

Abstract Constraint Programming on GPU

TALK AT ZHEJIANG UNIVERSITY, ZLAIRE, PROF. LIAO GROUP

Pierre Talbot

pierre.talbot@uni.lu

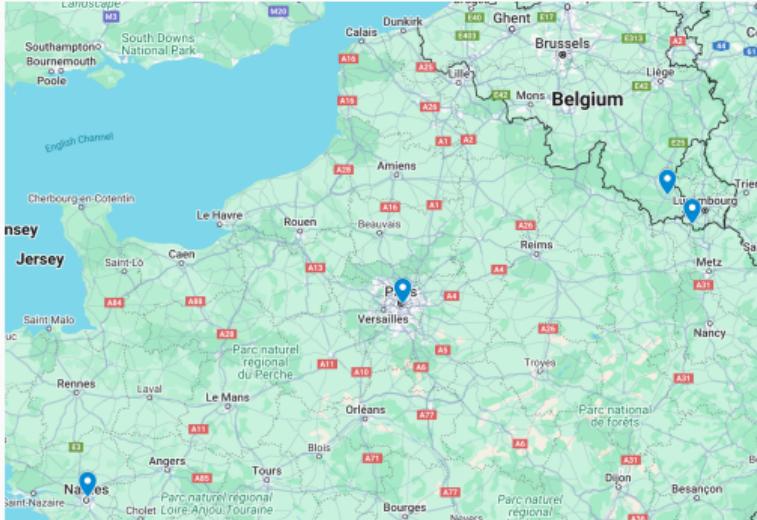
<https://ptal.github.io>

7th May 2025

University of Luxembourg



Who am I?



- 2014–2018: Ph.D., Sorbonne University, Paris
 - ▶ *Spacetime Programming: A Synchronous Language for Constraint Search.*
 - 2018–2019: Postdoc, University of Nantes.
 - ▶ *Abstract Domains for Constraint Programming.*
 - 2020–2023: Postdoc, University of Luxembourg
 - ▶ *A Lattice-Based Approach for GPU Programming.*
 - 2023–: Research scientist, University of Luxembourg.
 - ▶ *Abstract Satisfaction.*

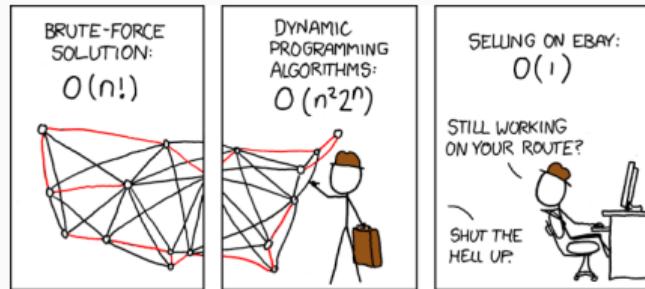
My Research in a Nutshell!

I research on the “fusion” of...

Constraint reasoning

+

Abstract interpretation
(and lattice theory)



that gives us **abstract satisfaction**.

My Research in a Nutshell!

I research on the “fusion” of...

WHY?

Accelerate constraint solving

HOW?

- Combining constraint solvers
- Constructing sound solving procedure over complex domains
- Constraint solving on GPUs

that gives us **abstract satisfaction**.

My Research in a Nutshell!

I research on the “fusion” of...

WHY?

Accelerate constraint solving

HOW?

- Combining constraint solvers
- Constructing sound solving procedure over complex domains
- Constraint solving on GPUs \Leftarrow TODAY

that gives us **abstract satisfaction**.

TEAM

I have the pleasure to co-supervise and work with several Ph.D. students and postdocs.



Pierre Talbot



Thibault Falque



Hedieh Haddad



Manuel Combarro



Yi-Nung Tsao



Tobias Fischbach



Angelica Rings



Hakan Hasan

- Thibault Falque, postdoctoral researcher on the project COMOC, 2024-2025.
- Hedieh Haddad, Ph.D. candidate, *Hyperparameter Optimization of Constraint Solver*, 2023-2025.
- Manuel Combarro, Ph.D. candidate, *Multiobjective Constraint Programming*, 2023-2026.
- Yi-Nung Tsao, Ph.D. candidate, *Verification of Neural Networks by Abstract Interpretation*, 2023-2026.
- Tobias Fischbach, Ph.D. candidate, *Optimization of Quantum Circuits*, 2023-2026.
- Angelica Rings, Bachelor student, *A Musical Game to Understand the Challenge of Collaboration in Computing*, 2024-2025.
- Hasan Hakan, Master student, *Multi-GPU Parallel Constraint Programming*, 2025.

Study Programmes

Master in High Performance Computing

[Overview](#)[Programme](#)[Career](#)[Testimonials](#)[Teaching staff](#)[Admissions](#)

Your outstanding career in high-performance computing

The Master in High Performance Computing (MHPC) is a distinctive programme at the intersection of parallel programming, hardware architecture, and artificial intelligence. We are training the next generation of HPC experts in Luxembourg and Europe. Besides to the MHPC, the EUMaster4HPC is another programme where students earn a dual degree from two of the eight universities of the EUMaster4HPC consortium. EUMaster4HPC has a different application procedure, so be sure to check out the dedicated website.



On the Menu

- **Abstract Satisfaction** (connection between logic and constraint reasoning)
- **Abstract Constraint Programming** (contribution, expressive reasoning framework)
- **Abstract Constraint Programming on GPU** (contribution, efficient reasoning framework)

Abstract Satisfaction

Syntax of First-Order Logic (FOL)

Let $S = \langle X, F, P \rangle$ be a *first-order signature* where X set of variables, F set of function symbols and P set of predicate symbols.

$$\langle t \rangle ::= x \quad \text{variable } x \in X \\ | \quad f(t, \dots, t) \quad \text{function } f \in F$$

$$\begin{aligned} \langle \varphi \rangle ::= & \ p(t, \dots, t) \quad \text{predicate } p \in P \\ | \quad & \neg \varphi \quad \text{negation} \\ | \quad & \varphi \diamond \varphi \quad \text{connector } \diamond \in \{\wedge, \vee, \Rightarrow, \Leftrightarrow\} \\ | \quad & \exists x, \ \varphi \quad \text{existential quantifier} \\ | \quad & \forall x, \ \varphi \quad \text{universal quantifier} \end{aligned}$$

Let Φ the set of well-formed formulas.

Semantics of FOL

A *structure A* is a tuple $(\mathbb{U}, \llbracket \cdot \rrbracket_F, \llbracket \cdot \rrbracket_P)$ where

1. \mathbb{U} is a non-empty set of elements—called the *universe of discourse*,
2. $\llbracket \cdot \rrbracket_F$ is a function mapping function symbols $f \in F$ with arity n to interpreted functions $\llbracket f \rrbracket_F : \mathbb{U}^n \rightarrow \mathbb{U}$, and
3. $\llbracket \cdot \rrbracket_P$ is a function mapping predicate symbols $p \in P$ with arity n to interpreted predicates $\llbracket p \rrbracket_P \subseteq \mathbb{U}^n$.

An *assignment* is a function $X \rightarrow \mathbb{U}$ mapping variables to values. We denote the set of assignment by **Asn**. Let $\rho \in \mathbf{Asn}$, we write $\rho[x \mapsto d]$ the assignment in which we updated the value of x by d in ρ .

Entailment

The syntax and semantics are related by the ternary relation $A \vDash_{\rho} \varphi$, called the *entailment*, where A is a structure, $\rho \in \mathbf{Asn}$ and $\varphi \in \Phi$. It is read as “the formula φ is satisfied by the assignment ρ in the structure A ”. We first give the interpretation function $\llbracket \cdot \rrbracket_{\rho}$ for evaluating the terms of the language:

$$\begin{aligned}\llbracket x \rrbracket_{\rho} &= \rho(x) \text{ if } x \in X \\ \llbracket f(t_1, \dots, t_n) \rrbracket_{\rho} &= \llbracket f \rrbracket_F(\llbracket t_1 \rrbracket_{\rho}, \dots, \llbracket t_n \rrbracket_{\rho})\end{aligned}$$

The relation \vDash is defined inductively as follows:

$A \vDash_{\rho} p(t_1, \dots, t_n)$	iff $(\llbracket t_1 \rrbracket_{\rho}, \dots, \llbracket t_n \rrbracket_{\rho}) \in \llbracket p \rrbracket_P$
$A \vDash_{\rho} \varphi_1 \wedge \varphi_2$	iff $A \vDash_{\rho} \varphi_1$ and $A \vDash_{\rho} \varphi_2$
$A \vDash_{\rho} \varphi_1 \vee \varphi_2$	iff $A \vDash_{\rho} \varphi_1$ or $A \vDash_{\rho} \varphi_2$
$A \vDash_{\rho} \neg \varphi$	iff $A \vDash_{\rho} \varphi$ does not hold
$A \vDash_{\rho} \exists x, \varphi$	iff there exists $d \in \mathbb{U}$ such that $A \vDash_{\rho[x \mapsto d]} \varphi$
$A \vDash_{\rho} \forall x, \varphi$	iff for all $d \in \mathbb{U}$, we have $A \vDash_{\rho[x \mapsto d]} \varphi$

