PROFESSOR: OK, in that case, let's begin in our usual way by going through a review of last time's lecture. Last time, we talked really about two calculational problems. One was the calculation of the age of the universe, taking into account a universe model which has matter, radiation, vacuum energy, and curvature. And we got the general formula. And then for the same type of cosmological model, we also calculated how one finds the brightness of a distant source-- the energy flux in terms of the redshift of that source.

So first, the age of the universe calculation-- that really just depends on the first-order Friedman equation, which I've rewritten here. We put three terms on the right hand side for the mass density-- a matter term, a radiation term, and a vacuum energy term. And we know-- and this is the important ingredient-- we know how each depend on the scale factor. Non-relativistic matter falls off like 1 over the cube of the scale factor. Radiation falls off like 1 over the fourth power of the scale factor. And vacuum energy is just constant.

Next step that we did was just to rewrite this equation, where we put in the explicit time dependence in the form of this x which is the ratio of a of t to the present value of the scale factor-- a of t 0. And furthermore, we expressed the matter density in terms of the present contribution to omega. And rewriting equation in that language, it takes that form.

And then, I pulled a fast one. I said we could also write this last term to look pretty much like the others. It just is a constant that falls off like 1 over a squared. So if you define omega sub k 0, which is exactly what you need to make this look like that, and in terms of omega sub k 0, all four terms have the same characteristic. They're just a constant times a power of x. So this is, then, the rewriting of the Friedman

equation one more time, just using this new definition of how we're going to treat the curvature of the universe.

And simply by looking at this formula and applying it to x equals 1, you can see that that becomes then 1 is equal to the sum of these omegas. And that can be thought of as a clearer, perhaps, definition of what omega sub k 0 is. It's just 1 minus all of the other contributions to omega. So it's how much the actual mass density of the universe differs from the critical density.

Then once we have this equation, which is the equation which tells us what x dot is as a function of x, we could just rewrite that by bringing dt to one side of the equation and dx to the other and integrating both sides. And that leads to our final result. The age of the universe is simply given by that integral.

And this is a very neat expression for the age of the universe in terms of the present value of the Hubble expansion rate and each contribution to omega in terms of its present value. And you just plug those into this formula. In general, you have to do the integral numerically, because the integral's a little too complicated to have an analytic expression. And that will give you the age of universe for any model that meets this description.

So any questions about that calculation before we go on? OK, very good.

The next calculation we did last time was the calculation of radiation flux versus redshift. And this is exactly what the astronomers were measuring in 1998 when they concluded that the universe was accelerating. They were looking at distant supernova type 1a explosions.

They made the assumption that all supernova type 1a explosions have the same intrinsic power output. That's based roughly on observation and guesswork. There's not really a good theory for it, so it's mostly a matter of being consistent with observations. But then they could calculate for any given model in terms of these different omegas what you expect in terms of received radiation as a function of redshift. And they compared their data with the models-- and I'll show you that data

shortly-- and found that the models only fit if one had a significant component of vacuum energy causing universe to accelerate.

So to do the calculation, we need a metric for the universe. And I considered only the closed universe case. There's also the flat case and the open case, which are similar. And you'll actually be asked to do those on the homework set.

So the metric for a closed universe can be written this way, where sine psi is the square root of k times r, to relate it to the other way-- the more standard way-- of writing the Robertson Walker metric. But for our purposes for this calculation, it's easiest to do it this way, because we're going to be interested in radio trajectories of photons. And this metric simplifies the radial direction as much as it can be. It's just d psi squared.

Oh, it's the computer that froze. You never know with Windows. I think we're in business now.

Back to where we were. We have the metric. Now what we want to do is imagine a light source being received by a detector. And we put the light source in the center of our coordinate system. We put the detector at some distance corresponding to psi equals psi sub D, where psi is our radial coordinate and psi sub D is the radial coordinate of the detector.

We imagine a whole sphere with the same radius as the detector, because we expect the source to be spherically symmetric. And therefore, the light emitted by the source will be uniformly spread over that sphere. And that will allow us to calculate how much of it will hit the detector.

The fraction hitting the detector will just be the area of the detector divided by the area of the sphere. The area of the detector is whatever it is. We call it capital A. The area of the sphere is 4 pi times the radius of the sphere. And the radius of the sphere in physical coordinates is the scale factor squared times the sine squared of psi sub D, coming from the metric. It's the radius that appears in the angular part that counts, because it's the angles that we're integrating over to get the area of the

sphere. So the radius is just a tilde squared times sine squared is the radius squared.

Then we also need to remember something we've said a number of times previously in this class, which is that when the photons travel from the source to the detector, their intensity is suppressed by two powers of 1 plus z, two powers of the redshift. And one of those factors in 1 plus z comes from redshifting each photon. The frequency of each photon is redshifted, and that means that the energy of each photon is redshifted-- goes down by a factor of 1 plus z.

But in addition, the rate of arrival of the photons is essentially a clock which is also time dilated. So the rate of arrival of the photons as seen by the observer is suppressed by another factor of 1 plus z. So putting all that together, the received energy flux, which is the power received divided by the area, is just the power emitted by the source divided by 4 pi. We get this factor of 1 plus z squared, due to what we just discussed. And then the a squared of t sine squared psi sub D. So it's just the total power times that fraction that we receive times the two factors of 1 over 1 plus z.

And this then is essentially the final answer, except we want to know how to evaluate a tilde squared of t 0 and sine squared of psi sub D in terms of things that we more directly measure. So to do that, a tilde of t 0 turns out to be easy, because it really is just related by the definition of omega sub k 0 to omega sub k 0. So this formula is just a rewriting of the definition of omega sub k 0.

