PROFESSOR: We do have a little bit of finishing to do because we didn't quite finish the dynamics of homogeneous expansion last time. So we'll begin by finishing that after a brief review of where we were last time. And then we'll move on to discuss non-euclidean spaces, which I hope will be still the bulk of today's lecture.

OK. In that case, let's get going. Again, as I said, I want to begin by just reviewing some of the things we talked about last time. And you should consider this a good opportunity to ask questions if you discover there are things that you're not really sure you understood as well as you'd like.

We were talking about the evolution of a closed universe. And to summarize that calculation we first re-shuffled the first order Friedmann equation, the equation for a dot over a quantity squared, by bringing all the d t's to one side and all the d a's on the other side after doing a little bit of rescaling. And we got this equation. Which we then said we can integrate. And the integral from time will go from time 0 to some arbitrary final time that we called t tilde sub f, where the "tilde" indicates that we multiplied by c. And the sub f means it's the final time of our calculation.

And on the other side we have to integrate with corresponding limits of integration. Corresponding to t equals 0 is a equals zero. So the lower limit of integration is 0. And the final limit of integration is just the value of a tilde at the final time, whatever that is.

Then we discovered that we could actually do the integral on the right if we made a substitution. And in the lecture last time we did it in two stages. But the combined substitution is just to replace a tilde by cosine theta, according to this formula, if we combine the two substitutions that we made last time.

And if we do that we could integrate the right hand side. And the integral then gives us t tilde f is equal to the integral of that, which is just this expression. And the expression below that is just the substitution itself. How to relate a tilde to theta, according to substitution which gave us that formula. So these two formulas together allow us to determine t tilde sub f, and a tilde sub f in terms of theta and alpha. And once we had that we realized we no longer needed to keep those sub f's because that was really just a way of keeping track of our notation during the calculation. It holds for any theta sub f. So it holds for any theta.

So then we just rewrote those same formulas without the sub f's. And here we wrote it removing the tildes, replacing them by their definitions. And this was the final answer. This describes the evolution of a closed universe expressed in this parametric form. That is, we were not able to explicitly write a as a function of t, which is what would have liked to determine how the expansion varied as a function of time. But instead we introduced the auxiliary variable theta, often called the development angle. And in terms of theta we can express both t and a, and thereby indirectly have an unambiguous relationship between a and t.

Any questions about that?

OK. Then we noticed that those were, in fact, the equations of a cycloid. And I won't go through the argument again, but the key point is that the evolution of our closed universe scale factor, as a function of time, is described by what would happen if you had a disk rolling on the horizontal axis with a point marked on the disk which was initially straight down, and then as the disk rolls that point traces out a cycloid.

And that is exactly the evolution of a closed universe. It starts at size zero, goes to a maximum size, and then contracts again in a completely symmetric way. The contraction phase is the mirror image of the expansion phase.

And then we went on to calculate the age of a closed universe. And that was really more of an exercise in algebra than anything else. The age is really given by this formula to start with. This expresses the age, t, in terms of alpha and theta. And the only problem with that is that nobody really knows what alpha and theta mean.

It's much more useful to have an expression for the age in terms of things that astronomers directly measure. And one needs two things to replace alpha and theta. And the kinds of things astronomers directly measure are things like the Hubble parameter and the mass density, where they often express the mass density in terms of omega where omega is the mass density divided by the critical density. And that's what I chose to do in the formulas that we went on to derive.

So our goal is simply to take that formula for t and figure out how to express the alpha and the theta that appear in it so that we can get an expression in terms of h and omega. And to do that, we'll go one step at a time.

We started by just writing down the Friedmann equation. And the Friedmann equation has an a in it. And everything else in it is rho, which can be expressed in terms of omega without any trouble, and h itself. So everything else in it consists of variables that we're accepting to be part of our final answer. So we could solve the Friedmann equations for a or a tilde and that will allow us to eliminate any a tilde's from our expressions.

Next, alpha was originally defined in terms of this formula which only involves rho and a tilde. We now know what to do with a tilde. We could substitute that formula. And rho, we always knew what to do with. We could express that in terms of omega. So we can instantly solve that equation and get an equation for alpha in terms of the quantities that we want to appear in our final answer.

And there's still one more thing we need. We need an expression for theta. And theta we can get by looking at the other of those two parametric equations, the one that's not the "t equals" equation, but rather the "a equals" equation. And in this equation we know everything except theta itself. So we could substitute for a over root k from up here- that. And on the right hand side we can place alpha by that expression and then the 1 minus cosine theta stays. And now we could solve this equation for theta.

I might just mention that, in lecture last time I actually mis-copied this equation. I forgot the factor of omega in the numerator. So if any of you were taking notes you can go back and correct your notes. But it's correct in the printed notes. And now it's correct on a screen.

So this you could either solve for cosine theta-- initially what I did was to solve for cosine theta. But then it turned out to be more useful to know what sine theta is, because sine theta appears in that parametric expression for the age. So if you know cosine theta you can, of course, get sine theta because sine theta is just the square root of 1 minus cosine squared theta. And that's what we did.

And that's how we got a square root here. And since square roots can have either sign, there's a plus or minus there which depends on where you are in the evolution of our universe. The right hand side here is always positive because we define that square root symbol to mean the positive square root, where the plus or minus, in the end, tells you the sine. And sine theta itself-- we know what theta does. It goes from 0 to 2 pi. So sine theta starts out positive and after theta crosses pi, sine theta is negative for the second half of the evolution. Meaning, sine theta is negative for the contracting phase and positive for the expanding phase.

