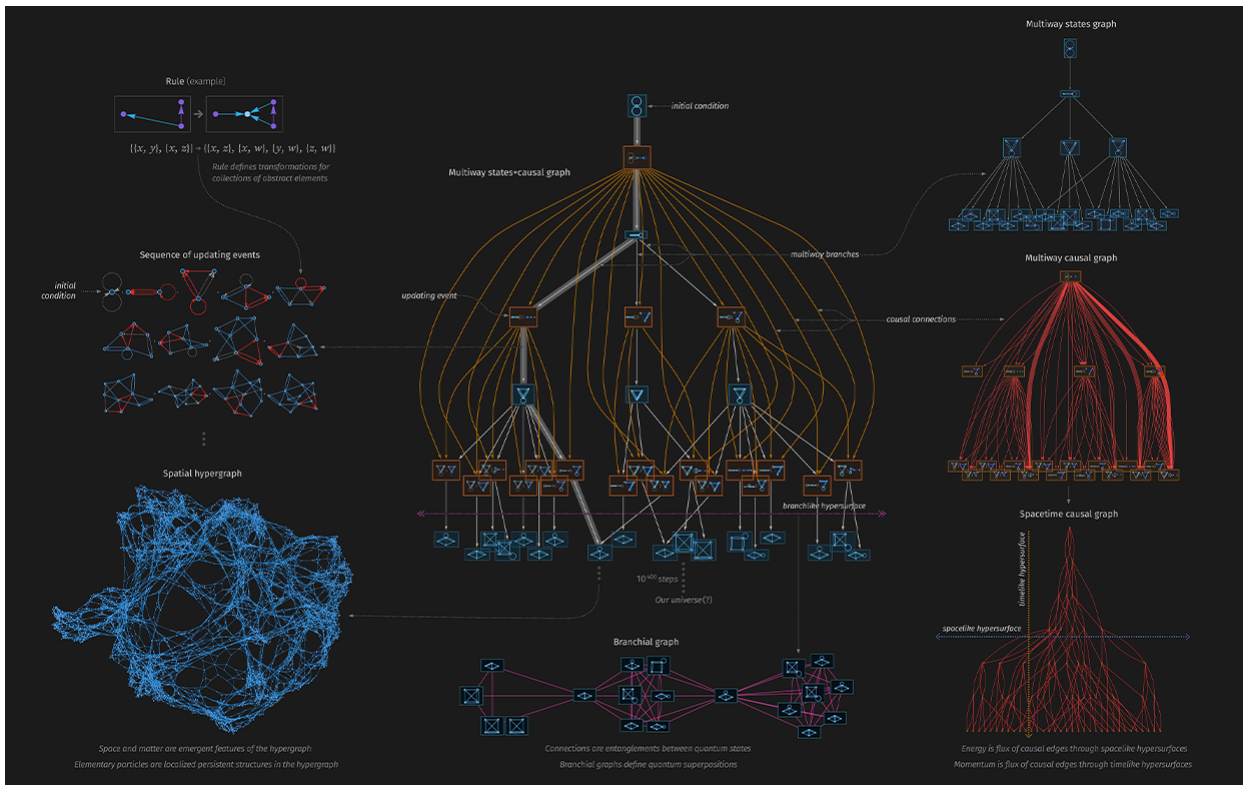≡ S ̲TEPHEN̲ W ̲OLFRAM̲ │ *Writings*

Finally We May Have a Path to the Fundamental Theory of Physics… and It's Beautiful

April 14, 2020

Website: **Wolfram Physics Project**

Technical Intro: **A Class of Models with the Potential to Represent Fundamental Physics**

How We Got Here: **The Backstory of the Wolfram Physics Project**

I Never Expected This

It's unexpected, surprising—and for me incredibly exciting. To be fair, at some level I've been working towards this for nearly 50 years. But it's just in the last few months that it's finally come together. And it's much more wonderful, and beautiful, than I'd ever imagined.

In many ways it's the ultimate question in natural science: How does our universe work? Is there a fundamental theory? An incredible amount has been figured out about physics over the past few hundred years. But even with everything that's been done—and it's very impressive—we still, after all this time, don't have a truly fundamental theory of physics.

Back when I used do theoretical physics for a living, I must admit I didn't think much about trying to find a fundamental theory; I was more concerned about what we could figure out based on the theories we had. And somehow I think I imagined that if there was a fundamental theory, it would inevitably be very complicated.

But in the early 1980s, when I started studying the computational universe of simple programs I made what was for me a very surprising and important discovery: that even when the underlying rules for a system are extremely simple, the behavior of the system as a whole can be essentially arbitrarily rich and complex.

And this got me thinking: Could the universe work this way? Could it in fact be that underneath all of this richness and complexity we see in physics there are just simple rules? I soon realized that if that was going to be the case, we'd in effect have to go underneath space and time and basically everything we know. Our rules would have to operate at some lower level, and all of physics would just have to emerge.

By the early 1990s I had a definite idea about how the rules might work, and by the end of the 1990s I had figured out quite a bit about their implications for space, time, gravity and other things in physics—and, basically as an example of what one might be able to do with science based on studying the computational universe, I devoted nearly 100 pages to this in my book *A New Kind of Science*.

I always wanted to mount a big project to take my ideas further. I tried to start around 2004. But pretty soon I got swept up in building Wolfram|Alpha, and the Wolfram Language and everything around it. From time to time I would see physicist friends of mine, and I'd talk about my physics project. There'd be polite interest, but basically the feeling was that finding a fundamental theory of physics was just too hard, and only kooks would attempt it.

It didn't help that there was something that bothered me about my ideas. The particular way I'd set up my rules seemed a little too inflexible, too contrived. In my life as a computational language designer I was constantly thinking about abstract systems of rules.

And every so often I'd wonder if they might be relevant for physics. But I never got anywhere. Until, suddenly, in the fall of 2018, I had a little idea.

It was in some ways simple and obvious, if very abstract. But what was most important about it to me was that it was so elegant and minimal. Finally I had something that felt right to me as a serious possibility for how physics might work. But wonderful things were happening with the Wolfram Language, and I was busy thinking about all the implications of finally having a full-scale computational language.

But then, at our annual Summer School in 2019, there were two young physicists (Jonathan Gorard and Max Piskunov) who were like, "You just have to pursue this!" Physics had been my great passion when I was young, and in August 2019 I had a big birthday and realized that, yes, after all these years I really should see if I can make something work.

So—along with the two young physicists who'd encouraged me—I began in earnest in October 2019. It helped that—after a lifetime of developing them—we now had great computational tools. And it wasn't long before we started finding what I might call "very interesting things". We reproduced, more elegantly, what I had done in the 1990s. And from tiny, structureless rules out were coming space, time, relativity, gravity and hints of quantum mechanics.

We were doing zillions of computer experiments, building intuition. And gradually things were becoming clearer. We started understanding how quantum mechanics works. Then we realized what energy is. We found an outline derivation of my late friend and mentor Richard Feynman's path integral. We started seeing some deep structural connections between relativity and quantum mechanics. Everything just started falling into place. All those things I'd known about in physics for nearly 50 years—and finally we had a way to see not just what was true, but why.

I hadn't ever imagined anything like this would happen. I expected that we'd start exploring simple rules and gradually, if we were lucky, we'd get hints here or there about connections to physics. I thought maybe we'd be able to have a possible model for the first 10^{-100} seconds of the universe, but we'd spend years trying to see whether it might actually connect to the physics we see today.

In the end, if we're going to have a complete fundamental theory of physics, we're going to have to find the specific rule for our universe. And I don't know how hard that's going to be.

I don't know if it's going to take a month, a year, a decade or a century. A few months ago I would also have said that I don't even know if we've got the right framework for finding it.

But I wouldn't say that anymore. Too much has worked. Too many things have fallen into place. We don't know if the precise details of how our rules are set up are correct, or how simple or not the final rules may be. But at this point I am certain that the basic framework we have is telling us fundamentally how physics works.

It's always a test for scientific models to compare how much you put in with how much you get out. And I've never seen anything that comes close. What we put in is about as tiny as it could be. But what we're getting out are huge chunks of the most sophisticated things that are known about physics. And what's most amazing to me is that at least so far we've not run across a single thing where we've had to say "oh, to explain that we have to add something to our model". Sometimes it's not easy to see how things work, but so far it's always just been a question of understanding what the model already says, not adding something new.

At the lowest level, the rules we've got are about as minimal as anything could be. (Amusingly, their basic structure can be expressed in a fraction of a line of symbolic Wolfram Language code.) And in their raw form, they don't really engage with all the rich ideas and structure that exist, for example, in mathematics. But as soon as we start looking at the consequences of the rules when they're applied zillions of times, it becomes clear that they're very elegantly connected to a lot of wonderful recent mathematics.

There's something similar with physics, too. The basic structure of our models seems alien and bizarrely different from almost everything that's been done in physics for at least the past century or so. But as we've gotten further in investigating our models something amazing has happened: we've found that not just one, but many of the popular theoretical frameworks that have been pursued in physics in the past few decades are actually directly relevant to our models.

I was worried this was going to be one of those "you've got to throw out the old" advances in science. It's not. Yes, the underlying structure of our models is different. Yes, the initial approach and methods are different. And, yes, a bunch of new ideas are needed. But to make everything work we're going to have to build on a lot of what my physicist friends have been working so hard on for the past few decades.

And then there'll be the physics experiments. If you'd asked me even a couple of months ago when we'd get anything experimentally testable from our models I would have said it was far away. And that it probably wouldn't happen until we'd pretty much found the final rule. But it looks like I was wrong. And in fact we've already got some good hints of bizarre new things that might be out there to look for.

OK, so what do we need to do now? I'm thrilled to say that I think we've found a path to the fundamental theory of physics. We've built a paradigm and a framework (and, yes, we've built lots of good, practical, computational tools too). But now we need to finish the job. We need to work through a lot of complicated computation, mathematics and physics. And see if we can finally deliver the answer to how our universe fundamentally works.

It's an exciting moment, and I want to share it. I'm looking forward to being deeply involved. But this isn't just a project for me or our small team. This is a project for the world. It's going to be a great achievement when it's done. And I'd like to see it shared as widely as possible. Yes, a lot of what has to be done requires top-of-the-line physics and math knowledge. But I want to expose everything as broadly as possible, so everyone can be involved in—and I hope inspired by—what I think is going to be a great and historic intellectual adventure.

Today we're officially launching our Physics Project. From here on, we'll be livestreaming what we're doing—sharing whatever we discover in real time with the world. (We'll also soon be releasing more than 400 hours of video that we've already accumulated.) I'm posting all my working materials going back to the 1990s, and we're releasing all our software tools. We'll be putting out bulletins about progress, and there'll be educational programs around the project.

Oh, yes, and we're putting up a Registry of Notable Universes. It's already populated with nearly a thousand rules. I don't think any of the ones in there yet are our own universe— though I'm not completely sure. But sometime—I hope soon—there might just be a rule entered in the Registry that has all the right properties, and that we'll slowly discover that, yes, this is it—our universe finally decoded.
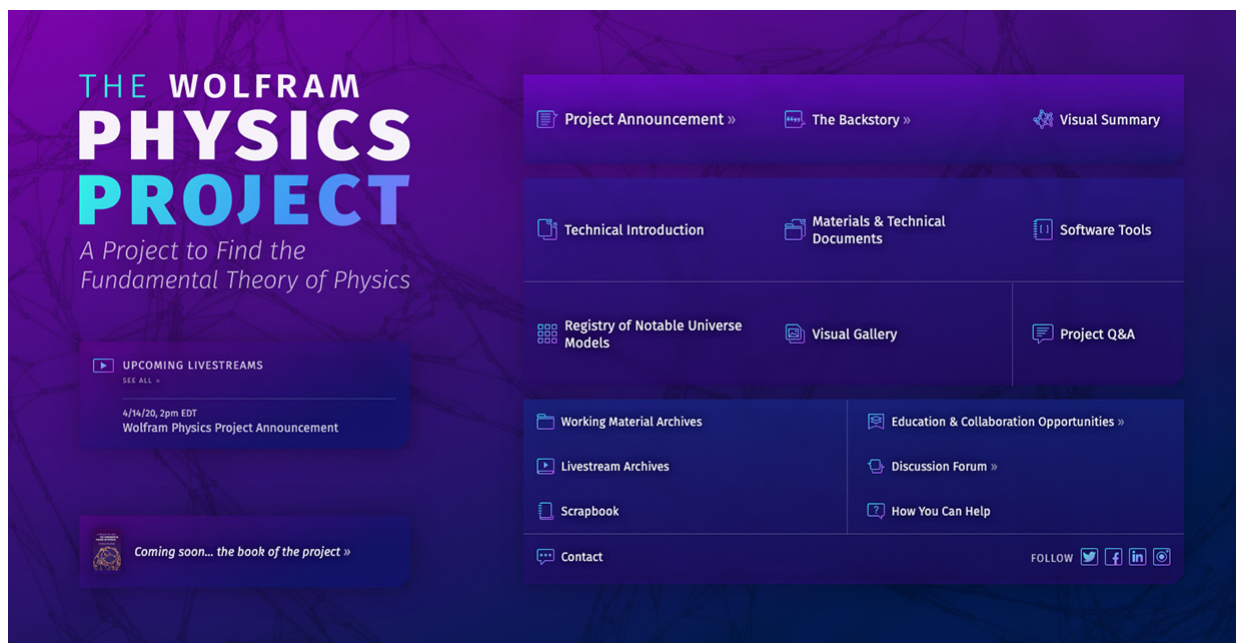
How It Works

OK, so how does it all work? I've written a 448-page technical exposition (yes, I've been busy the past few months!). Another member of our team (Jonathan Gorard) has written two 60-page technical papers. And there's other material available at the project website. But here I'm going to give a fairly non-technical summary of some of the high points.

It all begins with something very simple and very structureless. We can think of it as a collection of abstract relations between abstract elements. Or we can think of it as a hypergraph—or, in simple cases, a graph.

We might have a collection of relations like

{{1, 2}, {2, 3}, {3, 4}, {2, 4}}

that can be represented by a graph like
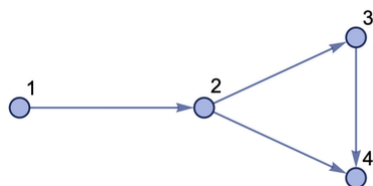
All we're specifying here are the relations between elements (like {2,3}). The order in which we state the relations doesn't matter (although the order within each relation does matter). And when we draw the graph, all that matters is what's connected to what; the actual layout on the page is just a choice made for visual presentation. It also doesn't matter what the

elements are called. Here I've used numbers, but all that matters is that the elements are distinct.

OK, so what do we do with these collections of relations, or graphs? We just apply a simple rule to them, over and over again. Here's an example of a possible rule:

$$\{\{x, y\}, \{x, z\}\} \rightarrow \{\{x, z\}, \{x, w\}, \{y, w\}, \{z, w\}\}$$

What this rule says is to pick up two relations—from anywhere in the collection—and see if the elements in them match the pattern $\{\{x,y\},\{x,z\}\}$ (or, in the Wolfram Language, $\{\{x\_,y\_\}, \{x\_,z\_\}\}$), where the two x's can be anything, but both have to be the same, and the y and z can be anything. If there's a match, then replace these two relations with the four relations on the right. The w that appears there is a new element that's being created, and the only requirement is that it's distinct from all other elements.

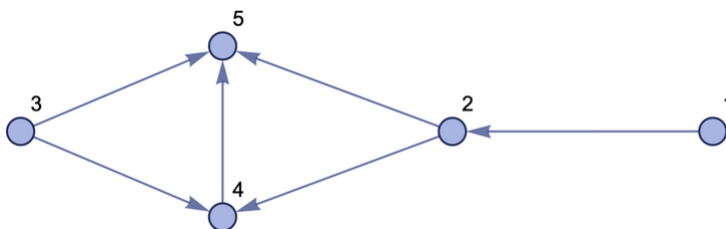We can represent the rule as a transformation of graphs:
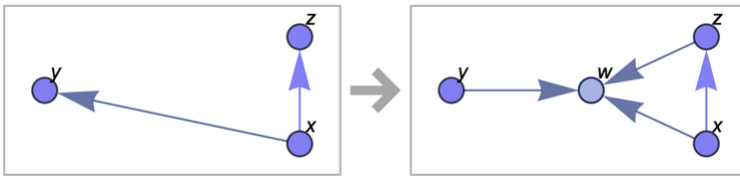
Now let's apply the rule once to:

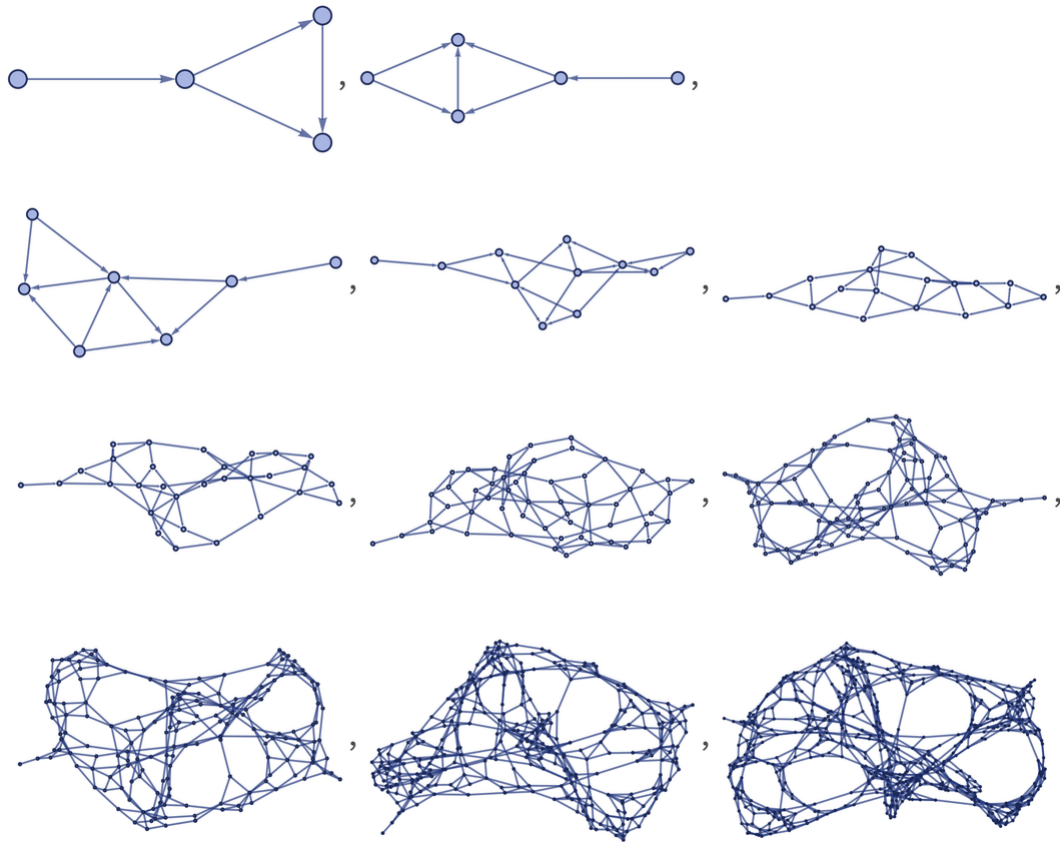$$\{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{2, 4\}\}$$

The $\{2,3\}$ and $\{2,4\}$ relations get matched, and the rule replaces them with four new relations, so the result is:

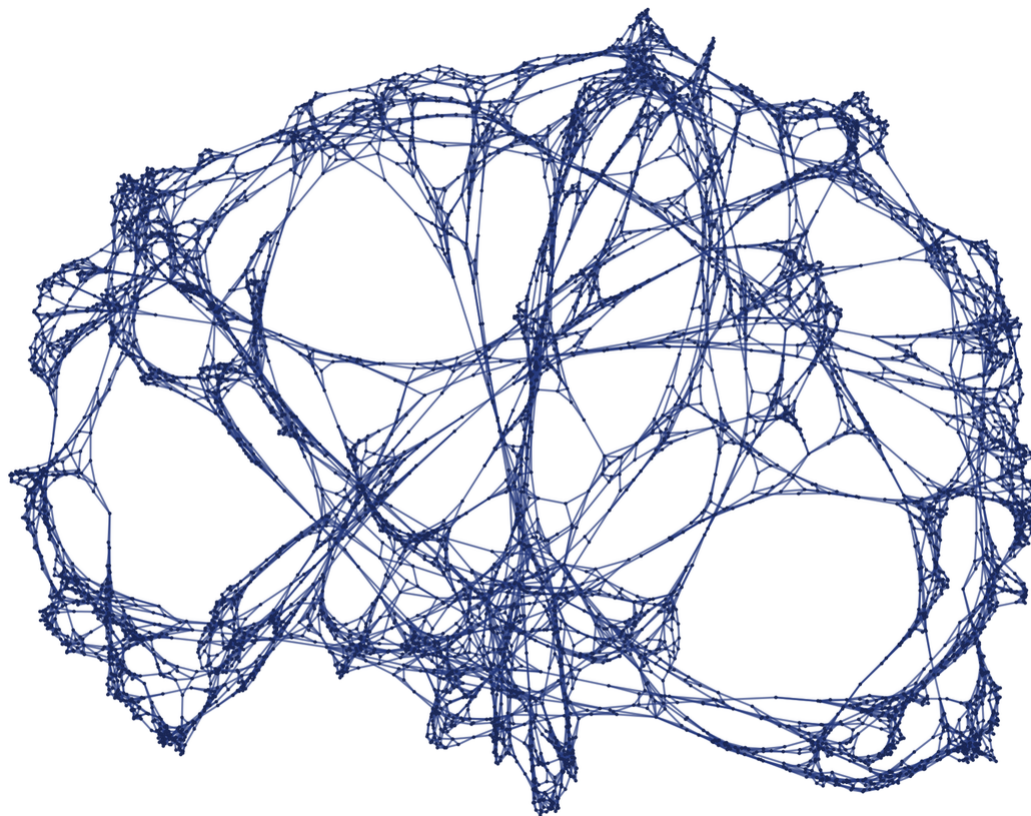$$\{\{1, 2\}, \{3, 4\}, \{2, 4\}, \{2, 5\}, \{3, 5\}, \{4, 5\}\}$$

We can represent this result as a graph (which happens to be rendered flipped relative to the graph above):

OK, so what happens if we just keep applying the rule over and over? Here's the result:

Let's do it a few more times, and make a bigger picture:

What happened here? We have such a simple rule. Yet applying this rule over and over again produces something that looks really complicated. It's not what our ordinary intuition tells us should happen. But actually—as I first discovered in the early 1980s—this kind of intrinsic, spontaneous generation of complexity turns out to be completely ubiquitous among simple rules and simple programs. And for example my book *A New Kind of Science* is about this whole phenomenon and why it's so important for science and beyond.

But here what's important about it is that it's what's going to make our universe, and everything in it. Let's review again what we've seen. We started off with a simple rule that just tells us how to transform collections of relations. But what we get out is this complicated-looking object that, among other things, seems to have some definite shape.

We didn't put in anything about this shape. We just gave a simple rule. And using that simple rule a graph was made. And when we visualize that graph, it comes out looking like it has a definite shape.

If we ignore all matter in the universe, our universe is basically a big chunk of space. But what *is* that space? We've had mathematical idealizations and abstractions of it for two thousand years. But what really is it? Is it made of something, and if so, what?

Well, I think it's very much like the picture above. A whole bunch of what are essentially abstract points, abstractly connected together. Except that in the picture there are 6704 of these points, whereas in our real universe there might be more like 10^{400} of them, or even many more.

All Possible Rules

We don't (yet) know an actual rule that represents our universe—and it's almost certainly not the one we just talked about. So let's discuss what possible rules there are, and what they typically do.

One feature of the rule we used above is that it's based on collections of "binary relations", containing pairs of elements (like {2,3}). But the same setup lets us also consider relations with more elements. For example, here's a collection of two ternary relations:

{{1, 2, 3}, {3, 4, 5}}

We can't use an ordinary graph to represent things like this, but we can use a hypergraph—a construct where we generalize edges in graphs that connect pairs of nodes to "hyperedges" that connect any number of nodes:
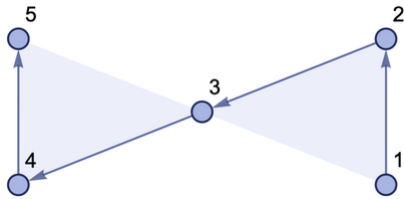

(Notice that we're dealing with directed hypergraphs, where the order in which nodes appear in a hyperedge matters. In the picture, the "membranes" are just indicating which nodes are connected to the same hyperedge.)