Concrete Domain

Given a structure A , we define the *concrete interpretation function* as:

$$\begin{aligned}\llbracket . \rrbracket^b : \Phi &\rightarrow \mathcal{P}(\mathbf{Asn}) \\ \llbracket \varphi \rrbracket^b &= \{\rho \in \mathbf{Asn} \mid A \models_\rho \varphi\}\end{aligned}$$

- We call the *concrete domain* the set $\mathcal{P}(\mathbf{Asn})$ with $\llbracket . \rrbracket^b$.
- A *solution* of the formula φ is an assignment $s \in \llbracket \varphi \rrbracket^b$.
- **Example** in the theory of standard integer arithmetics (and $X = \{x, y\}$):

$$\begin{aligned}\llbracket x < y \wedge x \geq 0 \rrbracket^b &= \{ \\ &\quad \{x \mapsto 0, y \mapsto 1\} \\ &\quad \{x \mapsto 0, y \mapsto 2\} \\ &\quad \dots \\ &\quad \{x \mapsto 1, y \mapsto 2\} \\ &\quad \dots \\ &\quad \}\end{aligned}$$

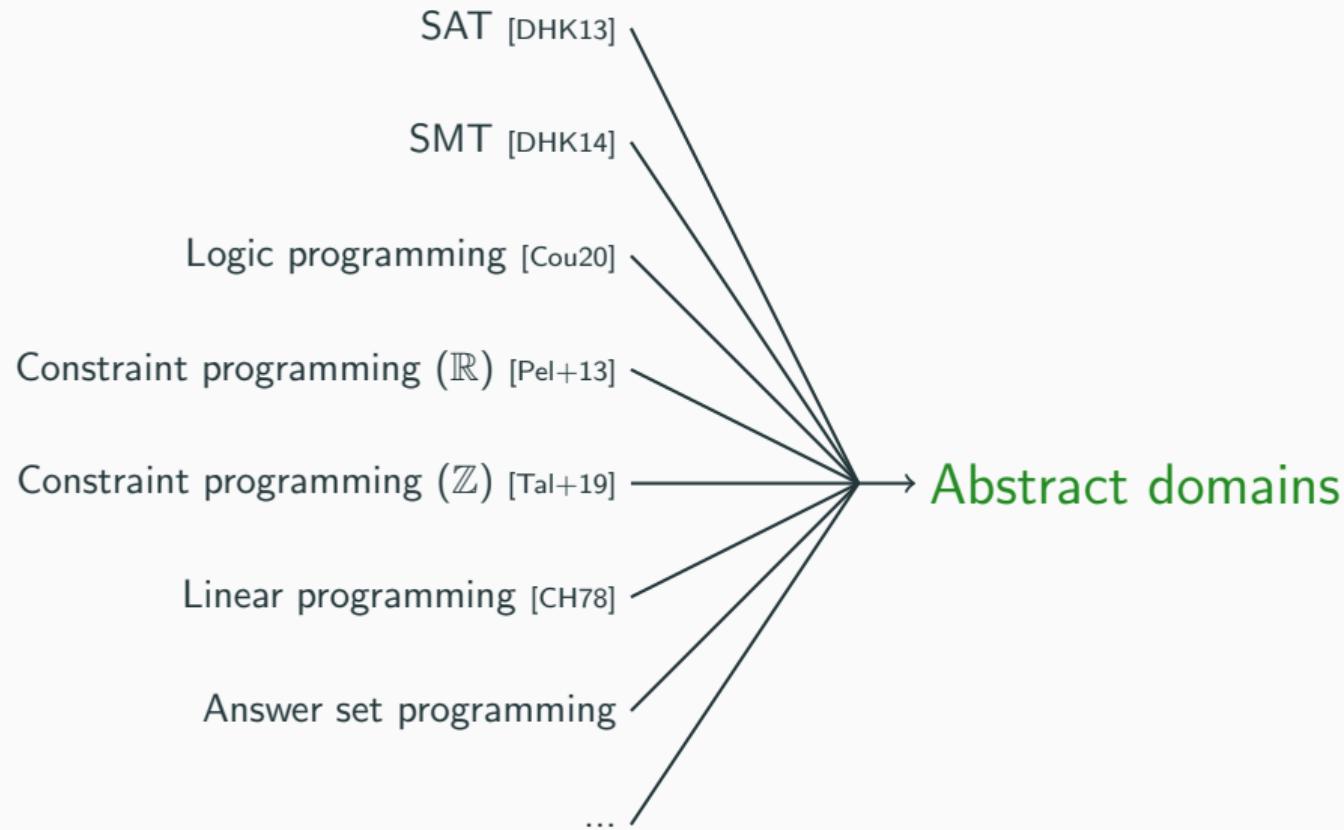
One Problem, Many Communities, Many Formalisms

Many communities emerged to solve the same problem: find ρ such that $A \models_{\rho} \varphi$.

BUT they (generally) focus on different fragments of FOL:

- Propositional fragment (SAT): $(a \vee b) \wedge (\neg b \vee c)$ with $a, b, c \in \{0, 1\}$.
- Pseudo-Boolean fragment: $\sum_{1 \leq i \leq n} c_i * a_i \leq c_0$ with $a_i \in \{0, 1\}$ and c_i some integers constants.
- Linear programming (LP): $\sum_{1 \leq i \leq n} c_i * b_i \leq b_0$ with $b_i \in \mathbb{R}$ and c_i some real constants.
- Integer linear programming (ILP): $\sum_{1 \leq i \leq n} c_i * b_i \leq b_0$ with $b_i \in \mathbb{Z}$ and c_i some integer constants.
- Mixed integer linear programming (MILP): $\sum_{1 \leq i \leq n} c_i * b_i \leq b_0$ with $b_i \in \mathbb{Z}$ or $b_i \in \mathbb{R}$ and c_i some integer or real constants.
- Uninterpreted fragment (logic programming).
- Discrete constraint programming: $\langle X, D, C \rangle$ with $D_i \in \mathcal{P}_f(\mathbb{Z})$.
- Continuous constraint programming: $\langle X, D, C \rangle$ with $D_i \in \mathcal{I}(\mathbb{R})$.
- Satisfiability modulo theories (SMT).
- ...

One Theory to Rule Them All?



*What is an **abstract domain**?*

It is a lattice with some operations.

What is an *abstract domain*?

It is a lattice with some operations.

What is a *lattice*?

A tuple $\langle S, \sqsubseteq, \sqcup, \sqcap, \perp, \top \rangle$ where S is a set.

What is an *abstract domain*?

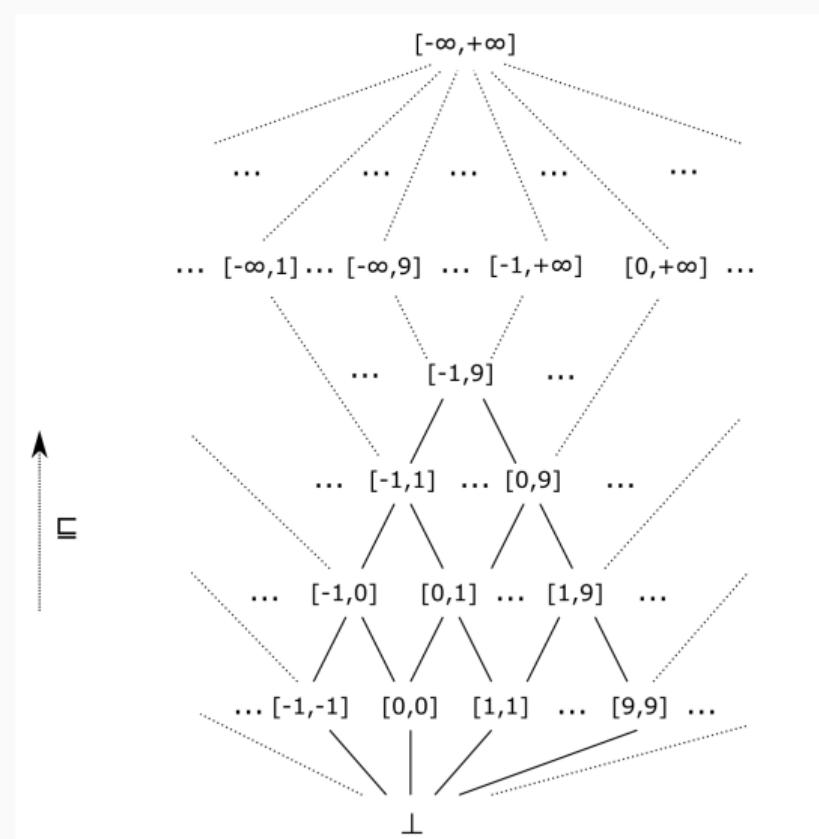
It is a lattice with some operations.

What is a *lattice*?

A tuple $\langle S, \sqsubseteq, \sqcup, \sqcap, \perp, \top \rangle$ where S is a set.

