PROFESSOR: OK. Good morning, everybody. Welcome to lecture 12 of A286. I can't think of any announcements for today, but let me begin by asking if there are any questions either about logistics or about physics.

OK. In that case, let's get started. I want to begin by having a rapid run through of the things we talked about last time just to firm everything up in our minds and get us ready to go on.

So, last time we were talking about non-euclidean geometry in a serious way. We began by considering the surface of a sphere, just a two-dimensional sphere embedded in a three-dimensional space, described by the simple equation x squared plus y squared plus z squared equals R squared.

We said that if we wanted to talk about the surface itself, we'd want to have coordinates for the surface and not just speak of things in terms of x, y, and z. So we introduced the standard polar coordinates-- theta and phi, which are related to x, y, and z by these fairly well-known equations.

Then we wanted to know the metric in terms of our new variables theta and phi, which is the main goal-- to figure out the metric. So we first considered varying the two variables one at a time. By varying theta, we see that the point described by theta phi would sweep out a circle whose radius is R, and the angle subtended is d theta. So the arc length is just R times d theta. So for varying theta, the arc length is given by that simple equation.

Similarly, we went on to ask ourselves what happens when we vary phi. As we vary phi, the point described by theta comma phi again sweeps out a circle, but this time it's a circle in the horizontal plane whose radius is not R but whose radius has this projection factor that's R times sine theta. So the angle is again-- excuse me, the arc length is again d phi, the angles, times the radius, but the radius is R sine theta. So ds, the total arc length, is R times sine theta d phi.

Then, to put them together, we notice that these two variations are orthogonal to each other, which you could see pretty directly from the diagram. So if we do both of them at the same time and ask what's the total length of the displacement, it's just a simple application of the Pythagorean theorem. And we get the sum of the squares.

So, varying theta gives us Rd theta. Varying phi gives us R sine theta d phi. And putting them together, we just get ds squared is the sum of the squares of those, which is R squared times d theta squared plus sine squared theta d phi squared. And that's the standard metric for the service of a sphere in polar coordinates.

So that was a warm up. What we really want to do is to elevate that problem one more dimension, and then we have a model for the universe. We can use the same method to construct a three-dimensional space, which is a three-dimensional surface of a sphere embedded in four euclidean dimensions, and that becomes a perfectly viable homogeneous, isotropic, non-euclidean metric that can describe a universe and, in particular, describes the type of universe called a closed universe.

So to do that, we introduce one more axis, w. And we consider the sphere described by x squared plus y squared plus z squared plus w squared equals R squared. So it's a three-dimensional surface of a sphere in four dimensions. We then need to introduce one more variable to describe points on the surface, and we introduce this in the form of a new angle. The new angle I chose to call psi.

And we measure that angle from the new axis, from the w-axis. So the new angle psi is simply the angle from the w-axis, which means that the projection of our vector from the origin to the point in the w direction is just the R cosine psi and the projection into the x, y, z subspace is R times sine psi. And the four equations that describe x, y, z, and w are shown there.

And all we did is we set the w-coordinate equal to R times cosine psi, which is just the statement that psi measures the angle from the w-axis. Nothing more.

And then we multiplied x, y, and z by a factor of sine psi, so that now we still have maintained the condition that x squared plus y squared plus z squared plus w squared is equal to R squared, which you can prove directly by manipulating this using the famous identity sine squared plus cosine squared equals 1. Nothing more profound than that.

So, we're now ready to go ahead and find the new metric, and this time it'll really be something nontrivial, something you didn't already know from high school. The new displacement is to vary psi. If we vary psi, it's really the same story as we've seen before except in a different plane. ds is just equal to R times d psi. I guess, as we vary psi, the point described by these coordinates makes a full circle of radius R.

OK, now what we want to do is put all this together. If we vary psi, we know that ds is equal to R times deep psi. If we just vary theta or phi, it's the same thing we had before. We don't need to rethink it. All we need to do is remember there's an extra factor of sine psi in front of all [INAUDIBLE] in the XYZ subspace.

So if you vary theta or phi, ds squared is just equal to what we had before for the metric multiplied by the extra factor of sine squared of psi. Then to put them together, if we assume for the moment that they are orthogonal to each other, then we just add the sum of the squares. And that is the right answer. But I'll justify it in a minute.

But jumping ahead and making the assumption that these separate displacements are always orthogonal to each other, ds squared is then just the sum of the squares, and we get this matrix to describe our closed universe in terms of the variables psi, theta, and phi.

To prove this orthogonality, which is crucial for believing that result, I gave an argument last time, and I'll outline again on the slides here. We can consider the two displacement vectors that we're trying to show to be orthogonal. dR sub psi is a four-dimensional vector, which represents the displacement of the point being described by these coordinates when psi is changed to psi plus deep psi, infinitesimal change in the psi coordinate.

Similarly, I'm going to let dR sub theta be the displacement vector that the point described by these coordinates undergoes when theta is varied by an infinitesimal amount, d theta. And what we're trying to show is that these two are orthogonal to each other. So if we do them both, the magnitude of the change is just the sum of squares, the square root of the sum of squares.

So, first looking at dR sub theta, we notice that dR sub theta has no w-component. And to make that clear, we should go back a couple slides and look at how w is defined. w is defined as R times cosine psi. So if we vary theta, w doesn't change. It doesn't depend on theta.

