Overcoming Intractable Complexity in MetiTarski: An Automatic Theorem Prover for Real-Valued Functions

Prof. Lawrence C Paulson, University of Cambridge

Computability And Complexity In Analysis, 24–27 June 2012

real quantifier elimination (QE)

$$\exists x \, [ax^2 + bx + c = 0]$$

$$\iff$$

$$b^2 \geq 4ac \wedge (c = 0 \vee a \neq 0 \vee \cancel{b^2 > 4ac})$$

$$b \neq 0$$

The equivalent quantifier-free formula can be messy...

real QE: some history

- Tarski (1948): A first-order RCF formula can be replaced by an equivalent, quantifier-free one.
- Implies the decidability of RCF
- … and also the decidability of Euclidean geometry.

RCF (*real-closed field*): any field elementarily equivalent to the reals

QE is expensive!

- Tarski's algorithm has *non-elementary* complexity! There are usable algorithms by Cohen, Hörmander, etc.
- The key approach: *cylindrical algebraic decomposition* (Collins, 1975)
- But quantifier elimination can yield a **huge** quantifier-free formula
- ... *doubly exponential* in the number of quantifiers (Davenport and Heintz, 1988)

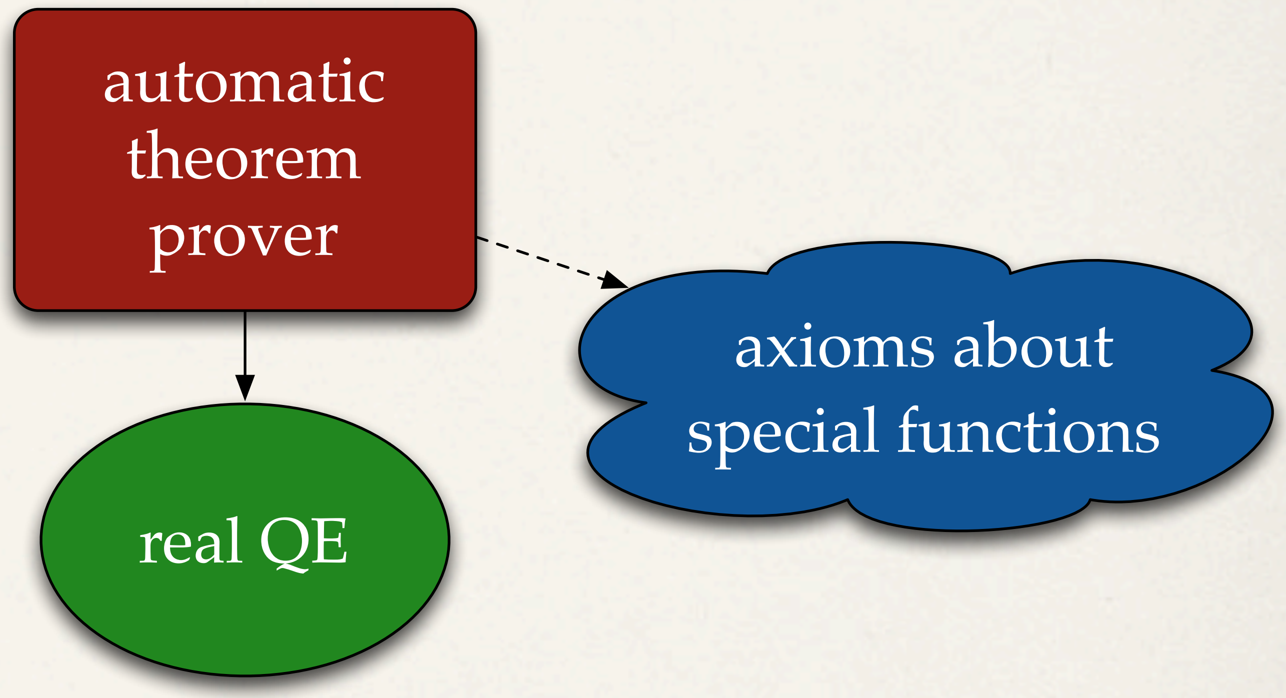
No efficient algorithm can exist. Do we give up?
Of course not...

