Verifying Automated Reasoning Results

Marijn J.H. Heule

Carnegie Mellon University

http://www.cs.cmu.edu/~mheule/15816-f24/

https://github.com/marijnheule/proof-demo

Automated Reasoning and Satisfiability, October 2, 2024

Outline

Introduction

Proof Checking

Proof Systems and Formats

Certified Checking

Applications

Conclusions

Introduction

Proof Checking

Proof Systems and Formats

Certified Checking

Applications

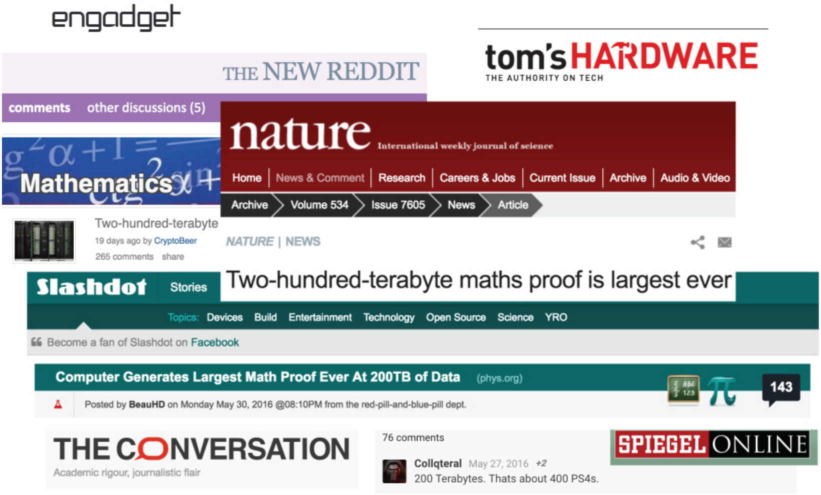
Conclusions

Motivation for Validating Proofs of Unsatisfiability

SAT solvers may have errors and only return yes/no.

- Documented bugs in SAT, SMT, and QSAT solvers; [Brummayer and Biere, 2009; Brummayer et al., 2010]
- Competition winners have contradictory results (HWMCC winners from 2011 and 2012)
- Implementation errors often imply conceptual errors;
- Proofs now mandatory for the annual SAT Competitions;
- Mathematical results require a stronger justification than a simple yes/no by a solver. UNSAT must be verifiable.

"The Largest Math Proof Ever"



engadget

tom's HARDWARE
THE AUTHORITY ON TECH

THE NEW REDDIT

comments other discussions (5)

Mathematics

Two-hundred-terabyte

19 days ago by CryptoBeer

265 comments share

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video

Archive > Volume 534 > Issue 7605 > News > Article

NATURE | NEWS

Two-hundred-terabyte maths proof is largest ever

Slashdot Stories

Topics: Devices Build Entertainment Technology Open Source Science YRO

Become a fan of Slashdot on Facebook

Computer Generates Largest Math Proof Ever At 200TB of Data (phys.org)

143

Posted by **BeauHD** on Monday May 30, 2016 @08:10PM from the red-pill-and-blue-pill dept.

THE CONVERSATION
Academic rigour, journalistic flair

76 comments

Collqteral May 27, 2016 +2
200 Terabytes. Thats about 400 PS4s.

SPIEGEL ONLINE

Demo: Validating Results

```
git clone https://github.com/marijnheule/proof-demo
```

Introduction

Proof Checking

Proof Systems and Formats

Certified Checking

Applications

Conclusions

Resolution Rule and Resolution Chains

Resolution Rule

$$\frac{C \lor x \qquad \overline{x} \lor D}{C \lor D}$$

- Or equivalently: $C \lor D := (C \lor x) \bowtie (\overline{x} \lor D)$
- Many SAT techniques can be simulated by resolution.

Resolution Rule and Resolution Chains

Resolution Rule

$$\frac{C \vee x \qquad \overline{x} \vee D}{C \vee D}$$

- Or equivalently: $C \vee D := (C \vee x) \bowtie (\overline{x} \vee D)$
- Many SAT techniques can be simulated by resolution.

A resolution chain is a sequence of resolution steps.
The resolution steps are performed from left to right.

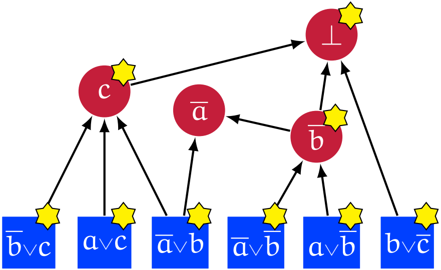
Example

- $(c) := (\overline{a} \vee \overline{b} \vee c) \bowtie (\overline{a} \vee b) \bowtie (a \vee c)$
- $(\overline{a} \vee c) := (\overline{a} \vee b) \bowtie (a \vee c) \bowtie (\overline{a} \vee \overline{b} \vee c)$
- The order of the clauses in the chain matter

Resolution Proofs versus Clausal Proofs

Consider $\Gamma := (\overline{b} \vee c) \wedge (a \vee c) \wedge (\overline{a} \vee b) \wedge (\overline{a} \vee \overline{b}) \wedge (a \vee \overline{b}) \wedge (b \vee \overline{c})$

A resolution graph of Γ is:



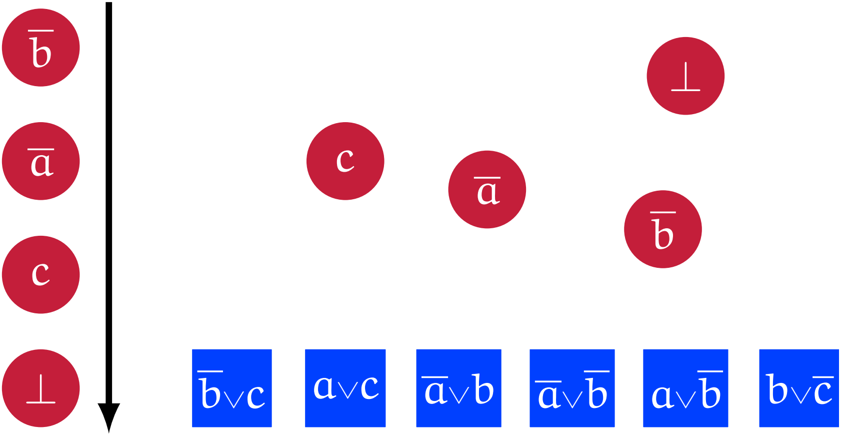
A resolution proof consists of all nodes and edges of the resolution graph

- Graphs from SAT solvers have ~ 400 incoming edges per node
- Resolution proof logging can heavily increase memory usage $(\times 100)$

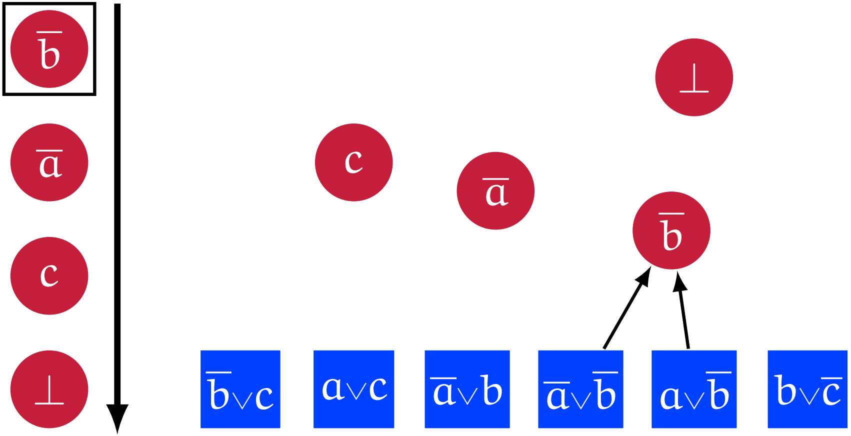
A clausal proof is a list of all nodes sorted by topological order

- Clausal proofs are easy to emit and relatively small
- Clausal proof checking requires to reconstruct the edges (costly)