To figure out what psi is, we want to integrate along the line of sight to be able to figure out the time of emission in terms of psi. And that time of emission could then be related to the redshift, because the redshift is just the ratio of the scale factors between reception and emission. So we look first at the metric. And say we're going to be looking at null geodesics in the radial direction. And null means ds squared equals 0, and that's minus c squared dt squared plus a tilde squared of t times d psi squared. And that implies immediately that the psi dt is just equal to the speed of light divided by a tilde.

And then we can get the total increment in psi between the source and us by integrating between the time of emission-- the time of the source-- to the present time-- t sub 0. And then it's just a matter of changing variables to express the variable of integration. Instead of st, we could express it as z-- the redshift itself. And that brings in a factor of h, because h is a dot over a. And I showed the manipulations last time, but it brings in a factor of h.

But we know what h is as a function of z. It comes from the Friedman equation. And that then gives us an expression for psi of z sub s as an integral over z. And writing in what h of z is and what a tilde of z is from that expression, the expression for psi of z becomes the equation that's boxed. Just a matter of algebraic substitutions involving the Friedman equation, which determines what h of z is.

And then putting everything together, J is just given by this expression, where all I've done is to substitute the expression for a tilde of c 0. And sine squared psi is still here, but it gets evaluated according to that formula. And putting these together, we have a complete calculation of the received radiation flux as a function of cosmological parameters-- the omegas and the h 0-- and the redshift of the source. And that's the end of the calculation. And that's where we finished last time. So any questions about that calculation? OK, fine.

In that case, moving on, the next thing I wanted to show you was some real data. So here are some real data from one of those two teams that made the original announcements in 1998. This is from the High-Z Supernova Search Team. And I should write some definitions on the blackboard.

The vertical axis there is essentially brightness. But you wouldn't expect the astronomers to just call it brightness, because they like to use fancier words. So they write it as little m minus capital M-- measured in magnitudes, they put in parentheses. And little m minus capital M has the name, it's called the distance modulus, meaning it's a way of measuring distance. They think of brightness as a way of measuring distance, which indeed is what it's being used for.

And it's defined as 5 times the logarithm base 10 of d sub L over 1 megaparsec,

which means the luminosity distance-- I'll define this in more detail in a second-- d sub L is the luminosity distance-- distance as inferred from the luminosity. And they're measuring it in megaparsecs and taking the logarithm base 10. And then, by convention, there's an offset here of 25. Why not? So this is the definition of the distance modulus.

And d sub L is defined by the relationship of what J would be in a flat Euclidean universe if you were receiving that luminosity. So J is equal to the actual power output of the source divided by 4 pi d sub L squared. This defines d sub L. So d sub L is the distance that that source would have to be at in a static Euclidean universe for you to see it with the brightness that you actually see.

This, I guess, completes the definitions, but we can put these together. And m minus M is then equal to minus 5/2 times the logarithm base 10 of 4 pi J times 1 megaparsec squared, divided by the actual power output of the source, and then, of course, plus 25. So this relates this distance modulus to the energy flux and the power output of the original source.

There's also on this slide the acronym MLCS. MLCS stand for multi-color light curve shape. And what that refers to is the High-Z Supernova Search Team invented a method of compensating, to some extent, for small variations in the actual power output of the supernovae type 1a. Instead of assuming that they all have exactly the same brightness, they discovered by looking at nearby supernovae of this type that there's a correlation between the absolute brightness of the supernovae and the shape of the light curve-- that is, light versus time. So they were careful to measure the light versus time for the supernovae that they used in this study. And they used that as a way of applying a small correction to what they interpreted as the intrinsic brightness of each supernova.

And the results are these points. [INAUDIBLE] the top are the raw points, and three different curves for three different models. And they characterize the models in the same way we would-- in terms of different contributions to omega. So the top model is the cosmological constant dominated model, where omega sub lambda, which is

what we've been calling omega sub vac is 0.76. And it's a flat model, so 0.24 for omega matter. And radiation is ignorable.

They compared that with the middle model of these three, which was a model that had no vacuum energy, and omega matter of 0.2. That was essentially the dominant model at the time, the belief that the universe was open then had about a critical density of 1/5 or 1/4. And then they also compared it with a model where omega was 1-- entirely of matter with no vacuum energy. And that was this dashed curve, which is the lower of these three curves.

And when the data is just plotted, it's a little hard to see how much difference there is between the three curves. So they re-plotted the data, plotting the middle curve as a straight line by construction. And then they plotted deviations from that line. And they did that for both the theoretical curves and the data. And in this magnified picture, you can see a little bit better that this top curve fits things the best.

And that's what they call the lambda CDM model. It corresponds to omega m equals 0.24. Omega lambda equals 0.76. So it's the model with a cosmological constant, with a vacuum energy. And lambda CDM stands for lambda and cold, dark matter. And cold, dark matter is just what we've been calling non-relativistic matter.

So the claim is that these data points, even though there's a fair amount of scatter, fit the top curve much, much better than they fit the middle curve or the bottom curve. And statistically, that's true. It really is a much better fit, even though by eye, it's not that clear what's going on. I think by eye it looks clear that the top one fits it better than others, but it's not that clear how important the difference is. But nonetheless, the astronomers were thoroughly convinced that this was a real effect.

There was considerable discussion about possible systematic errors. And I guess next, I'll say a few words about that. First of all, I should maybe just clarify a little bit better what's being seen. What's being seen is that for a given redshift, this curve, which basically shows brightness in a funny, funny way, where dimmer is upward, larger values of little m minus M-- there's a minus sign in this formula-- means a dimmer galaxy, one that looks further away. And basically, when astronomers see

this, they think distance. So larger values means further away.