So if dR sub theta has no w-component, it means that when we take the dot product of dR psi with the dR theta, we want to show that this is 0 to show that they're orthogonal. The w-components won't enter, because one of the two w-components is 0, and the dot product is a sum of the product of the x-components plus the product of the y-components plus the product of the z-components plus the product of the w-components. So w-compontents only enter as a product of the two w-components. So as long as one of them is 0, there's no contribution there.

So the four-dimensional dot product reduces to a three-dimensional dot product. And here I'm introducing a peculiar notation when I put a subscript-- a superscript, rather-- 3 in a vector. I just mean take the first three components and ignore the fourth and think of it as a 3 vector.

So the dot product that we're trying to calculate now is just the dot product of two 3 vectors-- the one that we get when we vary psi and the one that we get when we vary theta.

Next thing to notice is that we can look at the properties of these two vectors. And dR psi, the vector we get when we vary psi, I claim is in the radial direction in this three-dimensional subspace. And we can see that by looking again at these formulas that relate the angles to the Cartesian coordinates. When we vary psi, sine psi changes, but sine psi multiplies x, y, and z all by the same amount.

So sine psi changes. It changes x, y, and z proportionally. And if you change x, y, and z proportionally, it means you're moving in the radial direction in this three-dimensional subspace.

On the other hand, dR theta is what we get when we vary theta. And from the beginning, theta was defined in a way that parametrized the sphere. So varying theta only moves you along the sphere. It does not change your distance from the origin.

So varying theta is purely tangential. So we have a dot product between a radial vector and a tangential vector, and those are always orthogonal to each other. So we get a dot product of 0, as claimed. So the two original four vectors are orthogonal to each other, which is what we're trying to prove.

OK, everybody happy with that? It is a crucial step. You haven't really gotten the results unless you know these vectors are orthogonal.

OK, almost done now. We then later in lecture talked about the implications of general relativity, and here we didn't prove what we were claiming. We just admitted that there are some things in general relativity that we're just going to have to assume, and this really is almost the only one.

General relativity tells us how matter causes space to curve. And it does that in the form of what are called the Einstein field equations. And we're not going to learn the Einstein field equations. That's the subject of a general relativity course. So we're just going to have to assume what general relativity tells us about how space curves, and in particular in this instance, what it tells us is that the radius of curvature R-- this R that we've introduced into our metric-- is the radius of curvature of the space, is related to the matter and motion by R squared being equal to a squared of t divided by k.

And we did argue last time that, that k in the denominator really is necessary just to make the units turn out right. So we really know by dimensional analysis that this formula has to hold up to some factor. The fact that the factor is 1 is a fact about general relativity, which we're not showing at this point.

When one puts this back into the metric to express the metric in terms of a of t, we find finally that the metric can be written as shown in the box here, and this is the last equation, where I've made a substitution of variables.

I replaced the angle psi by a radial coordinate little r, which is defined to be sine of psi divided by the square root of k. And this form of the metric is what's called the Robertson-Walker metric. And it's a famous form of the metric. This is what people normally use.

So that finishes everything we said last time, I think. Any questions? Yes.

AUDIENCE: What is the motivation for saying that the-- where you can describe space as a three-dimensional sphere in a four space. Is it because it's only real geometry that, where there's isotropian [INAUDIBLE]?

PROFESSOR: Yes, that's right. I was going to be saying that shortly, but yes. This metric and its open universe counterpart and flat space together make up the most general possible metric, which is homogeneous and isotropic.

At this stage, I'm really not going to claim that anyway, but at this stage, what we do know is that this metric is homogeneous and isotropic. And certainly what we're trying to construct is metrics, which are homogeneous and isotropic. But this also is actually the only possibility within that small class.

Any other questions?

OK. In that case, we will continue on the blackboard. So what we have derived so far is the metric for a closed universe. Maybe I'll start by getting on the blackboard the same formula that's up there, just so I can see it better. Even though you can probably see equally well either way.

A closed universe is described by ds squared is equal to a squared of t times dr squared over 1 minus kr squared plus r squared d theta squared plus sine squared theta d phi squared. And to relate this variable r to our previous definitions, little r is equal to the sine of psi divided by the square root of k.

OK. Question back there?

AUDIENCE: Psi is-- sorry this is just [INAUDIBLE] think. Psi is which angle?

PROFESSOR: OK, the question is psi is which angle. Psi is the angle we introduced when we went from two-dimension sphere embedded in three dimensions to one dimension higher. Psi is the angle from the new axis. The angle from the w axis.

AUDIENCE: OK.

PROFESSOR: OK. So, we've covered a lot of ground here. We have our first non-euclidean metric that's visibly important. But we know from our work on the Newtonian model of the universe that this little k doesn't have to be positive. It can be positive, negative, or 0.

If k is 0, the metric is actually just the metric of a flat space. But when k is negative, it's a case that we haven't talked about yet. So we want to know, what will we write for a metric if k were negative? And the answer to that turns out to be perfectly simple.

This formula has a k in it. Lots of times in our experience-- I'm sure we all know-- when we write an equation for one sine of a variable, we find that the same equation works even if the variable has the other sine. So if you're buying and selling stocks, if the stocks go up and the stocks go down, you can use the same equations. The price today is the price yesterday plus the increment, and the increment could be positive or negative. But that equation-- price today equals price yesterday plus income-- it still works. And same thing here.