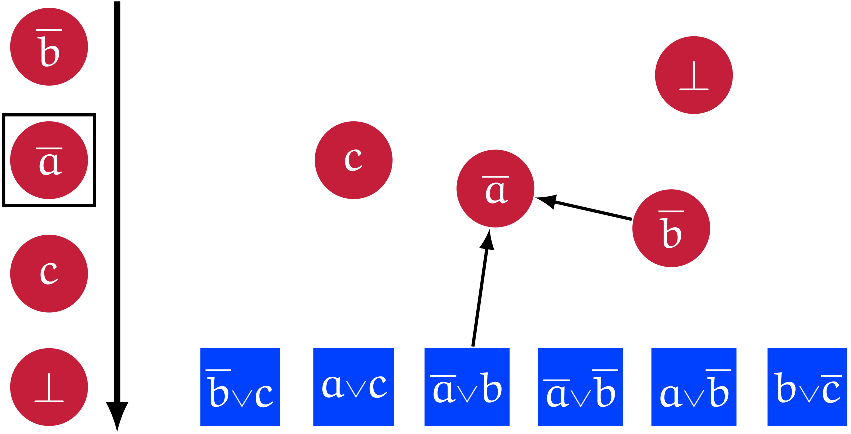
Clausal Proof: Checker has to reconstruct resolution edges



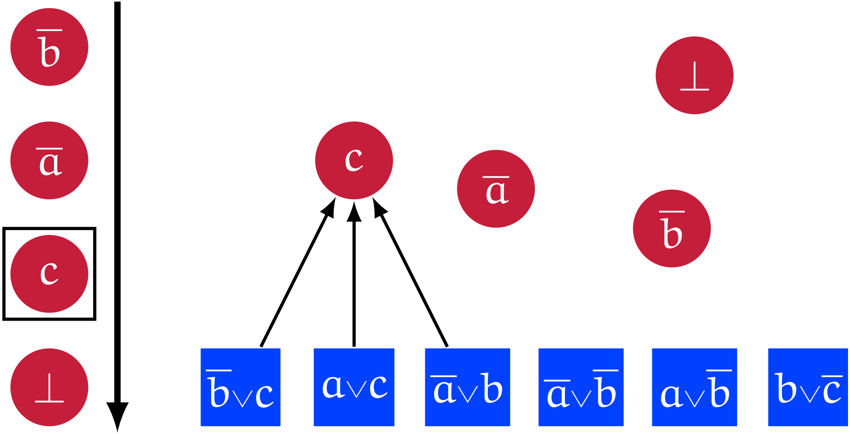
Clausal Proof: Checker has to reconstruct resolution edges



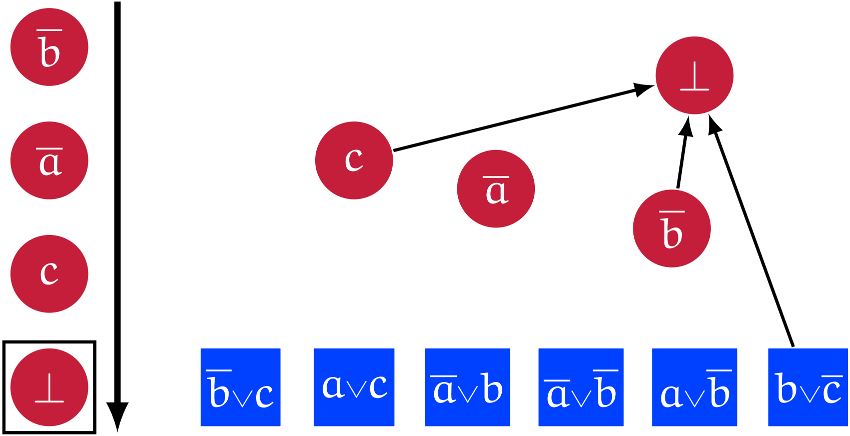
Clausal Proof: Checker has to reconstruct resolution edges



Clausal Proof: Checker has to reconstruct resolution edges



Clausal Proof: Checker has to reconstruct resolution edges



Reconstruct Edges Efficiently: Reverse Unit Propagation

- Unit propagation (UP) satisfies unit clauses by assigning their literal to true (until fixpoint or a conflict).
- Let F be a formula. A clause C is implied by F via UP (denoted by $F \vdash_1 C$) if UP on $F \land \neg C$ results in a conflict.

Example

$$\begin{aligned}\Gamma \;=\; & (a \lor b \lor \overline{c}) \land (\overline{a} \lor \overline{b} \lor c) \land (b \lor c \lor \overline{d}) \land (\overline{b} \lor \overline{c} \lor d) \;\land \\ & (a \lor c \lor d) \land (\overline{a} \lor \overline{c} \lor \overline{d}) \land (\overline{a} \lor b \lor d) \land (a \lor \overline{b} \lor \overline{d})\end{aligned}$$

Reconstruct Edges Efficiently: Reverse Unit Propagation

- Unit propagation (UP) satisfies unit clauses by assigning their literal to true (until fixpoint or a conflict).
- Let F be a formula. A clause C is implied by F via UP (denoted by $F \vdash_1 C$) if UP on $F \land \neg C$ results in a conflict.

Example

$$\begin{aligned}\Gamma = {} & (a \lor b \lor \bar{c}) \land (\bar{a} \lor \bar{b} \lor c) \land (b \lor c \lor \bar{d}) \land (\bar{b} \lor \bar{c} \lor d) \land \\ & (a \lor c \lor d) \land (\bar{a} \lor \bar{c} \lor \bar{d}) \land (\bar{a} \lor b \lor d) \land (a \lor \bar{b} \lor \bar{d})\end{aligned}$$

clause	$(a \lor b)$
units	$\bar{a} \land \bar{b}$

Reconstruct Edges Efficiently: Reverse Unit Propagation

- Unit propagation (UP) satisfies unit clauses by assigning their literal to true (until fixpoint or a conflict).
- Let F be a formula. A clause C is implied by F via UP (denoted by $F \vdash_1 C$) if UP on $F \wedge \neg C$ results in a conflict.

Example

$$\begin{aligned}\Gamma = {} & (a \vee b \vee \overline{c}) \wedge (\overline{a} \vee \overline{b} \vee c) \wedge (b \vee c \vee \overline{d}) \wedge (\overline{b} \vee \overline{c} \vee d) \wedge \\ & (a \vee c \vee d) \wedge (\overline{a} \vee \overline{c} \vee \overline{d}) \wedge (\overline{a} \vee b \vee d) \wedge (a \vee \overline{b} \vee \overline{d})\end{aligned}$$

clause	$(a \vee b)$	$(a \vee b \vee \overline{c})$
units	$\overline{a} \wedge \overline{b}$	$\overline{c}$

Reconstruct Edges Efficiently: Reverse Unit Propagation

- Unit propagation (UP) satisfies unit clauses by assigning their literal to true (until fixpoint or a conflict).
- Let F be a formula. A clause C is implied by F via UP (denoted by $F \vdash_1 C$) if UP on $F \land \neg C$ results in a conflict.

Example

$$\begin{aligned}\Gamma = {} & (a \lor b \lor \bar{c}) \land (\bar{a} \lor \bar{b} \lor c) \land (b \lor c \lor \bar{d}) \land (\bar{b} \lor \bar{c} \lor d) \land \\ & (a \lor c \lor d) \land (\bar{a} \lor \bar{c} \lor \bar{d}) \land (\bar{a} \lor b \lor d) \land (a \lor \bar{b} \lor \bar{d})\end{aligned}$$

clause	$(a \lor b)$	$(a \lor b \lor \bar{c})$	$(b \lor c \lor \bar{d})$
units	$\bar{a} \land \bar{b}$	$\bar{c}$	$\bar{d}$

Reconstruct Edges Efficiently: Reverse Unit Propagation

- Unit propagation (UP) satisfies unit clauses by assigning their literal to true (until fixpoint or a conflict).
- Let F be a formula. A clause C is implied by F via UP (denoted by $F \vdash_1 C$) if UP on $F \land \neg C$ results in a conflict.

