

# Mathematics in the Age of the Turing Machine

Thomas C. Hales\*

University of Pittsburgh  
hales@pitt.edu



**Fig. 1.** Alan Turing (image source [Ima13])

*“And when it comes to mathematics, you must realize that this is the human mind at the extreme limit of its capacity.” (H. Robbins)*

*“... so reduce the use of the brain and calculate!” (E. W. Dijkstra)*

*“The fact that a brain can do it seems to suggest that the difficulties [of trying with a machine] may not really be so bad as they now seem.” (A. Turing)*

---

\* Research supported in part by NSF grant 0804189 and the Benter Foundation.

## 1 Computer Calculation

### 1.1 a panorama of the status quo

Where stands the mathematical endeavor?

In 2012, many mathematical utilities are reaching consolidation. It is an age of large aggregates and large repositories of mathematics: the arXiv, Math Reviews, and euDML, which promises to aggregate the many European archives such as Zentralblatt Math and Numdam. Sage aggregates dozens of mathematically oriented computer programs under a single Python-scripted front-end.

Book sales in the U.S. have been dropping for the past several years. Instead, online sources such as Wikipedia and Math Overflow are rapidly becoming students' preferred math references. The Polymath blog organizes massive mathematical collaborations. Other blogs organize previously isolated researchers into new fields of research. The slow, methodical deliberations of referees in the old school are giving way; now in a single stroke, Tao blogs, gets feedback, and publishes.

Machine Learning is in its ascendancy. *LogAnswer* and *Wolfram Alpha* answer our elementary questions about the quantitative world; *Watson* our *Jeopardy* questions. *Google Page* ranks our searches by calculating the largest eigenvalue of the largest matrix the world has ever known. *Deep Blue* plays our chess games. The million-dollar-prize-winning *Pragmatic Chaos* algorithm enhances our *Netflix searches*. The major proof assistants now contain tens of thousands of formal proofs that are being mined for hints about how to prove the next generation of theorems.

Mathematical models and algorithms rule the quantitative world. Without applied mathematics, we would be bereft of Shor's factorization algorithm for quantum computers, Yang-Mills theories of strong interactions in physics, invisibility cloaks, Radon transforms for medical imaging, models of epidemiology, risk analysis in insurance, stochastic pricing models of financial derivatives, RSA encryption of sensitive data, Navier-Stokes modeling of fluids, and models of climate change. Without it, entire fields of engineering from Control Engineering to Operations Research would close their doors. The early icon of mathematical computing, Von Neumann, divided his final years between meteorology and hydrogen bomb calculations. Today, applications fuel the economy: in 2011 rankings, the first five of the "10 best jobs" are math or computer related: software engineer, mathematician, actuary, statistician, and computer systems analyst [CC111].

Computers have rapidly become so pervasive in mathematics that future generations may look back to this day as a golden dawn. A comprehensive survey is out of the question. It would almost be like asking for a summary of applications of symmetry to mathematics. Computability – like symmetry – is a wonderful structural property that some mathematical objects possess that makes answers flow more readily wherever it is found. This section gives many examples that give a composite picture of computers in mathematical research, showing that computers are neither the panacea that the public at large might imagine, nor the evil that the mathematical purist might fear. I have deliberately selected many examples from pure mathematics, partly because of my own background and partly to correct the conventional wisdom that couples computers with applied mathematics and blackboards with pure mathematics.

## 1.2 Birch and Swinnerton-Dyer conjecture

I believe that the Birch and Swinnerton-Dyer conjecture is the deepest conjecture ever to be formulated with the help of a computer [BSD65]. The Clay Institute has offered a one-million dollar prize to anyone who settles it.

Let  $E$  be an elliptic curve defined by an equation  $y^2 = x^3 + ax + b$  over the field of rational numbers. Motivated by related quantities in Siegel's work on quadratic forms, Birch and Swinnerton-Dyer set out to estimate the quantity

$$\prod N_p/p, \quad (1)$$

where  $N_p$  is the number of rational points on  $E$  modulo  $p$ , and the product extends over primes  $p \leq P$  [Bir02]. Performing experiments on the EDSAC II computer at the Computer laboratory at Cambridge University during the years 1958–1962, they observed that as  $P$  increases, the products (1) grow asymptotically in  $P$  as

$$c(E) \log^r P,$$

for some constant  $c$ , where  $r$  is the Mordell-Weil rank of  $E$ ; that is, the maximum number of independent points of infinite order in the group  $E(\mathbb{Q})$  of rational points. Following the suggestions of Cassels and Davenport, they reformulated this numerical asymptotic law in terms of the zeta function  $L(E, s)$  of the elliptic curve. Thanks to the work of Wiles and subsequent extensions of that work, it is known that  $L(E, s)$  is an entire function of the complex variable  $s$ . The Birch and Swinnerton-Dyer conjecture asserts that the rank  $r$  of an elliptic curve over  $\mathbb{Q}$  is equal to the order of the zero of  $L(E, s)$  at  $s = 1$ .

A major (computer-free) recent theorem establishes that the Birch and Swinnerton-Dyer conjecture holds for a positive proportion of all elliptic curves over  $\mathbb{Q}$  [BS10]. This result, although truly spectacular, is mildly misleading in the sense that the elliptic curves of high rank rarely occur but pose the greatest difficulties.

## 1.3 Sato-Tate

The Sato-Tate conjecture is another major conjecture about elliptic curves that was discovered by computer. If  $E$  is an elliptic curve with rational coefficients

$$y^2 = x^3 + ax + b,$$

then the number of solutions modulo a prime number  $p$  (including the point at infinity) has the form

$$1 + p - 2\sqrt{p} \cos \theta_p.$$

for some real number  $0 \leq \theta_p \leq \pi$ . In 1962, Sato, Nagashima, and Namba made calculations of  $\theta_p$  on a Hitachi HIPAC 103 computer to understand how these numbers are distributed as  $p$  varies for a fixed elliptic curve  $E$  [Sch]. By the spring of 1963, the evidence suggested  $\sin^2 \theta$  as a good fit of the data (Figure 2). That is, if  $P(n)$  is the set of the first  $n$  primes, and  $f : [0, \pi] \rightarrow \mathbb{R}$  is any smooth test function, then for large  $n$ ,

$$\frac{1}{n} \sum_{p \in P(n)} f(\theta_p) \quad \text{tends to} \quad \frac{2}{\pi} \int_0^\pi f(\theta) \sin^2 \theta \, d\theta.$$

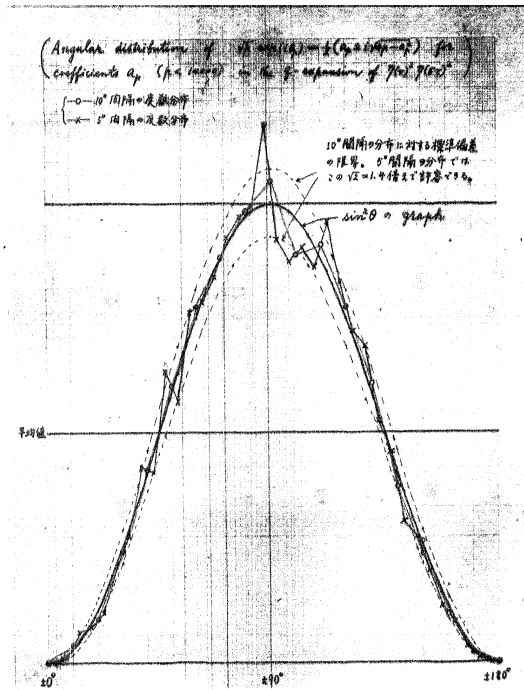


Fig. 2. Data leading to the Sato-Tate conjecture (image source [SN])

The Sato-Tate conjecture (1963) predicts that this same distribution is obtained, no matter the elliptic curve, provided the curve does not have complex multiplication. Tate, who arrived at the conjecture independently, did so without computer calculations.

Serre interpreted Sato-Tate as a generalization of Dirichlet's theorem on primes in arithmetic progression, and gave a proof strategy of generalizing the analytic properties of  $L$ -functions used in the proof of Dirichlet's theorem [Ser68]. Indeed, a complete proof of Sato-Tate conjecture has now been found and is based on extremely deep analytic properties of  $L$ -functions [Car07]. The proof of the Sato-Tate conjecture and its generalizations has been one of the most significant recent advances in number theory.

#### 1.4 transient uses of computers

It has become common for problems in mathematics to be first verified by computer and later confirmed without them. Some examples are the construction of sporadic groups, counterexamples to a conjecture of Euler, the proof of the Catalan conjecture, and the discovery of a formula for the binary digits of  $\pi$ .

Perhaps the best known example is the construction of sporadic groups as part of the monumental classification of finite simple groups. The sporadic groups are the 26 finite simple groups that do not fall into natural infinite families. For example, Lyons

(1972) predicted the existence of a sporadic group of order

$$2^8 \cdot 3^7 \cdot 5^6 \cdot 7 \cdot 11 \cdot 31 \cdot 37 \cdot 67.$$

In 1973, Sims proved the existence of this group in a long unpublished manuscript that relied on many specialized computer programs. By 1999, the calculations had become standardized in group theory packages, such as GAP and Magma [HS99]. Eventually, computer-free existence and uniqueness proofs were found [MC02], [AS92].

Another problem in finite group theory with a computational slant is the inverse Galois problem: is every subgroup of the symmetric group  $S_n$  the Galois group of a polynomial of degree  $n$  with rational coefficients? In the 1980s Malle and Matzat used computers to realize many groups as Galois groups [MM99], but with an infinite list of finite groups to choose from, non-computational ideas have been more fruitful, such as Hilbert irreducibility, rigidity, and automorphic representations [KLS08].

Euler conjectured (1769) that a fourth power cannot be the sum of three positive fourth powers, that a fifth power cannot be the sum of four positive fifth powers, and so forth. In 1966, a computer search [LP66] on a CDC 6600 mainframe uncovered a counterexample

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5,$$

which can be checked by hand (I dare you). The two-sentence announcement of this counterexample qualifies as one of the shortest mathematical publications of all times. Twenty years later, a more subtle computer search gave another counterexample [Elk88]:

$$2682440^4 + 15365639^4 + 18796760^4 = 20615673^4.$$

The Catalan conjecture (1844) asserts that the only solution to the equation

$$x^m - y^n = 1,$$

in positive integers  $x, y, m, n$  with exponents  $m, n$  greater than 1 is the obvious

$$3^2 - 2^3 = 1.$$

That is, 8 and 9 are the only consecutive positive perfect powers. By the late 1970s, Baker's methods in diophantine analysis had reduced the problem to an astronomically large and hopelessly infeasible finite computer search. Mihăilescu's proof (2002) of the Catalan conjecture made light use of computers (a one-minute calculation), and later the computer calculations were entirely eliminated [Mih04], [Met03].

Bailey, Borwein, and Plouffe found an algorithm for calculating the  $n$ th binary digit of  $\pi$  directly: it jumps straight to the  $n$ th digit without first calculating any of the earlier digits. They understood that to design such an algorithm, they would need an infinite series for  $\pi$  in which powers of 2 controlled the denominators. They did not know of any such formula, and made a computer search (using the PSLQ lattice reduction algorithm) for any series of the desired form. Their search unearthed a numerical identity

$$\pi = \sum_{n=0}^{\infty} \left( \frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right) \left( \frac{1}{16} \right)^n,$$

which was then rigorously proved and used to implement their binary-digits algorithm.

### 1.5 Rogers-Ramanujan identities

The famous Rogers-Ramanujan identities

$$1 + \sum_{k=1}^{\infty} \frac{q^{k^2+ak}}{(1-q)(1-q^2)\cdots(1-q^k)} = \prod_{j=0}^{\infty} \frac{1}{(1-q^{5j+a+1})(1-q^{5j-a+4})}, \quad a = 0, 1.$$

can now be proved by an almost entirely mechanical procedure from Jacobi's triple product identity and the  $q$ -WZ algorithm of Wilf and Zeilberger that checks identities of  $q$ -hypergeometric finite sums [Pau94]. Knuth's foreword to a book on the WZ method opens, "Science is what we understand well enough to explain to a computer. Art is everything else we do." Through the WZ method, many summation identities have become a science [PWZ96].

### 1.6 packing tetrahedra

Aristotle erroneously believed that regular tetrahedra tile space: "It is agreed that there are only three plane figures which can fill a space, the triangle, the square, and the hexagon, and only two solids, the pyramid and the cube" [AriBC]. However, centuries later, when the dihedral angle of the regular tetrahedron was calculated:

$$\arccos(1/3) \approx 1.23 < 1.25664 \approx 2\pi/5,$$

it was realized that a small gap is left when five regular tetrahedra are grouped around a common edge (Figure 3). In 1900, in his famous list of problems, Hilbert asked "How can one arrange most densely in space an infinite number of equal solids of given form, e.g., spheres with given radii or regular tetrahedra ...?"

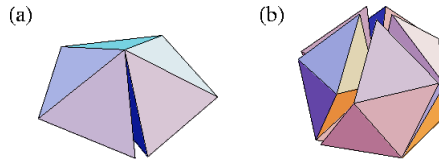
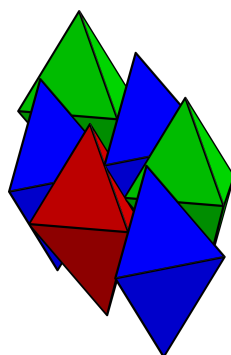


Fig. 3. Regular tetrahedra fail to tile space (image source [Doy11]).

Aristotle notwithstanding, until recently, no arrangements of regular tetrahedra with high density were known to exist. In 2000, Betke and Henk developed an efficient computer algorithm to find the densest lattice packing of a general convex body [BH00]. This opened the door to experimentation [CT06]. For example, the algorithm can determine the best lattice packing of the convex hull of the cluster of tetrahedra in Figure 3. In rapid succession came new record-breaking arrangements of tetrahedra, culminating in what is now conjectured to be the best possible [CEG10]. (See Figure 4.) Although Chen had the panache to hand out Dungeons and Dragons tetrahedral dice to the audience for a hands-on modeling session during her thesis defense, the best arrangement

was found using Monte Carlo experiments. In the numerical simulations, a finite number of tetrahedra are randomly placed in a box of variable shape. The tetrahedra are jiggled as the box slowly shrinks until no further improvement is possible. Now that a precise conjecture has been formulated, the hardest part still remains: to give a proof.



**Fig. 4.** The best packing of tetrahedra is believed to be the Chen-Engel-Glotzer arrangement with density  $4000/4671 \approx 0.856$  (image source [CEG10]).

