A Switching Lemma Primer[*]

Paul Beame
University of Toronto[†]
10 King's College Road
Toronto, Ontario
Canada M5S 1A4
beame@cs.washington.edu

April 94

Abstract

We present simplified proofs for a variety of switching lemmas. Our arguments extend recent switching lemma improvements to show, using simple counting techniques, that under restriction, DNF (resp. CNF) formulas can be represented by small height decision trees. We also show how the lemmas are applied in different contexts. In addition to reproving existing results, we also present a proof of a new switching lemma for restrictions that are partial q-matchings.

1 Introduction

Arguments using restrictions – partial assignments to input variables – to simplify unbounded fan-in Boolean circuits have been quite successful for obtaining lower bounds on circuit size and depth ([FSS81, Ajt83, Sip83, Yao85, Hås87, Ajt89, Lyn86, Bea90]), oracles to separate complexity classes ([Yao85, Hås87, Cai86, Ko91], see also [FSS81, Sip83]), lower bounds on time, processors, and memory of PRAMs ([BH89, Bea90]), as well as on the complexity of proofs in bounded-depth proof systems ([Ajt88, Ajt90, BPU91, BIK$^+$92, KPW91, PBI93, BP93].) The combinatorial essence of these arguments is what is known as a *switching lemma* which is used to show that an AND of small OR's can be written (possibly approximately) as an OR of small AND's (or vice versa) if an appropriate restriction is applied. These switching lemmas allow one to reduce the depth of the formulas by 1 at the expense of reducing the number of variables. The key to their application

[*] Research supported by NSF grant CCR-8858799.

[†] On leave from the University of Washington.

is that they simplify the formulas without simplifying the function being computed too much. To this end, one chooses a family of restrictions that is tailored to the function being computed and argues that some member of the family has the desired properties.

Although switching lemmas were found originally by Furst, Saxe, and Sipser [FSS81] (and independently in a different guise by Ajtai [Ajt83],) the model for the most powerful of these switching lemmas is that of Håstad [Hås87]. The argument uses the probabilistic method – one argues that the probability that a restriction from the family fails to have the desired properties is strictly less than 1. In the argument one considers, for example, an OR of small AND's (DNF formula with short terms) and considers each term in turn. The basic idea is that a term that is falsified by a restriction does not contribute any variables to the AND of small OR's and for each term that is not falsified it is more likely that the term is satisfied (and thus the whole formula is fixed to a constant) than that any variable is contributed in the AND of small OR's.

One complication in Håstad's argument is that one must deal with the bias induced on the probability space by the observations of terms that are falsified by the restriction. To handle this in as simple a manner as possible, Håstad in fact proves the switching lemma conditioned on the event that some arbitrary function's value is fixed to 0. He then argues that such conditioning can only bias the outcome in his favour. For the fully independent restrictions that he first considers this is argument is fairly easy. However, with increased complexity of the families of restrictions, such arguments become increasingly non-trivial. Furthermore, because the argument is recursive in nature, in some cases there are other conditionings that complicate things further.

Recently there has been a simplification of the Håstad technique that is based on a similar intuition but avoids this kind of conditioning. Woods (private communication) observed that by considering the unwinding of the recursive argument (which he viewed as a game between two players, one choosing the long term and the other choosing the restriction) and observing only part of what happens in the recursive argument one could obtain the probability bounds without explicitly handling the bias induced by an arbitrary function's value being fixed to 0.

Then, Razborov [Raz93] described a variation on the Håstad technique that allows one to obtain similar bounds to the conditional probability argument using only a counting argument. Again the critical feature of this variation is that one never explicitly considers the bias caused by the fact that certain terms are falsified by the restriction but now it also avoids the game-theoretic structure of Woods' argument. The formal argument is that a bad restriction – one under which the input DNF formula is not sufficiently simplified – can be mapped to an element of a small set in such a way that knowledge of the formula permits one to reconstruct the original bad restriction from the image of this map and thus the number of such bad restrictions is small. The actual calculations in the argument are quite similar to those of the conditional probability argument if one were to assume that the conditioning did not cause difficulty. One other interesting difference is that, whereas in the conditional probability argument the conditioning on the value of an arbitrary function being forced to 0 forces one to allow the number of unset variables to vary, in the counting argument it is actually advantageous to fix the number of unset variables.

In the following, we present the arguments for several switching lemmas for varied probability distributions, some of which require new variations on the structure of the counting argument. Razborov followed Håstad's original proof by showing that with high probability a DNF formula with short terms has only short maxterms after restriction.

In fact, Håstad's argument, as is the case with many of the switching lemma arguments mentioned above, more naturally proves a statement of the form that with high probability a DNF formula with short terms has a small height decision tree after restriction. This is a slightly stronger statement than the standard switching lemma phrasing since a small height decision tree allows one to obtain a short DNF formula for the negation of the formula which is essentially what is desired. This formulation of a switching lemma was first used by Cai [Cai86]. Several authors have subsequently noted that Håstad's argument also works in this fashion.

We modify Razborov's argument to prove results about decision trees, in part because it produces a more natural argument but also because in the case of the q-matching restrictions described in section 5 it is not clear how one could adapt Razborov's argument to the analogue of maxterms.

2 Decision tree version of the Håstad switching lemma

A *restriction* on a set of Boolean variables $\{x_i \mid i \in I\}$ is a map $\rho : I \to \{0, 1, *\}$. The result of its action on a Boolean function f is a Boolean function $f\lceil_\rho$ which is the result of substituting $\rho(i)$ for x_i for all places where $\rho(i) \neq *$. We say that all variables x_i such that $\rho(i) = *$ are *unset* and the resulting function becomes a function of the unset variables in the obvious way.

Define $\mathcal{R}_n^\ell$ to be the set of all restrictions ρ on a domain of n variables that have exactly ℓ unset variables. Håstad's switching lemma states that for any function f that is representable in disjunctive normal form (DNF) with short terms, then for almost all restrictions $\rho \in \mathcal{R}_n^\ell$, $f\lceil_\rho$ has a small height decision tree.

Fix some function f representable as a DNF formula F and assume that there is a total order on the terms of F as well as on the indices of the variables. A restriction ρ is applied to F in order, so that $F\lceil_\rho$ is the DNF formula whose terms consist of those terms of F that are not falsified by ρ, each shortened by removing any variables that are satisfied by ρ, and taken in the order of occurrence of the original terms on which they are based.

The *canonical decision tree for F*, $T(F)$ is defined inductively as follows:

1. If F is the constant function 0 or 1 (contains no terms or has an empty first term, respectively) then $T(F)$ consists of a single leaf node labelled by the appropriate constant value.
2. If the first term C_1 of F is not empty then let F' be the remainder of F so that $F = C_1 \vee F'$. Let K be the set of variables appearing in C_1. The tree $T(F)$ starts with a complete binary tree for K, which queries the variables in K in the order induced by the order on the indices. Each leaf v_σ in the tree is associated with a restriction σ which sets the variables of K according to the path from the root to v_σ. For each σ we replace the leaf node, v_σ, by the subtree $T(F\lceil_\sigma)$. (Note that for the unique σ which satisfies C_1 the leaf v_σ will remain a leaf and be labelled 1. For all other choices of σ, the tree that replaces v_σ is $T(F\lceil_\sigma) = T(F'\lceil_\sigma)$.)

We'll show that for any DNF formula F, for an appropriately chosen restriction ρ, the height of $T(F\lceil_\rho)$, $|T(F\lceil_\rho)|$, is small with high probability. This lemma is a switching lemma in the spirit of [Hås87] because it will allow us to obtain a DNF formula with short terms for $\neg F\lceil_\rho$ by taking the terms corresponding to the paths in $T(F\lceil_\rho)$ that have leaf labels 0. (We do not optimize the constants here. For improved constants see the discussion at the end of this section.)

Lemma 1: (Håstad Switching Lemma) Let F be a DNF formula in n variables with terms of length at most r. For $s \geq 0$, $\ell = pn$, and $p \leq 1/7$,

$$\frac{|\{\rho \in \mathcal{R}_n^\ell \ : \ |T(F\lceil_\rho)| \geq s\}|}{|\mathcal{R}_n^\ell|} < (7pr)^s.$$

The proof of this switching lemma is a small modification of Razborov's simplified proof of Håstad's switching lemma and uses a counting argument rather than complicated reasoning involving conditional probability. The property of the restriction family that is critical to the argument was clearly necessary in Håstad's argument but is implicit here: For any assignment of values to a set of variables and any s, it is exponentially more likely in s that a randomly chosen restriction agrees with the assignment than that it leaves s variables unset.

Before giving the proof of the switching lemma we give the following definition. Let $stars(r,s)$ to be the set of all sequences $\beta = (\beta_1, \ldots, \beta_k)$ such that for each j, $\beta_j \in \{*, -\}^r \setminus \{-\}^r$ and such that the total number of *'s in all the β_j is s. There is an easy bound of $|stars(r,s)| \leq 2^{s-1}r^s$ but we can also prove:

Lemma 2: $|stars(r,s)| < (r/\ln 2)^s$.

Proof: For convenience in the proof we shall include the empty string in $stars(r,0)$ which would otherwise be empty. We shall show by induction on s that $|stars(r,s)| \leq \gamma^s$ for $(1+1/\gamma)^r = 2$; the statement of the lemma follows by using $1 + x < e^x$ for $x \neq 0$.

The base case $s = 0$ follows trivially. Now suppose that $s > 0$. It is easy to see from the definition that for any $\beta \in stars(r,s)$, if β_1 has $i \leq s$ *'s then $\beta = (\beta_1, \beta')$ where $\beta' \in stars(r, s-i)$. (For $i = s$ we have used our augmentation of $stars(r,0)$.) There are $\binom{r}{i}$ choices of β_1 so

$$\begin{aligned}
|stars(r,s)| &= \sum_{i=1}^{\min(r,s)} \binom{r}{i} |stars(r,s-i)| \\
&\leq \sum_{i=1}^{r} \binom{r}{i} \gamma^{s-i} \\
&= \gamma^s \sum_{i=1}^{r} \binom{r}{i} (1/\gamma)^i \\
&= \gamma^s[(1+1/\gamma)^r - 1] \\
&= \gamma^s
\end{aligned}$$

by the inductive hypothesis and the definition of γ. □

Proof: (Håstad Switching Lemma) We only need to consider $s > 0$. Let $S \in \mathcal{R}_n^\ell$ be the set of restrictions ρ such that $|T(F\lceil_\rho)| \geq s$. As in Razborov's argument we obtain a bound on $|S|/|\mathcal{R}_n^\ell|$ by defining a 1-1 map from S to a small set. The proof is somewhat different because we are interested in the height of decision trees for $F\lceil_\rho$ rather than the length of maxterms of $F\lceil_\rho$.

