Understanding the Power of Clause Learning

Paul Beame Henry Kautz Ashish Sabharwal

Presentation: Aurèle Barrière

03/10/18

Context

Published in IJCAI in 2003

Motivation

- Deciding SAT as quickly as possible has many uses.
- DPLL-based algorithms work well.
- The most efficient algorithms are just variants of DPLL with clause learning.
- No formal work to explain the success of clause learning.

The SAT decision problem

SAT

Given a Conjunctive Normal Form (CNF) formula, decide if there exists an assignment that satisfies it.
NP-complete.

Techniques

- DPLL and its variants, **clause learning** and **restarts**
- General Resolution
- Regular Resolution
- ...

Davis-Putnam-Logemann-Loveland (DPLL)

DPLL algorithm

Branch on variables (following a **branching order**).

Apply **unit propagation** and **pure literal** rules.

If a conflict is found, backtrack to the last choice.

Clause Learning

At each conflict, draw a conflict graph then learn a new clause.

Very efficient.

Restarts

At any point during the execution of the algorithm, forget all your choices and start from the beginning, while keeping the learned clauses.

Resolution Proof Systems

General Resolution

$$(x \vee A) \wedge (\neg x \vee B) \wedge C \quad \sim_{\mathsf{SAT}} \quad (A \vee B) \wedge C$$

Resolving on x. Applying the resolution rule and getting to the empty clause means the formula is unsatisfiable.

General Resolution is not efficiently implementable

It's hard to find the good choices for the variables to resolve on.

Regular Resolution Variant

Any variable can be resolved upon at most once.
Used in the original DP in 1960.

Comparing SAT methods for unsatisfiable formulas

Proof Complexity [Cook and Reckhow, 1977]

Compare the size of the shortest proof of unsatisfiability.

Size of the Shortest Proof?

- Resolution systems: number of times we applied the resolution rule.
- DPLL-based: number of choices made (including the ones leading to conflicts).

Unsatisfiable Formulas Only

Comparing SAT methods for unsatisfiable formulas

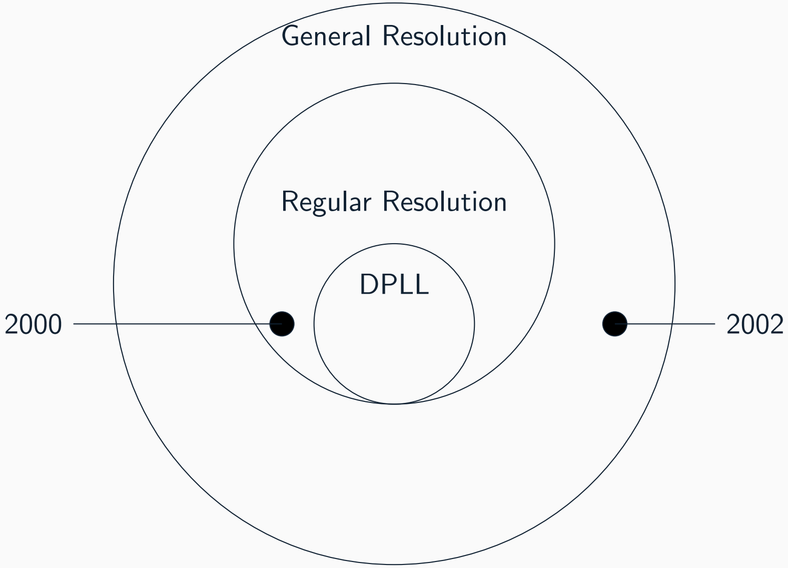
Best-Case Scenario

We are not evaluating heuristics, only the technique to prove unsatisfiability.
This framework doesn't account for the difficulty of making the right choices.
This framework doesn't account for the time to apply rules.

Relevance for satisfiable formulas

[Achlioptas, Beame, Molloy, 2001] shows that good complexity results on unsatisfiable formulas imply good results on satisfiable formulas as well.

