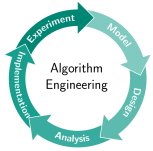
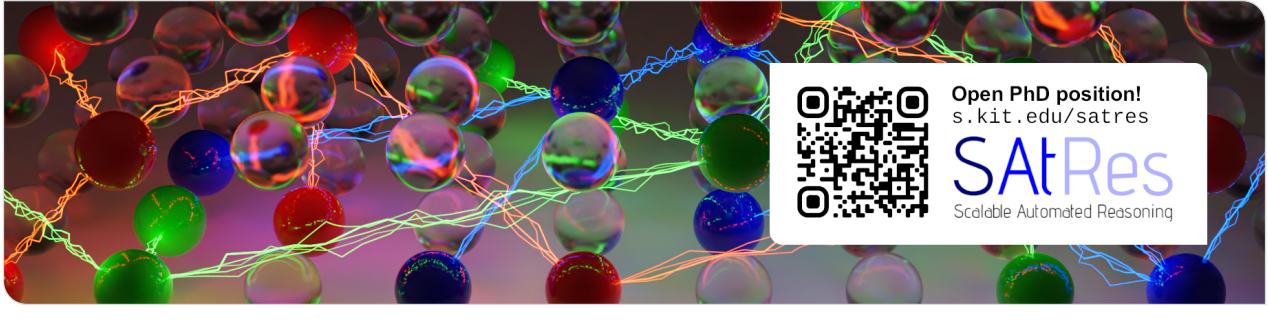
Karlsruher Institut für Technologie



MallobSat: Scalable SAT Solving by Clause Sharing

Journal of Artificial Intelligence Research (JAIR) article · Pragmatics of SAT 2024, Pune, India

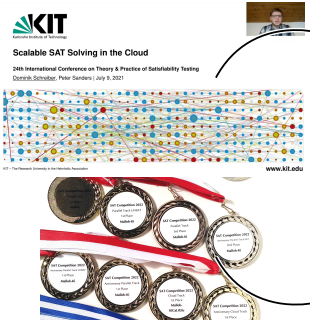
Dominik Schreiber, Peter Sanders | August 20, 2024



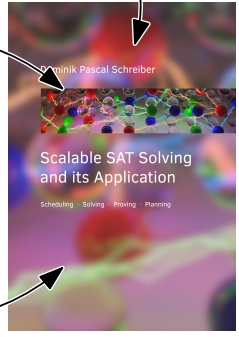
Context

SAT'21 paper "Scalable SAT Solving in the Cloud",
SAT Competition '20–23

Context

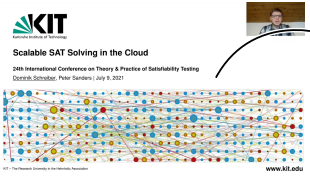


SAT'21 paper "Scalable SAT Solving in the Cloud", SAT Competition '20–23

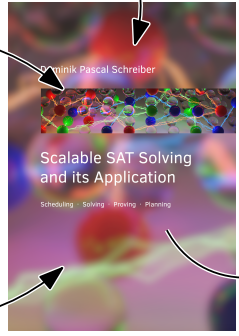


2023: Dissertation

Context



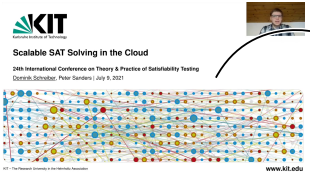
SAT'21 paper "Scalable SAT Solving in the Cloud", SAT Competition '20–23



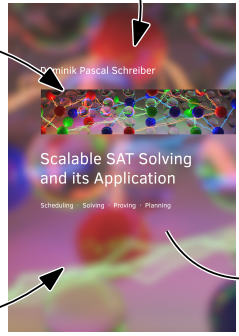
2023: Dissertation

August 2024: JAIR article "MallobSat: Scalable SAT Solving by Clause Sharing"

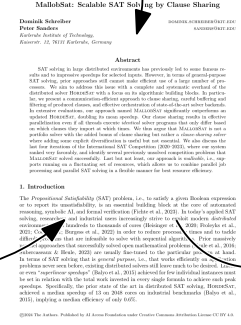
Context



SAT'21 paper "Scalable SAT Solving in the Cloud", SAT Competition '20–23



2023: Dissertation


MallobSat: Scalable SAT Solving by Clause Sharing

Dominik Schreiber
Peter Sanders

Abstract

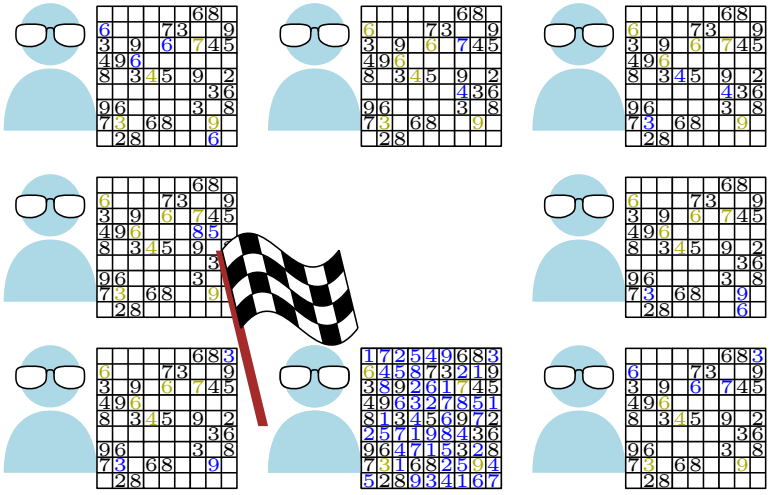
1. Introduction

August 2024: JAIR article "MallobSat: Scalable SAT Solving by Clause Sharing"

- LIVE

PoS'24 presentation

Reminder from Two Talks Earlier



Cooperative portfolio

- All experts work on original problem independently
- Brief meetings to exchange crucial insights
- Insights accelerate solving
- Only one expert needs to find a solution!