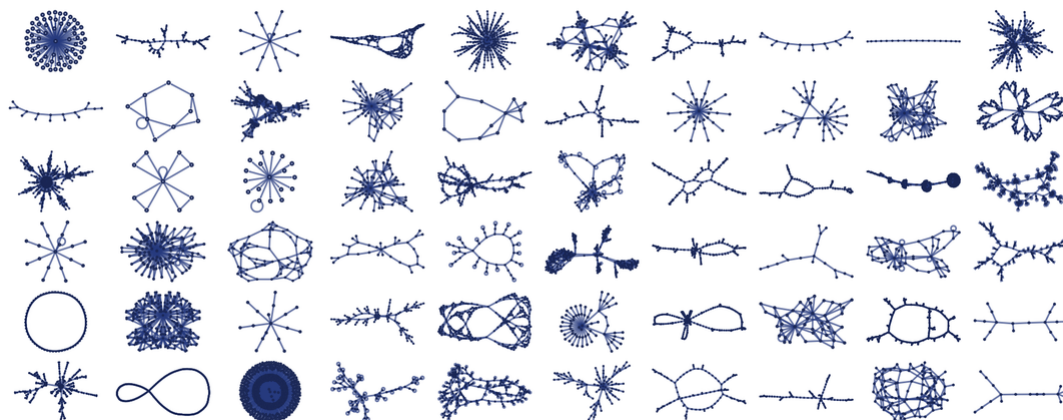We can make rules for hypergraphs too:

$$\{\{x, y, z\}\} \rightarrow \{\{w, w, y\}, \{w, x, z\}\}$$
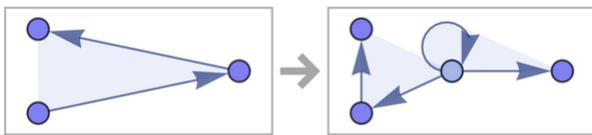
And now here's what happens if we run this rule starting from the simplest possible ternary hypergraph—the ternary self-loop $\{\{0,0,0\}\}$:
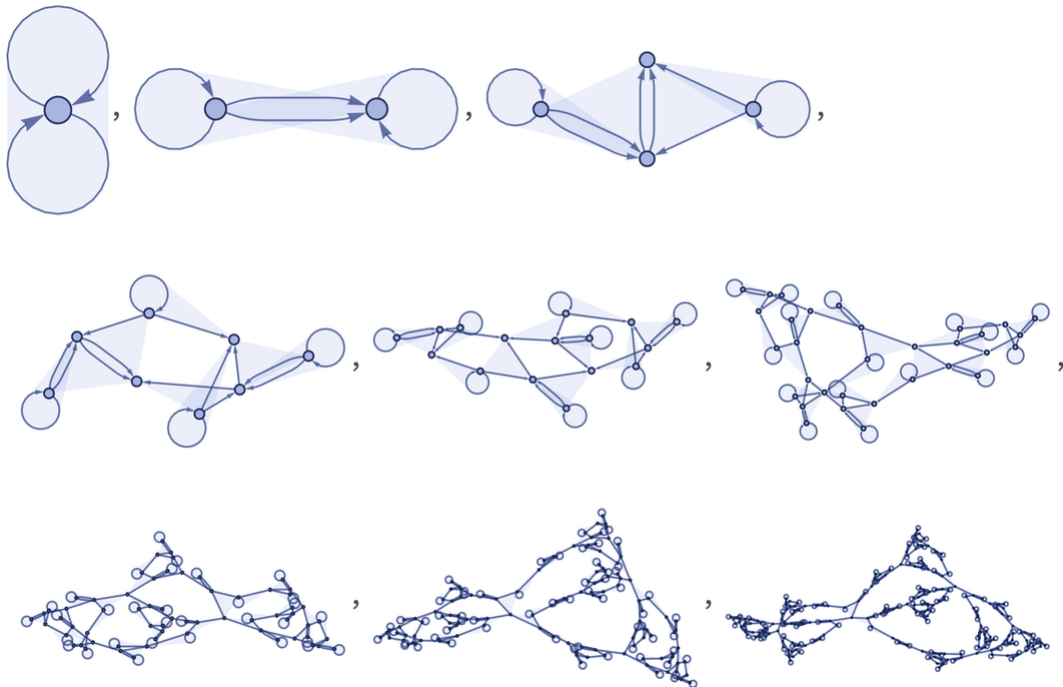

Alright, so what happens if we just start picking simple rules at random? Here are some of the things they do:

Somehow this looks very zoological (and, yes, these models are definitely relevant for things other than fundamental physics—though probably particularly molecular-scale construction). But basically what we see here is that there are various common forms of behavior, some simple, and some not.

Here are some samples of the kinds of things we see:
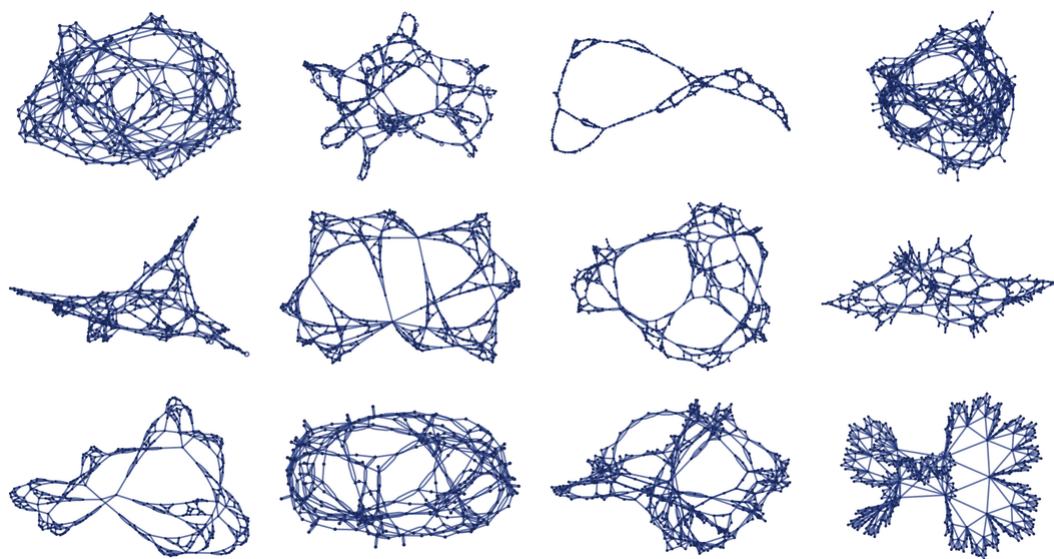
And the big question is: if we were to run rules like these long enough, would they end up making something that reproduces our physical universe? Or, put another way, out in this computational universe of simple rules, can we find our physical universe?

A big question, though, is: How would we know? What we're seeing here are the results of applying rules a few thousand times; in our actual universe they may have been applied 10^{500} times so far, or even more. And it's not easy to bridge that gap. And we have to work it from both sides. First, we have to use the best summary of the operation of our universe that what we've learned in physics over the past few centuries has given us. And second, we have to go as far as we can in figuring out what our rules actually do.

And here there's potentially a fundamental problem: the phenomenon of computational irreducibility. One of the great achievements of the mathematical sciences, starting about three centuries ago, has been delivering equations and formulas that basically tell you how a system will behave without you having to trace each step in what the system does. But many years ago I realized that in the computational universe of possible rules, this very often isn't possible. Instead, even if you know the exact rule that a system follows, you may still not be able to work out what the system will do except by essentially just tracing every step it takes.

One might imagine that—once we know the rule for some system—then with all our computers and brainpower we'd always be able to "jump ahead" and work out what the system would do. But actually there's something I call the Principle of Computational Equivalence, which says that almost any time the behavior of a system isn't obviously simple, it's computationally as sophisticated as anything. So we won't be able to "outcompute" it—and to work out what it does will take an irreducible amount of computational work.

Well, for our models of the universe this is potentially a big problem. Because we won't be able to get even close to running those models for as long as the universe does. And at the outset it's not clear that we'll be able to tell enough from what we can do to see if it matches up with physics.

But the big recent surprise for me is that we seem to be lucking out. We do know that whenever there's computational irreducibility in a system, there are also an infinite number of pockets of computational reducibility. But it's completely unclear whether in our case those pockets will line up with things we know from physics. And the surprise is that it seems a bunch of them do.
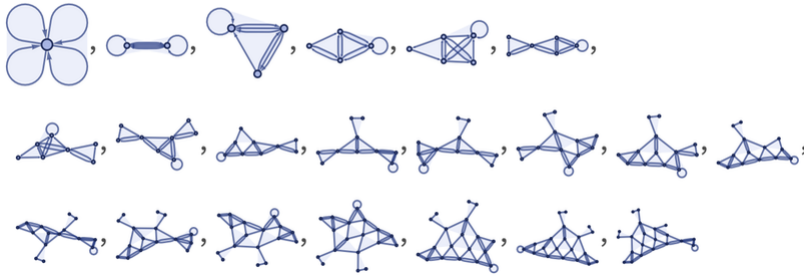
What Is Space?

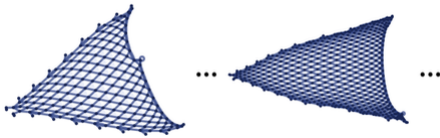Let's look at a particular, simple rule from our infinite collection:

$$\{\{x, y, y\}, \{z, x, u\}\} \rightarrow \{\{y, v, y\}, \{y, z, v\}, \{u, v, v\}\}$$
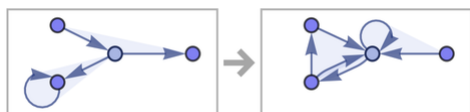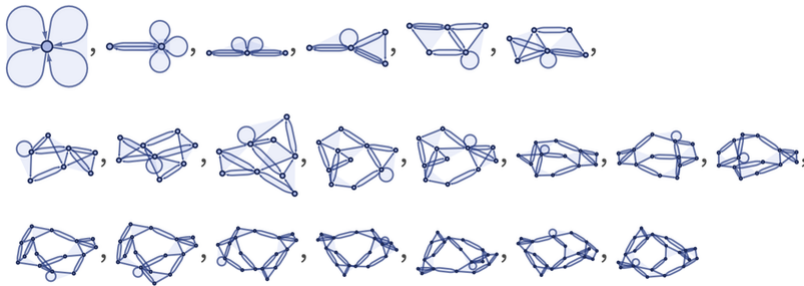

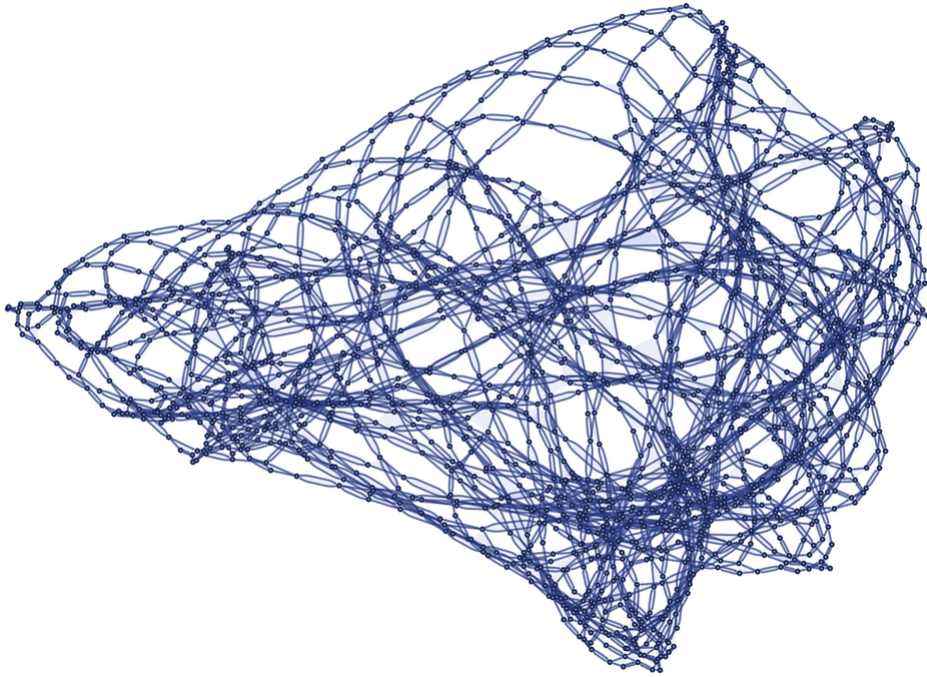Here's what it does:

And after a while this is what happens:

It's basically making us a very simple "piece of space". If we keep on going longer and longer it'll make a finer and finer mesh, to the point where what we have is almost indistinguishable from a piece of a continuous plane.
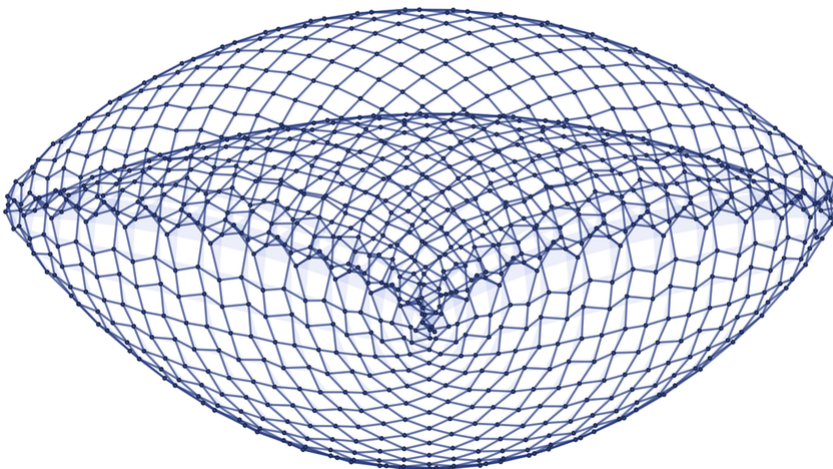
Here's a different rule:
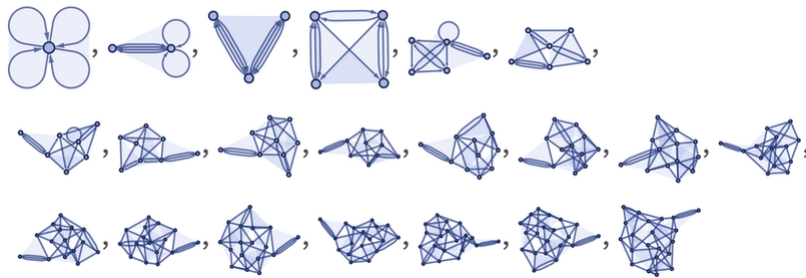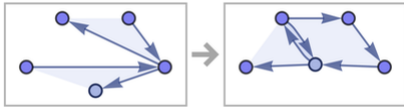
$$\{\{x, x, y\}, \{z, u, x\}\} \rightarrow \{\{u, u, z\}, \{v, u, v\}, \{v, y, x\}\}$$

It looks it's "trying to make" something 3D. Here's another rule:

$$\{\{x, y, z\}, \{u, y, v\}\} \rightarrow \{\{w, z, x\}, \{z, w, u\}, \{x, y, w\}\}$$

Isn't this strange? We have a rule that's just specifying how to rewrite pieces of an abstract hypergraph, with no notion of geometry, or anything about 3D space. And yet it produces a hypergraph that's naturally laid out as something that looks like a 3D surface.

Even though the only thing that's really here is connections between points, we can "guess" where a surface might be, then we can show the result in 3D:
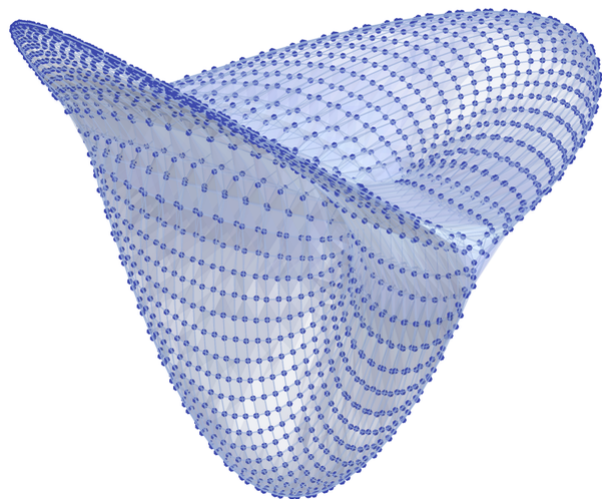
If we keep going, then like the example of the plane, the mesh will get finer and finer, until basically our rule has grown us—point by point, connection by connection—something that's like a continuous 3D surface of the kind you might study in a calculus class. Of course, in some sense, it's not "really" that surface: it's just a hypergraph that represents a bunch of abstract relations—but somehow the pattern of those relations gives it a structure that's a closer and closer approximation to the surface.

And this is basically how I think space in the universe works. Underneath, it's a bunch of discrete, abstract relations between abstract points. But at the scale we're experiencing it, the pattern of relations it has makes it seem like continuous space of the kind we're used to. It's a bit like what happens with, say, water. Underneath, it's a bunch of discrete molecules bouncing around. But to us it seems like a continuous fluid.

Needless to say, people have thought that space might ultimately be discrete ever since antiquity. But in modern physics there was never a way to make it work—and anyway it was much more convenient for it to be continuous, so one could use calculus. But now it's looking like the idea of space being discrete is actually crucial to getting a fundamental theory of physics.

The Dimensionality of Space

A very fundamental fact about space as we experience it is that it is three-dimensional. So can our rules reproduce that? Two of the rules we just saw produce what we can easily recognize as two-dimensional surfaces—in one case flat, in the other case arranged in a certain shape. Of course, these are very bland examples of (two-dimensional) space: they are effectively just simple grids. And while this is what makes them easy to recognize, it also means that they're not actually much like our universe, where there's in a sense much more going on.

So, OK, take a case like:
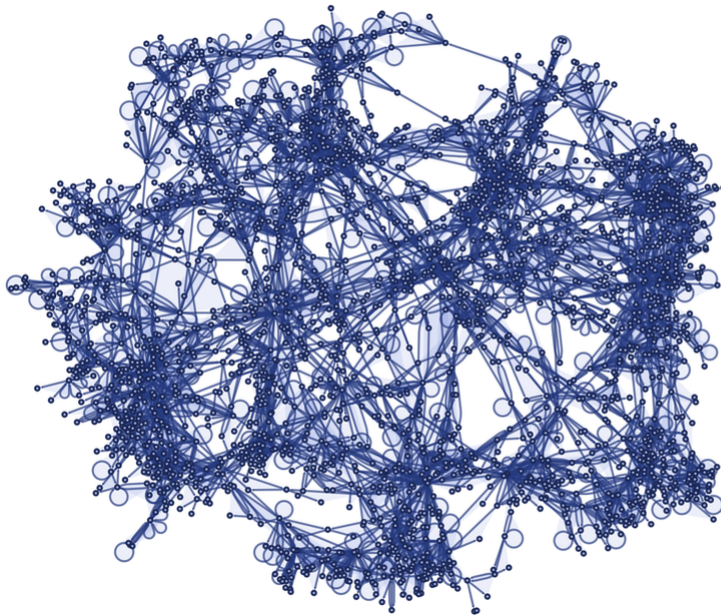
If we were to go on long enough, would this make something like space, and, if so, with how many dimensions? To know the answer, we have to have some robust way to measure dimension. But remember, the pictures we're drawing are just visualizations; the underlying structure is a bunch of discrete relations defining a hypergraph—with no information about coordinates, or geometry, or even topology. And, by the way, to emphasize that point, here is the same graph—with exactly the same connectivity structure —rendered four different ways:
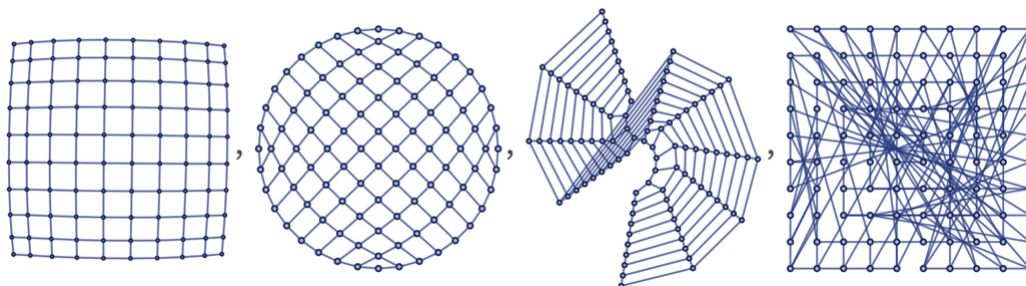
But getting back to the question of dimension, recall that the area of a circle is πr^2; the volume of a sphere is $\frac{4}{3}\pi r^3$. In general, the "volume" of the d-dimensional analog of a sphere is a constant multiplied by r^d. But now think about our hypergraph. Start at some point in the hypergraph. Then follow r hyperedges in all possible ways. You've effectively made the analog of a "spherical ball" in the hypergraph. Here are examples for graphs corresponding to 2D and 3D lattices:
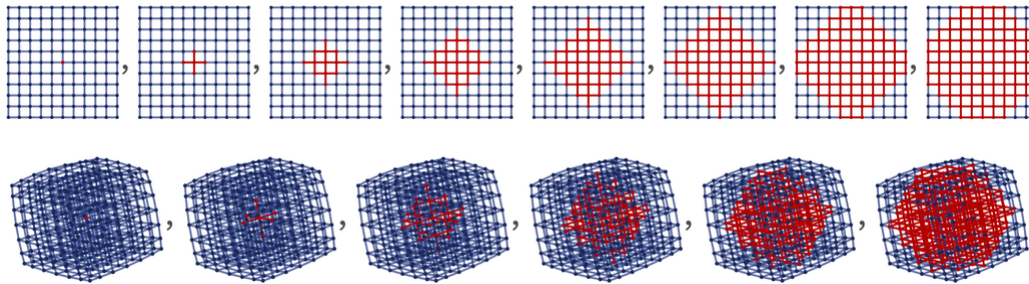
And if you now count the number of points reached by going "graph distance r" (i.e. by following r connections in the graph) you'll find in these two cases that they indeed grow like r^2 and r^3.
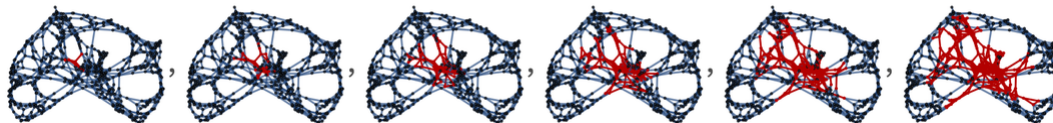
So this gives us a way to measure the effective dimension of our hypergraphs. Just start at a particular point and see how many points you reach by going r steps:

Now to work out effective dimension, we in principle just have to fit the results to r^d. It's a bit complicated, though, because we need to avoid small r (where every detail of the hypergraph is going to matter) and large r (where we're hitting the edge of the hypergraph) —and we also need to think about how our "space" is refining as the underlying system evolves. But in the end we can generate a series of fits for the effective dimension—and in this case these say that the effective dimension is about 2.7:
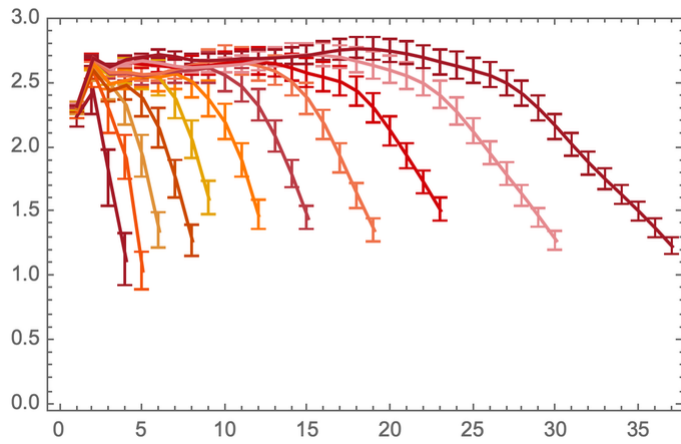
If we do the same thing for
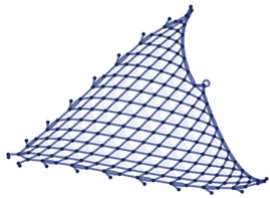
it's limiting to dimension 2, as it should:
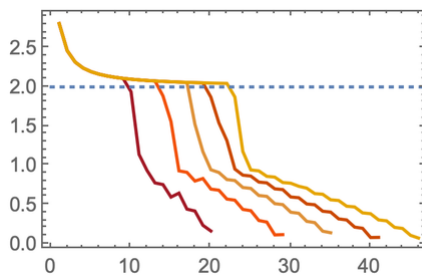
What does the fractional dimension mean? Well, consider fractals, which our rules can easily make:

$$\{\{x, y, z\}\} \rightarrow \{\{x, u, w\}, \{y, v, u\}, \{z, w, v\}\}$$
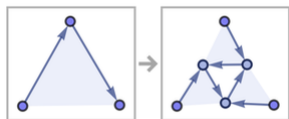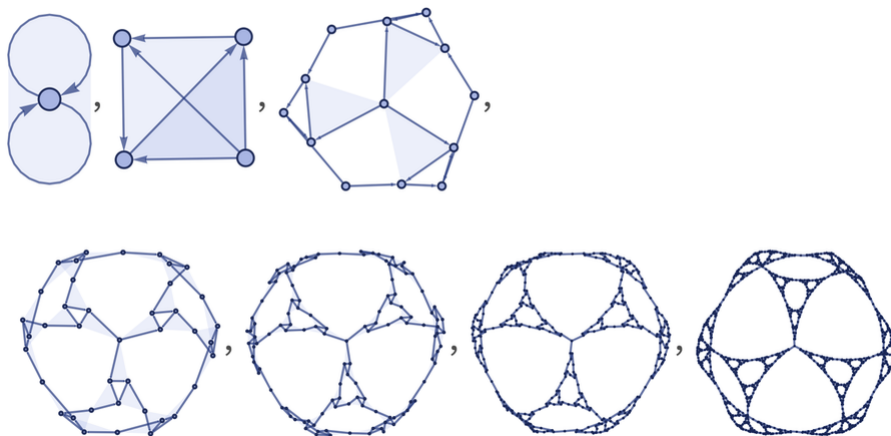
If we measure the dimension here we get 1.58—the usual fractal dimension $\log_2(3)$ for a Sierpiński structure:
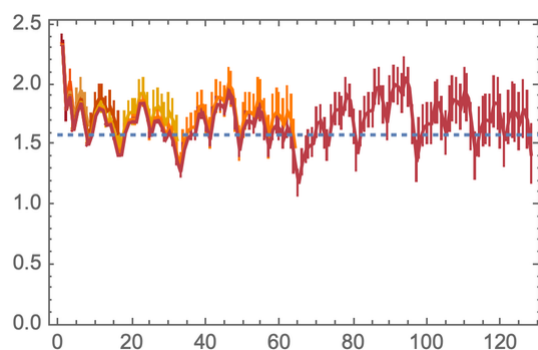
Our rule above doesn't create a structure that's as regular as this. In fact, even though the rule itself is completely deterministic, the structure it makes looks quite random. But what our measurements suggest is that when we keep running the rule it produces something that's like 2.7-dimensional space.