Can real QE solve *even harder* problems? —with exp, ln, etc.?

- Decision procedures exist for some fragments… probably
- … but trigonometric functions obviously destroy decidability.
- The alternative? Stop looking for decision procedures. Employ **heuristics**…

idea: combine real QE with theorem proving

- To prove statements involving real-valued special functions.
- This *theorem-proving* approach delivers machine-verifiable evidence to justify its claims.
- Based on heuristics, it often finds proofs—but with no assurance of getting an answer.
- Real QE will be called as a **decision procedure**.



But why call something intractable as a subroutine??

- This is basic research. Theorem proving for real-valued functions has been largely unexplored.
- There could be many applications in science and engineering.
- High complexity does not imply uselessness. As with the boolean satisfiability (SAT) problem.

Another example: Higher-order unification is only semi-decidable...

but it is the foundation of Isabelle, a well-known interactive theorem prover.

MetiTarski: an automatic theorem prover coupled with RCF decision procedures

- *Objective*: to prove first-order statements involving real-valued functions such as exp, ln, sin, cos, $\tan^{-1}$, …
- *Method*: **resolution** theorem proving augmented with
  - **axioms** bounding these functions by rational functions
  - **heuristics** to isolate function occurrences and create RCF problems
  - … to be solved using QE tools: QEPCAD, Mathematica, Z3, etc.

the basic idea

Our approach involves replacing functions by *rational function upper or lower bounds*.

We end up with *polynomial inequalities*: in other words, RCF problems

... and first-order formulae involving +, −, × and ≤ (on reals) are **decidable**.

Real QE and *resolution theorem proving* are the core technologies.

A Simple Proof:

$$\forall x\ |e^x - 1| \le e^{|x|} - 1$$

negating the claim

$$e^{|c|} < 1 + |e^c - 1|$$

absolute value

$$0 \le c \vee e^{-c} < 1 + |e^c - 1|$$

absolute value

$$1 \le e^c \vee 0 \le c \vee e^{-c} < 2 - e^c$$

lower bound: $1 - c \le e^{-c}$

$$1 \le e^c \vee 0 \le c \vee e^c < 1 + c$$

lower bound: $1 + c \le e^c$

$$1 \le e^c \vee 0 \le c$$

$0 \le c \Rightarrow 1 \le e^c$

$$1 \le e^c$$

absolute value

$$e^{|c|} < e^c \cancel{\vee e^c < 1}$$

absolute value, etc.

$$c < 0$$

$$\square$$

Some MetiTarski Theorems

$$
\begin{aligned}
0 < t \wedge 0 < v_f \implies & ((1.565 + .313 v_f)\cos(1.16t) \\
& + (.01340 + .00268 v_f)\sin(1.16t))e^{-1.34t} \\
& - (6.55 + 1.31 v_f)e^{-.318t} + v_f + 10 \geq 0
\end{aligned}
$$

$$
\begin{aligned}
0 \leq x \wedge x \leq 1.46 \times 10^{-6} \implies & \\
& (64.42\sin(1.71 \times 10^6 x) - 21.08\cos(1.71 \times 10^6 x))e^{9.05 \times 10^5 x} \\
& + 24.24 e^{-1.86 \times 10^6 x} > 0
\end{aligned}
$$

$$0 \leq x \wedge 0 \leq y \implies y \tanh(x) \leq \sinh(yx)$$

Each is proved in a few seconds!

some bounds for ln

- based on the continued fraction for ln(x+1)
- *much* more accurate than the Taylor expansion
- Simplicity can be exchanged for accuracy.
- With these, the maximum degree we use is 8.