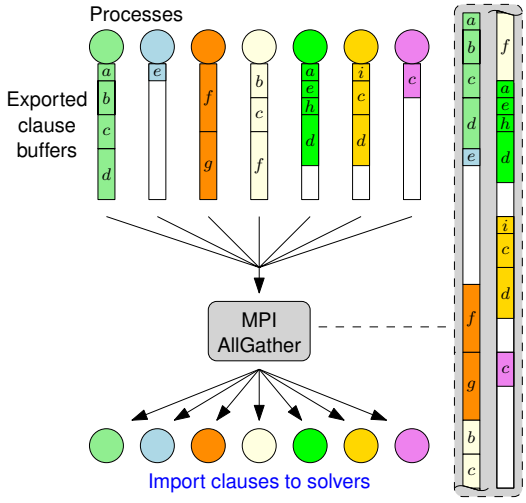
Parallel SAT

- Experts ≡ diversified sequential SAT solver threads
- Shared information ≡ conflict clauses

Distributed SAT: Prior State of the Art

Massively parallel solver **HordeSat** [Balyo et al. 2015]

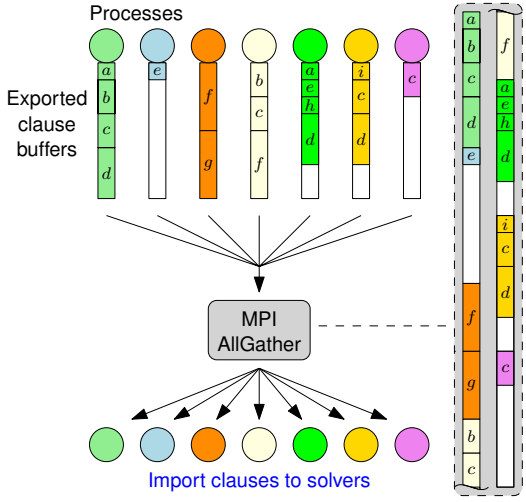
- Modular interface to plug in sequential solvers
- Periodic clause exchange
  - Concatenation of fixed-size clause buffers
  - Duplicates, unused space in buffers
    ⇒ often low number of distinct shared clauses



Distributed SAT: Prior State of the Art

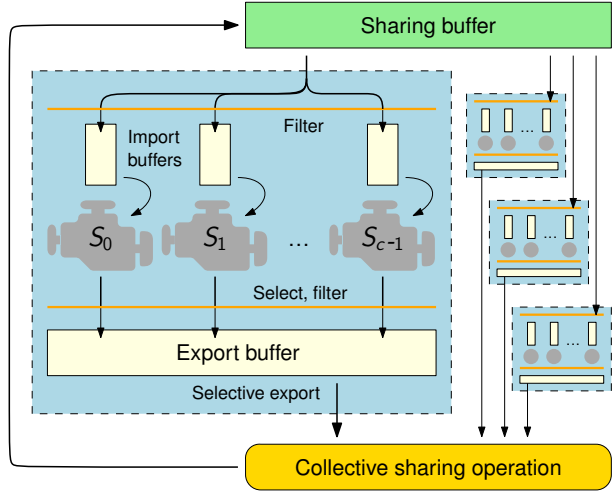
Massively parallel solver **HordeSat** [Balyo et al. 2015]

- Modular interface to plug in sequential solvers
- Periodic clause exchange
  - Concatenation of fixed-size clause buffers
  - Duplicates, unused space in buffers
    $\Rightarrow$ often low number of distinct shared clauses
- Experiments with $\leq$ 2048 cores
  - Surprisingly good scaling for difficult instances
  - Median speedup at 2048 cores: 13 (efficiency 0.6%)



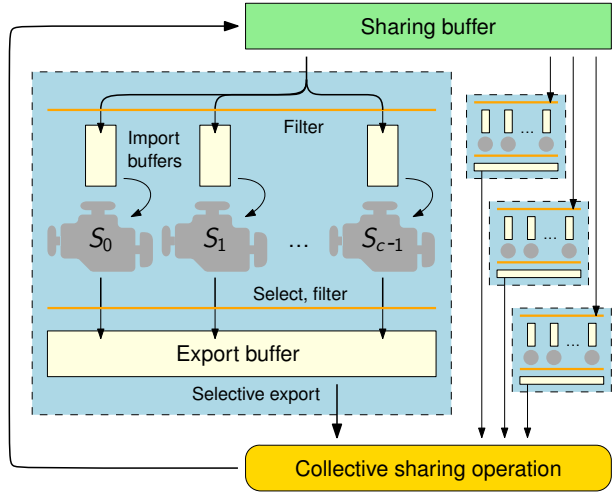
Design Decisions of MallobSat

- **Fundament:** HordeSat
  - Two-level hybrid parallelization
  - Periodic all-to-all clause sharing



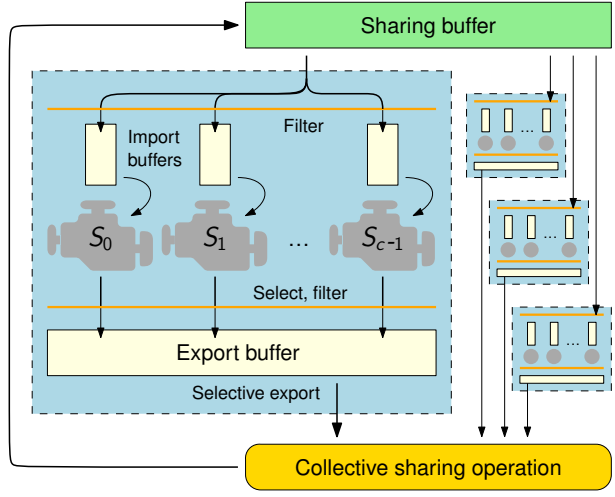
Design Decisions of MallobSat

- **Fundament:** HordeSat
  - Two-level hybrid parallelization
  - Periodic all-to-all clause sharing
- Fix a certain sharing volume, spend it on the globally most useful **distinct** clauses



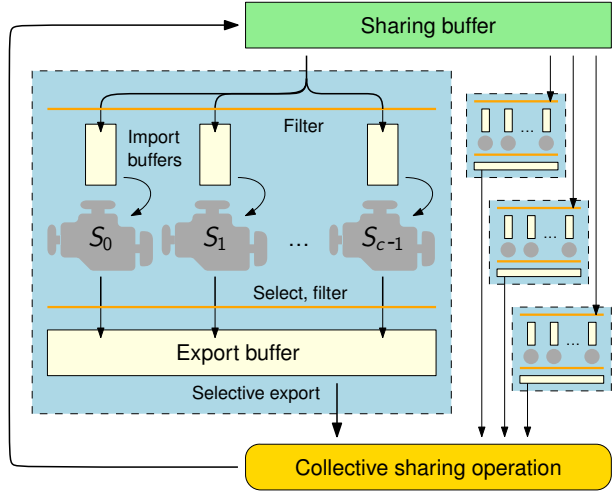
Design Decisions of MALLOBSAT

- **Fundament:** HORDESAT
  - Two-level hybrid parallelization
  - Periodic all-to-all clause sharing
- Fix a certain sharing volume, spend it on the globally most useful **distinct** clauses
- Prioritize clauses by clause length



