Solving Mathematical Challenges with Symbolic AI

Marijn J.H. Heule

Carnegie Mellon University

27th European Conference on Artificial Intelligence

Santiago de Compostela October 22, 2024

AI for Mathematics

A.I. Is Coming for Mathematics, Too

For thousands of years, mathematicians have adapted to the latest advances in logic and reasoning. Are they ready for artificial intelligence?

Move Over, Mathematicians, Here Comes AlphaProof

A.I. is getting good at math — and might soon make a worthy collaborator for humans.

AI for Mathematics

A.I. Is Coming for Mathematics, Too

For thousands of years, mathematicians have adapted to the latest advances in logic and reasoning. Are they ready for artificial intelligence?

Move Over, Mathematicians, Here Comes AlphaProof

A.I. is getting good at math — and might soon make a worthy collaborator for humans.

Mathematics is the perfect playground to get AI right

- Symbolic AI offers essential logic-based reasoning
- Highly trustworthy results thanks to (formal) proofs

50 Years of Successes in Computer-Aided Mathematics

1976 Four-Color Theorem

1998 Kepler Conjecture

2010 "God's Number = 20": Optimal Rubik's cube strategy

2014 Boolean Erdős discrepancy problem

2016 Boolean Pythagorean triples problem

2018 Schur Number Five

2019 Keller's Conjecture

2022 Packing Number of Square Grid

2023 Empty Hexagon in Every 30 Points

50 Years of Successes in Computer-Aided Mathematics

1976 Four-Color Theorem

1998 Kepler Conjecture

2010 "God's Number = 20": Optimal Rubik's cube strategy

2014 Boolean Erdős discrepancy problem (using a SAT solver)

2016 Boolean Pythagorean triples problem (using a SAT solver)

2018 Schur Number Five (using a SAT solver)

2019 Keller's Conjecture (using a SAT solver)

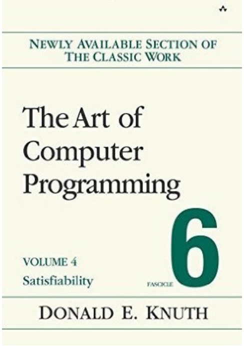
2022 Packing Number of Square Grid (using a SAT solver)

2023 Empty Hexagon in Every 30 Points (using a SAT solver)

Breakthrough in SAT Solving in the Last 30 Years

Satisfiability (SAT) problem: Can a Boolean formula be satisfied?

mid '90s: formulas solvable with thousands of variables and clauses
now: formulas solvable with millions of variables and clauses



Edmund Clarke: *"a key technology of the 21st century"* [Biere, Heule, vanMaaren, Walsh '09/'21]

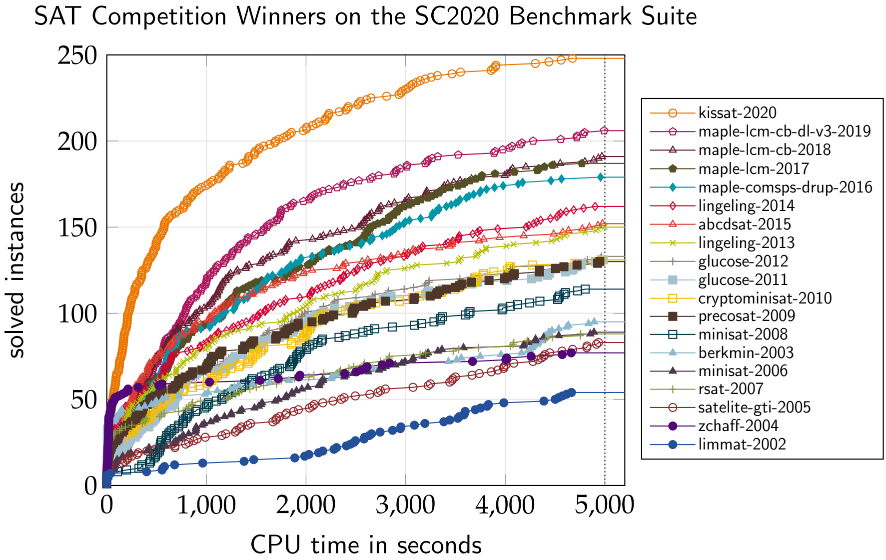
Donald Knuth: *"evidently a killer app, because it is key to the solution of so many other problems"* [Knuth '15]

Naive SAT Solving: Truth Table

$$\Gamma := (p \vee \overline{q}) \wedge (q \vee r) \wedge (\overline{r} \vee \overline{p})$$

p	q	r	falsifies	eval(Γ)
0	0	0	$(q \vee r)$	0
0	0	1	—	1
0	1	0	$(p \vee \overline{q})$	0
0	1	1	$(p \vee \overline{q})$	0
1	0	0	$(q \vee r)$	0
1	0	1	$(\overline{r} \vee \overline{p})$	0
1	1	0	—	1
1	1	1	$(\overline{r} \vee \overline{p})$	0

