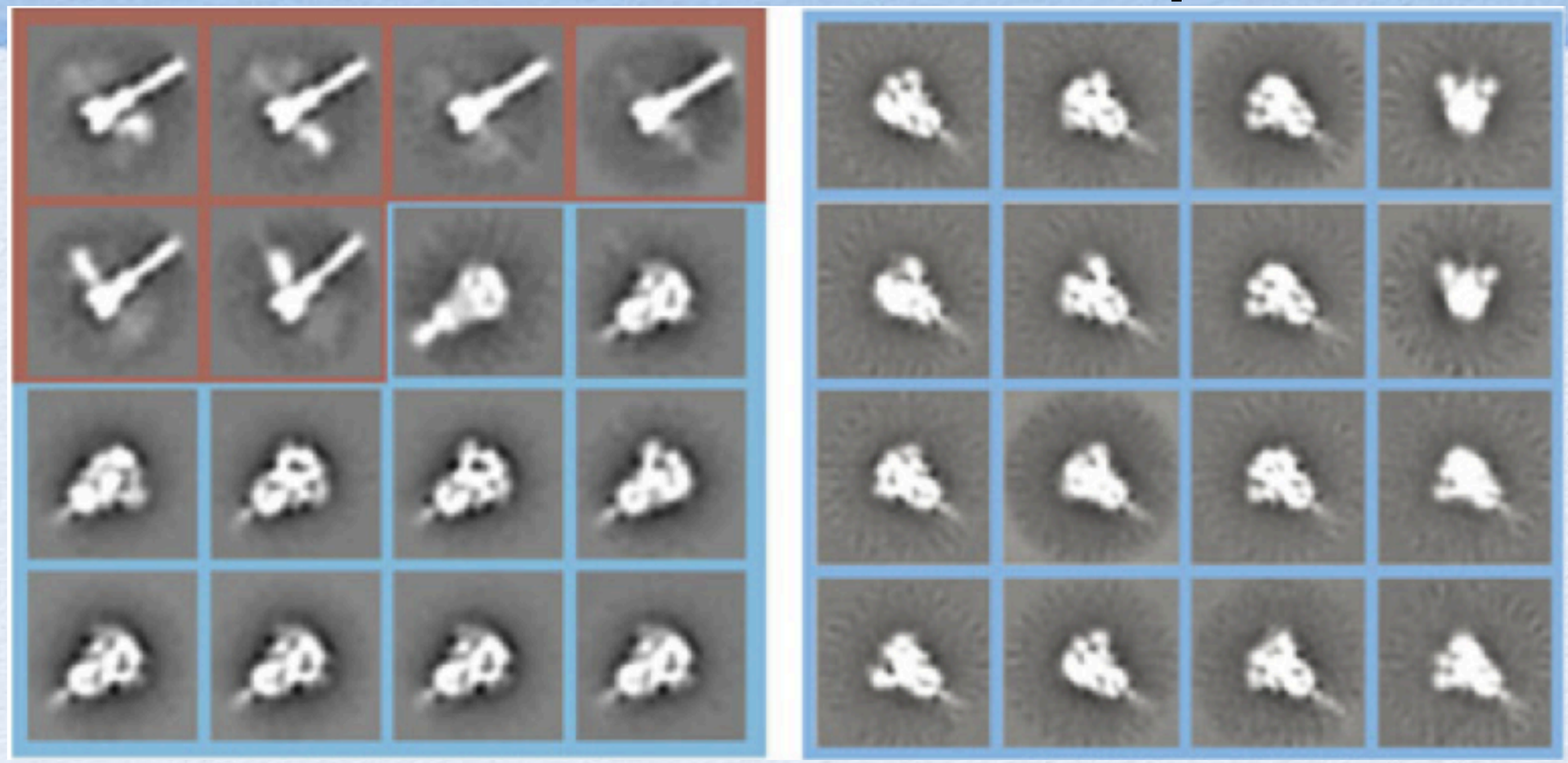
Phil Nelson
University of Pennsylvania

Inference in biological physics

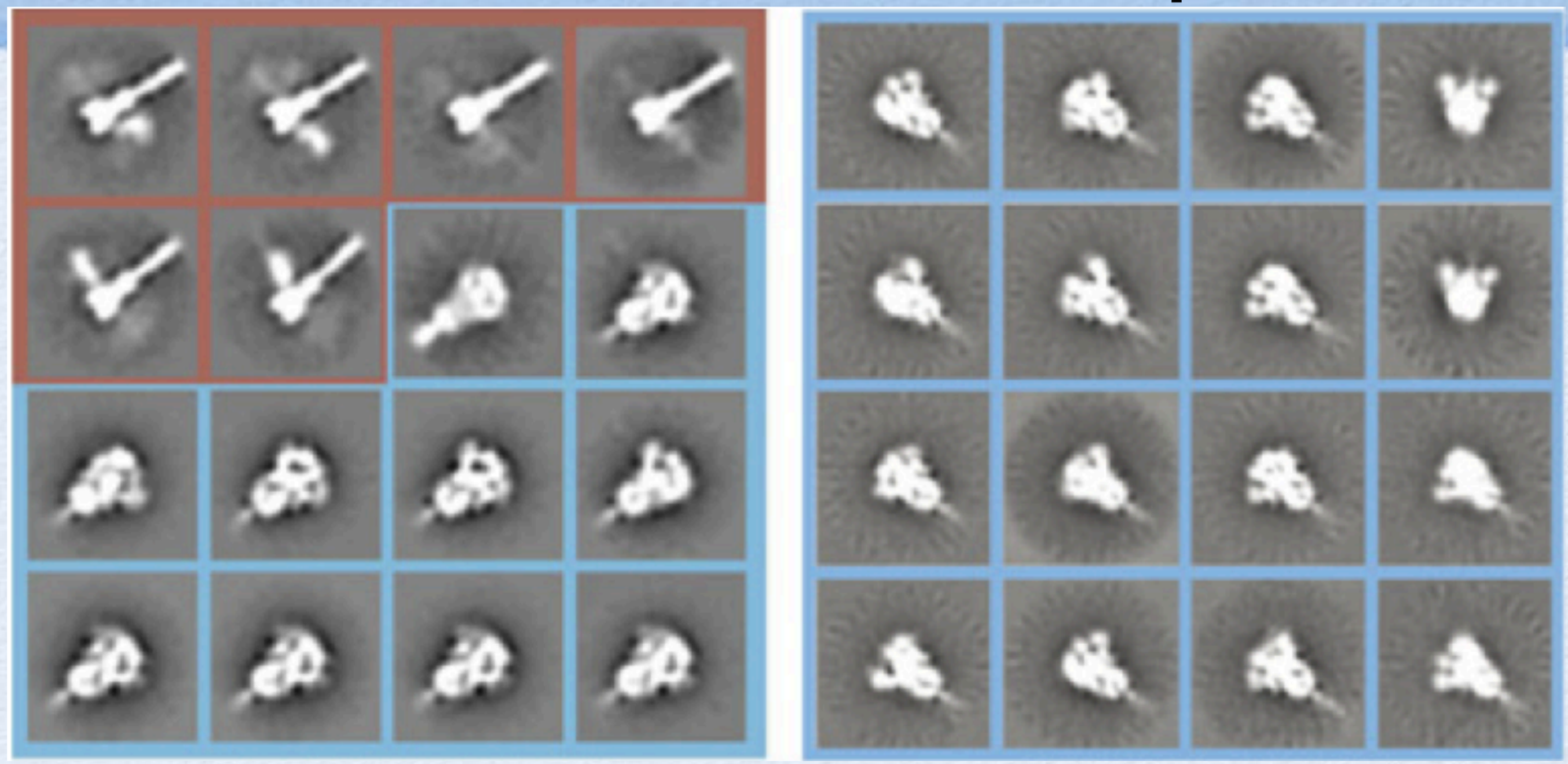
For these slides see:
www.physics.upenn.edu/~pcn

Image courtesy Mark Bates.

Is basic research important?

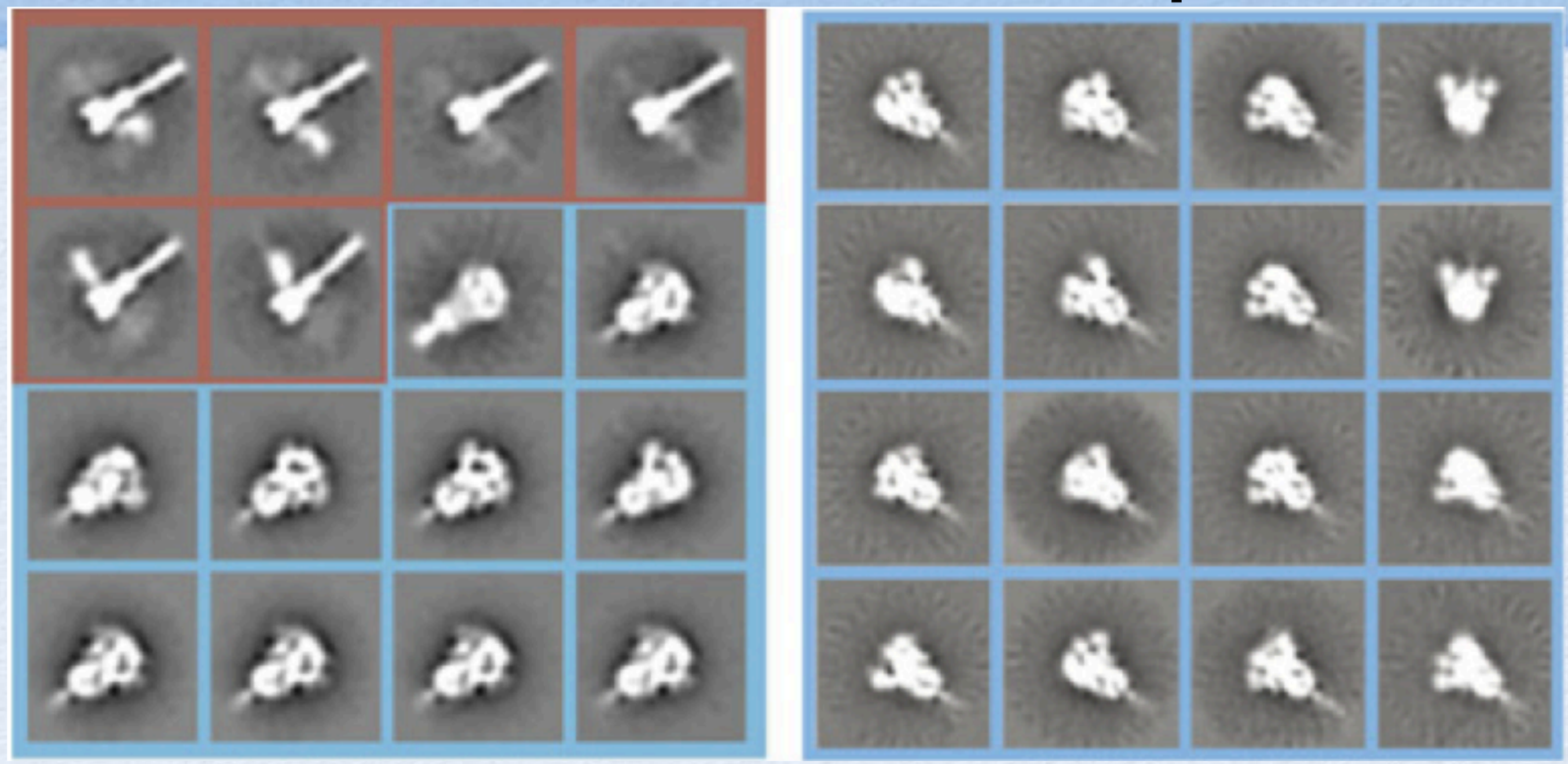


Is basic research important?



Spike protein conformations. Classes of images extracted from many copies of S from the severe acute respiratory syndrome coronavirus (SARS-CoV-1). *Left:* Natural form. Two quite different conformations are seen. *Right:* Corresponding images from a mutant designed to stabilize the pre-fusion conformation. Pallesen, J, et al. 2017. Proc. Natl. Acad. Sci. USA, **114.**

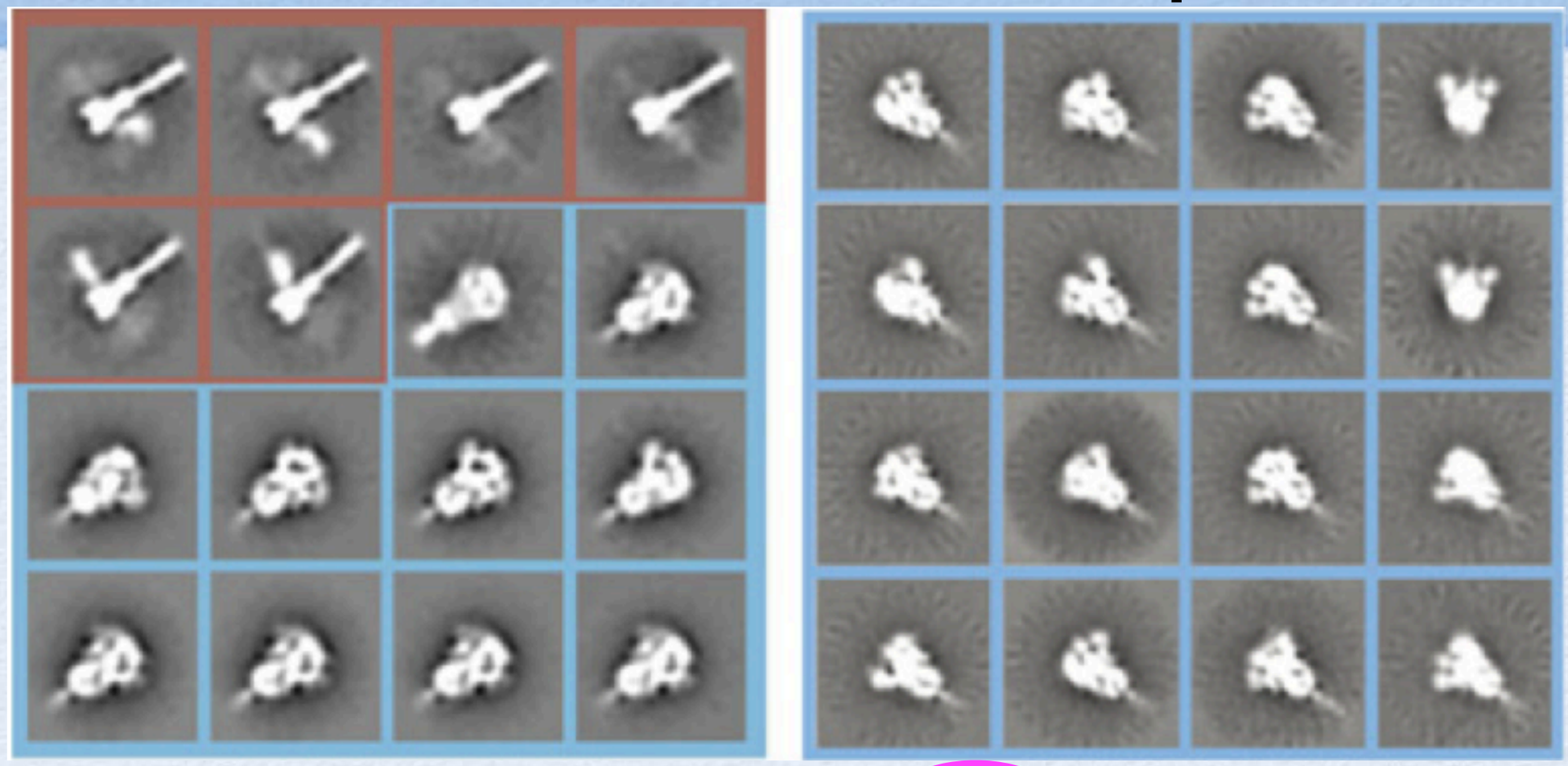
Is basic research important?



How did analogous images get made just a few weeks after SARS-CoV2 sequence was found? It takes forever to crystallize a new protein! And anyway, crystallography can't handle conformational heterogeneity – which is the whole point here.

Spike protein conformations. Classes of images extracted from many copies of S from the severe acute respiratory syndrome coronavirus (SARS-CoV-1). *Left:* Natural form. Two quite different conformations are seen. *Right:* Corresponding images from a mutant designed to stabilize the pre-fusion conformation. Pallesen, J, et al. 2017. Proc. Natl. Acad. Sci. USA, **114.**

Is basic research important?



How did analogous images get made just a few weeks after SARS-CoV2 sequence was found? It takes forever to crystallize a new protein! And anyway, crystallography can't handle conformational heterogeneity – which is the whole point here.

Spike protein conformations. Classes of images extracted from many copies of S from the severe acute respiratory syndrome coronavirus (SARS-CoV-1). *Left:* Natural form. Two quite different conformations are seen. *Right:* Corresponding images from a mutant designed to stabilize the pre-fusion conformation. Pallesen, J, et al. 2017. Proc. Natl. Acad. Sci. USA, **114.**

Part 1

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Part 1

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Conditional probability tells us what we can conclude from data,

Part 1

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Conditional probability tells us what we can conclude from data, *and*

Part 1

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Conditional probability tells us what we can conclude from data,
and
we live in a world with boatloads of data,

Part 1

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Conditional probability tells us what we can conclude from data,
and
we live in a world with boatloads of data,
but

Part 1

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Conditional probability tells us what we can conclude from data,
and
we live in a world with boatloads of data,
but
conditional probability is not hardwired into our intuition,

Part 1

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Conditional probability tells us what we can conclude from data,
and
we live in a world with boatloads of data,
but
conditional probability is not hardwired into our intuition,
so

Part 1

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

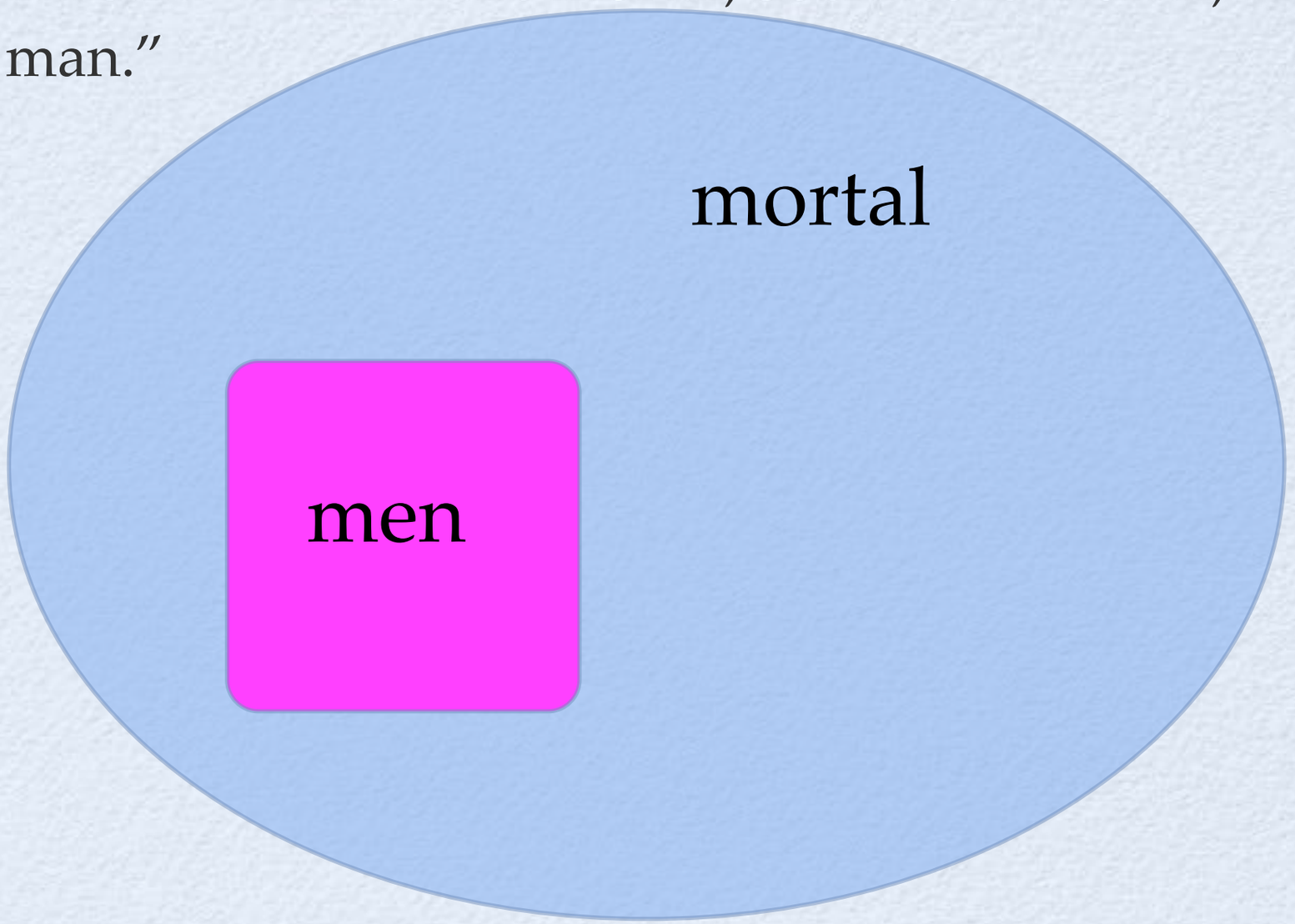
Conditional probability tells us what we can conclude from data,
and
we live in a world with boatloads of data,
but
conditional probability is not hardwired into our intuition,
so
we need to systematize it via the Bayes formula.

Inference

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

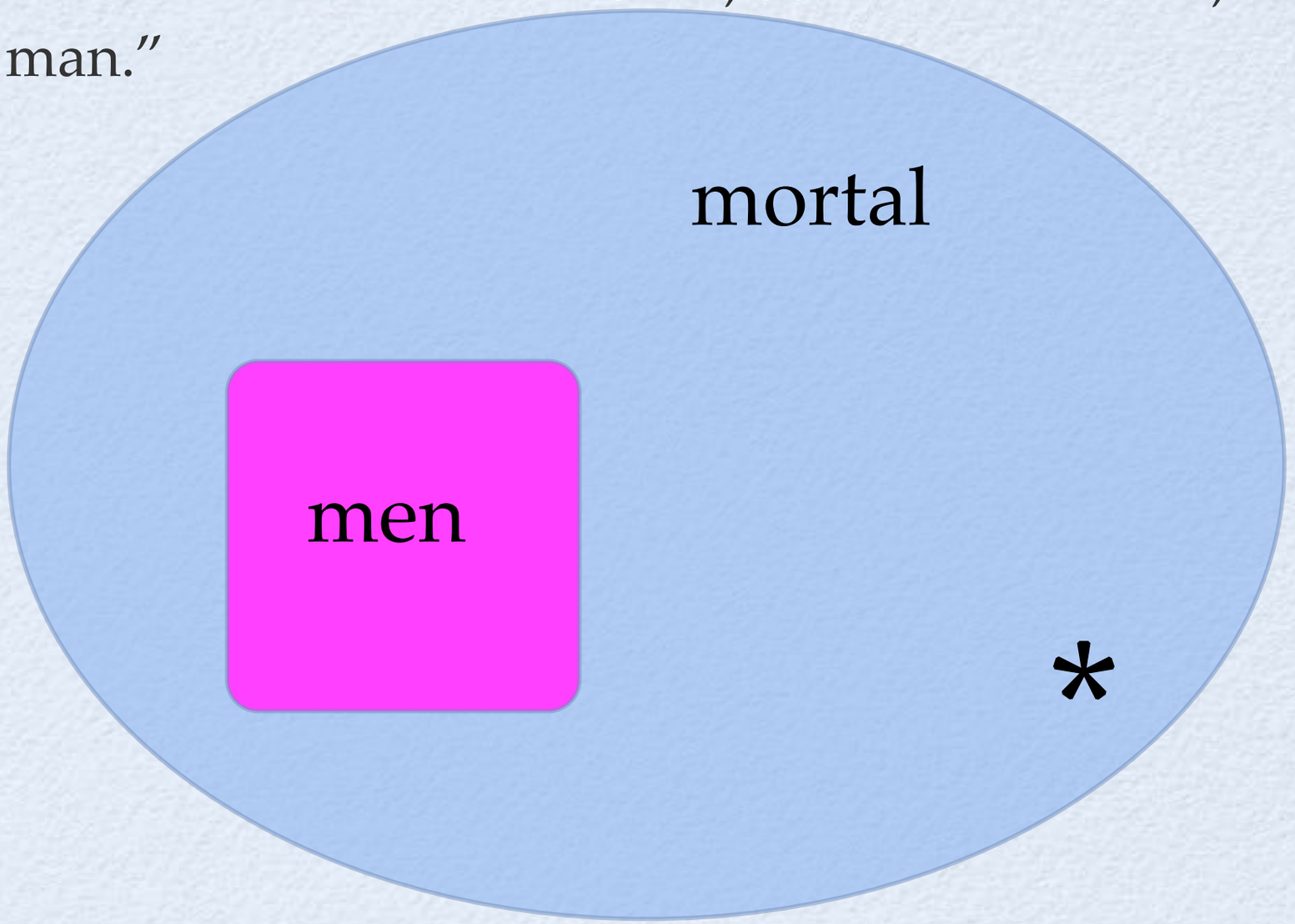
Inference

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



Inference

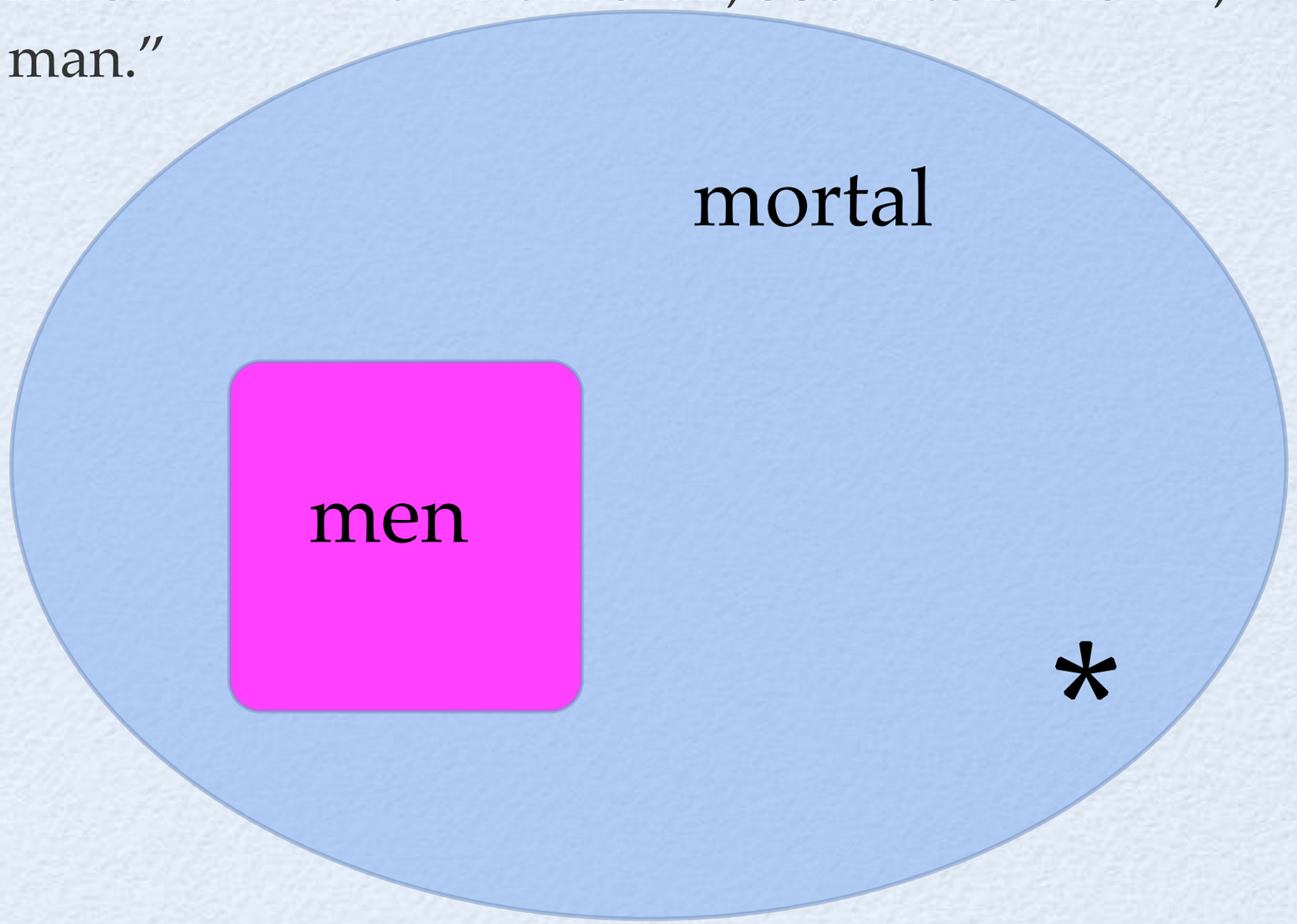
Suppose I stood here and said “all men are mortal; Socrates is mortal; therefore Socrates is a man.”



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

Inference

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.

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

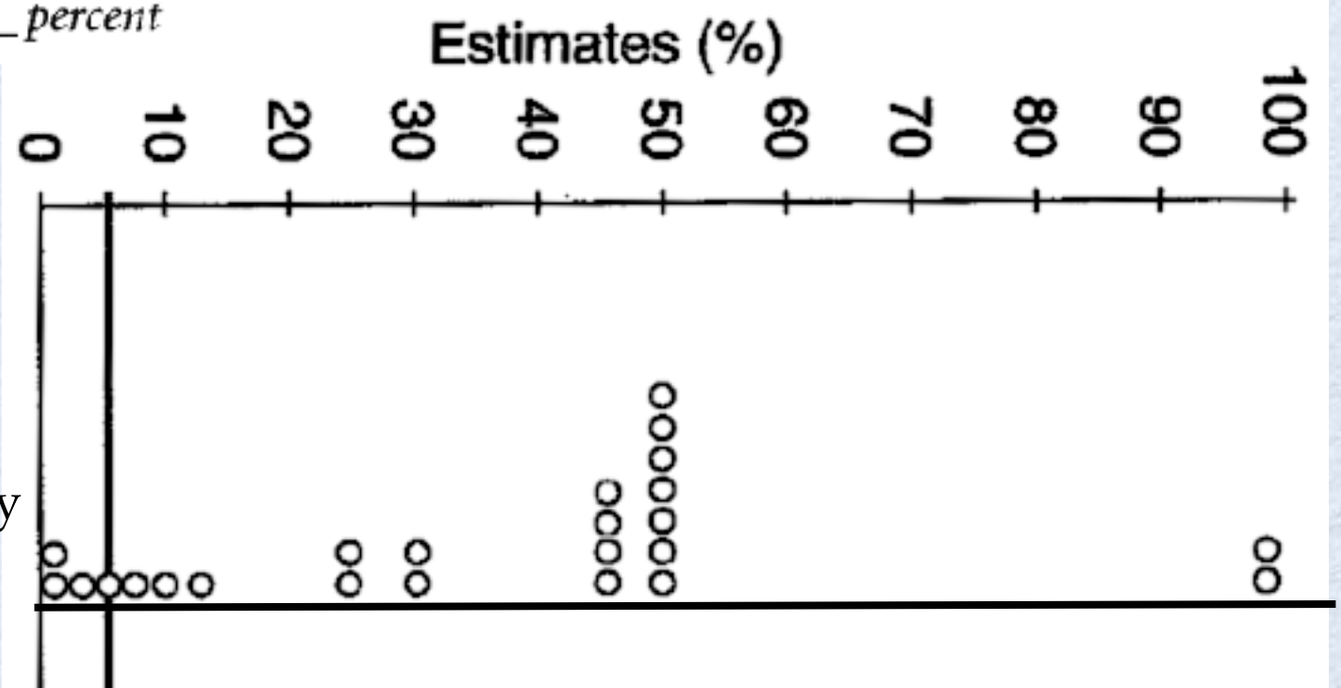
G. Gigerenzer, *Calculated risks*

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:



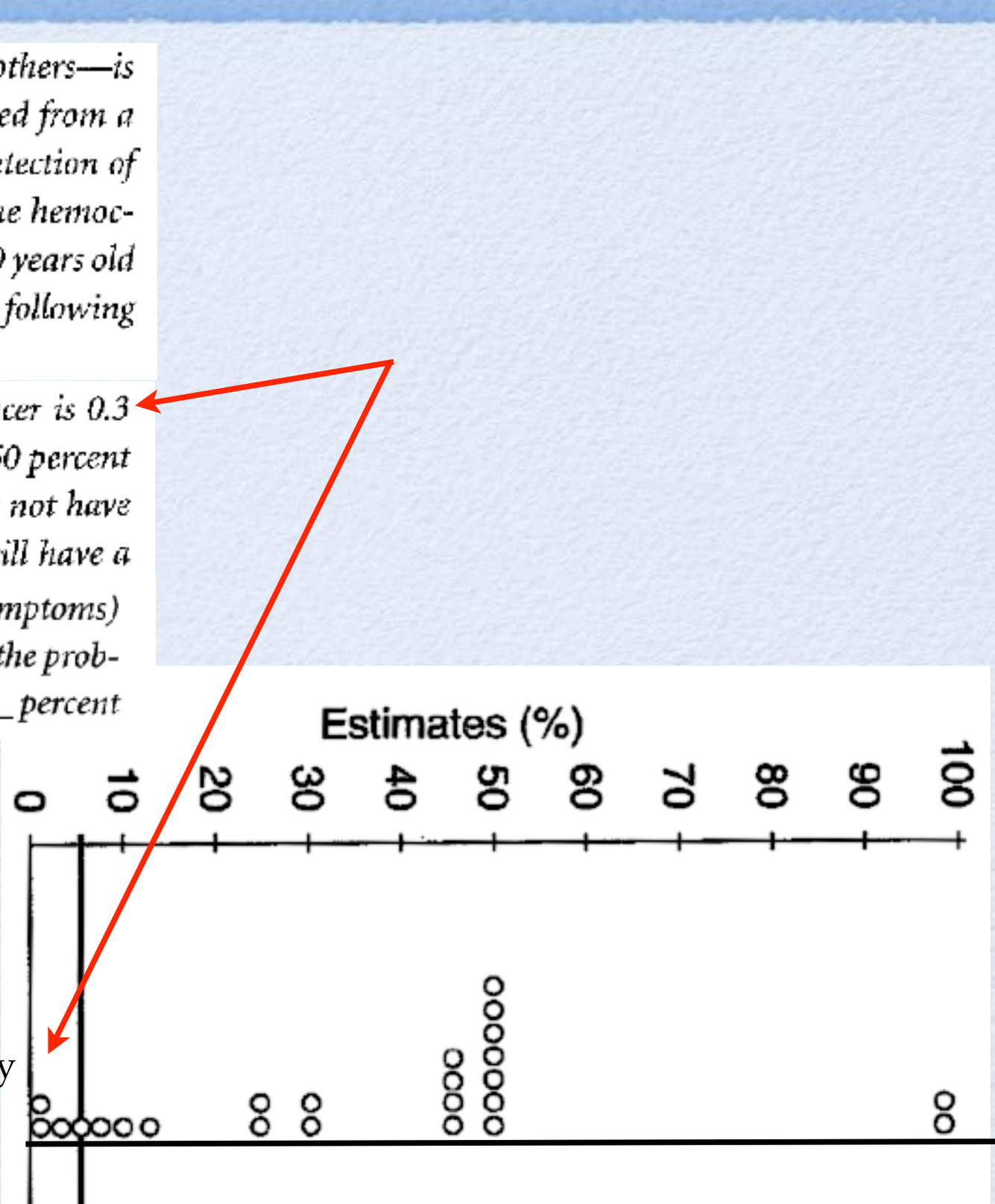
G. Gigerenzer, *Calculated risks*

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:



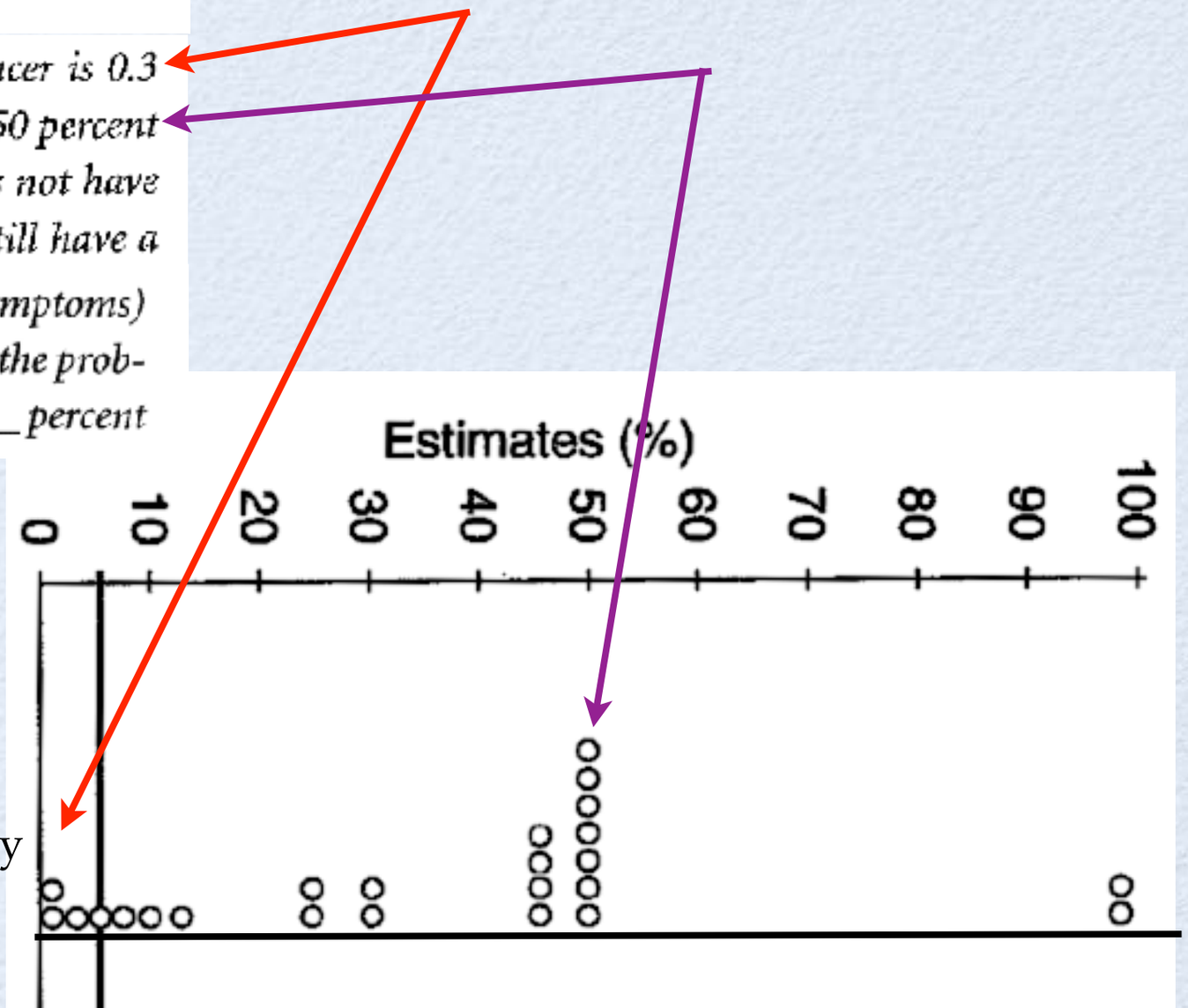
G. Gigerenzer, *Calculated risks*

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:



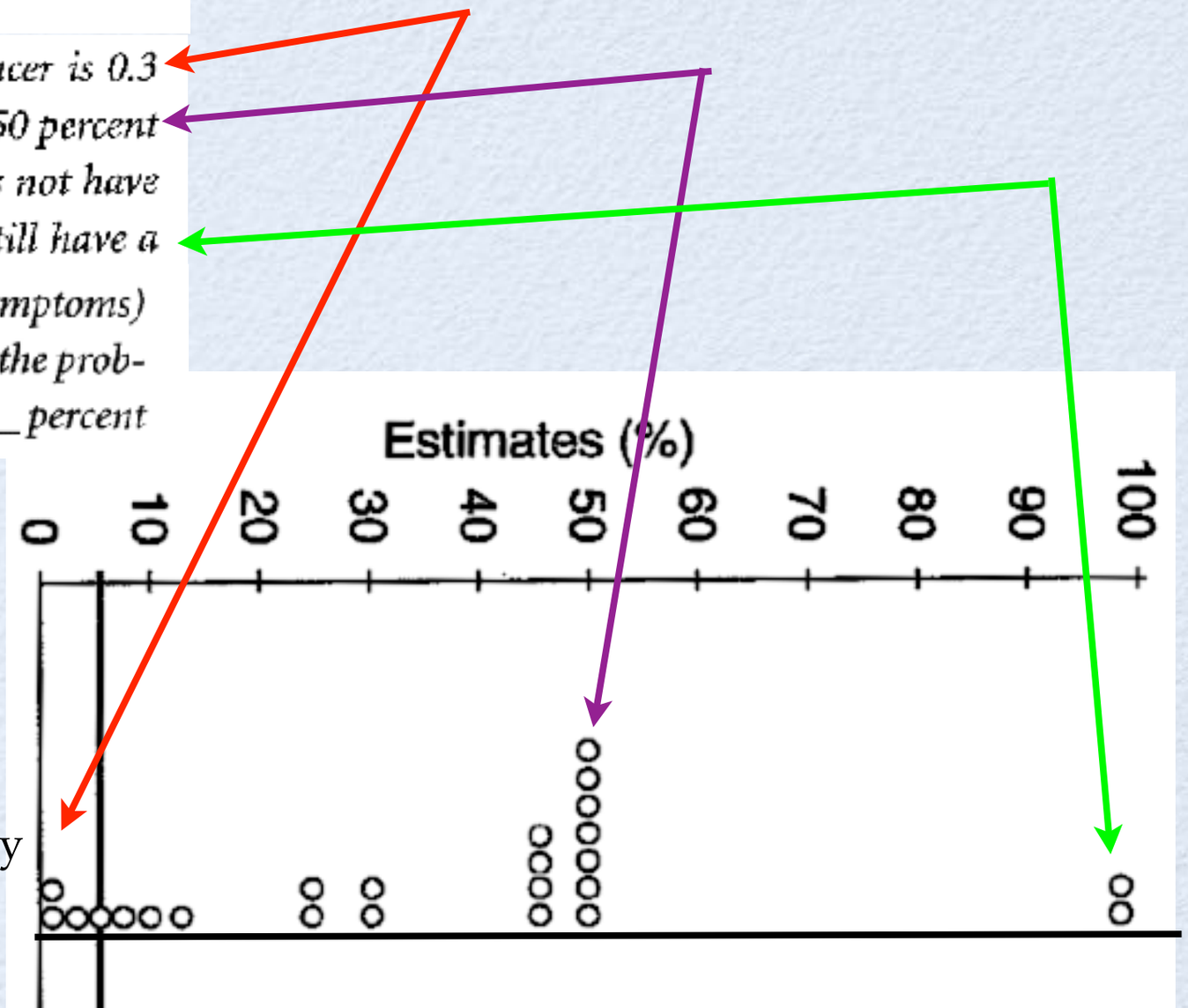
G. Gigerenzer, *Calculated risks*

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

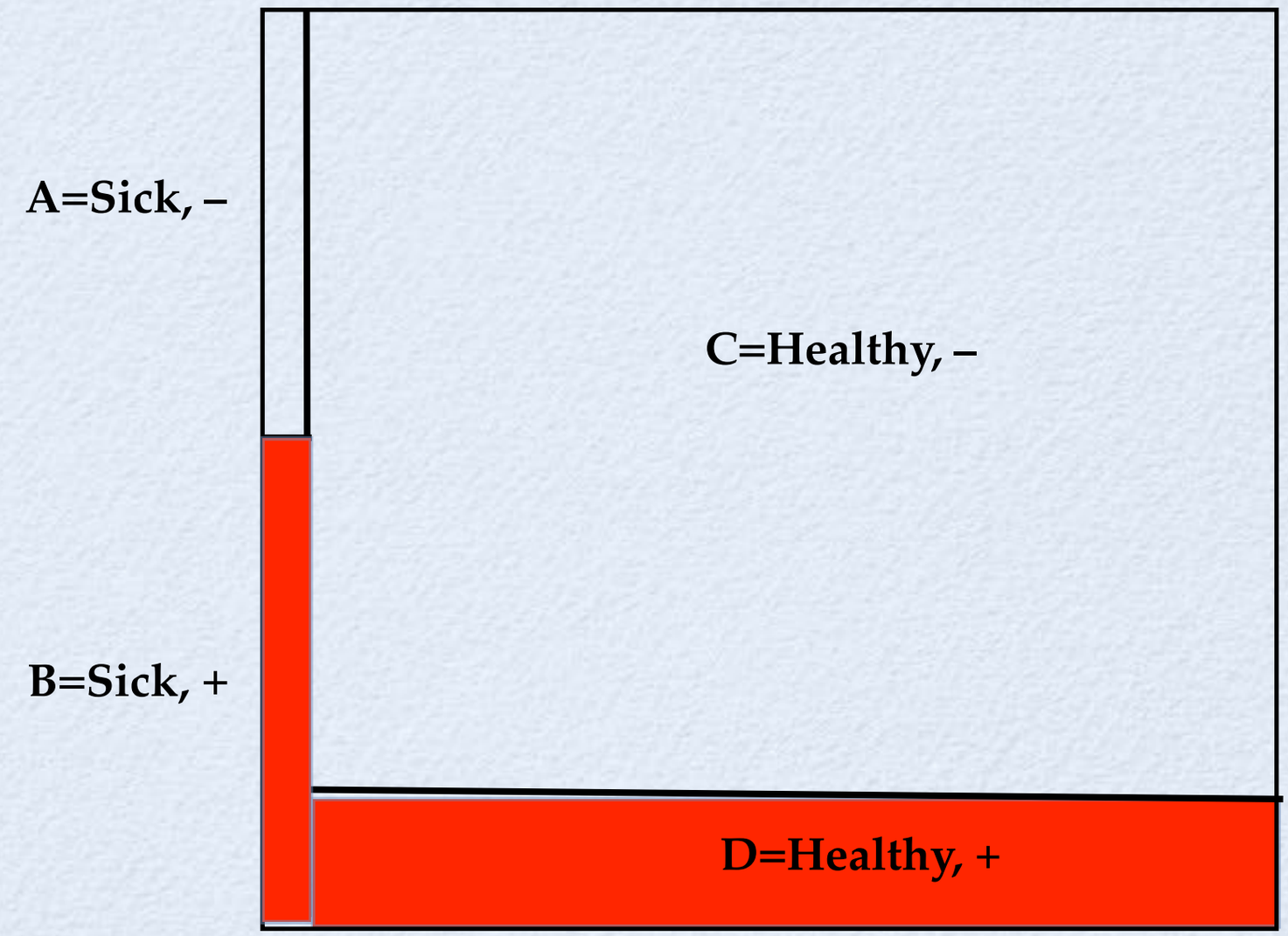
Here are the replies of 24 practicing physicians, who had an average of 14 years of professional experience:



G. Gigerenzer, *Calculated risks*

Work it out

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



Work it out

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

Work it out

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

Work it out

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

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

Work it out

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

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

Work it out

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

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

Work it out

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

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

Work it out

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

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

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

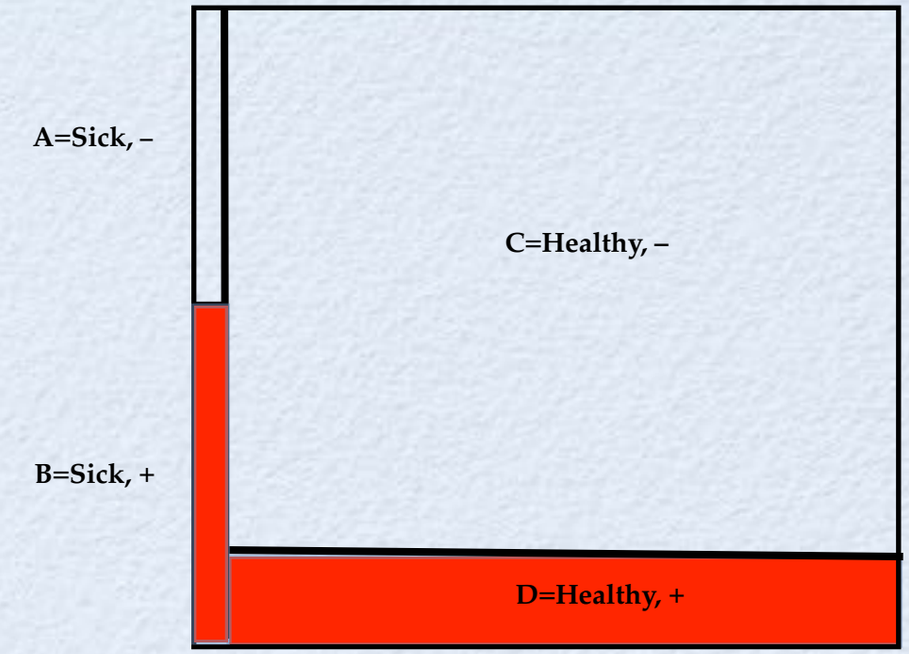
Work it out

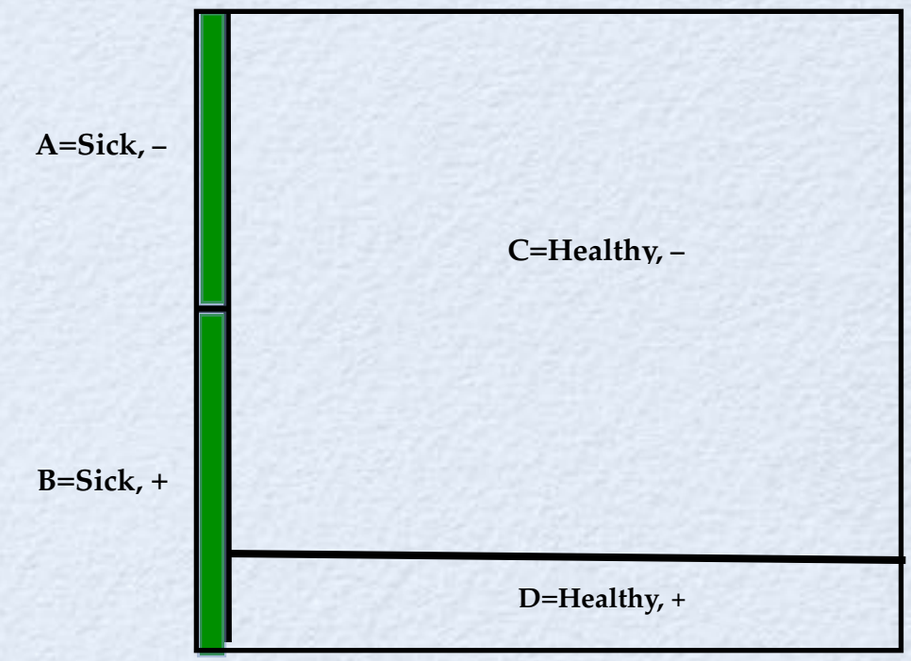
We are asked for $\mathcal{P}(\text{sick} \mid +) = B/(B+D)$.