Of course, 2.7 is not 3, and presumably this particular rule isn't the one for our particular universe (though it's not clear what effective dimension it'd have if we ran it 10^{100} steps). But the process of measuring dimension shows an example of how we can start making "physics-connectable" statements about the behavior of our rules.

By the way, we've been talking about "making space" with our models. But actually, we're not just trying to make space; we're trying to make everything in the universe. In standard current physics, there's space—described mathematically as a manifold—and serving as a kind of backdrop, and then there's everything that's in space, all the matter and particles and planets and so on.

But in our models there's in a sense nothing but space—and in a sense everything in the universe must be "made of space". Or, put another way, it's the exact same hypergraph

that's giving us the structure of space, and everything that exists in space.

So what this means is that, for example, a particle like an electron or a photon must correspond to some local feature of the hypergraph, a bit like in this toy example:
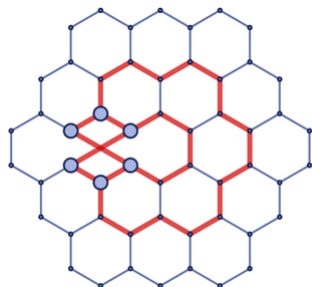
To give a sense of scale, though, I have an estimate that says that 10^{200} times more "activity" in the hypergraph that represents our universe is going into "maintaining the structure of space" than is going into maintaining all the matter we know exists in the universe.

Curvature in Space & Einstein's Equations

Here are a few structures that simple examples of our rules make:
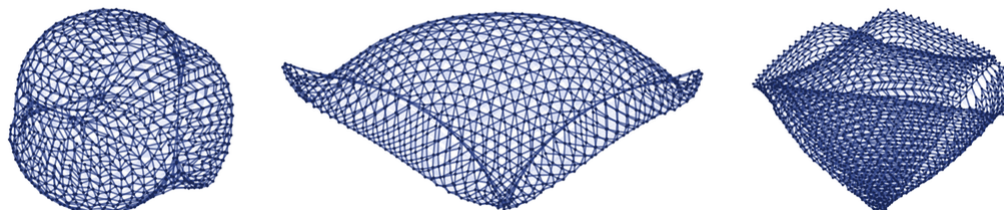
But while all of these look like surfaces, they're all obviously different. And one way to characterize them is by their local curvature. Well, it turns out that in our models, curvature is a concept closely related to dimension—and this fact will actually be critical in understanding, for example, how gravity arises.

But for now, let's talk about how one would measure curvature on a hypergraph. Normally the area of a circle is πr^2. But let's imagine that we've drawn a circle on the surface of a sphere, and now we're measuring the area on the sphere that's inside the circle:
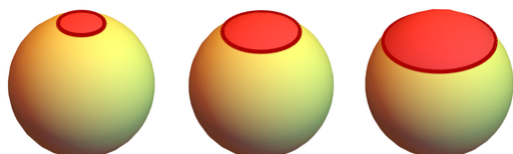
This area is no longer πr^2. Instead it's $\pi r^2 \left(1 - \dfrac{r^2}{12\,a^2} + \dfrac{r^4}{360\,a^4} - \ldots \right)$, where a is the radius of

the sphere. In other words, as the radius of the circle gets bigger, the effect of being on the sphere is ever more important. (On the surface of the Earth, imagine a circle drawn around the North Pole; once it gets to the equator, it can never get any bigger.)

If we generalize to d dimensions, it turns out the formula for the growth rate of the volume

is $r^d \left(1 - \dfrac{r^2}{6\,(d+2)}\,R + \ldots \right)$, where R is a mathematical object known as the Ricci scalar

curvature.

So what this all means is that if we look at the growth rates of spherical balls in our hypergraphs, we can expect two contributions: a leading one of order r^d that corresponds to effective dimension, and a "correction" of order r^2 that represents curvature.

Here's an example. Instead of giving a flat estimate of dimension (here equal to 2), we have something that dips down, reflecting the positive ("sphere-like") curvature of the surface:
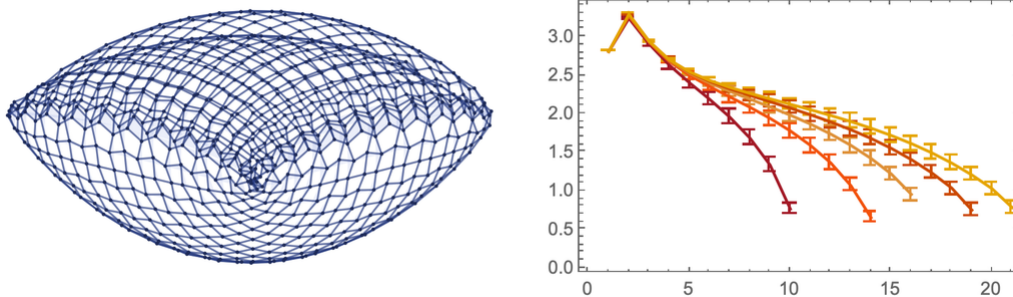
What is the significance of curvature? One thing is that it has implications for geodesics. A geodesic is the shortest distance between two points. In ordinary flat space, geodesics are just lines. But when there's curvature, the geodesics are curved:
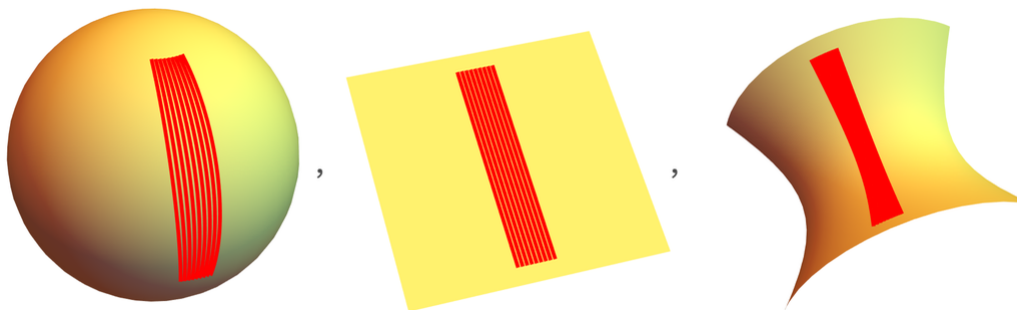
In the case of positive curvature, bundles of geodesics converge; for negative curvature they diverge. But, OK, even though geodesics were originally defined for continuous space (actually, as the name suggests, for paths on the surface of the Earth), one can also have them in graphs (and hypergraphs). And it's the same story: the geodesic is the shortest path between two points in the graph (or hypergraph).

Here are geodesics on the "positive-curvature surface" created by one of our rules:
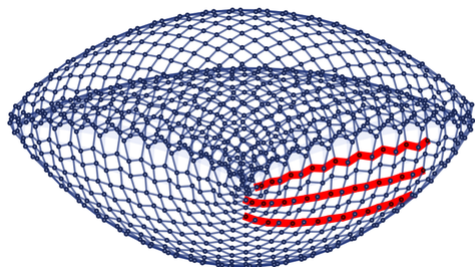
And here they are for a more complicated structure:
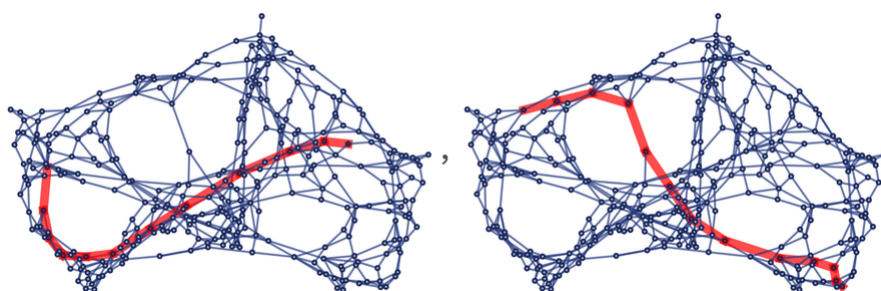
Why are geodesics important? One reason is that in Einstein's general relativity they're the paths that light (or objects in "free fall") follows in space. And in that theory gravity is associated with curvature in space. So when something is deflected going around the Sun, that happens because space around the Sun is curved, so the geodesic the object follows is also curved.

General relativity's description of curvature in space turns out to all be based on the Ricci scalar curvature R that we encountered above (as well as the slightly more sophisticated Ricci tensor). But so if we want to find out if our models are reproducing Einstein's equations for gravity, we basically have to find out if the Ricci curvatures that arise from our hypergraphs are the same as the theory implies.

There's quite a bit of mathematical sophistication involved (for example, we have to consider curvature in space+time, not just space), but the bottom line is that, yes, in various limits, and subject to various assumptions, our models do indeed reproduce Einstein's equations. (At first, we're just reproducing the vacuum Einstein equations, appropriate

when there's no matter involved; when we discuss matter, we'll see that we actually get the full Einstein equations.)

It's a big deal to reproduce Einstein's equations. Normally in physics, Einstein's equations are what you start from (or sometimes they arise as a consistency condition for a theory): here they're what comes out as an emergent feature of the model.

It's worth saying a little about how the derivation works. It's actually somewhat analogous to the derivation of the equations of fluid flow from the limit of the underlying dynamics of lots of discrete molecules. But in this case, it's the structure of space rather than the velocity of a fluid that we're computing. It involves some of the same kinds of mathematical approximations and assumptions, though. One has to assume, for example, that there's enough effective randomness generated in the system that statistical averages work. There is also a whole host of subtle mathematical limits to take. Distances have to be large compared to individual hypergraph connections, but small compared to the whole size of the hypergraph, etc.

It's pretty common for physicists to "hack through" the mathematical niceties. That's actually happened for nearly a century in the case of deriving fluid equations from molecular dynamics. And we're definitely guilty of the same thing here. Which in a sense is another way of saying that there's lots of nice mathematics to do in actually making the derivation rigorous, and understanding exactly when it'll apply, and so on.

By the way, when it comes to mathematics, even the setup that we have is interesting. Calculus has been built to work in ordinary continuous spaces (manifolds that locally approximate Euclidean space). But what we have here is something different: in the limit of an infinitely large hypergraph, it's like a continuous space, but ordinary calculus doesn't work on it (not least because it isn't necessarily integer-dimensional). So to really talk about it well, we have to invent something that's kind of a generalization of calculus, that's for example capable of dealing with curvature in fractional-dimensional space. (Probably the closest current mathematics to this is what's been coming out of the very active field of geometric group theory.)

It's worth noting, by the way, that there's a lot of subtlety in the precise tradeoff between changing the dimension of space, and having curvature in it. And while we think our universe is three-dimensional, it's quite possible according to our models that there are at

least local deviations—and most likely there were actually large deviations in the early universe.

Time

In our models, space is defined by the large-scale structure of the hypergraph that represents our collection of abstract relations. But what then is time?

For the past century or so, it's been pretty universally assumed in fundamental physics that time is in a sense "just like space"—and that one should for example lump space and time together and talk about the "spacetime continuum". And certainly the theory of relativity points in this direction. But if there's been one "wrong turn" in the history of physics in the past century, I think it's the assumption that space and time are the same kind of thing. And in our models they're not—even though, as we'll see, relativity comes out just fine.

So what then is time? In effect it's much as we experience it: the inexorable process of things happening and leading to other things. But in our models it's something much more precise: it's the progressive application of rules, that continually modify the abstract structure that defines the contents of the universe.

The version of time in our models is in a sense very computational. As time progresses we are in effect seeing the results of more and more steps in a computation. And indeed the phenomenon of computational irreducibility implies that there is something definite and irreducible "achieved" by this process. (And, for example, this irreducibility is what I believe is responsible for the "encrypting" of initial conditions that is associated with the law of entropy increase, and the thermodynamic arrow of time.) Needless to say, of course, our modern computational paradigm did not exist a century ago when "spacetime" was introduced, and perhaps if it had, the history of physics might have been very different.

But, OK, so in our models time is just the progressive application of rules. But there is a subtlety in exactly how this works that might at first seem like a detail, but that actually turns out to be huge, and in fact turns out to be the key to both relativity and quantum mechanics.

At the beginning of this piece, I talked about the rule

$$\{\{x, y\}, \{x, z\}\} \rightarrow \{\{x, z\}, \{x, w\}, \{y, w\}, \{z, w\}\}$$
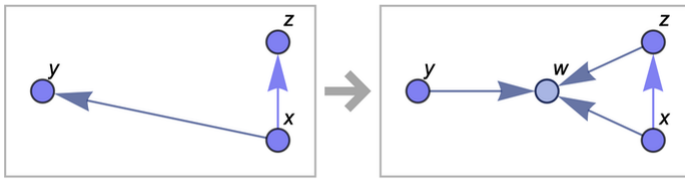
and showed the "first few steps" in applying it
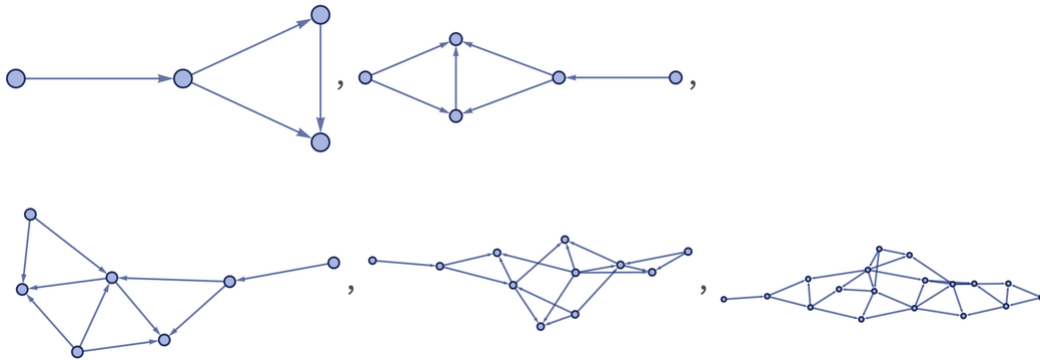
But [how exactly did the rule get applied]? What is "inside" these steps? The rule defines how to take two connections in the hypergraph (which in this case is actually just a graph) and transform them into four new connections, creating a new element in the process. So each "step" that we showed before actually consists of several individual "updating events" (where here newly added connections are highlighted, and ones that are about to be removed are dashed):
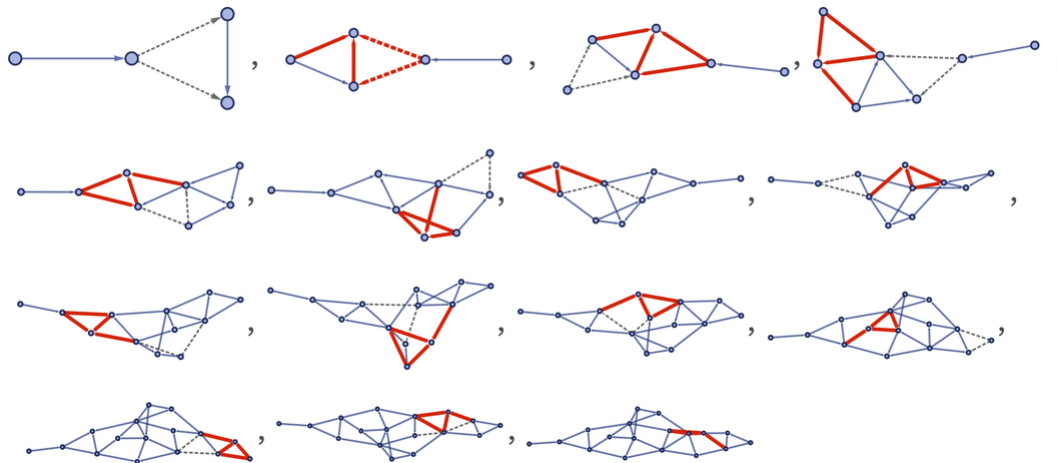
But now, here is the crucial point: this is not the only sequence of updating events consistent with the rule. The rule just says to find two adjacent connections, and if there are several possible choices, it says nothing about which one. And a crucial idea in our model is in a sense just to do all of them.

We can represent this with a [graph that shows all possible paths]:
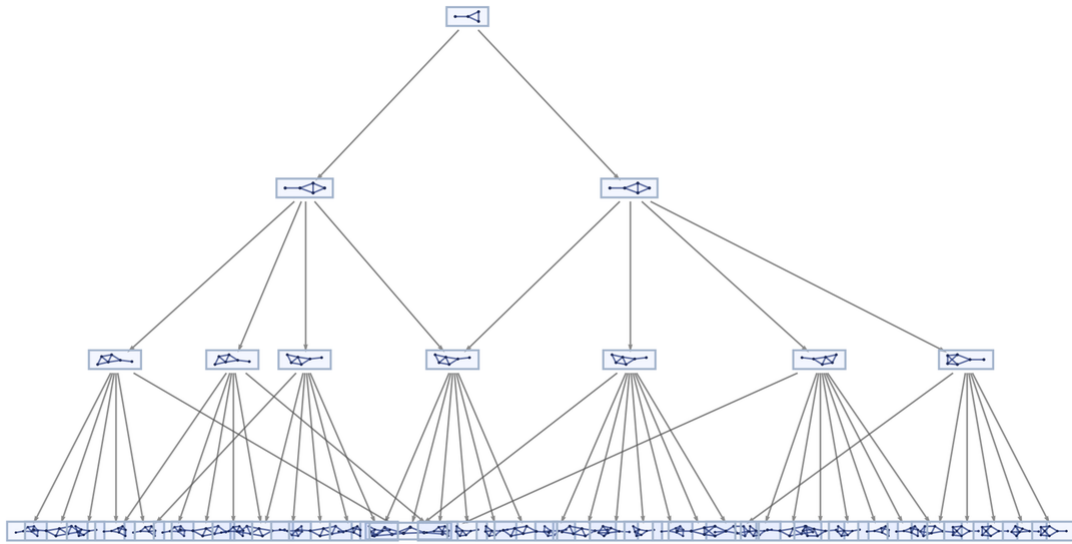
For the very first update, there are two possibilities. Then for each of the results of these, there are four additional possibilities. But at the next update, something important happens: two of the branches merge. In other words, even though we have done a different sequence of updates, the outcome is the same.

Things rapidly get complicated. Here is the graph after one more update, now no longer trying to show a progression down the page:
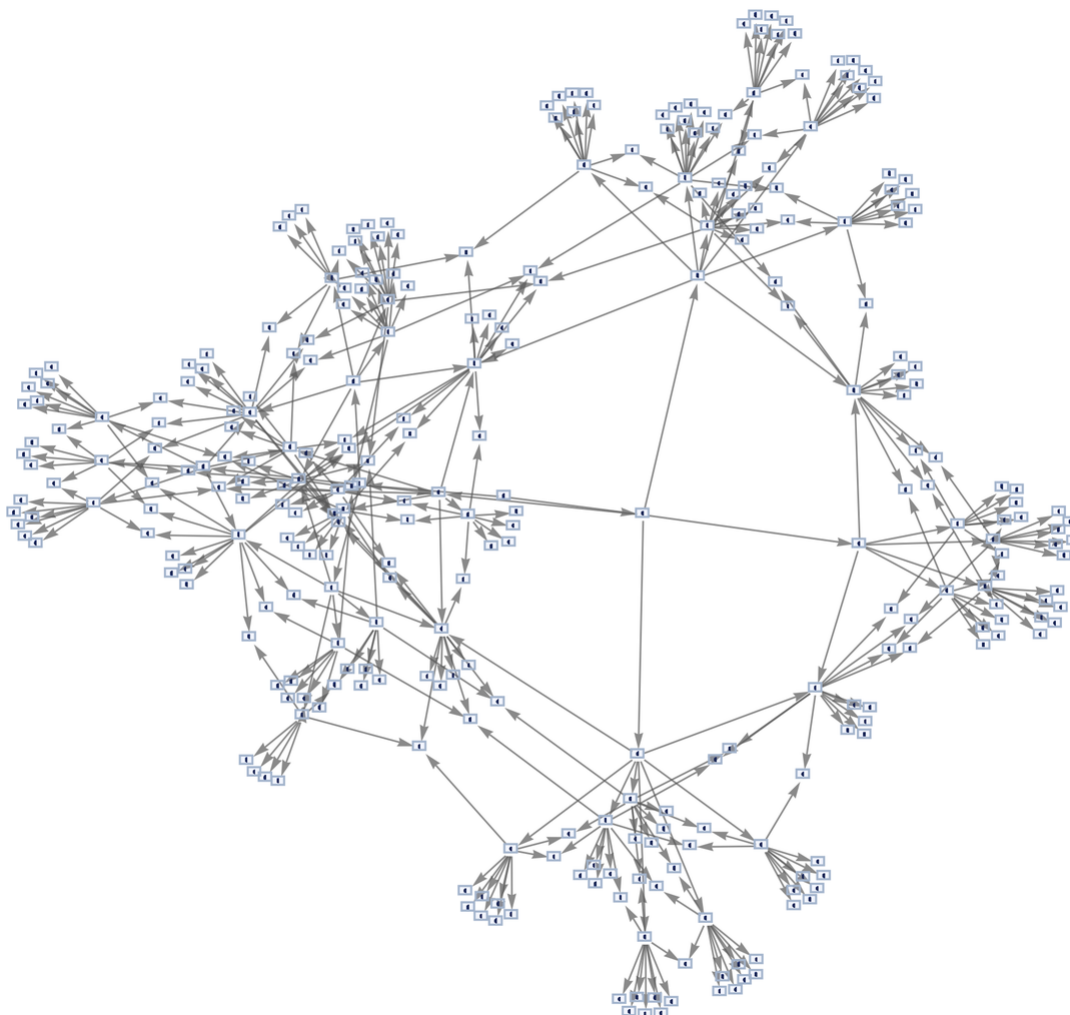
So how does this relate to time? What it says is that in the basic statement of the model there is not just one path of time; there are many paths, and many "histories". But the model—and the rule that is used—determines all of them. And we have seen a hint of something else: that even if we might think we are following an "independent" path of history, it may actually merge with another path.

It will take some more discussion to explain how this all works. But for now let me say that what will emerge is that time is about causal relationships between things, and that in fact, even when the paths of history that are followed are different, these causal relationships can end up being the same—and that in effect, to an observer embedded in the system, there is still just a single thread of time.

The Graph of Causal Relationships

In the end it's wonderfully elegant. But to get to the point where we can understand the elegant bigger picture we need to go through some detailed things. (It isn't terribly

surprising that a fundamental theory of physics—inevitably built on very abstract ideas—is somewhat complicated to explain, but so it goes.)

To keep things tolerably simple, I'm not going to talk directly about rules that operate on hypergraphs. Instead I'm going to talk about rules that operate on strings of characters. (To clarify: these are not the strings of string theory—although in a bizarre twist of "pun-becomes-science" I suspect that the continuum limit of the operations I discuss on character strings is actually related to string theory in the modern physics sense.)