Example

$$\begin{aligned}\Gamma = {} & (a \lor b \lor \bar{c}) \land (\bar{a} \lor \bar{b} \lor c) \land (b \lor c \lor \bar{d}) \land (\bar{b} \lor \bar{c} \lor d) \land \\ & (a \lor c \lor d) \land (\bar{a} \lor \bar{c} \lor \bar{d}) \land (\bar{a} \lor b \lor d) \land (a \lor \bar{b} \lor \bar{d})\end{aligned}$$

clause	$(a \lor b)$	$(a \lor b \lor \bar{c})$	$(b \lor c \lor \bar{d})$	$(a \lor c \lor d)$
units	$\bar{a} \land \bar{b}$	$\bar{c}$	$\bar{d}$	$\perp$

Reconstruct Edges Efficiently: Reverse Unit Propagation

- Unit propagation (UP) satisfies unit clauses by assigning their literal to true (until fixpoint or a conflict).
- Let F be a formula. A clause C is implied by F via UP (denoted by $F \vdash_1 C$) if UP on $F \land \neg C$ results in a conflict.

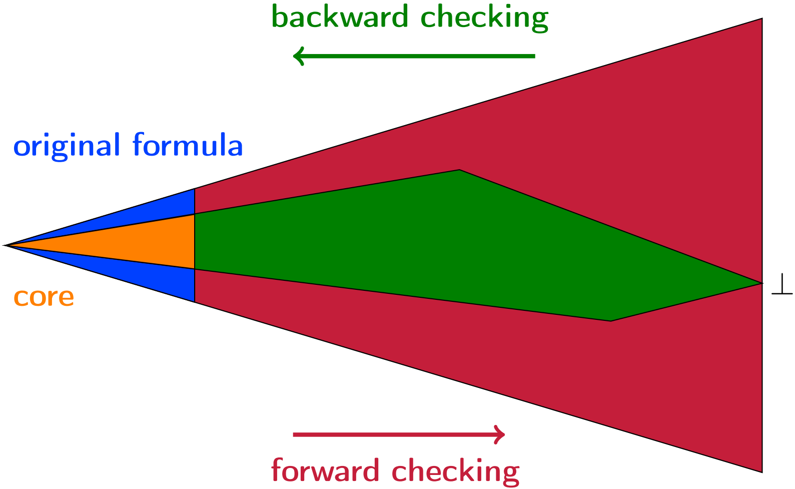
Example

$$\begin{aligned}\Gamma \;=\; & (a \lor b \lor \bar{c}) \land (\bar{a} \lor \bar{b} \lor c) \land (b \lor c \lor \bar{d}) \land (\bar{b} \lor \bar{c} \lor d) \;\land \\ & (a \lor c \lor d) \land (\bar{a} \lor \bar{c} \lor \bar{d}) \land (\bar{a} \lor b \lor d) \land (a \lor \bar{b} \lor \bar{d})\end{aligned}$$

clause	$(a \lor b)$	$(a \lor b \lor \bar{c})$	$(b \lor c \lor \bar{d})$	$(a \lor c \lor d)$
units	$\bar{a} \land \bar{b}$	$\bar{c}$	$\bar{d}$	$\bot$

$$\dfrac{\dfrac{(a \lor c \lor d) \qquad (b \lor c \lor \bar{d})}{(a \lor b \lor c)} \qquad (a \lor b \lor \bar{c})}{(a \lor b)}$$

Forward vs Backward Proof Checking



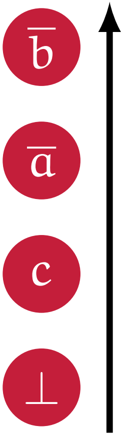
Improvement I: Backwards Checking

Goldberg and Novikov proposed checking the refutation backwards [DATE 2003]:

- start by validating the empty clause;
- mark all lemmas using conflict analysis;
- only validate marked lemmas.

Advantage: validate fewer lemmas.

Disadvantage: more complex.



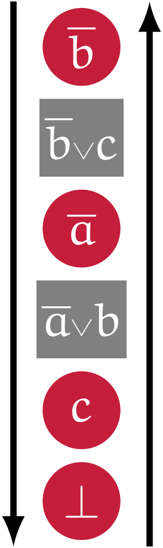
Improvement II: Clause Deletion

We proposed to extend clausal proofs with deletion information [STVR 2014]:

- clause deletion is crucial for efficient solving;
- emit learning and deletion information;
- proof size might double;
- checking speed can be reduced significantly.

Clause deletion can be combined with backwards checking [FMCAD 2013]:

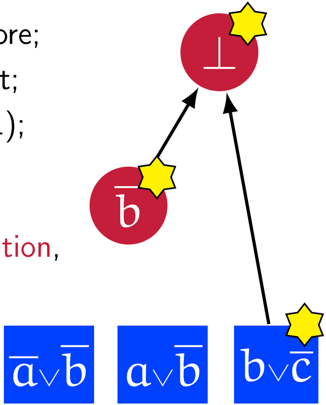
- ignore deleted clauses earlier in the proof;
- optimize clause deletion for trimmed proofs.



Improvement III: Core-first Unit Propagation

We propose a new unit propagation variant:

1. propagate using clauses already in the core;
2. examine non-core clauses only at fixpoint;
3. if a non-core unit clause is found, goto 1);
4. otherwise terminate.

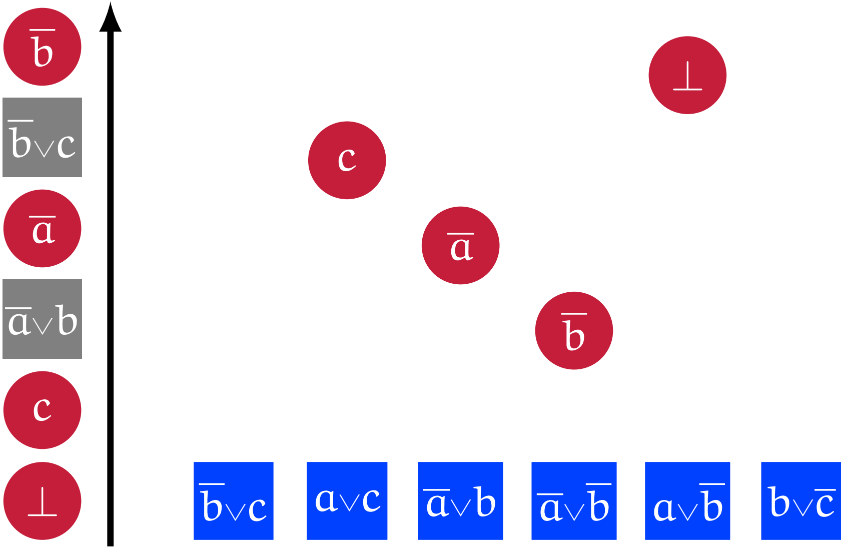


The variant, called Core-first Unit Propagation, can reduce checking costs considerably.

Fast propagation in a checker is different than fast propagation in a SAT solver.

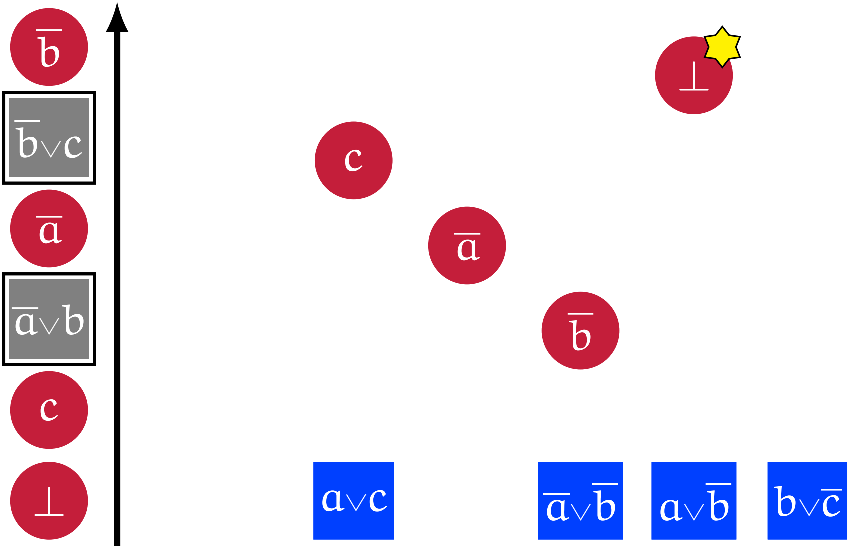
Also, the resulting core and proof are smaller