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

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 $\mathcal{P}(\text{sick} \mid +) = B/(B+D)$.

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

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

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

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

Still need this

Posterior estimate (desired)

Likelihood (given, 50%)

Prior estimate (given, 0.3%)

In words

In words

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

In words

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

"The probability that *X* is true given the data"

In words

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

"The probability that X is true given the data"

is

"The probability that the data you *did* observe *would have been observed* in a world where X is true"

In words

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

"The probability that X is true given the data"

is

"The probability that the data you *did* observe *would have been observed* in a world where X is true"

times

"The prior probability of X"

In words

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

"The probability that X is true given the data"

is

"The probability that the data you *did* observe *would have been observed* in a world where X is true"

times

"The prior probability of X"

and

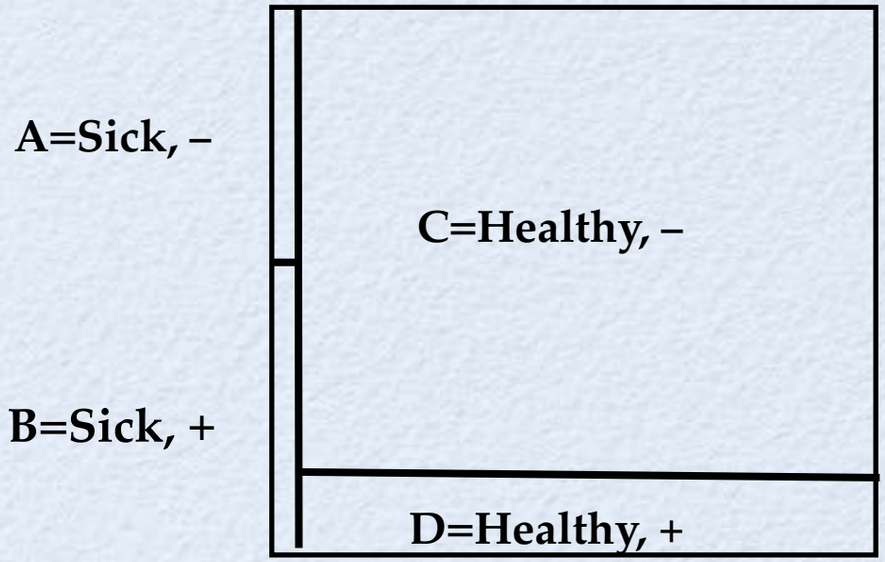
"A normalization factor."

Finish working it out

Bayes Formula:

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

Is that last factor really important?
P(sick) was given, but we also need:



Finish working it out

Bayes Formula:

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

Is that last factor really important?

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

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

Finish working it out

Bayes Formula:

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

Is that last factor really important?

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

$$\mathcal{P}(+) = B + D$$

$$= \frac{B}{A+B}(A+B) + \frac{D}{C+D}(C+D)$$

$$= \mathcal{P}(+|\text{sick})\mathcal{P}(\text{sick}) + \mathcal{P}(+|\text{healthy})\mathcal{P}(\text{healthy})$$

$$= (0.5)(0.003) + (0.03)(0.997) \approx 0.03$$

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

Yes, it's important: in this made-up example a positive test result means only a 5% chance you're sick. *Not 97%*.

Part 2

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Part 2

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

You can specifically label molecules of interest,

Part 2

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

You can specifically label molecules of interest,
and

Part 2

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

You can specifically label molecules of interest,
and
you can watch them going about their cellular business, in video,

Part 2

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

You can specifically label molecules of interest,
and
you can watch them going about their cellular business, in video,
but

Part 2

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

You can specifically label molecules of interest,
and
you can watch them going about their cellular business, in video,
but
everything is blurred out to 200nm by diffraction,

Part 2

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

You can specifically label molecules of interest,
and
you can watch them going about their cellular business, in video,
but
everything is blurred out to 200nm by diffraction,
so

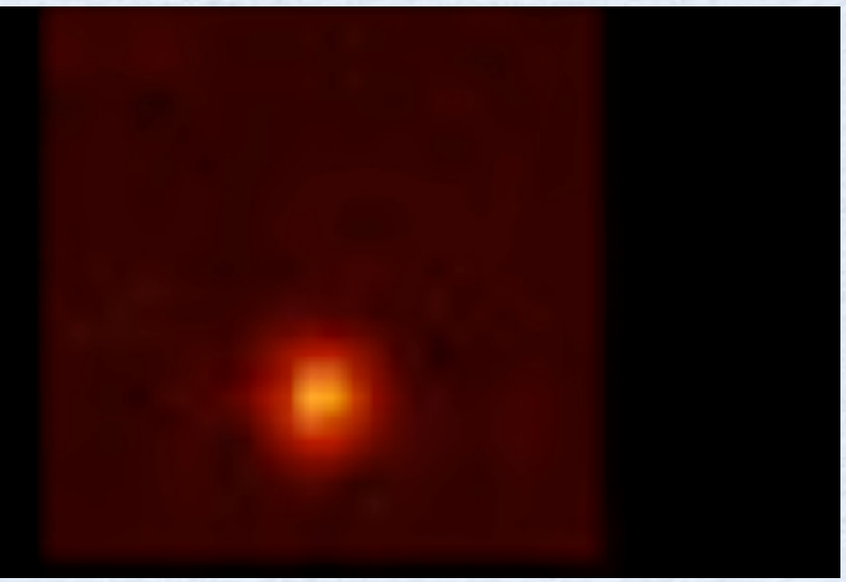
Part 2

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

You can specifically label molecules of interest,
and
you can watch them going about their cellular business, in video,
but
everything is blurred out to 200nm by diffraction,
so
how can you observe nanometer-scale motions and structures?

Superresolution microscopy

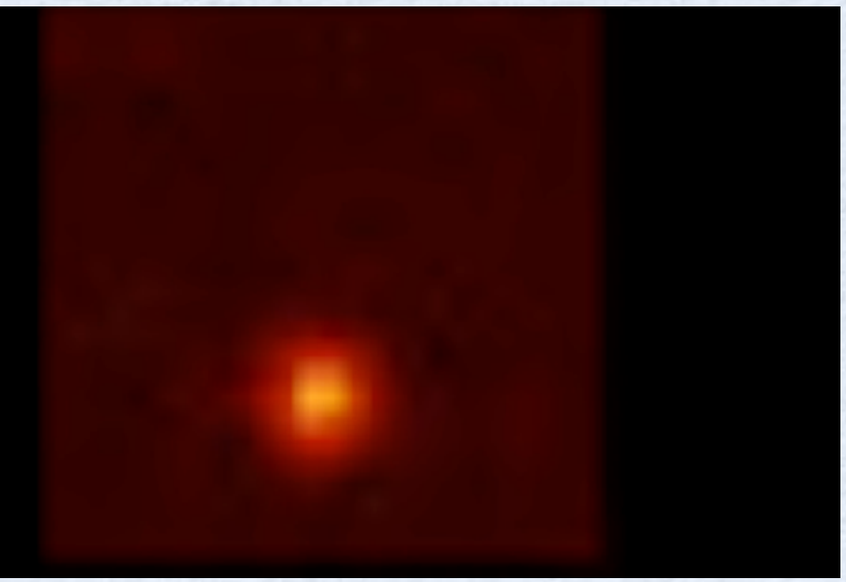
How does one measure myosin steps to within a few nm accuracy using visible light? The diffraction-limited spot is at least 200 nm wide!
The key point is to realize that although we cannot resolve *two* spots closer than this, sometimes all we want is to detect *motion* of *one* spot.



A. Yildiz, et al. Science 2003. **Precursors:** *M. K. Cheezum, W. F. Walker, W. H. Guilford, Biophys. J. 81, 2378 (2001). R. E. Thompson, D. R. Larson, W. W. Webb, Biophys. J. 82, 2775 (2002).*

Superresolution microscopy

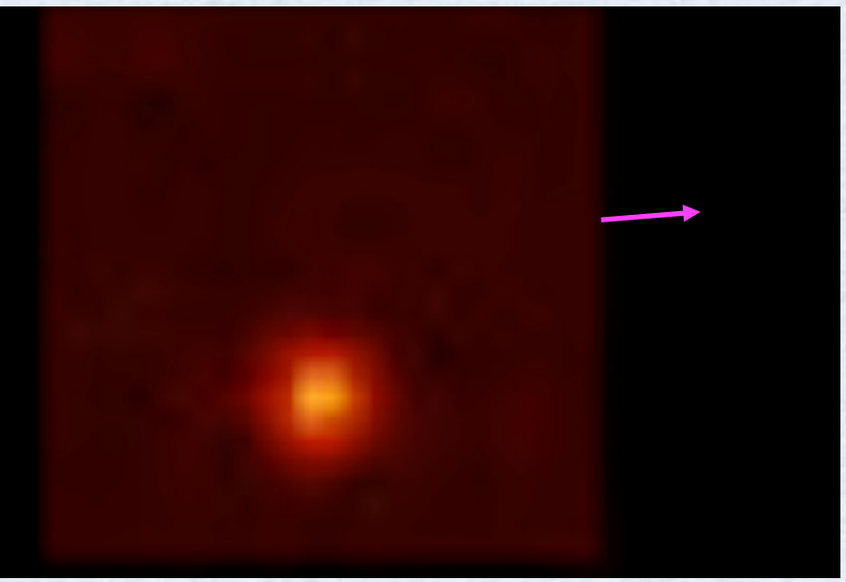
How does one measure myosin steps to within a few nm accuracy using visible light? The diffraction-limited spot is at least 200 nm wide!
The key point is to realize that although we cannot resolve *two* spots closer than this, sometimes all we want is to detect *motion* of *one* spot.

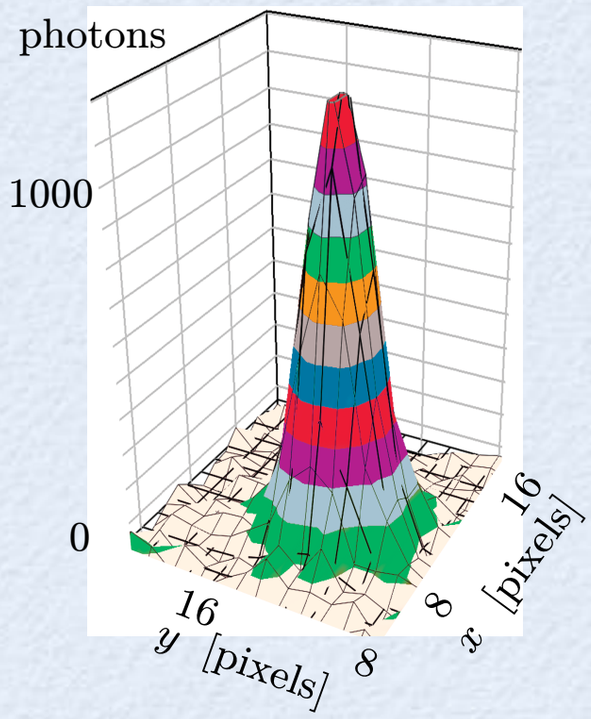


A. Yildiz, et al. Science 2003. **Precursors:** *M. K. Cheezum, W. F. Walker, W. H. Guilford, Biophys. J. 81, 2378 (2001). R. E. Thompson, D. R. Larson, W. W. Webb, Biophys. J. 82, 2775 (2002).*

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.

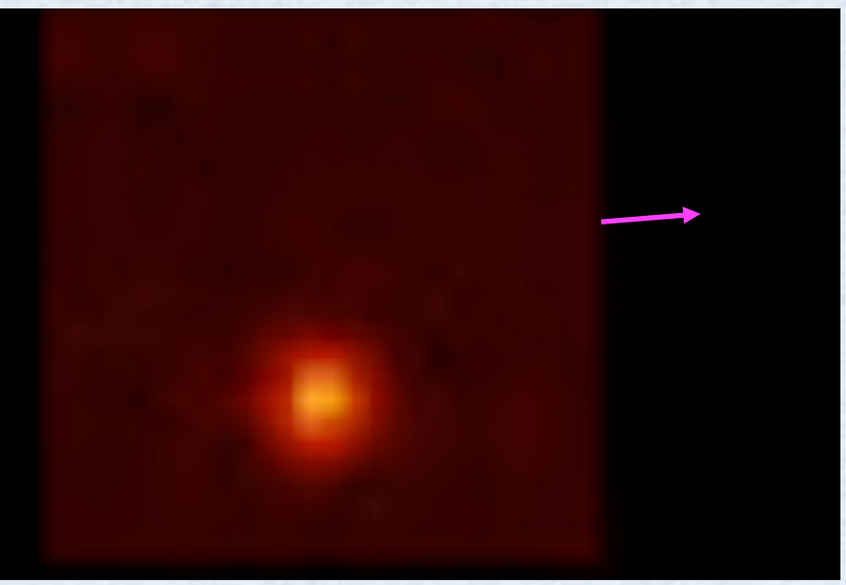


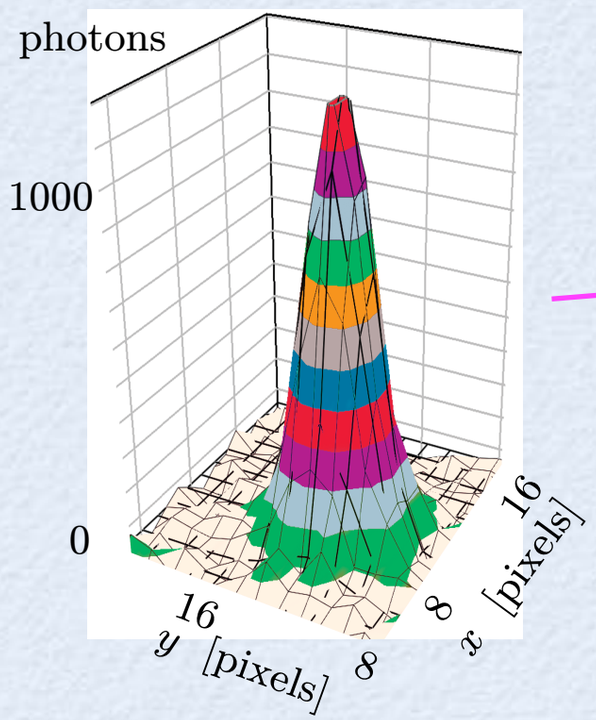


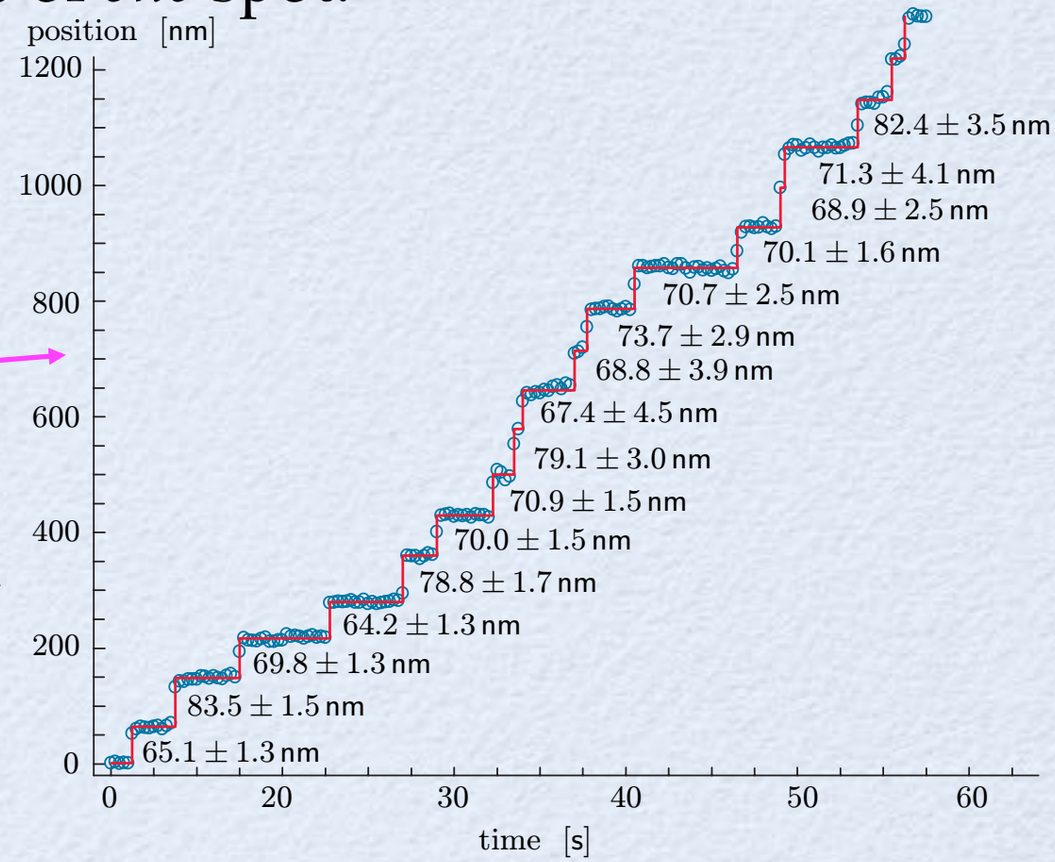
A. Yildiz, et al. Science 2003. **Precursors:** *M. K. Cheezum, W. F. Walker, W. H. Guilford, Biophys. J. 81, 2378 (2001). R. E. Thompson, D. R. Larson, W. W. Webb, Biophys. J. 82, 2775 (2002).*

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.



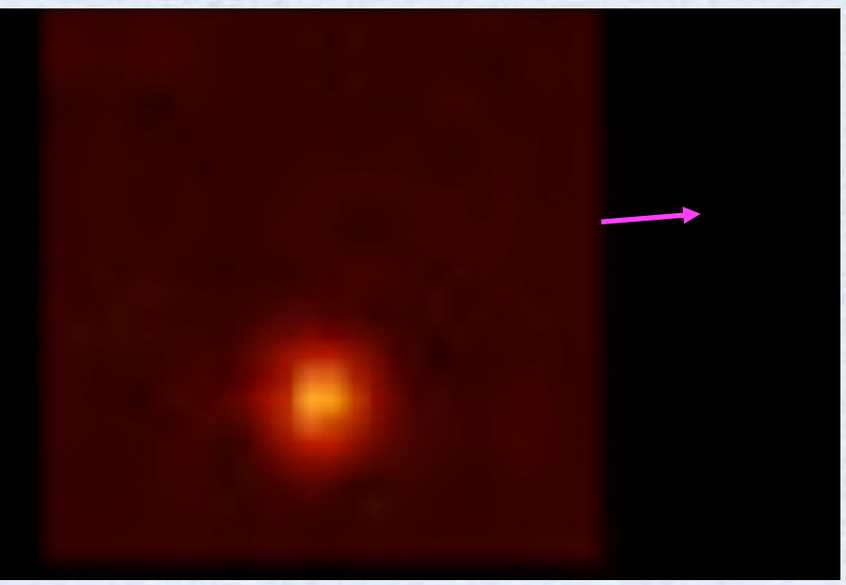


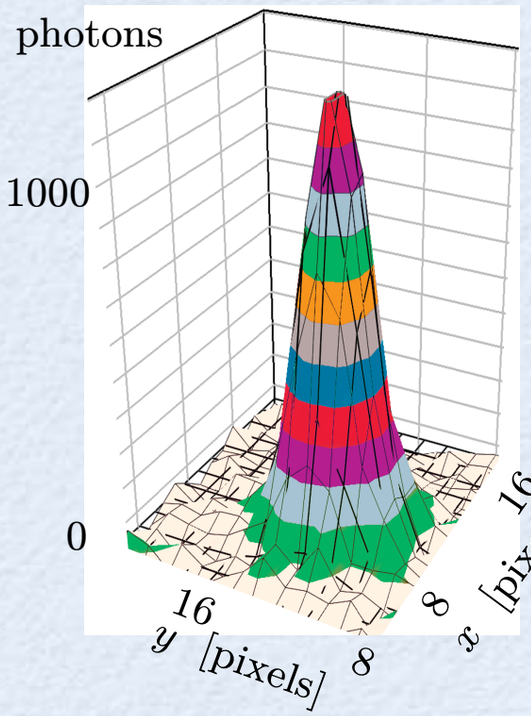


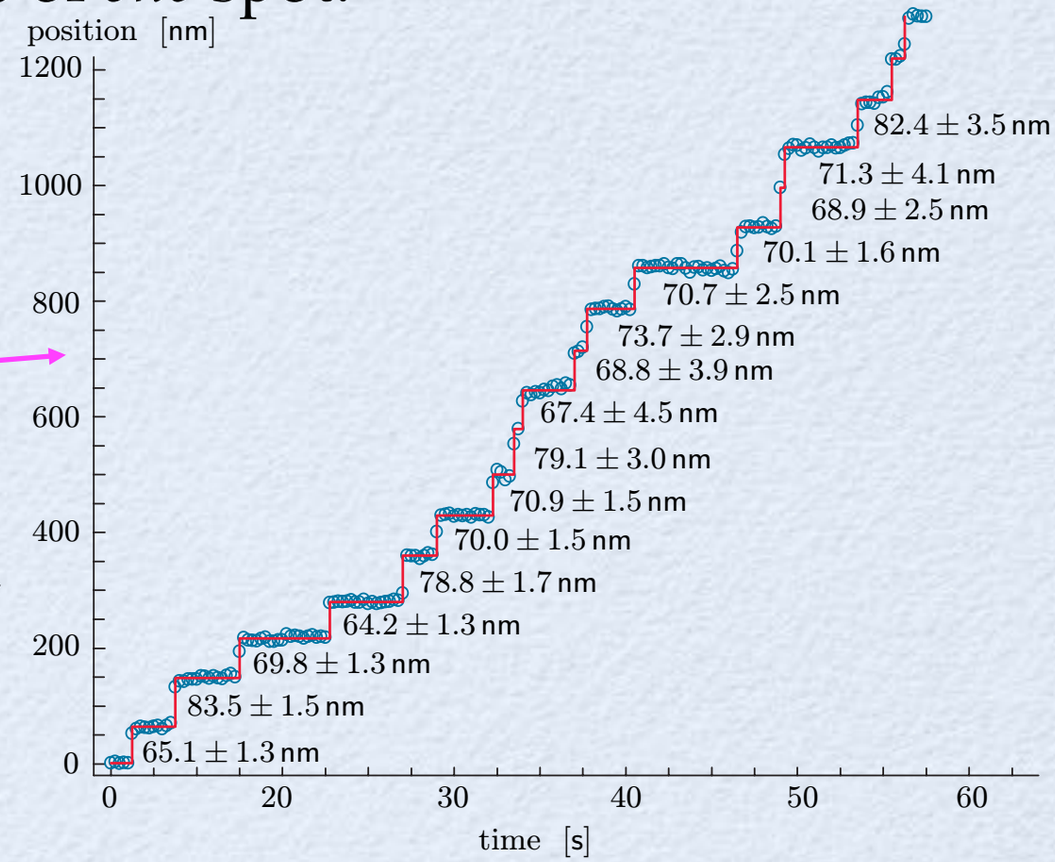
A. Yildiz, et al. Science 2003. **Precursors:** *M. K. Cheezum, W. F. Walker, W. H. Guilford, Biophys. J. 81, 2378 (2001). R. E. Thompson, D. R. Larson, W. W. Webb, Biophys. J. 82, 2775 (2002).*

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.

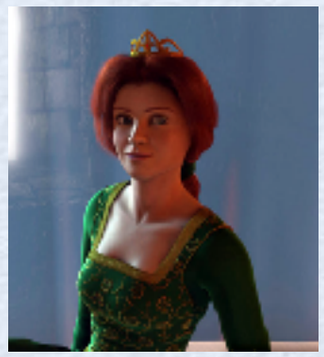






Fluorescence Imaging at One Nanometer Accuracy... but what principle does it rest on? Can it be improved?

F.I.O.N.A.



A. Yildiz, et al. Science 2003. **Precursors:** *M. K. Cheezum, W. F. Walker, W. H. Guilford, Biophys. J. 81, 2378 (2001). R. E. Thompson, D. R. Larson, W. W. Webb, Biophys. J. 82, 2775 (2002).*

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

The posterior probability is

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

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

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,
and the denominator)

likelihood is the product of
independent terms

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

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, and the denominator)

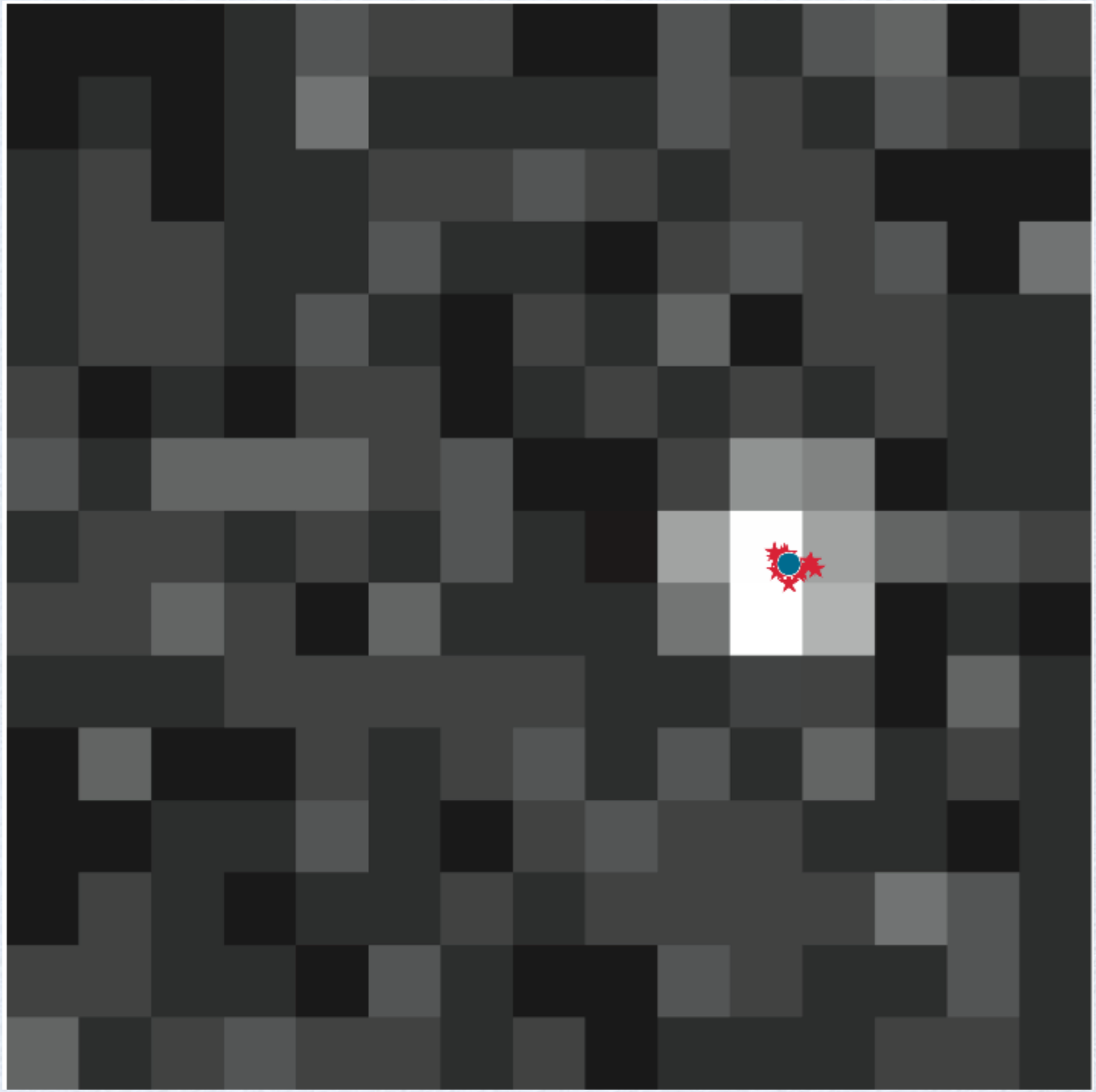
likelihood is the product of independent terms

Its log is simple:

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

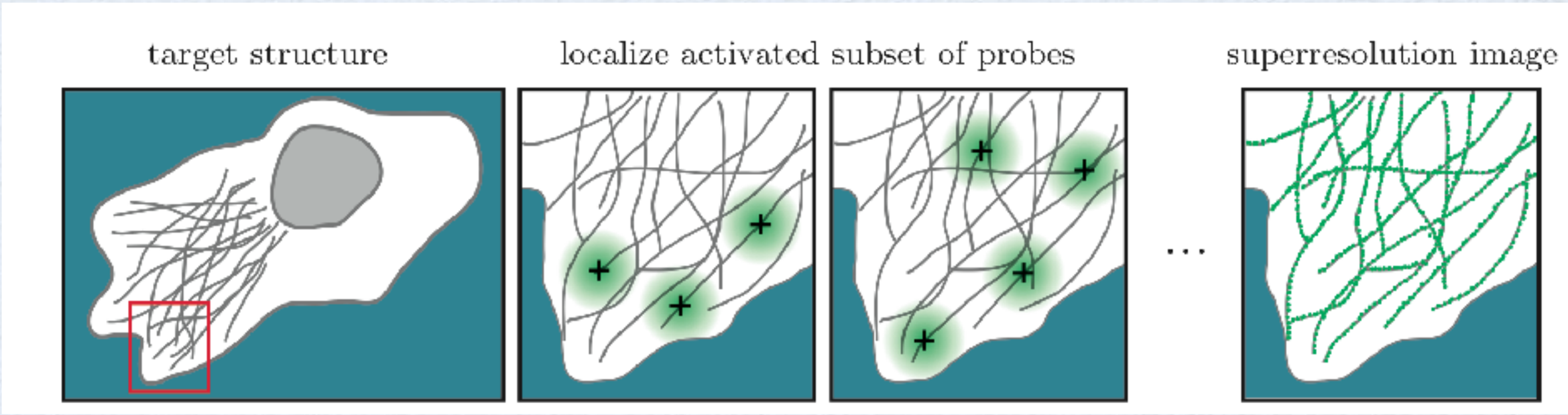
We wish to maximize this function over x_*, holding σ and all the data $\{x_1,\ldots,x_M\}$ fixed. The beauty of this approach is that it can be generalized to include a more accurate point-spread function, background, etc.

Same principle, with some extra realism: Even with real-world complications you can get not only sub-diffraction, but even *sub-pixel* resolution, by maximizing likelihood.

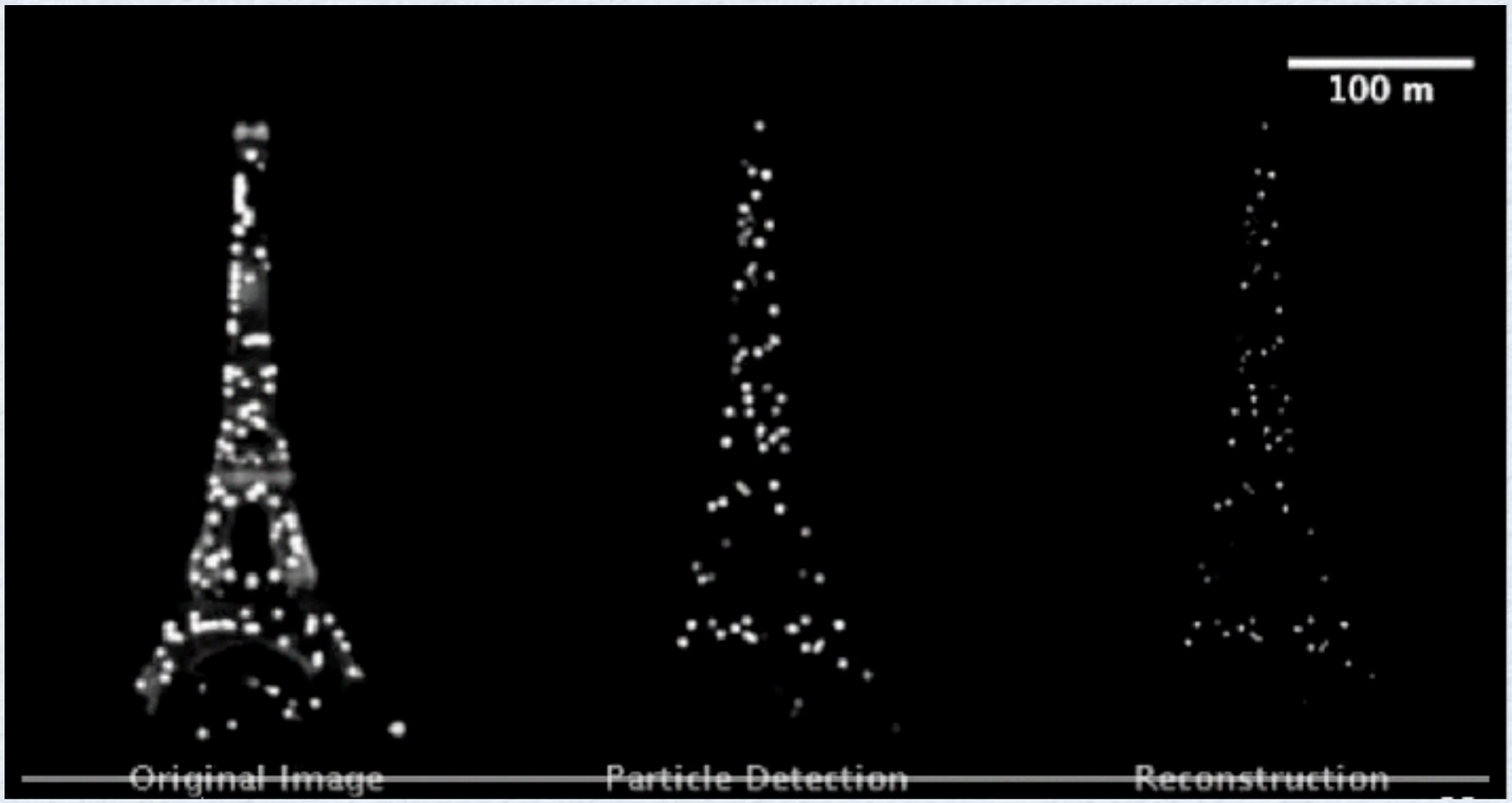


From P. Nelson, ***From Photon to Neuron: Light, Imaging, Vision*** (Princeton, 2017).

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

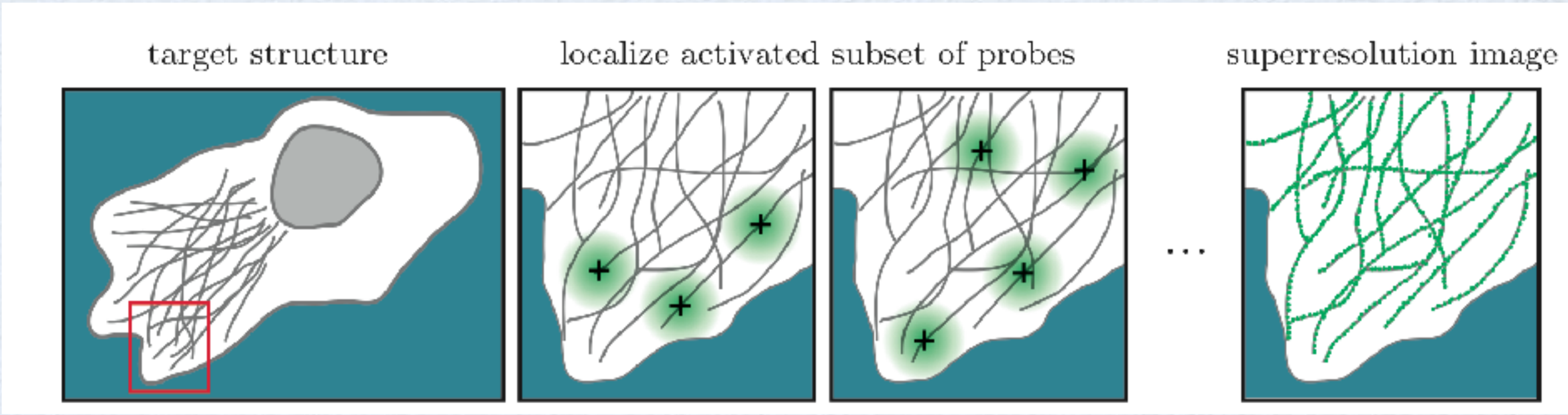


P. Nelson, ***Physical models of living systems*** (2/e, 2022)

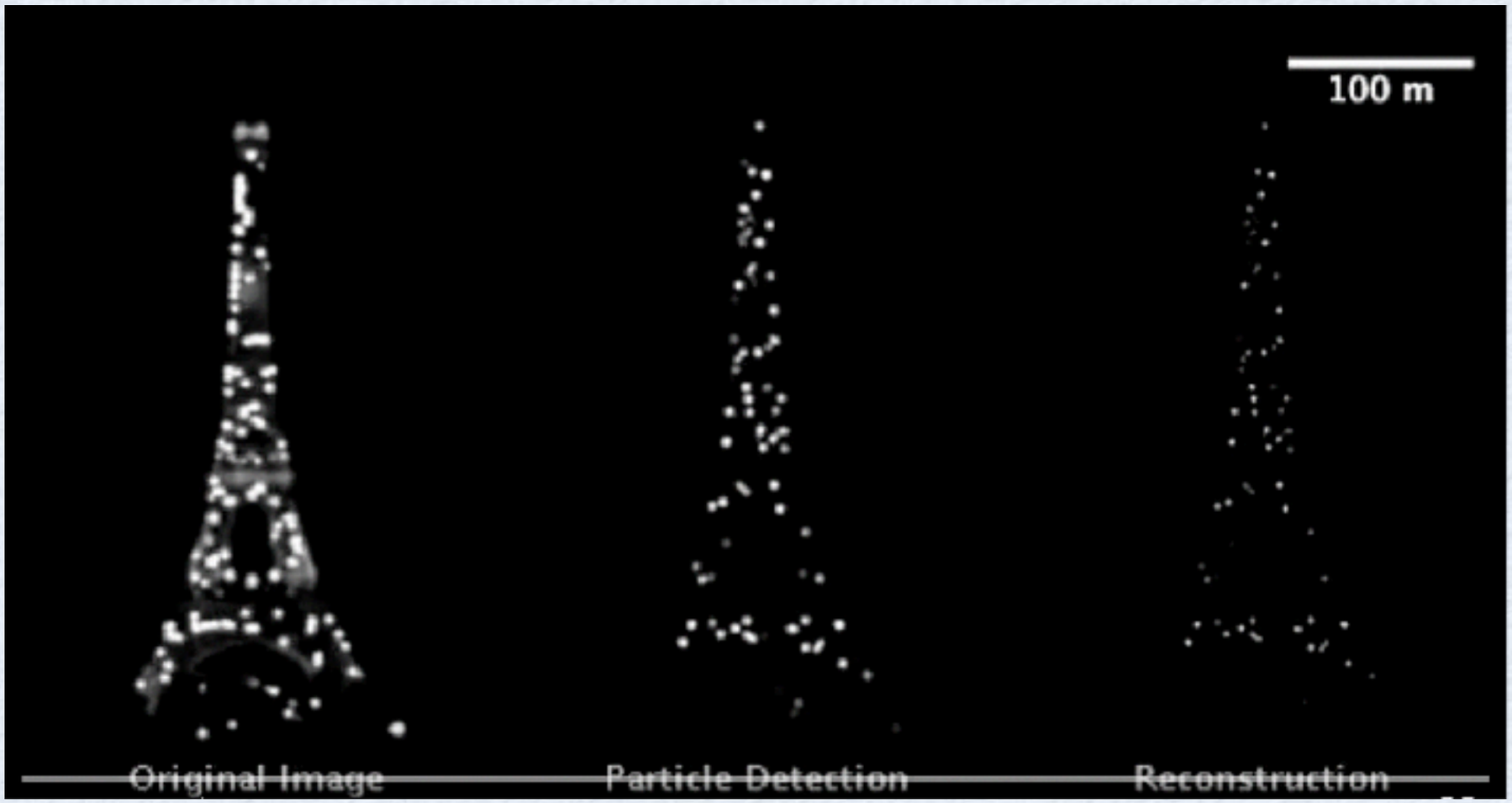


https://www.youtube.com/watch?v=RE70GuMCzww

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



P. Nelson, ***Physical models of living systems*** (2/e, 2022)



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

Part 3

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Part 3

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We'd like to know the spatial orientation of a molecule in real time,

Part 3

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We'd like to know the spatial orientation of a molecule in real time, *and*

Part 3

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We'd like to know the spatial orientation of a molecule in real time,
and
polarized TIRF microscopy can deliver that information,

Part 3

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We'd like to know the spatial orientation of a molecule in real time,
and
polarized TIRF microscopy can deliver that information,
but

Part 3

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We'd like to know the spatial orientation of a molecule in real time,
and
polarized TIRF microscopy can deliver that information,
but
a cruel tradeoff must be made between orientation accuracy and time resolution,

Part 3

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We'd like to know the spatial orientation of a molecule in real time,
and
polarized TIRF microscopy can deliver that information,
but
a cruel tradeoff must be made between orientation accuracy and time resolution,
so

Part 3

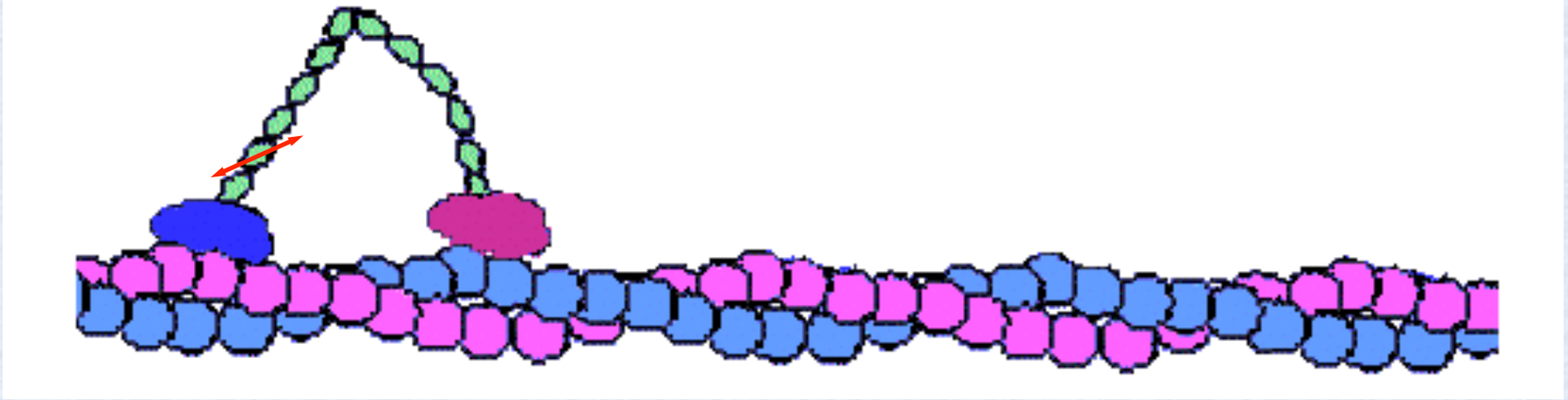
1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We'd like to know the spatial orientation of a molecule in real time,
and
polarized TIRF microscopy can deliver that information,
but
a cruel tradeoff must be made between orientation accuracy and time resolution,
so
we need to find the changepoints in order to optimize that tradeoff.

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.