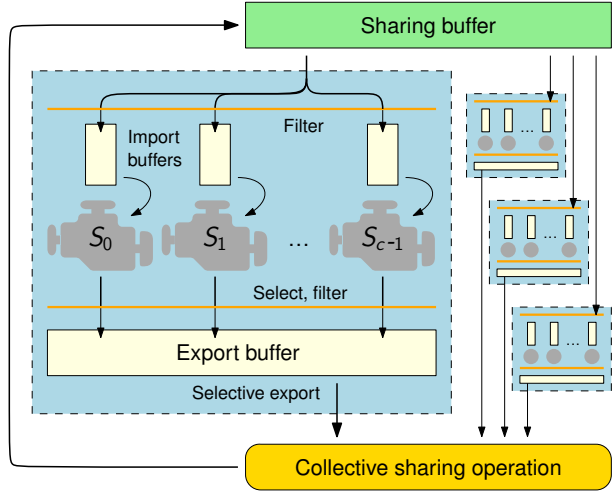
Design Decisions of MallobSat

- **Fundament:** HordeSat
  - Two-level hybrid parallelization
  - Periodic all-to-all clause sharing
- Fix a certain sharing volume, spend it on the globally most useful **distinct** clauses
- Prioritize clauses by clause length
- Minimize clause turnaround times



Design Decisions of MallobSat

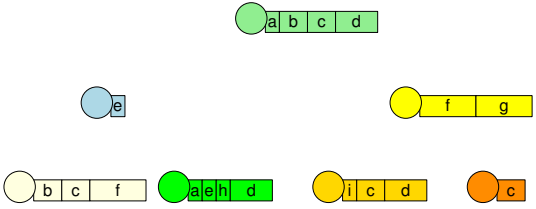
- **Fundament:** HordeSat
  - Two-level hybrid parallelization
  - Periodic all-to-all clause sharing
- Fix a certain sharing volume, spend it on the globally most useful **distinct** clauses
- Prioritize clauses by clause length
- Minimize clause turnaround times
- Support fluctuating workers (*malleability*)



Clause sharing: Our approach

Exchange of useful clauses

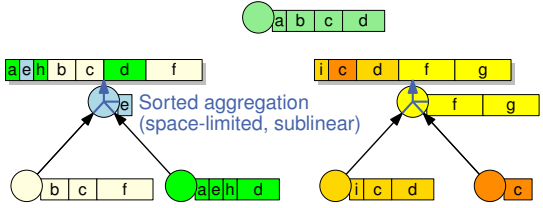
1.



Clause sharing: Our approach

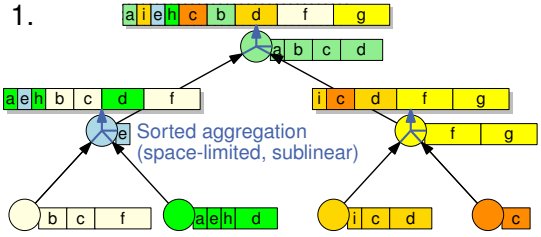
Exchange of useful clauses

1.