We will define a 1-1 map

$$S \quad \rightarrow \quad \mathcal{R}_n^{\ell-s} \times stars(r,s) \times 2^s.$$

Let $F = C_1 \vee C_2 \vee \ldots$. Suppose that $\rho \in S$ and let π be the restriction associated with the lexicographically first path in $T(F\lceil_\rho)$ that has length $\geq s$ (any way of canonically associated such a long path will do.) Trim the last few variables set in π along the path from the root so that $|\pi| = s$. We use the formula F and π to determine the image of ρ. The image of ρ is defined by following the path π in the canonical decision tree for $F\lceil_\rho$ and using the structure of that tree (see Figure 1.)

Let C_{ν_1} be the first term of F that is not set to 0 by ρ. Then $C_{\nu_1}\lceil_\rho$ will be the first term in $F\lceil_\rho$. Since $|\pi| > 0$, such a term must exist and will not be the empty term. Let K be the set of variables in $C_{\nu_1}\lceil_\rho$ and let σ_1 be the unique restriction of the variables in K that satisfies $C_{\nu_1}\lceil_\rho$. Let π_1 be the portion of π that sets the variables in K. We have two cases based on whether or not $\pi_1 = \pi$.

1: If $\pi_1 \neq \pi$ then by the construction of π, π_1 sets all the variables in K. Note also that $C_{\nu_1}\lceil_{\rho\sigma_1} = 1$ but since $\pi_1 \neq \pi$, $\pi_1 \neq \sigma_1$, and thus $C_{\nu_1}\lceil_{\rho\pi_1} = 0$.

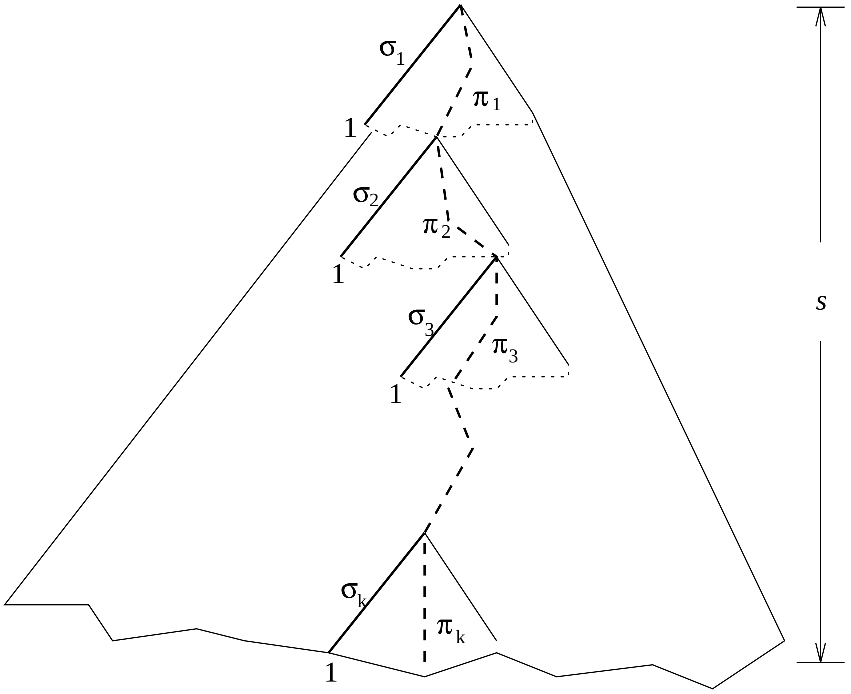

Figure 1: Canonical decision tree $T(F\restriction_\rho)$

2: If $\pi_1 = \pi$ then it is possible that π does not set all of the variables in K. In this case we shorten σ_1 to the variables in K that appear in π_1. Now all we know is that $C_{\nu_1}\lceil_{\rho\sigma_1} \neq 0$.

Define $\beta_1 \in \{*, -\}^k$ based on the fixed ordering of the variables in term C_{ν_1} by letting the j-th component of β_1 be $*$ if and only if the j-th variable in C_{ν_1} is set by σ_1. Note that since $C_{\nu_1}\lceil_\rho$ is not the empty term there is at least one $*$ in β_1. From C_{ν_1} and β_1 we can reconstruct σ_1.

Now, by the definition of $T(F\lceil_\rho)$, $\pi \setminus \pi_1$ labels a path in the canonical tree $T(F\lceil_{\rho\pi_1})$. If $\pi_1 \neq \pi$, we repeat the above argument, with $\pi \setminus \pi_1$ in place of π, $\rho\pi_1$ in place of ρ and find a term C_{ν_2} which is the first term of F not set to 0 by $\rho\pi_1$. Based on this we generate π_2, σ_2, and β_2 as before. We repeat this process until the round k in which $\pi_1\pi_2...\pi_k = \pi$.

Let $\sigma = \sigma_1\sigma_2...\sigma_k$. We finally define $\delta \in \{0,1\}^s$ to be a vector that indicates for each variable set by π (which are the same as those set by σ) whether it is set to the same value as σ sets it.

The image of ρ under the 1-1 map we define is a triple, $\langle \rho\sigma_1...\sigma_k, (\beta_1, ..., \beta_k), \delta\rangle$. Clearly $\rho\sigma = \rho\sigma_1...\sigma_k \in \mathcal{R}_n^{\ell-s}$ and $(\beta_1, ..., \beta_k) \in stars(r, s)$ so the map is as required.

It remains to show that the map we have just defined is indeed 1-1. To do this, as in Razborov's argument, we show how to recover ρ from its image. The reconstruction is iterative. In the general stage of the reconstruction we will have recovered $\pi_1, ..., \pi_{i-1}$, $\sigma_1, ..., \sigma_{i-1}$, and will have constructed $\rho\pi_1...\pi_{i-1}\sigma_i...\sigma_k$. Recall that for $i < k$, $C_{\nu_i}\lceil_{\rho\pi_1...\pi_{i-1}\sigma_i} = 1$ and $C_j\lceil_{\rho\pi_1...\pi_{i-1}\sigma_i} = 0$ for all $j < \nu_i$. This clearly also holds when we append $\sigma_{i+1}...\sigma_k$ to the restriction. When $i = k$, something similar occurs except the only guarantee is that $C_{\nu_i}\lceil_{\rho\pi_1...\pi_{k-1}\sigma_k} \neq 0$. Thus we can recover ν_i as the index of the first term of F that is not set to 0 by $\rho\pi_1...\pi_{i-1}\sigma_i...\sigma_k$.

Now, based on C_{ν_i} and β_i we can determine σ_i. Since we know $\sigma_1, ..., \sigma_i$, using the vector δ we can determine π_i. We can now change $\rho\pi_1...\pi_{i-1}\sigma_i...\sigma_k$ to $\rho\pi_1...\pi_{i-1}\pi_i\sigma_{i+1}...\sigma_k$ using the knowledge of π_i and σ_i. Finally, given all the values of the π_i we can reconstruct ρ.

Now we compute the value $|S|/|\mathcal{R}_n^\ell|$:

$|\mathcal{R}_n^\ell| = \binom{n}{\ell}2^{n-\ell}$ so

$$\frac{|\mathcal{R}_n^{\ell-s}|}{|\mathcal{R}_n^\ell|} = \frac{\ell^{(s)}}{(n-\ell+s)^{(s)}} \cdot 2^s \leq \frac{(2\ell)^s}{(n-\ell)^s}.$$

Applying the bounds we obtain

$$\frac{|S|}{|\mathcal{R}_n^\ell|} \leq \frac{|\mathcal{R}_n^{\ell-s}|}{|\mathcal{R}_n^\ell|} \cdot |stars(r, s)| \cdot 2^s$$

$$
\begin{aligned}
&\leq \left(\frac{4\ell r}{(n-\ell)\ln 2}\right)^s \\
&= (\frac{4pr}{(1-p)\ln 2})^s
\end{aligned}
$$

for $\ell = pn$. For $p < 1/7$ this is at most $(7pr)^s$. □

It is worth noting that in avoiding conditional probability we do not obtain bounds that are quite as strong as those obtained by Håstad. It is possible to obtain somewhat better bounds than described above by combining the information in $stars(r, s)$ and δ since, except for $i = k$, $\sigma_i \neq \pi_i$ and thus δ must contain at least one 1 in the seqment associated with each σ_i. In fact, by choosing without loss of generality a long branch π that does not have a leaf labelled 1, this is true even in the case that $i = k$. In that case we can replace Lemma 2 by a similar argument that produces a bound of α^s on the number of different encodings of both $stars(r, s)$ and δ where α is the solution of $(1 + 2/\alpha)^r - (1 + 1/\alpha)^r = 1$. This produces a final result very close to Håstad's bounds but it has a $1-p$ in the denominator as opposed to a $1+p$. The gap here seems to depend on the fact that we have fixed the number of stars as opposed to allowing it to vary. We chose to separate $stars(r, s)$ from δ in our argument above since $stars(r, s)$ is useful in the other switching lemma proofs.

3 The Clique Switching Lemma

The restrictions in $\mathcal{R}_n^\ell$ set the values of the variables independently of each other. They were appropriate for proving lower bounds on the parity function. In this section we consider restrictions that are tailored for proving lower bounds on the clique function for small values of the clique size. The switching lemma for these restrictions that we re-derive was originally proved by Beame [Bea90]. The technique here is more general than that of the previous section and was suggested to us by Johan Håstad although the general outline is the same. This switching lemma was originally proven when 0's and 1's are not equally likely but we first derive it for that case which is the appropriate value for the clique problem for $O(\log n)$ size cliques.

The variables for the restrictions are the graph edge variables $\{x_e \mid e \in [n]^2\}$. Let $\mathcal{C}_n^\ell$ be the set of all restrictions ρ chosen with the following property: There is a set $V \subseteq [n]$ of size ℓ such that every $e \in V^2$, $\rho(e) = *$ is unset and for every $e \in [n]^2 - [V]^2$, $\rho(e) \neq *$.

