log-fun-ineq-e-weak

$$\forall x \in (0, 12), y \in (-\infty, \infty)$$
$$xy \leq \frac{1}{5} + x\,\ln(x) + e^{y-1}$$

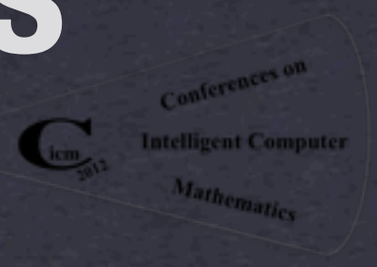
sin-3425b

$$\forall x \in (0, \infty), y \in (-\infty, \infty)$$
$$(x < y \wedge y^2 < 6) \Rightarrow \frac{\sin(y)}{\sin(x)} \leq 10^{-4} + \frac{y - \frac{1}{6}y^3 + \frac{1}{120}y^5}{x - \frac{1}{6}x^3 + \frac{1}{120}x^5}$$

CONVOI2-sincos

$$\forall t \in (0, \infty), v \in (0, \infty)$$
$$((1.565 + 0.313\,v)\cos(1.16\,t) + (0.01340 + 0.00268\,v)\sin(1.16\,t))\,e^{-1.34\,t}$$
$$-(6.55 + 1.31\,v)\,e^{-0.318\,t} + v + 10 \geq 0$$

Real Algebraic Strategies for MetiTarski Proofs



Grant Passmore (Cambridge & Edinburgh), Lawrence Paulson (Cambridge), Leo de Moura (MSR, Redmond)

MetiTarski: A theorem prover for real-valued special functions

- Many applications in mathematics and engineering require reasoning about real-valued special functions such as *sin, cos, tan, log, exp, arcsin, ...*

- MetiTarski is a prover for special function inequalities combining *resolution theorem proving* and *decision procedures for real algebra* (i.e., for the *theory of real closed fields* or **RCF**)

Hold it right there...

- We know Th(R,+,*,<,0,1) a.k.a. **RCF** is decidable.
- But is this *extended decision problem* solvable?
- No! Consider the following simple example, bearing in mind that Th(Q,+,*,<,0,1) is undecidable (AEA fragment: Julia Robinson; AE: Bjorn Poonen):

$$\forall r \in \mathbb{R}$$
$$r \in \mathbb{Q} \quad iff$$
$$\exists y, z \in \mathbb{R}(ry = z \ \wedge \ y \neq 0 \ \wedge$$
$$sin(y) = 0 \ \wedge \ sin(z) = 0).$$

MetiTarski is Incomplete

- As it works over an undecidable theory, MetiTarski is necessarily incomplete
- Thus, MetiTarski employs heuristic methods
- These heuristics are, however, *systematic* with a rather compelling story, as we'll see
- Despite this incompleteness, MetiTarski is remarkably powerful

Some Example MetiTarski Theorems

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

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

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

Each is proved in a few seconds!

In this talk

- We'll present some key improvements to MetiTarski's heuristic proof search
- These improvements centre around how MetiTarski makes use of an RCF decision procedure: RCF reasoning is often a bottleneck as decision procedures are hyper-exponential
- To understand these improvements, we must understand more about how MetiTarski works

MetiTarski at 30,000 Feet

GOAL: TO PROVE INEQUALITIES INVOLVING SIN, COS, LOG, EXP, ...

- Transcendental function occurrences can be replaced by *rational function* upper and lower bounds (e.g., using continued fraction expansions)

- Eventually, pure polynomial (`algebraic') inequality subproblems can be derived -- These can be handled by an RCF decision procedure

- All done systematically through extensions to a superposition calculus (and prover)

- Let's see in more detail...