OK, so let's say we have the rule:

{A → BBB, BB → A}

This rule says that anywhere we see an A, we can replace it with BBB, and anywhere we see BB we can replace it with A. So now we can generate what we call the multiway system for this rule, and draw a "multiway graph" that shows everything that can happen:
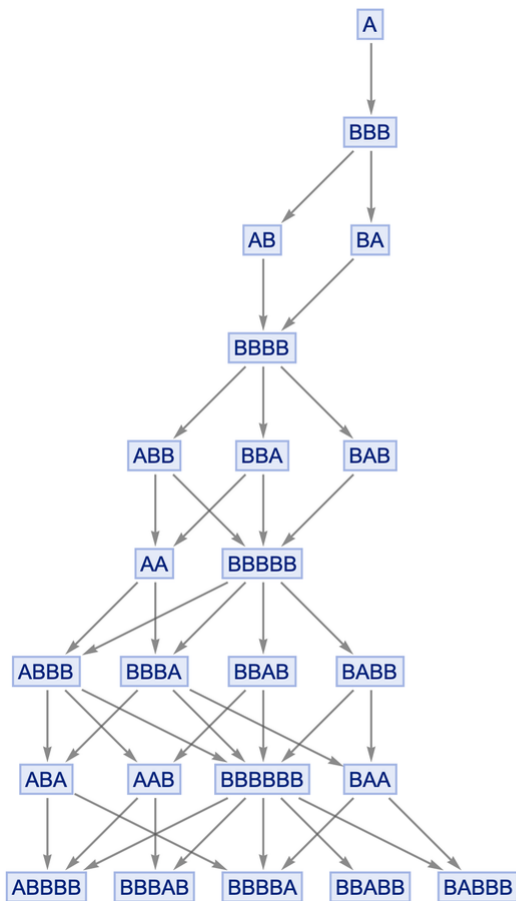
At the first step, the only possibility is to use A→BBB to replace the A with BBB. But then there are two possibilities: replace either the first BB or the second BB—and these choices

give different results. On the next step, though, all that can be done is to replace the A—in both cases giving BBBB.

So in other words, even though we in a sense had two paths of history that diverged in the multiway system, it took only one step for them to converge again. And if you trace through the picture above you'll find out that's what always happens with this rule: every pair of branches that is produced always merges, in this case after just one more step.

This kind of balance between branching and merging is a phenomenon I call "causal invariance". And while it might seem like a detail here, it actually turns out that it's at the core of why relativity works, why there's a meaningful objective reality in quantum mechanics, and a host of other core features of fundamental physics.

But let's explain why I call the property causal invariance. The picture above just shows what "state" (i.e. what string) leads to what other one. But at the risk of making the picture more complicated (and note that this is incredibly simple compared to the full hypergraph case), we can annotate the multiway graph by including the updating events that lead to each transition between states:
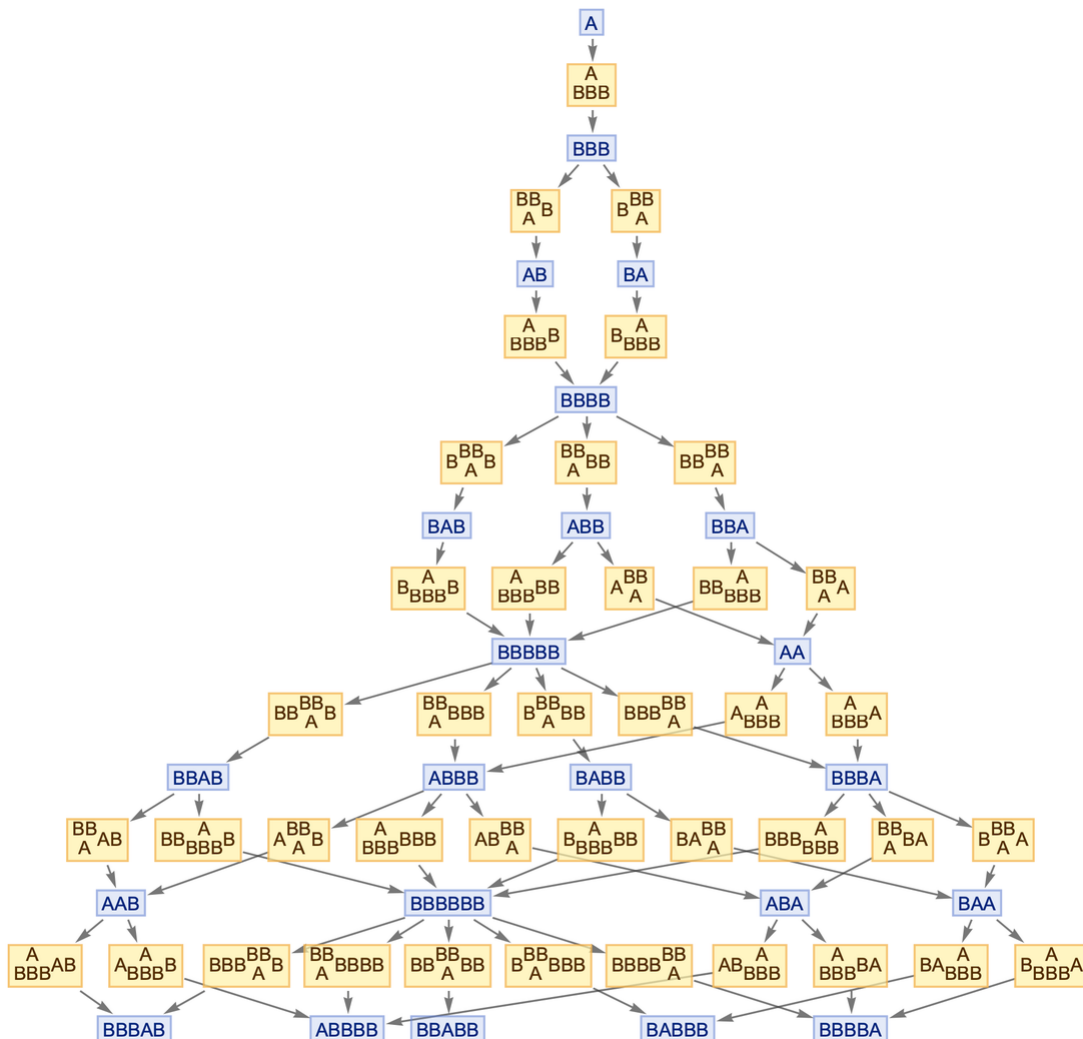
But now we can ask the question: what are the causal relationships between these events? In other words, what event needs to happen before some other event can happen? Or, said another way, what events must have happened in order to create the input that's needed for some other event?

Let us go even further, and annotate the graph above by showing all the causal dependencies between events:
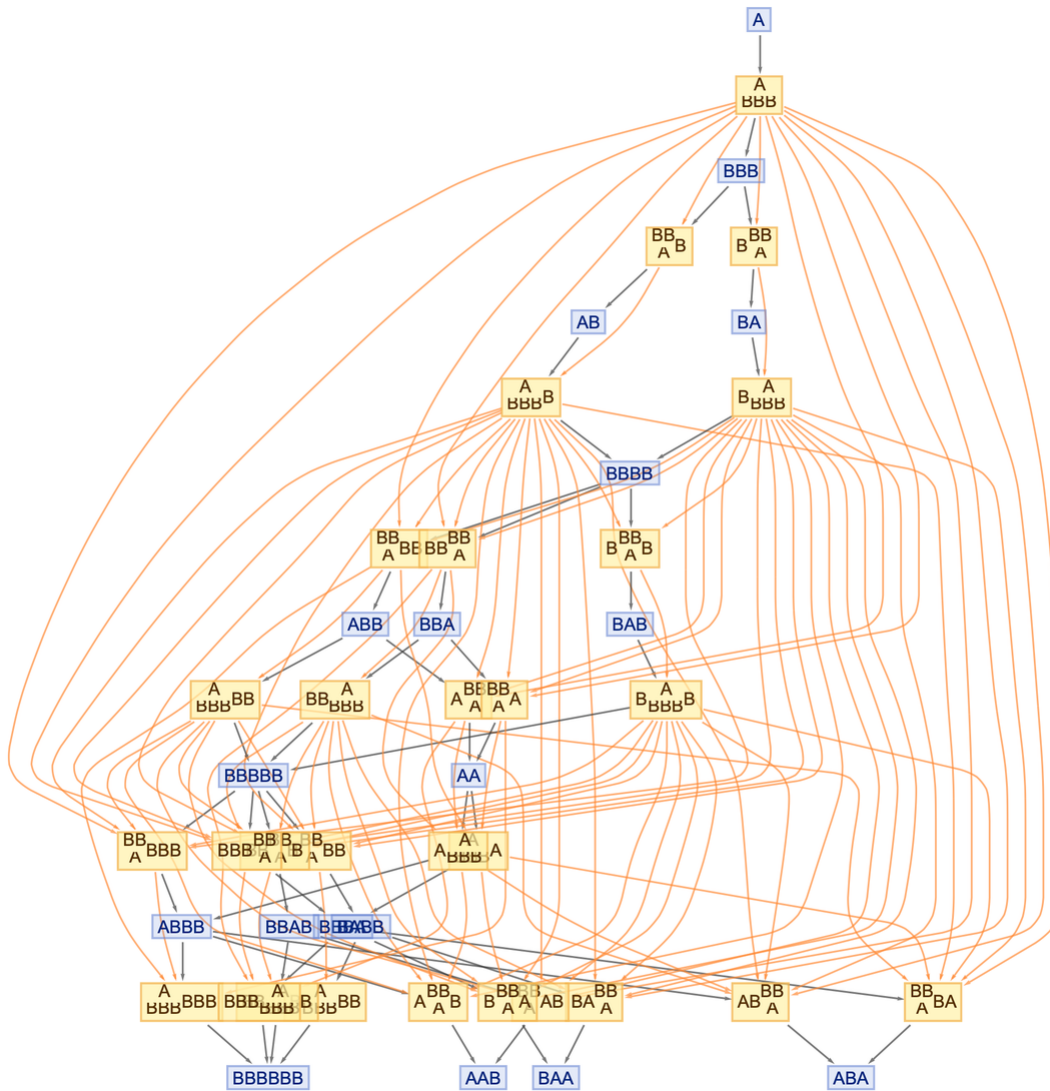
The orange lines in effect show which event has to happen before which other event—or what all the causal relationships in the multiway system are. And, yes, it's complicated. But note that this picture shows the whole multiway system—with all possible paths of history— as well as the whole network of causal relationships within and between these paths.

But here's the crucial thing about causal invariance: it implies that actually the graph of causal relationships is the same regardless of which path of history is followed. And that's why I originally called this property "causal invariance"—because it says that with a rule like this, the causal properties are invariant with respect to different choices of the sequence in which updating is done.

And if one traced through the picture above (and went quite a few more steps), one would find that for every path of history, the causal graph representing causal relationships between events would always be:
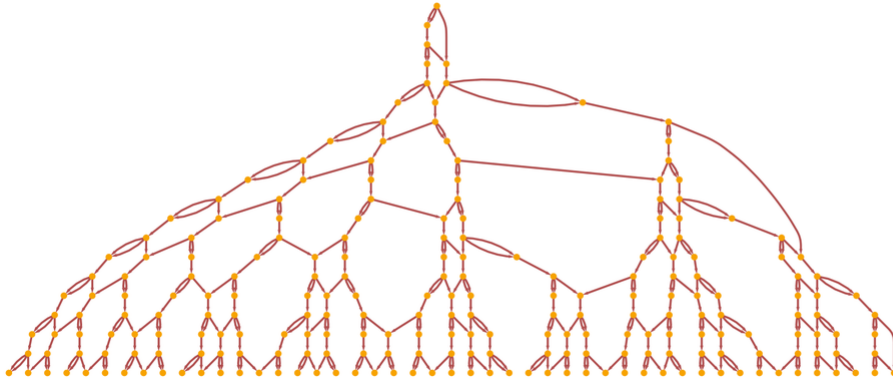
or, drawn differently,
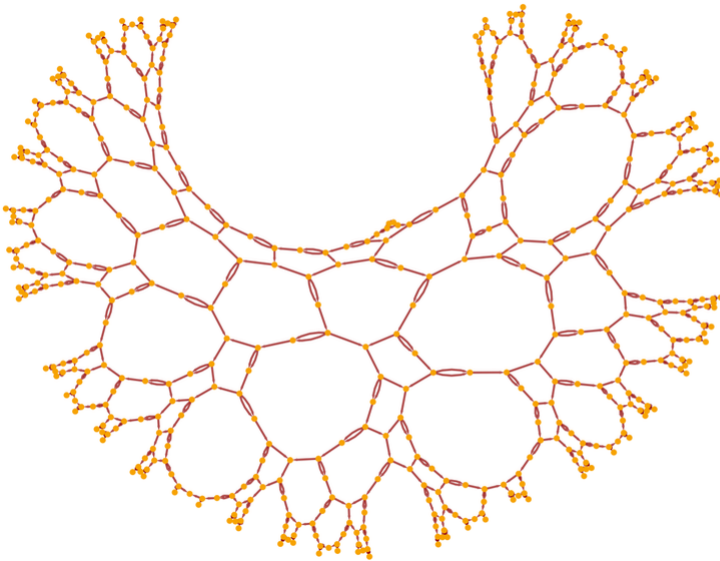
The Importance of Causal Invariance

To understand more about causal invariance, it's useful to look at an [even simpler example](#): the case of the rule BA→AB. This rule says that any time there's a B followed by an A in a string, swap these characters around. In other words, this is a rule that tries to sort a string into alphabetical order, two characters at a time.

Let's say we start with BBBAAA. Then here's the multiway graph that shows all the things that can happen according to the rule:
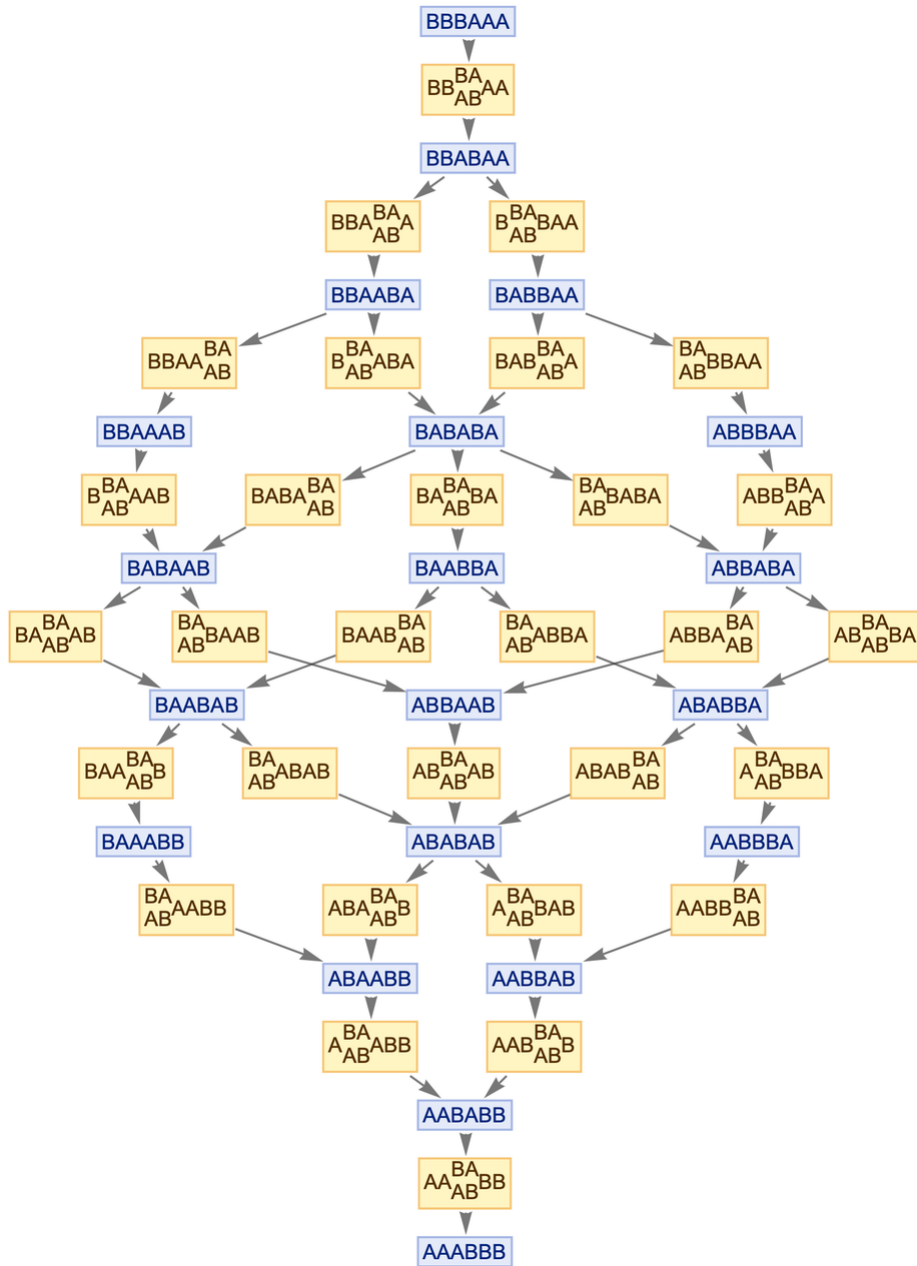
There are lots of different paths that can be followed, depending on which BA in the string the rule is applied to at each step. But the important thing we see is that at the end all the paths merge, and we get a single final result: the sorted string AAABBB. And the fact that we get this single final result is a consequence of the causal invariance of the rule. In a case like this where there's a final result (as opposed to just evolving forever), causal invariance basically says: it doesn't matter what order you do all the updates in; the result you'll get will always be the same.

I've introduced causal invariance in the context of trying to find a model of fundamental physics—and I've said that it's going to be critical to both relativity and quantum mechanics. But actually what amounts to causal invariance has been seen before in various different guises in mathematics, mathematical logic and computer science. (Its most

common name is "confluence", though there are some technical differences between this and what I call causal invariance.)

Think about expanding out an algebraic expression, like $(x + (1 + x)^2)(x + 2)^2$. You could expand one of the powers first, then multiply things out. Or you could multiply the terms first. It doesn't matter what order you do the steps in; you'll always get the same canonical form (which in this case Mathematica tells me is $4 + 16x + 17x^2 + 7x^3 + x^4$). And this independence of orders is essentially causal invariance.

Here's one more example. Imagine you've got some recursive definition, say f[n_]:=f[n-1]+f[n-2] (with f[0]=f[1]=1). Now evaluate f[10]. First you get f[9]+f[8]. But what do you do next? Do you evaluate f[9], or f[8]? And then what? In the end, it doesn't matter; you'll always get 55. And this is another example of causal invariance.

When one thinks about parallel or asynchronous algorithms, it's important if one has causal invariance. Because it means one can do things in any order—say, depth-first, breadth-first, or whatever—and one will always get the same answer. And that's what's happening in our little sorting algorithm above.

OK, but now let's come back to causal relationships. Here's the multiway system for the sorting process annotated with all causal relationships for all paths:
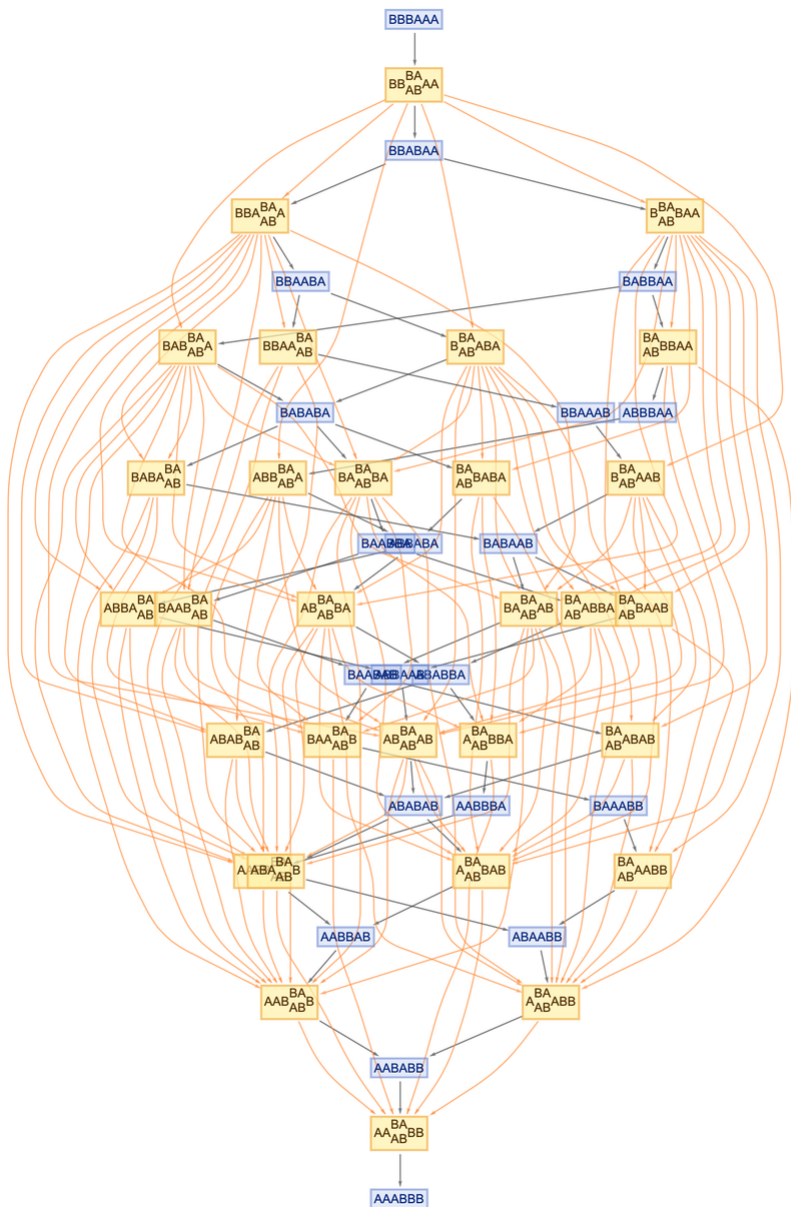
And, yes, it's a mess. But because there's causal invariance, we know something very important: this is basically just a lot of copies of the same causal graph—a simple grid:
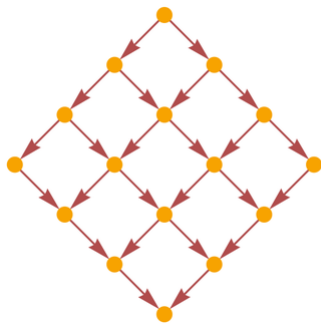
(By the way—as the picture suggests—the cross-connections between these copies aren't trivial, and later on we'll see they're associated with deep relations between relativity and

quantum mechanics, that probably manifest themselves in the physics of black holes. But we'll get to that later…)

OK, so every different way of applying the sorting rule is supposed to give the same causal graph. So here's one example of how we might apply the rule starting with a particular initial string:
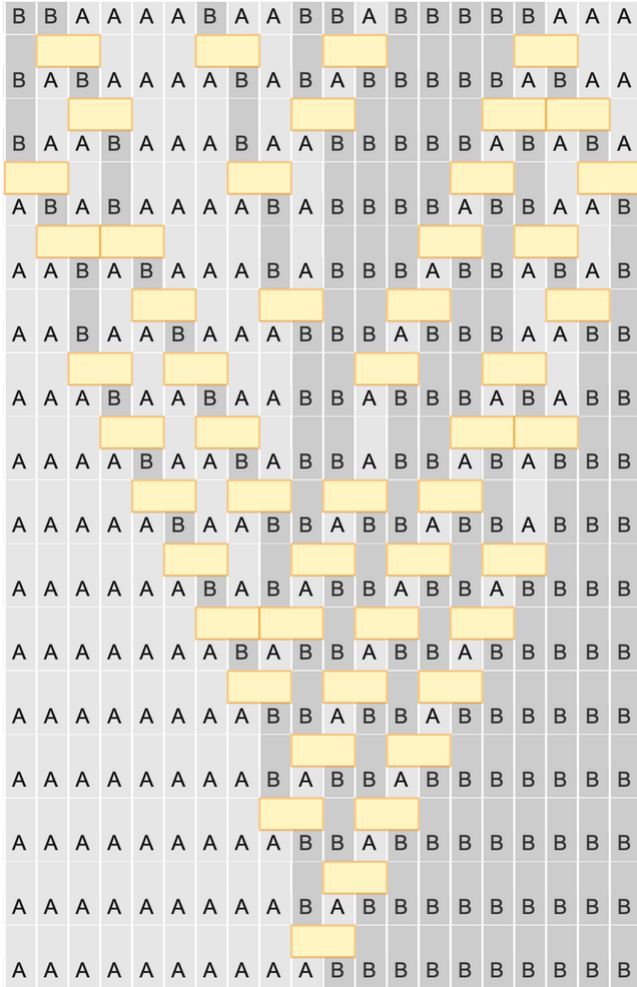
But now let's show the graph of causal connections. And we see it's just a grid:
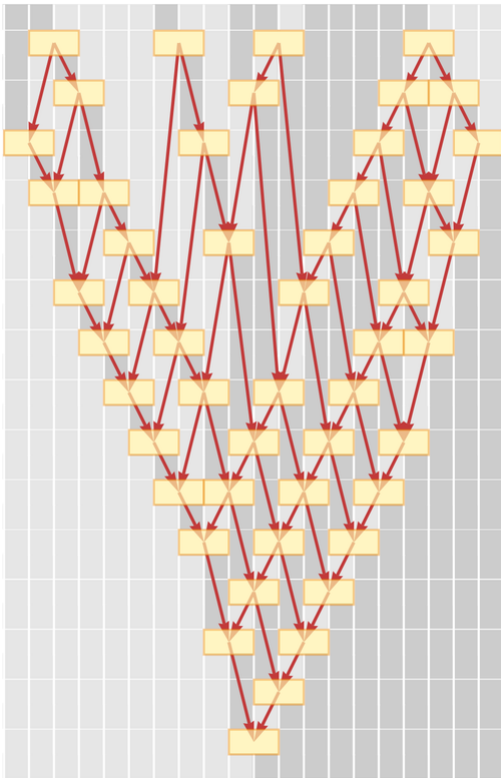
Here are three other possible sequences of updates:
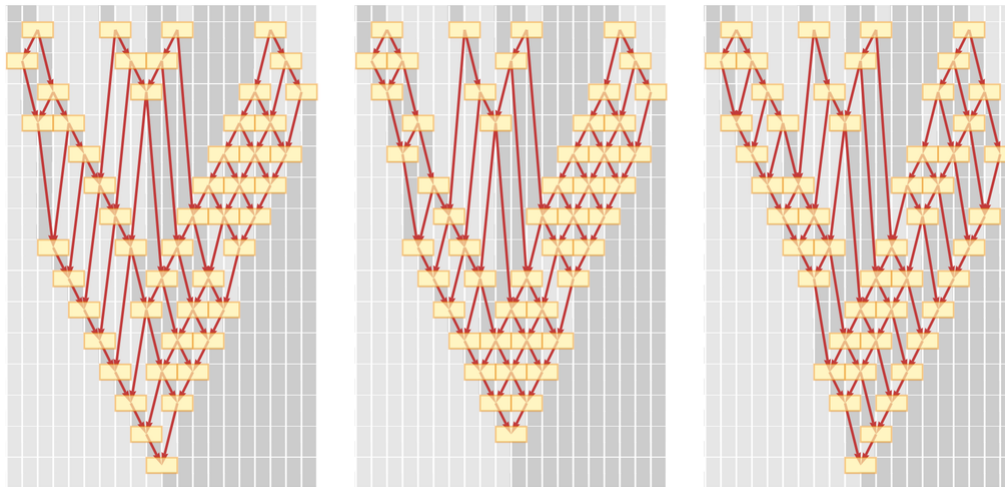
But now we see causal invariance in action: even though different updates occur at different times, the graph of causal relationships between updating events is always the same. And having seen this—in the context of a very simple example—we're ready to talk about special relativity.