We then put all that together into the formula for CT, which is alpha theta minus sine theta. The alpha becomes this factor. And the theta minus sine theta becomes the arc sine of this expression, minus or plus that expression, where this is just sine theta from that formula. And it's minus sine theta, which is why the plus minus becomes minus or plus. The signs get a little tricky, but it's, in principle, straightforward. It all just follows from this formula. And if you know the sine of theta you have all your problems solved.

So we put all that together into a table. This is just a copy of the same formula that we had on the previous slide. Theta is what appears in this right hand column. It's indicated as the sine inverse of some expression. It refers to that expression-- an abbreviation. So we know what theta does. Theta just goes from 0 to 2 pi in quadrants. At least, it pays to divide it up into quadrants.

Sine theta is always positive for the first half and negative for the second half. And that means that we have the plus sign, or the upper sign, for the first half of the motion and the minus sign for the second half of the motion. Omega we could just calculate in terms of theta. No problem about filling in the omega column. And we know that we're expanding for the first half and contracting for the second half, so that really completes the table.

And the important thing when you're actually using this formula is to decide what theta is. And once you have that, the sine of that-- the inverse sine of that-- well, no. Excuse me. Theta itself appears there, and the sine of theta appears there. And theta itself you have to figure out which is the relevant value in terms of where you are in the evolution. The point is that sine theta itself goes-- does not uniquely determine what theta is.

OK. That, I think, completes our discussion of the evolution of closed universes. I think it completes everything that we did last time. So are there any questions? OK. Good.

To finish our discussion of the evolution of matter-dominated universes, we go on to discuss open universes. And open universes are really the same algebra as closed universes. They just differ by the sine of k. Because one doesn't like to deal with negative numbers, I defined kappa to be equal to minus k, so that for our open universes kappa was positive while k was negative.

Then I used a different substitution for a tilde. Instead of a tilde being a divided by the square root of k, which in this case would be the square root of a negative number-- which, one doesn't like to deal with imaginary numbers if one doesn't have any need to. So instead, for the open universe I'm defining a tilde to be a divided by the square root of kappa, so that a tilde is again real.

Then I'm going to skip all the algebra here. There's a little bit more of it shown in the printed lecture notes. But there are no new concepts here. Everything is really the same, conceptually, as it was for the closed universe.

One difference is that this time, instead of getting trigonometric functions, we find that we get hypergeometric functions. Hyper-- yeah. Hyper trigonometric functions, I think, is the right word. That is, sinhs and coshs instead of sines and cosines. The formulas are very similar.

These are the formulas we get for the open universe case, compared to those formulas for the closed universe case. We get a change in the order of-- instead of theta minus sine theta, we get sinh theta minus theta. But that's what you have to get if it's going to turn out to be positive. Sine theta is always less than theta. So this is a positive quantity. Sinh theta is always greater than theta, so this is a positive quantity.

And same for the second lines. Cosh theta minus 1 is always positive, and one minus cosine theta is always positive. So you really know which order to write them in just by knowing that you want to write down something that's positive.

So these formulas describe the evolution of the open universe exactly the same way as those formulas describe the evolution of a closed, matter-dominated universe.

So any questions? OK. Next step, just repeating what we did for the closed universe, we can derive a formula for the age of an open universe. And again, it's really just a matter of substituting into the formula we already have to be able to re-express it in terms of useful quantities. Which, again, I choose to be the Hubble expansion rate and omega.

And here I've put together all three formulas for the age. The flat universe, the first one we did, where t is just h inverse if we bring it to the other side, times 2/3. And the open universe on the top, and the closed universe on the bottom.

Now one of the, perhaps surprising, things that one finds here is that all three of these expressions look fairly different from each other. And you might think that that would give some kind of a jagged, discontinuous curve. But you can go ahead and plot it, which I did. And there's the plot. It's one nice, smooth curve.

And we won't go into this in detail, but many of you have had courses in complex functions, functions of a complex variable. If you know about functions of a complex variable you can tell that these are, in fact, all the same function. That is, if omega is, say, bigger than 1, this formula can be evaluated straightforwardly. It involves only things like square roots of positive numbers.

But you could also try evaluating this formula for omega bigger than 1. And then you have square roots of negative numbers appearing. But square roots of negative numbers are OK if you know about complex numbers. They're just purely imaginary. And then you get trigonometric functions, and even inverse trigonometric functions, or inverse hyperbolic trig functions of imaginary arguments. But all those are well-defined.

And if you work through what the definitions are, the top line really is identical to the bottom line. Those really are the same function. And that's why one expects that, when you plot them they will join together smoothly, as they clearly do.

The point in the middle here is the first age that we derived for the flat universe where omega is equal to 1 and h t is just equal to 2/3. That is, t is equal to 2/3 h inverse for a flat universe, which is the middle dot. And the closed universes are on the right and open universes are on the left.

OK. Any questions about these age calculations? Yes?

AUDIENCE: I noticed that, for the open solution, there's a case where you get some imaginary numbers?

PROFESSOR: If you use these formulas you don't get any imaginary numbers. But if you tried to use, say, the bottom formula for a value of omega less than 1 it would give you imaginary numbers. And it would, in fact, if you trace through what those imaginary numbers do, it would give you the formula on the top. So it's all consistent. It's most straightforward to use the formula on the top for omega less than 1 and the formula on the bottom for omega greater than 1. And then one never confronts imaginary numbers.

AUDIENCE: OK.

PROFESSOR: Any other questions? OK. Where are we going next?

Finally, just actually one last graph on the evolution of matter-dominated universes, which is the final form of a of t for a matter-dominated universe. If you re-scale things by dividing a over root k by alpha, and c t over alpha-- if you look back at our equations-- let me go back to where the equations were. We're really just graphing these equations.

If I divide this equation by alpha, I just get a pure function of theta with dimensionless variables. And similarly here, if I divide a over root k by alpha I just get 1 minus cosine theta, which is a pure number. So that's what I've chosen to do. And the dimensionality works the same for the open case as well.

