Computability and Complexity 2022/23 March 29, 2023

Lecture 20

Lecturer: Jakob Nordström *Scribe: Susanna F. de Rezende*

1 Introduction to This Lecture

In this lecture we will review another one of the classic results in proof complexity, namely the lower bound by Haken [Hak85] showing that that the resolution proof system has no clue how to count. We start by explaining what we mean by this by writing down a (contradictory) CNF encoding of the well-known pigeonhole principle. The rest of the lecture will then be spent on proving that the length of resolution refutations of such CNF formulas must scale exponentially with the formula size.

2 Pigeonhole Principle Formulas

The *pigeonhole principle* states that $n+1$ pigeons do not fit into n pigeonholes if each pigeon needs its own hole. More formally, in its most basic version the pigeonhole principle states that there is no injective relation from $[n+1]$ to $[n]$. In order to be be able to investigate this principle from the point of view of proof complexity, we can encode the negation of the principle, i.e., the claim that $n+1$ pigeons in fact do fit into n pigeonholes with at most one pigeon per hole, as a family of (unsatisfiable) *pigeonhole principle (PHP) formulas* in conjunctive normal form.

We will define the PHP formulas for any number of pigeons $m > n$, although when m grows much larger than n it only makes the formulas easier to refute (which is intuitively clear, since the more pigeons we claim can be crammed into just n holes, the more obviously false the statement becomes). Once we have given the formal definition, for the rest of this lecture we will fix $m = n+1$.

For any positive integers m and n such that $m > n$ we define the formula PHP^m_n as follows. We have variables $x_{i,j}$, for $i \in [m], j \in [n]$, where we think of i as ranging over pigeons and j as ranging over pigeonholes. The intended interpretation of $x_{i,j}$ is that this variable is true if and only if pigeon i flies into hole j. We have clauses of two types:

Pigeon axioms For all $i \in [m]$ the clause

$$P^i = \bigvee_{j=1}^{n} x_{i,j} \tag{2.1a}$$

encodes that "pigeon i flies into some hole."

Hole axioms For all $i, i' \in [m]$, $i < i'$, and all $j \in [n]$ the clause

$$H^{i,i'}_j = \overline{x}_{i,j} \vee \overline{x}_{i',j} \tag{2.1b}$$

encodes that "pigeons i and i' do not both fly to hole j."

The formula PHP^m_n is defined as

$$PHP^m_n = \bigwedge_{i=1}^{m} P^i \wedge \bigwedge_{i'=2}^{m} \bigwedge_{i=1}^{i'-1} \bigwedge_{j=1}^{n} H^{i,i'}_j \, . \tag{2.2}$$

The pigeonhole principle formulas are probably hands-down *the* most studied formula family in all of proof complexity. There is even a (very readable) survey paper [Raz02] dedicated exclusively to this family of formulas.

PHP formulas come in several different flavours. If we want to think of pigeons flying into holes not as a relation but as a mapping, which is the way we usually reason about the pigeonhole principle, then we can add axiom clauses forbidding pigeons from going into multiple holes. And if we want to be fair to the holes, we can also require that every hole should get at least one pigeon. This yields the following additional types of clauses:

Functionality axioms For all $i \in [m]$ and all $j, j' \in [n]$, $j < j'$, the clause

$$F^i_{j,j'} = \overline{x}_{i,j} \lor \overline{x}_{i,j'} \tag{2.3a}$$

encoding that "pigeon i does not fly both to hole j and hole j'."

Onto axioms For all $j \in [n]$ the clause

$$S_j = \bigvee_{i=1}^{m} x_{i,j} \tag{2.3b}$$

encoding that "some pigeon flies to hole j."

By adding all clauses on the form (2.3a) and/or (2.3b) to the formula in (2.2), we obtain the *functional pigeonhole principle (FPHP)*, *onto pigeonhole principle*, or *onto functional pigeonhole principle (Onto-FPHP)*, respectively. Note that another way of describing Onto-FPHP formulas is that they claim that there is a perfect matching in a complete bipartite graph $K_{m,n}$ with m vertices on the left and n vertices on the right, and for this reason they are sometimes referred to as *perfect matching formulas*.

Despite the fact that PHP formulas encode such a basic combinatorial principle, and despite that they have been so heavily studied, there are still many (fascinating) open problems concerning these formulas in proof complexity. When it comes to the resolution proof system, however, PHP formulas are fairly well understood. In this lecture, we will focus on the "vanilla version" (2.2) of the PHP formulas for $m = n + 1$ pigeons and prove that they are exponentially hard for resolution. This result, due to Haken [Hak85], is one of the celebrated early theorems in proof complexity.

Theorem 2.1 ([Hak85]). *Resolution refutations of PHP^{n+1}_n require length* $\exp(\Omega(n))$.

Note that PHP^{n+1}_n has $\Theta(n^2)$ variables, $\Theta(n^3)$ clauses, and size $\Theta(n^3)$. Therefore, in terms of formula size $N = \Theta(n^3)$ Theorem 2.1 yields a lower bound of $\exp\left(\Omega\left(\sqrt[3]{N}\right)\right)$.

