What Is Ultimately Possible in Physics?

Stephen Wolfram

This essay uses insights from studying the computational universe to explore questions about possibility and impossibility in the physical universe and in physical theories. It explores the ultimate limits of technology and of human experience, and their relation to the features and consequences of ultimate theories of physics.

The history of technology is littered with examples of things that were claimed to be impossible—but later done. So what is genuinely impossible in physics? There is much that we will not know about the answer to this question until we know the ultimate theory of physics. And even when we do—assuming it is possible to find it—it may still often not be possible to know what is possible.

Let's start, though, with the simpler question of what is possible in mathematics.

In the history of mathematics, particularly in the 1800s, many "impossibility results" were found [1, p. 1137]. Squaring the circle. Trisecting an angle. Solving a quintic equation. But these were not genuine impossibilities. Instead, they were in a sense only impossibilities at a certain level of mathematical technology.

It is true, for example, that it is impossible to solve any quintic—if one is only allowed to use square roots and other radicals. But it is perfectly possible to write down a finite formula for the solution to any quintic in terms, say, of elliptic functions [2]. And indeed, by the early 1900s, there emerged the view that there would ultimately be no such impossibilities in mathematics. And that instead it would be possible to build more and more sophisticated formal structures that would eventually allow any imaginable mathematical operation to be done in some finite way.

Yes, one might want to deal with infinite series or infinite sets. But somehow these could be represented symbolically, and everything about them could be worked out in some finite way.

In 1931, however, it became clear that this was not correct. For Gödel's Theorem [3] showed that in a sense mathematics can never be reduced to a finite activity. Starting from the standard axiom system for arithmetic and basic number theory, Gödel's Theorem showed that there are questions that cannot be guaranteed to be answered by any finite sequence of mathematical steps—and that are therefore "undecidable" with the axiom system given.

One might still have thought that the problem was in a sense one of "technology": that one just needed stronger axioms, and then everything would be possible. But Gödel's Theorem showed that no finite set of axioms can ever be added to cover all possible questions within standard mathematical theories.

At first, it wasn't clear how general this result really was. There was a thought that perhaps something like a transfinite sequence of theories could exist that would render everything possible—and that perhaps this might even be how human minds work.

But then in 1936 along came the Turing machine [4], and with it a new understanding of possibility and impossibility. The key was the notion of universal computation: the idea that a single universal Turing machine could be fed a finite program that would make it do anything that any Turing machine could do.

In a sense this meant that however sophisticated one's Turing machine technology might be, one would never be able to go beyond what any Turing machine that happened to be universal can do. And so if one asked a question, for example, about what the behavior of a Turing machine could be after an infinite time (say, does the machine ever reach a particular "halt" state), there might be no possible systematically finite way to answer that question, at least with any Turing machine.

But what about something other than a Turing machine?

Over the course of time, various other models of computational processes were proposed. But the surprising point that gradually emerged was that all the ones that seemed at all practical were ultimately equivalent. The original mathematical axiom system used in Gödel's Theorem was also equivalent to a Turing machine. And so were all other reasonable models of what might constitute not only a computational process, but also a way to set up mathematics.

There may be some quite different way to set up a formal system than the way it is done in mathematics. But at least within mathematics as we currently define it, we can explicitly prove that there are impossibilities. We can prove that there are things that are genuinely infinite, and cannot meaningfully be reduced to something finite.

We know, for example, that there are polynomial equations involving integers where there is no finite mathematical procedure that will always determine whether the equations have solutions [5]. It is not—as with the ordinary quintic equation—that with time some more sophisticated mathematical technology will be developed that allows solutions to be found. It is instead that within mathematics as an axiomatic system, it is simply impossible for there to be a finite general procedure.

So in mathematics there is in a sense "genuine impossibility".

Somewhat ironically, however, mathematics as a field of human activity tends to have little sense of this. And indeed there is a general belief in mathematics—much more so than in physics—that with time essentially any problem of "mathematical interest" will be solved.

