PROFESSOR: I'd like to begin by reviewing the last lecture, where we introduced the idea of non-Euclidean spaces. The important idea here is that general relativity describes gravity as a distortion of space and time. And that idea becomes crucial in cosmology, so we want to understand what that's about. As I explained last time, there are really two aspects to this general subject. There's the question of, how do we treat curved spacetimes and how do particles, for example, behave when they move through curved spacetimes. That we will be doing. There's also the important question of how does matter distort the spacetime, which is the Einstein field equations. That we will not do. That we will leave for the general relativity course that you may or may not want to take at some point, or maybe you already have.

The idea of non Euclidean geometry really goes back to Euclid himself, and the fifth postulate. The distinction between Euclidean geometry and what is generally called non Euclidean geometry is entirely in the fifth postulate. We never bothered changing the first four postulates. This fifth postulate is explained better in the next slide, where we have a diagram. The fifth postulate says that if a straight line falls on to other straight lines, such that the sum of the two included angles is less than pi, then these two lines will meet on the same side where the sum of the angles is less than pi, and will not meet on the other side.

This postulate was something that attracted attention, really from the very beginning, because it seems like a much more complicated postulate than the other four. And for many years, mathematicians tried to derive this postulate from the others, thinking that it couldn't really be a fundamental postulate. But that never worked, and eventually in the 17 and 1800s, mathematicians realized that the postulate was independent. And you could either assume it's true or assume it's false. And you get different versions of geometry in the two cases. We learned last time that there are four other ways, at least, of stating the things that are equivalent to the fifth postulate, and they're diagrammed here.

If a straight line intersects one of two parallel lines, it will also intersect the other. If one has a straight line and another point, there is one and only one line through that point, parallel to the original line. If one has a figure, one can construct a figure which is similar to it, of any size. And finally, the famous statement about sum of the angles that make up the vertices of a triangle. If there exists just one triangle for which it's 180 degrees, then that's equivalent to the fifth postulate, and you can prove that every triangle has 180 degrees.

The fifth postulate was questioned in a serious way by Giovanni Geralamo Saccheri in the 16, 1700s, who wrote a detailed study of what geometry would be like if the fifth postulate were false. He wrote this believing that the fifth postulate must be true. And he was looking for a contradiction, which he never found.

Things went further in the later 1700s, with Gauss and Bolyai and Lobachevski, who independently developed the geometry that we call Gauss Bolyai Lobachevski geometry, which is a two dimensional geometry, non Euclidean. It corresponds to what we now call an open universe, that we'll be learning about in more detail later today.

The Gauss Bolyai Lobachevski geometry was treated purely axiomatically by the three authors that I just mentioned. But it was given a coordinate representation by Felix Klein in 1870, which was really the first demonstration that it really existed. When one treats it axiomatically, one still always has the possibility that some contradiction could be found someplace. But by the time you put it into algebraic equations, then it becomes as consistent as our understanding of the real numbers, which we have a lot of confidence in, even though I don't think mathematicians really know how to prove the consistency of anything. But we have a lot of confidence in this kind of mathematics.

So by 1870, it was absolutely clear that this open geometry, this non-Euclidean geometry, was a perfectly consistent, is a perfectly consistent, formulation of geometry.

An important development coming out of Kline's work is that the idea about how one describes geometry changed dramatically. Prior to Kline, essentially all of geometry was done in the same way that Euclid did it, by writing down axioms and then proving theorems. Kline realized that you could gain a lot of mileage by taking advantage of our understanding of algebra and calculus by describing things in terms of functions. And in particular, geometry is described by giving a distance function between points.

This was further developed by Gauss, who realized that distances are additive. So if distances mean anything like what we think distances mean, it would be sufficient to describe the distance between any two arbitrarily close points. And then if you want to know the distance between two distant points, you draw a line between them, and measure the length of that line by adding up an infinite number of infinitesimal segments. So the idea that distances need only be defined infinitesimally was very crucial to our current understanding of geometry.

Gauss also introduced another important idea, which is a restriction on what that infinitesimal distance function should look like. Gauss proposed that it should always have the same quadratic form that it has for Euclidean distances. For Euclidean distances, the Pythagorean theorem tells us that the distance between any two points is the sum of the squares of the coordinate distances. And for non-Euclidean geometry, we generalize that by allowing each term in this quadratic expansion to have its own prefactor, and those prefactors could be functions of position. So g sub xx of xy is just a function of x and y. And g sub xy is another function of x and y. And g sub yy is another function of x and y. And the distance function is taken as the sum of those three terms.

The important feature that that quadratic form corresponds to, which was noticed by Gauss, is that if the distance function has that form, it means that even though the space is not Euclidean and will not obey, in general, the axioms of Euclidean geometry-- and in particular the fifth postulate-- it is still true that in a very tiny neighborhood, it will resemble Euclidean geometry, where the resemblance will become more and more exact as you confine yourself to tinier and tinier neighborhoods.

And we're kind of aware of this in everyday life. The surface of the earth is approximately spherical-- we'll ignore little things like mountains and roads and bumps, and pretend the surface of the Earth is spherical. Nonetheless, the surface of the Earth looks flat to us. And the reason it looks flat is that we see only a tiny little bit of it. And a tiny piece of a curved surface always looks flat. And mathematically speaking, the way to introduce enough assumptions to validate that conclusion is to assume that the local distance function is a quadratic function of this form.

