Physics considerations for a simulated reality

Max Hodak

1 Why build a simulated reality?

Video games have a powerful natural appeal. Our physical world imposes harsh constraints: atoms are difficult to work with compared to bits, and the scarcity of both materials and the capacity to transform them leads to a tough environment. By contrast, video games can be anything we want them to be: players can fly, use magic, or even have their own planet, all the while exercising mental faculties in just as satisfying a way as they would in our base reality. As computer processing power continues to improve and the space of games we can build expands, a natural opportunity arises to build simulated realities that are as engaging and emotionally immersive as our base reality, but with fewer constraints and a greater capacity for easier creative expression.

Video game design is challenging on multiple levels. Scripted worlds may be engaging for a short period of time, but ultimately suffer from the limited hand-crafted information embedded by the developers for the player to experience. It is possible to build open sandbox worlds that support a great degree of creative expression (for example, Minecraft or SimCity), but game mechanic design is a difficult art. Most variations are neither engaging nor fun, and all of them are characterized by severe limits to their depth.

It's often been observed that immersion in virtual worlds is far more related to psychological and emotional depth than graphical fidelity.[1] It is an important point that in considering the physics of a simulated reality, we are *not* principally concerned with the quality of the graphics, though presumably they would follow as "for free." We are principally concerned with the range and sophistication of the emotionally satisfying creativity that the world supports.

The physics of our world enables a vast range of creativity to be expressed and its results enjoyed. A central part of our argument is that, in developing a "complete" virtual world, one that players will find engaging for indefinite periods of time, the primary challenge is not game design as it is classically imagined, but that of defining a suitable physics system. We conjecture that there is a relationship between the depth of a virtual world and the sophistication of its underlying physics.

Any non-trivial universe must contain information, and we can phrase the central problem in world design as one of writing down the information that defines a universe; informally, we can state that a universe without information has zero depth. We can start from a premise that information in any universe is found in two forms: matter and its interactions.

2 Naturalness

Before we think about physics as it is usually understood, let's briefly look at video game "physics" as it exists today. The physics of a game is encoded in its source code, and implemented by the game engine. The logic laid out in the source determines the complete range of things possible in the game environment.

This physics is probably built from a lot of nonlinear branching logic like "if" and "switch" statements. Also, the logic is probably structured in a way such that the phenomena you tend to get in the game are generally "discrete". For example, when you walk up to a door suddenly the character "E" has a new behavior, or how picking up an object is an all-or-nothing event. This logic has a lot of complicated structure that is irreducible to simpler unifying principles. Most of the information about the world is encoded specifically and is incompressible. It is highly "fine-tuned".

On the other hand, essentially everything that we understand day-to-day about our world is understood in terms of a small set of very deep things. The color of the sky is nowhere hard-coded into our universe; it emerges automatically from quantum field theory. There is some intuitive feeling of improved "smoothness" over the game logic example.

This is a very informal statement. There's no real specific claim about the structure of our physics being made here. All we have right now is a vague intuitive sense of the difference between "natural" universes and heavily fine-tuned ones.

2.1 Is naturalness important?

We conjecture that universes built from more natural physics, if they work at all, will be much richer for the users than universes built from highly fine-tuned physics. This is because an essential measure of the depth of the world is the range, sophistication and coherency of the phenomena it allows, and we hypothesize that a universe in which the main effects we interact with are emergent is likely to have a much wider range of interesting things to do versus a universe in which the range of things to do is directly prescribed.

At first glance, it seems much easier to build a fine-tuned universe than a natural one. Certainly it is easier to build a regular video game than write down a field theory that spontaneously produces a compelling video game.

Game developers can construct huge MMOs full of quests and social mechanics, but eventually the players will exhaust them, and lose interest and leave. Worse, since the players can from early on feel these limits, however gently, they fail to fully invest themelves in the world. With limited attention and capacity for emotional labor, most people will only construct a life and career in the most promising reality available. Mechanics are known that can generate immense variability, but even these eventually become stale and lose relevance to the multifacteted ambitious human in search of novelty and accomplishment.