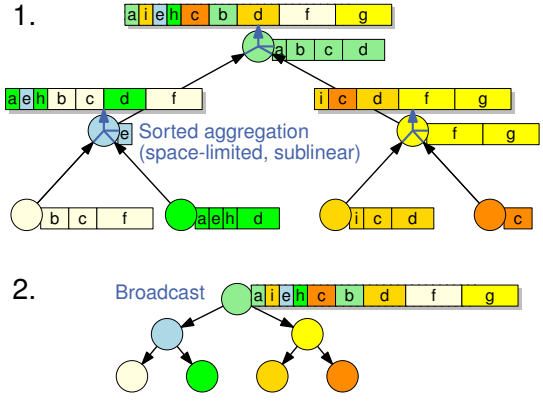
Clause sharing: Our approach

Exchange of useful clauses



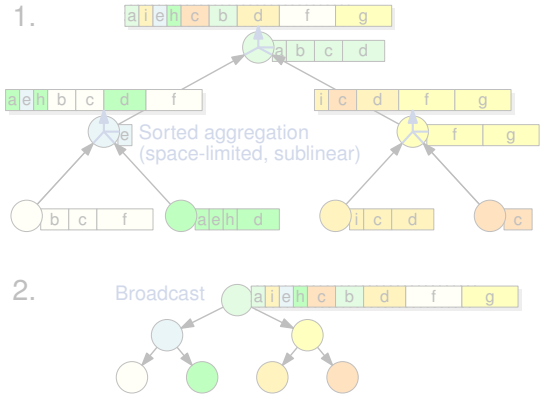
Clause sharing: Our approach

Exchange of useful clauses

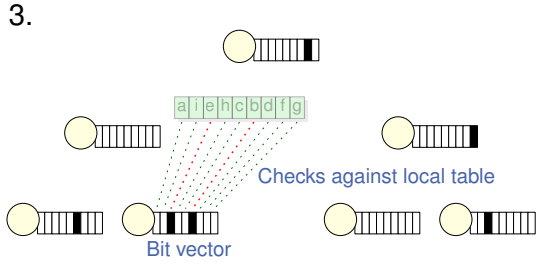


Clause sharing: Our approach

Exchange of useful clauses

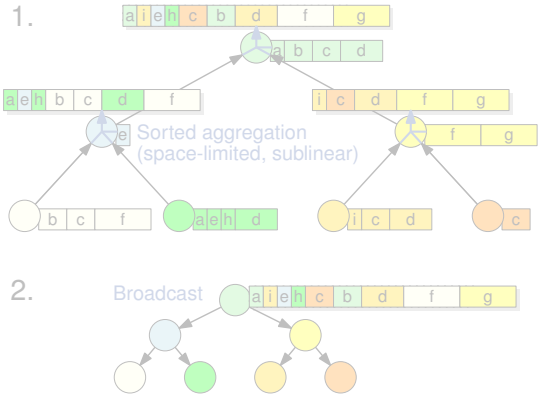


Filtering of recently shared clauses

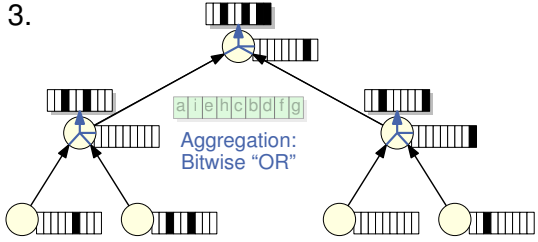


Clause sharing: Our approach

Exchange of useful clauses

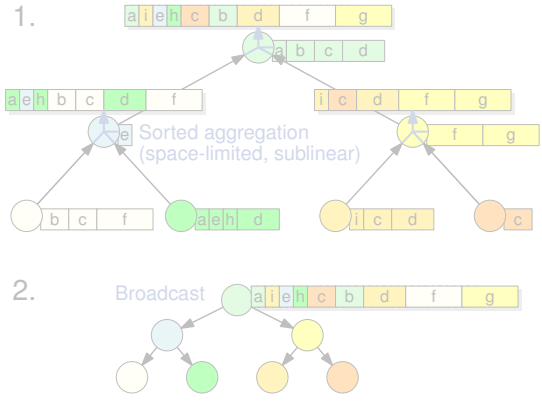


Filtering of recently shared clauses

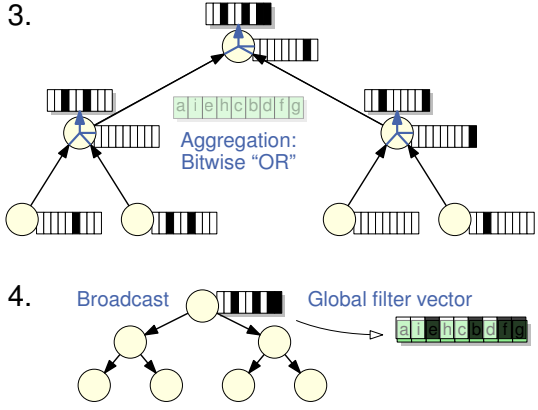


Clause sharing: Our approach

Exchange of useful clauses



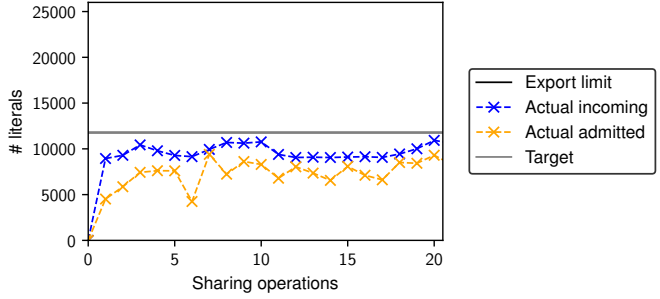
Filtering of recently shared clauses



Enforcing a Sharing Volume

We want to share L literals per sharing but may only get $L' < L$ successfully shared literals. Why?

1. Processes didn't produce, export enough clauses
2. Duplicate clauses were detected and eliminated during aggregation
3. Distributed filter blocked some of the transmitted clauses

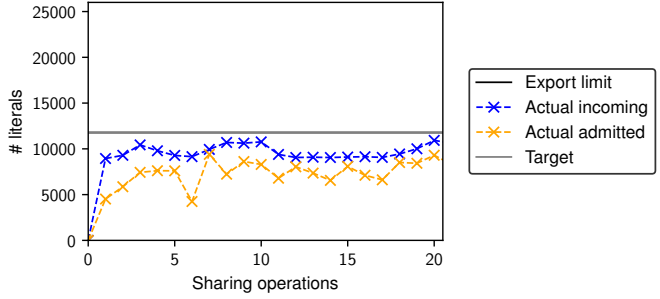


Enforcing a Sharing Volume

We want to share L literals per sharing but may only get $L' < L$ successfully shared literals. Why?

1. Processes didn't produce, export enough clauses
2. Duplicate clauses were detected and eliminated during aggregation
3. Distributed filter blocked some of the transmitted clauses

Fix: Elastic compensation for sharing volume unused for algorithmic reasons (2 , 3)

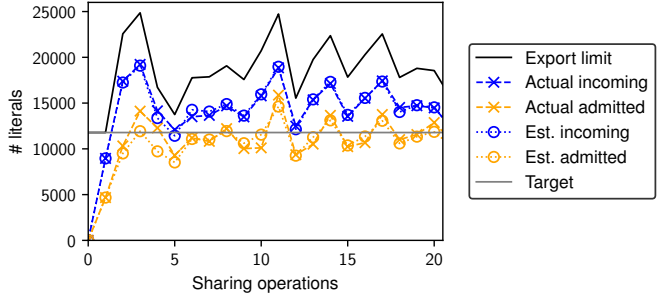


Enforcing a Sharing Volume

We want to share L literals per sharing but may only get $L' < L$ successfully shared literals. Why?

1. Processes didn't produce, export enough clauses
2. Duplicate clauses were detected and eliminated during aggregation
3. Distributed filter blocked some of the transmitted clauses

Fix: Elastic compensation for sharing volume unused for algorithmic reasons (2 , 3)



Handling LBD Values

- Seq. solving: central metric for whether to keep a clause
- **But:** LBD found by solver A not necessarily meaningful for solver B! $\rightarrow$ not as “global” as clause length

Handling LBD Values

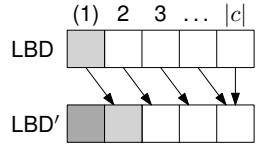
- Seq. solving: central metric for whether to keep a clause
- **But:** LBD found by solver A not necessarily meaningful for solver B! $\rightarrow$ not as “global” as clause length
- Some solvers keep clauses with LBD 2 indefinitely
  — but expect a single solver’s clause volume!
  $\Rightarrow$ Growing overhead (time, space) from low-LBD clauses

Handling LBD Values

- Seq. solving: central metric for whether to keep a clause
- **But:** LBD found by solver A not necessarily meaningful for solver B! $\rightarrow$ not as “global” as clause length
- Some solvers keep clauses with LBD 2 indefinitely — but expect a single solver’s clause volume!
  $\Rightarrow$ Growing overhead (time, space) from low-LBD clauses

Our current approach: Increment each LBD before import

- Maintains LBD-based prioritization of clauses
- Solver keeps more control over its LBD-2-clauses



	Median RAM	PAR-2
Orig. LBD	108.8 GiB	75.7
Reset LBD	95.6 GiB	74.3
LBD++	97.3 GiB	72.9

768 cores $\times$ 349 instances $\times$ 300 s

Portfolio & Diversification

Solver backends:

- LINGELING + YALSAT local search solver
- GLUCOSE + SYRUP clause sharing code
- CADICAL
- KISSAT

Diversification:

- Cycle through solver configuration options: restart intervals, pre–/inprocessing techniques, ...
- Sparse random variable phases
- Seeds, input shuffling, Gaussian noise for numeric parameters

Pitfalls in Parallel Solver Evaluation

Balyo et al. (2016):

"*Experiments showed that* HordeSAT *can achieve superlinear average speedup on hard benchmarks.*"

Pitfalls in Parallel Solver Evaluation

Balyo et al. (2016):
"*Experiments showed that* HordeSAT *can achieve superlinear average speedup on hard benchmarks.*"

Superlinear speedups: erratic,
often due to running time variance
or some crucial solver configuration
the sequential solver is missing

Pitfalls in Parallel Solver Evaluation

Arithmetic average of speedups:
no statistical meaning –
use geometric mean or median

Balyo et al. (2016):
"*Experiments showed that* HordeSAT *can achieve superlinear average speedup on hard benchmarks.*"

Superlinear speedups: erratic,
often due to running time variance
or some crucial solver configuration
the sequential solver is missing

