previous slide expressed $\mathbf{P}(\text{spikes} \mid \text{data})$ as:

$$\mathbf{K} \times (\text{likelihood}) \times (\text{prior}) = \mathbf{K}\ \mathbf{P}(\text{data} \mid \text{spikes})\ \mathbf{P}(\text{spikes})$$

Here "spikes" refers to a collection of spike archetypes $\mu_1, \ldots$ occurring at times $t_1, \ldots$ with amplitudes $A_1, \ldots$ relative to the amplitude of the corresponding archetype.

To get the **prior**, $\mathbf{P}(\text{spikes})$, assume that for a single spike it has the form

$$P^{\text{cell}}(\mu)\, P^{\text{time}}(t)\, P^{\text{ampl}}(A|\mu)$$

The three factors are respectively the popularity of this neuron, uniform in time, and a Gaussian reflecting its typical amplitude and amplitude variability. We get these priors from the data subset used in clustering.

To get the **likelihood function** $\mathbf{P}(\text{data} \mid \text{spikes})$, suppose that the data consist of one archetype, plus noise. And suppose that the noise is some **Gaussian,** independent of which spikes fired. We know all about this Gaussian from our measurement of noise covariance.

Then the likelihood is that distribution, evaluated at the difference between the actual waveform and the idealized one. [Pouzat et. al. 2002]

Priors co
itself--an
benefit o
and can b

Bayesian idea, II

We start with an experimental trace ("data").
We find its peak (absolute minimum), and start looking for a spike there.

We ask for the **likelihood ratio** between the hypotheses of no spike versus one spike of given type, at given time, with given amplitude.

★To compute the likelihood of no spike, evaluate the noise distribution on the trace.

★To compute the probability of one spike, choose a spike archetype and a value of **t,** the spike time. Holding the "data" fixed, **the probability is now a Gaussian function in the remaining parameter A,** so it's fast and easy to marginalize over **A**.

[Nuts and Bolts]

Let $V_\alpha(t)$ be measured voltage, electrode α and $F_{\mu\alpha}(t)$ be archetype waveform of type μ. Define the deviation $[\delta\mathbf{V}]_{\alpha t} = V_\alpha(t) - AF_{\mu\alpha}(t - t_1)$

Then the probability that one spike, of type μ, is present is

The noise covariance

$$P(\text{spikes}\,|\,\text{data}) = K_\mu \exp\left[-\frac{(A-\gamma_\mu)^2}{2\sigma_\mu^2} - \frac{1}{2}(\delta\mathbf{V})^{\mathrm{t}}\mathsf{C}^{-1}(\delta\mathbf{V})\right]$$

which is a Gaussian in A. So it's easy to marginalize over A: **just complete the square!**
[Here $K_\mu = P^{\text{cell}}(\mu)P^{\text{time}}(t_1)\left(2\pi\sigma_\mu^2\right)^{-1/2}$ doesn't depend on A.]

JS Prentice, J Homann, KD Simmons, G Tkacik, V Balasubramanian, PCN, PLoS ONE 6(7): e19884 (2011).

[Nuts and Bolts]

Let $V_\alpha(t)$ be measured voltage, electrode α and $F_{\mu\alpha}(t)$ be archetype waveform of type μ. Define the deviation $[\delta\mathbf{V}]_{\alpha t} = V_\alpha(t) - AF_{\mu\alpha}(t - t_1)$

Then the probability that one spike, of type μ, is present is

The noise covariance

$$P(\text{spikes}\,|\,\text{data}) = K_\mu \exp\left[-\frac{(A-\gamma_\mu)^2}{2\sigma_\mu^2} - \frac{1}{2}(\delta\mathbf{V})^{\mathrm{t}}\mathsf{C}^{-1}(\delta\mathbf{V})\right]$$

which is a Gaussian in A. So it's easy to marginalize over A: **just complete the square!**
[Here $K_\mu = P^{\text{cell}}(\mu)P^{\text{time}}(t_1)\left(2\pi\sigma_\mu^2\right)^{-1/2}$ doesn't depend on A.]

Next, we sweep over a range of $\mathbf{t}$ to find the best value of likelihood ratio for this spike type. [We only check $\mathbf{t}$ values close to the peak of the event.]

JS Prentice, J Homann, KD Simmons, G Tkacik, V Balasubramanian, PCN, PLoS ONE 6(7): e19884 (2011).

[Nuts and Bolts]

Let $V_\alpha(t)$ be measured voltage, electrode α and $F_{\mu\alpha}(t)$ be archetype waveform of type μ. Define the deviation $[\delta\mathbf{V}]_{\alpha t} = V_\alpha(t) - AF_{\mu\alpha}(t - t_1)$

Then the probability that one spike, of type μ, is present is

The noise covariance

$$P(\text{spikes}\,|\,\text{data}) = K_\mu \exp\left[-\frac{(A-\gamma_\mu)^2}{2\sigma_\mu^2} - \frac{1}{2}(\delta\mathbf{V})^{\mathrm{t}}\mathsf{C}^{-1}(\delta\mathbf{V})\right]$$

which is a Gaussian in A. So it's easy to marginalize over A: **just complete the square!**
[Here $K_\mu = P^{\text{cell}}(\mu)P^{\text{time}}(t_1)\left(2\pi\sigma_\mu^2\right)^{-1/2}$ doesn't depend on A.]

Next, we sweep over a range of $\mathbf{t}$ to find the best value of likelihood ratio for this spike type. [We only check $\mathbf{t}$ values close to the peak of the event.]

Then we choose the winner among spike types.

JS Prentice, J Homann, KD Simmons, G Tkacik, V Balasubramanian, PCN, PLoS ONE 6(7): e19884 (2011).

[Nuts and Bolts]

Let $V_\alpha(t)$ be measured voltage, electrode α and $F_{\mu\alpha}(t)$ be archetype waveform of type μ. Define the deviation $[\delta\mathbf{V}]_{\alpha t} = V_\alpha(t) - AF_{\mu\alpha}(t - t_1)$

Then the probability that one spike, of type μ, is present is

The noise covariance

$$P(\text{spikes}\,|\,\text{data}) = K_\mu \exp\left[-\frac{(A-\gamma_\mu)^2}{2\sigma_\mu^2} - \frac{1}{2}(\delta\mathbf{V})^{\mathrm{t}}\mathsf{C}^{-1}(\delta\mathbf{V})\right]$$

which is a Gaussian in A. So it's easy to marginalize over A: **just complete the square!**
[Here $K_\mu = P^{\text{cell}}(\mu)P^{\text{time}}(t_1)\left(2\pi\sigma_\mu^2\right)^{-1/2}$ doesn't depend on A.]

Next, we sweep over a range of $\mathbf{t}$ to find the best value of likelihood ratio for this spike type. [We only check $\mathbf{t}$ values close to the peak of the event.]