$$\frac{x-1}{x} \leq \ln x \leq x-1$$

$$\frac{(1+5x)(x-1)}{2x(2+x)} \leq \ln x \leq \frac{(x+5)(x-1)}{2(2x+1)}$$

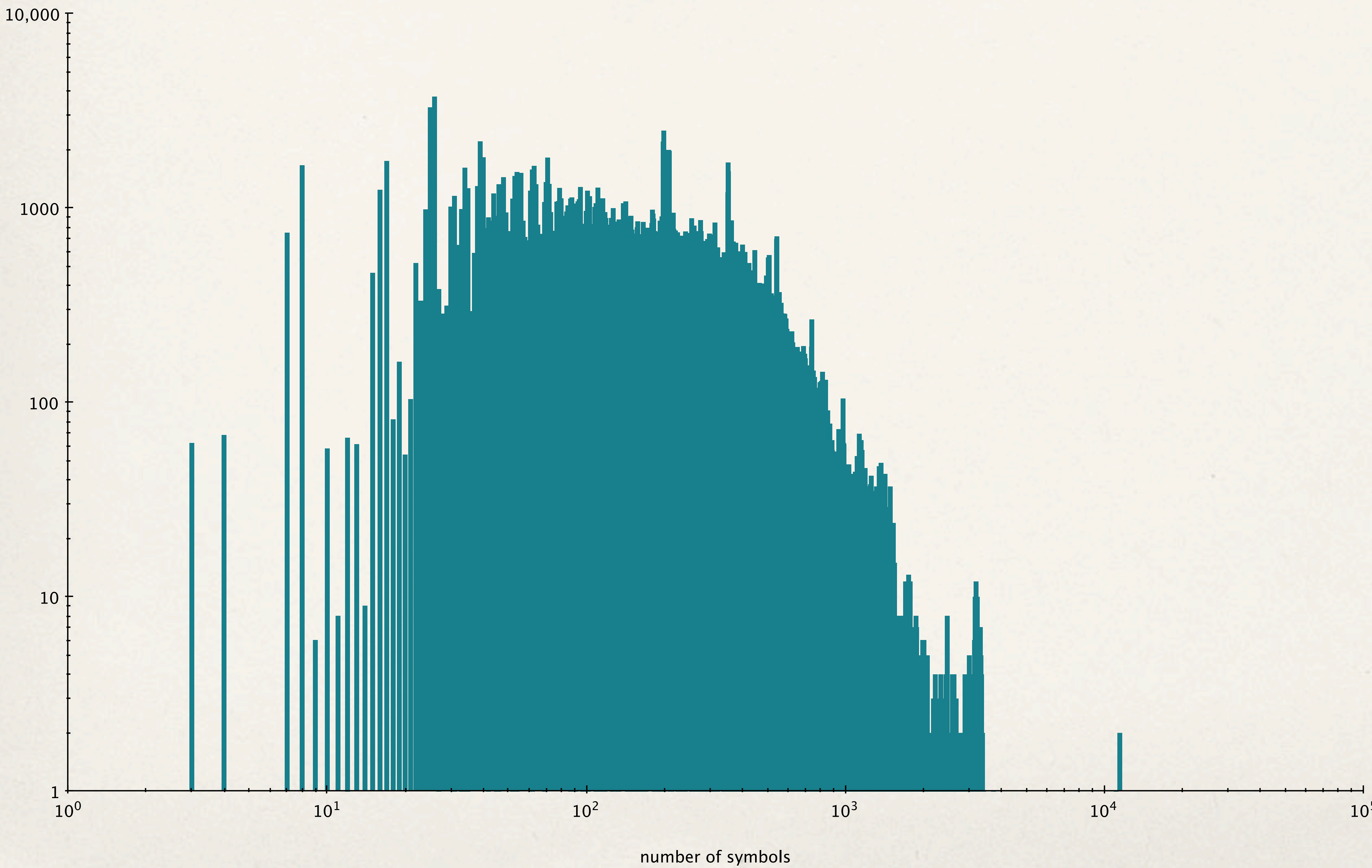
bounds for other functions

- a mix of *continued fraction* approximants and truncated *Taylor series*, etc, modified to suit various argument ranges and accuracies
- a tiny bit of **built-in knowledge** about signs, for example, $\exp(x) > 0$
- NO fundamental mathematical knowledge, for example, the geometric interpretation of trigonometric functions
- MetiTarski can reason about any function that has well-behaved *upper and lower bounds* as rational functions.