There is some essential relationship to be elucidated between the compressibility of a universe's physics and the depth of the universe, assuming an anthropic principle.

2.2 Non-smooth physics is difficult to make highly expressive

To make this a little clearer, imagine a virtual environment with a ball placed at a height of $z = 1$ above the $x, y = 0$ origin of an infinite plane. When time is started, the ball experiences an acceleration in the $-z$ direction and proceeds to fall under this force until it reaches $z = 0$ (the surface of the plane), at which time its momentum vector is transformed by $p_z = -1 * p_z$.

In this world, we have directly defined two material entities (the ball and the plane) and two interactions (the apparent gravity and the the momentum change of the ball striking the plane). That information defines the physics of this world, and while it admits a non-zero range of game mechanics, the lack of depth becomes quickly apparent. We cannot deform the shape of the ball; we cannot cause its bouncing to suddenly become damped or cause the plane to impart new energy into the ball when they interact. We cannot change the color of anything in this universe, and we cannot make any new things.

We can of course do any of these things by writing additional logic into the game engine. It is easy to see, though, that the complexity of this problem grows very quickly with the depth of the world. It is straightforward to make the ball deformable so that when it collides with the plane it is briefly squished into an oblate spheroid. But this is a superficial change unless we also make the ball composed of some definite kind of matter, and by extension a definite density and a set of forces that hold it together. Then, instead of simply detecting a collision with the surface and inverting the z dimension of the momentum vector, we can calculate the force imparted on the plane from the kinetic energy of the ball accounting for the deformation of its material and propagate an opposing force back into the material of the ball and allow the change in momentum to fall out naturally. Doing this requires us to also build the plane out of a definite material and ensure that our logic is sensical for the interaction of the material ball with the material surface.

Suppose we now wanted to divide the ball in two and use two balls. This requires further special-cased logic in the game engine to support; nothing we have defined yet automatically allows it. Implementing this in a way that only allows dividing the material ball but not dividing the material surface is less satisfactory than an approach that allows dividing any material object in the world.

If the $z = 0$ plane is an infinite surface, this brings challenges: what does it mean to divide an infinite plane? Are we satisfied with the creation of voids in its surface? Do we want a way to remove all of the (infinite) material from the plane out from some defined point? Further, we have special-case defined the apparent force of gravity on the ball: does that force also apply to the plane? If not, we need to define some new mechanism by which gravity acts on the ball (or subdivided balls) but not the plane.

The devoted game developer can add a huge number of rules to their engine to get any specific desired behavior, but the problem of getting broadly general emergent behavior in circumstances not specifically imagined by the developer quickly becomes an intricate theoretical challenge.

The richness of the world as enabled by its physics is of central importance in building a world that is as richly creative, or more so, as our base reality. In lieu of a good theoretical motivation for such a physics system, games will remain comparatively shallow and unrewarding over a long enough timescale.

2.3 Depth

So far we have informally used a concept of "depth" to mean some metric of the how long a simulated reality remains engaging. While we don't necessarily have a very strong motivation for the best way to sharpen this idea, for now let's define the depth λ as the rate constant of a Poisson process that corresponds to the arrival of the average user's last accession of the world on a regular basis. (For example, the last time they log in and then fail to log in again within a week.) There is likely a better metric out there, but this can serve as an operational definition to facilitate a sufficiently specific discussion.

We can now update our definition of the central problem in this paper to one of defining a group of interactions that operates on some informational structure such that λ is maximized. (In the limit, towards ∞.)

