

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR: Today, as I mentioned to you in last lecture, we're going to really be focusing in quite some depth into this paper by Sunney Xie. So I would say that it is one of my all time favorite papers. And in particular from the standpoint of discussing a paper in class, I think it is absolutely wonderful.

It is, I think, clearly written. It explains why that did all these things. And they checked all sorts of possible sources of perhaps being lead astray. And I think it was a huge amount of work, and it was a technical tour de force when it came out.

So before this, I would say single molecule biophysics, by which I mean both single molecule fluorescence, i.e., detection, as well as single molecule manipulation, were almost exclusively in vitro techniques. So we took purified components, and then we studied the fluorescence, or the mechanical properties and so forth, of these molecules in the equivalent of a test tube.

But really, in glass slides, where we just had to purified components, so no know living cells. And I think we got a lot of insight into the dynamics of molecular motors, transcription, translation, and so forth. And I think that many of us in the field thought that this paper was essentially not possible.

I did my Ph.D in the single molecule area, from of 2002 to 2005, I graduated. And indeed, I did a little bit of work in this area of single molecule fluorescence, and I was basically unsuccessful. Even just doing this kind of an in vitro setting. My lab is [INAUDIBLE], we did primarily single molecule manipulation. We were playing with the single molecule fluorescence . And eventually, people in the lab got it to work. But I would say that my foray into it was maybe unsuccessful.

So I had a very healthy respect for the challenges that are involved in doing single

molecule fluorescence. And the thought of doing this in live cells was very scary. And I'd say that many of us thought it was not going to work. And indeed, this project was one-- and the general goal of studying a single molecule dynamics in living cells, is something that's Sunney's group had been working on, I think, for many years.

And it indeed was very hard. But then there were these two papers they came out, both from Sunney's lab actually. And they came out both, I think, in January of 2006, one in science, one in Nature, demonstrating not one way of doing this, but rather two ways of getting single molecule dynamics inside living cells.

So today, obviously, we're going to be primarily talking about this paper by [INAUDIBLE]. But if you're interested in these things, I encourage you to check out [INAUDIBLE] paper, which also published, [INAUDIBLE], at the same time. And that was based on a microfluidic assay, where instead of doing the single molecule fluorescence within cells, instead by trapping the cells in small volumes, and then using more traditional enzymatic assays, such as this beta [INAUDIBLE] assay, enclosed in a small volume, it's almost possible to study, once again, these sort of bursting dynamics in E. coli. And also they did it in yeast, and demonstrated that it's kind of a generally practical assay.

So if you're interested in those papers, I encourage you to check it out. But for me, this was really an eye-opening thing. So I graduated with my Ph.D. in December of 2005, and then I went to a conference in Cambridge, England, where where Sunney presented this work. And I think that it really blew many of our minds, this idea that you could start to get this sort of data within live cells.

And indeed, Sunney's group over the next five years did a whole series of what I'd consider to be beautiful studies, probing, for example, the dynamics of this [INAUDIBLE] repressor binding, unbindings onto this promoter, the search process. Yeah. A whole slew of, I think, really beautiful things. So we're not going to have the chance to go over all those papers in this class, but I encourage you to look at them.

Can somebody say what the primary challenge is with doing single molecule fluorescence in these live cells? So why is it that I did not think that this was going to work? And now, you're going to have to give an argument that ends up not being true. But why is it that this is such a hard thing to do. Yeah?

AUDIENCE: [INAUDIBLE] laser at the cell but you can't kill the cell?

PROFESSOR: Right. OK. So one is that there's laser, the cell, and then there's a big question mark. Is this going to be OK? And indeed, we certainly know that at one limit, it's not going to be OK.

If you take the lasers at Los Alamos National Lab, you can vaporize the cell. So it's certainly enough power, and the cell's going to be dead for sure. And so the question is maybe, oh, can you dial down the laser power enough to get-- and indeed, this is something that they talk about, their strategy in this paper.

Other challenges, problems?

AUDIENCE: Many molecules.

PROFESSOR: Right. so the principle of many molecules. So we have to figure out some way of separating them, either temporarily or spatially. Indeed in this paper, they actually do both. We say many molecules. And of course, we have to decide what we mean by this. Because ultimately, we're interested in doing single molecule measurements.

But then, of course, of the plural of single is many. How many is too many for us to study, and so forth? The question here is maybe like, how to separate, right? What are other challenges in this?

AUDIENCE: Diffusion.

PROFESSOR: Right. There's diffusion. So we're going to talk more about this, for sure. Well, actually, all of these things, we're going to talk about. Diffusion. And why is this a problem, though? Right. So it's diffusion is maybe fast. And so this is going to end

up being relevant for kind of signal to noise reasons. So what's the signal and what's the noise?

AUDIENCE: Autofluorescence.

PROFESSOR: Right. So there's autofluorescence. So in particular, this noise is autofluorescence from what?

AUDIENCE: From the cell.

PROFESSOR: From the cell. Right. And what's the signal, just to be clear here?

AUDIENCE: Photons from [INAUDIBLE].

PROFESSOR: Right. So it's photons from, in this case, the GFP-like molecule. Yeah? So we want to do single molecule measurements. We want to be able to measure or detect the fluorescence coming from this single fluorescent protein.

Now the question is, if it's a single molecule, does that mean that it's going to send out just a single photon? No. Maybe they come out as single photons. But we can detect them. Now the challenge in, in some ways surprising, is not that the number of photons is so small. Does anybody have any rough sense of maybe how many photons are we collecting from each of these?

AUDIENCE: Many thousands.

PROFESSOR: Yes. I'd say many thousands. In particular-- so we'll say many thousands. It could be even 10 to the 4 per second or so. It depends on the laser intensity. Many thousands of photons collected. Yes?