Bounds for e^x

- Transcendental functions can be approximated by rational functions; these can yield families of upper and lower bounds
- E.g., via Taylor series or continued fractions
- Typically, several formulas are needed to cover a range of intervals. For example:

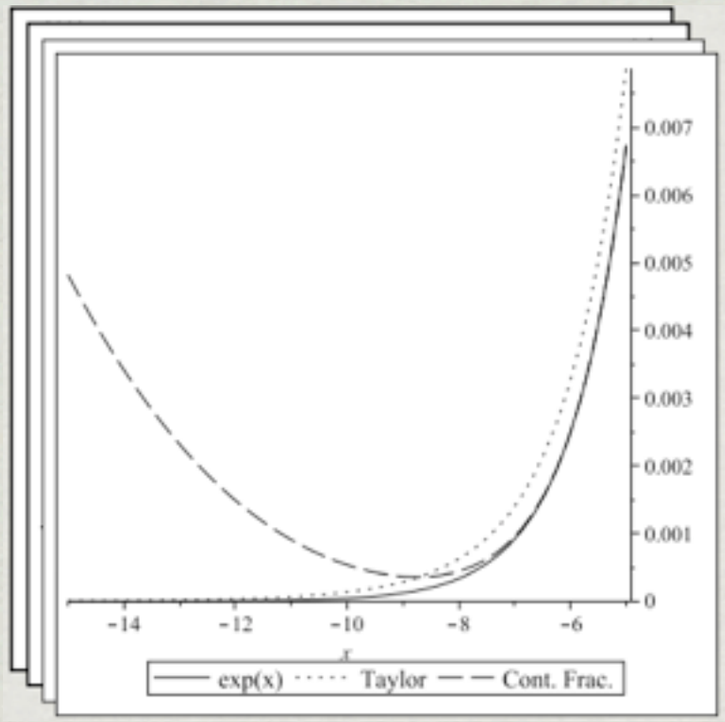
$$e^x \geq 1 + x + \ldots + x^n/n! \qquad (n \text{ odd})$$

$$e^x \leq 1 + x + \ldots + x^n/n! \qquad (n \text{ even}, x \leq 0)$$

$$e^x \leq 1/(1 - x + x^2/2! - x^3/3!) \qquad (x < 1.596)$$

Building up good families of transcendental function bounds

- ...takes *a lot* of work!
- A huge effort has gone into building up bounds which are useful to MetiTarski's applications.
- For this talk, let's just accept them as given.



CFE MUCH BETTER THAN TS HERE!

$$\frac{(1+19x+10x^2)(x-1)}{3x(3+6x+x^2)} \le \ln x \le \frac{(x^2+19x+10)(x-1)}{3(3x^2+6x+1)}$$

Resolution: A Primer

- MetiTarski produces proofs in an extended *superposition (i.e., `modern resolution')* calculus.
- Resolution provers work with *clauses:* disjunctions of *literals* (atoms or their negations).
- They seek to *contradict* the *negation* of the goal.
- Each step combines two clauses and yields new clauses, which are *simplified* and perhaps kept.
- If the *empty clause* is produced, we have found the desired contraction.

From Metis to MetiTarski

- MetiTarski extends a superposition calculus and prover (Metis) in many ways:
- algebraic literal deletion (using RCF procedure),
- algebraic redundancy checking (subsumption),
- formula normalisation and simplification,
- modified Knuth-Bendix ordering,
- case-splitting, `dividing out products,' ...

Algebraic Literal Deletion

- MetiTarski keeps a data-structure of all *ground*, *algebraic* clauses - an *algebraic context*
- Any literal inconsistent with the *algebraic context* can be **deleted**!
- This is one of the key uses of an ***RCF decision procedure:*** *to recognise when we may delete literals from derived clauses.*
- Deleting literals brings us closer to the empty clause!

Literal Deletion Example

Algebraic Context:

$ax^2 + bx + c = 0,$
$yz = 1,$
$z > 0.$

Clause:

L1 ∨
$y = 0$ ∨
$(b^2)z - 4acz < 0.$

Literal Deletion

Refined Clause:

L1

On RCF Decisions

- In searching for a proof of a transcendental function inequality...

MetiTarski generates a sequence of RCF subproblems (sometimes *tens of thousands*).

These subproblems are in the Existential fragment of RCF, *ExRCF.*

RCF decisions only contribute to a MetiTarski proof when they *refute* an ExRCF subproblem.

