Time-Space Tradeoffs in Proof Complexity: Superpolynomial Lower Bounds for Superlinear Space

Chris Beck

Princeton University

Joint work with

Paul Beame & Russell Impagliazzo,
Jakob Nordström & Bangsheng Tang

SAT & SAT Solvers

- SAT is central to both theory and practice
- In the last ten years, there has been a revolution in practical SAT solving. Modern SAT solvers can sometimes solve practical instances with millions of variables.
- Best current solvers use a Backtracking approach pioneered by DPLL ’62, plus an idea called Clause Learning developed in Chaff ‘99.

SAT & SAT Solvers

- DPLL search requires very little memory
- Clause learning adds new clauses to the CNF every time the search backtracks
  - Uses **lots** of memory to try to beat DPLL.
  - In practice, **must** use heuristics to guess which clauses are "important" and store only those. Hard to do well! Memory becomes a bottleneck.
- **Question:** Is this inherent? Or can the right heuristics avoid the memory bottleneck?

SAT Solvers and Proof Complexity

- How can we get lower bounds for SAT Solvers?
- Analyzing search heuristics is very hard! Instead, give that away. Focus on the proofs.
- If a CNF φ only has Resolution proofs of size t, then t lower bounds runtime for "ideal" solver
- Amazingly, we can get sharp bounds this way!
- Explicit CNFs known with exponential size lower bounds. [Haken, Urquhart, Chvátal & Szemeredi...]

Resolution Proof System

- Proof lines are clauses, one simple proof step

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

.

- Proof is a sequence of clauses each of which is
  - an original clause, or
  - follows from previous clauses via resolution step
- A CNF is UNSAT iff can derive empty clause $\perp$
- Generated by CDCL SAT solvers on UNSAT runs.

SAT Solvers and Proof Complexity

- More recently, researchers want to investigate memory bottleneck for DPLL + Clause Learning
- We know

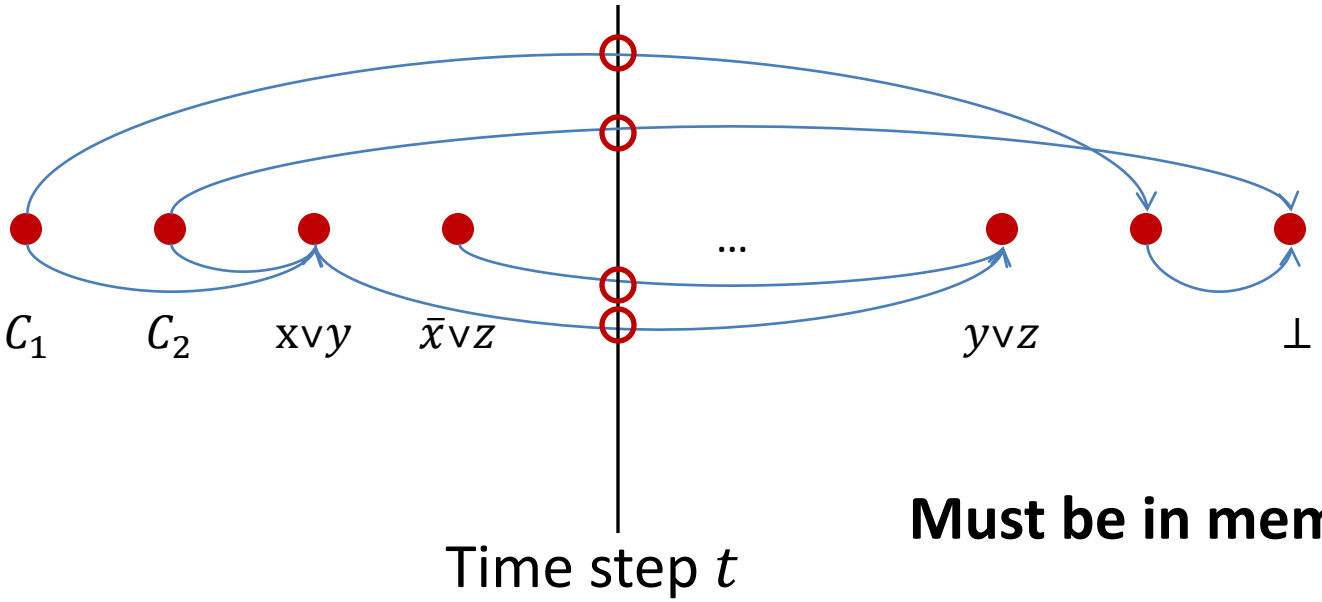
Proof Size $\leq$ Time for Ideal SAT Solver,

and similarly, we will define *Clause Space*

Clause Space $\leq$ Memory for Ideal SAT Solver,

and prove lower bounds in terms of this.

Space in Resolution: Clause Space



- Informally:
- *Clause space of a proof = Number of clauses you need to hold in memory at once in order to carry out the proof.*

[Esteban, Torán '99]

Clause Space of a proof = $\max_t$ (# active clauses).

Lower Bounds on Space?

- Generic Upper Bound: All UNSAT formulas on n vars have DPLL refutation in space $\leq n$.
  - Sharp lower bounds are known for explicit tautologies. [ET'99, ABRW'00, T'01, AD'03]
- So although we can get tight results for space, we can't show superpolynomial space is needed this way – need to think about size-space tradeoffs.
- In this direction: [Ben-Sasson, Nordström '10] Pebbling formulas with proofs in Size $\mathbf{O(n)}$, Space $\mathbf{O(n)}$, but Space $\mathbf{O(n/\log n)} \Rightarrow$ Size $\mathbf{\exp(n^{\Omega(1)})}$.
- **But,** this is still only for sublinear space.

Size-Space Tradeoffs

Theorem: [Beame, B., Impagliazzo'12]
For any $k > 0$, there are formulas of size n s.t.

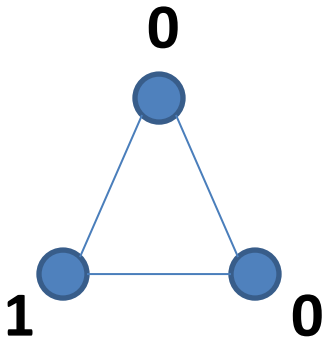
- There is a proof in $Size, Space \approx n^k$
- For any proof,

$$Size \geq \left(\frac{n^{.57k}}{Space}\right)^{\log\log n/\log\log\log n}$$

- Eli Ben-Sasson asks formally: "Does there exist c such that any CNF with a refutation of size T also has a refutation of size T^c in space $O(n)$?"