AUDIENCE: That's before [INAUDIBLE].

PROFESSOR: Yes. That's right. And indeed, the stronger the intensity of the laser light that you illuminate with, the faster you're going to collect the photons, but in general, it won't increase the total number of photons that you collect. So in these sorts of situations, you might get, say-- we'll say 10 to the 4, plus or minus in order of magnitude,

photons per second.

And they might last for, depending on how-- for 30 seconds or so. And of course, we'll look at the actual numbers in this paper. But in many of these situations-- times 10 to 100 seconds.

In this case, it actually bleached faster. Right. But we'll see. Well, you might be able to get also-- if you use organic dyes and other-- right. But this gives you some sense of that there's a fair number of photons that you could, in principle, collect from a single molecule.

Now, of course, you might be worried, well, these are the photons that you shine on your camera. But then your camera won't pick up all of them. The way that we think about this is by what's known as the quantum efficiency. Quantum efficiency tells us basically this is the fraction of photons detected.

But with modern cameras, actually, this thing is approximately one. So it's 0.9 maybe with modern cameras, which is for our purposes, basically one. Which means that you can detect, actually, the majority of the photons that are hitting your camera.

From that standpoint, the number of photons is not actually the problem. You can collect many thousands of photons. So the problem is really detecting that signal over the background signal. Over the autofluorescence of the cell.

And indeed, if you look at the figure, figure one, you can very clearly see the autofluorescence of the cell. So the fluorescence where there's the cell is indeed much larger than where there's no cell. The autofluorescence is, I'd say, the primary challenge here.

Of course, there are many, many others. Right. So this is potentially a big problem. And in order to get around that, there are all these other strategies that the author's going to implement. Yes?

AUDIENCE: What is the [INAUDIBLE]?

PROFESSOR: Yeah. So many things are weakly fluorescent, I think is the short answer. And this also depends upon, for example, cells are more autofluorescent if you grow them in a rich media than in minimal media, for some mysterious reason.

Yeah. But it's really just that there are many things that are weakly fluorescent. And it's just there are a lot of molecules in a cell.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah, right. In the case that you put in a fluorescent protein or a fluorescent dye, then it has rather well-defined absorption and emission spectra. And each individual absorber emitter indeed has a well-defined absorption emission profile. But then in the cell, there are just many, many, many of them, which means that there's rather what you want to call an absorption of broadband emission. So indeed, in some wavelengths, it's worse than in others. But it's not that you get well-defined peaks the way you do for a single kind of [INAUDIBLE].

AUDIENCE: I guess I was just wondering [INAUDIBLE].

PROFESSOR: Yes. And indeed, in this case, they're exciting with a laser. So at least on the excitation, it's as precise as you can hope for. And indeed, on the emission they will use a filter. So they're not going to be absorbed. It's probably a few tens of nanometers that they're looking at. So in that sense, they are filtering out, but still, there is autofluorescence.

All right. Now getting at this question of the single molecule fluorescence, the limitations, diffusion, and so forth, it's always valuable to have a sense of scale in anything that you're ever doing. So just let's wake up by reminding ourselves, how big is a protein? Right. So, typical protein. Typical protein. And for now, we'll say e.g. For example, GFP or whatever.

All right. We're not going to give you very much time to think about this. But I just want to make sure that we all keep track of senses of scale in the world. Ready. Three, two, one. All right. So we've got some B's, C's, D's. All right.

Wow. We got a lot of surprisingly wide wide range, actually. [LAUGHING] OK. Well, we also have the mirror image problem over here. OK, right. So, indeed, I'll say this is a typical protein size. So, it's a few nanometers. Depending on, there are some that get longer, especially if you're thinking about a long-- there are some structural-- well, you know. Of course, if you're talking about filaments, they can be-- but if you're talking about a typical globular protein, it's a few nanometers in diameter.

The question is, let's say that this is a fluorescent protein, or GFP. And now what we do is we look at it. And we're going to get some fluorescent spot. So this is plotting the intensity as a function of position. Right?

So this is the intensity I is a function of position X . Now the question is, what is going to be the size of the spot? We conveniently have some size scales up on the board. I'll let us think about this for eight seconds. All right. Ready. Three, two, one.

OK. We have a majority of the group that is saying that indeed, it's going to be D . So this is what's known as a diffraction limited spot. And this is a fundamental physics limitation, that if you are imaging something with light that is of some wavelength, λ , this is of order λ over 2-- it depends on numerical aperture or projective and so forth. But you know.

So order of λ . A little bit less maybe. And indeed, the wavelength of the light that's being used here is-- they're exciting with 500 something. And then let me just-- 514. OK. So indeed, what happens is that we shine in light. So this is the λ incident that is 514.

Now, there's going to be some GFP that looks like this. And then, we're going to get out λ emission. All right. Question, is the emitted light going to be equal? Is the wavelength going to be 514 nanometers? Yes or no? Ready? Three, two, one.

AUDIENCE: No.

PROFESSOR: No. λ emission. Is it going to be greater than or less than λ incident? Ready? Three, two, one.

AUDIENCE: Greater than.

PROFESSOR: Greater than. So this thing is greater than 514. Now, if you just have scattering off something-- so let's say that we had a gold particle, and we shine-- what's going to be the relationship between lambda incident and lambda emission?

Is it possible to have the same wavelength come out as you put in of some object? Yes. If you just have something, a mirror, you can get back-- so it is possible. But in general, there's going to be some dissipation. It's a question of how much and so forth.

But certainly for something like fluorescence, you have a higher energy photon, and you have a photon that's emitted. And of course, energy goes as $1/\lambda$.

