Does the proof stack up?

Think peer review takes too long? One mathematician has waited four years to have his paper refereed, only to hear that the exhausted reviewers can't be certain whether his proof is correct. George Szpiro investigates.

Grocers the world over know the most efficient way to stack spheres — but a mathematical proof for the method has brought reviewers to their knees.

Just under five years ago, Thomas Hales made a startling claim. In an e-mail he sent to dozens of mathematicians, Hales declared that he had used a series of computers to prove an idea that has evaded certain confirmation for 400 years. The subject of his message was Kepler's conjecture, proposed by the German astronomer Johannes Kepler, which states that the densest arrangement of spheres is one in which they are stacked in a pyramid — much the same way as grocers arrange oranges.

Soon after Hales made his announcement, reports of the breakthrough appeared on the front pages of newspapers around the world. But today, Hales's proof remains in limbo. It has been submitted to the prestigious *Annals of Mathematics*, but is yet to appear in print. Those charged with checking it say that they believe the proof is correct, but are so exhausted with the verification process that they cannot definitively rule out any errors. So when Hales's manuscript finally does appear in the *Annals*, probably during the next year, it will carry an unusual editorial note — a statement that parts of the paper have proved impossible to check.

At the heart of this bizarre tale is the use of computers in mathematics, an issue that has split the field. Sometimes described as a 'brute force' approach, computer-aided proofs often involve calculating thousands of possible outcomes to a problem in order to produce the final solution. Many mathematicians dislike this method, arguing that it is inelegant. Others criticize it for not offering any insight into the problem under consideration. In 1977, for example, a computer-aided proof was published for the four-colour theorem, which states that no more than four colours are needed to fill in a map so that any two adjacent regions have different colours[1,2]. No errors have been found in the proof, but some mathematicians continue to seek a solution using conventional methods.

Pile-driver

Hales, who started his proof at the University of Michigan in Ann Arbor before moving to the University of Pittsburgh, Pennsylvania, began by reducing the infinite number of possible stacking arrangements to 5,000 contenders. He then used computers to calculate the density of each arrangement. Doing so was more difficult than it sounds. The proof involved checking a series of mathematical inequalities using specially written computer code. In all, more than 100,000 inequalities were verified over a ten-year period.

Robert MacPherson, a mathematician at the Institute for Advanced Study in Princeton, New Jersey, and an editor of the *Annals*, was intrigued when he heard about the proof. He wanted to ask Hales and his graduate student Sam Ferguson, who had assisted with the proof, to submit their finding for publication, but he was also uneasy about the computer-based nature of the work.

The *Annals* had, however, already accepted a shorter computer-aided proof — the paper, on a problem in topology, was published this March[3]. After sounding out his colleagues on the journal's editorial board, MacPherson asked Hales to submit his paper. Unusually, MacPherson assigned a dozen mathematicians to referee the proof — most journals tend to employ between one and three. The effort was led by Gábor Fejes Tóth of the Alfréd Rényi Institute of Mathematics in Budapest, Hungary, whose father, the mathematician László Fejes Tóth, had predicted in 1965 that computers would one day make a proof of Kepler's conjecture possible.

It was not enough for the referees to rerun Hales's code — they had to check whether the programs did the job that they were supposed to do. Inspecting all of the code and its inputs and outputs, which together take up three gigabytes of memory space, would have been impossible. So the referees limited themselves to consistency checks, a reconstruction of the thought processes behind each step of the proof, and then a

study of all of the assumptions and logic used to design the code. A series of seminars, which ran for full academic years, was organized to aid the effort.

But success remained elusive. Last July, Fejes Tóth reported that he and the other referees were 99% certain that the proof is sound. They found no errors or omissions, but felt that without checking every line of the code, they could not be absolutely certain that the proof is correct.
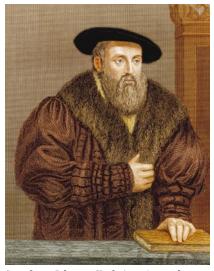
For a mathematical proof, this was not enough. After all, most mathematicians believe in the conjecture already — the proof is supposed to turn that belief into certainty. The history of Kepler's conjecture also gives reason for caution. In 1993, Wu-Yi Hsiang, then at the University of California, Berkeley, published a 100-page proof of the conjecture in the *International Journal of Mathematics*[4]. But shortly after publication, errors were found in parts of the proof. Although Hsiang stands by his paper, most mathematicians do not believe it is valid.