Pitfalls in Parallel Solver Evaluation

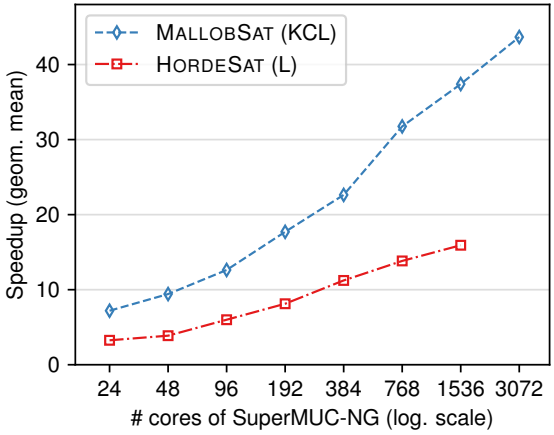
Arithmetic average of speedups:
no statistical meaning –
use geometric mean or median

Balyo et al. (2016):
"*Experiments showed that* HordeSAT *can achieve superlinear average speedup on hard benchmarks.*"

Superlinear speedups: erratic,
often due to running time variance
or some crucial solver configuration
the sequential solver is missing

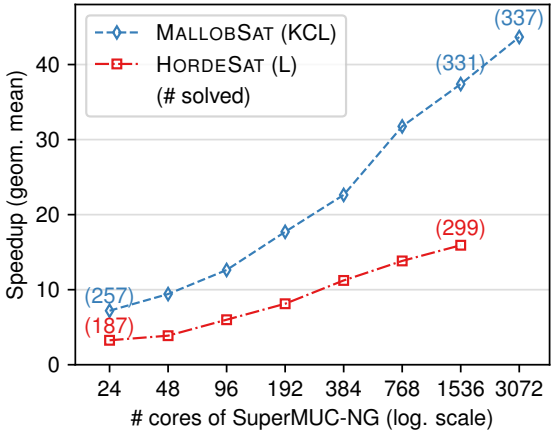
Counting sequential timeouts as
solved at the timeout:
makes speedups difficult to interpret,
especially for huge timeouts

Scaling



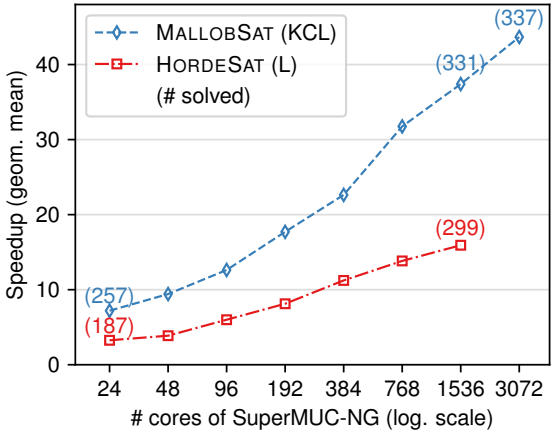
400 problems from SAT Comp. 2021 · Seq. baseline KISSAT_MAB-HYWALK · Seq. limit 32 h (331 solved) · Par. limit 300 s

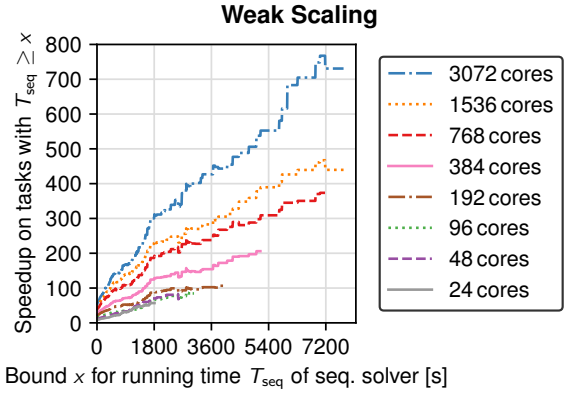
Scaling



400 problems from SAT Comp. 2021 · Seq. baseline KissAT_MAB-HyWalk · Seq. limit 32 h (331 solved) · Par. limit 300 s

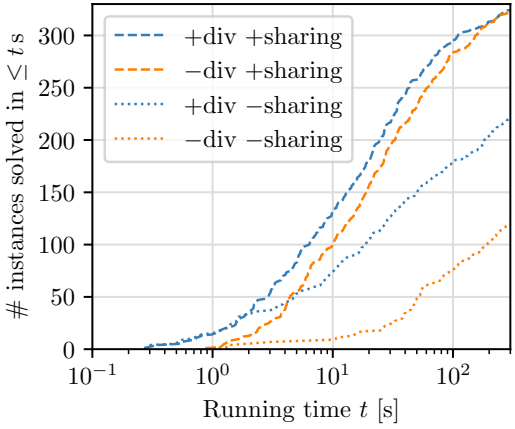
Scaling





400 problems from SAT Comp. 2021 · Seq. baseline KissatMAB-HyWalk · Seq. limit 32 h (331 solved) · Par. limit 300 s

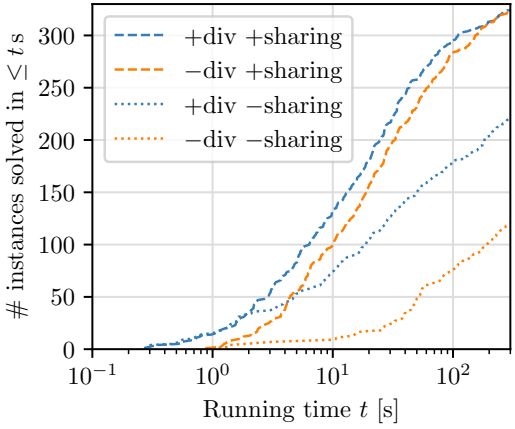
Impact of Diversification, Sharing @ 768 Cores



- Without sharing, diversification is highly effective

349 problems from SAT Comp. 2022 · KCL portfolio

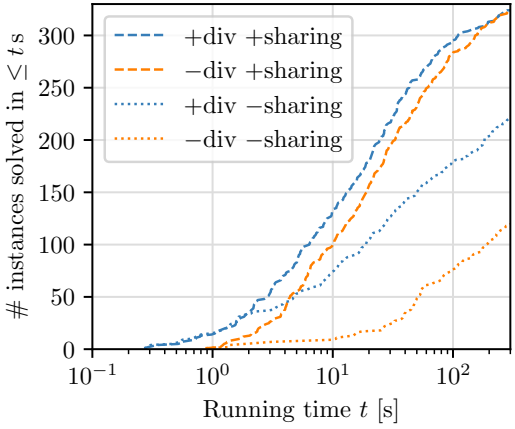
Impact of Diversification, Sharing @ 768 Cores



- Without sharing, diversification is highly effective
- With sharing: Only ≈ 40 distinct solver programs across 768 cores still perform competitively?!