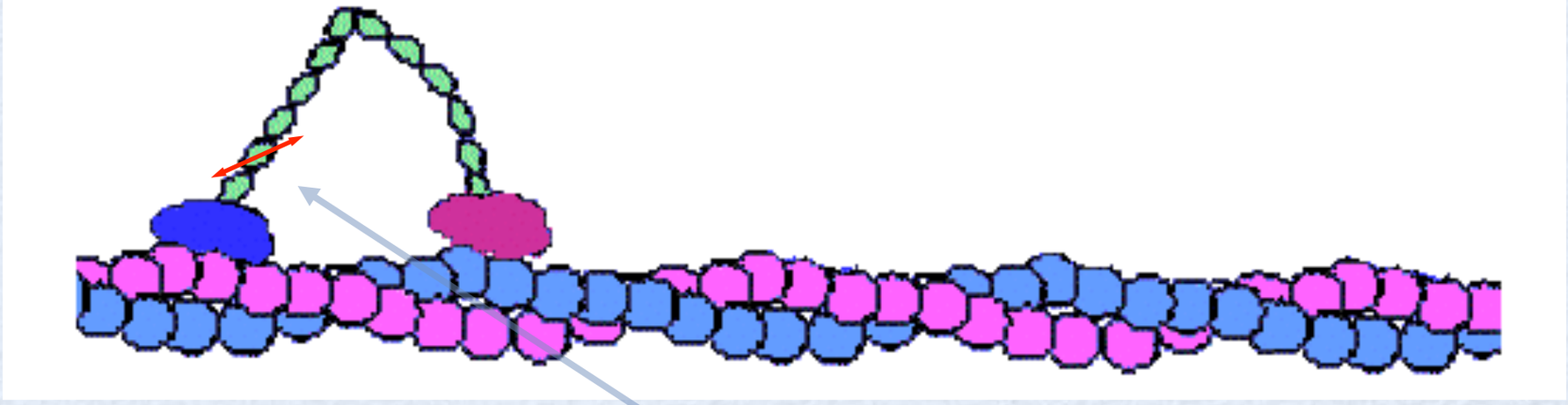


JF Beausang, Yale Goldman, PN

Myosin V stepping

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

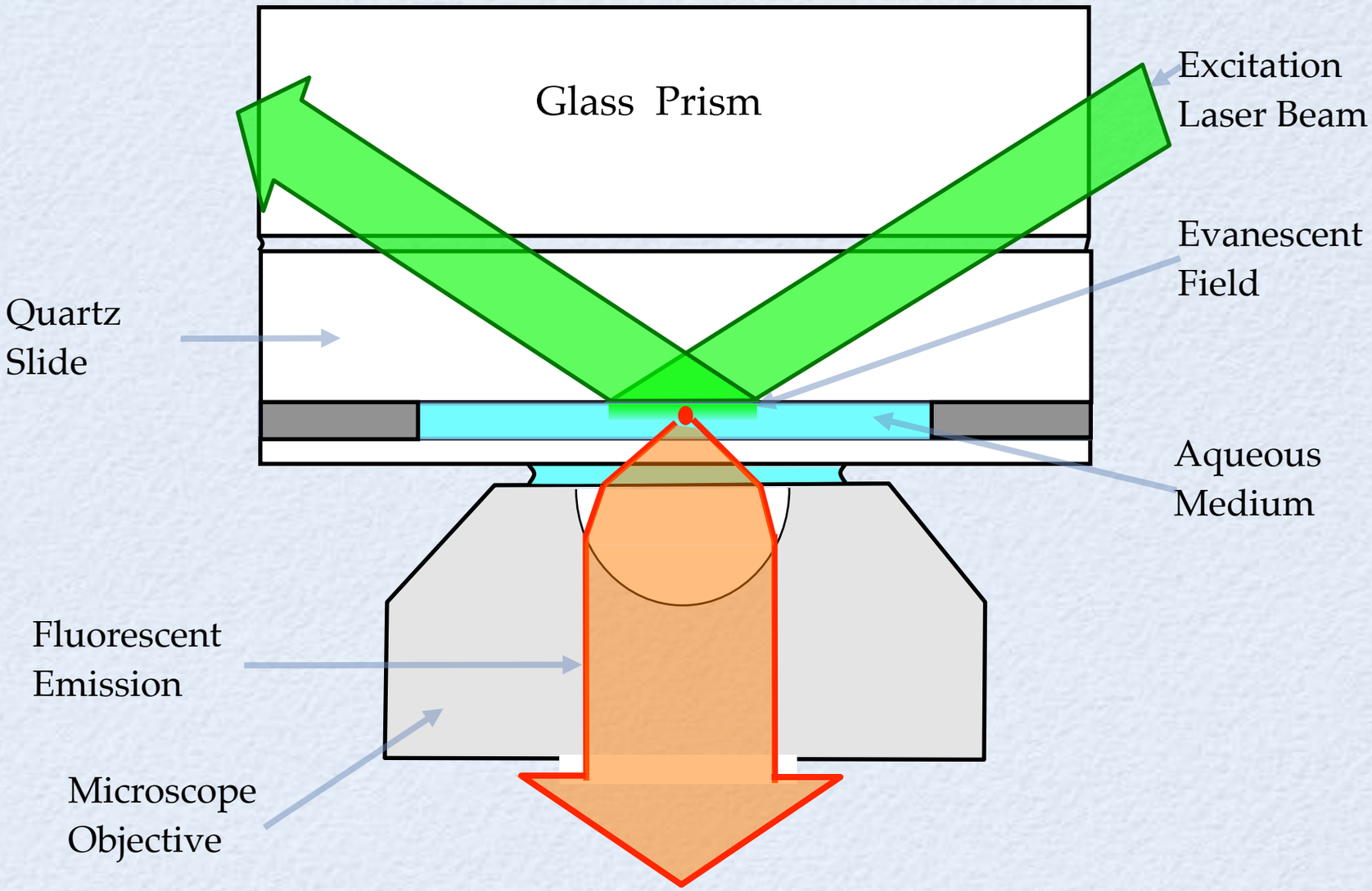
We'd like to know things like: How does it walk? What are the steps in the kinetic pathway? What is the geometry of each state?
One classic approach is to monitor the position in space of a marker (e.g. a bead) attached to the motor. But this does not address the geometry of each state.



JF Beausang, Yale Goldman, PN

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.

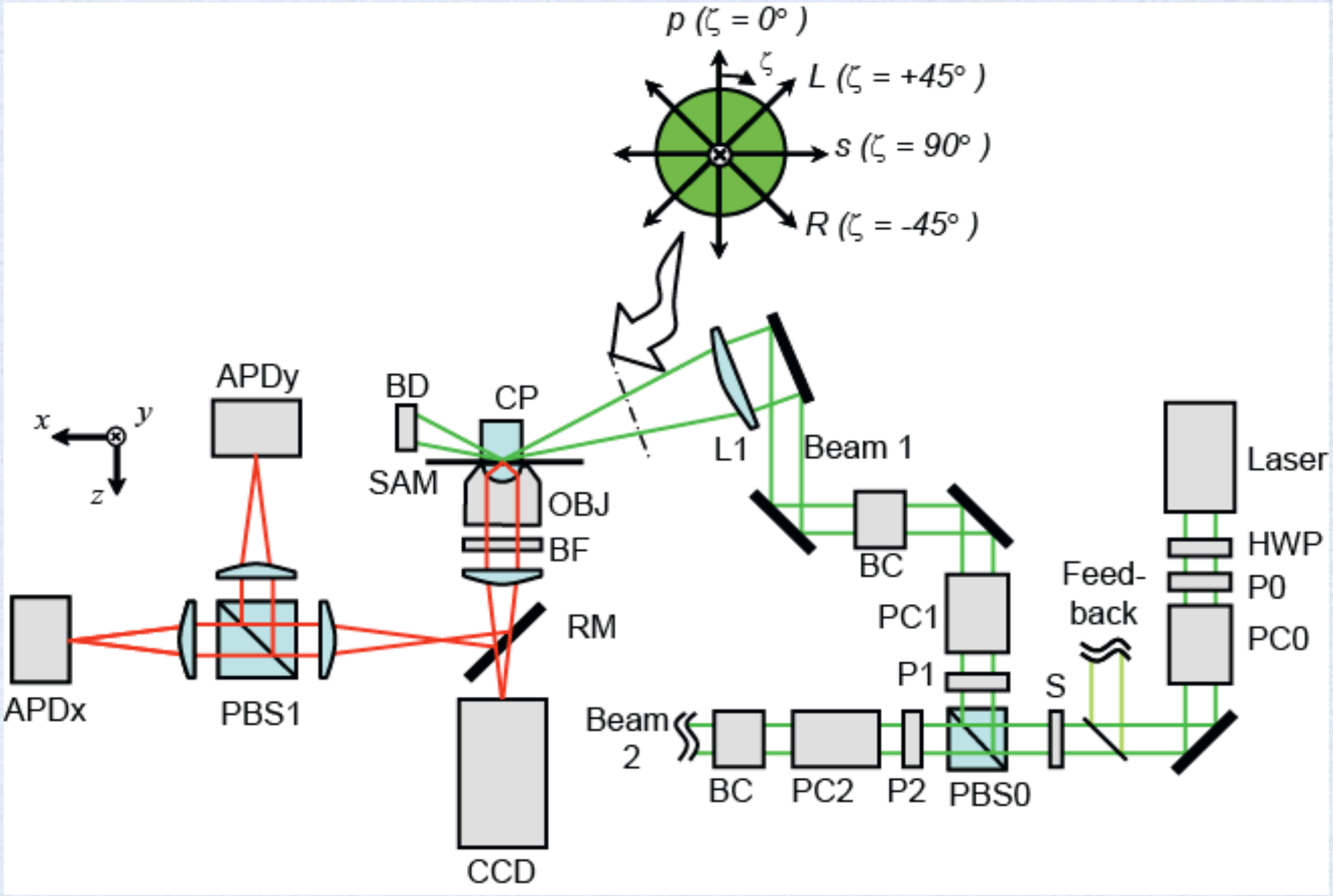
Polarized total internal reflection fluorescence microscopy



Fluorescence illumination by the evanescent wave eliminates a lot of noise, and importantly, maintains the polarization of the incident light.

To read out the orientation, we send in polarized light and see how many fluorescence photons, in each polarization, emerge. In this experiment a total of 8 different incoming polarizations was used.

pol-TIRF setup



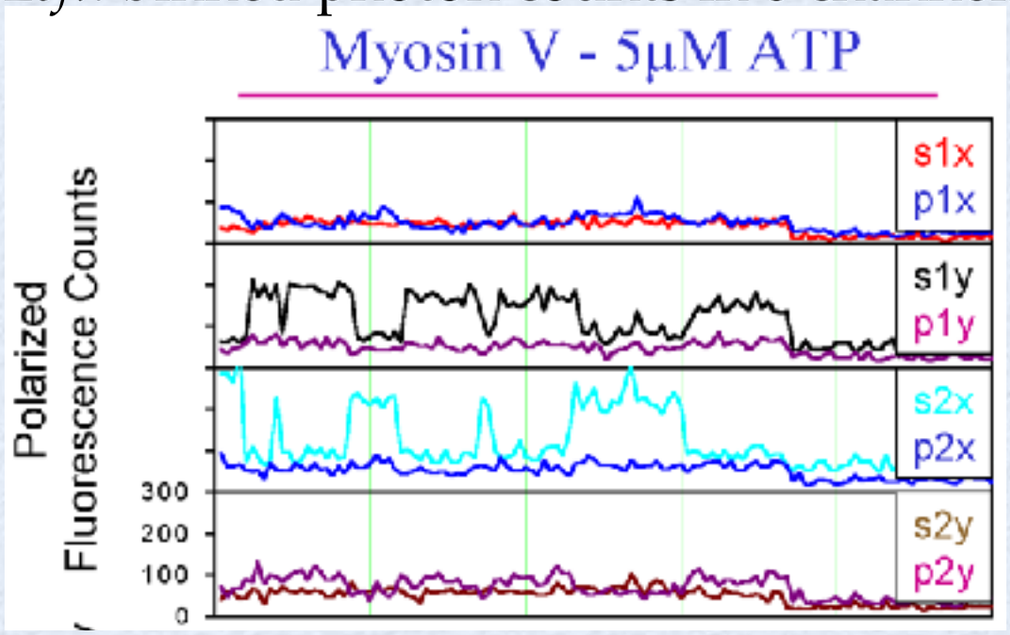
For each of the 8 incoming beams, outgoing photons were analyzed into 2 polarizations, for a total fo 16 channels.

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.*** Once we've got that, then we can use quantum mechanics to determine the orientation of the molecule in space.

Unfortunately, existing analyses gave noisy rate determinations. That in turn led to poor determinations of orientation – garbage in/garbage out.

Left: binned photon counts in 8 channels.



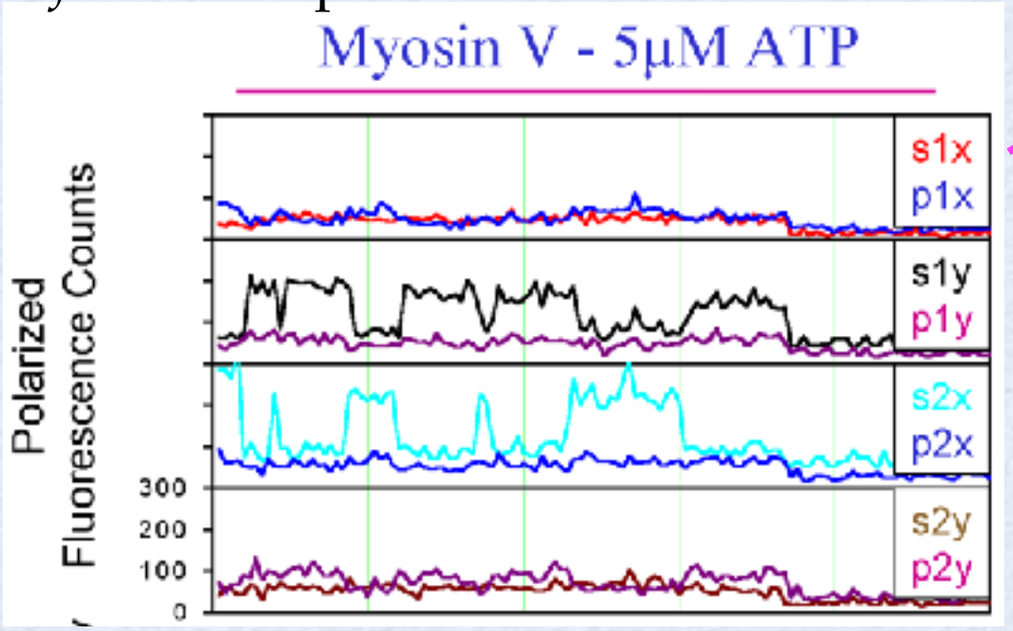
JN Forkey et al. Nature 2003

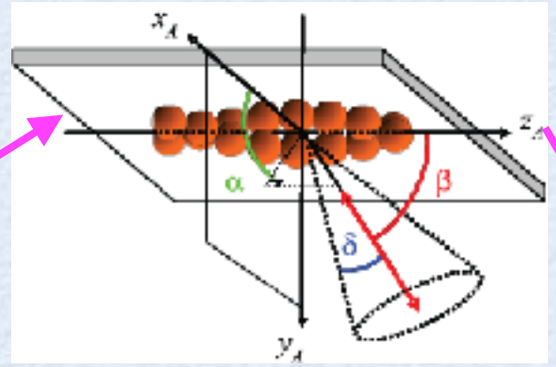
Previous state of the art

For our purposes, the upshot is that: ***We need to know the arrival rates of photons in each of several channels.*** Once we've got that, then we can use quantum mechanics to determine the orientation of the molecule in space.

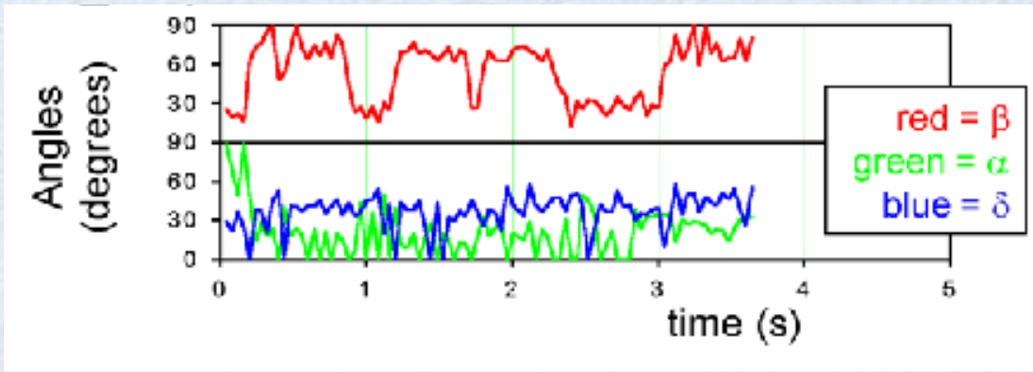
Unfortunately, existing analyses gave noisy rate determinations. That in turn led to poor determinations of orientation – garbage in/garbage out.

Left: binned photon counts in 8 channels.





Right: Polar and azimuthal angles of the fluorescent label, inferred from data on the left.



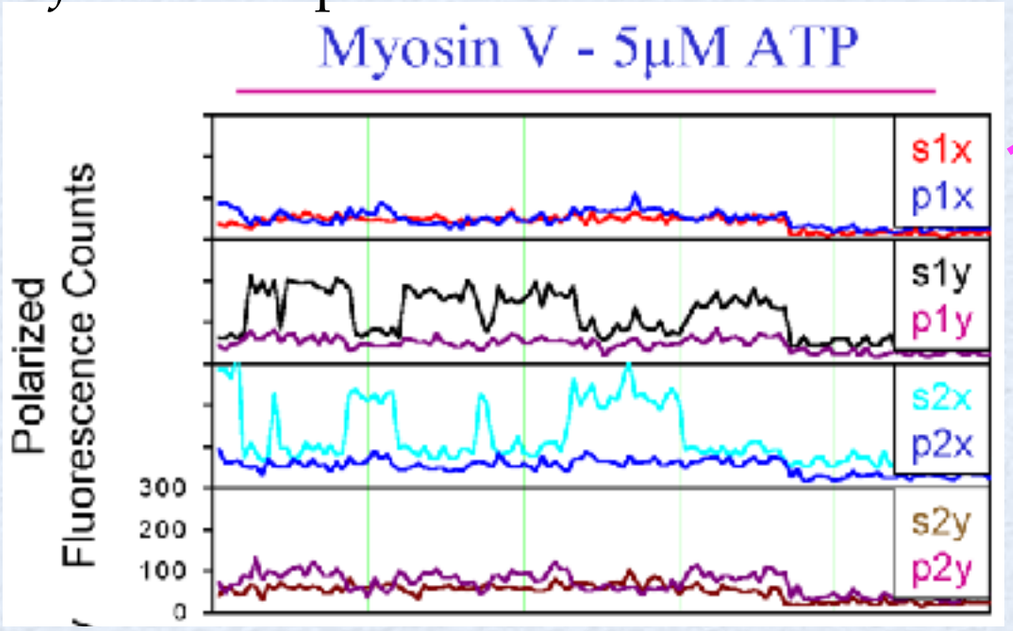
JN Forkey et al. Nature 2003

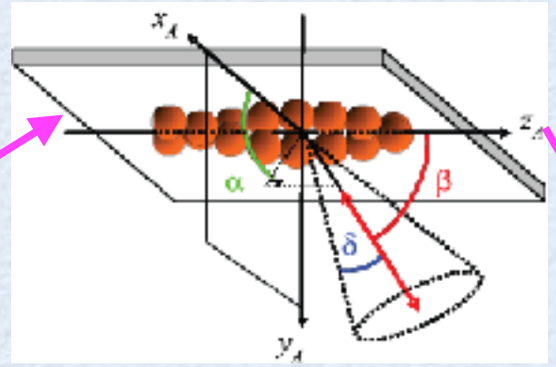
Previous state of the art

For our purposes, the upshot is that: ***We need to know the arrival rates of photons in each of several channels.*** Once we've got that, then we can use quantum mechanics to determine the orientation of the molecule in space.

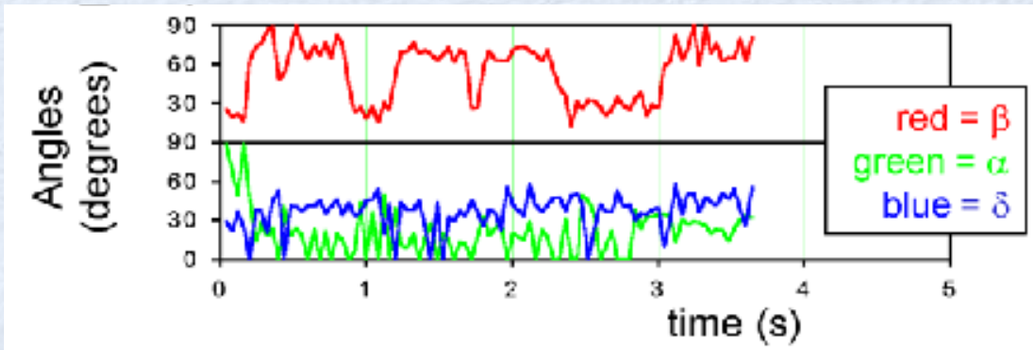
Unfortunately, existing analyses gave noisy rate determinations. That in turn led to poor determinations of orientation – garbage in/garbage out.

Left: binned photon counts in 8 channels.





Right: Polar and azimuthal angles of the fluorescent label, inferred from data on the left.



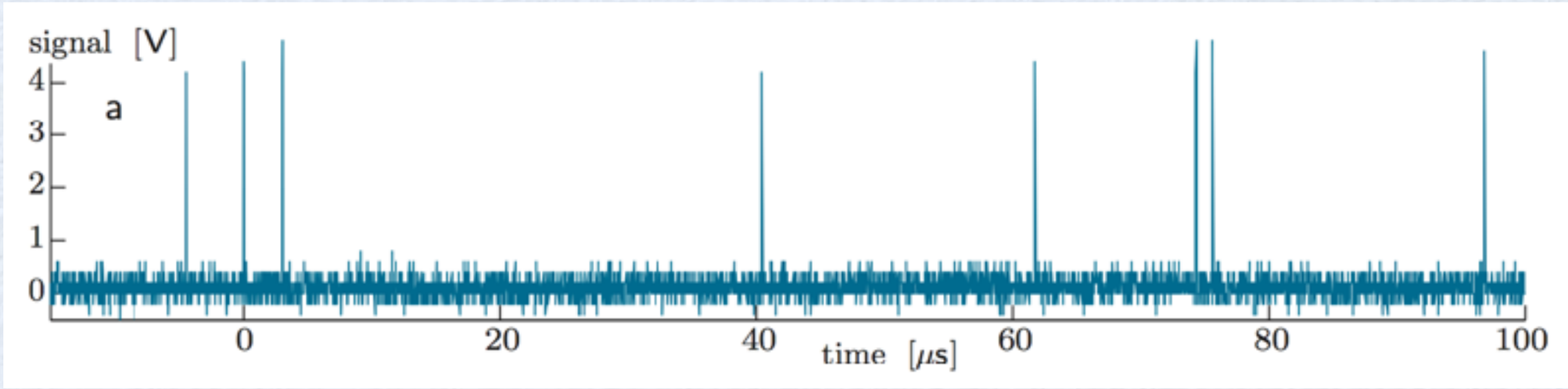
Noisy rate estimates lead to noisy orientation estimates.

Also, a state transition will generally happen in the middle of a time bin, spoiling our estimation of rates in that entire bin.

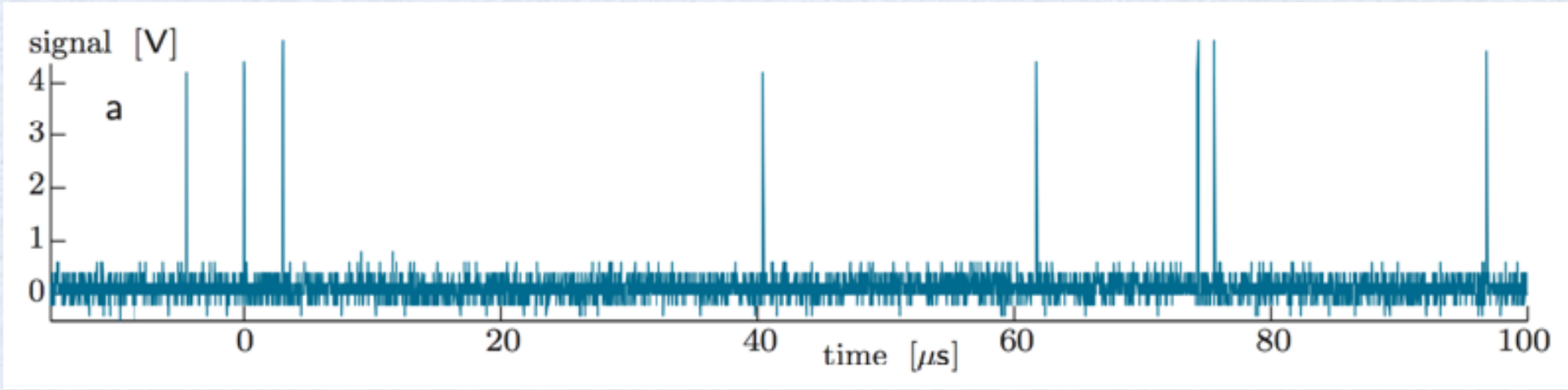
Moreover, you could easily miss a short-lived state -- e.g. the elusive diffusive-search step (if it exists). *Can we do better?*

JN Forkey et al. Nature 2003

The problem is that only a few thousand photons can be observed – makes it hard to estimate the rate:

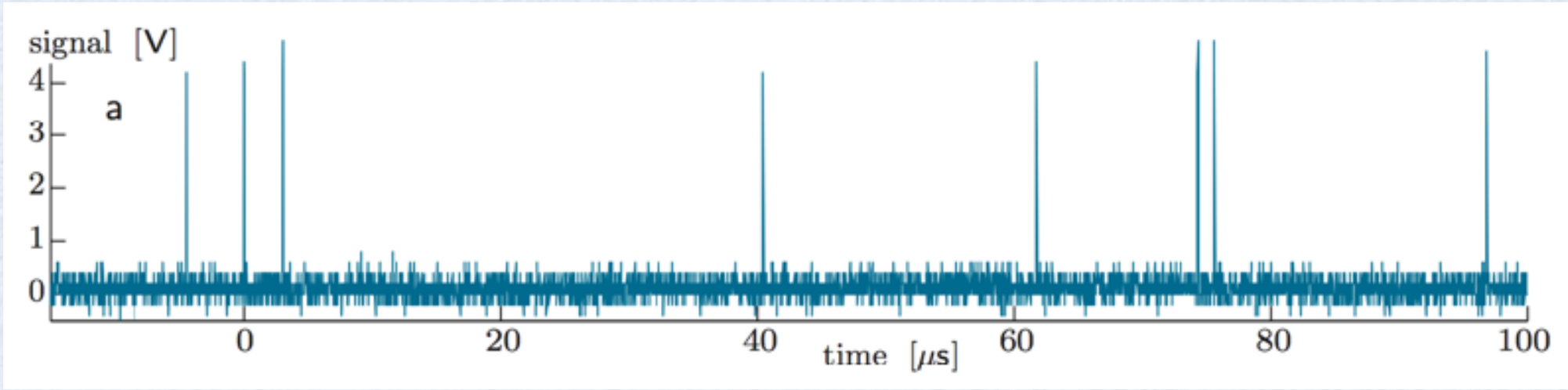


The problem is that only a few thousand photons can be observed – makes it hard to estimate the rate:



So we have a tough choice:

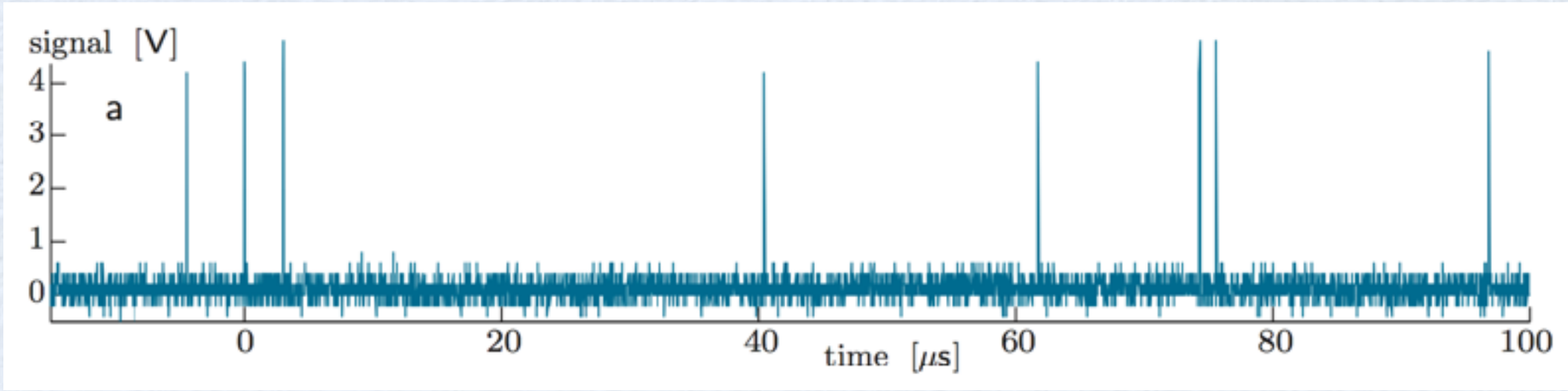
The problem is that only a few thousand photons can be observed – makes it hard to estimate the rate:



So we have a tough choice:

- Longer time bins degrade our ability to observe transient states, get kinetics, etc.

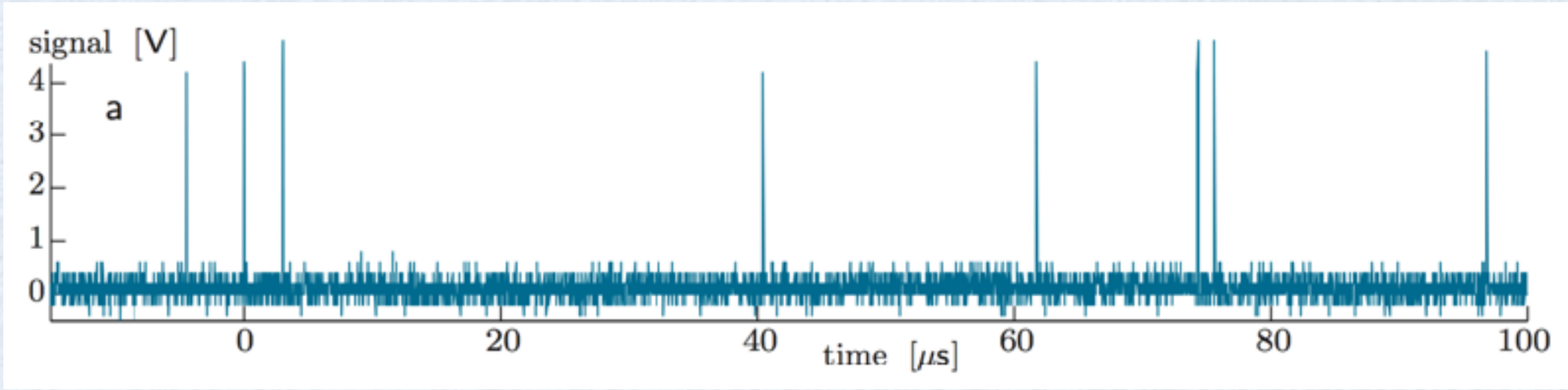
The problem is that only a few thousand photons can be observed – makes it hard to estimate the rate:



So we have a tough choice:

- Longer time bins degrade our ability to observe transient states, get kinetics, etc.
- Shorter time bins give worse relative standard deviation for our rate estimate.

The problem is that only a few thousand photons can be observed – makes it hard to estimate the rate:

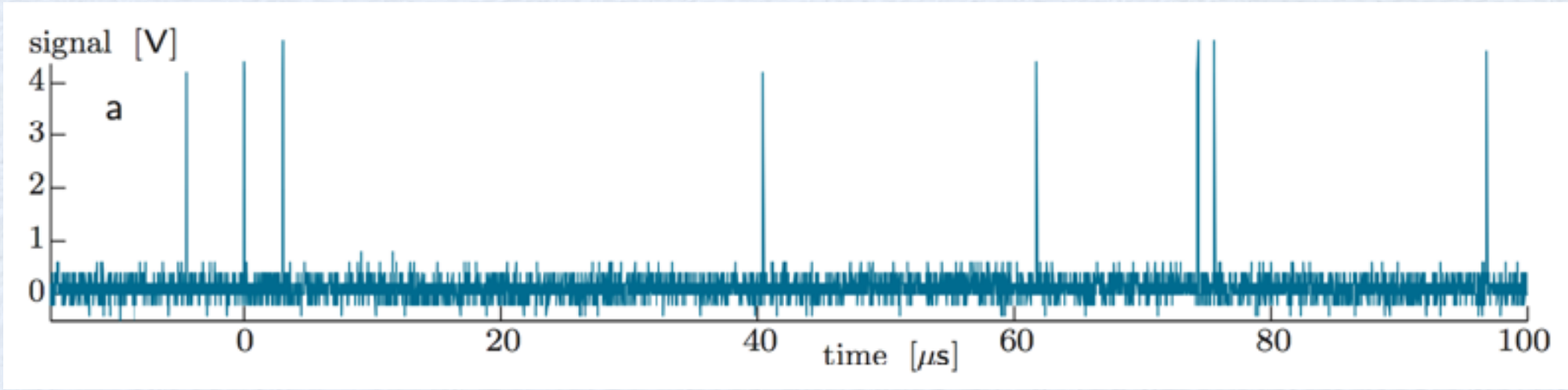


So we have a tough choice:

- Longer time bins degrade our ability to observe transient states, get kinetics, etc.
- Shorter time bins give worse relative standard deviation for our rate estimate.

Can we evade the cruel logic of photon statistics? If only we could find the changepoints *first*, then use the *entire durations* between consecutive changepoints as our windows–the biggest choice possible! That would lead to the best possible estimate of photon rates, and hence the best possible estimate of orientation.

The problem is that only a few thousand photons can be observed – makes it hard to estimate the rate:



So we have a tough choice:

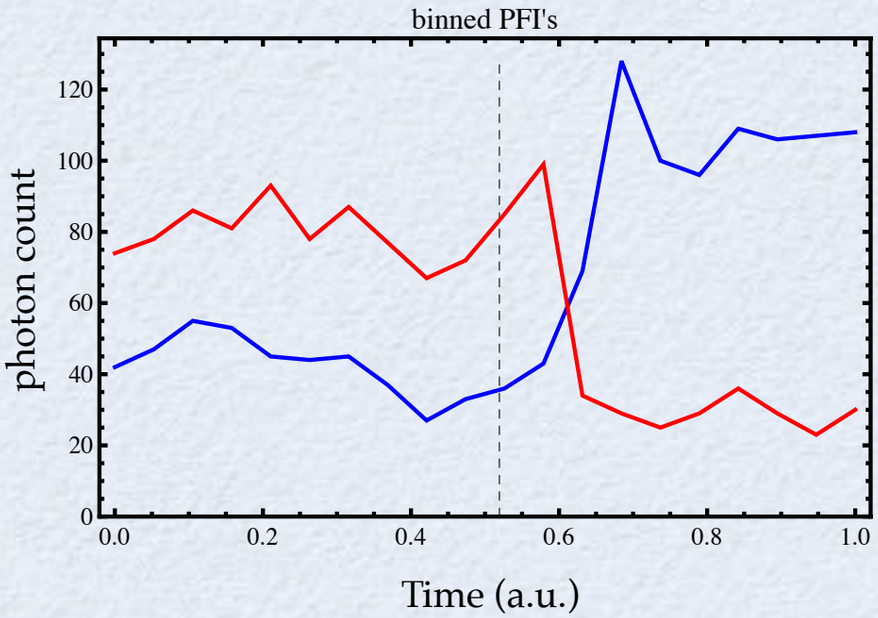
- Longer time bins degrade our ability to observe transient states, get kinetics, etc.
- Shorter time bins give worse relative standard deviation for our rate estimate.

Can we evade the cruel logic of photon statistics? If only we could find the changepoints *first*, then use the *entire durations* between consecutive changepoints as our windows–the biggest choice possible! That would lead to the best possible estimate of photon rates, and hence the best possible estimate of orientation.

But seems a chicken-and-egg problem: I need changepoints to find orientation, but the changepoints are themselves defined as changes in... orientation!

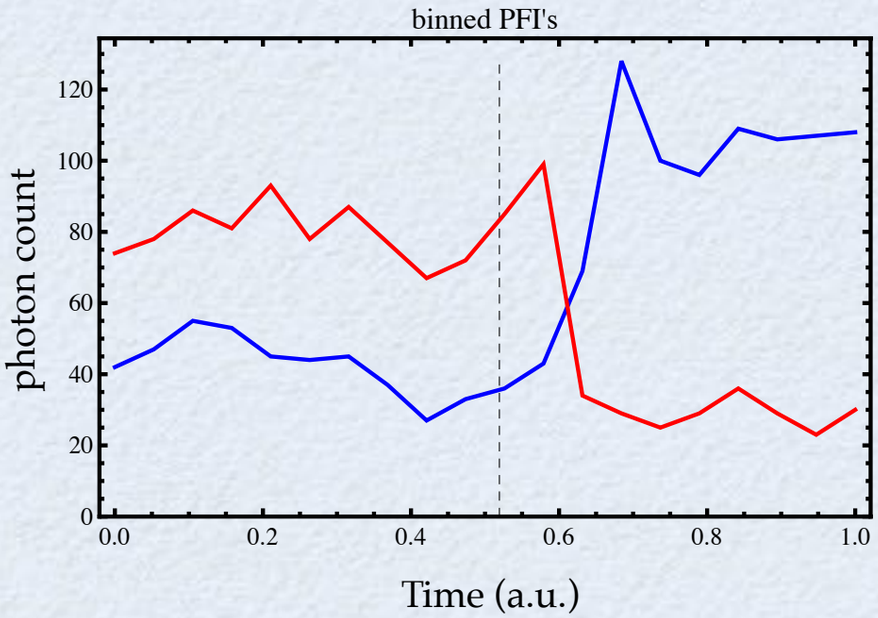
Here is some real experimental data. For simplicity, we look at only two channels. Only 1200 photons were observed in each.

Here is some real experimental data. For simplicity, we look at only two channels. Only 1200 photons were observed in each.



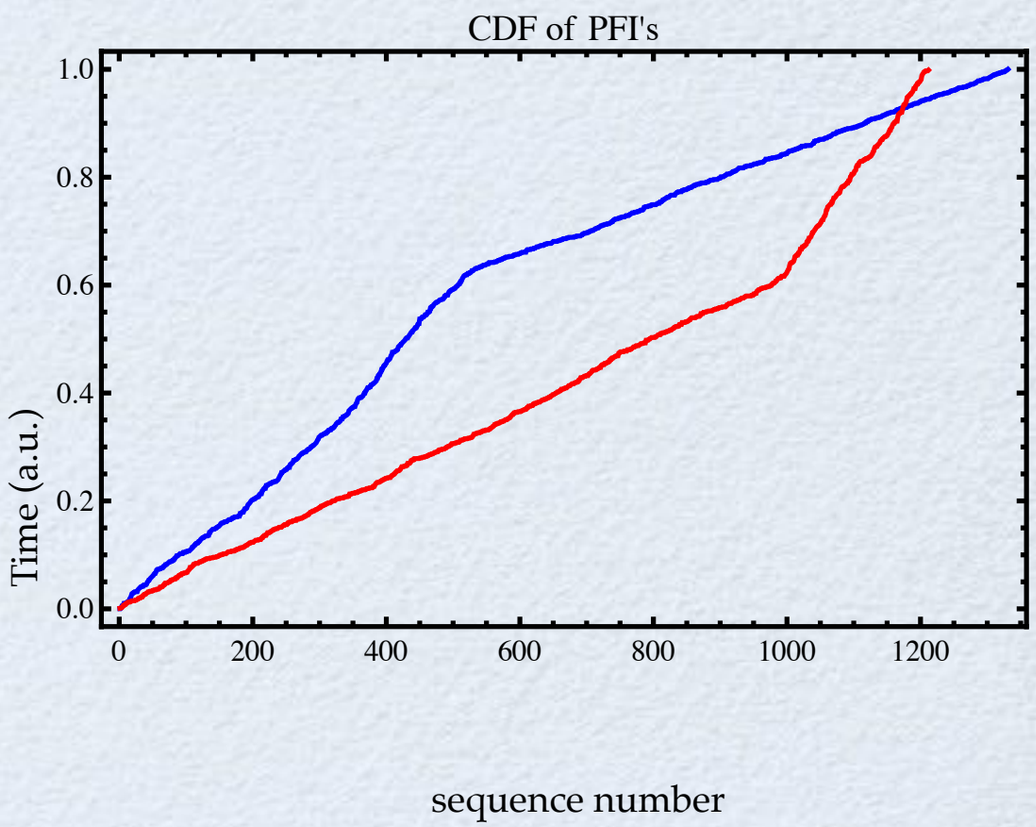
When we classify the photons by polarization and bin them (here 20 bins were used), that reveals a definite changepoint. But when exactly did it occur? Probably not at the dashed line shown, but how can we be more precise?

Here is some real experimental data. For simplicity, we look at only two channels. Only 1200 photons were observed in each.



When we classify the photons by polarization and bin them (here 20 bins were used), 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?

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, suddenly 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 PN, Meth. Enzymol. 487:431 (2011).

Systematize/generalize

- ***Why did that trick succeed?*** How did we extract such great time resolution from such cruddy data? What *principle* is at work?
- ***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.

Systematize/generalize

- ***Why did that trick succeed?*** How did we extract such great time resolution from such cruddy data? What *principle* is at work?
- ***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 channel (e.g. one photon polarization bin).

Systematize/generalize

- ***Why did that trick succeed?*** How did we extract such great time resolution from such cruddy data? What *principle* is at work?
- ***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 channel (e.g. one photon polarization bin).

Suppose that in total time T we catch N photons at times t_1,... t_N.

Systematize/generalize

- ***Why did that trick succeed?*** How did we extract such great time resolution from such cruddy data? What *principle* is at work?
- ***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 channel (e.g. one photon polarization bin).

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'.

Systematize/generalize

- ***Why did that trick succeed?*** How did we extract such great time resolution from such cruddy data? What *principle* is at work?
- ***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 channel (e.g. one photon polarization bin).

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'.

Systematize/generalize

- ***Why did that trick succeed?*** How did we extract such great time resolution from such cruddy data? What *principle* is at work?
- ***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 channel (e.g. one photon polarization bin).

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'.

To do this, we again need 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:

Systematize/generalize

- ***Why did that trick succeed?*** How did we extract such great time resolution from such cruddy data? What *principle* is at work?
- ***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 channel (e.g. one photon polarization bin).

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'.

To do this, we again need 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:

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

Systematize/generalize

- ***Why did that trick succeed?*** How did we extract such great time resolution from such cruddy data? What *principle* is at work?
- ***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 channel (e.g. one photon polarization bin).

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'.

To do this, we again need 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:

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

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'.

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

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

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

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

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

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'.

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

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

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

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

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

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'.

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

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

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

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

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

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'.

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

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

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

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

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

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'.

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

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

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

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

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

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'*.

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

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

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

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

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

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'.

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

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

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

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

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

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'.

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

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

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

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

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

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'.

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

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

Now: Divide the N photons into n that arrived before the putative changepoint, and n'=N-n that arrived after.

Take the limit $\Delta t \to 0$:

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

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

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'.

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

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

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

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

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

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'.

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

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

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

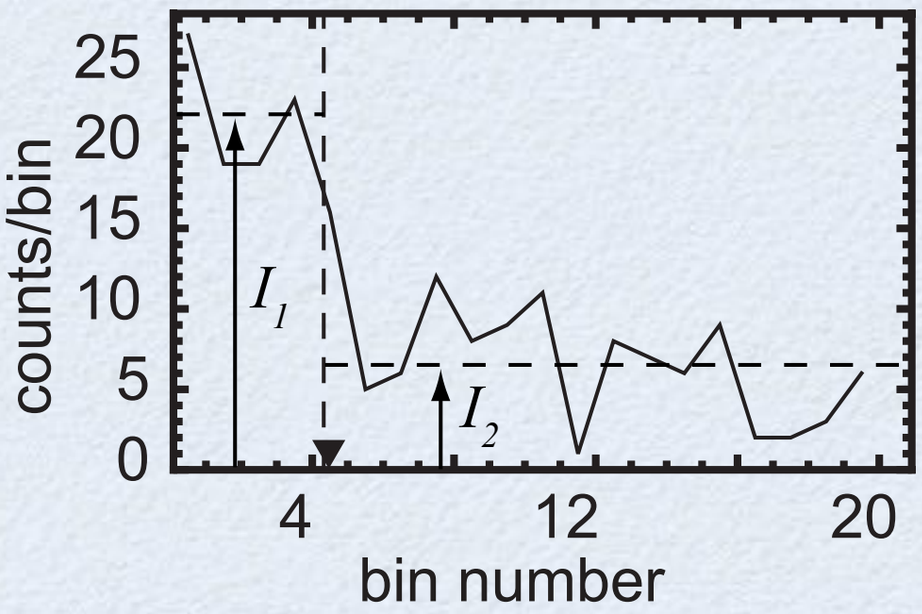
$$\mathcal{P} \approx N \log(\Delta t) + n \log R + n' \log R' - \left(\frac{t_*}{\Delta t} - n\right)\left(R\Delta t\right) - \left(\frac{T - t_*}{\Delta t} - 1 - (N - n)\right)\left(R'\Delta t\right)$$

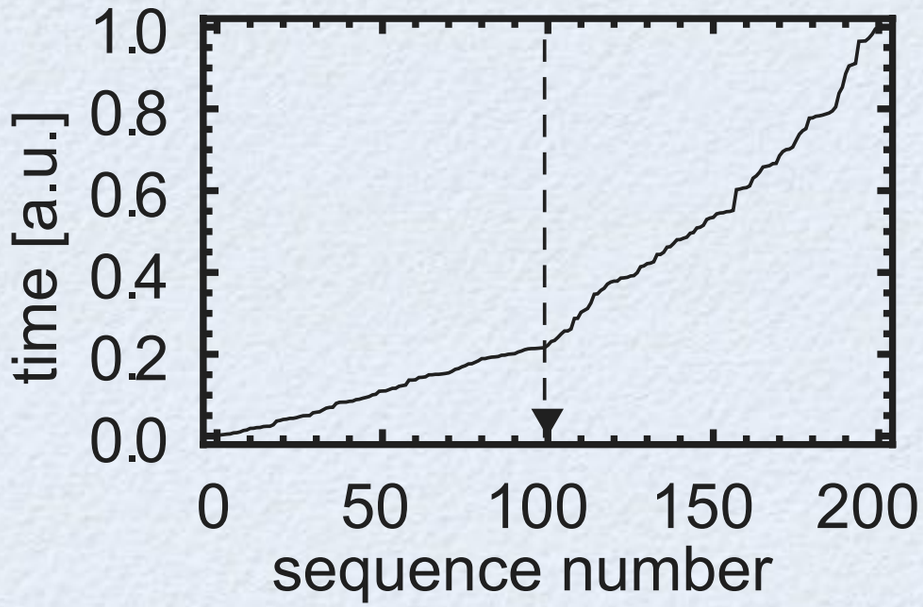
$$\approx \text{const} + n \log R + n' \log R' - Rt_* - R'(T - t_*)$$

Maximize this first over R and R':

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

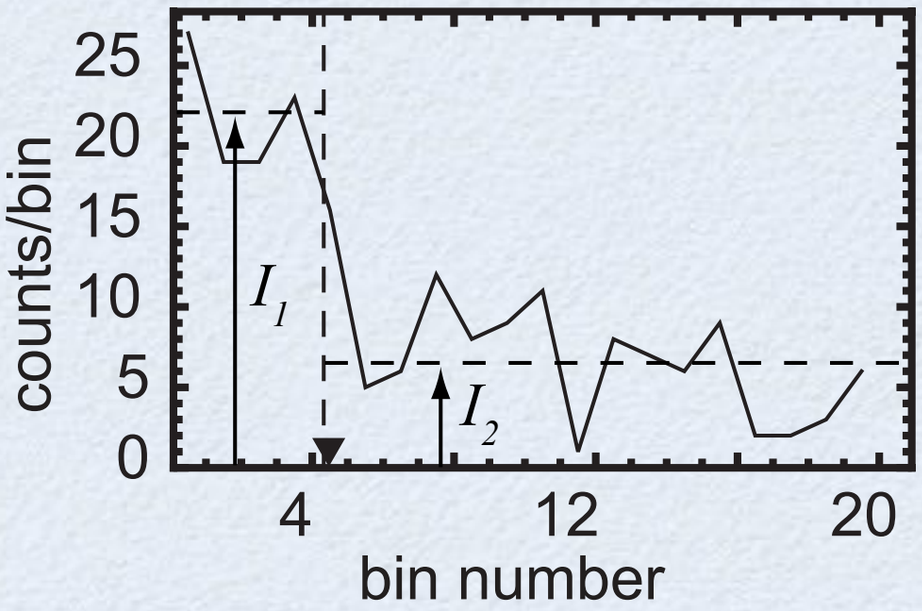
OK, **duh**, that was no surprise! But it does explain why we can just lay a ruler along the cumulative plot to get our best estimate of the before and after rates.
More interestingly, we can substitute these optimal rates into the formula for $\mathcal{P}$ to find the likelihood as a function of putative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





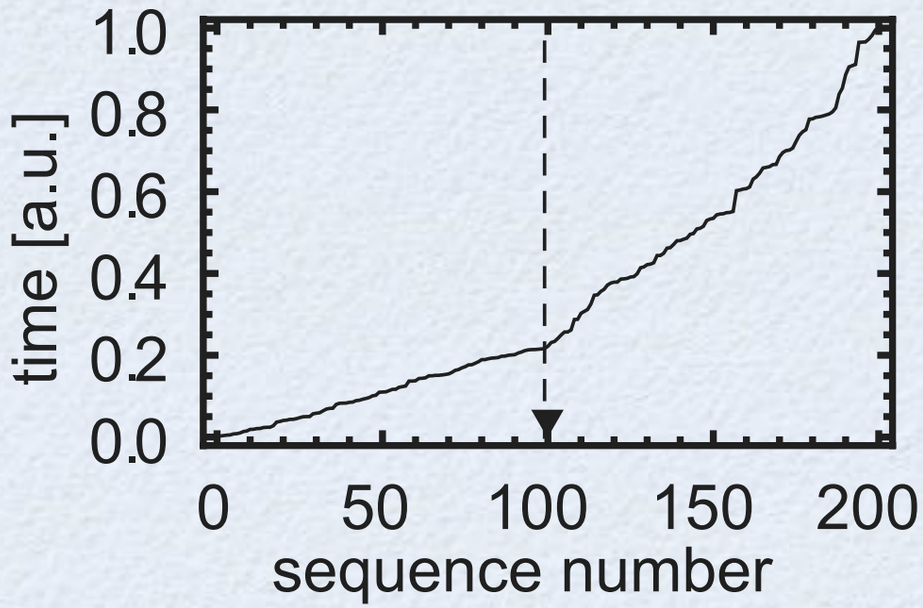
Left: Realistic, but fake, data, shown in traditional binned form and in the improved version.