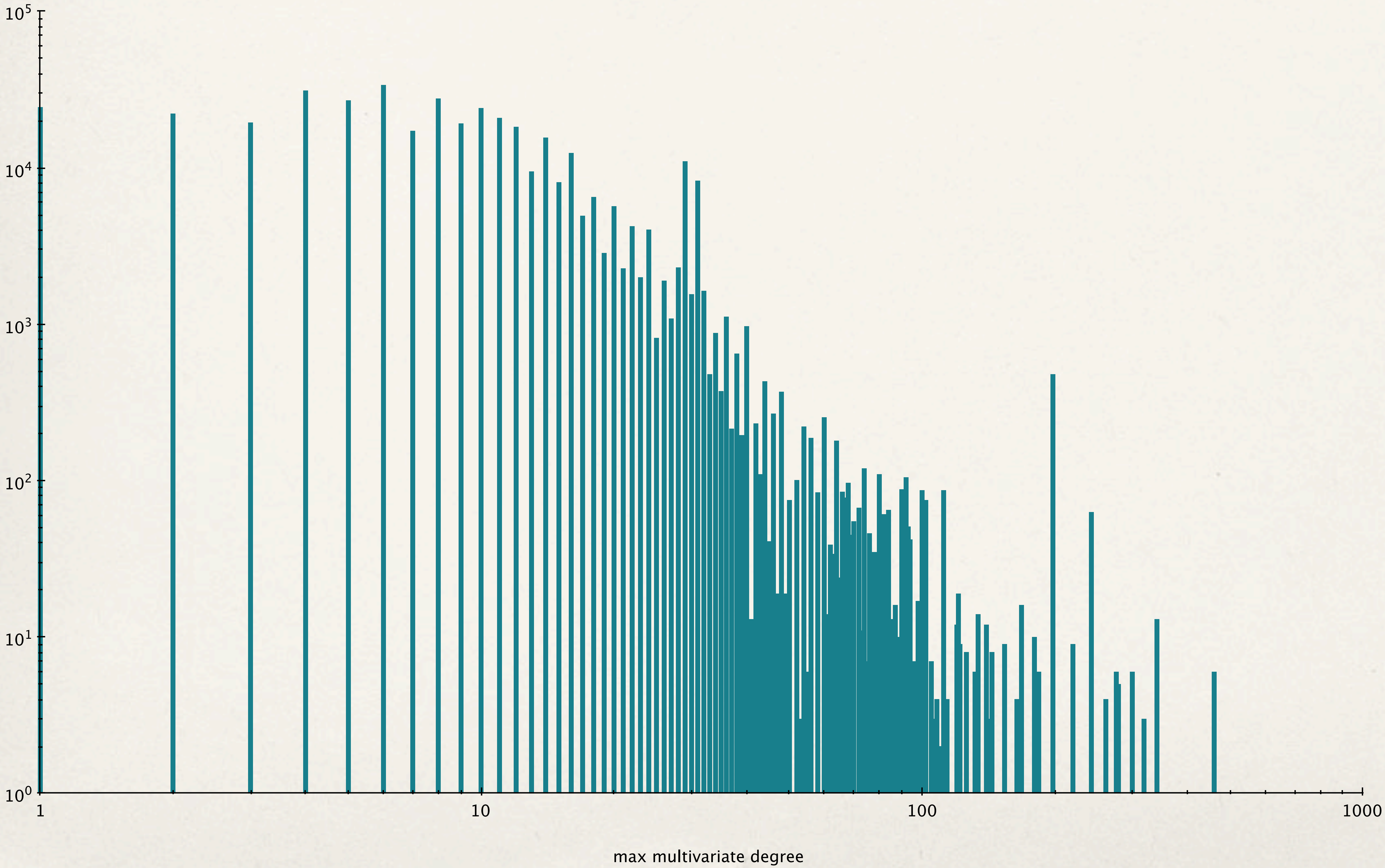
statistics about the RCF problems

- 400,000 RCF problems generated from 859 MetiTarski problems.
- Number of *symbols*: in some cases, 11,000 or more!
- Maximum *degree*: up to 460!
- But… number of *variables*? Typically just 1. No more than 8.