A large part of the reason for this belief is that known examples of undecidable—or effectively impossible—problems tend to be complicated and contrived, and seem to have little to do with problems that could be of mathematical interest. My own work [1] in exploring generalizations of mathematics gives strong evidence that undecidability is actually much closer at hand—and that in fact its apparent irrelevance is merely a reflection of the narrow historical path that mathematics as a field has followed [1, sect. 12.9].

In a sense, the story is always the same—and to understand it sheds light on some of what might be impossible in physics. The issue is computation universality. Just where is the threshold for computation universality?

For once it is possible to achieve computation universality within a particular type of system or problem, it follows that the system or problem is in a sense as sophisticated as any other—and it is impossible to simplify it in any general way. And what I have found over and over again is that universality—and traces of it—occur in vastly simpler systems and problems than one might ever have imagined [1, chap. 11; 6, 7].

Indeed, my guess is that a substantial fraction of the famous unsolved problems in mathematics today are not unsolved because of a lack of mathematical technology—but because they are associated with universality, and so are fundamentally impossible to solve.

But what of physics?

Is there a direct correspondence of mathematical impossibility with physical impossibility? The answer is that it depends what physics is made of. If we can successfully reduce all of physics to mathematics, then mathematical impossibility in a sense becomes physical impossibility.

In the first few decades of the modern study of computation, the various models of computation that were considered were thought of mainly as representing processes—mechanical, electronic or mathematical—that a human engineer or mathematician might set up. But particularly with the rise of models like cellular automata (e.g. [8]), the question increasingly arose of how these models—and computational processes they represent—might correspond to the actual operation of physics.

The traditional formulation of physics in terms of partial differential equations—or quantized fields—makes it difficult to see a correspondence. But the increasing implementation of physical models on computers has made the situation somewhat clearer.

There are two common technical issues. The first is that traditional physics models tend to be formulated in terms of continuous variables. The second is that traditional physics models tend not to say directly how a system should behave—but instead just to define an equation which gives a constraint on how the system should behave.

In modern times, good models of physical systems have often been found (e.g. [1, chap. 8]) that are more obviously set up like traditional digital computations—with discrete variables, and explicit progression with time. But even traditional physical models are in many senses computational. For we know that even though there are continuous variables and equations to solve, there is an immense amount that we can work out about traditional physical models using, for example, *Mathematica* [9].

Mathematica obviously runs on an ordinary digital computer. But the point is that it can symbolically represent the entities in physical models. There can be a variable x that represents a continuous position, but to *Mathematica* it is just a finitely represented symbol, that can be manipulated using finite computational operations.

There are certainly questions that cannot obviously be answered by operating at a symbolic level—say about the precise location of some idealized particle represented by a real number. But when we imagine constructing an experiment or an apparatus, we specify it in a finite, symbolic, way. And we might imagine that then we could answer all questions about its behavior by finite computational processes.

But this is undoubtedly not so. For it seems inevitable that within standard physical theories there is computation universality. And the result is that there will be questions that are impossible to answer in any finite way. Will a particular three-body gravitational system (or an idealized solar system) be stable forever? Or have some arbitrarily complicated form of instability?

Of course, it could be even worse.

If one takes a universal Turing machine, there are definite kinds of questions that cannot in general be answered about it—an example being whether it will ever reach a halt state from a given input. But at an abstract level, one can certainly imagine constructing a device that can answer such questions: doing some form of "hypercomputation" (e.g. [10,11]). And it is quite straightforward to construct formal theories of whole hierarchies of such hypercomputations.

The way we normally define traditional axiomatic mathematics, such things are not part of it. But could they be part of physics? We do not know for sure. And indeed within traditional mathematical models of physics, it is a slippery issue.

In ordinary computational models like Turing machines, one works with a finite specification for the input that is given. And so it is fairly straightforward to recognize when some long and sophisticated piece of computational output can really be attributed to the operation of the system, and when it has somehow been slipped into the system through the initial conditions for the system.