If k it happens to be negative, there's nothing wrong with this formula. It, in fact, describes an open universe just as well as it describes a closed universe.

Now notice, however, that things are a little bit tricky. If you look at the equation that I wrote immediately below here, if k is negative, we have the square root of a negative number here, so the denominator would be imaginary, and what would that say about r and psi? It would obviously confuse us.

So you really have to write the metric and the correct form before you can just change the sign of k. If we had written the metric in terms of psi and not made the substitution, we could just as well have written the metric for a closed universe as a squared of t divided by k times d psi squared plus sine squared psi d theta squared plus sine squared theta d phi squared.

This is an alternative metric for the closed universe. It's, in fact, where we started. We then made a substitution, replacing sine psi by little r. If we had the metric in this form and we said, well, let k be negative instead of positive, then notice that a squared is certainly positive, so we'd have a negative number out here times things which are also manifestly positive. We would have a negative definite metric instead of a positive definite metric.

So we could not change the sign of little k in this formula and get what we want. So you have to careful. It doesn't always work. But it does work when you write the metric in this form.

Now since it doesn't always work, and since we haven't really made any sound arguments yet, I'd like to spend a little time describing-- I'm not going to do the calculation because it's too messy, but I'd like spend little time describing how you would show that this metric works for an open universe.

So first thing to recognize is-- what do we actually mean when we say it "works"? Can somebody tell me what I probably mean when I say that? Yes.

AUDIENCE: It doesn't have any glaring contradictions?

PROFESSOR: Doesn't have any glaring contradictions. Yeah, that's good, but we can be more specific, especially since I'm going to try to describe how we would actually show it, and it's a little hard to show that something doesn't have glaring contradictions.

What do we actually care about in constructing these metrics? Yes?

AUDIENCE: Does the goal that they hold well in limits, i.e., the flat universes?

PROFESSOR: That the physics will hold well in certain limits, like they should approach the flat universe and limit. We certainly do want that to happen, but there is something else that we want that doesn't involve taking limits, because you have different things which all approach the same limit, of course.

Making sure an answer approaches the right limits is a good way to test the answer, because most wrong answers will not have the right limits. But merely knowing you have the right limits does not prove that you have the right answer. Yes?

AUDIENCE: It could also reflect an isotropic and homogeneous non-euclidean?

PROFESSOR: Exactly. Exactly. What we're looking for is a homogeneous and isotropic non-euclidean space, because that's what we know about our universe. It's homogeneous and isotropic, and we're trying to build a mathematical model of those facts.

So we want homogeneity and isotropy. If we limit isotropy to isotropy about the origin, which is enough if we're going to later prove homogeneity, which will prove that all points are equivalent, isotropy about the origin is obvious here, because the angular part is just exactly what we had for a sphere in three euclidean dimensions. So it behaves on angles exactly like a euclidean problem, so we know that it's isotropic.

If you point out that algebraically as you look at it, it's not obvious that it's isotropic. We just know where that expression came from. It came from the sphere.

In terms of theta and phi, it's not manifestly isotropic. And that's because in choosing theta and phi, we chose a special point, the North Pole, to measure our angle theta from. And the choice of that special point for our coordinate system broke the isotropy. But we know that deep down it is isotropic.

And that idea, that you can have such isotropy without having manifest isotropy is also crucial to how homogeneity plays out in this metric. I claim, and we know really, that this metric is homogeneous. At least we know where it came from. It came, again, from the surface of a sphere one dimension higher than just the angular part. And then spherical picture--it's obviously homogeneous.

But nonetheless, in building our coordinate system, we had to break the homogeneity. We chose a special point-- again, what we might call the North Pole-- in this case, the point where w has its maximum value and made that point special. The point where w had its maximum value in the x, y, z, w space is the point which is now the origin of this coordinate system.

So, if we wanted to prove that this metric really is homogeneous, we would like to prove it for the k equals minus 1 case. But let's first, imaging, what would we do if we wanted to prove that it was homogeneous for the k equals plus 1 case or k positive case? The case that we really think we do understand. The closed universe.

For the closed universe, this does not look homogeneous. It looks like the origin is special. r equals 0 is special. But we know that it came from the sphere, and if somebody asked us to prove that, that metric was homogeneous-- more particularly, somebody might, for example, challenge us to construct a coordinate transformation, which would preserve the metric and map some arbitrary point r 0 theta 0 phi 0 to the origin. We might undertake that challenge.

It's a lot of work. We're not going to actually do it. And I promise I'll never ask you to do it. But I want to talk a little bit about how we would do it, because we do have a method, which we know will work. And knowing that we have a method that works is all we really need to know.

So suppose we wanted a coordinate transformation that preserves the form of the metric and maps some arbitrary point, and I'll give the coordinates to this arbitrary point a name. I'll call it r 0, theta 0, and phi 0. These are coordinates of a point.

So we're going to map this arbitrary point to the origin. Notice that this is a concrete statement about homogeneity. If we can map an arbitrary point to the origin while preserving the metric, we're really proving that an arbitrary point is equivalent to the origin. And if an arbitrary point is equivalent to the origin, then all points are equivalent to the origin and equivalent to each other. We're done. That proves homogeneity.