Then we choose the winner among spike types.

If the winner's likelihood ratio is good enough (bigger than about 1), we say there's a spike here. **That's the absolute criterion I promised earlier.**

JS Prentice, J Homann, KD Simmons, G Tkacik, V Balasubramanian, PCN, PLoS ONE 6(7): e19884 (2011).

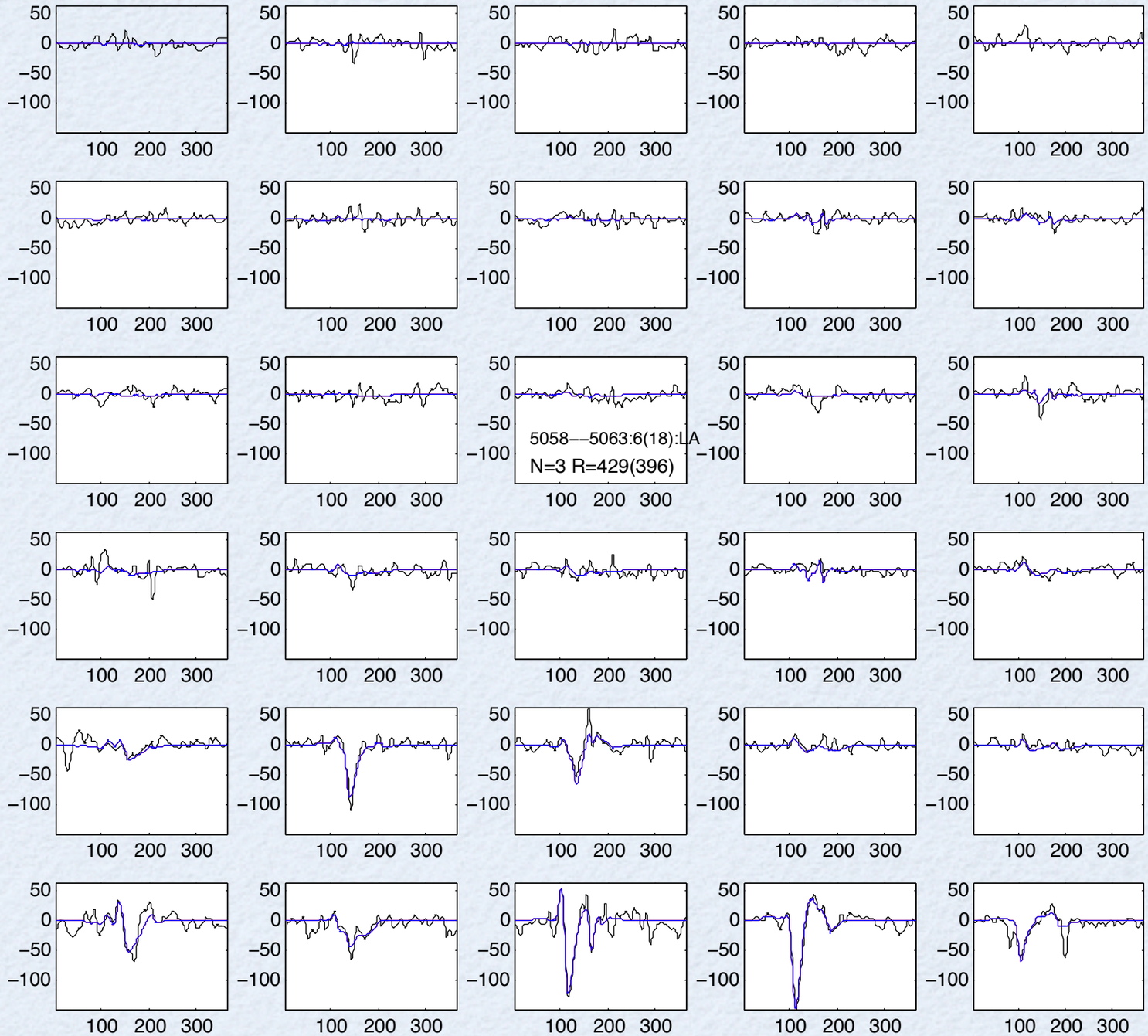
Test our assumptions

Can we really assume that the spikes from a particular cell differ only in overall amplitude? We took many events that contained a single spike of each type. Point by point in time, we subtracted the scaled shifted archetype and found the residual (on each channel).

Green: the archetype itself.
Red: mean deviation from archetype.
Blue: std deviation from archetype.
We really do subtract spikes pretty completely.