Now, this is useful because you can actually spectrally separate things. I just want to highlight, though, that if this separation is of order 300 nanometers and our protein is-- that [INAUDIBLE] our GFP. Nicely drawn. Not even quite to scale. To scale, it's actually even smaller.

It's a factor of 100 in size. This thing is only 3 nanometers in size, 300 nanometers wide. Now, it's important to be clear about what this means, this diffraction limited spot. The first thing to note is that what it means is that if you have two proteins-- so for example, I add another one over here-- that it's going to be very hard to tell that we have those two proteins next to each other. Because the resulting fluorescence pattern will look essentially the same.

Will it be exactly the same? What's going to change? The intensity. The intensity, you expect to go up by a factor of two, absent some interaction between. You can, in principle, get interactions there. But let's for now assume that there's no interaction.

Right. Then the intensity wouldn't need to go up by a factor of two. But unless you're very careful about all of your optics and so forth, it's actually a challenge to use this intensity alone to distinguish these things. It's only-- so the statement with a diffraction-- you need these two proteins to be separated by something like λ

over 2 in order for you to start to see the separation.

Right. Because then you have something that looks like this. Something that looks like this. And then the sum of those two, indeed-- so you say, all right, well, it looks like there are two molecules there.

Now, I just want to-- and the notion of a lot of these so-called super resolution techniques is figuring out a way to distinguish these things, and we'll maybe say something about that in a moment. But I just want to highlight that if we come back to the situation where we have a single protein there. Now the question is, how accurately can we tell where that protein is, if we know that there's just a single protein there?

Now in particular, this size of the spot is telling us something, but it might not be quite as strong of a limitation as it appears at first glance. And that's because in this case, well maybe I'll bring it back, what we see is a big spot. 300 nanometers, kind of wide.

But if we see this and we know it's just a single protein there, I mean, could the protein be over here? No. If the protein were over here, then the spot would be over there right. So actually, even though the size of the spot is 300 nanometers, in principle, if you want to know where that protein is, if you know there's just a single protein, well, in that case what you want to know is, where's the center of that distribution?

And that problem, well, the width of the distribution is relevant. But there's something else that's also very relevant. And quite generally, if you measure some quantity n times, and you want to know-- so you're measuring the height of men entering the army-- if you want to know the mean, what is it that determines your uncertainty around the mean?

Right. There's the sample size. Now, does the width of the distribution enter? Yeah. So in general, your uncertainty in the mean is going to go with the width of the distribution, sigma of whatever, divided by-- what do I put down here? Root of N ,

where N is the number of samples that we take.

What this is saying is that as we sample this distribution more and more, does the standard deviation of the distribution, does it go to 0? No. These are all trivial statements, but I can't tell you how many times I see this getting confused.

OK. So if you measure many, many, many times, you get a very beautiful distribution. Right. The width of this region of the sigma, that you get very accurately. It's true also that your uncertainty in the width, that actually does go to 0. But the width of it, the width doesn't go to 0. The width of the distribution.

But your uncertainty in the mean, that goes down as $1/\sqrt{N}$. And what is N in the case our detection business here? The number of photons. Right?

Now, of course, in the actual experiment, we don't get precisely this distribution. Instead, it's kind of sort of quantized somehow spatially, because we're actually detecting it on a CCD chip. All right. So you can go and do the math, figure out everything. But actually, that's not as much a limitation as you might have expected.

In many cases, the pixel size on the image plane is actually something like 100 nanometers. So this distribution that measure, although in principle it looks like this, what you actually measure is something that looks like-- well, maybe I should-- something like that. Right. Because you have discrete pixels on the CCD.

And it feels that that should just kind of totally screw you. But if you go and do the math, you find it's not as bad as you might expect. So broadly, you do get essentially something that goes as $1/\sqrt{N}$, where N is the number of photons. And if you collect 10 to the 4 photons, that actually, it's a lot of photons.

So if we want to know the uncertainty in the center of our distribution, well, this thing, we might have something that's of order 300 nanometers here. We take the square root of 10 to the 4.

All right. So we get to divide by something like 100. And these are all very rough numbers. But the point is that we can get down to nanometer resolution in terms of

the uncertainty that which we know the mean of that distribution. Yes?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah, yeah. Right. So again, it's surprising. The thing is that even with just two, in principle, we can be very sensitive. Right? I mean, actually, sometimes people actually do just put it on like quadrant photo detector, where you really only get essentially binary information. But even with just two, if I say OK, well, it looks like this, or if it looks a little bit like that, if there's no error in our measurements there, then you can actually get that location very well. Yeah.

It's surprising. I'm not going to like. Yeah, but even with this quantization of some sort that's due to the CCD, you can still get down to nanometer resolution. Your resolution is worse than it would be if you knew actually exactly where each photon was hitting, but it's not very sensitive, actually.

And indeed, in the presence of-- and this is a highly technical comment. But in the presence of read noise and other kinds of noise in the CCD, actually, in many cases, it's actually better to have somewhat larger pixels, again, than you would expect.

So these balancing many different things. People have thought carefully about this stuff. But in the end, 100 nanometer pixel is actually fine. And just to be clear, it's 100 nanometers at the sample plane. So it's typically of order 10 microns size on the camera itself. So the physical size of each of the pixels on the cameras, 10 microns within a factor of 2, between 5 and 20, but then you get 100x typically magnification at the sample point.

So to be clear, 10 microns divided by 100 is 100 nanometers. Is everybody following? OK I don't want to-- all right. The key thing-- three, you can ignore. But the key thing to notice here is this thing that's here, which is nanometer resolution. And it's been known for decades that this is, in principle, possible and so forth. But I think that within the realm of single molecule biophysics, it was really popularized in some very nice papers by Ahmet Yildiz, et al, where they attached single molecules

onto the heads of various motors as they were walking along tracks, and showing that these motors were walking kind of like this, by catching fluorophores here, and then you could really just see it, see them walking it.