And what Gauss originally proved is that if the distance function is of that form, it is always true that in a tiny neighborhood, to an arbitrary accuracy, you could define new coordinates-- x prime and y prime in the notation of this diagram-- where in terms of the new coordinates in the tiny neighborhood, the distances are just the Euclidean distances. ds squared equals dx prime squared plus dy prime squared. And that's a very crucial fact that we will be making use of, Einstein made use of, in the context of general relativity, which we'll be getting to.

OK, that finishes the review. Any questions about anything that we talked about last time? OK, great. Now what we want to do is to go on to apply these ideas in detail. In one of them in particular, build up a full description of closed and open universes today. And we are going to begin by giving a mathematical description of the simplest non-Euclidean geometry that we have available, which is just the surface of a sphere-- a two dimensional sphere embedded in three dimensions. That is, a two dimensional surface embedded in a three dimensional space, as is intended to be shown in that diagram.

The sphere is described simply by x squared plus y squared equals plus z squared equals r squared, where x, y, and z are just Euclidean coordinates. So in this case, our curved space, which is the surface of the sphere, can be embedded in a

Euclidean space of one higher dimension. That's not always the case. We should not pretend that that will always be the case. But when it is the case, it allows us to study that curved surface in a very straightforward way, because everything is really determined by the Euclidean geometry of the space in which this sphere is embedded.

Nonetheless, when we're done formalizing our description of the surface of the sphere, the goal will be to concentrate on what Gauss called the "inner properties," namely the properties of the surface itself. And we will try to pretend that the three dimensional space never even existed. It won't be required for anything that we'll be left with, once we have a solid description of the surface itself. And this will be very important for what we'll be doing later. So it is important to get in touch with the idea that we're going to study this sphere, making use of the fact that it could be embedded in three Euclidean dimensions. But in the end, we want to think of it as a two dimensional geometry, which is non Euclidean.

OK, so our goal will be to write down the distance function for some coordinization of the surface of the sphere. I should say at the beginning, when this picture makes it most obvious, that one of the reasons we might be interested in the surface of a sphere, if we're interested in cosmology, is that we know that we're trying to build cosmological models that are consistent with homogeneity and isotropy. Because we discussed earlier those are, to a very good approximation, valid features of the universe that we're living in. So the surface of a sphere has those properties. It's certainly homogeneous, in the sense that any point on the surface of a sphere will look exactly like any other point. If you were living on that sphere, and you didn't have any other landmarks, you'd have no way of knowing where on the sphere you were.

Furthermore, it's isotropic-- same in all directions. And when I say that, it's important that I really mean it in the context of the two dimensional surface, not the three dimensional geometry. So the three dimensional geometry is isotropic. If you sat at the center of that sphere and looked any direction in three dimensions, everything would look the same.

But that's not the isotropy that's important for us. We want to imagine ourselves as two dimensional creatures living on the surface. And then you can imagine that if you were a two dimensional creature living on the surface-- so you happen to be at the North Pole, because that's the easiest to describe-- you could imagine looking around in a circle, 360% available, and the world that you'd be living in would look exactly the same in all directions on the surface. And that's the isotropy that's important to us, because the two dimensional surface here is what we're soon going to generalize to be our three dimensional world. And it's isotropy within that world that we're talking about.

OK, so, first thing we wanted to do is to put coordinates on our two dimensional surface. If we want to ultimately forget the third dimension and live in the surface, we want to have coordinates to use in the surface. It's a two dimensional surface, so it should have two coordinates. And when we use the usual coordinization of a sphere, polar coordinates, well, it will be two angles, theta and phi. And there are some different dimensions that are used in different books, but I think almost all physics book use these conventions.

Theta is an angle measured from the z-axis, and phi is an angle measured by taking the point that you're trying to describe, which is that dot there, projecting it down into the xy plane, and in the xy plane, measuring the angle from the x-axis. So that's phi. And theta and phi are the polar coordinates describing a point on the surface of the sphere. And what we want to do is describe the distance function in terms of those polar coordinates-- that's our goal.

OK, to describe the distance function, what we want to imagine is two infinitesimally nearby points-- one described by coordinates theta and phi, and one described by theta plus d theta and phi plus d phi. So we have one point described by theta and phi, and another point described by theta plus d theta, phi plus d phi. So the coordinate changes are just d theta and d phi.

What we want to know is how much distance is undergone by moving from the first point to the second point. And the easiest way to see it is to make the changes one at a time. So first, we can just vary theta. And if we just vary theta, we see that the point described by theta and phi moves along a great circle, which goes through the z-axis. And the distance that the point goes is just an arc length as a piece of that great circle. And since this attended angle is d theta, and the radius is r, the distance of that arc length really just follows from the definition of an angle in radians. The arc length is r times d theta, and that's really the definition of d theta in radians. So if we vary theta only, the distance ds is just equal to r times d theta. Everybody happy with that?