Checking: Backwards + Core-first + Deletion



Core-first unit propagation results in smaller cores and proofs

Checking: Backwards + Core-first + Deletion



$\overline{b}$

$\overline{b} \vee c$

$\overline{a}$

$\overline{a} \vee b$

c

$\bot$

$\bot$

c

$\overline{a}$

$\overline{b}$

$a \vee c$

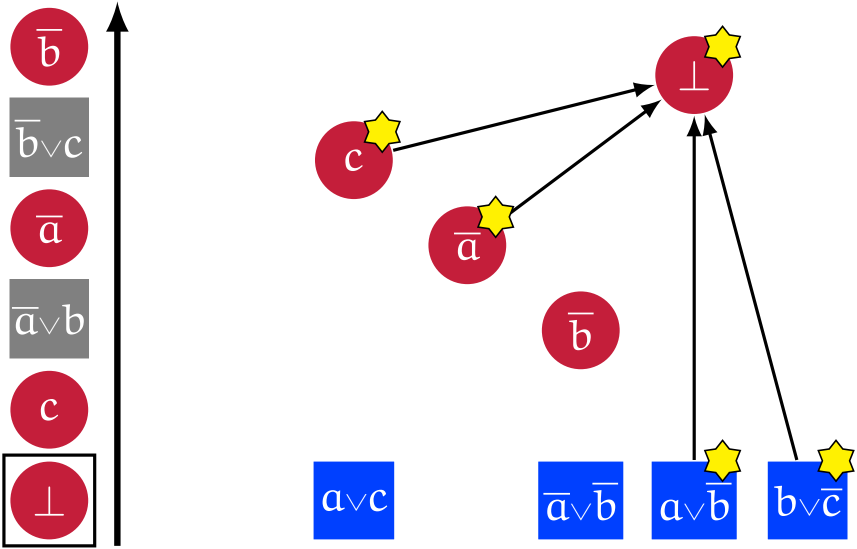
$\overline{a} \vee \overline{b}$

$a \vee \overline{b}$

$b \vee \overline{c}$

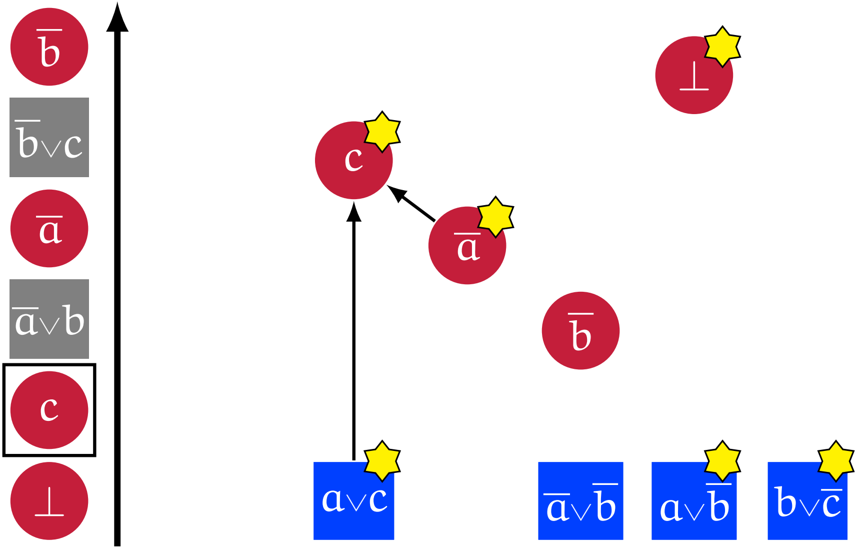
Core-first unit propagation results in smaller cores and proofs

Checking: Backwards + Core-first + Deletion



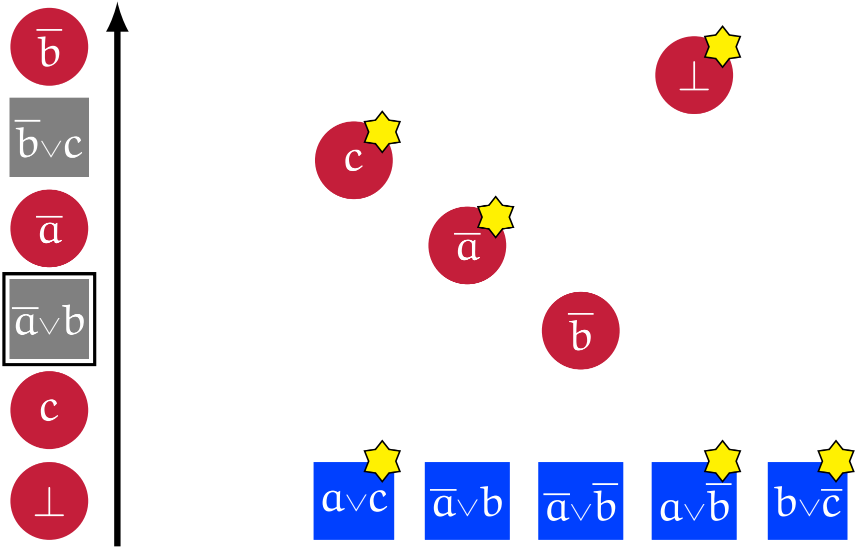
Core-first unit propagation results in smaller cores and proofs

Checking: Backwards + Core-first + Deletion



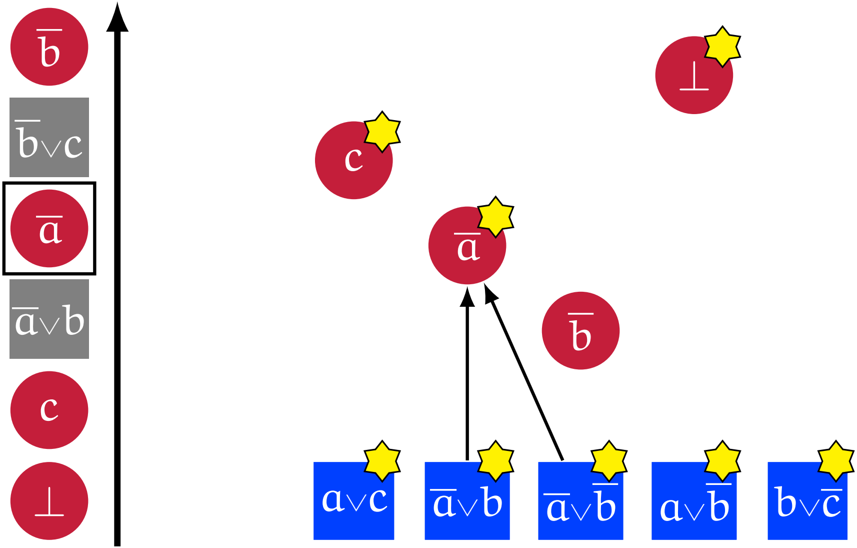
Core-first unit propagation results in smaller cores and proofs

Checking: Backwards + Core-first + Deletion



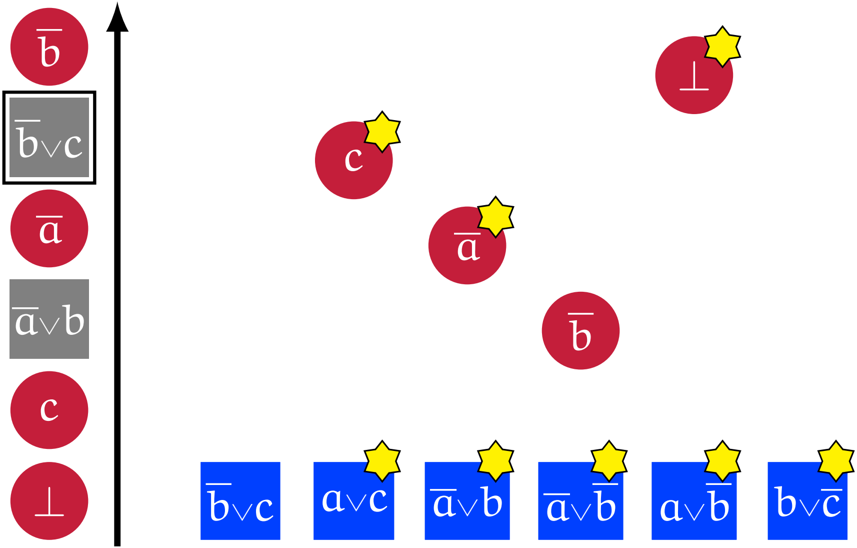
Core-first unit propagation results in smaller cores and proofs