Tseitin Tautologies

Given an undirected graph $G = (V, E)$, and a function $\chi: V \to \mathbb{F}_2$, define a CSP:



Boolean variables: $\forall e \in E$ $\quad x_e$

Parity constraints: $\forall v \in V$ $\quad \sum_{e \ni v} x_e \equiv \chi(v)$

(linear equations)

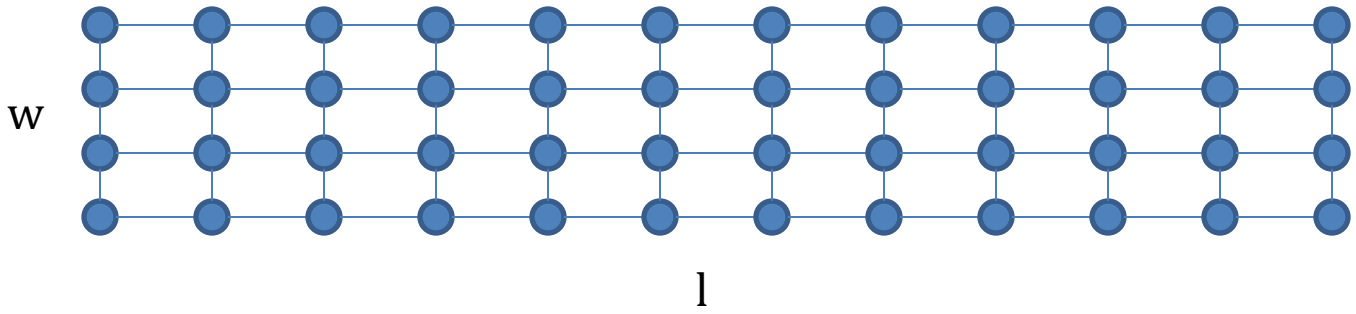
When χ has **odd** total parity, CSP is UNSAT.

Tseitin Tautologies

- When χ odd, G connected, corresponding CNF is called a Tseitin tautology. [Tseitin '68]
- Specifics of χ don't matter, only total parity. The graph is what determines the hardness.
- Known to be hard with respect to Size and Space when G is a constant degree expander. [Urquhart '87, Torán '99]
- **This work:** Tradeoffs on $\boldsymbol{w \times l}$ **grid**, $\boldsymbol{l \gg w}$, and similar graphs, using isoperimetry.

Tseitin formula on Grid

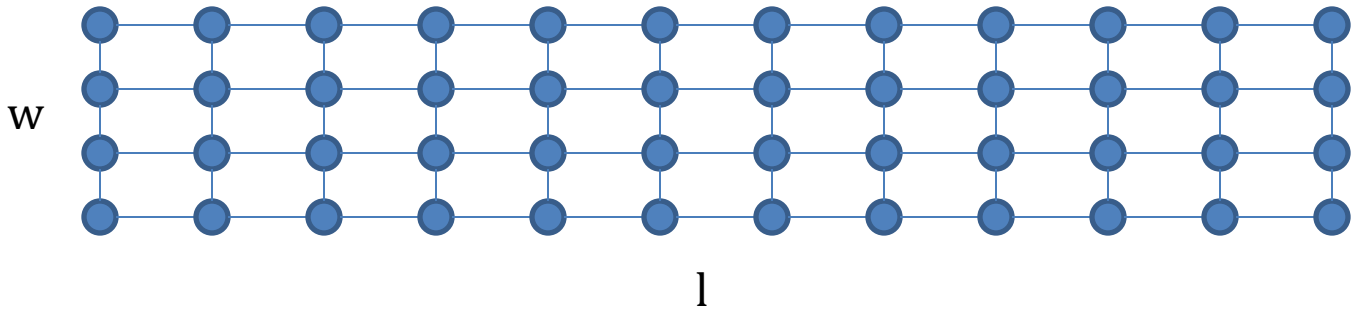
- Take a Tseitin formula on $w \times l$ grid, $l=4w^2$



- How can we build a resolution refutation?

Tseitin formula on Grid

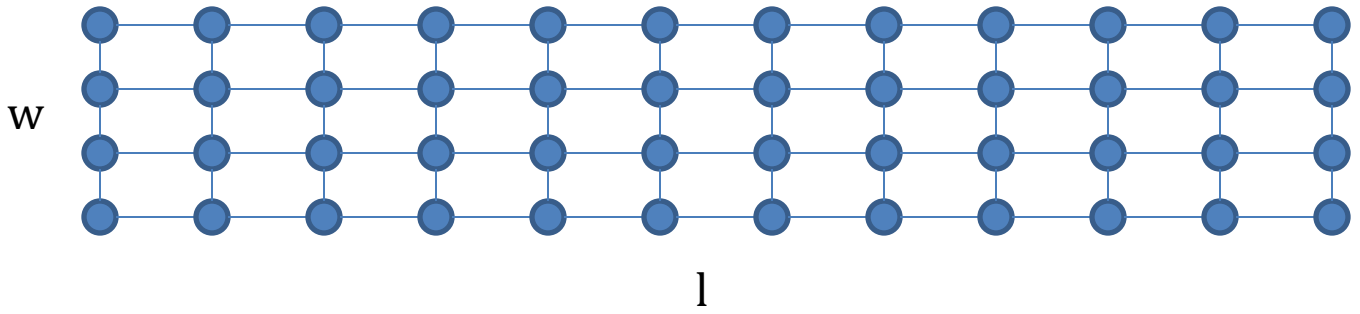
- One idea: Divide and conquer
- Think of DPLL, repeatedly bisecting the graph



- Each time we cut, one component is unsat. So after branching $w \log |V|$ times, get a violated clause. Idea leads to a tree-shaped proof with Space $\approx w \log l$, Size l^w.

Tseitin formula on Grid

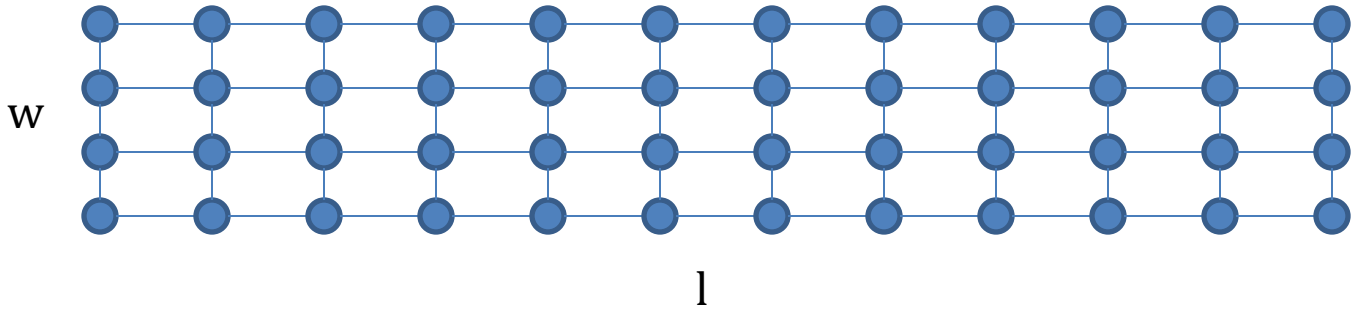
- 2nd idea: Mimic linear algebra refutation
- If we add all eqns in some order, get 1 = 0.



- A linear equation on k variables corresponds to 2^k clauses. Resolution can simulate a sum of two k-variable equations with $2^{O(k)}$ steps.