So what's being seen is that these distant supernovae are a little bit dimmer than what you would expect in either of the other two models, either of the models that do not have vacuum energy. And the amount by which they're dimmer is a few tenths of a magnitude. And each tenth of a magnitude corresponds to about 10% in brightness. So what they're saying is that these distant if supernovae, if we assume they really fit this curve, are 20% to 30% dimmer than you would have expected in other models.

It might be worth saying a little bit about why dimmer is the right sign to correspond to acceleration, which is by no means totally obvious, I don't think. So we're plotting--

AUDIENCE: What year is this?

PROFESSOR: What year? This was old data that was published in 1998. It has gotten better. Now it's much more unambiguous that this works.

So this is distance as inferred by brightness. So this is basically what's being plotted. If one thinks about a fixed z in which way that you go-- up or down-- I find that totally cryptic. I don't really parse that very well in my own head. But it's much clearer if you think about the other way. You could think about a galaxy-- or a supernova in this case-- at a fixed distance, and ask, suppose I compare different models-- ones that accelerate and models that don't accelerate.

So if we fix the distance and say, what would we expect for the redshift of a given galaxy, in an accelerating model versus a non-accelerating model-- remember, the redshift is basically a measure of the velocity, or at least it's strongly influenced by the velocity of the object. So if the universe is accelerating, it means that the universe was expanding slower in the past than you would have thought otherwise. It's speeded up to reach its present expansion rate. So an accelerating universe is a universe that was expanding slower in the past. And slower in the past means that a galaxy at a given distance would have been moving slower, and hence would have

had a lower value of z.

So the effect of acceleration for a given distance-- we'll fix the distance-- should be to move the line that way, towards lower z. And by moving the dot that way, it puts it above the curve. So it's the same as shifting things up, which is the more natural way of describing what's seen in the graph. The points are higher than the curve.

So the bottom line, though, is that what they're saying is distant supernovae are 20%, 30% dimmer than you might have thought. And from that, they want to infer that the universe is accelerating, which is a rather dramatic conclusion. So naturally, you want to ask, are there other things that can cause supernovae to look dimmer? And of course, there are other things that can cause supernovae to look dimmer than you might have thought.

And there are two main ideas that were discussed at the time. One of them is just plain dust. If you're looking at something through a dusty atmosphere, it looks dimmer than it would otherwise. And that is a genuine possibility that was strongly considered.

The arguments against dust were mainly twofold. The first is that dust very rarely absorbs uniformly across the spectrum. Dust usually-- depending on the size of the dust grains-- absorbs more blue light than red light, leaving more red light coming through. So the effect of seeing something through dust is normally to cause it to look more red.

And this reddening was not seen. The spectrum of the light from the existing supernovae was analyzed very carefully. And the spectrum of the distant ones looked just like the spectrum of the nearby ones-- appropriately redshifted, of course, but otherwise not distorted in any way. There was no sign of this reddening.

Now, it's possible to have what the astronomers refer to as gray dust, which is by definition dust that absorbs uniformly across the spectrum of what you're looking at. But the grains have to be unusually large. And nobody was ever able to figure out a source for dust grains of that sort. So based partly on theoretical grounds and partly

on what nobody has ever found, there's no evidence for dust grains that would possibly cause dimming that would look this way, that would be dimming that was uniform across the spectrum. Yes?

AUDIENCE: How do you tell the difference between reddened light from dust and redshifted light?

PROFESSOR: OK, how do you tell the difference between reddened light from dust and plain old redshift? The difference is that the plain old redshift uniformly shifts everything by the same factor. So the whole spectrum is just moved down uniformly towards the red. This reddening effect really means that the blue part of the spectrum is depressed relative to the red part. So the shape of the spectrum is changed.

So one argument is that we don't see reddening, and we don't know any way to make dust that would be gray. The second argument is that if dust was a major factor, presumably most of the dust that would be relevant would be dust in the same galaxy as the supernova explosion itself, because there's not that much dust in intergalactic space. And if dust in the galaxy of the supernova itself were relevant, then-- let me draw a little picture here.

So if dust in what's called the host galaxy-- the galaxy which has the supernova in it-- then you would have a picture where there would be a ball of dust filling the galaxy. And the supernova that you're looking at might be there, or it might be there. And let's say we're looking from over here. So depending on where the supernova was in the galaxy, we would see very different amounts of intervening dust. And if dust were causing this dimming, it would mean we would be seeing a significant scatter in the amount of dimming depending on where the supernova happened to be in its host galaxy.

And that spread was not seen. The spread that one sees in that curve could be measured and calibrated against known uncertainties in the brightness of supernovae and then the detection apparatus. And the spread that was seen was just what you expect without any additional spread associated with a dusty galaxy acting as the host. So no evidence for the spread of brightnesses that would be

expected from a dusty host.

Another item that was considered-- these are the main arguments against dust-- another argument that was considered, another possible source of dimming, is galactic evolution. And there, the main effect that people worried about was the production of heavy chemical elements during the life of a galaxy. As you've certainly learned about from your reading-- I don't know if we've talked about it in class or not-- the early universe was essential all hydrogen and helium. Heavier elements were made later in stars that produce supernovae explosions. And these supernovae explosions gradually cause galaxies to become more and more enriched with heavy elements. And by heavy, I mean anything heavier than helium. And that could affect, in principle, the behavior of supernovae explosions.

So the evidence against that was simply that every other characteristic that astronomers could measure of these supernovae in the distant galaxies looked exactly like what was seen for nearby galaxies. So no evidence for any kind of evolution was seen. And there are many properties you could measure that are independent of distance, like the shape of the spectrum and things like that, and the pattern of the light curve versus time.