Known Results

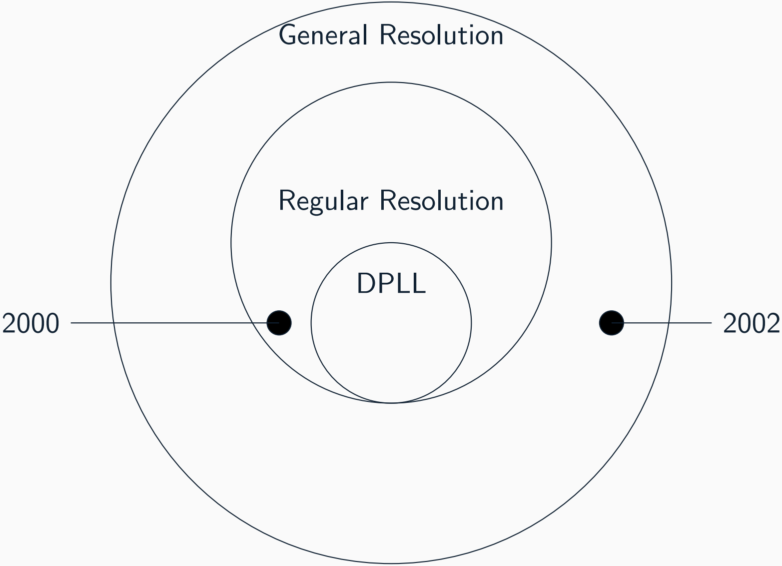


Contribution

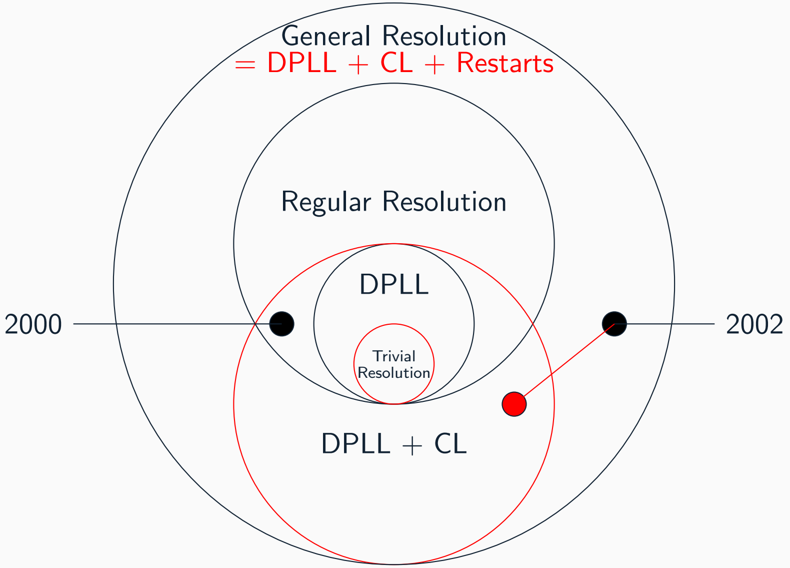
Contributions

- The first formal analysis of Clause Learning
- A new learning scheme, **FirstNewCut**
- **DPLL + clause learning** can be better than **Regular Resolution**
- **DPLL + clause learning + restarts** is equivalent **to General Resolution**
- Experimental Results

Known Results



New Results


General Resolution
= DPLL + CL + Restarts
Regular Resolution
DPLL
2000
2002
Trivial Resolution
DPLL + CL

A cut of the conflict graph

At each conflict, cutting the conflict graph gives you a clause to learn with equivalent satisfiability.
Here we consider determinist schemes that only use the variables used in the conflict.

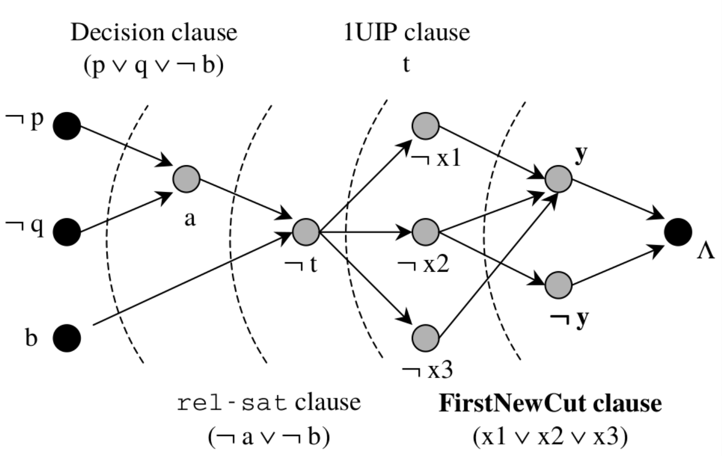
Existing schemes

- Decision clause
- `rel-sat`
- 1UIP, 2UIP...
- ...