Any questions about why it's in principle possible to get nanometer resolution in this process?

AUDIENCE: Doesn't this assume that the protein is 100% static?

PROFESSOR: Yes.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah. Yeah, indeed. Right now, I'm assuming that this thing is constant. And the question is like, how much movement is a problem? And so then you have to-- you know.

AUDIENCE: [INAUDIBLE].

PROFESSOR: That's right. So indeed, often, you're trading off spatial resolution for temporal resolution. And also, the intensity of your laser, and so forth. But in these sort of experiments, I think that what they did is they slowed down the motors quite a lot. So it was limiting ATP.

So indeed, the motors in those experiments, I think they were taking steps of order every second or 10 seconds. I mean, it was as slow as you can go, and still, yes. And then at each location, I think they were collecting 10 to the 4 or a few 10 to the 4 photons. And incidentally in this case, these are the photons that are being collected. And typically, you would only be collecting 10%, 15% of the photons. Because the photons are actually being emitted everywhere. But you only collect the ones that go back to your objective.

All right. I just want to make one comment about the super resolution techniques that have been spreading. So the question here is, well, let's say that you have two proteins next to each other. What can you do?

Now, the basic idea of all these super resolution techniques is that if we know that we have a signal for only one protein, then we can actually figure out where it is. So what you need to do is figure out a way so you just have one at a time emitting. So there are various schemes to make it so that these proteins can either turn on or off.

Now, what you can do is if just one turns on, you got some photons, you say OK, this protein goes over here. Then if later, this other protein becomes fluorescent, now you can figure out where that is. And so you do this basic super resolution localization multiple times, and then you can identify where things are.

Yeah. So all the microscopy guys really like to have fun acronyms. So these guys, when they did it, they called it FIONA. So this is-- is it DreamWorks? Or this is where the green ogre like that and then the red head?

AUDIENCE: Shrek.

PROFESSOR: Shrek? All right. So Shrek. And so Fiona was the redhead. And this stands for fluorescent imaging with 1 nanometer accuracy. And then indeed, a group at UCSF then developed SHREK, which is simultaneous high resolution imaging-- something. OK. I can't remember how it ended. But, yeah.

So the super resolution techniques, they call them-- so Xiaowei Zhuang at Harvard called hers STORM, stochastic reconstruction of something or another. So this is the Zhuang method. And then Eric Betzig called his PALM, which stood still for something else.

I don't know. But in particular, Betzig-- there's a long history of the hard core microscopists, somehow like, developing their techniques in their garages. I don't know what it is, but there have been a number of these cases. And Betzig I think was one of them.

Now he's at Janelia Farm, HHMI, and has been developing all sorts of advanced microscopy techniques. So I think that that [INAUDIBLE] did not develop hers in the garage. But um--

AUDIENCE: And they're all based on the same principle--

PROFESSOR: Yeah. It's all about temporal. It's a question of how you're getting them to turn on and off.

AUDIENCE: It sounds like [INAUDIBLE] acronyms.

PROFESSOR: Oh do they also have?

AUDIENCE: ROSY, COZY, NOSY.

PROFESSOR: Yeah. Right, right. For the different sequences or something? Yeah, yeah. All of these acronyms, maybe I just never came up with a good one, so then I--

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yes. Indeed. I always like when somebody uses acronyms that I don't know, I always like to say, oh, all these TLAs are tricky, or whatever. And then I say, it's three letter acronym.

AUDIENCE: [LAUGHING]

PROFESSOR: I very much like self-referential humor. OK. So that's the idea of the super resolution techniques. Any questions about that before we kind of get back actually to the paper? OK. Now in this whole discussion, as was pointed out, we've been assuming that the protein is not moving around during our imaging time.

So one of the major challenges of doing this whole business in live cells is not only is there a lot of autofluorescence, but in addition, you can't necessarily wait 10 seconds to localize where this thing is, because it will have moved somewhere else right. And in particular, diffusion is a problem.

Can somebody remind us what the authors did in order to get around the diffusion problems? I'm sorry what was it?

AUDIENCE: [INAUDIBLE].

PROFESSOR: They attached it to the membrane. And why does that help?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah. That's perfect. Proteins diffuse slower. And I think depending on the organism, there's more or less diffusion and so forth right. But what they did is they anchored to the membrane to reduce diffusion. I'll just say it reduces diffusion.

And indeed, just from a back of the envelope calculation, you can convince yourself that you probably are going to need to do this. So in particular, let's ask-- in this paper, actually, they image the fluorophore for 0.1 seconds, right? Does that sound right? So the image collected, so Δt is equal to 0.1 sets.

So the question is, how far will a protein typically diffuse in 0.1 seconds? Well, this is why we have diffusion calculations. Right. First of all, the diffusion coefficient, we're going to talk more about diffusion in a few weeks. But you should also in principal be able to calculate how these things go.

So in general, this is going to be a kT over some γ , which tells us how hard-- so kT is thermal energy. OK. So, thermal. And at room temperature kT is around 4.1 piconewton nanometers in some unit. There are many different ways you can write that.

Whereas γ tells us just how hard it is to push something. In particular, if you push them with some force, it will move with some velocity. Now is this consistent with freshman mechanics? No. It's not. OK. Is that a problem? So why is it that I'm writing this?

AUDIENCE: Solvents exerting a force [INAUDIBLE]?

PROFESSOR: Right. So solvents exerting a force. That's true. But in the case of freshman mechanics, when we're pushing blocks, it's also true that the other things are exerting forces. The tables. But we still write down F is equal to MA . So it's not just that other things are exerting forces. Yeah?