349 problems from SAT Comp. 2022 · KCL portfolio

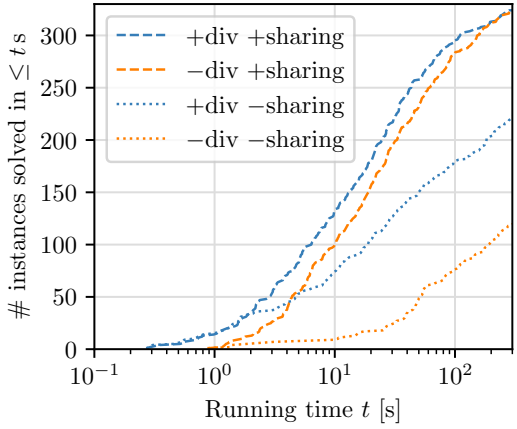
Impact of Diversification, Sharing @ 768 Cores



- Without sharing, diversification is highly effective
- With sharing: Only $\approx$ 40 distinct solver programs across 768 cores still perform competitively?!
  - Threads receive shared clauses at differing points in time
  - "Butterfly effect" $\Rightarrow$ deviating exploration
  - Clause sharing as distributed search space pruning

349 problems from SAT Comp. 2022 · KCL portfolio

Impact of Diversification, Sharing @ 768 Cores



- Without sharing, diversification is highly effective
- With sharing: Only ≈ 40 distinct solver programs across 768 cores still perform competitively?!
  - Threads receive shared clauses at differing points in time
  - "Butterfly effect" ⇒ deviating exploration
  - Clause sharing as distributed search space pruning
- Similar findings @ 3072 cores
  - Default CaDiCaL with primitive diversification (seeds, phases) performs competitively
  - Fully diversified portfolio without clause sharing does not

349 problems from SAT Comp. 2022 · KCL portfolio

MallobSat: A Portfolio Solver?

Prevalent concept in literature: **Portfolio solver** *with clause sharing* / Clause-sharing portfolio

- "*each thread runs a different SAT solver on the same instance[, which] in combination with clause-sharing leads to surprisingly good performance for small portfolio sizes*" – Ozdemir et al., 2021

MallobSat: A Portfolio Solver?

Prevalent concept in literature: **Portfolio solver** *with clause sharing* / Clause-sharing portfolio

- "*each thread runs a different SAT solver on the same instance[, which] in combination with clause-sharing leads to surprisingly good performance for small portfolio sizes*" – Ozdemir et al., 2021

Our view, based on empirical observations:

MallobSat is a Clause-sharing solver *with diversification*

- Clause sharing = main driver of scalability
- Adding explicit diversification is beneficial but not essential
- Applicability to other solvers?

MallobSat: Further Notes

MallobSat ...

- ... was the best cloud solver 2020–23 and among the top parallel solvers 2021–23
- ... was able to solve 22/100 previously unsolved instances within 20 min @ 3072 cores
- ... performs well in on-demand settings coupled with malleable job scheduling
  - Solve hundreds of instances at once, redistributing resources based on perceived difficulty

MallobSat: Further Notes

MallobSat ...

- ... was the best cloud solver 2020–23 and among the top parallel solvers 2021–23
- ... was able to solve 22/100 previously unsolved instances within 20 min @ 3072 cores
- ... performs well in on-demand settings coupled with malleable job scheduling
  - Solve hundreds of instances at once, redistributing resources based on perceived difficulty
- ... supports incremental SAT solving (CaDiCaL, Lingeling only; no proofs)
- ... supports proof checking (CaDiCaL only; non-incremental only)
  – more on Thursday, 10:30 AM!

Wrap-Up

Testimonials

"*Mallob-mono is now, by a **wide** margin, the most powerful SAT solver on the planet.*" —Byron Cook, Amazon Science, 2021

https://www.amazon.science/blog/automated-reasonings-scientific-frontiers

Best cloud solver @ International SAT Competition 2020–2023

Mallob(Sat) @ GitHub

github.com/
domschrei/mallob

JAIR article

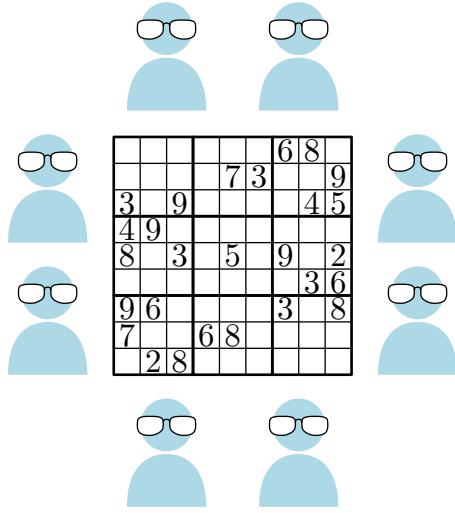
jair.org/index.php/
jair/article/view/15827

Parallel Logical Reasoning

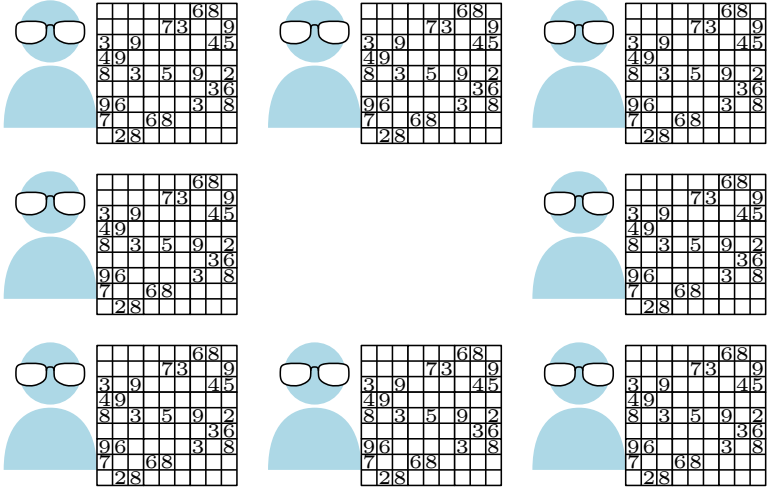
The assembly of logicians

- Complex logic puzzle
- *n* logic experts want to solve the puzzle
- Experts tend to work the best undisturbed

How to coordinate our experts?