But traditional mathematical models of physics tend to have parameters that are specified in terms of real numbers. And in the infinite sequence of digits in a precise real number, one can in principle pack all sorts of information—including, for example, tables of results that are beyond what a Turing machine can compute. And by doing this, it is fairly easy to set things up so that traditional mathematical models of physics appear to be doing hypercomputation.

But can this actually be achieved with anything like real, physical, components?

I doubt it. For if one assumes that any device one builds, or any experiment one does, must be based on a finite description, then I suspect that it will never be possible to set up hypercomputation within traditional physical models [1, sect. 12.4 and notes].

In systems like Turing machines, there is a certain robustness and consistency to the notion of computation. Large classes of models, initial conditions and other setups are equivalent at a computational level. But when hypercomputation is present, details of the setup tend to have large effects on the level of computation that can be reached, and there do not seem to be stable answers to questions about what is possible and not.

In traditional mathematical approaches to physics, we tend to think of mathematics as the general formalism, which in some special case applies to physics. But if there is hypercomputation in physics, it implies that in a sense we can construct physical tools that give us a new level of mathematics—and that answer problems in mathematics, though not by using the formalism of mathematics. And while at every level there are analogs of Gödel's Theorem, the presence of hypercomputation in physics would in a sense overcome impossibilities in mathematics, for example giving us ways to solve all integer equations.

So could this be how our universe actually works?

From existing models in physics we do not know. And we will not ultimately know until we have a fundamental theory of physics.

Is it even possible to find a fundamental theory of physics? Again, we do not know for sure. It could be—a little like in hypercomputation—that there will never be a finite description for how the universe works. But it is a fundamental observation—really the basis for all of natural science—that the universe does show order, and does appear to follow definite laws.

Is there in a sense some complete set of laws that provide a finite description for how the whole universe works? We will not know for sure until or unless we find that finite description—the ultimate fundamental theory.

One can argue about what that theory might be like. Is it perhaps finite, but very large? Like the operating system of one of today's computers. Or is it not only finite, but actually quite small? Like a few lines of computer code. We do not yet know.

Looking at the complexity and richness of the physical universe as we now experience it, we might assume that a fundamental theory—if it exists—would have to reflect all that complexity and richness, and itself somehow be correspondingly complex. But I have spent many years studying what is in effect a universe of possible theories—the computational universe of simple programs. And one of the clear conclusions is that in that computational universe it is easy to find immense complexity and richness, even among extremely short programs with extremely simple structure [1].

Will we actually be able to find our physical universe in this computational universe of possible universes? I am not sure. But certainly it is not obvious that we will not be able to do so. For already in my studies of the computational universe, I have found candidate universes that I cannot exclude as possible models of our physical universe (e.g. [12, 13]).

If indeed there is a small ultimate model of our physical universe, it is inevitable that very few familiar features of our universe as we normally experience it will be visible in that model [1, sect. 9.5]. For in a small model, there is in a sense no room to specify, say, the number of dimensions of space, the conservation of energy or the spectrum of particles. Nor probably is there any room to have anything that corresponds directly to our normal notion of space or time [1, sects. 9.6 - 9.11].

Quite what the best representation for the model should be I am not sure. And indeed it is inevitable that there will be many seemingly quite different representations that only with some effort can be shown to be equivalent.

A particular representation that I have studied involves setting up a large number of nodes, connected in a network, and repeatedly updated according to some local rewrite rule [1, chap. 9]. Within this representation, one can in effect just start enumerating possible universes, specifying their initial conditions and updating rules. Some candidate universes are very obviously not our physical universe. They have no notion of time, or no communication between different parts, or an infinite number of dimensions of space, or some other obviously fatal pathology.

But it turns out that there are large classes of candidate universes that already show remarkably suggestive features. For example, any universe that has a notion of time with a certain robustness property turns out in an appropriate limit to exhibit Special Relativity [1, sect. 9.13]. And even more significantly, any universe that exhibits a certain conservation of finite dimensionality—as well as generating a certain level of effective microscopic randomness—will lead on a large scale to spacetime that follows Einstein's equations for General Relativity [1, sect. 9.15].