### 1.7 the Kepler conjecture

Hilbert's 18th problem asks to find dense packings of both spheres and regular tetrahedra. The problem of determining the best sphere packing in three dimensions is the Kepler conjecture. Kepler was led to the idea of density as an organizing principle in nature by observing the tightly packed seeds in a pomegranate. Reflecting on the hexagonal symmetry of snowflakes and honeycombs, by capping each honeycomb cell with a lid of the same shape as the base of the cell, he constructed a closed twelve-sided cell that tiles space. Kepler observed that the familiar pyramidal cannonball arrangement is obtained when a sphere is placed in each capped honeycomb cell (Figure 5). This he believed to be the densest packing.

L. Fejes Tóth proposed a strategy to prove Kepler's conjecture in the 1950s, and later he suggested that computers might be used. The proof, finally obtained by Ferguson and me in 1998, is one of the most difficult nonlinear optimization problems ever rigorously solved by computer [Hal05b]. The computers calculations originally took about 2000 hours to run on Sparc workstations. Recent simplifications in the proof have reduced the runtime to about 20 hours and have reduced the amount of customized code by a factor of more than 10.

### 1.8 the four-color theorem

The four-color theorem is the most celebrated computer proof in the history of mathematics. The problem asserts that it is possible to color the countries of any map with at



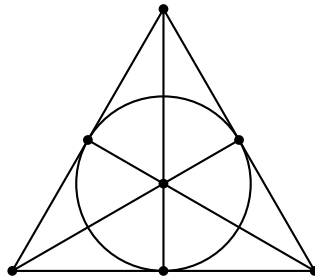
**Fig. 5.** An optimal sphere packing is obtained by placing one sphere in each three-dimensional honeycomb cell (image source [RhD11]).

most four colors in such a way that contiguous countries receive different colors. The proof of this theorem required about 1200 hours on an IBM 370-168 in 1976. So much has been written about Appel and Haken's computer solution to this problem that it is pointless to repeat it here [AHK77]. Let it suffice to cite a popular account [Wil02], a sociological perspective [Mac01], the second generation proof [RSST97], and the culminating formal verification [Gon08].

### 1.9 projective planes

A finite projective plane of order  $n > 1$  is defined to be a set of  $n^2 + n + 1$  lines and  $n^2 + n + 1$  points with the following properties:

1. Every line contains  $n + 1$  points;
2. Every point is on  $n + 1$  lines;
3. Every two distinct lines have exactly one point of intersection;
4. Every two distinct points lie on exactly one line.



**Fig. 6.** The Fano plane is a finite projective plane of order 2.



The definition is an abstraction of properties that evidently hold for  $\mathbb{P}^2(\mathbb{F}_q)$ , the projective plane over a finite field  $\mathbb{F}_q$ , with  $q = n$ , for any prime power  $q$ . In particular, a finite projective plane exists whenever  $n$  is a positive power of a prime number (Figure 6).

The conjecture is that every finite projective plane of order  $n > 1$  is a prime power. The smallest integers  $n > 1$  that are *not* prime powers are

$$6, 10, 12, 14, 15, \dots$$

The brute force approach to this conjecture is to eliminate each of these possibilities in turn. The case  $n = 6$  was settled in 1938. Building on a number of theoretical advances [MST73], Lam eliminated the case  $n = 10$  in 1989, in one of the most difficult computer proofs in history [LTS89]. This calculation was executed over a period of years on multiple machines and eventually totaled about 2000 hours of Cray-1A time.

Unlike the computer proof of the four-color theorem, the projective plane proof has never received independent verification. Because of the possibilities of programming errors and soft errors (see Section 3.5), Lam is unwilling to call his result a proof. He writes, “From personal experience, it is extremely easy to make programming mistakes. We have taken many precautions, . . . Yet, I want to emphasize that this is only an experimental result and it desperately needs an independent verification, or better still, a theoretical explanation” [Lam91].

Recent speculation at *Math Overflow* holds that the next case,  $n = 12$ , remains solidly out of computational reach [Hor10].

### 1.10 hyperbolic manifolds

Computers have helped to resolve a number of open conjectures about hyperbolic manifolds (defined as complete Riemannian manifolds with constant negative sectional curvature  $-1$ ), including the proof that the space of hyperbolic metrics on a closed hyperbolic 3-manifold is contractible [GMT03], [Gab10].

### 1.11 chaos theory and strange attractors

The theory of chaos has been one of the great success stories of twentieth century mathematics and science. Turing<sup>1</sup> expressed the notion of chaos with these words, “quite small errors in the initial conditions can have an overwhelming effect at a later time. The displacement of a single electron by a billionth of a centimetre at one moment might make the difference between a man being killed by an avalanche a year later, or escaping” [Tur50]. Later, the metaphor became a butterfly that stirs up a tornado in Texas by flapping its wings in Brazil.

Thirteen years later, Lorenz encountered chaos as he ran weather simulations on a Royal McBee LGP-30 computer [Lor63]. When he reran an earlier numerical solution with what he thought to be identical initial data, he obtained wildly different results.

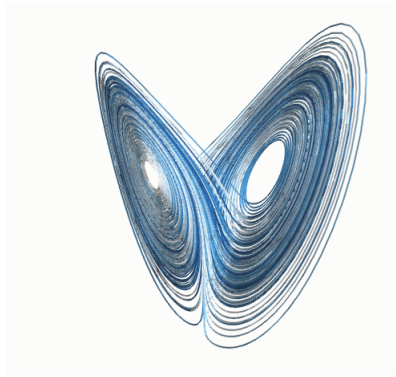
<sup>1</sup> For early history, see [Wol02, p. 971]. Turing vainly hoped that digital computers might be insulated from the effects of chaos.

He eventually traced the divergent results to a slight discrepancy in initial conditions caused by rounding in the printout. The *Lorenz oscillator* is the simplified form of Lorenz's original ordinary differential equations.

A set  $A$  is *attracting* if it has a neighborhood  $U$  such that

$$A = \bigcap_{t \geq 0} f_t(U),$$

where  $f_t(x)$  is the solution of the dynamical system (in present case the Lorenz oscillator) at time  $t$  with initial condition  $x$ . That is,  $U$  flows towards the attracting set  $A$ . Simulations have discovered attracting sets with strange properties such as non-integral Hausdorff dimension and the tendency for a small slab of volume to quickly spread throughout the attractor.



**Fig. 7.** The Lorenz oscillator gives one of the most famous images of mathematics, a *strange attractor* in dynamical systems (image source [Aga13]).

Lorenz conjectured in 1963 that his oscillator has a strange attractor (Figure 7). In 1982, the Lax report cited soliton theory and strange attractors as two prime examples of the “discovery of new phenomena through numerical experimentation,” and calls such discovery perhaps the most “significant application of scientific computing” [Lax82]. Smale, in his list of 18 “Mathematical Problems for the Next Century” made the fourteenth problem to present a rigorous proof that the dynamics of the Lorenz oscillator is a strange attractor [Sma98] with various additional properties that make it a “geometric Lorenz attractor.”

Tucker has solved Smale's fourteenth problem by computer [Tuc02] [Ste00]. One particularly noteworthy aspect of this work is that chaotic systems, by their very nature, pose particular hardships for rigorous computer analysis. Nevertheless, Tucker implemented the classical Euler method for solving ordinary differential equations with particular care, using interval arithmetic to give mathematically rigorous error bounds. Tucker has been awarded numerous prizes for this work, including the Moore Prize (2002) and the EMS Prize (2004).

Smale’s list in general envisions a coming century in which computer science, especially computational complexity, plays a much larger role than during the past century. He finishes the list with the open-ended philosophical problem that echoes Turing: “*What are the limits of intelligence, both artificial and human?*”

### 1.12 4/3

Mandelbrot’s conjectures in fractal geometry have resulted in two Fields Medals. Here he describes the discovery of the  $4/3$ -conjecture made in [Man82]. “The notion that these conjectures might have been reached by pure thought – with no picture – is simply inconceivable... I had my programmer draw a very big sample [Brownian] motion and proceeded to play with it.” He goes on to describe computer experiments that led him to enclose the Brownian motion into black clusters that looked to him like islands with jagged coastlines (Figure 8). “[I]nstantly, my long previous experience with the coastlines of actual islands on Earth came handy and made me suspect that the boundary of Brownian motion has a fractal dimension equal to  $4/3$ ” [Man04].

This conjecture, which Mandelbrot’s trained eye spotted in an instant, took 18 years to prove [LSW01].



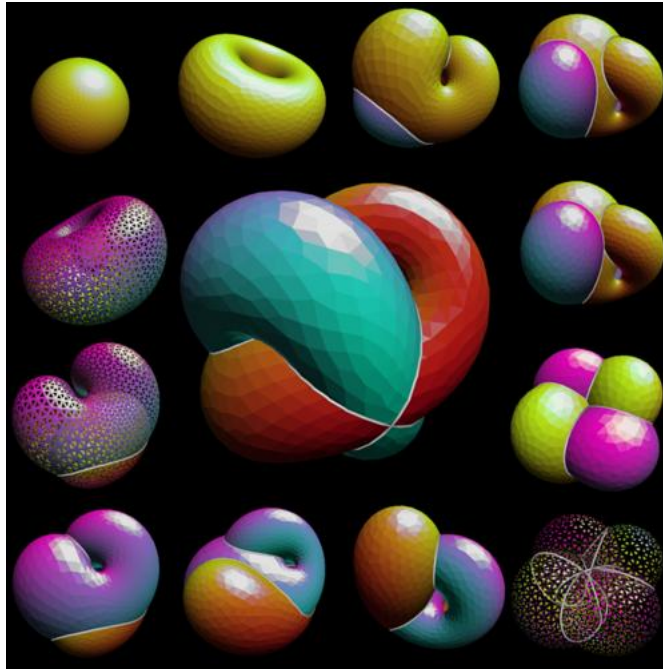
**Fig. 8.** A simulation of planar Brownian motion. Mandelbrot used “visual inspection supported by computer experiments” to formulate deep conjectures in fractal geometry (image generated from source code at [LSW01]).

### 1.13 sphere eversion visualization

Smale (1958) proved that it is possible to turn a sphere inside out without introducing any creases.<sup>2</sup> For a long time, this paradoxical result defied the intuition of experts. R. Bott, who had been Smale’s graduate advisor, refused to believe it at first. Levy writes that trying to visualize Smale’s mathematical argument “is akin to describing

<sup>2</sup> I am fond of this example, because The Scientific American article [Phi66] about this theorem was my first exposure to “real mathematics” as a child.

what happens to the ingredients of a soufflé in minute detail, down to the molecular chemistry, and expecting someone who has never seen a soufflé to follow this ‘recipe’ in preparing the dish” [Lev95].



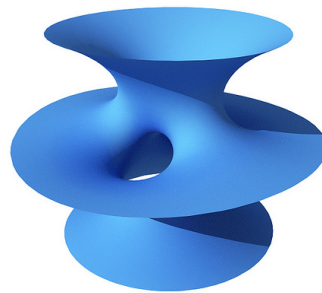
**Fig. 9.** Computer-generated stages of a sphere eversion (image source [Op111]).

It is better to see and taste a soufflé first. The computer videos of this theorem are spectacular. Watch them on YouTube! As we watch the sphere turn inside out, our intuition grows. The computer calculations behind the animations of the first video (the Optiverse) start with a sphere, half inverted and half right-side out [SFL]. From halfway position, the path of steepest descent of an energy functional is used to calculate the unfolding in both directions to the round spheres, with one fully inverted (Figure 9). The second video is based on Thurston’s “corrugations” [LMM94]. As the name suggests, this sphere eversion has undulating ruffles that dance like a jellyfish, but avoids sharp creases. Through computers, understanding.

Speaking of Thurston, he contrasts “our amazingly rich abilities to absorb geometric information and the weakness of our innate abilities to convey spatial ideas... We effortlessly look at a two-dimensional picture and reconstruct a three-dimensional scene, but we can hardly draw them accurately” [Pit11]. As more and more mathematics migrates to the computer, there is a danger that geometrical intuition becomes buried under a logical symbolism.

### 1.14 minimal surface visualization

Weber and Wolf [WW11] report that the use of computer visualization has become “commonplace” in minimal surface research, “a conversation between visual aspects of the minimal surfaces and advancing theory, each supporting the other.” This started when computer illustrations of the Costa surface (a particular minimal surface, Figure 10) in the 1980s revealed dihedral symmetries of the surface that were not seen directly from its defining equations. The observation of symmetry turned out to be the key to the proof that the Costa surface is an embedding. The symmetries further led to a conjecture and then proof of the existence of other minimal surfaces of higher genus with similar dihedral symmetries. As Hoffman wrote about his discoveries, “The images produced along the way were the objects that we used to make discoveries. They are an integral part of the process of doing mathematics, not just a way to convey a discovery made without their use” [Hof87].



**Fig. 10.** The Costa surface launched an era of computer exploration in minimal surface theory (image source [San12]).

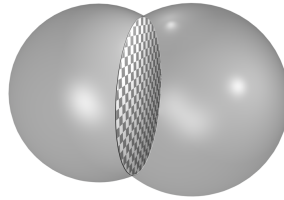
### 1.15 double bubble conjecture

Closely related to minimal surfaces are surfaces of constant mean curvature. The mean curvature of a minimal surface is zero; surfaces whose mean curvature is constant are a slight generalization. They arise as surfaces that are minimal subject to the constraint that they enclose a region of fixed volume. Soap bubble films are surfaces of constant mean curvature.

The isoperimetric inequality asserts that the sphere minimizes the surface area among all surfaces that enclose a region of fixed volume. The double bubble problem is the generalization of the isoperimetric inequality to two enclosed volumes. What is the surface minimizing way to enclose two separate regions of fixed volume? In the nineteenth century, Boys [Boy90] and Plateau observed experimentally that the answer should be two partial spherical bubbles joined along a shared flat disk (Figure 11). The size of the

shared disk is determined by the condition that angles should be  $120^\circ$  where the three surfaces meet. This is the *double bubble conjecture*.

The first case of double bubble conjecture to be established was that of two equal volumes [HHS95]. The proof was a combination of conventional analysis and computer proof. Conventional analysis (geometric measure theory) established the existence of a minimizer and reduced the possibilities to a small number of figures of revolution, and computers were to analyze each of the cases, showing in each case by interval analysis either that the case was not a local minimizer or that its area was strictly larger than the double bubble. Later theorems proved the double bubble conjecture in the general unequal volume case without the use of computers [HMRR00].



**Fig. 11.** The optimality of a double bubble was first established by computer, using interval analysis (image source [Tsi13]).