Example: Interval Lattice

- $S \triangleq \{[a, b] \mid a \in \mathbb{Z} \cup \{-\infty\}, b \in \mathbb{Z} \cup \{\infty\}, a \leq b\} \cup \{\perp\}$
- $[a, b] \sqsubseteq [c, d] \Leftrightarrow a \geq c \wedge b \leq d$
- $\top \triangleq [-\infty, \infty]$
- $[a, b] \sqcap [c, d] \triangleq [\max\{a, c\}, \min\{b, d\}]$



Simple Logic of Intervals

- **Logic:** $\Phi \triangleq x \leq k \mid x \geq k \mid \Phi \wedge \Phi \mid \Phi \vee \Phi.$ (only 1 variable)

Simple Logic of Intervals

- **Logic:** $\Phi \triangleq x \leq k \mid x \geq k \mid \Phi \wedge \Phi \mid \Phi \vee \Phi.$ (only 1 variable)
- **Abstract interpretation:**
 - $\llbracket x \leq k \rrbracket \triangleq [-\infty, k]$
 - $\llbracket x \geq k \rrbracket \triangleq [k, \infty]$
 - $\llbracket \varphi \wedge \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcap \llbracket \varphi' \rrbracket$
 - $\llbracket \varphi \vee \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcup \llbracket \varphi' \rrbracket$

Simple Logic of Intervals

- **Logic:** $\Phi \triangleq x \leq k \mid x \geq k \mid \Phi \wedge \Phi \mid \Phi \vee \Phi.$ (only 1 variable)
- **Abstract interpretation:**
 - $\llbracket x \leq k \rrbracket \triangleq [-\infty, k]$
 - $\llbracket x \geq k \rrbracket \triangleq [k, \infty]$
 - $\llbracket \varphi \wedge \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcap \llbracket \varphi' \rrbracket$
 - $\llbracket \varphi \vee \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcup \llbracket \varphi' \rrbracket$
- **Example:**
 - $\llbracket (x \leq 10 \wedge x \geq 0) \vee (x \geq 5) \rrbracket$
 - $\llbracket x \leq 10 \wedge x \geq 0 \rrbracket \sqcup \llbracket x \geq 5 \rrbracket$
 - $(\llbracket x \leq 10 \rrbracket \sqcap \llbracket x \geq 0 \rrbracket) \sqcup \llbracket x \geq 5 \rrbracket$
 - $([-\infty, 10] \sqcap [0, \infty]) \sqcup [5, \infty]$
 - $[0, 10] \sqcup [5, \infty]$
 - $[0, \infty]$

Simple Logic of Intervals

- **Logic:** $\Phi \triangleq x \leq k \mid x \geq k \mid \Phi \wedge \Phi \mid \Phi \vee \Phi.$ (only 1 variable)

- **Abstract interpretation:**

- $\llbracket x \leq k \rrbracket \triangleq [-\infty, k]$
- $\llbracket x \geq k \rrbracket \triangleq [k, \infty]$
- $\llbracket \varphi \wedge \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcap \llbracket \varphi' \rrbracket$
- $\llbracket \varphi \vee \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcup \llbracket \varphi' \rrbracket$

- **Example:**

- $\llbracket (x \leq 10 \wedge x \geq 0) \vee (x \geq 5) \rrbracket$
- $\llbracket x \leq 10 \wedge x \geq 0 \rrbracket \sqcup \llbracket x \geq 5 \rrbracket$
- $(\llbracket x \leq 10 \rrbracket \sqcap \llbracket x \geq 0 \rrbracket) \sqcup \llbracket x \geq 5 \rrbracket$
- $([-\infty, 10] \sqcap [0, \infty]) \sqcup [5, \infty]$
- $[0, 10] \sqcup [5, \infty]$
- $[0, \infty]$

- **Soundness:** $\llbracket \varphi \rrbracket^b \subseteq \llbracket \varphi \rrbracket$ (compute all solutions).

- **Completeness:** $\llbracket \varphi \rrbracket^b \supseteq \llbracket \varphi \rrbracket$ (compute only solutions).

Intervals are not complete: $\llbracket x \leq 10 \vee x \geq 15 \rrbracket = [-\infty, \infty]$ (intervals cannot represent “holes”).

What About Multiple Variables?

We lift interval to a function $X \rightarrow Itv$ mapping variables to intervals where Itv is the interval lattice.

Now, we can define (with $x \in X$ any variable):

- $\llbracket x \leq k \rrbracket \triangleq \{x \mapsto [-\infty, k]\}.$
- $\llbracket x \geq k \rrbracket \triangleq \{x \mapsto [k, \infty]\}.$
- ...

Example: $\llbracket x \leq 0 \wedge y \geq 0 \rrbracket = \{x \mapsto [-\infty, 0], y \mapsto [0, \infty]\}.$

What About Multiple Variables?

We lift interval to a function $X \rightarrow Itv$ mapping variables to intervals where Itv is the interval lattice.

Now, we can define (with $x \in X$ any variable):

- $\llbracket x \leq k \rrbracket \triangleq \{x \mapsto [-\infty, k]\}.$
- $\llbracket x \geq k \rrbracket \triangleq \{x \mapsto [k, \infty]\}.$
- ...

Example: $\llbracket x \leq 0 \wedge y \geq 0 \rrbracket = \{x \mapsto [-\infty, 0], y \mapsto [0, \infty]\}.$

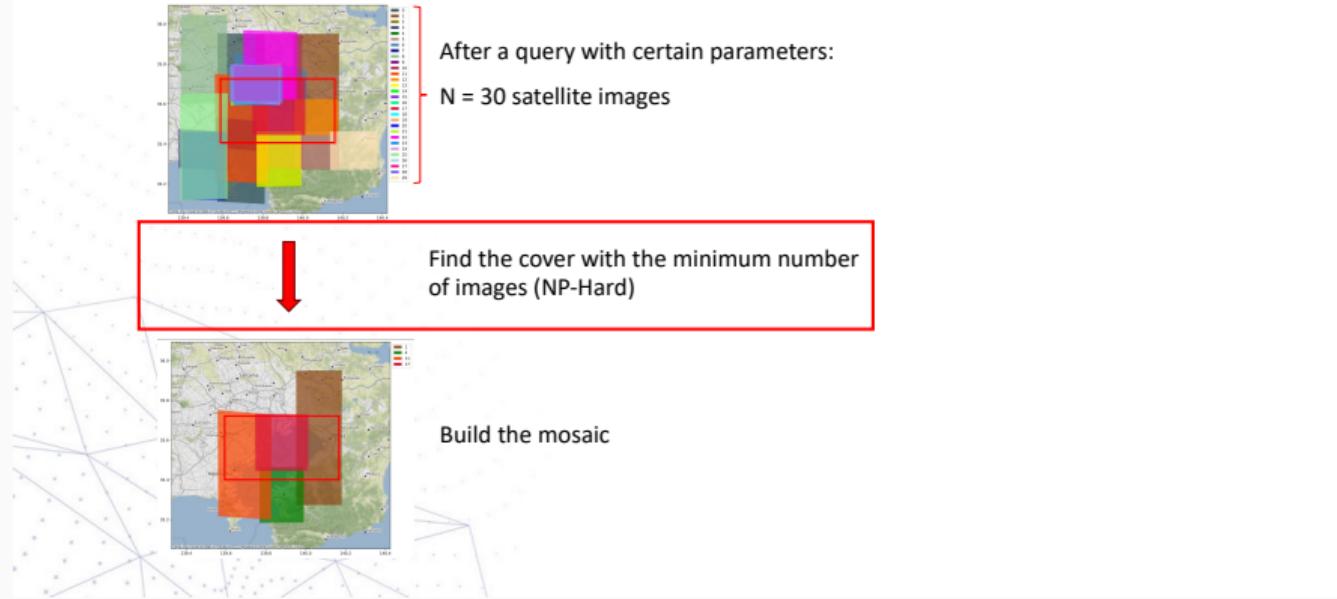
How to compute solutions of more expressive logic?

Abstract Constraint Programming

Constraint programming: FOL without quantifiers, $\mathbb{U} = \mathbb{Z}$ and arithmetic constraints.

- **Declarative paradigm:** specify your problem and let the computer solves it for you.
- **Many applications:** scheduling, bin-packing, hardware design, satellite imaging, ...
- **Constraint programming** is one approach to solve such combinatorial problems.
- Other approaches include SAT, linear programming, SMT, MILP, ASP,...

Satellite image mosaic



1

¹Combarro et al., Constraint Model for the Satellite Image Mosaic Selection Problem, CP 2023

Constraint model of satellite imaging in MiniZinc:

The screenshot shows the MiniZinc IDE interface. At the top is a toolbar with icons for New model, Open, Save, Copy, Cut, Paste, Undo, Redo, Shift left, Shift right, and Run. The Run button is currently selected. To the right of the Run button is a dropdown menu for Solver configuration, set to Gecode 6.3.0. Below the toolbar, there are two tabs: 'satellite.mzn' and 'satellite1.dzn'. The code editor displays the following MiniZinc code:

```
4 int: universe;
5
6 set of int: IMAGES = 1..images;
7 set of int: UNIVERSE = 1..universe;
8
9 array[IMAGES] of set of int: sets;
10 array[IMAGES] of int: costs;
11
12 constraint forall(u in UNIVERSE)(
13   exists(i in IMAGES)(taken[i] /\ u in sets[i]));
14
15 array[IMAGES] of var bool: taken;
16
17 solve minimize sum(i in IMAGES)(costs[i] * taken[i]);
```

Below the code editor is an 'Output' section. It contains three buttons: 'Hide all', 'dzn', and 'default', with 'dzn' being the active button. The output window shows the results of running the solver:

```
Running satellite.mzn, satellite1.dzn
taken = [true, true, false, true, true, false];
-----
=====
Finished in 114msec.
```

Constraint Network

Constraint Network

Let X be a finite set of variables and C be a finite set of constraints.