RCF is a bottleneck

- Though decidable, RCF is fundamentally infeasible
- RCF quantifier elimination is inherently doubly exponential (Davenport-Heintz)
- ExRCF has a theoretical exponential speed-up over RCF, but this hasn't been realised in practice
- Currently, best practical ExRCF methods are based on algebraic methods underlying full RCF QE (and are still doubly exponential in worst case)

Motivating Hypotheses

- By studying the structure of the sequences of RCF subproblems MetiTarski generates, we can devise specialised RCF proof methods which outperform ``off the shelf'' RCF proof methods on these sequences of RCF subproblems.

- By making use of these specialised RCF proof methods during MetiTarski's proof search, we can significantly improve MetiTarski's performance.

Main Contributions

- *Model sharing*: the use of past models for SAT ExRCF subproblems to satisfy subsequently encountered ones.
- The observation that *polynomial factorisation* is in practice a waste of time for MetiTarski's RCF subproblems; *disabling it* leads to serious gains. *(And this specialisation can't be done with some tools, e.g., Mathematica's Partial CAD!)*

Model sharing

- Let F_1, ..., F_k be the sequence of RCF subproblems generated by MetiTarski during its search for a proof of P.
- F_i only contributes to a MetiTarski proof when F_i is unsatisfiable over R^n,
- Many of the F_i share common subexpressions with each other.

Q: How often do F_i, F_{i+k} share a model?

A simple running example

$$\forall x \in (-8, 5) \quad \max(\sin(x), \sin(x+4), \cos(x)) > 0.$$

IN SEARCHING FOR A PROOF, METITARSKI WILL...

- make use of axioms for *sin*, *max* and *cos*,
- find a proof with **600 *steps***,
- when pretty-printed to a text-file at 75 columns per line, this **proof is 12,453 lines**.
- ...what about the RCF component?

A simple running example

$$\forall x \in (-8, 5) \quad \max(\sin(x), \sin(x+4), \cos(x)) > 0.$$

IN SEARCHING FOR A PROOF, RCF SUBPROBLEMS GENERATED...

- total number of RCF inferences used in proof: **62**
- total number of RCF subproblems generated: **2,776**
- of these, **2,221** are SAT, thus cannot contribute to MetiTarski's proof!
- max total deg: **24**; ave total deg: **3.53**; max coefficient bit-width: **103**; ave coefficient bit-width: **21.03**
- how much time was wasted on these SAT problems?

A simple running example

$$\forall x \in (-8, 5) \quad \max(\sin(x), \sin(x+4), \cos(x)) > 0.$$

IN SEARCHING FOR A PROOF, RCF SUBPROBLEMS GENERATED...

- **2,221** of **2,776** RCF subproblems are SAT
- Let's analyse them using Mathematica's **`Reduce[]`** command, a state-of-the-art RCF decision method.
- To decide all 2,776: **253.33 sec**
- To decide the 2,221 SAT ones: **185.28 sec**
- Thus, over **70%** of RCF time was spent on SAT RCF subproblems which can't contribute to MetiTarski's proof!

Such results are typical

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%

sin-3425b

$\forall x \in (0,\infty), y \in (-\infty,\infty)$

$$(x < y \wedge y^2 < 6) \Rightarrow \frac{\sin(y)}{\sin(x)} \leq 10^{-4} + \frac{y - \frac{1}{6}y^3 + \frac{1}{120}y^5}{x - \frac{1}{6}x^3 + \frac{1}{120}x^5}$$