Progress of SAT Solvers



Satisfiability and Mathematics

Symbolic AI Proofs

Future and Challenges

Schur's Theorem [Schur 1916]

Will any coloring of the positive integers with red and blue result in a monochromatic solution of $a + b = c$?

$$1 + 1 = 2 \qquad 1 + 2 = 3 \qquad 1 + 3 = 4$$
$$1 + 4 = 5 \qquad 2 + 2 = 4 \qquad 2 + 3 = 5$$

Schur's Theorem [Schur 1916]

Will any coloring of the positive integers with red and blue result in a monochromatic solution of $a+b=c$?

$$1+1=2 \qquad 1+2=3 \qquad 1+3=4$$
$$1+4=5 \qquad 2+2=4 \qquad 2+3=5$$

Schur's Theorem [Schur 1916]

Will any coloring of the positive integers with red and blue result in a monochromatic solution of $a + b = c$?

$$1 + 1 = 2 \qquad 1 + 2 = 3 \qquad 1 + 3 = 4$$
$$1 + 4 = 5 \qquad 2 + 2 = 4 \qquad 2 + 3 = 5$$

Schur's Theorem [Schur 1916]

Will any coloring of the positive integers with red and blue result in a monochromatic solution of $a + b = c$?

$$1 + 1 = 2 \qquad 1 + 2 = 3 \qquad 1 + 3 = 4$$
$$1 + 4 = 5 \qquad 2 + 2 = 4 \qquad 2 + 3 = 5$$

Schur's Theorem [Schur 1916]

Will any coloring of the positive integers with red and blue result in a monochromatic solution of $a + b = c$? Yes

$$1 + 1 = 2 \qquad 1 + 2 = 3 \qquad 1 + 3 = 4$$
$$1 + 4 = 5 \qquad 2 + 2 = 4 \qquad 2 + 3 = 5$$

Schur's Theorem [Schur 1916]

Will any coloring of the positive integers with red and blue result in a monochromatic solution of $a + b = c$? Yes

$1 + 1 = 2$	$1 + 2 = 3$	$1 + 3 = 4$
$1 + 4 = 5$	$2 + 2 = 4$	$2 + 3 = 5$

Theorem (Schur's Theorem)

For every positive integer k, there exists a number $S(k)$, such that $[1, S(k)]$ can be colored with k colors while avoiding a monochromatic solution of $a + b = c$ with $a, b, c \leq S(k)$, while this is impossible for $[1, S(k)+1]$.

$S(1) = 1, S(2) = 4, S(3) = 13, S(4) = 44$ [Baumert 1965].

Schur's Theorem [Schur 1916]

Will any coloring of the positive integers with red and blue result in a monochromatic solution of $a + b = c$? Yes

$1 + 1 = 2$	$1 + 2 = 3$	$1 + 3 = 4$
$1 + 4 = 5$	$2 + 2 = 4$	$2 + 3 = 5$

Theorem (Schur's Theorem)

For every positive integer k, there exists a number $S(k)$, such that $[1, S(k)]$ can be colored with k colors while avoiding a monochromatic solution of $a + b = c$ with $a, b, c \leq S(k)$, while this is impossible for $[1, S(k)+1]$.

$S(1) = 1, S(2) = 4, S(3) = 13, S(4) = 44$ [Baumert 1965].

We show that $S(5) = 160$ [Heule 2018].

Schur's Theorem [Schur 1916]

Will any coloring of the positive integers with red and blue result in a monochromatic solution of $a + b = c$? Yes

$1 + 1 = 2$	$1 + 2 = 3$	$1 + 3 = 4$
$1 + 4 = 5$	$2 + 2 = 4$	$2 + 3 = 5$

Theorem (Schur's Theorem)

For every positive integer k, there exists a number $S(k)$, such that $[1, S(k)]$ can be colored with k colors while avoiding a monochromatic solution of $a + b = c$ with $a, b, c \leq S(k)$, while this is impossible for $[1, S(k)+1]$.

$S(1) = 1, S(2) = 4, S(3) = 13, S(4) = 44$ [Baumert 1965].

We show that $S(5) = 160$ [Heule 2018]. Proof: 2 petabytes

Pythagorean Triples Problem (I) [Ronald Graham, early 80's]

Will any coloring of the positive integers with red and blue result in a monochromatic Pythagorean Triple $a^2 + b^2 = c^2$?