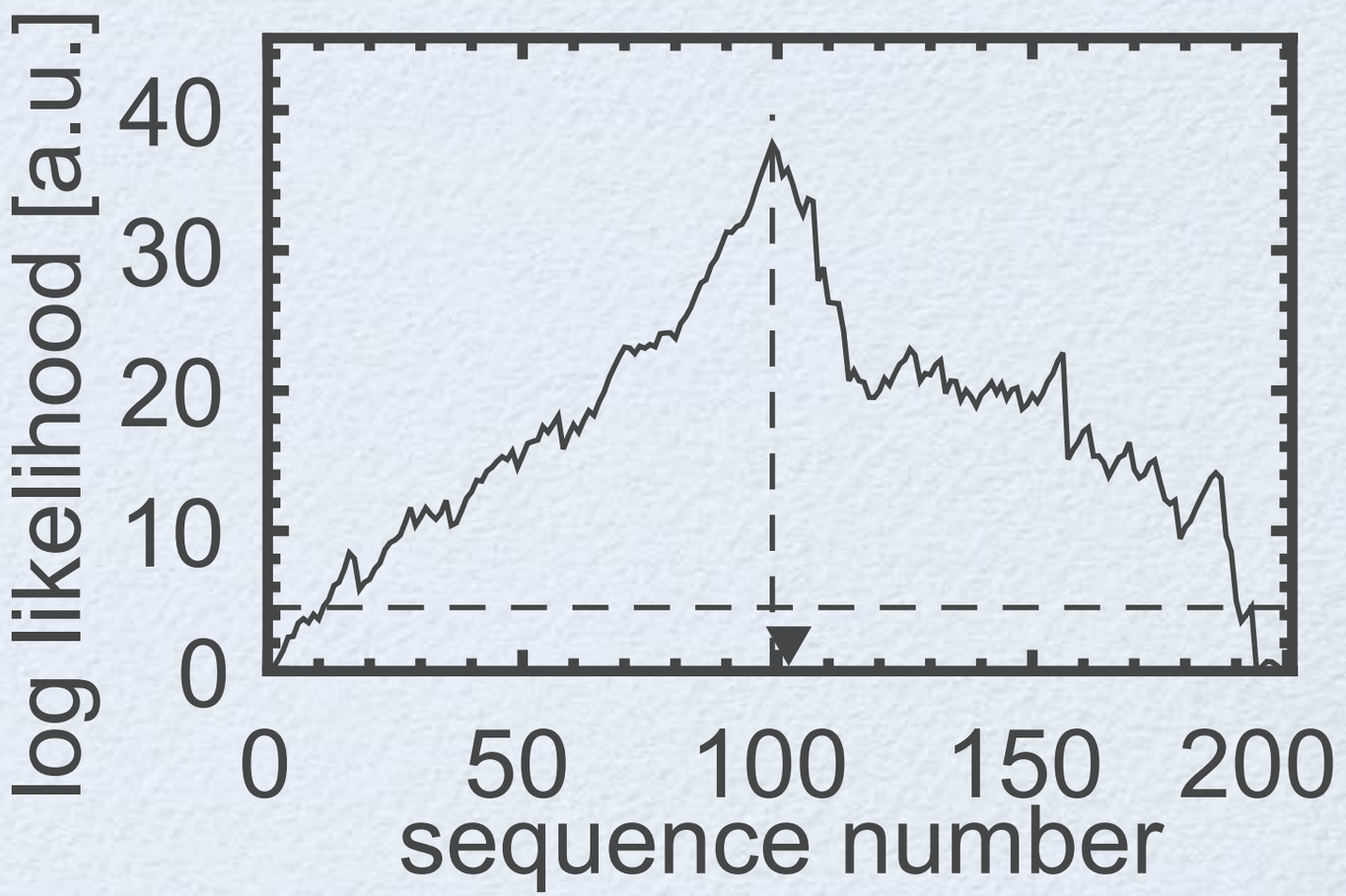
JF Beausang, YE Goldman, and PN, Meth. Enzymol. 487:431 (2011).



Left: Realistic, but fake, data, shown in traditional binned form and in the improved version.

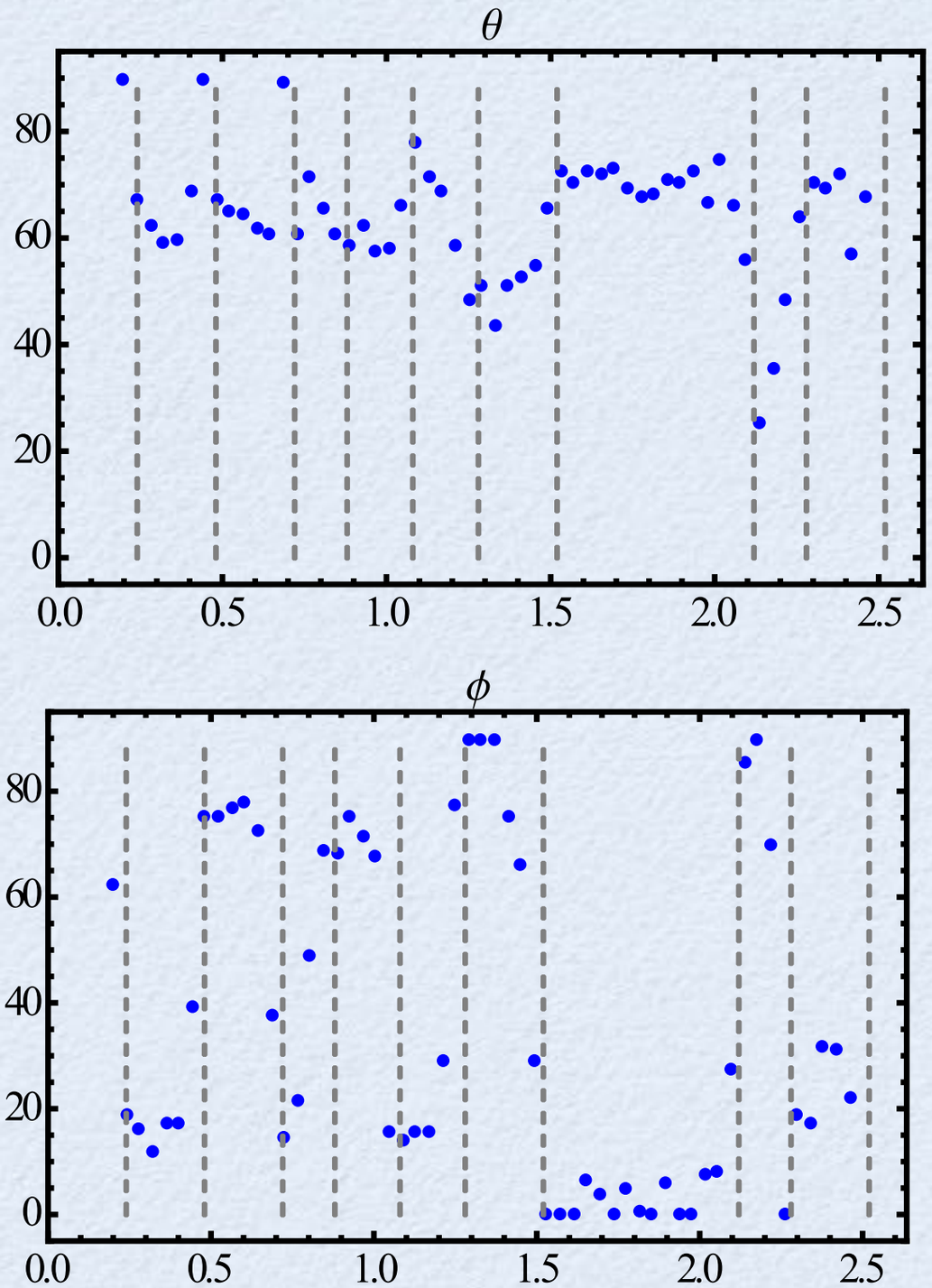
Below: Likelihood function for placement of the changepoint. Dashed line, maximum-likelihood point. Black triangle: Actual changepoint used to generate the simulated data. The analysis found a robust changepoint, even though there were a total of just 200 photons in the entire dataset.





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

Payoff



Oh, yes--it also works on 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.

Our approach first finds changepoints, shown as dashed lines.

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

Payoff

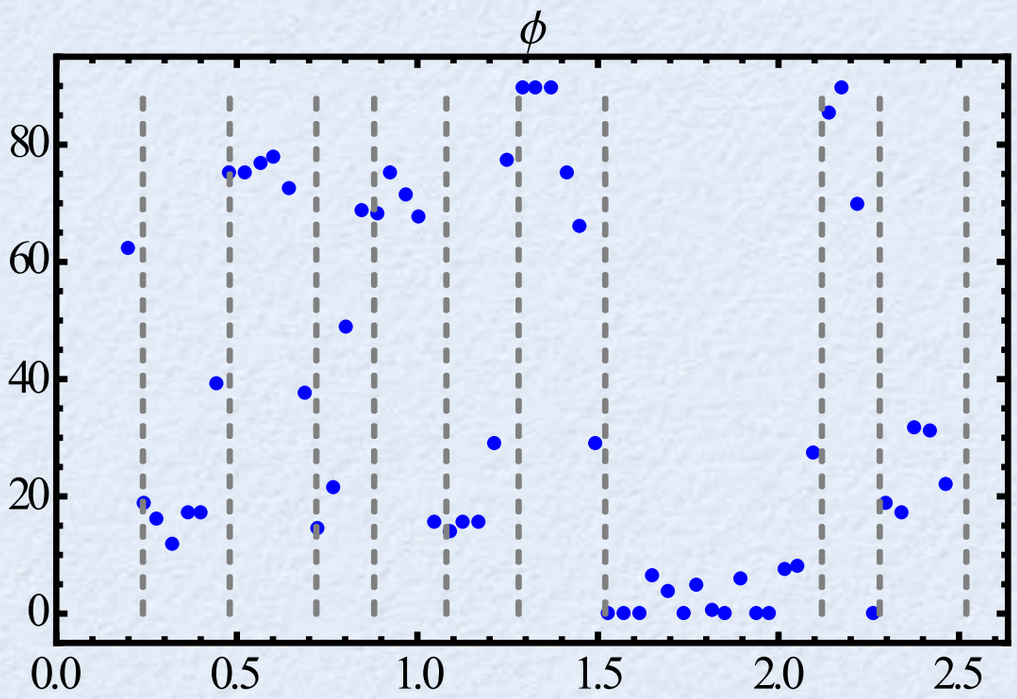
Oh, yes--it also works on multiple-channel data, data with many different changepoints...

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.

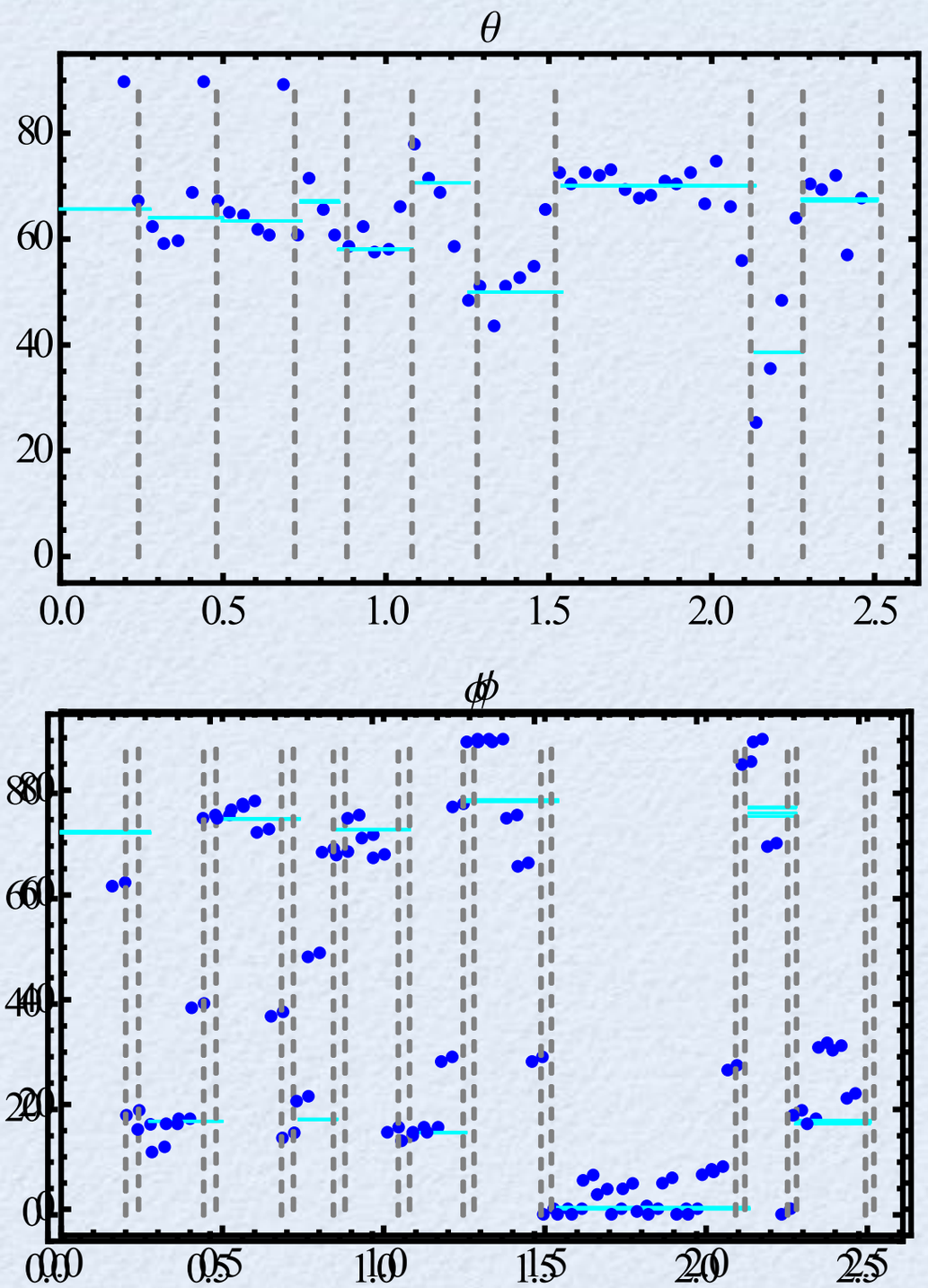
One problem is that if a transition happens in the middle of a time bin, then the inferred orientation in that time bin can be crazy.

Our approach first finds changepoints, shown as dashed lines.



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

Payoff



Oh, yes--it also works on 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.

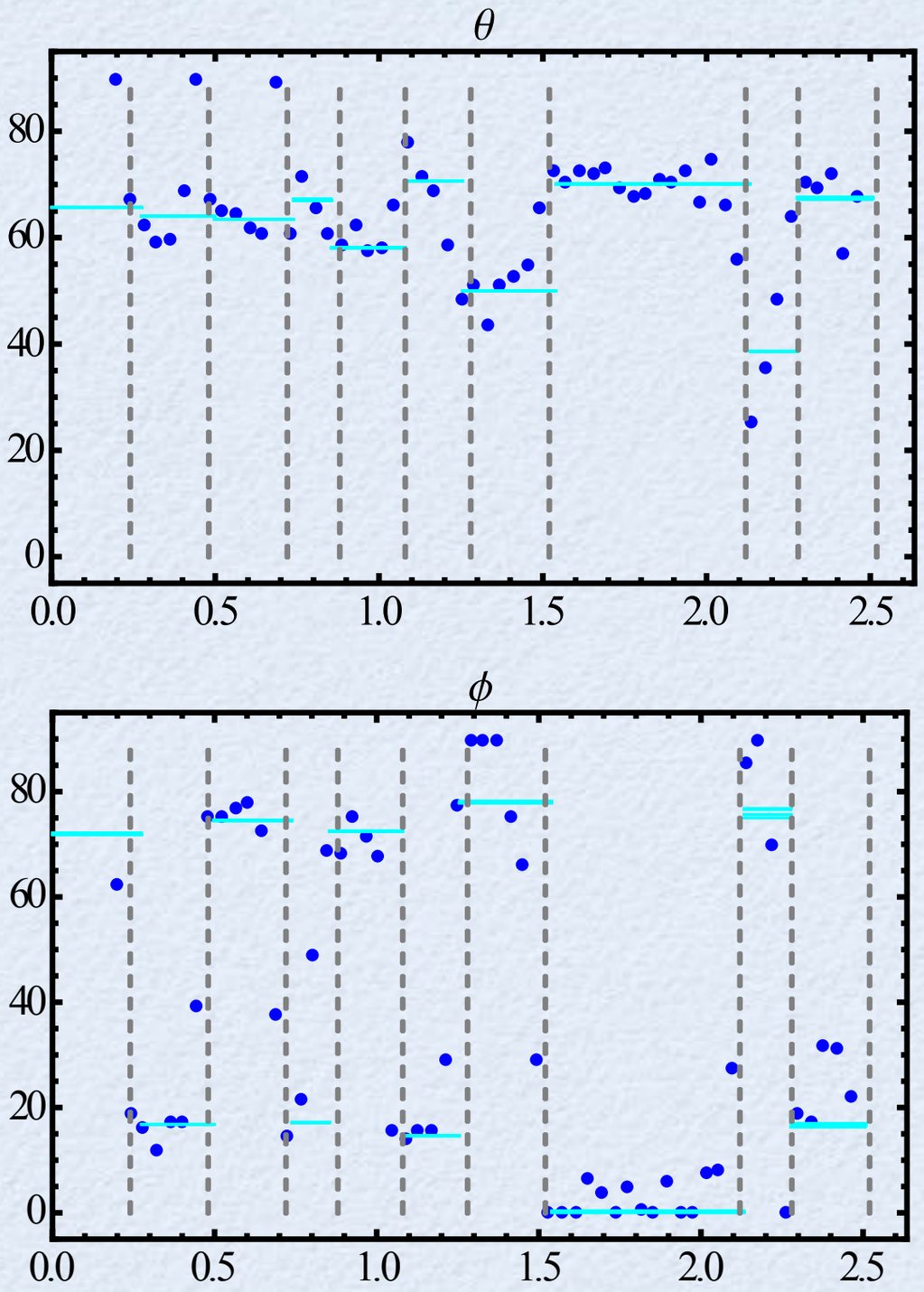
That procedure leads to the rather noisy dots shown here.

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.

Our approach first finds changepoints, shown as dashed lines.

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

Payoff



Oh, yes--it also works on 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.

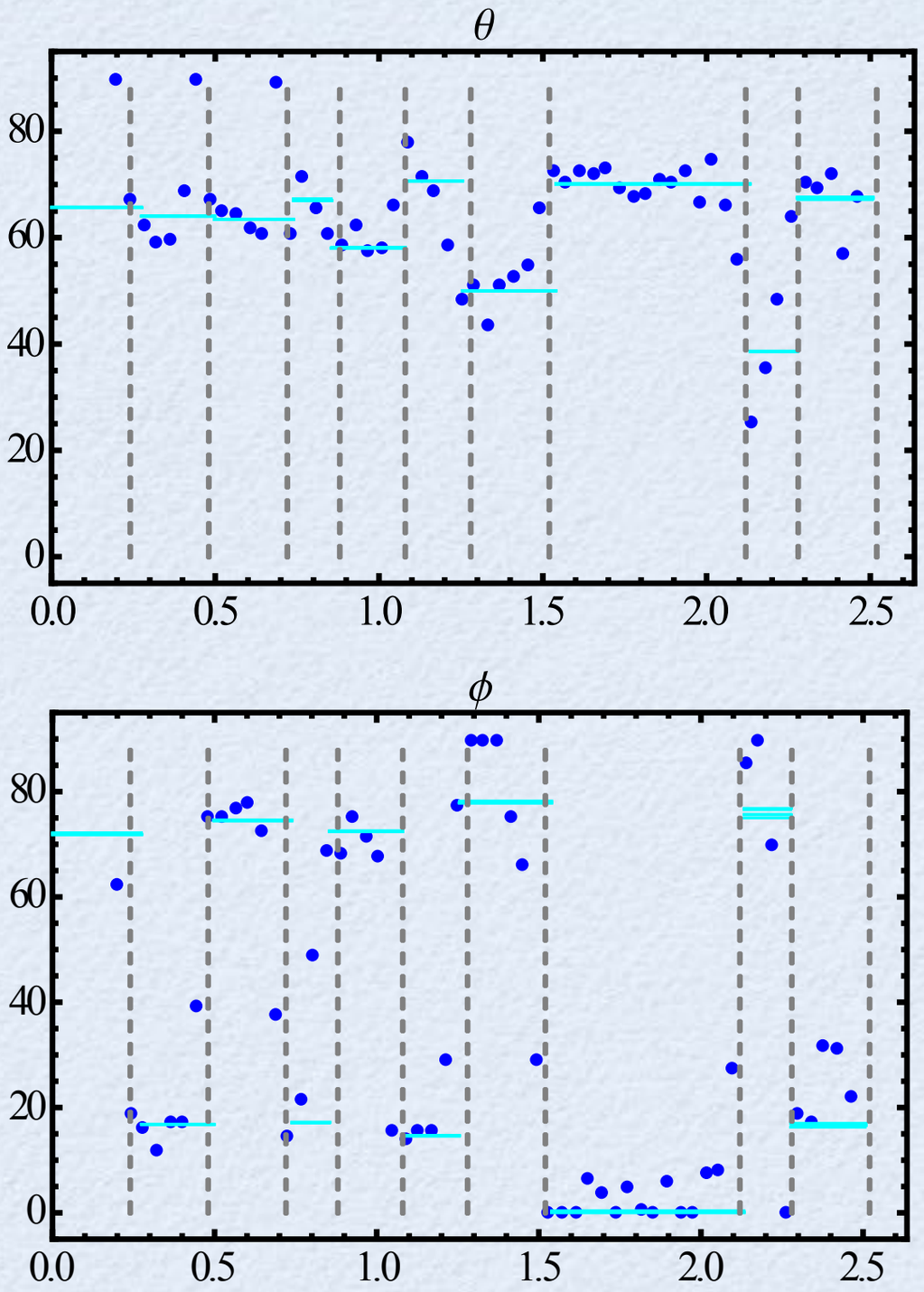
That procedure leads to the rather noisy dots shown here.

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.

Our approach first finds changepoints, shown as dashed lines.

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

Payoff



Oh, yes--it also works on multiple-channel data, data with many different changepoints...

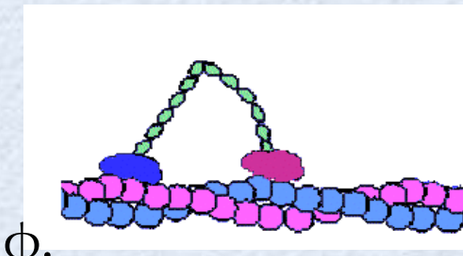
Previously, people would take data from multiple polarizations, bin it, and pipe the inferred intensities into a maximum-likelihood estimator of the orientation of the fluorophore.

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.

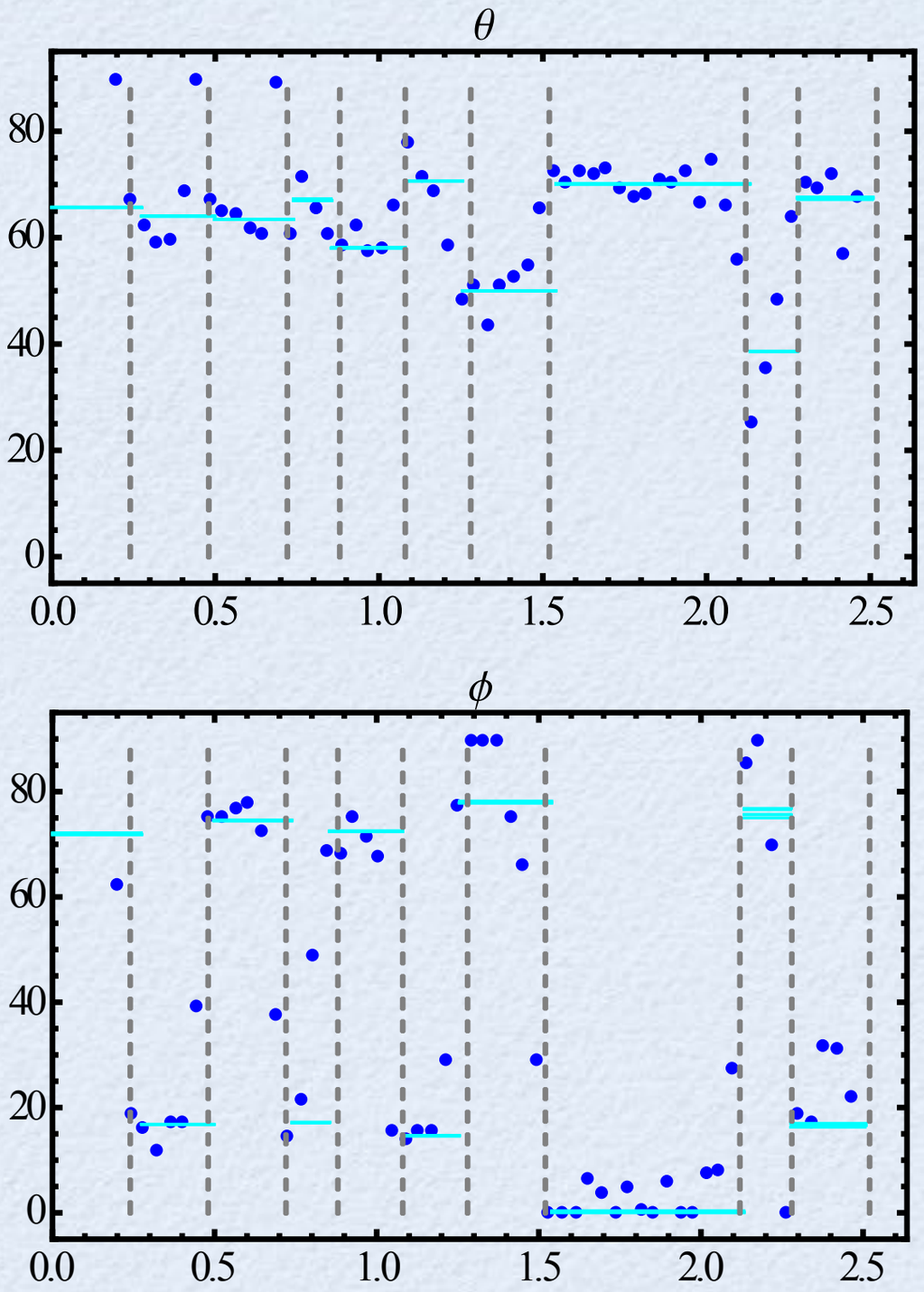
Our approach first finds changepoints, shown as dashed lines.

Then the solid lines shown 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 PN, Meth. Enzymol. (2011); JF Beausang, DY Shroder, PN, and YE Goldman, Biophys J (2013).

Payoff



Oh, yes--it also works on multiple-channel data, data with many different changepoints...

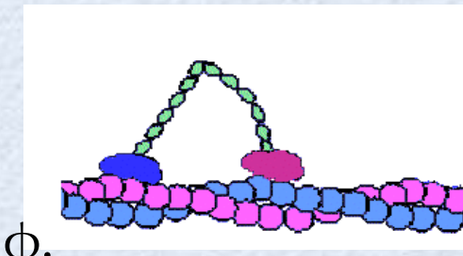
Previously, people would take data from multiple polarizations, bin it, and pipe the inferred intensities into a maximum-likelihood estimator of the orientation of the fluorophore.

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.

Our approach first finds changepoints, shown as dashed lines.

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

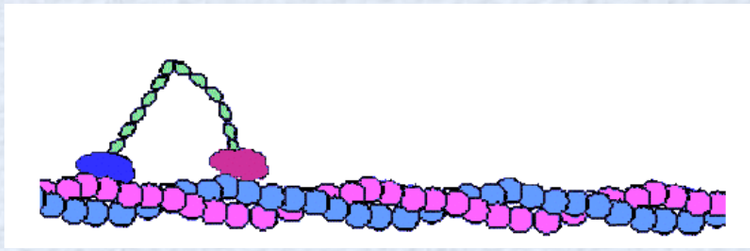


We got a ***50-fold improvement*** in time resolution for finding changepoints, compared to the binning method, ***without changing the apparatus.***

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

Summary Part 3

- 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.

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

Summary Part 3

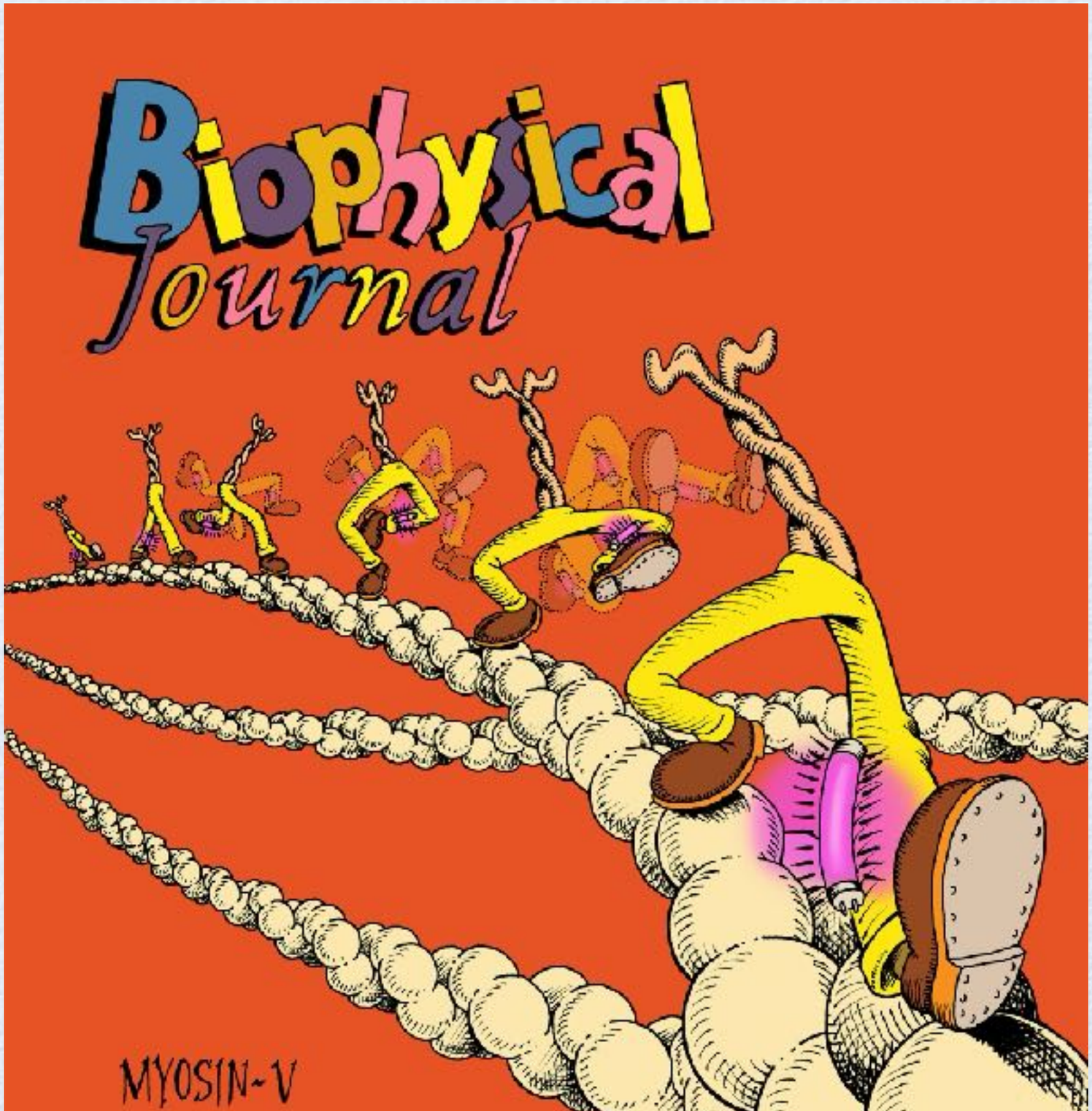
- When you only get a million photons, you'd better make every photon count.

- 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, PN, and YE Goldman, Biophys J (2013).



Part 4

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Part 4

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Sometimes your model's prediction is a probability distribution (really always);
and

Part 4

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Sometimes your model's prediction is a probability distribution (really always);
and
Single-molecule biophysical techniques give you individual data points for individual molecular transactions;
but

Part 4

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Sometimes your model's prediction is a probability distribution (really always);
and
Single-molecule biophysical techniques give you individual data points for individual molecular transactions;
but
Many of us grew up binning data, then least-squares fitting it, which destroys some of its information content, distorts relative importance of different parts of the data, etc.
so

Part 4

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Sometimes your model's prediction is a probability distribution (really always);
and
Single-molecule biophysical techniques give you individual data points for individual molecular transactions;
but
Many of us grew up binning data, then least-squares fitting it, which destroys some of its information content, distorts relative importance of different parts of the data, etc.
so
The fact that that that's often unnecessary is potentially interesting, even beyond the scope of today's applications.

Why [background]

Shalev, M. and Baasov, T. (2014) Med Chem. Commun, 5(8):1092-1105. Loudon, J.A. (2013) J Bioanal Biomed, 5:079-096. Nadeem Siddiqui, and Nahum Sonenberg PNAS 2016;113:12353-12355

Why [background]

- The ribosome has various "proofreading" steps. They've been studied closely in bacterial ribosome – ***not so much yet in eukaryotes.***

Shalev, M. and Baasov, T. (2014) Med Chem. Commun, 5(8):1092-1105. Loudon, J.A. (2013) J Bioanal Biomed, 5:079-096. Nadeem Siddiqui, and Nahum Sonenberg PNAS 2016;113:12353-12355

Why [background]

- The ribosome has various "proofreading" steps. They've been studied closely in bacterial ribosome – ***not so much yet in eukaryotes.***
- There are ~7000 genetically transmitted disorders, ~ 11% of which are nonsense mutations like cystic fibrosis (CF), Duchenne muscular dystrophy (DMD), etc. which specifically involve transition of a valid codon to a "stop" codon.

Shalev, M. and Baasov, T. (2014) Med Chem. Commun, 5(8):1092-1105. Loudon, J.A. (2013) J Bioanal Biomed, 5:079-096. Nadeem Siddiqui, and Nahum Sonenberg PNAS 2016;113:12353-12355

Why [background]

- The ribosome has various "proofreading" steps. They've been studied closely in bacterial ribosome – ***not so much yet in eukaryotes.***
- There are ~7000 genetically transmitted disorders, ~ 11% of which are nonsense mutations like cystic fibrosis (CF), Duchenne muscular dystrophy (DMD), etc. which specifically involve transition of a valid codon to a "stop" codon.
- Symptoms in patients can be alleviated even with just small amount of full length protein.

Shalev, M. and Baasov, T. (2014) Med Chem. Commun, 5(8):1092-1105. Loudon, J.A. (2013) J Bioanal Biomed, 5:079-096. Nadeem Siddiqui, and Nahum Sonenberg PNAS 2016;113:12353-12355

Why [background]

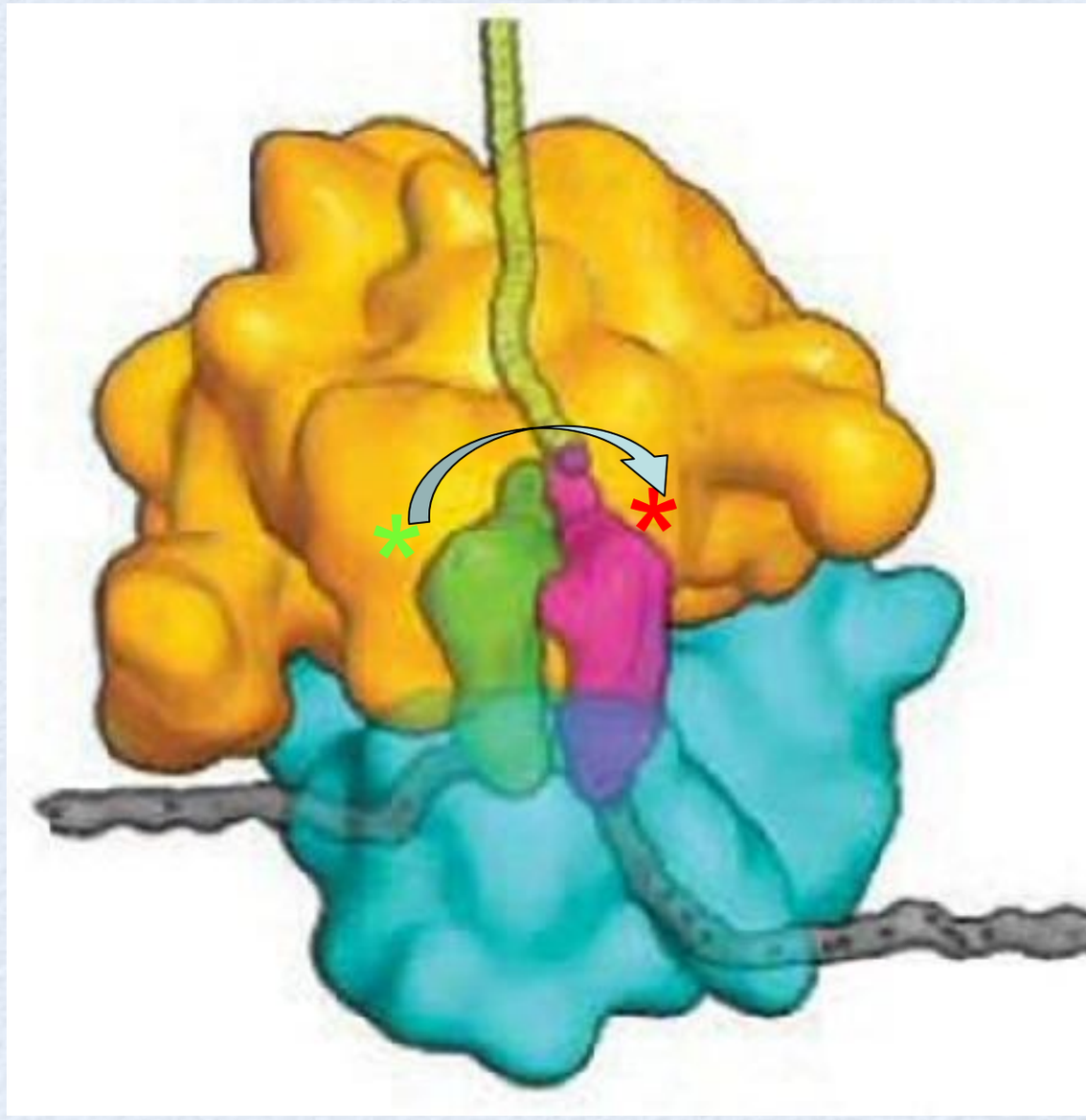
- The ribosome has various "proofreading" steps. They've been studied closely in bacterial ribosome – ***not so much yet in eukaryotes.***
- There are ~7000 genetically transmitted disorders, ~ 11% of which are nonsense mutations like cystic fibrosis (CF), Duchenne muscular dystrophy (DMD), etc. which specifically involve transition of a valid codon to a "stop" codon.
- Symptoms in patients can be alleviated even with just small amount of full length protein.
- Drugs such as Ataluren hold promise for helping ribosome to chug through this particular "stop." How do they work?

I won't answer, but it would be good to know as much as possible about the working cycle of the eukaryotic ribosome.

Shalev, M. and Baasov, T. (2014) Med Chem. Commun, 5(8):1092-1105. Loudon, J.A. (2013) J Bioanal Biomed, 5:079-096. Nadeem Siddiqui, and Nahum Sonenberg PNAS 2016;113:12353-12355

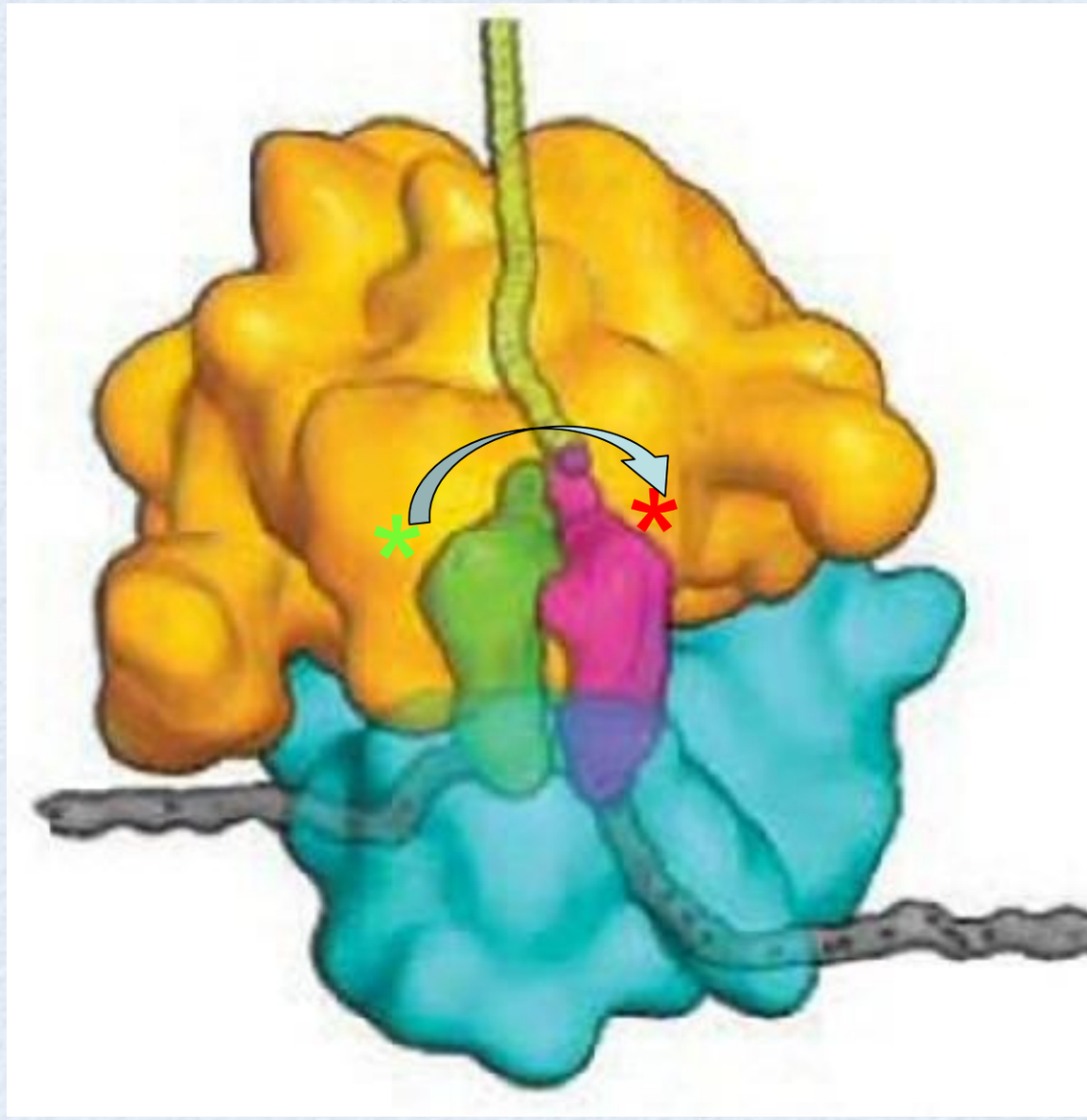
Experiment and puzzle

Single-molecule Fluorescence Resonance Energy Transfer (smFRET) tells exactly when two specifically labeled molecules are spatially close (high transfer) or not (low transfer). Hundreds, even thousands of molecules can be simultaneously monitored yielding individual time courses.



Experiment and puzzle

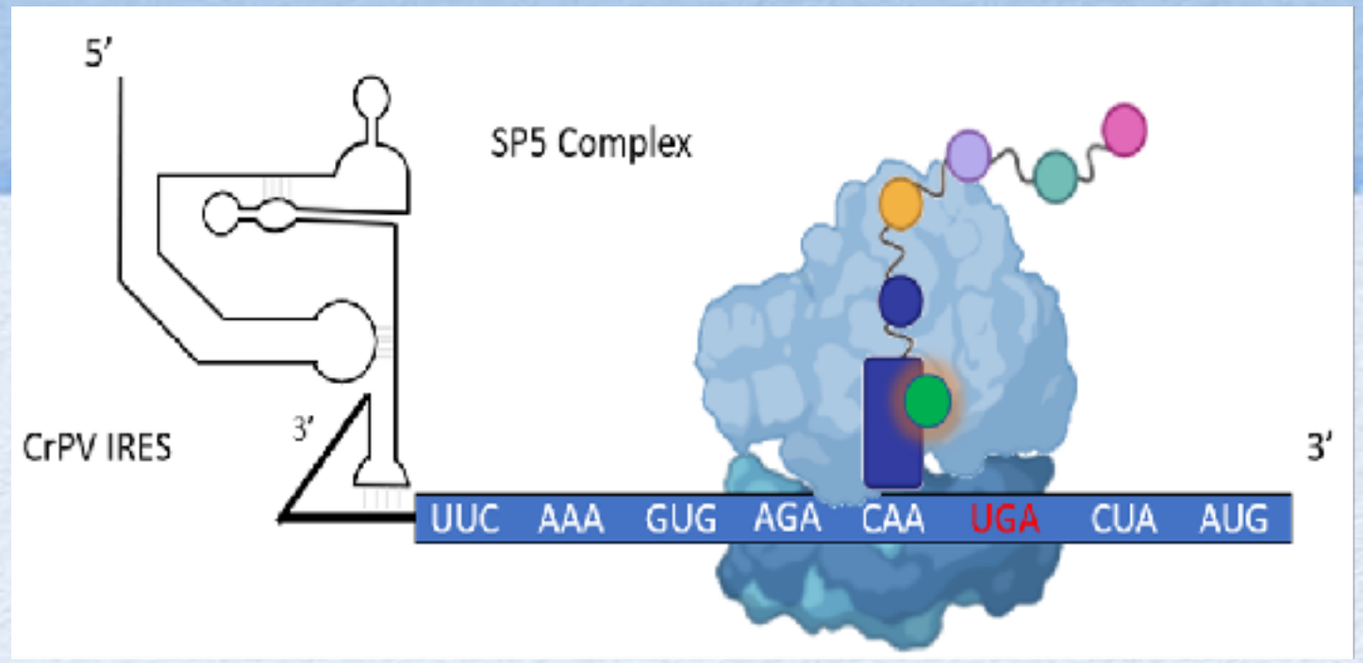
Single-molecule Fluorescence Resonance Energy Transfer (smFRET) tells exactly when two specifically labeled molecules are spatially close (high transfer) or not (low transfer). Hundreds, even thousands of molecules can be simultaneously monitored yielding individual time courses.