say more and better

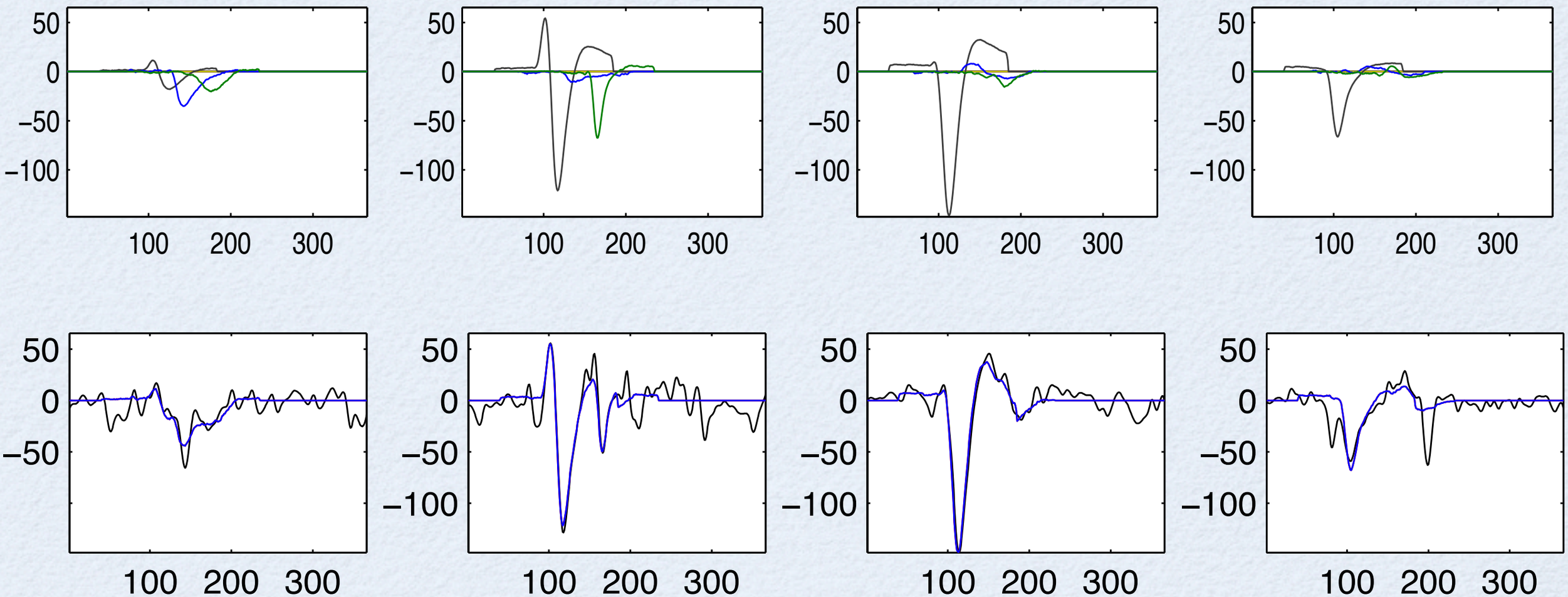
Successfully fit overlaps



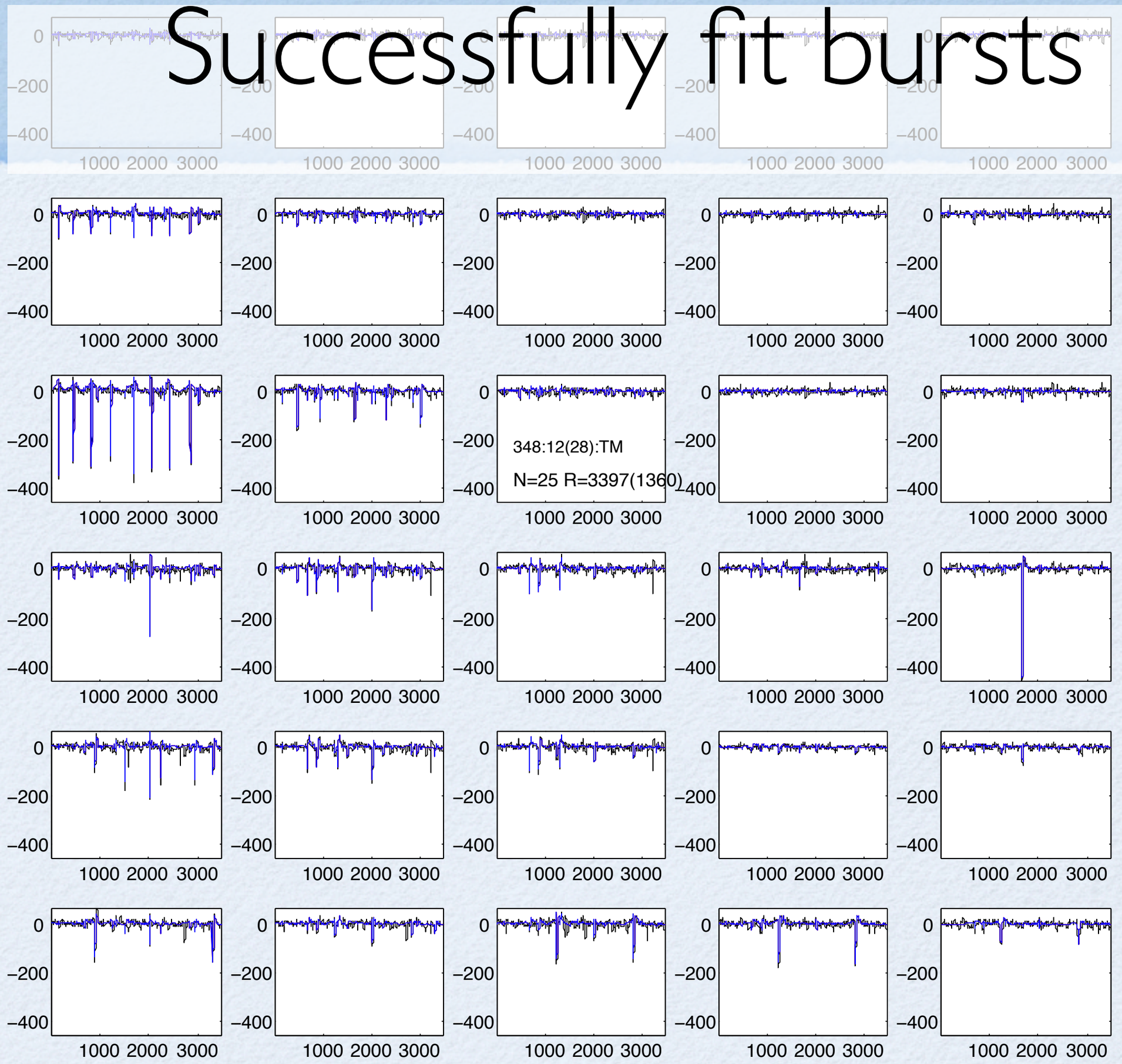
Successfully fit overlaps

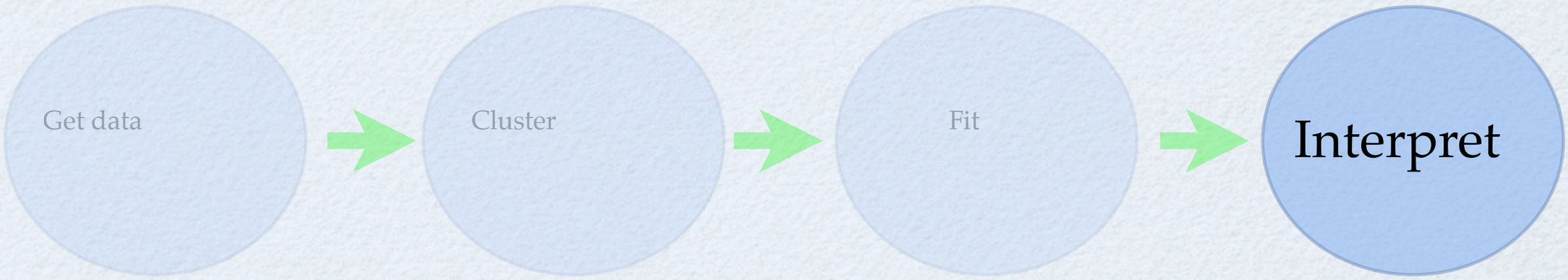
Successfully fit overlaps

Top: Closeup of four channels, showing three fit archetypes found by the algorithm.
Bottom: sum of those fits (color) versus actual data (black).