$3^2 + 4^2 = 5^2$	$6^2 + 8^2 = 10^2$	$5^2 + 12^2 = 13^2$	$9^2 + 12^2 = 15^2$
$8^2 + 15^2 = 17^2$	$12^2 + 16^2 = 20^2$	$15^2 + 20^2 = 25^2$	$7^2 + 24^2 = 25^2$
$10^2 + 24^2 = 26^2$	$20^2 + 21^2 = 29^2$	$18^2 + 24^2 = 30^2$	$16^2 + 30^2 = 34^2$
$21^2 + 28^2 = 35^2$	$12^2 + 35^2 = 37^2$	$15^2 + 36^2 = 39^2$	$24^2 + 32^2 = 40^2$

Pythagorean Triples Problem (I) [Ronald Graham, early 80's]

Will any coloring of the positive integers with red and blue result in a monochromatic Pythagorean Triple $a^2 + b^2 = c^2$?

$3^2 + 4^2 = 5^2$	$6^2 + 8^2 = 10^2$	$5^2 + 12^2 = 13^2$	$9^2 + 12^2 = 15^2$
$8^2 + 15^2 = 17^2$	$12^2 + 16^2 = 20^2$	$15^2 + 20^2 = 25^2$	$7^2 + 24^2 = 25^2$
$10^2 + 24^2 = 26^2$	$20^2 + 21^2 = 29^2$	$18^2 + 24^2 = 30^2$	$16^2 + 30^2 = 34^2$
$21^2 + 28^2 = 35^2$	$12^2 + 35^2 = 37^2$	$15^2 + 36^2 = 39^2$	$24^2 + 32^2 = 40^2$

Best lower bound: a bi-coloring of $[1, 7664]$ s.t. there is no monochromatic Pythagorean Triple [Cooper & Overstreet 2015].

Myers conjectures that the answer is No [PhD thesis, 2015].

Pythagorean Triples Problem (II) [Ronald Graham, early 80's]

Will any coloring of the positive integers with red and blue result in a monochromatic Pythagorean Triple $a^2 + b^2 = c^2$?

A bi-coloring of $[1, n]$ is encoded using Boolean variables p_i with $i \in \{1, 2, \ldots, n\}$ such that $p_i = 1$ $(= 0)$ means that i is colored red (blue). For each Pythagorean Triple $a^2 + b^2 = c^2$, two clauses are added: $(p_a \vee p_b \vee p_c)$ and $(\overline{p}_a \vee \overline{p}_b \vee \overline{p}_c)$.

Pythagorean Triples Problem (II) [Ronald Graham, early 80's]

Will any coloring of the positive integers with red and blue result in a monochromatic Pythagorean Triple $a^2 + b^2 = c^2$?

A bi-coloring of $[1, n]$ is encoded using Boolean variables p_i with $i \in \{1, 2, \ldots, n\}$ such that $p_i = 1$ $(= 0)$ means that i is colored red (blue). For each Pythagorean Triple $a^2 + b^2 = c^2$, two clauses are added: $(p_a \vee p_b \vee p_c)$ and $(\overline{p}_a \vee \overline{p}_b \vee \overline{p}_c)$.

Theorem ([Heule, Kullmann, and Marek (2016)])

$[1, 7824]$ *can be bi-colored s.t. there is no monochromatic Pythagorean Triple. This is impossible for* $[1, 7825]$.

Pythagorean Triples Problem (II) [Ronald Graham, early 80's]

Will any coloring of the positive integers with red and blue result in a monochromatic Pythagorean Triple $a^2 + b^2 = c^2$?

A bi-coloring of $[1, n]$ is encoded using Boolean variables p_i with $i \in \{1, 2, \ldots, n\}$ such that $p_i = 1$ $(= 0)$ means that i is colored red (blue). For each Pythagorean Triple $a^2 + b^2 = c^2$, two clauses are added: $(p_a \vee p_b \vee p_c)$ and $(\overline{p}_a \vee \overline{p}_b \vee \overline{p}_c)$.

Theorem ([Heule, Kullmann, and Marek (2016)])

$[1, 7824]$ *can be bi-colored s.t. there is no monochromatic Pythagorean Triple. This is impossible for* $[1, 7825]$.

4 CPU years computation, but 2 days on cluster (800 cores)

Pythagorean Triples Problem (II) [Ronald Graham, early 80's]

Will any coloring of the positive integers with red and blue result in a monochromatic Pythagorean Triple $a^2 + b^2 = c^2$?

A bi-coloring of $[1, n]$ is encoded using Boolean variables p_i with $i \in \{1, 2, \ldots, n\}$ such that $p_i = 1$ $(= 0)$ means that i is colored red (blue). For each Pythagorean Triple $a^2 + b^2 = c^2$, two clauses are added: $(p_a \vee p_b \vee p_c)$ and $(\overline{p}_a \vee \overline{p}_b \vee \overline{p}_c)$.