A *constraint network* is a pair $P = \langle d, C \rangle$ such that $d \in X \rightarrow Itv$ is the *domain* of the variables where Itv is the set of intervals.

Note: It is just a "format" to represent quantifier-free logical formulas where variables have bounded domains.

Example

$$\langle \{x \mapsto [0, 2], y \mapsto [2, 3]\}, \{x \leq y - 1\} \rangle$$

A solution is $\{x \mapsto 0, y \mapsto 2\}$.

Constraints with Multiple Variables

- We already have: $\llbracket x \leq k \rrbracket \triangleq \{x \mapsto [-\infty, k]\}.$
- Also: $\llbracket x = k \rrbracket \triangleq \{x \mapsto [k, k]\}.$

How to interpret $\llbracket x = y \rrbracket$?

Constraints with Multiple Variables

- We already have: $\llbracket x \leq k \rrbracket \triangleq \{x \mapsto [-\infty, k]\}$.
- Also: $\llbracket x = k \rrbracket \triangleq \{x \mapsto [k, k]\}$.

How to interpret $\llbracket x = y \rrbracket$?

- Unfortunately, without more information on x and y , we must set $\llbracket x = y \rrbracket \triangleq \{x \mapsto [-\infty, \infty], y \mapsto [-\infty, \infty]\}$, which is the same than ignoring the constraint...

Constraints with Multiple Variables

- We already have: $\llbracket x \leq k \rrbracket \triangleq \{x \mapsto [-\infty, k]\}$.
- Also: $\llbracket x = k \rrbracket \triangleq \{x \mapsto [k, k]\}$.

How to interpret $\llbracket x = y \rrbracket$?

- Unfortunately, without more information on x and y , we must set $\llbracket x = y \rrbracket \triangleq \{x \mapsto [-\infty, \infty], y \mapsto [-\infty, \infty]\}$, which is the same than ignoring the constraint...

Solution: give more information to the interpretation function.

- $\mathcal{I}[\cdot] \in \Phi \times (X \rightarrow \text{Itv}) \rightarrow (X \rightarrow \text{Itv})$
- $\mathcal{I}[x = y]d \triangleq \{x \mapsto d(x) \sqcap d(y), y \mapsto d(x) \sqcap d(y)\}$

Example: Let $d = \{x \mapsto [0, 5], y \mapsto [5, 10]\}$, then $\mathcal{I}[x = y]d = \{x \mapsto [5, 5], y \mapsto [5, 5]\}$.

How to Deal with Conjunction?

- Before, we had $\llbracket \varphi \wedge \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcap \llbracket \varphi' \rrbracket$.
- Now, we can lift this function to

$$\mathcal{I}[\varphi \wedge \varphi']d \triangleq \mathcal{I}[\varphi]d \sqcap \mathcal{I}[\varphi']d$$

How to Deal with Conjunction?

- Before, we had $\llbracket \varphi \wedge \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcap \llbracket \varphi' \rrbracket$.
- Now, we can lift this function to

$$\mathcal{I}[\varphi \wedge \varphi']d \triangleq \mathcal{I}[\varphi]d \sqcap \mathcal{I}[\varphi']d$$

Problem: d must be copied... inefficient

How to Deal with Conjunction?

- Before, we had $\llbracket \varphi \wedge \varphi' \rrbracket \triangleq \llbracket \varphi \rrbracket \sqcap \llbracket \varphi' \rrbracket$.
- Now, we can lift this function to

$$\mathcal{I}[\varphi \wedge \varphi']d \triangleq \mathcal{I}[\varphi]d \sqcap \mathcal{I}[\varphi']d$$

Problem: d must be copied... inefficient

- Instead, we can use functional composition:

$$\mathcal{I}[\varphi \wedge \varphi']d \triangleq (\mathcal{I}[\varphi] \circ \mathcal{I}[\varphi'])d$$

Computing Solutions of Constraint Network

A constraint network $\langle d, C \rangle$ is a conjunctive collection of constraints. So we can compute the set of solutions using:

$$\mathcal{I}[\![c_1 \wedge c_2 \wedge \dots \wedge c_n]\!] = \mathcal{I}[\![c_1]\!] \circ \mathcal{I}[\![c_2]\!] \circ \dots \circ \mathcal{I}[\![c_n]\!]$$

Example: Let $\langle d, \{x = y, y = z\} \rangle$ be a constraint network with $d = \{x \mapsto [2, 2], y \mapsto [1, 2], z \mapsto [0, 2]\}$, then:

$$\begin{aligned} & \mathcal{I}[\![x = y \wedge y = z]\!]d \\ = & (\mathcal{I}[\![x = y]\!] \circ \mathcal{I}[\![y = z]\!])d \\ = & \mathcal{I}[\![x = y]\!](\mathcal{I}[\![y = z]\!](d)) \\ = & \mathcal{I}[\![x = y]\!](\{x \mapsto [2, 2], y \mapsto [1, 2], z \mapsto [0, 2]\}) \\ = & \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [1, 2]\} \end{aligned}$$

Computing Solutions of Constraint Network

A constraint network $\langle d, C \rangle$ is a conjunctive collection of constraints. So we can compute the set of solutions using:

$$\mathcal{I}[\![c_1 \wedge c_2 \wedge \dots \wedge c_n]\!] = \mathcal{I}[\![c_1]\!] \circ \mathcal{I}[\![c_2]\!] \circ \dots \circ \mathcal{I}[\![c_n]\!]$$

Example: Let $\langle d, \{x = y, y = z\} \rangle$ be a constraint network with $d = \{x \mapsto [2, 2], y \mapsto [1, 2], z \mapsto [0, 2]\}$, then:

$$\begin{aligned} & \mathcal{I}[\![x = y \wedge y = z]\!]d \\ = & (\mathcal{I}[\![x = y]\!] \circ \mathcal{I}[\![y = z]\!])d \\ = & \mathcal{I}[\![x = y]\!](\mathcal{I}[\![y = z]\!](d)) \\ = & \mathcal{I}[\![x = y]\!](\{x \mapsto [2, 2], y \mapsto [1, 2], z \mapsto [1, 2]\}) \\ = & \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [1, 2]\} \end{aligned}$$

We are not very precise... $z = [1, 2]$ instead of $z = [2, 2]$.

Computing the Greatest Fixpoint

- $d_1 = \{x \mapsto [2, 2], y \mapsto [1, 2], z \mapsto [0, 2]\}.$
- $d_2 = \mathcal{I}[\![x = y \wedge y = z]\!]d_1 = \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [1, 2]\}.$

Computing the Greatest Fixpoint

- $d_1 = \{x \mapsto [2, 2], y \mapsto [1, 2], z \mapsto [0, 2]\}.$
- $d_2 = \mathcal{I}[\![x = y \wedge y = z]\!]d_1 = \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [1, 2]\}.$
- **More precision?** We can apply the function again!
- $d_3 = \mathcal{I}[\![x = y \wedge y = z]\!]d_2 = \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [2, 2]\}.$

Computing the Greatest Fixpoint

- $d_1 = \{x \mapsto [2, 2], y \mapsto [1, 2], z \mapsto [0, 2]\}.$
- $d_2 = \mathcal{I}[\![x = y \wedge y = z]\!]d_1 = \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [1, 2]\}.$
- **More precision?** We can apply the function again!
- $d_3 = \mathcal{I}[\![x = y \wedge y = z]\!]d_2 = \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [2, 2]\}.$
- **Again?** $d_3 = \mathcal{I}[\![x = y \wedge y = z]\!]d_3$, nothing changed! We reached a *fixpoint*.

Computing the Greatest Fixpoint

- $d_1 = \{x \mapsto [2, 2], y \mapsto [1, 2], z \mapsto [0, 2]\}.$
- $d_2 = \mathcal{I}[\![x = y \wedge y = z]\!]d_1 = \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [1, 2]\}.$
- **More precision?** We can apply the function again!
- $d_3 = \mathcal{I}[\![x = y \wedge y = z]\!]d_2 = \{x \mapsto [2, 2], y \mapsto [2, 2], z \mapsto [2, 2]\}.$
- **Again?** $d_3 = \mathcal{I}[\![x = y \wedge y = z]\!]d_3$, nothing changed! We reached a *fixpoint*.

For all formulas φ , $\mathcal{I}[\![\varphi]\!]$ is a *monotone function*.

Hence, we are guaranteed to find the greatest fixpoint, which is unique (Tarski theorem).