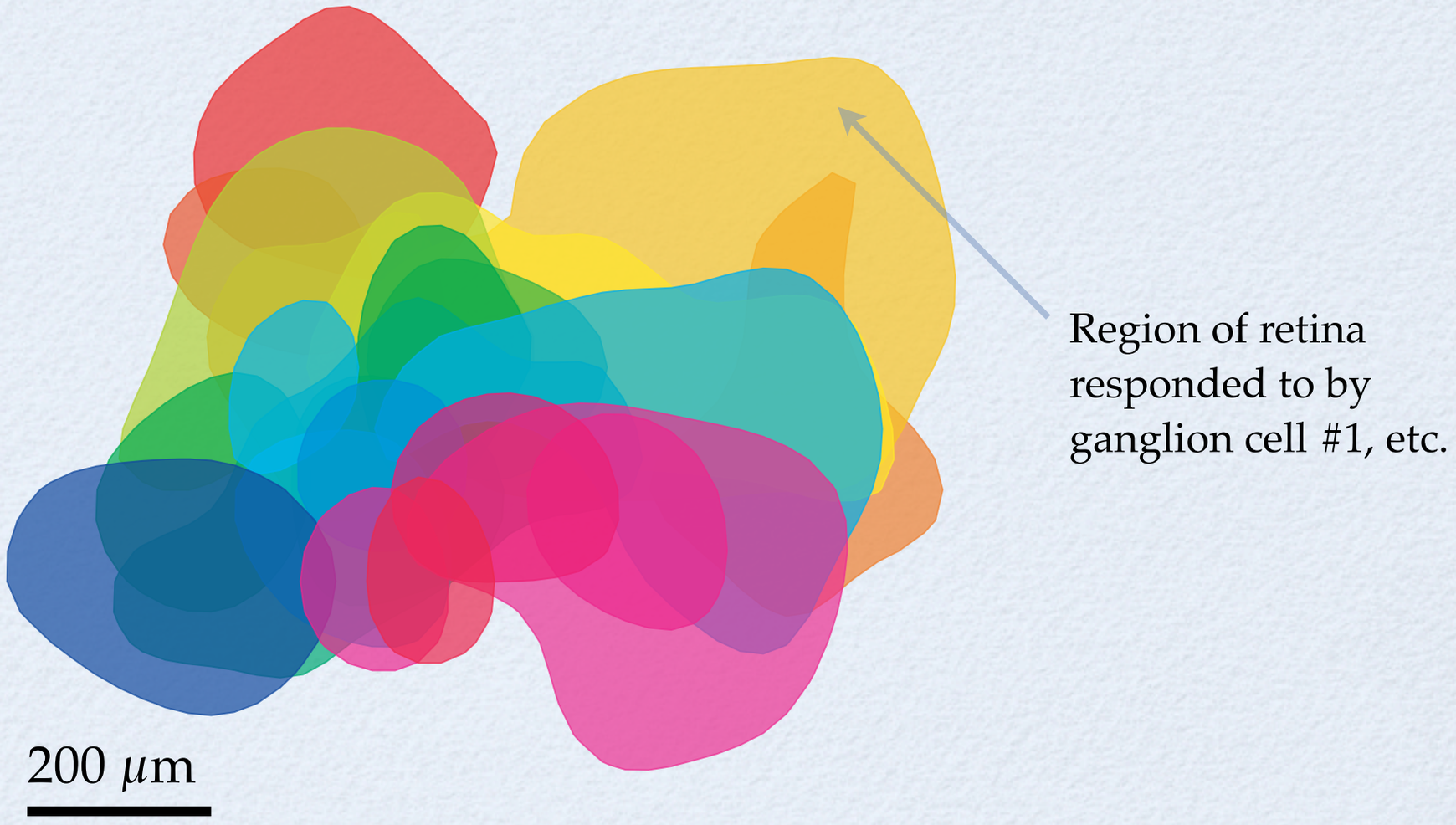
Successfully fit bursts




Get data
Cluster
Fit
Interpret

Each cell has a receptive field...

... and they tile the whole visual field. MEA recording is **high throughput**: We got dozens of cells all at once. Here are cells from just one functional group, "on cells." Each putative receptive field is a single connected region of image space.



KD Simmons, JS Prentice, G Tkacik, J Homann, PCN, V Balasubramanian, submitted.

Receptive fields

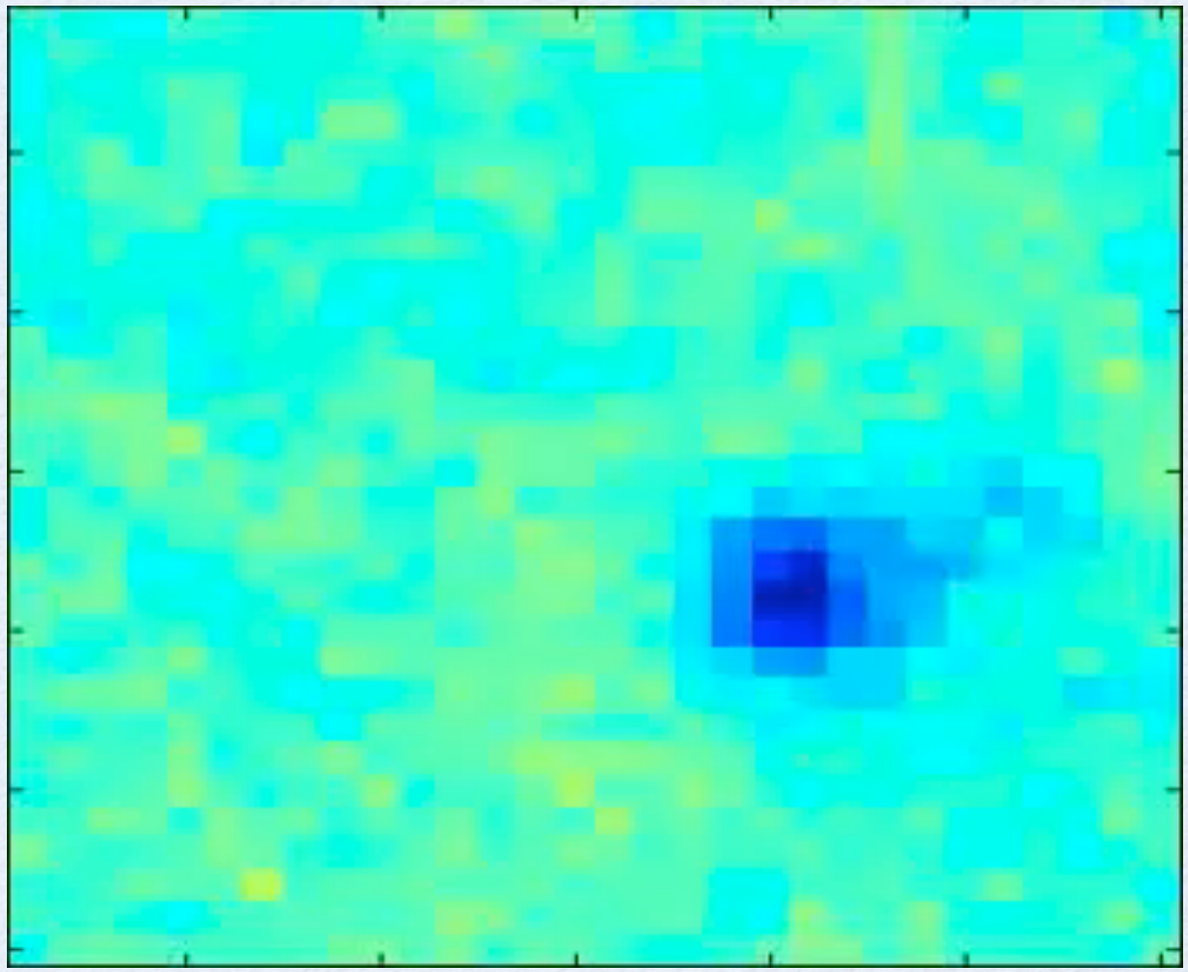
Once you've got the spike trains, you can find receptive fields etc. Here's a typical spike-triggered average.