OK. Now if we vary phi, it's slightly more complicated, but not much. If we vary phi, the point being described would-- if you vary phi all the way around-- make a circle around the z-axis. That circle does not have radius r. That's the one thing that may be a little bit surprising, until you look at the picture and see that it's true. The radius of that circle is r times sine theta. So in particular, if theta were 0, if you're up around the North Pole, going around that circle would just be going around the point. 0 radius. And you have maximum radius when you're at the equator, going all the way around. So again, we're going in a circle through an angle-- in this case d phi. So the arc length is just the angle times the radius. But the radius is r times sine theta, not r itself. So when we vary phi only, ds is equal to r times sine theta times d phi. Any questions?

OK, now, the next important thing to notice is that these two variations that we made are orthogonal to each other. When we varied phi, we moved in the horizontal plane. There's only motion in the x and y directions when we vary phi. When we vary theta, we move in the vertical direction. And those two vectors are orthogonal, as you can see from the diagram. And because we have two orthogonal distances that we're adding up, and because we're in underlying Euclidean space here-- we can think of those distances as being distances in the three dimensional Euclidean space that we're embedded in-- we get to use the Pythagorean theorem. So putting together these two variations, we get ds squared is equal to an overall factor of r squared, times d theta squared plus sine squared theta, d phi squared. And that formula then describes the metric on the surface of a sphere. And it describes a non-Euclidean geometry.

And once we have that metric, we can forget the three dimensional picture that we've been drawing, and just think of a world in which there are two coordinates-- theta and phi-- with that distance function. And that's the way we want to be able to think about it. OK, everybody happy?

OK, I want to mention, because it is perhaps useful in other cases, depending on your taste of how you like to solve problems-- the description I just gave of deriving this formula was geometric, that is, we drew pictures and wrote down the answer based on visualizing the pictures. But it can also be done purely algebraically. To do it purely algebraically, one would first write down formulas that relayed the angular coordinates-- I should go back to slides. What's going on here? Is my computer frozen? [INAUDIBLE] Yes?

AUDIENCE: For that formula, are you assuming that d theta and d phi are really small so that you can [INAUDIBLE] the triangle?

PROFESSOR: Yes, these are infinitesimal separations only. That's the key idea of Gauss. And yes, we're making use of that. This formula will not hold if d theta and d phi were large angles. Holds only when they're infinitesimal. Yes?

AUDIENCE: And then similiarly, we can use the line integral for calculating the distances based on this metric?

PROFESSOR: Yes, that's right. If we wanted to know the distances between two finitely separated points, we would construct a line between them, and then integrate along that line. And by line what we mean is the path of shortest distance, which we'll be learning about more next time, probably. Those are not necessarily easy to calculate. In this case, they're calculable, but not really easy. Any other questions?

OK, so what I wanted to do was to look at the definition of the coordinate system, as now shown on the screen. And from that, we can write down the relationship between x, y, and z, and theta and phi. And those relationships are that x is equal to r times sine theta times cosine phi. y is equal to r times sine theta times sine phi.

And z is equal to r times cosine theta. And once one writes those formulas, then one can just use straightforward calculus to get the metric, without needing to draw any pictures at all. You may have wanted to draw pictures to get these formulas, but once you have these formulas, you can get that by straightforward calculus. I'll sketch the calculation without writing it out in full.

But given this formula for x, we can write down what dx is by calculus, by chain rule. So dx-- it's a function of two variables. So it would be the partial of x, with respect to theta, times d theta, plus the partial derivative of x, with respect to phi, times d phi. And we'll go work out what these partial derivatives are if I differentiate this with respect to theta. The sine theta turns into cosine theta.

So the first term becomes r cosine theta, cosine phi, d theta. And then plus, from here we have the partial of x with respect to phi. We just differentiate this expression with respect to phi. The derivative of cosine phi is minus sine phi. So the plus sign becomes a minus sign-- r sine theta, sine phi d phi.

And then I won't continue, but we could do the same thing for dy and dz. And once we have expressions for dx, dy, and dz, we can calculate ds squared, using the fact again that all of this is embedded in Euclidean space. That's where we're starting, although in the end, we want to forget that Euclidean space. But we could still make use of it here, and write ds squared is equal to dx squared, plus dy squared, plus dz squared.

One can then plug in the expression that we have for dx, in terms of d theta and d phi, and the analogous expressions that I'm not writing down for dy and dz. And when one puts them in here, one makes lots of use of the identity that cosine squared plus sine squared equals 1 And after using that identity a number of times, what you get when you just put together this algebra is exactly what we had before-- r squared times d theta squared, plus sine squared theta, d phi squared.

So the important point is that once you have the identities that relate these two different coordinate systems, and if you know the distance function in the xyz coordinate system, you're home free, as far as geometry is concerned. One could just use calculus from there on if one wants to. Although usually the geometric pictures make things easier. OK, any questions about that?

OK in that case, we are now ready to move on, having discussed the two dimensional surface of a sphere embedded in three Euclidean dimensions. The next thing I'd like to do-- and this really will be our closed universe cosmology-- we're going to discuss the three dimensional surface of a sphere in four Euclidean dimensions by analogy. The previous exercise was a warm- up. Now we want to introduce a four dimensional Euclidean space.