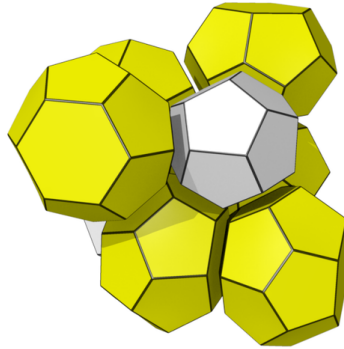
The natural extension of the double bubble conjecture from two bubbles to an infinite bubbly foam is the Kelvin problem. The problem asks for the surface area minimizing partition of Euclidean space into cells of equal volume. Kelvin's conjecture – a tiling by slight perturbations of truncated octahedra – remained the best known partition until a counterexample was constructed by two physicists, Phelan and Weaire in 1993 (Figure 12). The counterexample exists not as a physical model, nor as an exact mathematical formula, but only as an image generated from a triangular mesh in the *Surface Evolver* computer program. By default, the counterexample has become the new conjectural answer to the Kelvin problem, which I fully expect to be proved someday by computer.

### 1.16 kissing numbers

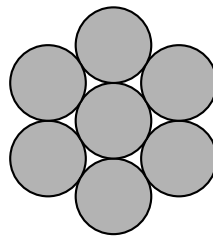
In the plane, at most six pennies can be arranged in a hexagon so that they all touch one more penny placed at the center of the hexagon (Figure 13). Odlyzko and Sloane, solved the corresponding problem in dimension 8: at most 240 nonoverlapping congruent balls can be arranged so that they all touch one more at the center.

Up to rotation, a unique arrangement of 240 exists. To the cognoscenti, the proof of this fact is expressed as one-line certificate:

$$\left(t - \frac{1}{2}\right)t^2\left(t + \frac{1}{2}\right)^2(t + 1).$$



**Fig. 12.** The Phelan-Weaire foam, giving the best known partition of Euclidean space into cells of equal volume, was constructed with Surface Evolver software. This foam inspired the bubble design of the Water Cube building in the 2008 Beijing Olympics (image source [PW111]).



**Fig. 13.** In two dimensions, the kissing number is 6. In eight dimensions, the answer is 240. The proof certificate was found by linear programming.

(For an explanation of the certificates, see [PZ04].) The certificate was produced by a linear programming computer search, but once the certificate is in hand, the proof is computer-free.

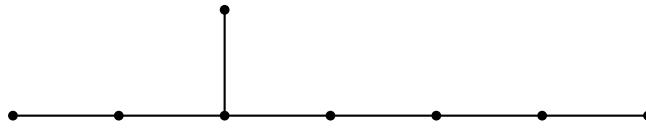
As explained above, six is the *kissing number* in two dimensions, 240 is the kissing number in eight dimensions. In three dimensions, the kissing number is 12. This three-dimensional problem goes back to a discussion between Newton and Gregory in 1694, but was not settled until the 1950s. A recent computer proof makes an exhaustive search through nearly 100 million combinatorial possibilities to determine exactly how much the twelve spheres must shrink to accommodate a thirteenth [MT10]. Bachoc and Valentin were recently awarded the SIAG/Optimization prize for their use of semi-definite programming algorithms to establish new proofs of the kissing number in dimensions 3, 4, 8 and new bounds on the kissing number in various other dimensions [BV08].

### 1.17 digression on $E_8$

It is no coincidence that the calculation of Odlyzko and Sloane works in dimension 8. Wonderful things happen in eight dimensional space and again in 24 dimensions.

Having mentioned the 240 balls in eight dimensions, I cannot resist mentioning some further computer proofs. The centers of the 240 balls are vectors whose integral linear combinations generate a lattice in  $\mathbb{R}^8$ , known as the  $E_8$  lattice (Figure 14).

There is a packing of congruent balls in eight dimensions that is obtained by centering one ball at each vector in the  $E_8$  lattice, making the balls as large as possible without overlap. Everyone believes that this packing in eight dimensions is the densest possible, but this fact currently defies proof. If the center of the balls are the points of a lattice, then the packing is called a *lattice packing*. Cohn and Kumar have a beautiful computer assisted proof that the  $E_8$  packing is the densest of all lattice packings in  $\mathbb{R}^8$  (and the corresponding result in dimension 24 for the Leech lattice). The proof is based on the Poisson summation formula. Pfender and Ziegler's account of this computer-assisted proof won the Chauvenet Prize of the MAA for writing [PZ04].



**Fig. 14.** The  $E_8$  lattice is generated by eight vectors in  $\mathbb{R}^8$  whose mutual angles are  $120^\circ$  or  $90^\circ$  depending on whether the corresponding dots are joined by a segment are not.

The 240 vectors that generate the  $E_8$  lattice are the *roots* of a  $240+8$  dimensional Lie group (also called  $E_8$ ); that is, a differentiable manifold that has the analytic structure of a group. All simple Lie groups were classified in the nineteenth century.<sup>3</sup> They fall into infinite families named alphabetically,  $A_n$ ,  $B_n$ ,  $C_n$ ,  $D_n$ , with 5 more exceptional cases that do not fall into infinite families  $E_6$ ,  $E_7$ ,  $E_8$ ,  $F_4$ ,  $G_2$ . The exceptional Lie group<sup>4</sup> of highest dimension is  $E_8$ .

The long-term *Atlas Project* aims to use computers to determine all unitary representations of real reductive Lie groups [Atl]. The 19-member team focused on  $E_8$  first, because everyone respects the formidable  $E_8$ . By 2007, a computer had completed the character table of  $E_8$ . Since there are infinitely many irreducible characters and each character is an analytic function on (a dense open subset of) the group, it is not clear without much further explanation what it might even mean for a computer to output the full character table as a 60 gigabyte file [Ada11]. What is significant about this work is that it brings the computer to bear on some abstract parts of mathematics that have been traditionally largely beyond the reach of concrete computational description,

<sup>3</sup> I describe the families over  $\mathbb{C}$ . Each complex Lie group has a finite number of further real forms.

<sup>4</sup> For decades,  $E_8$  has stood for the ultimate in speculative physics, whether in heterotic string theory or a “theory of everything.” Last year,  $E_8$  took a turn toward the real world, when  $E_8$  calculations predicted neutron scattering experiments with a cobalt niobate magnet [BG11].



including infinite dimensional representations of Lie groups, intersection cohomology and perverse sheaves. Vogan's account of this computational project was awarded the 2011 Conant Prize of the AMS [Vog07].

While on the topic of computation and representation theory, I cannot resist a digression into the  $P$  versus  $NP$  problem, the most fundamental unsolved problem in mathematics. In my opinion, attempts to settle  $P$  versus  $NP$  from the axioms of ZFC are ultimately as ill-fated as Hilbert's program in the foundations of math (which nonetheless spurred valuable partial results such as the decision procedures of Presburger and Tarski), but if I were to place faith anywhere, it would be in Mulmuley's program in *geometric complexity theory*. The program invokes geometric invariant theory and representation theoretic invariants to tease apart complexity classes: if the irreducible constituents of modules canonically associated with two complexity classes are different, then the two complexity classes are distinct. In this approach, the determinant and permanent of a matrix are chosen as the paradigms of what is easy and hard to compute, opening up complexity theory to a rich algebro-geometric structure [Mul11], [For09].

### 1.18 future computer proofs

Certain problems are natural candidates for computer proof: the Kelvin problem by the enumeration of the combinatorial topology of possible counterexamples; the search for a counterexample to the two-dimensional Jacobian conjecture through the minimal model program [Bor09]; resolution of singularities in positive characteristic through an automated search for numerical quantities that decrease under suitable blowup; existence of a projective plane of order 12 by constraint satisfaction programming; the optimality proof of the best known packing of tetrahedra in three dimensions [CEG10]; Steiner's isoperimetric conjecture (1841) for the icosahedron [Ste41]; and the Reinhardt conjecture through nonlinear optimization [Hal11]. But proceed with caution! Checking on our zeal for brute computation, computer-generated patterns can sometimes fail miserably. For example, the sequence:

$$\left\lfloor \frac{2}{2^{1/n} - 1} \right\rfloor - \left\lfloor \frac{2n}{\log 2} \right\rfloor, \quad n = 1, 2, 3, \dots$$

starts out as the zero sequence, but remarkably first gives a nonzero value when  $n$  reaches 777, 451, 915, 729, 368 and then again when  $n = 140, 894, 092, 055, 857, 794$ . See [Sta07].

At the close of this first section, we confess that a survey of mathematics in the age of the Turing machine is a reckless undertaking, particularly if it almost completely neglects software products and essential mathematical algorithms – the Euclidean algorithm, Newton's method, Gaussian elimination, fast Fourier transform, simplex algorithm, sorting, Schönhage-Strassen, and many more. A starting point for the exploration of mathematical software is KNOPPIX/Math, a bootable DVD with over a hundred free mathematical software products (Figure 15) [Ham08]. Sage alone has involved over 200 developers and includes dozens of other packages, providing an open-source Python scripted alternative to computer algebra systems such as Maple and Mathematica.

|                  |   |
|------------------|---|
| $\text{\TeX}$    | Active-DVI, AUCT $\text{\TeX}$ , $\text{\TeX}$ macs, Kile, Whizzy $\text{\TeX}$   |
| computer algebra | Axiom, CoCoA4, GAP, Macaulay2, Maxima, PARI/GP, Risa/Asir, Sage, Singular, Yacas  |
| numerical calc   | Octave, Scilab, FreeFem++, Yorick   |
| visualization    | 3D-XplorMath-J, Dynagraph, GANG, Geomview, gnuplot, JavaView, K3DSurf   |
| geometry         | C.a.R, Dr.Geo, GeoGebra, GEONExT, KidsCindy, KSEG   |
| programming      | CLISP, Eclipse, FASM, Gauche, GCC, Haskell, Lisp Prolog, Guile, Lazarus, NASM, Objective Caml, Perl, Python, Ruby, Squeak |

Fig. 15. Some free mathematical programs on the Knoppix/Math DVD [Ham08].

## 2 Computer Proof

Proof assistants represent the best effort of logicians, computer scientists, and mathematicians to obtain complete mathematical rigor by computer. This section gives a brief introduction to proof assistants and describes various recent projects that use them.

The first section described various computer calculations in math, and this section turns to computer reasoning. I have never been able to get used to it being the mathematicians who use computers for calculation and the computer scientists who use computers for proofs!

### 2.1 design of proof assistants

A formal proof is a proof that has been checked at the level of the primitive rules of inference all the way back to the fundamental axioms of mathematics. The number of primitive inferences is generally so large that it is quite hopeless to construct a formal proof by hand of anything but theorems of the most trivial nature. McCarthy and de Bruijn suggested that we program computers to generate formal proofs from high-level descriptions of the proof. This suggestion has led to the development of proof assistants.

A *proof assistant* is an interactive computer program that enables a user to generate a formal proof from its high-level description. Some examples of theorems that have been formally verified by proof assistants appear in Figure 16. The computer code that implements a proof assistant lists the fundamental axioms of mathematics and gives procedures that implement each of the rules of logical inference. Within this general framework, there are enormous variations from one proof assistant to the next. The feature table in Figure 16 is reproduced from [Wie06]. The columns list different proof assistants, HOL, Mizar, etc.

Since it is the one that I am most familiar with, my discussion will focus largely on a particular proof assistant, *HOL Light*, which belongs to the *HOL* family of proof assistants. *HOL* is an acronym for Higher-Order Logic, which is the underlying logic of these proof assistants. A fascinating account of the history of *HOL* appears in [Gor00]. In 1972, R. Milner developed a proof-checking program based on a deductive system

LCF (for Logic of Computable Functions) that had been designed by Dana Scott a few years earlier. A long series of innovations (such as goal-directed proofs and tactics, the ML language, enforcing proof integrity through the type system, conversions and theorem continuations, rewriting with discrimination nets, and higher-order features) have led from LCF to HOL.

| Year | Theorem                   | Proof System | Formalizer     | Traditional Proof |
|------|---------------------------|--------------|----------------|-------------------|
| 1986 | First Incompleteness      | Boyer-Moore  | Shankar        | Gödel             |
| 1990 | Quadratic Reciprocity     | Boyer-Moore  | Russinoff      | Eisenstein        |
| 1996 | Fundamental - of Calculus | HOL Light    | Harrison       | Henstock          |
| 2000 | Fundamental - of Algebra  | Mizar        | Milewski       | Brynski           |
| 2000 | Fundamental - of Algebra  | Coq          | Geuvers et al. | Kneser            |
| 2004 | Four Color                | Coq          | Gonthier       | Robertson et al.  |
| 2004 | Prime Number              | Isabelle     | Avigad et al.  | Selberg-Erdős     |
| 2005 | Jordan Curve              | HOL Light    | Hales          | Thomassen         |
| 2005 | Brouwer Fixed Point       | HOL Light    | Harrison       | Kuhn              |
| 2006 | Flyspeck I                | Isabelle     | Bauer-Nipkow   | Hales             |
| 2007 | Cauchy Residue            | HOL Light    | Harrison       | classical         |
| 2008 | Prime Number              | HOL Light    | Harrison       | analytic proof    |
| 2012 | Odd Order Theorem         | Coq          | Gonthier       | Feit-Thompson     |

**Fig. 16.** Examples of Formal Proofs, adapted from [Hal08].

Without going into full detail, I will make a few comments about what some of the features mean. Different systems can be commended in different ways: HOL Light for its small trustworthy kernel, Coq for its powerful type system, Mizar for its extensive libraries, and Isabelle/HOL for its support and usability.

**small proof kernel.** If a proof assistant is used to check the correctness of proofs, who checks the correctness of the proof assistant itself? De Bruijn proposed that the proofs of a proof assistant should be capable of being checked by a short piece of computer code – something short enough to be checked by hand. For example, the kernel of the proof assistant HOL Light is just 430 lines of very readable computer code. The architecture of the system is such that if these 430 lines are bug free then it is incapable<sup>5</sup> of generating a theorem that hasn't been properly proved.

