Chapter 20

The Fastest Way of Computing All Universes

Jürgen Schmidhuber

IDSIA, Galleria 2, 6928 Manno-Lugano, Switzerland
University of Lugano & SUPSI, Switzerland.

Is there a short and fast program that can compute the precise history of our universe, including all seemingly random but possibly actually deterministic and pseudo-random quantum fluctuations? There is no physical evidence against this possibility. So let us start searching! We already *know* a short program that computes all constructively computable universes in parallel, each in the asymptotically fastest way. Assuming ours is computed by this optimal method, we can predict that it is among the fastest compatible with our existence. This yields testable predictions.

Note: This paper extends an overview of previous work[51–54,58,59] presented in a survey for the German edition of Scientific American.[61]

1. Introduction

In the 1940s, Konrad Zuse already speculated that our universe is computable by a deterministic computer program (Horst Zuse, personal communication, 2006), like the virtual worlds of today's video games. In 1967 he published the first scientific paper on this idea,[77] soon to be followed by his book *Calculating Space*,[78] focusing on cellular automata as computational devices. We shall see that contrary to common belief, Zuse's hypothesis is compatible with all known observations of quantum physics. Since computable universes are much simpler than non-computable ones, and since one should prefer simple explanations over complex ones, we shall accept his hypothesis as long as there is no evidence to the contrary.

Somewhat surprisingly, there must then exist a very short and in a sense optimally fast algorithm that not only computes the entire history of our own universe, but also those of all other logically possible universes. If the computation of our world indeed is indeed based on such an optimal method, then we may derive non-trivial predictions about its future. I will also briefly discuss some philosophical and theological consequences of this view.

2. Simplicity and Complexity

An object is simple if it has a short description that can be quickly transcribed into the object. For example, the image of a fractal structure[42] may seem complex due to its wealth of detail. But in reality it is simple, as it can be completely generated by a very short and fast program. Therefore it has low algorithmic information or Kolmogorov complexity, defined as the length of the shortest program that computes it.[2,14–16,18,24,25,34–36,38,39,53,66,70,71,76,79] This length hardly depends on the chosen programming language, since programs written in one language can be translated into equivalent programs of another language through a compiler[26,75] of constant, program-independent size.

Is the past and future history of our entire universe simple or complex in this sense? Is there perhaps a very short program that calculates it, including us as observers? This program would have to yield not only the known physical laws but also determine and explain every single seemingly random elementary event. The noblest goal of physics would be to find it.

3. No Problems with Non-Computable Real Numbers

Or is the universe perhaps not computable at all, because it somehow contains or depends on non-computable numbers? As of today there is no compelling reason whatsoever to assume that.

Most physicists are indeed convinced that the universe is quantized by smallest discrete units of time and space and energy. On the other hand, they like to predict macroscopic phenomena using calculus based on the axioms of real numbers, and most real numbers are not even computable (because there are uncountably many real numbers,[12] but only countably many finite programs, such as the non-halting program computing all digits of π). Even quantum physicists who are ready to give up the assumption of a continuous universe usually do take for granted the existence of continuous probability distributions on their discrete universes, and Stephen Hawking explicitly said: *"Although there have been suggestions that space-time may have a discrete structure I see no reason to abandon the continuum theories that have been so successful."* Note, however, that all physicists in fact have only manipulated discrete symbols, thus generating finite, describable proofs of their results derived from enumerable axioms. That real numbers really *exist* in a way transcending the finite symbol strings used by everybody may be a figment of imagination[52] — compare Brouwer's constructive mathematics[3,7] and the Löwenheim-Skolem Theorem[41,68] which implies that any first order theory with an uncountable model such as the real numbers also has a countable model. As Kronecker put it: *"Die ganze Zahl schuf der liebe Gott, alles Übrige ist Menschenwerk"* ("God created the integers, all else is the work of man"[10]). Kronecker greeted with scepticism Cantor's celebrated insight[12] that there are uncountably many real numbers, mathematical objects Kronecker believed did not even exist.

Anyway, calculus does yield very good macro-level approximations of whatever discrete computable processes may really be happening on the microscopic level.

4. No Problems with Uncertainty Principle

Obviously the universe at least partially obeys simple program-like rules: apples fall to the ground again and again in similar ways; all electrons apparently act the same. Many quantum physicists, however, believe that the history of the universe also includes an incredible number of principally unpredictable, random events on the quantum level.[58] If that were true, then it would *not* have a short description, since truly random, irregular data has maximal Kolmogorov complexity, being incompressible by definition.