So, suppose we wanted to do this. How would we do it? The point is that knowing how we got this metric from the sphere allows us to go back to sphere and rotate the sphere and rotate back and derive the coordinate transformation that we want. And I'll just describe that in slightly more detail.

What we would first do is-- I claim we can do it in three steps, each of which we know how to do although they're messy. I guess I'll start over here.

So step one is to find the x, y, z, w-coordinates that go with this point. Because once we have the x, y, z, w-coordinates, we're in our four-dimensional space where we know how to do rotations.

So the first thing we do is we just find the corresponding x, y, z, w-coordinates where x0 is just the x-coordinate that goes with r0, theta0, and phi0. And y0 is the y-coordinate that goes with r0, theta0, and phi0. And z0 is the z-coordinate, and w0 is the w-coordinate.

And these are just points we had before. I'm writing them symbolically, but we know how to express the Cartesian coordinates in terms of the angles. Yes?

AUDIENCE: Do we have a psi0 as well?

PROFESSOR: r0 replaced psi0. r0 is the sine of psi divided by root k. So we only need three coordinates. We could have different choices of what we call them. We could've used psi here.

The reason I'm using little r is that I want to, when I'm done, describe what we would do if k were a negative. And if k were a negative, we already said that psi does not actually work. We have to use different coordinates in order to smoothly write an open universe coordinate system. Yes?

AUDIENCE: How is that not like cheating? Because, I mean, you did define r with the root k, and now we're just kind of ignoring them.

PROFESSOR: Well that's exactly-- the question is why is this not cheating? And the reason it's not cheating is because of what I'm about to show you. What I'm saying really is that just setting k negative is something you might expect to probably work. I think you have good grounds to expect it to probably work. Now what we're talking about is how to actually show that it works. Yes?

AUDIENCE: Professor, what's the necessity of defining w as opposed to, I don't know, t? The traditional, like, t--

PROFESSOR: OK. The question is why did I call the fourth variable w and not t. The answer is that the variable we're talking about here is not time. It's another spatial coordinate. So, for that reason I think it's better to call it w than to call it t.

Of course, needless to say, the name of a variable doesn't have any actual significance. So I certainly could have called it t equally well, but I think that would have caused some confusion by people thinking it was time, which it's not. Any other questions?

OK. So, what I'm outlining is the steps that you would use to prove that this is homogeneous. I'm doing it for the closed case. Now the point is that if we can do it for the open case, we'll prove that the open metric is what we want it to be-- homogeneous and isotropic. And that's the only criteria that we have for goodness of a metric.

So, continuing with the closed case in mind, the first step of doing this mapping, to map some arbitrary point to the origin, is to first find its x, y, z, w-coordinates, its Cartesian coordinates. Once we have the Cartesian coordinates, we know that in the four-dimensional space we can perform ordinary euclidean rotations.

Rotations in four dimensions [INAUDIBLE], not three dimensions. And even rotations in three dimensions are not all that simple, but nonetheless in principle, we know how to do rotations in four dimensions. And we know what we're trying to do in the four-dimensional picture. We're trying to rotate this point to the origin of our coordinate system.

And the origin of our coordinate system-- the way we've done this mapping-- is to make w equals capital R, w equal to its maximum value, the center of our coordinate system. That's where psi was equal to 0, and now where our new variable little r is equal to 0. So we'd like to map this point, whatever it is, to the point where w has a maximum value, and the other coordinates all vanish.

So we can do that. We can find a rotation that does that. It's not even unique, because you can always rotate about the final axis.

But in any case, we imagine that we can do that. And that's step two-- is to find the right rotation. And a general rotation is a linear transformation, so you can write it as x prime, y prime, z prime, w prime, as a four vector, is equal to some 4 by 4 rotation matrix times the original four coordinates.

So an equation of this form would describe a four-dimensional rotation. And we in particular want the four-dimensional rotation which maps-- maybe I'll write it as a similar matrix equation-- what we want is that the matrix, when it operates on x0, y0, z0, and w0, which remember are just the four coordinates that correspond to our original point, r0, theta0, phi0, we want this to map into 0, 0, 0, r, which is the four-dimensional description of the origin of our new coordinate system.

And then finally, step three is the obvious one. Once we've found the transformation that maps to the origin, now we just go back to our original angular coordinates. So, now we set r prime equal to the radius function of x prime, y prime, z prime, and w prime. And this just means the r-coordinate-- that corresponds to those four euclidean coordinates and similarly for the other variables.

Theta prime is the theta function of x prime, y prime, z prime, and w prime. And phi prime is equal to the phi function. All these are functions that we know. I just don't want to write them out explicitly, because that's a lot of work.

So it's phi of x prime, y prime, z prime, w prime. So now with the three steps, we have our mapping. We could start with an arbitrary point, perform the rotation, and then calculate the angular variables again.

Ok, do people understand what I'm talking about here? Oh good. OK. And the good things is I promise I won't make you do it. And I've never done it either, to be honest. But it's obvious that we can do it.

And if we can do this, this would prove and especially demonstrate homogeneity. It would be a mapping that would map an arbitrary point to the origin, proving that, that arbitrary point was equivalent to the origin as far as the metric is concerned.

And now my claim, I mean, I'm not really going to prove this either, but it's a claim that could be verified by going through things. And it also seems highly plausible-- that if you looked at each of these steps-- these are all just algebra steps, these are all just algebraic equations-- that they would work just as well for negative k as they will for positive k.