So all those characteristics that astronomers can measure seem to be exactly the same for the very distant supernovae which happened billions of years ago, and the more nearby ones that happened recently. And furthermore, among the nearby ones, there's a big spread of abundances of heavy chemical elements, just because different galaxies have had different histories. So among the nearby ones, you could look for is there an effect caused by the relative abundance of heavy elements, and astronomers didn't find any. So there was no sign that galactic evolution could be playing a role here, even though one does need to worry about it.

So the point is that distant supernovae 1a look like nearby ones. I'll call that a in my outline. And b is that among the nearby 1a's, heavy element abundance had no perceptible effect. So the dominant opinion gradually shifted, and now I think it's

almost 100% that this acceleration is real. The acceleration, by the way, is further confirmed by measurements of fluctuations in the cosmic background radiation measurements that have been done by some ground-based experiments, and also the satellite experiments of WMAP and now Planck, which measure the anisotropies-- the ripples-- in the cosmic background radiation.

It's hard to see what those ripples would have to do with the amount of vacuum energy. But it does turn out-- and we'll talk more about this a little bit later-- that we really do have a detailed theory of what makes these ripples. We can calculate what the spectrum of those ripples should look like. And the calculations depend on parameters which include the amount of vacuum energy. And in order to make things work, one does have to put in essentially exactly the same amount of vacuum energy as has been detected in these supernova 1a observations. So everything fits together very tightly. And I think now, just about everybody is convinced that the universe really is accelerating.

The acceleration could, in principle, have at least two different causes that we can talk about. One is vacuum energy, which is the one that I'm focusing on, which is the simplest explanation. The other possibility that is discussed in the literature is something called quintessence, which is a made-up word. And what it refers to is the possibility that the acceleration of the universe today could be caused by a mechanism which is really in principle exactly the same as what we talk about for inflation in the early universe and will be talking about later.

Specifically, there could be a slowly evolving scalar field which is essentially uniform throughout the universe, and changing slowly with time so it looks like it's a constant. And it could be the energy density of that scalar field that is looking to us as if it were vacuum energy. But that's the minority point of view. And that introduces extra parameters that don't seem to be necessary. But it's up for grabs. Nobody really knows.

OK, any questions about what we just talked about? In that case, let me go on to my next topic, which is I want to talk a little bit more about the physics of vacuum

energy. What is it that we understand about it, and why is it that most physicists say it's the least understood issue in physics? We really don't understand vacuum energy, even though we do understand why it might be nonzero. Where we're totally at a loss is trying to make any sense out of the value of the energy density that is actually observed.

So where does vacuum energy come from in a quantum field theory? There are basically, I would say, three contributions. Maybe I should say in quantum field theory.

The other context in which this might be discussed would be string theory. I may or may not say something about string theory, but I won't say much. But in quantum field theory, there are basically, I think, three contributions. The first is the easiest to understand, which is quantum fluctuations in bosonic fields, where the best example is the photon, or the electromagnetic field.

Now, in a classical vacuum, e and b-- the the electric and magnetic fields-- would just be 0, because that's the lowest possible energy density. But just as you are probably aware that there's an uncertainty principle in quantum mechanics which tells you that the momentum and position of a particle cannot be well-defined at the same time, it is also true that e and b cannot be well-defined at the same time. So the uncertainty principles applied to the field theory imply that e and b cannot just be 0 and stay 0. E and b are constantly fluctuating. And that means that there's energy associated with those fluctuations. And the mathematics of it is actually incredibly simple.

If one imagines the fields inside a box, to be able to at least avoid the infinity of space, the fields inside a box could be described in terms of standing waves, where each standing wave is either a half wavelength or a full wavelength across. And by the way, you'll be doing a homework problem on this. And each standing wave has the physics of a harmonic oscillator. It oscillates sinusoidally with time, the wave. And when one works out the mathematics, and even the quantum mechanics, it's exactly the same as a harmonic oscillator.

So each standing wave has a zero-point energy. You may know that the zero-point energy of a harmonic oscillator is not 0, but it's 1/2 h bar omega, or 1/2 h nu, depending on whether you're using nu or omega to describe the frequency of the oscillator. So each standing wave contributes 1/2 h bar omega.

And then the problem is how many standing waves are there? And the answer is, there's an infinite number of them, because there's no limit to how short the wavelength can be. So there's no limit to how many ups and downs you can have in your standing wave from one end of the box to the next. So the answer you get is infinite. It diverges.

Now, the fact that it diverges at short distances can be used as an excuse for getting the problem wrong. Obviously, it's wrong. The answer's not infinite. But we have an excuse, because we certainly know there are wavelengths that are short enough that we don't understand the physics at those length scales anymore. We're basing everything on extrapolating from wavelengths that we can actually measure in the laboratory.

So one could imagine that there's some wavelength beyond which everything we're saying here is nonsense, and we don't have to keep adding up 1/2 h bar omega anymore, because the arguments that justify the 1/2 h bar omega no longer apply. So we can use that as a cutoff for the calculation. And a typical cutoff-- by typical, I mean typical in arguments that physicists talk about, so typical in physics speak. So a cutoff that's often invoked here is the Planck scale, which is the square root of h bar times G divided by c cubed. And that has units of length, and it's equal to about 1.6 times 10 to the minus 33 centimeters.

And what makes the scale significant is it's the scale at which we expect the effects of quantum gravity to start to be important. And we know that this quantum field theory that we're talking about does not include the effects of gravity. And we don't really even know how to modify it so that it would include the effects of gravity. So the quantum effects of gravity are still something of a mystery. So it makes sense to cut the theory off, if not earlier, at least at the Planck scale. Yes?