Deriving Special Relativity

It's a typical first instinct in thinking about doing science: you imagine doing an experiment on a system, but you—as the "observer"—are outside the system. Of course if you're

thinking about modeling the whole universe and everything in it, this isn't ultimately a reasonable way to think about things. Because the "observer" is inevitably part of the universe, and so has to be modeled just like everything else.

In our models what this means is that the "mind of the observer", just like everything else in the universe, has to get updated through a series of updating events. There's no absolute way for the observer to "know what's going on in the universe"; all they ever experience is a series of updating events, that may happen to be affected by updating events occurring elsewhere in the universe. Or, said differently, all the observer can ever observe is the network of causal relationships between events—or the causal graph that we've been talking about.

So as toy model let's look at our BA→AB rule for strings. We might imagine that the string is laid out in space. But to our observer the only thing they know is the causal graph that represents causal relationships between events. And for the BA→AB system here's one way we can draw that:
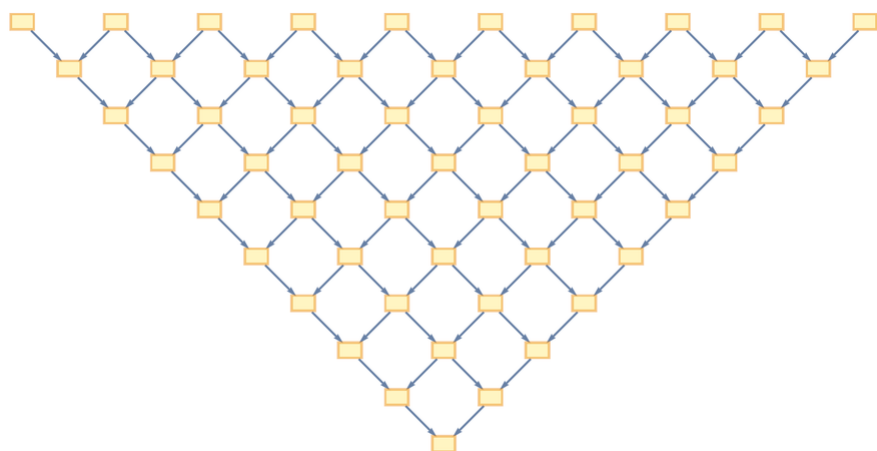
But now let's think about how observers might "experience" this causal graph. Underneath, an observer is getting updated by some sequence of updating events. But even though that's "really what's going on", to make sense of it, we can imagine our observers setting up internal "mental" models for what they see. And a pretty natural thing for observers like us to do is just to say "one set of things happens all across the universe, then another, and so on". And we can translate this into saying that we imagine a series of "moments" in time, where things happen "simultaneously" across the universe—at least with some convention for defining what we mean by simultaneously. (And, yes, this part of what we're doing is basically following what Einstein did when he originally proposed special relativity.)

Here's a possible way of doing it:
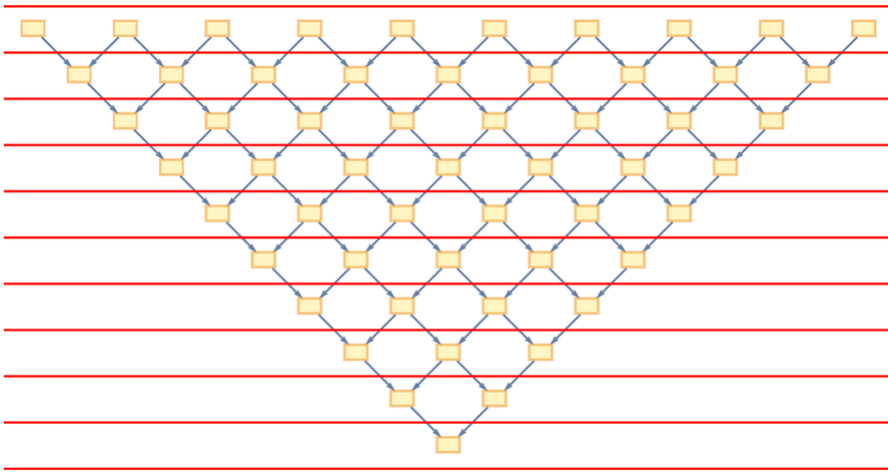
One can describe this as a "foliation" of the causal graph. We're dividing the causal graph into leaves or slices. And each slice our observers can consider to be a "successive moment in time".

It's important to note that there are some constraints on the foliation we can pick. The causal graph defines what event has to happen before what. And if our observers are going to have a chance of making sense of the world, it had better be the case that their notion of the progress of time aligns with what the causal graph says. So for example this foliation wouldn't work—because basically it says that the time we assign to events is going to disagree with the order in which the causal graph says they have to happen:
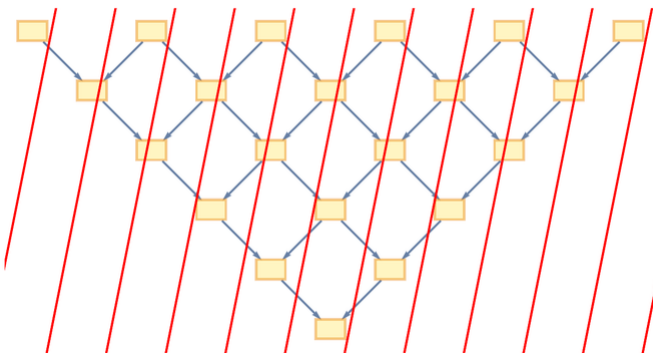
But, so given the foliation above, what actual order of updating events does it imply? It basically just says: as many events as possible happen at the same time (i.e. in the same slice of the foliation), as in this picture:
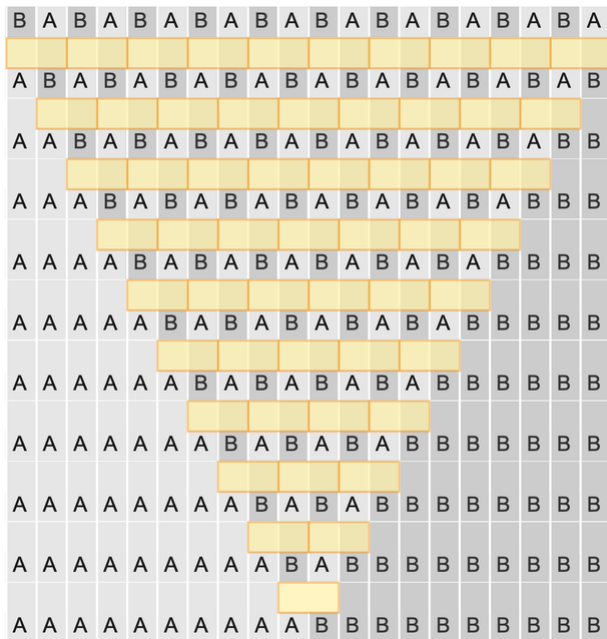
OK, now let's connect this to physics. The foliation we had above is relevant to observers who are somehow "stationary with respect to the universe" (the "cosmological rest frame"). One can imagine that as time progresses, the events a particular observer experiences are ones in a column going vertically down the page:
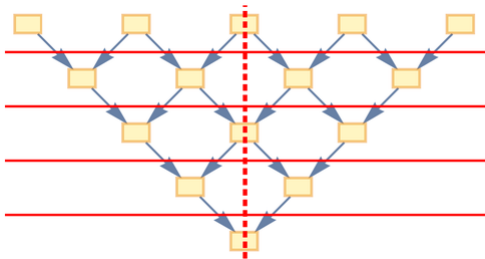
But now let's think about an observer who is uniformly moving in space. They'll experience a different sequence of events, say:
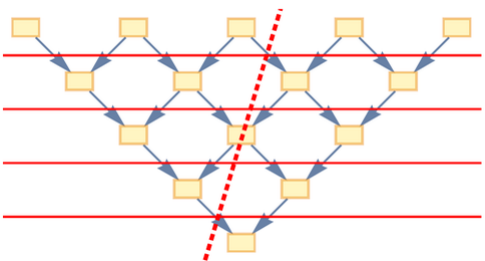
And that means that the foliation they'll naturally construct will be different. From the "outside" we can draw it on the causal graph like this:
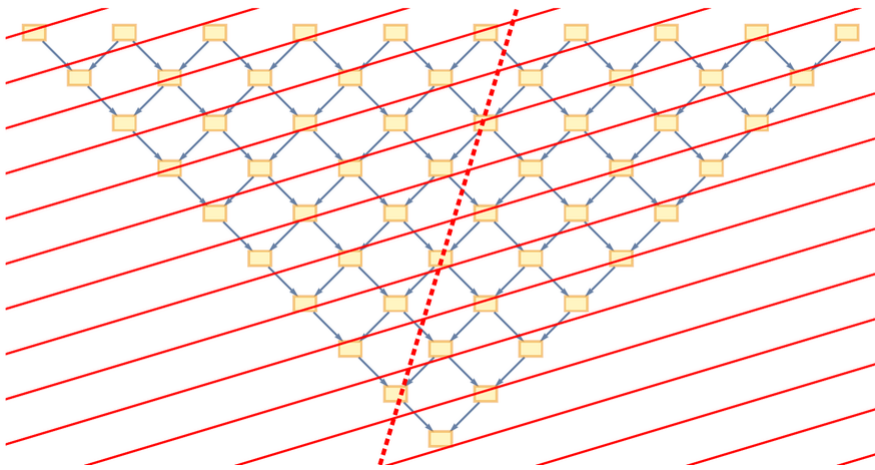
But to the observer each slice just represents a successive moment of time. And they don't have any way to know how the causal graph was drawn. So they'll construct their own version, where the slices are horizontal:
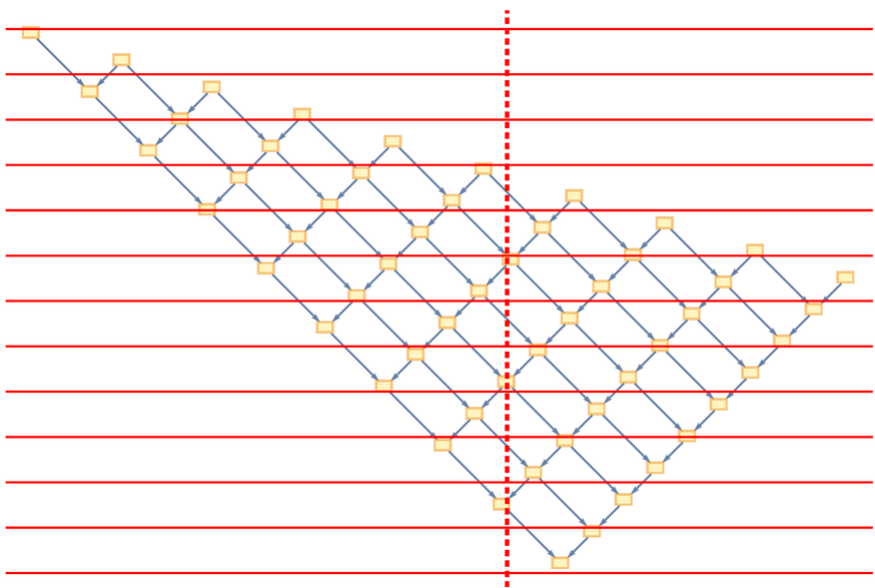
But now there's a purely geometrical fact: to make this rearrangement, while preserving the basic structure (and here, angles) of the causal graph, each moment of time has to sample fewer events in the causal graph, by a factor of $\sqrt{1 - \beta^2}$ where β is the angle that represents the velocity of the observer.

If you know about special relativity, you'll recognize a lot of this. What we've been calling foliations correspond directly to relativity's "reference frames". And our foliations that represent motion are the standard inertial reference frames of special relativity.

But here's the special thing that's going on here: we can interpret all this discussion of foliations and reference frames in terms of the actual rules and evolution of our underlying

system. So here now is the evolution of our string-sorting system in the "boosted reference frame" corresponding to an observer going at a certain speed:

And here's the crucial point: because of causal invariance it doesn't matter that we're in a different reference frame—the causal graph for the system (and the way it eventually sorts the string) is exactly the same.

In special relativity, the key idea is that the "laws of physics" work the same in all inertial reference frames. But why should that be true? Well, in our systems, there's an answer: it's a consequence of causal invariance in the underlying rules. In other words, from the property of causal invariance, we're able to derive relativity.

Normally in physics one puts in relativity by the way one sets up the mathematical structure of spacetime. But in our models we don't start from anything like this, and in fact space and

time are not even at all the same kind of thing. But what we can now see is that—because of causal invariance—relativity emerges in our models, with all the relationships between space and time that that implies.

So, for example, if we look at the picture of our string-sorting system above, we can see relativistic time dilation. In effect, because of the foliation we picked, time operates slower. Or, said another way, in the effort to sample space faster, our observer experiences slower updating of the system in time.

The speed of light c in our toy system is defined by the maximum rate at which information can propagate, which is determined by the rule, and in the case of this rule is one character per step. And in terms of this, we can then say that our foliation corresponds to a speed 0.3 c. But now we can look at the amount of time dilation, and it's exactly the amount

$$\gamma = 1 \bigg/ \sqrt{1 - \frac{v^2}{c^2}}$$ that relativity says it should be.

By the way, if we imagine trying to make our observer go "faster than light", we can see that can't work. Because there's no way to tip the foliation at more than 45° in our picture, and still maintain the causal relationships implied by the causal graph.

OK, so in our toy model we can derive special relativity. But here's the thing: this derivation isn't specific to the toy model; it applies to any rule that has causal invariance. So even though we may be dealing with hypergraphs, not strings, and we may have a rule that shows all kinds of complicated behavior, if it ultimately has causal invariance, then (with various technical caveats, mostly about possible wildness in the causal graph) it will exhibit relativistic invariance, and a physics based on it will follow special relativity.

What Is Energy? What Is Mass?

In our model, everything in the universe—space, matter, whatever—is supposed to be represented by features of our evolving hypergraph. So within that hypergraph, is there a way to identify things that are familiar from current physics, like mass, or energy?

I have to say that although it's a widespread concept in current physics, I'd never thought of energy as something fundamental. I'd just thought of it as an attribute that things (atoms, photons, whatever) can have. I never really thought of it as something that one could identify abstractly in the very structure of the universe.

So it came as a big surprise when we recently realized that actually in our model, there is something we can point to, and say "that's energy!", independent of what it's the energy of. The technical statement is: energy corresponds to the flux of causal edges through spacelike hypersurfaces. And, by the way, momentum corresponds to the flux of causal edges through timelike hypersurfaces.

OK, so what does this mean? First, what's a spacelike hypersurface? It's actually a standard concept in general relativity, for which there's a direct analogy in our models. Basically it's what forms a slice in our foliation. Why is it called what it's called? We can identify two kinds of directions: spacelike and timelike.

A spacelike direction is one that involves just moving in space—and it's a direction where one can always reverse and go back. A timelike direction is one that involves also progressing through time—where one can't go back. We can mark spacelike (—) and timelike (--) hypersurfaces in the causal graph for our toy model:
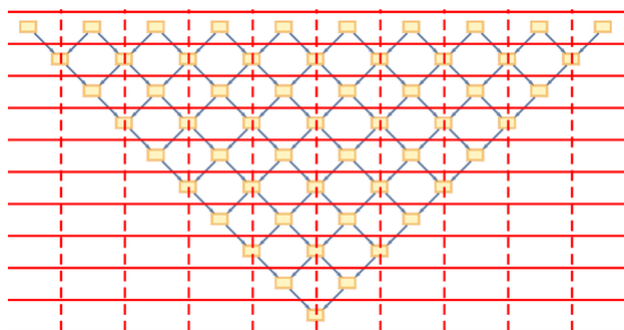
(They might be called "surfaces", except that "surfaces" are usually thought of as 2-dimensional, and our 3-space + 1-time dimensional universe, these foliation slices are 3-dimensional: hence the term "hypersurfaces".)

OK, now let's look at the picture. The "causal edges" are the causal connections between events, shown in the picture as lines joining the events. So when we talk about a "flux of causal edges through spacelike hypersurfaces", what we're talking about is the net number of causal edges that go down through the horizontal slices in the pictures.

In the toy model that's trivial to see. But here's a causal graph from a simple hypergraph model, where it's already considerably more complicated:
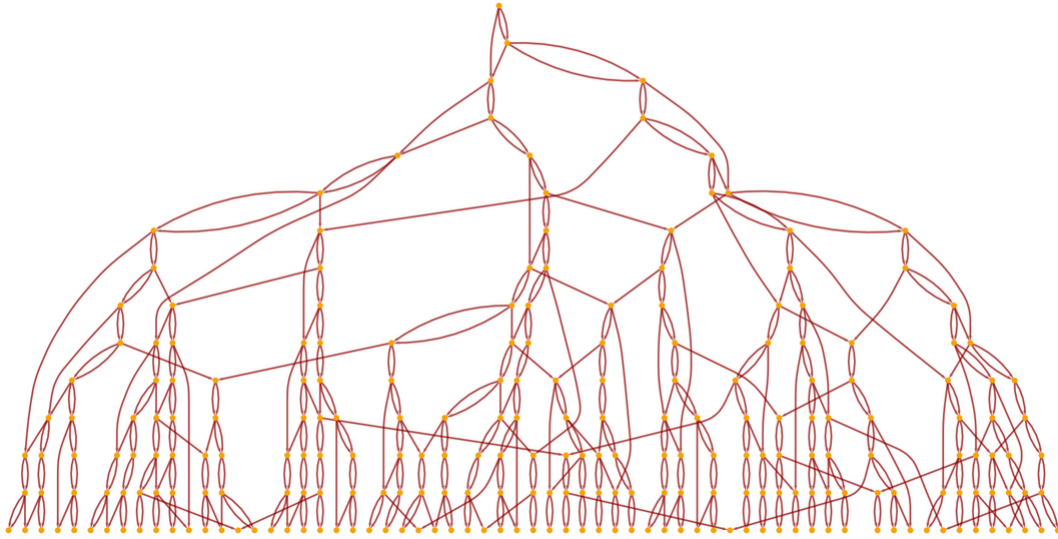
(Our toy-model causal graph starts from a line of events because we set up a long string as the initial condition; this starts from a single event because it's starting from a minimal initial condition.)

But when we put a foliation on this causal graph (thereby effectively defining our reference frame) we can start counting how many causal edges go down through successive ("spacelike") slices:
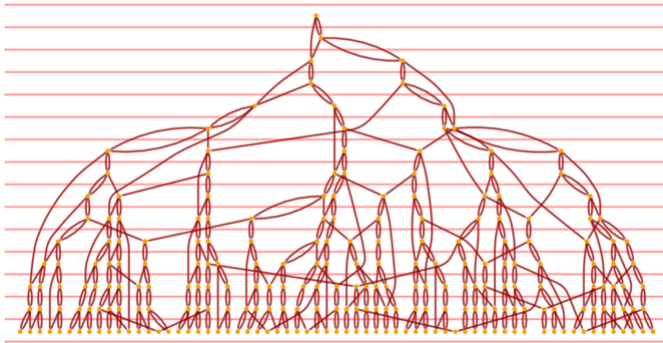
We can also ask how many causal edges go "sideways", through timelike hypersurfaces:
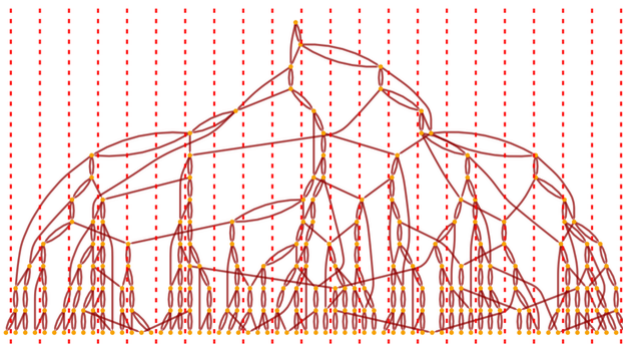
OK, so why do we think these fluxes of edges correspond to energy and momentum? Imagine what happens if we change our foliation, say tipping it to correspond to motion at

some velocity, as we did in the previous section. It takes a little bit of math, but what we find out is that our fluxes of causal edges transform with velocity basically just like we saw distance and time transform in the previous section.

In the standard derivation of relativistic mechanics, there's a consistency argument that energy has to transform with velocity like time does, and momentum like distance. But now we actually have a structural reason for this to be the case. It's a fundamental consequence of our whole setup, and of causal invariance. In traditional physics, one often says that position is the conjugate variable to momentum, and energy to time. And that's something that's burnt into the mathematical structure of the theory. But here it's not something we're burning in; it's something we're deriving from the underlying structure of our model.

And that means there's ultimately a lot more we can say about it. For example, we might wonder what the "zero of energy" is. After all, if we look at one of our causal graphs, a lot of the causal edges are really just going into "maintaining the structure of space". So if in a sense space is uniform, there's inevitably a uniform "background flux" of causal edges associated with that. And whatever we consider to be "energy" corresponds to the fluctuations of that flux around its background value.

By the way, it's worth mentioning what a "flux of causal edges" corresponds to. Each causal edge represents a causal connection between events, that is in a sense "carried" by some element in the underlying hypergraph (the "spatial hypergraph"). So a "flux of causal edges" is in effect the communication of activity (i.e. events), either in time (i.e. through spacelike hypersurfaces) or in space (i.e. through timelike hypersurfaces). And at least in some approximation we can then say that energy is associated with activity in the hypergraph that propagates information through time, while momentum is associated with activity that propagates information in space.

There's a fundamental feature of our causal graphs that we haven't mentioned yet—that's related to information propagation. Start at any point (i.e. any event) in a causal graph. Then trace the causal connections from that event. You'll get some kind of cone (here just in 2D):
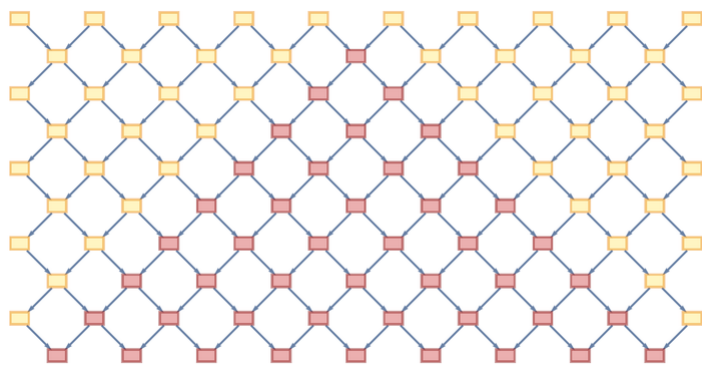
The cone is more complicated in a more complicated causal graph. But you'll always have something like it. And what it corresponds to physically is what's normally called a light cone (or "forward light cone"). Assuming we've drawn our causal network so that events are somehow laid out in space across the page, then the light cone will show how information (as transmitted by light) can spread in space with time.

When the causal graph gets complicated, the whole setup with light cones gets complicated, as we'll discuss for example in connection with black holes later. But for now, we can just say there are cones in our causal graph, and in effect the angle of these cones represents the maximum rate of information propagation in the system, which we can identify with the physical speed of light.

And in fact, not only can we identify light cones in our causal graph: in some sense we can think of our whole causal graph as just being a large number of "elementary light cones" all knitted together. And, as we mentioned, much of the structure that's built necessarily goes into, in effect, "maintaining the structure of space".