Constraint propagation is an approach to compute efficiently the *greatest fixpoint*:

$$\mathbf{gfp}_d \mathcal{I}[\![c_1]\!] \circ \dots \circ \mathcal{I}[\![c_n]\!]$$

Propagate and Search

The main algorithm behind constraint solvers:

```
function SOLVE( $d, \{c_1, \dots, c_n\}$ )
   $d \leftarrow \text{gfp}_d \mathcal{I}[\![c_1]\!] \circ \dots \circ \mathcal{I}[\![c_n]\!]$ 
  if  $\forall x \in X, d(x) = [v, v]$  then return  $\{d\}$ 
  else if  $\exists x \in X, d(x) = \perp$  then return  $\{\}$ 
  else
     $\langle d_1, \dots, d_n \rangle \leftarrow \text{split}(d)$ 
    return  $\bigcup_{i=0}^n \text{solve}(d_i, C)$ 
  end if
end function
```

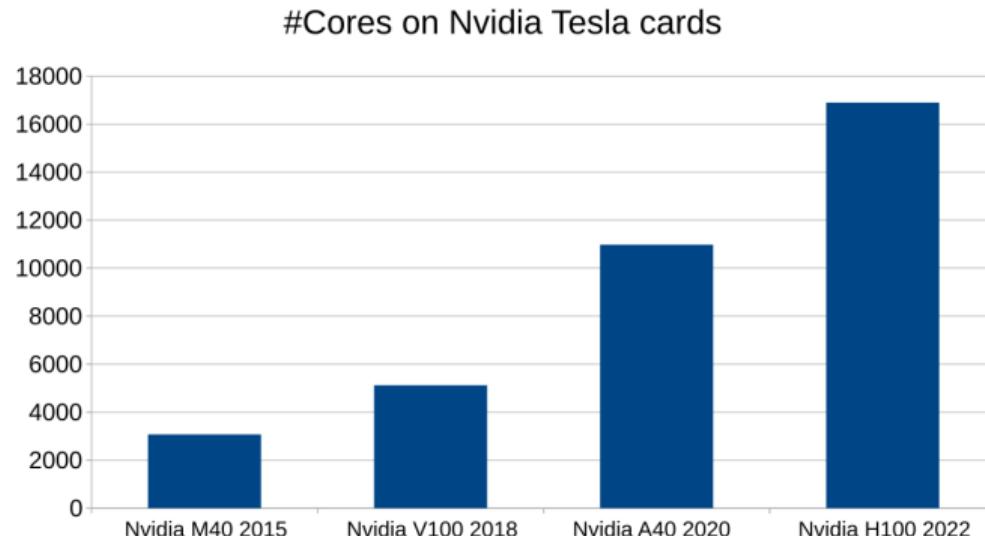
Thanks to the split function, the algorithm is **sound and complete**.

Constraint Programming on GPU

Why Constraint Programming on GPU?

Why CP on GPU?

CPU clock speed is stagnating, GPU #cores is increasing quickly each year.



Easy speed-up: same code but faster.

Why CP on GPU?

- Machine learning (deep learning, reinforcement learning, ...) has seen tremendous speed-ups (e.g. 100x, 1000x) by using GPU.
- Some (sequential) optimizations on CPU are made irrelevant if we can explore huge state space faster.

Can we replicate the success of GPU on machine learning applications to combinatorial optimization?

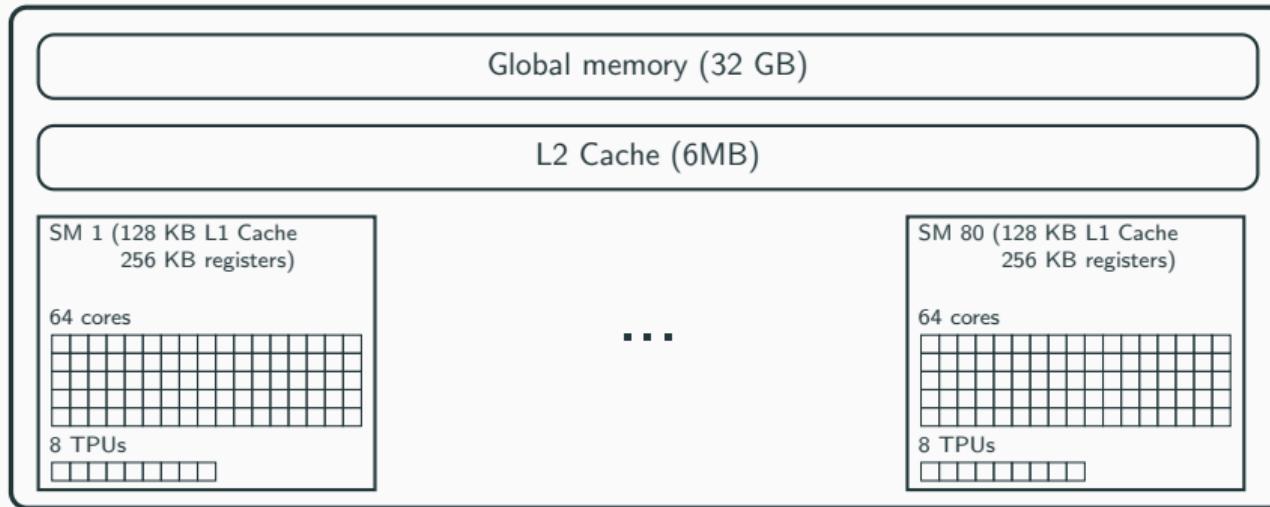
Constraint Programming on GPU

The rest of this talk:

- GPU Architecture.
- Challenges of Constraint Programming on GPU.
- Parallel Model of Computation.
- Ternary Constraint Network.

GPU Architecture

(Simplified) Architecture of the GPU Nvidia V100



5120 cores on a single V100 GPU @ 1290MHz

Whitepaper: <https://images.nvidia.com/content/volta-architecture/pdf/volta-architecture-whitepaper.pdf>

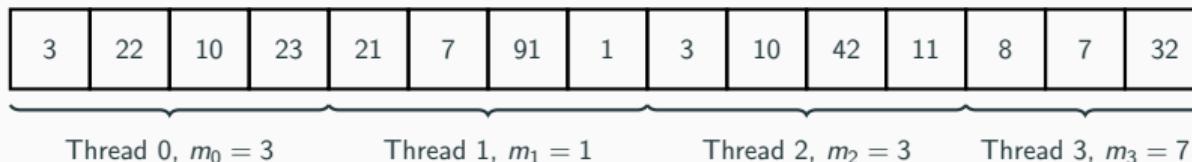
Programming Challenges

- **Memory coalescence:** the way to access the data is important (factor 10).
- **Thread divergence:** each thread within a warp (group of 32 threads) should execute the same instructions.
- **Memory allocation** (dynamic data structures): costly on GPU, everything is generally pre-allocated.
- **Other limitations:** small cache, limited number of lines of code, limited STL...

Example: Find the Minimum in an Array (CPU Style)

Each thread computes its local min (*map*), then we compute the min of all local min (*reduce*).

- Map:

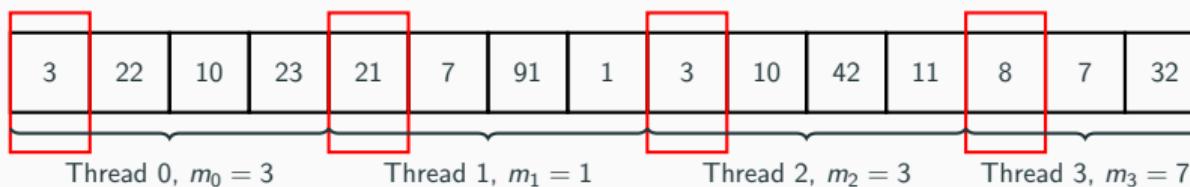


- Reduce: $\min([3, 1, 3, 7]) = 1$.

Example: Find the Minimum in an Array (CPU Style)

Iteration 1:

- Map:

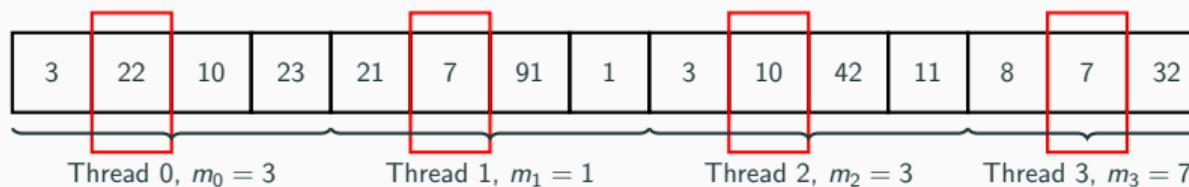


- Reduce: $\min([3, 1, 3, 7]) = 1$.

Example: Find the Minimum in an Array (CPU Style)

Iteration 2:

- Map:

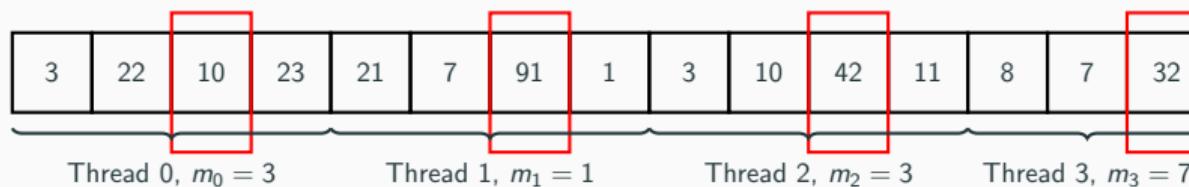


- Reduce: $\min([3, 1, 3, 7]) = 1$.