Schematic showing ribosome assembled on an mRNA with a UGA (stop) codon positioned in the A-site of the ribosome. To visualize every binding event, my colleagues made a FRET pair consisting of ternary complex in solution with a donor fluorophore and already-incorporated tRNA in the ribosome with an acceptor fluorophore in the P-site.

Experiment

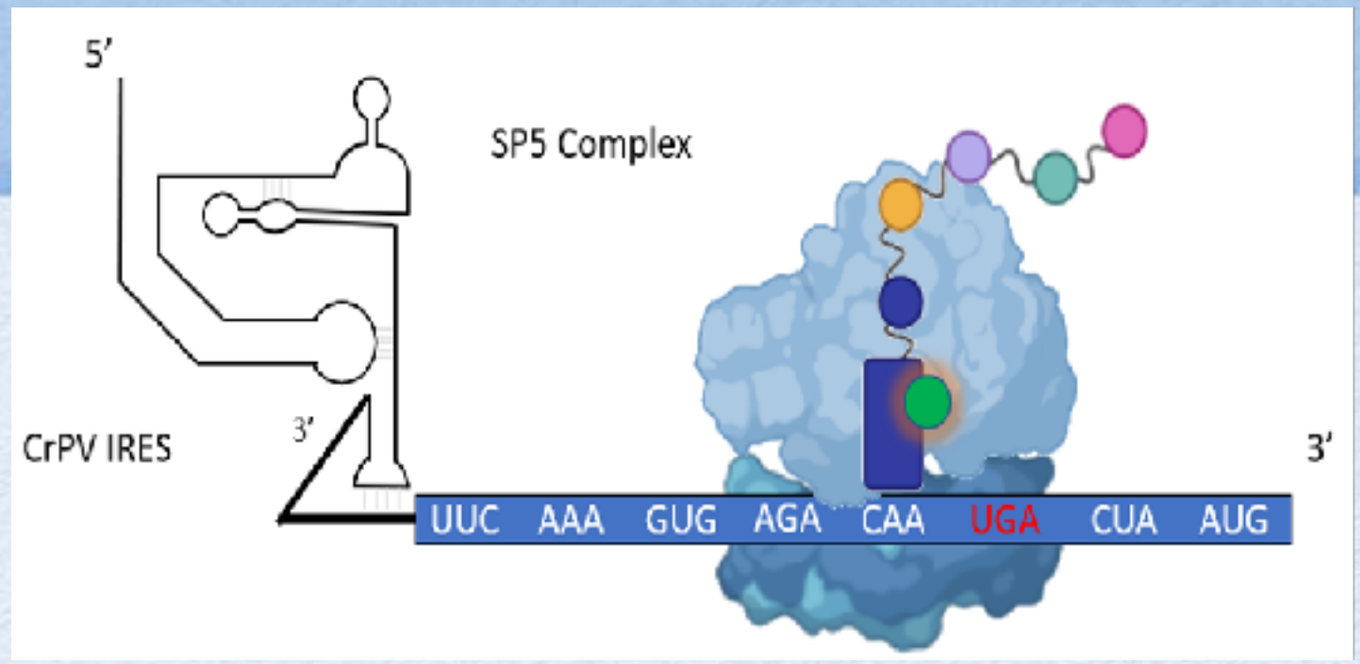
tRNA is supplied solution in the form of "ternary complex," or "TC." It samples the A-site of ribosome, binding transiently until eventually it is (wrongly) bound stably. *FRET lets us see individual binding and unbinding events with high time resolution.*

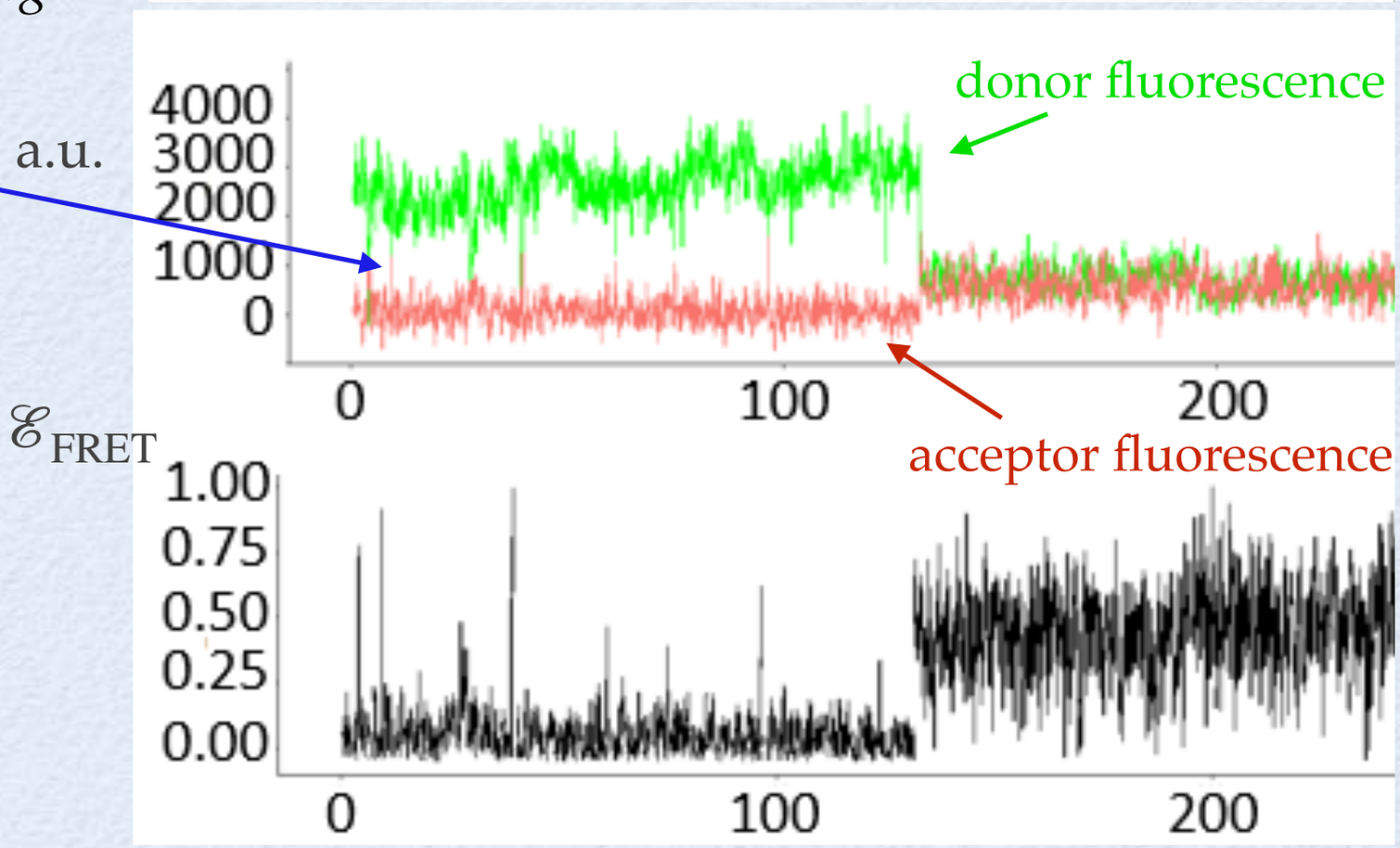


Experiment by Clark Fritsch, Arpan Bhattacharya, Martin Ng, Hong Li, Barry S. Cooperman, Yale E. Goldman

Experiment

tRNA is supplied solution in the form of "ternary complex," or "TC." It samples the A-site of ribosome, binding transiently until eventually it is (wrongly) bound stably. *FRET lets us see individual binding and unbinding events with high time resolution.*

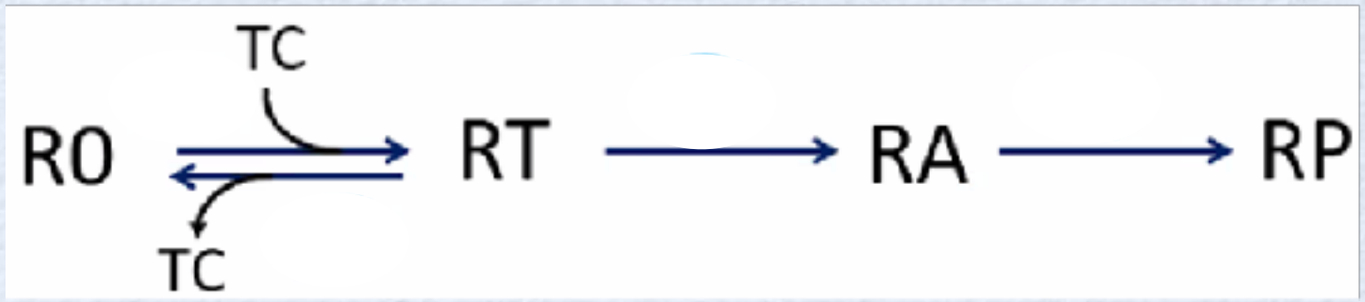




Representative single-molecule trace collected to study eukaryotic tRNA selection on ribosomes programmed on a near-cognate mRNA.

Experiment by Clark Fritsch, Arpan Bhattacharya, Martin Ng, Hong Li, Barry S. Cooperman, Yale E. Goldman

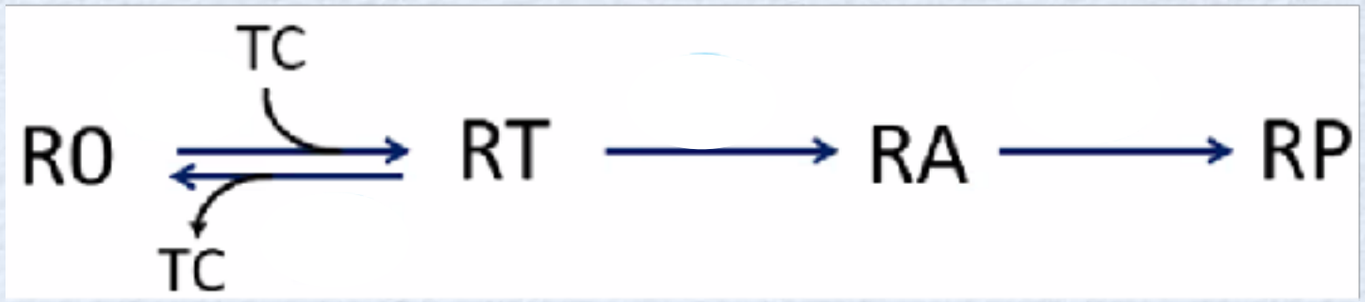
Uh-oh



In the simplest model, of course initial binding should be faster if ternary complex (TC) is more abundant. But every binding event is predicted to be independent of every other one, and in particular:

- The distribution of waiting times to bind near-cognate TC should be the same for every attempt.
- The distribution of the number of attempts before stable binding should be independent of TC concentration.

Uh-oh



In the simplest model, of course initial binding should be faster if ternary complex (TC) is more abundant. But every binding event is predicted to be independent of every other one, and in particular:

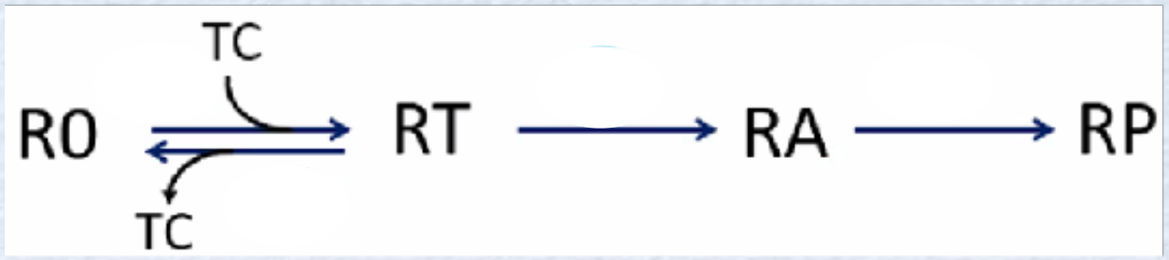
- The distribution of waiting times to bind near-cognate TC should be the same for every attempt.
- The distribution of the number of attempts before stable binding should be independent of TC concentration.

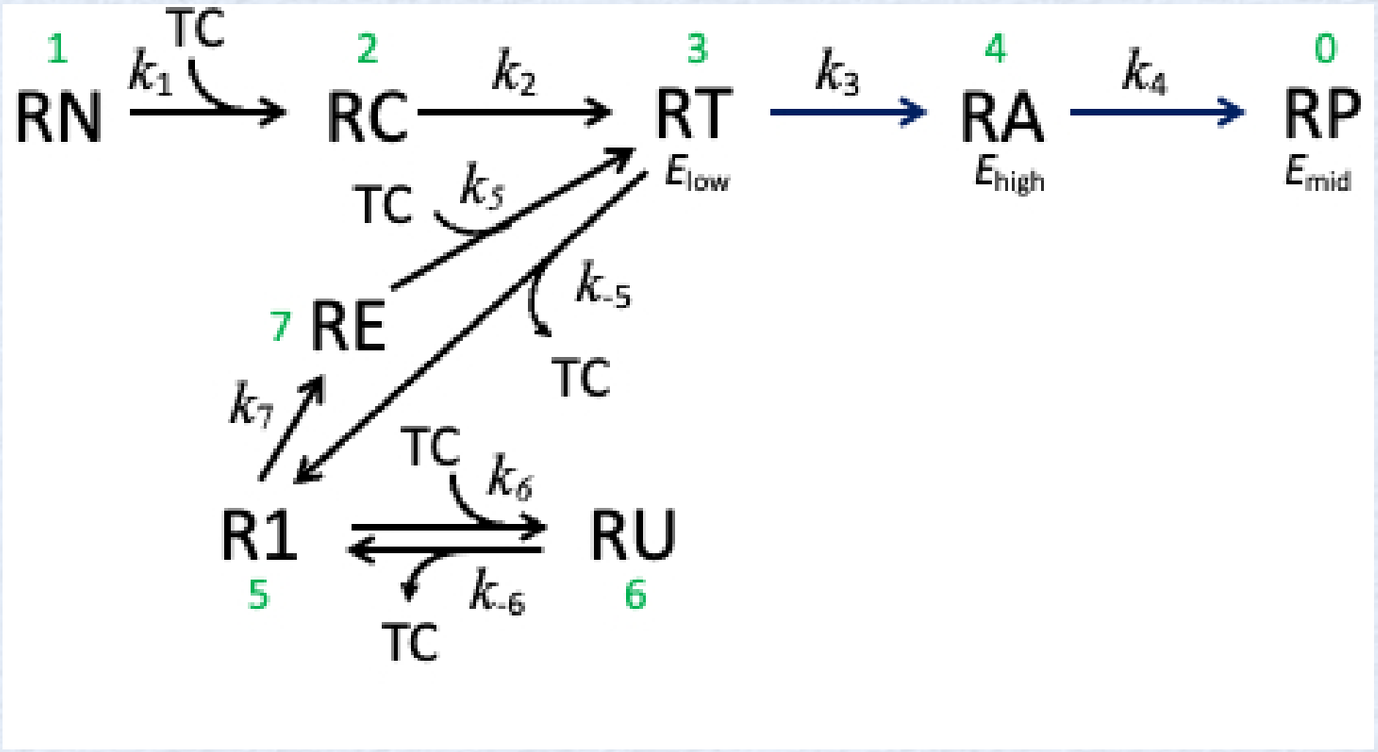
Both of those predictions were found to be false.

- The distribution of waiting times for near-cognate TC to bind the first time (single exponential) was *qualitatively different from subsequent times* (double exponential).
- The mean number of attempts before stable binding of near-cognate TC was an *increasing function of ternary complex concentration*.

Revised proposal for kinetic cycle

Instead of:



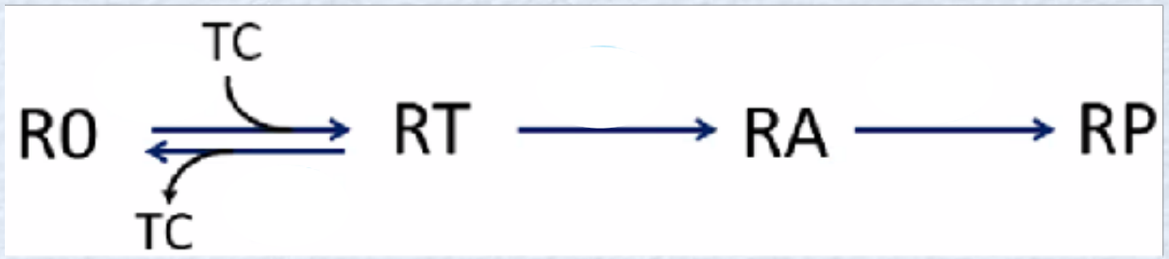


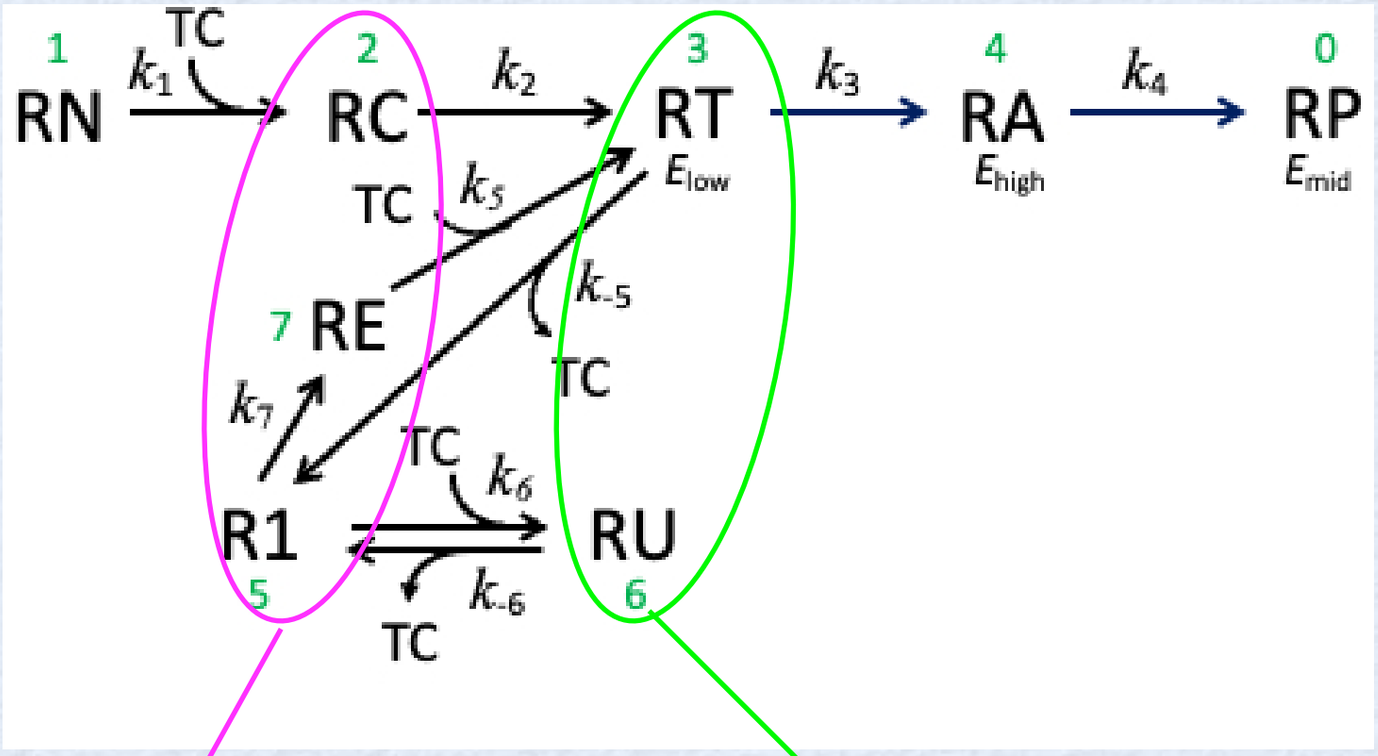
Hypothesize a new side-branch with a dead-end, as the main route for the tentatively bound ternary complex to be rejected from RT.

I won't attempt to argue for this model; I will try to work out its experimental signatures.

Revised proposal for kinetic cycle

Instead of:





zero-FRET (unbound) states

higher-FRET (bound) states

Hypothesize a new side-branch with a dead-end, as the main route for the tentatively bound ternary complex to be rejected from RT.

I won't attempt to argue for this model; I will try to work out its experimental signatures.

On fitting

PN, *Physical models of living systems 2nd ed*. (2022).

On fitting

The power of single-molecule methods is that instead of ensemble-averaged data, we have the *actual duration of every* binding event, thousands of them (and similarly unbinding events).

PN, *Physical models of living systems 2nd ed*. (2022).

On fitting

The power of single-molecule methods is that instead of ensemble-averaged data, we have the *actual duration of every* binding event, thousands of them (and similarly unbinding events).

PN, *Physical models of living systems 2nd ed*. (2022).

On fitting

The power of single-molecule methods is that instead of ensemble-averaged data, we have the *actual duration of every* binding event, thousands of them (and similarly unbinding events).

If we can get a model to predict the probability density function of those durations in terms of a few parameters, then we can compute the likelihood of an experimental dataset in terms of those parameters, that is, the probability that the data we *did* observe *would have been* observed in a world with certain values of the parameters.

PN, *Physical models of living systems 2nd ed*. (2022).

On fitting

The power of single-molecule methods is that instead of ensemble-averaged data, we have the *actual duration of every* binding event, thousands of them (and similarly unbinding events).

If we can get a model to predict the probability density function of those durations in terms of a few parameters, then we can compute the likelihood of an experimental dataset in terms of those parameters, that is, the probability that the data we *did* observe *would have been* observed in a world with certain values of the parameters.

PN, *Physical models of living systems 2nd ed*. (2022).

On fitting

The power of single-molecule methods is that instead of ensemble-averaged data, we have the *actual duration of every* binding event, thousands of them (and similarly unbinding events).

If we can get a model to predict the probability density function of those durations in terms of a few parameters, then we can compute the likelihood of an experimental dataset in terms of those parameters, that is, the probability that the data we *did* observe *would have been* observed in a world with certain values of the parameters.

Then we maximize over parameters, holding the data fixed, to get the values best supported by the data.

PN, *Physical models of living systems 2nd ed*. (2022).

On fitting

The power of single-molecule methods is that instead of ensemble-averaged data, we have the *actual duration of every* binding event, thousands of them (and similarly unbinding events).

If we can get a model to predict the probability density function of those durations in terms of a few parameters, then we can compute the likelihood of an experimental dataset in terms of those parameters, that is, the probability that the data we *did* observe *would have been* observed in a world with certain values of the parameters.

Then we maximize over parameters, holding the data fixed, to get the values best supported by the data.

PN, *Physical models of living systems 2nd ed*. (2022).

On fitting

The power of single-molecule methods is that instead of ensemble-averaged data, we have the *actual duration of every* binding event, thousands of them (and similarly unbinding events).

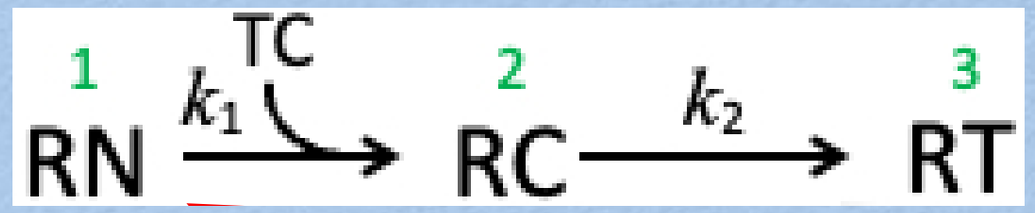
If we can get a model to predict the probability density function of those durations in terms of a few parameters, then we can compute the likelihood of an experimental dataset in terms of those parameters, that is, the probability that the data we *did* observe *would have been* observed in a world with certain values of the parameters.

Then we maximize over parameters, holding the data fixed, to get the values best supported by the data.

It's easy in the familiar case of a simple mass-action binding model, in which the probability per time for binding ternary complex is a rate constant times the concentration.

PN, *Physical models of living systems 2nd ed*. (2022).

Initial binding



In the model, first binding is a Michaelis–Menten type process. Luckily smart people have already worked out the PDF of completion times: Let

$$B' = (\kappa_1 + k_{-1} + k_2)/2,\ A = \sqrt{(B')^2 - \kappa_1 k_2}.$$

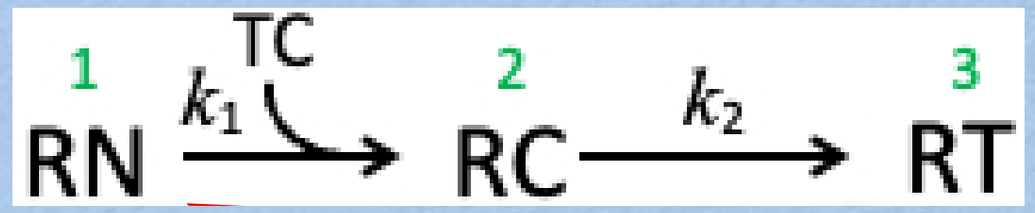
$$\wp(t_1) = \frac{\kappa_1 k_2}{2A} \mathrm{e}^{(A-B')t_1} \left(1 - \mathrm{e}^{-2At_1}\right)$$

Kou et al, J. Phys. Chem. B 2005, 109, 19068--19081.

This distribution can be used to define a likelihood function that determines k_1 and k_2 from t_1 data. We have several sets of t_1 values, each with a different, but known, [TC].

[*First unbinding* is easier—no concentration dependence.]

Initial binding



In the model, first binding is a Michaelis–Menten type process. Luckily smart people have already worked out the PDF of completion times: Let

$$B' = (\kappa_1 + k_{-1} + k_2)/2,\ A = \sqrt{(B')^2 - \kappa_1 k_2}.$$

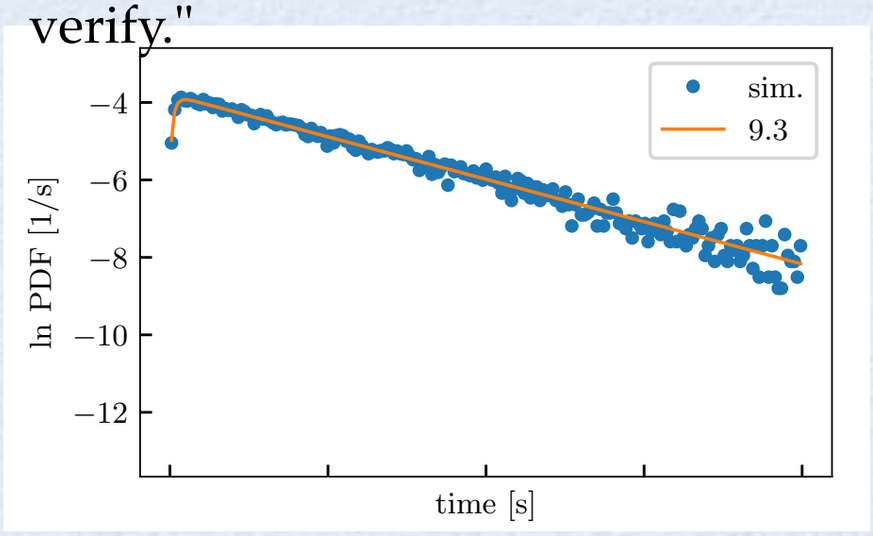
$$\wp(t_1) = \frac{\kappa_1 k_2}{2A} \mathrm{e}^{(A-B')t_1} \left(1 - \mathrm{e}^{-2At_1}\right)$$

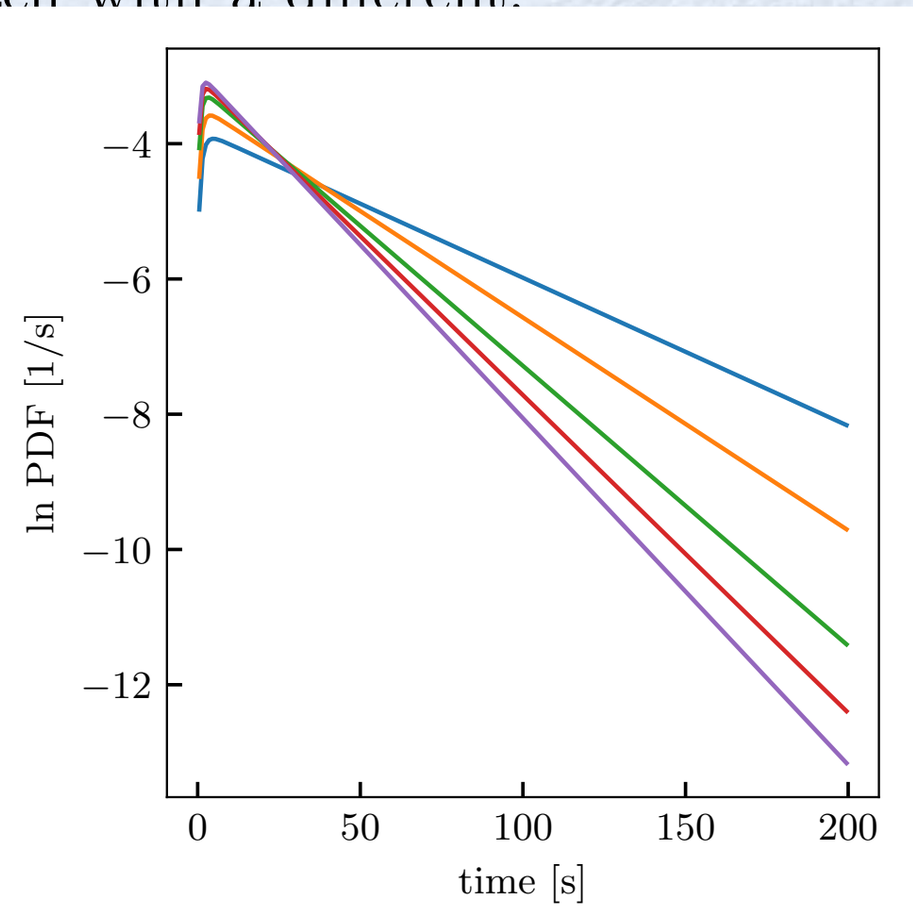
Kou et al, J. Phys. Chem. B 2005, 109, 19068--19081.

This distribution can be used to define a likelihood function that determines k_1 and k_2 from t_1 data. We have several sets of t_1 values, each with a different, but known, [TC].

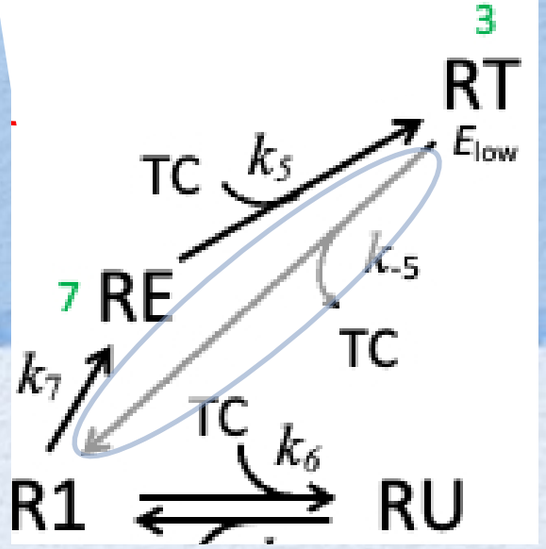
[*First unbinding* is easier—no concentration dependence.]

Right: This PDF is shown for various [TC] and illustrative *k* values.
Below: It's easy to confirm the result by simulation. "Trust but verify."

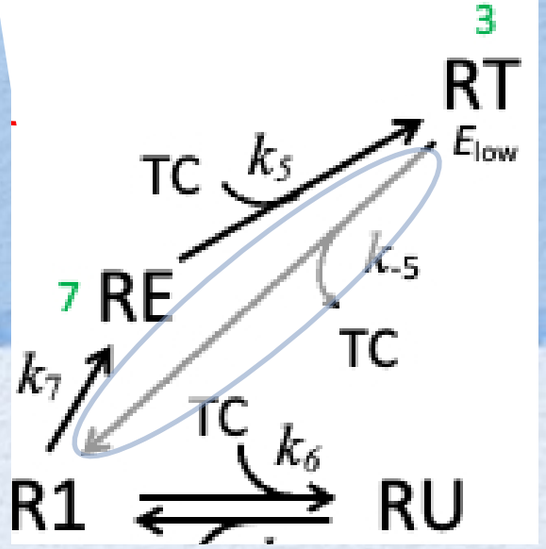




[Subsequent binding]

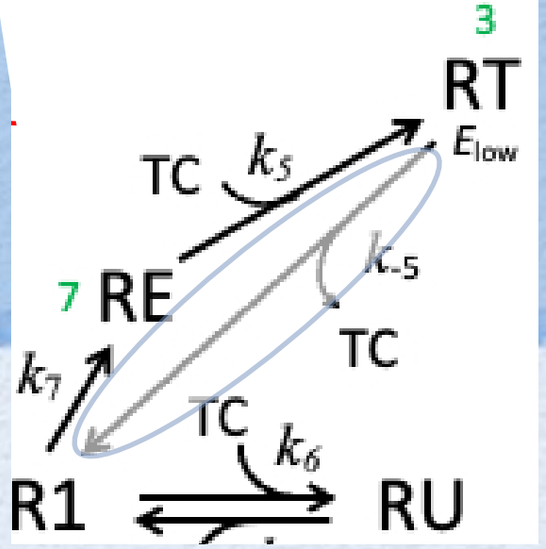


[Subsequent binding]



Every unbinding brings us to state R1. Because RT and RU are both high-FRET states, we want the distribution of the first-passage time to *either one*. This one I had to work out for myself.

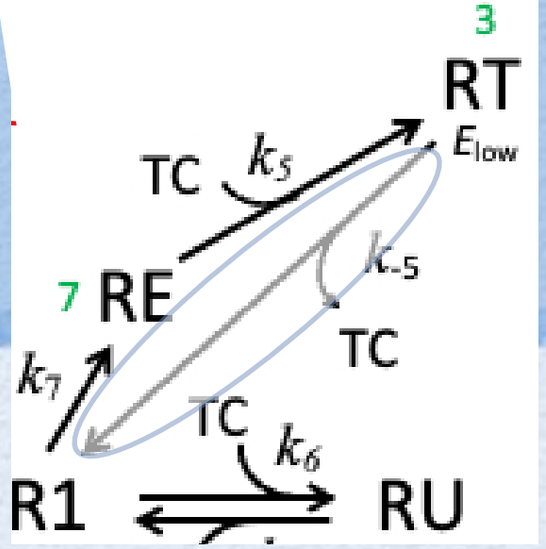
[Subsequent binding]



Every unbinding brings us to state R1. Because RT and RU are both high-FRET states, we want the distribution of the first-passage time to *either one*. This one I had to work out for myself.

This time the formulas are too long to decently display, but it comes down to convolving two steps for the upper pathway, then finding probability per unit time for first arrival at either R1 or RU, given that the event is "sampling," that is, known to not be the final binding.

[Subsequent binding]



Every unbinding brings us to state R1. Because RT and RU are both high-FRET states, we want the distribution of the first-passage time to *either one*. This one I had to work out for myself.

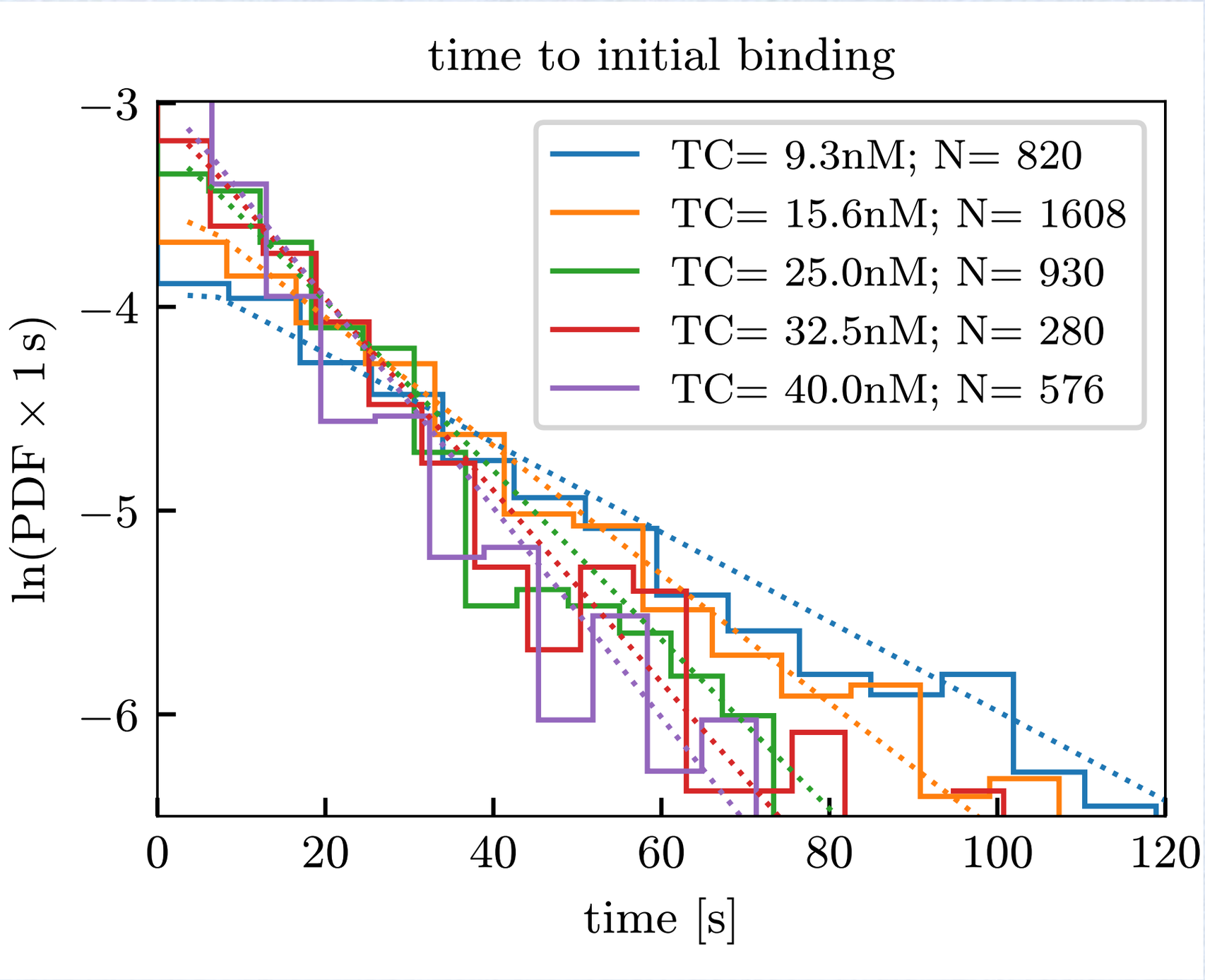
This time the formulas are too long to decently display, but it comes down to convolving two steps for the upper pathway, then finding probability per unit time for first arrival at either R1 or RU, given that the event is "sampling," that is, known to not be the final binding.

We are using *all* the data in the likelihood function. If we wish, we can then look at reduced statistics to get a human-viewable look at some aspects of our fit.

Results: First-binding waiting time

So great – we got our model to divulge its PDF. Can the model actually fit real experimental data? It's a tall order – lots of data, just a few fit parameters to get a global fit. – highly overdetermined, which means highly falsifiable.

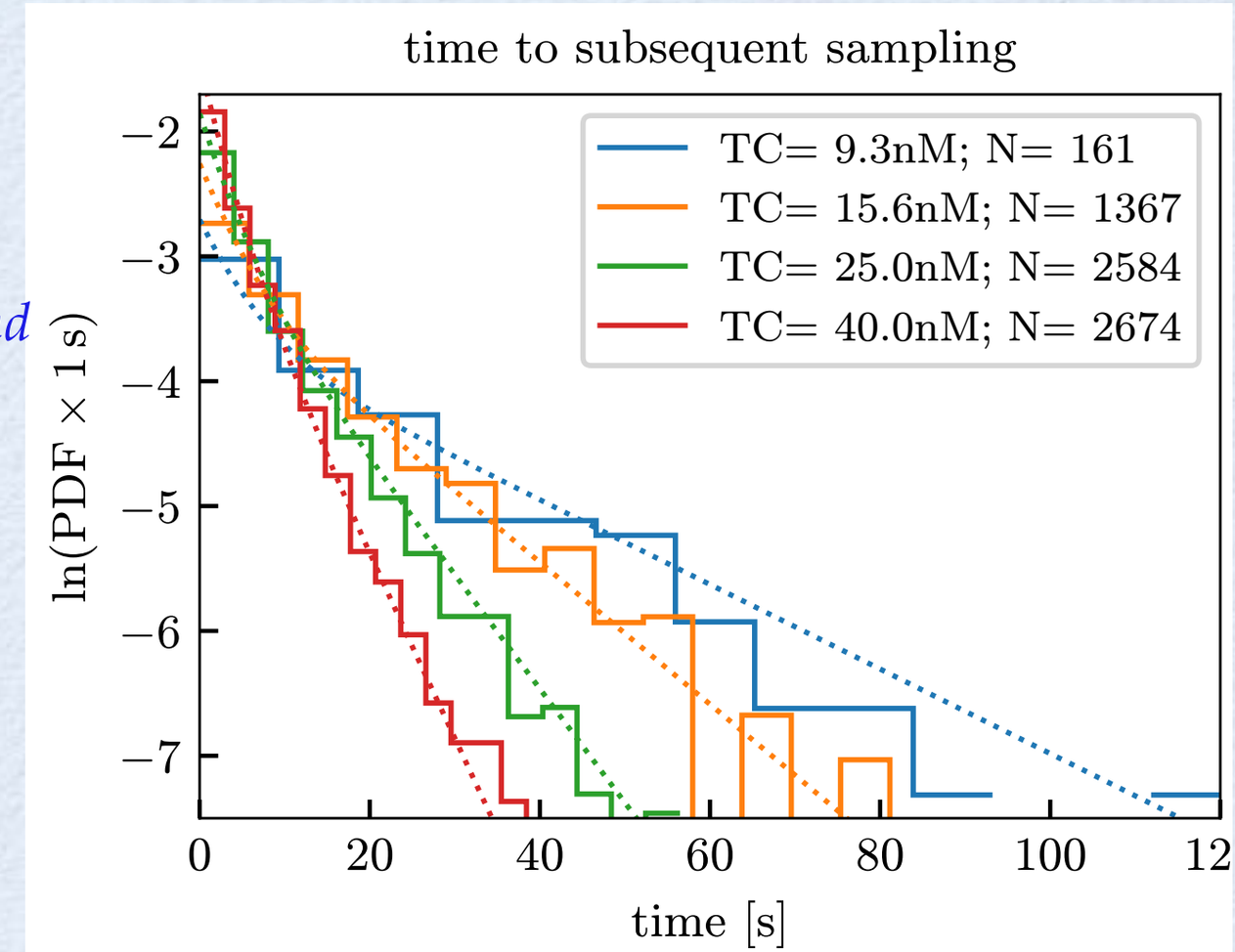
Initial binding looks pretty good and determines some rate constants.



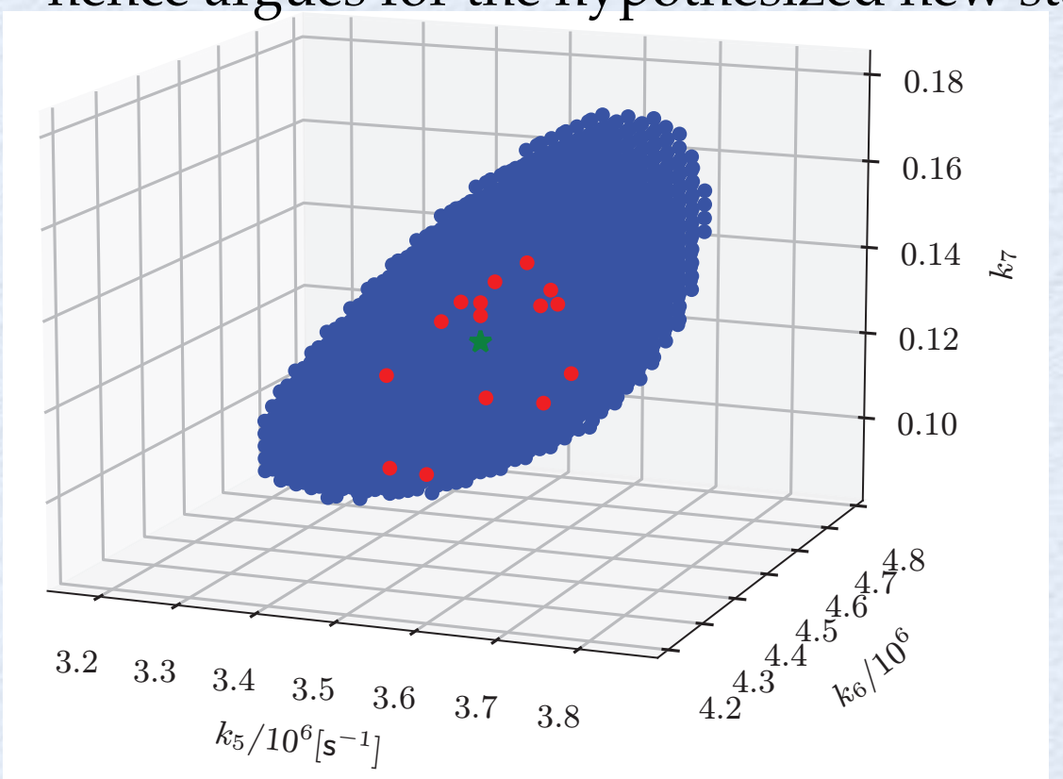
Clark Fritsch, et al., submitted 2023

Results: Subsequent binding

It's good to have a lot of data, so that we can see deep into the telltale tails of the distributions – the transition from one exponential to the other.
Also highly overdetermined, also looks pretty good, and determines more rate constants (green star below).