For these restrictions instead of simply the length of terms or the height of decision trees we consider the maximum number of different endpoints that appear in a term or along any path. We call this the *vertex length* of a term and denote the vertex height of a decision tree T by $|T|_v$.

Lemma 3: Let F be a DNF formula in the graph edge variables with terms with vertex length at most r. For $s \geq 0$, $\ell = pn$, and $p \leq 1/(r2^{r+s})$,

$$\frac{|\{\rho \in \mathcal{C}_n^\ell \ : \ |T(F\lceil_\rho)|_v \geq s\}|}{|\mathcal{C}_n^\ell|} < \frac{8}{3}(2^{r+s-1}pr)^s.$$

Proof: Let $S \in \mathcal{C}_n^\ell$ be the set of restrictions ρ such that $|T(F\lceil_\rho)|_v \geq s$. Let $F = C_1 \vee C_2 \vee \ldots$. Suppose that $\rho \in S$ and let π be the restriction associated with the lexicographically first path in $T(F\lceil_\rho)$ that has vertex length $\geq s$. Note that for each path in a canonical decision tree for a DNF formula there is a natural partition of the path into blocks which correspond to the variables contributed by a single term of the formula. Thus π is naturally split into blocks. For this switching lemma we trim π in a different way from Lemma 1 partly because we cannot necessarily find a prefix of π that has vertex length exactly s and because of other considerations that will become clearer later. Instead we trim π at the first block boundary such that the prefix of π at that block boundary has vertex length at least s. Let the vertex length of the trimmed π be s' and note that $s \leq s' \leq s + r - 1$.

As before we use the formula F and π to determine the image of ρ. However, rather than a 1-1 map we map each element of S to a *set* of elements of the range so that the image sets are disjoint.

Let e be the number of variables in π. Note that $s'/2 \leq e \leq \binom{s'}{2}$. We will let $S_{e,s'}$ be the set of $\rho \in S$ whose associated π has e variables and vertex length s' and count each set $S_{e,s'}$ separately. The map we define will be from $S_{e,s'}$ to subsets of $\mathcal{C}_n^{\ell-s'} \times stars(r,s') \times 2^e$.

The basic procedure for defining σ_i and π_i is exactly the same as in the proof of Lemma 1. Note that $\pi_1, ..., \pi_k$ are in fact the blocks of π. One difference is that because we trimmed π at a block boundary, even σ_k is such that $C_{\nu_k}\lceil_{\rho\pi_1\pi_2\ldots\pi_{k-1}\sigma_k} = 1$. Another is that, instead of the vector β coding which variables are starred by ρ in each relevant term, we code which *vertices* in these terms are in the set V of vertices unset by ρ. Since each original term has vertex length at most r and π has a total of s' vertices this may also be coded by a string β in $stars(r,s')$.

σ and π are restrictions on the same set of e variables and as in Lemma 1 the differences between σ and π are encoded by a bit string β of length e.

To continue the analogy with Lemma 1 we would like to map ρ to $\rho\sigma$ as in the argument above. However, unlike the situation there, $\rho\sigma$ is not a member of $\mathcal{C}_n^{\ell-s'}$. Since π touches exactly s' vertices unset by ρ, so does σ and thus there is still a clique of variables of size $\ell - s'$ that is unset by $\rho\sigma$. The idea that works is to *extend* $\rho\sigma$ to be a member of $\mathcal{C}_n^{\ell-s'}$ by setting extra variables. However, instead of doing a single extension we extend $\rho\sigma$ to the set of all possible extensions in $\mathcal{C}_n^{\ell-s'}$. More precisely, we map

$$\rho \in S_{e,s'} \to \{\langle\rho\sigma\tau, \beta, \delta\rangle \mid \rho\sigma\tau \in \mathcal{C}_n^{\ell-s'}\}.$$

In order to do the accounting we need distinct ρ to have disjoint image sets. We accomplish this by showing that we can decode any of the triples $\langle \rho\sigma\tau, \beta, \delta\rangle$ to obtain ρ. The basic structure of this argument is essentially the same as in the switching lemma above, except that we must also be able to remove the extension τ. It is important to note that for each i, because $C_{\nu_i}\upharpoonright_{\rho\pi_1\ldots\pi_{i-1}\sigma_i} = 1$, it also the case that C_{ν_i} will be the first term such that $C_{\nu_i}\upharpoonright_{\rho\pi_1\ldots\pi_{i-1}\sigma_i\ldots\sigma_k\tau} = 1$. This guarantees that knowing $\pi_1, \ldots, \pi_{i-1}, \sigma_1, \ldots, \sigma_{i-1}$ we can determine C_{ν_i} and, using β, we can then determine σ_i.

Once all of σ has been determined, we can immediately recover τ since every variable set by τ either has both endpoints touching the vertices touched by σ or has one endpoint touching σ and the other touching the set of unset vertices. Finally, since we have determined σ and τ, we can recover ρ from $\rho\sigma\tau$. Thus the image sets are disjoint.

It remains to determine the sizes of the relevant sets involved in the maps.

Easy calculation gives $|\mathcal{C}_n^\ell| = \binom{n}{\ell}2^{\binom{n}{2}-\binom{\ell}{2}}$. Thus

$$\begin{aligned}
\frac{|\mathcal{C}_n^{\ell-s'}|}{|\mathcal{C}_n^\ell|}\times |stars(r,s')| \times 2^e &\leq \frac{\binom{n}{\ell-s'}}{\binom{n}{\ell}}2^{\binom{\ell}{2}-\binom{\ell-s'}{2}}\cdot (r/\ln 2)^{s'}\cdot 2^e \\
&\leq (\ell/(n-\ell))^{s'}2^{\binom{\ell}{2}-\binom{\ell-s'}{2}}\cdot (r/\ln 2)^{s'}\cdot 2^e \\
&= 2^{\binom{\ell}{2}-\binom{\ell-s'}{2}}\cdot\left(\frac{r\ell}{(n-\ell)\ln 2}\right)^{s'}\cdot 2^e
\end{aligned}$$

If $\rho \in S_{e,s'}$, then there $\binom{\ell}{2} - \binom{\ell-s'}{2} - e$ variables that must be set by τ. Since there are two values for each variable there are $d_{e,s'} = 2^{\binom{\ell}{2}-\binom{\ell-s'}{2}-e}$ choices of τ. Thus

$$\begin{aligned}
\frac{|S_{e,s'}|}{|\mathcal{C}_n^\ell|} &\leq \frac{|\mathcal{C}_n^{\ell-s'}|}{d_e|\mathcal{C}_n^\ell|}\cdot |stars(r,s')|\cdot 2^e \\
&= \left(\frac{r\ell}{(n-\ell)\ln 2}\right)^{s'}\cdot 2^{2e}
\end{aligned}$$

It is clear that for each s', the bounds for the terms $|S_{e,s'}|/|\mathcal{C}_n^\ell|$ form a geometric sequence with ratio 4 and largest term corresponding to $e = \binom{s'}{2} = s'(s'-1)/2$. Thus

$$\sum_e \frac{|S_{e,s'}|}{|\mathcal{C}_n^\ell|} \leq \frac{4}{3}\left(\frac{r\ell}{(n-\ell)\ln 2}\right)^{s'}\cdot 2^{s'(s'-1)} = \frac{4}{3}\left(\frac{2^{s'-1}r\ell}{(n-\ell)\ln 2r}\right)^{s'}.$$

Now if we set $\ell = pn$ this becomes $\frac{4}{3}[2^{s'-1}pr/((1-p)\ln 2)]^{s'} \leq \frac{4}{3}(2^{s'}pr)^{s'}$. Note that $s' \leq r+s-1$ so this is at most $\frac{4}{3}(2^{s+r-1}pr)^{s'}$. Thus when we sum this over all choices

s' we obtain another geometric sequence with ratio $2^{s+r-1}pr$ and given $p \leq 1/(r2^{s+r})$ the ratio is $1/2$ and the largest term has $s' = s$. Thus

$$\frac{|S|}{|\mathcal{C}_n^\ell|} = \sum_{e,s'} \frac{|S_{e,s'}|}{|\mathcal{C}_n^\ell|} \leq \frac{8}{3}(2^{s+r-1}pr)^s.$$

□

4 Imbalanced 1's and 0's

In the arguments of the two previous sections, when variables were set by the restriction they were equally likely to be set to 0 or 1. In this section we consider how they might be set with imbalanced probability. This is necessary to get bounds for the clique problem. However, the basic idea is simpler in the case of fully independent restrictions so we show how it would go in that case.

The essential idea here is to give weights to the restrictions which reflect the probability of the restriction being chosen as a random member of the set. Suppose that in the context of Håstad's switching lemma we wanted to argue that there is a restriction which is strongly biased towards setting bits to 1 and keeps the decision tree height small. The basic switching Lemma does not give us this information because as n gets large it is much more unlikely to get a restriction with even a constant factor bias than the probability of failure.

Thus consider the distribution where we choose a random member of $\mathcal{R}_n^\ell$ so that the probability of a restriction ρ is proportional to $q^a(1-q)^b$ where a is the number of values ρ sets to 1 and b is the number of values that ρ sets to 0. Clearly, the original uniform choice of ρ corresponds to $q = 1/2$.

The map corresponding to Lemma 1 is exactly the same as before. The only difference is that we associate a *weight* with each restriction ρ. If ρ assigns a 1's and b 0's then let the weight of ρ be $q^a(1-q)^b$. Thus the total weight of $\mathcal{R}_n^\ell$ is $\binom{n}{\ell}$. Instead of simply counting the set S, we count the total weight of the members of S as a proportion of the total weight of $\mathcal{R}_n^\ell$.

To do this we compute the total weight of the image set $\mathcal{R}_n^{\ell-s} \times stars(r,s) \times 2^s$ where the weight of a triple is the weight of the restriction it contains. To convert this to a bound on the total weight of S we must divide this by the total weight of $\mathcal{R}_n^\ell$ times the factor by which the map decreases the weight of a restriction.