AUDIENCE: They're always in a continuum?

PROFESSOR: Always in a continuum.

AUDIENCE: They're always surrounded by [INAUDIBLE].

PROFESSOR: OK. Yeah. It's always surrounded-- but I'm actually surrounded by fluid now too. You know, you can wave your arms and feel it.

AUDIENCE: [INAUDIBLE].

PROFESSOR: OK, right. So it comes down to the viscosity. Indeed, this whole thing about being a low Reynolds number, we're going to talk about this in much more detail in a few weeks, when we think about how bacteria swim, and so forth. But I just want to mention that this is because we're at this low Reynolds number, where the so-called inertial forces, like momentum are negligible.

Inertia forces are negligible. So then it's really, this is in some ways more like Aristotelian physics, but it ends up being true for small objects in viscous liquids. And indeed, this thing it scales as the radius. So in principle, we can actually calculate roughly how the diffusion coefficient is going to behave as a function of size. The object and so forth.

But I'll just tell you that for a protein size object in the cell, you might get something like 10 micron squared per second. Now, already just from units, you can see how the kind of typical diffusion distance has to scale with time. And in particular, you're going to get that the typical kind of distance that you go in a typical distance-- we'll say square root, because it's the square root of 2 times D times time.

So you can see if you multiply the time by the D, then you end up with a micron squared, so you have to take a square root to get something that's a characteristic distance. And indeed, this is the kind of math that I can do. So this is an order one micron.

And how big is an E. coli cell?

AUDIENCE: One micron?

PROFESSOR: One micron, roughly, right? So it might be a couple microns long. A bit less than a micron in width. And what this is saying is that 0.1 seconds, which is our exposure time on our camera, you would expect something like GFP to diffuse around roughly the cell volume. And maybe not the entire one, but a fair fraction of it. What this is saying is that the diffusion really would be a problem, even with this relatively short exposure time. Yeah?

AUDIENCE: [INAUDIBLE]?

PROFESSOR: Yes. So this is assuming that the cytoplasm has a viscosity that's maybe an order of magnitude larger than water. And that's just because the inside is chock full of proteins and so forth. Now, there's a lot of discussion of what the mechanism is of diffusion and transport inside cells. It may depend on the size.

It's a very complicated area. But for our purposes, this is a reasonable way to think about it. But indeed, the viscosity of the cytoplasm, you'd expect to be significantly more than the viscosity of water. Yeah?

AUDIENCE: [INAUDIBLE] lower the [INAUDIBLE] time?

PROFESSOR: Lower the?

AUDIENCE: Yeah. Like two orders of magnitude instead of one?

PROFESSOR: Right. So in principle, we could. There are technical issues on various sides. So of course, you have to say, oh well, a typical camera, just the shutter of opening, shutting, that actually has some limit. But you can get around that using kind of strob-- you know, there are fancy things you can do.

But there's just a more fundamental thing here, which is-- so this is already 100 milliseconds. If you go down to like say, 1 millisecond, then it's true that the protein won't be able to diffuse very far, but then you also just don't collect any photons.

The number of photons you collect scales linearly with the time, right? So at some

point, it's just that you really don't get very many photons. And then again, you have this extra problem of distinguishing the fluorescence from the autofluorescence.

And I just want to maybe mention one more thing. In Figure 1, you can actually see how big the fluorescence intensity of this Venus protein is, as compared to the autofluorescence. And you see that if you decrease that exposure time by even one order of magnitude, you wouldn't be able to see them over the background. Right? And that's true, even though they won't have had a chance to diffuse. It's just that you don't have enough signal.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Oh. Right. So you can also increase the intensity. Yeah, that's right. And there is some limit to how much you can increase the intensity of the laser, just because there is some cycling time of the protein in terms of, you excite it, and then it takes some time before it's going to emit. So that actually sets a fundamental limit.

Yeah, I don't know enough about the details of this in the sense of whether maybe it would've been possible for them to try to adjust these various parameters to do it in some other way. But these are all the things you have to consider. Question?

OK. So what we have now is some sense-- OK, we need to maybe anchor into the membrane to reduce the diffusion. All right. So they did that. And we'll maybe say something more about this anchoring process in a moment. But first, I want to make sure that we're all on the same page in understanding their arguments for why this is a single molecule that they're looking at. Can somebody remind us their primary evidence that this is a single molecule?

So question is, a molecule, we'll say single fluorophore, just to-- question mark. How do we know?

AUDIENCE: The intensity drops off.

PROFESSOR: Right. The intensity drops off suddenly. So if you look at the intensity as a function of time, what you see is that it looks like, and then. So it's actually more noise, but

this is just to-- now this is what it looks like for a single molecule, but we should also be clear of what it would look like if it were many molecules.

So this is if it's a single molecule. And this is what we call a bleach event. And so the molecule dies for one reason or another. So it goes, some oxygen reaction, something. Now the question is, what happens if it said there are many molecules?

AUDIENCE: [INAUDIBLE].

PROFESSOR: OK. Now, what we want to do is imagin-- let's say that we shined light on a bead containing fluorescence, containing fluorescent molecules. So I'm going to give us some options. It could be many molecules. OK. We could. All right. I'll give you a choice those three.

AUDIENCE: So we're talking about [INAUDIBLE]?

PROFESSOR: Yes. I'm asking if instead-- what happens in the microscope is you see a spot. And the spot is always huge. Right? 300 nanometers. So there could be one molecule there, or there could be 100. You could fit 1,000 in there. No problem. Right? 10 by 10 by 10 molecules? That still is only 30 nanometers. That's still much smaller than diffraction limited spot.