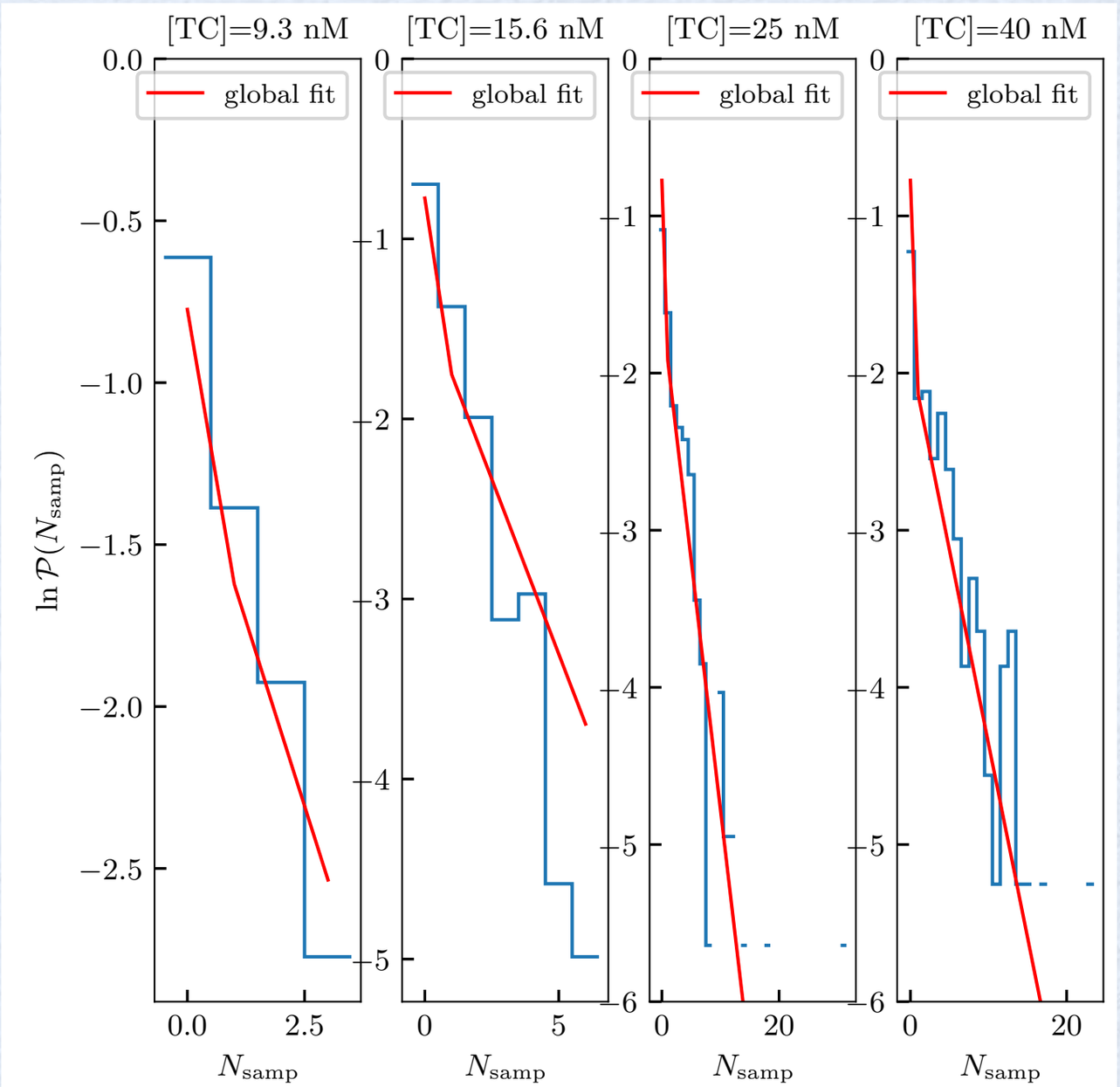
Moreover, bootstrap replicates of the experimental data (red) define a cloud of credible rate values that *excludes infinity* and hence argues for the hypothesized new state:



Clark Fritsch, et al., submitted 2023

An acid test

But once we've found our best version of the model, can it *also* explain *other, different* phenomena that it *wasn't* trained on? We asked it to predict in detail the probability distribution for the number of attempts before stable binding. With no additional adjustment we got:



Clark Fritsch, et al., submitted 2023

Part 4: Summary

Let's return to the the qualitative observed, surprising, phenomena that motivated the model:

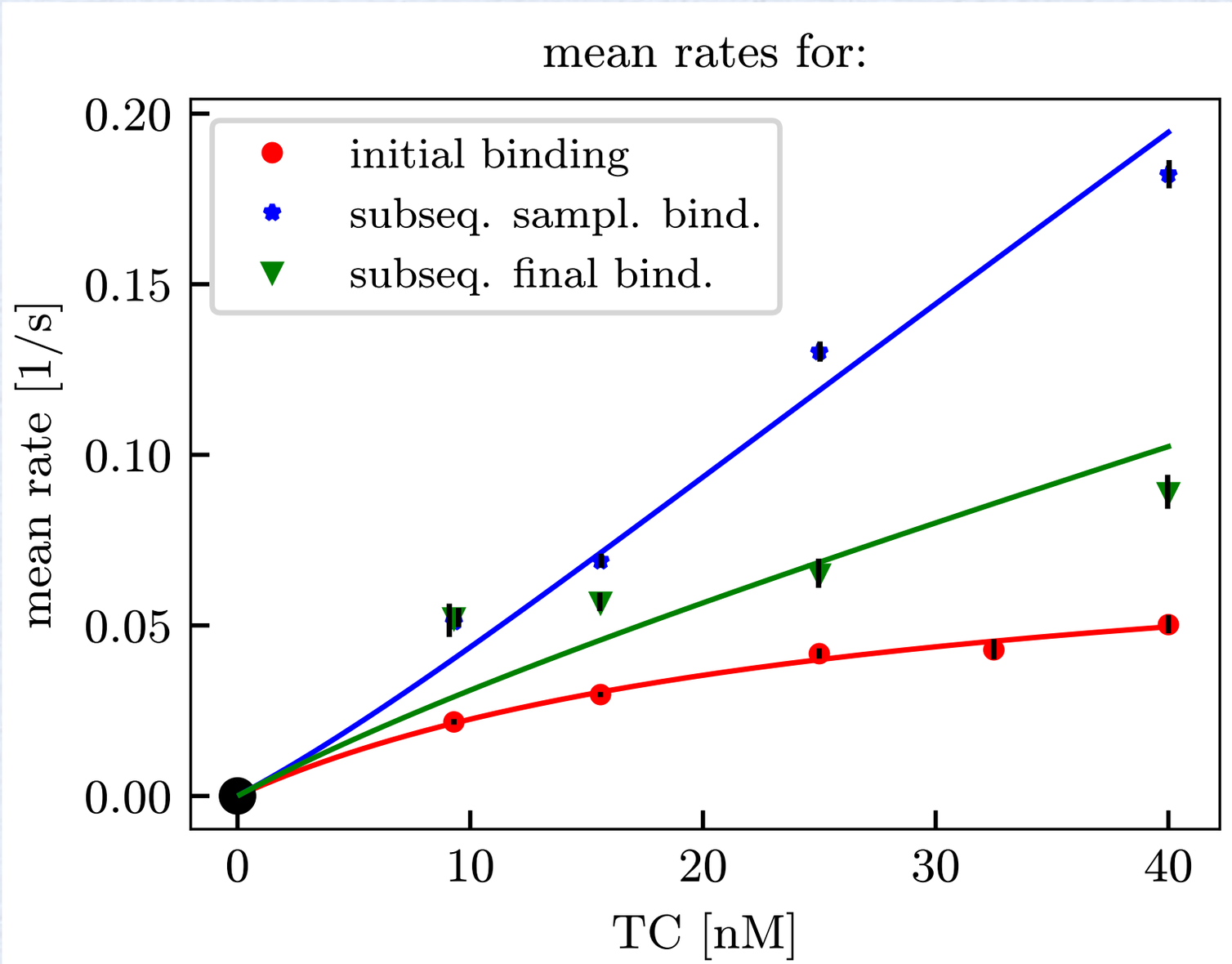
- The distribution of waiting times for near-cognate TC to bind the first time was different from subsequent times. **<-- looks pretty good**

Clark Fritsch, et al., submitted 2023

Part 4: Summary

Let's return to the the qualitative observed, surprising, phenomena that motivated the model:

- The distribution of waiting times for near-cognate TC to bind the first time was different from subsequent times. **<-- looks pretty good**



Clark Fritsch, et al., submitted 2023

Part 5

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

Part 5

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We've can improve noisy images by aligning and averaging,

Part 5

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We've can improve noisy images by aligning and averaging,
and

Part 5

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We've can improve noisy images by aligning and averaging,
and
that procedure has an impeccable probabilistic foundation,

Part 5

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We've can improve noisy images by aligning and averaging,
and
that procedure has an impeccable probabilistic foundation,
but

Part 5

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We've can improve noisy images by aligning and averaging,
and
that procedure has an impeccable probabilistic foundation,
but
it's not obvious how to align images at ultra-low SNR,

Part 5

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We've can improve noisy images by aligning and averaging,
and
that procedure has an impeccable probabilistic foundation,
but
it's not obvious how to align images at ultra-low SNR,
so

Part 5

1. Inference
2. Superresolution
3. Changepoint
4. Ribosome
5. CryoEM

We've can improve noisy images by aligning and averaging,
and
that procedure has an impeccable probabilistic foundation,
but
it's not obvious how to align images at ultra-low SNR,
so
we need to deploy slightly heavier artillery, following F. Sigworth 1999.

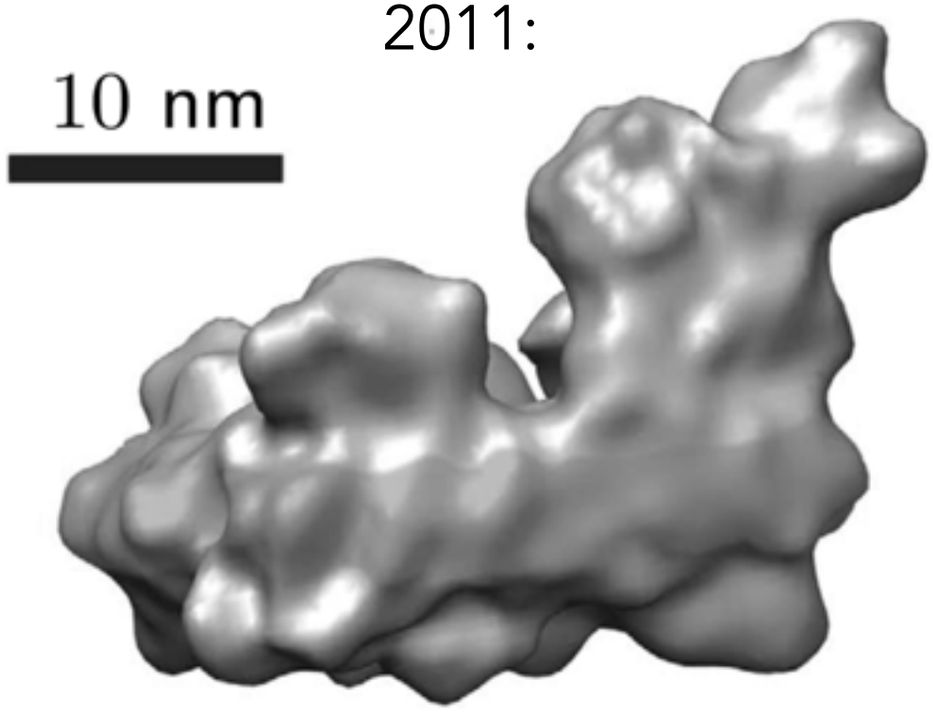
"The revolution will not be crystallized"

2011:

https://www.nobelprize.org/prizes/chemistry/2017/advanced-information/

"The revolution will not be crystallized"

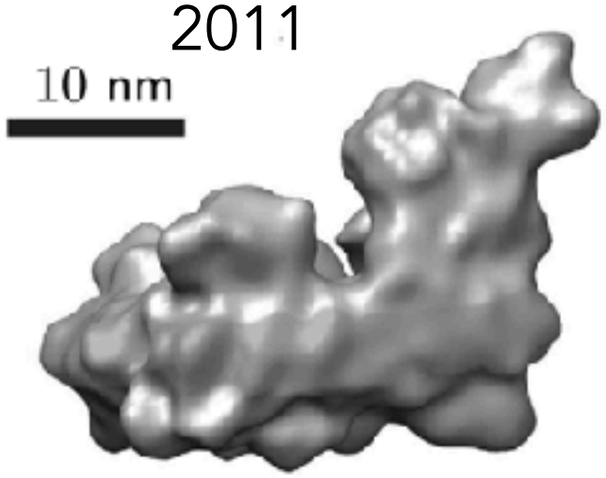
2011:



https://www.nobelprize.org/prizes/chemistry/2017/advanced-information/

"The revolution will not be crystallized"

2011



2016

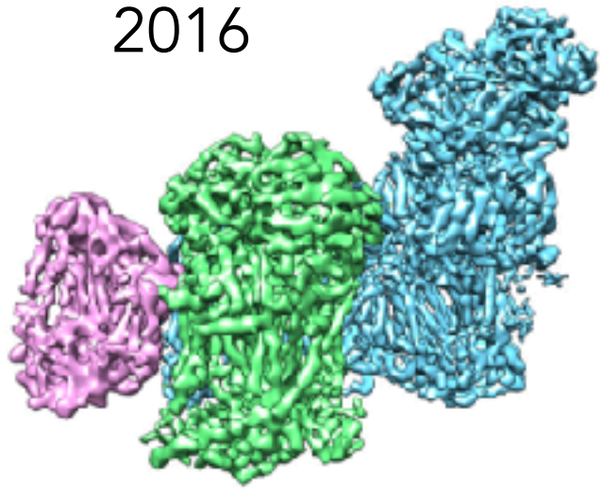
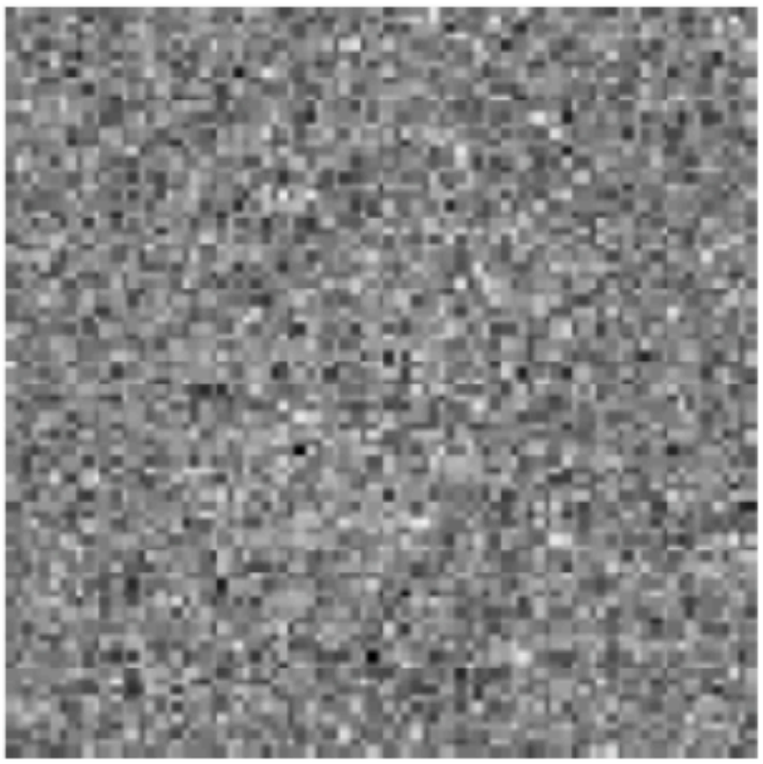
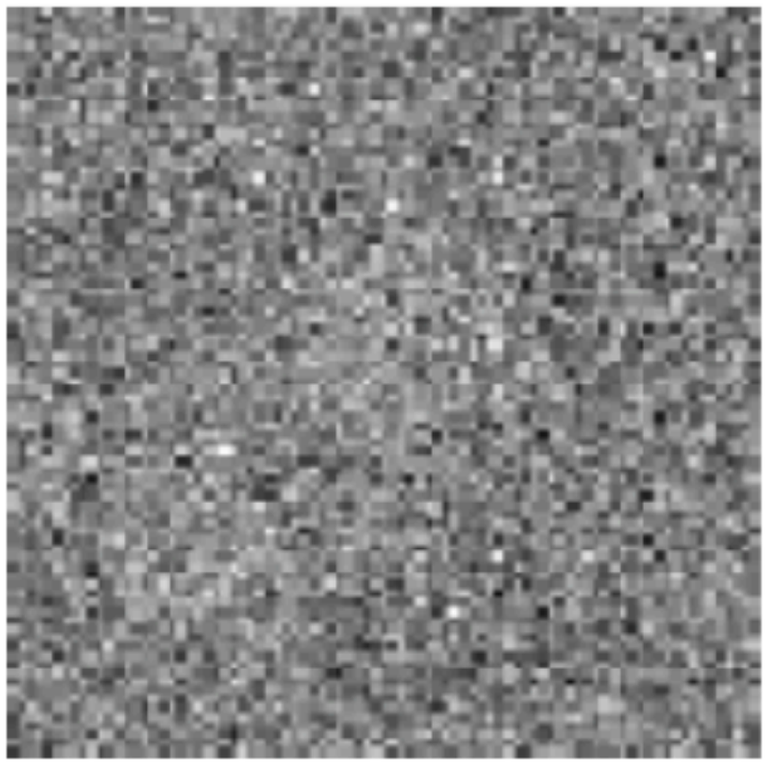


Figure 12.2: [Reconstructions from electron micrographs.] (a) The electron-transport chain components in a mitochondrial supercomplex $I_1III_2IV_1$, as determined in 2011 (b) The same complex as determined 2016. Subcomplexes I, III and IV are shown in blue, green and pink, respectively. [(a): From Althoff et al., 2011. (b): From Letts et al., 2016.]

https://www.nobelprize.org/prizes/chemistry/2017/advanced-information/

A challenge


When you try to image a *single molecule*, naturally your contrast is very low– lots of background.
Here are 3 examples of the raw images taken from the thousands in a typical cryo-EM setup. There's something hiding here. It's not a tiger, but we still need to find it.

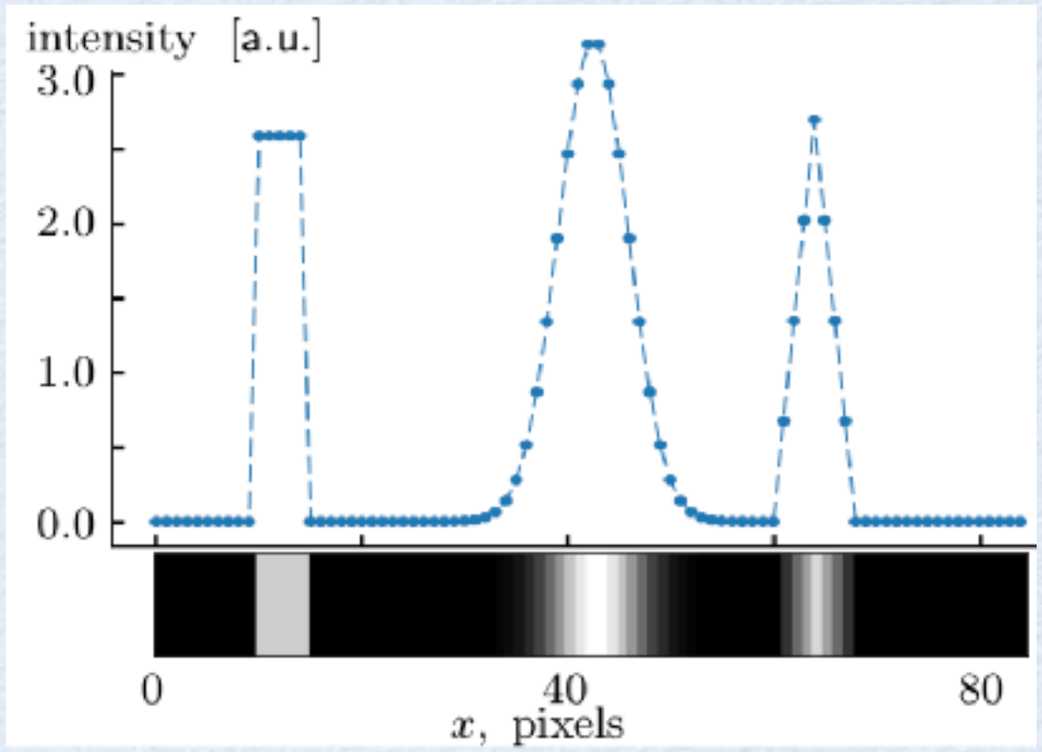
Scheres, S. H. W., et al. (2005). Journal of Molecular Biology, 348(1), 139–149. http://doi.org/10.1016/j.jmb.2005.02.031

1d Warmup

Here are three "objects": First with sharp edges, second and third with softer edges:

As intensity:

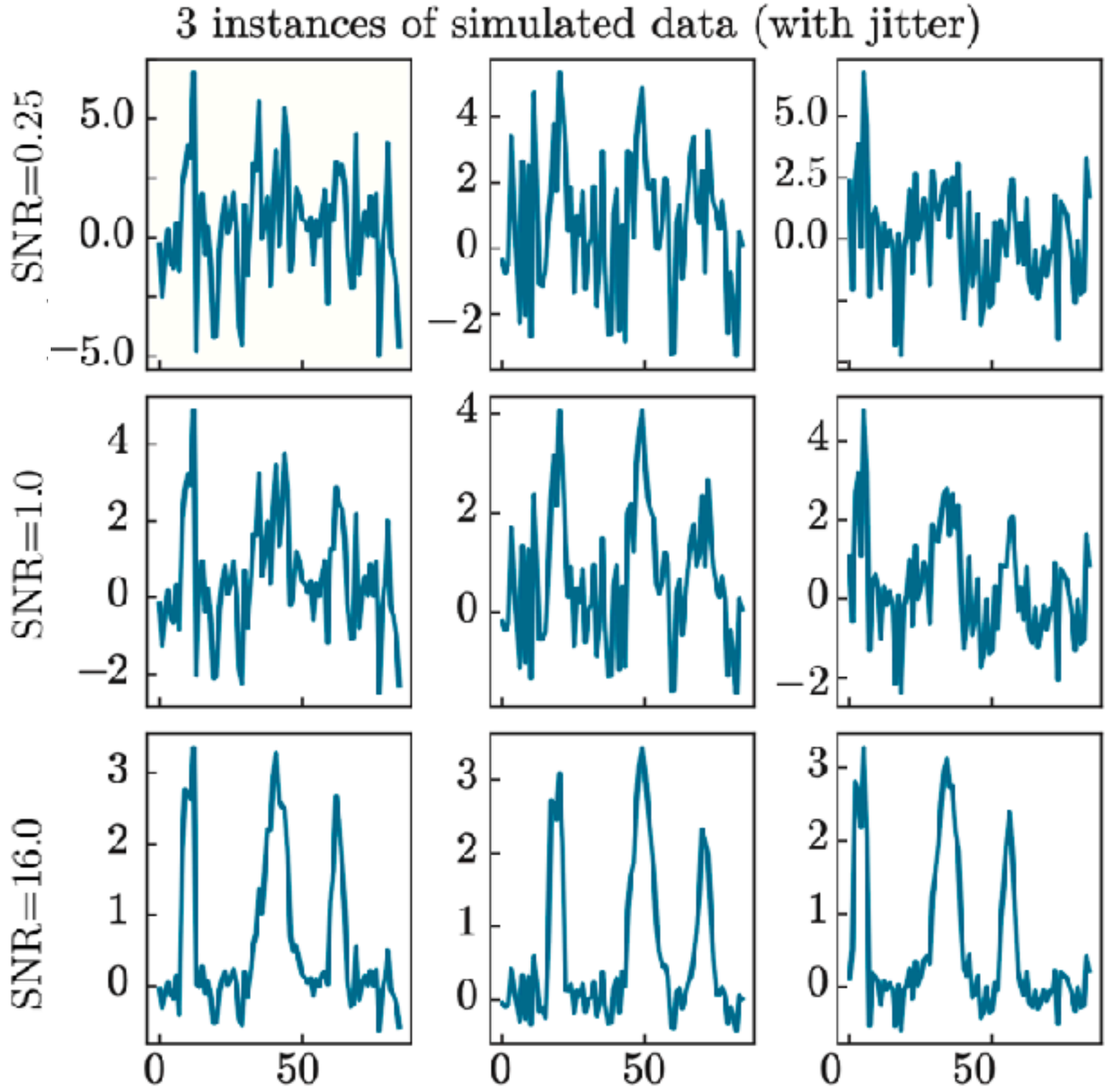
As raster:



PN, *Physical models of living systems* (2/e, 2022)

Reasonable but doesn't work (1D)

"If you've got many imprecise measurements, average them." (Wisdom of crowds)

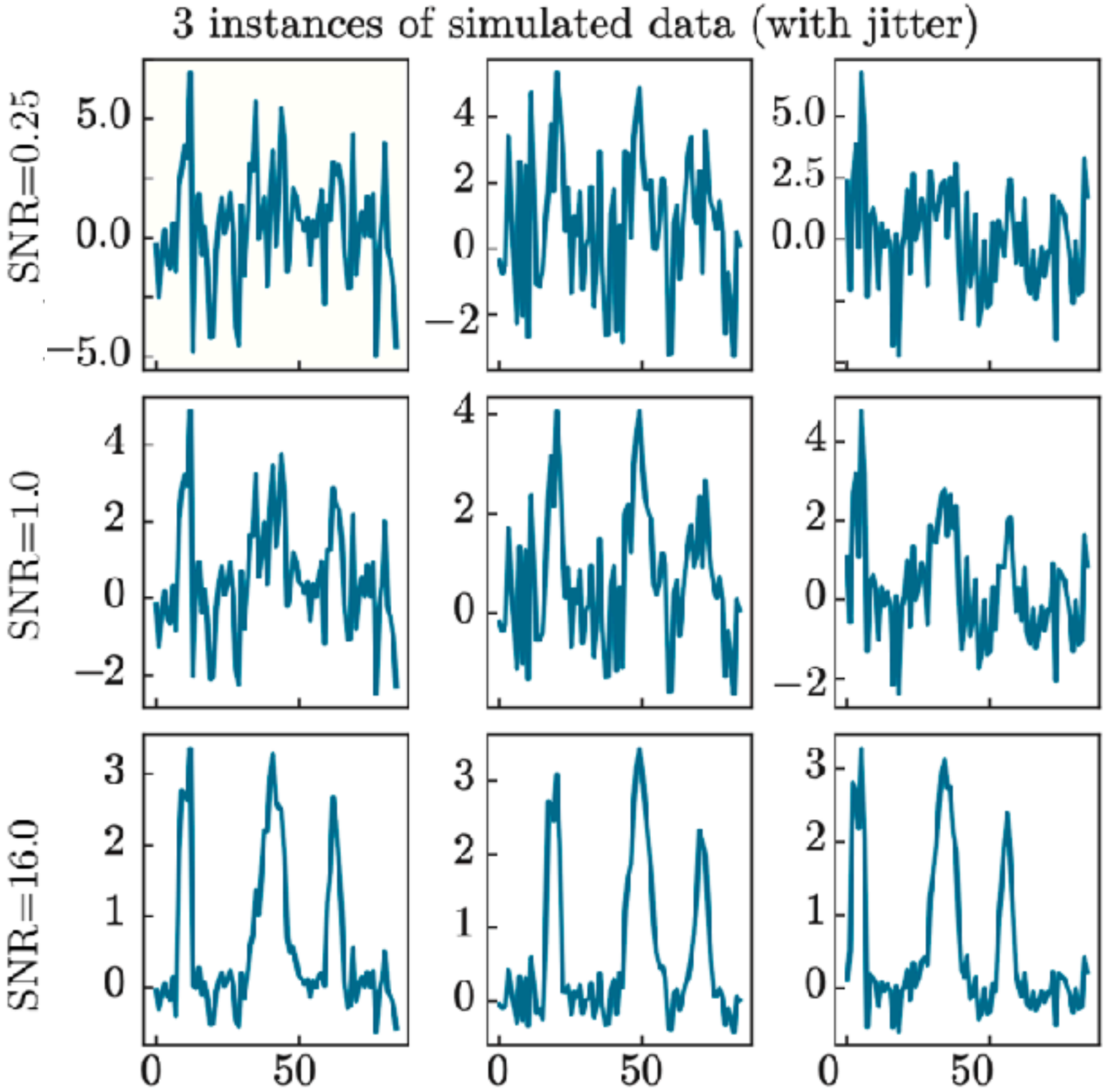


PN, *Physical models of living systems* (2/e, 2022)

Reasonable but doesn't work (1D)

"If you've got many imprecise measurements, average them." (Wisdom of crowds)

Even averaging over 1500 instances doesn't help enough. It's a chicken/egg problem. To reduce noise by averaging, we must first align the samples. But to align the samples, we must first reduce the noise!

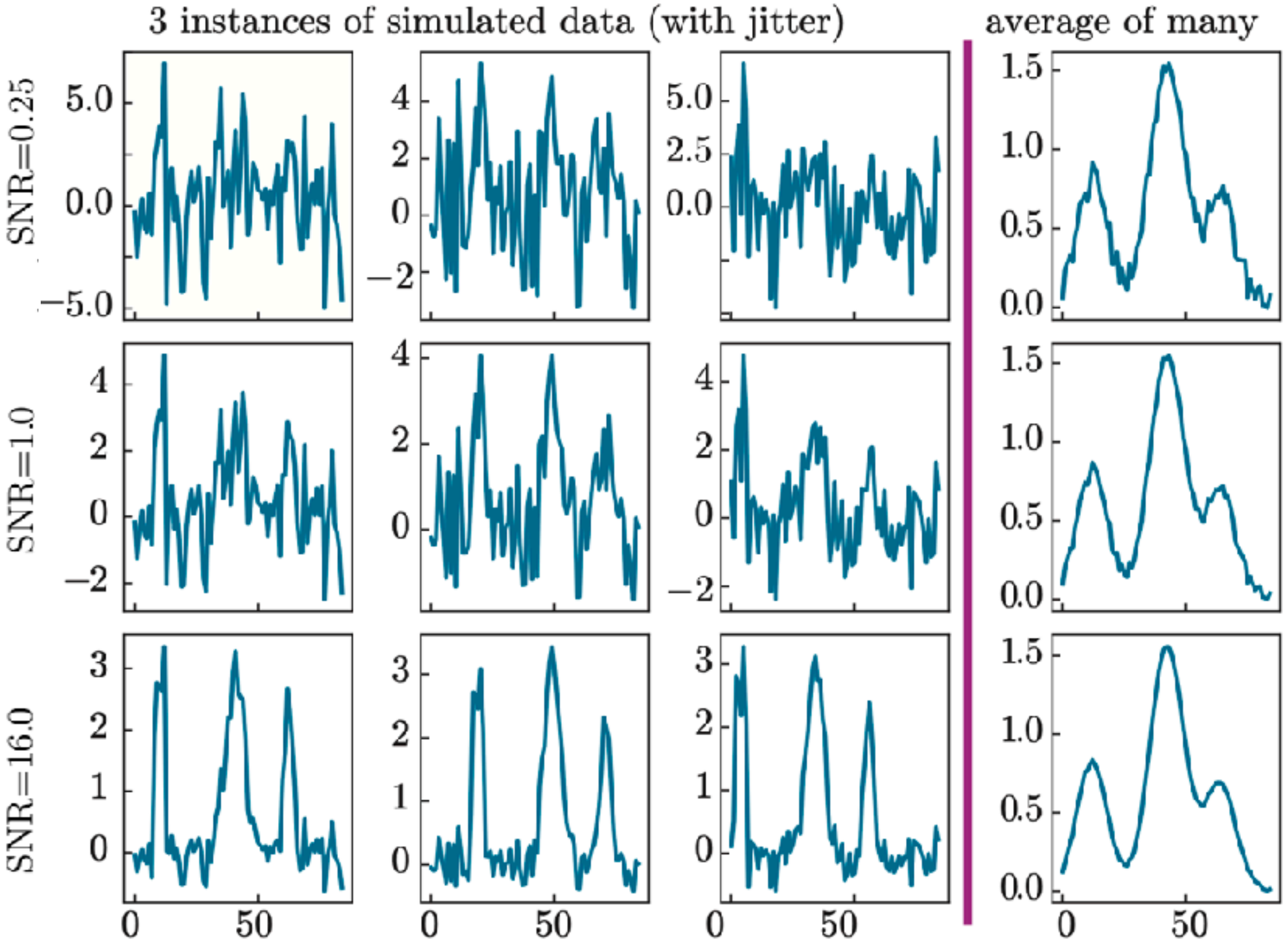


PN, *Physical models of living systems* (2/e, 2022)

Reasonable but doesn't work (1D)

"If you've got many imprecise measurements, average them." (Wisdom of crowds)

Even averaging over 1500 instances doesn't help enough. It's a chicken/egg problem. To reduce noise by averaging, we must first align the samples. But to align the samples, we must first reduce the noise!



PN, *Physical models of living systems* (2/e, 2022)

Unfair advantage

Unfair advantage

One approach is to throw it all at some huge deep neural net and hope something useful comes out. That's "data-driven" approach, and it's all the rage.

Unfair advantage

One approach is to throw it all at some huge deep neural net and hope something useful comes out. That's "data-driven" approach, and it's all the rage.

But we can do much better if we know something about the biology and physics of *what generated those data*. Let's exploit any unfair advantage we may have. What we know is that:

Unfair advantage

One approach is to throw it all at some huge deep neural net and hope something useful comes out. That's "data-driven" approach, and it's all the rage.

But we can do much better if we know something about the biology and physics of *what generated those data*. Let's exploit any unfair advantage we may have. What we know is that:

- Each pixel has shot noise *independent of other pixels*.

Unfair advantage

One approach is to throw it all at some huge deep neural net and hope something useful comes out. That's "data-driven" approach, and it's all the rage.

But we can do much better if we know something about the biology and physics of *what generated those data*. Let's exploit any unfair advantage we may have. What we know is that:

- Each pixel has shot noise *independent of other pixels*.
- Each image has been *rigidly translated* by an unknown amount relative to every other image but otherwise represents the same PDF for electron arrivals.

Unfair advantage

One approach is to throw it all at some huge deep neural net and hope something useful comes out. That's "data-driven" approach, and it's all the rage.

But we can do much better if we know something about the biology and physics of *what generated those data*. Let's exploit any unfair advantage we may have. What we know is that:

- Each pixel has shot noise *independent of other pixels*.
- Each image has been *rigidly translated* by an unknown amount relative to every other image but otherwise represents the same PDF for electron arrivals.

Why is this information so helpful? Because in 1D, the right shift to align each instance with the others is a *single number,* determined *globally by the entire image,* which means we have a lot of data to determine it to good accuracy.

Unfair advantage

One approach is to throw it all at some huge deep neural net and hope something useful comes out. That's "data-driven" approach, and it's all the rage.

But we can do much better if we know something about the biology and physics of *what generated those data*. Let's exploit any unfair advantage we may have. What we know is that:

- Each pixel has shot noise *independent of other pixels*.
- Each image has been *rigidly translated* by an unknown amount relative to every other image but otherwise represents the same PDF for electron arrivals.

Why is this information so helpful? Because in 1D, the right shift to align each instance with the others is a *single number,* determined *globally by the entire image,* which means we have a lot of data to determine it to good accuracy.

Unfair advantage

One approach is to throw it all at some huge deep neural net and hope something useful comes out. That's "data-driven" approach, and it's all the rage.

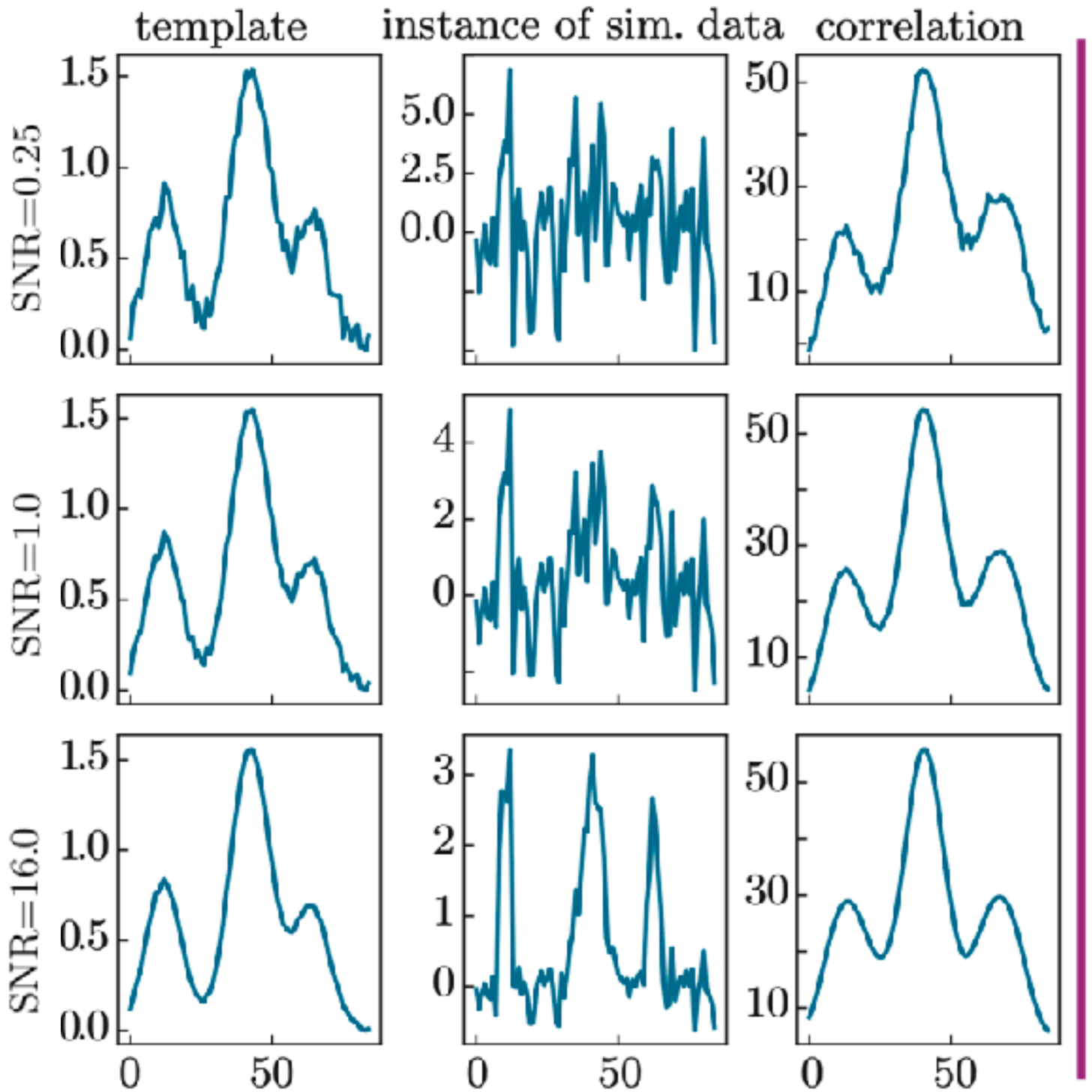
But we can do much better if we know something about the biology and physics of *what generated those data*. Let's exploit any unfair advantage we may have. What we know is that:

- Each pixel has shot noise *independent of other pixels*.
- Each image has been *rigidly translated* by an unknown amount relative to every other image but otherwise represents the same PDF for electron arrivals.

"Physics"

Why is this information so helpful? Because in 1D, the right shift to align each instance with the others is a *single number,* determined *globally by the entire image,* which means we have a lot of data to determine it to good accuracy.

Cross-correlation: 1D

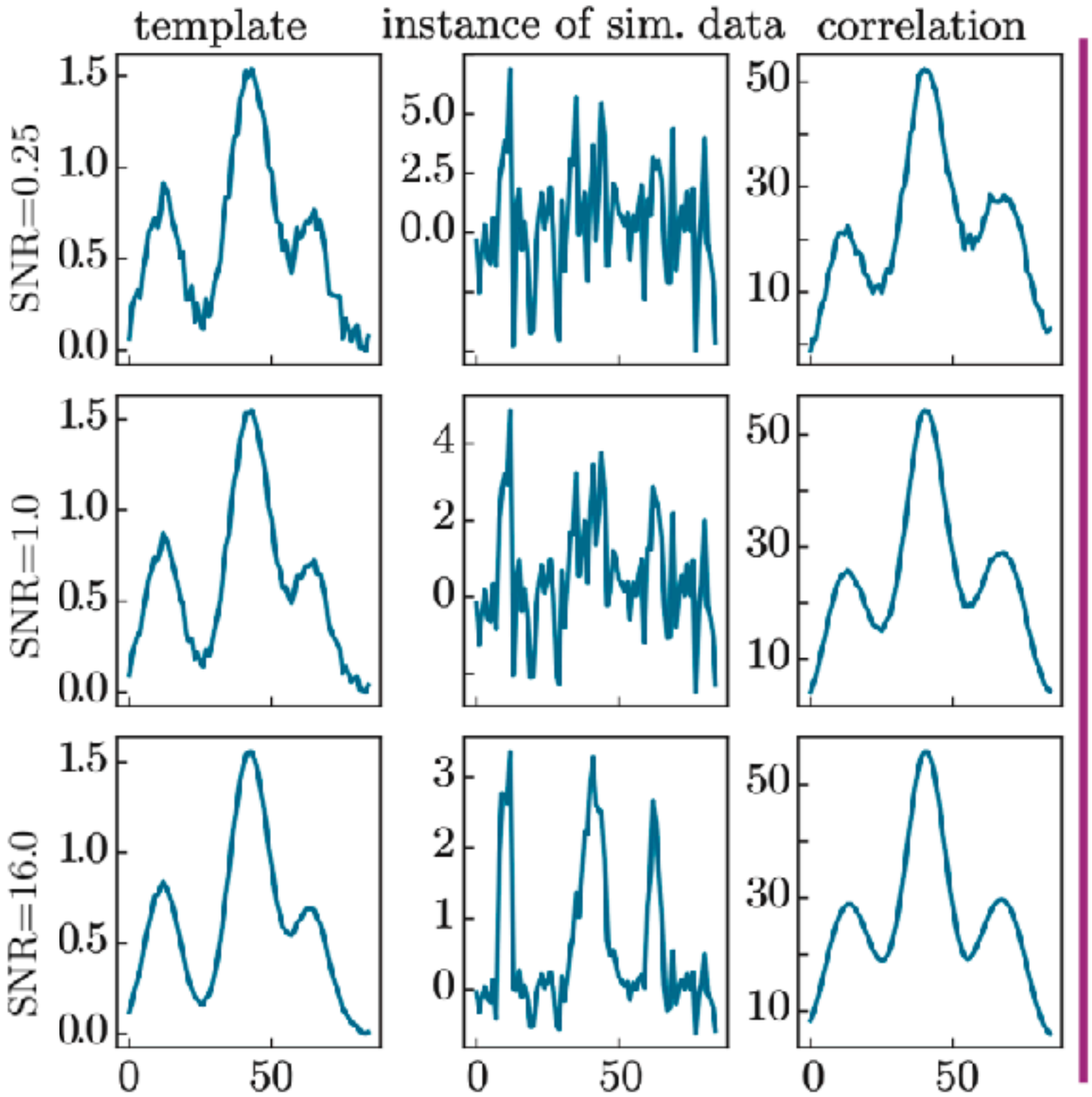


PN, *Physical models of living systems* (2/e, 2022)

Cross-correlation: 1D

The same 1500 instances as before.

We can translate each instance of the noisy data by an amount that optimizes its correlation with a "template," *then* average the instances point by point.
This is more successful than naive averaging, though still not great at low SNR (soon we will do better).

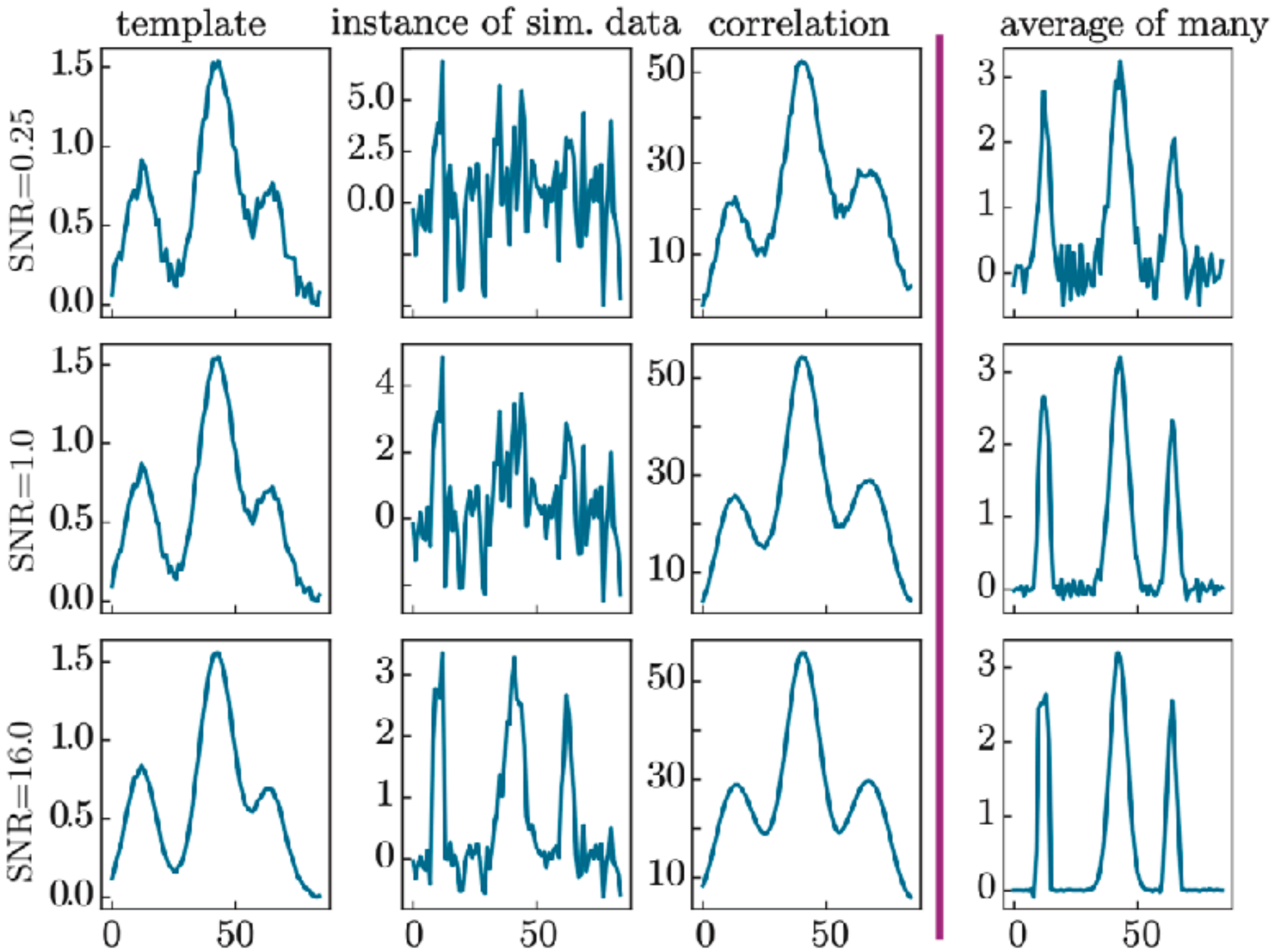


PN, *Physical models of living systems* (2/e, 2022)

Cross-correlation: 1D

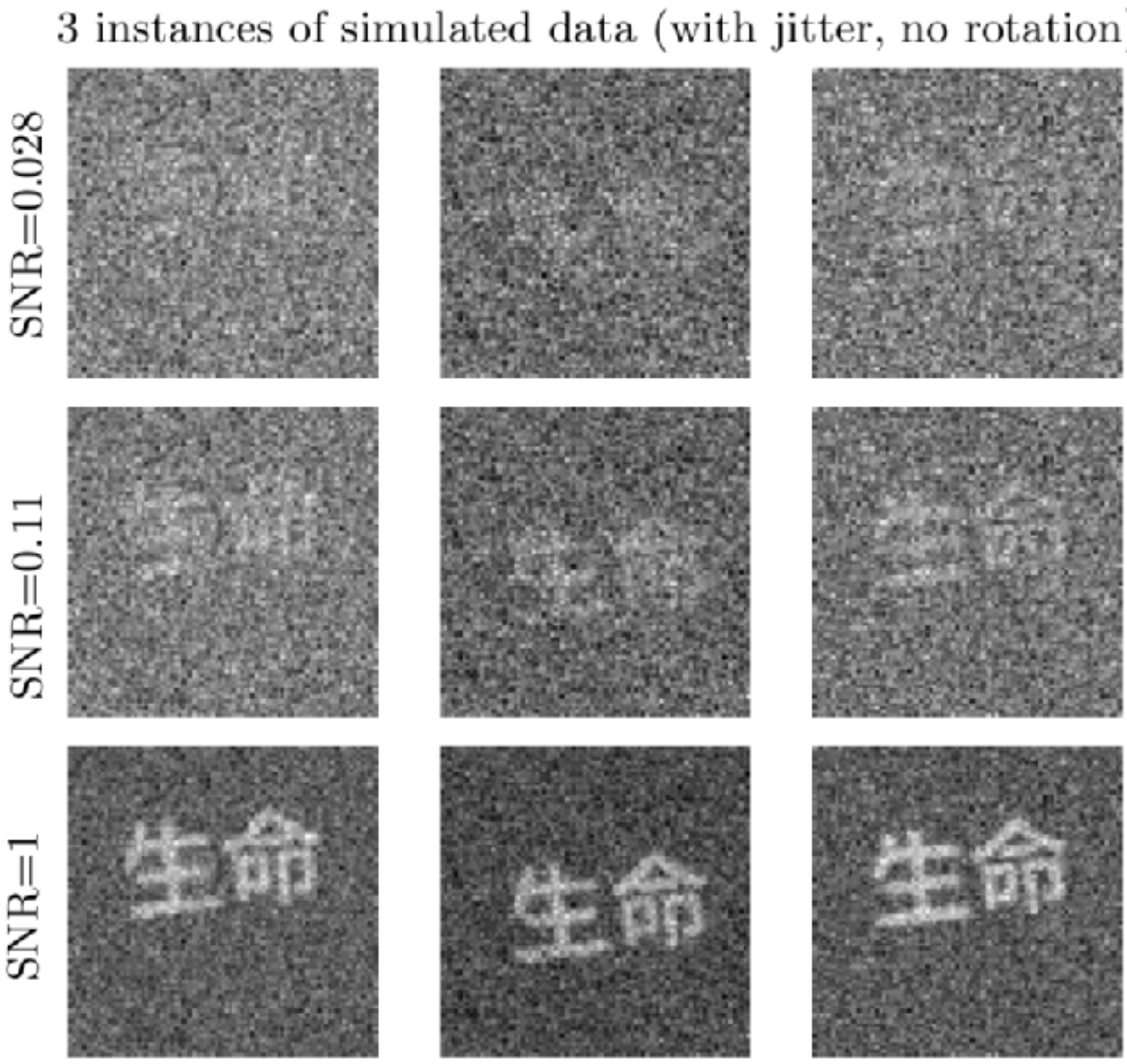
The same 1500 instances as before.