Tseitin formula on Grid

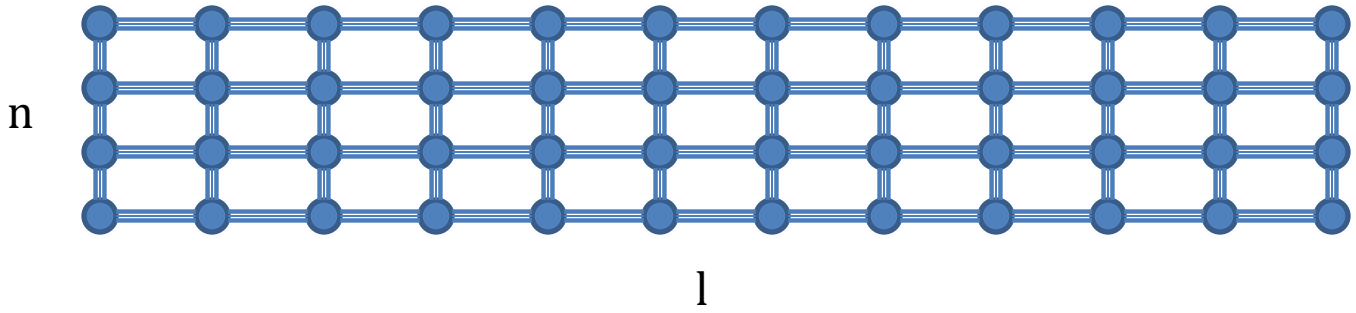
- If we add the lin eqn's in column order, then any intermediate equation has at most n vars.



- Get a proof of Size $\approx \#vertices \cdot 2^w$, Space $\approx 2^w$. This can also be thought of as *dynamic programming* version of 1st proof.

Tseitin formula on Grid

- If $w = n, l = \text{poly}(n)$, formula has size $\text{poly}(n)$, and you can have time and space $\approx 2^n$, or time n^n and space $poly(n)$. "Savitch-like" savings.



- Our theorem shows that quasipolynomial blowup in size when the space is below 2^n is *necessary* for the proof systems we studied.
- For technical reasons, work with "doubled" grid.

Warmup Proof

- Our proof builds on existing size lower bound techniques using random restrictions. [Haken, Beame Pitassi ‘95].
- To illustrate the ideas behind our result, we’ll first review some details of the Beame Pitassi result, then show how to build on it to get a size/space tradeoff.

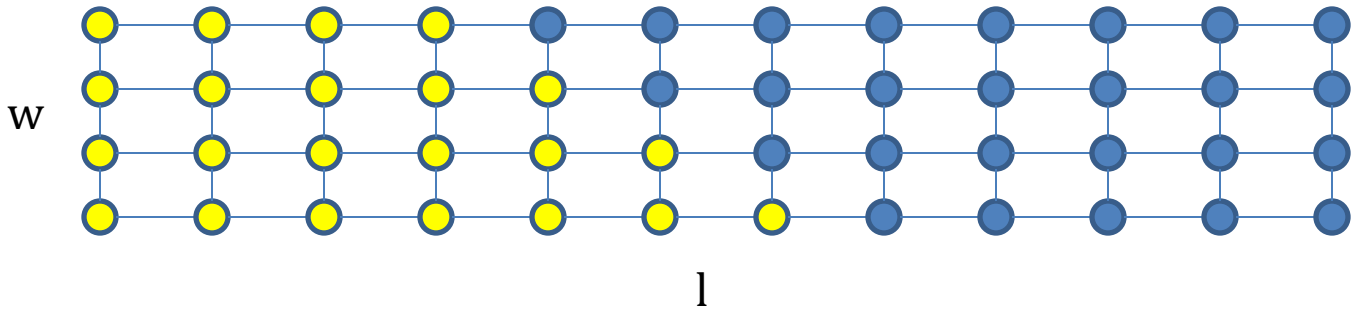
Warmup Proof

- First, we show that any refutation of the 1x grid formula must contain *at least one wide clause*.

- Then, we use a random restriction argument to "boost" this, showing that proofs of 2x grid must contain *many wide clauses*, and hence are large.

Warmup Proof: One Wide Clause

- Observation: Any roughly balanced cut in the $\mathbf{w \times l}$ **grid**, has at least $\mathbf{w}$ crossing edges.



(Precise: Any cut with $\geq w^2$ points on either side.)

- Want to use this to show that proofs of 1x grid formula require a clause of width $\geq \boldsymbol{w}$.

Warmup Proof: One Wide Clause

- Strategy: For any proof which uses all of the axioms, there must exist a statement which relies exactly on about half of the axioms.

- Formally: Define a "complexity measure" on clauses, $\mu(C)$, which is the size of the smallest subset of vertices such that the corresponding axioms logically imply C.

Warmup Proof: One Wide Clause

- μ is a *sub-additive complexity measure*:
  - $\mu(\text{initial clause}) = 1$,
  - $\mu(\bot) = \#\ vertices$,
  - $\mu(C) \leq \mu(C_1) + \mu(C_2)$, when $C_1, C_2 \vdash C$.

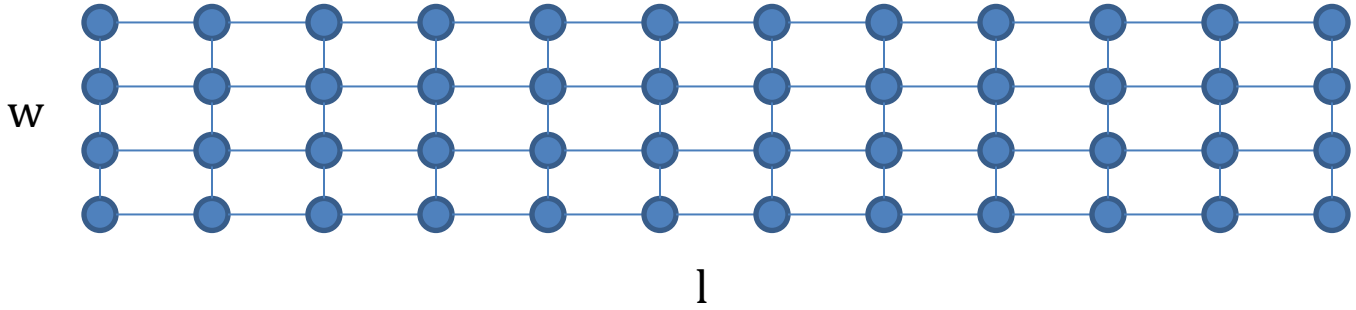
- Important property: Let S be a minimal subset of vertices whose axioms imply C. Then every edge on the boundary of S appears in C.

Warmup Proof: One Wide Clause

- Take any proof of 1x grid formula. At the start of the proof, all clauses have small μ. At the end of the proof, the final clause has large μ. Since μ at most doubles in any one step, there is at least one C such that $\frac{1}{3} \leq \frac{\mu(C)}{|V|} \leq \frac{2}{3}$.
- Let S be minimal subset of the vertices which imply C. Since S represents a balanced cut, its boundary is large; C has $\geq w$ variables.