So it allows you to draw a plot which is just independent of parameters. All the parameters are absorbed into the way the axes are defined, which are both defined as dimensionless numbers. And in that case the closed universe survives for a duration of 2 pi. The axis here, at least, has the same duration as theta. It's not actually theta, because t is not linear in theta. But this point does change by 2 pi as theta goes from 0 to 2 pi.

And one can see here, the three possible curves. A closed universe which starts and then falls back, a flat universe which goes off to the right and has actually a constant slope as you go out here, and an open universe which behaves slightly differently. Actually, I think I was wrong when I said the flat universe has a constant slope. But the open and the flat case are similar to each other. They both go off to infinity, but in different ways.

And all three of them merge as you go backwards in time. That's not something that might have been obvious before we wrote down the equations. But in very early times all universes look like they're flat universes, if you go to early enough times.

And that actually is an important point which we'll talk about later in terms of what's called the flatness problem of cosmology. That basically is the consequences of that fact.

Yes?

AUDIENCE: Why didn't you just say all of them looks like open universes [INAUDIBLE]?

PROFESSOR: [LAUGHS]

AUDIENCE: I mean, what's special about flat universes?

PROFESSOR: Well, actually, there is something special about flat. Which is that, if we plotted omega as the function of time all of them approach omega equals 1 as time goes to 0. So there is a very definite meaning of saying that they're all approaching flat, rather than they're all approaching open or closed. Good question, though. Because from this graph you really can't tell anything special about the flat.

Any other questions? Yes?

AUDIENCE: So does this mean that the open and flat solutions will extend indefinitely?

PROFESSOR: Yeah. The open and flat solutions extend indefinitely in time. That's right, they do. And one can see that from the formulas or the graph.

Yes?

AUDIENCE: So for plotting omega as a function of time, I'm a little bit confused as to how it changes as-- for example, an open universe expands, or a closed universe-- because it seemed like, from our calculations, that when the universe was expanding, omega was increasing? Is that-- at least for a closed universe?

PROFESSOR: That is true for a closed universe during the expanding phase. For a closed universe during the expanding phase, omega does increase. It starts out as 1 and then it rises to infinity when the universe reaches its maximum size and is about to turn around and go back.

Because it doesn't mean the mass density increases. That's maybe what's confusing you. The actual mass density always decreases as it expands, but the critical density decreases even faster. So the ratio, omega, actually rises for a closed universe as it expands.

For an open universe it's the reverse. For an open universe omega starts out being 1 at early times, as it does for any matter-dominated universe. And as the universe expands, omega gets smaller and smaller for an open universe.

And it follows-- I don't have them on slides here, but we do have in the notes formulas that we derived, that we did on the blackboard, that give us omega as a function of theta. And those formulas, you can just look at them and see how omega behaves as the universe evolves. Because as the universe evolves, theta just increases. And those formulas do trivially show what I said about how omega evolves.

Any other questions? Yes?

AUDIENCE: Why a over alpha root k, but for a flat universe k is 0?

PROFESSOR: That's right. I didn't really say that, but for the flat universe there's an arbitrary normalization that one chooses in drawing this graph. And it was really an arbitrary choice for me to draw the flat case to join on smoothly with the open and closed cases. I could have put any constant in front of t to the 2/3. And you're right, then they would not necessarily mesh unless I chose that constant in the right way.

Yes?

AUDIENCE: Based on that, is there a particular reason you decided to chose [INAUDIBLE] a flat universe looks like this? Is there a particular thing you're trying to show in choosing [INAUDIBLE]?

PROFESSOR: OK. The question is, is there a particular reason why I chose the normalization that I chose? If I did not choose it, it would only differ by an overall factor. So it would still look-- the flat line by itself would look indistinguishable, really. It would just be higher or lower. So the only question is how it meshes with the others. And since the flat case really is the borderline between the other two cases, and since this constant that appears in front of t to the 2/3 has no physical meaning whatever, it seems the sensible thing to do is to plot it so that it looks like the limit of an open or closed universe. Because physically it is the same as the limit of an open or a closed universe coming from either side.

Any other questions? Those are good questions.

OK. In that case, we are finished with the evolution of matter-dominated universes, and ready to start talking about non-euclidean spaces.

So what we'll be doing next is kind of a mini introduction to general relativity, which how non-euclidean spaces enter physics. Now, needless to say, general relativity is an entire course separate from this course. And of course that even has more prerequisites that this course has, so we're not going to duplicate what would be taught in the general relativity class.

But it turns out that the discussion of general relativity does, in fact, divide pretty cleanly into two major issues. And we will be dealing with one of those issues but not the other. In particular, what we will be doing in this class is learning how the formulas of general relativity is used to describe curved spaces. And we will learn how particles move in curved spaces. So we'll be able to analyze trajectories in any curbed space if somebody tells you what the curved space itself looks like.

What we will not be doing is we won't even attempt to describe how general relativity predicts that matter should cause space to curve. That would be left entirely for the GR course that you may or may not take. But it will not be discussed here.

There's only one point where we will need a result of that sort. We'll want to know how the matter in our Friedman Roberson Walker universe affects the curvature. And there I would just give you the result. I'll try to make it sound plausible. But I won't make any pretense being able to drive that result.

That is, we will not be able to drive how much space curves as a consequence of the matter that's in it. But we will write down the answer so you at least know what the answer is for a homogeneous isotropic universe. So here's a picturesque slide about curve space.

Four years ago, I think it was, a postdoc here named Mustafa Amin gave this lecture for me because I was out of town. And he had much more colorful transparencies than I ever do. So I'll be using some of his transparencies here. And this is one of his opening slides.