Example: Find the Minimum in an Array (CPU Style)

Iteration 3:

- Map:

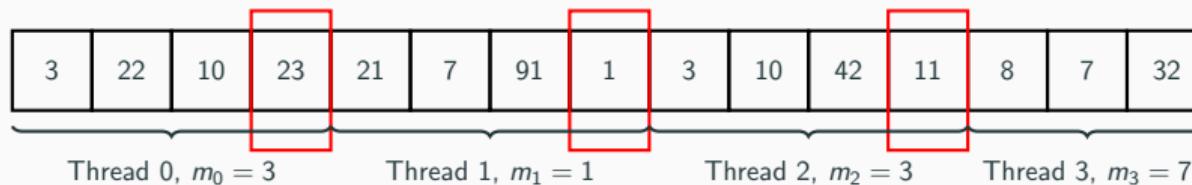


- Reduce: $\min([3, 1, 3, 7]) = 1$.

Example: Find the Minimum in an Array (CPU Style)

Iteration 4:

- Map:



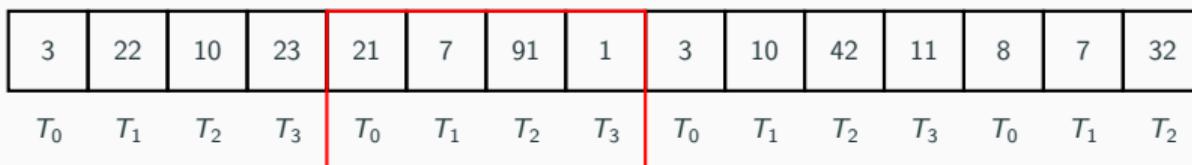
- Reduce: $\min([3, 1, 3, 7]) = 1$.

Optimization: Coalesced Memory Accesses

3	22	10	23	21	7	91	1	3	10	42	11	8	7	32
T_0	T_1	T_2	T_3	T_0	T_1	T_2	T_3	T_0	T_1	T_2	T_3	T_0	T_1	T_2

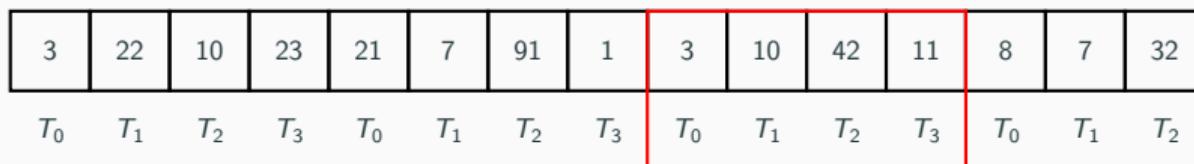
Very important: up to an order of magnitude faster (10x).

Optimization: Coalesced Memory Accesses



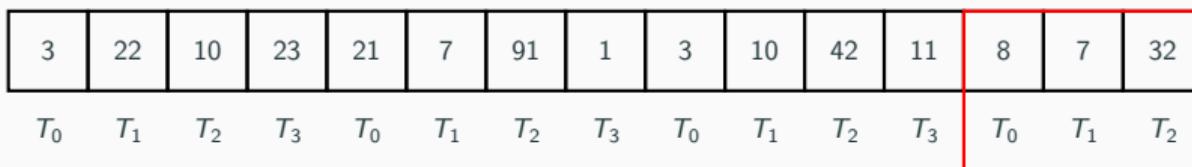
Very important: up to an order of magnitude faster (10x).

Optimization: Coalesced Memory Accesses



Very important: up to an order of magnitude faster (10x).

Optimization: Coalesced Memory Accesses

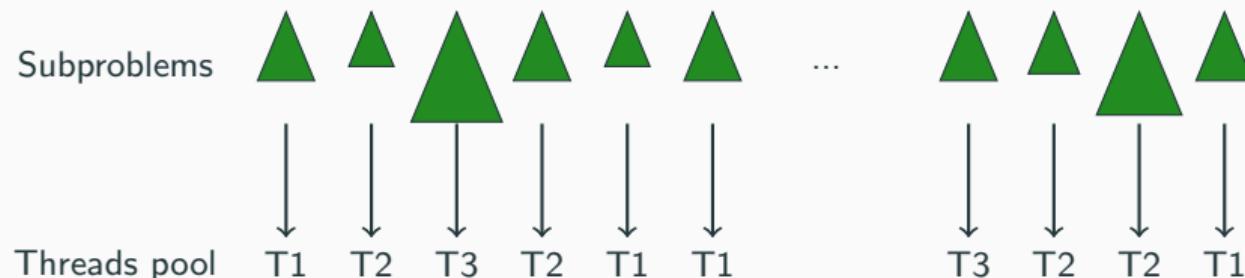


Very important: up to an order of magnitude faster (10x).

Challenges of Constraint Programming on GPU

On CPU: Embarrassingly Parallel Search (EPS)²

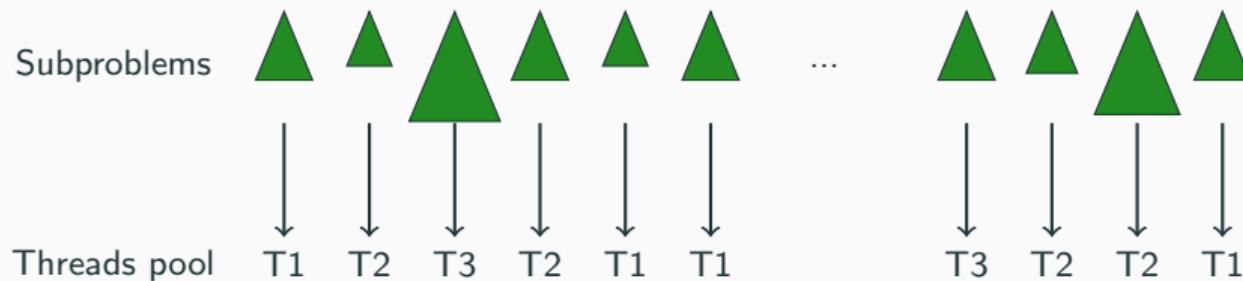
- **Idea:** Divide the problem into many subproblems beforehand (e.g. $N \times 30$ with N the number of threads).
- **Intuition:** Statistically, there is little chance a subproblem takes longer than the sum of the other subproblems.



²A. Malapert et al., 'Embarrassingly Parallel Search in Constraint Programming', JAIR, 2016

On CPU: Embarrassingly Parallel Search (EPS)

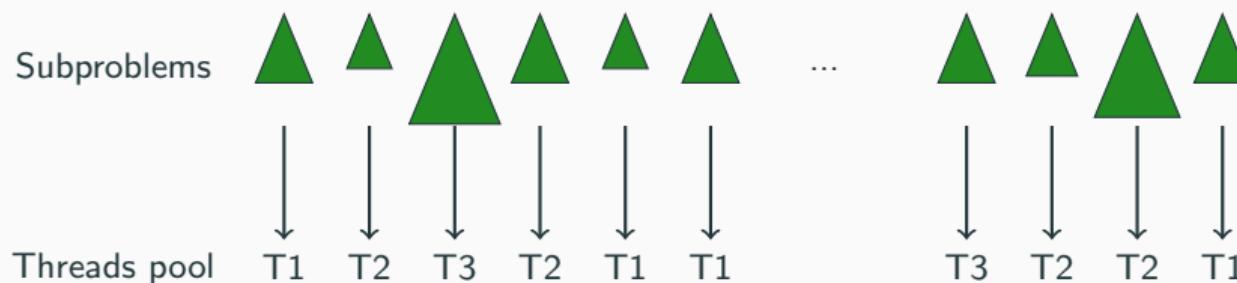
- **Idea:** Divide the problem into many subproblems beforehand (e.g. $N \times 30$ with N the number of threads).
- **Intuition:** Statistically, there is little chance a subproblem takes longer than the sum of the other subproblems.



⇒ **Other approach:** In modern solvers (e.g., Choco, OR-Tools), they use a portfolio approach (e.g., different *split* strategy on the *same problem*).

On CPU: Embarrassingly Parallel Search (EPS)

- **Idea:** Divide the problem into many subproblems beforehand (e.g. $N \times 30$ with N the number of threads).
- **Intuition:** Statistically, there is little chance a subproblem takes longer than the sum of the other subproblems.



On GPU architectures, 1 subproblem per thread is not efficient (limited cache).
⇒ Need to parallelize propagation: $\text{gfp}_d \mathcal{I}[\![c_1]\!] \circ \dots \circ \mathcal{I}[\![c_n]\!]$.

Where is the Challenge?

Parallelizing $\mathbf{gfp}_d \mathcal{I}[c_1] \circ \dots \circ \mathcal{I}[c_n]$ is challenging because constraints share variables, and we have typical *shared state memory* issues such data races and inefficiencies.