It is possible to show (and it is not so hard) that resolution never needs length more than $\exp(\mathrm{O}(N))$, where N is the formula size. One might ask whether there exist formulas such that any refutation must have exponential length measured in terms of formula size or whether perhaps any formula of large enough size N can be refuted in length $\exp(\mathrm{o}(N))$. The answer is that the upper bound is the correct one. It has been shown in, e.g., [Urq87, CS88, BW01] that there are formulas of size N with truly exponential lower bounds $\exp(\Omega(N))$ on resolution refutation length—i.e., tightly matching the worst-case $\exp(\mathrm{O}(N))$ upper bound up to constant factors in the exponent. But in this lecture we will be satisfied with proving lower bounds for pigeonhole principle formulas.

To prove Theorem 2.1 we need to show that there exists a $\delta > 0$ such that for large enough $n_0 \in \mathbb{N}^+$ it holds that if $n \geq n_0$, then any resolution refutation of PHP^{n+1}_n requires $2^{\delta n}$ steps. The challenge is that as $n \to \infty$ we get infinitely many possible refutations of PHP^{n+1}_n. We need to show that no such refutation can be shorter than $2^{\delta n}$. How can one prove such a thing? This is not obvious at all, and indeed Haken's paper [Hak85] is considered to be a major breakthrough. We will not quite follow Haken's paper, however, but will instead present a nicer and cleaner (and cuter) proof due to Pudlák [Pud00].

Pudlák presents the lower bound in terms of a two-person game. This game can be played over any unsatisfiable CNF formula F, but in this lecture we will define it only for PHP formulas.

3 Pudlák's Prosecutor–Defendant Game for PHP Formulas

Pudlák's game is played by two players: *Defendant*, who claims there is a way to fit $n + 1$ pigeons into n holes; and *Prosecutor*, who wants to convict Defendant of lying. This should be a clear-cut case, but we have two problems. To begin with, this is a jury trial, and the jury will only be convinced by obviously, explicitly contradictory answers. On top of that, after a long and successful career Prosecutor is starting to get old and forgetful, and so although he has prepared the case very carefully, in order to aid his memory he needs a book with super-explicit instructions how to cross-examine Defendant.

During the trial, Prosecutor will ask questions of the type "does pigeon i fly into pigeonhole j?" It is important to note that here i and j are not generic variables but always concrete numbers. Defendant answers "yes" or "no" to every question. Prosecutor then makes a note of Defendant's answer in a *record*. Unfortunately, Defendant can see Prosecutor's record, and so always has full information about exactly what Prosecutor knows at each point in time, and can use this when choosing how to answer. For reasons that will become clearer later Prosecutor can also purposely forget some notes on the record. If so, Defendant will know that a previously given answer has been forgotten and can choose to answer differently next time the same question is asked.

Every record R consists of the questions-and-answers that Prosecutor remembers at a particular point during play, and looks something like

$$R = \{(i_1, j_1, \text{yes}),\ (i_2, j_2, \text{no}),\ (i_3, j_3, \text{no}),\ (i_4, j_4, \text{no}),\ (i_5, j_5, \text{yes}),\ \ldots\} \tag{3.1}$$

(for concrete numbers i_ℓ, j_ℓ). Prosecutor's goal is to convince the jury and, as we already mentioned, this requires an explicit contradiction. This means Prosecutor has to be able to produce a record R containing some contradiction of one of the following types:

1. (i_1, j, yes) and (i_2, j, yes), $i_1 \neq i_2$, both belong to R, i.e., Defendant says that two different pigeons fly into the same pigeonhole; or

2. $(i, 1, \text{no}), (i, 2, \text{no}), \ldots, (i, n, \text{no})$ all belong to R, i.e., Defendant says that a pigeon does not fly into any pigeonhole.

Let us call such records *winning records*.

An *instruction book* (a.k.a. *Prosecutor strategy*) contains instructions what Prosecutor should do next based on the current record. For every record R that can be reached while Prosecutor plays according to this strategy, the book should contain a page with the record R together with an associated instruction what Prosecutor should do for his particular record. The instruction can be an *ask* instruction "ask whether pigeon i^* flies into hole j^*" or a *forget* instruction, for instance, "forget notes $\{(i_2, j_2, \text{no}), (i_5, j_5, \text{yes})\}$" (for the record in (3.1)). Observe that forget instructions can forget several notes, while ask instructions can only ask one question. It is important to note that Prosecutor can never forget *records* but only *notes* in records, and such a forget instruction creates another record, for which there must be another entry in the book with instructions of what to do.