**automating calculations.** Mathematical argument involves both calculation and proof. The foundations of logic often specify in detail what constitutes a mathematical proof (a

<sup>5</sup> I exaggerate. Section 3 goes into detail about trust in computers.

| <i>proof assistant</i>                   | HOL | Mizar | PVS | Coq | Otter/Ivy | Isabelle/Isar | Alfa/Agda | ACL2 | PhoX | IMPS | Metamath | Theorema | Lego | Nuprl | $\Omega$ mega | B method | Mimilog |
|--|-----|-------|-----|-----|-----------|---------------|-----------|------|------|------|----------|----------|------|-------|---------------|----------|---------|
| small proof kernel ('proof objects')     | +   | -     | -   | +   | +         | +             | +         | -    | +    | -    | +        | -        | +    | -     | +             | -        | +       |
| calculations can be proved automatically | +   | -     | +   | +   | +         | +             | -         | +    | +    | +    | -        | +        | +    | +     | +             | +        | +       |
| extensible/programmable by the user      | +   | -     | +   | +   | -         | +             | -         | -    | -    | -    | -        | -        | -    | +     | +             | -        | +       |
| powerful automation                      | +   | -     | +   | -   | +         | +             | -         | +    | -    | +    | -        | +        | -    | -     | +             | +        | -       |
| readable proof input files               | -   | +     | -   | -   | -         | +             | -         | +    | -    | -    | +        | +        | -    | -     | -             | -        | -       |
| constructive logic supported             | -   | -     | -   | +   | -         | +             | +         | -    | -    | -    | +        | -        | +    | +     | -             | -        | +       |
| logical framework                        | -   | -     | -   | -   | -         | +             | -         | -    | -    | -    | +        | -        | -    | -     | -             | -        | -       |
| typed                                    | +   | +     | +   | +   | -         | +             | +         | -    | +    | +    | -        | -        | +    | +     | +             | -        | +       |
| decidable types                          | +   | +     | -   | +   | -         | +             | +         | -    | +    | +    | -        | -        | +    | -     | +             | -        | +       |
| dependent types                          | -   | +     | +   | +   | -         | -             | +         | -    | -    | -    | -        | -        | +    | +     | -             | -        | -       |
| based on higher order logic              | +   | -     | +   | +   | -         | +             | +         | -    | +    | +    | -        | +        | +    | +     | +             | -        | -       |
| based on ZFC set theory                  | -   | +     | -   | -   | -         | +             | -         | -    | -    | -    | +        | -        | -    | -     | -             | +        | -       |
| large mathematical standard library      | +   | +     | +   | +   | -         | +             | -         | -    | -    | +    | -        | -        | -    | +     | -             | -        | -       |

**Fig. 17.** Features of proof assistants [Wie06]. The table is published by permission from Springer Science Business Media B.V.

sequence of logical inferences from the axioms), but downgrade calculation to second-class status, requiring every single calculation to undergo a cumbersome translation into logic. Some proof assistants allow *reflection* (sometimes implausibly attributed to *Poincaré*), which admits as proof the output from a verified algorithm (bypassing the expansive translation into logic of each separate execution of the algorithm) [Poi52, p. 4], [Bar07].

**constructive logic.** The law of excluded middle  $\phi \vee \neg\phi$  is accepted in classical logic, but rejected in constructive logic. A proof assistant may be constructive or classical. A box (*A Mathematical Gem*) shows how HOL Light becomes classical through the introduction of an axiom of choice.

*A Mathematical Gem – Proving the Excluded Middle*

The logic of HOL Light is intuitionistic until the axiom of choice is introduced and classical afterwards. By a result of Diononescu [Bee85], choice and extensionality imply the law of excluded middle:

$$\phi \vee \neg\phi.$$

The proof is such a gem that I have chosen to include it as the only complete proof in this survey article. Consider the two sets of booleans

$$P_1 = \{x \mid (x = \text{false}) \vee ((x = \text{true}) \wedge \phi)\} \quad \text{and} \\ P_2 = \{x \mid (x = \text{true}) \vee ((x = \text{false}) \wedge \phi)\}.$$

The sets are evidently nonempty, because  $\text{false} \in P_1$  and  $\text{true} \in P_2$ . By choice, we may pick  $x_1 \in P_1$  and  $x_2 \in P_2$ ; and by the definition of  $P_1$  and  $P_2$ :

$$(x_1 = \text{false}) \vee (x_1 = \text{true}), \quad (x_2 = \text{false}) \vee (x_2 = \text{true}).$$

We may break the proof of the excluded middle into four cases, depending on the two possible truth values of each of  $x_1$  and  $x_2$ .

**Cases**  $(x_1, x_2) = (\text{true}, \text{true})$ ,  $(x_1, x_2) = (\text{true}, \text{false})$ : By the definition of  $P_1$ , if  $x_1 = \text{true}$ , then  $\phi$ , so  $\phi \vee \neg\phi$ .

**Case**  $(x_1, x_2) = (\text{false}, \text{false})$ : Similarly, by the definition of  $P_2$ , if  $x_2 = \text{false}$ , then  $\phi$ , so also  $\phi \vee \neg\phi$ .

**Case**  $(x_1, x_2) = (\text{false}, \text{true})$ : If  $\phi$ , then  $P_1 = P_2$ , and the choices  $x_1$  and  $x_2$  reduce to a single choice  $x_1 = x_2$ , which contradicts  $(x_1, x_2) = (\text{false}, \text{true})$ . Hence  $\phi$  implies false; which by the definition of negation gives  $\neg\phi$ , so also  $\phi \vee \neg\phi$ . *Q.E.D.*

**logical framework.** Many different systems of logic arise in computer science. In some proof assistants the logic is fixed. Other proof assistants are more flexible, allowing different logics to be plugged in and played with. The more flexible systems implement a meta-language, a *logical framework*, that gives support for the implementation of multiple logics. Within a logical framework, the logic and axioms of a proof assistant can themselves be formalized, and machine translations<sup>6</sup> can be constructed between different foundations of mathematics [IR11].

**type theory.** Approaching the subject of formal proofs as a mathematician whose practice was shaped by Zermelo-Fraenkel set theory, I first treated types as nothing more than convenient identifying labels (such real number, natural number, list of integers, or boolean) attached to terms, like the PLU stickers on fruit that get peeled away before consumption. Types are familiar from programming languages as a way of identifying what data structure is what. In the simple type system of HOL Light, to each term is affixed a unique type, which is either a primitive type (such as the boolean type *bool*), a type variable ( $A, B, C, \dots$ ), or inductively constructed from other types with the arrow constructor ( $A \rightarrow B, A \rightarrow (bool \rightarrow C)$ , etc.). There is also a way to create subtypes of existing types. If the types are interpreted naively as sets, then  $x:A$  asserts that the term  $x$  is a member of  $A$ , and  $f : A \rightarrow B$  asserts that  $f$  is a member of  $A \rightarrow B$ , the set of functions from  $A$  to  $B$ .

In untyped set theory, it is possible to ask ridiculous questions such as whether the real number  $\pi = 3.14\dots$ , when viewed as a raw set, is a finite group. In fact, in a random exploration of set theory, like a monkey composing sonnets at the keyboard, ridiculous questions completely overwhelm all serious content. Types organize data on the computer in meaningful ways to cut down on the static noise in the system. The question about  $\pi$  and groups is not well-typed and cannot be asked. Russell's paradox also disappears:  $X \notin X$  is not well-typed. For historical reasons, this is not surprising: Russell and Whitehead first introduced types to overcome the paradoxes of set theory, and from there, through Church, they passed into computer science.

Only gradually have I come to appreciate the significance of a comprehensive *theory of types*. The type system used by a proof assistant determines to a large degree how much of a proof the user must contribute and how much the computer automates behind the scenes. The type system is *decidable* if there is a decision procedure to determine the type of each term.

A type system is *dependent* if a type can depend on another term. For example, Euclidean space  $\mathbb{R}^n$ , depends on its dimension  $n$ . For this reason, Euclidean space is most naturally implemented in a proof assistant as a dependent type. In a proof assistant such as HOL Light that does not have dependent types, extra work is required to develop a Euclidean space library.

---

<sup>6</sup> My long term Flyspeck project seeks to give a formal proof of the Kepler conjecture [Hal05a]. This project is now scattered between different proof assistants. Logical framework based translations between proof assistants gives me hope that an automated tool may assemble the scattered parts of the project.

## 2.2 propositions as types

I mentioned the naive interpretation of each type  $A$  as a set and a term  $x:A$  as a member of the set. A quite different interpretation of types has had considerable influence in the design of proof assistants. In this “terms-as-proofs” view, a type  $A$  represents a proposition and a term  $x:A$  represents a proof of the proposition  $A$ . A term with an arrow type,  $f : A \rightarrow B$ , can be used to construct a proof  $f(x)$  of  $B$  from a proof  $x$  of  $A$ . In this interpretation, the arrow is logical implication.

A further interpretation of types comes from programming languages. In this “terms-as-computer-programs” view, a term is a program and the type is its specification. For example,  $f : A \rightarrow B$  is a program  $f$  that takes input of type  $A$  and returns a value of type  $B$ .

By combining the “terms as proofs” with the “terms as computer programs” interpretations, we get the famous *Curry-Howard correspondence* that identifies proofs with computer programs and identifies each proposition with the type of a computer program. For example, the most fundamental rule of logic,

$$\frac{A, \quad A \rightarrow B}{B}, \quad (\textit{modus ponens})$$

(from  $A$  and  $A$ -implies- $B$  follows  $B$ ) is identified with the function application in a computer program; from  $x:A$  and  $f : A \rightarrow B$  we get  $f(x):B$ . To follow the correspondence is to extract an executable computer program from a mathematical proof. The Curry-Howard correspondence has been extremely fruitful, with a multitude of variations, running through a gamut of proof systems in logic and identifying each with a suitable programming domain.

## 2.3 proof tactics

In some proof assistants, the predominant proof style is a backward style proof. The user starts with a *goal*, which is a statement to be proved. In interactive steps, the user reduces the goal to successively simpler goals until there is nothing left to prove.

Each command that reduces a goal to simpler goals is called a *tactic*. For example, in the proof assistant HOL Light, there are about 100 different commands that are tactics or higher-order operators on tactics (called tacticals). Figure 18 shows the most commonly used proof commands in HOL Light. The most common tactic is *rewriting*, which takes a theorem of the form  $a = b$  and substitutes  $b$  for an occurrence of  $a$  in the goal.

In the Coq proof assistant, the tactic system has been streamlined to an extraordinary degree by the *SSReflect* package, becoming a model of efficiency for other proof assistants to emulate, with an extremely small number of tactics such as the *move* tactic for bookkeeping, one for rewriting, ones for forward and backward reasoning, and another for case analysis [GM11], [GMT11]. The package also provides support for exploiting the computational content of proofs, by integrating logical reasoning with efficient computational algorithms.

| <i>name</i>  | <i>purpose</i>  | <i>usage</i> |
|--------------|---|--------------|
| THEN         | combine two tactics into one                              | 37.2%        |
| REWRITE      | use $a = b$ to replace $a$ with $b$ in goal               | 14.5%        |
| MP_TAC       | introduce a previously proved theorem                     | 4.0%         |
| SIMP_TAC     | rewriting with conditionals                               | 3.1%         |
| MATCH_MP_TAC | reduce a goal $b$ to $a$ , given a theorem $a \implies b$ | 3.0%         |
| STRIP_TAC    | (bookkeeping) unpackage a bundled goal                    | 2.9%         |
| MESON_TAC    | apply first-order reasoning to solve the goal             | 2.6%         |
| REPEAT       | repeat a tactic as many times as possible                 | 2.5%         |
| DISCH_TAC    | (bookkeeping) move hypothesis to the assumption list      | 2.3%         |
| EXISTS_TAC   | instantiate an existential goal $\exists x \dots$         | 2.3%         |
| GEN_TAC      | instantiate a universal goal $\forall x \dots$            | 1.4%         |

**Fig. 18.** A few of the most common proof commands in the HOL Light proof assistant

## 2.4 first-order automated reasoning

Many proof assistants support some form of automated reasoning to relieve the user of doing rote logic by hand. For example, Table 18 lists *meson* (an acronym for Loveland’s Model Elimination procedure), which is HOL Light’s tactic for automated reasoning [Har09, Sec. 3.15], [Har96]. The various automated reasoning tools are generally *first-order* theorem provers. The classic resolution algorithm for first-order reasoning is illustrated in a box (*Proof by Resolution*).



*Proof by Resolution*

Resolution is the granddaddy of automated reasoning in first-order logic. The resolution rule takes two disjunctions

$$P \vee A \quad \text{and} \quad \neg P' \vee B$$

and concludes

$$A' \vee B',$$

where  $A'$  and  $B'$  are the specializations of  $A$  and  $B$ , respectively, under the *most general unifier* of  $P$  and  $P'$ . (Examples of this in practice appear below.)

This box presents a rather trivial example of proof by resolution, to deduce the easy theorem asserting that every infinite set has a member. The example will use the following notation. Let  $\emptyset$  be a constant representing the empty set and constant  $c$  representing a given infinite set. We use three unary predicates  $e$ ,  $f$ ,  $i$  that have interpretations

$$e(X) \text{ "X is empty"}, \quad f(X) \text{ "X is finite"}, \quad i(X) \text{ "X is infinite."}$$

The binary predicate ( $\in$ ) denotes set membership. We prove  $i(c) \Rightarrow (\exists z.z \in c)$  "an infinite set has a member" by resolution.

To argue by contradiction, we introduce the hypothesis  $i(c)$  and the negated conclusion  $\neg(Z \in c)$  as axioms. Here are the axioms that we allow in the deduction. The axioms have been preprocessed, stripped of quantifiers, and written as a disjunction of literals. Upper case letters are variables.

| <i>Axiom</i>   | <i>Informal Description</i>                |
|--|--|
| 1. $i(c)$  | Assumption of desired theorem.             |
| 2. $\neg(Z \in c)$                                     | Negation of conclusion of desired theorem. |
| 3. $e(X) \vee (u(X) \in X)$                            | A nonempty set has a member.               |
| 4. $e(\emptyset)$                                      | The empty set is empty.                    |
| 5. $f(\emptyset)$                                      | The empty set is finite.                   |
| 6. $\neg i(Y) \vee \neg f(Y)$                          | A set is not both finite and infinite.     |
| 7. $\neg e(U) \vee \neg e(V) \vee \neg i(U) \vee i(V)$ | Weak indistinguishability of empty sets.   |

Here are the resolution inferences from this list of axioms. The final step obtains the desired contradiction.

| <i>Inference</i>   | <i>Resolvent</i>                     |
|--|--------------------------------------|
| 8. (resolving 2,3, unifying $X$ with $c$ and $u(X)$ with $Z$ ) | $e(c)$                               |
| 9. (resolving 7,8, unifying $U$ with $c$ )                     | $\neg e(V) \vee \neg i(c) \vee i(V)$ |
| 10. (resolving 1,9)  | $\neg e(V) \vee i(V)$                |
| 11. (resolving 4,10, unifying $V$ with $\emptyset$ )           | $i(\emptyset)$                       |
| 12. (resolving 6,11, unifying $Y$ with $\emptyset$ )           | $\neg f(\emptyset)$                  |
| 13. (resolving 12,5)   | $\perp$                              |

*Q.E.D.*

Writing about first-order automated reasoning, Huet and Paulin-Mohring [BC04] describe the situation in the early 1970s as a “catastrophic state of the art.” “The standard mode of use was to enter a conjecture and wait for the computer’s memory to exhaust its capacity. Answers were obtained only in exceptionally trivial cases.” They go on to describe numerous developments (Knuth-Bendix, LISP, rewriting technologies, LCF, ML, Martin-Löf type theory, NuPrl, Curry-Howard correspondence, dependent types, etc.) that led up to the Coq proof assistant. These developments led away from first-order theorem proving with its “thousands of unreadable logical consequences” to a highly structured approach to theorem proving in Coq.