Warmup Proof: Many Clauses

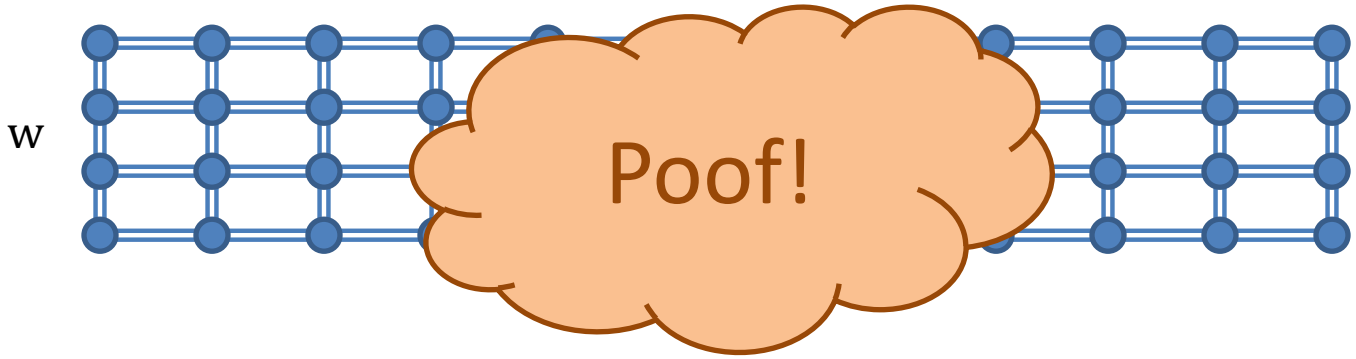
- Consider a *random restriction* for 2x grid which for each edge pair, sets one of the two at random to a random constant.



- Then, formula always simplifies to the 1x grid.
- However, for any clause C, the probability that its restriction is width > w is at most $(3/4)^w$.

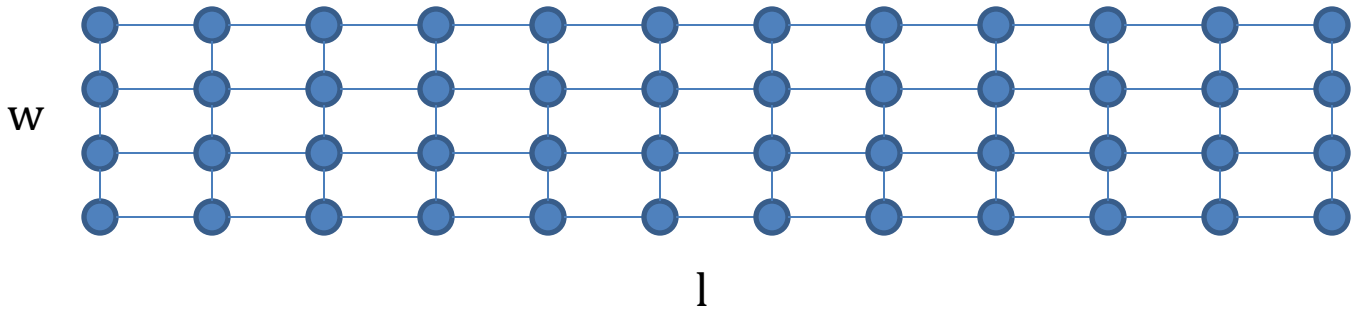
Warmup Proof: Many Clauses

- Suppose there was a proof of the 2x grid with fewer than $(4/3)^w$ clauses. Hit it with a random restriction to obtain a proof of 1x grid.



Warmup Proof: Many Clauses

- Suppose there was a proof of the 2x grid with fewer than $(4/3)^w$ clauses. Hit it with a random restriction to obtain a proof of 1x grid.



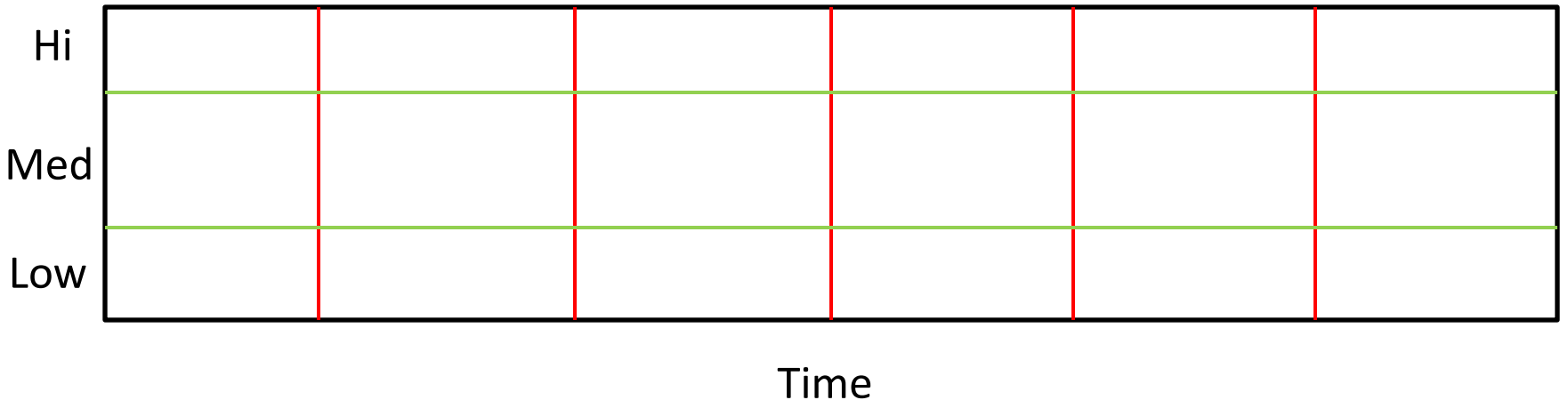
- With nonzero probability, no width w clauses remain, a contradiction to the first step.

Size Space Tradeoff

- Generic Recipe for Time Space Tradeoffs:
  - Divide (computation) into epochs. If epochs are small, not much "progress" happens in any one.
  - If the space is small, and #epochs is small, not much progress can be carried between epochs.
- To get Tradeoffs for Resolution, our first step is to make a very simple argument about epochs and progress, kind of like this recipe, then use restrictions to boost it to something strong.

