SAT/SMT Solvers
and
Applications

Vijay Ganesh
University of Waterloo
Winter 2013
The Goals of this Course

Introduction to SAT/SMT Solvers

- What are constraint solvers (e.g., Boolean SAT and SMT solvers)?
- Why should you care?
- Theoretical aspects (e.g., proof theoretic treatment of modern SAT algorithm)
- Practical aspects (e.g., Important SAT/SMT solver heuristics)

Applications

- Symbolic execution and solver-based automated bug finding
- Solver-based program analysis/synthesis
- Solver-based model-checking and theorem provers
- Debugging and equivalence-checking tools
- Applications in AI
- Solvers and computer security
- Solver-based modeling and requirements analysis
Goals of this Course

Lecture Schedule

- ✔ Modern CDCL SAT solvers: core ideas (VG)
- ✔ Modern SMT Solvers: core ideas (VG)
- ✔ MiniSAT: Anatomy of a modern CDCL SAT solver (VG and students)
- ✔ Symbolic execution and solver-based automated bug finding (VG and students)
- ✔ Solver-based program analysis/synthesis (Frank Tip)
- ✔ Solver-based model-checking and theorem provers (Shoham Ben-David)
- ✔ Debugging and equivalence-checking tools (VG and students)
- ✔ Applications in AI (Peter Van Beek)
- ✔ Solver-based modeling and requirements analysis (Czarnecki and Rayside)
- ✔ Solvers and computer security (Mahesh Tripunitara)
- ✔ Solvers and programming languages (VG and students)
What is a Constraint Solver?

Engineer/Mathematician’s point of view

☑ A “method” that takes as input a math formula, and produces a solution

☑ Examples: solving linear equations over the reals, polynomials, quadratic, Boolean logic,...

☑ Computing zeros of a polynomial

Theoretical computer scientist/logician’s point of view

☑ A computer program called as a satisfiability procedure that solves a specific kind of decision problem, namely, the SAT problem

☑ The input formula is in a specified logic (e.g., Boolean, first-order, reals, integers,...)

☑ Output of a satisfiability procedure

☑ UNSAT, if input has no satisfying assignments
☑ SAT, otherwise
Before modern conception of logic (Before Boole and Frege)

- From Babylon to late 1800’s: Huge amount of work on methods to solve (find roots of) polynomials over reals, integers,...
- System of linear equations over the reals (Chinese methods, Cramer’s method, Gauss elimination)
- These methods were typically not complete (e.g., worked for a special class of polynomials)

After modern conception of logic

- Systems of linear inequalities over the integers are solvable (Presburger, 1927)
- Peano arithmetic is undecidable (hence, not solvable) (Godel, 1931)
- First-order logic is undecidable (hence, not solvable) (Turing, 1936. Church, 1937)
- A exponential-time algorithm for Boolean SAT problem (Davis, Putnam, Loveland, Loggeman in 1962)
- Systems of Diophantine equations are not solvable (Matiyasevich. 1970)
- Boolean SAT problem is NP-complete (Cook 1971)
- Many efficient, scalable SAT procedures since 1971 for a variety of mathematical theories
Foundation of Software Engineering

Logic Abstractions of Computation

- Formal Methods
- Program Analysis
- Logics (Boolean,...)
- Automatic Testing
- Program Synthesis
Software Engineering & SAT/SMT Solvers
An Indispensable Tactic for Any Strategy

Formal Methods
Program Analysis

Automatic Testing
Program Synthesis

SE Goal: Reliable/Secure Software
Software Engineering & SAT/SMT Solvers
An Indispensable Tactic for Any Strategy

Formal Methods
Program Analysis

SAT/SMT Solvers

Automatic Testing
Program Synthesis
Software Engineering using Solvers

Engineering, Usability, Novelty

Program Reasoning Tool

Program Specification

Logic Formulas

SAT/SMT Solver

Program is Correct?
or Generate Counterexamples (Test cases)

SAT/UNSAT
SAT/SMT Solver Research Story

A 1000x Improvement

- Solver-based programming languages
- Compiler optimizations using solvers
- Solver-based debuggers
- Solver-based type systems
- Solver-based concurrency bugfinding
- Solver-based synthesis
- Bio & Optimization

- Concolic Testing
- Program Analysis
- Equivalence Checking
- Auto Configuration

- Bounded MC
- Program Analysis
- AI

1,000,000 Constraints

100,000 Constraints

10,000 Constraints

1,000 Constraints


Why Should You Care?

**SAT/SMT solver user**
- Use solver as a black-box
- More importantly, solver algorithms are influencing algorithms in other areas
- Synthesis: Sketching by Armando Solar-Lezama uses SAT techniques
- Analysis: Combining static and dynamic analysis has a flavor of the SAT algorithm
- Model-checking: The IC3 algorithm integrates SAT deeply into the model-checker

**Some SAT/SMT solvers users at Waterloo**
- Imeson, Tripunitara, Garg are using programmatic SAT for hardware security
- Krzysztof Czarnecki’s group is using solvers for auto-configuration
- Derek Rayside’s group is using solvers for software modelling through Alloy
- Frank Tip’s group is using solvers for analysis
- Peter Van Beek’s group works on CSPs, closely related to SAT
- Venkat Raman is working on complexity-theoretic aspects of the SAT problem
- Lin Tan’s group is looking into symbolic-execution/solver-based bug finding
Brief overview of relevant Logic Concepts

What is a Logic?