So this is what we want to talk about-- curved space as illustrated in that nifty picture. So I want to begin with a kind of an historical introduction. To be honest, I'm pretty much following the logic of the first chapter of Steve Weinberg's General Relativity book.

Non-Euclidean geometry of course starts by thinking about Euclidean geometry and then how one might be move away from it. And historically, there's kind of a clear cut path, which was followed. Euclid based his geometry, as described in Euclid's elements, in terms of five postulates.

The first of which is that a straight line segment can be drawn joining any two points. Sounds clear enough. Second, any straight line segment can be extended indefinitely in a straight line. Also sounds obvious, which is what Euclid was banking on.

Third, given a straight line segment, a circle can be drawn having the segment as a radius and one endpoint as the center. That also-- you can imagine yourself doing that-- seems straightforward enough. But then we come to the fifth postulate, which still sounds pretty obvious. But it's certainly much more complicated than the others.

The fifth postulate says that if a straight line falling on two straight lines makes the interior angle on the same side are less than two right angles, the two straight lines-- if produced indefinitely-- meet on that-- I think this is mis-typed. Mustafa typed it. I guess he's typed this one, too. This one shows the picture.

Yeah, this should be on that side in which the angles are less than two right angles. So let me just explain it, independent of the text. The question is, what happens if you have one line-- the line that's shown more or less vertical here-- and two lines that cross it such that the interior angles-- here shown as a and b-- are on one side less than pi. Less than two right angles is the way you could describe it.

Then, as you can see from the picture, these lines will meet on this side and will not meet on that side. And that's what the postulate says that under those circumstances where two lines cross a given line such that the sum of the two angles adds up to less than pi on one side that the lines will be on that side and will not be on the other side. Yes?

AUDIENCE: What was his motivation for making this the fifth postulate? It seems kind of arbitrary.

PROFESSOR: OK, the question is what was Euclid's motivation for making this the fifth postulate? Well, I have to admit I haven't had many conversations with Euclid so I'm not sure I know the answer to that question. But it was what was needed to basically complete geometry as we know it. So much of what you've learned in geometry would not be there if there was not something equivalent to this postulate.

But actually what I'm going to be talking about next is there has been discovered there are a number of substitutes for this fifth postulate. Mathematicians studied for a long time whether or not this postulate could be derived from the others since it seems so much more complicated than the others.

And there was a strong desire during thousands of years really-- at least over 1,000 years-- among mathematicians to try to prove the fifth postulate from the first four postulates. And nobody ever succeeded in doing that. And we now are pretty clear that it's not possible to do that. That the postulate is independent of the other postulates.

It was discovered along the way that there are a number of equivalent statements to fifth postulate. And you could equally well have chosen any one of these four statements that are illustrated in these pictures. And I'll give you words one by one for what these alternative versions of the fifth postulate would be.

A, up here, illustrated there, says that if a straight line intersects one of two parallels, meaning two parallel lines-- so this is the one line intersecting two parallel lines. If it intersects one of them as the heavy part of the line shows, then the theorem says that if you continue that line it will always intersect the other. And certainly obvious from the picture, that's the way it works. But that's equivalent to the fifth postulate and not provable from the other four postulates.

A second statement-- b is the one that's illustrated there-- is the one that I remember learning when I was in high school, I think, which says that if you have one line and another line parallel to it-- or rather, I'm sorry another point off the line that there's one and only one line through that point parallel to the original line. So that is yet another statement of this famous fifth postulate.

Number c is less obvious. But it turns out that once you go away from Euclidean geometry, your space always has a built in scale. So things are not scalable. One example I might mention at this point of occurred space is, say, the surface of a sphere. And the service of a sphere has some fixed size.

So if you have a figure of one size, and you wanted, on the surface of this sphere, to make a figure 5 times bigger, it might not fit on the sphere anymore. So you can't always make a figure bigger on the surface of the sphere. In fact, you could never make a figure bigger without bending in some way on the surface of the sphere.

So that gives rise to this third statement of the fifth postulate, which is that, given any figure, there exists a figure similar to it of any size. And by similar it means that for polygons they're similar if the corresponding angles are equal to each other as they're supposed to be on those two images. And the corresponding sides are proportional to each other.

So a similar figure is just a blow up-- a rescaling-- of the original figure. And you can only do it if the fifth postulate holds. You can do it on a flat space but not on a curved space. And I think that is a less obvious version of the fifth postulate.

And finally, the fifth postulate is linked to the behavior of triangles. We all learned in

Euclidean geometry that the sum of the three angles of a triangle is 180 degrees. That is a crucial theorem of Euclidean geometry that depends directly on the fifth postulate and is, in fact, equivalent to the fifth postulate.

So you could assume it and forget the fifth postulate and still prove everything. So the fact that-- actually, I'm sorry. It is equivalent to the fifth postulate. But you don't need to assume that it's true for every triangle to prove the fifth postulate. It turns out that sufficient and mathematicians always look for the minimum axiom to be able to prove what they want to prove.

The minimum version of the axiom for triangles is to simply assume that there's just one triangle who's angles add up 180 degrees. And if there exist just one triangle whose angles add up to 180 degrees, then the fifth postulate has to hold-- turns out-- which is not so obvious, again. So these are all different, equivalent ways of staying in the fifth postulate.

But it's pretty clear now that it's not possible to prove the fifth postulate from the first four. Any questions about that? OK, so a little bit of history now.

The first person, apparently, to seriously explore the possibility that the fifth postulate might be wrong was a Jesuit priest named Giovanni Geralamo Saccheri. I'm sure other people can pronounce it much better than I can. And in 1733, which is the same year he died, he published a study of what geometry would be like if the fifth postulate did not hold. And he titled it- this is a lot title, which I don't know how to pronounce really, but in English it's apparently translated as Euclid Freed of Every Flaw, treating the fifth postulate as kind of a flaw in Euclid's axiomization of geometry.