As mentioned, an instruction book always has a page for every possible record that can arise during play. The record on the page must match exactly the current record (except that the order of the notes is irrelevant). That is, if there is an entry in the book with a record R and an instruction to ask whether pigeon i^* flies into hole j^*, then there have to exist two records $R \cup \{(i^*, j^*, \text{yes})\}$ and $R \cup \{(i^*, j^*, \text{no})\}$ taking care of the two possible answers from Defendant. We say that a Prosecutor strategy is *complete* if Prosecutor can win against any Defendant using the strategy, meaning that however Defendant answers, the cross-examination will eventually end with a winning record. A strategy can be incomplete, however, in that Prosecutor gets stuck in infinite play if Defendant answers in particular ways. For instance, maybe Prosecutor chooses to forget information that later turns out to be crucial to force a contradiction.

Why would Prosecutor forget? The reason to forget is to reduce number of pages in the instruction book. It is not very impressive if Prosecutor needs to flip through several thousands of pages just to figure out which question to ask next. Prosecutor's objective is to minimize the size of a complete strategy/instruction book, i.e., the number of distinct records/pages in it.

4 Extracting a Prosecutor Strategy from a Resolution Refutation

Good, so the Pudlák game is clearly a cute game, but why are we doing this? The reason is that a short resolution refutation gives a complete Prosecutor strategy with few records. Hence, if we prove that no complete Prosecutor strategy can have few records, then we can conclude that there is no short resolution refutation.

Lemma 4.1. *If there exists a resolution refutation $\pi : PHP_n^{n+1} \vdash \bot$ of length $L(\pi) = L$, then there exists a complete Prosecutor strategy in the Prosecutor–Defendant game on PHP_n^{n+1} with $\mathrm{O}(L)$ records.*

Proof. Given a resolution refutation $\pi : PHP_n^{n+1} \vdash \bot$ in length $L(\pi) = L$, we will show how to construct a Prosecutor strategy with $\mathrm{O}(L)$ records.

Consider the DAG representation G_π of π, where every vertex v is labelled by a clause in the refutation. Prosecutor constructs the strategy by walking from the sink, which is labelled by $\bot$, towards the sources. The contradictory empty clause $\bot$ was derived by resolving $x_{i,j}$ and $\overline{x}_{i,j}$ for some $x_{i,j}$, and Prosecutor will let this correspond to a record with the empty partial assignment together with an instruction to ask whether pigeon i flies to hole j. Prosecutor then moves to the clause/vertex falsified by Defendant's answer, i.e., if Defendant answers "yes" Prosecutor moves to vertex $\overline{x}_{i,j}$ and otherwise to $x_{i,j}$.

In general, Prosecutor will maintain the invariant that at the clause/vertex D_i in the refutation it holds that the corresponding record contains the minimal partial assignment falsifying D_i. Suppose that we have $D_i = B \lor C$ where this clause was derived by resolving the premises $B \lor x_{i,j}$ and $C \lor \overline{x}_{i,j}$ over $x_{i,j}$. Then the instruction associated to that record is to ask whether pigeon i flies to hole j and move to the clause falsified by Defendant's answer. In a second step, Prosecutor then forgets any notes/assignments that are not needed to falsify the premise clause. This clearly maintains the invariant.

In this way, Prosecutor can construct a strategy by processing all vertices in G_π, and with a little bit of care we can see that every clause in π corresponds to at most two records. Any sequence of answers from Defendant will yield a walk backwards through G_π. Sooner or later, Prosecutor will reach a source labelled by an axiom of PHP_n^{n+1}. By the invariant, this axiom clause is falsified by the record at that point in the game. If we look at the way explicit contradictions were defined above, it so happens that the record falsifying the axiom clauses constitutes an explicit contradictions (namely, a falsified hole axiom $H_j^{i_1,i_2}$ corresponds to a contradiction of type 1 and a falsified pigeon axiom P^i corresponds to a contradiction of type 2). The lemma follows. □

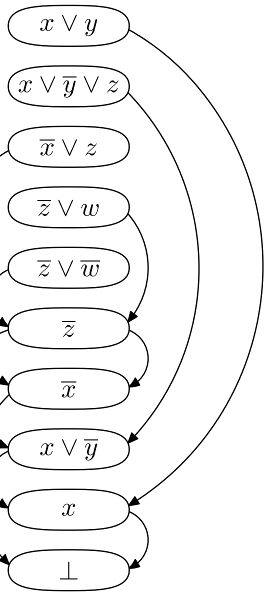
Example 4.2. Let us illustrate Lemma 4.1 by applying it to the CNF formula and resolution refutation that we had as an example a couple of lectures back, and that we illustrate again in Figure 1 for the convenience of the reader. Strictly speaking, we cannot do this since we only defined the Prosecutor–Defendant game for PHP formulas, but the reader most likely has already realized that the question "does pigeon i fly to hole j?" is just a complicated way of saying that Prosecutor asks Defendant about the value of the variable $x_{i,j}$, and that "explicit contradictions" are just falsified axiom clauses. So let us show how the refutation in Figure 1 helps Prosecutor to win this generalized game played on the formula refuted in Figure 1