exp-problem-9

$\forall x \in (0,\infty)$

$$\frac{1 - e^{-2x}}{2x(1 - e^{-x})^2} - \frac{1}{x^2} \leq \frac{1}{12}$$

log-fun-ineq-e-weak

$\forall x \in (0,12), y \in (-\infty,\infty)$

$$xy \leq \frac{1}{5} + x\ln(x) + e^{y-1}$$

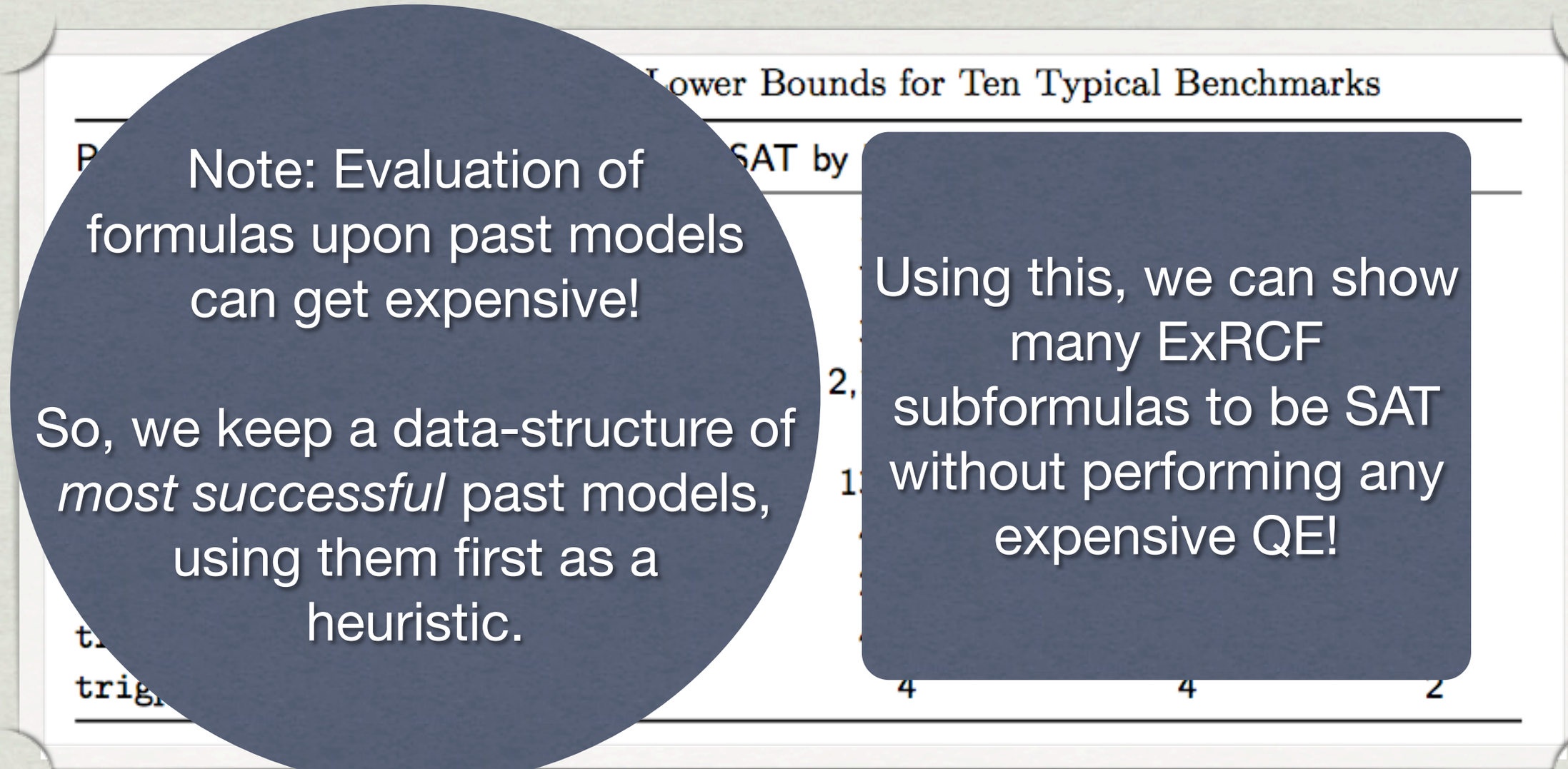
What about sharing models?