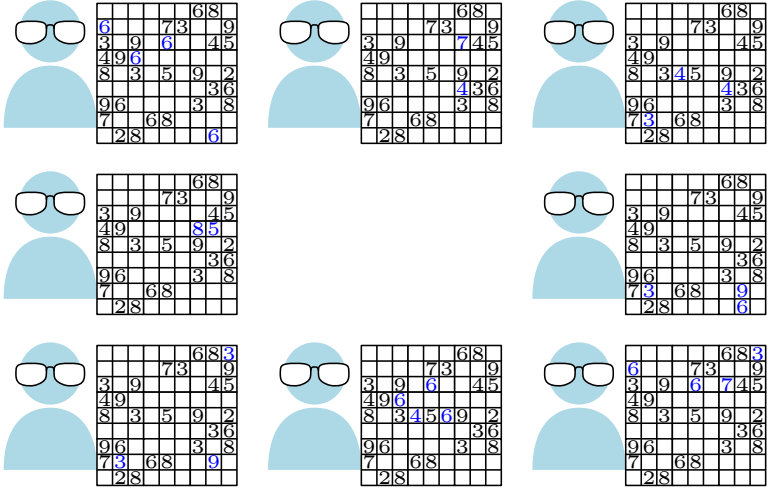
Pure Portfolio



Parallel portfolio

- All experts work on original problem independently

Pure Portfolio



Parallel portfolio

- All experts work on original problem independently
- Different approaches lead to different insights
- Only one expert needs to find a solution!

Pure Portfolio: Oracle view vs. Speedup view

Virtual Best Solver (VBS) / Oracle

Consider n algorithms $A_1, \ldots, A_n$ where for each input x, algorithm A_i has run time $T_{A_i}(x)$.
The Virtual Best Solver (VBS) for $A_1, \ldots, A_n$ has run time $T^*(x) = \min\{T_{A_1}(x), \ldots, T_{A_n}(x)\}$.

Pure Portfolio: Oracle view vs. Speedup view

Virtual Best Solver (VBS) / Oracle

Consider n algorithms $A_1, \ldots, A_n$ where for each input x, algorithm A_i has run time $T_{A_i}(x)$.
The Virtual Best Solver (VBS) for $A_1, \ldots, A_n$ has run time $T^*(x) = \min\{T_{A_1}(x), \ldots, T_{A_n}(x)\}$.

Optimist: A pure portfolio simulates the VBS using parallel processing!

- On idealized hardware, we "select" best sequential solver for each instance

Pure Portfolio: Oracle view vs. Speedup view

Virtual Best Solver (VBS) / Oracle

Consider n algorithms $A_1, \ldots, A_n$ where for each input x, algorithm A_i has run time $T_{A_i}(x)$.
The Virtual Best Solver (VBS) for $A_1, \ldots, A_n$ has run time $T^*(x) = \min\{T_{A_1}(x), \ldots, T_{A_n}(x)\}$.

Optimist: A pure portfolio simulates the VBS using parallel processing!

- On idealized hardware, we "select" best sequential solver for each instance

Parallel speedup

Given parallel algorithm P and input x, the speedup of P is defined as $s_P(x) = T_Q(x)/T_P(x)$
where Q is the best available sequential algorithm.

Pure Portfolio: Oracle view vs. Speedup view

Virtual Best Solver (VBS) / Oracle

Consider n algorithms $A_1, \ldots, A_n$ where for each input x, algorithm A_i has run time $T_{A_i}(x)$.
The Virtual Best Solver (VBS) for $A_1, \ldots, A_n$ has run time $T^*(x) = \min\{T_{A_1}(x), \ldots, T_{A_n}(x)\}$.

Optimist: A pure portfolio simulates the VBS using parallel processing!

- On idealized hardware, we "select" best sequential solver for each instance

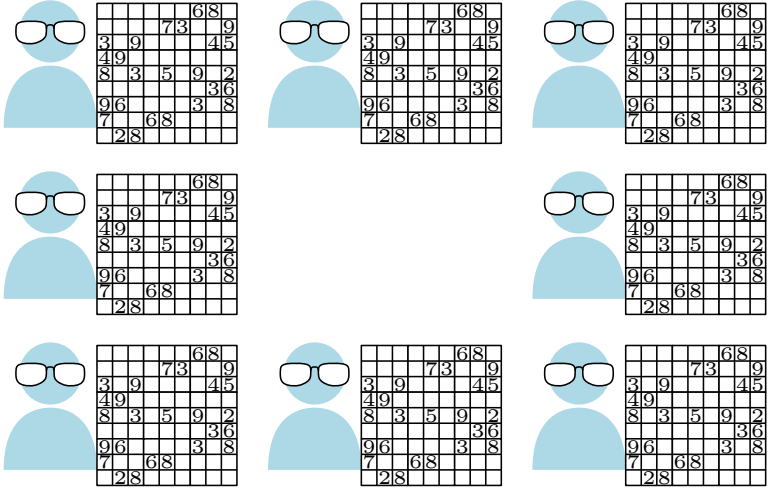
Parallel speedup

Given parallel algorithm P and input x, the speedup of P is defined as $s_P(x) = T_Q(x)/T_P(x)$
where Q is the best available sequential algorithm.

Pessimist: A pure portfolio never achieves actual speedups!

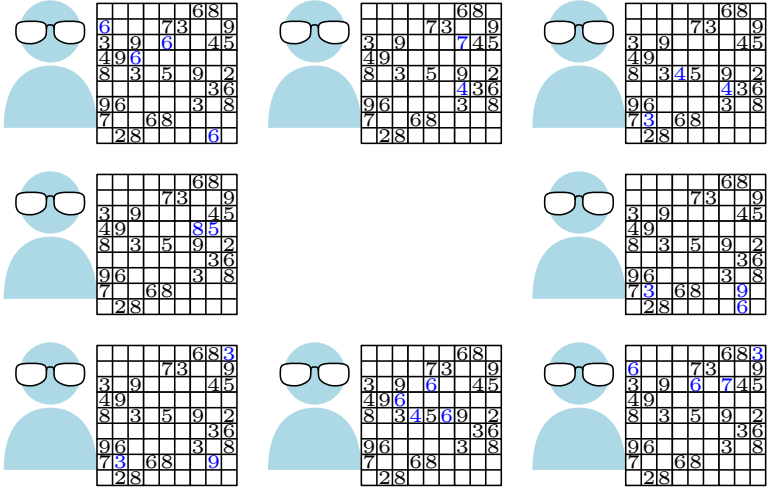
- There is always a sequential algorithm performing at least as well
- Consequence: Not resource efficient, not scalable

Cooperative Portfolio



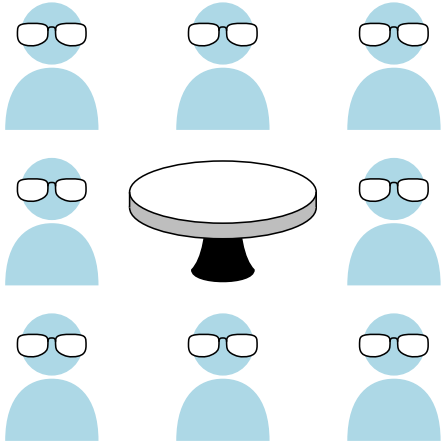
- All experts work on original problem independently