Here physicists like to refer to Werner Heisenberg (1901-1976), whose famous uncertainty principle[31] says that an observer cannot simultaneously precisely determine impulse and location of a physical object. For example, to measure the state of an electron, one needs to shoot other particles at it, thus changing its state. To mathematically quantify the resulting uncertainty, quantum mechanics replaces precise deterministic predictions by probabilistic ones. Many physicists believe this uncertainty to be not only a practical measurement problem, but a fundamental property of nature, claiming that God does not obey Albert Einstein's famous quote: *Gott würfelt nicht (God does not play dice)*. According to this view, history would not be pre-determined, and neither compactly describable nor precisely predictable, not even in principle.

It is possible, however, to imagine a computer-generated, pseudorandom,[20] *totally deterministic* world that makes its inhabitants believe that it is partially random and only partially observable, thanks to

Heisenberg-like observation limits.[58] A hypothetical programmer of this world could interrupt the computation at any time, dump the current storage into a file, and analyze every little detail, including precise impulse and location of every bitstring-encoded elementary particle.[51] Later he could continue the program's execution, without any internal observer even noticing the pause.

5. No Problems with Bell's Inequality

Quantum physics seems weird. Two entangled particles may be separated by light years, but they somehow seem to immediately "feel" whether one of them is measured, yielding a correlated measurement. Einstein viewed this *spooky action at a distance* as a proof of quantum physics' incompleteness.

A famous inequality of John Stewart Bell (1928-1990) shows that if observers and observations are statistically independent in a certain sense, then there is no local physical rule to explain such spooky effects, even if each particle had unknown internal variables to store information about events that occurred when its entangled particle was still close.[4]

In deterministically computable universes, however, Bell's assumption of independent observers and observations is void and irrelevant. Bell himself was well aware of this.[58]

6. Occam and the Search for the Shortest Program

Most scientists appreciate the rule of William Occam (1280-1347): Among all hypothesis explaining the observations, favor the simplest one. In modern terms: Among all programs reproducing or compressing the observations, favor the shortest one. The principle is widely accepted not only in the inductive sciences such as physics,[34,39] but even in the fine arts.[50] I will later sharpen it a bit, taking into account not only program size[2,14–16,18,24,25,34–36,38,39,53,66,70,71,76,79] but also computation time.[54]

7. What Can be Computed Constructively?

So far we have seen that no physical observations contradict Zuse's hypothesis of a computable universe. Even prominent physicists such as 1999 Nobel laureate Gerard 't Hooft take it seriously.[73] Now we have to clarify, however, what exactly is constructively computable at all.

Let us consider traditional computers that take a binary input program such as 10011010100..., process it by an internal mechanism, and produce a growing number of output bits. The output could encode the evolution of some universe, for example, the total space-time of ours, or even an entire multiverse (many parallel, partially interacting universes) in the sense of Everett.[21,22]

Note that a computed universe history does not have to correspond to incrementally computed *local time steps* like in certain examples provided in my previous publications[51–54]—maybe our standard concepts of time do not even make sense in a given computable universe. But we insist that the output yields a complete representation of every detail of the universe or multiverse in question, without any loss of information.

An additional "viewer program" may facilitate the interpretation of output bitstrings, reminiscent of video games that come with a computer graphics interface to visualize bitstrings in the computer's memory which encode game states.

In traditional computer science, each output bit is viewed as being final and unmodifiable. It turns out, however, that many possible output bitstrings (and thus universes) are compactly describable only if we relax this view, and allow non-halting programs[8,9,23,27,30,32,46,48] to edit their former outputs on occasion[52,53] (compare functions in the *arithmetic hierarchy*[48] and the concept of Δ^0_n-describability, e.g., [39, p. 46-47]).

I defined[52,53] the set of *formally describable* or constructively limit-computable bitstrings x: those x that have a (possibly non-halting) finite program *converging* towards x — after some time each bit of x has to stop changing, that is, each prefix of x becomes fixed after finite (but in general unknowable) time.

For example, let us us assume the n-th output bit is 1 if the n-th program in a list of all possible programs halts, where n is a natural number. This output sequence has a very compact input program which systematically enumerates all possible programs and runs them in interleaving fashion; whenever a program in the list (say, the m-th) halts, the m-th output bit (initialized by 0) becomes 1. Every prefix of the infinite output will converge at some point. But we do not know when, otherwise we could solve the generally unsolvable halting problem.[26,75]

It turns out that a given universe such as ours might have a very short explanation or description on a machine that can edit its former outputs, but not on a traditional machine. In fact, there are more or less powerful variants of output-editing machines which vary in their expressiveness, some being able to compactly encode certain universes that need long codes on others. For example, the enumerable "number of wisdom" Ω[11,16,69,72] cannot be compressed on traditional Turing Machines,[75] but on so-called *Enumerable Output Machines*.[45,52,53] These distinctions are technically very important, but not central to the present overview; the interested reader is referred to.[52,53]