What is a mathematical theory?

Notions of models, truth and proof

What is the connection between truth and proof?

What are the satisfiability and validity problems?

What is a decision procedure?

What is a satisfiability procedure?

What is the connection between satisfiability, validity and proof?

What is meant by soundness and completeness?
Logic, model, truth, assignments

- Study of valid modes of reasoning (inductive, deductive, ...)
- Formal language (e.g., Boolean logic, first-order logic,...)
- Rules for constructing well-formed formulas
- An associated proof system (axioms, inference rules,...)
- Model
  - Interpretation of connectives, functions, predicates over a domain
  - True, false
  - Assignment: Mapping of variables to elements of the domain
The SAT/SMT Problem

- Rich logics (Modular arithmetic, Arrays, Strings,...)
- NP-complete, PSPACE-complete,...
- Practical, scalable, usable, automatic
- Enable novel software reliability approaches
The SAT/SMT Problem

- Closely related to the Validity Problem
- Soundness, completeness, termination
- Connecting model theory and proof theory

Logic

Formula

(q \lor p \lor \neg r)
(q \lor \neg p \lor r)
...

Solver

SAT
UNSAT
Lecture Outline

Points already covered

- Motivation for SAT/SMT solvers in software engineering
- High-level description of the SAT/SMT problem & logics
- Defined logic, models, truth, proofs, SAT procedure, soundness, completeness

Rest of the lecture

- Modern CDCL SAT solver architecture & techniques
- SAT/SMT-based applications
- Future of SAT/SMT solvers
- Some history (who, when,...) and references sprinkled throughout the talk
- Non-CDCL SAT techniques
A **literal** $p$ is a Boolean variable $x$ or its negation $\neg x$.

A **clause** $C$ is a disjunction of literals: $x_2 \lor \neg x_{41} \lor x_{15}$

A **CNF** is a conjunction of clauses: $(x_2 \lor \neg x_{41} \lor x_{15}) \land (x_6 \lor \neg x_2) \land (x_3 \lor \neg x_4 \lor \neg x_6)$

All Boolean formulas assumed to be in **CNF**

**Assignment** is a mapping (binding) from variables to Boolean values (True, False).

A **unit clause** $C$ is a clause with a single unbound literal

The **SAT-problem** is:

- Find an assignment s.t. each input clause has a true literal (aka input formula has a solution or is SAT)
- OR establish input formula has no solution (aka input formula is UNSAT)

The Input formula is represented in **DIMACS Format**:
```
c DIMACS
p cnf 6 3
2 -1 5 0
6 -2 0
3 -4 -6 0
```
The Basic Solver

DPLL(Θcnf, assign) {
    Propagate unit clauses;
    if ”conflict”: return FALSE;
    if ”complete assign”: return TRUE;
    ”pick decision variable x”;
    return DPLL(Θcnf | x=0, assign[x=0]) || DPLL(Θcnf | x=1, assign[x=1]);
}

• Propagate (Boolean Constant Propagation):
  • Propagate inferences due to unit clauses
  • Most time in solving goes into this

• Detect Conflict:
  • Conflict: partial assignment is not satisfying

• Decide (Branch):
  • Choose a variable & assign some value

• Backtracking:
  • Implicitly done by the recursion
Modern CDCL SAT Solver Architecture

Key Steps and Data-structures

Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decide()

TopLevel Conflict?

Return UNSAT

BackJump()

Key steps

- Decide()
- Propagate() (BCP: Boolean constraint propagation)
- Conflict analysis and learning()
- Backjump()
- Forget()
- Restart()

CDCL: Conflict-Driven Clause-Learning

- Conflict analysis is a key step
- Results in learning a conflict clause
- Prunes the search space

Key data-structures (State):

- Stack or trail of partial assignments (AT)
- Input clause database
- Conflict clause database
- Conflict graph
- Decision level (DL) of a variable

Vijay Ganesh

Wednesday, 9 January, 13
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

Input SAT Instance

- Propagate (Boolean Constant Propagation):
  - Propagate inferences due to unit clauses
  - Most time in solving goes into this

- Conflict Analysis:
  - Analyze the reason (learn conflict clause)
  - Conflict clause blocks the non-satisfying
    and a large set of other 'no-good' assignments

- Decide:
  - Choose a variable & assign some value (decision)
  - Each decision is a decision level
  - Imposes dynamic variable order

- Conflict?:
  - partial assignment is not satisfying

- Decision Level (DL): variable natural number

- BackJump:
  - Undo the decision(s) that caused no-good assignment
  - Assign 'decision variables' different values
  - Go back several decision levels

Marques-Silva & Sakallah (1999)

Marques-Silva & Sakallah (1999)

Backtrack (Davis, Putnam, Loveland, Logemann 1962)
Input SAT Instance

Propagate()
(BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Decide()

TopLevel Conflict?