AUDIENCE: So I would imagine what we're doing, in order to say that we have a standing wave, we have to have a box. And then in order to realize the fact that the universe may be large, you just take the limit as the box gets large. But is it really OK to do that? I mean, to treat an infinite system as the limit of a finite system?

PROFESSOR: OK, the question is-- what we're going to be doing here is I talked about putting the standing waves in a box. And then at the end, we're going to take the limit as the box gets bigger and bigger. And the question is, is that really a valid way of treating the infinite space? And the answer is, in this case, it is.

I'm not sure how solid an argument I can make. Certainly what one does find is what you'd expect, that as you make the box bigger and bigger, the energy that you get is proportional to the size of the box. So you're calculating an energy density. And probably the most precise thing I can say at the moment is that if it were not true, if the answer you got really depended on the way in which the space was infinite, then you'd be learning something about the infinite universe by doing an experiment in the lab, which is a little far-fetched. That is, if you do an experiment in a lab, it really doesn't tell you anything about whether the universe is infinite or turns back on itself and is closed.

And calculations certainly do show that you get the same-- you could do, for example, a closed universe without a box. And you get the same energy density, as long as the universe was big, as we're getting this way. So I think there's a pretty solid calculational evidence that what you get does not depend on the box. Yes?

AUDIENCE: Going off that question, do we use the maximum size of our box as the size of our observable universe, then?

PROFESSOR: OK, the question is, what do we use as the maximum size of the box? Is it the size of the observable universe? The answer really is that what you find is that you get an energy density that's independent of the size of the box, as long as the box is big. And it's that energy density that we're looking for. We don't claim to know anything about the total energy. And we don't really need to know anything about the total energy. Everything that we formulated here in terms of energy densities.

Now, the catch is, that if one puts in this cutoff and takes into account only the energies of 1/2 h bar omega going up to this cutoff and stopping there-- or down to the cutoff if one's thinking of length as the measure-- you could then ask, do we get an energy density that's in any way close to what the astronomers tell us the vacuum energy actually is? And the answer is emphatically no. We don't get anything close. We're in fact off by about 120 orders of magnitude, which even in cosmology is a significant embarrassment, which is why physicists consider this question of the vacuum energy density to be such an incredible mystery. We really have no idea how to get a number as small as what we observe.

Let us go on to talk about other contributions, because they are certainly important in the way we think about things. So far I have number one, right? So next comes two. And that is the quantum fluctuations of Fermi fields, where the best-known example here is the electron.

Now, in quantum field theory, I should point out that all particles are described by fields, not just the photon. The electron is described by a field also. It's called the electron field. And because the electron is a fermion and not a boson, the electron field has somewhat different properties than bosonic fields, reflecting the fact that the fermions themselves obey the exclusion principle.

It turns out that for fermions, there are also quantum fluctuations. They're also of order 1/2 h bar omega. But actually, it's a little bit different. They're in some sense h bar omega and not 1/2. But what's peculiar is that for electrons, the contribution is negative.

And the origin of this negativity I think has a fairly simple explanations, although the explanation is not ever given, actually. The exploration that's used in quantum field theory books involves looking at equation 47 and seeing that there's an anticommutator there. And because the fields anticommute, there's a minus sign. And that means the energy is negative. And that is basically the way it's described in textbooks. That certainly says where the minus sign appears in which equation. But I don't think it's really an explanation of what the minus sign is talking about.

But I think there is an explanation of what the minus sign is talking about, which goes back to the old picture that Dirac himself introduced when he first invented the Dirac equation. When Dirac first invented the Dirac equation, he was trying to interpret it more or less in the same language as the Schrodinger equation. We don't quite do that anymore.

But in doing that, Dirac discovered that his Dirac equation, which was the natural relativistic generalization of the Schrodinger equation to a particle which has spin 1/2, which I'm not sure how Dirac knew it had spin 1/2, but in any case, it's the equation for a particle of spin 1/2. And what he found was that if you just look at the energy the spectrum that the equation itself gives you, it's symmetric about 0. So if we plot energy going this way, if there's a state here, there's also a state there at negative energy. And if there's a state there, there's another state there, exactly opposite it. It's completely symmetric up and down.

Now, the interpretation that Dirac gave to that was not that there are a lot of ways of making negative energy. He realized that the vacuum is by definition the state of lowest possible energy. And if you can lower the energy by adding a particle to these negative energy states, that would mean that there'd be a way of lowering the energy, and the state would not be the vacuum.

So the vacuum, Dirac proposed, is the state in which all of these negative energy levels are filled. And the action of putting all these x's on the picture is often called filling the Dirac sea. S-E-A-- sea, where sea refers to this ocean of negative energy states, which is infinite. It just keeps going down. You can imagine filling all of them to describe the vacuum.

Then if you ask what is the physics after you've done that-- what are the possible excitations of the vacuum, what states does this theory contain other than the vacuum? And the answer is that there could be occupations of these positive energy states, and those are called electrons. It's also possible to remove-- if you put in the right amount of energy-- one of the negative energy states, which is filled, but we could take away the particle that's there. And the absence of a particle there-

- a hole in the negative energy sea-- is a positron. So electrons are there. The e plus is a hole in the Dirac sea.

Now, the difficulty with this picture, and the reason why it's not often use these days, is that it makes it look like there's an intrinsic difference between electrons and positrons. Nonetheless, Dirac was perfectly aware that when you went through the math, they were completely symmetric. The fact that you described it this way is just really a feature of your description, but it doesn't make any measurable difference. So a positron really is just a perfect image of an electron, but with the opposite charge, with otherwise all the same physical properties. And there's ways of describing it where you don't make this distinction between particles and holes.