8. No Problem with Talk About Incomputable Things

Observers inhabiting a computable universe may talk about mathematical paradoxons and things that are incomputable in a sense, such as the halting probability of a universal Turing machine, which is closely related to Gödel's incompleteness theorem.[11,16,26,69,72,75] This does not involve any inconsistencies.[51] For example, the processes that correspond to our brain firing patterns and the sound waves they provoke by controlling our voices may still correspond to computable substrings of our universe's evolution. The same holds for talk about inconsistent worlds in which, say, time travel is possible.

9. The Fastest Way of Computing All Universes

In 1996 I pointed out that there is a very short algorithm that computes all possible universes, as long as they are computable.[51] In a certain sense this (non-halting) algorithm is also extremely fast, as I emphasized in 2000.[52,54] Let me write down a variant that does not consume excessive storage space (here $l(p)$ denotes the length of program p, a bitstring):

Algorithm 1 Algorithm FAST

  for $i := 1, 2, \ldots$ ***do***
    Run each program p *with* $l(p) \leq i$ *for at most* $2^{i-l(p)}$ *steps and reset storage modified by* p
  end for

That is, in phase i, FAST generates all universes computable by some program p satisfying $l(p) + log\ t(p) \leq i$, where $t(p)$ is the runtime of p, and *log* denotes the binary logarithm. True, phase $i + 1$ will repeat everything done in phase i, but that is not an essential efficiency problem: every phase costs roughly as much as all previous phases taken together, that is, we lose only a factor of 2 or so of computation time, but gain a lot by not having to store all intermediate results of previously executed, only partially finished programs, which would cost exponentially growing storage space.

It is easy to see that FAST will generate the n-th bit of each universe as quickly as if it were computed by this universe's fastest program, save for a constant factor that does not depend on n. Following standard practice of theoretical computer science, we may therefore call FAST the *asymptotically fastest* way of computing all computable universes. For any God-like *Great Programmer*,[51] FAST offers a natural, optimally efficient way of computing all logically possible worlds.

If our universe is one of the computable ones, then FAST will eventually produce a detailed representation of its first few billion years of local time (note that nearly 14 billion years have passed since the big bang).

10. The Fastest and Truest Version of Our World?

Since there are many programs computing one and the same universe (history), our optimal algorithm FAST (Section 9) will generate many copies of ours, and many histories that start like ours (but possibly continue in different ways). At any given time in the execution of FAST, the most advanced copies will be those computable by short and fast programs. Since we exist, we already know that at least one of the programs has computed enough to enable our existence, following the weak anthropic principle.[1,13] But which of the many? A little bit of thought shows: With high probability it will be one of the shortest and fastest compatible with our existence! For a more detailed analysis, see previous work.[52,54]

Following this argumentation, we are already part of one of the simplest, fastest, non-random worlds compatible with our very being, simply because even the optimal FAST needs much more time to compute truly random events as parts of any universe's history. Computationally, randomness is extremely expensive in terms of both time and space. It does not fit the Occam's razor criterion at all.

But even if our universe's history included a huge number of truly random quantum events, one question would arise immediately: Besides the physical laws, which is the simplest and fastest pseudo-random generator needed to compute a *similar*, less random world? In a philosophical sense, wouldn't this world be the *truest* version of our world, reflecting its true essence, thanks to its lack of arbitrariness?

11. Predictions Based on the Fastest Way

If whoever is generating our universe is using algorithm FAST (Section 9) to deal with computational resource constraints in an optimal way, we can make non-trivial predictions.

For example, all seemingly random events (such as beta decay of neutrons) actually must follow some pseudo-random rule, waiting to be discovered by some grad student at CERN or elsewhere. Perhaps current physicists are like observers seeing the second billion digits in the decimal expansion of π, which at first glance look very random (for example, every 3 digit sequence occurs roughly once in a thousand 3 digit subsequences) but is actually highly regular, since it can be computed by a short program.

One somewhat depressing prediction is that quantum computation, a subject of much current excitement,[5,19,40,44] will never work well and never scale to large problems. Sure, FAST will run many programs that compute multiverse-like universes, obeying known laws of quantum mechanics and allowing for quantum computers (which can be fully simulated on traditional computers). However, the FAST-generated programs that compute our history so far *and* permit the expected effects of quantum computing will cost much more computational effort than others that are also computing our history in a less computationally expensive way. That is, under FAST they are very unlikely. That is, it is very unlikely that we are inhabiting a multiverse where quantum computation will be able to solve non-trivial problems. A pity!