After the referees' reports had been considered, Hales says that he received the following letter from MacPherson: "The news from the referees is bad, from my perspective. They have not been able to certify the correctness of the proof, and will not be able to certify it in the future, because they have run out of energy … One can speculate whether their process would have converged to a definitive answer had they had a more clear manuscript from the beginning, but this does not matter now."

Pyramid power: Thomas Hales believes that computers will succeed where humans have failed in verifying his proof.

Star player: Johannes Kepler's conjecture has kept mathematicians guessing for 400 years.

The last sentence lets some irritation shine through. The proof that Hales delivered was by no means a polished piece. The 250-page manuscript consisted of five separate papers, each a sort of lab report that Hales and Ferguson filled out whenever the computer finished part of the proof. This unusual format made for difficult reading. To make matters worse, the notation and definitions also varied slightly between the papers.

Rough but ready

MacPherson had asked the authors to edit their manuscript. But Hales and Ferguson did not want to spend another year reworking their paper. "Tom could spend the rest of his career simplifying the proof," Ferguson said when they completed their paper. "That doesn't seem like an appropriate use of his time." Hales turned to other challenges, using traditional methods to solve the 2,000-year-old honeycomb conjecture, which states that of all conceivable tiles of equal area that can be used to cover a floor without leaving any gaps, hexagonal tiles have the shortest perimeter[5]. Ferguson left academia to take a job with the US Department of Defense.

Faced with exhausted referees, the editorial board of the *Annals* decided to publish the paper — but with a cautionary note. The paper will appear with an introduction by the editors stating that proofs of this type, which involve the use of computers to check a large number of mathematical statements, may be impossible to review in full. The matter might have ended there, but for Hales, having a note attached to his proof was not satisfactory.

This January, he launched the Flyspeck project, also known as the Formal Proof of Kepler. Rather than rely on human referees, Hales intends to use computers to verify every step of his proof. The effort will require the collaboration of a core group of about ten volunteers, who will need to be qualified mathematicians and willing to donate the computer time on their machines. The team will write programs to deconstruct each step of the proof, line by line, into a set of axioms that are known to be correct. If every part of the code can be broken down into these axioms, the proof will finally be verified.

Those involved see the project as doing more than just validating Hales's proof. Sean McLaughlin, a graduate student at New York University, who studied under Hales and has used computer methods to solve other mathematical problems, has already volunteered. "It seems that checking computer-assisted proofs is almost impossible for humans," he says. "With luck, we will be able to show that problems of this size can be subjected to rigorous verification without the need for a referee process."

But not everyone shares McLaughlin's enthusiasm. Pierre Deligne, an algebraic geometer at the Institute for Advanced Study, is one of the many mathematicians who do not approve of computer-aided proofs. "I believe in a proof if I understand it," he says. For those who side with Deligne, using computers to remove human reviewers from the refereeing process is another step in the wrong direction.

Despite his reservations about the proof, MacPherson does not believe that mathematicians should cut themselves off from computers. Others go further. Freek Wiedijk, of the Catholic University of Nijmegen in the Netherlands, is a pioneer of the use of computers to verify proofs. He thinks that the process could become standard practice in mathematics. "People will look back at the turn of the twentieth century and say 'that is when it happened'," Wiedijk says.

Whether or not computer-checking takes off, it is likely to be several years before Flyspeck produces a result. Hales and McLaughlin are the only confirmed participants, although others have expressed an interest. Hales estimates that the whole process, from crafting the code to running it, is likely to take 20 person-years of work. Only then will Kepler's conjecture become Kepler's theorem, and we will know for sure whether we have been stacking oranges correctly all these years. ■

1. Appel, K. & Haken, W. *Illinois J. Math.* **21,** 429–490 (1977).
2. Appel, K., Haken, W. & Koch, J. *Illinois J. Math.* **21,** 491–567 (1977).
3. Gabai, D., Meyerhoff, G. R. & Thurston, N. *Ann. Math.* **157,** 335–431 (2003).
4. Hsiang, W.-Y. *Int. J. Math.* **4,** 739–831 (1993).
5. Hales, T. C. *Discrete Comput. Geom.* **25,** 1–22 (2001).

Flyspeck
▶ www.math.pitt.edu/~thales/flyspeck/index.html

S. TERRY/SPL

UNIV. PITTSBURGH CIDDE