So our sphere now will obey the equation x squared plus y squared plus z squared plus-- we need a new letter for our new dimension in our four dimensional space, and I'm using w. x, y, z, and w. So that equation describes a three dimensional sphere in a four dimensional Euclidean space. And our goal is to do the same thing to that equation that we just did to the two dimensional sphere embedded in a three dimensional space.

Now of course it becomes much harder to visualize anything. If you ask how do we do visualize things in four dimensions, I think probably the best answer is that we usually don't. And what we're going to take advantage of mainly is that once you know how to write things in terms of equations, you don't usually have to visualize things. Or if you do, you can usually get by, by visualizing subspaces of the full space.

If you want to visualize the full sphere in some rational way, I think the crutch that I usually use, and that most people use, is if you have just one extra dimension, in this case w, try to think of w as a time coordinate. Even though it's not really a time coordinate, it still gives you a way of visualizing things. So if we think of w as a time coordinate here, the smallest possible value of w would be minus r. The maximum possible value of w would be plus r, consistent with this constraint, consistent with being on the sphere.

So when w is equal to minus r, the other coordinates have to be 0 to be on the surface of the sphere. So you can think of that as a sphere that's just appearing at time minus 1, with initially 0 radius. Then as w increases, x squared plus y squared plus z squared increases. So you could think of it sphere that starts at 0 size, gets bigger, gets as big as r in radius when w equals 0 and then gets smaller again and disappears. So that's one way of thinking of this four dimensional sphere. Yes?

AUDIENCE: Is that kind of like looking at the cross sections of the sphere?

PROFESSOR: Is that kind of like looking at the cross sections of the sphere? Exactly, yes. One is looking at cross sections of the sphere-- successive cross sections at successive values of w. And if you do it in succession, it makes w act like a time coordinate. But you're right. The fact that w is a time coordinate is kind of irrelevant. You could imagine just drawing them on a piece of paper in any order. And for any fixed value of w, you're seeing a cross section of what this sphere looks like. That is how the xyz coordinates behave for a fixed value of w.

Now to coordinatize the surface of our sphere. Last time we used two coordinates, because we had a two dimensional surface. This time we're going to want to use three coordinates, because this is a three dimensional surface that we're describing. And that means that we need at least one new coordinate. And the new coordinate that I'm going to introduce will be another angle, which I'm going to call psi. And the angle I'm going to define, as in this diagram, as the angle from the new axis, the w axis. So psi is the angle of any arbitrary point to the w axis.

And therefore the w coordinate itself is going to be r times cosine psi, just by projecting. And the square root of the sum of x squared plus y squared plus z squared is then the other component of that vector. And beyond the sphere, it's easy to see that the square root of x squared plus y squared plus z squared has to be r times sine psi.

OK, now we still need two more coordinates. This is only one coordinate-- we want to have three. But the two other coordinates will just be our old friends, theta and phi We're going to keep theta and phi. And in order to keep them, what we'll imagine doing is that for any point on the surface of this three dimensional sphere of the four dimensional space, we could imagine just ignoring the w coordinate.

And then we have x, y, and z-coordinates, and we can just ask what are the values of theta and phi that would go with those x, y, and z-coordinates. So theta and phi are just defined by quote, "projecting" the original point into the three dimensional xyz space, which just means ignore the w coordinates, look at xyz, and ask what would be the angular coordinates, theta and phi, for those values of x,y, and z.

And it's easy to take those words I just said and turn them into equations. We like to write down the analog of these equations. But we want to have four equations now that will specify x, y, z, and w as a function of our three angles, psi, theta, and phi. But that's not hard.

I'll show you at the bottom with w. w we already said is just r times the cosine of psi. Just coming from the fact that psi is defined as the angle from the point to the w-axis, and that's enough as you see from the picture to imply that w is equal to r times cosine psi. The other point, x,y, and z, really just follow by induction from what we already know.

Each of these formulas will hold in the three dimensional subspace, except that r, the radius of a sphere in the three dimensional subspace, is not r anymore but is r times sine psi. So x is equal to r times sine psi times what it was already-- sine theta cosine phi. y is equal to 4 times sine psi times sine theta sine phi. And z is equal to r times sine psi times cosine theta. So I just take each of these equations and multiply them by sine psi to get the new equations.

And you can straightforwardly check-- if we take x squared plus y squared plus z squared plus w squared here, it makes successive use of the identity that sine squared plus cosine squared equals 1. We'll be able to show that x squared plus y squared plus z squared plus w squared equals r squared, like it's supposed to. OK, so is everybody happy with this coordinatization?

OK, I should mention, by the way, that if you ever have the need to describe a sphere in 26 dimensions or whatever, this process easily iterates, once you've known how to do it once. That is, every time you add a new dimension, you invent a new letter for the new axis. You define a new angle, which is the angle from that access. And then the new coordinatization is just to set the new coordinate equal to r times the cosine of the new angle. And then take all the old equations and put in an extra factor of the sine of the new angle. And you got it. So you could do that as many times as you want, if you want to describe a very high dimensional sphere.