I first made this prediction a decade ago.[52] Since then, nobody has been able to make quantum computation scale. For example, the biggest number to be factored by any existing quantum computer is still 15.

12. How to Find our Universe's Program

Algorithm FAST (Section 9) computes all universes, not just ours. But what we'd really like to know is the program that computes ours and nothing else. That would be the world's essential formula, the holy grail of theoretical physics. How to find it? It turns out that the optimal way of searching for it is closely related to FAST. It goes like this:

Take any sequence of physical observations, and run FAST until one of the executed programs (written in a universal programming language) reproduces the data.

This is essentially Levin's universal search algorithm[37] applied to physics. Since it is only asymptotically optimal, it can be greatly accelerated under certain conditions by methods such as the Optimal Ordered Problem Solver,[56] which may use *partial*, incomplete reproductions of the data as intermediate subgoals, and then continue the search by re-using previous subgoal-achieving programs, thus possibly dramatically reducing the constant slowdown ignored by the asymptotic notion of optimality.[56]

13. Always Slower Than the Universe Itself

Note that if somebody indeed found the shortest and fastest program of our world, this would not necessarily help to figure out the future faster than by waiting for it happen. The computer on which to run this program would have to be built within our universe, and as a small part of the latter would be unable to run as fast as the universe itself.

14. Math v Computation?

Rather than pursuing the computability-oriented path layed out in,[51] Tegmark (back then at LMU Munich) suggested what at first glance seems to be an alternative ensemble of possible universes based on an informally defined set of "self-consistent mathematical structures"[74] — compare also Marchal's and Bostrom's theses.[6,43] It is not quite clear whether Tegmark wanted to include universes that are *not* formally describable according to our definition mentioned in Section 7. It is well-known, however, that for any set of mathematical axioms there is a program that lists all provable theorems in order of the lengths of their shortest proofs encoded as bitstrings. Hence Tegmark's view[74] seems in a certain sense encompassed by the algorithmic approach.[51] The latter offers several conceptual advantages though: (1) It provides the appropriate framework for issues of information-theoretic complexity traditionally ignored in pure mathematics, and imposes natural complexity-based orderings on the possible universes and subsets thereof.[51–53] (2) It taps into a rich source of theoretical insights on computable probability distributions relevant for establishing priors on possible universes. Such priors are needed for making probabilistic predictions concerning our own particular universe.[51–53] Although Tegmark suggests that "*... all mathematical structures are a priori given equal statistical weight*" (Ref. 74, p. 27), there is no way of assigning equal nonvanishing probability to all (infinitely many) mathematical structures. Hence we really need something like the complexity-based weightings discussed in in earlier papers.[51–53] (3) The algorithmic approach is the obvious framework for questions of temporal complexity such as those discussed in this paper, e.g., "what is the most efficient way of simulating all universes?"[52,54]

15. Optimal Artificial Intelligence in Computable Universes

The fully self-referential[26] Gödel machine[63] is a Universal Artificial Intelligence (AI)[55,57,60,62,64] that is at least theoretically optimal in a certain sense. It may interact with some initially unknown, partially observable environment to maximize future expected utility or reward by solving arbitrary user-defined computational tasks. Its initial algorithm is not hardwired; it can completely rewrite itself without essential limits apart from the limits of computability, provided a proof searcher embedded within the initial algorithm can first prove that the rewrite is useful, according to the formalized utility function taking into account the limited computational resources. Self-rewrites may modify / improve the proof searcher itself, and can be shown to be *globally optimal*, relative to Gödel's well-known fundamental restrictions of provability.[26] To make sure the Gödel machine is at least *asymptotically* optimal even before the first self-rewrite, we may initialize it by Hutter's non-self-referential but *asymptotically fastest algorithm for all well-defined problems* Hsearch,[33] which uses a hardwired brute force proof searcher and (justifiably) ignores the costs of proof search. Assuming discrete input/output domains $X/Y \subset B^*$, a formal problem specification $f : X \to Y$ (say, a functional description of how integers are decomposed into their prime factors), and a particular $x \in X$ (say, an integer to be factorized), Hsearch orders all proofs of an appropriate axiomatic system by size to find programs q that for all $z \in X$ provably compute $f(z)$ within time bound $t_q(z)$. Simultaneously it spends most of its time on executing the q with the best currently proven time bound $t_q(x)$. Remarkably, Hsearch is as fast as the *fastest* algorithm that provably computes $f(z)$ for all $z \in X$, save for a constant factor smaller than $1+\epsilon$ (arbitrary real-valued $\epsilon > 0$) and an f-specific but x-independent additive constant.[33] Given some problem, the Gödel machine may decide to replace its Hsearch initialization by a faster method suffering less from large constant overhead, but even if it doesn't, its performance won't be less than asymptotically optimal.