And by doing these same series of manipulations for negative k, you would prove that the Robertson-Walker metric for negative k is homogeneous, which is our goal. We already know it's isotropic. If we can prove it's homogeneous, we're home free. It's [INAUDIBLE]. Yes?

PROFESSOR: Does the necessity of placing the origin at r, like big R, on w is that just because that thing is expanding? Or-- I'm kind of having trouble understanding why it wouldn't be just straight 0s. Like, why there's that [INAUDIBLE] value [INAUDIBLE]?

PROFESSOR: OK, the question is why does the origin look like this as opposed to just being all 0s. The answer is that all 0s is not even in our space. Because, remember the space we're interested in is the surface of the sphere. And the surface of the sphere obeys x squared plus y squared plus z squared plus w squared equals r squared.

So if all the coordinates were 0, it's not part of our space at all. So we're using this four-dimensional space, the embedding space, to make things simple. But in the end, we're only interested in the three-dimensional surface. So the origin of our coordinate system for that three-dimensional surface had better be in the three-dimensional surface.

So of course, other choices we could have made-- we could have put it anywhere we want on the surface. Choosing to put it where w has its maximum value is just an arbitrary convention. Yes?

AUDIENCE: So you were saying that most metrics are not homogeneous?

PROFESSOR: Oh yeah. Sure. Most metrics are not homogeneous. Most objects are not around.

AUDIENCE: When we're doing this math to show that the Robertson-Walker was homogeneous, it didn't seem that we use the exact form of the Robertson-Walker metric at all in it. We just said, it is possible to do these--

PROFESSOR: Well, no. We did use the form when we made this rotation. We used the form that, in the euclidean formulation, we knew that it was rotationally invariant. And the rotational invariance in the euclidean formulation is homogeneity and guarantees homogeneity, but it's a special property.

If it was ellipsoidal shaped instead of spherical, when you rotated it, it would not be invariant. If it had any bumps or lumps when you rotated it, it would not be invariant. Yes, Aviv.

AUDIENCE: I feel like there should be a fourth step where you show that the metric doesn't change forms [INAUDIBLE].

PROFESSOR: Yeah. We do need to know the metric does not change form, but I think we do have a guaranteed-- maybe we should've said some words. I don't think it's really a fourth step in the sense that I don't think it requires anymore algebra. But the point is that we know that the metric is invariant, that the four-dimensional metric is invariant under this rotation, which was really the only non-trivial step. And otherwise, besides the rotation, all we did is we went from the r theta phi variables to the utility euclidean variables, and then we went back from the euclidean variables to the r theta phi variables.

But we already know how to go from euclidean variables to r theta phi variables, and it results in that metric. And it will still result in that metric when, if a prime is on all of the coordinates.

AUDIENCE: So you don't have to say, like, suppose we have a [INAUDIBLE] byproduct. And do the mapping, figure out what the ds squared is in terms of r theta phi to show that it's the same? Do we have to do that [INAUDIBLE]?

PROFESSOR: OK, the question is do we have to explicitly show that ds squared is the same for the new variables as it was for the old variables. We certainly want to be convinced that that's true, and we certainly want to have an argument which convinces us.

But I would claim that if you think about what underlies these steps, I only gave a schematic description of them. If you think about what underlies these steps, I think it's implicit that the form of the metric is what we had. The form of the metric that we had was completely dictated by the transformation, which expressed r theta and phi in terms of x, y, z, and w.

And as long as you know the metric in x, y, z, and w, and that's the euclidean metric both before and after our rotation, then when you use the same equations to go from x, y, z, w to r, theta, and phi, you'll always get the same metric. So I think we are guaranteed by this process to get a metric for our new variables, r prime, theta prime, and phi prime, which has exactly the same form as the original metric. Because it's the same calculation again. The only difference is that this time the variables all have primes on them.

And the crucial step, the step that was nontrivial, is the fact that this rotation did not change the metric. That's where the homogeneity was built in, that we started with a sphere that we knew was rotationally invariant. And this whole calculation just extracts that homogeneity that we built in from the beginning. Yes?

AUDIENCE: For k negative, it's r squared over k, so is r negative?

PROFESSOR: No. r is still positive when expressed in terms of x, y, z, and w. Let me think if I can show that. I'll show that next time. It's a little involved, but it will be positive.

I might add that if we look at this formula and ask what's going on, what's going on-- we don't necessarily need to know this-- but what's going on is that in going from the closed case to the open case, k goes from positive to negative and therefore square root of k becomes imaginary. But psi also becomes imaginary.

To describe the relationship between the closed metric and the open metric, if you're using psi, you have to say that for the closed metric, you'll use real values of psi, and for the open metric, you'll use imaginary values of psi. And that makes r real and makes this formula work.

And you could also then see how this formula works. If psi is assigned imaginary values, then deep psi squared is negative, so this negative sign cancels that negative sign, and you again get a positive definite metric. Yes?

AUDIENCE: So how can we choose the imaginary [INAUDIBLE] of sine and psi. Is that just to reflect or is that a choice that we're making for our model?

PROFESSOR: OK. Yeah. The question is why do we choose to use the imaginary value of psi. And the answer is perhaps that I failed to state all the conditions we're interested in when I said what properties we want this metric to have. We want the metric that we're seeking to be homogeneous and isotropic, as we said. What we didn't say, but what I kind of had in the back of my mind as an assumption, is that the metric should also be positive definite.