First-order theorem proving has developed significantly over the years into sophisticated software products. They are no longer limited to “exceptionally limited cases.” Many different software products compete in an annual competition (CASC), to see which can solve difficult first-order problems the fastest. The LTB (large theory batch) division of the competition includes problems with thousands of axioms [PSST08]. Significantly, this is the same order of magnitude as the total number of theorems in a proof assistant. What this means is that a first-order theorem provers have reached the stage of development that they might be able to give fully automated proofs of new theorems in a proof assistant, working from the full library of previously proved theorems.

**sledgehammer.** The Sledgehammer tactic is Paulson’s implementation of this idea of full automation in the Isabelle/HOL proof assistant [Pau10]. As the name ‘Sledgehammer’ suggests, the tactic is all-purpose and powerful, but demolishes all higher mathematical structure, treating every goal as a massive unstructured problem in first-order logic. If  $L$  is the set of all theorems in the Isabelle/HOL library, and  $g$  is a goal, it would be possible to hand off the problem  $L \implies g$  to a first-order theorem prover. However, success rates are dramatically improved, when the theorems in  $L$  are first assessed by heuristic rules for their likely relevance for the goal  $g$ , in a process called *relevance filtering*. This filtering is used to reduce  $L$  to an axiom set  $L'$  of a few hundred theorems that are deemed most likely to prove  $g$ .

The problem  $L' \implies g$  is stripped of type information, converted to a first-order, and fed to first-order theorem provers. Experiments indicate that it is more effective to feed a problem in parallel into multiple first-order provers for a five-second burst than to hand the problem to the best prover (Vampire) for a prolonged attack [Pau10], [BN10]. When luck runs in your favor, one of the first-order theorem provers finds a proof.

The reconstruction of a formal proof from a first-order proof can encounter hurdles. For one thing, when type information is stripped from the problem (which is done to improve performance), soundness is lost. “In unpublished work by Urban, MaLAREa [a machine learning program for relevance ranking] easily proved the full Sledgehammer test suite by identifying an inconsistency in the translated lemma library; once MaLAREa had found the inconsistency in one proof, it easily found it in all the others” [Pau10], [Urb07]. Good results have been obtained in calling the first-order prover repeatedly to find a smaller set of axioms  $L'' \subset L'$  that imply the goal  $g$ . A manageably sized set  $L''$  is then passed to the metis tactic<sup>7</sup> in Isabelle/HOL, which constructs a formal proof  $L'' \implies g$  from scratch.

<sup>7</sup> Metis is a program that automates first-order reasoning [Met].

Böhme and Nipkow took 1240 proof goals that appear in several diverse theories of the Isabelle/HOL system and ran sledgehammer on all of them [BN10]. The results are astounding. The success rate (of obtaining fully reconstructed formal proofs) when three different first-order provers run for two-minutes each was 48%. The proofs of these same goals by hand might represent years of human labor, now fully automated through a single new tool.

Sledgehammer has led to a new style of theorem proving, in which the user is primarily responsible for stating the goals. In the final proof script, there is no explicit mention of sledgehammer. Metis proves the goals, with sledgehammer operating silently in the background to feed metis with whatever theorems it needs. For example, a typical proof script might contain lines such as [Pau10]

**hence** “ $x \subseteq \text{space } M$ ”  
**by** (metis sets into space lambda system sets)

The first line is the goal that the user types. The second line has been automatically inserted into the proof script by the system, with the relevant theorems `sets`, `into` etc. selected by Sledgehammer.

## 2.5 computation in proof assistants.

One annoyance of formal proof systems is the difficulty in locating the relevant theorems. At last count, HOL Light had about 14,000 theorems and nearly a thousand procedures for proof construction. Larger developments, such as Mizar, have about twice as many theorems. Good search tools have somewhat relieved the burden of locating theorems in the libraries. However, as the formal proof systems continue to grow, it becomes ever more important to find ways to use theorems without mentioning them by name.

As an example of a feature which commendably reduces the burden of memorizing long lists of theorem names, I mention the `REAL_RING` command in HOL Light, which is capable of proving any system of equalities and inequalities that holds over an arbitrary integral domain. For example, I can give a one-line formal proof of an isogeny  $(x_1, y_1) \mapsto (x_2, y_2)$  of elliptic curves: if we have a point on the first elliptic curve:

$$\begin{aligned} y_1^2 &= 1 + ax_1^2 + bx_1^4, \\ x_2 y_1 &= x_1, \\ y_2 y_1^2 &= (1 - bx_1^4), \\ y_1 &\neq 0 \end{aligned}$$

then  $(x_2, y_2)$  lies on a second elliptic curve

$$y_2^2 = 1 + a'x_2^2 + b'x_2^4,$$

where  $a' = -2a$  and  $b' = a^2 - 4b$ . In the proof assistant, the input of the statement is as economical as what I have written here. We expect computer algebra systems to be capable of checking identities like this, but to my amazement, I found it *easier* to check this isogeny in HOL Light than to check it in *Mathematica*.

The algorithm works in the following manner. A universally quantified system of equalities and inequalities holds over all integral domains if and only if it holds over all fields. By putting the formula in conjunctive normal form, it is enough to prove a finite number of polynomial identities of the form:

$$(p_1 = 0) \vee \cdots \vee (p_n = 0) \vee (q_1 \neq 0) \vee \cdots \vee (q_k \neq 0). \quad (2)$$

An element in a field is zero, if and only if it is not a unit. Thus we may rewrite each polynomial equality  $p_i = 0$  as an equivalent inequality  $1 - p_i z_i \neq 0$ . Thus, without loss of generality, we may assume that  $n = 0$ ; so that all disjuncts are inequalities. The formula (2) is logically equivalent to

$$(q_1 = 0) \wedge \cdots \wedge (q_k = 0) \implies \text{false}.$$

In other words, it is enough to prove that the zero set of the ideal  $I = (q_1, \dots, q_n)$  is empty. For this, we may use Gröbner bases<sup>8</sup> to prove that  $1 \in I$ , to certify that the zero set is empty.

Gröbner basis algorithms give an example of a *certificate-producing procedure*. A formal proof is obtained in two stages. In the first stage an unverified algorithm produces a certificate. In the second stage the proof assistant analyzes the certificate to confirm the results. Certificate-producing procedures open the door to external tools, which tremendously augment the power of the proof assistant. The meson is procedure implemented this way, as a search followed by verification. Other certificate-producing procedures in use in proof assistants are linear programming, SAT, and SMT.

Another praiseworthy project is Kaliszyk and Wiedijk's implementation of a computer algebra system on top of the proof assistant HOL Light. It combines the ease of use of computer algebra with the rigor of formal proof [KW07]. Even with its notational idiosyncrasies (& and # as markers of real numbers, Cx as a marker of complex numbers, `ii` for  $\sqrt{-1}$ , and `--` for unary negation), it is the kind of product that I can imagine finding widespread adoption by mathematicians. Some of the features of the system are shown in Figure 19.

## 2.6 formalization of finite group theory

The Feit-Thompson theorem, or odd-order theorem, is one of the most significant theorems of the twentieth century. (For his work, Thompson was awarded the three highest honors in the mathematical world: the Fields Medal, the Abel Prize, and the Wolf Prize.) The Feit-Thompson theorem states that every finite simple group has even order, except for cyclic groups of prime order. The proof, which runs about 250 pages, is extremely technical. The Feit-Thompson theorem launched the endeavor to classify all finite simple groups, a monumental undertaking that consumed an entire generation of group theorists.

Gonthier's team has formalized the proof of the Feit-Thompson theorem [Gon12]. To me as a mathematician, nothing else that has been done by the formal proof community compares in splendor to the formalization of this theorem. Finally, we are doing

<sup>8</sup> Kaliszyk's benchmarks suggest that the Gröbner basis algorithm in the proof assistant Isabelle runs about twenty times faster than that of HOL Light.

```

In1 := (3 + 4 DIV 2) EXP 3 * 5 MOD 3
Out1 := 250
In2 := vector [&2; &2] - vector [&1; &0] + vec 1
Out2 := vector [&2; &3]
In3 := diff (diff (\x. &3 * sin (&2 * x) + &7 + exp (exp x)))
Out3 := \x. exp x pow 2 * exp (exp x) + exp x * exp (exp x) + -- &12 * sin (&2 * x)
In4 := N (exp (&1)) 10
Out4 := #2.7182818284 + ... (exp (&1)) 10 F
In5 := 3 divides 6 /\ EVEN 12
Out5 := T
In6 := Re ((Cx (&3) + Cx (&2) * ii) / (Cx (-- &2) + Cx (&7) * ii))
Out6 := &8 / &53

```

**Fig. 19.** Interaction with a formally verified computer algebra system [KW07].

real mathematics! The project formalized two books, [BG94] and [Pet00], as well as a significant body of background material.

The structures of abstract algebra – groups, rings, modules, algebras, algebraically closed fields and so forth – have all been laid out formally in the Coq proof assistant. Analogous algebraic hierarchies appear in systems such as OpenAxiom, MathScheme, Mizar, and Isabelle; and while some of these hierarchies are elaborate, none have delved so deeply as the development for Feit-Thompson. It gets multiple abstract structures to work coherently together in a formal setting. “The problem is not so much in capturing the semantics of each individual construct but rather in having all the concepts working together well” [GMR07].

```

Structure finGroupType Type := FinGroupType {
  element :> finType;
  1 : element;
  -1 : element → element;
  * : element → element → element;
  unitP : ∀ x, 1 * x = x;
  invP : ∀ x, x-1 * x = 1;
  mulP : ∀ x1 x2 x3, x1 * (x2 * x3) = (x1 * x2) * x3
}.

```

**Fig. 20.** The structure of a finite group [GMR07].

The definition of a finite group in Coq is similar to the textbook definition, expressed in types and structures (Figure 20). It declares a finite type called `element` that is the group carrier or domain. The rest of the structure specifies a left-unit element `1`, a left-inverse <sup>-1</sup> and an associative binary operation (`*`).

Other aspects of Gonthier’s recent work can be found at [Gon11], [GGMR09], [BGBP08]. Along different lines, a particularly elegant organization of abstract algebra and category theory is obtained with type classes [SvdW11].

## 2.7 homotopy type theory

The simple type theory of HOL Light is adequate for real analysis, where relatively few types are needed – one can go quite far with natural numbers, real numbers, booleans, functions between these types, and a few functionals. However, the dependent type theory of Coq is better equipped than HOL Light for the hierarchy of structures from groups to rings of abstract algebra. But even Coq’s type theory is showing signs of strain in dealing with abstract algebra. For instance, an unpleasant limitation of Coq’s theory of types is that it lacks the theorem of extensionality for functions: if two functions take the same value for every argument, it *does not* follow that the two functions are equal.<sup>9</sup> The gymnastics to solve the problem of function extensionality in the context of the Feit-Thompson theorem are found in [GMR07].

A lack of function extensionality is an indication that equality in type theory may be misconceived. Recently, *homotopy type theory* has exploded onto the scene, which turns to homotopy theory and higher categories as models of type theory [HTT11]. It is quite natural to interpret a dependent type (viewed as a family of types parametrized by a second type) topologically as a fibration (viewed as a family of fibers parametrized by a base space) [AW09]. Voevodsky took the homotopical notions of equality and equivalence and translated them back into type theory, obtaining the *univalence axiom* of type theory, which posits what types are equivalent [Voe11], [PW12], [KLV12b], [KLV12a]. One consequence of the univalence axiom is the theorem of extensionality for functions. Another promising sign for computer theorem-proving applications is that the univalence axiom appears to preserve the computable aspects of type theory (unlike for instance, the axiom of choice which makes non-computable choices) [LH]. We may hope that some day there may be a back infusion of type-theoretic proofs into homotopy theory.

## 2.8 language of mathematics

Ganesalingam’s thesis is the most significant linguistic study of the language of mathematics to date [Gan09], [Gan10]. Ganesalingam was awarded the 2011 Beth Prize for the best dissertation in Logic, Language, or Information. Although this research is still at an early stage, it suggests that the mechanical translation of mathematical prose into formal computer syntax that faithfully represents the semantics is a realistic hope for the not-to-distant future.

The linguistic problems surrounding the language of mathematics differ in various ways from those of say standard English. A mathematical text introduces new definitions and notations as it progresses, whereas in English, the meaning of words is generally fixed from the outset. Mathematical writing freely mixes English with symbolic expressions. At the same time, mathematics is self-contained in a way that English can

<sup>9</sup> HOL Light avoids this problem by positing extensionality as a mathematical axiom.

never be; to understand English is to understand the world. By contrast, the meaning in a carefully written mathematical text is determined by Zermelo-Fraenkel set theory (or your favorite foundational system).

Ganesalingam’s analysis of notational syntax is general enough to treat quite general mixfix operations generalizing infix (e.g. +), postfix (e.g. factorial !), and prefix (cos). He analyzes subscripted infix operators (such as a semidirect product  $H \rtimes_{\alpha} N$ ), multi-symbolled operators (such as the three-symbolled  $[ \ : \ ]$  operator for the degree  $[K : k]$  of a field-extension), prefixed words ( $R$ -module), text within formulas  $\{(a, b) \mid a \text{ is a factor of } b\}$ , unusual script placement  ${}^L G$ , chained relations  $a < b < c$ , ellipses  $1 + 2 + \dots + n$ , contracted forms  $x, y \in \mathbb{N}$ , and exposed formulas (such as “for all  $x > 0, \dots$ ” to mean “for all  $x$ , if  $x > 0$ , then  $\dots$ ”).

The thesis treats what is called the formal mode of the language of mathematics – the language divested of all the informal side-remarks. The syntax is treated as a context-free grammar, and the semantics are analyzed with a variant of *discourse representation theory*, which in my limited understanding is something very similar to first-order logic; but different in one significant aspect: it provides a theory of pronoun references; or put more precisely, a theory of what may be the “legitimate antecedent for anaphor.”