Theorem ([Heule, Kullmann, and Marek (2016)])

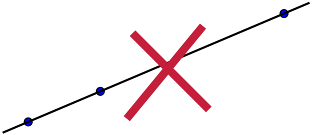
$[1, 7824]$ *can be bi-colored s.t. there is no monochromatic Pythagorean Triple. This is impossible for* $[1, 7825]$.

4 CPU years computation, but 2 days on cluster (800 cores)

200 terabytes proof, but validated with verified checker

An Empty Hexagon in Every Set of 30 Points

Geometry and SAT: Shapes in point sets without three points on a line

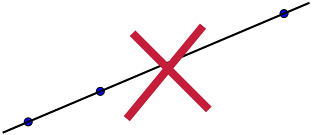


k-hole: empty k-point convex shape

An Empty Hexagon in Every Set of 30 Points

Geometry and SAT: Shapes in point sets without three points on a line

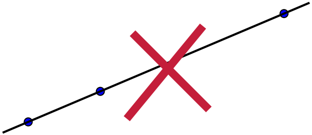
k-hole: empty k-point convex shape



- Every set of 5 points contains in a 4-hole [Klein '32]

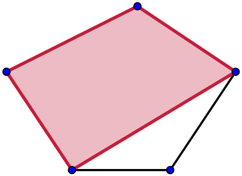
An Empty Hexagon in Every Set of 30 Points

Geometry and SAT: Shapes in point sets without three points on a line



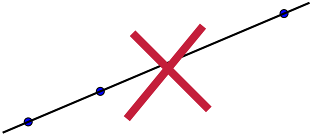
k-hole: empty k-point convex shape

- Every set of 5 points contains in a 4-hole [Klein '32]



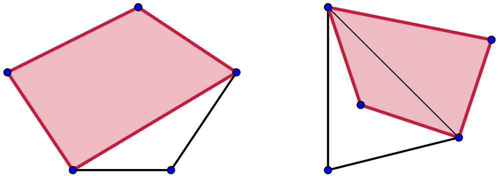
An Empty Hexagon in Every Set of 30 Points

Geometry and SAT: Shapes in point sets without three points on a line



k-hole: empty k-point convex shape

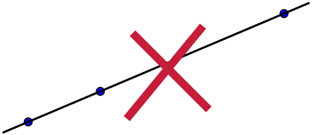
- Every set of 5 points contains in a 4-hole [Klein '32]



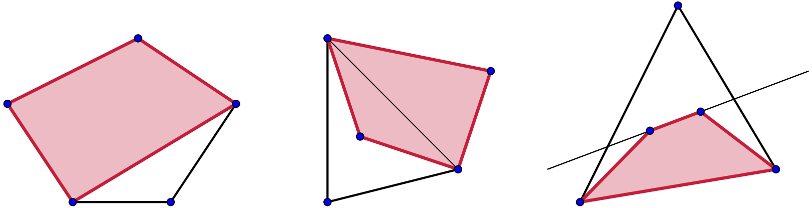
An Empty Hexagon in Every Set of 30 Points

Geometry and SAT: Shapes in point sets without three points on a line

k-hole: empty k-point convex shape

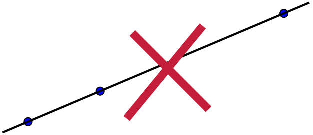


- Every set of 5 points contains in a 4-hole [Klein '32]



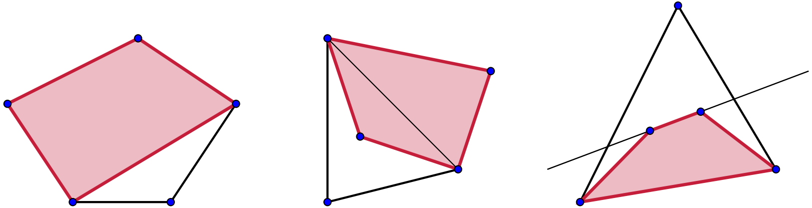
An Empty Hexagon in Every Set of 30 Points

Geometry and SAT: Shapes in point sets without three points on a line



k-hole: empty k-point convex shape

- Every set of 5 points contains in a 4-hole [Klein '32]

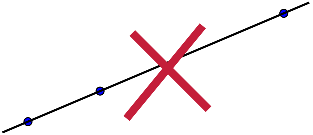


- Every set of 10 points contains in a 5-hole [Harborth '78]
- 7-holes can always be avoided [Horton '83]