Prosecutor starts by asking about the value of x. Suppose Defendant answers 1, i.e., true. Then Prosecutor writes down $x = 1$ in the record, moves to the clause $\overline{x}$, and asks about z. Defendant would immediately lose if answering 0, since this would falsify the axiom $\overline{x} \lor z$, so suppose the answer is 1. This makes Prosecutor write down $z = 1$ and move to the clause $\overline{z}$. Since information about x is no longer needed, Prosecutor now forgets the assignment $x = 0$, keeping only $\{z = 1\}$ as the record. Finally, Prosecutor asks about w. Now Defendant is in trouble, since for an answer $w = 0$ Prosecutor obtains a contradiction to the axiom clause $\overline{z} \lor w$ and an answer $w = 1$ falsifies $\overline{z} \lor \overline{w}$.

It is important to understand that although in any single round of the game Prosecutor might only need a small number of records, the strategy needs to be able to deal with all scenarios that can arise and so will in general contain many more records. To see another possible way of playing the game, consider when Prosecutor starts by asking about the value of x, as will always be the case in this strategy, but gets the answer 0 from the defendant. Then Prosecutor writes down $x = 0$ in the record, moves to the clause x,

1.	$x \lor y$	Axiom
2.	$x \lor \overline{y} \lor z$	Axiom
3.	$\overline{x} \lor z$	Axiom
4.	$\overline{z} \lor w$	Axiom
5.	$\overline{z} \lor \overline{w}$	Axiom
6.	$\overline{z}$	Res$(4, 5)$
7.	$\overline{x}$	Res$(3, 6)$
8.	$x \lor \overline{y}$	Res$(2, 6)$
9.	x	Res$(1, 8)$
10.	$\bot$	Res$(7, 9)$

(a) Resolution refutation as an annotated list.



(b) Resolution refutation as a DAG.

Figure 1: Example resolution refutation.

and asks about y. Defendant would immediately lose if answering 0, since this would falsify the axiom $x \lor y$, so suppose the answer is 1. This makes Prosecutor write down $y = 1$, move to the clause $x \lor \overline{y}$, and ask about z. Again, Defendant would lose immediately if answering 0, so let us assume the answer is 1. Given this answer, Prosecutor moves to the clause $\overline{z}$, forgets the assignments $\{x = 0, y = 1\}$ and keeps only $\{z = 1\}$. Note that this is a record we encountered also in the previous round of the game, but this time we reached it in a different way. The question asked for this record is the same, namely about the value of w, and we know from the discussion above that both ways of answering will lead to a winning position for Prosecutor.

5 Resolution Lower Bounds from Clever Defendant Strategies

In view of Lemma 4.1, all that is left to prove in order to obtain the lower bound in Theorem 2.1 is that any complete Prosecutor strategy for PHP^{n+1}_n must require a large number of records.

Lemma 5.1. *There is a $\delta > 0$ such that for large enough $n \in \mathbb{N}^+$ any Prosecutor instruction book for PHP^{n+1}_n must have at least $2^{\delta n}$ pages/records.*

To prove this lemma it is not enough to just consider a single round of play, as we already discussed in Example 4.2, since Prosecutor can always win a game played on PHP^{n+1}_n by asking at most $n(n+1)$ questions to obtain a complete truth value assignment. Thus, we have to consider several rounds and prove that in order to always win, Prosecutor needs to be able to deal with exponentially many scenarios/records. In order to achieve this, we must devise clever Defendant strategies.

What might be a smart strategy for Defendant? We can start by observing that, intuitively, answering “yes” to a question “does pigeon i fly to hole j?” gives away lots of information (all there is to know about pigeon i), whereas answering “no” is not so informative (it just means that there is one less hole where the pigeon could have gone, but there will typically be many more remaining). So perhaps answering “no” as often as possible might be a good Defendant strategy to attempt?

Defendant strategy 1. Defendant answers “no” to questions “does pigeon i fly to hole j?” as long as there is some other free hole for pigeon i and only answers “yes” when pigeon i has to go into hole j to avoid immediate contradiction.

Unfortunately, this idea does not work—it is not hard to see that Prosecutor has a strategy with just $\mathrm{O}(n)$ records to win in this case. Prosecutor can ask about pigeon 1 and all pigeon holes j except the nth one, and Defendant will always answer "no". Then Prosecutor asks about pigeon w and all pigeon holes j except the nth one, and again the Defendant answers "no" to all questions. But now asking about pigeon 1 and 2 and hole n leads to some winning record. (Note that this will not be a complete strategy, though, but it is sufficient to win against this Defendant and so this idea cannot yield the lower bound that we are after.)

If we want to force Prosecutor to be able to deal with exponentially many different cases, one idea could be to use randomization to have Defendant play according to one of exponentially many different partial matchings of pigeons to holes.

Defendant strategy 2. Defendant chooses n out of $n+1$ pigeons randomly and matches pigeons to holes randomly, yielding a matching $\mathcal{M}$. Defendant then answers all questions consistently with $\mathcal{M}$ (i.e., responds "yes" to the question "does pigeon i fly to hole j?" if and only if i is matched to j in $\mathcal{M}$).