Table 3. Model Sharing Lower Bounds for Ten Typical Benchmarks

Problem	# SAT	# SAT by MS	# $\mathbb{Q}$ Models	# Successful
`CONVOI2-sincos`	194	168	9	7
`exp-problem-9`	731	720	11	7
`log-fun-ineq-e-weak`	323	305	24	18
`max-sin-2`	2,221	2,172	37	37
`sin-3425b`	72	64	8	6
`sqrt-problem-13-sqrt3`	1403	1350	26	21
`tan-1-1var-weak`	458	445	13	9
`trig-squared3`	549	280	15	11
`trig-squared4`	637	497	21	16
`trigpoly-3514-2`	934	4	4	2

In max-sin-2: 2,172 SAT using only 37 rational models!

What about sharing models?



In max-sin-2: 2,172 SAT using only 37 rational models!

Polynomial Irreducibility

Table 4. Factorisation in RCF Subproblems for Typical Univariate Benchmarks

Problem	# Factor	# Irreducible	% Runtime
asin-8-sqrt2	7791	5975 (76.7%)	22.4%
atan-problem-2-sqrt-weakest21	65304	63522 (97.3%)	55.4%
atan-problem-2-weakest21	9882	8552 (86.5%)	2.2%
cbrt-problem-5a	88986	61068 (68.6%)	38.6%
cbrt-problem-5b-weak	138861	25107 (18.0%)	53.1%
cos-3411-a-weak	150354	138592 (92.1%)	53.9%
ellipse-check-2-weak2	5236	3740 (71.4%)	88.7%
ellipse-check-3-ln	1724	1284 (74.4%)	86.7%
ellipse-check-3-weak	12722	9464 (74.3%)	77.9%

% RUNTIME FOR Z3'S NLSAT EXRCF DECISION METHOD

introducing Strategy 1

model sharing

+

omitting the standard test for *irreducibility*

= Strategy 1

introducing Strategy 1

model sharing

+

```
(and-then simplify purify-arith
          propagate-values elim-term-ite
          solve-eqs tseitin-cnf simplify
          (using-params nlsat
                        :factor false
                        :algebraic-min-mag 256))
```