We should say something about the range of these angles. The original angle phi-- maybe I should go back a few slides now. The original angle phi, as you can see from the slide, goes around the xy plane, so it has a range of 0 to 2 pi. The original angle theta is an angle from the z-axis, and the furthest you could never be away from pointing towards an axis is pointing away from it. And that's pi, not 2 pi. So theta has a range of 0 to pi.

And similarly, psi is also defined as an angle from an axis. So again, the furthest you could ever be away from pointing towards an axis is pointing away from it. So psi, like theta, will have a range of 0 to pi. And if you ever need to coordinatize a 26 dimensional sphere, as I just mentioned, you keep adding new angles. Each of the new angles goes from 0 to pi.

Thus each new angle was introduced as an angle between the point that you're trying to describe and the new axis. So, they're all angles like theta and psi. So 0 is less than phi, is less than 2 pi. But 0 is-- these should be less than or equal to's-- 0 is less than or equal to theta, is less than or equal to pi. And 0 is less than or equal to psi, is less than or equal to pi.

OK, next we want to get the metric of our three dimensional spherical surface embedded in our four dimensional Euclidean space. And I'm going to do it by the geometric sort of way. I'll try to just motivate the pieces. There'll be some cross here between actually algebra and geometry.

Let's see, first I should mention that once you have this, you can get the answer by the same brute force process that we describe, but didn't really carry out here. That's tedious, but it's pretty well guaranteed to get you the right answer if you're careful enough, and does not require drawing any pictures or having any the visualization of the geometry, so it does have some advantages. But I will not do it that way. I will do it in a geometric sort of way, because I think it's easier to understand the geometric sort of way.

So here goes. As we did over there, we will vary our coordinates one at a time, and then see how we can combine the different variations. Again we'll find that they're orthogonal to each other. So I'll be able to combine them just by adding the squares. But we don't necessarily know that the beginning.

So let's start by varying psi, the new coordinate. And there, things are very simple because our new coordinate is just defined as the angle from an axis. So if we vary psi, the point in question just makes a circle around the origin, and the circle has radius, capital R. So the variation if I vary psi is just r times d psi. So ds is equal to r times d psi. What could be simpler?

OK, now we'll imagine varying either theta or phi or both. And since we already pretty well understand this three dimensional subspace-- this is the previous problem-- I'm going to talk about varying them simultaneously. So vary theta and phi. Well then we know that ds squared is really given by this formula. We are just varying theta and phi in a three dimensional space. The fourth direction doesn't change in this case.

But the radius involved is not what we originally called r. But the radius in the three dimensional space is r times sine psi. That's the square root of x squared plus y squared plus z squared. So what we get is ds squared is equal to R squared times sine squared psi times d theta squared plus sine squared theta, d phi squared. Everybody happy with that?

OK, now I'm going to first jump ahead and then come back and justify what we're doing. But if these variations are orthogonal-- which I will argue shortly that they are, so I'm not doing this for nothing-- if the variations are orthogonal, then we just add the sum of squares using the generalized Pythagorean theorem. In this case, Pythagorean theorem in four dimensions-- that's four Euclidean dimensions, so we should be able to use it.

So what we get for our final answer is ds squared is an overall factor of r squared, times d psi squared plus sine squared psi, times d theta squared, plus sine squared theta, d phi squared. I just added the sum of the squares. Now I need to justify this orthogonality.

OK, to do that, let me introduce a vector notation in the four dimensional space of x,y z, and w. OK, we're justifying this in the Euclidean embedding space. And the Euclidean embedding space of these transformations are orthogonal.

So let me imagine varying psi. And then I could construct a four dimensional vector dr-- I'll call it sub psi because it arises from varying psi. So this is the four dimensional vector that describes the motion of this point r as psi is varied.

And first let me just give these components names. I'll call it-- rather than repeat the r, I'm just going to call this d psi sub x, d psi sub y, d psi sub z, and d psi sub w. This is just by definition. I'm just naming the components of that vector. And since the vector already has a subscript, I don't want to have two subscripts. So I've changed the name of the vector for writing as components.

And similarly, here I will just vary one of these two angles, theta and phi I'll just vary theta, and let you know that varying psi is no different, and you can easily see that. So if I vary theta, the variation of r when I vary theta will be dr sub theta. And its components I will just call d theta x, d theta y, d theta z, and d theta w. So these are just definitions. I haven't said any actual facts yet. But I've defined these two vectors, and given names to their components.

And now we want to look at them and take their dot product. Their dot product is just a four dimensional Euclidean dot product. So the dot product is just dx times d theta d psi y d theta y plus d psi z times d theta z plus d-- blah. I think we should write it.

So this actually is now a fact about Euclidean geometry in four dimensions. The dot product of these two vectors is just equal to the product of the x components, plus the product of their y components, plus the product of their z components, plus the product of their w components. And now what we want to do is to look at this sum and argue they're 0, because if they're orthogonal, the dot product of two vectors should be 0.

OK, so let me first look at the dr sub theta vector. OK, what do we know about it? Well, we know that when we vary theta, from these formulas, w does not change-- and we can easily see that from the picture as well. So d theta sub w equals 0. And since these are all products, that means that this last term vanishes, no matter what d psi w is. So we know we don't need to worry about that term. We only need to worry about these three terms.