This makes Prosecutor's life slightly harder, but ultimately does not work either —Prosecutor has a counter-strategy with $\mathrm{O}(n^2)$ records. (We leave this as an exercise.) But what about mixing the two strategies?

Defendant strategy 3 (successful). Defendant chooses $n/4$ pigeons uniformly at random and assigns them to $n/4$ pigeon holes chosen uniformly at random to obtain a partial matching that we will denote $\mathcal{M}_{\text{init}}$. (We assume for simplicity that 4 divides n, and also that all divisions of n that will follow produce integers—this is without loss of generality since it is not hard to adjust the details to deal with this formally).

Let $\mathrm{dom}(\mathcal{M})$ denote the set of pigeons matched by a partial matching $\mathcal{M}$. Then $\mathcal{M}$ corresponds to a partial truth value assignment $\rho_{\mathcal{M}}$ in the natural way by letting

$$\rho_{\mathcal{M}}(x_{i,j}) = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{M} \\ 0 & \text{if } (i,j') \in \mathcal{M} \text{ for } j' \neq j, \\ * & \text{if } i \notin \mathrm{dom}(\mathcal{M}) \text{ (where $*$ just denotes that the variable is left unassigned).} \end{cases} \tag{5.1}$$

We say that a partial matching $\mathcal{M}$ is *consistent* with a record R if for every note (i,j,yes) in R it holds that $\rho_{\mathcal{M}}(x_{i,j}) = 1$ and for every note (i,j,no) in R it holds that $\rho_{\mathcal{M}}(x_{i,j}) \neq 1$ (i.e., $\rho_{\mathcal{M}}(x_{i,j}) = 0$ or $\rho_{\mathcal{M}}(x_{i,j}) = *$). Throughout the game, Defendant will maintain a partial matching $\mathcal{M}$ such that $\mathcal{M} \supseteq \mathcal{M}_{\text{init}}$ and $\rho_{\mathcal{M}}$ is consistent with the current Prosecutor record R (but in general $\mathcal{M}$ will contain extra information that is not known by Prosecutor).

It Prosecutor asks the question "does pigeon i fly to hole j?" Defendant answers "yes" if $(i,j) \in \mathcal{M}$ and "no" otherwise, and then considers whether to update the matching $\mathcal{M}$ as described next.

Let us say that record R *assigns* hole j to pigeon i if R contains the note (i,j,yes) and that it *prohibits* hole j for pigeon i if it contains the note (i,j,no). If for the pigeon i that Prosecutor just asked about it holds that $i \notin \mathrm{dom}(\mathcal{M})$ (in which case Defendant answered "no") then Defendant checks the number of prohibited holes for i in the current record. If this number is at least $n/2$, Defendant chooses some hole j^* not prohibited for i and consistent with $\mathcal{M}$ (i.e., not occupied by any other pigeons) and sets $\mathcal{M} = \mathcal{M} \cup \{(i,j^*)\}$ (we can let j^* be the smallest index of such a hole if we want a concrete number). If it is not possible to update the partial matching $\mathcal{M}$ to accommodate pigeon i, then Defendant gives up.[1]

If Prosecutor forgets some notes in the record R to obtain R', then Defendant looks at every pigeon $i \in \mathrm{dom}(\mathcal{M}) \setminus \mathrm{dom}(\mathcal{M}_{\text{init}})$ such that some note $(i,j,\mathrm{yes/no})$ was forgotten. If i does not have an assigned hole according to R' and the number of holes prohibited for i by R' is strictly less than $n/2$, then Defendant removes i from the matching $\mathcal{M}$. Note, however, that pigeons in the initial random matching $\mathcal{M}_{\text{init}}$ are never removed.

Let us say that a pigeon i is *thoroughly examined* in a record R if R assigns a hole to i or forbids at least $n/2$ holes for i, i.e., if R contains (i,j,yes) for some hole j or $(i,j_1,\mathrm{no}), (i,j_2,\mathrm{no}), \ldots, (i,j_\ell,\mathrm{no})$

[1] Strictly speaking, we do not have any "give up" move for Defendant, but we could just define it as Defendant answering 0 to any question until the game is over, or whatever else fixed way of play that we like.

for at least $n/2$ distinct holes j_ℓ. What this means is that Prosecutor has been forced to write down a lot of information about pigeon i. We claim that in order to win Prosecutor has to have lots of information about lots of pigeons on record simultaneously.

Lemma 5.2. *Before Defendant gets convicted when playing according to Strategy 3, Prosecutor must create a record R with $n/4$ thoroughly examined pigeons.*

Proof. As long as Defendant follows the strategy the answers are consistent with a partial matching of pigeons to holes, and hence cannot contradict any pigeon or hole constraints. This means that before conviction an update of $\mathcal{M}$ to accommodate some pigeon $i^* \notin \mathrm{dom}(\mathcal{M})$ must have failed. How did this happen? It must have been the case that at that point in time the current record R prohibited exactly $n/2$ holes for pigeon i^* but that there was no available free hole in Defendant's partial matching $\mathcal{M}$.