Step One: Plot μ vs. Time

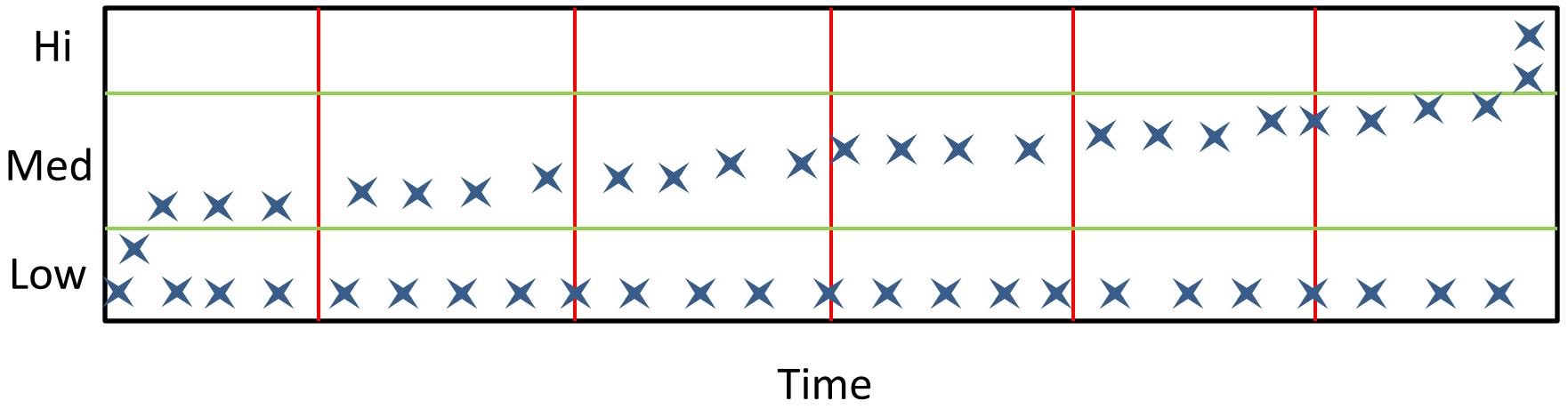
- Fix a proof of 1x grid. Here we plot μ vs. time. Red lines are breakpoints between epochs.



- Say that C is *medium complexity* if $w^2 < \mu(C) < |V|/2$, and high or low otherwise.

Step One: Plot μ vs. Time

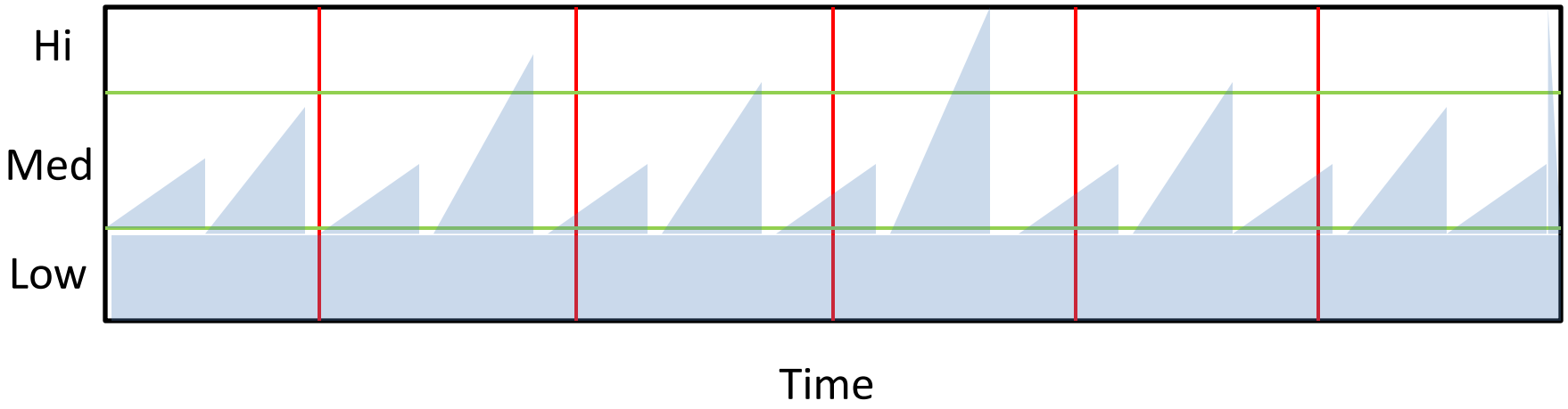
- Example Plot: “High Space” proof.



- Progress occurs roughly at a constant rate, and most clauses are of similar complexity values at any time step.

Step One: Plot μ vs. Time

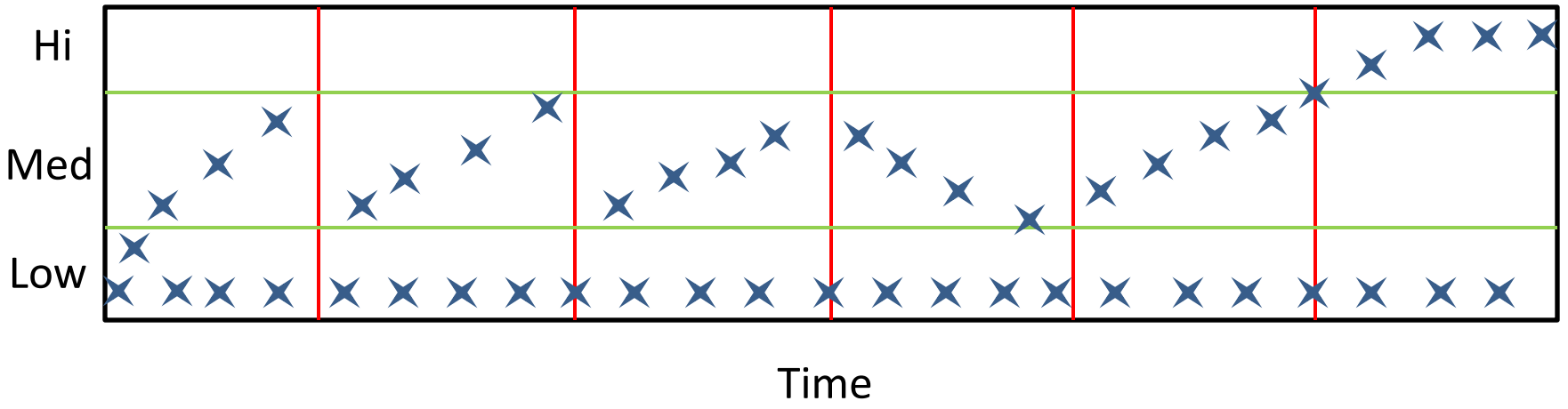
- Example Plot: “Divide & Conquer” proof (crude approximation)



- Will have clauses of many different complexities at all steps of proof

Step One: Plot μ vs. Time

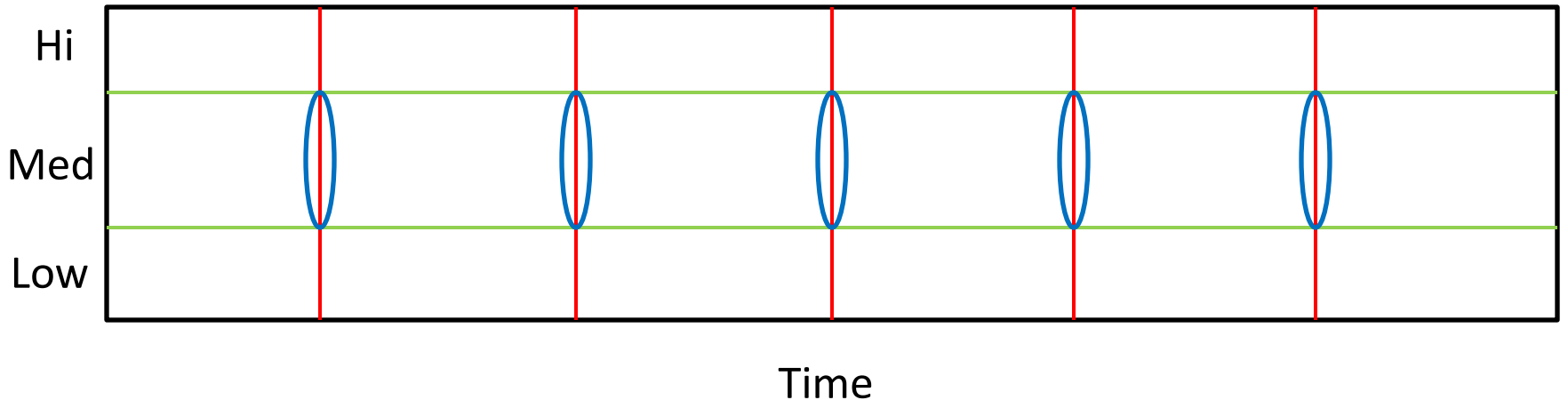
- Example Plot: “weird proof”



- The only constraints we have:
  Complexity starts low, ends high, and at most doubles in one step.