But the particle hole way is I think the easiest way of understanding where the negative energy is coming from. The negative energy came by saying that the energy was 0 before we filled any of these levels. And as you fill the negative energy sea, you're lowering energy all the time. And it's that contribution which makes up the infinite negative contribution coming from the Fermi fields. And the algebra is certainly exactly right. The energy that people write down for the negative energy of the Fermi fields-- what they get by anticommuting two operators in equation 37-- is exactly the expression you get for what it takes to fill the Dirac sea. Yes?

AUDIENCE: Are we pretty confident that the smallness of the vacuum energy can't come from cancellations between the bosonic and the Fermi fields?

PROFESSOR: OK, the question is, are we confident that the cancellation cannot come from the cancellations between the Fermi fields and the bosonic fields. No, we're all confident that it cannot come from that. it very likely does come from that. But we are confident that we have no idea why that happens. And therefore, it's a big mystery.

Certainly our ignorance allows for any answer. Because we have a positive infinite contribution, we're just going to cut off and make it large. And we're going to have a negative contribution, which we're going to cut off and make it large in magnitude but negative. And then we're going to add them, and we have no idea what we're

going to get. But the fact that we get something that gets incredibly close to 0, and not something that's at all the same magnitude as the pieces you're adding together-- the positive piece or the negative piece-- means there's something going on that we don't understand. There's a cancellation that's happening that we cannot explain.

Now, I should maybe add that there's one context where we would expect a cancellation. And that is, there are theories that are what are called supersymmetric, which have a perfect symmetry between bosons and fermions, which would relate the positive energy from the photon to the infinite negative energy you would get from particles called photinos, which would be the supersymmetric partner of the photons-- a spin 1/2 particle that's a mirror image of a photon but has a fermionic character. So in an exactly supersymmetric theory, you would get an exact cancellation between the positive and the negative contributions. And the answer has to be 0 in an exactly supersymmetric theory.

However, the world is clearly not exactly supersymmetric. This photino has never been seen. And there'd be a particle called the selectron, which would be the scalar partner of the electron, which also has not been seen. And every known particle would have a partner, which has not been seen. There are no supersymmetric pairs which are known.

So supersymmetry is still a possibility as a broken symmetry of nature. And a lot of people think-- for pretty good reasons, I thin-- that it's very likely that the world does have an underlying supersymmetry. But as long as the supersymmetry is broken, it no longer guarantees this cancellation. And you could estimate what the mismatch is.

And it does make things a little bit better here. If we just take this Planck scale cutoff, we miss an energy density by a factor of about 10 to the 120. If we apply supersymmetry and make an estimate of what the supersymmetry breaking scale is and what effect that has on the mismatch of these calculations, then it gets reduced. Instead of being 120 order of magnitude problems, it's got a 50 order of magnitude

problem, which is a lot better, but not good enough.

Now, I do want to mention a third contribution here for completeness. The third one is likely be finite, so it's not as problematic as the other two.

AUDIENCE: [INAUDIBLE]

PROFESSOR: Same thing. Planck scale.

AUDIENCE: Oh, OK.

PROFESSOR: The third contribution is that some fields are believed to have nonzero values in the vacuum. And the famous example of that is the Higgs field, for which the particle associated with the Higgs was discovered a year ago at CERN, after over 50 years of looking for it. And the Higgs field is maybe the only field that's part of the standard model that has a nonzero expectation value, a nonzero value in the vacuum. But in more sophisticated theories like grand unified theories, there are many more fields that have nonzero values in the vacuum. So that's a likely extension of our standard model of particle physics.

So the bottom line is that it's easy for particle physicists to understand why the vacuum energy should be nonzero, but damned hard to have any idea of why it has the value that it has. We'll talk maybe at the end of the course about the possibility that the value of the vacuum energy density is, quote, "anthropically selected." That is one possible explanation, which maybe shows how desperate physicists are to look for an explanation here.

One possible explanation begins with ideas from string theory, where string theory tells us that there isn't just one kind of vacuum, but in fact, a huge number of different types of vacuum, perhaps 10 to the 500. And that would mean that if there were sort of random values for these infinite numbers that get cut off, that get cut off with different values-- and there are other ways of looking at the vacuum energy in string theories-- you'd expect coming out of string theory that the typical vacuum energy would be about the same as what you get when you cut off the quantum fluctuations of the electromagnetic field at the Planck scale. That is, the typical

vacuum energy coming out of a string theory would be at the Planck scale, which is this huge number compared to what we observe.

But string theory would be predict that there would be a spread of numbers going essentially from plus the Planck scale to minus the Planck scale, with everything in between. There'd be a tiny fraction of those vacua that would have a very small vacuum energy like what we observe. That's what you'd expect from string theory-- a large number, but a tiny fraction of vacua that would be in that integral.

And then the only problem would be to explain why we might likely be living in such an unusually small fraction of the set of all possible vacuums. And the answer to that that's discussed is that it may be anthropically selected. That is, life may only form when the vacuum energy is incredibly small.

And that is not built entirely from whole cloth. There is some physics behind that. We know that this vacuum energy affects the Friedman equation, which means it affects the expansion rate of the universe. So if we had a Planck scale vacuum energy, that would cause the universe to essentially blow apart at the time scale of the Planck scale, which is about 10 to the minus 40 something seconds, due to the huge repulsion that would be created by that positive vacuum energy.