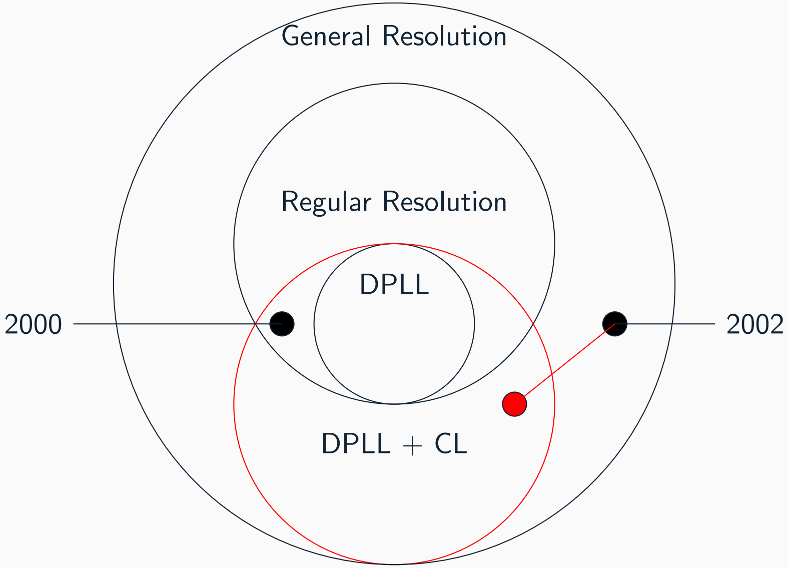
FirstNewCut

Non-trivial, non previously known clause closest to the conflict.

Only used for the proofs



Clause Learning can be better than Regular



Clause Learning can be better than Regular

Goal

Find a family of CNF unsatisfiable formulas, such that it is solved in exponential size by Regular Resolution, but in polynomial size by DPLL + clause learning.

Method

- Start with a family of function that separates Regular and General Resolution.
- Define a transformation $PT(f, \pi)$ such that solving $PT(f, \pi)$ with DPLL + CL is as long as with General Resolution. But the Regular Resolution is the same size as for f.

Clause Learning can be better than Regular - Transformation

f unsatisfiable formula with polynomial size General Resolution proof π and exponential size Regular Resolution proof.

Defining $PT(f, \pi)$

All clauses of f and, for each $C \in \pi$, for each $x \in C$, the clause $(\neg x \vee t_C)$ where t_c is a new variable.

Solving with Regular Resolution

"regularity of resolution proofs is preserved under restriction"

Exponential size proof.

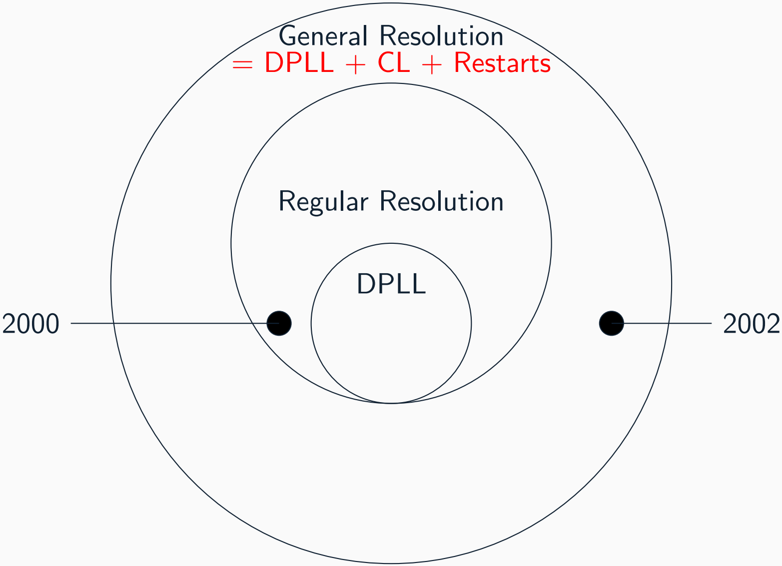
Solving with DPLL+CL

Branch on $\neg t_C$ for each $C \in \pi$.