Two Possibilities

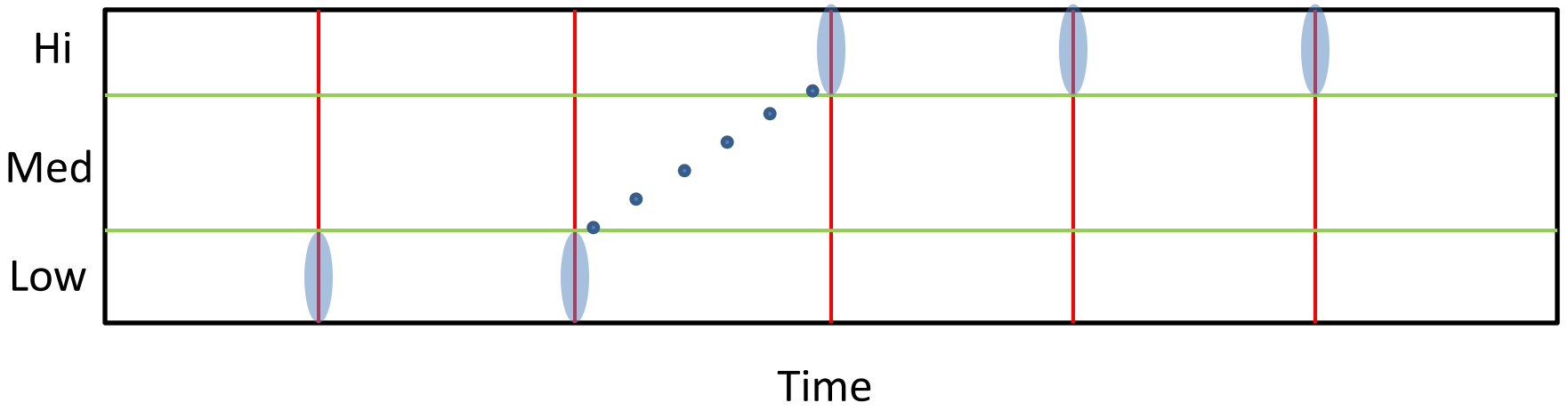
- **Either**, a medium clause appears in memory during **one** breakpoint between epochs,



- Intuitive meaning: Some “work” is being shared between two epochs.

Two Possibilities

- **Or**, all breakpoints only have Hi and Low.



- Must have an epoch which starts Low ends Hi, and so has $\log l/w$ clauses of superincreasing μ values, by subadditivity. (Does a lot of work)

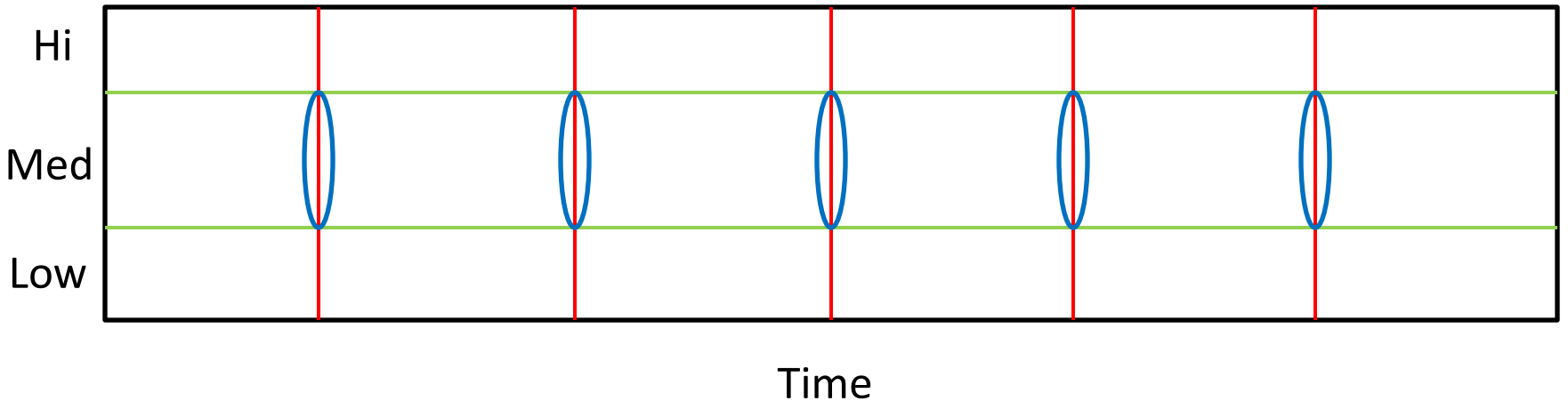
Second Step

- Now that we have our weak statement about progress and epochs, it is time to boost it.

- Fix a proof of 2x grid. If the space and size are small, we want to show that after a restriction both progress scenarios are unlikely, which implies a contradiction.

Probability of First Scenario

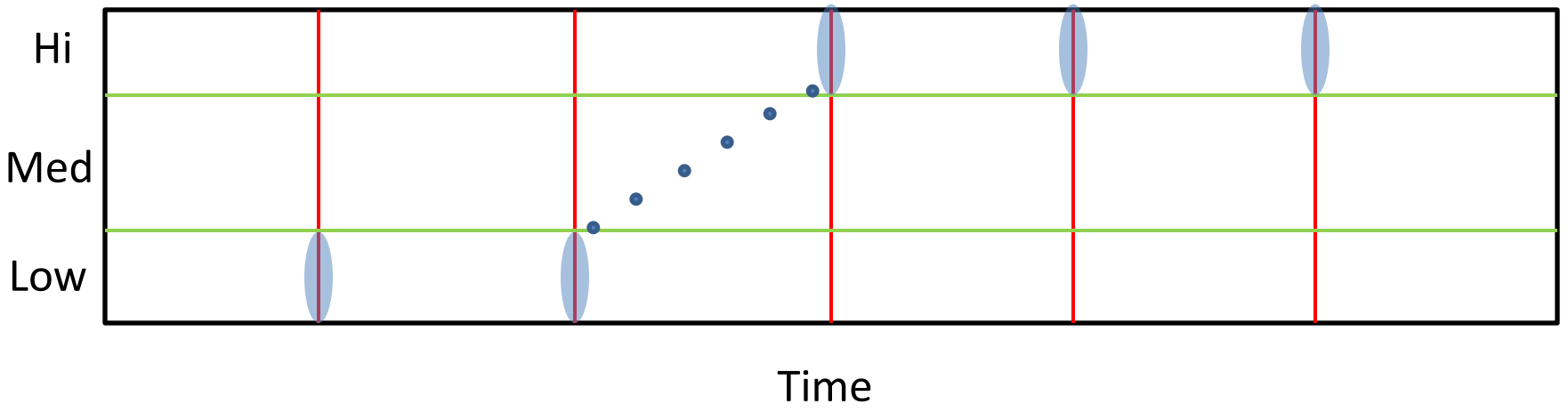
- **Either**, a medium clause appears in memory during **one** breakpoint between epochs,



- If $\#epochs = m, space = S$, probability is at most $mS(3/4)^w$, by a union bound.