All of this implies that there already exists the blueprint of a Universal AI which will solve almost all problems almost as quickly as if it already knew the best (unknown) algorithm for solving them, because almost all imaginable problems are big enough to make the additive constant negligible.

The only motivation for *not* quitting computer science research right now is that many real-world problems are so small and simple that the ominous constant slowdown (potentially relevant at least before the first Gödel machine self-rewrite) is *not* negligible. Nevertheless, the ongoing efforts at scaling universal AIs down to the rather few *small* problems are very much informed by the new millennium's theoretical insights[57,60,62,64] mentioned above, and may soon yield practically feasible yet still general problem solvers for physical systems with highly restricted computational power, say, a few trillion instructions per second, roughly comparable to a human brain power.

Simultaneously, our non-universal but still rather general fast deep / recurrent neural networks have already started to outperform traditional pre-programmed methods: they recently collected a string of 1st ranks in many important visual pattern recognition benchmarks, e.g., IJCNN traffic sign competition, NORB, CIFAR10, MNIST, three ICDAR handwriting competitions.[17,29,65] Here we greatly profit from ongoing advances in computing hardware, using GPUs (mini-supercomputers normally used for video games) 100 times faster than today's CPU cores, and a million times faster than PCs of 20 years ago, complementing the recent abovementioned progress in the theory of mathematically optimal universal problem solvers.[65]

16. Potential Criticism

Philosophers tend to create theories inspired by recent scientific developments. For instance, Heisenberg's uncertainty principle and Gödel's incompleteness theorem greatly influenced modern philosophy. Are algorithmic Theories of Everything (TOEs) and the "Great Programmer Religion"[51,52] just another reaction to recent developments, some in hindsight obvious byproduct of the advent of good virtual reality? (As they say: *For a man with a hammer, everything looks like a nail.*) Will they soon become obsolete, as so many previous philosophies? I find it hard to imagine so, even without a boost to be expected for algorithmic TOEs in case someone should indeed discover a simple subroutine responsible for certain physical events hitherto believed to be irregular. After all, algorithmic theories of the describable do encompass everything we will ever be able to talk and write about. Other things are simply beyond description.

17. Who Can Accept the Computable Real World?

Many researchers in the field of artificial life simulate the evolution of artificial beings adapting to their artificial environments, e.g.,.[28,47,67] Most of them do not have any problems with the idea of a computable real world in Zuse's sense. After all, a good simulation is not distinguishable from reality. Children with experience in virtual realities and video games also tend to find the idea of a computable universe more natural than their parents.

In our universe the raw computational power per cent will keep increasing by a factor of 100-1000 per decade, with no end in sight. As a consequence, realism and appeal of virtual realities will keep increasing dramatically, making the presented thoughts[51–54] more and more acceptable for the masses.

Remarkably, it is especially the quantum physicists who sometimes reject such ideas,[58] albeit without being able to justify their scepticism too well by facts.

Einstein, perhaps the greatest of all physicists, did not believe in nondeterminism, as already mentioned. For a long time his view has been unpopular among quantum physicists. But now it does not seem unreasonable to predict that it will experience a rennaissance. First, because there is no physical evidence against it. Second, because it greatly simplifies the description of the world's history in the framework of computability theory, without necessitating a gigantic amount of information for describing a vast number of truly random quantum-level events.

As long as nobody can show that the universe is indeed partially random, scientists are obliged to search for a short program that computes all the apparent randomness and therefore reveals it as pseudo-randomness.[58] If the process that calculates us makes optimally efficient use of the resources of some higher-level universe, we should expect this program to be not only short but also fast.[52,54]

18. Consequences for Philosophy and Theology

The theory of computable universes provides a purely rational and technologically oriented access to basic questions of philosophy and theology.[51]

At least in principle, everybody could become some sort of God by programming the algorithm FAST (Section 9) on a computer, systematically creating all constructively computable universes, including ours.

In some of them, programmers occasionally will intervene in the worlds computed by their programs, reminiscent of well-known religious role models. In some, the computable contents of simulated brains will be occasionally copied from one computable world to another, implementing variants of heaven or hell.

Beings evolved in some of the simulated universes will again build computers to simulate universes, in recursively nested fashion.[51] This begs the question: Where does the computer of the top universe in the hierarchy come from? It must remain open for now.