Contributions

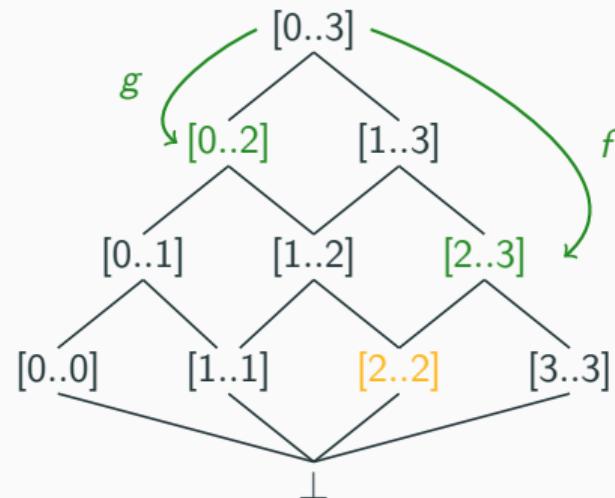
- **New parallel model of computation** to execute propagators in parallel²:
 $\mathbf{gfp}_d \mathcal{I}[c_1] \parallel \dots \parallel \mathcal{I}[c_n]$
- **Ternary constraint network**: representation of constraints suited for GPU architectures³.
- **First general constraint solver fully executing on GPU.**
⇒ **Open-source**: Publicly available on <https://github.com/ptal/turbo>.

²P. Talbot et al., *A Variant of Concurrent Constraint Programming on GPU*, AAAI, 2022.

³P. Talbot, *Ternary Constraint Network for Efficient Integer Bound Propagation on GPU*, submitted, 2025.

Parallel Model of Computation

Parallel Model of Computation



- $f(x) \triangleq x \sqcap [2..\infty]$ models the constraint $x \geq 2$.
- $g(x) \triangleq x \sqcap [-\infty..2]$ models the constraint $x \leq 2$.
- Parallel execution: $f \parallel g = [2..2]$

Example of Parallel Propagation

Let's consider $\mathcal{I}[x \leq 4 \wedge x \leq 5] = \mathcal{I}[x \leq 4] \parallel \mathcal{I}[x \leq 5]$

Memory:

$$x = [-\infty, \infty]$$

Propagators:

$$\begin{array}{ll} \boxed{x} \leftarrow [-\infty, 4] & (\mathcal{I}[x \leq 4]) \\ \parallel & \\ \boxed{x} \leftarrow [-\infty, 5] & (\mathcal{I}[x \leq 5]) \end{array}$$

Example of Parallel Propagation

Let's consider $\mathcal{I}[x \leq 4 \wedge x \leq 5] = \mathcal{I}[x \leq 4] \parallel \mathcal{I}[x \leq 5]$

Memory:

$$x = [-\infty, \boxed{?}]$$

Propagators:

$$\begin{array}{ll} \boxed{x} \leftarrow [-\infty, 4] & (\mathcal{I}[x \leq 4]) \\ \parallel & \\ \boxed{x} \leftarrow [-\infty, 5] & (\mathcal{I}[x \leq 5]) \end{array}$$

Issue 1: data race? Parallel update of the same variable: upper bound of x .

Example of Parallel Propagation

Let's consider $\mathcal{I}[x \leq 4 \wedge x \leq 5] = \mathcal{I}[x \leq 4] \parallel \mathcal{I}[x \leq 5]$

Memory:

$$x = [-\infty, 5]$$

Propagators:

$$\begin{array}{ll} x \leftarrow [-\infty, 4] & (\mathcal{I}[x \leq 4]) \\ \parallel & \\ x \leftarrow [-\infty, 5] & (\mathcal{I}[x \leq 5]) \end{array}$$

Issue 1: data race? Parallel update of the same variable: upper bound of x .

⇒ **Solution:** Use atomic load and store!

Example of Parallel Propagation

Let's consider $\mathcal{I}[x \leq 4 \wedge x \leq 5] = \mathcal{I}[x \leq 4] \parallel \mathcal{I}[x \leq 5]$

Memory:

$$x = [-\infty, 5]$$

Propagators:

$$\begin{array}{ll} x \leftarrow [-\infty, 4] & (\mathcal{I}[x \leq 4]) \\ \parallel & \\ x \leftarrow [-\infty, 5] & (\mathcal{I}[x \leq 5]) \end{array}$$

Issue 1: data race? Parallel update of the same variable: upper bound of x .

⇒ **Solution:** Use atomic load and store!

⇒ **In CUDA:** Integer load and store are atomic by default!

Example of Parallel Propagation

Let's consider $\mathcal{I}[x \leq 4 \wedge x \leq 5] = \mathcal{I}[x \leq 4] \parallel \mathcal{I}[x \leq 5]$

Memory:

$x = [-\infty, 5]$

Propagators:

$x \leftarrow [-\infty, 4] \quad (\mathcal{I}[x \leq 4])$
|| $x \leftarrow [-\infty, 5] \quad (\mathcal{I}[x \leq 5])$

Issue 1: data race? Parallel update of the same variable: upper bound of x .

⇒ **Solution:** Use atomic load and store!

⇒ **In CUDA:** Integer load and store are atomic by default!

Issue 2: nondeterminism? x can be equal to $[-\infty, 4]$ or $[-\infty, 5]$ depending on the order of execution.

Example of Parallel Propagation

Let's consider $\mathcal{I}[x \leq 4 \wedge x \leq 5] = \mathcal{I}[x \leq 4] \parallel \mathcal{I}[x \leq 5]$

Memory:

$$x = [-\infty, 4]$$

Propagators:

$$\begin{array}{ll} x \leftarrow [-\infty, 4] & (\mathcal{I}[x \leq 4]) \\ \parallel & \\ x \leftarrow [-\infty, 5] & (\mathcal{I}[x \leq 5]) \end{array}$$

Issue 1: data race? Parallel update of the same variable: upper bound of x .

⇒ **Solution:** Use atomic load and store!

⇒ **In CUDA:** Integer load and store are atomic by default!

Issue 2: nondeterminism? x can be equal to $[-\infty, 4]$ or $[-\infty, 5]$ depending on the order of execution.

⇒ **Solution:** Use a fixpoint loop.

Example of Parallel Propagation

Let's consider $\mathcal{I}[x \leq 4 \wedge x \leq 5] = \mathcal{I}[x \leq 4] \parallel \mathcal{I}[x \leq 5]$

Memory:

$$x = [-\infty, \boxed{5}]$$

Propagators:

$$\begin{array}{ll} x \leftarrow [-\infty, 4] & (\mathcal{I}[x \leq 4]) \\ \parallel & \\ x \leftarrow [-\infty, 5] & (\mathcal{I}[x \leq 5]) \end{array}$$

Issue 1: data race? Parallel update of the same variable: upper bound of x .

⇒ **Solution:** Use atomic load and store!

⇒ **In CUDA:** Integer load and store are atomic by default!

Issue 2: nondeterminism? x can be equal to $[-\infty, 4]$ or $[-\infty, 5]$ depending on the order of execution.

⇒ **Solution:** Use a fixpoint loop.

Issue 3: progress? What if $\mathcal{I}[x \leq 5]$ is always “winning”?

Example of Parallel Propagation

Let's consider $\mathcal{I}[x \leq 4 \wedge x \leq 5] = \mathcal{I}[x \leq 4] \parallel \mathcal{I}[x \leq 5]$

Memory:

$$x = [-\infty, 4]$$

Propagators:

$$\begin{array}{ll} x \leftarrow [-\infty, 4] & (\mathcal{I}[x \leq 4]) \\ \parallel & \\ x \leftarrow [-\infty, 5] & (\mathcal{I}[x \leq 5]) \end{array}$$

Issue 1: data race? Parallel update of the same variable: upper bound of x .

⇒ **Solution:** Use atomic load and store!

⇒ **In CUDA:** Integer load and store are atomic by default!

Issue 2: nondeterminism? x can be equal to $[-\infty, 4]$ or $[-\infty, 5]$ depending on the order of execution.

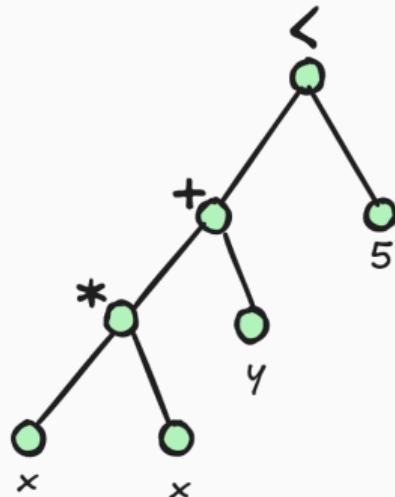
⇒ **Solution:** Use a fixpoint loop.

Issue 3: progress? What if $\mathcal{I}[x \leq 5]$ is always “winning”?

⇒ **Solution:** Write in the memory only if the value is strictly lower ($[x] = v$ iff $v < [x]$).

Ternary Constraint Network

Representation of Propagators



- Represented using `shared_ptr` and variant data structures.
⇒ Uncoalesced memory accesses.
- Code similar to an interpreter:

```
switch(term.index()) {  
    case IVar:  
    case INeg:  
    case IAdd:  
    case IMul:  
    // ...
```

⇒ Thread divergence.