Checking: Backwards + Core-first + Deletion



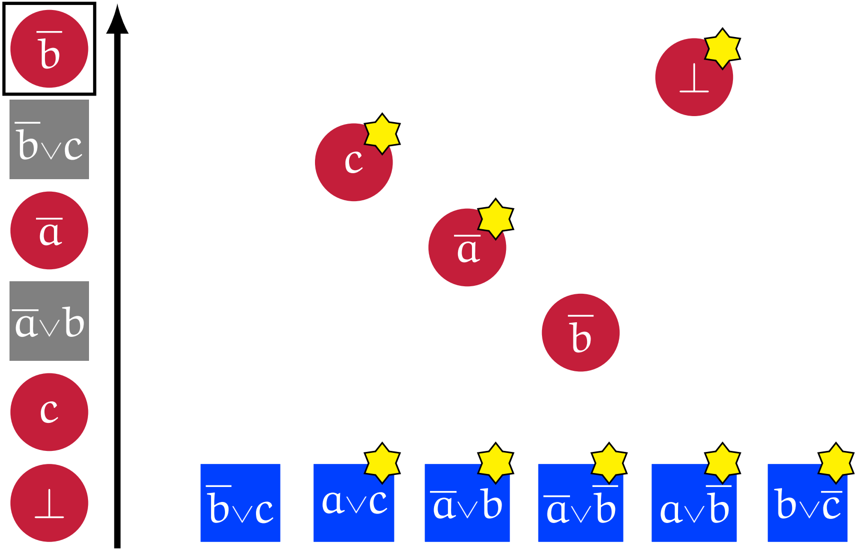
Core-first unit propagation results in smaller cores and proofs

Checking: Backwards + Core-first + Deletion



Core-first unit propagation results in smaller cores and proofs

Checking: Backwards + Core-first + Deletion



Core-first unit propagation results in smaller cores and proofs

DRAT (Deletion Resolution Asymmetric Tautology)

Drawbacks of resolution:

- For many seemingly simple formulas, there are only resolution proofs of exponential size.
- State-of-the-art solving techniques are not succinctly expressible.

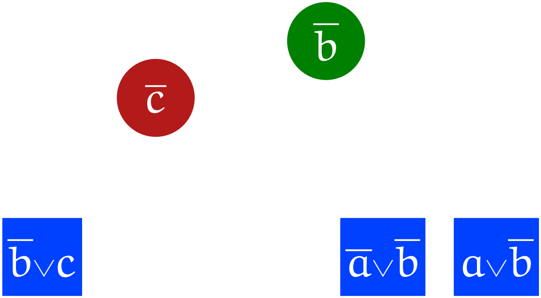
A clause $(C \lor x)$ is a resolution asymmetric tautology (RAT) on x w.r.t. a CNF formula Γ if for every clause $(D \lor \overline{x}) \in \Gamma$, the resolvent $C \lor D$ is implied by Γ via unit-propagation, i.e., $\Gamma \vdash_1 C \lor D$.

Popular example of a clausal proof system: DRAT

- DRAT allows the addition of RATs to a formula.
  - RATs are not necessarily implied by the formula.
  - But RATs are redundant: their addition preserves satisfiability.
  - Clause deletion may introduce clause addition options (interference)

DRAT Example

A clause $(C \vee x)$ is a resolution asymmetric tautology (RAT) on x w.r.t. a CNF formula Γ if for every clause $(D \vee \overline{x}) \in \Gamma$, the resolvent $C \vee D$ is implied by Γ via unit-propagation, i.e., $\Gamma \vdash_1 C \vee D$.



$\overline{b}$

$\overline{c}$

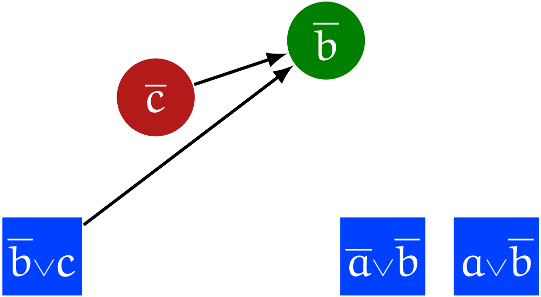
$\overline{b} \vee c$

$\overline{a} \vee \overline{b}$

$a \vee \overline{b}$

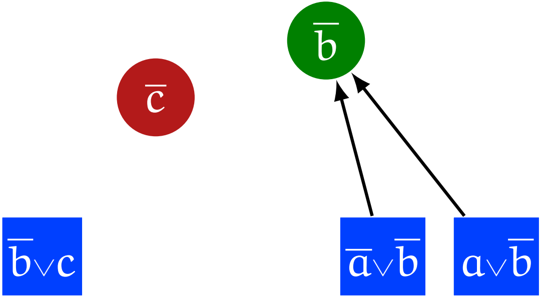
DRAT Example

A clause $(C \vee x)$ is a resolution asymmetric tautology (RAT) on x w.r.t. a CNF formula Γ if for every clause $(D \vee \overline{x}) \in \Gamma$, the resolvent $C \vee D$ is implied by Γ via unit-propagation, i.e., $\Gamma \vdash_1 C \vee D$.



DRAT Example

A clause $(C \vee x)$ is a resolution asymmetric tautology (RAT) on x w.r.t. a CNF formula Γ if for every clause $(D \vee \overline{x}) \in \Gamma$, the resolvent $C \vee D$ is implied by Γ via unit-propagation, i.e., $\Gamma \vdash_1 C \vee D$.



Demo: DRAT step

```
git clone https://github.com/marijnheule/proof-demo
```

Introduction

Proof Checking

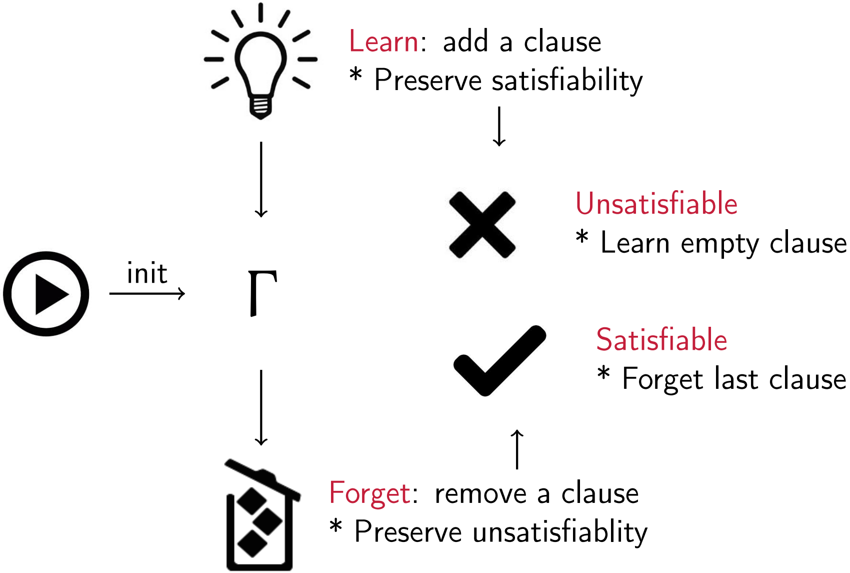
Proof Systems and Formats

Certified Checking

Applications

Conclusions

Clausal Proof System [Järvisalo, Heule, and Biere 2012]



init

Γ

Learn: add a clause
* Preserve satisfiability

Unsatisfiable
* Learn empty clause

Satisfiable
* Forget last clause

Forget: remove a clause
* Preserve unsatisfiablity

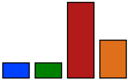
Ideal Properties of a Proof System for SAT Solvers