= Strategy 1

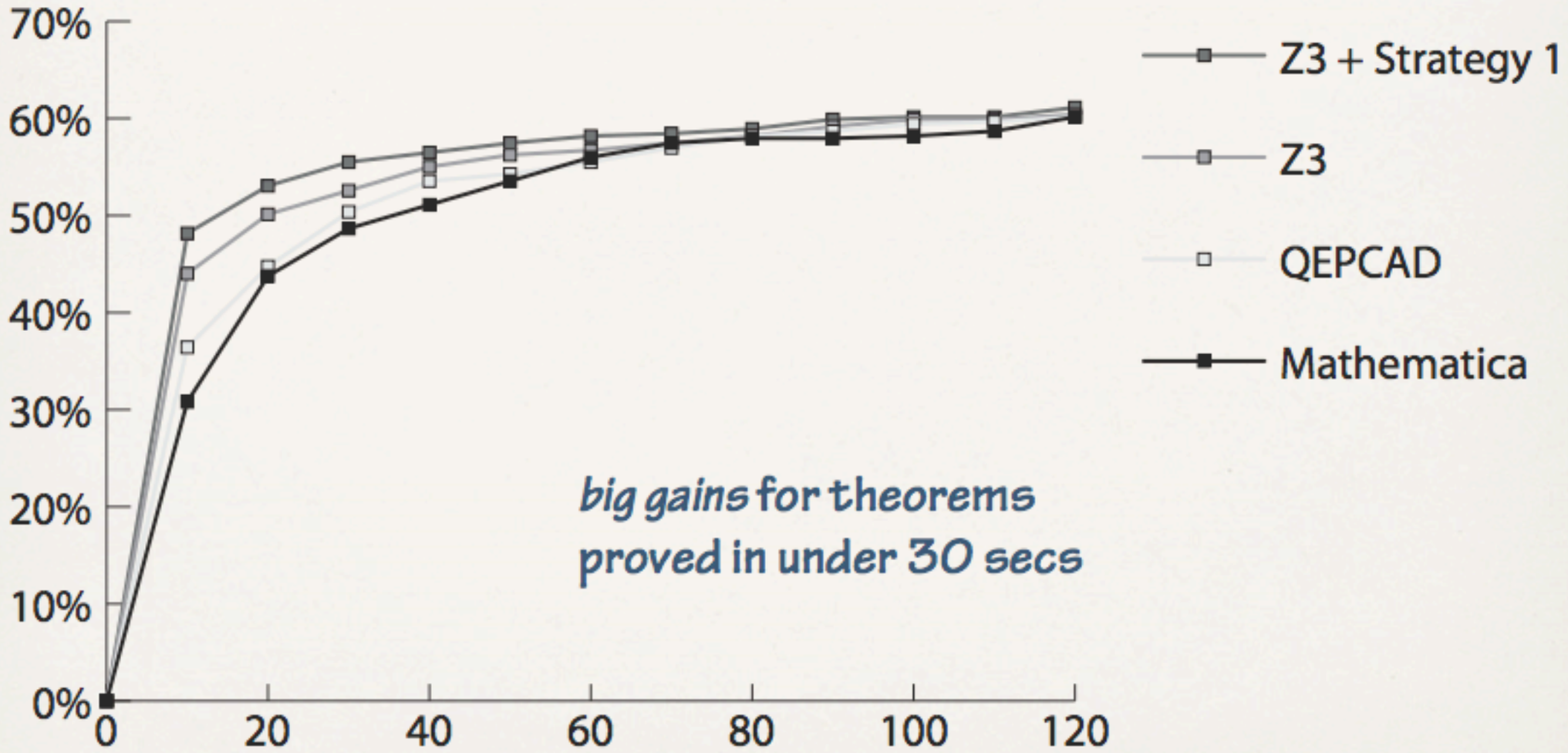
Introducing the ExRCF solvers

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

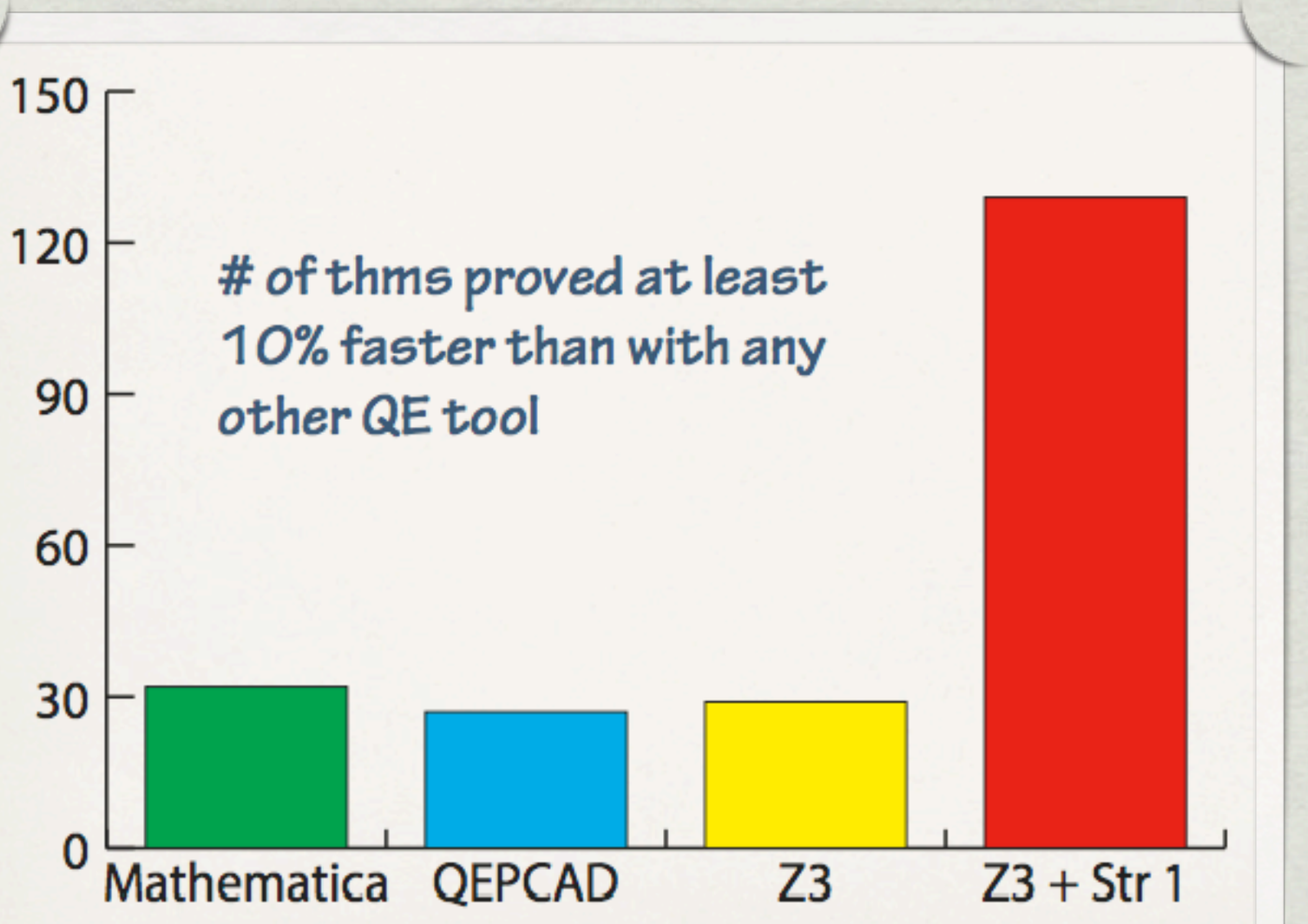
Mathematica (Wolfram research)
Much faster than QEPCAD for 3–4 variables