The fact that there are mathematically optimal ways of creating and computing all the logically possible worlds, however, opens a new and exciting field hardly discussed in today's mainstream philosophy and theology.

19. Acknowledgments

At the age of 17 my brother Christof Schmidhuber declared that the universe is *sum of all math*, inhabited by observers who are mathematical substructures (private communication, Munich, 1981). As he went on to become a theoretical physicist at LMU Munich, discussions with him about the relation between superstrings and bitstrings became a source of inspiration for writing both the first paper[51] and later ones[52–54] based on computational complexity theory, which seems to provide the natural setting for his more math-oriented ideas (private communication, Munich 1981-86; Caltech 1987-93; Princeton 1994-96; Berne/Geneva 1997–; compare his notion of *"mathscape"*[49]). I believe that Christof's early discussions with Munich-based scientists and students were the reason why such ideas emerged in Munich. Furthermore, his 1997 remarks on similarities and differences between Feynman path integrals and "the sum of all computable universes" and his resulting dissatisfaction with the lack of a discussion of temporal aspects in the 1996 paper[51] triggered the papers[52,54] on temporal complexity.

References

1. J. D. Barrow and F. J. Tipler. *The Anthropic Cosmological Principle*. Clarendon Press, Oxford, 1986.
2. Y. M. Barzdin. Algorithmic information theory. In *Encyclopaedia of Mathematics*, volume 1, pages 140–142. Reidel, Kluwer Academic Publishers, 1988.
3. M. Beeson. *Foundations of Constructive Mathematics*. Springer-Verlag, Heidelberg, 1985.

4. J. S. Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38:447–452, 1966.
5. C. H. Bennett and D. P. DiVicenzo. Quantum information and computation. *Nature*, 404(6775):256–259, 2000.
6. N. Bostrom. Observational selection effects and probability. Dissertation, Dept. of Philosophy, Logic and Scientific Method, London School of Economics, 2000.
7. L. E. J. Brouwer. Over de Grondslagen der Wiskunde. Dissertation, Doctoral Thesis, University of Amsterdam, 1907.
8. M. S. Burgin. Inductive Turing machines. *Notices of the Academy of Sciences of the USSR (translated from Russian)*, 270(6):1289–1293, 1991.
9. M. S. Burgin and Y. M. Borodyanskii. Infinite processes and super-recursive algorithms. *Notices of the Academy of Sciences of the USSR (translated from Russian)*, 321(5):800–803, 1991.
10. F. Cajori. *History of mathematics (2nd edition)*. Macmillan, New York, 1919.
11. C. S. Calude. Chaitin Ω numbers, Solovay machines and Gödel incompleteness. *Theoretical Computer Science*, 2001.
12. G. Cantor. Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Crelle's Journal für Mathematik*, 77:258–263, 1874.
13. B. Carter. Large number coincidences and the anthropic principle in cosmology. In M. S. Longair, editor, *Proceedings of the IAU Symposium 63*, pages 291–298. Reidel, Dordrecht, 1974.
14. G. J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. *Journal of the ACM*, 16:145–159, 1969.
15. G. J. Chaitin. A theory of program size formally identical to information theory. *Journal of the ACM*, 22:329–340, 1975.
16. G. J. Chaitin. *Algorithmic Information Theory*. Cambridge University Press, Cambridge, 1987.
17. D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In *Intl. Joint Conference on Artificial Intelligence IJCAI*, pages 1237–1242, 2011.
18. T. M. Cover, P. Gács, and R. M. Gray. Kolmogorov's contributions to information theory and algorithmic complexity. *Annals of Probability Theory*, 17:840–865, 1989.
19. D. Deutsch. *The Fabric of Reality*. Allen Lane, New York, NY, 1997.
20. T. Erber and S. Putterman. Randomness in quantum mechanics – nature's ultimate cryptogram? *Nature*, 318(7):41–43, 1985.
21. H. Everett III. 'Relative State' formulation of quantum mechanics. *Reviews of Modern Physics*, 29:454–462, 1957.
22. H. Everett, III. The many-worlds interpretation of quantum mechanics. Princeton University Press, Princeton, 1986.
23. R. V. Freyvald. Functions and functionals computable in the limit. *Transactions of Latvijas Vlasts Univ. Zinatn. Raksti*, 210:6–19, 1977.
24. P. Gács. On the symmetry of algorithmic information. *Soviet Math. Dokl.*, 15:1477–1480, 1974.