Easy to Emit

Compact

Checked Efficiently

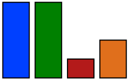
Expressive



Resolution Proofs

Zhang and Malik, 2003

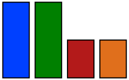
Van Gelder, 2008; Biere, 2008



Clausal Proofs

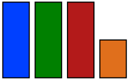
Goldberg and Novikov, 2003

Van Gelder, 2008



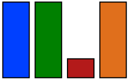
Clausal proofs + deletion

Heule, Hunt, Jr., Wetzler [STVR'14]



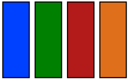
Optimized clausal proof checker

Heule, Hunt, Jr., and Wetzler [FMCAD'13]



Clausal RAT proofs

Heule, Hunt, Jr., Wetzler [CADE'13]



DRAT proofs (RAT + deletion)

Wetzler, Heule, Hunt, Jr. [SAT'14]

Ideal Properties of a Proof System for SAT Solvers

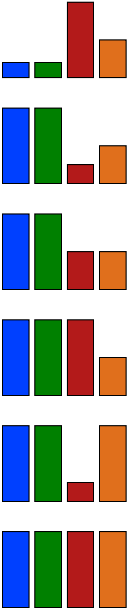
Easy to Emit

Compact

Checked Efficiently

Expressive

Verified



Resolution Proofs
Zhang and Malik, 2003
Van Gelder, 2008; Biere, 2008

Clausal Proofs
Goldberg and Novikov, 2003
Van Gelder, 2008

Clausal proofs + deletion
Heule, Hunt, Jr., Wetzler [STVR'14]

Optimized clausal proof checker
Heule, Hunt, Jr., and Wetzler [FMCAD'13]

Clausal RAT proofs
Heule, Hunt, Jr., Wetzler [CADE'13]

DRAT proofs (RAT + deletion)
Wetzler, Heule, Hunt, Jr. [SAT'14]

Proof Formats: The Input Format DIMACS

$$E := (\overline{b} \vee c) \wedge (a \vee c) \wedge (\overline{a} \vee b) \wedge (\overline{a} \vee \overline{b}) \wedge (a \vee \overline{b}) \wedge (b \vee \overline{c})$$

The input format of SAT solvers is known as DIMACS

- header starts with `p cnf` followed by the number of variables (n) and the number of clauses (m)
- the next m lines represent the clauses
- positive literals are positive numbers
- negative literals are negative numbers
- clauses are terminated with a 0

```
p cnf 3 6
-2  3 0
 1  3 0
-1  2 0
-1 -2 0
 1 -2 0
 2 -3 0
```

Most proof formats use a similar syntax.

Proof Formats: Proofs with Hints

LRAT is the most popular resolution-style format (RAT also supported)

$$E := (\overline{b} \vee c) \wedge (a \vee c) \wedge (\overline{a} \vee b) \wedge (\overline{a} \vee \overline{b}) \wedge (a \vee \overline{b}) \wedge (b \vee \overline{c})$$

LRAT orders clauses by unit propagation, ending with falsified clause

- formula implicit, index by input order

$$\begin{array}{rcl}
\langle\text{trace}\rangle & = & \{\langle\text{clause}\rangle\} \\
\langle\text{clause}\rangle & = & \langle\text{pos}\rangle\langle\text{literals}\rangle\langle\text{clsidx}\rangle \\
\langle\text{literals}\rangle & = & \{\langle\text{lit}\rangle\}\ \text{"0"} \\
\langle\text{clsidx}\rangle & = & \{\langle\text{pos}\rangle\}\ \text{"0"} \\
\langle\text{lit}\rangle & = & \langle\text{pos}\rangle \mid \langle\text{neg}\rangle \\
\langle\text{pos}\rangle & = & \text{"1"} \mid \text{"2"} \mid \cdots \mid \langle\text{maxidx}\rangle \\
\langle\text{neg}\rangle & = & \text{"}-\text{"}\langle\text{pos}\rangle
\end{array}$$

```
1 -2  3 0 0
2  1  3 0 0
3 -1  2 0 0
4 -1 -2 0 0
5  1 -2 0 0
6  2 -3 0 0
```

```
7 -2  0 4 5 0
8  3  0 3 2 1 0
9  0  8 7 6 0
```

Proof Formats: Proofs with Hints Example

LRAT is the most popular resolution-style format (RAT also supported)

$$E := (\overline{b} \lor c) \land (a \lor c) \land (\overline{a} \lor b) \land (\overline{a} \lor \overline{b}) \land (a \lor \overline{b}) \land (b \lor \overline{c})$$

LRAT orders clauses by unit propagation, ending with falsified clause

- formula implicit, index by input order

The clauses 1 to 6 are input clauses

Clause 7 is the resolvent of 5 and 4:

- $(\overline{b}) := (a \lor \overline{b}) \bowtie (\overline{a} \lor \overline{b})$

Clause 8 is the resolvent of 1, 2 and 3:

- $(c) := (\overline{b} \lor c) \bowtie (\overline{a} \lor b) \bowtie (a \lor c)$

Clause 9 is the resolvent of 6, 7 and 8:

- $\bot := (b \lor \overline{c}) \bowtie (\overline{b}) \bowtie (c)$

```
1 -2  3 0 0
2  1  3 0 0
3 -1  2 0 0
4 -1 -2 0 0
5  1 -2 0 0
6  2 -3 0 0
```

```
7 -2  0 4 5 0
8  3  0 3 2 1 0
9  0  8 7 6 0
```

Proof Formats: Clausal Proofs

RUP and extensions is the most popular clausal-style format.

$$E := (\overline{b} \vee c) \wedge (a \vee c) \wedge (\overline{a} \vee b) \wedge (\overline{a} \vee \overline{b}) \wedge (a \vee \overline{b}) \wedge (b \vee \overline{c})$$

RUP is much more compact than LRAT because it lacks hints

- formula not included as well

$$
\begin{aligned}
\langle\text{proof}\rangle &= \{\langle\text{lemma}\rangle\} \\
\langle\text{lemma}\rangle &= \langle\text{delete}\rangle\{\langle\text{lit}\rangle\}\ \texttt{"0"} \\
\langle\text{delete}\rangle &= \texttt{""} \mid \texttt{"d"} \\
\langle\text{lit}\rangle &= \langle\text{pos}\rangle \mid \langle\text{neg}\rangle \\
\langle\text{pos}\rangle &= \texttt{"1"} \mid \texttt{"2"} \mid \cdots \mid \langle\text{maxidx}\rangle \\
\langle\text{neg}\rangle &= \texttt{"-"}\langle\text{pos}\rangle
\end{aligned}
$$

```
-2  0
 3  0
 0
```

$$E \wedge (b) \vdash_1 \bot$$

$$E \wedge (\overline{b}) \wedge (\overline{c}) \vdash_1 \bot$$

$$E \wedge (\overline{b}) \wedge (c) \vdash_1 \bot$$

Proof Formats: Binary Formats

There are various cheap compression techniques to shrink proofs:

- Use 4 bytes per literal instead storing the ascii characters
- Sort literals in clauses and store the delta between literals
- Use a variable byte encoding for literals

encoding	**example** (prefix pivot $\text{lit}_1 \ldots \text{lit}_{k-1}$ end)	**#bytes**
ascii	`d 6278 -3425 -42311 9173 22754 0\n`	33
sascii	`d 6278 -3425 9173 22754 -42311 0\n`	33
4byte	`64 0c310000 c31a0000 8f4a0100 aa470000 c4b10000 00000000`	25
s4byte	`64 0c310000 c31a0000 aa470000 c4b10000 8f4a0100 00000000`	25
ds4byte	`64 0c310000 c31a0000 e82c0000 1a6a0000 cb980000 00000000`	25
vbyte	`64 8c62c335 8f9505aa 8f01c4e3 0200`	15
svbyte	`64 8c62c335 aa8f01c4 e3028f95 0500`	15
dsvbyte	`64 8c62c335 e8599ad4 01cbb102 00`	14

Proof Formats: Beyond Checking

Clausal Proof checkers can produce many additional results:

- Clausal core, e.g. useful for MUS computation, MaxSAT
  DRAT-trim option: `-c CORE`
- Extract a resolution proof, e.g. useful for interpolation
  DRAT-trim option: `-r RESPROOF`
- Proof minimization: removing redundant lemmas and literals
  DRAT-trim option: `-l OPTPROOF`