It is worth emphasizing that the models I am discussing are in a sense much more complete than models one usually studies in physics. For traditionally in physics, it might be considered quite adequate to find equations one of whose solutions successfully represents some feature of the universe. But in the models I have studied the concept is to have a formal system which starts from a particular initial state, then explicitly evolves so as to reproduce in every detail the precise evolution of our universe.

One might have thought that such a deterministic model would be excluded by what we know of quantum mechanics. But in fact the detailed nature of the model seems to make it quite consistent with quantum mechanics. And for example its network character makes it perfectly plausible to violate Bell's inequalities at the level of a large-scale limit of three-dimensional space [1, sect. 9.16].

So if in fact it turns out to be possible to find a model like this for our universe, what does it mean?

In some sense it reduces all of physics to mathematics. To work out what will happen in our universe becomes like working out the digits of pi: it just involves progressively applying some particular known algorithm.

Needless to say, if this is how things work, we will have immediately established that hypercomputation does not happen in our universe. And instead, only those things that are possible for standard computational systems like Turing machines can be possible in our universe.

But this does not mean that it is easy to know what is possible in our universe. For this is where the phenomenon of computational irreducibility [1, sect. 12.6] comes in.

When we look at the evolution of some system—say a Turing machine or a cellular automaton—the system goes through some sequence of steps to determine its outcome. But we can ask whether perhaps there is some way to reduce the computational effort needed to find that outcome—some way to computationally reduce the evolution of the system.

And in a sense much of traditional theoretical physics has been based on the assumption that such computational reduction is possible. We want to find ways to predict how a system will behave, without having to explicitly trace each step in the actual evolution of the system.

But for computational reduction to be possible, it must in a sense be the case that the entity working out how a system will behave is computationally more sophisticated than the system itself.

In the past, it might not have seemed controversial to imagine that humans, with all their intelligence and mathematical prowess, would be computationally more sophisticated than systems in physics. But from my work on the computational universe, there is increasing evidence for a general Principle of Computational Equivalence [1, chap. 12], which implies that even systems with very simple rules can have the same level of computational sophistication as systems constructed in arbitrarily complex ways.

And the result of this is that many systems will exhibit computational irreducibility, so that their processes of evolution cannot be "outrun" by other systems—and in effect the only way to work out how the systems behave is to watch their explicit evolution.

This has many implications—not the least of which is that it can make it very difficult even to identify a fundamental theory of physics.

For let us say that one has a candidate theory—a candidate program for the universe. How can we find out whether that program actually is the program for our universe? If we just start running the program, we may quickly see that its behavior is simple enough that we can in effect computationally reduce it—and readily prove that it is not our universe.

But if the behavior is complex—and computationally irreducible—we will not be able to do this. And indeed as a practical matter in actually searching for a candidate model for our universe, this is a major problem. And all one can do is to hope that there is enough computational reducibility that one manages to identify known physical laws within the model universe.

It helps that if the candidate models for the universe are simple enough, then there will in a sense always be quite a distance from one model to another—so that successive models will tend to show very obviously different behavior. And this means that if a particular model reproduces any reasonable number of features of our actual universe, then there is a good chance that within the class of simple models, it will be essentially the only one that does so.

But, OK. Let us imagine that we have found an ultimate model for the universe, and we are confident that it is correct. Can we then work out what will be possible in the universe, and what not?

Typically, there will be certain features of the universe that will be associated with computational reducibility, and for which we will readily be able to identify simple laws that define what is possible, and what is not.

Perhaps some of these laws will correspond to standard symmetries and invariances that have already been found in physics. But beyond these reducible features, there lies an infinite frontier of computational irreducibility. If we in effect reduce physics to mathematics, we still have to contend with phenomena like Gödel's Theorem. So even given the underlying theory, we cannot work out all of its consequences.