But let's look more closely at our light cones. There are causal edges on their boundaries that in effect correspond to propagation at the speed of light—and that, in terms of the underlying hypergraph, correspond to events that "reach out" in the hypergraph, and "entrain" new elements as quickly as possible. But what about causal edges that are "more vertical"? These causal edges are associated with events that in a sense reuse elements in the hypergraph, without involving new ones.

And it looks like these causal edges have an important interpretation: they are associated with mass (or, more specifically, rest mass). OK, so the total flux of causal edges through spacelike hypersurfaces corresponds to energy. And now we're saying that the flux of causal edges specifically in the timelike direction corresponds to rest mass. We can see what happens if we "tip our reference" frames just a bit, say corresponding to a velocity $v \ll c$.

Again, there's a small amount of math, but it's pretty easy to derive formulas for momentum (p) and energy (E). The speed of light c comes into the formulas because it defines the ratio of "horizontal" (i.e. spacelike) to "vertical" (i.e timelike) distances on the causal graph. And for v small compared to c we get:

$$p = m\,v + \ldots$$

$$E = m\,c^2 + \frac{1}{2}\,m\,v^2 + \ldots$$

So from these formulas we can see that just by thinking about causal graphs (and, yes, with a backdrop of causal invariance, and a whole host of detailed mathematical limit questions that we're not discussing here), we've managed to derive a basic (and famous) fact about the relation between energy and mass:

$$E = m\,c^2$$

Sometimes in the standard formalism of physics, this relation by now seems more like a definition than something to derive. But in our model, it's not just a definition, and in fact we can successfully derive it.

General Relativity & Gravity

Earlier on, we talked about how curvature of space can arise in our models. But at that point we were just talking about "empty space". Now we can go back and also talk about how curvature interacts with mass and energy in space.

In our earlier discussion, we talked about constructing spherical balls by starting at some point in the hypergraph, and then following all possible sequences of r connections. But now we can do something directly analogous in the causal graph: start at some point, and follow possible sequences of t connections. There's quite a bit of mathematical trickiness, but essentially this gets us "volumes of light cones".

If space is effectively d-dimensional, then to a first approximation this volume will grow like t^{d+1}. But like in the spatial case, there's a correction term, this time proportional to the so-

called Ricci tensor $R_{\mu\nu}$. (The actual expression is roughly $t^{d+1}\left(1 - \dfrac{1}{6}\, t_i\, t_j\, R_{ij} + \ldots\right)$ where the t_i are timelike vectors, etc.)

OK, but we also know something else about what is supposed to be inside our light cones: not only are there "background connections" that maintain the structure of space, there are also "additional" causal edges that are associated with energy, momentum and mass. And in the limit of a large causal graph, we can identify the density of these with the so-called energy-momentum tensor $T_{\mu\nu}$. So in the end we have two contributions to the "volumes" of our light cones: one from "pure curvature" and one from energy-momentum.

Again, there's some math involved. But the main thing is to think about the limit when we're looking at a very large causal graph. What needs to be true for us to have d-dimensional space, as opposed to something much wilder? This puts a constraint on the growth rates of our light cone volumes, and when one works everything out, it implies that the following equation must hold:

$$R_{\mu\nu} - \frac{1}{2}\, R\, g_{\mu\nu} = \sigma\, T_{\mu\nu}$$

But this is exactly Einstein's equation for the curvature of space when matter with a certain energy-momentum is present. We're glossing over lots of details here. But it's still, in my view, quite spectacular: from the basic structure of our very simple models, we're able to derive a fundamental result in physics: the equation that for more than a hundred years has passed every test in describing the operation of gravity.

There's a footnote here. The equation we've just given is without a so-called cosmological term. And how that works is bound up with the question of what the zero of energy is, which in our model relates to what features of the evolving hypergraph just have to do with the "maintenance of space", and what have to do with "things in space" (like matter).

In existing physics, there's an expectation that even in the "vacuum" there's actually a formally infinite density of pairs of virtual particles associated with quantum mechanics. Essentially what's happening is that there are always pairs of particles and antiparticles being created, that annihilate quickly, but that in aggregate contribute a huge effective energy density. We'll discuss how this relates to quantum mechanics in our models later.

But for now let's just recall that particles (like electrons) in our models basically correspond to locally stable structures in the hypergraph.

But when we think about how "space is maintained" it's basically through all sorts of seemingly random updating events in the hypergraph. But in existing physics (or, specifically, quantum field theory) we're basically expected to analyze everything in terms of (virtual) particles. So if we try to do that with all these random updating events, it's not surprising that we end up saying that there are these infinite collections of things going on. (Yes, this can be made much more precise; I'm just giving an outline here.)

But as soon as we say this, there is an immediate problem: we're saying that there's a formally infinite—or at least huge—energy density that must exist everywhere in the universe. But if we then apply Einstein's equation, we'll conclude that this must produce enough curvature to basically curl the universe up into a tiny ball.

One way to get out of this is to introduce a so-called cosmological term, that's just an extra term in the Einstein equations, and then posit that this term is sized so as to exactly cancel (yes, to perhaps one part in 10^{60} or more) the energy density from virtual particles. It's certainly not a pretty solution.

But in our models, the situation is quite different. It's not that we have virtual particles "in space", that are having an effect on space. It's that the same stuff that corresponds to the virtual particles is actually "making the space", and maintaining its structure. Of course, there are lots of details about this—which no doubt depend on the particular underlying rule. But the point is that there's no longer a huge mystery about why "vacuum energy" doesn't basically destroy our universe: in effect, it's because it's what's making our universe.

Black Holes, Singularities, etc.

One of the big predictions of general relativity is the existence of black holes. So how do things like that work in our models? Actually, it's rather straightforward. The defining feature of a black hole is the existence of an event horizon: a boundary that light signals can't cross, and where in effect causal connection is broken.

In our models, we can explicitly see that happen in the causal graph. Here's an example:
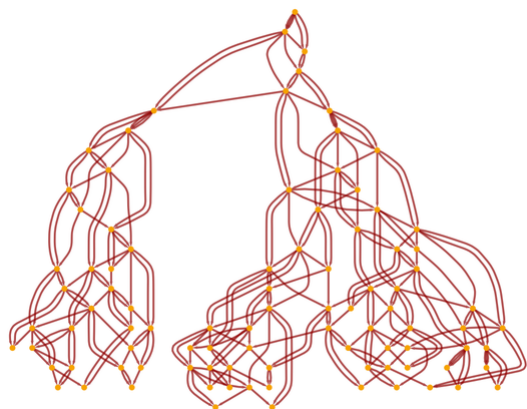
At the beginning, everything is causally connected. But at some point the causal graph splits —and there's an event horizon. Events happening on one side can't influence ones on the other, and so on. And that's how a region of the universe can "causally break off" to form something like a black hole.

But actually, in our models, the "breaking off" can be even more extreme. Not only can the causal graph split; the spatial hypergraph can actually throw off disconnected pieces—each of which in effect forms a whole "separate universe":
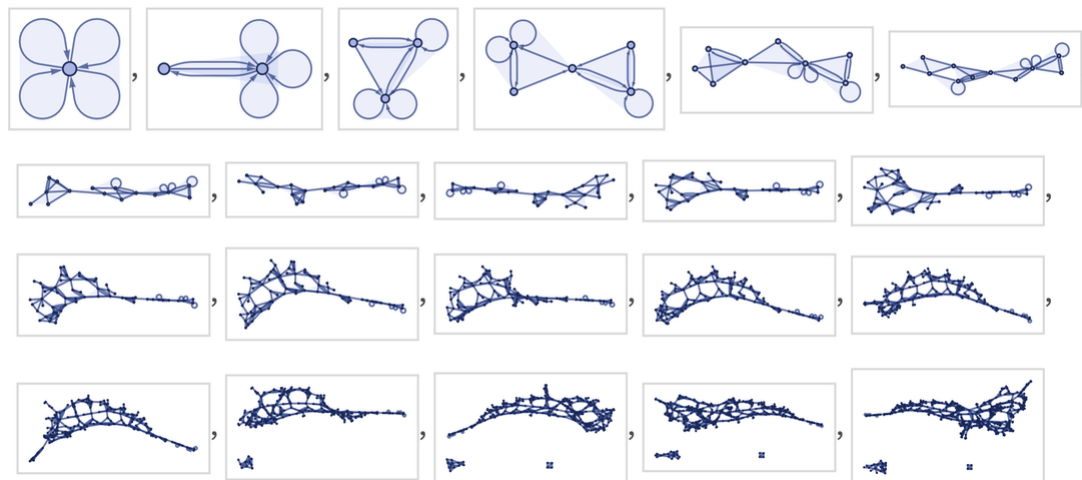
By the way, it's interesting to look at what happens to the foliations observers make when there's an event horizon. Causal invariance says that paths in the causal graph that diverge should always eventually merge. But if the paths go into different disconnected pieces of the causal graph, that can't ever happen. So how does an observer deal with that? Well, basically they have to "freeze time". They have to have a foliation where successive time slices just pile up, and never enter the disconnected pieces.

It's just like what happens in general relativity. To an observer far from the black hole, it'll seem to take an infinite time for anything to fall into the black hole. For now, this is just a phenomenon associated with the structure of space. But later we'll see that it's also the

direct analog of something completely different: the process of measurement in quantum mechanics.

Coming back to gravity: we can ask questions not only about event horizons, but also about actual singularities in spacetime. In our models, these are places where lots of paths in a causal graph converge to a single point. And in our models, we can immediately study questions like whether there's always an event horizon associated with any singularity (the "cosmic censorship hypothesis").

We can ask about other strange phenomena from general relativity. For example, there are closed timelike curves, sometimes viewed as allowing time travel. In our models, closed timelike curves are inconsistent with causal invariance. But we can certainly invent rules that produce them. Here's an example:
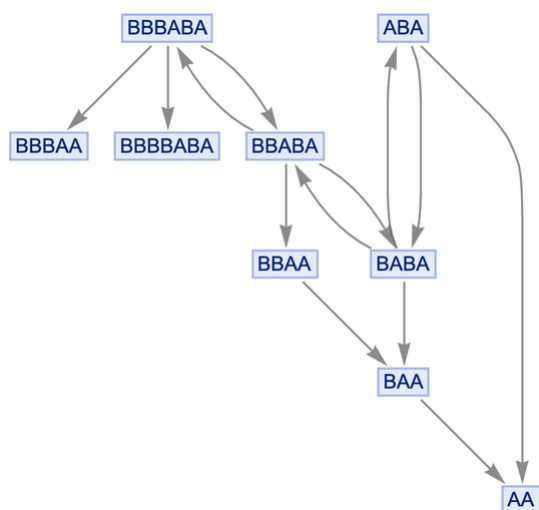
We start from one "initial" state in this multiway system. But as we go forward we can enter a loop where we repeatedly visit the same state. And this loop also occurs in the causal graph. We think we're "going forward in time". But actually we're just in a loop, repeatedly returning to the same state. And if we tried to make a foliation where we could describe time as always advancing, we just wouldn't be able to do it.

Cosmology

In our model, the universe can start as a tiny hypergraph—perhaps a single self-loop. But then—as the rule gets applied—it progressively expands. With some particularly simple rules, the total size of the hypergraph has to just uniformly increase; with others it can fluctuate.

But even if the size of the hypergraph is always increasing, that doesn't mean we'd necessarily notice. It could be that essentially everything we can see just expands too—so in effect the granularity of space is just getting finer and finer. This would be an interesting resolution to the age-old debate about whether the universe is discrete or continuous. Yes, it's structurally discrete, but the scale of discreteness relative to our scale is always getting smaller and smaller. And if this happens fast enough, we'd never be able to "see the discreteness"—because every time we tried to measure it, the universe would effectively have subdivided before we got the result. (Somehow it'd be like the ultimate calculus epsilon-delta proof: you challenge the universe with an epsilon, and before you can get the result, the universe has made a smaller delta.)

There are some other strange possibilities too. Like that the whole hypergraph for the universe is always expanding, but pieces are continually "breaking off", effectively forming black holes of different sizes, and allowing the "main component" of the universe to vary in size.

But regardless of how this kind of expansion works in our universe today, it's clear that if the universe started with a single self-loop, it had to do a lot of expanding, at least early on. And here there's an interesting possibility that's relevant for understanding cosmology.

Just because our current universe exhibits three-dimensional space, in our models there's no reason to think that the early universe necessarily also did. There are very different things that can happen in our models:
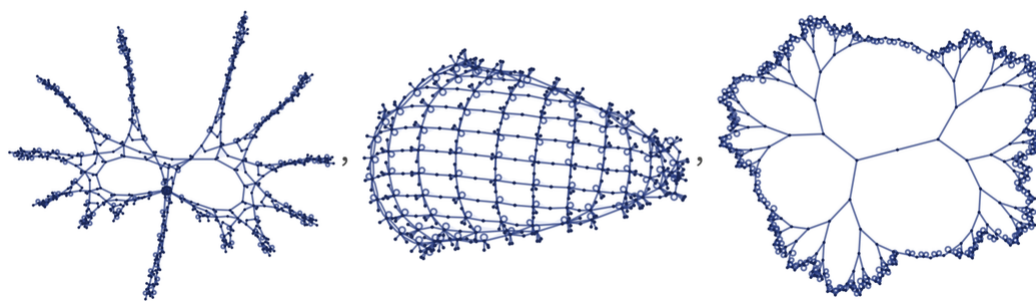
In the first example here, different parts of space effectively separate into non-communicating "black hole" tree branches. In the second example, we have something like ordinary—in this case 2-dimensional—space. But in the third example, space is in a sense very connected. If we work out the volume of a spherical ball, it won't grow like r^d; it'll grow exponentially with r (e.g. like 2^r).

If we look at the causal graph, we'll see that you can effectively "go everywhere in space", or affect every event, very quickly. It'd be as if the speed of light is infinite. But really it's because space is effectively infinite dimensional.

In typical cosmology, it's been quite mysterious how different parts of the early universe managed to "communicate" with each other, for example, to smooth out perturbations. But if the universe starts effectively infinite-dimensional, and only later "relaxes" to being finite-dimensional, that's no longer a mystery.

So, OK, what might we see in the universe today that would reflect what happened extremely early in its history? The fact that our models deterministically generate behavior that seems for all practical purposes random means that we can expect that most features of the initial conditions or very early stages of the universe will quickly be "encrypted", and effectively not reconstructable.

But it's just conceivable that something like a breaking of symmetry associated with the first few hypergraphs might somehow survive. And that suggests the bizarre possibility that— just maybe—something like the angular structure of the cosmic microwave background or the very large-scale distribution of galaxies might reflect the discrete structure of the very early universe. Or, in other words, it's just conceivable that what amounts to the rule for the universe is, in effect, painted across the whole sky. I think this is extremely unlikely, but it'd certainly be an amazing thing if the universe were "self-documenting" that way.

Elementary Particles—Old and New

We've talked several times about particles like electrons. In current physics theories, the various (truly) elementary particles—the quarks, the leptons (electron, muon, neutrinos, etc.), the gauge bosons, the Higgs—are all assumed to intrinsically be point particles, of zero size. In our models, that's not how it works. The particles are all effectively "little lumps of space" that have various special properties.

My guess is that the precise list of what particles exist will be something that's specific to a particular underlying rule. In cellular automata, for example, we're used to seeing complicated sets of possible localized structures arise:
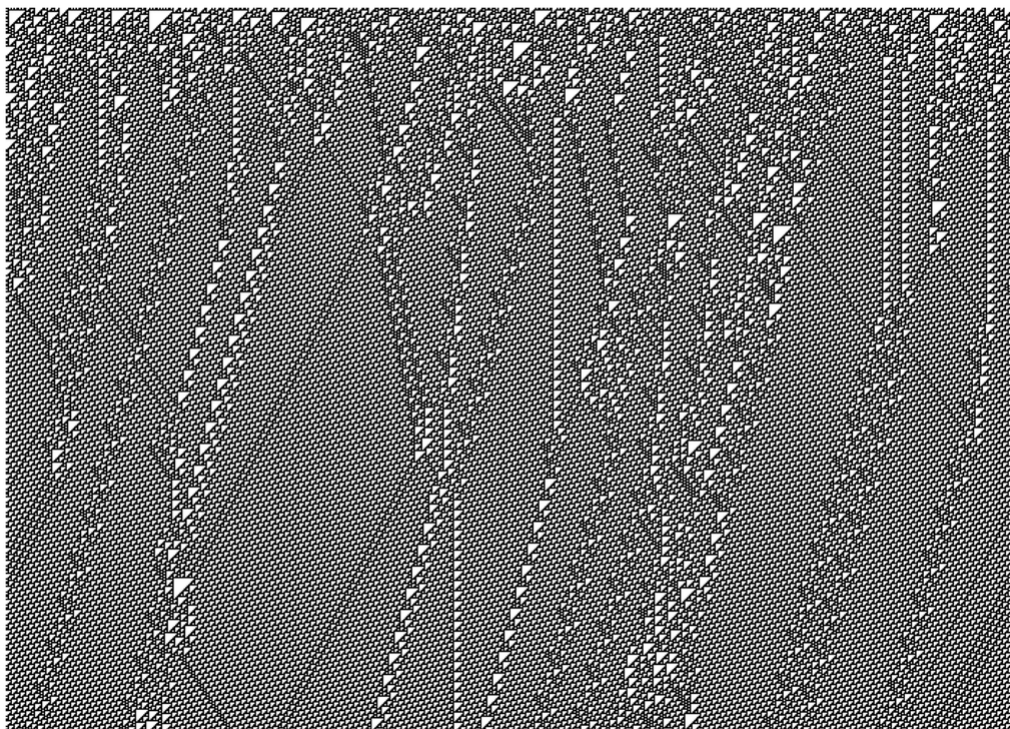
In our hypergraphs, the picture will inevitably be somewhat different. The "core feature" of each particle will be some kind of locally stable structure in the hypergraph (a simple analogy might be that it's a lump of nonplanarity in an otherwise planar graph). But then there'll be lots of causal edges associated with the particle, defining its particular energy and momentum.

Still, the "core feature" of the particles will presumably define things like their charge, quantum numbers, and perhaps spin—and the fact that these things are observed to occur in discrete units may reflect the fact that it's a small piece of hypergraph that's involved in defining them.

It's not easy to know what the actual scale of discreteness in space might be in our models. But a possible (though potentially unreliable) estimate might be that the "elementary length" is around 10^{-93} meters. (Note that that's very small compared to the Planck length $\sim 10^{-35}$ meters that arises essentially from dimensional analysis.) And with this elementary length, the radius of the electron might be 10^{-81} meters. Tiny, but not zero. (Note that current experiments only tell us that the size of the electron is less than about 10^{-22} meters.)

One feature of our models is that there should be a "quantum of mass"—a discrete amount that all masses, for example of particles, are multiples of. With our estimate for the

elementary length, this quantum of mass would be small, perhaps 10^{-30}, or 10^{36} times smaller than the mass of the electron.

And this raises an intriguing possibility. Perhaps the particles—like electrons—that we currently know about are the "big ones". (With our estimates, an electron would have 10^{35} hypergraph elements in it.) And maybe there are some much smaller, and much lighter ones. At least relative to the particles we currently know, such particles would have few hypergraph elements in them—so I'm referring to them as "oligons" (after the Greek word ὀλίγος for "few").

What properties would these oligons have? They'd probably interact very very weakly with other things in the universe. Most likely lots of oligons would have been produced in the very early universe, but with their very weak interactions, they'd soon "drop out of thermal equilibrium", and be left in large numbers as relics—with energies that become progressively lower as the universe expands around them.

So where might oligons be now? Even though their other interactions would likely be exceptionally weak, they'd still be subject to gravity. And if their energies end up being low enough, they'd basically collect in gravity wells around the universe—which means in and around galaxies.

And that's interesting—because right now there's quite a mystery about the amount of mass seen in galaxies. There appears to be a lot of "dark matter" that we can't see but that has gravitational effects. Well, maybe it's oligons. Maybe even lots of different kinds of oligons: a whole shadow physics of much lighter particles.

The Inevitability of Quantum Mechanics

"But how will you ever get quantum mechanics?", physicists would always ask me when I would describe earlier versions of my models. In many ways, quantum mechanics is the pinnacle of existing physics. It's always had a certain "you-are-not-expected-to-understand-this" air, though, coupled with "just-trust-the-mathematical-formalism". And, yes, the mathematical formalism has worked well—really well—in letting us calculate things. (And it almost seems more satisfying because the calculations are often so hard; indeed, hard enough that they're what first made me start using computers to do mathematics 45 years ago.)

Our usual impression of the world is that definite things happen. And before quantum mechanics, classical physics typically captured this in laws—usually equations—that would tell one what specifically a system would do. But in quantum mechanics the formalism involves any particular system doing lots of different things "in parallel", with us just seeing samples—ultimately with certain probabilities—of these possibilities.

And as soon as one hears of a model in which there are definite rules, one might assume that it could never reproduce quantum mechanics. But, actually, in our models, quantum mechanics is not just possible; it's absolutely inevitable. And, as we'll see, in something I consider quite beautiful, the core of what leads to it turns out to be the same as what leads to relativity.

OK, so how does this work? Let's go back to what we discussed when we first started talking about time. In our models there's a definite rule for updates to make in our hypergraphs, say:
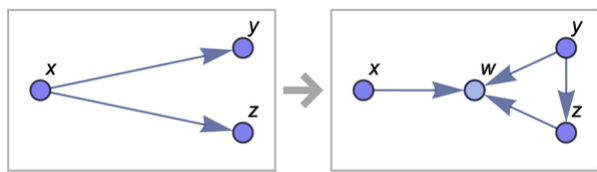
But if we've got a hypergraph like this:
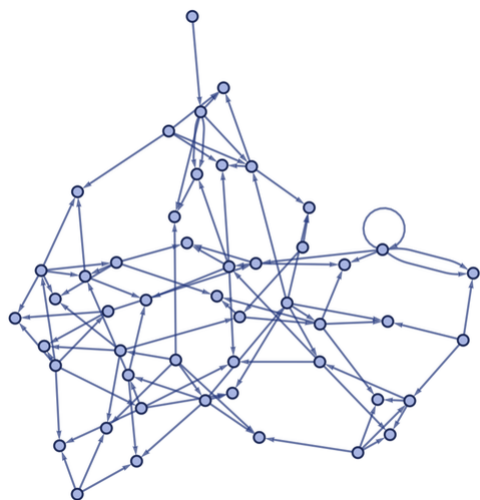
there will usually be many places where this rule can be applied. So which update should we do first? The model doesn't tell us. But let's just imagine all the possibilities. The rule tells us what they all are—and we can represent them (as we discussed above) as a multiway system—here illustrated using the simpler case of strings rather than hypergraphs:
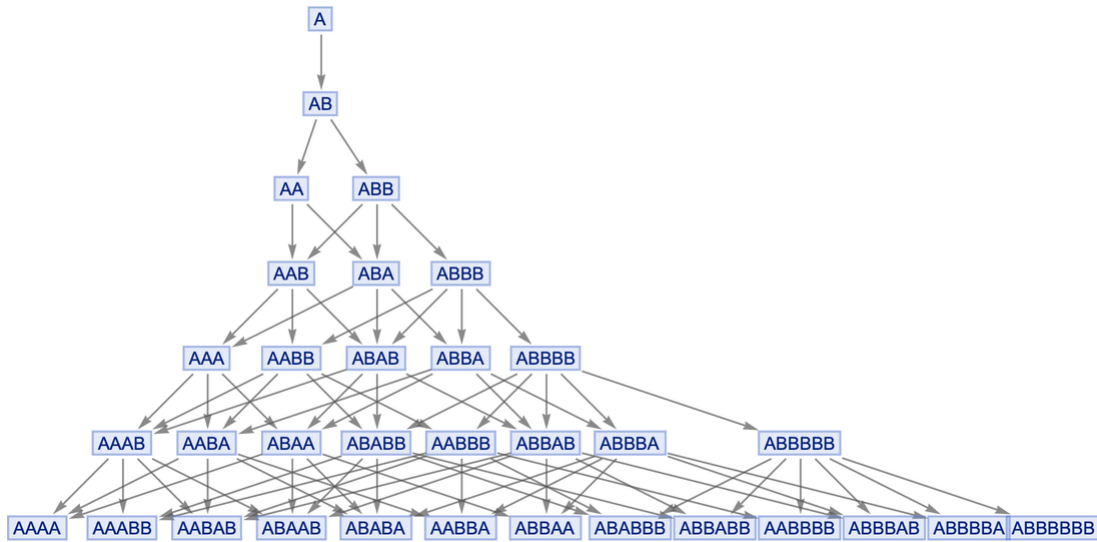
Each node in this graph now represents a complete state of our system (a hypergraph in our actual models). And each node is joined by arrows to the state or states that one gets by applying a single update to it.

If our model had been operating "like classical physics" we would expect it to progress in time from one state to another, say like this:

But the crucial point is that the structure of our models leaves us no choice but to consider multiway systems. The form of the whole multiway system is completely determined by the rules. But—in a way that is already quite reminiscent of the standard formalism of quantum mechanics—the multiway system defines many different possible paths of history.