25. P. Gács. On the relation between descriptional complexity and algorithmic probability. *Theoretical Computer Science*, 22:71–93, 1983.
26. K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.
27. E. M. Gold. Limiting recursion. *Journal of Symbolic Logic*, 30(1):28–46, 1965.
28. F. J. Gomez, J. Schmidhuber, and R. Miikkulainen. Efficient non-linear control through neuroevolution. *Journal of Machine Learning Research JMLR*, 9:937–965, 2008.
29. Alex Graves and Jüergen Schmidhuber. Offline handwriting recognition with multidimensional recurrent neural networks. In *Advances in Neural Information Processing Systems 21*. MIT Press, Cambridge, MA, 2009.
30. A. Gregorczyk. On the definitions of computable real continuous functions. *Fundamenta Mathematicae*, 44:61–71, 1957.
31. W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 33:879–893, 1925.
32. G. Hotz, G. Vierke, and B. Schieffer. Analytic machines. Technical Report TR95-025, Electronic Colloquium on Computational Complexity, 1995. http://www.eccc.uni-trier.de/eccc/.
33. M. Hutter. The fastest and shortest algorithm for all well-defined problems. *International Journal of Foundations of Computer Science*, 13(3):431–443, 2002. (On J. Schmidhuber's SNF grant 20-61847).
34. M. Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin, 2005. (On J. Schmidhuber's SNF grant 20-61847).
35. A. N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:1–11, 1965.
36. L. A. Levin. On the notion of a random sequence. *Soviet Math. Dokl.*, 14(5):1413–1416, 1973.
37. L. A. Levin. Universal sequential search problems. *Problems of Information Transmission*, 9(3):265–266, 1973.
38. L. A. Levin. Laws of information (nongrowth) and aspects of the foundation of probability theory. *Problems of Information Transmission*, 10(3):206–210, 1974.
39. M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications (2nd edition)*. Springer, 1997.
40. S. Lloyd. Ultimate physical limits to computation. *Nature*, 406:1047–1054, 2000.
41. L. Löwenheim. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, 76:447–470, 1915.
42. B. Mandelbrot. *The Fractal Geometry of Nature*. Freeman and Co., San Francisco, 1982.
43. B. Marchal. *Calculabilité, Physique et Cognition*. PhD thesis, L'Université des Sciences et Technologies De Lilles, 1998.
44. R. Penrose. *The Emperor's New Mind*. Oxford University Press, 1989.
45. J. Poland. A coding theorem for enumerable output machines. *Information*

Processing Letters, 91(4):157–161, 2004.
46. H. Putnam. Trial and error predicates and the solution to a problem of Mostowski. *Journal of Symbolic Logic*, 30(1):49–57, 1965.
47. T. S. Ray. An approach to the synthesis of life. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 371–408. Addison Wesley Publishing Company, 1992.
48. H. Rogers, Jr. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, New York, 1967.
49. C. Schmidhuber. Strings from logic. Technical Report CERN-TH/2000-316, CERN, Theory Division, 2000. http://xxx.lanl.gov/abs/hep-th/0011065.
50. J. Schmidhuber. Low-complexity art. *Leonardo, Journal of the International Society for the Arts, Sciences, and Technology*, 30(2):97–103, 1997.
51. J. Schmidhuber. A computer scientist's view of life, the universe, and everything. In C. Freksa, M. Jantzen, and R. Valk, editors, *Foundations of Computer Science: Potential - Theory - Cognition*, volume 1337, pages 201–208. Lecture Notes in Computer Science, Springer, Berlin, 1997, submitted 1996.
52. J. Schmidhuber. Algorithmic theories of everything. Technical Report IDSIA-20-00, quant-ph/0011122, IDSIA, Manno (Lugano), Switzerland, 2000. Sections 1-5: see Ref. 53; Section 6: see Ref. 54.
53. J. Schmidhuber. Hierarchies of generalized Kolmogorov complexities and nonenumerable universal measures computable in the limit. *International Journal of Foundations of Computer Science*, 13(4):587–612, 2002.
54. J. Schmidhuber. The Speed Prior: a new simplicity measure yielding near-optimal computable predictions. In J. Kivinen and R. H. Sloan, editors, *Proceedings of the 15th Annual Conference on Computational Learning Theory (COLT 2002)*, Lecture Notes in Artificial Intelligence, pages 216–228. Springer, Sydney, Australia, 2002.
55. J. Schmidhuber. Towards solving the grand problem of AI. In P. Quaresma, A. Dourado, E. Costa, and J. F. Costa, editors, *Soft Computing and complex systems*, pages 77–97. Centro Internacional de Mathematica, Coimbra, Portugal, 2003. Based on Ref. 57.
56. J. Schmidhuber. Optimal ordered problem solver. *Machine Learning*, 54:211–254, 2004.
57. J. Schmidhuber. The new AI: General & sound & relevant for physics. In B. Goertzel and C. Pennachin, editors, *Artificial General Intelligence*, pages 175–198. Springer, 2006. Also available as TR IDSIA-04-03, arXiv:cs.AI/0302012.
58. J. Schmidhuber. Randomness in physics. *Nature*, 439(3):392, 2006. Correspondence.
59. J. Schmidhuber. The Computational Universe. Review of *Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos*, by S. Lloyd. *American Scientist*, 2006.
60. J. Schmidhuber. 2006: Celebrating 75 years of AI - history and outlook: the next 25 years. In M. Lungarella, F. Iida, J. Bongard, and R. Pfeifer, editors, *50 Years of Artificial Intelligence*, volume LNAI 4850, pages 29–41. Springer