The total weight of the image set is $\binom{n}{\ell-s} \cdot |stars(r,s)| \cdot 2^s$. The decrease in the weight of restriction ρ under the map is simply the weight of σ. Since $|\sigma| = s$, the total decrease is at worst $\min\{q, 1-q\}^s$. Thus the total weight of S as a fraction of the weight of $\mathcal{R}_n^\ell$ is at most

$$\begin{aligned}
&\frac{\binom{n}{\ell-s}}{\binom{n}{\ell}\cdot\min\{q,1-q\}^s}\cdot(r/\ln 2)^s\cdot 2^s\\
&\quad\le\quad\left(\frac{2r\ell}{(n-\ell)\min\{q,1-q\}\ln 2}\right)^s\\
&\quad=\quad\left(\frac{2pr}{(1-p)\min\{q,1-q\}\ln 2}\right)^s.
\end{aligned}$$

where $\ell = pn$. For $q = 1/2$ this gives exactly the same bounds as before. If we assume that $q \le 1/2$ and if $p \le 1/7$ then this bound is less than $(4pr/q)^s$.

Now for $q(1-q)n$ growing with n, it is almost certain that the restriction will assign a fraction of 1's roughly proportional to q and so one can obtain imbalanced restrictions that cause the decision tree height to be small.

An alternative argument for obtaining imbalanced restrictions that keep the decision tree height small is to fix the exact number of 1's that a restriction assigns at a q fraction of the variables set; call the set of such restrictions $\mathcal{R}_n^{\ell,q}$. Rather than a weighting argument, one again obtains a counting argument in this case, i.e. one argues that the set $S \subseteq \mathcal{R}_n^{\ell,q}$ of bad restrictions is a small proportion of the total number of restrictions. For example, instead of replacing a 2^s by $(1/\min\{q,1-q\})^s$ in the calculations, one ends up replacing 2^s by $\sum_{0\le i\le s}\binom{n-\ell+s}{q(n-\ell)+i}/\binom{n-\ell}{q(n-\ell)}$. For $q(n-\ell)+s \le \frac{1}{2}(n-\ell+s)$ the largest term has $i = s$ and in this case, instead of the directly calculated $(1/q)^s$ factor, one notes that

$$\binom{n-\ell+s}{q(n-\ell)+s}/\binom{n-\ell}{q(n-\ell)} = (n-\ell+s)^{(s)}/(q(n-\ell)+s)^{(s)} \le [(n-\ell)/(q(n-\ell))]^s = (1/q)^s.$$

Finally, consider the restrictions in $\mathcal{C}_n^\ell$ with an imbalanced probability of 0's and 1's; call the distribution $\mathcal{C}_n^{\ell,q}$. We'll compute the new bounds here with a weighting argument as opposed to fixing the proportion of 0's and 1's. This time however, we need to compute the net decrease that the map produces in the *total* weight of the set of restrictions that ρ maps to. The decrease in weight for any individual restriction $\rho\sigma\tau$ in the image is due to the weight of $\sigma\tau$. However, since the set of restrictions includes all possible restrictions on the variables of τ, the decrease in the total weight under the map is simply $\min\{q,1-q\}^{|\sigma|} = \min\{q,1-q\}^e$. Thus instead of 2^{2e} in the numerator in the calculations one obtains $(2/\min\{q,1-q\})^e$ and thus if $q \le 1/2$, the final bound is less than $\frac{8}{3}((2/q)^{(s+r-1)/2}pr)^s$ provided that $p \le 1/(r(2/q)^{(r+s)/2})$

The bounds one obtains by choosing restrictions from $\mathcal{R}_n^\ell$ with a bias towards 0 or 1 can usually be proven just as easily by a reduction argument. However in the case of $\mathcal{C}_n^{\ell,q}$ these biased restrictions are critical for obtaining lower bounds for the $Clique_n^k$ problem for $k \le \log n$.

5 Switching Lemma for q-Matching Restrictions

We now apply these new techniques to give a switching lemma for a family of restrictions where the variables are q-dimensional hyperedges and the restrictions are partial matchings. These restrictions for $q = 2$ (and similar ones) were used to provide exponential lower bounds for the lengths of proofs of tautologies related to the propositional pigeonhole principle in bounded-depth Frege proof systems ([Ajt88, Ajt90, BIK$^+$92, KPW91, PBI93, BP93]).

There is a significant difference from the switching lemmas that we have previously considered since the decision trees involved do not exactly compute the original function but rather *represent* that function in a certain natural sense. Because of the difference in the sets of variables and the kind of decision trees, we write out the entire switching lemma argument in full rather than refer to the previous lemmas.

The original arguments to show this switching lemma in the case $q = 2$ were quite complicated because the conditioning on an arbitrary function being forced to 0 was problematic. In fact, without certain bounds on the probability of $*$, it was no longer the case with these restrictions that the conditioning could only help. Here, however, the argument is fairly simple overall and is only about 20% of the length of the original proof.

We now give the definitions for partial q-dimensional matching restrictions. The q-variables over D will be the set $\{x_e : \ e \subseteq D, \ |e| = q\}$. If $e = \{i_1, \ldots, i_q\}$ then $i_1, \ldots, i_q$ will be called the *endpoints* of x_e. A *q-term over D* is defined to be a conjunction of the form $\bigwedge \Gamma$, where Γ is a set of variables over D such that distinct variables in Γ have distinct endpoints. The q-terms are in 1-1 correspondence with the partial q-matchings on the set D. Given a partial q-matching, π, let $term(\pi)$ denote the q-term associated with π. The emit size of a q-term $\bigwedge \Gamma$ is $|\Gamma|$. An OR of q-terms where each q-term is of size at most t will be called a *t-disjunction.* When q is understood from the context we simply call a q-term a term. A *truth assignment* φ over D is any total assignment of $\{0, 1\}$ to the variables over D. Let $D' \subseteq D$. A truth assignment φ over D is a *q-matching on D'* if for each $i \in D'$ there is a unique $e \subseteq D$ such that $\varphi(P_e) = 1$.

If Y is a q-term or a set of variables, then $v(Y)$ denotes the set of endpoints of variables in Y.

For $|D| = qn + 1$, define $\mathcal{M}^{\ell}_{D,q}$ to be the set of all partial q-matchings on D, ρ, which match all but $q\ell + 1$ nodes of D.

Every ρ in $\mathcal{M}^{\ell}_{D,q}$ determines a unique *restriction*, r, of the variables over D as follows.

$$r(e) = \begin{cases} 1 & \text{if } e \in \rho \\ 0 & \text{if there is an } e' \in \rho \text{ such that } e' \neq e \\ & \quad \text{and } e \cap e' \neq \emptyset \\ * & \text{otherwise} \end{cases}$$

If r is the restriction obtained from ρ, we will refer to both the restriction and the partial

q-matching by ρ. For a Boolean formula F in the q-variables and a partial q-matching ρ, F restricted by ρ will be denoted by $F\lceil_\rho$.

We say that two partial q-matchings σ and τ are *compatible* if $\sigma \cup \tau$ is also a partial q-matching. When viewed as restrictions, we use the notation $\sigma\tau$ to denote the restriction defined by the partial q-matching $\sigma \cup \tau$.

A *q-matching decision tree* over domain D is defined as follows. It is a rooted tree where each interior node v is labelled by a query $i \in D$ and each edge is labelled by some q-edge $e \subseteq D$, $|e| = q$ with $i \in e$. Leaves are labelled with either "0" or "1". For each interior node v labelled by $i \in D$, there is exactly one out-edge labelled e for each $e \subseteq D$, $|e| = q$ such that $i \in e$ and no endpoint of e is contained in the label of any edge on the path from the root to v. The label of an interior node v may not appear in any edge label on the path from the root to v. Thus the set of edge labels on any path defines a partial q-matching.

A matching decision tree T over D *represents* a function f over domain D provided that for all leaf nodes $v \in T$, if we let σ be the partial q-matching defined by the path in T from the root to v then for all truth assignments α over D that are q-matchings on $v(\sigma)$ and satisfy σ, $f(\alpha)$ is equal to the label of v.

If F is a DNF formula that we say that T *refines and represents* F over domain D if T represents F as a function over D and furthermore for every path in T with leaf label 1 there is a term in F which is forced to 1 by the partial matching σ in T that reaches that leaf.

For any q-matching decision tree T let $disj(T)$ be the disjunction which has as its terms $term(\pi)$ for each partial q-matching π defined by some path p in T that ends in a leaf labelled 1. Note that if T has height t, then $disj(T)$ is a t-disjunction.

Note that if T represents f over D then the tree T^c obtained by switching the 1's and 0's labelling the leaves of T represents $\neg f$.

We assume that there is a total order on the elements of D. Let $K \subseteq D$. Then $Proj_D[K]$ is the set of all minimal partial q-matchings over D which involve all of the elements of K.

We define the *complete q-matching tree* for $K \subseteq D$ over D inductively as follows. If K consists of a single node $k \in D$, then label the root "k", and create $\binom{|D|-1}{q-1}$ edges adjacent to the root, labelled by those $e \subseteq D$ such that $k \in e$ respectively. Otherwise, suppose that k is the largest element of K and let $K' = K \setminus \{k\}$. Assume that we have created the complete q-matching tree for K'; we will now extend it to a complete tree for K. This is done by extending each leaf node v_ℓ as follows. Let p_ℓ be the path from the root to v_ℓ. The edge labellings along p_ℓ define a partial q-matching involving all elements of K'. If this partial q-matching does not include k, then label v_ℓ by k, and add new edges leading out of v_ℓ, one for every possible q-edge containing k that results in a partial q-matching extending the partial q-matching along p_ℓ. Otherwise, if k is involved in the partial matching, leave

v_ℓ unlabelled. Note that each path of the complete tree over K will be labelled by some $\sigma \in \mathit{Proj}_D[K]$.

Given a disjunction F over D, assume that F has a total order on its terms and an order on the variables within each term. A restriction ρ is applied to F in order, so that $F\upharpoonright_\rho$ is the q-term disjunction whose terms consist of those q-terms of F that are not falsified by ρ, each shortened by removing any variables that are satisfied by ρ, and taken in the order of occurrence of the original terms on which they are based.

The *canonical decision tree for F over D*, $T_D(F)$ is defined inductively as follows:

1. If F is the constant function 0 or 1 (contains no terms or has an empty first term, respectively) then $T_D(F)$ consists of a single leaf node labelled by the appropriate constant value.