And conversely, if there was a huge negative vacuum energy on the order of the Planck scale, the universe would just implode on a time scale of order of the Planck scale-- 10 to the minus 40 something seconds. So assuming that life takes billions of years to evolve and assuming nothing else about life, one can conclude that life can only exist in the very narrow band of possible vacuum energy densities which are incredibly small, like the one that we're living in. So it could be that we're here only because there isn't any life anyplace else. So all living things see a very, very small value of this vacuum energy density, even though if you plunk yourself down at a random place in this multiverse, you'd be likely to see a vacuum energy that's near the Planck scale.

OK, I'm done talking about this for now. Any further questions about it before we leave the topic?

I had suggested that we go on to talk about problems with the conventional big bang model, but, there is actually something else I wanted to do. I don't know how long it will take exactly, but I have a little historical interlude to talk about here.

We've been talking about the Friedman equations and how they're modified by the cosmological constant, which of course is an item that was very dear to Einstein's heart. So I'd like to tell you a little history story about Albert Einstein and Alexander Friedman, which I think is very interesting. The punchline of the story is that Einstein made pretty much of a fool out of himself on this.

And the reason why I like the story is maybe twofold. One is, I find it very comforting to know that even perhaps the greatest physicist of all time can make dumb mistakes just like the rest of us make dumb mistakes. I think that's a very comforting thing to keep in mind. And the other moral of the story is, I think, the importance of trying to be open-minded about issues in physics. Einstein was very much convinced that the universe was static, and so convinced that, in fact, he really made stupid mistakes trying to defend his static universe. So this will be a story of such a mistake.

So those are the two people. Friedman was a Russian natural-- he was really a meteorologist. They didn't really have that many theoretical physicists back in those days. But as a meteorologist, he was an expert in solving partial differential equations, and got himself interested in general relativity, which was a new theory at this point. And in 1922, he published an actual physics paper, I think the first physics paper he ever published, and one of two. He wrote basically two papers about the Friedman equations-- one for closed universes, and one for open universes.

So the first of those papers was published in June 29, 1922 in the premier physics journal of the day-- the *Zeitschrift fur Physik*, a German journal. And almost immediately-- or a few months later, when Einstein noticed this article-- Einstein submitted a comment about the article claiming that the article was entirely wrong, just mathematically wrong. And the article was titled "Remark on the Work of A. Friedmann 'On the Curvature of Space' " by A. Einstein, Berlin, received September

18, 1922. Looking at these dates-- the original article was received in June 1922, and Einstein was responding by September 18, a few months later.

And this is a translation, which comes from a book called *Cosmological Constants*, which is basically a book of famous articles in cosmology, like Friedman's, and all these original articles. It's a great book if you can still get a copy of it. It's no doubt out of print. It was written by Jeremy Bernstein and Gary Feinberg. And I'm taking the translation from there, because this was written in German. I don't know German.

"The works cited contains a result concerning a non-stationary world which seems suspect to me. Indeed, those solutions do not appear compatible with the field equations." And I guess A is the label of the field equations as they appeared in Friedman's paper.

"From the field equation, it follows necessarily that the divergence of the matter tensor Tik vanishes." That is, energy momentum is conserved as a four-vector quantity. "This along with ansatzes C and D"-- equations from the paper-- leads, according to Einstein, to an equation which we can all recognize the meaning of-- the partial of rho with respect to x sub 4-- time-- is 0. Einstein convinced himself that the equations of general relativity led to the conclusion that rho cannot change with time.

And he then goes on to say "which together with 8 implies that the world radius R"-- that's the scale factor. That's what we call a of t-- "is constant in time. The significance of the work, therefore, is to demonstrate this constancy." All Friedman does once you correct his equations, according to Einstein, was prove that the only cosmological solution is rho equals a constant, which was Einstein's static solution. This was entirely wrong-- no basis whatever in mathematics. But it took a while before Einstein got himself straightened out. And he did actually publish this.

The sequence of events was, June 29, Friedman submits his paper. September 18, Einstein submits his rebuttal to the paper. Friedman didn't learn about this until the following December. Friedman had a friend who played a key role in this story-- Yrui

Krutkov, who was visiting in Berlin during this time. And Friedman actually learned from Krutkov that Einstein had submitted a rebuttal.

So Friedman apparently was able to track it down and read it. And he wrote a detailed letter to Einstein explaining to Einstein what he got wrong, which is a gutsy thing to do, but Friedman was right in this case. But Einstein was traveling and actually never read the letter, at least not until much later.

Then the following May, Krutkov and Einstein are both at a conference in Leiden, a conference that they were both attending, which was a farewell lecture by Lorentz, who was retiring at that time. So they met and started talking, and continued talking. And we know most about it from a series of letters that Krutkov wrote to his sister back in Saint Petersburg.

And according to those letters-- and I'm now quoting from a rather lovely book called *Alexander A. Friedmann-- The Man who Made the Universe Expand*, by Tropp, Frenkel, and Chernin. Krutkov wrote to his sister that on Monday, May 7, 1923, "I was reading, together with Einstein, Friedman's article in the *Zeitschrift fur Physik.* And then on May 18, he wrote, "I defeated Einstein in the argument about Friedmann. Petrograd's honor is saved!" Petrograd is what we now call Saint Petersburg, and where they were all from-- that is, Friedman and Krutkov.

And then shortly after that, on May 31, Einstein submitted a retraction of his refutation of Friedman's paper. And the retraction is-- again, I'm quoting from *Cosmological Constant*, which translates all these nice papers into English. Einstein wrote, very briefly, "I have in an earlier note criticized the cited work-- Friedmann 1922. My objection rested however, as Mr. Krutkov off in person and a letter from Mr. Friedmann convinced me, on a calculational error.