Cooperative Portfolio



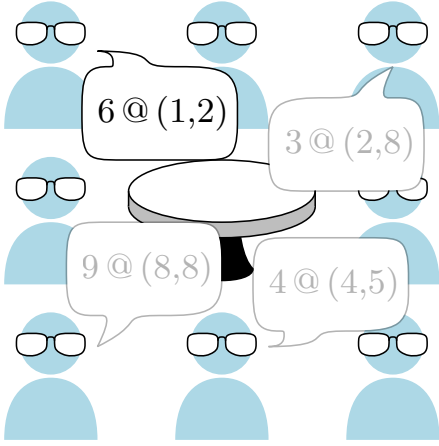
- All experts work on original problem independently

Cooperative Portfolio



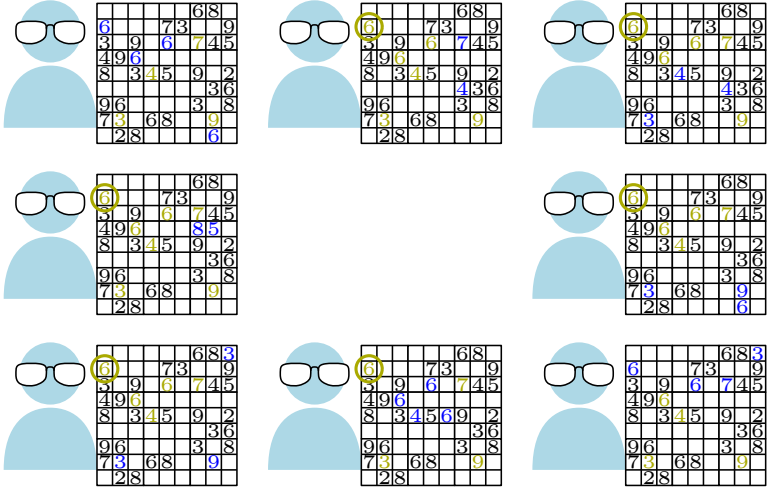
- All experts work on original problem independently
- Brief meetings to exchange crucial insights

Cooperative Portfolio



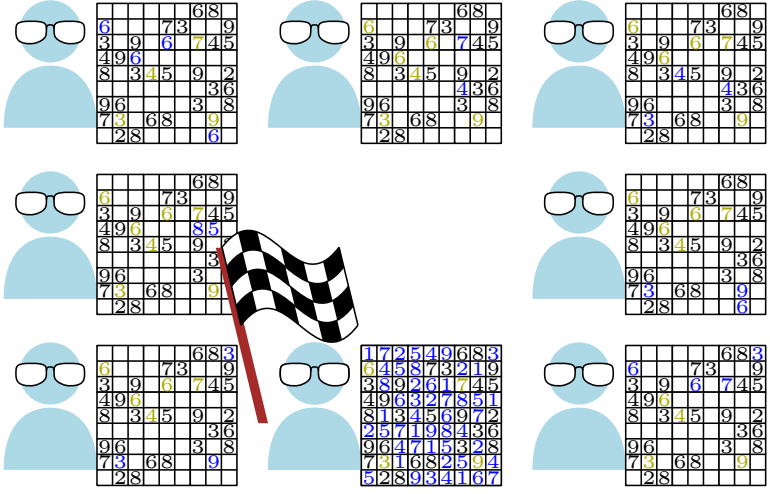
- All experts work on original problem independently
- Brief meetings to exchange crucial insights

Cooperative Portfolio



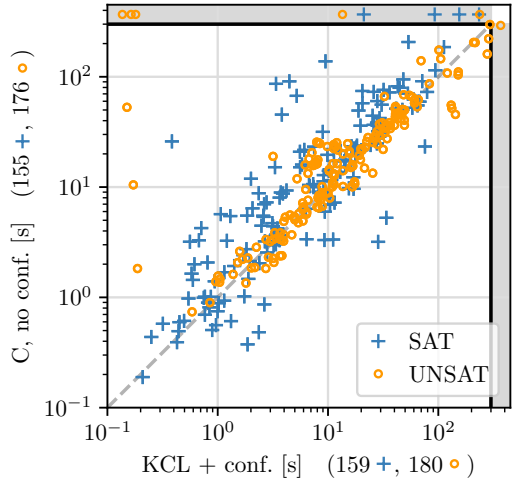
- All experts work on original problem independently
- Brief meetings to exchange crucial insights
- Insights accelerate solving

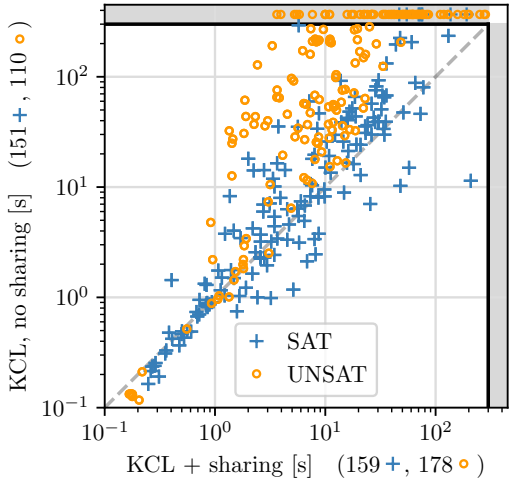
Cooperative Portfolio



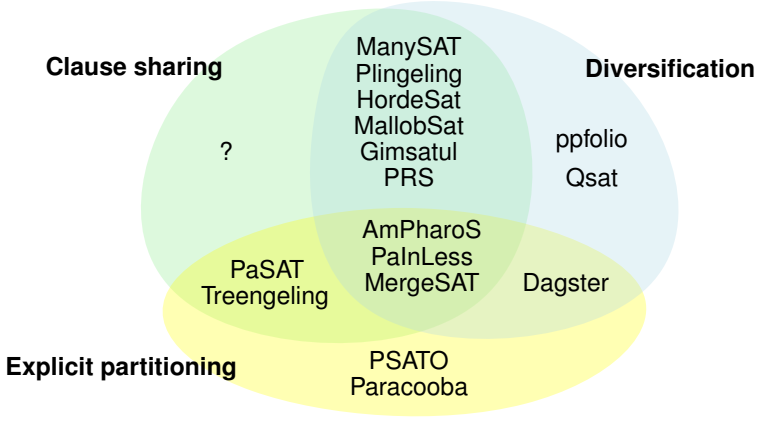
- All experts work on original problem independently
- Brief meetings to exchange crucial insights
- Insights accelerate solving
- Only one expert needs to find a solution!

Impact of Diversification, Sharing @ 3072 Cores





Parallel SAT Landscape



Parallel SAT Landscape