Probability of Second Scenario

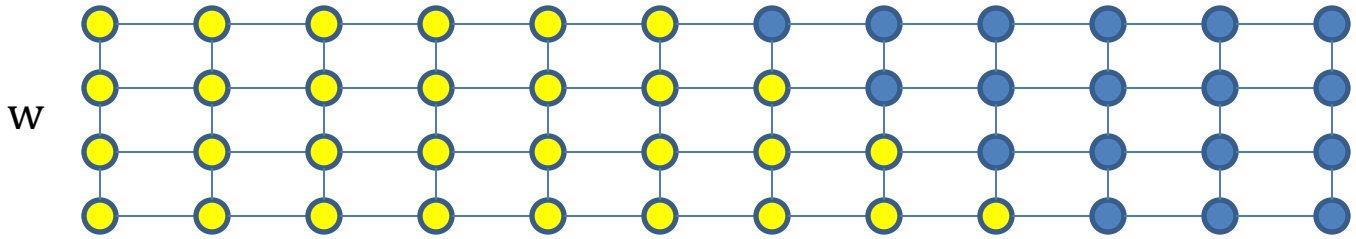
- **Or**, some epoch has k clauses of super increasing med. complexities, $k = \log {}^{l}/_{v}$.



- Idea: Proceed by union bound over all k-tuples of clauses in an epoch. What is the collective width of k clauses of such complexity?

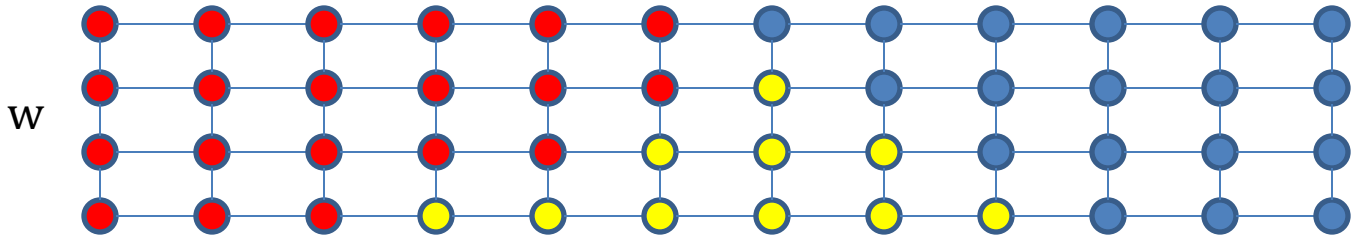
Isoperimetry in the Grid

- Observation: If we have k medium subsets of the grid of superincreasing sizes, have at least kw edges in the union of their boundaries.



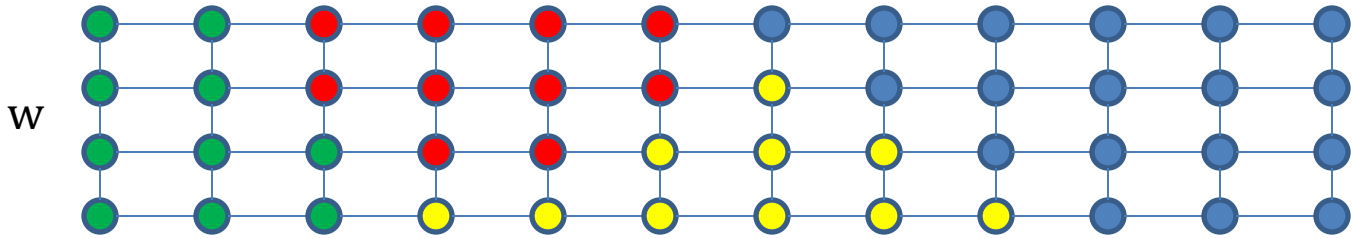
Isoperimetry in the Grid

- Observation: If we have k medium subsets of the grid of superincreasing sizes, have at least kw edges in the union of their boundaries.



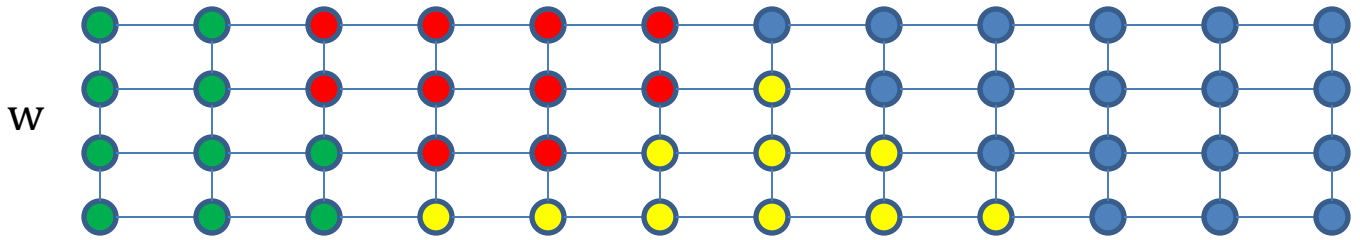
Isoperimetry in the Grid

- Observation: If we have k medium subsets of the grid of superincreasing sizes, have at least kw edges in the union of their boundaries.