Since there are no free holes, all the $n/2$ holes that are not prohibited for pigeon i^* in R must be matched to another pigeon in $\mathcal{M}$. Thus, we have $|\mathrm{dom}(\mathcal{M})| \geq n/2$, which implies

$$|\mathrm{dom}(\mathcal{M}) \setminus \mathrm{dom}(\mathcal{M}_{\text{init}})| \geq |\mathrm{dom}(\mathcal{M})| - |\mathrm{dom}(\mathcal{M}_{\text{init}})| \geq n/2 - n/4 = n/4\,. \tag{5.2}$$

But a pigeon i is in $\mathrm{dom}(\mathcal{M}) \setminus \mathrm{dom}(\mathcal{M}_{\text{init}})$ only if R either assigns a hole to pigeon i or prohibits $n/2$ holes for it, i.e., only if pigeon i is thoroughly examined. The lemma follows. □

We say that a record R with at least $n/4$ thoroughly examined pigeons is *informative*. By Lemma 5.2, for any choice of $\mathcal{M}_{\text{init}}$ a round of play passes through an informative record R. This record has to be consistent with $\mathcal{M}_{\text{init}}$, since Defendant always gives answers that are consistent with $\mathcal{M}_{\text{init}}$. We want to prove that any complete instruction book (i.e., one with which Prosecutor can win against any Defendant) must contain at least $2^{\delta n}$ distinct informative records.

We will prove the lower bound on the number of distinct informative records by showing that during a particular round of the game it is very unlikely that we see any fixed informative record.

Claim 5.3. For any arbitrary fixed informative record R it holds that

$$\Pr[R \text{ and } \mathcal{M}_{\text{init}} \text{ are consistent}] \leq 2^{-\delta n}\,,$$

where the probability is over the random choice of an initial partial matching $\mathcal{M}_{\text{init}}$.

If we can prove Claim 5.3, then we are done. To see this, note that for any $\mathcal{M}_{\text{init}}$ that Defendant chooses, Prosecutor will build at least one informative record R before winning. In the first informative record, Defendant's inductive update of $\mathcal{M}$ has not yet failed, and so Defendant has answered consistently with $\mathcal{M}_{\text{init}}$ up to this point and R must be consistent with $\mathcal{M}_{\text{init}}$. This yields the sequence of inequalities (where the probabilities are over $\mathcal{M}_{\text{init}}$):

$$\begin{aligned}
1 &= \Pr[\exists \text{ informative } R \text{ such that game passes through } R] && \text{[by Lemma 5.2]} \\
&\leq \sum_{\text{informative } R} \Pr[\text{game passes through } R] && \text{[by a union bound]} \\
&\leq \sum_{\text{informative } R} \Pr[R \text{ consistent with } \mathcal{M}_{\text{init}}] && \text{[by the reasoning above]} \qquad (5.3)\\
&\leq 2^{-\delta n} \cdot (\#\text{ informative records } R) && \text{[by Claim 5.3]} \\
&\leq 2^{-\delta n} \cdot (\text{total \# records } R)\,,
\end{aligned}$$

which is just another way of saying that Prosecutor's instruction book must contain at least $2^{\delta n}$ records. So all that remains for us to do is to establish Claim 5.3.

Proof of Claim 5.3. Let $\mathcal{I}_R$ denote the set of thoroughly examined pigeons in R. For any fixed pigeon, the probability that it is chosen as one of the matched pigeons in $\mathcal{M}_{\text{init}}$ is $1/4$.[2] We know that $|\mathcal{I}_R| \geq n/4$

[2] If we want to be really picky, the probability is actually $\frac{n/4}{n+1} = 1/(4 + \frac{1}{n})$, but again it is not hard to adjust the calculations that will follow to deal with this.

and thus the expected size of the intersection $\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})$ is

$$\mathrm{E}[|\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})|] = \sum_{i \in \mathcal{I}_R} \Pr[i \in \mathrm{dom}(\mathcal{M}_{\text{init}})] \geq \frac{n}{16} \tag{5.4}$$

by linearity of expectation. Not only is the expected size of the intersection linear in n, but it can also be shown that the actual size of the intersection is extremely likely to be close to the expected value. By something that is called *concentration of measure*, and that we will talk a little bit more about below, it follows that except with exponentially small probability it holds that

$$|\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})| \geq \frac{n}{32} \, . \tag{5.5}$$

For now, let us just accept that this is so and continue our line of argument.