2. If the first term C_1 of F is not empty then let F' be the remainder of F so that $F = C_1 \vee F'$. Let $K = v(C_1)$. We start with the complete q-matching tree for K. The paths of this tree correspond exactly to elements of $\mathit{Proj}_D[K]$. Let v_σ be the leaf node corresponding to the path labelled by $\sigma \in \mathit{Proj}_D[K]$. To obtain $T_D(F)$, for each σ we replace the leaf node, v_σ, by the subtree $T_{D\upharpoonright_\sigma}(F\upharpoonright_\sigma)$. (Note that for the unique element $\sigma \in \mathit{Proj}_D(K)$ which satisfies C_1 the leaf label of v_σ will be 1. For all other choices of σ, $T_{D\upharpoonright_\sigma}(F\upharpoonright_\sigma) = T_{D\upharpoonright_\sigma}(F'\upharpoonright_\sigma)$.)

$T_D(F)$ clearly refines and represents F over D. We'll show that for appropriately chosen restriction ρ the height of $T_D(F\upharpoonright_\rho)$, $|T_D(F\upharpoonright_\rho)|$, is small with high probability. This lemma is a switching lemma in the spirit of [Hås87] because it will allow us to obtain a disjunction that approximates the negation of F by representing F by a matching decision tree T and then taking $\mathit{disj}(T^c)$.

Lemma 4: Let D be a set with $qn+1$ elements. Let F be an r-disjunction over D. If $s \geq 0$ and $6 \leq \ell = pn \leq (n/r)^{1/q^2}/e$ then

$$\frac{|\{\rho \in \mathcal{M}^\ell_{D,q} \ : \ |T_{D\upharpoonright_\rho}(F\upharpoonright_\rho)| \geq s\}|}{|\mathcal{M}^\ell_{D,q}|} < (4e^q r^{1/q} p^q n^{q-1/q})^s.$$

Proof: We only need to consider $s > 0$. Let $S \in \mathcal{M}^\ell_{D,q}$ be the set of restrictions ρ such that $|T_{D\upharpoonright_\rho}(F\upharpoonright_\rho)| \geq s$. As in Lemma 1 we obtain a bound on $|S|/|\mathcal{M}^\ell_{D,q}|$ by defining a 1-1 map from S to a small set.

We will define a 1-1 map

$$S \ \ \rightarrow \bigcup_{s/q \leq j \leq s} \mathcal{M}^{\ell-j}_{D,q} \times \mathit{stars}(r,j) \times \Delta_j$$

where Δ_j is a set of size $\binom{q\ell+1}{q-1}^s$.

Let $F = C_1 \vee C_2 \vee \ldots$. Suppose that $\rho \in S$ and let π be the partial q-matching labelling the lexicographically first path in $T_{D\restriction_\rho}(F\restriction_\rho)$ that has length $\geq s$. Trim the last few q-edges of π along the path from the root so that $|\pi| = s$. We use the formula F and π to determine the image of ρ. Let C_{ν_1} be the first term of F that is not set to 0 by ρ. Then $C_{\nu_1}\restriction_\rho$ will be the first term in $F\restriction_\rho$. Since $|\pi| > 0$, such a term must exist and is not the empty term. Let $K = v(C_{\nu_1}\restriction_\rho)$ and let σ_1 be the unique partial q-matching in $Proj_{D\restriction_\rho}[K]$ that satisfies $C_{\nu_1}\restriction_\rho$. Let π_1 be the portion of π that touches K. We have two cases based on whether or not $\pi_1 = \pi$.

1: If $\pi_1 \neq \pi$ then by the construction of π, $\pi_1 \in Proj_{D\restriction_\rho}[K]$. Note also that $C_{\nu_1}\restriction_{\rho\sigma_1}= 1$ but since $\pi_1 \neq \pi$, $\pi_1 \neq \sigma_1$, and thus $C_{\nu_1}\restriction_{\rho\pi_1}= 0$.

2: If $\pi_1 = \pi$ then it is possible that $v(\pi)$ does not contain all of K. In this case we shorten σ_1 so that it is the unique element of $Proj_{D\restriction_\rho}[K']$ that does not falsify $C_{\nu_1}\restriction_\rho$ where $K' = v(\pi_1) \cap K$.

Note that in either case $|\pi_1| \leq q|\sigma_1|$.

Define β_1 to be the vector of length r based on the fixed ordering of the variables in term f_{ν_1} by letting the j-th component of β_1 be $*$ if and only if The j-th variable in C_{ν_1} is in $v(\sigma_1)$. Note that since $C_{\nu_1}\restriction_\rho$ is not the empty term then there is at least one $*$ in β_1. From C_{ν_1} and β_1 we can reconstruct σ_1.

Now, by the definition of $T_{D\restriction_\rho}(F\restriction_\rho)$, $\pi \setminus \pi_1$ labels a path in the canonical tree $T_{D\restriction_{\rho\pi_1}}(F\restriction_{\rho\pi_1})$. If $\pi_1 \neq \pi$, we repeat the above argument, with $\pi \setminus \pi_1$ in place of π, $\rho\pi_1$ in place of ρ and find a term C_{ν_2} which is the first term of F not set to 0 by $\rho\pi_1$. Based on this we generate π_2, σ_2, β_2, as before. We repeat this process until the round k in which $\pi_1\pi_2 ... \pi_k = \pi$.

For each i, π_i matches all elements of $v(\sigma_i)$, so the $\sigma_1, \ldots, \sigma_k$ are mutually compatible and thus $\sigma_1 ... \sigma_k = \sigma_1 \cup \cdots \cup \sigma_k$. The image of ρ under the 1-1 map we define is a triple, $\langle \rho\sigma_1 ... \sigma_k, (\beta_1, ..., \beta_k), \delta \rangle$ where δ is defined below. Let $\sigma = \sigma_1 ... \sigma_k$ and $j = |\sigma|$. Clearly $\rho\sigma = \rho\sigma_1 ... \sigma_k \in \mathcal{M}_{D,q}^{\ell - j}$ and $(\beta_1, ..., \beta_k) \in stars(r, j)$.

We now define the information $\delta \in \Delta_j$. This will encode the relationships between all the σ_i and π_i. Since the elements of $v(\pi)$ are all nodes unset by ρ, every element of $v(\pi)$ is either in $v(\sigma)$ or among the $q(\ell - j) + 1$ nodes unset by $\rho\sigma$. To encode π we consider a fixed numbering of these $q\ell + 1$ nodes: the qj nodes in $v(\sigma)$ are numbered $1, ..., qj$ in the order $v(\sigma_1) < v(\sigma_2) < ... < v(\sigma_k)$ and the nodes unset by $\rho\sigma$ are numbered $qj + 1, ..., q\ell + 1$. For each i, to specify π_i we list the q-edges in π_i in order of their smallest numbered elements. The first such edge will contain the smallest numbered node in $v(\sigma_i)$, the next will contain that smallest numbered node in $v(\sigma)$ not touched by the first edge and so on. Thus to specify π it is only necesary to the numbers of the $q - 1$ other elements in the q-edge. For each such edges there are $\binom{q\ell+1}{q-1}$ choices of these elements. Note that by construction, for each i, $v(\pi_i) \cap v(\sigma) = v(\sigma_i)$. δ is the vector of these specifications, one per q-edge of π. Thus the image of the map is as required.

It remains to show that the map we have just defined is indeed 1-1. To do this, as in Lemma 1, we show how to recover ρ from its image. The reconstruction is iterative. In the general stage of the reconstruction we will have recovered $\pi_1, ..., \pi_{i-1}$, $\sigma_1, ..., \sigma_{i-1}$, and will have constructed $\rho\pi_1...\pi_{i-1}\sigma_i...\sigma_k$. Recall that for $i < k$, $C_{\nu_i}\upharpoonright_{\rho\pi_1...\pi_{i-1}\sigma_i} = 1$ and $C_j\upharpoonright_{\rho\pi_1...\pi_{i-1}\sigma_i} = 0$ for all $j < \nu_i$. This clearly also holds when we append $\sigma_{i+1}...\sigma_k$ to the restriction. When $i = k$, something similar occurs except the only guarantee is that $C_{nu_i}\upharpoonright_{\rho\pi_1...\pi_{k-1}\sigma_k} \neq 0$. Thus we can recover ν_i as the index of the first term of F that is not set to 0 by $\rho\pi_1...\pi_{i-1}\sigma_i...\sigma_k$.

Now, based on C_{ν_i} and β_i we can determine σ_i. Since we know $\sigma_1, ..., \sigma_i$ we can examine the entries in the vector δ associated with each of the vertices in $v(\sigma_i)$. At this point, although $\sigma_{i+1}, ..., \sigma_k$ are still undetermined, π_i can still be determined since π_i does not touch any of the vertices these restrictions touch.

We can now change $\rho\pi_1...\pi_{i-1}\sigma_i...\sigma_k$ to $\rho\pi_1...\pi_{i-1}\pi_i\sigma_{i+1}...\sigma_k$ using the knowledge of π_i and σ_i. Finally, given all the values of the π_i we can reconstruct ρ.

Now we compute the value $|S|/|\mathcal{M}^{\ell}_{D,q}|$. We can describe an element of $\mathcal{M}^{\ell}_{D,q}$ by choosing $q(n-\ell)$ elements from D in order and then ignoring the order within each of the $n-\ell$ groups of q elements and between these $n-\ell$ groups. Thus $|\mathcal{M}^{\ell}_{D,q}| = (qn+1)!/(q\ell+1)!(q!)^{n-\ell}(n-\ell)!$. We now compute

$$\begin{aligned}
\frac{|\mathcal{M}^{\ell-j}_{D,q}|}{|\mathcal{M}^{\ell}_{D,q}|} &= \frac{(q\ell+1)^{(qj)}}{(q!)^j(n-\ell+j)^{(j)}} \\
&\leq \frac{(q\ell+1)^{qj}}{(q!)^j(n-\ell)^j}.
\end{aligned}$$