Saccheri was, in fact, convinced that the fifth postulate was true. He didn't really want to consider the possibility that it was false. But he was nonetheless exploring the possibility that it was false because he understood the concept of a proof by contradiction.

He was looking for a contradiction to be able to prove that mathematics would not be consistent if you assume the fifth postulate was false. And therefore, the fifth postulate would be proven to be true.

So he was exploring what would happen if the fifth postulate was false, looking all the time to find some inconsistency, and was not able to find any. So he considered all of this a failure. But he published it anyway. And that's the publication front page.

The next person to enter the stage-- or actually three people together, but I only have a nice transparency on Gauss. Gauss, Bolyai, and Lobachevsky went on to seriously explore the possibility of geometry without the fifth postulate, actually assuming that the fifth postulate is false, developing what we call GBL, Gauss-Bolyai-Lobachevsky geometry.

Gauss was a German mathematician. These were the years he lived. He, in fact, was born the son of a poor, working class parents, which I found a little surprising. We kind of think of scholars in those early years as being gentleman who were part of the nobility. But Gauss was not but, nonetheless, went on to be one of the most important mathematicians of his age.

One of the other things that surprised me and to be honest I just learned all this from Wikipedia. I'm no real expert on the history. But they gave a list of Gauss' students. And they included that Richard Dedekind, Bernhard Riemann, Peter Gustav Lejeune Dirichlet, which is the name I remember, Kirchhoff, and Mobius. So quite a list of famous mathematicians.

So I have to admit, when I read that, I was just about to send off an email to all of my former students saying, look, what's happening here? You're not competing at all. [LAUGHING] But I decided not to do that. It would be impolite. And who knows? Maybe 100 years from now my students will be as famous as these guys. You never know. We can plan, I hope.

OK, so the other guys involved in this and they were all working independently were Bolyai who was, I think, a Prussian military man, actually, and Lobachevsky who was also a professional mathematician working in the university. The three of them independently developed a geometry in which the fifth postulate was assumed to be false.

There are two ways it could be false. In version with the triangles, for example, a triangle could have more than or less than 180 degrees.

Since there were assuming that the fifth postulate was false, it meant they had to be assuming that every version that we just talked about of the fifth postulate is false because they are all equivalent to each other and these people realize that. So in particular in the Gauss-Bolyai-Lobachevsky geometry, there are infinitely many lines parallel to a given line.

And no figures of different size are similar. And the sum of the angles of a triangle are always less than 180 degrees or pi, depending on whether you're a radian person or a degree person.

Now I should mention here that the surface of the sphere is, in fact, a perfectly good example of the non-Euclidean geometry. But for some reason it was not taken seriously by mathematicians until long after these guys were doing their work.

And part of the reason, I guess, is that the surfaces three evaluates not just one of Euclid's axioms but two if we go back to Euclid's axioms. The second of Euclid's axioms said that any straight line segment can be extended indefinitely in a straight line. And the surface of the sphere and the analog of the straight line is a great circle.

And if you extend the great circle, it comes back on itself. So the surface of the sphere violates the fifth postulate. But it also violates the second postulate. But still perfectly consistent geometry, and it is a non-Euclidean geometry.

And on the surface of this sphere, these statements all kind of reverse. Instead of having infinitely many lines parallel to a given line, you have no lines that are parallel to a given line. Remember, lines are great circles and lines are parallel if they never meet. But any two great circles meet. So there are no lines parallel to a given line in the geometry of the surface of the sphere.

It's, again, true that no figures of different size are similar. That has to be true for any any geometry where the fifth postulate was false. The last one, again, has a choice. And it's the opposite choice for a sphere as for the Gauss-Bolyai-Lobachevsky geometry.

In the Gauss-Bolyai-Lobachevsky Bolyai- geometry, the sum of the angles is always less than 180 degrees. If you picture a triangle and a sphere, you can imagine that the edges look like they bulge. And the sum of the angles on the surface of the sphere of a triangle are always greater than 180 degrees.

The easiest way to convince yourself that that's true in at least one important case is to imagine a triangle. Here's a sphere. Everybody see the sphere? There's the North Pole. There's an equator. Imagine a triangle where one vertex is at the North Pole and the liver disease or on the equator and the triangle looks like this.

And these are both 90 degree angles down here. So you already have 180 degrees. And any angle you have on top just adds to the 180 degrees putting you above the Euclidean value of 180 degrees. So for a sphere it's always the opposite of this greater than 180 degrees.

OK, continuing with Gauss, Bolyai and Lobachevsky, their work was based on exploring the axioms of Euclid, assuming the reverse of the fifth postulate in any one of its forms. And they proved theorems and wrote, sort of, their own versions of Euclid's elements.

But that still left open the question whether or not all of this was really consistent. That is, from the thinking of Saccheri that we already talked about, there's always the chance that you might find the contradiction even if you haven't found a contradiction yet.

And Gauss, Bolyai and Lobachevsky didn't really have any way of answering that. They didn't really have any way of knowing whether if they continued further they might find some contradiction. So what they did certainly extracted the properties of this non-Euclidean geometry. But it did not really prove that the non-Euclidean geometry was consistent.

That happened slightly later in an argument by Felix Klein, who was the same Klein as the Klein bottle, by the way. And that was his famous paper on this was published in 1870 somewhat later than the early work that we're talking about.

And what he did is he gave an actual construction of the GBL geometry. And by construction I mean in terms of coordinates using coordinate geometry ideas, which were originally developed by Descartes. That's why we call them Cartesian coordinate systems.