Berlin / Heidelberg, 2007. Preprint available as arXiv:0708.4311.

61. J. Schmidhuber. Alle berechenbaren Universen (All computable universes). *Spektrum der Wissenschaft Spezial (German edition of Scientific American)*, (3):75–79, 2007.
62. J. Schmidhuber. New millennium AI and the convergence of history. In W. Duch and J. Mandziuk, editors, *Challenges to Computational Intelligence*, volume 63, pages 15–36. Studies in Computational Intelligence, Springer, 2007. Also available as arXiv:cs.AI/0606081.
63. J. Schmidhuber. Ultimate cognition *à la* Gödel. *Cognitive Computation*, 1(2):177–193, 2009.
64. J. Schmidhuber. The new AI is general and mathematically rigorous. *Front. Electr. Electron. Eng. China*, 2010.
65. J. Schmidhuber, D. Ciresan, U. Meier, J. Masci, and A. Graves. On fast deep nets for AGI vision. In *Fourth Conference on Artificial General Intelligence (AGI)*, 2011.
66. C. P. Schnorr. Process complexity and effective random tests. *Journal of Computer Systems Science*, 7:376–388, 1973.
67. K. Sims. Evolving virtual creatures. In Andrew Glassner, editor, *Proceedings of SIGGRAPH '94 (Orlando, Florida, July 1994)*, Computer Graphics Proceedings, Annual Conference, pages 15–22. ACM SIGGRAPH, ACM Press, jul 1994. ISBN 0-89791-667-0.
68. T. Skolem. Logisch-kombinatorische Untersuchungen über Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen. *Skrifter utgit av Videnskapsselskapet in Kristiania, I, Mat.-Nat. Kl.*, N4:1–36, 1919.
69. T. Slaman. Randomness and recursive enumerability. Technical report, Univ. of California, Berkeley, 1999. Preprint, http://www.math.berkeley.edu/~slaman.
70. R. J. Solomonoff. A formal theory of inductive inference. Part I. *Information and Control*, 7:1–22, 1964.
71. R. J. Solomonoff. Complexity-based induction systems. *IEEE Transactions on Information Theory*, IT-24(5):422–432, 1978.
72. R. M. Solovay. A version of Ω for which ZFC can not predict a single bit. In C. S. Calude and G. Păun, editors, *Finite Versus Infinite. Contributions to an Eternal Dilemma*, pages 323–334. Springer, London, 2000.
73. G. 't Hooft. Quantum gravity as a dissipative deterministic system. Technical Report SPIN-1999/07/gr-gc/9903084, http://xxx.lanl.gov/abs/gr-qc/9903084, Institute for Theoretical Physics, Univ. of Utrecht, and Spinoza Institute, Netherlands, 1999. Also published in *Classical and Quantum Gravity 16*, 3263.
74. M. Tegmark. Is "the theory of everything" merely the ultimate ensemble theory? *Annals of Physics*, 270:1–51, 1998.
75. A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, Series 2*, 41:230–267, 1936.
76. V. A. Uspensky. Complexity and entropy: an introduction to the theory of

Kolmogorov complexity. In O. Watanabe, editor, *Kolmogorov complexity and computational complexity*, pages 85–102. EATCS Monographs on Theoretical Computer Science, Springer, 1992.

77. K. Zuse. Rechnender Raum. *Elektronische Datenverarbeitung*, 8:336–344, 1967.
78. K. Zuse. *Rechnender Raum*. Friedrich Vieweg & Sohn, Braunschweig, 1969. English translation: *Calculating Space,* MIT Technical Translation AZT-70-164-GEMIT, Massachusetts Institute of Technology (Proj. MAC), Cambridge, Mass. 02139, Feb. 1970.
79. A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the algorithmic concepts of information and randomness. *Russian Math. Surveys*, 25(6):83–124, 1970.