We can translate each instance of the noisy data by an amount that optimizes its correlation with a "template," *then* average the instances point by point.
This is more successful than naive averaging, though still not great at low SNR (soon we will do better).



PN, *Physical models of living systems* (2/e, 2022)

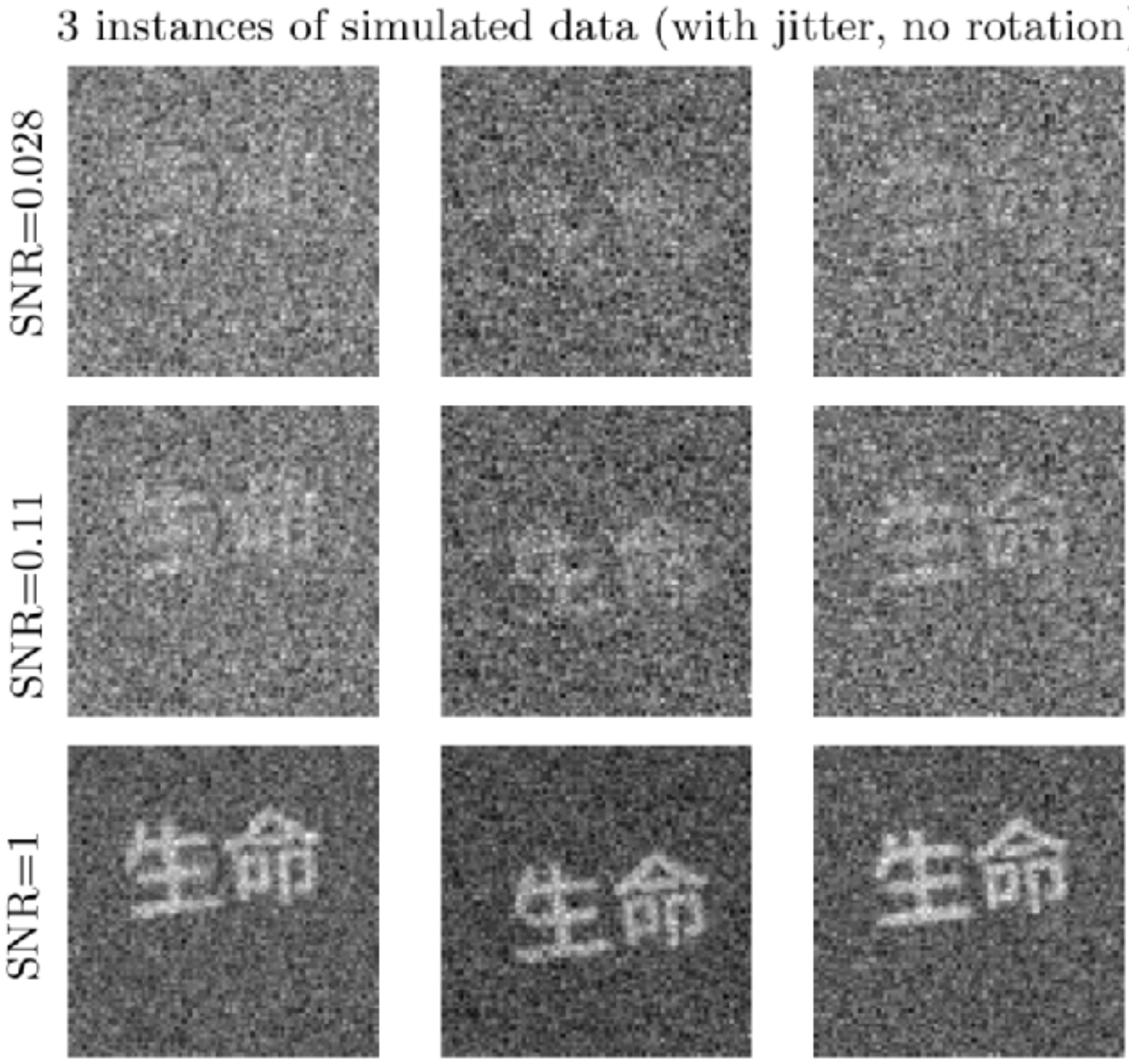
Same problem in 2D



PN, *Physical models of living systems* (2/e, 2022)

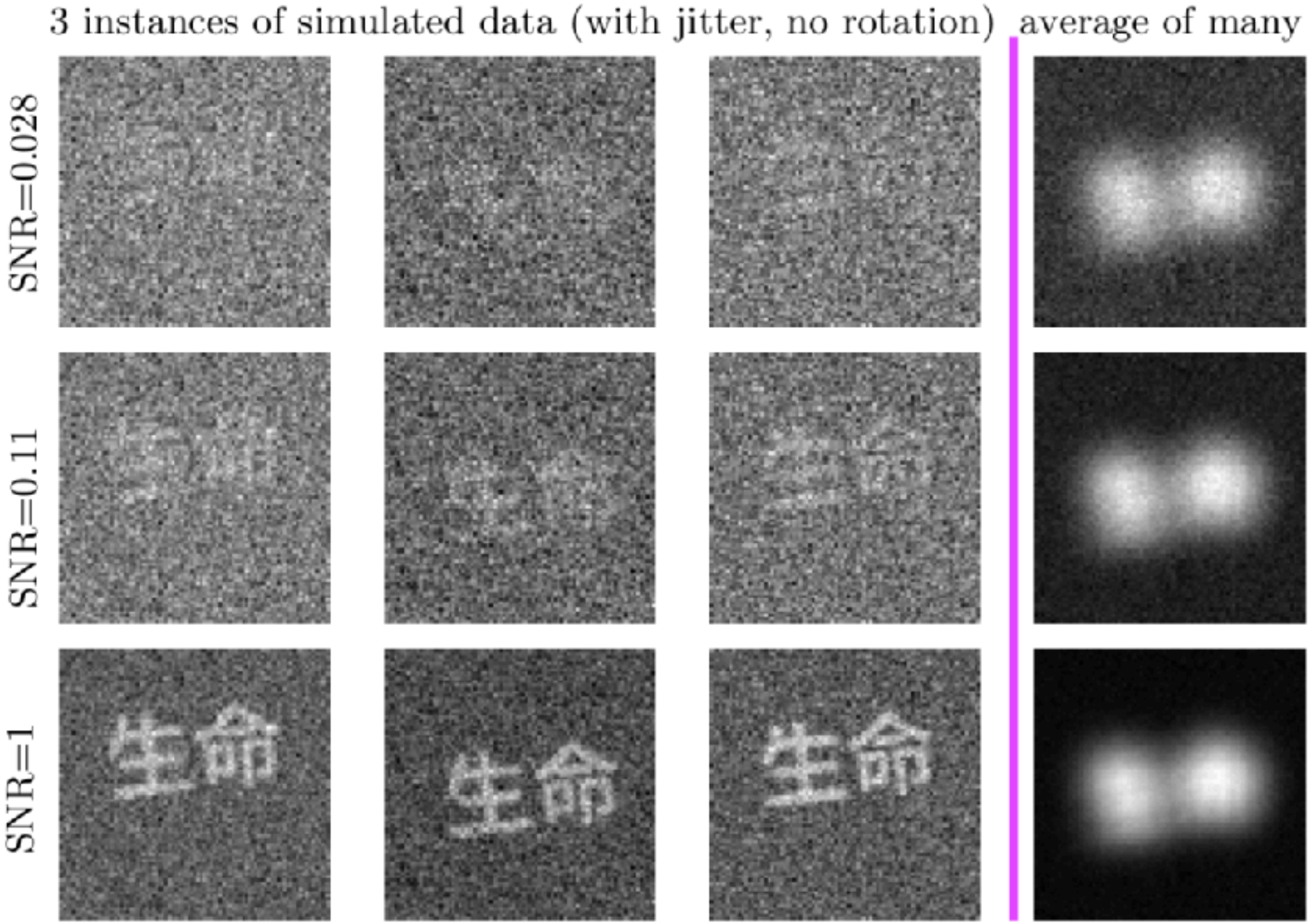
Same problem in 2D

Once again, jitter (random translations) spoils our ability to apply wisdom of crowds directly. Try averaging 1500 simulated experimental images:



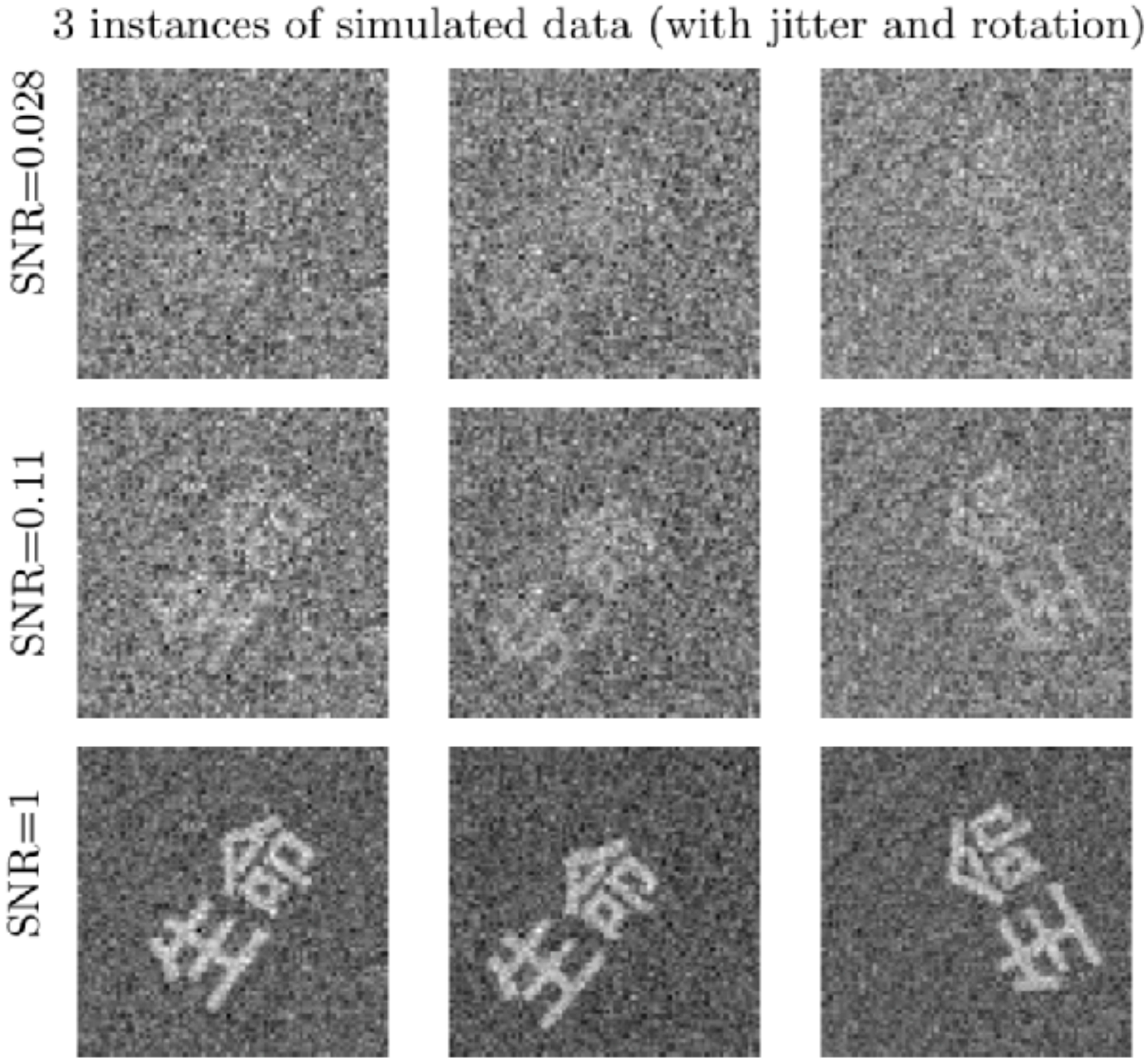
Same problem in 2D

Once again, jitter (random translations) spoils our ability to apply wisdom of crowds directly. Try averaging 1500 simulated experimental images:



PN, *Physical models of living systems* (2/e, 2022)

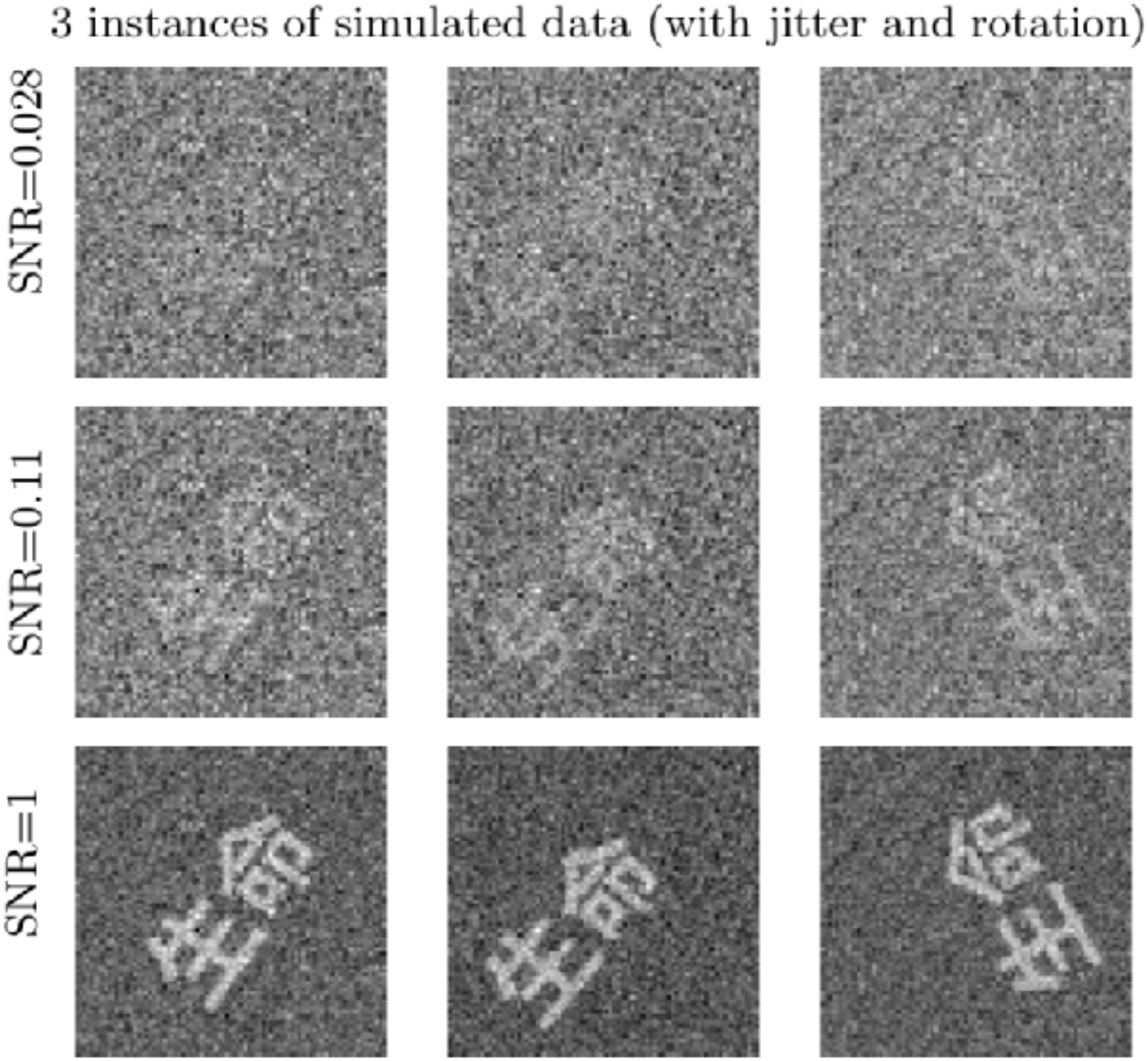
Wait – It's even worse



PN, *Physical models of living systems* (2/e, 2022)

Wait – It's even worse

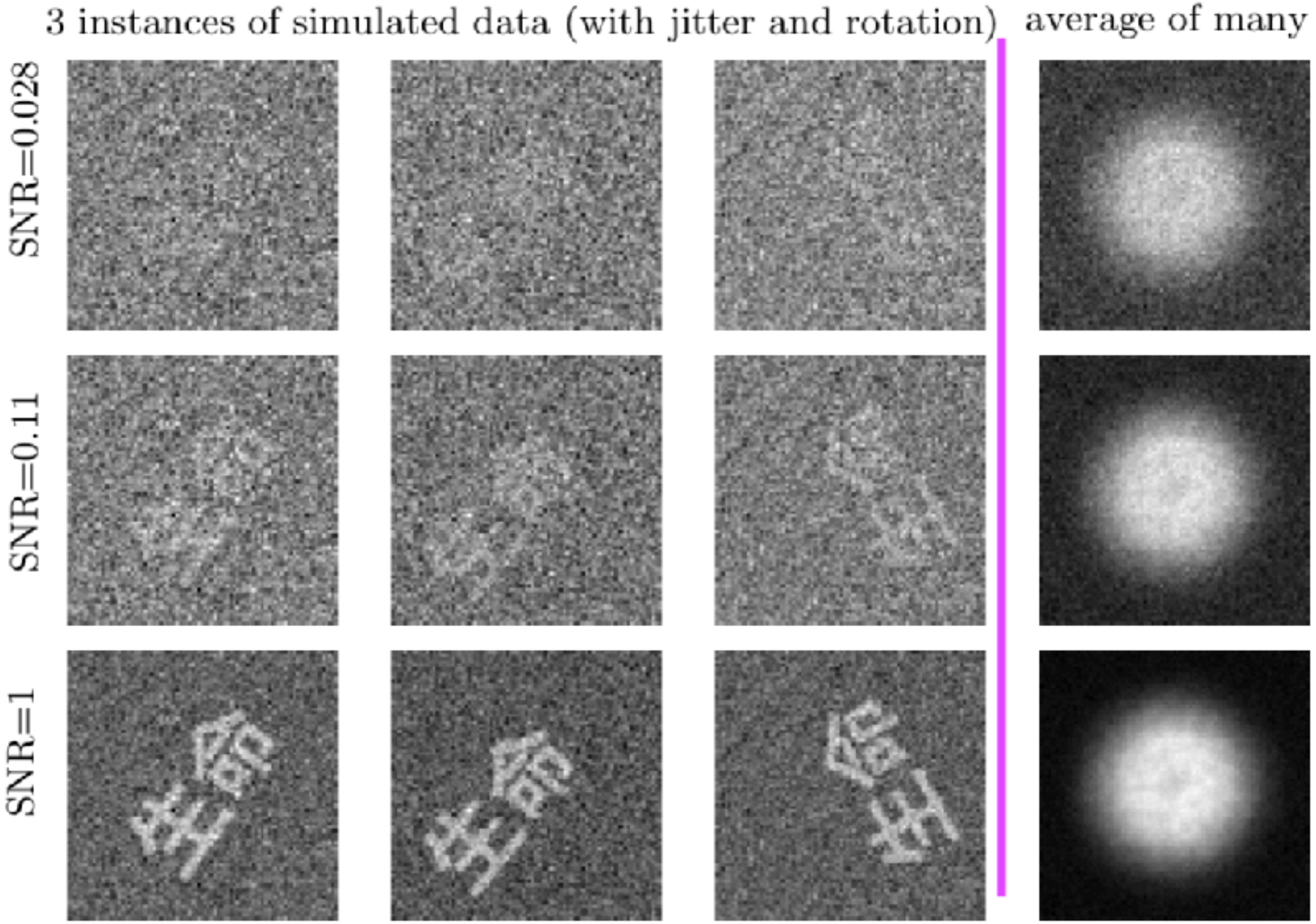
In 2D, and even moreso in 3D, jitter also includes random *rotations*. Again it's a chicken/egg problem. To reduce noise by averaging, we must first align the samples. But to align the samples, we must first reduce the noise!



PN, *Physical models of living systems* (2/e, 2022)

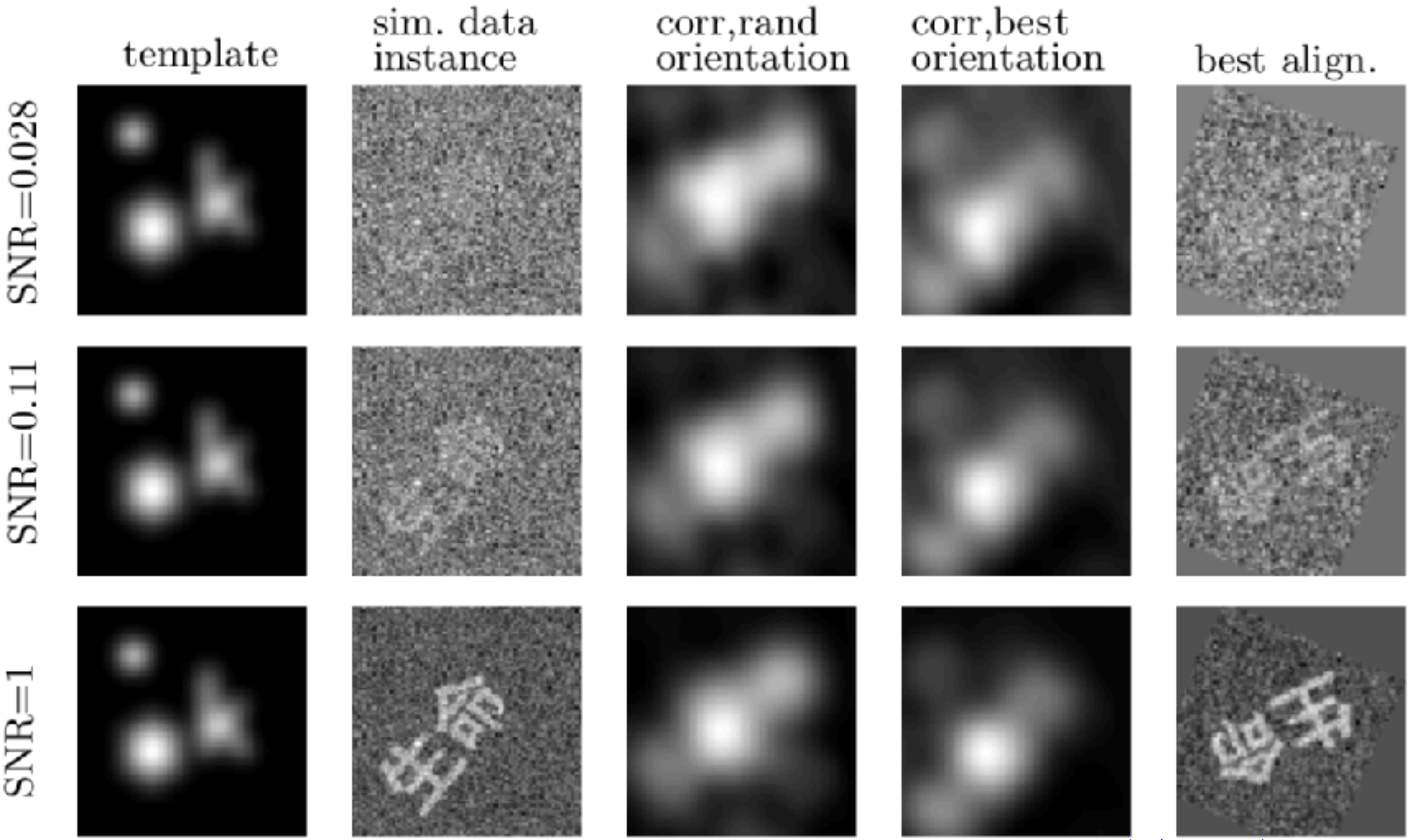
Wait – It's even worse

In 2D, and even moreso in 3D, jitter also includes random *rotations*. Again it's a chicken/egg problem. To reduce noise by averaging, we must first align the samples. But to align the samples, we must first reduce the noise!



PN, *Physical models of living systems* (2/e, 2022)

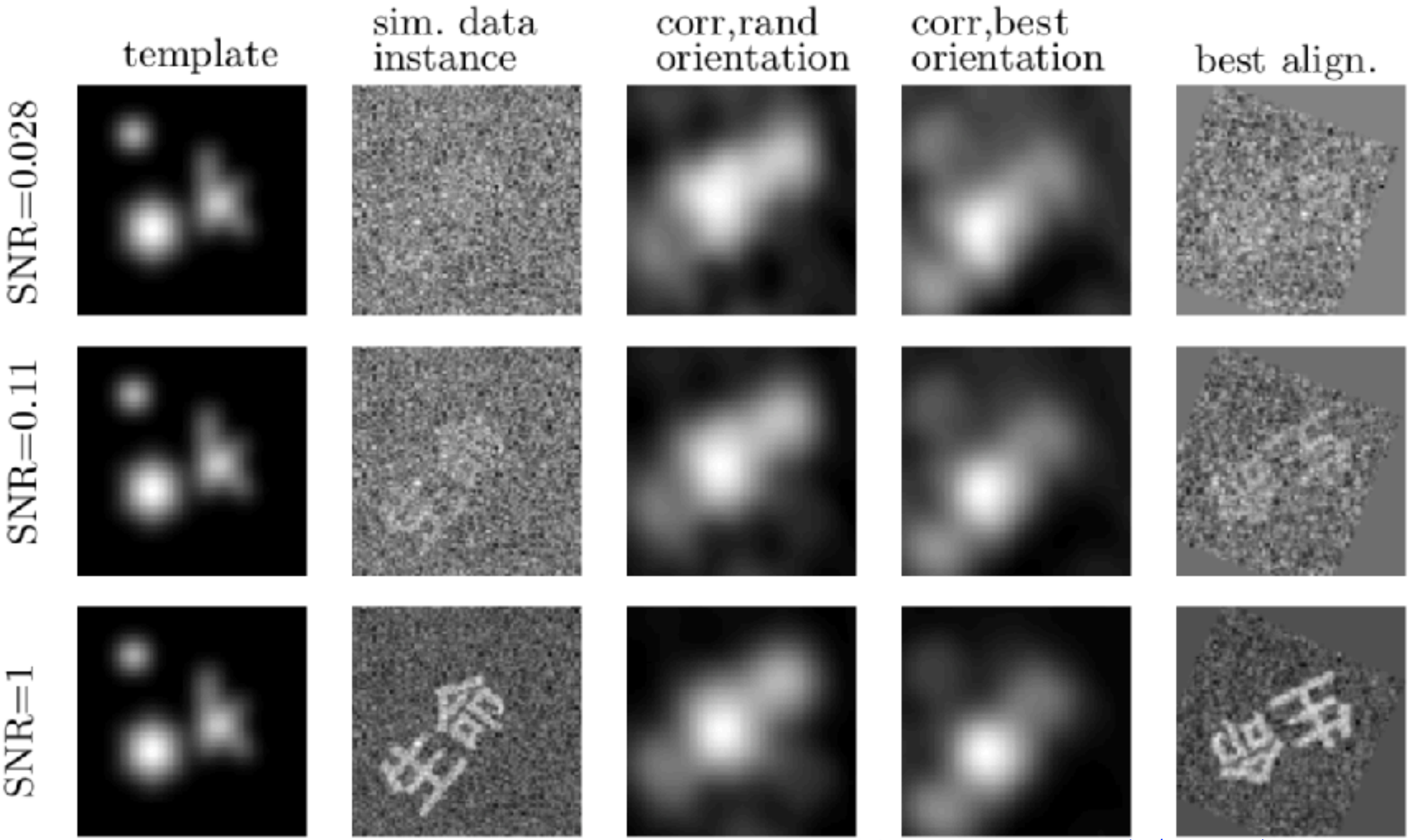
Cross-correlation: 2D



PN, *Physical models of living systems* (2/e, 2022)

Cross-correlation: 2D

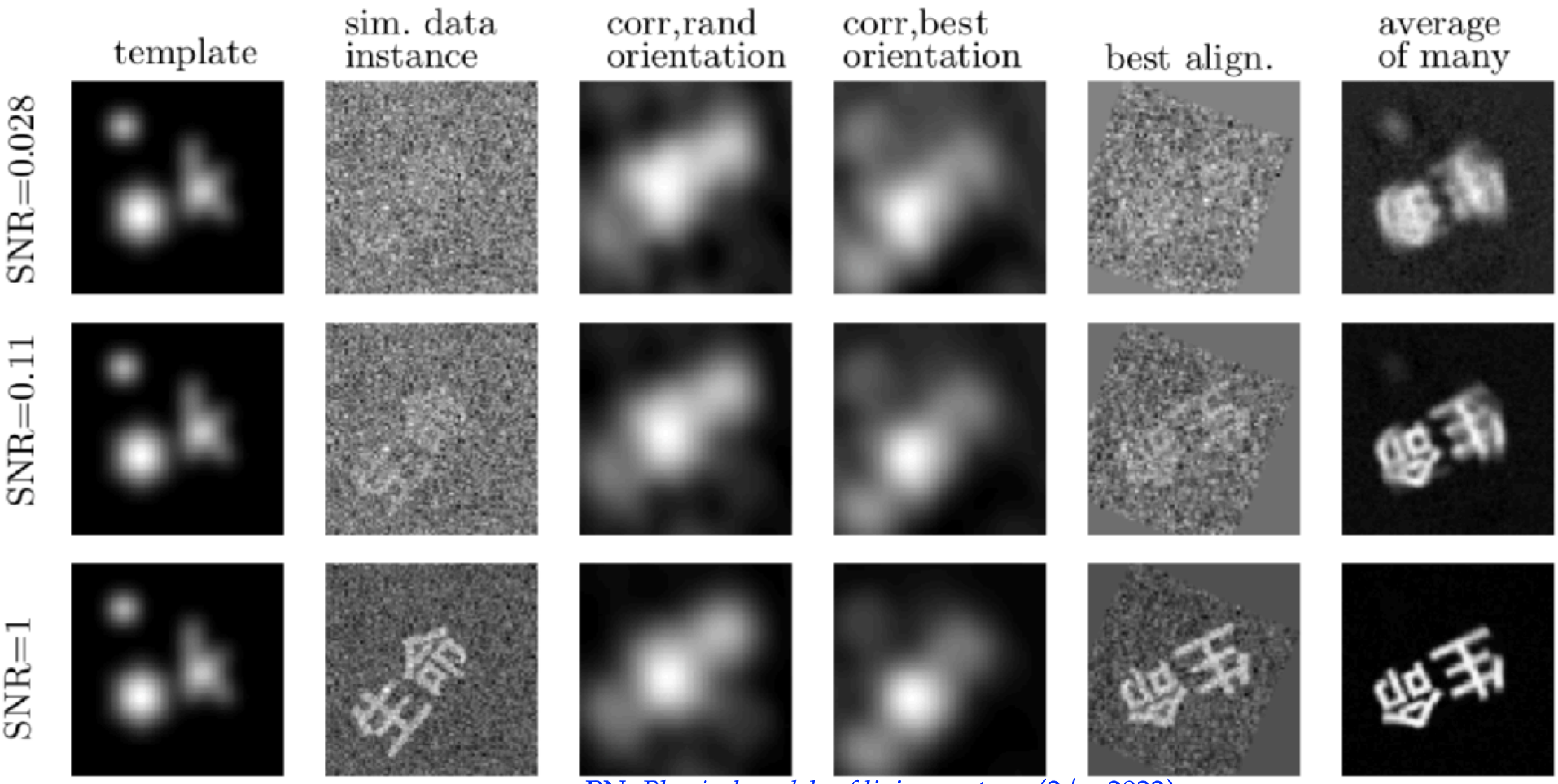
We translate each instance of the noisy data by an amount that optimizes its cross-correlation with a template, for *each of many possible rotations* of the template. Then we choose the rotation that gave the biggest peak in the cross-correlation function. Then we shift *and rotate* each data instance to *undo* the shift and rotation we found, *prior* to averaging the instances point by point.
This is far more successful than naive averaging – but still not great at low SNR.



PN, *Physical models of living systems* (2/e, 2022)

Cross-correlation: 2D

We translate each instance of the noisy data by an amount that optimizes its cross-correlation with a template, for *each of many possible rotations* of the template. Then we choose the rotation that gave the biggest peak in the cross-correlation function. Then we shift *and rotate* each data instance to *undo* the shift and rotation we found, *prior* to averaging the instances point by point.
This is far more successful than naive averaging – but still not great at low SNR.

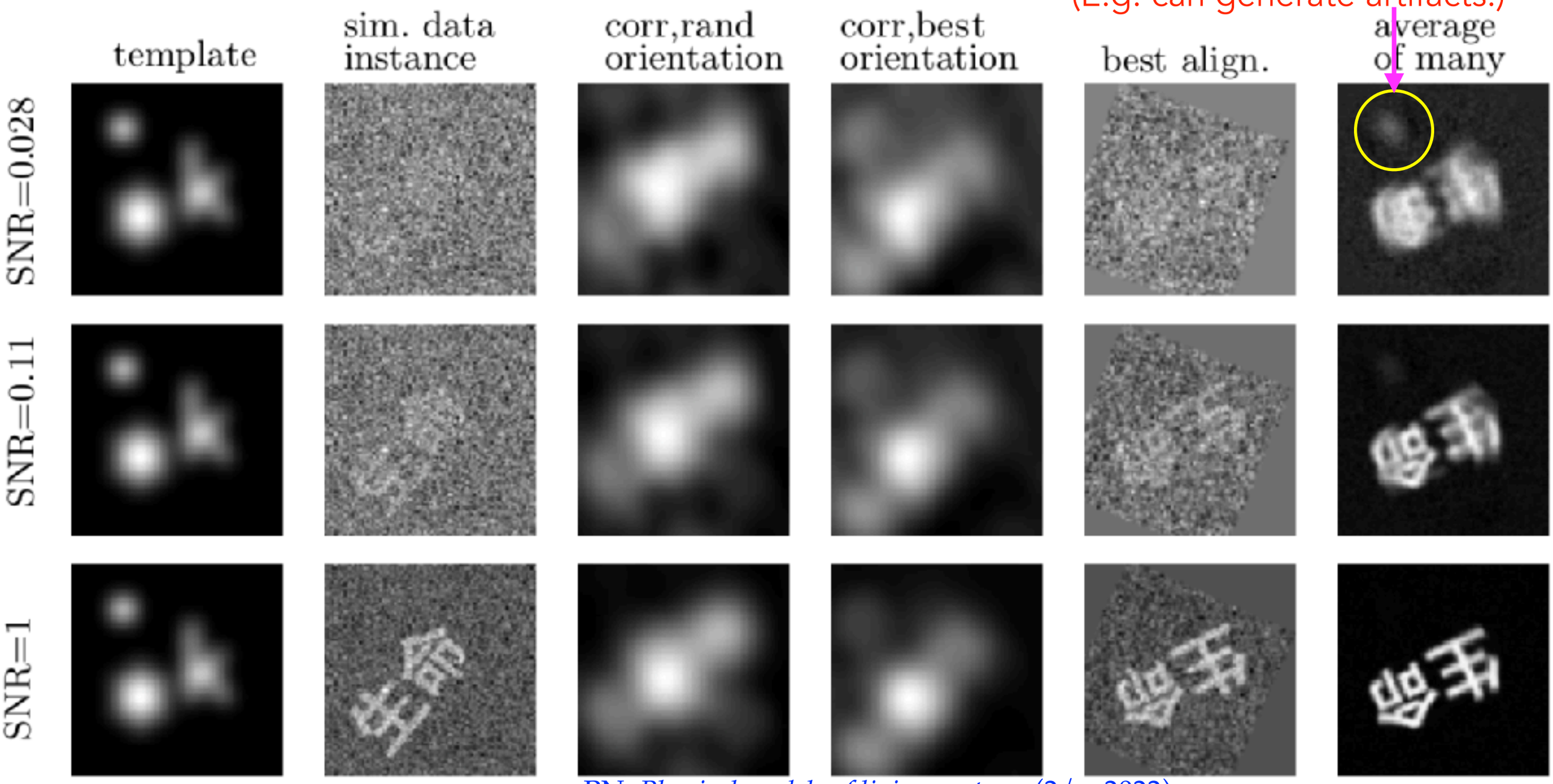


PN, *Physical models of living systems* (2/e, 2022)

Cross-correlation: 2D

We translate each instance of the noisy data by an amount that optimizes its cross-correlation with a template, for *each of many possible rotations* of the template. Then we choose the rotation that gave the biggest peak in the cross-correlation function. Then we shift *and rotate* each data instance to *undo* the shift and rotation we found, *prior* to averaging the instances point by point.
This is far more successful than naive averaging – but still not great at low SNR.

(E.g. can generate artifacts.)

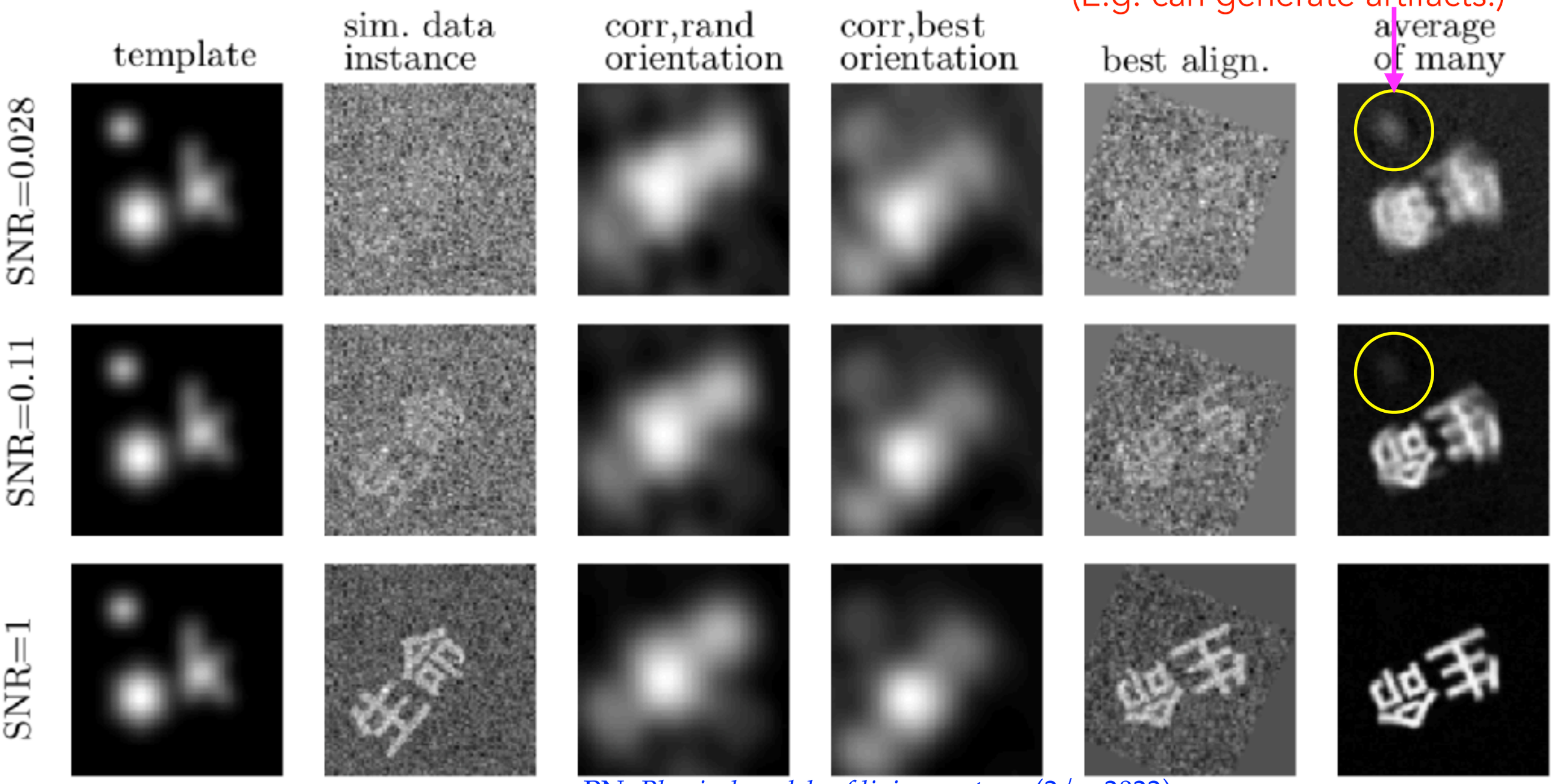


PN, *Physical models of living systems* (2/e, 2022)

Cross-correlation: 2D

We translate each instance of the noisy data by an amount that optimizes its cross-correlation with a template, for *each of many possible rotations* of the template. Then we choose the rotation that gave the biggest peak in the cross-correlation function. Then we shift *and rotate* each data instance to *undo* the shift and rotation we found, *prior* to averaging the instances point by point.
This is far more successful than naive averaging – but still not great at low SNR.

(E.g. can generate artifacts.)



PN, *Physical models of living systems* (2/e, 2022)

Back to the unfair advantage

We saw how to pluck information out of a sea of noise with the help of alignment by cross-correlation. That's great, but:

Why did it work as well as it did? Is there some principled foundation?

Why didn't it work *better* than that, and what alternative might outperform it?

Back to the unfair advantage

We saw how to pluck information out of a sea of noise with the help of alignment by cross-correlation. That's great, but:

Why did it work as well as it did? Is there some principled foundation?

Why didn't it work *better* than that, and what alternative might outperform it?

Again:

- Each pixel has shot noise *independent of other pixels*.
- Each image has been *rigidly translated* (and in 2D also rotated) relative to every other image, but otherwise represents the same PDF for electron arrivals.

Back to the unfair advantage

We saw how to pluck information out of a sea of noise with the help of alignment by cross-correlation. That's great, but:

Why did it work as well as it did? Is there some principled foundation?

Why didn't it work *better* than that, and what alternative might outperform it?

Again:

- Each pixel has shot noise *independent of other pixels*.
- Each image has been *rigidly translated* (and in 2D also rotated) relative to every other image, but otherwise represents the same PDF for electron arrivals.

Why is this information so helpful? In 1D, the right shift to align each instance with the others is a *single number,* determined *globally by the entire image,* which means we have a lot of data to determine it to good accuracy.

In a succinct formula, our generative data model is:

$$\text{exp. image} = \text{Shift}_q(\overrightarrow{A}) + \overrightarrow{\Xi}$$

and we wish to infer the true image $\overrightarrow{A}$ from a collection of experimental images masked by noise $\overrightarrow{\Xi}$. It's **another Bayesian inference problem.**

Sigworth's insight

exp. image = $\text{Shift}_q(\overrightarrow{A}) + \overrightarrow{\Xi}$

FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.

Sigworth's insight

exp. image = $\text{Shift}_q(\overrightarrow{A}) + \overrightarrow{\Xi}$

was to realize that:

FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.

Sigworth's insight

exp. image = $\text{Shift}_q(\overrightarrow{A}) + \overrightarrow{\Xi}$

was to realize that:

- The alignment of each image (q) is itself a *random variable* (i.e. unknown).

FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.

Sigworth's insight

exp. image = $\text{Shift}_q(\overrightarrow{A}) + \overrightarrow{\Xi}$

was to realize that:

- The alignment of each image (q) is itself a *random variable* (i.e. unknown).
- So actually we should be asking about its *probability distribution,* which is not fully represented by any single "best" choice.

FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.

Sigworth's insight

exp. image = $\mathrm{Shift}_q(\overrightarrow{A}) + \overrightarrow{\Xi}$

was to realize that:

- The alignment of each image (q) is itself a *random variable* (i.e. unknown).
- So actually we should be asking about its *probability distribution,* which is not fully represented by any single "best" choice.
- Not surprisingly, that posterior involves the cross-correlation.

FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.

Sigworth's insight

exp. image = $\mathrm{Shift}_q(\overrightarrow{A}) + \overrightarrow{\Xi}$

was to realize that:

- The alignment of each image (q) is itself a *random variable* (i.e. unknown).
- So actually we should be asking about its *probability distribution,* which is not fully represented by any single "best" choice.
- Not surprisingly, that posterior involves the cross-correlation.
- But *we don't really care about the alignment;* all we want for our science is the best possible estimate of the true image, given the data.

FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.

Sigworth's insight

exp. image = $\text{Shift}_q(\overrightarrow{A}) + \overrightarrow{\Xi}$

was to realize that:

- The alignment of each image (q) is itself a *random variable* (i.e. unknown).
- So actually we should be asking about its *probability distribution,* which is not fully represented by any single "best" choice.
- Not surprisingly, that posterior involves the cross-correlation.
- But *we don't really care about the alignment;* all we want for our science is the best possible estimate of the true image, given the data.
- In jargon: **We want the posterior distribution of images given the data, *"marginalized" over the alignment.***

FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.

What to optimize

A = unknown image pixel values

$$\wp(\overleftrightarrow{A} \mid \text{data}) = \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi \, \wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\} \mid \text{data}).$$

$\mathbf{q}_i$, φ_i = unknown shift and rotation of experimental image *i, both uninteresting "nuisance variables"* so we marginalized them.