Isoperimetry in the Grid

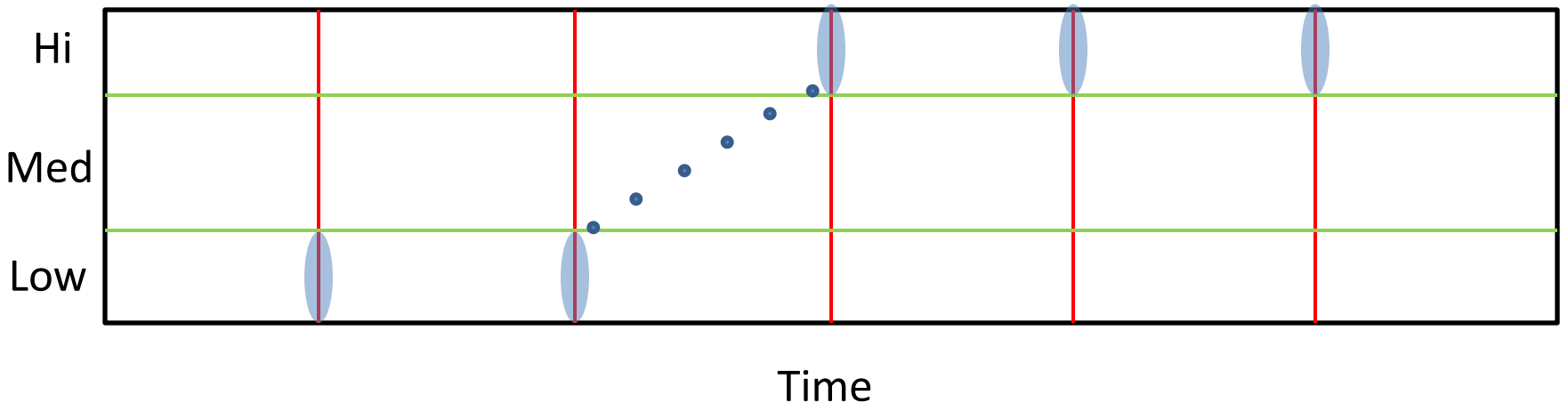
- Observation: If we have k medium subsets of the grid of superincreasing sizes, have at least kw edges in the union of their boundaries.



- Key Lemma: For any k clauses, the probability that their restrictions superincreasing medium complexities is $\exp(-\Omega(kw))$.

Two Possibilities

- **Or**, some epoch has $k = \log {}^{l}/_{v}$ clauses of super increasing med. complexities.



- Corollary: Probability of this is at most

$$m \cdot (T/m)^k \cdot \exp(-\Omega(kw))$$

Time Space Tradeoff

- Conclude:

$$1 \leq mS\exp\left(-\Omega(w)\right) + m(T/m)^k \exp(-\Omega(kw))$$

- We won't do the calculation, but this idea can give a nontrivial tradeoff of the form

$$\text{Size}\cdot\text{Space} \geq 2^{(2+c')w}$$

Full Result

- To get the full result, need to boost from a more elaborate first step argument, subdividing epochs recursively into subepochs.
- For convenience, we partition the medium complexity clauses into "levels" of the form

$$\{C : k < \mu(C) \leq 2k\}$$

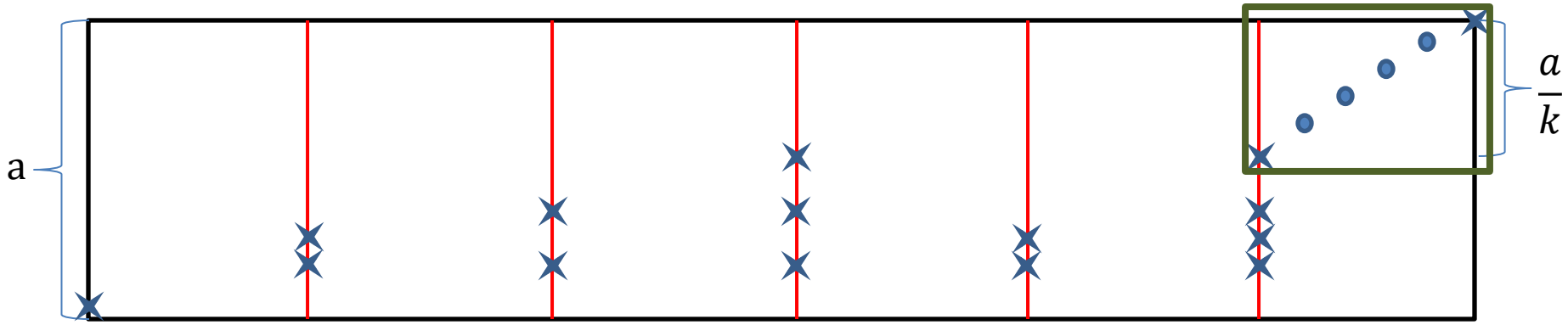
  for some k.
- So, clauses of superincreasing complexities are of distinct levels. As a short hand, we also say they are of distinct complexities.

Full Result

- Consider a proof of 1x grid, divide into m equal epochs, recursively divide those into m subepochs etc. to a recursive height of h.
- Claim: Let $g = \log(l/w)$ denote the total number of medium complexity levels.
  - Either, there is an epoch with at least k complexities at the breakpoints between its children,
  - Or, a bottom epoch has $\geq g/k^{h-1}$ complexities

Proof: Induction

"Zoom in" here and repeat the argument



- If an epoch contains a clause at the end of level b, but every clause at start is level $\leq b - a$, (so the epoch "makes a progress"),
- and the breakpoints of its m children epochs contain together $\leq k$ complexity levels,
- then some child epoch makes a/k progress.

Full Result

- Theorem: [BBI'12]

  For any size T space S refutation of Tseitin on a 2x grid of dimensions $w \times l, l \geq 4w^2$, we have

$$T \geq \left(\frac{\exp(\Omega(w))}{S}\right)^k,$$

  where $k^k = \log(l/w)$, that is, $k \geq \frac{\log\log w}{\log\log\log w}$.

  Moreover, $T, S \approx \exp(w)$, and $T \approx l^w, S \approx w \log l$ are achievable.

Polynomial Calculus

- In recent work of myself, Jakob Nordstrom, and Bangsheng Tang, we generalized this result to PCR, losing only constants.
- Starting Point:
  In a well known-result of [BGIP], PC-degree bounds are known for mod q Tseitin which basically match resolution width, as long as we work over a field different from q.

Polynomial Calculus

- Size lower bounds via Random Restrictions:
  - Begin with a Tseitin formula with {0,1} valued variables x_i. If a PCR proof has few monomials, there is a restriction such that it is low degree.
  - Apply a change of variables to {+1,-1} valued y_i. This doesn't increase the degree, but the Tseitin axioms become binomials. The proof can now be restructured to consist entirely of binomials. A medium complexity binomial must be high degree, for similar reasons as clauses, qed.

Polynomial Calculus

- Size Space Tradeoffs in [BBI] style?
  - Difficulties: Once we change variables, we lose control of Size, Space. Also, there are no meaningful space/degree tradeoffs for binomials.
- Solution in Brief:
  - Show that for any small collection of x_i monomials, it is unlikely that it is possible that its restriction translates to binomials of many different complexities.

Open Questions

- More than quasi-polynomial separations?
  - For Tseitin formulas upper bound for small space is only a log n power of the unrestricted size
  - Candidate formulas? Are these even possible?
- Tight result for Tseitin? This question has also spawned a difficult graph pebbling question, which we can discuss offline.
- Can we get tradeoffs for Cutting Planes? Monotone Circuits? Frege subsystems?

Thanks!

Final Tradeoff

- Tseitin formulas on $K_n \otimes P_l$ for $l = 2n^4$
  - are of size $\boldsymbol{2^{2n} \cdot poly(n)}$
  - have resolution refutations in Size, Space $\approx 2^{n^2}$
  - have $n^2 \log n$ space refutations of Size $2^{n^2 \log n}$
  - and any resolution refutation for Size T and Space S requires $T > (2^{0.58\, n^2}/S)^{\omega(1)}$

If space is at most $2^{n^2}/2$ then size blows up by a super-polynomial amount