You can construct other metrics which are homogeneous and isotropic but not positive definite. In fact, you would if you let psi be real and let k be negative. You'd have a negative definite metric, which would still be homogeneous and isotropic.

So to enforce all three properties, you have to use some imagination. And the easiest way to do it is to write the metric in the magical form where you can just let k go to minus k, and that's the Robertson-Walker form here.

And if you write it this way, when you let k go to minus k, it becomes negative definite, and you have to scratch your head. And if you really clever, you might say, well, if I assign negative-- excuse me, if I assign imaginary values to psi, it'll become positive definite again. That works, but it's less straightforward. Yes?

AUDIENCE: Do we want it to be positive definite so that it's visible?

PROFESSOR: Exactly. We want it to be positive definite so it's visible.

AUDIENCE: Is there any way we can have a negative definite for a multidimensional space that's reflects a 3D positive definite space?

PROFESSOR: Is there anyway in higher dimensional space or something that we can have it maybe have mixed signs or be negative. Well in fact, we will see shortly, because we're going to add in time, time will occur with the opposite sign, and it will not be positive definite anymore. But that will nonetheless correspond to real physics.

AUDIENCE: Right. So why do we have to force that ds squared into spaces?

PROFESSOR: Because now we are talking only about space, and certainly for our universe, space is positive.

Now, I might add since you brought this up, that in relativity, there's no clear distinction between space and time, so you might wonder why should I be saying that space is positive and time is negative. And perhaps I'm oversimplifying a bit when I say that space is positive and time is negative.

But what is a requirement for general relativity to match our universe, and therefore a requirement that we impose on the general relativity theory, is that the metric have three positive eigenvalues and one negative eigenvalue. And that's how it's described, and that's called the signature of the metric-- the number of positive eigenvalues and the number of negative eigenvalues.

And reality clearly has-- well, I shouldn't say clearly. String theory is more dimensions-- but to describe our macroscopic world, clearly, we have quantities that we intuitively identify as three space dimensions in one time dimension. And the metric that describes that is a metric whose signature is three positive eigenvalues and one negative eigenvalue. Although, some people reverse the sign conventions and say it's three negative and 1 positive, which works just as well. You can define the metric with either sign. But this is the one that we're using, the one that corresponds to space being positive.

OK, any other questions?

OK. In that case, let's move onward. What did I want to talk about next? OK.

I wanted to write on the blackboard a statement which came about earlier due to questioning but hasn't been written on the blackboard yet, which is that there's a theorem which says that the most general possible three-dimensional metric, which is homogeneous and isotropic, is this space. So any three-dimensional, homogeneous, and isotropic space can be described by the Robertson-Walker metric.

We are not going to prove this, but it is a theorem. If you want to see a proof, there's a proof, for example, in Steve Weinberg's gravitation and cosmology textbook. And in a lot of other books, I'm sure. But we'll take it for granted.

These are certainly the only homogeneous and isotopic spaces that we know how to construct. And in fact, it's the only ones that exist.

Now, I emphasize that if the space is homogeneous and isotropic, obviously the metric does not have to look exactly like that, because you can choose different coordinates. We could take these coordinates and make some arbitrary transformation and make the metric look incredibly ugly. It would still be a homogeneous and isotropic space. But the claim is that any homogeneous and isotropic space can be written in metric that looks exactly like this with a proper choice of coordinates.

OK. Next thing I want to discuss is the size of these universes. Size in the generalized sense of, actually the question of whether it's positive-- whether it's infinite or finite. Notice that our closed universe, one can see from the embedding in four euclidean dimensions, is finite. The surface of a sphere is finite.

And one can also see it from the form of the metric, although one has to think a little bit about how exactly it works. If we start at the origin and let r get bigger, clearly something funny's going to happen when r is equal to 1 over the square root of k. That is, when kr squared is equal to 1, this metric will become singular.

If one goes back to the angular description in terms of psi, it's clearer what's going on there. When kr squared is 1, that's exactly where sine psi is one. And that just means you've reached the equator of your sphere.

Quick picture.

Size measured from the w-axis. The equator corresponds to sine psi equals 1. And then if you continue, psi gets bigger up to pi, but r starts getting smaller again. So r is double valued in the sense that there are two latitudes at which r has the same value-- one of the northern hemisphere and one in the southern hemisphere.

But in any case, r starts from 0, goes to some maximum value, and then goes back to 0. Everything is finite. And one could integrate and find the total volume, and it's finite. And I think it was or will be a homework problem where you do exactly that integral at the volume of a closed universe.

On the other hand, if k-- let's first set it equal to 0. If k is 0, we have here just the euclidean metric and polar coordinates describing flat space. And there's no limit on r. r can become as large as you want.

You can hypothesize that space somehow ends, but we don't believe-- we don't know of any end to space. And there are-- in any case, a precise way you could describe what it would mean for space to end in general relativity and the usual postulates of general relativity is that , that doesn't happen-- that space doesn't just have an arbitrary end.

So the flat case is an infinite space when k is equal to 0. It's just an infinite euclidean space. Similarly, if k is negative, then nothing funny happens as r increases. So there's no reason not to let r increase to infinity. Anything else would just be putting in an arbitrary wall into space without any motivation for believing that such a wall is there.