Let $\mathcal{A}$ be the event "$\mathcal{M}_{\text{init}}$ and R are consistent" and let $\mathcal{B}$ be the event "$|\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})| < \frac{n}{32}$". To prove the lower bound in Claim 5.3 we can use the easy bound

$$\begin{aligned} \Pr[\mathcal{A}] &= \Pr\left[\mathcal{A}\middle|\mathcal{B}\right] \cdot \Pr[\mathcal{B}] + \Pr\left[\mathcal{A}\middle|\overline{\mathcal{B}}\right] \cdot \Pr[\overline{\mathcal{B}}] \\ &\leq \Pr[\mathcal{B}] + \Pr\left[\mathcal{A}\middle|\overline{\mathcal{B}}\right] \end{aligned} \tag{5.6}$$

from probability theory to deduce that

$$\begin{aligned} &\Pr[\mathcal{M}_{\text{init}} \text{ and } R \text{ are consistent}] \leq \\ &\quad \Pr\left[|\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})| < \frac{n}{32}\right] + \Pr\left[\mathcal{M}_{\text{init}} \text{ and } R \text{ consistent} \,\middle|\, |\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})| \geq \frac{n}{32}\right] \end{aligned} \tag{5.7}$$

So it is enough to show that

$$\Pr[\mathcal{B}] = \Pr\left[|\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})| < \frac{n}{32}\right] \leq 2^{-\delta' n} \tag{5.8}$$

for some $\delta' > 0$ and that

$$\Pr\left[\mathcal{A}\middle|\overline{\mathcal{B}}\right] \leq \Pr\left[\mathcal{M}_{\text{init}} \text{ and } R \text{ consistent} \,\middle|\, |\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})| \geq \frac{n}{32}\right] \leq 2^{-\delta'' n}, \tag{5.9}$$

for some $\delta'' > 0$ to establish Claim 5.3 that $\Pr[\mathcal{M}_{\text{init}}$ and R are consistent$] \leq 2^{-\delta' n} + 2^{-\delta'' n} \leq 2^{-\delta n}$, for some appropriately chosen $\delta > 0$ depending on δ' and δ'', from which Lemma 5.1 follows.

The probability in (5.8), which is the concentration of measure claimed above, can be estimated as

$$\begin{aligned} \Pr\left[|\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})| < \frac{n}{32}\right] &\leq \frac{\sum_{i=0}^{n/32-1} \binom{n/4}{i}\binom{n+1-n/4}{n/4-i}}{\binom{n+1}{n/4}} \\ &\leq \frac{\frac{n}{32}\binom{n/4}{n/32}\binom{3n/4+1}{7n/32}}{\binom{n+1}{n/4}} \\ &\leq 2^{-\delta' n} \end{aligned} \tag{5.10}$$

for some suitably chosen $\delta' > 0$, where the second-to-last inequality follows since the numerator is maximized for large i and the last inequality follows from a series of calculation using Stirling's formula

$$\sqrt{2\pi m}\left(\frac{m}{e}\right)^m e^{\frac{1}{12m+1}} < m! < \sqrt{2\pi m}\left(\frac{m}{e}\right)^m e^{\frac{1}{12m}} \tag{5.11}$$

which we will not perform here, because they are fairly standard and also quite tedious (but they are a useful exercise for readers who have not seen concentration of measure calculations before).

Instead of doing the formal calculations, let us give an intuitive argument why (5.8) holds. We obtain pigeons in $\mathcal{I}_R \cap \mathrm{dom}(\mathcal{M}_{\text{init}})$ by picking pigeons to include in $\mathrm{dom}(\mathcal{M}_{\text{init}})$ randomly one by one. Every time, the chance of hitting a pigeon in $\mathcal{I}_R$ is approximately $1/4$ and we do this $n/4$ times. If these events were independent, then they would be a series of coin flips of a biased coin, and it is well known that if we flip ℓ times a coin that has probability p of coming up heads, then the number of heads we actually see will be sharply concentrated around the expectation $p\ell$ (which in our experiment here is $n/16$). This is *not* a formal argument, since in our experiment we do not have independent events, but the dependencies are not too bad and so "morally" the argument above still works. However, the actual calculations to prove the inequality in (5.10) would just take too much time and effort for us to do at this point.

Let us instead proceed to prove the bound in (5.9), where we are giving a lower bound on the probability that $\mathcal{M}_{\text{init}}$ and R are consistent under the assumption that $|\mathcal{I}_R \cap \operatorname{dom}(\mathcal{M}_{\text{init}})| \geq n/32$. For every pigeon $i \in \mathcal{I}_R$ it holds that the record R either

1. assigns a hole j to pigeon i, or

2. prohibits $n/2$ distinct holes $j_1, j_2, \ldots, j_{n/2}$ for pigeon i.

In order for R and $\mathcal{M}_{\text{init}}$ to be consistent, the pigeons in $\mathcal{I}_R \cap \operatorname{dom}(\mathcal{M}_{\text{init}})$ must have been randomly matched in $\mathcal{M}_{\text{init}}$ consistently with these restrictions. Consider the pigeons in $\mathcal{I}_R \cap \operatorname{dom}(\mathcal{M}_{\text{init}})$ one by one. Let us say that a pigeon $i \in \operatorname{dom}(\mathcal{M}_{\text{init}})$ is *consistent* with R) if for the pair $(i, j) \in \mathcal{M}_{\text{init}}$ it holds that the note (i, j, yes) is consistent with R.