Return SAT

BackJump()

Return UNSAT

• Propagate (Boolean Constant Propagation):
  • Propagate inferences due to unit clauses
  • Most time in solving goes into this

Marques-Silva & Sakallah (1999)

Backtrack (Davis, Putnam, Loveland, Logemann 1962)
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

1. **Input SAT Instance**
2. **Propagate() (BCP)**
   - No Conflict?
   - All Vars Assigned?
   - **Conflict Analysis()**
   - **Detect Conflict?**
   - **Return SAT**
   - **Decide()**
   - **Top Level Conflict?**
   - **Return UNSAT**
   - **BackJump()**

### Key Points:
- **Propagate (Boolean Constant Propagation):**
  - Propagate inferences due to unit clauses
  - Most time in solving goes into this
- **Detect Conflict?**
  - Conflict: partial assignment is not satisfying
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

TopLevel Conflict?

Return SAT

Decide()

BackJump()

Return UNSAT

• Propagate (Boolean Constant Propagation):
  • Propagate inferences due to unit clauses
  • Most time in solving goes into this

• Detect Conflict?
  • Conflict: partial assignment is not satisfying

Input SAT Instance

• Propagate (Boolean Constant Propagation):
  • Propagate inferences due to unit clauses
  • Most time in solving goes into this

• Detect Conflict?
  • Conflict: partial assignment is not satisfying
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decide()

TopLevel Conflict?

Return UNSAT

BackJump()

• Propagate (Boolean Constant Propagation):
  • Propagate inferences due to unit clauses
  • Most time in solving goes into this

• Detect Conflict?
  • Conflict: partial assignment is not satisfying

• Decide (Branch):
  • Choose a variable & assign some value (decision)
  • Basic mechanism to do search
  • Imposes dynamic variable order
  • Decision Level (DL): variable ⇒ natural number

Marques-Silva & Sakallah (1999)
Marques-Silva & Sakallah (1999)
Backtrack (Davis, Putnam, Loveland, Logemann 1962)
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

- **Propagate (Boolean Constant Propagation):**
  - Propagate inferences due to unit clauses
  - Most time in solving goes into this

- **Detect Conflict?**
  - Conflict: partial assignment is not satisfying

- **Decide (Branch):**
  - Choose a variable & assign some value (decision)
  - Basic mechanism to do search
  - Imposes dynamic variable order
  - Decision Level (DL): variable $\Rightarrow$ natural number

- **Conflict analysis and clause learning:**
  - Analyze the reason (learn conflict clause)
  - Conflict clause blocks the non-satisfying & a large set of other 'no-good' assignments
  - Marques-Silva & Sakallah (1999)

- **BackJump:**
  - Undo the decision(s) that caused no-good assignment
  - Assign 'decision variables' different values
  - Go back several decision levels
  - Marques-Silva & Sakallah (1999)
  - Backtrack (Davis, Putnam, Loveland, Logemann 1962)
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

- **Propagate:**
  - Propagate inferences due to unit clauses
  - Most time in solving goes into this

- **Detect Conflict?**
  - Conflict: partial assignment is not satisfying

- **Decide (Branch):**
  - Choose a variable & assign some value (decision)
  - Each decision is a decision level
  - Imposes dynamic variable order
  - Decision Level (DL): variable ⇒ natural number

- **Conflict analysis and clause learning:**
  - Compute assignments that lead to conflict (analysis)
  - Construct conflict clause blocks the non-satisfying & a large set of other ‘no-good’ assignments (learning)
  - Marques-Silva & Sakallah (1996)
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

Input SAT Instance

- **Propagate:**
  - Propagate inferences due to unit clauses
  - Most time in solving goes into this

- **Detect Conflict?**
  - Conflict: partial assignment is not satisfying

- **Decide (Branch):**
  - Choose a variable & assign some value (decision)
  - Each decision is a decision level
  - Imposes dynamic variable order
  - Decision Level (DL): variable ⇒ natural number

- **Conflict analysis and clause learning:**
  - Compute assignments that lead to conflict (analysis)
  - Construct conflict clause blocks the non-satisfying & a large set of other ‘no-good’ assignments (learning)
  - Marques-Silva & Sakallah (1996)

BackJump()
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Res SAT

Decide()

TopLevel Conflict?

Return UNSAT

BackJump()

- Propagate:
  - Propagate inferences due to unit clauses
  - Most time in solving goes into this

- Detect Conflict?
  - Conflict: partial assignment is not satisfying

- Decide:
  - Choose a variable & assign some value (decision)
  - Each decision is a decision level
  - Imposes dynamic variable order
  - Decision Level (DL): variable \rightarrow natural number

- Conflict analysis and clause learning:
  - Compute assignments that lead to conflict (analysis)
  - Construct conflict clause blocks the non-satisfying & a large set of other ‘no-good’ assignments (learning)
  - Marques-Silva & Sakallah (1996)

- Conflict-driven BackJump:
  - Undo the decision(s) that caused no-good assignment
  - Assign ‘decision variables’ different values
  - Go back several decision levels
  - Backjump: Marques-Silva, Sakallah (1999)
  - Backtrack: Davis, Putnam, Loveland, Logemann (1962)
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

- **Propagate:**
  - Propagate inferences due to unit clauses
  - Most time in solving goes into this

- **Detect Conflict?**
  - Conflict: partial assignment is not satisfying

- **Decide:**
  - Choose a variable & assign some value (decision)
  - Each decision is a decision level
  - Imposes dynamic variable order
  - Decision Level (DL): variable $\Rightarrow$ natural number

- **Conflict analysis and clause learning:**
  - Compute assignments that lead to conflict (analysis)
  - Construct conflict clause blocks the non-satisfying & a large set of other ‘no-good’ assignments (learning)
  - Marques-Silva & Sakallah (1996)

- **Conflict-driven BackJump:**
  - Undo the decision(s) that caused no-good assignment
  - Assign ‘decision variables’ different values
  - Go back several decision levels
  - Backjump: Marques-Silva, Sakallah (1999)
  - Backtrack: Davis, Putnam, Loveland, Logemann (1962)
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

Input SAT Instance

Propagate()
(BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decide()

TopLevel Conflict?