3 Matter

Let us now turn this question around. If we wanted to construct a simulated reality, what is the simplest thing that could work? We are searching for the intersection of being similar enough to our reality that the simulated world is easy for humans to reason about and engage with while also being easy to run on digital computers. This second constraint is a key source of the challenge: nature has no problem with calculations we cannot easily perform (our universe has no issue solving large n-body interactions, or for example, implementing the principle of least action), and so we cannot simply write an emulator for the physics of our universe. (For a more detailed review of this problem, see [2].)

Any material universe must contain some kind of *stuff*. The simplest form of stuff is simply some kind of undividable atom. If we are to have multiple kinds of stuff in our universe, we must add some additional information to our our atoms. One way to embed this information is to associate each atom with an integer, a_1, that defines its type. If we define the space using Euclidian coordinates, then we can define for every point in space a matter operator that tell us whether that point is empty or has an atom of some integer type. In the simplest case with only one type of matter, a_1 can be restricted to $[0, 1]$, where the matter operator returns zero if the point is empty and one if it is occupied.

Here appears an issue that must be resolved: if the space is continuous, then the definition above is unsatisfactory because it assumes each atom is infinitely small, and if space is discrete, the minimum length scale must be the size of our atoms and the shape of our atoms must reflect the metric of the space.

(If we constructed a discrete space with a much finer resolution than our atoms, our atoms would have a spatially-extended form with some density function. This is unsatisfying because it implies that atoms are further subdivideable: each atom is a cloud of stuff that can be detected separately, presumably with properties and interactions of their own.)

Alternatively, we could define the matter operator as a probability of observing an atom at a given set of spatial coordinates at a specified time. With either continuous space or discrete space possesing a fine grain relative to the atoms, this is required to preserve the concept of unitary atoms as our fundamental building block of matter. It is worth noting that this also implies, in some sense, an uncertainty principle (though not necessarily complementarity) and in effect a spatial "foam" at the minimum observable length scale, which is distinct from the scale of our atoms.

4 Computation

Let us now add a constraint that our simulation should allow the construction of universal computers. To do this, we can add rules that govern how neighboring atoms of various types interact to transform information.

We would like our computer to be multi-use; that is, not consumed by performing computation. This requires the keeping of state beyond simply the definition of the colored atoms and their configuration. This information cannot be stored in the set of interactions; it must be stored somehow in the matter.

We can see that this is true because the computations we wish to carry out will vary with time and each new desired computation may contain novel information, whereas the set of interactions, and the information they contain, is fixed upon creation. (This assumes, of course, that our physics is in fact not time-varying; or if it is, that it is not receiving significant information from some external system during its evolution. This seems safe to assume since if it is false, our physics becomes not clearly definable and, maybe more importantly, we exit the domain over which our central challenge of constructing interesting physics makes sense.) The information for the computation must come from somewhere, and if it comes from the set of interactions the foundations on which we are building become poorly behaved.

Thus, we must introduce an additional degree of freedom to our atoms. The simplest way to do this is to associate a second integer number, a_2, with each atom. We can define interactions that do not mutate the computer itself but instead act on transforable state within it. For simplicity, we can restrict a_2 to the values of 0 and 1 (a single bit), but this is not fundamental.

Note that this information doesn't have to be stored in the atoms necessarily, but it must be somewhere. A second type of matter (call it a 'tape') that interacts with the first type of matter would be equivalent.

Solving a problem would then be a matter of either configuring appropriate a_2 values on each atom or applying a prepared tape to the first kind of matter and allowing time to run. With appropriate interactions defined, as time increases the configuration of this second kind of information will trend towards the solution to the problem posed in the initial conditions of the system.

With a sufficient set of interactions (yet to be defined) and at least two bits of effective information per atom, we can construct a universe that contains reusable computers.

5 Time

Now another question emerges: how do we represent time? The naïve approach of handling time implicitly as a series of state changes provides no explicit representation at all; instead, we mimic the effect of time using some global clock.

However, this is problematic as it creates a single global reference frame: no matter how you choose the users' spatial coordinate systems, there is always only a single clock available.