If we ask a finite question, then at least in principle there will be a finite computational process to answer that question—though in practice we might be quite unable to run it. But to know what is possible, we also have to address questions that are in some sense not finite.

Imagine that we want to know if macroscopic spacetime wormholes are possible.

It could be that we can use some computationally reducible feature of the universe to answer this.

But it could also be that we will immediately be confronted with computational irreducibility—and that our only recourse will for example be to start enumerating configurations of material in the universe to see if any of them end up evolving to wormholes.   And it could even be that the question of whether any such configuration—of any size—-exists could be formally undecidable, at least in an infinite universe.

But what about all those technologies that have been discussed in science fiction?

Just as we can imagine enumerating possible universes, so also we can imagine enumerating possible things that can be constructed in a particular universe. And indeed from our experience in exploring the computational universe of simple programs, we can expect that even simple constructions can readily lead to things with immensely rich and complex behavior.

But when do those things represent useful pieces of technology?

In a sense, the general problem of technology is to find things that can be constructed in nature, and then to match them with human purposes that they can achieve (e.g. [1, sect. 9.11 and 9.10]). And usually when we ask whether a particular type of technology is possible, what we are effectively asking is whether a particular type of human purpose can be achieved in practice. And to know this can be a surprisingly subtle matter, which depends almost as much on understanding our human context as it does on understanding features of physics.

Take for example almost any kind of transportation.

Earlier in human history, pretty much the only way to imagine that one would successfully achieve the purpose of transporting anything would be explicitly to move the thing from one place to another. But now there are many situations where what matters to us as humans is not the explicit material content of a thing, but rather the abstract information that represents it. And it is usually much easier to transport that information, often at the speed of light.

So when we say "will it ever be possible to get from here to there at a certain speed" we need to have a context for what would need to be transported. In the current state of human evolution, there is much that we do that can be represented as pure information, and readily transported. But we ourselves still have a physical presence, whose transportation seems like a different issue.

No doubt, though, we will one day master the construction of atomic-scale replicas from pure information. But more significant, perhaps our very human existence will increasingly become purely informational—at which point the notion of transportation changes, so that just transporting information can potentially entirely achieve our human purposes.

There are different reasons for saying that things are impossible.

One reason is that the basic description of what should be achieved makes no sense. For example, if we ask "can we construct a universe where 2 + 2 = 5?", this makes no sense. From the very meaning of the symbols in 2 + 2 = 5, we can deduce that it can never be satisfied, whatever universe we are in.

There are other kinds of questions where at least at first the description seems to make no sense.

Like "is it possible to create another universe?" Well, if the universe is defined to be everything, then by definition the answer is obviously "no". But it is certainly possible to create simulations of other universes; indeed, in the computational universe of possible programs we can readily enumerate an infinite number of possible universes.

For us as physical beings, however, these simulations are clearly different from our actual physical universe. But consider a time in the future when the essence of the human condition has been transferred to purely informational form. At that time, we can imagine transferring our experience to some simulated universe, and in a sense existing purely within it—just as we now exist within our physical universe.

And from this future point of view, it will then seem perfectly possible to create other universes.

So what about time travel? There are also immediate definitional issues here. For at least if the universe has a definite history—with a single thread of time—the effect of any time travel into the past must just be reflected in the whole actual history that the universe exhibits.

We can often describe traditional physical models—for example for the structure of spacetime—by saying that they determine the future of a system from its past. But ultimately such models are just equations that connect different parameters of a system. And there may well be configurations of the system in which the equations cannot readily be seen just as determining the future from the past.

Quite which pathologies can occur with particular kinds of setups may well be undecidable, but when it seems that the future affects the past what is really being said is just that the underlying equations imply certain consistency conditions across time. And when one thinks of simple physical systems, such consistency conditions do not seem especially remarkable. But when one combines them with human experience—with its features of memory and progress—they seem more bizarre and paradoxical.

In some ancient time, one might have imagined that time travel for a person would consist in projecting them—or some aspect of them—far into the future. And indeed today when one sees writings and models that were constructed thousands of years ago for the afterlife, there is a sense in which that conception of time travel has been achieved.