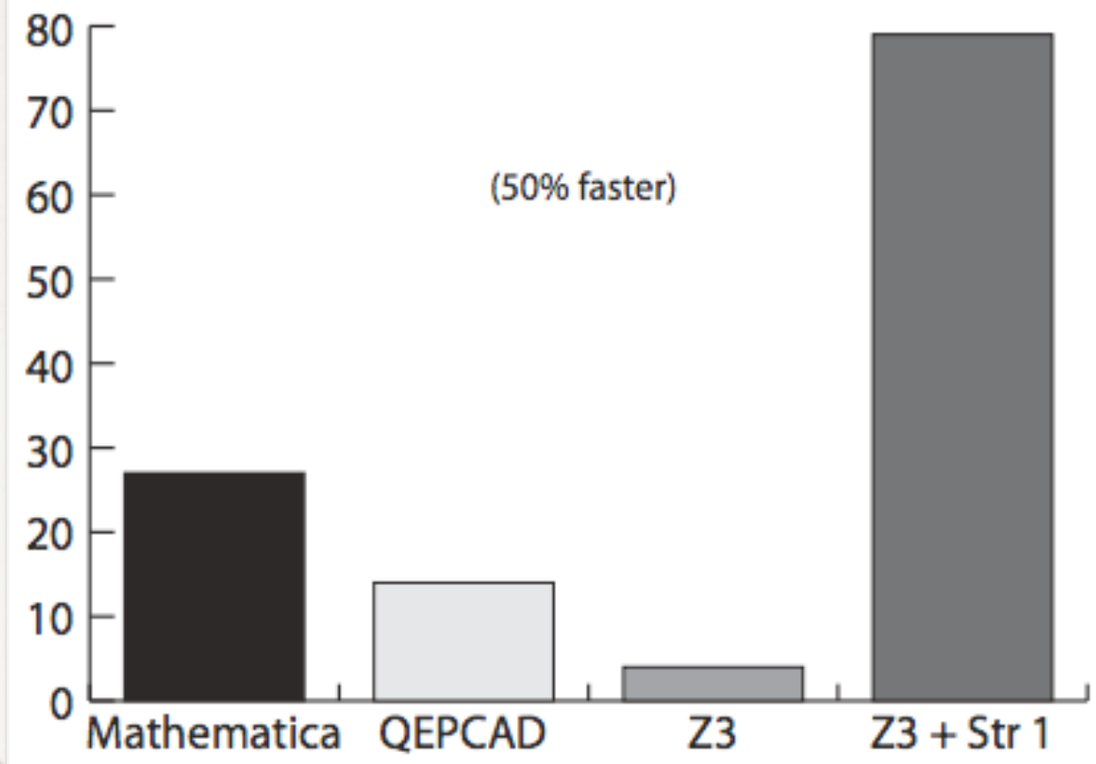
Z3 (de Moura, Microsoft Research)
An SMT solver with non-linear reasoning.