Return UNSAT

BackJump()

{3, 6, -7, 8}
{1, 4, 7}
{-8, 4}
{-1, -3, 8}
{-3, -4, -8}
{-1, -2, 3, 4, -6}

Unit clause
(BCP)

{3, 6, -7, 8}
{1, 4, 7}
{-8, 4}
{-1, -3, 8}
{-3, -4, -8}
{-1, -2, 3, 4, -6}

CONFLICT!
(Trigger to analyze & backjump)

Another unit clause
(more BCP)

{3, 6, -7, 8}
{1, 4, 7}
{-8, 4}
{-1, -3, 8}
{-3, -4, -8}
{-1, -2, 3, 4, -6}

CONFLICT!
(Trigger to analyze & backjump)
Modern CDCL SAT Solver Architecture
Decide() Details: VSIDS Heuristic

- Decide() or Branching():
  - Choose a variable & assign some value (decision)
  - Imposes dynamic variable order (Malik et al. 2001)

- How to choose a variable:
  - VSIDS heuristics
  - Each variable has an activity
  - Activity is bumped additively, if variable occurs in conflict clause
  - Activity of all variables is decayed by multiplying by const < 1
  - Next decision variable is the variable with highest activity
  - Over time, truly important variables get high activity
  - This is pure magic, and seems to work for many problems

Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decide()

TopLevel Conflict?

Return UNSAT

BackJump()
Modern CDCL SAT Solver Architecture

Propagate() Details: Two-watched Literal Scheme

Input SAT Instance

Propagate()
(BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decide()

TopLevel Conflict?

Return UNSAT

BackJump()

{3, 6, -7, 8}
{1, 4, 7}
{-8, 4}
{-1, -3, 8}
{-3, -4, -8}
{-1, -2, 3, 4, -6}

{3, 6, -7, 8}
{1, 4, 7}
{-8, 4}
{-1, -3, 8}
{-3, -4, -8}
{-1, -2, 3, 4, -6}

Watched Literal | Watcher List
-1 \{1, 3, 8\}
-3 \{-1, -3, 8\}
8 \{-1, 3, 8\}

Watched Literal | Watcher List
-1 \{-1, -3, 8\}
-3 ...
8 \{-1, 3, 8\}

The constraint propagates 8
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

Input SAT Instance

Propagate() (BCP)

No Conflict?

Conflicts?

Return SAT

Decide()

All Vars Assigned?

Conflicts Analysis()

Return UNSAT

BackJump()

Basic Backtracking Search

- Flip the last decision
- Try setting 1 to False
- Highly inefficient
- No learning from mistakes

Unit clause (BCP)

More unit clauses (more BCP)

CONFLICT! (Trigger to analyze & backjump)

Wednesday, 9 January, 13
Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decision Level (DL)
  - Map from Boolean variables in input to natural numbers
  - All unit clauses in input & resultant propagations get DL = 0
  - Every decision var gets a DL in increasing order >= 1
  - All propagations due to decision var at DL=x get the DL=x

Conflict Graph (CG) or Implication Graph
  - Directed Graph that records decisions & propagations
  - Vertices: literals, Edge: unit clauses

Conflict Clause (CC)
  - Clause returned by Conflict Analysis(), added to conflict DB
  - Implied by the input formula
  - A cut in the CG
  - Prunes the search

Assignment Trail (AT)
  - A stack of partial assignment to literals, with DL info

Return UNSAT

BackJump()
Modern CDCL SAT Solver Architecture

Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \( \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\} \)

Current decision: \( \{X_1 = 1@6\} \)

Clause DB

\( W_1 = (\neg X_1 + X_2) \)
\( W_2 = (\neg X_1 + X_3 + X_9) \)
\( W_3 = (\neg X_2 + \neg X_3 + X_4) \)
\( W_4 = (\neg X_4 + X_5 + X_{10}) \)
\( W_5 = (\neg X_4 + X_6 + X_{11}) \)
\( W_6 = (\neg X_5 + \neg X_6) \)
\( W_7 = (X_1 + X_7 + \neg X_{12}) \)
\( W_8 = (X_1 + X_8) \)
\( W_9 = (\neg X_7 + \neg X_8 + \neg X_{13}) \)
Current Assignment Trail: \( \{ X_9 = 0 @ 1, X_{10} = 0 @ 3, X_{11} = 0 @ 3, X_{12} = 1 @ 2, X_{13} = 1 @ 2, \ldots \} \)

Current decision: \( \{ X_1 = 1 @ 6 \} \)