Now what do we know about those three terms? If you look at dr sub theta, and look at its three spacial components-- x,y, and z-- from here, we could see that varying theta does the same thing to x, y, and z as it did over here, except a different overall factor out front. So in particular, what I want to point out is that varying theta does not change x squared plus y squared plus z squared. It leaves it constant.

So if we think of the xyz space, we could imagine a sphere in the xyz space, and varying theta always causes a variation that's tangential to that sphere. It never moves in the radial direction. So the three vector, defined by d theta x, d theta y, d theta z, is tangential in the three dimensional subspace xyz. Just as it was when we didn't have a w coordinate. The w coordinate doesn't change anything here. So that's a little bit [? soluable. ?] Are people happy with that? Do you know what I'm talking about?

OK. Now we want to look at dr sub psi, and it will have a w component-- d psi sub w-- but we don't care about that. We know we don't care about it because that piece already dropped out of our expression. So we want to know what this vector looks like in the xyz space. We don't care about what it looks like in the w space.

So in the xyz space, we can look at these formulas here. As we vary psi, x, y, and z could change. But they all change by the same factor-- whatever factor psi changes by. The same sign psi appears in all three lines. So changing psi can only multiply x, y, and z all by the same factor.

And what that means is that if you think of this geometrically in the xyz space, varying psi moves the point only in the radial direction. If you multiply all of the coordinates by a constant, you are just moving in the radial direction. So dr psi has the property that when we look at only its x, y, and z components-- d psi x d psi y, d psi z, it is radial in the three dimensional subspace.

So, the sum of these three terms-- this is what we're trying to evaluate-- is the dot product of a radial vector and a tangential vector. And the dot product of a radial vector and the tangential vector will always be 0, because there are orthogonal to each other. Sorry for the overlap here, but equal 0, and that's because radial is perpendicular to tangential. OK, everybody happy with that?

OK if so, we have important result now. We have derived the metric for the three dimensional surface of a four dimensional sphere embedded in four Euclidean dimensions. And that, in fact, is precisely the closed universe of cosmology. It's the homogeneous isotropic description of a closed universe.

OK, next thing I want to point out is just a definition. An important feature of non Euclidean geometry and general relativity-- because they're connected to each other-- is that there never is a unique, useful coordinate system. And Euclidean spaces, there is a unique, useful coordinate system. It's the Cartesian system. Sometimes it's also useful to use polar coordinates or something else, but by and large the Cartesian coordinate system is the natural description of Euclidean spaces. And the coordinates of a Cartesian coordinate system really are distances. Once, however, you go from Euclidean geometry to non Euclidean geometry-- from flat spaces to curved spaces-- you're usually in a situation where there just is no natural coordinate system.

When we invented this psi, theta, phi, we really made a number of arbitrary choices there. We could have defined things quite differently if we wanted to. So in general, one has to deal with the fact that the coordinates no longer represent distances, and therefore there's a lot of arbitrariness in the way you choose the coordinates in the first place.

So in particular, we could think of psi equals zero as the center of our new coordinate system, with coordinates psi, theta, and phi. psi equals 0 corresponds to being along the w axis, so it's a unique point. When you say that psi is equal to 0, it no longer matters what theta and phi are. You're at the point w equals r and x, y, and z equals 0. So we can think of that as the origin of our new coordinate system. And we can think of then the value of psi as measuring how far we are from that origin. So psi will become our radial coordinate.

So thinking of it as our sphere, psi equals 0 we might think of as the North Pole this sphere. But we're also going to think of it as the origin of psi, theta, phi space. And we will sometimes use other radial variables-- other variables for the distance from the origin. So in particular, another coordinate that's very commonly used is u, which is just defined to be the sine of the angle psi. And notice that the sine of the angle psi shows up in a lot of equations. So taking that as our natural variable is a reasonable thing to do on occasion. Both are useful.

If we do use this, then we can rewrite the metric in terms of u, instead of psi. And in order to do that, we just have to know how du relates to d psi, because the metric is written in terms of the differentials of the coordinates. So that's easy to calculate. du would be equal to cosine of psi d psi. But if we're trying to rewrite the metric solely in terms of u, we don't want to have to divide or multiply by cosine psi, because that's written in terms of psi. But we could express cosine psi in terms of u, because if u equals sine psi, then cosine psi is the square root of 1 minus u squared.

So I can rewrite this as the square root of 1 minus u squared times d psi. And then d psi squared, which is what appears in our metric, can be rewritten just using that as du squared divided by 1 minus u squared. And the full metric now, in terms of the u theta and psi coordinates, can be written as r squared times du squared over 1 minus u squared, plus u squared times d theta squared plus sine squared theta d phi squared. So this is another way of writing the metric for this three dimensional sphere embedded in four Euclidean dimensions. Any questions? Yes?

AUDIENCE: The value of u doesn't uniquely determine a point on the sphere, right? Because

[INAUDIBLE].

PROFESSOR: Very good point. In case you didn't hear the question, it was pointed out that the value of u, unlike the value of psi, is not uniquely indicate a point, because on the entire sphere, there are two points u for every value of sine psi-- one in the northern hemisphere and one in the southern hemisphere, if we think of hemispheres as dividing whether w is positive or negative.