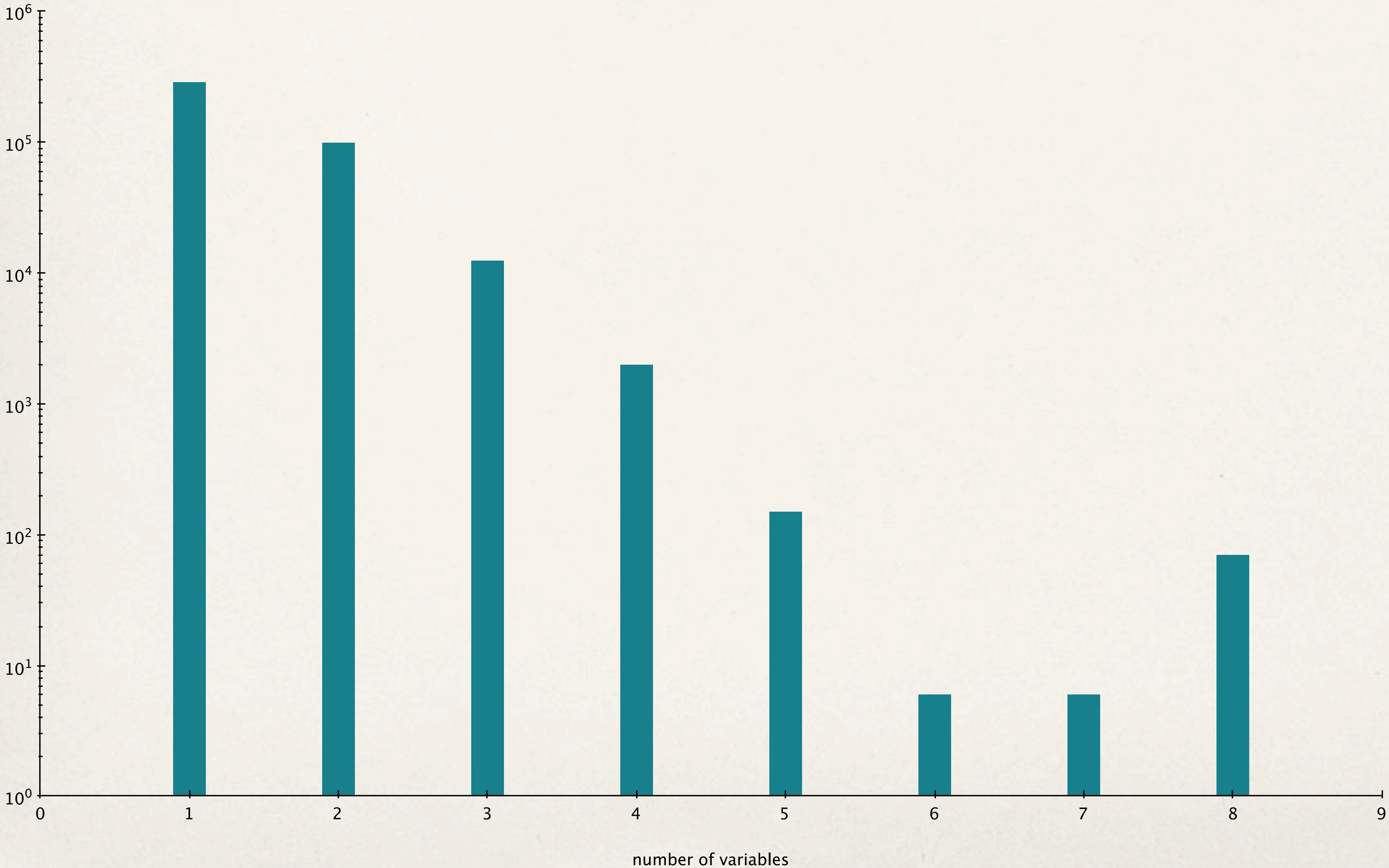
distribution of problem sizes (in symbols)



distribution of polynomial degrees (multivariate)



distribution of problem dimensions



introducing the QE solvers

QEPCAD (Hoon Hong, C. W. Brown et al.)
Venerable. Very fast for univariate problems.