A major issue in Ganesalingam’s thesis is the resolution of ambiguity. For example, in the statement

$$P \text{ is prime} \tag{3}$$

the term ‘prime’ may mean prime number, prime ideal, or prime manifold. His solution is to attach type information to terms (in the sense of types as discussed above). The reading of (3) depends on the type of  $P$ , variously a number, a subset of a ring, or a manifold. In this analysis, resolution of ambiguity becomes a task of a type inference engine.

Because of the need for type information, Ganesalingam raises questions about the suitability of Zermelo-Fraenkel set theory as the ultimate semantics of mathematics. A number of formal-proof researchers have been arguing in favor of typed foundational systems for many years. It is encouraging that there is remarkable convergence between Ganesalingam’s linguistic analysis, innovations in the Mizar proof assistant, and the development of abstract algebra in Coq. For example, in various camps we find ellipses (aka big operators), mixfix operators, type inference, missing argument inference mechanisms, and so forth. Also see [Hoe11] and [Pas07]. Mathematical abuses of notation have turned out to be rationally construed after all!

## 2.9 looking forward

Let’s take the long term view that the longest proofs of the last century are of insignificant complexity compared to what awaits. Why would we limit our creative endeavors to 10,000 page proofs when we have tools that allow us to go to a million pages or more? So far it is rare for a computer proof has defied human understanding. No human has been able to make sense of an unpublished 1500 page computer-generated proof

about Bruck loops<sup>10</sup> [PS08]. Eventually, we will have to content ourselves with fables that approximate the content of a computer proof in terms that humans can comprehend.

Turing’s great theoretical achievements were to delineate what a computer can do in the concept of a universal Turing machine, to establish limits to what a computer can do in his solution to the *Entscheidungsproblem*, and yet to advocate nonetheless that computers might imitate all intelligent activity. It remains a challenging research program: to show that one limited branch of mathematics, computation, might stand for all mathematical activity.

In the century since Turing’s birth, the computer has become so ubiquitous and the idea of computer as brain so commonplace that it bears repeating that we must still think very long and hard about how to construct a computer that can imitate a living, thinking mathematician.

Proof assistant technology is still under development in labs; far more is needed before it finds widespread adoption. Ask any proof assistant researcher, and you will get a sizable list of features to implement: more automation, better libraries, and better user interfaces! Wiedijk discusses ten design questions for the next generation of proof assistants, including the type system, which axiomatic foundations of mathematics to use, and the language of proof scripts [Wie10b].

Everyone actively involved in proof formalization experiences the incessant barrage of problems that have been solved multiple times before and that other users will have to solve multiple times again, because the solutions are not systematic. To counter this, the DRY “Don’t Repeat Yourself” principle of programming, formulated in [HT00], has been carried to a refreshing extreme by Carette in his proof assistant design. For example, in his designs, a morphism is defined only once, eliminating the need for separate definitions of a morphism of modules, of algebras, of varieties, and so forth. Carette’s other design maxims include “math has a lot of structure; use it” and “abstract mathematical structures produce the best code” [CES11]. Indeed, mathematicians turn to abstraction to bring out relevant structure. This applies to computer code and mathematical reasoning alike. American Math Society guidelines for mathematical writing apply directly to the computer: “omit any computation which is routine. . . . Merely indicate the starting point, describe the procedure, and state the outcome” [DCF<sup>+</sup>62] (except that computations should be automated rather than entirely omitted).

We need to separate the concerns of construction, maintenance, and presentation of proofs. The construction of formal proofs from a mathematical text is an extremely arduous process, and yet I often hear proposals that would increase the labor needed to formalize a proof, backed by secondary goals such as ease of maintenance, elegance of presentation, fidelity to printed texts, and pedagogy.<sup>11</sup> Better to avail ourselves of automation that was not available in the day of paper proofs, and to create new mathemat-

---

<sup>10</sup> The theorem states that Bruck loops with abelian inner mapping group are centrally nilpotent of class two.

<sup>11</sup> To explain the concept of *separation of concerns*, Dijkstra tells the story of an old initiative to create a new programming language that failed miserably because the designers felt that the new language had to look just like *FORTRAN* to gain broad acceptance. “The proper technique is clearly to postpone the concerns for general acceptance until you have reached a result of such a quality that it deserves acceptance” [Dij82].



ical styles suited to the medium, with proofs that variously look like a computer-aided design session, a functional program, or a list of hypotheses as messages in gmail. The most pressing concern is to reduce the skilled labor it takes a user to construct a formal proof from a pristine mathematical text.

The other concerns of proof transformation should be spun off as separate research activities: refactored proofs, proof scripts optimized for execution time, translations into other proof assistants, natural language translations, natural language abstracts, probabilistically checkable proofs, searchable metadata extracts, and proof mining.

For a long time, proof formalization technology was unable to advance beyond the mathematics of the 19th century, picking classical gems such as the Jordan curve theorem, the prime number theorem, or Dirichlet’s theorem on primes in arithmetic progressions. With the Feit-Thompson theorem, formalization has risen to a new level, by taking on the work of a Fields medalist.

At this level, there is an abundant supply of mathematical theorems to choose from. A Dutch research agenda lists the formalization of Fermat’s Last Theorem as the first in a list of “Ten Challenging Research Problems for Computer Science” [Ber05]. Hesselink predicts that this one formalization project alone will take about “fifty years, with a very wide margin.” Small pieces of the proof of Fermat, such as class field theory, the Langlands-Tunnell theorem, or the arithmetic theory of elliptic curves would be a fitting starting point. The aim is to develop technologies until formal verification of theorems becomes routine at the level of Atiyah-Singer index theorem, Perelman’s proof of the Poincaré conjecture, the Green-Tao theorem on primes in arithmetic progression, or Ngô’s proof of the fundamental lemma.

Starting from the early days of Newell, Shaw, and Simon’s experiments, researchers have dreamed of a general-purpose mechanical problem solver. Generations later, after untold trials, it remains an unwavering dream. I will end this section with one of the many proposals for a general problem solving algorithm. Kurzweil breaks general problem solving into three phases:

1. State your problem in precise terms.
2. Map out the contours of the solution space by traversing it recursively, within the limits of available computational resources.
3. Unleash an evolutionary algorithm to configure a neural net to tackle the remaining leaves of the tree.

He concludes, “And if all of this doesn’t work, then you have a difficult problem indeed” [Kur99]. Yes, indeed we do! Some day, energy and persistence will conquer.

### 3 Issues of Trust

We all have first-hand experience of the bugs and glitches of software. We exchange stories when computers run amok. Science recently reported the story of a textbook “The Making of a Fly” that was on sale at Amazon for more than 23 million dollars [Sci11]. The skyrocketing price was triggered by an automated bidding war between two sellers, who let their algorithms run unsupervised. The textbook’s author, Berkeley professor Peter Lawrence, said he hoped that the price would reach “a billion.” An overpriced textbook on the fly is harmless, except for students who have it as a required text.

But what about the Flash Crash on Wall Street that brought a 600 point plunge in the Dow Jones in just 5 minutes at 2:41 pm on May 6, 2010? According to the New York Times [NYT10], the flash crash started when a mutual fund used a computer algorithm “to sell \$4.1 billion in futures contracts.” The algorithm was designed to sell “without regard to price or time... [A]s the computers of the high-frequency traders traded [futures] contracts back and forth, a ‘hot potato’ effect was created.” When computerized traders backed away from the unstable markets, share prices of major companies fluctuated even more wildly. “Over 20,000 trades across more than 300 securities were executed at prices more than 60% away from their values just moments before” [SEC10]. Throughout the crash, computers followed algorithms to a T, to the havoc of the global economy.

#### 3.1 mathematical error

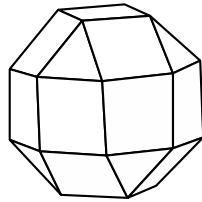
Why use computers to verify mathematics? The simple answer is that carefully implemented proof checkers make fewer errors than mathematicians (except J.-P. Serre).

Incorrect proofs of correct statements are so abundant that they are impossible to catalogue. Ralph Boas, former executive editor of Math Reviews, once remarked that proofs are wrong “half the time” [Aus08]. Kempe’s claimed proof of the four-color theorem stood for more than a decade before Heawood refuted it [Mac01, p. 115]. “More than a thousand false proofs [of Fermat’s Last Theorem] were published between 1908 and 1912 alone” [Cor10]. Many published theorems are like the hanging chad ballots of the 2000 U.S. presidential election, with scrawls too ambivalent for a clear yea or nay. One mathematician even proposed to me that a new journal is needed that unlike the others only publishes reliable results. Euclid gave us a method, but even he erred in the proof of the very first proposition of the Elements when he assumed without proof that two circles, each passing through the other’s center, must intersect. The concept that is needed to repair the gap in Euclid’s reasoning is an intermediate value theorem. This defect was not remedied until Hilbert’s ‘Foundations of Geometry.’

Examples of widely accepted proofs of false or unprovable statements show that our methods of proof-checking are far from perfect. Lagrange thought he had a proof of the parallel postulate, but had enough doubt in his argument to withhold it from publication. In some cases, entire schools have become sloppy, such as the Italian school of algebraic geometry or real analysis before the revolution in rigor towards the end of the nineteenth century. Plemelj’s 1908 accepted solution to Hilbert’s 21st problem on the monodromy of linear differential equations was refuted in 1989 by

Bolibruch. Auslander gives the example of a theorem<sup>12</sup> published by Waraskiewicz in 1937, generalized by Choquet in 1944, then refuted with a counterexample by Bing in 1948 [Aus08]. Another example is the approximation problem for Sobolev maps between two manifolds [Bet91], which contains a faulty proof of an incorrect statement. The corrected theorem appears in [HL03]. Such examples are so plentiful that a Wiki page has been set up to classify them, with references to longer discussions at Math Overflow [Wik11], [Ove09], [Ove10].

Theorems that are calculations or enumerations are especially prone to error. Feynman laments, “I don’t notice in the morass of things that something, a little limit or sign, goes wrong... I have mathematically proven to myself so many things that aren’t true” [Fey00, p. 885]. Elsewhere, Feynman describes two teams of physicists who carried out a two-year calculation of the electron magnetic moment and independently arrived at the same predicted value. When experiment disagreed with prediction, the discrepancy was eventually traced to an arithmetic error made by the physicists, whose calculations were not so independent as originally believed [Fey85, p. 117]. Pontryagin and Rokhlin erred in computing stable homotopy groups of spheres. Little’s tables of knots from 1885 contains duplicate entries that went undetected until 1974. In enumerative geometry, in 1848, Steiner counted 7776 plane conics tangent to 5 general plane conics, when there are actually only 3264. One of the most persistent blunders in the history of mathematics has been the misclassification (or misdefinition) of convex Archimedean polyhedra. Time and again, the pseudo rhombic cuboctahedron has been overlooked or illogically excluded from the classification (Figure 21) [Grü11].



**Fig. 21.** Throughout history, the pseudo rhombic cuboctahedron has been overlooked or misclassified.

### 3.2 In HOL Light we trust

To what extent can we trust theorems certified by a proof assistant such as HOL Light? There are various aspects to this question. Is the underlying logic of the system consistent? Are there any programming errors in the implementation of the system? Can a devious user find ways to create bogus theorems that circumvent logic? Are the underlying compilers, operating system, and hardware reliable?

As mentioned above, formal methods represent the best cumulative effort of logicians, computer scientists and mathematicians over the decades and even over the

<sup>12</sup> The claim was that every homogeneous plane continuum is a simple closed curve.

centuries to create a trustworthy foundation for the practice of mathematics, and by extension, the practice of science and engineering.

### 3.3 a network of mutual verification

John Harrison repeats the classical question “*Quis custodiet ipsos custodes*” – who guards the guards [Har06]? How do we prove the correctness of the prover itself? In that article, he proves the consistency of the HOL Light logic and the correctness of its implementation in computer code. He makes this verification in HOL Light itself! To skirt Gödel’s theorem, which implies that HOL Light – if consistent – cannot prove its own consistency, he gives two versions of his proof. The first uses HOL Light to verify a weakened version of HOL Light that does not have the axiom of infinity. The second uses a HOL Light with a strengthened axiom of infinity to verify standard HOL Light.

Recently, Adams has implemented a version of HOL called HOL Zero. His system has the ability to import mechanically proofs that were developed in HOL Light [Ada09]. He imported the self-verification of HOL Light, to obtain an external verification. You see where this is going. As mechanical translation capabilities are developed for proof assistants, it becomes possible for different proof assistants to share consistency proofs, similar to the way that different axiomatic systems give relative consistency proofs of one another. We are headed in the direction of knowing that if the logic or implementation of one proof assistant has an error, then all other major proof assistants must fail in tandem. Other self-verification projects are Coq in Coq (Coc) and ACL2 in ACL2 (Milawa) [Bar98], [Dav09].

### 3.4 hacking HOL

Of course, every formal verification project is a verification of an abstract model of the computer code, the computer language, and its semantics. In practice, there are gaps between the abstract model and implementation.

This leaves open the possibility that a hacker might find ways to create an unauthorized theorem; that is, a theorem generated by some means other than the rules of inference of HOL Logic. Indeed, there are small openings that a hacker can exploit.<sup>13</sup> Adams maintains a webpage of known vulnerabilities in his system and offers a cash bounty to anyone who uncovers a new vulnerability.

These documented vulnerabilities need to be kept in perspective. They lie at the fringe of the most reliable software products ever designed. Proof assistants are used to verify the correctness of chips and microcode [Fox03], operating system kernels [KAE<sup>+</sup>10],

<sup>13</sup> For example, strings are mutable in HOL Light’s source language, Objective CAML, allowing theorems to be maliciously altered. Also, Objective CAML has *object magic*, which is a way to defeat the type system. These vulnerabilities and all other vulnerabilities that I know would be detected during translation of the proof from HOL Light to HOL Zero. A stricter standard is Pollack consistency, which requires a proof assistant to avoid the appearance of inconsistency [Ada09], [Wie10a]. For example, some proof assistants allow the substitution of a variable whose name is a meaningless sequence of characters ‘ $n < 0 \wedge 0$ ’ for  $t$  in  $\exists n. t < n$  to obtain a Pollack-inconsistency  $\exists n. n < 0 \wedge 0 < n$ .

compilers [Ler06], safety-critical software such as aircraft guidance systems, security protocols, and mathematical theorems that defeat the usual refereeing process.

Some take the view that nothing short of absolute certainty in mathematics gives an adequate basis for science. Poincaré was less exacting<sup>14</sup>, only demanding the imprecision of calculation not to exceed experimental error. As Harrison reminds us, “a foundational death spiral adds little value” [Har10].