Let us first give a quick argument that conveys the intuition. Pigeons of type 1 have to hit exactly the right hole in $\mathcal{M}_{\text{init}}$, so the probability for such pigeons of being consistent is approximately $1/n \ll 1/2$. Pigeons of type 2 have to avoid half of the holes, so the probability of being consistent for them is at most $1/2$. Since the number of pigeons that have to be mapped consistently is at least $n/32$, if all of these events were independent, then we would get a total probability of being correct of at most $(1/2)^{n/32}$ which is exponentially small. Done!

Well, actually we are not done, since unfortunately we do not have independence. Calculating more carefully (but not estimating very precisely, because we do not need to), we let the event $\mathcal{C}_i$ be "the ith pigeon in $\mathcal{I}_R \cap \operatorname{dom}(\mathcal{M}_{\text{init}})$ is mapped consistently with R by $\mathcal{M}_{\text{init}}$." We use another standard fact

$$\Pr\left[\bigcap_{i=1}^{\ell} \mathcal{C}_i\right] = \prod_{i=1}^{\ell} \Pr\left[\mathcal{C}_i \;\middle|\; \bigcap_{j=1}^{i-1} \mathcal{C}_j\right] \tag{5.12}$$

from probability theory and derive for the $(i+1)$st pigeon that

1. the probability that an $(i+1)$st pigeon of type 1 hits the right hole given that previous pigeons are consistent is at most $\frac{1}{n-i}$, and

2. the probability that an $(i+1)$st pigeon of type 2 avoids $n/2$ prohibited holes given that previous pigeons are consistent is at most $\frac{n/2}{n-i}$.

Since $i \leq n/4$, both probabilities above are less than $2/3$ and we can apply (5.12) to obtain that the probability that all pigeons are consistent is less than $(2/3)^{n/32} \leq 2^{\delta'' n}$ for some $\delta'' > 0$. This establishes the inequality in (5.9). As discussed above, we can now conclude the proof of Claim 5.3 by plugging (5.8) and (5.9) into (5.7). □

6 Recap of Proof

Just to review quickly what happened in this lecture, we proved exponential lower bounds on the length of resolution refutations of pigeonhole principle (PHP) formulas. We presented our proof in the language of Pudlák's Prosecutor–Defendant game. In this game Defendant claims to have a satisfying assignment for an unsatisfiable CNF formula F and Prosecutor interrogates Defendant about what values are given to the variables in F according to this assignment. Prosecutor can always win this game, since F is unsatisfiable, but we proved that there exist good strategies for Defendant that forces Prosecutor to be able to deal with many different scenarios, i.e., many different partial assignments to $Vars(F)$, which implies lower bounds on resolution refutation length (since a short refutation can be used by Prosecutor to construct a strategy with few records/scenarios).

The key idea in the Defendant strategy for PHP formulas is to pick uniformly at random a matching of $n/4$ pigeons into $n/4$ holes and answer any questions from Prosecutor consistently with this matching. Clearly, there are exponentially many different ways to choose such a matching. What we just showed is that in any complete strategy Prosecutor has to write down a noticeable fraction of (information about) the random matching chosen by Defendant. This implies that Prosecutor has to be able to deal with exponentially many mutually inconsistent partial truth value assignments, which means that the strategy needs to have an exponential number of records.

Our proof used the language of probability theory to show that any particular Prosecutor record has only exponentially small probability of being helpful for Prosecutor for a particular random matching chosen by Defendant. This is just another way of implementing a counting argument saying that Prosecutor needs exponentially many records (since for any random matching there is a helpful record, so the probabilities over all such records has to sum up to 1). It is often the case in proof complexity (and in computational complexity in general) that counting arguments are expressed in terms of probabilities in this way, so it can be good to try to become used to this way of thinking.

References

[BW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. *Journal of the ACM*, 48(2):149–169, March 2001. Preliminary version in *STOC '99*.

[CS88] Vašek Chvátal and Endre Szemerédi. Many hard examples for resolution. *Journal of the ACM*, 35(4):759–768, October 1988.

[Hak85] Armin Haken. The intractability of resolution. *Theoretical Computer Science*, 39(2-3):297–308, August 1985.

[Pud00] Pavel Pudlák. Proofs as games. *American Mathematical Monthly*, pages 541–550, 2000.

[Raz02] Alexander A. Razborov. Proof complexity of pigeonhole principles. In *5th International Conference on Developments in Language Theory, (DLT '01), Revised Papers*, volume 2295 of *Lecture Notes in Computer Science*, pages 100–116. Springer, July 2002.

[Urq87] Alasdair Urquhart. Hard examples for resolution. *Journal of the ACM*, 34(1):209–219, January 1987.