Clause DB

\[
\begin{align*}
W_1 &= (\neg X_1 + X_2) \\
W_2 &= (\neg X_1 + X_3 + X_9) \\
W_3 &= (\neg X_2 + \neg X_3 + X_4) \\
W_4 &= (\neg X_4 + X_5 + X_{10}) \\
W_5 &= (\neg X_4 + X_6 + X_{11}) \\
W_6 &= (\neg X_5 + \neg X_6) \\
W_7 &= (X_1 + X_7 + \neg X_{12}) \\
W_8 &= (X_1 + X_8) \\
W_9 &= (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}
\]
Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \(\{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\}\)

Current decision: \(\{X_1 = 1@6\}\)

Clause DB

- \(W_1 = (\neg X_1 + X_2)\)
- \(W_2 = (\neg X_1 + X_3 + X_9)\)
- \(W_3 = (\neg X_2 + \neg X_3 + X_4)\)
- \(W_4 = (\neg X_4 + X_5 + X_{10})\)
- \(W_5 = (\neg X_4 + X_6 + X_{11})\)
- \(W_6 = (\neg X_5 + \neg X_6)\)
- \(W_7 = (X_1 + X_7 + \neg X_{12})\)
- \(W_8 = (X_1 + X_8)\)
- \(W_9 = (\neg X_7 + \neg X_8 + \neg X_{13})\)
Modern CDCL SAT Solver Architecture
Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, ...\}

Current decision: \{X_1 = 1@6\}

Clause DB

\begin{align*}
W_1 &= (\neg X_1 + X_2) \\
W_2 &= (\neg X_1 + X_3 + X_9) \\
W_3 &= (\neg X_2 + \neg X_3 + X_4) \\
W_4 &= (\neg X_4 + X_5 + X_{10}) \\
W_5 &= (\neg X_4 + X_6 + X_{11}) \\
W_6 &= (\neg X_5 + \neg X_6) \\
W_7 &= (X_1 + X_7 + \neg X_{12}) \\
W_8 &= (X_1 + X_8) \\
W_9 &= (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}
Conflict Analysis/Learn() Details: Implication Graph

- Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, ...\}
- Current decision: \{X_1 = 1@6\}

Clause DB

\[
\begin{align*}
W_1 &= (\neg X_1 + X_2) \\
W_2 &= (\neg X_1 + X_3 + X_9) \\
W_3 &= (\neg X_2 + \neg X_3 + X_4) \\
W_4 &= (\neg X_4 + X_5 + X_{10}) \\
W_5 &= (\neg X_4 + X_6 + X_{11}) \\
W_6 &= (\neg X_5 + \neg X_6) \\
W_7 &= (X_1 + X_7 + \neg X_{12}) \\
W_8 &= (X_1 + X_8) \\
W_9 &= (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}
\]
Modern CDCL SAT Solver Architecture

Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, ...\}

Current decision: \{X_1 = 1@6\}

Clause DB

\[
\begin{align*}
W_1 &= (\neg X_1 + X_2) \\
W_2 &= (\neg X_1 + X_3 + X_9) \\
W_3 &= (\neg X_2 + \neg X_3 + X_4) \\
W_4 &= (\neg X_4 + X_5 + X_{10}) \\
W_5 &= (\neg X_4 + X_6 + X_{11}) \\
W_6 &= (\neg X_5 + \neg X_6) \\
W_7 &= (X_1 + X_7 + \neg X_{12}) \\
W_8 &= (X_1 + X_8) \\
W_9 &= (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}
\]
Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, ...\}

Current decision: \{X_1 = 1@6\}

Clause DB

\[ W_1 = (\neg X_1 + X_2) \]
\[ W_2 = (\neg X_1 + X_3 + X_9) \]
\[ W_3 = (\neg X_2 + \neg X_3 + X_4) \]
\[ W_4 = (\neg X_4 + X_5 + X_{10}) \]
\[ W_5 = (\neg X_4 + X_6 + X_{11}) \]
\[ W_6 = (\neg X_5 + \neg X_6) \]
\[ W_7 = (X_1 + X_7 + \neg X_{12}) \]
\[ W_8 = (X_1 + X_8) \]
\[ W_9 = (\neg X_7 + \neg X_8 + \neg X_{13}) \]
Modern CDCL SAT Solver Architecture

Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, ...\}

Current decision: \{X_1 = 1@6\}

Clause DB

\begin{align*}
W_1 &= (\neg X_1 + X_2) \\
W_2 &= (\neg X_1 + X_3 + X_9) \\
W_3 &= (\neg X_2 + \neg X_3 + X_4) \\
W_4 &= (\neg X_4 + X_5 + X_{10}) \\
W_5 &= (\neg X_4 + X_6 + X_{11}) \\
W_6 &= (\neg X_5 + \neg X_6) \\
W_7 &= (X_1 + X_7 + \neg X_{12}) \\
W_8 &= (X_1 + X_8) \\
W_9 &= (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}
Modern CDCL SAT Solver Architecture

Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\}

Current decision: \{X_1 = 1@6\}

Clause DB

\[ W_1 = (\neg X_1 + X_2) \]
\[ W_2 = (\neg X_1 + X_3 + X_9) \]
\[ W_3 = (\neg X_2 + \neg X_3 + X_4) \]
\[ W_4 = (\neg X_4 + X_5 + X_{10}) \]
\[ W_5 = (\neg X_4 + X_6 + X_{11}) \]
\[ W_6 = (\neg X_5 + \neg X_6) \]
\[ W_7 = (X_1 + X_7 + \neg X_{12}) \]
\[ W_8 = (X_1 + X_8) \]
\[ W_9 = (\neg X_7 + \neg X_8 + \neg X_{13}) \]
Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\}