And one has to remember that r is not physical distance. So the fact that r can go to infinity doesn't necessarily make the space infinite, but you can calculate physical distance.

We can calculate the physical distance from the origin to the radius r, and we get that just by integrating the metric. The metric tells us what the actual physical length is for an infinitesimal segment. That's what the metric meant in the first place.

So the integral that we'd be doing of an a of t out front, and then we'll be integrating dr prime over the square root of 1 minus kr prime squared from 0 up to r. Remember k is negative. So this is a positive quantity under the square root. It's not going to cause any problems by vanishing on us.

And this is an integral, which is in fact, doable. And it's just equal to, still the a of t in front, which I think I left out in the notes, times an inverse hyperbolic cinch of the square root of minus kr over square root of minus k.

Remember k is negative. See these are all square roots of positive numbers. And that cinch function, the inverse cinch can get to be as large as one wants by letting r be as large as one wants. It grows without bound. And that means the physical distance grows without bound as r grows to infinity. And it grows faster than linear, I think.

I take that back. I'm not sure. Yes?

AUDIENCE: I guess I'm still confused because r is sine of psi over square root of k, so sine psi is bounded by 1 and negative 1, so if we're letting r go to infinity, how--

PROFESSOR: That formula only works for the closed case. r equals sine psi over root k. We can apply it to the open case if we let psi become imaginary. But then the bounds that you said no longer apply.

The sine of an imaginary variable is, in fact, the cinch of a real variable. OK?

Next thing I want to point out is that the Gauss-Bolyai-Lobachevski geometry that you did a homework problem about or it was an extra credit problem, so some of you did not. But we talked about the Gauss-Bolyai-Lobachevski geometry. That really is just an open Robertson-Walker, RW Robertson-Walker metric, but in two space dimensions. But it's completely analogous to the Robertson-Walker metric in three space dimensions that we're talking about here.

So you might recall that Felix Klein construction looked very complicated. That's because of the coordinates that he used. Those coordinates might be simple from some point of view, but from the point of view of illustrating homogeneity, they're very complicated coordinates. And anyway, to physicists, the Robertson-Walker open coordinate system is familiar, and the Felix Klein coordinate system is not.

OK. If there are no further questions about these spatial metrics, the next thing I want to talk about is adding time to the picture. Because in the end, we're going to be interested in a spacetime metric, because that's what general relativity is all about-- spacetime metrics.

OK. Everything is going to hinge on an important fact from special relativity, which we are going to assume but not prove, because most you have had courses about special relativity elsewhere. And for those you who have not, you can either, if you wish, read an appendix to lecture notes five, in which the fact that I'm about to show you is derived, or you could just assume it. Whichever you prefer, depending on how much time you have. This is not a course about special relativity. You're not required to learn how to derive the fact that I'm about to write.

And what I'm about to write starts with a definition given any two events. An event is a point in spacetime. Snapping my finger-- this clearly speaking event. It happens at a certain place in a certain time.

And every real events occupies some small volume of spacetime. An ideal event is a point in spacetime. And we'll be talking about ideal events. Which, our model, as I said, has points in spacetime.

So given any two events, one can talk about a separation between those events. And they will be separated in both space and time, although either one of those could be 0. But they're not both 0, or it's the same event.

And it's possible to define an interesting quantity, which is the difference between the x-coordinates of the two events. This will be the separation between two events, which I'll call a and b, and xa and xb are the x-coordinates of those two events.

And probably you all have enough imagination to guess that ya and yb are the y-coordinates of those two events. And za and zb are the z-coordinates of those events. And now here's a surprising one, if you haven't already seen it. We're going to have minus c squared times ta minus tb squared.

Now, this is all in special relativity. I maybe should clarify. We haven't gotten to general relativity or cosmology yet. But we need to understand something from special relativity first.

So in special relativity, it's natural to define that interval between two events. And the magical property, which is why we define this integral in the first place, is that if we had two different inertial observers, we could calculate how the coordinates as seen by one observer are related to the coordinates as seen by the other observer. And that's called the Lorentz transformation.

And one finds that this particular quantity will have exactly the same value to both observers always. The two observers will, in general, find different values for every one of the four quantities here. But when the four quantities are added up with a minus sign in front of the time term, the calculations would show that you get the same value for inertial observers.

So this quantity is called Lorentz invariant, meaning it's invariant under Lorentz transformations. And it makes it a very important thing to talk about because in the end, physical things have to be essentially Lorentz invariant because the laws of physics are Lorentz invariant. The laws of physics are the same in all Lorentz frames, so ultimately they have to involve quantities, which have some simple relationship from one Lorentz frame to another.

OK now, we'd still like to have a clearer notion, I think, of what this quantity means. It's defined by that equation and principle, but it would be nice if we had some understanding of what it means. And I think the easiest way to describe what it means is to look at special frames, even though the important feature of this quantity is that it has the same numerical value in all frames.

So the numerical value is the same in all frames, but some frames make it easier to interpret it. That's what I'm claiming. So, what that frame is depends on the value of s squared, or at least the sine of it. For sab squared greater than 0, which means it's dominated by the spatial terms there, because those are the positive ones. And therefore, the separation is called spacelike.

Some books put a hyphen between space and like and some don't. I don't.