Demo: Proof Mining

```
git clone https://github.com/marijnheule/proof-demo
```

Introduction

Proof Checking

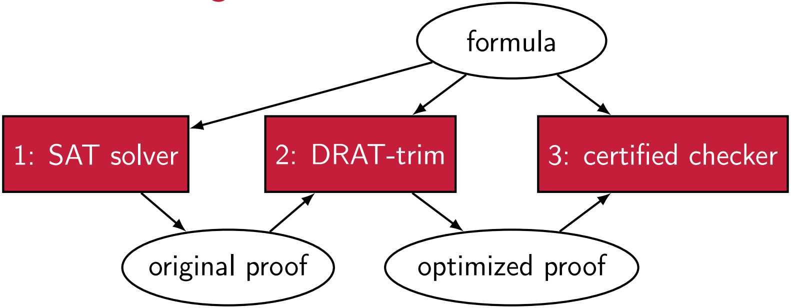
Proof Systems and Formats

Certified Checking

Applications

Conclusions

Certified Checking: Tool Chain



The proof of the Pythagorean Triples problem is almost 200 terabytes (DRAT) and has been validated in 16,000 CPU hours.

This proof has been certified using formally-verified checkers.

Certified Checking: ACL2-Based, SAT Proof Checker

We developed a mechanically verified, ACL2-based, proof checker for proofs of unsatisfiability.

Given files containing:

- the initial conjecture, as a set of clauses, and
- an ordered list of proof steps ending with the empty clause,

our mechanically verified, SAT proof checker attempts to confirm the veracity of each proof step.

Parsing is hard, while writing is easy.

- after verification, we emit a conjecture that can be compared to the initial conjecture.
- a common tool, such as `diff`, can do the comparison.

Certified Checking: Proof Claims

Basic Soundness.

```
(implies (and (formula-p formula)
              (refutation-p proof formula))
         (not (satisfiable formula))))
```

Soundness Plus Formula Confirmation.

```
(let ((formula
       (mv-nth 1 (proved-formula cnf-file clrat-file
                                 chunk-size debug
                                 nil ; incomplete-okp
                                 ctx state))))
  (implies formula
           (not (satisfiable formula))))
```

; Print proved formula, to `diff` *against input formula*

Certified Checking: Eliminate Complexity

Certified proof checking challenges:

- backward checking is complex and heavy on memory;
- unit propagation is expensive.

We eliminate both challenges by modifying the proof:

- an efficient unverified tool removes the redundancy, making forward checking as fast as backward checking;
- searching for units is replaced by hints to locate units;
- the modified proofs are not much larger;
- we do not need to trust the unverified tool.

Certified Checking: LRAT format

The LRAT format is syntactically similar to TraceCheck, however:

- The formula in **not** included in the proof
- Clause deletion support: $\langle \mathrm{pos} \rangle$“ `d` "$\langle \mathrm{clsidx} \rangle$
- Can express a RAT step: use negative cls to denote resolvent

DIMACS:

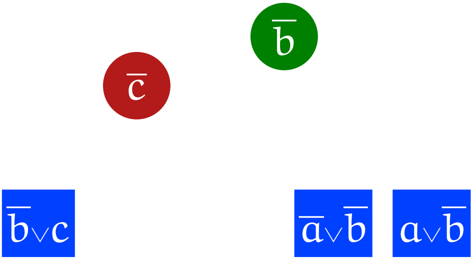
```
p cnf 3 3
-2  3 0
-1 -2 0
 1 -2 0
```

DRAT:

```
-3 0
```

LRAT:

```
4 -3 0 -1 2 3 0
```



Certified Checking: LRAT format

The LRAT format is syntactically similar to TraceCheck, however:

- The formula in **not** included in the proof
- Clause deletion support: $\langle\mathrm{pos}\rangle$“ `d` "$\langle\mathrm{clsidx}\rangle$
- Can express a RAT step: use negative cls to denote resolvent

DIMACS:

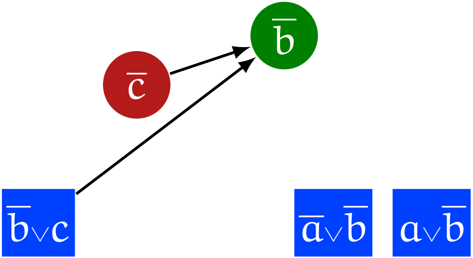
```
p cnf 3 3
-2  3 0
-1 -2 0
 1 -2 0
```

DRAT:

```
-3 0
```

LRAT:

```
4 -3 0 -1 2 3 0
```



Certified Checking: LRAT format

The LRAT format is syntactically similar to TraceCheck, however:

- The formula in **not** included in the proof
- Clause deletion support: $\langle \text{pos} \rangle$“ d "$\langle \text{clsidx} \rangle$
- Can express a RAT step: use negative cls to denote resolvent

DIMACS:

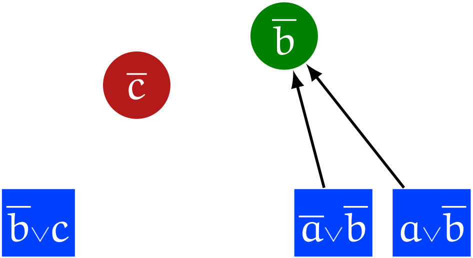
```
p cnf 3 3
-2  3 0
-1 -2 0
 1 -2 0
```

DRAT:

```
-3 0
```

LRAT:

```
4 -3 0 -1 2 3 0
```



Introduction

Proof Checking

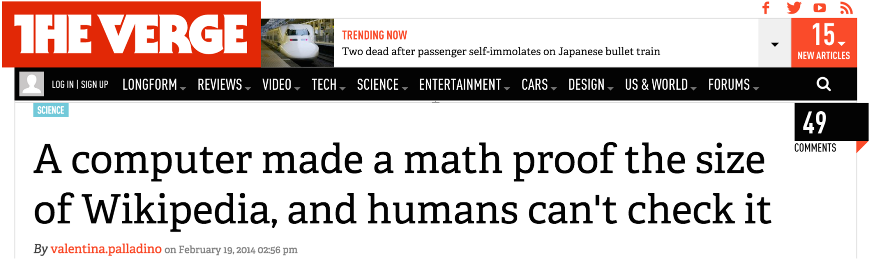
Proof Systems and Formats

Certified Checking

Applications

Conclusions

Applications: Erdős Discrepancy Conjecture



TRENDING NOW

Two dead after passenger self-immolates on Japanese bullet train

A computer made a math proof the size of Wikipedia, and humans can't check it

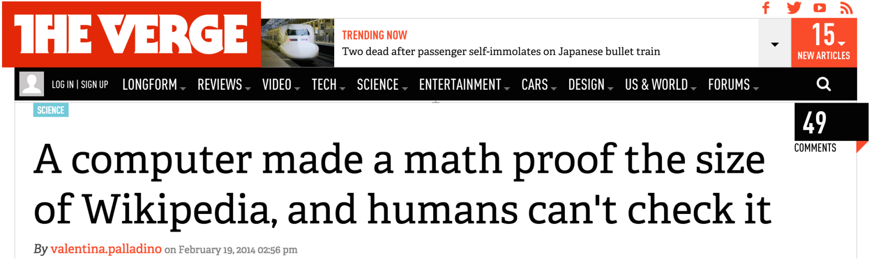
By valentina.palladino on February 19, 2014 02:56 pm

Erdős Discrepancy Conjecture was recently solved using SAT.

The conjecture states that there exists no infinite sequence of -1, +1 such that for all d, k holds that $(x_i \in \{-1, +1\})$:

$$\left| \sum_{i=1}^{k} x_{id} \right| \leq 2$$

Applications: Erdős Discrepancy Conjecture



Two dead after passenger self-immolates on Japanese bullet train

A computer made a math proof the size of Wikipedia, and humans can't check it

By valentina.palladino on February 19, 2014 02:56 pm

Erdős Discrepancy Conjecture was recently solved using SAT.