How interesting--guinea pig retina has a lot of these highly anisotropic receptive fields. The "surround" doesn't surround the "center"!

Receptive fields

Once you've got the spike trains, you can find receptive fields etc. Here's a typical spike-triggered average.

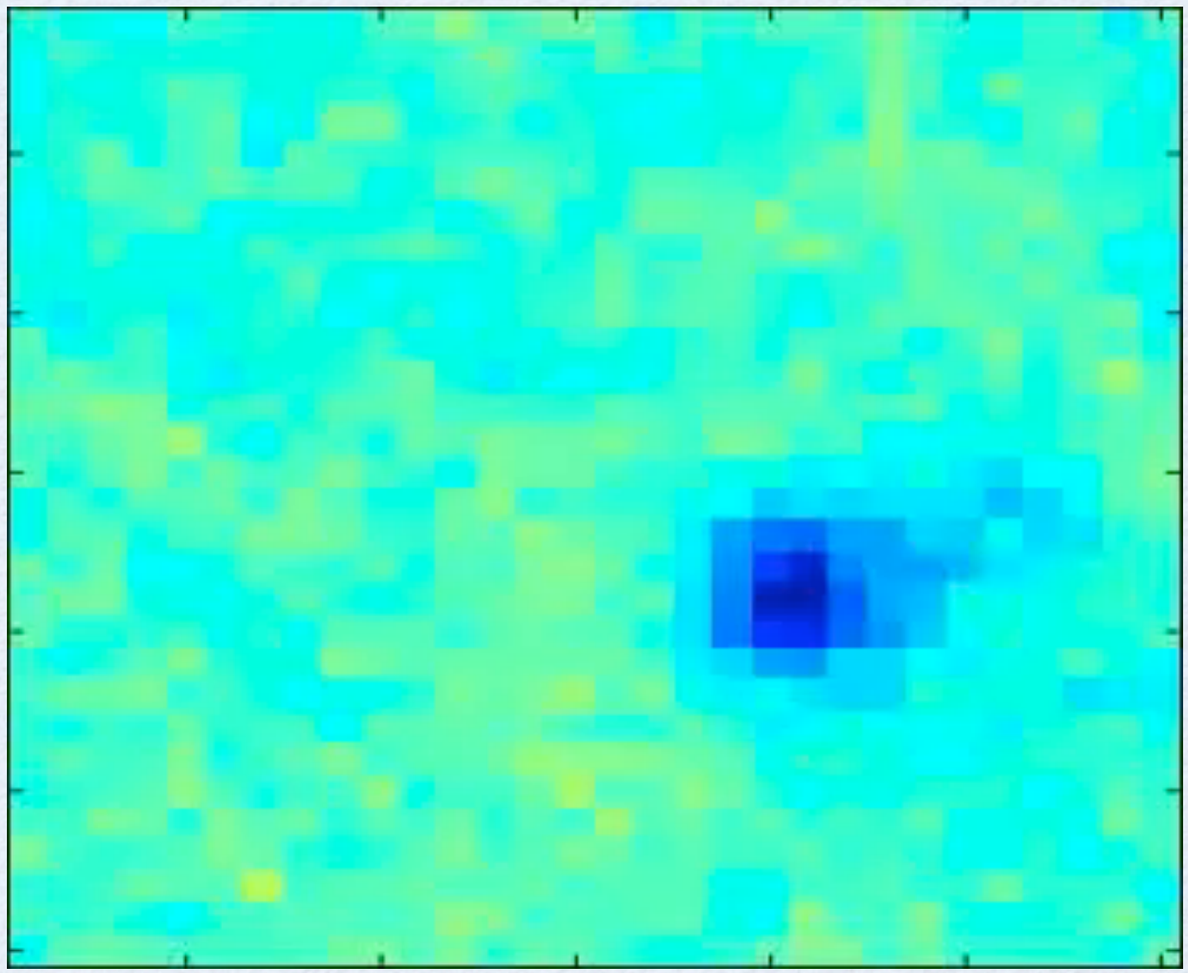
How interesting--guinea pig retina has a lot of these highly anisotropic receptive fields. The "surround" doesn't surround the "center"!



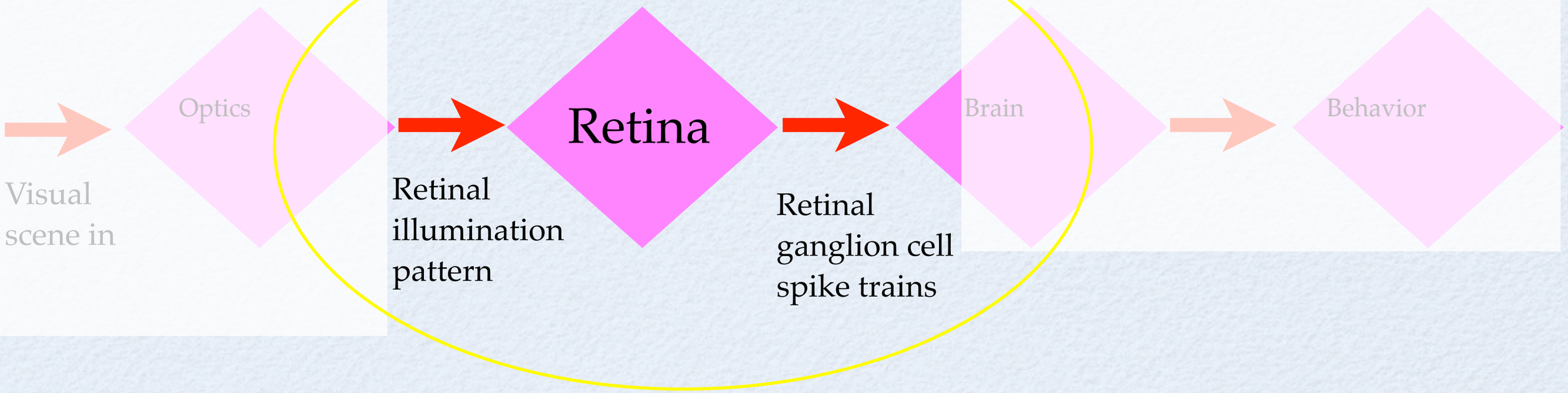
Receptive fields

Once you've got the spike trains, you can find receptive fields etc. Here's a typical spike-triggered average.