A lot of gaussians building up a cross-correlation

What to optimize

A = unknown image pixel values

$$\wp(\overleftrightarrow{A} \mid \text{data}) = \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi \, \wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\} \mid \text{data}).$$

$\mathbf{q}_i$, φ_i = unknown shift and rotation of experimental image *i, both uninteresting "nuisance variables"* so we marginalized them.

$$= \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi \, \wp(\text{data} \mid \overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\}) \frac{\wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\})}{\wp(\text{data})}.$$

A lot of gaussians building up a cross-correlation

What to optimize

A = unknown image pixel values

$$\wp(\overleftrightarrow{A} \mid \text{data}) = \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi\, \wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\} \mid \text{data}).$$

$\mathbf{q}_i$, φ_i = unknown shift and rotation of experimental image *i, both uninteresting "nuisance variables"* so we marginalized them.

$$= \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi\, \wp(\text{data} \mid \overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\}) \frac{\wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\})}{\wp(\text{data})}.$$

likelihood factorizes

prior factorizes

A lot of gaussians building up a cross-correlation

don't need this constant

What to optimize

A = unknown image pixel values

$$\wp(\overleftrightarrow{A} \mid \text{data}) = \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi\, \wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\} \mid \text{data}).$$

$\mathbf{q}_i$, φ_i = unknown shift and rotation of experimental image *i, both uninteresting "nuisance variables"* so we marginalized them.

$$= \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi\, \wp(\text{data} \mid \overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\}) \frac{\wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\})}{\wp(\text{data})}.$$

likelihood factorizes

prior factorizes

$$= C \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi\, \wp(\overleftrightarrow{X}_1 \mid \overleftrightarrow{A}, \boldsymbol{q}_1, \varphi_1) \cdots \wp(\overleftrightarrow{X}_N \mid \overleftrightarrow{A}, \boldsymbol{q}_N, \varphi_N)\, \mathrm{e}^{-\|\boldsymbol{q}_1\|^2/(2\sigma_q^2)} \ldots \mathrm{e}^{-\|\boldsymbol{q}_N\|^2/(2\sigma_q^2)}$$

A lot of gaussians building up a cross-correlation

don't need this constant

What to optimize

A = unknown image pixel values

$$\wp(\overleftrightarrow{A} \mid \text{data}) = \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi\, \wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\} \mid \text{data}).$$

$\mathbf{q}_i$, φ_i = unknown shift and rotation of experimental image *i, both uninteresting "nuisance variables"* so we marginalized them.

$$= \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi\, \wp(\text{data} \mid \overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\}) \frac{\wp(\overleftrightarrow{A}, \{\boldsymbol{q}_i, \varphi_i\})}{\wp(\text{data})}.$$

likelihood factorizes

prior factorizes

$$= C \sum_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N} \int \mathrm{d}^N\varphi\, \wp(\overleftrightarrow{X}_1 \mid \overleftrightarrow{A}, \boldsymbol{q}_1, \varphi_1) \cdots \wp(\overleftrightarrow{X}_N \mid \overleftrightarrow{A}, \boldsymbol{q}_N, \varphi_N)\, \mathrm{e}^{-\|\boldsymbol{q}_1\|^2/(2\sigma_q^2)} \ldots \mathrm{e}^{-\|\boldsymbol{q}_N\|^2/(2\sigma_q^2)}$$

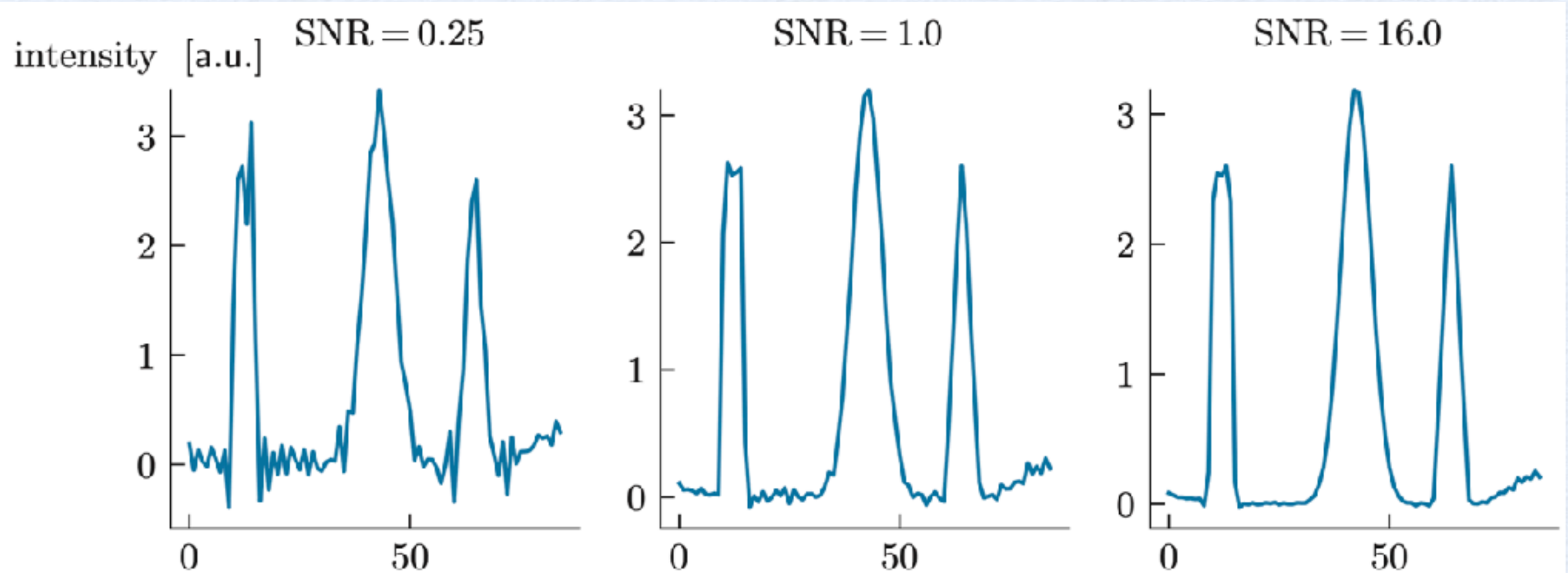
A lot of gaussians building up a cross-correlation

don't need this constant

So the whole integral factorizes! Not hard to estimate these integrals one by one, then optimize over A.

1D image reconstructions obtained by maximizing posterior probability

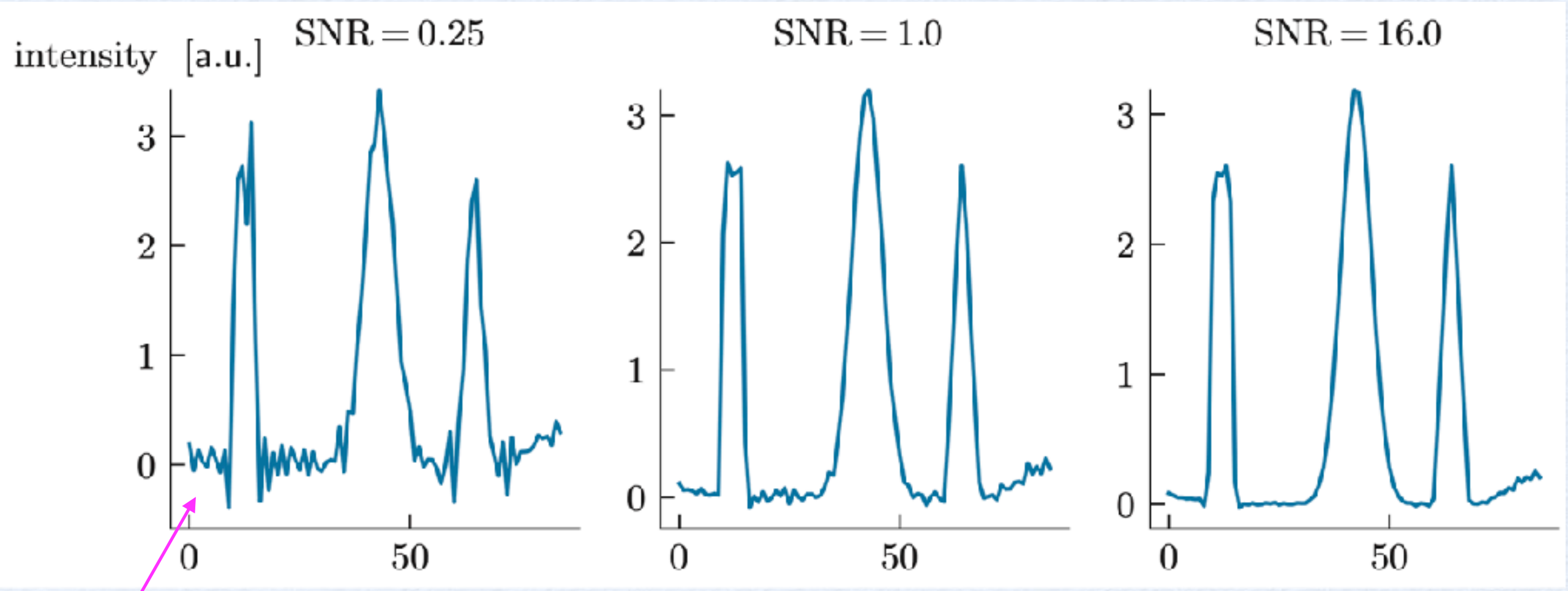
Applying the method the *same 1500 simulated data instances as before* gives much more successful reconstruction of the underlying "image" then the cross-correlation method at low SNR:



PCN, https://repository.upenn.edu/physics_papers/656/ ; https://www.physics.upenn.edu/biophys/PMLS2e/

1D image reconstructions obtained by maximizing posterior probability

Applying the method the *same 1500 simulated data instances as before* gives much more successful reconstruction of the underlying "image" then the cross-correlation method at low SNR:



It's incredible when you recall how terrible the individual "images" looked!

PCN, https://repository.upenn.edu/physics_papers/656/ ; https://www.physics.upenn.edu/biophys/PMLS2e/

2D image reconstructions obtained by maximizing posterior probability

Applying the method to the *same 1500 simulated data instances as before:*

SNR=0.028

SNR=0.11

SNR=1.0

Much better than cross-correlation at low SNR. In particular, even at the lowest SNR **the artifact found earlier is absent,** even though the algorithm used the same data and started with the same initial guess (template). Later refinements grew into the RELION algorithm and successors cryoSPARC, cisTEM, et al.

PCN, *Physical models of living systems* (2/e, 2022); Implementing an algorithm due to *FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.*

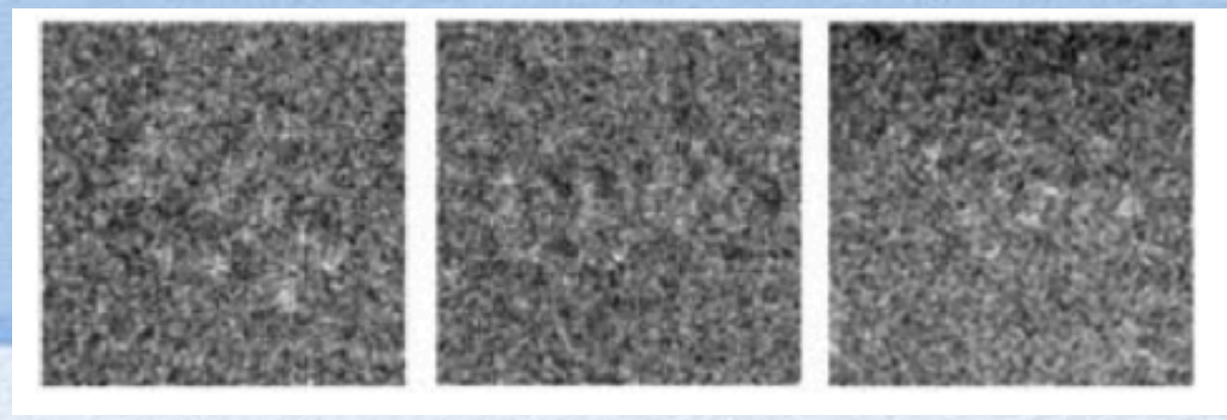
2D image reconstructions obtained by maximizing posterior probability

Applying the method to the *same 1500 simulated data instances as before:*

Much better than cross-correlation at low SNR. In particular, even at the lowest SNR **the artifact found earlier is absent,** even though the algorithm used the same data and started with the same initial guess (template). Later refinements grew into the RELION algorithm and successors cryoSPARC, cisTEM, et al.

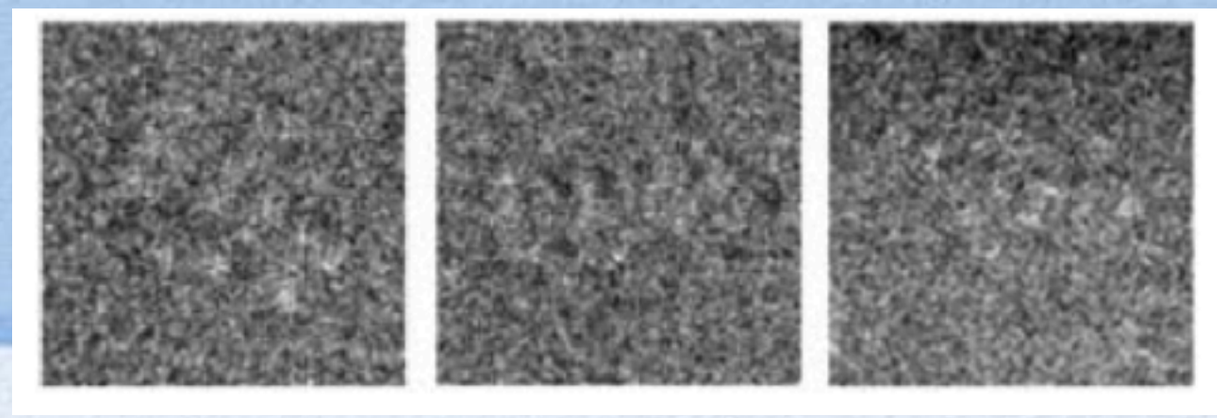
PCN, *Physical models of living systems* (2/e, 2022); Implementing an algorithm due to *FJ Sigworth (1998). Journal of Structural Biology, 122(3), 328–339.*

Summary part 5



[To deal with sample heterogeneity, add another discrete variable allowing each image to be probabilistically assigned to one of several conformational classes.]

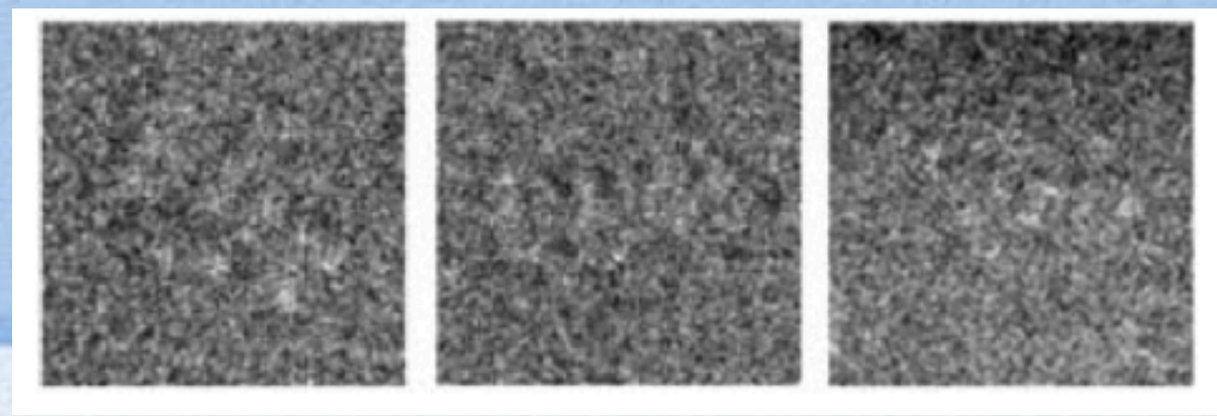
Summary part 5



- Bad: Naively average many noisy images.

[To deal with sample heterogeneity, add another discrete variable allowing each image to be probabilistically assigned to one of several conformational classes.]

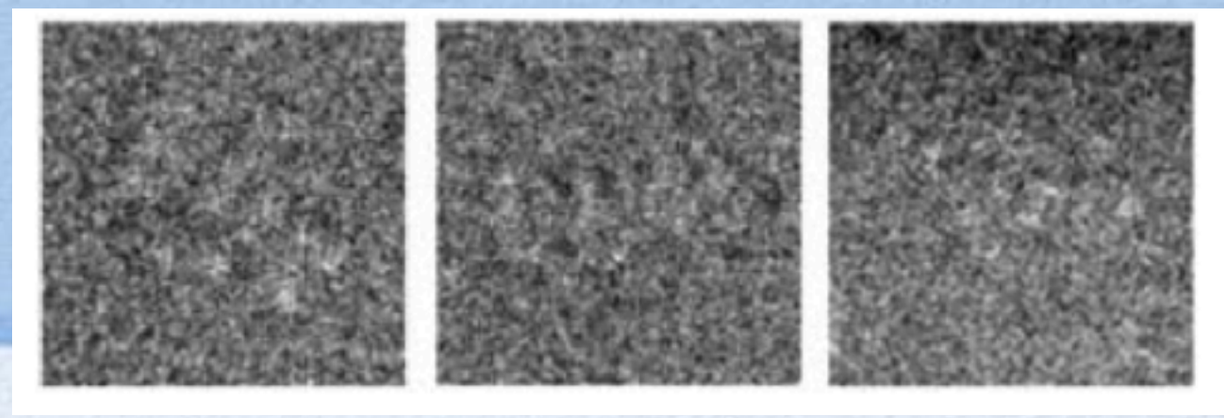
Summary part 5



- Bad: Naively average many noisy images.
- Better: For each noisy experimental image, select the one rigid motion that best aligns it to a guess; then average over all experimental images.

[To deal with sample heterogeneity, add another discrete variable allowing each image to be probabilistically assigned to one of several conformational classes.]

Summary part 5



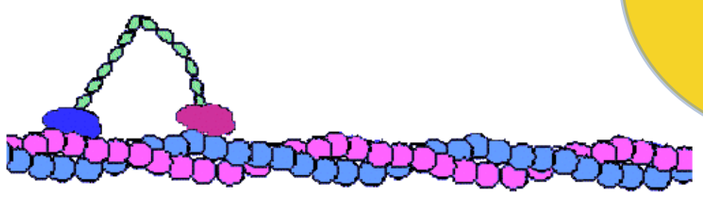
- Bad: Naively average many noisy images.
- Better: For each noisy experimental image, select the one rigid motion that best aligns it to a guess; then average over all experimental images.
- Much better: For each experimental image, instead of *one* winner make a probability distribution over *all* rigid motions and find the weighted average; then also average over experimental images.

[To deal with sample heterogeneity, add another discrete variable allowing each image to be probabilistically assigned to one of several conformational classes.]

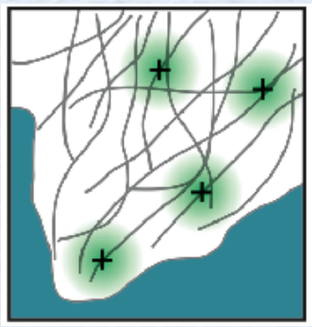
Full circle

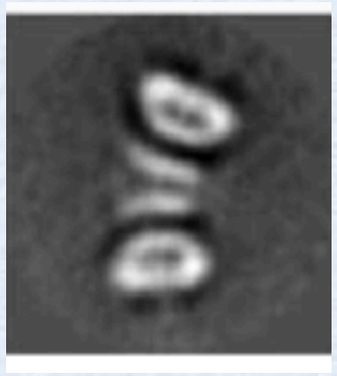
Medical test

Changepoint Analysis



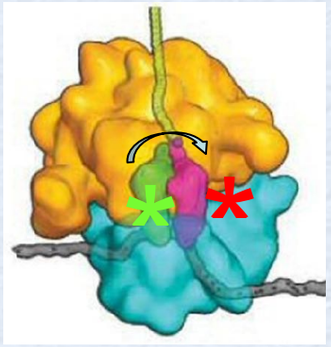
Superresolution



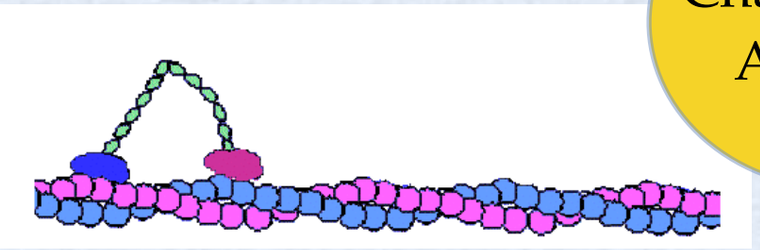


cryoEM

Enzyme kinetics



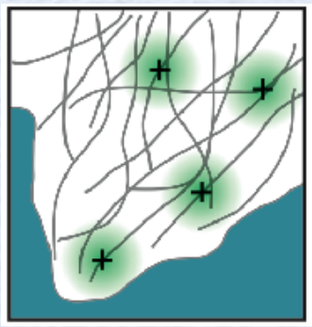
Full circle



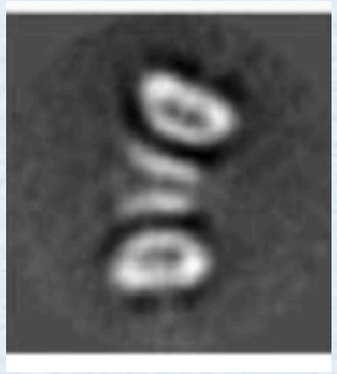
Changepoint Analysis

Medical test

Superresolution

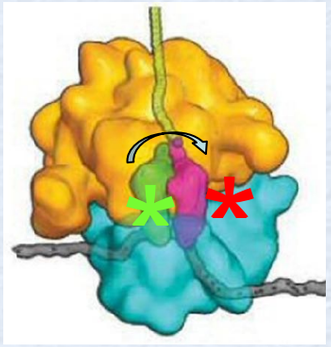


Bayes formula

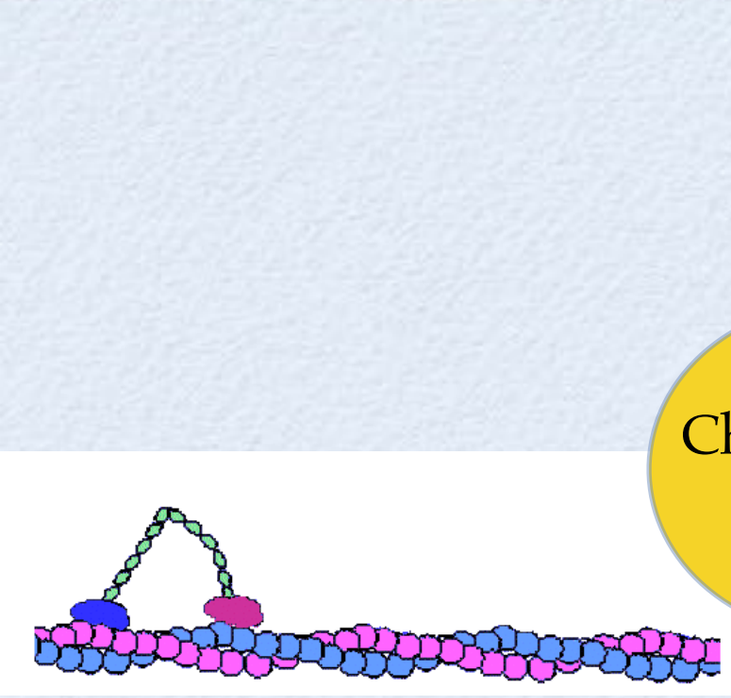


cryoEM

Enzyme kinetics



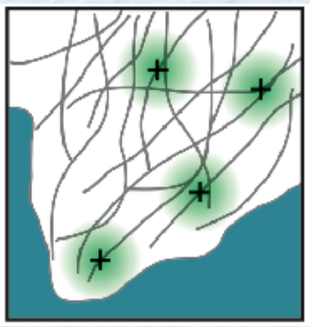
Full circle



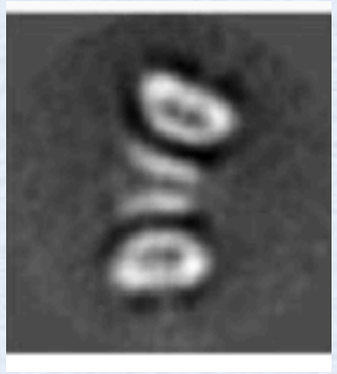
Medical test

Changepoint Analysis

Superresolution

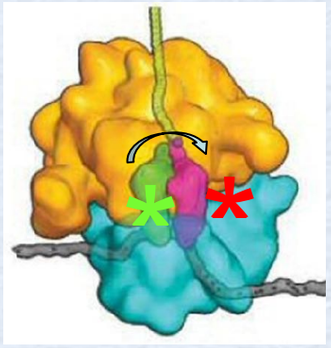


Bayes formula



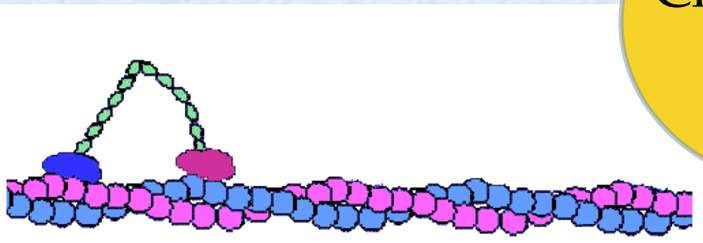
cryoEM

Enzyme kinetics



Theory can cut across apparently different kinds of experiment, offering useful methods to one domain from another without having to reinvent everything. Physicists are pretty good at this – when we're part of a *team* involving life scientists.

Full circle

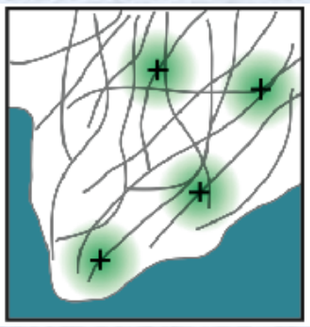


Medical test

Changepoint Analysis

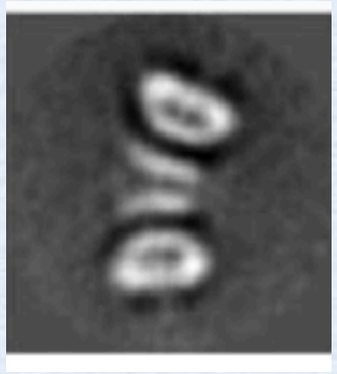
A ***physical idea*** – photon theory – helped us to extract what was going on.

Superresolution



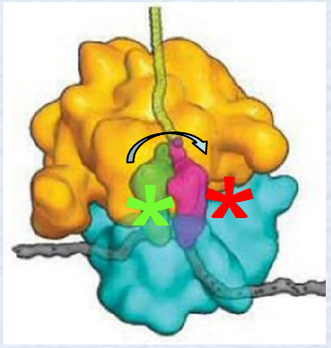
Bayes formula

A ***physical idea*** – additive shot noise plus rigid motion of a well-defined object – helped us to extract the structure.



cryoEM

Enzyme kinetics



A ***physical idea*** – first-passage time – helped us to extract what was going on.

Theory can cut across apparently different kinds of experiment, offering useful methods to one domain from another without having to reinvent everything. Physicists are pretty good at this – when we're part of a *team* involving life scientists.

Go long

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

Go long

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

Some of the ideas we have encountered are things that many scientists describe as “beautiful.” What does that mean?

Go long

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

Some of the ideas we have encountered are things that many scientists describe as "beautiful." What does that mean?

There are as many definitions as there are scientists, but I think many would agree that part of the answer is that a beautiful physical idea is ***surprising yet inevitable;*** it may also be ***simple yet unexpectedly general.***

Go long

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

Some of the ideas we have encountered are things that many scientists describe as "beautiful." What does that mean?

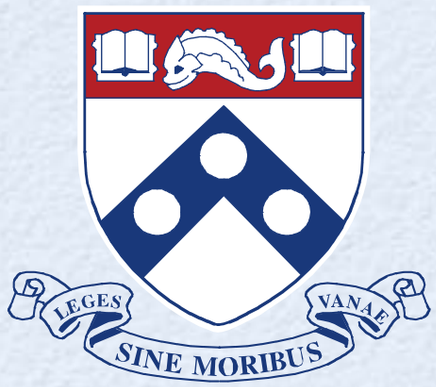
There are as many definitions as there are scientists, but I think many would agree that part of the answer is that a beautiful physical idea is ***surprising yet inevitable***; it may also be ***simple yet unexpectedly general.***

For example, *maximizing posterior probability* has those qualities. We've seen how it is a general framework for many kinds of scientific inference, replacing and extending a grab-bag of seemingly unrelated methods.

Thanks and further reading

John Beausang, Clark Fritsch; Yale Goldman. Sophie Lohmann, Monika Makurath, Fereshteh Memarian; Fred Sigworth.

NSF CMMI



University of Pennsylvania

For posterior-maximization applied to optical superresolution:
P. Nelson, *From Photon to Neuron* Princeton Univ. Press.

For posterior-maximization applied to cryo-EM:
P. Nelson, *Physical models of living systems: Probability, simulation, dynamics*. Second Ed.
`https://www.physics.upenn.edu/biophys/PMLS2e/`

Also
Jesse Kinder and P. Nelson, *Student's guide to Python for physical modeling*. Second Ed. Princeton Univ Press, August 2021.

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

Details

back

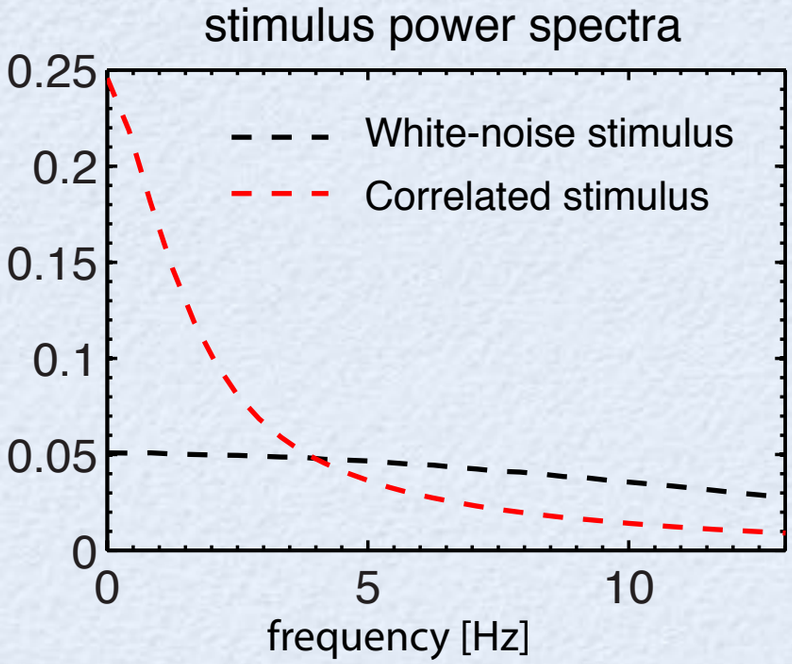
150x180μm recording spot
= 5x6 array of electrodes spaced 30μm (similar to RGC spacing).

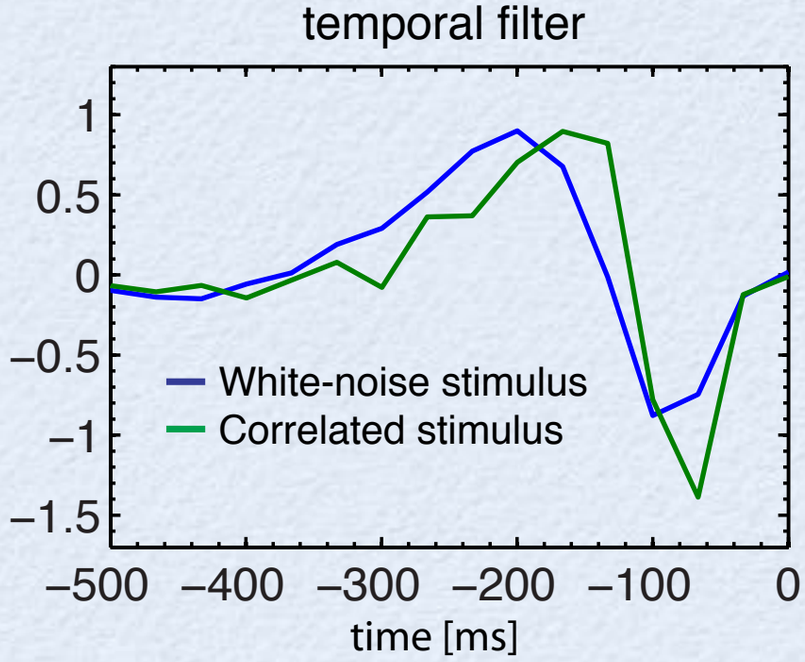
[Data taken at 10kHz. Noise ~30μV. Big spikes ~400μV. Others go all the way down to the noise floor. Prior to analysis, filter out slow baseline drift. Also apply a spatial decorrelating filter, deduced from statistics of noise, to sharpen the “image” spatially.]

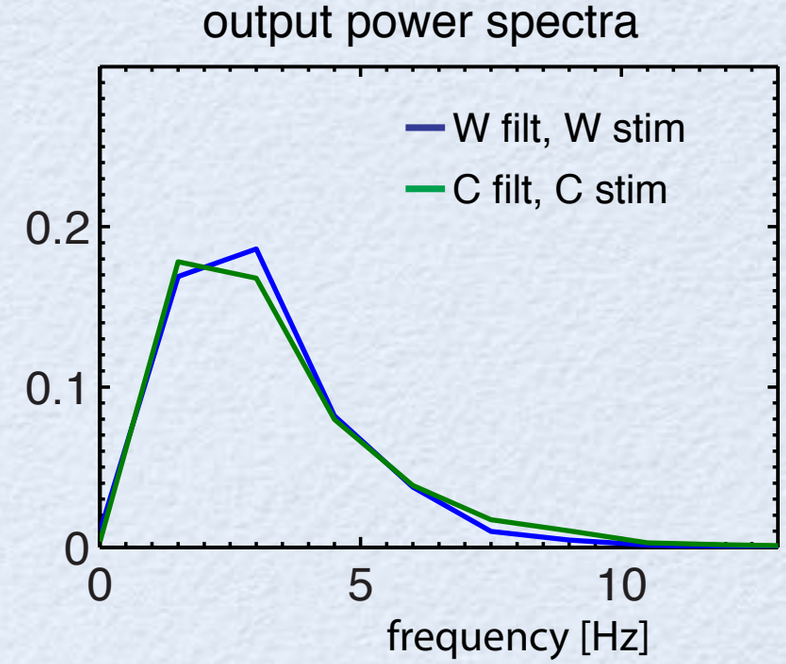
Back

Adaptive decorrelation, (temporal)

The retina dynamically adjusts its signal processing in response to statistical properties of recently-viewed scenes, as predicted on information-theoretic grounds.





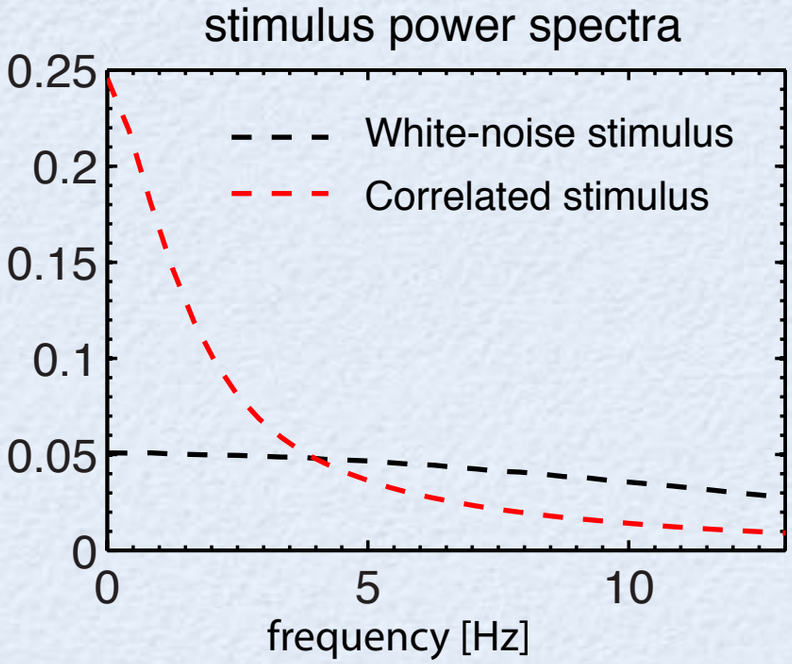


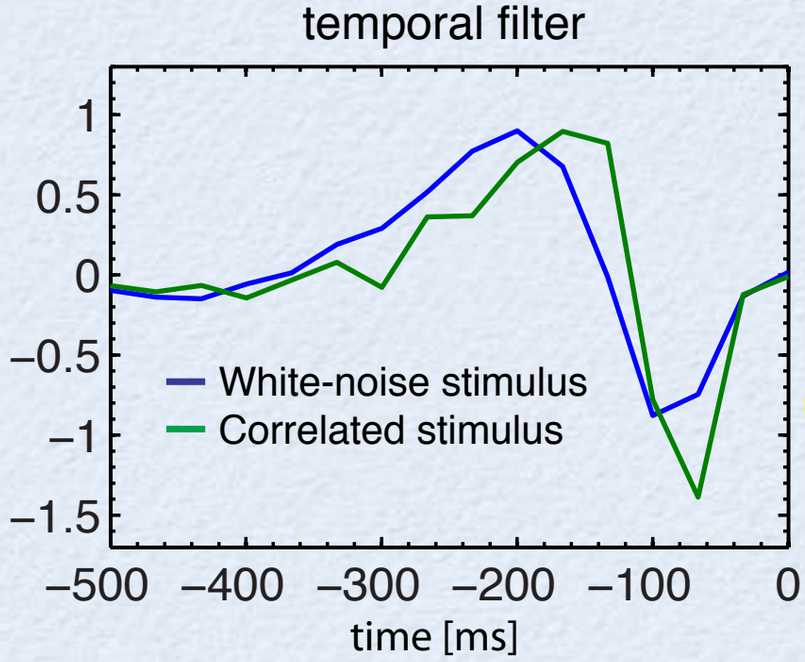
Here a particular OFF ganglion cell maintains a constant amount of temporal correlation in its output, regardless of the amount of correlation in its visual stimulus.

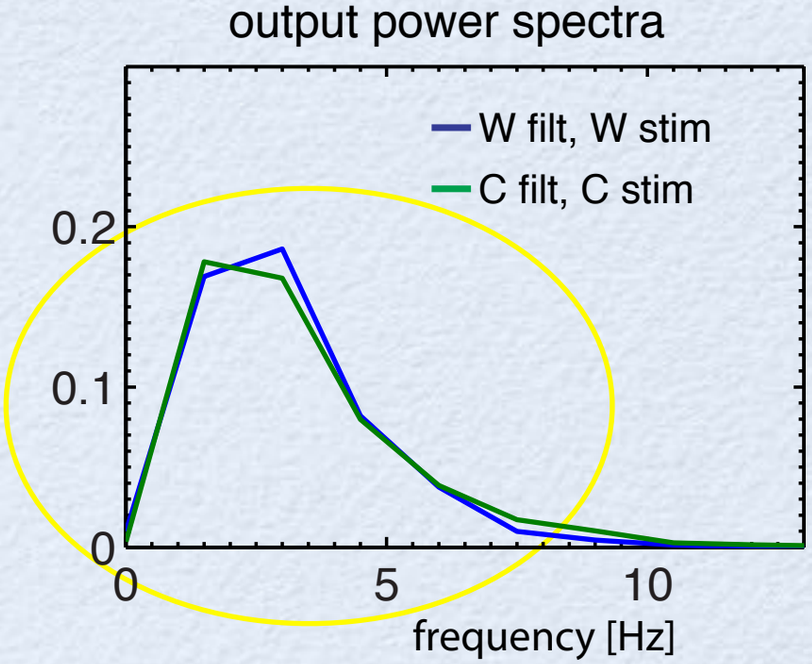
KD Simmons, JS Prentice, G Tkacik, J Homann, H Yee, S Palmer, PCN, V Balasubramanian, PLoS Comput Biol. (2013)

Adaptive decorrelation, (temporal)

The retina dynamically adjusts its signal processing in response to statistical properties of recently-viewed scenes, as predicted on information-theoretic grounds.







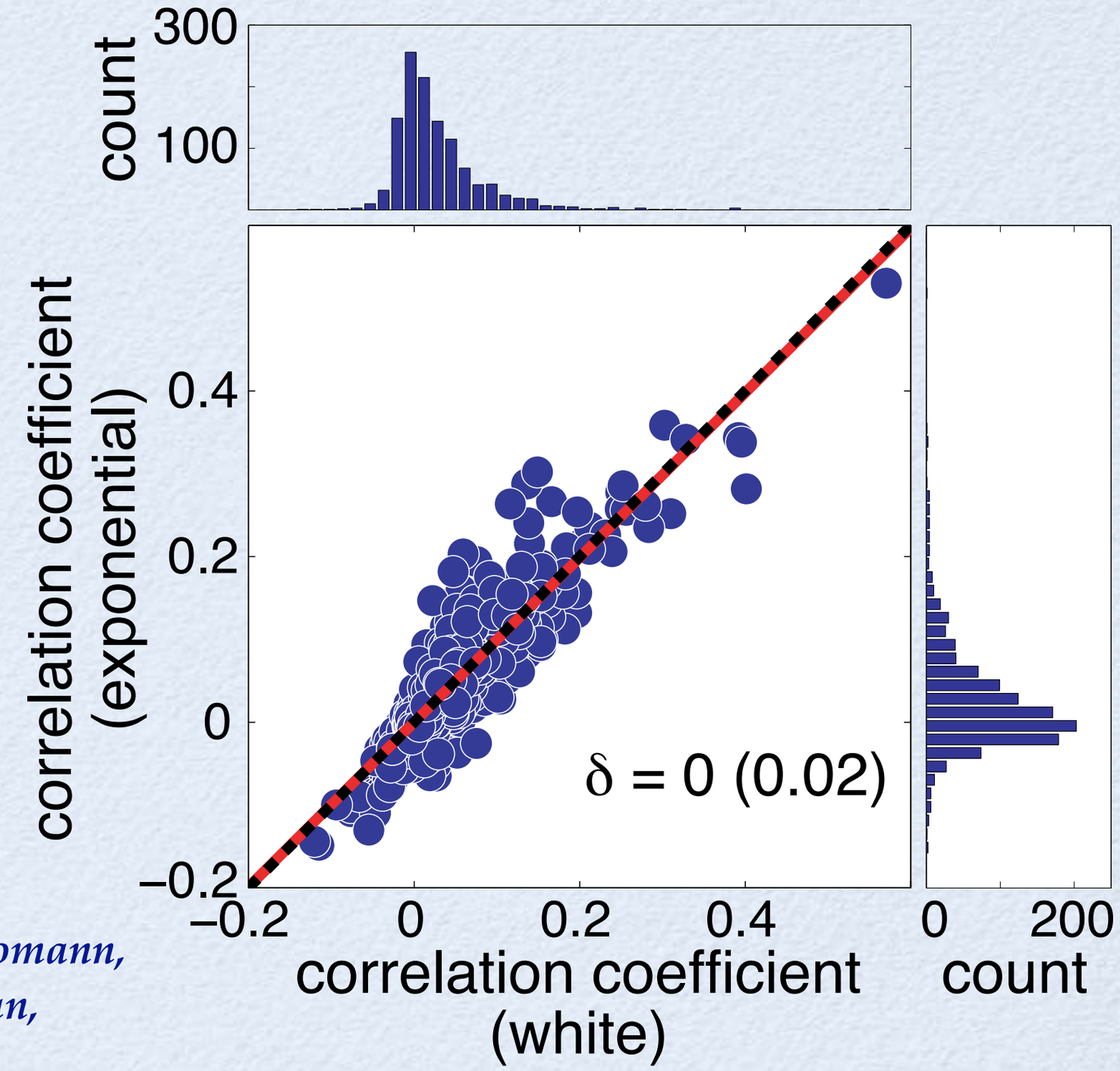
Here a particular OFF ganglion cell maintains a constant amount of temporal correlation in its output, regardless of the amount of correlation in its visual stimulus.

KD Simmons, JS Prentice, G Tkacik, J Homann, H Yee, S Palmer, PCN, V Balasubramanian, PLoS Comput Biol. (2013)

back

Adaptive decorrelation, (spatial)

Also at the multi-cell level, after adaptation the degree of correlation between any two ganglion cells is nearly unchanged when we change the correlation strength in the stimulus.



KD Simmons, JS Prentice, G Tkacik, J Homann, H Yee, S Palmer, PCN, V Balasubramanian, PLoS Comput Biol. (2013)

[Nuts and Bolts]

Let $V_\alpha(t)$ be measured voltage, electrode α and $F_{\mu\alpha}(t)$ be template 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 rest of the prior — amplitude prior — noise covariance

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

which is a Gaussian in A. So it's easy to marginalize over A: **just complete the square!**
[Here $K_\mu = \mathcal{P}^{\text{cell}}(\mu)\mathcal{P}^{\text{time}}(t_1)(2\pi\sigma_\mu^2)^{-1/2}$ 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 template 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 rest of the prior — amplitude prior — noise covariance

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

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

Next, we sweep over a range of t to find the best value of likelihood ratio for this spike type. [We only check 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 template 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 rest of the prior — amplitude prior — noise covariance

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

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

Next, we sweep over a range of t to find the best value of likelihood ratio for this spike type. [We only check 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 template 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 rest of the prior — amplitude prior — noise covariance

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

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

Next, we sweep over a range of t to find the best value of likelihood ratio for this spike type. [We only check 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 an absolute criterion. We know we're done** when this test fails.

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

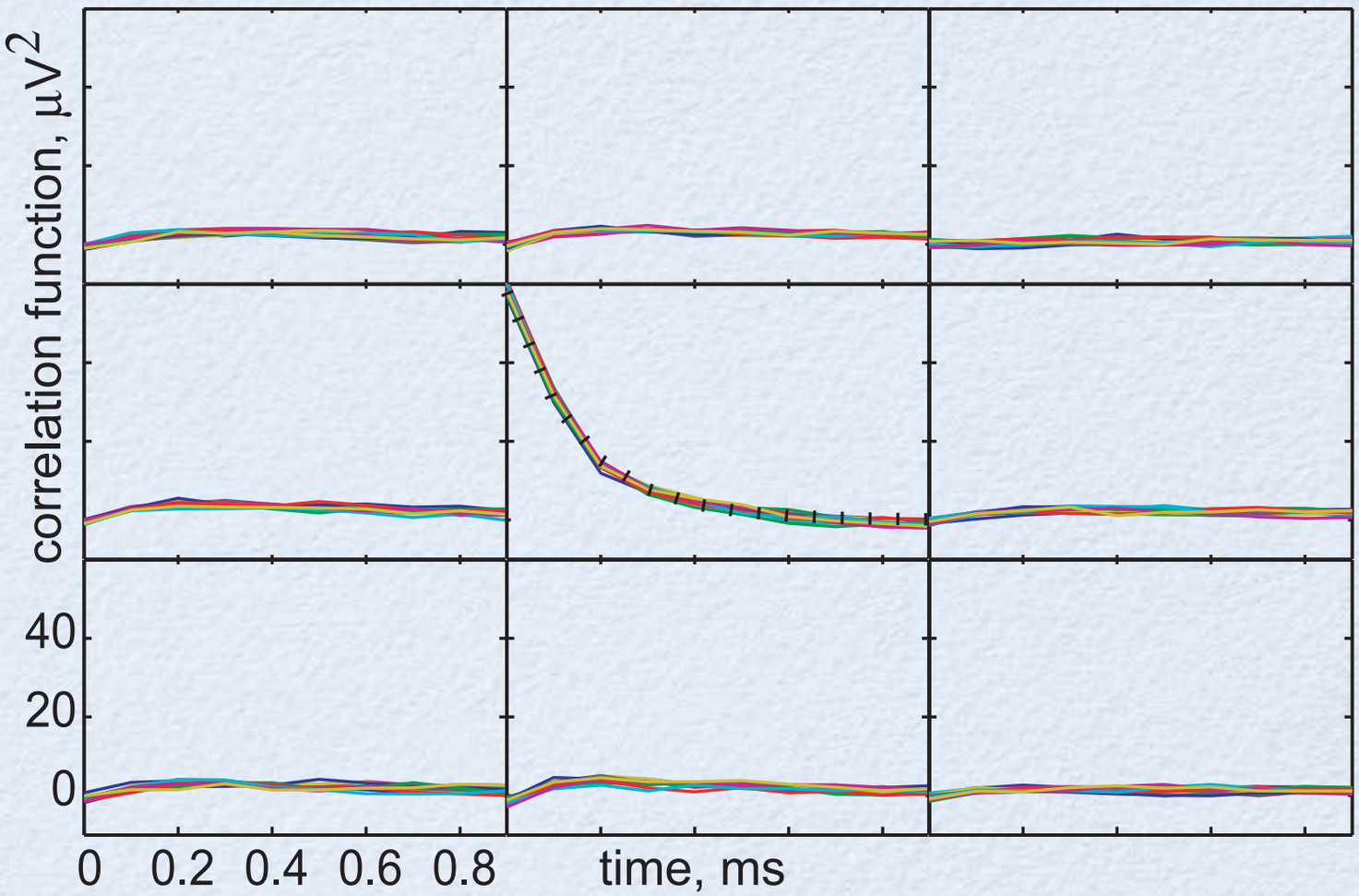
[Nuts and Bolts: Noise covariance]

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 one channel with nearby channels (after doing an initial spatial filter, which we also obtained from data).

We see that the selected channel 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.]



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