So in fact, that's right. If we use the u coordinate, we should remember that if we want to talk about the whole sphere, we should remember that the whole sphere is twice as big as what we see if we just let u vary between 0 and its maximum value, 1. u equals 1 corresponds to the equator of this sphere. Another point which I'd like to make now that we've written it this way is that writing it this way is the easiest way to see-- although we can see it in other ways as well-- that if u is very small, if we look right in the vicinity of the origin of our new coordinate system, if u is very small, 1 minus u squared is very close to 1. The square of a small quantity is extra small, so this denominator is extra close to 1. And that means that for very small values of u, what we have is du squared plus u squared times this quantity. And this is just polar coordinates in Euclidean space.

So for u very small, we do see that we have a local Euclidean space. And that, if you might remember, was one of the key points about writing the metric as the sum of squares in the first place. And it's true about every point, although the coordinate system here only makes it obvious about the origin. But we know that the space actually is homogeneous, from the way we constructed it.

So what's true about the origin is true about any point. So the coordinate system, the metric around any point, if you look close enough in the vicinity of the point, looks like a Euclidean space, which is what we expected from the very beginning, but we can see it very explicitly here.

OK. So far, this is just geometry. But this will be a model for a homogeneous, isotropic universe. We know it's homogeneous, we know it's isotropic from the way we discussed it. When we write the metric this way, it's obvious that it's isotropic about the origin, because this construction we know is just polar coordinates, and we know polar coordinates don't really single out any direction, even though manifestly they look like they do, because you're measuring angles from the z-axis. But we know that really describes an isotropic sphere.

It's not obvious that this formula describes a homogeneous space, because it makes it look like u equals 0 is a special point. But we do know that it did correspond to a homogeneous space from the way we constructed it. It really is just a three dimensional sphere embedded in four Euclidean dimensions. And if you think of it as the sphere, there clearly is no special point on the sphere.

So the homogeneity is a feature of this metric, but a hidden feature. It's hard to see how you would transform coordinates to, for example, put a different point at the origin. But it is possible, and we know it's possible, because of the way we originally constructed. And if we had to do it, we go back to the original construction and actually do it.

That is, if I told you I wanted some other point in that system to be the origin, you could trace it all way back to the four dimensional Euclidean space, and figure out how you have to do a rotation in that four dimensional Euclidean space to make the point that I told you I wanted to be the origin to actually be the origin.

So you'd be able to do that. It would be some work, but you would in fact know how to do that. We really do understand that the space is homogeneous, which is guaranteed by our original construction.

OK. Finish with the basic geometry-- one more chance to ask any questions about it. OK, next I want to go on now to talk about how this fits into general relativity. And here we are going to be confronting-- actually the only place we ever will confront-- the issue of how matter causes space to curve, which is the aspect of general relativity that we're not really going to do it all.

So I'll basically just be giving you the answer. Although we will in fact know enough to narrow down the range of possible answers, pretty much. But there will be a fudge factor that I'll have to just tell you the right answer for.

So what I want to do now is to make a connection between this formalism and the model that we already discussed of the expanding universe whose dynamics we derived using Newtonian mechanics. So using internet Newtonian mechanics, we introduced a scale factor a of t. And convinced ourselves that a dot over a squared is equal to 8 pi over 3 g rho, minus kc squared over a squared.

And furthermore, that this a of t describes the relationship between physical distances and coordinate distances. Namely, if we have objects that are at rest in this expanding universe, comoving objects is the phrase usually used to describe that.

If we have comoving objects, the comoving objects will sit at fixed coordinates in our coordinate system. And the distance between any two of them will be some fixed distance-- l sub c-- a coordinate distance. But the physical distance will vary with time, proportional to this scale factor. So the physical distance between any two points will be a function of time, which is the scale factor times the time-independent coordinate distance.

OK, if we look at our metric for this sphere, and say we're going to assume that this is going to be the metric that describes the space that we're describing over here, then clearly this r that sits out front rescales all of the distances. All the distances are proportional to r, just as all the distances here are proportional to a. So that could only work if r is proportional to a. So that's our key conclusion here-- that r is going to have to vary with time, and be proportional to a.

But we can even say a little bit more than that, because we can look at dimensionality. And here comes in handy that I insisted from the beginning in introducing this idea of a notch. The notch helps us here to get this formula right. The units of r-- r is just distance-- so the units of r are just distance units-- and I'll pretend that we're using meters. It doesn't really matter what actual units we're using, so I'll call this m for meters.

On the other hand, a of t comes from this formula, where physical distances are measured in meters, but coordinate distances are measured in notches, so a is meters per notch, as we've said many times before. So a of t is meters per notch. So that tells us something about this constant of proportionality-- it has to have the right units to turn meters per notch into meters. That is, it has to have units of notches.

Where did we get notches from? The other thing that we know is that the little k that appears in the Friedman equation, which we know is a constant-- we also know it's a constant that has units of 1 over notches squared, as we worked out some time ago. Units of k is 1 over notches squared. And that's the only thing around that we can find that has units of notches, so we're going to use that to make the units turn out right in this equation.

So to get the units right, we can write it as r of t is equal to some constant-- actually, it's more convenient to square this-- r squared of t is equal to some constant times a squared of t, divided by k. And now that constant is dimensionless. The units are all built into the a's and the k's and everything else.