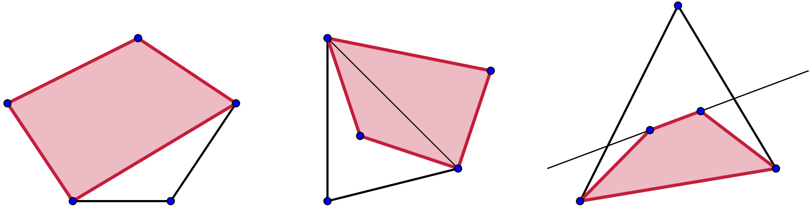
An Empty Hexagon in Every Set of 30 Points

Geometry and SAT: Shapes in point sets without three points on a line

k-hole: empty k-point convex shape



- Every set of 5 points contains in a 4-hole [Klein '32]



- Every set of 10 points contains in a 5-hole [Harborth '78]
- 7-holes can always be avoided [Horton '83]
- Every set of 30 points contains in a 6-hole (using SAT) [Heule & Scheucher 2023]

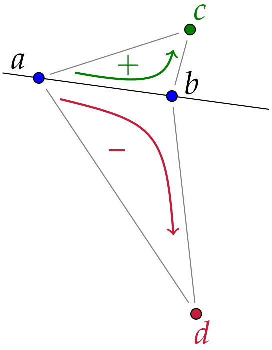
Orientation Variables and Realizability Constraints

No explicit coordinates of points

Instead for every triple $a < b < c$, one orientation variable $O_{a,b,c}$ to denote whether point c is above the line ab

Not all assignments are realizable

- Constraints can eliminate many unrealizable assignments



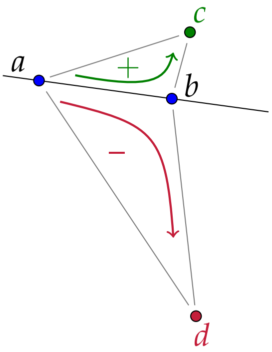
Orientation Variables and Realizability Constraints

No explicit coordinates of points

Instead for every triple $a < b < c$, one orientation variable $O_{a,b,c}$ to denote whether point c is above the line ab

Not all assignments are realizable

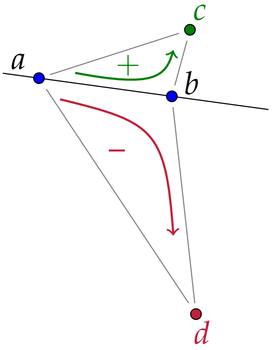
- Constraints can eliminate many unrealizable assignments



Hole variables $H_{a,b,c}$ denote whether triangle abc is empty

- Defined using orientation variables

Orientation Variables and Realizability Constraints



No explicit coordinates of points

Instead for every triple $a < b < c$, one orientation variable $O_{a,b,c}$ to denote whether point c is above the line ab

Not all assignments are realizable

- Constraints can eliminate many unrealizable assignments

Hole variables $H_{a,b,c}$ denote whether triangle abc is empty

- Defined using orientation variables

Many possible SAT encodings

- Big impact on performance
- Machine learning can help!

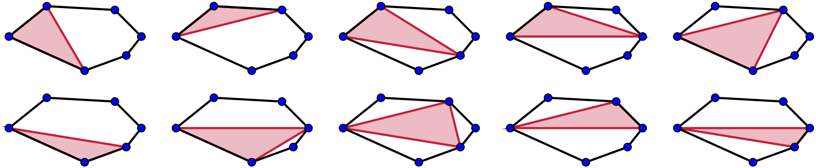
Empty Hexagon Encoding

Given 6 points, how many empty triangles with these points guarantee an empty hexagon (possibly among other points)?

Empty Hexagon Encoding

Given 6 points, how many empty triangles with these points guarantee an empty hexagon (possibly among other points)?

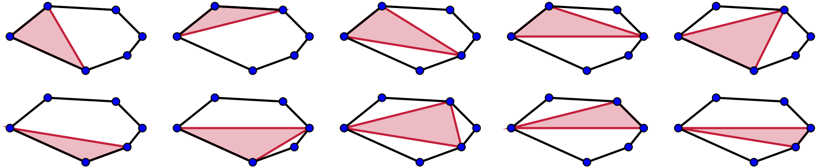
If the points may not be in convex position: 10



Empty Hexagon Encoding

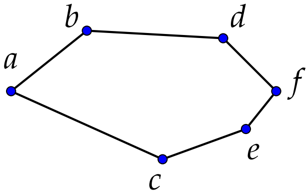
Given 6 points, how many empty triangles with these points guarantee an empty hexagon (possibly among other points)?