Now applying our bounds we obtain

$$\begin{aligned}
\frac{|S|}{|\mathcal{M}^{\ell}_{D,q}|} &\leq \sum_{s/q \leq j} \frac{|\mathcal{M}^{\ell-j}_{D,q}|}{|\mathcal{M}^{\ell}_{D,q}|} \cdot |stars(r,j)| \cdot |\Delta_j| \\
&\leq \sum_{s/q \leq j} \frac{(q\ell+1)^{qj}}{(q!)^j(n-\ell)^j}(r/\ln 2)^j \binom{q\ell+1}{q-1}^s \\
&= \binom{q\ell+1}{q-1}^s \sum_{s/q \leq j} \left(\frac{(q\ell+1)^q r}{q!(n-\ell)\ln 2}\right)^j \\
&\leq \left(\frac{(q\ell+1)^{q-1} q}{q!}\right)^s \sum_{s/q \leq j} \left(\frac{(q\ell+1)^q r}{q!(n-\ell)\ln 2}\right)^j
\end{aligned}$$

$$\leq \quad \left(\frac{e^q(q\ell+1)^{q-1}q}{q^q}\right)^s \sum_{s/q\leq j}\left(\frac{e^q(q\ell+1)^q r}{q^q(n-\ell)\ln 2}\right)^j$$

since $q! \geq (q/e)^q$. Thus

$$\begin{aligned}
\frac{|S|}{|\mathcal{M}^{\ell}_{D,q}|} &\leq [e^q\ell^{q-1}(1+1/q\ell)^{q-1}]^s \sum_{s/q\leq j}\left(\frac{(e\ell)^q(1+1/q\ell)^q r}{(n-\ell)\ln 2}\right)^j \\
&\leq [e^{1+(q-1)/q\ell}(e\ell)^{q-1}]^s \sum_{s/q\leq j}\left(\frac{e^{1/\ell}(e\ell)^q r}{(n-\ell)\ln 2}\right)^j \\
&< [(e^{1/6})^{(q-1)/q}e(e\ell)^{q-1}]^s \sum_{s/q\leq j}\left(\frac{e^{1/6}(e\ell)^q r}{(n-\ell)\ln 2}\right)^j . \qquad (1)
\end{aligned}$$

since $\ell \geq 6$. Because $6 \leq \ell \leq (n/r)^{1/q^2}/e$ it must be the case that $r \leq n/(6e)^{q^2}$ and thus it follows that

$$\frac{e^{1/6}(e\ell)^q r}{(n-\ell)\ln 2} \leq 2\frac{(e\ell)^q r}{n} \leq 2(r/n)^{1-1/q} \leq 2/(6e)^{q(q-1)} \leq 1/(12e^2) < 1/88.$$

and thus the sum in line 1 is bounded by a geometric series with ratio $< 1/88$. From this we derive that it is less than 1.02 times its largest term which occurs when $j = s/q$ so

$$\begin{aligned}
\frac{|S|}{|\mathcal{M}^{\ell}_{D,q}|} &\leq 1.02\left(\frac{e^{1+1/6}(e\ell)^q r}{(n-\ell)^{1/q}(\ln 2)^{1/q}}\right)^s \\
&\leq \left(\frac{1.02e^{1+1/6}e^q p^q n^{q-1/q} r^{1/q}}{(1-p)^{1/q}(\ln 2)^{1/q}}\right)^s
\end{aligned}$$

for $p = \ell/n$. Now since $q \geq 2$ and $\ell \leq (n/r)^{1/q^2}/e \leq n^{1/4}/e$ we have

$$\frac{1.02e^{1+1/6}}{(1-p)^{1/q}(\ln 2)^{1/q}} \leq 4.$$

Thus we can bound $|S|/|\mathcal{M}^{\ell}_{D,q}|$ above by $(4e^q p^q r^{1/q} n^{q-1/q})^s$. □

6 Applying the Switching Lemmas

For completeness we include an overview of how these switching lemmas are applied to produce lower bounds for bounded-depth circuits. To do this we first define unbounded-depth circuits.

An *unbounded fan-in circuit* with basis $\mathcal{B}$ on a set of n Boolean input variables is given by a directed acyclic graph with each source node labelled by an input from $\{x_i \mid i \in [n]\}$ with a single sink node and with each non-source node labelled an element from the set $\mathcal{B}$ of allowable Boolean operations. We call the labelled node a *gate*. Each gate of the circuit computes a Boolean function in the obvious way and the function computed by the circuit is the function computed by the output gate. For this section we will identify the gates and the functions computed at them.

It is standard to allow labels on the inputs to a circuit to be both inputs and their negations and let $\mathcal{B} = \{\vee, \wedge\}$. The size of such a circuit is the number of its gates and the depth of a circuit is the length of the longest path from any input to the output in this definition. However, we will find it more convenient to give an equivalent definition where $\mathcal{B} = \{\vee, \neg\}$ where $\neg$ only labels nodes of in-degree 1. For these circuits we take the equivalent notion of size as the number of $\vee$-gates. We define the depth of a gate to be the maximum number of $\vee$-gates on a path from any input to the gate and the depth of a circuit to the depth of its output node.

6.1 Parity

We first apply Lemma 1 to obtain a slightly weaker form of the parity lower bound in [Hås87].

Lemma 5: Let $n_{i+1} = n/14(14 \log_2 S)^{-i}$ for $0 \leq i \leq d-1$. If C is a circuit of size S and depth d then for every i, $1 \leq i \leq d$, if $n_i \geq \log_2 S$ there is a restriction $\rho_i \in \mathcal{R}_n^{n_i}$ such that for every gate g of C of depth at most i, $g\lceil_{\rho_i}$ is computed by a decision tree of height $\leq \log_2 S$.

Proof: We first note that if we can prove the result for the $\vee$-gates then it will follow immediately for the $\neg$-gates since a decision tree for $\neg g$ is exactly the same as that for g except that the leaf labels 1 and 0 are reversed. We now prove the result for the $\vee$-gates by induction:

Base Case: The inputs to an $\vee$-gate at depth 1 are merely inputs or negations of inputs which we can think of as DNF terms of size 1. Let $p = 1/14$ and note that $n_1 = np$. By Lemma 1, for each $\vee$-gate at depth 1, less than a $(7p)^{\log_2 S} = 1/S$ fraction of all restrictions ρ_1 in $\mathcal{R}_n^{n_1}$ fail to keep the decision tree height of the $\vee$-gate at most $\log_2 S$. Since there are at most S $\vee$-gates at depth 1 there is at least one restriction ρ that keeps the height of all the decision trees at most $\log_2 S$. Since this is already true for gates of depth 0 the base case holds.

Induction Step: Suppose that there is a restriction $\rho_i \in \mathcal{R}_n^{n_i}$ so that for all gates g of depth at most i, $g\lceil_{\rho_i}$ has decision tree height at most $\log_2 S$. Therefore these functions can expressed by DNF formulas with term length at most $\log_2 S$. Consider any $\vee$-gate

g at depth $i+1$. All the inputs to this gate have depth at most i so after ρ_i is applied they can be expressed as DNF formulas with term length at most $\log_2 S$. Thus after ρ_i is applied $g\lceil_{\rho_i}$ itself may be expressed as a DNF formula with term length at most $\log_2 S$. We now let $p = 1/(14 \log_2 S) = n_{i+1}/n_i$. By Lemma 1 less than a $(7p \log_2 S)^{\log_2 S} = 1/S$ fraction of all restrictions $\pi \in \mathcal{R}_{[n_i]}^{n_{i+1}}$ fail to keep the decision tree height of $(g\lceil_{\rho_i})\lceil_{\pi} = g\lceil_{\rho_i \pi}$ bounded by $\log_2 S$. As in the base case, since there at most S $\vee$-gates of depth $i+1$, there is some fixed restriction π such that for all gates at depth $i+1$, applying $\rho_i \pi$ leaves their decision tree height at most $\log_2 S$. Since $\rho_i \pi \in \mathcal{R}_n^{n_{i+1}}$, we let $\rho_{i+1} = \rho_i \pi$ and, observing that all gates of depth less than $i+1$ maintain their small decision tree height, we see that ρ_{i+1} satisfies the conditions of the lemma. □

Theorem 6: Any unbounded fan-in circuit of depth d computing the parity function requires size at least $2^{\frac{1}{14} n^{1/(d-1)}}$.

Proof: Let S be the size of such a circuit C. The most straightforward way to apply Lemma 5 to the parity function would be to apply the restriction ρ_d to the output of the circuit computing the parity function and thus argue that $n_d \leq \log_2 S$ since a restricted parity function must have decision tree height equal to the number of variables unset. However, the lower bound in that case is slightly inferior to our claimed bound.

Instead, fix the restriction $\rho_{d-1} \in \mathcal{R}_n^{n_{d-1}}$ for C. Then, as in the proof of Lemma 5, for the unique $\vee$-gate of depth d, $g\lceil_{\rho_{d-1}}$ is expressible as a DNF formula with term length at most $\log_2 S$. Clearly g is either parity or its negation. It follows that after ρ_{d-1} is applied the term length of g must still be n_{d-1}. Thus $\log_2 S \geq n_{d-1} = (n/14)(14 \log_2 S)^{-d+2}$. From this we obtain $(\log_2 S)^{d-1} \geq n/14^{d-1}$, $\log_2 S \geq n^{1/(d-1)}/14$, and so $S \geq 2^{\frac{1}{14} n^{1/(d-1)}}$ as required. □

In the argument above we fixed the restriction in an iterative fashion, with one piece fixed per level. Another way to show that the restriction exists is simply to choose the final restriction first. As we'll see in the case of distributions with 0 and 1 not equally likely, this method has advantages.

6.2 Clique

The main virtue of this approach is that we can now apply known properties of random assignments to find restrictions with extra properties from among those that keep the decision tree height small.

Lemma 7: Let $k \leq \log_2 n$, $q = n^{-3/k} \leq 1/2$, and for $i \geq 0$ let $\ell_i = n/(n^{5i\sqrt{(\log_n 6Sd)/k}})$. Let C be a circuit of size S and depth d computing a function f and suppose that $\ell_d \geq \sqrt{k(\log n 6Sd)}$. For ρ chosen at random from $\mathcal{C}^{\ell_d, q}$, with probability at least $1/2$, $f\lceil_\rho$, is representable by a decision tree with vertex height $\leq \sqrt{k(\log_n 6Sd)}$.

Proof: Choose a restriction ρ at random from $\mathcal{C}_n^{\ell_d,q}$. Choose a random ordering of the vertices set by ρ and order the edge variables set by ρ so that an edge is listed in order of the endpoint that occurs first in this ordering. For $1 \le i \le d$, let π_i be the restriction in $\mathcal{C}_{\ell_{i-1}}^{\ell_i,q}$ that agrees with ρ on the edge variables associated with vertices $\ell_i - \ell_{i-1}$ in this ordering. Note that the probability of ρ is the same as that of the joint probability of the set of π_i's. Let ρ_0 be the empty restriction and $\rho_{i+1} = \rho_i \pi_i$ for $i \ge 0$.