A very interesting dicohotomy emerges from this! Which is: this is no problem at all if the observers are entirely contained within the simulated world. It is easy to simulate relativisitic effects for observers created by the simulation and which only exist within it, who are happy to take the software's accounting of their timeline at face value.

On the other hand, if our goal is to create a simulation that *we* can visit, with our physical brains rooted in this reality, the existence of an unavoidable global clock is a significant issue.

Explicitly modeling spacetime, with observers as cursors within it, runs into similar issues. It does not provide an easy way for multiple observers from our reality to access "first-class" relativistic experience within the simulation, though it has an aesthetic appeal over implicit time.

It seems that any reality populated by multiple consciousnesses that also exist in a "deeper" reality (e.g., the world when you take off the head-mounted display) must get their time from the "lowest level" available. It is not clear how multiple observers with distinct reference frames each anchored to our external "base reality" would work in a relativistic simulation.

5.1 Non-relativistic universes

If we accept this constraint, we find ourselves restricted down to the space of physics of non-relativistic universes.

This immediately causes some very weird issues: for one, the speed of light (information) becomes infinite as change propagates instantaneously. An implication of this is that the time must be discrete, since there is no mechanism for continuous evolution. (In each moment, light must fill the whole space instantaneously.) Were this not true, two users could coordinate in the base reality to transmit information superluminally and create paradoxes. For example, the user in the "future" could communicate price movements in an in-world market to the user in the "past". Though a game can obviously simulate travel through space, it is impossible to simulate travel through time with multiple players in a way that has any meaningful consequences.

It's not enough to say there is a *different*, faster speed of light, because we have to follow a new posulate of relativity: that our physics is the same for arbitrary informational connectivity. That is, regardless of where in the universe the two players are, or what reference frame each is using, they might be sitting right next to each other in the real world and able to compare observations.

These universes contain no space-time relationship, with profound effects.

Any simulation we build, to be experienced by us, and which contains multiple users, must have the consequences of relativity embedded in some other kind of engineered interaction. Further, we cannot keep all of the effects: regardless of how we construct our interactions, time dilation can't be included.

There is a clear relationship between the independence of the clocks of the users and the degree to which it can resemble a relativistic universe like ours. However, it's certainly possible (as many game engines have shown so far) to simulate enough for a locally believable world.

5.2 Gravitation

One of the physics we are most eager to try and simulate is ours, as presently understood by general relativity and quantum field theory. We can note that the issues above affect *only* the physics arising from GR; the effects of QFT, if adapted, seem perfectly compatible with our singular notion of time. Given that we cannot use GR to get gravitation in our simulation, it seems like a developing the physics of gravity separately from the rest of the system would be a good idea.

6 Observers

Another lens through which to consider this family of problems is: once you have a universe filled with matter, what point do you choose as your origin for the purpose of beginning to run your simulation? From what point in spacetime do you begin evaluating?

If the universe begins preconfigured with a spatial extent, it is difficult to choose any one point over another. However, it is apparent this choice matters in that concrete observations will vary depending on how the starting point is chosen. (For example, the results of diffusive equations will in general depend on their boundary conditions, which can vary depending on the choice of "boundary.")

There are several possible ways we could imagine to resolve this. First, we could say we will start everywhere simultaneously and achieve convergence across the entire universe within each 'frame' given a discrete, implicit model of time. Second, we could start the universe at absolute zero and slowly introduce energy until it becomes warm and dynamic, while minimizing the impact of our arbitrary choice of starting point. (*Hand wavily*, in some sense kind of a renormalization-like approach; we end up with a low-energy effective limit solution.) Third, our universe could begin with expansion from a singularity, allowing us to choose a natural starting point for evaluation that avoids the issue of starting point choice in the first place.

We already know that time must be discrete in our simulation; even if our physics didn't require it, that would have to be true regardless for anything running as software on a digital computer.

The issue with the first solution is that it requires solving the *entire* universe jointly on every timestep, which is expensive. It is desirable to choose a solution that allows us to lazily evaluate our universe only as the information is needed.