So of course, if you see something, and the spot is 10 microns in width, you'd be pretty confident that either your optics suck, or you're looking at many molecules. But if you see diffraction limited spot, then it's not so obvious. And so the question is, if you plot the intensity of a diffraction limited spot as a function of time, how do you know that it's a single molecule?

They claim, oh well, it's because of this. But it's always good to be clear. What would it look like if it were not a single molecule, but instead it were a collection of molecules> Let's go ahead and vote. Ready? Three, two, one.

So we have a fair number of different responses. It seems to be a split across the room is the only problem. So I'm not going to have you discuss, because I think your neighbors typically agree with you. But in this case, it's going to be C. And this,

it was an attempt of mine of drawing-- what is it? Exponential distribution. So you'll see exponential decay.

So this is typical of processes where something is happening at a constant rate over time, and then you're seeing this thing go away. So this could be, for example, radioactivity is the classic thing we always talk. These are random events. We thought the radiation coming off of some source is a function of time that's going to decay exponentially.

Similarly here, now this is-- again, intensity is a function of time. In the case of many molecules, we get this thing that look like C. Now, there's going to be some time scale here which is telling us about the typical time for this bleach event, which we're told is what?

250 milliseconds. And indeed, this is telling us, actually, that they're already illuminating these guys at pretty high intensity. Because 250 milliseconds is not that long. Yes?

AUDIENCE: [INAUDIBLE]?

PROFESSOR: Yes.

AUDIENCE: Is it really a function common photons [INAUDIBLE]?

PROFESSOR: It's not really a function of the-- it ends up being a function of the number of photons that are emitted, but that's basically because you're in kind of some ground state. You excite up to this other state. And then, you get this relaxation to a lower state.

So this is the energy of the absorbed photon. This is the energy of the emitted photon. The idea is that each time you go around this cycle, that's one emission cycle, there's some probability that's small-- 1 in 10 to the 5, or something like that-- that it reacts with oxygen, or something that causes it to go to the start. And it's in principle, irreversible state.

Of course, the dynamics of these things can be more complicated, but that's the [INAUDIBLE] way of thinking about it. What that means is that that's more or less a

constant number of photons that you're going to get out. There are many cases where this approximation fails.

This single step bleaching is kind of a classic signature of the fact that you're looking at a single fluorescent molecule. Now, there are secondary arguments for why this is a single [INAUDIBLE] Venus they're looking at was what?

What was their supporting evidence? Yes?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah. The intensity matched what?

AUDIENCE: [INAUDIBLE].

PROFESSOR: That's right. So what they said is, well, all right. This intensity matched what they measured on a slide. Just the molecule. But I'd say this really is, I would say, supporting evidence, because the intensity of the fluorescence can just be different in different environments.

I think that this is kind of thing that it's-- there are many ways that this can fail. Right, so I think that in general, we consider this to be the gold standard. All right. Now in their experimental setup, there's something that's, I think, very nice that they do, which is, if you look at the-- now this is the we'll say, laser illumination. Here, this is kind of on, and this is off.

What they do is every three minutes, they illuminate. And then, this is not to scale. This was 1.2 seconds. So we might want to even-- there's a separation in there. So they illuminate for 1.2 seconds, and they collect the light for the first 0.1 seconds. This is the period where they collect for 0.1.

Can somebody tell us why they might possibly want to do this? Shine more light on the sample than they need to? They're not going to analyze that data. So this is an intentional bleaching step. So this part here is to bleach. And we'll see that this is actually essential for the way that they're collecting their data.

Given everything that we've just said, you should be able to tell me what fraction of the molecules will not be bleached. That survive bleaching. Survive the so-called bleaching step. You can ignore my writing. You should be able to think about it on your own.

I'll go ahead, and I'll give you 30 seconds to think about this. And I think all of the information that you need is in principle written up on the board. Yes?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Right. I'm assuming that they start out being fluorescent. Now over time, we're illuminating them with this laser light. So eventually, they will bleach. But perhaps some of them have survived for that entire 1.2 seconds. So if you looked at it after this bleaching step, it would still be fluorescent. Maybe another 30 seconds, because--

Do you need more time? Let me see where we are, just so that we can get a sense of things. Ready? Three, two, one. All right. Great. We're roughly uniformly distributed between-- all right. Perfect. So this is an opportunity to turn to your neighbor and discuss.

You should, even with back of the envelope calculations in your head, be able to get roughly where this. But you're also welcome, if you want, to double check. Pull out your iPhone. You can use the Google calculator. I give you just a minute or two to turn to a neighbor. You should certainly be able to find somebody that disagrees with you.

[STUDENT CHATTER]

PROFESSOR: Why don't we go ahead and reconvene? It's OK if you're still kind of confused by this. I've partly doing this to encourage you to review your probability distributions, because we are going to dive into them rather strongly over the next week. And so it's good-- if you can't quite remember how these are going to work, it's good to start reviewing now.

All right. Can I just see where we are? Ready? Three, two, one. OK. So we have some B's and C's. All right. I'm pretty confident it's B, although I'm a little bit worried now. All right, so what's going to happen? So basically, the probability that this thing survives as a function of time is the same as, essentially, the decay and the overall intensity fluorescence for many molecules over time.

Because this plot here is really a plot of the fraction of the molecules that have survived as a function of time. Right. So indeed, the probability of survival for an exponential process like this is a function of time, t , is going to be equal to e to the minus t over some constant [INAUDIBLE].