But now there is a mystery. If there are always all these different possible paths of history, how is it that we ever think that definite things happen in the world? This has been a core mystery of quantum mechanics for a century. It turns out that if one's just using quantum mechanics to do calculations, the answer basically doesn't matter. But if one wants to "really understand what's going on" in quantum mechanics, it's something that definitely does matter.

And the exciting thing is that in our models, there's an obvious resolution. And actually it's based on the exact same phenomenon—causal invariance—that gives us relativity.

Here's roughly how this works. The key point is to think about what an observer who is themselves part of the multiway system will conclude about the world. Yes, there are different possible paths of history. But—just as in our discussion of relativity—the only

aspect of them that an observer will ever be aware of is the causal relationships between the events they involve. But the point is that—even though when looked at from "outside" the paths are different—causal invariance implies that the network of relationships between causal events (which is all that's relevant when one's inside the system) will always be exactly the same.

In other words—much as in the case of relativity—even though from outside the system there may seem to be many possible "threads of time", from inside the system causal invariance implies that there's in a sense ultimately just one thread of time, or, in effect, one objective reality.

How does this all relate to the detailed standard formalism of quantum mechanics? It's a little complicated. But let me make at least a few comments here. (There's some more detail in my technical document; Jonathan Gorard has given even more.)

The states in the multiway system can be thought of as possible states of the quantum system. But how do we characterize how observers experience them? In particular, which states is the observer aware of when? Just like in the relativity case, the observer can in a sense make a choice of how they define time. One possibility might be through a foliation of the multiway system like this:
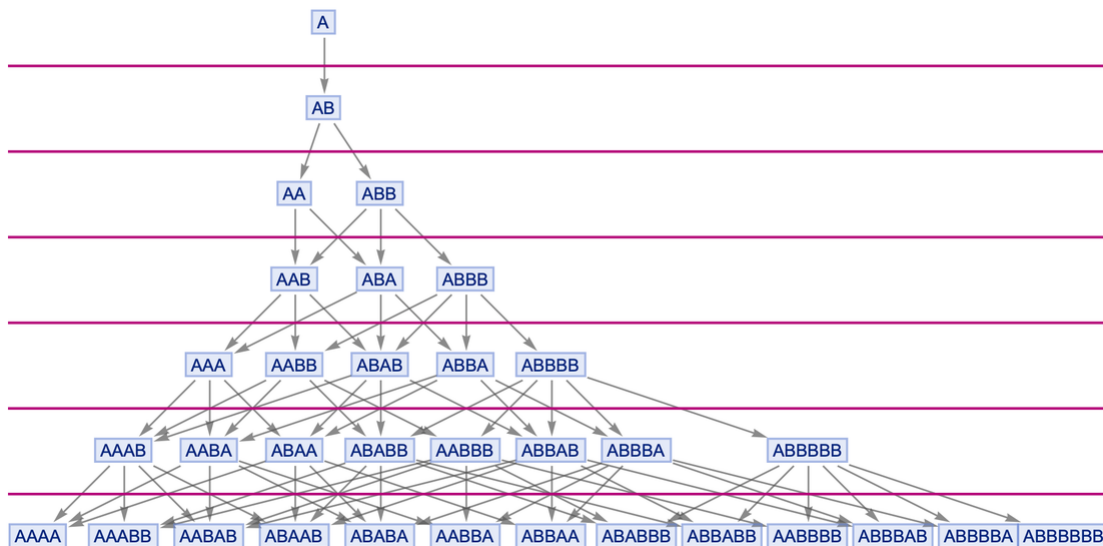
In the formalism of quantum mechanics, one can then say that at each time, the observer experiences a superposition of possible states of the system. But now there's a critical point. In direct analogy to the case of relativity, there are many different possible choices the observer can make about how to define time—and each of them corresponds to a different foliation of the multiway graph.

Again by analogy to relativity, we can then think of these choices as what we can call different "quantum observation frames". Causal invariance implies that as long they respect the causal relationships in the graph, these frames can basically be set up in any way we want. In talking about relativity, it was useful to just have "tipped parallel lines" ("inertial frames") representing observers who are moving uniformly in space.

In talking about quantum mechanics, other frames are useful. In particular, in the standard formalism of quantum mechanics, it's common to talk about "quantum measurement": essentially the act of taking a quantum system and determining some definite (essentially classical) outcome from it. Well, in our setup, a quantum measurement basically corresponds to a particular quantum observation frame.

Here's an example:
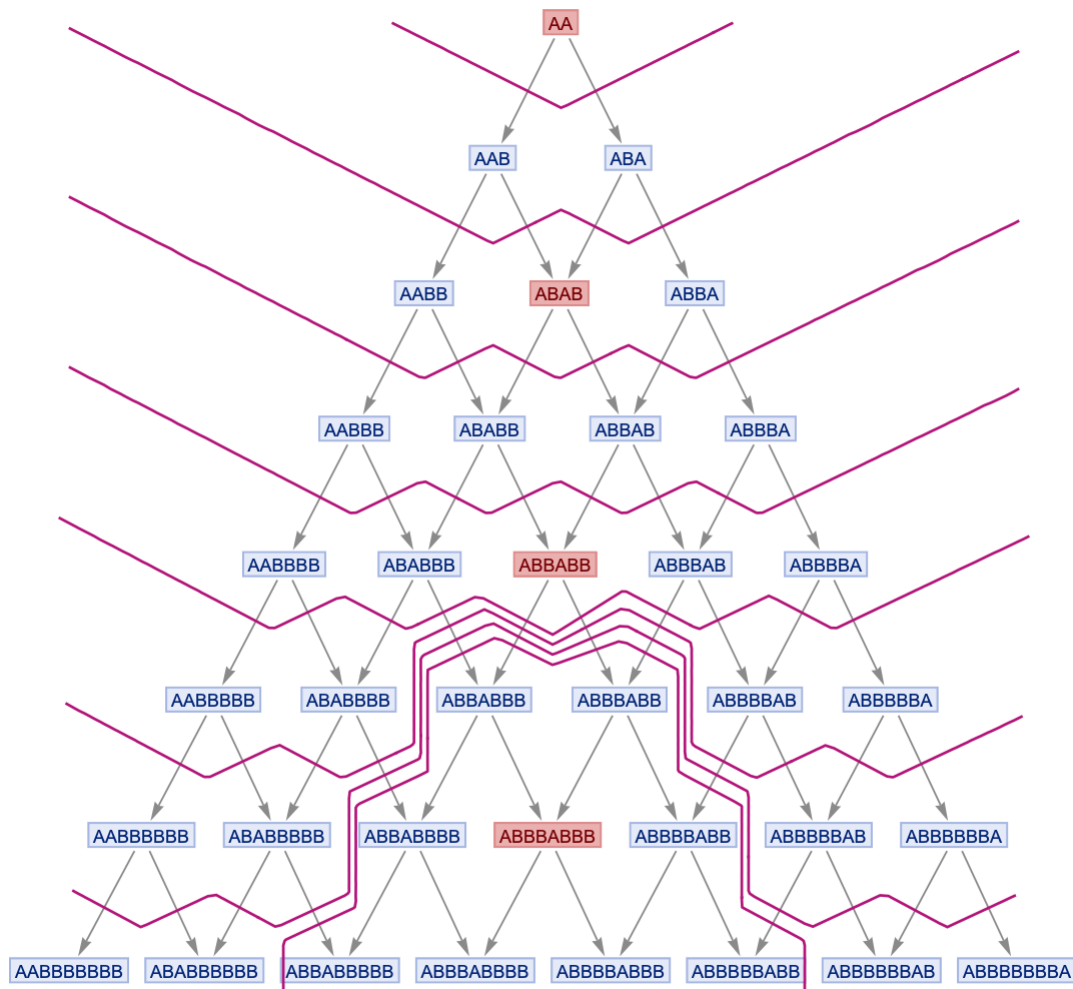
The successive pink lines effectively mark off what the observer is considering to be successive moments in time. So when all the lines bunch up below the state ABBABB what it means is that the observer is effectively choosing to "freeze time" for that state. In other

words, the observer is saying "that's the state I consider the system to be in, and I'm sticking to it". Or, put another way, even though in the full multiway graph there's all sorts of other "quantum mechanical" evolution of states going on, the observer has set up their quantum observation frame so that they pick out just a particular, definite, classical-like outcome.

OK, but can they consistently do that? Well, that depends on the actual underlying structure of the multiway graph, which ultimately depends on the actual underlying rule. In the example above, we've set up a foliation (i.e. a quantum observation frame) that does the best possible job in this rule at "freezing time" for the ABBABB state. But just how long can this "reality distortion field" be maintained?

The only way to keep the foliation consistent in the multiway graph above is to have it progressively expand over time. In other words, to keep time frozen, more and more quantum states have to be pulled into the "reality distortion field", and so there's less and less coherence in the system.

The picture above is for a very trivial rule. Here's a corresponding picture for a slightly more realistic case:
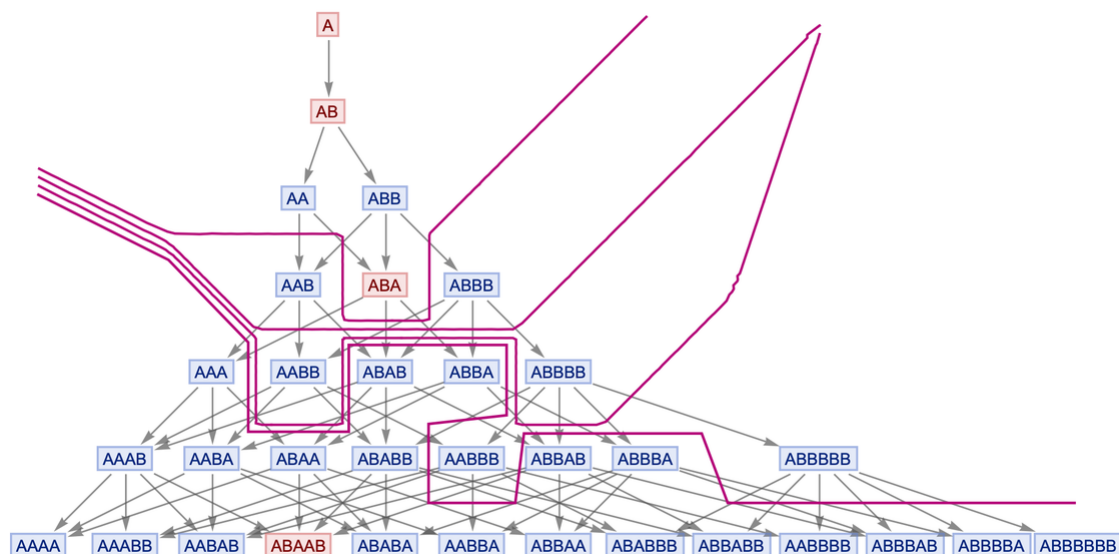
And what we see here is that—even in this still incredibly simplified case—the structure of the multiway system will force the observer to construct a more and more elaborate foliation if they are to successfully freeze time. Measurement in quantum mechanics has always involved a slightly uncomfortable mathematical idealization—and this now gives us a sense of what's really going on. (The situation is ultimately very similar to the problem of decoding "encrypted" thermodynamic initial conditions that I mentioned above.)

Quantum measurement is really about what an observer perceives. But if you are for example trying to construct a quantum computer, it's not just a question of having a qubit be perceived as being maintained in a particular state; it actually has to be maintained in that state. And for this to be the case we actually have to freeze time for that qubit. But here's a very simplified example of how that can happen in a multiway graph:
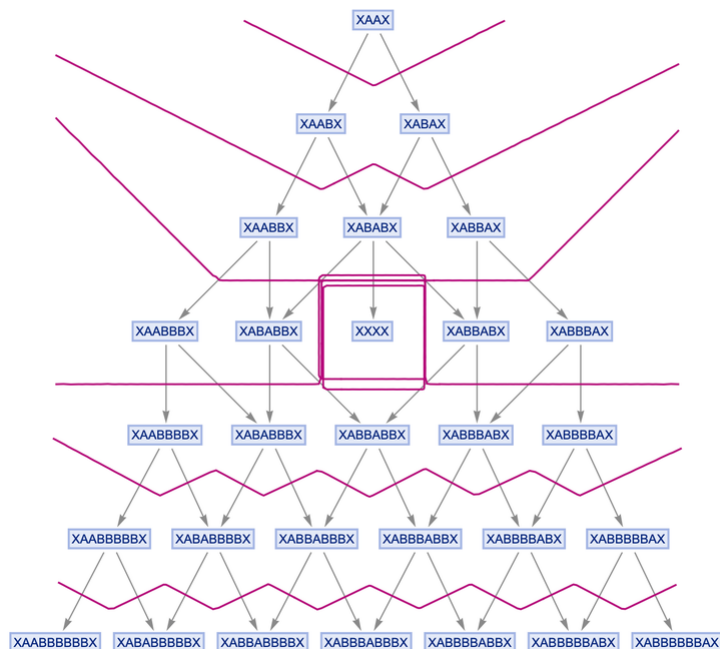
All this discussion of "freezing time" might seem weird, and not like anything one usually talks about in physics. But actually, there's a wonderful connection: the freezing of time we're talking about here can be thought of as happening because we've got the analog in the space of quantum states of a black hole in physical space.

The picture above makes it plausible that we've got something where things can go in, but if they do, they always get stuck. But there's more to it. If you're an observer far from a black hole, then you'll never actually see anything fall into the black hole in finite time (that's why black holes are called "frozen stars" in Russian). And the reason for this is precisely because (according to the mathematics) time is frozen at the event horizon of the black hole. In other words, to successfully make a qubit, you effectively have to isolate it in quantum space like things get isolated in physical space by the presence of the event horizon of a black hole.

General Relativity and Quantum Mechanics Are the Same Idea!

General relativity and quantum mechanics are the two great foundational theories of current physics. And in the past it's often been a struggle to reconcile them. But one of the

beautiful outcomes of our project so far has been the realization that at some deep level general relativity and quantum mechanics are actually the same idea. It's something that (at least so far) is only clear in the context of our models. But the basic point is that both theories are consequences of causal invariance—just applied in different situations.

Recall our discussion of causal graphs in the context of relativity above. We drew foliations and said that if we looked at a particular slice, it would tell us the arrangement of the system in space at what we consider to be a particular time. So now let's look at multiway graphs. We saw in the previous section that in quantum mechanics we're interested in foliations of these. But if we look at a particular slice in one of these foliations, what does it represent? The foliation has got a bunch of states in it. And it turns out that we can think of them as being laid out in an abstract kind of space that we're calling "branchial space".

To make sense of this space, we have to have a way to say what's near what. But actually the multiway graph gives us that. Take a look at this multiway graph:
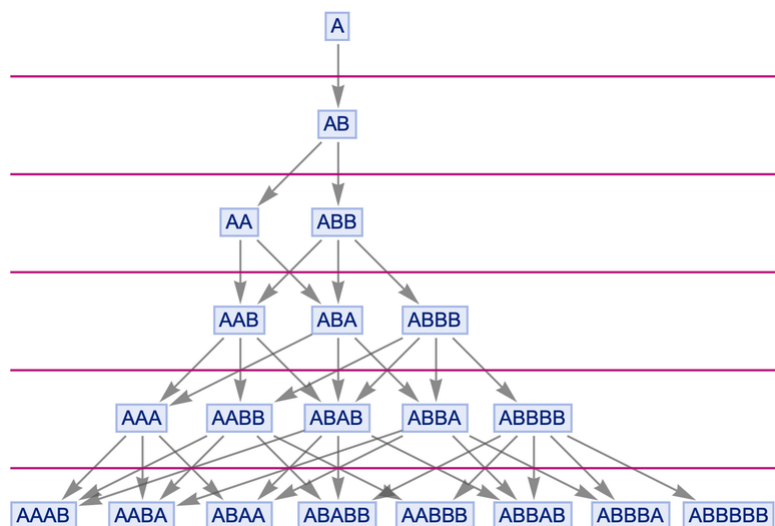
At each slice in the foliation, let's draw a graph where we connect two states whenever they're both part of the same "branch pair", so that—like AA and ABB here—they both come from the same state on the slice before. Here are the graphs we get by doing this for successive slices:
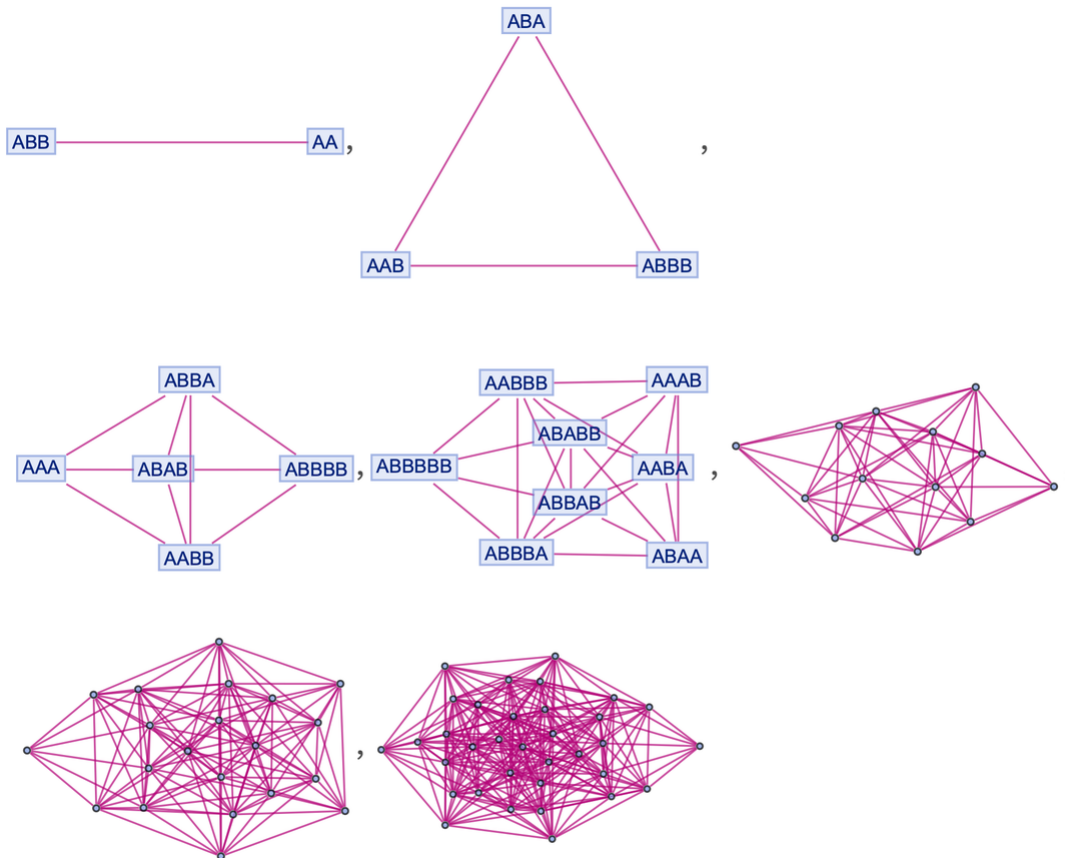
We call these branchial graphs. And we can think of them as representing the correlation— or entanglement—of quantum states. Two states that are nearby in the graph are highly entangled; those further away, less so. And we can imagine that as our system evolves, we'll get larger and larger branchial graphs, until eventually, just like for our original hypergraphs, we can think of these graphs as limiting to something like a continuous space.

But what is this space like? For our original hypergraphs, we imagined that we'd get something like ordinary physical space (say close to three-dimensional Euclidean space). But branchial space is something more abstract—and much wilder. And typically it won't even be finite-dimensional. (It might approximate a projective Hilbert space.) But we can still think of it mathematically as some kind of space.

OK, things are getting fairly complicated here. But let me try to give at least a flavor of how things work. Here's an example of a wonderful correspondence: curvature in physical space is like the uncertainty principle of quantum mechanics. Why do these have anything to do with each other?

The uncertainty principle says that if you measure, say, the position of something, then its momentum, you'll get a different answer than if you do it in the opposite order. But now

think about what happens when you try to make a rectangle in physical space by going in direction x first, then y, and then you do these in the opposite order. In a flat space, you'll get to the same place. But in a curved space, you won't:
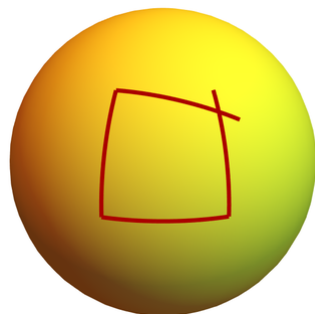
And essentially what's happening in the uncertainty principle is that you're doing exactly this, but in branchial space, rather than physical space. And it's because branchial space is wild—and effectively very curved—that you get the uncertainty principle.

Alright, so the next question might be: what's the analog of the Einstein equations in branchial space? And again, it's quite wonderful: at least in some sense, the answer is that it's the path integral—the fundamental mathematical construct of modern quantum mechanics and quantum field theory.

This is again somewhat complicated. But let me try to give a flavor of it. Just as we discussed geodesics as describing paths traversed through physical space in the course of time, so also we can discuss geodesics as describing paths traversed through branchial space in the course of time. In both cases these geodesics are determined by curvature in the corresponding space. In the case of physical space, we argued (roughly) that the presence of excess causal edges—corresponding to energy—would lead to what amounts to curvature in the spatial hypergraph, as described by Einstein's equations.

OK, so what about branchial space? Just like for the spatial hypergraph, we can think about the causal connections between the updating events that define the branchial graph. And we can once again imagine identifying the flux of causal edges—now not through spacelike hypersurfaces, but through branchlike ones—as corresponding to energy. And—much like in the spatial hypergraph case—an excess of these causal edges will have the effect of producing what amounts to curvature in branchial space (or, more strictly, in branchtime—the analog of spacetime). But this curvature will then affect the geodesics that traverse branchial space.

In general relativity, the presence of mass (or energy) causes curvature in space which causes the paths of geodesics to turn—which is what is normally interpreted as the action of the force of gravity. But now we have an analog in quantum mechanics, in our branchial space. The presence of energy effectively causes curvature in branchial space which causes the paths of geodesics through branchial space to turn.

What does turning correspond to? Basically it's exactly what the path integral talks about. The path integral (and the usual formalism of quantum mechanics) is set up in terms of complex numbers. But it can just as well be thought of in terms of turning through an angle. And that's exactly what's happening with our geodesics in branchial space. In the path integral there's a quantity called the action—which is a kind of relativistic analog of energy —and when one [works things out more carefully](#), our fluxes of causal edges correspond to the action, but are also exactly what determine the rate of turning of geodesics.

It all fits together beautifully. In physical space we have Einstein's equations—the core of general relativity. And in branchial space (or, more accurately, multiway space) we have Feynman's path integral—the core of modern quantum mechanics. And in the context of our models they're just different facets of the same idea. It's an amazing unification that I have to say I didn't see coming; it's something that just emerged as an inevitable consequence of our simple models of applying rules to collections of relations, or hypergraphs.

Branchial Motion and the Entanglement Horizon

We can think of motion in physical space as like the process of exploring new elements in the spatial hypergraph, and potentially becoming affected by them. But now that we're talking about branchial space, it's natural to ask whether there's something like motion there too. And the answer is that there is. And it's basically exactly the same kind of thing: but instead of exploring new elements in the spatial hypergraph, we're exploring new elements in the branchial graph, and potentially becoming affected by them.

There's a way of talking about it in the standard language of quantum mechanics: as we move in branchial space, we're effectively getting "entangled" with more and more quantum states.

OK, so let's take the analogy further. In physical space, there's a maximum speed of motion —the speed of light, c. So what about in branchial space? Well, in our models we can see

that there's also got to be a maximum speed of motion in branchial space. Or, in other words, there's a maximum rate at which we can entangle with new quantum states.

In physical space we talk about light cones as being the regions that can be causally affected by some event at a particular location in space. In the same way, we can talk about entanglement cones that define regions in branchial space that can be affected by events at some position in branchial space. And just as there's a causal graph that effectively knits together elementary light cones, there's something similar that knits together entanglement cones.

That something similar is the multiway causal graph: a graph that represents causal relationships between all events that can happen anywhere in a multiway system. Here's an example of a multiway causal graph for just a few steps of a very simple string substitution system—and it's already pretty complicated:
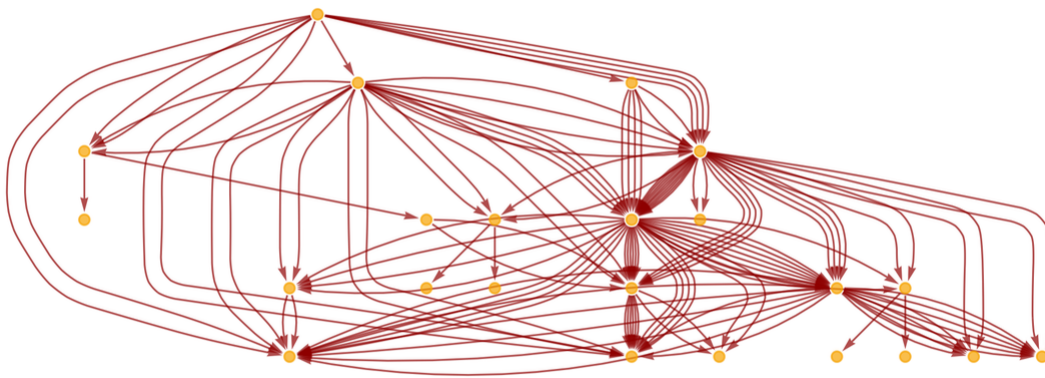
But in a sense the multiway causal graph is the most complete description of everything that can affect the experience of observers. Some of the causal relationships it describes represent spacelike connections; some represent branchlike connections. But all of them are there. And so in a sense the multiway causal graph is where relativity and quantum mechanics come together. Slice one way and you'll see relationships in physical space; slice another way and you'll see relationships in branchial space, between quantum states.