And there's a theorem that says that if the separation between two events is spacelike, there always exists a Lorentz frame, an inertial frame, in which the two events happen simultaneously.

Backwards E is a There Exists symbol. Then there exists an inertial frame in which a and b are simultaneous. In that frame, we could look at what that formula tells us sab squared is.

Since they're simultaneous, ta equals tb, and therefore the last term does not contribute. So in that frame, s squared is just xa minus xb squared plus ya minus yb squared plus za minus zb squared. And we know what that is. That's just the euclidean length, the euclidean distance between the two points.

So, in this frame, sab squared is just equal to the distance between events squared. Or you take the square root because they're positive numbers. You could say sab is equal to the distance between the two events. So when sab squared is positive, sab is just the distance between the two events in the frame in which they're simultaneous.

If sab squared is not positive, it could be negative or 0. Let me go to the negative case first. For sab squared less than 0, for it to be less than 0, it means that this expression is dominated by the time term because that's the negative term. And therefore the separation is called timelike.

And again, there's a theorem. The theorem says that if the separation between two events is timelike, there exists a frame in which it happened in the same location. If they're at the same location, we could again look at that equation that defines sab squared and ask what form does it take in this special frame where the two events are the same location.

That means the first three terms are all 0 because-- same location. It means in that frame, sab squared is negative, and it's just minus c squared times the time separation squared.

So in that frame, sab squared is equal to minus c squared tau ab squared, where tau ab is just equal to the time separation. So when sab squared is negative, its meaning is as minus c squared times the square of the time separation in the frame where the two events happened at the same place.

Now, this notion of the two events happening at the same place has a particularly simple intuition if the two events that we're talking about happen on the same object, like two flashes of the same strobe if that strobe is moving at a constant velocity. Otherwise, all bets are off.

But if that strobe is moving at a constant velocity so that the frame of the strobe is an inertial frame, then the frame of the strobe is, in fact, the frame in which the two events happened at the same place. They both happened at the bulb of the strobe light, which in the frame of the strobe is just one point.

So this time interval, which the sab squared measures, is simply the time interval as measured by the object itself, is measured by an observer following the strobe. And if we place the strobe by a person with a wristwatch, this notion of time, which is called proper time, is simply the time measured by the person's wristwatch. A clock that follows the object so that anything that happened to that object happens at the same location.

So if events happen to the same object, ta ab is just the time interval measured by that object. And as you give these things names for the spacelike case, sab is often called the proper distance between the events, and ta ab is the proper time interval between the events.

OK. One more case to do, which is, if it's not positive or negative, there's only one remaining choice, which is it's got to be 0. If sab squared is 0, then again looking back at the original definition, it means that the spatial piece is equal to minus c squared times the time piece so that they all cancel.

If you think about it, that's precisely the statement that these two events are located in just the right situation so the light beam that leaves one will just arrive at the other. Because it says that some of the first three terms, which is the distance squared, is equal to c squared times the time interval squared.

And that just says that something travels at the speed of light. It could travel that distance in that time and go from point a to point b or vice versa. Only one or the other, not both. But it's always one or the other.

So, for that reason, the interval is called lightlike. And it means that a light pulse can travel from a to b, or I could have interchanged a or b. Everything is squared. It doesn't matter which is which.

Now, there's a peculiar thing here. You would think that if a light pulse can travel from a to b, there would still be some relevant measure to how far apart a and b are. However, what we're basically seeing here is that if a and b are lightlike separated, in any given reference frame you could talk about what the time interval is and that will be equal to the space interval up to a factor of c.

But if we imagine looking at this at different frames, different inertial frames, these two points can get arbitrarily close together or arbitrarily far apart, depending on what frame we look at them in. There is no Lorentz invariant measure of how far apart they look. The Lorentz invariant-- the only Lorentz invariant measure simply tells us that they're lightlike related to each other.

And this leads to some very peculiar issues when you try to prove rigorous theorems about relativity. You can't really say whether two lightlike points, two lightlike separated points are close or far. Because there's no real meaning for them to be close or far.

OK. Let me just say one more fact about special relativity, and then we'll quit for today and come back on Thursday and then talk about how to extend this into general relativity. OK, there's a question. Yes?

AUDIENCE: Just really quick. For the Lorentz invariant to equal to 0, does that mean that the objects should be moving at the speed of light relative to each other? Is it like that?

PROFESSOR: OK. The question is, if the separation is lightlike, s squared is 0. Does that mean that these two objects are moving at the speed of light relative to each other or something like that? No, it does not. It only talks about their positions. It doesn't say anything about the motion of these objects. It's only a statement about their x and t-coordinates at some instant.

OK, let me still write one more equation on the blackboard to kind of finish the special relativity part of the discussion. In the end, we are interested in the metric. And what makes a metric a little bit different from a distance function is that metrics refer to infinitesimal distances. So we're going to want to know the infinitesimal form of that.

And it's obvious, so it's nothing to make a big deal about. But I think it's worth writing on the blackboard. The infinitesimal form of that equation is that ds squared is equal to dx squared plus dy squared plus dz squared minus c squared dt squared where dxdy, dz and dt are the infinitesimal coordinate differences between two events.

And it's in that form that we'll be beginning from and taking off into the world of general relativity and the metric of general relativistic spacetimes. So we'll continue with this on Thursday.