How interesting--guinea pig retina has a lot of these highly anisotropic receptive fields. The "surround" doesn't surround the "center"!

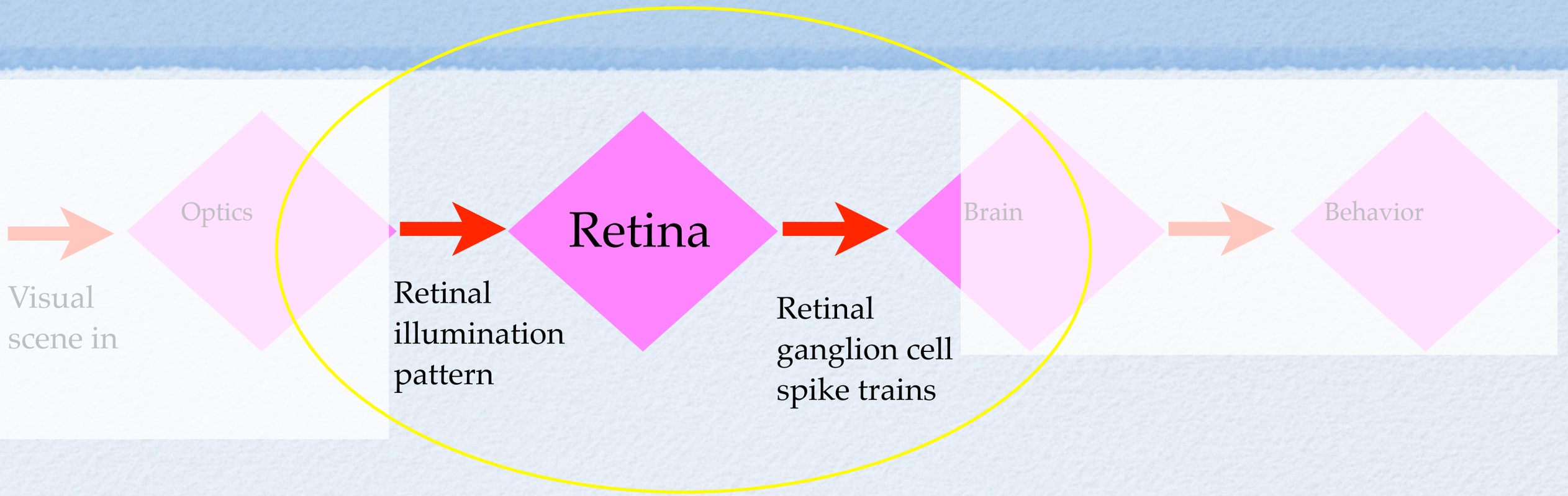


Takehome Part III

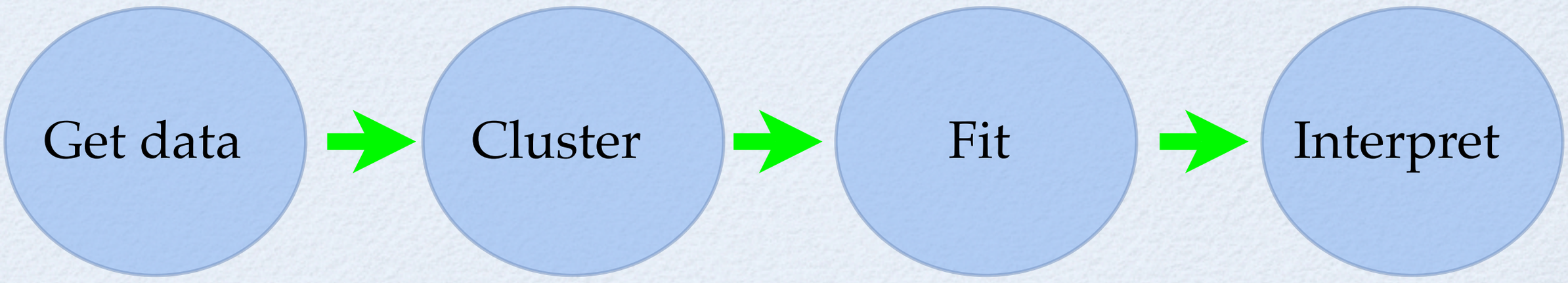


I described how we identify the individual ganglion cell signals from a hash of noise and overlapping real signals:

Takehome Part III



I described how we identify the individual ganglion cell signals from a hash of noise and overlapping real signals:

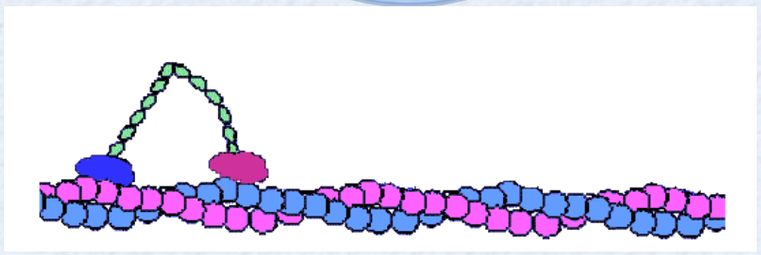


Full circle

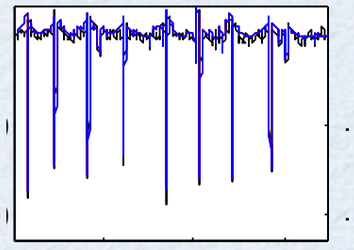
Medical tests

OK, I was a scatterbrain and gave you three talks. But wait -- if I can fill in the spaces

Changepoint Analysis



Multi-Electrode Array



Full circle

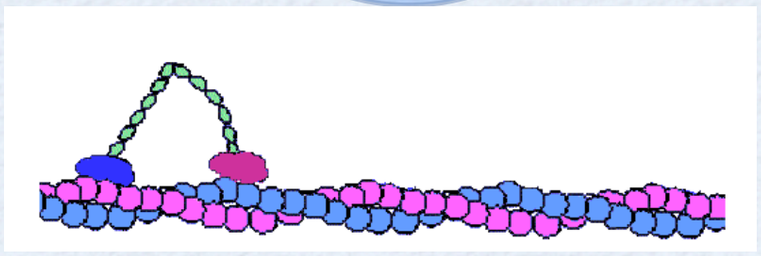
OK, I was a scatterbrain and gave you three talks. But wait -- if I can fill in the spaces