### 3.5 soft errors

Mathematicians often bring up the “cosmic ray argument” against the use of computers in math. Let’s look at the underlying science.

A soft error in a computer is a transient error that cannot be attributed to permanent hardware defects nor to bugs in software. Hard errors – errors that can be attributed to a lasting hardware failure – also occur, but at rates that are ten times smaller than soft errors [MW04]. Soft errors come from many sources. A typical soft error is caused by cosmic rays, or rather by the shower of energetic neutrons they produce through interactions in the earth’s atmosphere. A nucleus of an atom in the hardware can capture one of these energetic neutrons and throw off an alpha particle, which strikes a memory circuit and changes the value stored in memory. To the end user, a soft error appears as a gremlin, a seemingly inexplicable random error that disappears when the computer is rebooted and the program runs again.

As an example, we will calculate the expected number of soft errors in one of the mathematical calculations of Section 1.17. The Atlas Project calculation of the  $E_8$  character table was a 77 hour calculation that required 64 gigabytes RAM [Ada07]. Soft errors rates are generally measured in units of failures-in-time (FIT). One FIT is defined as one error per  $10^9$  hours of operation. If we assume a soft error rate of  $10^3$  FIT per Mbit, (which is a typical rate for a modern memory device operating at sea level<sup>15</sup> [Tez04]), then we would expect there to be about 40 soft errors in memory during the calculation:

$$\frac{10^3 \text{ FIT}}{1 \text{ Mbit}} \cdot 64 \text{ GB} \cdot 77 \text{ hours} = \frac{10^3 \text{ errors}}{10^9 \text{ hours Mbit}} \cdot (64 \cdot 8 \cdot 10^3 \text{ Mbit}) \cdot 77 \text{ hours} \approx 39.4 \text{ errors.}$$

This example shows that soft errors can be a realistic concern in mathematical calculations. (As added confirmation, the  $E_8$  calculation has now been repeated about 5 times with identical results.)

In software that has been thoroughly debugged, soft errors become the most significant source of error in computation. Although there are numerous ways to protect against soft errors with methods such as repeated calculations and error-correcting codes, hardware redesign carries an economic cost. In fact, soft errors are on the rise through miniaturization: a smaller circuit generally has a lower capacitance and responds to less energetic alpha particles than a larger circuit.

<sup>14</sup> “Il est donc inutile de demander au calcul plus de précision qu’aux observations; mais on ne doit pas non plus lui en demander moins” [Poi92].

<sup>15</sup> The soft error rate is remarkably sensitive to elevation; a calculation in Denver produces about three times more soft errors than the same calculation on identical hardware in Boston.

Soft errors are depressing news in the ultra-reliable world of proof assistants. Alpha particles rain on perfect and imperfect software alike. In fact, because the number of soft errors is proportional to the execution time of a calculation, by being slow and methodical, the probability of a soft error during a calculation inside a proof assistant can be much higher than the probability when done outside.

Soft errors and susceptibility to hacking have come to be more than a nuisance to me. They alter my philosophical views of the foundations of mathematics. I am a computational formalist – a formalist who admits physical limits to the reliability of any verification process, whether by hand or machine. These limits taint even the simplest theorems, such as our ability to verify that  $1 + 1 = 2$  is a consequence of a set of axioms. One rogue alpha particle brings all my schemes of perfection to nought. The rigor of mathematics and the reliability of technology are mutually dependent; math to provide ever more accurate models of science, and technology to provide ever more reliable execution of mathematical proofs.

## 4 Concluding Remarks

To everyone who has made it this far in this essay, I highly recommend MacKenzie's book [Mac01]. It written by a sociologist with a fine sensitivity to mathematics. The author received the Robert K. Merton Award of the American Sociological Association in 2003 for this book.

A few years ago, a special issue of the Notices of the AMS presented a general introduction to formal proofs [Hal08], [Har08], [Gon08], [Wie08]. I also particularly recommend the body of research articles by Harrison, Gonthier, and Carette.

I thank Adams (both Jeff and Mark), Urban, Carette, Kapulkin, Harrison, and Manfredi for conversations about ideas in this article.

## References

- Ada07. J. Adams, *Atlas of Lie groups and representations*, 2007, MIT colloquium slides, <http://atlas.math.umd.edu/talks/boston.pdf>.
- Ada09. M. Adams, *Importing proofs into HOL Zero*, private communication, 2009.
- Ada11. J. Adams, *Computing global characters*, <http://www.liegroups.org/papers/characters.pdf>, 2011.
- Aga13. Agarzago, accessed 2013, [http://en.wikipedia.org/wiki/Lorenz\\_attractor](http://en.wikipedia.org/wiki/Lorenz_attractor).
- AHK77. K. Appel, W. Haken, and J. Koch, *Every planar map is four colorable*, Illinois Journal of Mathematics **21** (1977), 439–567.
- AriBC. Aristotle, *On the heaven*, translated by J.L. Stocks, <http://classics.mit.edu/Aristotle/heavens.html>, 350BC.
- AS92. M. Aschbacher and Y. Segev, *The uniqueness of groups of Lyons type*, J. AMS **5** (1992), no. 1, 75–98.
- Atl. *Atlas of Lie groups and representations*, <http://www.liegroups.org/>.
- Aus08. Joseph Auslander, *On the roles of proof in mathematics*, Proof and other dilemmas (Bonnie Gold and Roger A. Simons, eds.), MAA, 2008.
- AW09. S. Awodey and M. Warren, *Homotopy theoretic models of identity types*, Mathematical Proceedings of the Cambridge Philosophical Society **146** (2009), 45–55.
- Bar98. B. Barras, *Verification of the interface of a small proof system in Coq*, LNCS, vol. 1512, pp. 28–45, 1998.
- Bar07. H. Barendregt, *Foundations of mathematics from the perspective of computer verification*, Springer Verlag, 2007.
- BC04. Y. Bertot and P. Castéran, *Interactive theorem proving and program development Coq'Art: The calculus of inductive constructions*, Springer, 2004.
- Bee85. M. Beeson, *Foundations of constructive mathematics: metamathematical studies*, Springer-Verlag, 1985.
- Ber05. J. Bergstra, *Nationale onderzoeksagenda informatie en communicatietechnologie (noag-ict) 2005–2010*, 2005.
- Bet91. Fabrice Bethuel, *The approximation problem for Sobolev maps between two manifolds*, Acta Math. **167** (1991), no. 3–4, 153–206.
- BG94. H. Bender and G. Glauberman, *Local analysis for the odd order theorem*, LMS, vol. 188, Cambridge University Press, 1994.

- BG11. D. Borthwick and S. Garibaldi, *Did a 1-dimensional magnet detect a 248-dimensional Lie algebra?*, Notices of the AMS **58** (2011), 1055–1066.
- BGBP08. Y. Bertot, G. Gonthier, S. Ould Biha, and I. Pasca, *Canonical big operators*, LNCS, vol. 5170, pp. 86–101, Springer, 2008.
- BH00. U. Betke and M. Henk, *Densest lattice packings of 3-polytopes*, Comput. Geom. **16** (2000), 157–186.
- Bir02. B. J. Birch, *In lieu of birthday greetings*, pp. 1–30, Cambridge Univ. Press, 2002.
- BN10. Sascha Böhme and Tobias Nipkow, *Sledgehammer: Judgement day*, Automated Reasoning (IJCAR 2010) (J. Giesl and R. Hähnle, eds.), LNCS, vol. 6173, Springer, 2010, pp. 107–121.
- Bor09. A. Borisov, *A geometric approach to the two-dimensional Jacobian conjecture*, <http://arxiv.org/abs/0912.4803>, 2009.
- Boy90. C. V. Boys, *Soap bubbles*, (Dover reprint), 1890.
- BS10. M. Bhargava and A. Shankar, *Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0*, arXiv:1007.0052v1 [math.NT], 2010.
- BSD65. B. J. Birch and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves, II*, J. für die reine und angew. Math. **218** (1965), 79–108.
- BV08. C. Bachoc and F. Vallentin, *New upper bounds for kissing numbers from semidefinite programming*, J. Amer. Math. Soc. **21** (2008), 909–924.
- Car07. H. Carayol, *La conjecture de Sato-Tate [d’après Clozel, Harris, Shepherd-Barron, Taylor]*, Sémin. Bourbaki **59** (2006–07), 345–391.
- CC111. 10 best jobs 2011, 2011, <http://www.careercast.com/jobs-rated/10-best-jobs-2011>.
- CEG10. E. R. Chen, Michael Engel, and Sharon C. Glotzer, *Dense crystalline dimer packings of regular tetrahedra*, Disc. Comp. Geom. **44** (2010), 253–280, <http://arxiv.org/abs/1001.0586>.
- CES11. J. Carette, M. Els Sheikh, and S. Smith, *A generative geometric kernel*, Proceedings of the 20th ACM SIGPLAN workshop on Partial evaluation and program manipulation, PEPM ’11, 2011, <http://www.cas.mcmaster.ca/~carette/publications/pepm28p-carette.pdf>, pp. 53–62.
- Cor10. Leo Corry, *On the history of Fermat’s last theorem: fresh views on an old tale*, Mathematische Semesterberichte **57** (2010), no. 1, 123–138.
- CT06. J. H. Conway and S. Torquato, *Packing, tiling, and covering with tetrahedra*, PNAS **103** (July 11, 2006), 10612–10617.
- Dav09. J. Davis, *A self-verifying theorem prover*, Ph.D. thesis, University of Texas at Austin, 2009, .
- DCF+62. J. L. Doob, L. Carlitz, F. A. Ficken, G. Paranian, and N. E. Steenrod, *Manual for authors of mathematical papers*, vol. 68, 1962.
- Dij82. E. W. Dijkstra, *On the role of scientific thought*, Springer, 1982, <http://www.cs.utexas.edu/users/EWD/ewd04xx/EWD447.PDF>.
- Doy11. J. Doye, accessed 2011, <http://physchem.ox.ac.uk/~doye/jon/structures/Morse/paper/node3.html>.
- Elk88. N. Elkies, *On  $a^4 + b^4 + c^4 = d^4$* , Mathematics of Computation **51** (1988), no. 184, 825–835.
- Fey85. R. P. Feynman, *QED: the strange theory of light and matter*, Princeton University Press, 1985.
- Fey00. R. Feynman, *Selected papers of Richard Feynman*, World Scientific, 2000.
- For09. L. Fortnow, *The status of the P versus NP problem*, Communications of the ACM **52** (2009), no. 9, 78–86.



- Fox03. A. Fox, *Formal specification and verification of ARM6*, LNCS, vol. 2758, pp. 25–40, Springer, 2003.
- Gab10. D. Gabai, *Hyperbolic 3-manifolds in the 2000's*, Proceedings of the International Congress of Mathematicians, 2010.
- Gan09. M. Ganesalingam, *The language of mathematics*, Ph.D. thesis, Cambridge University, 2009, <http://people.pwf.cam.ac.uk/mg262/GanesalingamMdis.pdf>.
- Gan10. ———, *The language of mathematics*, slides <http://www.srcf.ucam.org/principia/files/ganesalingam.pdf>, 2010.
- GGMR09. F. Garillot, G. Gonthier, A. Mahboubi, and L. Rideau, *Packaging mathematical structures*, LNCS, vol. 5674, pp. 327–342, Springer, 2009.
- GM11. G. Gonthier and A. Mahboubi, *An introduction to small scale reflection in Coq*, <http://hal.inria.fr/inria-00515548/>, 2011.
- GMR07. G. Gonthier, A. Mahboubi, and L. Rideau, *A modular formalisation of finite group theory*, LNCS, vol. 4732, pp. 86–101, Springer, 2007.
- GMT03. D. Gabai, R. Meyerhoff, and N. Thurston, *Homotopy hyperbolic 3-manifolds are hyperbolic*, *Annals of Math.* **157** (2003), 335–431.
- GMT11. G. Gonthier, A. Mahboubi, and E. Tassi, *A small scale reflection extension for the Coq system*, <http://hal.archives-ouvertes.fr/inria-00258384/>, 2011.
- Gon08. G. Gonthier, *Formal proof – the four colour theorem*, *Notices of the AMS* **55** (December 2008), no. 11, 1382–1393.
- Gon11. ———, *Point-free, set-free concrete linear algebra*, LNCS, vol. 6898, pp. 103–118, Springer, 2011.
- Gon12. ———, <http://www.msr-inria.inria.fr/events-news/feit-thompson-proved-in-coq>, 2012.
- Gor00. M. Gordon, *From LCF to HOL: a short history*, Proof, language, and interaction: essays in honour of Robin Milner (2000), 169–185.
- Grü11. B. Grünbaum, *An enduring error*, The Best Writing on Mathematics 2010 (M. Piteci, ed.), Princeton University Press, 2011.
- Hal05a. T. C. Hales, *The Flyspeck project*, Dagstuhl seminar proceedings (T. Coquand and H. Lombardi, eds.), vol. 25, IBFI, 2005, pp. 489–507.
- Hal05b. T. C. Hales, *A proof of the Kepler conjecture*, *Annals of Mathematics* **162** (2005), 1065–1185.
- Hal08. T. C. Hales, *Formal proof*, *Notices of the AMS* **55** (December 2008), no. 11, 1370–1380.
- Hal11. ———, *On the Reinhardt conjecture*, <http://arxiv.org/abs/1103.4518>, 2011.
- Ham08. T. Hamada, *KNOPPIX/math: A live system for enjoying mathematics with computer*, *ACM Communications in Computer Algebra* **42** (2008), no. 3, <http://www.knoppix-math.org/>.
- Har96. J. Harrison, *Optimizing proof search in model elimination*, Proceedings of the 13th International Conference on Automated Deduction (CADE-13), LNCS, vol. 1104, Springer, 1996, pp. 313–327.
- Har06. ———, *Towards self-verification of HOL light*, proceedings of IJCAR 2006, Lect. Notes in Comp. Sci., vol. 4130, springer, 2006, pp. 177–191.
- Har08. ———, *Formal proof – theory and practice*, *Notices of the AMS* **55** (December 2008), no. 11, 1395–1406.
- Har09. ———, *Handbook of practical logic and automated reasoning*, Cambridge University Press, 2009.