AUDIENCE: Is the constant multiplied by that, or--

PROFESSOR: Yes. That equals sign was a big mistake. Constant times a squared over k. And that constant is now dimensionless. Because this is meters squared, this is meters squared per notch squared, and this is just per notch squared, so the notches cancel over here.

OK, now what this constant is really is the statement of how curved is our space-- r is really a measure the curvature of our space-- how curved is our space for a given description of what the matter is doing? a of t is directly related to the [INAUDIBLE] calculations you've already done.

So this clearly is a formula of exactly the type that I told you we weren't going to learn how to deal with. We're not capable in this course of describing the Einstein field equations, which determine how matter causes a space to curve.

So I'll just tell you the answer. The answer is that this constant is equal to 1. So it's certainly a simple answer, but we won't be able to derive it. So we end up with just r squared of t is equal to a squared of t, divided by k.

Now it may be useful at this point to remind ourselves what little k meant in the first place. We introduced little k in the context of describing a purely Newtonian model of an expanding universe, where we imagined just a finite sphere of matter expanding. And in doing that, we defined k to be equal to minus 2 e divided by c squared, where e itself was not a quantity that we proved was conserved. And it's related to an energy, but as we discussed, there are various ways that you could relate it to the energies of different pieces of the system.

But e was given by that expression. And if I put this into that, just to see more clearly how our discussion relates to our Newtonian discussion, we get r squared is equal to a squared of t times c squared over 2 e.

And I wrote it this way mainly to illustrate, or demonstrate, an important point, which is that our calculation was non-relativistic, but there's a c squared appearing in this formula. This c squared really just arose from our definitions. And if we had some other quantity here whose units were-- we have to sub units of meters per second for this formula to come out right. If we put some other velocity here, then this constant would not be 1 anymore, but something else. So saying that this constant is 1 is saying that this formula is meaningful. Putting the c squared there simplifies things.

And that in turn means that the curvature really is a relativistic effect. OK, we think of relativistic effects as effects that disappear as the speed of light goes to infinity. So this formula tells us that as the speed of light goes to infinity-- for fixed values of things like the mass density, which are buried in a squared of t-- as the speed of light goes to infinity for fixed values of the mass density, r squared goes to infinity.

Now infinity may sound like it's backwards, but it's the right way. 1 over r is really the curvature. r is the radius of curvature of the space. As r goes to infinity, our curved space looks more and more flat.

So we're saying that if you could imagine varying the speed of light, as you made the speed of light larger and larger, this space would become flatter and flatter. So this curvature of the space really is a relativistic effect, which is related to the fact that the speed of light is finite and not infinite. Yes?

AUDIENCE: Sorry, when we replace a with that, are we missing a minus sign?

PROFESSOR: Oh we might be, yeah. Minus sign now fixed. The point is that for the case we're talking about, e would be negative and k would be positive. So this formula needs an absolute value sign in it. Thank you.

OK, it may also be useful to relate r more directly to astronomical observables, which we can do, because we have the Friedman equation up there, which relates a to rho. And a dot over a is also the Hubble expansion rate, so that's h squared. So this formula tells exactly how to write a in terms of rho and h squared.

And in fact, it tells us how to write a over the square root of k in terms of rho and H squared. And that's exactly what r is-- it's a squared over k. So putting those equations together, we could write r is equal to c times the inverse of the Hubble expansion rate, over the square root of omega minus 1. Where omega you would call is equal to rho divided by rho sub c. And rho sub c is 3 h squared over 8 pi-- Newton's constant.

So this formula says two things. It says that the radius of curvature becomes infinite if c were infinite. And that says what I already said, that's a relativistic effect. It also tells you the radius of curvature goes to 0 as omega goes to 1. So omega approaching 1 is the flat universe case, which is what we've already mumbled about, but this formula shows it very directly. As omega approaches 1, for a fixed value of h and a fixed value of the speed of light, the radius of curvature goes to infinity. The space becomes more and more flat.

Look, I'm just going to write one more formula, which is really just a redefinition, but an important redefinition as far as where we're going to be going next. And then we'll finish today's lecture, and continue next week.

What I wanted to do is to put these definitions back into the metric itself. So we can write ds squared is equal to a squared of t divided by k-- which is what we previously called r squared-- times du squared over 1 minus u squared, plus u squared times d theta squared plus sine squared theta d phi squared. And now what I want to do is make one further redefinition of this radial variable, which, remember, initially was psi. Then we let u be equal to the sine of psi.

Now I'm going to make one further substitution. I'm going to let little r be equal to u divided by the square root of k, to bring this k inside. And that then is also equal to sine of psi, divided by the square root of k. And when we do that, the metric takes a slightly simpler form. ds squared is equal to a squared of t all by itself on the outside now.

And then this, after we factor in the k, becomes dr squared over 1 minus k r squared, plus r squared times d theta squared plus sine squared theta, d phi squared. And this is the form that the metric is usually written in. It's called the Robertson-Walker metric.

So we've only discussed closed universes. I had hoped to discuss closed and open, but open will in fact follow very quickly from what we already have. So we'll begin next time by discussing open universes.