If the points may not be in convex position: 10



If the points are in convex position:

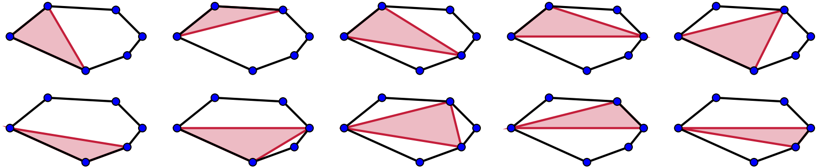
- Requires assignment to four orientation variables
- Includes info which points are above/below the line a to f



Empty Hexagon Encoding

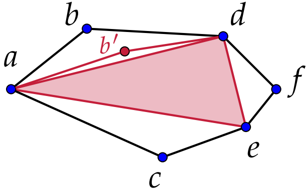
Given 6 points, how many empty triangles with these points guarantee an empty hexagon (possibly among other points)?

If the points may not be in convex position: 10



If the points are in convex position: 1

- Requires assignment to four orientation variables
- Includes info which points are above/below the line a to f

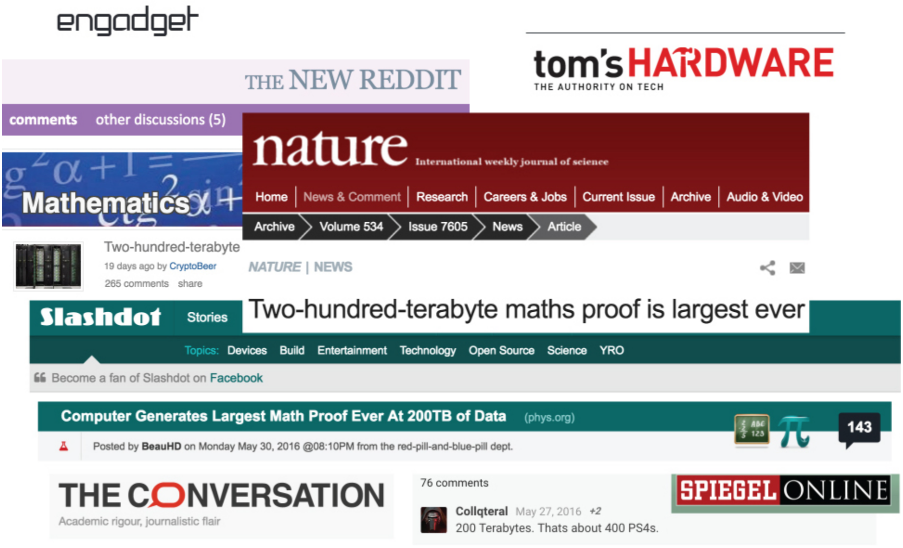


Satisfiability and Mathematics

Symbolic AI Proofs

Future and Challenges

Media: "The Largest Math Proof Ever"



Proof-Producing Tools: Arbitrarily Complex Solvers

Proof-producing tools with verified checkers is a powerful idea:

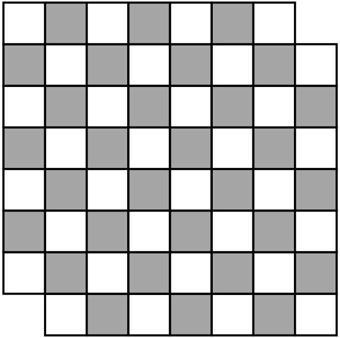
- Don't worry about correctness or completeness of tools;
- Facilitates making tools more complex and efficient; while
- Full confidence in results. [Heule, Hunt, Kaufmann, Wetzler '17]

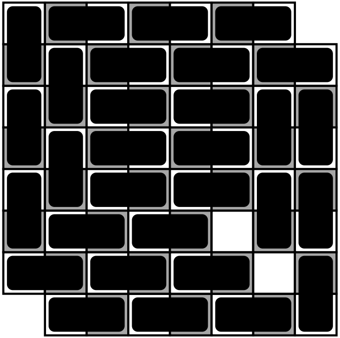
ACL2 Isabelle CAKEML LEAN

Formally verified checkers now also used in industry

Mutilated Chessboards: "A Tough Nut to Crack" [McCarthy]

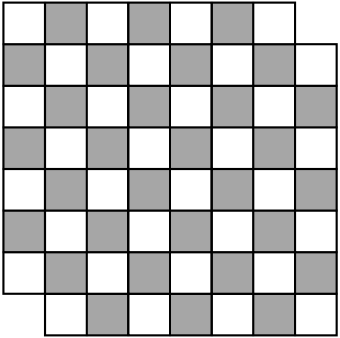
Can a chessboard be fully covered with dominos after removing two diagonally opposite corner squares?