With **FirstNewCut** you learn C. Backtrack on t_C.

Polynomial: for each branch on $\neg t_c$ you get one clause of π.

Clause Learning + Restarts equivalent to General Resolution



Clause Learning + Restarts equivalent to General Resolution

Goals: 2 simulations

- If f has a General polynomial proof, then it also has one with DPLL+CL+restarts.
- If f has a DPLL+CL+restarts polynomial proof, then it also has one with General Resolution.

First Simulation

When the resolution proof resolves $(A \vee x) \wedge (B \vee \neg x)$ to get C, branch on the negation of the literals of A and B. You learn C then restart.

Clause Learning + Restarts equivalent to General Resolution

Goals: 2 simulations

- If f has a General polynomial proof, then it also has one with DPLL+CL+restarts.
- If f has a DPLL+CL+restarts polynomial proof, then it also has one with General Resolution.

Second simulation

Let f_n with DPLL proof of size s.

A previous results shows that in at most n steps we can derive each learned clause with General Resolution.

The resulting proof has size $n \times s$.

Experimental Results

Solver	*Formula* layers	variables	*Runtime* (seconds) unsat.	satisfiable
DPLL	5	30	0.24	0.12
	6	42	110	0.02
	7	56	$>$ 24 hrs	0.07
	8	72	$>$ 24 hrs	$>$ 24 hrs
Branch sequence only	5	30	0.20	0.00
	6	42	105	0.00
	7	56	$>$ 24 hrs	0.00
	9	90	$>$ 24 hrs	$>$ 24 hrs
Clause learning only (original `zChaff`)	20	420	0.12	0.05
	40	1,640	59	36
	60	3,660	65	14
	65	4,290	‡	47
	70	4,970	‡	‡
Clause learning and branch sequence	40	1,640	0.04	0.04
	100	10,100	0.59	0.62
	500	250,500	254	288
	1,000	1,001,000	4,251	5,335
	1,500	2,551,500	21,103	‡

Table 1: `zChaff` on pebbling formulas. ‡ denotes out of memory

Conclusion

Clause Learning

- has potential for small proofs
- with restarts, has the same theoretical lower bound as General Resolution

Good Heuristics are still important!

Open Problems

- Which formula for which technique?
- Is Regular $\subsetneq$ DPLL+CL?

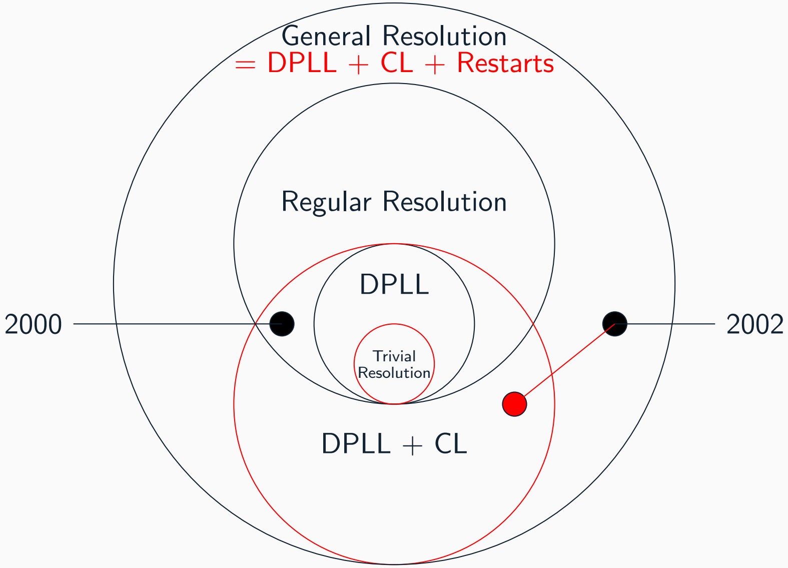
Overview

103 citations in 15 years. The following paper has 286.

Questions

- Experimental Results on transformed pebbling formulas?
- Non-deterministic instead of **FirstNewCut**?
- Satisfiable Formulas

New Results


General Resolution
= DPLL + CL + Restarts
Regular Resolution
DPLL
2000
2002
Trivial
Resolution
DPLL + CL