Ternary Normal Form (TNF)

We simplify the representation of constraints to ternary constraints of the form:

- $x = y \text{ } <\text{op}> z$ where x, y, z are variables.
- The operators are $\{+, /, *, \text{mod}, \text{min}, \text{max}, \leq, =\}$.

Example

The constraint $x + y \neq 2$ is represented by:

```
t1 = x + y  
ZERO = (t1 = TWO)           equivalent to false  $\Leftrightarrow$  (t1 = 2)
```

where ZERO and TWO are two variables with constant values.

Bytecode Representation

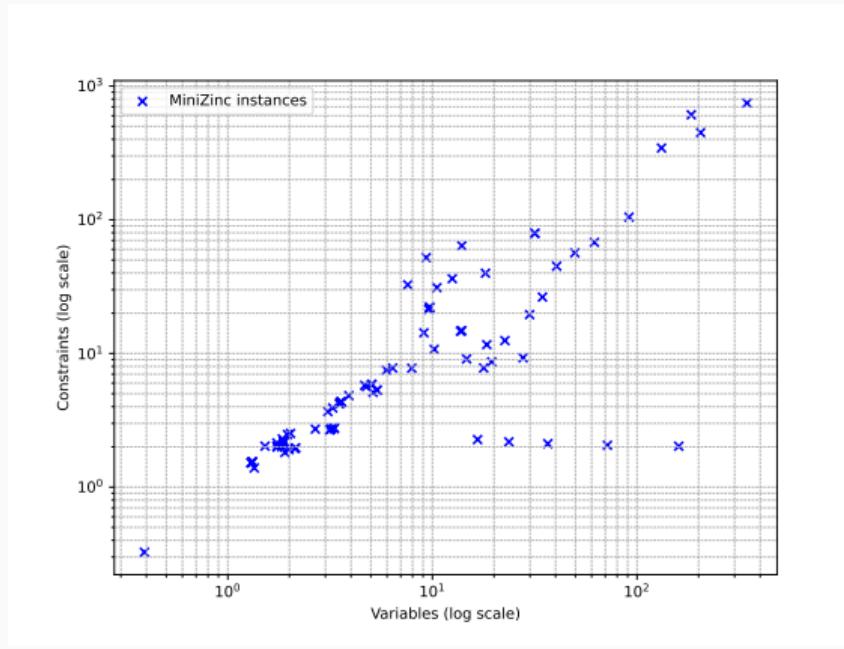
The ternary form of a propagator holds on 16 bytes:

```
struct bytecode_type {  
    int op;  
    int x;  
    int y;  
    int z;  
};
```

- **Uniform representation** of propagators in memory ⇒ coalesced memory accesses.
- Limited number of operators + sorting ⇒ reduced thread divergence.

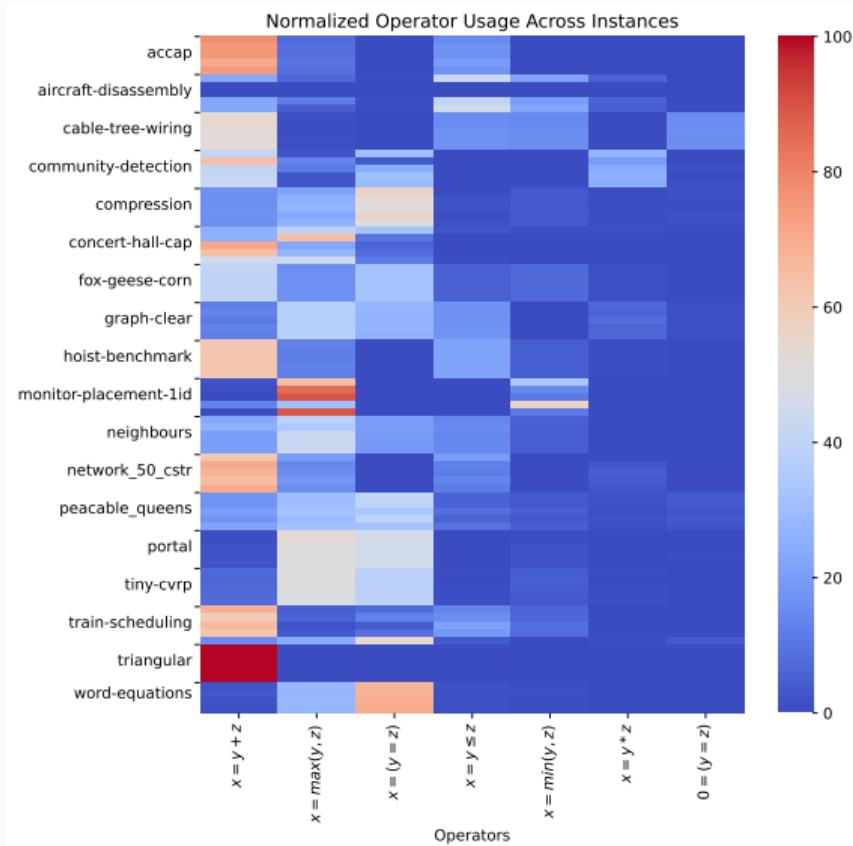
Drawback of TNF: increase in number of propagators and variables.

Benchmark on the MiniZinc Challenge 2024 (89 instances)



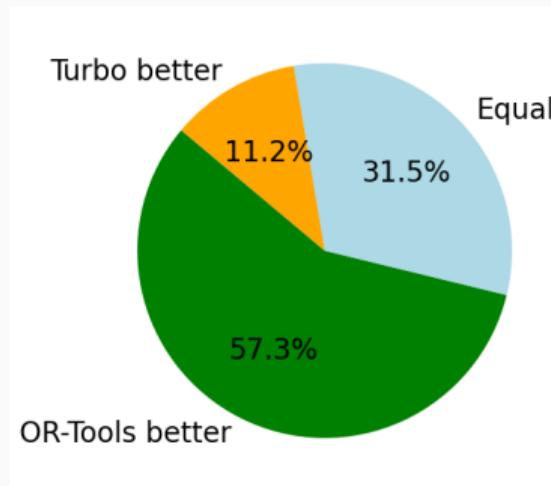
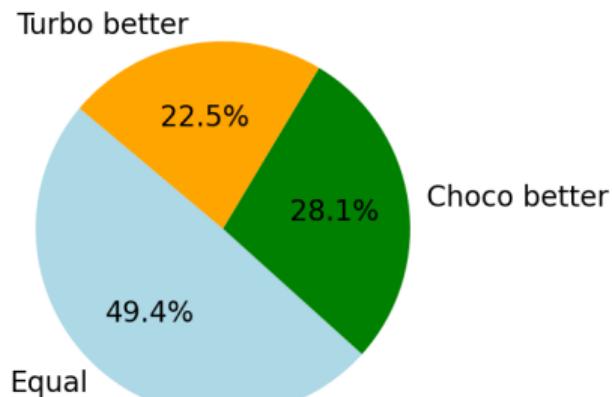
The **median increase** of variables is 4.76x and propagators is 4.33x.

Divergence?



Benchmarks: Turbo vs Choco v4.18

Comparison of the best objective values found (timeout: 20 mins, GPU: H100).



Conclusion

Conclusion: Theoretical Parallel Model of Computation

Data races occur rarely, so we should avoid working so much to avoid them.

Properties of the model

A Variant of Concurrent Constraint Programming on GPU (AAAI 2022)⁴.

- **Correct:** Proofs that $P; Q \equiv P||Q$, parallel and sequential versions produce the same results.
- **Restartable:** Stop the program at any time, and restart on partial data.
- **Weak memory consistency:** Very few requirements on the underlying memory model \Rightarrow wide compatibility across hardware, unlock optimization.

⁴<https://ptal.github.io/papers/aaai2022.pdf>

Conclusion: Practical Implementation

Turbo

- **Simple:** solving algorithms from 50 years ago.
⇒ no global constraints, nogoods learning, lazy clause generation, restart strategies, event-based propagation, trailing or recomputation-based state restoration and domain consistency.
- **Efficient:** Almost on-par with Choco (algorithmic optimization VS hardware optimization).



<https://github.com/ptal/turbo>

Conclusion: Practical Implementation

Turbo

- **Simple:** solving algorithms from 50 years ago.
⇒ no global constraints, nogoods learning, lazy clause generation, restart strategies, event-based propagation, trailing or recomputation-based state restoration and domain consistency.
- **Efficient:** Almost on-par with Choco (algorithmic optimization VS hardware optimization).



<https://github.com/ptal/turbo>

But...

- Still lagging behind CP+SAT solvers, and SAT learning is inherently sequential...
- There is hope: ACE (a pure CP solver) show CP-only is still competitive in XCSP3 compet.

Joint Projects?

- Abstract interpretation of abstract argumentation framework
 - Is it possible to use abstract interpretation for non-monotone logic?
 - Could be useful to combine abstract argumentation and FOL (not everything is non-monotone?).
- GPU-accelerated abstract argumentation
 - Does not seem to exist yet.
 - On CPU: "Cerutti, Federico, et al. *Exploiting parallelism for hard problems in abstract argumentation.*, AAAI-15"
 - Based on *strongly connected component* (SCC).
 - SCC is also useful for constraint reasoning (all-different constraint).
- Many smaller projects directly relevant to the topic of today (but not to abstract argumentation).