Mathematica (Wolfram research)
Much faster than QEPCAD for 3–4 variables

Z3 (de Moura, Microsoft Research)
An SMT solver with non-linear reasoning.

a heuristic: *model sharing*

- MetiTarski applies QE only to existential formulas, $\exists x\ \exists y\ \ldots$
- Many of these turn out to be satisfiable,…
- and many satisfiable formulas have the *same model*.
- By maintaining a list of "successful" models, we can show many RCF formulas to be satisfiable **without performing QE**.

… because most of our RCF problems are satisfiable…

Problem	All RCF		SAT RCF		% SAT	
	#	secs	#	secs	#	secs
CONVOI2-sincos	268	3.28	194	2.58	72%	79%
exp-problem-9	1213	6.25	731	4.11	60%	66%
log-fun-ineq-e-weak	496	31.50	323	20.60	65%	65%
max-sin-2	2776	253.33	2,221	185.28	80%	73%
sin-3425b	118	39.28	72	14.71	61%	37%
sqrt-problem-13-sqrt3	2031	22.90	1403	17.09	69%	75%
tan-1-1var-weak	817	19.5	458	7.60	56%	39%
trig-squared3	742	32.92	549	20.66	74%	63%
trig-squared4	847	45.29	637	20.78	75%	46%
trigpoly-3514-2	1070	17.66	934	14.85	87%	84%

In one example, 2172 of 2221 satisfiable RCF problems can be settled using model sharing, with only 37 separate models.