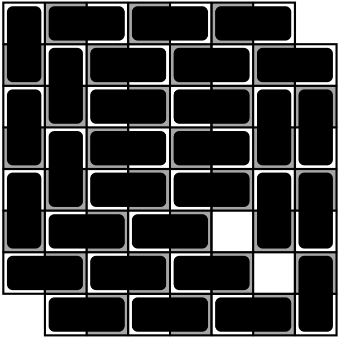




Mutilated Chessboards: "A Tough Nut to Crack" [McCarthy]

Can a chessboard be fully covered with dominos after removing two diagonally opposite corner squares?



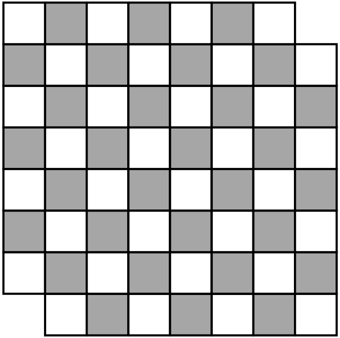


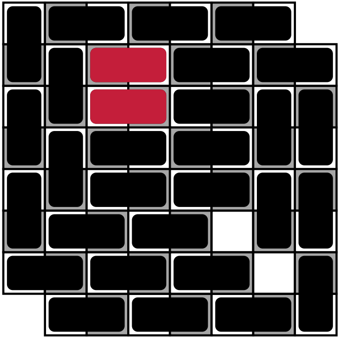
Easy to refute based on the following two observations:

- There are more white squares than black squares; and
- A domino covers exactly one white and one black square.

Mutilated Chessboards: "A Tough Nut to Crack" [McCarthy]

Can a chessboard be fully covered with dominos after removing two diagonally opposite corner squares?



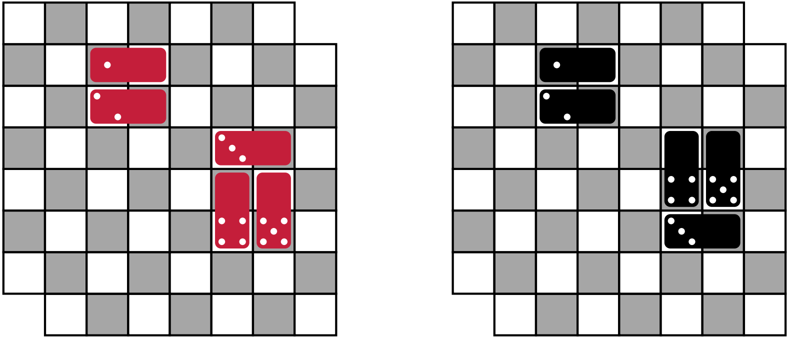


Easy to refute based on the following two observations:

- There are more white squares than black squares; and
- A domino covers exactly one white and one black square.

Symbolic AI is Local Reasoning

The chessboard pattern argument is challenging to find, but an alternative short argument can be found automatically...



Symbolic AI tools produce proofs that can be very different compared to human-made proofs for the same problem

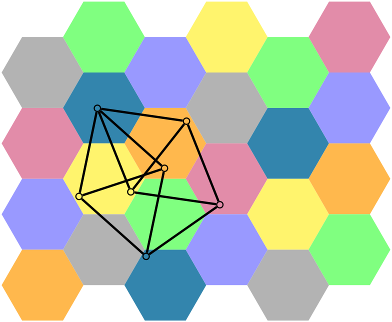
Satisfiability and Mathematics

Symbolic AI Proofs

Future and Challenges

Chromatic Number of the Plane (CNP)

The Hadwiger-Nelson problem (around 1950):
How many colors are required to color the plane such that each pair of points that are exactly 1 apart are colored differently?



- The Moser Spindle graph shows the lower bound of 4
- A coloring of the plane showing the upper bound of 7

CNP: First progress in decades

Recently enormous progress:

- Lower bound of 5 [DeGrey '18] based on a 1581-vertex graph
- This breakthrough started a polymath project
- Improved bounds of the fractional chromatic number of the plane

CNP: First progress in decades

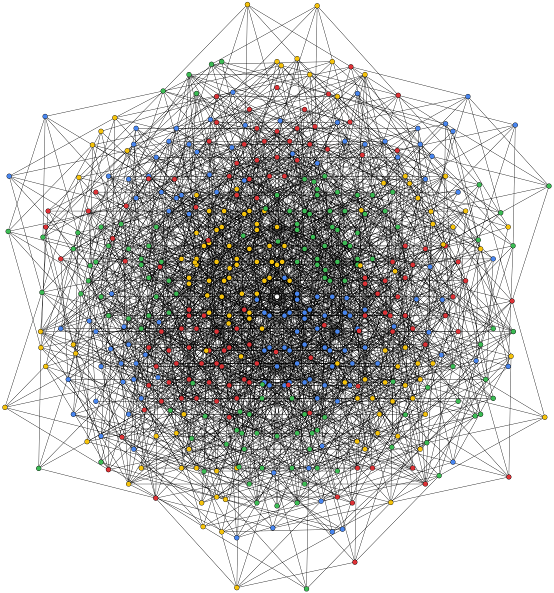
Recently enormous progress:

- Lower bound of 5 [DeGrey '18] based on a 1581-vertex graph
- This breakthrough started a polymath project
- Improved bounds of the fractional chromatic number of the plane

We found smaller graphs with SAT:

- 874 vertices on April 14, 2018
- 803 vertices on April 30, 2018
- 610 vertices on May 14, 2018

Proof Minimization: 510 Vertices [Heule 2021]



Beyond NP: The Collatz Conjecture

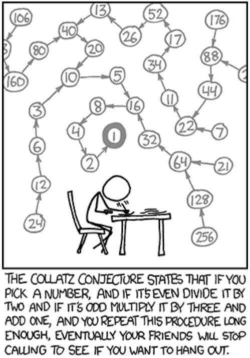
Resolving foundational algorithm questions

$$Col(n) = \begin{cases} n/2 & \text{if } n \text{ is even} \\ (3n+1)/2 & \text{if } n \text{ is odd} \end{cases}$$

Does $\texttt{while}(n > 1)\ n = Col(n)\texttt{;}$ terminate?

Find a non-negative function $fun(n)$ s.t.

$$\forall n > 1 : fun(n) > fun(Col(n))$$



source: xkcd.com/710

Beyond NP: The Collatz Conjecture

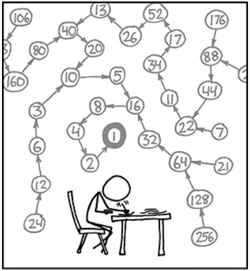
Resolving foundational algorithm questions

$$Col(n) = \begin{cases} n/2 & \text{if } n \text{ is even} \\ (3n+1)/2 & \text{if } n \text{ is odd} \end{cases}$$

Does $\texttt{while}(n > 1)\ n = Col(n);$ terminate?

Find a non-negative function $fun(n)$ s.t.

$$\forall n > 1 : fun(n) > fun(Col(n))$$



THE COLLATZ CONJECTURE STATES THAT IF YOU PICK A NUMBER, AND IF IT'S EVEN DIVIDE IT BY TWO AND IF IT'S ODD MULTIPLY IT BY THREE AND ADD ONE, AND YOU REPEAT THIS PROCEDURE LONG ENOUGH, EVENTUALLY YOUR FRIENDS WILL STOP CALLING TO SEE IF YOU WANT TO HANG OUT.

source: xkcd.com/710

Can we construct a function s.t. $fun(n) > fun(Col(n))$ holds?

$fun(3)$	$fun(5)$	$fun(8)$	$fun(4)$	$fun(2)$	$fun(1)$
5	4	3	2	1	0

Collatz Conjecture: Successes and Challenge

Success. Our tool proves termination of Farkas' variant:

$$F(n) = \begin{cases} \frac{n-1}{3} & \text{if } n \equiv 1 \pmod 3 \\ \frac{n}{2} & \text{if } n \equiv 0 \text{ or } n \equiv 2 \pmod 6 \\ \frac{3n+1}{2} & \text{if } n \equiv 3 \text{ or } n \equiv 5 \pmod 6 \end{cases}$$

Collatz Conjecture: Successes and Challenge

Success. Our tool proves termination of Farkas' variant:

$$F(n) = \begin{cases} \frac{n-1}{3} & \text{if } n \equiv 1 \pmod 3 \\ \frac{n}{2} & \text{if } n \equiv 0 \text{ or } n \equiv 2 \pmod 6 \\ \frac{3n+1}{2} & \text{if } n \equiv 3 \text{ or } n \equiv 5 \pmod 6 \end{cases}$$

Challenge ($500). An easier generalized Collatz problem is open:

$$H(n) = \begin{cases} \frac{3n}{4} & \text{if } n \equiv 0 \pmod 4 \\ \frac{9n+1}{8} & \text{if } n \equiv 7 \pmod 8 \\ \perp & \text{otherwise} \end{cases}$$

Takeaways

Successes, Advances, and Trust:

- A performance boost of symbolic AI technology allows solving challenges in mathematics
- Creative, but possibly gigantic proofs can be validated using formally-verified checkers
- Future: combine symbolic AI and ML

Challenges ready for symbolic AI + ML?

- Chromatic number of the plane
- Optimal matrix multiplication
- Hadamard conjecture
- Collatz conjecture
- ...