And what is t in tau? Well, t is this time 1.2 seconds. So we have e to the-- we'll write it like an exponent. So now we have a minus. We have 1.2 seconds divided by the lifetime in this condition, we're told is 250 milliseconds. So we can write 0.25 seconds. So this is approximately e to the minus 5, which indeed, is equal to what I thought is 0.8%. So it's around 1%,

Are there any questions about this logic or this calculation? Yes?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Oh, no, no. Sorry, sorry. No, no. Bleaching, this is a result of chemical inactivation of the fluorescent protein that is kind of induced by being in this excited state. So the idea is that if you're shining light on this fluorescent protein, then it's going to bleach at a rate that is kind of-- distribution of say, lifetime's going to be exponential with a time constant of 250 milliseconds.

AUDIENCE: [INAUDIBLE].

PROFESSOR: That's right. That's right. But it's a one-step process though. That's what we see here. It's not that that individual protein is getting worse and worse as it's being used. But it's really just that there's some rate that it-- some probability, of course, time that it cycles that it is, we'll say for now, irreversibly inactivated.

There are also all these so-called blinking events, where fluorescent molecules can

kind of temporarily go into a non-fluorescent state, and then they return, but right now, we're talking about these irreversible steps. Are there any other questions about what I mean by bleaching, or this calculation?

AUDIENCE: [INAUDIBLE].

PROFESSOR: So my claim is that 0.8% is very close, is approximately 1%.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yes, sorry. I did something funny here. Yeah, yeah. So it's useful to just kind of go through these calculations. And actually, I think that in going through them, you kind of really get to get a sense of why they design their experiments the way they did and everything.

So the idea is that what they're doing here every three minutes, they're looking at the cells, and they're asking, is there are a protein, or maybe more than one here? And then they try to kill all those fluorescent proteins. And then they look again three minutes later. Are there any new proteins that were made. Yes?

AUDIENCE: [INAUDIBLE]. You can't see it [INAUDIBLE].

PROFESSOR: That's right. You don't--

AUDIENCE: Do you calibrate the intensity [INAUDIBLE].

PROFESSOR: So what they're really doing is they're asking, how many spots do I see? So in that experiment, they don't know that it was a single molecule. Although I think that in many--

AUDIENCE: Then why do they talk about this [INAUDIBLE]?

PROFESSOR: Well, that's how they checked to make sure it was a single molecule. Although I think that actually, in principle here, I think they actually maybe do continue to look at them. That's not part of their analysis. So for many of the case, they actually do see that this molecule is here, and then it bleaches, and a different molecule was

here, and it bleached.

But what they're really asking is at the beginning, how many fluorescent proteins did I see? And I think the camera actually maybe was collecting still during that bleaching step. It's just that it wasn't kind of part of it. Their analysis is in some ways really just based on this. Or in other experiments, you can go look to confirm that it bleaches single step here.

OK. So we spent a long time talking like the general idea of how to design these experiments and so forth. I'm not going to say very much about the design of the experiments, except that they did a number things. They used this Venus protein that has a faster maturation than traditional GFP.

They also targeted it to the membrane, not by putting Venus into the membrane-- that would be tricky, I think-- but rather by attaching the Venus protein to another protein that is put in the membrane. And indeed, this TSR membrane protein, we're going to be talking about it in a couple of weeks when we're discussing the chemotaxis network that is in *E. coli* for how *E. coli* find food and so forth.

I think that in reading these papers, it's interesting. Sometimes, authors make kind of a side comment that just illuminates kind of how difficult everything was. And I think that they had a nice one in here, where they said that they were checking with TSR to make sure that the behavior of the TSR Venus and just Venus were similar in terms of the amount of fluorescence.

And then, they say, no notable difference was observed, indicating that the introduction of the TSR sequence does not change the yield of Venus production, which is not the case for many other membrane targeting sequences that we tested.

So this is like, a little add-on onto the sentence that is like, I mean, six months of somebody's life was dedicated to trying-- you can just imagine all the over coffee, their frustration. They tried all of these different things, and they always got-- and for an awful lot of these things, I would have still been very much interested in the study, even if the addition of TSR did change the kinetics. Because I think that still is

very interesting.

But they really wanted this to be just airtight, or maybe the referees need to be-- I don't know. But you can tell that they just went to a lot of work to try to find the thing where everything would be just right. Now, once they kind of described their setup, they had this wonderful paragraph, I think. They say, oh, these proteins, they're generating bursts, and the number in the bursts varies, and there's spread, and so forth.

And they very nicely tell us, with this data, we can ask four questions. And they say, do these gene expression bursts occur randomly in time? That's going to be yes. How many mRNA new molecules are responsible for each gene expression burst under the repressed conditions? One.

What is the distribution of the number of protein molecules in each burst? It's going to be geometrically distributed. And what is the origin of the temporal spread of the individual bursts? And now, I think that this is nice, just to give a reader a kind of like, heads up of where we're heading.

And the origin of the temporal spread is actually-- they're arguing is actually the Venus maturation time. So in that case, the fact that there's a finite time for maturation of the Venus ends up allowing them to measure the bursts in an interesting way.

Before we get into the details of that, though, I want to make sure that we're all clear about what they mean by a burst. Because this is something that is oddly-- it feels like it's the most trivial statement ever. But I think what we're going to find is that there's a lot of confusion about it.

This is a question, how is it that you go from the data to the quantities that they plot and that they're interested in? We want to get a sense of how many proteins are made in each one of these bursts? And so they have in Figure 3B, they plot the number of protein molecules produced. Number of proteins produced. It's a function of time.

And they have these things. And This was a cell division event. And here they say, we have 2, 4. And here at 25 minutes, we have this thing here, and it looks like. And then this thing goes on. And there's 50, we have another, and so forth.

This is a zoom in of Figure 3B, the top panel. So what I want to know is, what is the size of the first burst?