Current decision: \{X_1 = 1@6\}

Clause DB

\begin{align*}
W_1 & = (\neg X_1 + X_2) \\
W_2 & = (\neg X_1 + X_3 + X_9) \\
W_3 & = (\neg X_2 + \neg X_3 + X_4) \\
W_4 & = (\neg X_4 + X_5 + X_{10}) \\
W_5 & = (\neg X_4 + X_6 + X_{11}) \\
W_6 & = (\neg X_5 + \neg X_6) \\
W_7 & = (X_1 + X_7 + \neg X_{12}) \\
W_8 & = (X_1 + X_8) \\
W_9 & = (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}
Current Assignment Trail: \( \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\} \)

Current decision: \( \{X_1 = 1@6\} \)

Clause DB

\[
W_1 = (\neg X_1 + X_2) \\
W_2 = (\neg X_1 + X_3 + X_9) \\
W_3 = (\neg X_2 + \neg X_3 + X_4) \\
W_4 = (\neg X_4 + X_5 + X_{10}) \\
W_5 = (\neg X_4 + X_6 + X_{11}) \\
W_6 = (\neg X_5 + \neg X_6) \\
W_7 = (X_1 + X_7 + \neg X_{12}) \\
W_8 = (X_1 + X_8) \\
W_9 = (\neg X_7 + \neg X_8 + \neg X_{13})
\]
Modern CDCL SAT Solver Architecture
Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\}

Current decision: \{X_1 = 1@6\}

Clause DB

- \(W_1 = (\neg X_1 + X_2)\)
- \(W_2 = (\neg X_1 + X_3 + X_9)\)
- \(W_3 = (\neg X_2 + \neg X_3 + X_4)\)
- \(W_4 = (\neg X_4 + X_5 + X_{10})\)
- \(W_5 = (\neg X_4 + X_6 + X_{11})\)
- \(W_6 = (\neg X_5 + \neg X_6)\)
- \(W_7 = (X_1 + X_7 + \neg X_{12})\)
- \(W_8 = (X_1 + X_8)\)
- \(W_9 = (\neg X_7 + \neg X_8 + \neg X_{13})\)
Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\}

Current decision: \{X_1 = 1@6\}

Clause DB

\begin{align*}
W_1 &= (\neg X_1 + X_2) \\
W_2 &= (\neg X_1 + X_3 + X_9) \\
W_3 &= (\neg X_2 + \neg X_3 + X_4) \\
W_4 &= (\neg X_4 + X_5 + X_{10}) \\
W_5 &= (\neg X_4 + X_6 + X_{11}) \\
W_6 &= (\neg X_5 + \neg X_6) \\
W_7 &= (X_1 + X_7 + \neg X_{12}) \\
W_8 &= (X_1 + X_8) \\
W_9 &= (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}
Current Assignment Trail: \( \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\} \)

Current decision: \( \{X_1 = 1@6\} \)

Clause DB

- \( W_1 = (\neg X_1 + X_2) \)
- \( W_2 = (\neg X_1 + X_3 + X_9) \)
- \( W_3 = (\neg X_2 + \neg X_3 + X_4) \)
- \( W_4 = (\neg X_4 + X_5 + X_{10}) \)
- \( W_5 = (\neg X_4 + X_6 + X_{11}) \)
- \( W_6 = (\neg X_5 + \neg X_6) \)
- \( W_7 = (X_1 + X_7 + \neg X_{12}) \)
- \( W_8 = (X_1 + X_8) \)
- \( W_9 = (\neg X_7 + \neg X_8 + \neg X_{13}) \)
Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, ...\}

Current decision: \{X_1 = 1@6\}

Clause DB

\[ W_1 = (\neg X_1 + X_2) \]
\[ W_2 = (\neg X_1 + X_3 + X_9) \]
\[ W_3 = (\neg X_2 + \neg X_3 + X_4) \]
\[ W_4 = (\neg X_4 + X_5 + X_{10}) \]
\[ W_5 = (\neg X_4 + X_6 + X_{11}) \]
\[ W_6 = (\neg X_5 + \neg X_6) \]
\[ W_7 = (X_1 + X_7 + \neg X_{12}) \]
\[ W_8 = (X_1 + X_8) \]
\[ W_9 = (\neg X_7 + \neg X_8 + \neg X_{13}) \]
Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\}

Current decision: \{X_1 = 1@6\}

Clause DB

\begin{align*}
W_1 &= \neg X_1 + X_2 \\
W_2 &= \neg X_1 + X_3 + X_9 \\
W_3 &= \neg X_2 + \neg X_3 + X_4 \\
W_4 &= \neg X_4 + X_5 + X_{10} \\
W_5 &= \neg X_4 + X_6 + X_{11} \\
W_6 &= \neg X_5 + \neg X_6 \\
W_7 &= X_1 + X_7 + \neg X_{12} \\
W_8 &= X_1 + X_8 \\
W_9 &= \neg X_7 + \neg X_8 + \neg X_{13}
\end{align*}
Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, ...\}