So what he gave as a description of the Gauss-Bolyai-Lobachevsky geometry was a space that consisted of a disk with coordinates x and y just as Descartes would have done. He restricts himself to the inside of the disk. So he restricts himself to x squared plus y squared less than 1. And what he gave was a function of two points in this disk where the function represents the distance between those two points.

He decided that distances don't have to be Euclidean distances if we're trying to explore non-Euclidean geometry. So he invented his own distance function. And it's pretty complicated looking. The function he came up with was that the cosh of the distance between points 1 and 2 divided by sum number a-- and a could be any number-- is equal to 1 minus x1 times x2-- these are the two x-coordinates for points 1 and 2-- minus y1 y2 over the square root of 1 minus x1 squared minus y1 squared.

Sorry, it's the product of two square roots in the denominator here. And the second square root is the same thing for the two coordinates 1 minus x2 squared minus y2 squared. So this formula is certainly not obvious to anybody. But Klein figured out that this actually does-- the geometry described by these formulas-- reproduce completely the postulates of the Gauss-Bolyai-Lobachevsky geometry, including the failure of Euclid's fifth postulate.

And since this boils down to just algebra, if algebra is consistent, it proves that the Gauss-Bolyai-Lobachevsky geometry is consistent. Now, as I understand it, nobody can prove that any of this actually is consistent. People prove relative consistency. So in the assumption that algebra is consistent, theorems about the real numbers are consistent. Felix Klein was able to prove that the Gauss-Bolyai-Lobachevsky geometry is consistent.

And this was really the beginning of coordinate geometry. I'm sure all of you are rather familiar with coordinate geometry. It's a standard topic now in math courses and even in high school.

And this is really where it began. Euclid had no idea that it was any value in trying to represent geometric quantities as equations. Euclid did everything in terms of theorems. But this opened up a whole new door for how to discuss geometry.

So the geometry is a slide that just shows the same formula. And I guess this is supposed to be an image of the disk. One thing I should point out, which I forgot to point out, which is that although this disk looks finite, it really is an infinite space that's being described. And one can see that by looking carefully at the distance functions.

As either one of these two points-- point 1 or point 2-- approaches the boundary x square plus y square equals 1, this square root denominator blows up. So the distance between a point and another point that's approaching the boundary goes to infinity as that point approaches the boundary. So boundary's actually infinitely far away even though in coordinates they are still x squared plus y squared equals 1.

So this introduces another important concept, which we'll be using in general relativity, which is the coordinates don't represent distances. Distances could be very different from the way the coordinates look. So that boundary, even though it looks like it's right there on the blackboard, is actually very far away from the other points.

OK, so after Klein, the important new idea that Klein introduced was, first of all, the explicit construction but also the general idea that you can describe geometry not by giving postulates but instead by actually doing a construction where you've given names to the points in terms of coordinates and you describe what happens geometrically in terms of some distance function which describes the distance between two arbitrary points.

And Gauss went on to make two other very important observations about geometry, which has become essential to our current understanding of non-Euclidean geometry. So let me mention two other ideas that Gauss introduced.

The first one was the distinction between what he called inner and outer properties of a curved surface. His curved spaces were all two dimensional, so they were surfaces. So this is most easily described for say a surface of a sphere where we can all visualize what we're talking about. The surface of a sphere we visualize in three Euclidean dimensions and we think of its properties as being determined by that three dimensional space.

And the geometry of that three dimensional space. And of course the three dimensional space is Euclidean that we're embedding our sphere into. But the non-Euclidean aspects are all seen in the geometry of the two dimensional surface. Figures drawn on the surface as if the rest of the three dimensional space did not exist. And that is this key idea of inner versus outer properties. The outer properties of the sphere are properties that relate to the three dimensional space in which the sphere is embedded.

But what Gauss realized is that there's a perfectly well defined mathematics contained entirely in the two dimensional space of the surface of the sphere. You could discuss it without making any reference to the three dimensional space around it if you wanted to, it's just a little more complicated to be able to do that. But we will in fact be doing it explicitly very shortly. And all it amounts to from our point of view is assigning coordinates to the points on the surface of the sphere and the distance function for those coordinates.

And then one has a full description of this two dimensional space consisting of the surface of the sphere which no longer needs make any reference to the third dimension that you imagined the sphere embedded in. So the study of the properties of that two dimensional space is the study of the inner properties of the space. And if you care about how it's embedded, that's called the outer properties.

And Gauss made it clear in the way he described things that from his point of view, the real thing that mathematicians should be doing is studying the inner properties. The outer properties are not that interesting. So that's one key idea that Gauss introduced. And the second is the idea of what we now call a metric.

And there are really two pieces to this. The first of which is that instead of giving macroscopic distances, which is what Klein did, he told you how to write down a formula for the distance between any two points, Gauss realized that things could become more interesting if instead of trying to immediately write down a function for the distance between two arbitrary points, you can find your attention to very nearby points. And consider the distance between two far away points as just the length of a line that joins them where every little segment in the line is a short distance which is governed by the rules that you've made up for short distances.

So if you understand short distances, long distances are obtained just by adding. That's the basic idea. So only short distances are needed. And then there's another important idea which is that the short distances themselves for an interesting class of spaces should not be some arbitrary function of the coordinates of the two points but should in fact be a quadratic function. And in two dimensions, a quadratic function mean something like this, where gxx is just any function of x and y, x and y are the two coordinates of the space.

gxy is also just any function of x and u. gyy is any function and gxx multiplies the x squared, xy multiplies the x times dy and g sub yy multiplies dy squared. And there's no terms portion for dx by itself or dy by itself or dz by itself. And there's no terms proportional to the cubes of those quantities, it's all quadratic. That's the assumption. Now what underlies that assumption is not that all spaces have to have this property.