introducing Strategy 1

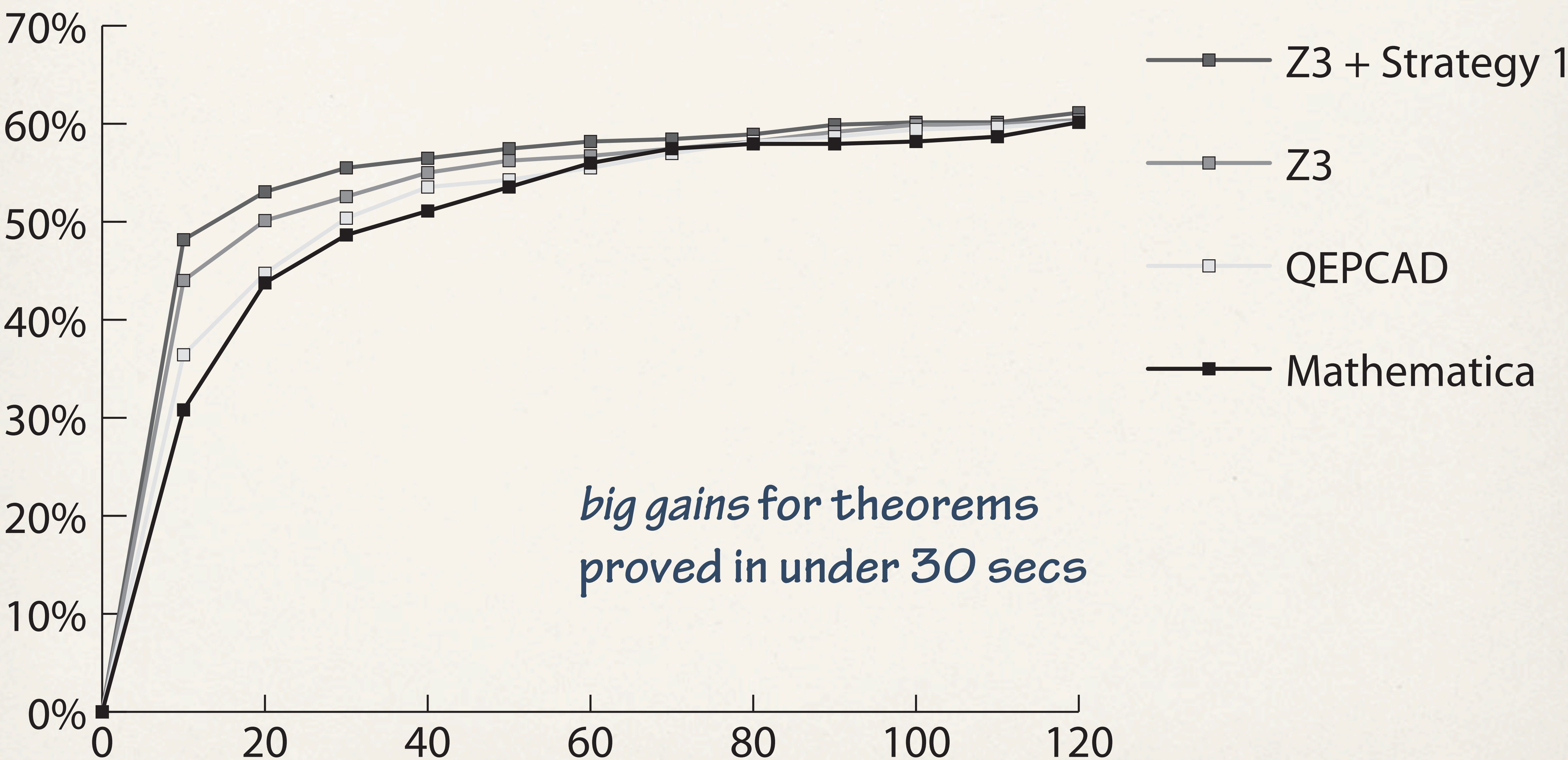
model sharing

+

omitting the standard test for *irreducibility*

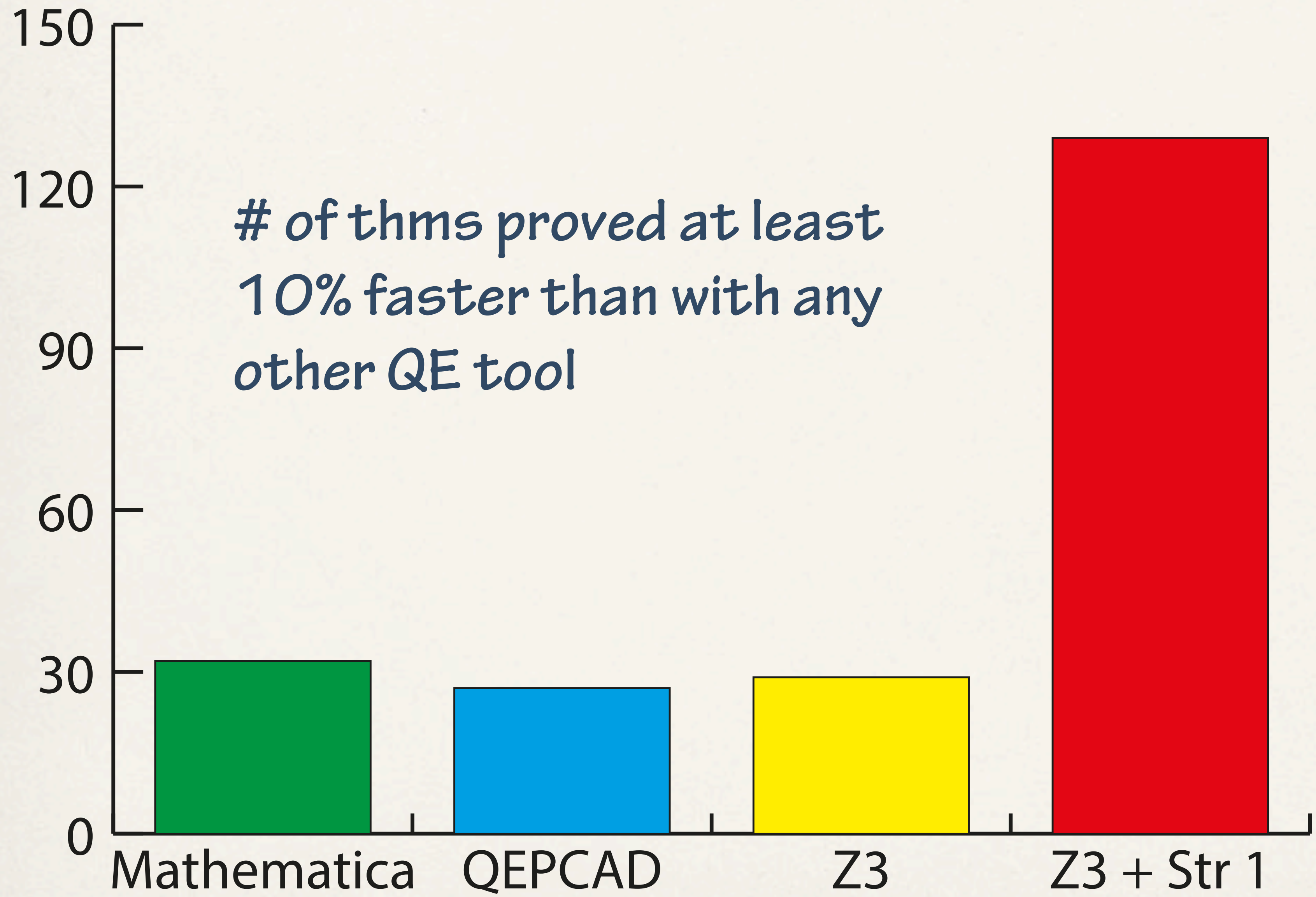
= Strategy 1

comparative results
(% proved in up to 120 secs)



big gains for theorems proved in under 30 secs

Strategy 1 finds the fastest proofs



possible applications

- *hybrid systems,* especially those involving transcendental functions
- showing stability of dynamical systems using *Lyapunov functions*
- real error analysis…?
- any application involving *ad hoc* real inequalities

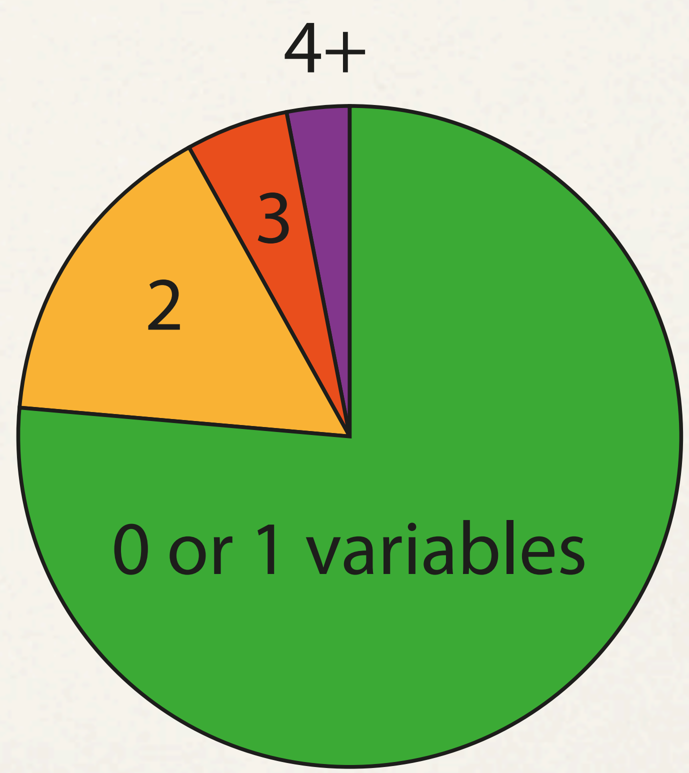
We are still looking...

inherent limitations

- Only non-sharp inequalities can be proved.
- Few MetiTarski proofs are mathematically elegant.
- Problems involving nested function calls can be very difficult.

research challenges

- Real QE is still much too slow! It's usually a serious bottleneck.
- We need to handle many more variables!
- Upper/lower bounds sometimes need *scaling* or *argument reduction*: how?
- How can we set the numerous options offered by RCF solvers?



conclusions

- Real QE is applicable now
- ... and there are ways to improve its performance.
- Nevertheless, its complexity poses continual difficulties.

the Cambridge team

James Bridge

Grant Passmore

Zongyan Huang

William Denman

acknowledgements

- *Edinburgh*: Paul Jackson; *Manchester*: Eva Navarro
- Assistance from C. W. Brown, A. Cuyt, I. Grant, J. Harrison, J. Hurd, D. Lester, C. Muñoz, U. Waldmann, etc.
- Behzad Akbarpour formalised most of the engineering examples.
- The research was supported by the Engineering and Physical Sciences Research Council [grant numbers EP/C013409/1,EP/I011005/1,EP/I010335/1].