This idea of "as the information is needed" demands a concept of an observer: broadly, something to require the information. An observer is something that takes a measurement.

Now, the obvious point is that our simulation is running on a computer, and the computer has all of the data in memory! This memory is a physical thing, which concretely contains all of the information, and with a sufficient device we could see the values of all of the bits in it. Thus, the computer running the simulation is itself an observer of the world, and one that is observing the *whole* world continously!

This also means that if there is no source of external randomness available, the simulation is entirely deterministic. All new information must come from the outside, for example, from the users sitting in their base reality manipulation the simulation. But regardless of whether external randomness is available, we become limited to physics which allow hidden-variable theories.

Using quantum computers, where the memory is composed of qubits and unobservable (non-teleportable), avoids this constraint.

We go back to the question of laziness. If the computer running the simulation is classical, it can only perform what we can call "Type I" laziness: instead of storing an explicit representation of the state of the universe, it can store generating functions which produce that state. These generating functions can be called asynchronously, but they compress space in exchange for time; the "framerate" of the simulation will be determined by the slowest measurement operation, which limits how lazy you can be.

If the computer is quantum, it can implement "Type II" laziness, in which the information representing the universe can be compressed (e.g., entangled) with no cost in time.

7 Entropy

Floating point math on digital computers is necessarily approximate: real numbers are defined to infinitely many digits, and computers have fixed memory. Since computers can only represent a small fraction of $\mathbb{R}$, errors are introduced on essentially every arithmetic operation.

As a sequence of floating point operations progresses, the accumulated error is generally chaotic.[3] The error never decreases, though it can stay constant through some operations that by chance or construction can be done exactly. We can consider this compounding error as a measure of entropy for our simulation, and it implies the existence of a principle similar in nature to our own second law of thermodynamics.

Note that while, to us, the compounding error is interpreted as a progressive loss of information, within the simulation it would merely imply a fundamental limit on measurement or ability to easily predict certain phenomena which become non-smooth, but not overall conservation violations.

8 Basis for further discussion

This has been a brief tour of some of the challenges of building an MMO "from scratch" that provides a very high degree of flexibility. Importantly, this analysis provides no guarantees that such a universe would be fun, which is an essential pre-requisite for any successful virtual world. However, the hand-design of mechanics in narrowly constructed universes suffer from a universal lack of depth, and this is a critical bottleneck for the realization of an compelling long-term alternative to the world most people inhabit today.

The main takeaways from this draft can be summarized as:

- Immersion is much more a factor of the creative expressiveness possible in a world than the "first-order" (eg, graphical) fidelity, though graphics are important
- How creatively satisfying a world is long-term is largely a function of the world's physics, broadly considered
- Defining physics that is very deep, recognizable as a material universe, and also efficient to run on von Neumann machines is non-trivial
- Much experimentation is required to explore physics systems from first principles, which is uncommon among game developers today
- Solving this challenge will open up a massive new frontier, directly enabling the creation of new cities and environments in a way that would be impossible on Earth with its intense power struggles and scarce resources

As HCI technology - from VR headsets to implantable neural interfaces - improves over the coming decades, the bottleneck will shift towards the challenge of constructing worlds worthy of being inhabited using such devices. Sure, they can always be used for the twenty hours of a $60 episodic game, or the couple months of obsession over the latest FPS or battle royale game, but the rewards if we can move beyond that will be immense, driving significant economic pressures to solve these problems and unlock new, infinite "internal" universes to complement our vast but harshly limiting "external" one.

References

1. Castronova, E. *Synthetic Worlds: The Business and Culture of Online Games* (University of Chicago Press, 2005).
2. Feynman, R. Simulating physics with computers. *Int J Theor Phys* (1982).
3. Palmore, J. & Herring, C. Computer arithmetic, chaos and fractals. *Physica D* (1990).