Let p_i be the probability that ρ_i fails to keep the vertex heights of the decision trees for the functions computed at gates of depth $\le i$ bounded by $\sqrt{k(\log_n 6Sd)}$. We'll show that $p_0 = 0$ and $p_i \le 1/2d$ for each $i \ge 1$ and derive the desired result. The proof is by an induction like that of Lemma 5. Again we do not need to worry about $\neg$-gates.

Base Case: $i = 0$. Since each gate at height 0 depends on only one variable, its associated function has vertex height 2. Thus $p_0 = 0$.

Induction Step: Suppose that the vertex height of the decision trees computing the functions associated with each gate of depth $\le i$ is at most $\sqrt{k(\log_n 6Sd)}$ after ρ_i is applied. Consider the function g associated with an $\vee$-gate at depth $i+1$. As in Lemma 5, $g\upharpoonright_{\rho_i}$ is expressible as a DNF formula with terms having vertex length at most $\sqrt{k(\log_n 6Sd)}$. Let $s = \sqrt{k(\log_n 6Sd)}$ and note that $p = \ell_{i+1}/\ell_i = n^{-5\sqrt{(\log_n 6Sd)/k}} = n^{-5s/k}$. Now applying the modified version of Lemma 3 from section 4, the total weight (probability) of restrictions $\pi \in \mathcal{C}_{\ell_i}^{\ell_{i+1},q}$ that fail to keep the vertex height of the decision tree for $(g\upharpoonright_{\rho_i})\upharpoonright_{\pi}$ at most $s = \sqrt{k(\log_n 6Sd)}$ is most

$$\begin{aligned}(8/3)((2/q)^{s-1/2}ps)^s &< 3[s(2n^{3/k})^{s-1/2}n^{-5s/k}]^s \\ &\le 3n^{4s^2/k}n^{-5s^2/k} = 3n^{-s^2/k} = 3n^{-(\log_n 6Sd)} = \frac{1}{2Sd}.\end{aligned}$$

Since there are at most S $\vee$-gates at depth $i+1$ in C, we have $p_{i+1} < \frac{1}{2d}$.

It now follows that the total failure probability is at most $1/2$ as required. □

Having randomly chosen the restriction at once, we can use the following simple property of random restrictions with probability q. (We include the proof merely for completeness. Much sharper bounds are available.)

Lemma 8: Let $k \ge 4$ and $q = n^{-3/k}$. With probability at least $3/4$, the size of the largest clique fixed by a random $\rho \in \mathcal{C}_n^{\ell_d,q}$ is at most $k-1$.

Proof: There are fewer than n^k vertex sets of size k that are potential cliques. For each such set the probability that all its edges are set to 1 by ρ is $q^{\binom{k}{2}}$. Thus the probability that some such set contains a clique on k nodes is at most

$$n^k q^{\binom{k}{2}} = (nq^{(k-1)/2})^k = (n^{1-\frac{3(k-1)}{2k}})^k = (n^{\frac{3}{2k}-\frac{1}{2}})^k \le n^{-\frac{k}{8}} < 1/4$$

for n sufficiently large. $\square$

Although the technique of choosing the whole restriction at once as in Lemma 7 makes Lemma 8 easy to state and prove, there is a slight degradation of the failure probability which leads to a slight reduction in the bound when compared with what might be obtained by choosing the restriction iteratively. To get this better bound one would need to prove a more detailed version of Lemma 8 which would say that each additional restriction π_i does not add much to the size of the largest clique generated so far. In any case, we can now derive the resulting lower bound.

Theorem 9: (Beame [Bea90]) Any unbounded fan-in circuit of depth d computing the $Clique_n^k$ function for $4 \leq k \leq \log_2 n$ requires size $S > n^{\frac{k}{100d^2}}/6d$.

Proof: Suppose that C computes the $Clique_n^k$ function in size S and depth d. From Lemma 7 with probability $\geq 1/2$ a randomly chosen restriction $\rho \in \mathcal{C}_n^{\ell_d,q}$ is such that $Clique_n^k\lceil_\rho$ is computable by a decision tree of vertex height at most $\sqrt{k(\log_n 6Sd)}$. Furthermore, with probability at least 3/4, by Lemma 8 a clique of at most $k-1$ nodes is set by ρ. Thus there is a restriction $\rho \in \mathcal{C}_n^{\ell_d,q}$ that has both properties.

Fix such a restriction ρ and suppose that $\ell_d \geq 2k$. In this case, $Clique_n^k\lceil_\rho$ is not identically 1 or 0. Any assignment that forces $Clique_n^k\lceil_\rho$ to 0 must set edge variables to 0 touching more than $\ell_d - k$ vertices unset by ρ. Thus the vertex height of the decision tree for $Clique_n^k\lceil_\rho$ must be more than k. In this case we must have $\sqrt{k(\log_n 6Sd)} > k$ or $\log_n 6Sd > k$ from which $S > n^k/6d$ which is much stronger than required.

Thus $\ell_d < 2k$ and so $n/(n^{5d\sqrt{(\log_n 6Sd)/k}}) < 2k$. Rewriting this we obtain $n^{5d\sqrt{(\log_n 6Sd)/k}} > n/2k$ or $5d\sqrt{(\log_n 6Sd)/k} > \log_n \frac{n}{2k} \geq 1/2$ for n sufficiently large. Squaring both sides we obtain $25d^2(\log_n 6Sd)/k > 1/4$ or $6Sd > n^{\frac{k}{100d^2}}$ which gives $S > n^{\frac{k}{100d^2}}/6d$ as required. $\square$

6.3 q-Matchings

We now give the analogue of Lemma 5 for the q-matching restrictions from $\mathcal{M}_{D,q}^\ell$. Unlike the switching lemma for $\mathcal{R}_n^\ell$ and $\mathcal{C}_n^\ell$, the conclusion of the switching lemma for $\mathcal{M}_{D,q}^\ell$ only says that the function after restriction may be represented by a small height q-matching decision tree as opposed to being exactly computed by such a tree. Thus we will not even be able to carry out the iterative restriction process to say that each gate of a small depth circuit in the q-matching variables can be represented by a small height decision tree. The statement will of necessity be somewhat weaker than that. First, in order to state the result we give a couple of definitions and a simple lemma.

If ρ is a partial q-matching restriction over D and T is a q-matching decision tree over D, then define $T\lceil_\rho$ to be the decision tree obtained from T by removing all paths which have a label that has been set to "0" by ρ, and contracting all edges whose labels are set to "1" by ρ. Note that for any partial q-matching restriction ρ over D, $disj(T\lceil_{\rho}) = disj(T)\lceil_{\rho}$.

Lemma 10: Let f be a boolean function over D and let T be a q-matching decision tree representing f over D. If ρ is a partial q-matching restriction over D, then $T\lceil_{\rho}$ is a q-matching decision tree representing $f\lceil_{\rho}$ over $D\lceil_{\rho}$.

Lemma 11: Let $m_0 = n$ and $m_{i+1} = (m_i/\log_2 S)^{1/q^2}/3e$ for $i \geq 0$ and suppose that $m_d \geq \log_2 S$. If C is a circuit of size S and depth d with q-matching variables over D then for every i, $0 \leq i \leq d$, there is a restriction $\rho_i \in \mathcal{M}_{D,q}^{m_i}$ such that for every node g of C (including the inputs) of depth at most i, there is an associated q-matching decision tree T_g^i of height less than $\log_2 S$ such that

(a) if g is an input x_e then T_g^i represents $x_e\lceil_{\rho_i}$ over $D\lceil_{\rho_i}$,

(b) if g is an $\vee$-gate with inputs $g_1, ..., g_k$ then T_g^i refines and represents $\bigvee_{j=1}^k disj(T_{g_j}^i)$ over $D\lceil_{\rho_i}$ and,

(c) if g is a $\neg$-gate with input h then $T_g^i = (T_h^i)^c$.

Proof: We prove the result by induction.

Base Case: $i = 0$. Let ρ_0 be the empty restriction. The only nodes of depth 0 are inputs and their negations. We associate input x_e with a decision tree $T_{x_e}^0$ of height 1 that queries the smallest numbered node in e and has a leaf label 1 on the unique branch labelled by e. Clearly this tree represents x_e over D as required. If $g = \neg x_e$ then let $T_g^0 = (T_{x_e}^0)^c$.

Induction Step: Assume that the lemma holds for $i \geq 0$. Consider an $\vee$-gate g at depth $i+1$ and suppose that the inputs to g are $g_1, ..., g_k$. The g_j all have depth at most i and thus have associated q-matching decision trees T_g^i of height $\leq \log_2 S$. Let $F_g^{i+1} = \bigvee_{j=1}^k disj(T_{g_j}^i)$ and notice that F_g^{i+1} is a $\log_2 S$-disjunction over $D\lceil_{\rho_i}$. Let $p = \frac{(m_i/\log_2 S)^{1/q^2}}{3em_i} = m_{i+1}/m_i$. By Lemma 4, for less than a $(4e^q(\log_2 S)^{1/q}p^q n^{q-1/q})^{\log_2 S} < 1/S$ fraction of all restrictions π in $\mathcal{M}_{D\lceil_{\rho_i},q}^{m_{i+1}}$ the height of the q-matching decision tree $T_{D\lceil_{\rho_i\pi}}(F_g^{i+1}\lceil_{\pi})$ is at least $\log_2 S$. Since there at most S $\vee$-gates g of depth $i+1$, there is some fixed restriction π such that for all gates at depth $i+1$, applying $\rho_i\pi$ leaves the height of $T_{D\lceil_{\rho_i\pi}}(F_g^i\lceil_{\pi})$ less than $\log_2 S$. Since $\pi \in \mathcal{M}_{D\lceil_{\rho_i},q}^{m_{i+1}}$, $\rho_{i+1} = \rho_i\pi \in \mathcal{M}_{D\lceil_{\rho_i},q}^{m_{i+1}}$ as required.

For each $\vee$-gate g at depth $i+1$ let $T_g^{i+1} = T_{D\lceil_{\rho_i\pi}}(F_g^{i+1}\lceil_{\pi})$. The associated tree for each $\neg$-gate at level $i+1$ is the complement of the tree of its input. For all $j \leq i$ and all gates h of depth j let $T_h^{i+1} = T_h^i\lceil_{\pi}$.