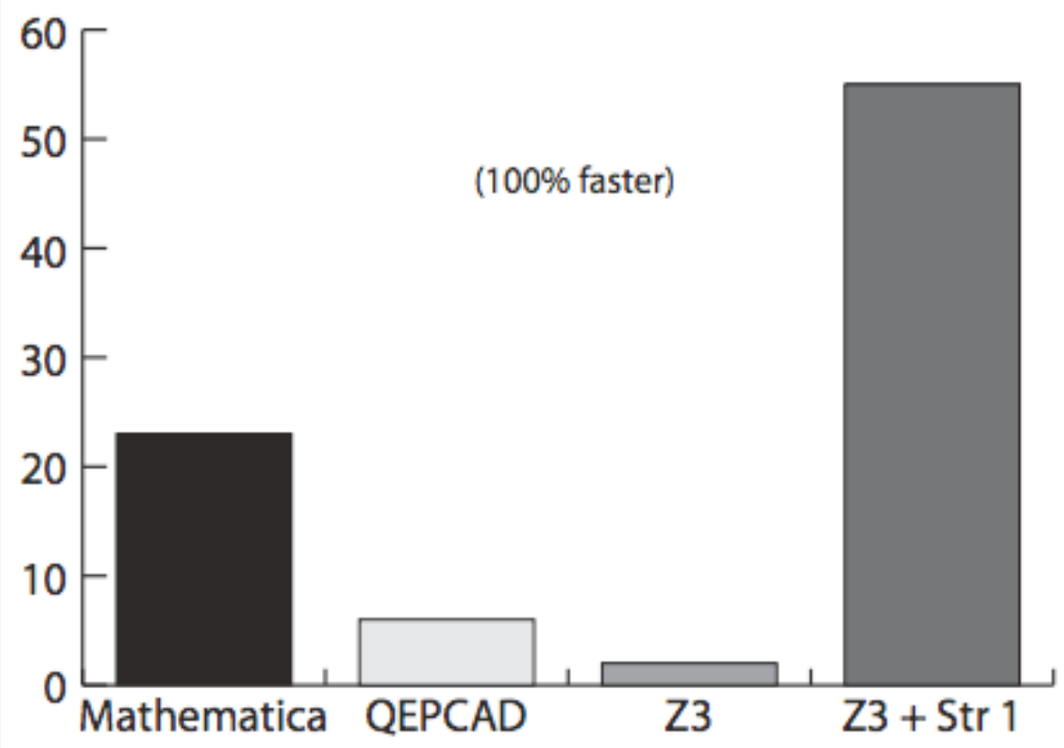
comparative results
(% proved up to 120sec)



Strategy 1 finds the fastest proofs




80
70
60
50
40
30
20
10
0
(50% faster)
Mathematica
QEPCAD
Z3
Z3 + Str 1


60
50
40
30
20
10
0
(100% faster)
Mathematica
QEPCAD
Z3
Z3 + Str 1

Conclusion

- By studying the structure of the ExRCF subproblems generated, we can devise specialised variants which vastly improve our results
- Expensive decision procedures shouldn’t be seen as only `black boxes,’ but should be specialised
- Authors of decision method tools should make it easy for users to specialise their procedures in this way (Z3 does so using a new *strategy language*)