This does narrow one down to a particular class of spaces. But one way of characterizing that class is that what Gauss understood is that that class, the class of spaces described by a metric like this, are precisely the class of spaces which are locally Euclidean, meaning that even though the surface is curved, any tiny little patch of it looks flat and can be covered by Euclidean coordinates, redefining coordinates in that one little patch only where the metric in the one little patch would just be the Euclidean metric given by the Pythagorean theorem of Euclidean geometry, dx prime squared plus dy squared, which would be the distance function for a flat space and ordinary Cartesian coordinates.

And it turns out that saying that this is true in every microscopic region is equivalent to saying that the metric over the entire space can be written as a quadratic form, meaning one that looks exactly like that. And spaces that have this property are now usually called Riemannian spaces, even though it was Gauss who first identified them. And all the spaces that we deal with in physics, in particular the spaces we deal with general relativity, will be Riemannian spaces.

Or sometimes they're called pseudo Riemannian spaces because time is treated a little bit differently in physics. And the word "Riemannian" was really built on spatial geometry. But the word "pseudo" only changes the fact that this becomes the Lorenz metric, if you know what that means, instead of the Euclidean metric. But the same idea holds, which is that the spaces that we're interested in are spaces which locally look exactly like flat space.

And that implies that globally, you can always write down a metric function which is quadratic, as Gauss said So the metric should be a quadratic function that specifies the distance only between infinitesimally separated points, not finitely separated points and should have the form of ds squared is equal to some function of x and y times dx squared plus a different function of x and y times dx times dy plus a different function of x and y times dy squared. So I'm just writing the same form that was on the side.

This is a very important formula. Incidentally, the two that appears here is only because when we write this in a different notation, this term will occur twice, once as dx times dy, and once as dy times dx. And the two times it occurs are equal to each other, so here they're just collected together with a factor of two in front. Yes.

STUDENT: Just to make sure, those three different functions you wrote, those subscripts aren't supposed to mean partial derivatives, right?

PROFESSOR: No, that's right. The subscripts and this expression only mean that this multiplies dx squared, so it has subscripts xx. And this multiplies dx times dy, so it has subscripts xy. That's what the subscripts mean and nothing more. So the subscripts only mean that they're the things that appear in that equation. Yes.

STUDENT: It seems like the metric is giving us distance in terms of an infintesimal displacement, but then a locally Euclidean space is already tangent infintesimally, so how are we relating the local metric with the global metric?

PROFESSOR: OK, the question is how do we relate the local metric which I say is Euclidean to the global metric? And the answer I think for now I will stick to just giving kind of a pictorial answer based on the picture here. That is once you know the distances between any two points in a tiny little patch here, it's then always possible to construct coordinates, here called x prime and y prime, such that the distance between any two points as calculated from the original metric, which is the one here, is exactly the same as the distance you get using this metric.

And the claim is that you can always define coordinates x prime and y prime which make that true. That claim is not absolutely obvious. But it's something you can probably visualize if you just imagine that every little tiny piece of this curved surface looks like it was just a flat surface and then a flat surface you know that you can write down a Cartesian coordinate system, which will have the Euclidean metric. But it's only an intuitive statement, proving it is actually harder. Yes.

STUDENT: [INAUDIBLE] bottom formula [INAUDIBLE] with positive curvature if we analogize to second derivative.

PROFESSOR: I'm sorry, say that again.

STUDENT: If we analogize gxx, gyy, in the bottom formula there, [INAUDIBLE] positive curvature [INAUDIBLE] second derivative--

PROFESSOR: Yeah OK. So you're asking does this mean that we have two states of positive curvature?

STUDENT: Bottom right.

PROFESSOR: Bottom right, oh this. These are Mustafa's slides. I forgot to say that. You can tell from the style, these are not my slides, these are Mustafa's slides. And I don't know what he meant by this. You're right, this does-- well you do want the metric to be positive definite, which is not the same as saying the curvature is positive. And I think this might just be the condition that the metric is positive definite, that this expression will always be positive.

Yeah. I'll bet that's what it is, I don't know for sure. I'll bet that's what that condition is about. And you do want that, the metric had better be positive definite for spatial geometries. In fact, in general relativity, where it's a space time metric, it will not be positive definite. For reasons that we'll talk about later. But for geometry, the metric should be positive definite. All distances should be positive. OK. So that ends my slides. So now I'll continue on the blackboard. OK next I wanted to say a few general comments about general relativity.

General relativity was of course invented by the Einstein in 1916. It's a theory that he was working on for about 10 years after he invented the theory of special relativity. To understand what's going on there, you want to recognize that special relativity is a theory of mechanics and electrodynamics which was designed to be consistent with the principle that the speed of light is always the same speed of light independent of the speed of the source of light or the speed of the observer of the light.

And it's of course not easy to do that, because you think that if you move relative to something else that's moving, that you would see its velocity change. So in order to make a theory that where the speed of light was an absolute invariant, Einstein had introduced a number of things. And we talked about those at the beginning of the course, the three primary effects built in to special relativity. Time dilation, length contraction, and new rules about simultaneity, and how things that look simultaneous to one observer will not look simultaneous to other observers in a very definite, well defined way.

So by inventing these rules, Einstein was able to devise a system which was consistent with the idea that the speed of light always looked the same to all observers. And at similar times, the Michelson Morley experiment seemed to show that that was in fact the case and ultimately there's a tremendous amount of evidence verifying that what the hypothesis that Einstein was pursuing is the right one for the way nature behaves. The speed of light is invariant.