The conjecture states that there exists no infinite sequence of -1, +1 such that for all d, k holds that ($x_i \in \{-1, +1\}$):

$$\left| \sum_{i=1}^{k} x_{id} \right| \leq 2$$

The DRAT proof was 13Gb and checked with the tool DRAT-trim [SAT14]

Applications: SAT Competitions

DRAT proof logging supported by all the top-tier solvers:

- e.g. Lingeling, MiniSAT, Glucose, and CryptoMiniSAT
- Proof logging is mandatory since SAT Competition 2013
- Formally-verified checking since SAT Competition 2017

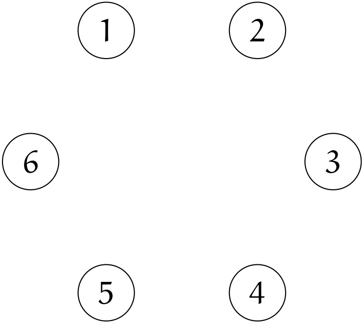
Example run of DRAT-trim on Erdős Discrepancy Proof

```
fud$ ./DRAT-trim EDP2_1161.cnf EDP2_1161.drat
c finished parsing
c detected empty clause; start verification via backward checking
c 23090 of 25142 clauses in core
c 5757105 of 6812396 lemmas in core using 469808891 resolution steps
c 16023 RAT lemmas in core; 5267754 redundant literals in core lemmas
s VERIFIED
```

Applications: Ramsey Numbers

Ramsey Number $R(k)$: What is the smallest n such that any graph with n vertices has either a clique or a co-clique of size k?

$$\begin{aligned} & R(3) = 6 \\ & R(4) = 18 \\ 43 \leq\ & R(5) \leq 48 \end{aligned}$$



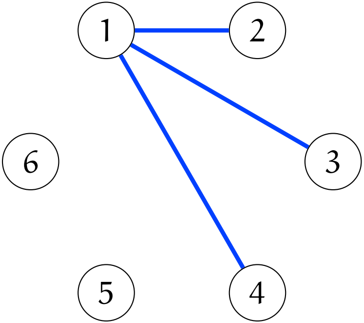
SAT solvers can determine that $R(4) = 18$ in 1 second using symmetry breaking; w/o symmetry breaking it requires weeks.

Symmetry breaking can be validated using DRAT [CADE'15]

Applications: Ramsey Numbers

Ramsey Number $R(k)$: What is the smallest n such that any graph with n vertices has either a clique or a co-clique of size k?

$$\begin{aligned} R(3) &= 6 \\ R(4) &= 18 \\ 43 \leq \quad R(5) &\leq 48 \end{aligned}$$



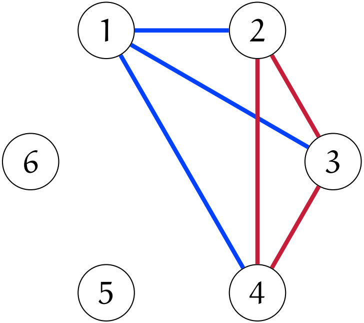
SAT solvers can determine that $R(4) = 18$ in 1 second using symmetry breaking; w/o symmetry breaking it requires weeks.

Symmetry breaking can be validated using DRAT [CADE'15]

Applications: Ramsey Numbers

Ramsey Number $R(k)$: What is the smallest n such that any graph with n vertices has either a clique or a co-clique of size k?

$$\begin{aligned} R(3) &= 6 \\ R(4) &= 18 \\ 43 \leq \quad R(5) &\leq 48 \end{aligned}$$



SAT solvers can determine that $R(4) = 18$ in 1 second using symmetry breaking; w/o symmetry breaking it requires weeks.

Symmetry breaking can be validated using DRAT [CADE'15]

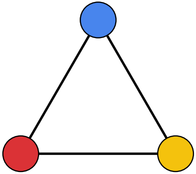
Demo: Certifying DRAT Proofs

```
git clone https://github.com/marijnheule/proof-demo
```

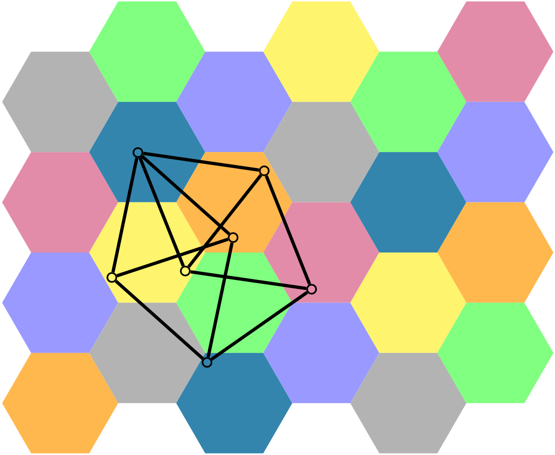
Chromatic Number of the Plane

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

The answer must be three or more because three points can be mutually 1 apart—and thus must be colored differently.



Bounds since the 1950s



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

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

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

業餘數學家為一道填色難題帶來突破！

2018/4/26 · TNL · 四色定理、填色難題、數學

Раскраска для математиков

Как покрасить плоскость?

Marijn Heule, a computer scientist at the University of Texas, Austin, found one with just 874 vertices. Yesterday he lowered this number to 826 vertices.

We found smaller graphs with SAT:

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

Propositional Proofs for Graph Validation and Shrinking

Checking that a unit-distance graph has chromatic number 5:

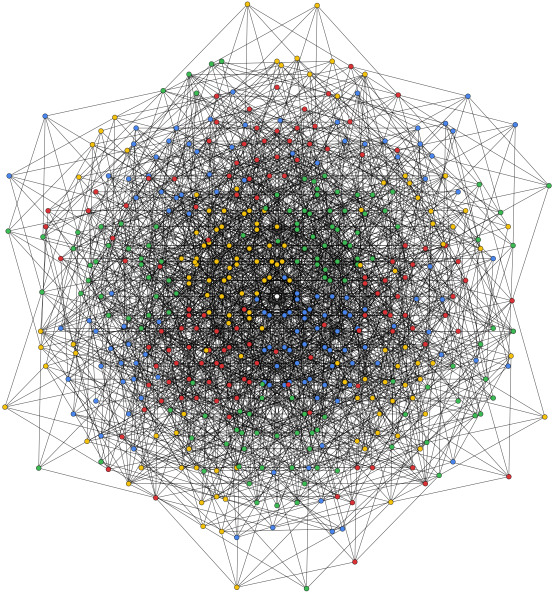
- Show that there exists a 5-coloring
- While there is no 4-coloring (formula is UNSAT)
- Unsatisfiable core represents a subgraph

SAT solvers find short proofs of unsatisfiability for these formulas

Proof minimization techniques allow further reduction

Combining the techniques allows finding much smaller graphs

Proof Minimization: 510 Vertices [Heule 2019]



Introduction

Proof Checking

Proof Systems and Formats

Certified Checking

Applications

Conclusions

Many options in DRAT-trim

```
usage: drat-trim [INPUT] [<PROOF>] [<option> ...]
  -h          print this command line option summary
  -c CORE     prints the unsatisfiable core to CORE
  -a ACTIVE   prints the active clauses to ACTIVE
  -l DRAT     prints the core lemmas to DRAT
  -L LRAT     prints the core lemmas to LRAT
  -r TRACE    prints resolution graph to TRACE
  -t <lim>    time limit in seconds (default 20000)
  -u          default unit propagation (no core)
  -f          forward mode for UNSAT
  -v          more verbose output
  -b          show progress bar
  -O          optimize proof till fixpoint
  -C          compress core lemmas (emit binary proof)
  -i          force binary proof parse mode
  -w          suppress warning messages
  -W          exit after first warning
  -p          run in plain mode (no deletion)
```

Conclusions

Verification of proofs of unsatisfiability is now mature:

- Practically all state-of-the-art SAT solvers support it;
- There exist formally-verified checkers in ACL2, Coq, Isabelle;
- Proofs exist of recently solved long-standing open problems;
- The SAT Competitions now require proof emission;
- The overhead of certification is reasonable.

Challenges:

- How to reduce the size of proofs on disk and in memory?
- What information can be mined from proofs?
- How to effectively deal with Gaussian elimination, cardinality resolution, and pseudo-Boolean reasoning? Use BDDs!