Current decision: \{X_1 = 1@6\}

**Clause DB**

1. \(W_1 = \neg X_1 + X_2\)
2. \(W_2 = \neg X_1 + X_3 + X_9\)
3. \(W_3 = \neg X_2 + \neg X_3 + X_4\)
4. \(W_4 = \neg X_4 + X_5 + X_{10}\)
5. \(W_5 = \neg X_4 + X_6 + X_{11}\)
6. \(W_6 = \neg X_5 + \neg X_6\)
7. \(W_7 = X_1 + X_7 + \neg X_{12}\)
8. \(W_8 = X_1 + X_8\)
9. \(W_9 = \neg X_7 + \neg X_8 + \neg X_{13}\)
Modern CDCL SAT Solver Architecture

Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\}

Current decision: \{X_1 = 1@6\}

Clause DB

\begin{align*}
W_1 &= (\neg X_1 + X_2) \\
W_2 &= (\neg X_1 + X_3 + X_9) \\
W_3 &= (\neg X_2 + \neg X_3 + X_4) \\
W_4 &= (\neg X_4 + X_5 + X_{10}) \\
W_5 &= (\neg X_4 + X_6 + X_{11}) \\
W_6 &= (\neg X_5 + \neg X_6) \\
W_7 &= (X_1 + X_7 + \neg X_{12}) \\
W_8 &= (X_1 + X_8) \\
W_9 &= (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}
Modern CDCL SAT Solver Architecture

Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \{X_9 = 0 @ 1, X_{10} = 0 @ 3, X_{11} = 0 @ 3, X_{12} = 1 @ 2, X_{13} = 1 @ 2, \ldots\}

Current decision: \{X_1 = 1 @ 6\}

Clause DB

\begin{align*}
W_1 &= (\neg X_1 + X_2) \\
W_2 &= (\neg X_1 + X_3 + X_9) \\
W_3 &= (\neg X_2 + \neg X_3 + X_4) \\
W_4 &= (\neg X_4 + X_5 + X_{10}) \\
W_5 &= (\neg X_4 + X_6 + X_{11}) \\
W_6 &= (\neg X_5 + \neg X_6) \\
W_7 &= (X_1 + X_7 + \neg X_{12}) \\
W_8 &= (X_1 + X_8) \\
W_9 &= (\neg X_7 + \neg X_8 + \neg X_{13})
\end{align*}

Conflict

Wednesday, 9 January, 13
Modern CDCL SAT Solver Architecture

Conflict Analysis/Learn() Details: Implication Graph

Current Assignment Trail: \( \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\} \)

Current decision: \( \{X_1 = 1@6\} \)

**Clause DB**

- \( W_1 = (\neg X_1 + X_2) \)
- \( W_2 = (\neg X_1 + X_3 + X_9) \)
- \( W_3 = (\neg X_2 + \neg X_3 + X_4) \)
- \( W_4 = (\neg X_4 + X_5 + X_{10}) \)
- \( W_5 = (\neg X_4 + X_6 + X_{11}) \)
- \( W_6 = (\neg X_5 + \neg X_6) \)
- \( W_7 = (X_1 + X_7 + \neg X_{12}) \)
- \( W_8 = (X_1 + X_8) \)
- \( W_9 = (\neg X_7 + \neg X_8 + \neg X_{13}) \)

**CONFLICT GRAPH**

- \( \)
Conflict Analysis/Learn() Details: Conflict Clause

Current Assignment Trail: \( \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\} \)

Current Decision: \( \{X_1 = 1@6\} \)

Simplest strategy is to traverse the conflict graph backwards until decision variables: conflict clause includes only decision variables \( (\neg X_1 + X_9 + X_{10} + X_{11}) \)

### Clause DB

- \( W_1 = (\neg X_1 + X_2) \)
- \( W_2 = (\neg X_1 + X_3 + X_9) \)
- \( W_3 = (\neg X_2 + \neg X_3 + X_4) \)
- \( W_4 = (\neg X_4 + X_5 + X_{10}) \)
- \( W_5 = (\neg X_4 + X_6 + X_{11}) \)
- \( W_6 = (\neg X_5 + \neg X_6) \)
- \( W_7 = (X_1 + X_7 + \neg X_{12}) \)
- \( W_8 = (X_1 + X_8) \)
- \( W_9 = (\neg X_7 + \neg X_8 + \neg X_{13}) \)
Modern CDCL SAT Solver Architecture

Conflict Analysis/Learn() Details: Conflict Clause

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots\}

Current Decision: \{X_1 = 1@6\}

Another strategy is to use First Unique Implicant Point (UIP):
Traverse graph backwards in breadth-first, expand literals of conflict, stop at first UIP

Clause DB

- \( W_1 = \neg X_1 + X_2 \)
- \( W_2 = \neg X_1 + X_3 + X_9 \)
- \( W_3 = \neg X_2 + \neg X_3 + X_4 \)
- \( W_4 = \neg X_4 + X_5 + X_{10} \)
- \( W_5 = \neg X_4 + X_6 + X_{11} \)
- \( W_6 = \neg X_5 + \neg X_6 \)
- \( W_7 = X_1 + X_7 + \neg X_{12} \)
- \( W_8 = X_1 + X_8 \)
- \( W_9 = \neg X_7 + \neg X_8 + \neg X_{13} \)
Conflict Analysis/Learn() Details: BackTrack