Now for each $\vee$-gate g at level $i+1$,

$$\begin{aligned} F_g^{i+1}\restriction_\pi &= (\bigvee_{j=1}^{k} disj(T_{g_j}^i))\restriction_\pi = \bigvee_{j=1}^{k} (disj(T_{g_j}^i))\restriction_\pi \\ &= \bigvee_{j=1}^{k} disj(T_{g_j}^i\restriction_\pi) = \bigvee_{j=1}^{k} disj(T_{g_j}^{i+1}) \end{aligned}$$

and thus $T_g^{i+1} = T_{D\restriction_{\rho_i\pi}}(F_g^{i+1}\restriction_\pi)$ refines and represents $\bigvee_{j=1}^k disj(T_{g_j}^{i+1})$ over $D\restriction_{\rho_i\pi} = D_{\rho_{i+1}}$ as required by the condition of the lemma.

It remains to show that all of the relationships at lower levels are maintained. First note that for each input x_e, since $T_{x_e}^i$ represents $x_e\restriction_{\rho_i}$ over $D\restriction_{\rho_i}$, $T_{x_e}^{i+1} = T_{x_e}^i\restriction_\pi$ represents $x_e\restriction_{\rho_i\pi} = x_e\restriction_{rho_{i+1}}$ over $D\restriction_{\rho_i\pi} = D\restriction_{\rho_{i+1}}$ as required. It is also easy to see that if $T_g^i = (T_h^i)^c$ then $T_g^{i+1} = T_g^i\restriction_\pi = (T_h^i)^c\restriction_\pi = (T_h^i\restriction_\pi)^c = (T_h^{i+1})^c$. Finally we see that for an $\vee$-gate g with inputs $g_1, ..., g_k$, since T_g^i refines and represents $\bigvee_{j=1}^k disj(T_{g_j}^i)$ over $D\restriction_{\rho_i}$, $T_g^{i+1} = T_g^i\restriction_\pi$ refines and represents $\bigvee_{j=1}^k disj(T_{g_j}^i)\restriction_\pi$ over $D\restriction_{\rho_i\pi} = D\restriction_{\rho_{i+1}}$; and that this is $\bigvee_{j=1}^k disj(T_{g_j}^{i+1})$ over $D\restriction_{\rho_{i+1}}$ as before. The lemma follows immediately. □

Corollary 12: If C is a circuit of depth d and size $S < 2^{n^{1/3q^{2d}}}/3$ in the q-matching variables over D then there is an m and a restriction $\rho \in \mathcal{M}_{D,q}^m$ such that for every node g of C (including the inputs), there is an associated q-matching decision tree T_g of height at most $\sqrt{m}$ such that

(a) if g is an input x_e then T_g represents $x_e\restriction_\rho$ over $D\restriction_\rho$,

(b) if g is an $\vee$-gate with inputs $g_1, ..., g_k$ then T_g refines and represents $\bigvee_{j=1}^k disj(T_{g_j})$ over $D\restriction_\rho$ and,

(c) if g is a $\neg$-gate with input h then $T_g = (T_h)^c$.

Proof: Consider the sequence of m_i from the statement of Lemma 11. Let $\delta_i = \sum_{j=0}^{i-1} q^{-2j}$. It is not hard to show by induction that $m_i = n^{q^{-2i}}/[3e(\log_2 S)^{q^{-2}}]^{\delta_i}$. Since $q \geq 2$, $\delta_i < 4/3$ so $m_i \geq n^{q^{-2i}}/[3e(\log_2 S)^{1/q^2}]^{4/3} > n^{1/q^{2i}}/(12\log_2 S)$. Now if $S < 2^{n^{1/3q^{2d}}}/3$ then $3\log_2 S < n^{1/3q^{2d}}$ and $27(\log_2 S)^3 < n^{1/q^{2d}}$ so $m_d > n^{1/q^{2d}}/(12\log_2 S) > (\log_2 S)^2$.

Thus we can apply Lemma 11 to C. Letting $m = m_d$ and $\rho = \rho_d$ produces the desired result. □

Corollary 13: With the somewhat smaller bound of $S < n^{n^{1/[5(2q^2)^d]}}/(2q)^{2d}$ the corollary above also holds with all the associated q-matching decision trees for gates having height at most $(2q)^{2d}\log_n S$ where m is at least $n^{4/[5(2q^2)^d]}/8$. Thus for constant d we can choose a restriction ρ leaving $qm+1$ nodes unset where $m = n^\gamma$ for some constant γ and so that the decision tree height is $O(\log_n S)$

Proof: We'll see that a version of Lemma 11 also holds with the height bound $s = (2q)^{2d}\log_n S$, $m'_0 = n$, and $m'_{i+1} = (m'_i/s)^{1/2q^2}/2e$. By the same reasoning as the corollary above we see that $m'_i = n^{1/(2q^2)^i}/[2es^{1/2q^2}]^{\epsilon_i}$ where $\epsilon_i = \sum_{j=0}^{i-1}(1/2q^2)^j$. Since $\epsilon_i < 8/7$ for all i, $m'_i \geq n^{1/(2q^2)^i}/(8s)$. By the condition of the corollary, $(2q)^{2d}\log_n S < n^{1/[5(2q^2)^d]}$, so $s < n^{1/[5(2q^2)^d]}$. Thus $m'_i \geq n^{4/[5(2q^2)^i]}/8$. and $(m'_i/s) \geq n^{1/[2(2q^2)^i]}$. Let $p = (m'_i/s)^{1/2q^2}/(2em'_i) = m'_{i+1}/m'_i$. Therefore applying Lemma 4 we get that the probability that we fail to find the appropriate restriction π for a given gate at depth $i+1$ is at most

$$\begin{aligned}
(6e^q p^q (m'_i)^{q-1/q} s^{1/q})^s &< (s/m'_i)^{s/2q^2} \\
&\leq (n^{1/[2(2q^2)^i]})^{-s/2q^2} \\
&= n^{-[(2q)^{2d-2}/(2q^2)^i]\log_n S} \\
&\leq 1/S^{2^{d-1}}
\end{aligned}$$

and thus the restriction π may always be found and the small height trees may be created at each step. The rest of the argument goes through exactly as before. □

7 Lower Bounds for CRCW PRAMs

The lower bounds proofs for CRCW PRAMs in [BH89, Bea90] can also be simplified in a similar manner to that described above for unbounded fan-in circuits. The key to the argument is to show at each step of the computation that, as a function of the input vector after a suitable restriction has been applied, the state of each processor and the contents of each memory cell may be described by a small height decision tree. (That is for each processor or memory cell, there is a decision tree of small height whose leaves are labelled by the potential states of that processor or contents of that memory cell.) The translations follow in a straightforward manner using the characterizations of CRCW PRAM computations in [BH89] and the arguments above.

Acknowledgements

The presentation of the results in this note was the result of discussions with several people, particularly Johan Håstad, Toni Pitassi, Russell Impagliazzo, and Steve Cook. Special thanks to Johan Håstad for suggesting the one-to-many approach for handling the Clique restrictions and to Sam Buss for suggesting a refinement to the proof of Lemma 4. Thanks also to Alasdair Urquhart for his comments on the paper.

References

[Ajt83] M. Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[Ajt88] M. Ajtai. The complexity of the pigeonhole principle. In *29th Annual Symposium on Foundations of Computer Science*, pages 346–355, White Plains, NY, October 1988. IEEE.

[Ajt89] M. Ajtai. First-order definability on finite structures. *Annals of Pure and Applied Logic*, 45:211–225, 1989.

[Ajt90] M. Ajtai. Parity and the pigeonhole principle. In *Feasible Mathematics*, pages 1–24. Birkhauser, 1990.

[Bea90] P. Beame. Lower bounds for recognizing small cliques on CRCW PRAM's. *Discrete Applied Mathematics*, 29(1):3–20, 1990.

[BH89] P. Beame and J. Håstad. Optimal bounds for decision problems on the CRCW PRAM. *Journal of the ACM*, 36(3):643–670, July 1989.

[BIK$^+$92] P. Beame, R. Impagliazzo, J. Krajíček, T. Pitassi, P. Pudlák, and A. Woods. Exponential lower bounds for the pigeonhole principle. In *Proceedings of the Twenty Fourth Annual ACM Symposium on Theory of Computing*, pages 200–220, Victoria, B.C., Canada, May 1992.

[BP93] P. Beame and T. Pitassi. An exponential separation between the matching principle and the pigeonhole principle. In *8th Annual IEEE Symposium on Logic in Computer Science*, Montreal, Quebec, June 1993.

[BPU91] S. Bellantoni, T. Pitassi, and A. Urquhart. Approximation and small depth Frege proofs. In *Proceedings, Structure in Complexity Theory, Sixth Annual Conference*, pages 367–391, Chicago, IL, June 1991. IEEE.

[Cai86] Jin-Yi Cai. With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, pages 21–29, Berkeley, California, May 1986.

[FSS81] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. In *22nd Annual Symposium on Foundations of Computer Science*, pages 260–270, Nashville, TN, October 1981. IEEE.

[Hås87] Johan Håstad. *Computational Limitations of Small-Depth Circuits.* ACM doctoral dissertation award, 1986. MIT Press, 1987.

[Ko91] Ker-I Ko. Separating the low and high hierarchies by oracles. *Information and Computation*, 90(2):156–177, 1991.

[KPW91] J. Krajíček, P. Pudlák, and A. Woods. Exponential lower bounds to the size of bounded-depth Frege proofs of the pigeonhole principle. Manuscript, 1991.

[Lyn86] J. Lynch. A depth-size tradeoff for Boolean circuits with unbounded fan-in. In Alan L. Selman, editor, *Structure in Complexity Theory*, volume 223 of *Lecture Notes in Computer Science*, pages 234–248, Berkeley, CA, June 1986. Springer-Verlag.

[PBI93] T. Pitassi, P. Beame, and R. Impagliazzo. Exponential lower bounds for the pigeonhole principle. *Computational Complexity*, 3(2):97–140, 1993.

[Raz93] A. Razborov. Bounded arithmetic and lower bounds in Boolean complexity. Submitted to Feasible Mathematics II, 1993.

[Sip83] M. Sipser. Borel sets and circuit complexity. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, pages 61–69, Boston, MA, April 1983.

[Yao85] A. C. Yao. Separating the polynomial hierarchy by oracles: Part I. In *26th Annual Symposium on Foundations of Computer Science*, pages 1–10, Portland, OR, October 1985. IEEE.