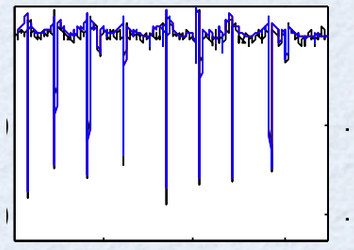
Medical tests

Changepoint Analysis

Poisson processes

Multi-Electrode Array





Full circle

OK, I was a scatterbrain and gave you three talks. But wait -- if I can fill in the spaces

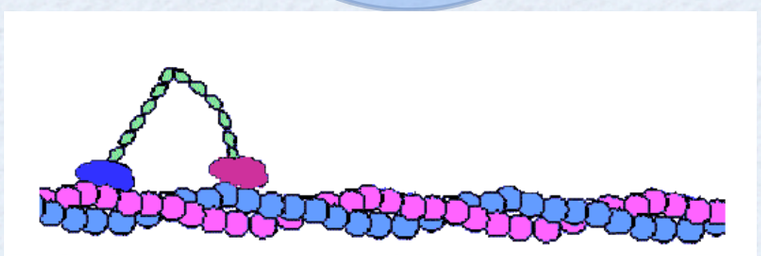
Medical tests

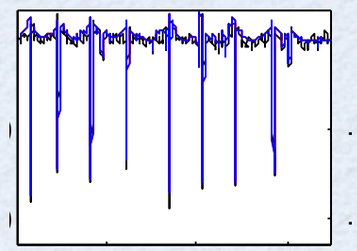
Bayes formula

Changepoint Analysis

Poisson processes

Multi-Electrode Array





Full circle

OK, I was a scatterbrain and gave you three talks. But wait -- if I can fill in the spaces

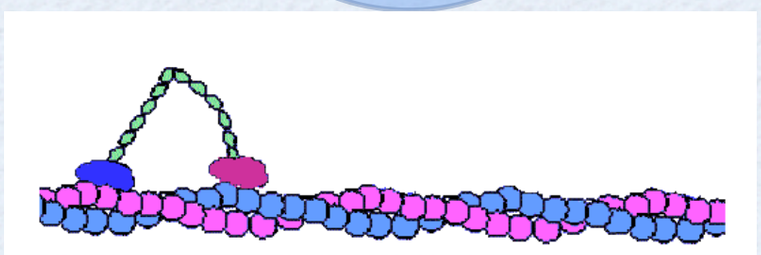
Medical tests

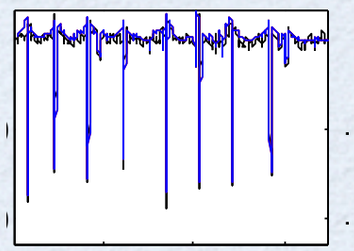
Bayes formula

Changepoint Analysis

Multi-Electrode Array

Poisson processes





Theory can cut across apparently different kinds of experiment, offering useful methods to one domain from another without having to reinvent everything.

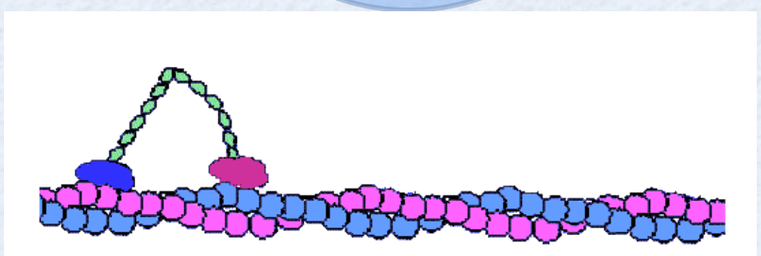
Wait, there's more

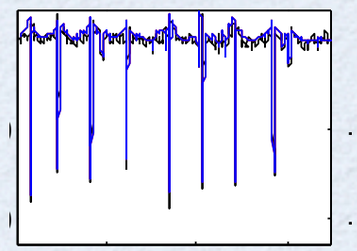
Bayes formula

Changepoint Analysis

Multi-Electrode Array

Poisson processes





Wait, there's more

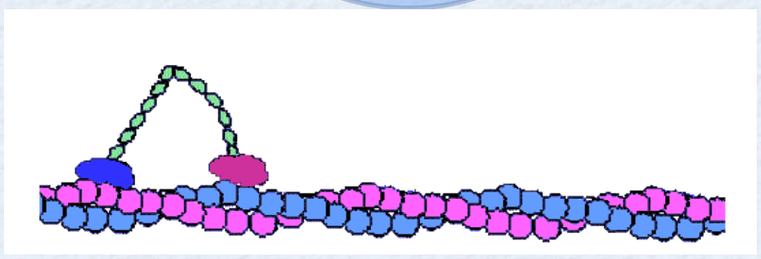
A ***physical model*** -- localized spreading of potential changes in solution -- helped us to extract what was going on.

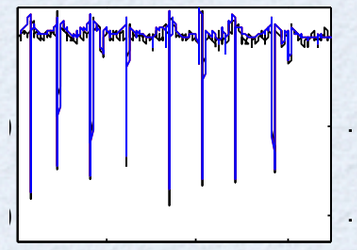
Bayes formula

Changepoint Analysis

Multi-Electrode Array

Poisson processes





Wait, there's more

A ***physical model*** -- photon theory -- helped us to extract what was going on.

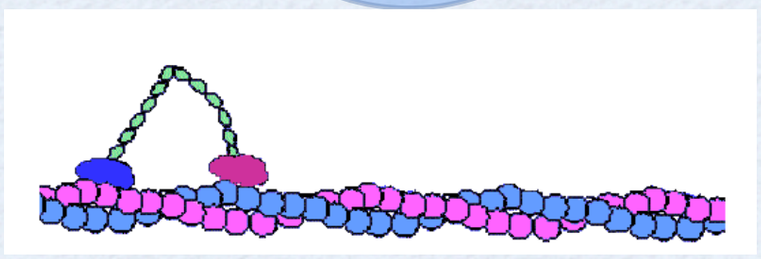
A ***physical model*** -- localized spreading of potential changes in solution -- helped us to extract what was going on.