Current Assignment Trail: \{X_9 = 0@1, X_{10} = 0@3, X_{11} = 0@3, X_{12} = 1@2, X_{13} = 1@2, \ldots \}

Current decision: \{X_1 = 1@6\}

Strategy: Closest decision level (DL) \leq\ current DL for which conflict clause is unit. Undo \{X_1 = 1@6\}

Clause DB

\begin{align*}
W_1 &= \neg X_1 + X_2 \\
W_2 &= \neg X_1 + X_3 + X_9 \\
W_3 &= \neg X_2 + \neg X_3 + X_4 \\
W_4 &= \neg X_4 + X_5 + X_{10} \\
W_5 &= \neg X_4 + X_6 + X_{11} \\
W_6 &= \neg X_5 + \neg X_6 \\
W_7 &= X_1 + X_7 + \neg X_{12} \\
W_8 &= X_1 + X_8 \\
W_9 &= \neg X_7 + \neg X_8 + \neg X_{13}
\end{align*}

CONFLICT GRAPH

\[X_9 = 0@1\]
\[X_{10} = 0@3\]
\[X_1 = 1@6\]
\[X_{11} = 0@3\]
\( \neg X_1 \) was implied literal, leading to another conflict described below.

Conflict clause: \((X_9 + X_{10} + X_{11} + \neg X_{12} + \neg X_{13})\)

BackJump strategy: Closest decision level (DL) \( \leq \) current DL for which conflict clause is unit. Undo \( \{X_{10} = 0@3\} \)
Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decide()

TopLevel Conflict?

Return UNSAT

BackJump()

- Restarts
  - Clear the Trail and start again
  - Start searching with a different variable order
  - Only Conflict Clause (CC) database is retained

- Forget: throw away less active learnt conflict clauses routinely
  - Routinely throw away very large CC
  - Logically CC are implied
  - Hence no loss in soundness/completeness
  - Time Savings: smaller DB means less work in propagation
  - Space savings
Modern CDCL SAT Solver Architecture

Why is SAT efficient?

Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Return UNSAT

Decide()

TopLevel Conflict?

BackJump()

- VSIDS branching heuristic and propagate (BCP)
- Conflict-Driven Clause-Learning (CDCL)
- Forget conflict clauses if DB goes too big
- BackJump
- Restarts
- All the above elements are needed for efficiency
- Deeper understanding lacking
- No predictive theory
Modern CDCL SAT Solver Architecture

Propagate(), Decide(), Analyze/Learn(), BackJump()

Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decide()

TopLevel Conflict?

Return UNSAT

BackJump()

- Conflict-Driven Clause-Learning (CDCL)
  (Marques-Silva & Sakallah 1996)

- Decide/branch and propagate (BCP)
  (Malik et al. 2001, Zabih & McAllester 1988)

- BackJump
  (McAllester 1980, Marques-Silva & Sakallah 1999)

- Restarts
  (Selman & Gomes 2001)

- Follows MiniSAT
  (Een & Sorensson 2003)
Soundness: A solver is said to be sound, if, for any input formula F, the solver terminates and produces a solution, then F is indeed SAT.

Proof: (Easy) SAT is returned only when all vars have been assigned a value (True, False) by Decide or BCP, and solver checks the solution.
Completeness: A solver is said to be complete, if, for any input formula $F$ that is SAT, the solver terminates and produces a solution (i.e., solver does not miss solutions).

Proof: (Harder)
- Backtracking + BCP + decide is complete (easy)
- Conflict clause is implied by input formula (easy)
- Only need to see backjumping does not skip assignments
  - Observe backjumping occurs only when conflict clause (CC) vars $<$ decision level (DL) of conflicting var
  - Backjumping to max(DL of vars in CC)
  - Decision tree rooted at max(DL of vars in CC)+1 is guaranteed to not satisfy CC
  - Hence, backjumping will not skip assignments
Modern CDCL SAT Solver Architecture
Soundness, Completeness & Termination

Input SAT Instance

Propagate() (BCP)

No Conflict?

All Vars Assigned?

Conflict Analysis()

Return SAT

Decision()

TopLevel Conflict?

Return UNSAT

BackJump()

Termination: Some measure decreases every iteration

Proof Sketch:

- Loop guarantees either conflict clause (CC) added OR assign extended

- CC added. What stops CC addition looping forever?
  - Recall that CC is remembered
  - No CC duplication possible
  - CC blocks UNSAT assign exploration in decision tree. No duplicate UNSAT assign exploration possible
  - Size of decision tree explored decreases for each CC add
References & Important SAT Solvers


7. zChaff SAT Solver by Lintao Zhang 2002.


9. MiniSAT Solver by Niklas Een and Niklas Sorenson 2005 - present

10. SAT Live: http://www.satlive.org/

11. SAT Competition: http://www.satcompetition.org/

1. SAT solvers are crucial for software engineering

2. Huge impact in formal methods, program analysis and testing

3. Key ideas that make SAT efficient
   1. Conflict-driven clause learning
   2. VSIDS (or similar) variable selection heuristics
   3. Backjumping
   4. Restarts

4. Techniques I didn’t discuss
   1. Survey propagation (belief propagation) by Selman & Gomes
   2. Works well for randomized SAT, not yet for industrial instances
   3. Physics-inspired
   4. Combining CDCL with survey propagation (?)