I am convinced that Mr. Friedmann's results are both correct and clarifying. They show that in addition to the static solution to the field equations, there are time varying solutions with a spatially symmetric structure." Anyway, the expanding universe that we now talk about. Einstein did have to admit, ultimately, that algebra is algebra, and you can't really futz with algebra. And the Einstein equations do not

imply that rho cannot change with time, and that Friedman was right.

There's an interesting twist on this retraction letter. This is just a photo of Einstein at this time period, and Krutkov. There's an interesting twist on the retraction letter, which is that the original draft still exists. I forget what museum it's in. But it's quoted in another marvelous book about this history called *The Invented Universe*, by Pierre Kerzberg. And I Xeroxed this from the book. And this is the original draft. And notice there are some cross-outs. And the last cross-out, which followed this explanation that there is this expanding solution-- in Einstein's original draft, he wrote but then crossed out "a physical significance can hardly be ascribed to them."

So his initial instinct, even after having been convinced that these were a valid solution to the equations, was to say that they couldn't possibly be physical, because they're not physical. The universe is static. But somehow, before he submitted it, he did realize that there wasn't actually any solid logic behind that reasoning. So logic did prevail, and he decided that he really had no right to say that the solution has no physical significance, which is a good thing, because now, of course, it is the solution that we consider physically significant-- the expanding solution of Friedman. So [INAUDIBLE] is a mystery, I think.

OK, we have just a couple minutes left in the class. So I think that is nearly enough time for me to at least introduce what I want to talk about next. What we'll be talking about next time-- and I'll just introduce it now-- are a set of two problems associated with the conventional big bang theory. And by the conventional bang big bang theory, I mean basically the theory we've been talking about, but in particular, the big bang theory without inflation, which we will be talking about later. But so far, we've been talking about the big bang theory without inflation.

And the two problems that we'll talk about are called the horizon or horizon homogeneity problem, and the flatness problem. Both of these are problems connected with the initial conditions necessary to make the model work. So this horizon homogeneity problem is a problem about trying to understand the uniformity of the observed universe, which we've just put in as part of our initial conditions. The

model that we've constructed was just completely homogeneous and isotropic from start to present.

The evidence for the uniformity of the universe shows up most strongly, as I think we said before, in the cosmic background radiation, which can be measured to fantastic precision. And this radiation is known to be uniform in all directions to an accuracy of one part in 100,000, which is really a phenomenal level of accuracy.

Now, what makes this hard to understand in the conventional big bang theory is that if instead of just putting it in as an assumption about the initial conditions, you try to get it out of any kind of dynamics, that turns out to be impossible. And in particular, a calculation that we'll do next time is we'll imagine tracing back photons from the cosmic background radiation arriving at the Earth today from two opposite directions in the sky. Now, the phenomenology is that those photons come with exactly the same temperature to an accuracy of one part in 100,000, and that's what we're trying to explain.

Now, we all do know that systems do come to a uniform temperature. If you heated the air in this room in a corner and then let the room stand, the heat would scatter throughout the room, and the room would come to a uniform temperature. If you take a hot slice of pizza out of the oven, it gets cool, as everybody knows.

So there is this so-called zeroth law of thermodynamics which says that everything tends to come to a uniform temperature. And it's a fair question to ask, can we perhaps explain the uniformity of the universe by invoking this zeroth law of thermodynamcs? Maybe the universe just had time to come to a uniform temperature.

But one can see immediately when one looks at details that that's not the case. Within the context of our conventional model of cosmology, the universe definitely did not have time to come to a uniform temperature. And the easiest way to drive that home will be a calculation that we will do first thing next time, which is that we will trace back photons coming from opposite directions in the sky and ask, what would it take for them to have been set equal to the same temperature when they

were first emitted?

And what we'll find is that when we trace them back to their emission sources, that those emissions took place at two points which were separated from each other by about 50 horizon distances. So assuming that physical influences are limited by the speed of light-- and according to everything that we know about the laws of physics, that's true-- there is no way that the emission of that photon coming from that direction could have had any causal connection with the emission of the photon coming from the other direction. So if the uniformity had to be set up by physical processes that happened after the initial singularity, there's just no way that that emission could have known anything about what was going on over there, and no way they could have arranged to be emitting photons at the same energy at the same time.

Now, everything does work if you're willing to just assume that everything started at uniform. But if you're not willing to assume that, and want to try to derive the uniformity of the universe as a dynamical consequence of processes in the early universe, there's just no way to do it in the conventional big bang theory because of this causality argument. And later, we'll see that inflation gets around that. Yes?

AUDIENCE: How do we know that the homogeneity wasn't just created when the universe was smaller, in such a way that the speed of light limit wouldn't be violated, and that it would just maintain [INAUDIBLE]?

PROFESSOR: OK, the question is, how do we know that the uniformity wasn't established when the universe was very small, and then the speed of light might not have to be violated? Well, the point is that if the dynamics is the conventional big bang model, what we'll show is that there's not really enough. No matter how early you imagine it happening, it still is 50 horizon distances apart. And there's no way that those points could've communicated, no matter how close you come to t equals 0.

Now, you are of course free to assuming anything you want about the singularity at t equals 0. So if you want to just assume that somehow the singularity homogenized everything, that's OK. But there's no theory behind it. That's just speculation. But it

is satisfactory speculation. There's nothing it contradicts. But the beauty of inflation is that it does, in fact, provide a dynamical explanation for how this uniformity could have been created, which, at least to many people, is better than just speculating that somehow it happened in the singularity.

OK, I think that's it for now. I will tell you about the other problem we'll talk about next time next time. And I will see you all next Tuesday.