And similarly, when one thinks of the past, the increasing precision with which molecular archaeology and the like can reconstruct things gives us something which at least at some time in history would have seemed tantamount to time travel.

Indeed, at an informational level—but for the important issue of computational irreducibility—we could reasonably expect to reconstruct the past and predict the future. And so if our human existence was purely informational, we would in some sense freely be able to travel in time.

The caveat of computational irreducibility is a crucial one, however, that affects the possibility of many kinds of processes and technologies.

We can ask, for example, whether it will ever be possible to do something like unscramble an egg, or in general in some sense to reverse time. The Second Law of thermodynamics has always suggested the impossibility of such things.

In the past, it was not entirely clear just what the fundamental basis for the Second Law might be. But knowing about computational irreducibility, we can finally see a solid basis for it [1, sect. 9.3]. The basic idea is just that in many systems the process of evolution through time in effect so "encrypts" the information associated with the initial conditions for the system that no feasible measurement or other process can recognize what they were. So in effect, it would take a Maxwell's Demon of immense computational power to unscramble the evolution.

In practice, however, as the systems we use for technology get smaller, and our practical powers of computation get larger, it is increasingly possible to do such unscrambling. And indeed that is the basis for a variety of important control systems and signal processing technologies that have emerged in recent years.

The question of just what kinds of effective reversal of time can be achieved by what level of technology depends somewhat on theoretical questions about computation. For example, if it is true that P ≠ NP, then certain questions about possible reversals will necessarily require immense computational resources.

. . .

[3 pages omitted for length reasons]

. . .

So in a sense what we will ultimately perceive as possible in physics depends more on the evolution of human purposes than it does on the details of the physical universe. In some ways this is a satisfying result. For it suggests that we will ultimately never be constrained in what we can achieve by the details of our physical universe. The constraints on our future will not be ones of physics, but rather ones of a deeper nature. It will not be that we will be forced to progress in a particular direction because of the specific details of the particular physical universe in which we live. But rather—in what we can view as an ultimate consequence of the Principle of Computational Equivalence—the constraints on what is possible will be abstract features of the general properties of the computational universe. They will not be a matter of physics—but instead of the general science of the computational universe.

References

1. S. Wolfram, *A New Kind of Science*, Wolfram Media, 2002.

2. Wolfram Research, "Solving the Quintic with *Mathematica,*" Poster ; M. Trott and V. Adamchik, library.wolfram.com/infocenter/ TechNotes/15, 1994.

3. K. Gödel, "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter System I," *Monatshefte für Math. u. physik*, **38**, 1931, pp. 173-198.

4. A. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," in *Proceedings of the London Mathematical Society*, Ser. 2, **42**, 1937, pp. 230-265.

5. Y. Matiyasevich, *Hilbert's Tenth Problem*, The MIT Press, 1993.

6. The Wolfram 2,3 Turing Machine Research Prize," www.wolframscience.com/prizes/tm23/, May 14, 2007.

7. A. Smith. "Universality of Wolfram's 2,3 Turing Machine," to appear in *Complex Systems*.

8. S. Wolfram, *Cellular Automata and Complexity: Collected Papers*, Addison-Wesley Publishing Company, 1994.

9. Wolfram Research, *Mathematica*, 1988-.

10. A. Turing, "Systems of logic based on ordinals," Proc. London Math. Soc., Ser. 2-45, 1939, pp. 161-228.

11. B. Copeland, "Hypercomputation,"*Minds and Machines*, **12**(4), 2002, pp. 461-502.

12. S. Wolfram, Talk given at the Emergent Gravity Conference, MIT , Aug 26, 2008.

13. S. Wolfram, Talk given at the JOUAL 2009 Workshop, CNR-Area, Jul 10, 2009.

14. Wolfram Research., *WolframTones*, tones.wolfram.com, 2005.

15. Wolfram | Alpha LLC, Wolfram | Alpha, www.wolframalpha.com, 2009.