So you can either look at my beautifully drawn illustration, or you can look at the paper in front of you. So this is a paper analyzing the size distribution of protein bursts observed in living cells. Right? That's the point of this paper. Now, the question is, from the data they're collecting, we want to know what is the size of the protein burst? The first protein burst.

Now, there's no calculation for you to do. There's not much of one. So I'm not going to give you maybe anymore time to figure this out. So let's see where we are. Ready? Three, two, one.

All right. So we got at least a majority of the group is saying that it's indeed 3. Now, the issue here is that the weight of the experimental design is working so that every three minutes, what we're asking is, how many Venus molecules kind of folded in that previous three minutes? And then, any of them that there are, we count, and then we kill them.

And then, the next. And indeed, what happened here is that every three minutes, they're asking this question. No proteins. And then here, they see one. Now, that's not yet a protein burst. That's maybe a protein verse. But it could be that we're in the middle of a protein burst. And indeed, what we see is that the next time point, the next three minutes, we see, oh, actually, now there's two new proteins that were produced in that next three minutes.

So indeed, this whole thing is a protein burst. So we got 1 plus 2. So that was the calculation I was referring to. And so what they're plotting is the distribution of these different protein burst sizes. Now, this is a small protein burst. They see some that get up to be 10, 15. And that corresponds to some of these cases, where they see

something that looks, for example, more like-- yeah, question?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah, right.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yes, exactly. And ultimately, first of all, the number of protein bursts per cell cycle, per hour in these conditions, is of order one. And then the width, the time, of a protein burst is five, seven minutes. Something like that typically.

So this gives you a sense of how frequently they will overlap. And indeed, what you expect from this is that 15% of them are actually that they see as one burst might actually have been two though. It's also worth mentioning that-- right. So what I just said is in the model, where you know that it's always just one mRNA that is produced each time, what they say is they think is that the promoter is tightly repressed by the Lac repressor.

Ever now and then, the Lac repressor falls off, and it's going to bind again. But some fraction of that time, when the repressor unbinds, you get the RNA p binding, and then you get a transcription event.

And I think that in general, it is just one mRNA that is produced there. So just a single RNA polymerase bound, and made an mRNA. But I'm sure that some fraction of the time, it was actually two that were produced during that time. And those would certainly show up as one protein burst. Right? Because the lifetime of the mRNAs in this situation is of order what?

Yeah. It was, I think, one and a half minutes maybe? It was short. Yeah. One and a half minutes. What that means is that on this time scale, if there were two mRNAs produced, they would look like the same mRNA.

But from this data what they conclude is that there's typically only one mRNA produced in each protein burst, and there's not that many protein bursts per hour of the cell division, so they won't overlap too much. But it's going to happen at some

rate. All right.

What they see is that the distribution of the protein bursts-- we said it was roughly one per cell division time, which they found was 55 minutes here. And they found that the number of protein bursts per cell cycle was distributed Poisson

So let me write this down somewhere. The number of protein bursts per cell cycle. So this was distributed as a Poisson with mean [INAUDIBLE] lambda of around 1.1.2. They call this n cycle. So I'll be consistent. So this n cycle to the 1.2.

Now, you guys-- the Poisson is a distribution that we're going to be spending a lot of time thinking about. So the normal way that we write it is that if it's the probability of observing some number n-- and this is a number n bursts per cycle in this case, p of n . We normally write it as a function of the mean lambda, where it's lambda to the n over n factorial.

And then for normalization, we have to write e to the minus lambda here. We're going to spend a lot of time thinking about the Poisson next class. So I would say that if it's been a while since you've thought about probability distributions, then you should play via textbook, Wikipedia, whatnot, with the Poisson, the exponential, the geometric, and also the gamma distributions. Because we're going to be using those in the next class.

Now, in this distribution, what it basically ends up being is that sometimes, you see zero bursts. Sometimes you see one. Every now and then, you see two. It's kind of what this means.

There's one other thing that is, I think, a bit tricky often, which is how they calculated that it was typically one mRNA that led to each of these proteins bursts. Can somebody remind us kind of experimentally what they had to do in order to get at that?

The average RNA per cell, right. Right, so they did this RT-PCR. So what they did, they reverse transcribed. They converted the mRNA into DNA, and they amplified to

get a sense of how much mRNA there was. And from that, the formula, when you first look at it, it feels kind of mysterious, or something like that.

But it's one of those things that you just have to keep track of like, units and so forth. So you can basically think about the number of mRNA. And this is indeed, this is per cell. But cell, this doesn't have units, right? But if we wanted the expectation value, the number of the mRNA that's going to be per cell, well, that's going to be given by the number of the mRNA per burst, times the number of bursts per unit time.

So this is some rate burst, and then also times the lifetime of the mRNA. Now in this formula, there's also the added factor, where they have like, the time of the cell cycle. But that's just because this could have been bursts per minute, lifetime of mRNA minutes.

But then, if you want to put in the extra term, then you have to say, oh, the cell cycle is 55 minutes. So then you have to do that conversion of time into the proper units. So that's what ends up happening.

Are there any questions about what happened here? Now, what we're going to do next lecture is kind of go through a simplified model of gene expression, where there's just some rate of mRNA formation, mRNA degradation, the mRNA makes protein, proteins get degraded. And then in that model, we want to try to understand how everything is distributed.

And we're going to relate that back to some of the experimental data in this paper. In particular, for example, this geometric distribution of protein burst sizes is something that you expect from the most basic simple model that you would have written down. So from that same point, it's not a surprise. It is often assumed this thing should be geometrically distributed, and it was. And that's wonderful.

From my standpoint, I think that even things that we assume to be true, we should still check to see if they are true. And in other cases, they may not be, and so forth. Are there any questions about this paper? No? OK. Then I will see you our next class.