To help see how this works here's a very toy version of a multiway causal graph:
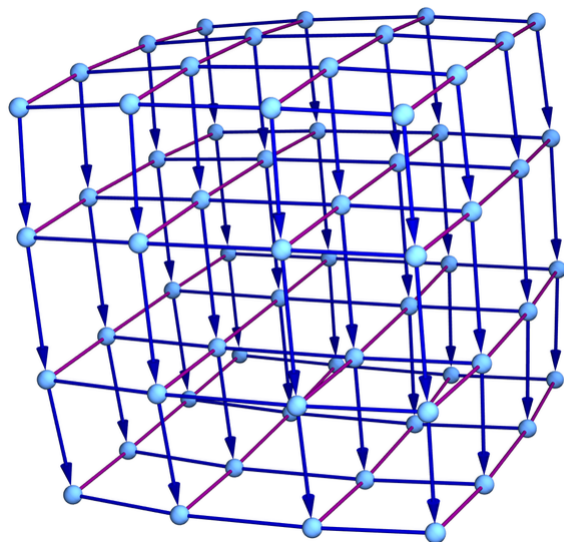
Each point is an event that happens in some hypergraph on some branch of a multiway system. And now the graph records the causal relationship of that event to other ones. In this toy example, there are purely timelike relationships—indicated by arrows pointing down—in which basically some element of the hypergraph is affecting its future self. But then there are both spacelike and branchlike relationships, where the event affects elements that are either "spatially" separated in the hypergraph, or "branchially" separated in the multiway system.

But in all this complexity, there's something wonderful that happens. As soon as the underlying rule has causal invariance, this implies all sorts of regularities in the multiway causal graph. And for example it tells us that all those causal graphs we get by taking different branchtime slices are actually the same when we project them into spacetime—and this is what leads to relativity.

But causal invariance has other consequences too. One of them is that there should be an analog of special relativity that applies not in spacetime but in branchtime. The reference frames of special relativity are now our quantum observation frames. And the analog of speed in physical space is the rate of entangling new quantum states.

So what about a phenomenon like relativistic time dilation? Is there an analog of that for motion in branchial space? Well, actually, yes there is. And it turns out to be what's sometimes called the quantum Zeno effect: if you repeatedly measure a quantum system fast enough it won't change. It's a phenomenon that's implied by the add-ons to the standard formalism of quantum mechanics that describe measurement. But in our models it just comes directly from the analogy between branchial and physical space.

Doing new measurements is equivalent to getting entangled with new quantum states—or to moving in branchial space. And in direct analogy to what happens in special relativity, as you get closer to moving at the maximum speed you inevitably sample things more slowly in time—and so you get time dilation, which means that your "quantum evolution" slows down.

OK, so there are relativistic phenomena in physical space, and quantum analogs in branchial space. But in our models these are all effectively facets of one thing: the multiway causal graph. So are there situations in which the two kinds of phenomena can mix? Normally there aren't: relativistic phenomena involve large physical scales; quantum phenomena tend to involve small ones.

But one example of an extreme situation where they can mix is black holes. I've mentioned several times that the formation of an event horizon around a black hole is associated with disconnection in the causal graph. But it's more than that. It's actually disconnection not only in the spacetime causal graph, but in the full multiway causal graph. And that means that there's not only an ordinary causal event horizon—in physical space—but also an "entanglement horizon" in branchial space. And just as a piece of the spatial hypergraph can get disconnected when there's a black hole, so can a piece of the branchial graph.

What does this mean? There are a variety of consequences. One of them is that quantum information can be trapped inside the entanglement horizon even when it hasn't crossed the causal event horizon—so that in effect the black hole is freezing quantum information "at its surface" (at least its surface in branchial space). It's a weird phenomenon implied by our models, but what's perhaps particularly interesting about it is that it's very much aligned with conclusions about black holes that have emerged in some of the latest work in physics on the so-called holographic principle in quantum field theory and general relativity.

Here's another related, weird phenomenon. If you pass the causal event horizon of a black hole, it's an inevitable fact that you'll eventually get infinitely physically elongated (or "spaghettified") by tidal forces. Well, something similar happens if you pass the entanglement horizon—except now you'll get elongated in branchial space rather than physical space. And in our models, this eventually means you won't be able to make a quantum measurement—so in a sense as an observer you won't be able to "form a classical thought", or, in other words, beyond the entanglement horizon you'll never be able to

"come to a definite conclusion" about, for example, whether something fell into the black hole or didn't.

The speed of light c is a fundamental physical constant that relates distance in physical space to time. In our models, there's now a new fundamental physical constant: the maximum entanglement speed, that relates distance in branchial space to time. I call this maximum entanglement speed ζ (zeta) (ζ looks a bit like a "tangled c"). I'm not sure what its value is, but a possible estimate is that it corresponds to entangling about 10^{102} new quantum states per second. And in a sense the fact that this is so big is why we're normally able to "form classical thoughts".

Because of the relation between (multiway) causal edges and energy, it's possible to convert ζ to units of energy per second, and our estimate then implies that ζ is about 10^5 solar masses per second. It's a big value, although conceivably not irrelevant to something like a merger of galactic black holes. (And, yes, this would mean that for an intelligence to "quantum grok" our galaxy would take maybe six months.)

Finding the Ultimate Rule

I'm frankly amazed at how much we've been able to figure out just from the general structure of our models. But to get a final fundamental theory of physics we've still got to find a specific rule. A rule that gives us 3 (or so) dimensions of space, the particular expansion rate of the universe, the particular masses and properties of elementary particles, and so on. But how should we set about finding this rule?

And actually even before that, we need to ask: if we had the right rule, would we even know it? As I mentioned earlier, there's potentially a big problem here with computational irreducibility. Because whatever the underlying rule is, our actual universe has applied it perhaps 10^{500} times. And if there's computational irreducibility—as there inevitably will be —then there won't be a way to fundamentally reduce the amount of computational effort that's needed to determine the outcome of all these rule applications.

But what we have to hope is that somehow—even though the complete evolution of the universe is computationally irreducible—there are still enough "tunnels of computational reducibility" that we'll be able to figure out at least what's needed to be able to compare with what we know in physics, without having to do all that computational work. And I have

to say that our recent success in getting conclusions just from the general structure of our models makes me much more optimistic about this possibility.

But, OK, so what rules should we consider? The traditional approach in natural science (at least over the past few centuries) has tended to be: start from what you know about whatever system you're studying, then try to "reverse engineer" what its rules are. But in our models there's in a sense too much emergence for this to work. Look at something like this:
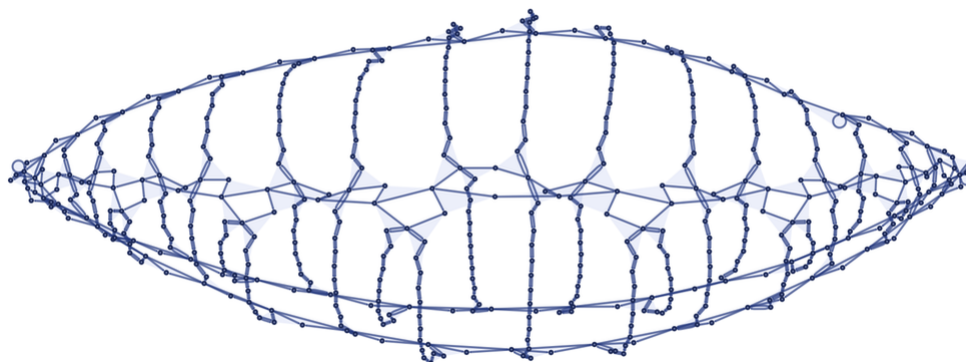
Given the overall form of this structure, would you ever figure that it could be produced just by the rule:

$$\{\{x, y, y\}, \{y, z, u\}\} \rightarrow \{\{u, z, z\}, \{u, x, v\}, \{y, u, v\}\}$$

Having myself explored the computational universe of simple programs for some forty years, I have to say that even now it's amazing how often I'm humbled by the ability of extremely simple rules to give behavior I never expected. And this is particularly common with the very structureless models we're using here. So in the end the only real way to find out what can happen in these models is just to enumerate possible rules, and then run them and see what they do.

But now there's a crucial question. If we just start enumerating very simple rules, how far are we going to have to go before we find our universe? Or, put another way, just how simple is the rule for our universe going to end up being?

It could have been that in a sense the rule for the universe would have a special case in it for every element of the universe—every particle, every position in space, etc. But the very fact

that we've been able to find definite scientific laws—and that systematic physics has even been possible—suggests that the rule at least doesn't have that level of complexity. But how simple might it be? We don't know. And I have to say that I don't think our recent discoveries shed any particular light on this—because they basically say that lots of things in physics are generic, and independent of the specifics of the underlying rule, however simple or complex it may be.

Why This Universe? The Relativity of Rules

But, OK, let's say we find that our universe can be described by some particular rule. Then the obvious immediate question would be: why that rule, and not another? The history of science—certainly since Copernicus—has shown us over and over again evidence that we're "not special". But if the rule we find to describe our universe is simple, wouldn't that simplicity be a sign of "specialness"?

I have long wondered about this. Could it for example be that the rule is only simple because of the way that we, as entities existing in our particular universe, choose to set up our ways of describing things? And that in some other universe, with some other rule, the entities that exist there would set up their ways of describing things so that the rule for their universe is simple to them, even though it might be very complex to us?

Or could it be that in some fundamental sense it doesn't matter what the rules for the universe are: that to observers embedded in a universe, operating according to the same rules as that universe, the conclusions about how the universe works will always be the same?

Or could it be that this is a kind of question that's just outside the realm of science?

To my considerable surprise, the paradigm that's emerging from our recent discoveries potentially seems to suggest a definite—though at first seemingly bizarre—scientific answer.

In what we've discussed so far we're imagining that there's a particular, single rule for our universe, that gets applied over and over again, effectively in all possible ways. But what if there wasn't just one rule that could be used? What if all conceivable rules could be used? What if every updating event could just use any possible rule? (Notice that in a finite universe, there are only ever finitely many rules that can ever apply.)

At first it might not seem as if this setup would ever lead to anything definite. But imagine making a multiway graph of absolutely everything that can happen—including all events for all possible rules. This is a big, complicated object. But far from being structureless, it's full of all kinds of structure.

And there's one very important thing about it: it's basically guaranteed to have causal invariance (basically because if there's a rule that does something, there's always another rule somewhere that can undo it).

So now we can make a rule-space multiway causal graph—which will show a rule-space analog of relativity. And what this means is that in the rule-space multiway graph, we can expect to make different foliations, but have them all give consistent results.

It's a remarkable conceptual unification. We've got physical space, branchial space, and now also what we can call rulial space (or just rule space). And the same overall ideas and principles apply to all of them. And just as we defined reference frames in physical space and branchial space, so also we can define reference frames in rulial space.

But what kinds of reference frames might observers set up in rulial space? In a typical case we can think of different reference frames in rulial space as corresponding to different description languages in which an observer can describe their experience of the universe.

In the abstract, it's a familiar idea that given any particular description language, we can always explicitly program any universal computer to translate it to another description language. But what we're saying here is that in rulial space it just takes choosing a different reference frame to have our representation of the universe use a different description language.

And roughly the reason this works is that different foliations of rulial space correspond to different choices of sequences of rules in the rule-space multiway graph—which can in effect be set up to "compute" the output that would be obtained with any given description language. That this can work ultimately depends on the fact that sequences of our rules can support universal computation (which the Principle of Computational Equivalence implies they ubiquitously will)—which is in effect why it only takes "choosing a different reference frame in rule space" to "run a different program" and get a different description of the observed behavior of the universe.

It's a strange but rather appealing picture. The universe is effectively using all possible rules. But as entities embedded in the universe, we're picking a particular foliation (or sequence of reference frames) to make sense of what's happening. And that choice of foliation corresponds to a description language which gives us our particular way of describing the universe.

But what is there to say definitely about the universe—independent of the foliation? There's one immediate thing: that the universe, whatever foliation one uses to describe it, is just a universal computer, and nothing more. And that hypercomputation is never possible in the universe.

But given the structure of our models, there's more. Just like there's a maximum speed in physical space (the speed of light c), and a maximum speed in branchial space (the maximum entanglement speed ζ), so also there must be a maximum speed in rulial space, which we can call ρ—that's effectively another fundamental constant of nature. (The constancy of ρ is in effect a reflection of the Principle of Computational Equivalence.)

But what does moving in rulial space correspond to? Basically it's a change of rule. And to say that this can only happen at a finite speed is to say that there's computational irreducibility: that one rule cannot emulate another infinitely fast. And given this finite "speed of emulation" there are "emulation cones" that are the analog of light cones, and that define how far one can get in rulial space in a certain amount of time.

What are the units of ρ? Essentially they are program length divided by time. But whereas in the theory of computation one typically imagines that program length can be scaled almost arbitrarily by different models of computation, here this is a measure of program length that's somehow fundamentally anchored to the structure of the rule-space multiway system, and of physics. (By the way, there'll be an analog of curvature and Einstein's equations in rulial space too—and it probably corresponds to a geometrization of computational complexity theory and questions like P?=NP.)

There's more to say about the structure of rulial space. For example, let's imagine we try to make a foliation in which we freeze time somewhere in rulial space. That'll correspond to trying to describe the universe using some computationally reducible model—and over time it'll get more and more difficult to maintain this as emulation cones effectively deliver more and more computational irreducibility.

So what does all this mean for our original goal—of finding a rule to describe our universe? Basically it's saying that any (computation universal) rule will do—if we're prepared to craft the appropriate description language. But the point is that we've basically already defined at least some elements of our description language: they are the kinds of things our senses detect, our measuring devices measure, and our existing physics describes. So now our challenge is to find a rule that successfully describes our universe within this framework.

For me this is a very satisfactory solution to the mystery of why some particular rule would be picked for our universe. The answer is that there isn't ultimately ever a particular rule; basically any rule capable of universal computation will do. It's just that—with some particular mode of description that we choose to use—there will be some definite rule that describes our universe. And in a sense whatever specialness there is to this rule is just a reflection of the specialness of our mode of description. In effect, the only thing special about the universe to us is us ourselves.

And this suggests a definite answer to another longstanding question: could there be other universes? The answer in our setup is basically no. We can't just "pick another rule and get another universe". Because in a sense our universe already contains all possible rules, so there can only be one of it. (There could still be other universes that do various levels of hypercomputation.)

But there is something perhaps more bizarre that is possible. While we view our universe—and reality—through our particular type of description language, there are endless other possible description languages which can lead to descriptions of reality that will seem coherent (and even in some appropriate definition "meaningful") within themselves, but which will seem to us to correspond to utterly incoherent and meaningless aspects of our universe.

I've always assumed that any entity that exists in our universe must at least "experience the same physics as us". But now I realize that this isn't true. There's actually an almost infinite diversity of different ways to describe and experience our universe, or in effect an almost infinite diversity of different "planes of existence" for entities in the universe—corresponding to different possible reference frames in rulial space, all ultimately connected by universal computation and rule-space relativity.

The Challenge of Language Design for the Universe

What does it mean to make a model for the universe? If we just want to know what the universe does, well, then we have the universe, and we can just watch what it does. But when we talk about making a model, what we really mean is that we want to have a representation of the universe that somehow connects it to what we humans can understand. Given computational irreducibility, it's not that we expect a model that will in any fundamental sense "predict in advance" the precise behavior of the universe down to every detail (like that I am writing this sentence now). But we do want to be able to point to the model—whose structure we understand—and then be able to say that this model corresponds to our universe.

In the previous section we said that we wanted to find a rule that we could in a sense connect with the description language that we use for the universe. But what should the description language for the rule itself be? Inevitably there is a great computational distance between the underlying rule and features of the universe that we're used to describing. So—as I've said several times here in different ways—we can't expect to use the ordinary concepts with which we describe the world (or physics) directly in the construction of the rule.

I've spent the better part of my life as a language designer, primarily building what's now the full-scale computational language that is the Wolfram Language. And I now view the effort to find a fundamental theory of physics as in many ways just another challenge in language design—perhaps even the ultimate such challenge.

In designing a computational language what one is really trying to do is to create a bridge between two domains: the abstract world of what is possible to do computationally, and the "mental" world of what people understand and are interested in doing. There are all sorts of computational processes that one can invent (say running randomly picked cellular automaton rules), but the challenge in language design is to figure out which ones people care about at this point in human history, and then to give people a way to describe these.

Usually in computational language design one is leveraging human natural language—or the more formal languages that have been developed in mathematics and science—to find words or their analogs to refer to particular "lumps of computation". But at least in the way I have done it, the essence of language design is to try to find the purest primitives that can be expressed this way.

OK, so let's talk about setting up a model for the universe. Perhaps the single most important idea in my effort to find a fundamental theory of physics is that the theory should be based on the general computational paradigm (and not, for example, specifically on mathematics). So when we talk about having a language in which to describe our model of the universe we can see that it has to bridge three different domains. It has to be a language that humans can understand. It has to be a language that can express computational ideas. And it has to be a language that can actually represent the underlying structure of physics.

So what should this language be like? What kinds of primitives should it contain? The history that has led me to what I describe here is in many ways the history of my attempts to formulate an appropriate language. Is it trivalent graphs? Is it ordered graphs? Is it rules applied to abstract relations?

In many ways, we are inevitably skating at the edge of what humans can understand. Maybe one day we will have built up familiar ways of talking about the concepts that are involved. But for now, we don't have these. And in a sense what has made this project feasible now is that we've come so far in developing ways to express computational ideas—and that through the Wolfram Language in particular those forms of expression have become familiar, at the very least to me.

And it's certainly satisfying to see that the basic structure of the models we're using can be expressed very cleanly and succinctly in the Wolfram Language. In fact, in what perhaps can be viewed as some sort of endorsement of the structure of the Wolfram Language, the models are in a sense just a quintessential example of transformation rules for symbolic expressions, which is exactly what the Wolfram Language is based on. But even though the structure is well represented in the Wolfram Language, the "use case" of "running the universe" is different from what the Wolfram Language is normally set up to do.

In the effort to serve what people normally want, the Wolfram Language is primarily about taking input, evaluating it by doing computation, and then generating output. But that's not what the universe does. The universe in a sense had input at the very beginning, but now it's just running an evaluation—and with all our different ideas of foliations and so on, we are sampling certain aspects of that ongoing evaluation.

It's computation, but it's computation sampled in a different way than we've been used to doing it. To a language designer like me, this is something interesting in its own right, with

its own scientific and technological spinoffs. And perhaps it will take more ideas before we can finish the job of finding a way to represent a rule for fundamental physics.

But I'm optimistic that we actually already have pretty much all the ideas we need. And we also have a crucial piece of methodology that helps us: our ability to do explorations through computer experiments. If we based everything on the traditional methodology of mathematics, we would in effect only be able to explore what we somehow already understood. But in running computer experiments we are in effect sampling the raw computational universe of possibilities, without being limited by our existing understanding.

Of course, as with physical experiments, it matters how we define and think about our experiments, and in effect what description language we use. But what certainly helps me, at least, is that I've now been doing computer experiments for more than forty years, and over that time I've been able to slowly refine the art and science of how best to do them.

In a way it's very much like how we learn from our experience in the physical world. From seeing the results of many experiments, we gradually build up intuition, which in turn lets us start creating a conceptual framework, which then informs the design of our language for describing things. One always has to keep doing experiments, though. In a sense computational irreducibility implies that there will always be surprises, and that's certainly what I constantly find in practice, not least in this project.

Will we be able to bring together physics, computation and human understanding to deliver what we can reasonably consider to be a final, fundamental theory of physics? It is difficult to know how hard this will be. But I am extremely optimistic that we are finally on the right track, and may even have effectively already solved the fascinating problem of language design that this entails.

Let's Go Find the Fundamental Theory!

OK, so given all this, what's it going to take to find the fundamental theory of physics? The most important thing—about which I'm extremely excited—is that I think we're finally on the right track. Of course, perhaps not surprisingly, it's still technically difficult. Part of that difficulty comes directly from computational irreducibility and from the difficulty of working out the consequences of underlying rules. But part of the difficulty also comes from the very success and sophistication of existing physics.

In the end our goal must be to build a bridge that connects our models to existing knowledge about physics. And there is difficult work to do on both sides. Trying to frame the consequences of our models in terms that align with existing physics, and trying to frame the (usually mathematical) structures of existing physics in terms that align with our models.

For me, one of the most satisfying aspects of our discoveries over the past couple of months has been the extent to which they end up resonating with a huge range of existing— sometimes so far seemingly "just mathematical"—directions that have been taken in physics in recent years. It almost seems like everyone has been right all along, and it just takes adding a new substrate to see how it all fits together. There are hints of string theory, holographic principles, causal set theory, loop quantum gravity, twistor theory, and much more. And not only that, there are also modern mathematical ideas—geometric group theory, higher-order category theory, non-commutative geometry, geometric complexity theory, etc.—that seem so well aligned that one might almost think they must have been built to inform the analysis of our models.

I have to say I didn't expect this. The ideas and methods on which our models are based are very different from what's ever been seriously pursued in physics, or really even in mathematics. But somehow—and I think it's a good sign all around—what's emerged is something that aligns wonderfully with lots of recent work in physics and mathematics. The foundations and motivating ideas are different, but the methods (and sometimes even the results) often look to be quite immediately applicable.

There's something else I didn't expect, but that's very important. In studying things (like cellular automata) out in the computational universe of simple programs, I have normally found that computational irreducibility—and phenomena like undecidability—are everywhere. Try using sophisticated methods from mathematics; they will almost always fail. It is as if one hits the wall of irreducibility almost immediately, so there is almost nothing for our sophisticated methods, which ultimately rely on reducibility, to do.

But perhaps because they are so minimal and so structureless our models for fundamental physics don't seem to work this way. Yes, there is computational irreducibility, and it's surely important, both in principle and in practice. But the surprising thing is that there's a remarkable depth of richness before one hits irreducibility. And indeed that's where many of our recent discoveries come from. And it's also where existing methods from physics and

mathematics have the potential to make great contributions. But what's important is that it's realistic that they can; there's a lot one can understand before one hits computational irreducibility. (Which is, by the way, presumably why we are fundamentally able to form a coherent view of physical reality at all.)

So how is the effort to try to find a fundamental theory of physics going to work in practice? We plan to have a centralized effort that will push forward with the project using essentially the same R&D methods that we've developed at Wolfram Research over the past three decades, and that have successfully brought us so much technology—not to mention what exists of this project so far. But we plan to do everything in a completely open way. We've already posted the full suite of software tools that we've developed, along with nearly a thousand archived working notebooks going back to the 1990s, and soon more than 400 hours of videos of recent working sessions.

We want to make it as easy for people to get involved as possible, whether directly in our centralized effort, or in separate efforts of their own. We'll be livestreaming what we do, and soliciting as much interaction as possible. We'll be running a variety of educational programs. And we also plan to have (livestreamed) working sessions with other individuals and groups, as well as providing channels for the computational publishing of results and intermediate findings.

I have to say that for me, working on this project both now and in past years has been tremendously exciting, satisfying, and really just fun. And I'm hoping many other people will be able to share in this as the project goes forward. I think we've finally got a path to finding the fundamental theory of physics. Now let's go follow that path. Let's have a blast. And let's try to make this the time in human history when we finally figure out how this universe of ours works!

Posted in: Big Picture, Computational Science, New Kind of Science, Physics

Join the discussion

+ 225 comments
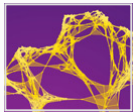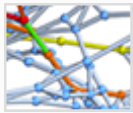
Related Writings
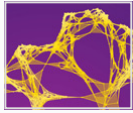
Exploring Rulial Space: The Case of Turing Machines
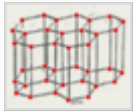June 9, 2020

Event Horizons, Singularities and Other Exotic Spacetime Phenomena
May 20, 2020

The Wolfram Physics Project:
The First Two Weeks
April 29, 2020

How We Got Here: The Backstory of the Wolfram Physics Project
April 14, 2020

Popular Categories

- Artificial Intelligence
- Big Picture
- Companies and Business
- Computational Science
- Computational Thinking
- Data Science
- Education
- Future Perspectives
- Historical Perspectives
- Language and Communication
- Life and Times
- Life Science

- Mathematica
- Mathematics
- New Kind of Science
- New Technology
- Personal Analytics
- Physics
- Software Design
- Wolfram|Alpha
- Wolfram|One
- Wolfram Language
- Other

Writings by Year

2020 | 2019 | 2018 | 2017 | 2016 | 2015 | 2014 | 2013 | 2012 | 2011 | 2010 | 2009 |
2008 | 2007 | 2006 | 2004 | 2003 | All