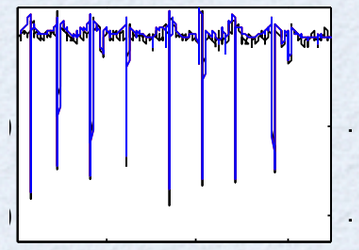
Bayes formula

Changepoint Analysis

Multi-Electrode Array

Poisson processes

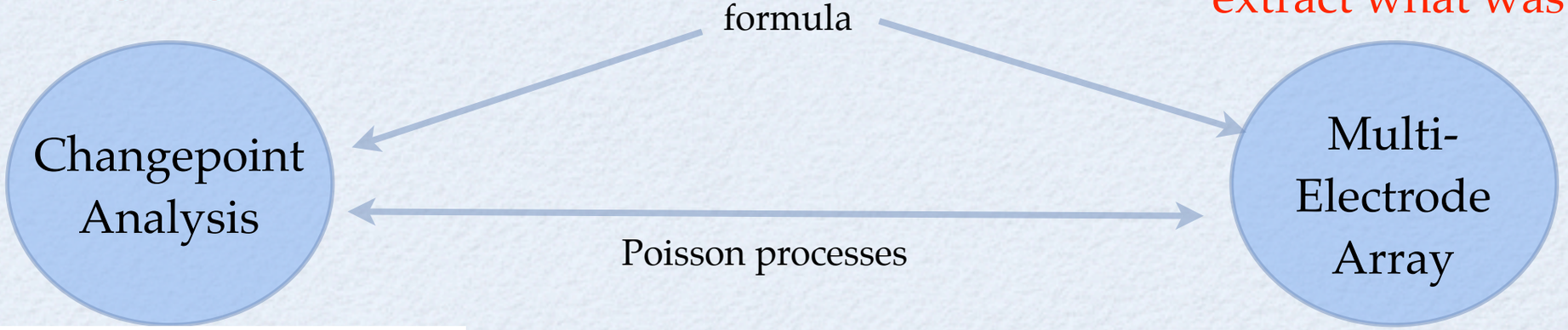


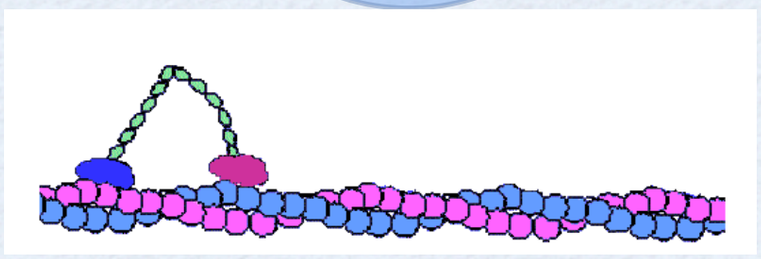


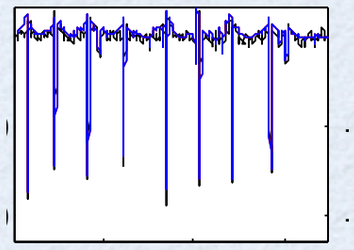
Wait, there's more

A ***physical model*** -- photon theory -- helped us to extract what was going on.

A ***physical model*** -- localized spreading of potential changes in solution -- helped us to extract what was going on.







There is something weirdly -- *unreasonably* -- effective about approaching biological systems with a physical model. I don't understand why. I don't *need* to understand why.

Go long

along the lines of, "clean up the bibliography"

Go long

Often, when we want to justify theory, we scratch our heads and say, "Well Hodgkin and Huxley was a big deal."

along the lines of, "clean up the bibliography"

Go long

Often, when we want to justify theory, 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 theory is something that happens every 50 years or so. It's also too reverent.

along the lines of, "clean up the bibliography"

Go long

Often, when we want to justify theory, 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 theory is something that happens every 50 years or so. It's also too reverent.

My point so far is that ***theory is needed every day***. It's our microscope; our Geiger counter; it helps us to see the invisible.

along the lines of, "clean up the bibliography"

Go long

Often, when we want to justify theory, 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 theory is something that happens every 50 years or so. It's also too reverent.

My point so far is that ***theory is needed every day***. It's our microscope; our Geiger counter; it helps us to see the invisible.
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).

along the lines of, "clean up the bibliography"

Go long

Often, when we want to justify theory, 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 theory is something that happens every 50 years or so. It's also too reverent.

My point so far is that ***theory is needed every day***. It's our microscope; our Geiger counter; it helps us to see the invisible.
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).

Another context in which theory enters laboratory discussions is, "W
theory to get this thing published. Go do some theory, run some ANOVA, whatever."

along the lines of, "clean up the bibliography"

Go long

Often, when we want to justify theory, 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 theory is something that happens every 50 years or so. It's also too reverent.

My point so far is that ***theory is needed every day***. It's our microscope; our Geiger counter; it helps us to see the invisible.
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).

Another context in which theory enters laboratory discussions is, "W
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. Particularly, a ***physical model*** can give a lot of dividends.

along the lines of, "clean up the bibliography"

Go long

Often, when we want to justify theory, 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 theory is something that happens every 50 years or so. It's also too reverent.

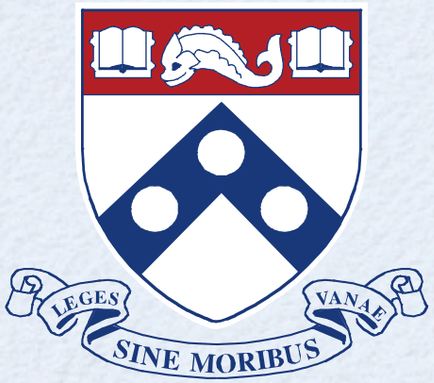
My point so far is that ***theory is needed every day***. It's our microscope; our Geiger counter; it helps us to see the invisible.
To emphasize that, I didn't select famous examples; instead I have told you about the two things I happen to be working on right now (a random choice, you'll agree).

Another context in which theory enters laboratory discussions is, "W
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. Particularly, a ***physical model*** can give a lot of dividends.

along the lines of, "clean up the bibliography"

We like to teach famous success stories in science, but we don't always remember to present them as showcases of the utility of physical modeling.

Thanks

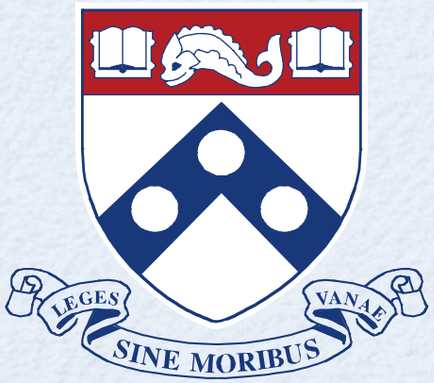


University of Pennsylvania

For these slides see:

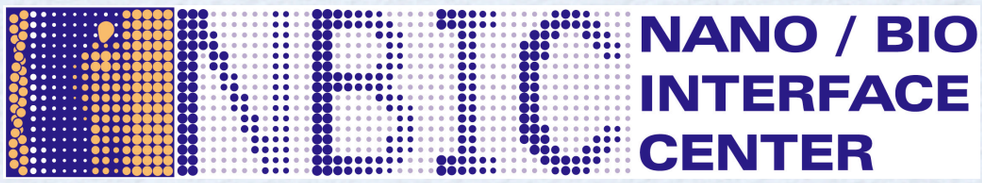
www.physics.upenn.edu/~pcn

Thanks



University of Pennsylvania

NSF DMR,
BIO, IBN



NSF NSEC

National Eye Institute
Training grant

Computational Neuroscience
Training grant

For these slides see:

www.physics.upenn.edu/~pcn