But what was lacking in Einstein's formulation of special relativity was any version of a consistent theory of gravity. Gravity had a well defined description known since Newton. But Newton's description was a description of a force at a distance. And that is intrinsically inconsistent with a theory like special relativity, which holds that simultaneity is itself relative. In Newtonian physics, the force of gravity is equal to Newton's constant times the product of the two masses that are interacting with each other divided by the square of the distance and then times some union vector pointing along the line that joins the two particles.

But for that formula to make any sense at all, you have to know where the two particles are at the same instant of time. And this r is the distance between the locations of the two particles at some instant of time. And this r hat is a unit vector that points along the line joining those two particles where you've pinned down where those particles are at one instant of time.

But we know from the very beginning that in special relativity the notion of two things happening at distant points at one instant of time is ambiguous. Different observers will see different notions of what it means for two events to happen at the same time across this distance. And that means there's no way to make sense out of Newton's Law of gravity in the context of special relativity. You can't modify it by just changing the way the force depends on distance. You really have to change the variables that it depends on from the beginning. Yes.

STUDENT: Is that equivalent to saying the other particle can't know how much mass of one is [INAUDIBLE] limit of the information of the speed of light?

PROFESSOR: Is that equivalent to saying that one particle can't know what the mass of the other particle is because of limitation. Yeah it's the distance, it's not the mass. Because the mass is preserved in time. So one particle could've measured the mass of the other particle at an earlier time and it would be reasonable to infer, given the laws of physics, that we know that would stay the same. But one particle has no way of where the other particle is at the same time.

And not only does the particle not have any way of knowing, but even an external observer can't know. Because different external observers will have different definitions of what it means to be at the same time. So there really is no way that this could work. Now I should mention that it is still possible to have action at a distance theories which are consistent with special relativity. And in fact, Maxwell's equations can be reformulated that way.

The easiest way to make things consistent with special relativity is to describe interactions in terms of fields. And that is really what Einstein did originally in special relativity. He was thinking about light, he was thinking about Maxwell's equations and he was thinking very explicitly about Maxwell's equations. And in the Maxwell description of electromagnetism, particles at a distance don't directly interact with each other. But rather, each particle interacts with the fields around it and that is a local interaction.

A particle interacts directly only with the fields at the same point. But then those fields obey wave equations that can propagate information. And the fact that you have an electron here can create a field which then exerts a force on an electron there. And the force on the electron here depends only on the electric fields here or the magnetic fields as well.

The particle is moving, but everything is completely local and the description of electromagnetism as given in the form of Maxwell's equations. And Maxwell's equations are completely relativistically invariant. And that was part of Einstein's-- it was really the key part of Einstein's motivation in constructing the theory of special relativity in the first place to make Maxwell's equations invariant. That they held in every frame. It is still possible though and worth recognizing that it's possible to reformulate electromagnetism as an action at a distance theory.

And it is in fact described that way in volume one of the Feynman lectures for those of you who've looked at the Feynman lectures. In order to make that work, you have to complicate things in a pretty significant way. So I'm going to draw here just a space time diagram. X and CT, CT going up that way, x going that way. So in this diagram the speed of light would be a 45 degree line. And let's suppose we have two particles traveling in this space.

A particle that I will call a and a particle that I will call b. Being very original with these names. If we wanted to know the force on particle a at a certain time t indicated by this dotted line, I guess I'll label t to make it as clear as possible. Feynman gives us a formula where we can determine it solely in terms of the motion of particle B without talking about fields at all. But the formula does not depend on where b is at the same time.

And it could not if this is going to be a relativistic description. But instead, the way this action and the distance formulation works is when imagine drawing a 45 degree line backwards, meaning a line that light could travel on, and one sees where that intersects the trajectory of particle B. And that time is t prime.

And the word that's used for that symbol is retarded time. It's an earlier time. It's exactly that time which has the property that if particle b emitted a light beam at that time, it would be arriving at particle a at just the time t that we're interested in the time when we're trying to calculate the force on particle a. And what Feynman gives you in volume one if you look at it is a very complicated formula that determines the force on particle a in terms of not only the position of this particle at time t prime but also its velocity and even its acceleration.

But if you do know the position, the velocity, and the acceleration of this particle, and of course the velocity of particle a, you can determine the force on a. Not obvious, but it's true. But that's certainly not the easiest way to formulate electromagnetism. And that's not the way most of us have learned, unless you've learned by starting by reading volume one of Feynman. But most of us learn Maxwell's equations as differential equations.

Where information is propagated by the field from one point to another. In the case of general relativity, one has the same problem. How can you describe something which-- and the only approximate you initially know is an action and the distance, how can you describe it in a way that's consistent with relativity? And the idea that simultaneity is not a well defined concept. So that was the problem was Einstein was wrestling with for 10 years, how to build a theory of gravity that would be consistent with the basic principles of special relativity.

And the result of those 10 years of cogitating is the theory that we call general relativity. And it's essentially a theory which describes gravity as a field theory similar to Maxwell's field theory where all interactions are local. Nothing interacts at a distance, but particles interact with fields at the same point, the fields can propagate information by obeying wave equations, and the fields then at a distant point can exert forces on other particles.

But in the case of general relativity, what Einstein concluded was that the fields that were relevant, the fields that described gravity, were in fact the metric of space and time. So general relativity is the field theory of the metric of space and time. And gravity is described solely as a distortion of space and time. And that's what general relativity is about and that's what we will be learning more about. Now as I said earlier, now I can say it perhaps more explicitly, what we will be learning about is how to describe the curvature of space time as general relativity describes it.

We will learn how that curvature affects things like the motions of particles. But we will not in this course learn how the presence of particles and masses affects the curvature of space time. That you'll have to take it in a relativity course to learn. OK, I think that's where we'll be stopping today. Just doesn't pay to start a new topic with a minute and a half left. But let me just ask if there are any questions. OK, well class over, I will see you folks in a week, because there's no class next Tuesday.