- Har10. ———, *On the cruelty of really doing formal proofs*, Principia Mathematica anniversary symposium, [www.srcf.ucam.org/principia/files/jrhslides.pdf](http://www.srcf.ucam.org/principia/files/jrhslides.pdf), 2010.
- HHS95. J. Hass, M. Hutchings, and R. Schlafly, *The double bubble conjecture*, *Elect. Res. Ann. AMS* **1** (1995), no. 3.
- HL03. Fengbo Hang and Fangua Lin, *Topology of Sobolev mappings. II*, *Acta Math.* **191** (2003), no. 1, 55–107.
- HMRR00. M. Hutchings, F. Morgan, M. Ritoré, and A. Ros, *Proof of the double bubble conjecture*, *Electronic Research Announcements of the AMS* **6** (2000), 45–49.
- Hoe11. J. Van Der Hoeven, *Towards semantic mathematical editing*, <http://hal.archives-ouvertes.fr/hal-00569351/fr/>, 2011.
- Hof87. D. Hoffman, *The computer-aided discovery of new embedded minimal surfaces*, *The Mathematical Intelligencer* **9** (1987), no. 3, 8–81.
- Hor10. M. Horn, *Projective plane of order 12*, *Math Overflow*, <http://mathoverflow.net/questions/38632/projective-plane-of-order-12>, 2010.
- HS99. G. Havas and C. C. Sims, *A presentation for the Lyons simple group*, <http://dimacs.rutgers.edu/~havas/TR0416.pdf>, 1999.
- HT00. A. Hunt and D. Thomas, *The pragmatic programmer*, Addison Wesley, 2000.
- HTT11. *Homotopy type theory*, 2011, <http://homotopytypetheory.org/>.
- Ima13. Wellcome Images, accessed 2013, <http://images.wellcome.ac.uk>, London.
- IR11. M. Iancu and F. Rabe, *Formalizing foundations of mathematics*, *Mathematical Structures in Computer Science* **21** (2011), no. 4, 883–911.
- KAE\*10. G. Klein, J. Andronick, K. Elphinstone, G. Heiser, D. Cock, P. Derrin, D. Elkaduwe, K. Engelhardt, R. Kolanski, M. Norrish, T. Sewell, H. Tuch, and S. Winwood, *sel4: Formal verification of an operating-system kernel*, *Communications of the ACM* **53** (2010), no. 6, 107–115.
- KLS08. C. Khare, M. Larsen, and G. Savin, *Functoriality and the inverse Galois problem*, *Comp. Math.* **144** (2008), 541–564.
- KLV12a. C. Kapulkin, P. F. Lumsdaine, and V. Voevodsky, *The simplicial model of univalent foundations*, arXiv:1211.2851, 2012.
- KLV12b. ———, *Univalence in simplicial sets*, arXiv:1203.2553, 2012.
- Kur99. R. Kurzweil, *The age of spiritual machines*, Penguin, 1999.
- KW07. C. Kaliszyk and F. Wiedijk, *Certified computer algebra on top of an interactive theorem prover*, *Proc. of the 14th Symposium on the Integration of Symbolic Computation and Mechanised Reasoning (Calculemus'07)* (Manuel Kauers, Manfred Kerber, Robert Miner, and Wolfgang Windsteiger, eds.), LNCS, vol. 4573, Springer Verlag, 2007, [http://score.cs.tsukuba.ac.jp/~kaliszyk/docs/kaliszyk\\_p04\\_calc.pdf](http://score.cs.tsukuba.ac.jp/~kaliszyk/docs/kaliszyk_p04_calc.pdf), pp. 94–105.
- Lam91. C. W. H. Lam, *The search for a finite projective plane of order 10*, *American Mathematical Monthly* **98** (1991), no. 4, 305 – 318.
- Lax82. P. Lax, *Report of the panel on large scale computing in science and engineering*, [www.pnl.gov/scales/docs/lax\\_report1982.pdf](http://www.pnl.gov/scales/docs/lax_report1982.pdf), 1982.
- Ler06. X. Leroy, *Formal certification of a compiler back-end, or: programming a compiler with a proof assistant*, 33rd ACM symposium on Principles of Programming Languages (2006), 42–54, <http://compcert.inria.fr/>.
- Lev95. S. Levy, *Making waves: A guide to the ideas behind “outside in”*, AK Peters, 1995, extract at <http://www.math.sunysb.edu/CDproject/OvUM/cool-links/www.geom.umn.edu/docs/outreach/oi/history.html>.

- LH. D. Licata and R. Harper, *Canonicity for 2-dimensional type theory*, preprint 2011, <http://www.cs.cmu.edu/~drl/pubs/lh112tt/lh112tt.pdf>.
- LMM94. S. Levy, D. Maxwell, and T. Munzner, *Outside in*, 1994, video, The Geometry Center (also on YouTube).
- Lor63. E. N Lorenz, *Deterministic nonperiodic flow*, J. Atmos. Sci. **20** (1963), 130–141.
- LP66. L. J. Lander and T. R. Parkin, *Counterexample to Euler's conjecture on sums of like powers*, Bulletin of the AMS (1966), 1079.
- LSW01. G. Lawler, O. Schramm, and W. Werner, *The dimension of the planar Brownian frontier is 4/3*, Math. Res. Letters **8** (2001), 401–411, <http://www.mrlonline.org/mrl/2001-008-004/2001-008-004-001.pdf>.
- LTS89. C. W. H. Lam, L. Thiel, and S. Swiercz, *The non-existence of finite projective planes of order 10*, Canadian journal of mathematics **41** (1989), no. 6, 1117–1123.
- Mac01. D. MacKenzie, *Mechanizing proof*, MIT Press, Cambridge, MA, 2001.
- Man82. B. Mandelbrot, *the fractal geometry of nature*, Freeman and Co., 1982.
- Man04. ———, *A theory of roughness*, [http://www.edge.org/3rd\\_culture/mandelbrot04/mandelbrot04\\_index.html](http://www.edge.org/3rd_culture/mandelbrot04/mandelbrot04_index.html), 2004.
- MC02. U. Meierfrankenfeld and C.W.Parker, *A computer-free construction of the Lyons group*, to appear, <http://www.math.msu.edu/~meier/Preprints/Ly/ly.pdf>, 2002.
- Met. *Metis*, <http://www.gilith.com/software/metis/index.html>.
- Met03. T. Metsänkylä, *Catalan's conjecture: another old Diophantine problem solved*, Bulletin AMS **41** (2003), no. 1, 43–57.
- Mih04. P. Mihăilescu, *Primary cyclotomic units and a proof of Catalan's conjecture*, J. für die reine und angew. Math. **572** (2004), 167–195.
- MM99. G. Malle and Matzat, *Inverse Galois theory*, Springer-Verlag, 1999.
- MST73. F. J. MacWilliams, N. J. A. Sloane, and J. G. Thompson, *On the existence of a projective plane of order 10*, J. Combin. Th. Ser. A **14** (1973), 66–78.
- MT10. O. R. Musin and A. S. Tarasov, *The strong thirteen spheres problem*, preprint <http://arxiv.org/abs/1002.1439>, February 2010.
- Mul11. K. Mulmuley, *On P vs. NP, and geometric complexity theory*, JACM **58** (April 2011).
- MW04. R. Mastipuram and E. C. Wee, *Soft errors' impact on system reliability*, EDN (2004).
- NYT10. *Lone \$4.1 billion sale led to 'flash crash' in May*, October 10 2010, New York Times <http://www.nytimes.com/2010/10/02/business/02flash.html>.
- Op111. accessed 2011, <http://new.math.uiuc.edu/optiverse/images.html>.
- Ove09. *Most interesting mathematics mistake?*, 2009, <http://mathoverflow.net/questions/879>.
- Ove10. *Widely accepted mathematical results that were later shown wrong?*, 2010, <http://mathoverflow.net/questions/35468>.
- Pas07. A. Paskevich, *The syntax and semantics of the ForTheL language*, <http://nevidal.org/download/forthe1.pdf>, 2007.
- Pau94. P. Paule, *Short and easy computer proofs of the Rogers-Ramanujan identities and of identities of similar type*, Electron. J. Combin. **1** (1994), 1–9.
- Pau10. L. Paulson, *Three years of experience with sledgehammer, a practical link between automatic and interactive theorem provers*, Paar-2010, Practical Aspects of Automated Reasoning, 2010.
- Pet00. T. Peterfalvi, *Character theory for the odd order theorem*, LMS, vol. 272, Cambridge University Press, 2000.

- Phi66. A. Phillips, *Turning a surface inside out*, Scientific American (1966), 112–120.
- Pit11. M. Pitici (ed.), *The best writing on mathematics 2010*, Princeton Univ. Press, 2011, forward by W. Thurston.
- Poi92. H. Poincaré, *Les méthodes nouvelles de la mécanique céleste*, Dover reprint, 1892.
- Poi52. ———, *Science and hypothesis*, Dover reprint, 1952.
- PS08. J. D. Phillips and D. Stanovský, *Using automated theorem provers in nonassociative algebra*, <http://www.karlin.mff.cuni.cz/~stanovsk/math/lpar08.pdf>, 2008.
- PSST08. A. Pease, G. Sutcliffe, N. Siegel, and S. Trac, *The annual SUMO reasoning prizes at CASC*, First International Workshop on Practical Aspects of Automated Reasoning, CEUR Workshop Proceedings, vol. 373, 2008.
- PW111. accessed 2011, [http://it.wikipedia.org/wiki/File:Foam\\_-\\_Weaire-Phelan\\_structure.png](http://it.wikipedia.org/wiki/File:Foam_-_Weaire-Phelan_structure.png).
- PW12. A. Pelayo and M. A. Warren, *Homotopy type theory and Voevodsky's univalent foundations*, arXiv:1210.5658, 2012.
- PWZ96. M. Petkovšek, H. S. Wilf, and D. Zeilberger, *A = B*, A. K. Peters, 1996.
- PZ04. F. Pfender and G. Ziegler, *Kissing numbers, sphere packings, and some unexpected proofs*, Notices of the AMS (2004), 873–883.
- RhD11. accessed 2011, [http://upload.wikimedia.org/wikipedia/commons/8/86/Rhombic\\_dodecahedra.jpg](http://upload.wikimedia.org/wikipedia/commons/8/86/Rhombic_dodecahedra.jpg).
- RSST97. Neil Robertson, Daniel Sanders, Paul Seymour, and Robin Thomas, *The four-colour theorem*, Journal of Combinatorial Theory, Series B **70** (1997), 2–44.
- San12. Anders Sandberg, accessed 2012, <http://www.flickr.com/photos/arenamontanus/8059864268/>.
- Sch. R. Schmidt, *The Sato-Tate conjecture*, [www.math.ou.edu/~rschmidt/satotate/page5.html](http://www.math.ou.edu/~rschmidt/satotate/page5.html).
- Sci11. *The \$23 million textbook*, Science **332** (2011), 647–648.
- SEC10. *Findings regarding the market events of May 6, 2010*, September 10, 2010, <http://sec.gov/news/studies/2010/marketevents-report.pdf>, Report of the Staffs of the Cftc and Sec to the Joint Advisory Committee on Emerging Regulatory Issues.
- Ser68. J.-P. Serre, *Abelian  $\ell$ -adic representations and elliptic curves*, WA Benjamin, 1968.
- SFL. J. M. Sullivan, G. Francis, and S. Levy, *the optiverse*, video <http://new.math.uiuc.edu/optiverse/> or YouTube.
- Sma98. S. Smale, *Mathematical problems for the next century*, Math. Intelligencer **20** (1998), no. 2, 7–15.
- SN. M. Sato and K. Namba, unpublished data.
- Sta07. R. Stanley, *Sequence A129935 in the on-line encyclopedia of integer sequences*, published electronically at <http://oeis.org/A129935>, 2007.
- Ste41. J. Steiner, *Über Maximum und Minimum bei den Figuren in der Ebene, auf der Kugelfläche und in Raume überhaupt*, C. R. Acad. Sci. Paris **12** (1841), 177–308.
- Ste00. I. Stewart, *The Lorenz attractor exists*, Nature **406** (31 Aug 2000), 948–949.
- SvdW11. B. Spitters and E. van der Weegen, *Type classes for mathematics in type theory*, <http://arxiv.org/abs/1102.1323>, 2011.
- Tez04. *Soft errors in electronic memory – a white paper*, 2004, Tezzaron Semiconductor, [http://www.tezzaron.com/about/papers/soft\\_errors\\_1\\_1\\_secure.pdf](http://www.tezzaron.com/about/papers/soft_errors_1_1_secure.pdf).
- Tsi13. Michael Tsirelson, accessed 2013, [http://en.wikipedia.org/wiki/File:Double\\_bubble.png](http://en.wikipedia.org/wiki/File:Double_bubble.png).
- Tuc02. W. Tucker, *A rigorous ODE solver and Smale's 14th problem*, Found. Comput. Math. **2** (2002), 53–117.

- Tur50. A. Turing, *Computing machinery and intelligence*, *Mind* **59** (1950), 433–460.
- Urb07. J. Urban, *MaLAREa: A metasytem for automated reasoning in large theories*, Proceedings of the CADE-21 Workshop on Empirically Successful Automated Reasoning in Large Theories (J. Urban, G. Sutcliffe, and S. Schultz, eds.), 2007, pp. 45–58.
- Voe11. V. Voevodsky, *Univalent foundations*, Mathematisches Forschungsinstitut Oberwolfach, [http://hottheory.files.wordpress.com/2011/06/report-11\\_2011.pdf](http://hottheory.files.wordpress.com/2011/06/report-11_2011.pdf), 2011.
- Vog07. D. Vogan, *The character table for  $E_8$* , *Notices of the AMS* **54** (2007), no. 9, 1022–1034.
- Wie06. F. Wiedijk (ed.), *The seventeen provers of the world*, Lecture Notes in Artificial Intelligence, Springer Verlag, 2006.
- Wie08. ———, *Formal proof – getting started*, *Notices of the AMS* **55** (December 2008), no. 11, 1408–1414.
- Wie10a. ———, *Pollack inconsistency*, ENTCS, 2010.
- Wie10b. ———, *the next generation of proof assistants*, <http://www.cs.ru.nl/F.Wiedijk/talks/lsfa.ps.gz>, slides, 2010.
- Wik11. *List of published incomplete proofs*, accessed 9/2011, [http://en.wikipedia.org/wiki/List\\_of\\_published\\_incomplete\\_proofs](http://en.wikipedia.org/wiki/List_of_published_incomplete_proofs).
- Wil02. R. Wilson, *Four colors suffice: How the map problem was solved*, Princeton University Press, 2002.
- Wol02. S. Wolfram, *A new kind of science*, Wolfram Media, 2002, <http://www.wolframscience.com/reference/notes/971c>.
- WW11. M. Weber and M. Wolf, *About the cover: early images of minimal surfaces*, *Bull. AMS* **48** (2011), no. 3, 457–460.