THE SHRINKAGE EXPONENT OF DE MORGAN FORMULAE IS 2

JOHAN HåSTAD
ROYAL INSTITUTE OF TECHNOLOGY
STOCKHOLM, SWEDEN

Abstract. We prove that if we hit a de Morgan formula of size L with a random restriction from R_p then the expected remaining size is at most $O(p^2(\log \frac{1}{p})^{3/2}L + p\sqrt{L})$. As a corollary we obtain a $\Omega(n^{3-o(1)})$ formula size lower bound for an explicit function in P. This is the strongest known lower bound for any explicit function in NP.

Key words. Lower bounds, formula size, random restrictions, computational complexity.

AMS subject classifications. 68Q25

1. Introduction. Proving lower bounds for various computational models is of fundamental value to our understanding of computation. Still we are very far from proving strong lower bounds for realistic models of computation, but at least there is more or less constant progress. In this paper we study formula size for formulae over the basis $\wedge$, $\vee$ and $\neg$. Our technique is based on random restrictions which were first defined and explicitly used in [2] although some earlier results can be formalized in terms of this type of random restrictions.

To create a random restriction in the space R_p we, independently for each variable, keep it as a variable with probability p and otherwise assign it the value 0 or 1 with equal probabilities $\frac{1-p}{2}$. Now suppose we have a function given by a de Morgan formula of size L. What will be the expected formula size of the induced function when we apply a random restriction from R_p? The obvious answer is that this size will be at most pL.

Subbotovskaya [11] was the first to observe that actually formulae shrink more. Namely she established an upper bound

$$O(p^{1.5}L + 1) \tag{1}$$

on the expected formula size of the induced function. This result allowed her to derive an $\Omega(n^{1.5})$ lower bound on the de Morgan formula size of the parity function.

This latter bound was superseded by Khrapchenko [12, 13] who, using a different method, proved a tight $\Omega(n^2)$ lower bound for the parity function. His result implied that the parity function shrinks by a factor $\theta(p^2)$, and provided an upper bound $\Gamma \leq 2$ on the *shrinkage exponent* Γ, defined as the least upper bound of all γ that can replace 1.5 in (1).

New impetus for research on the expected size of the reduced formula was given by Andreev [9] who, based upon Subbotovskaya's result, derived an $n^{2.5-o(1)}$ lower bound on the de Morgan formula size for a function in P. An inspection of the proof reveals that his method actually gives for the same function the bound $n^{\Gamma+1-o(1)}$.

New improvements of the lower bound on Γ followed. Nisan and Impagliazzo [5] proved that $\Gamma \geq \frac{21-\sqrt{73}}{8} \approx 1.55$. Paterson and Zwick [6], complementing the technique from [5] by very clever and natural arguments, pushed this bound further to $\Gamma \geq \frac{5-\sqrt{3}}{2} \approx 1.63$.

One can also define the corresponding notion for read-once formulae. For such formulae it was established in [3] that $\Gamma_{ro} = 1/\log_2(\sqrt{5}-1) \approx 3.27$. This result was made tight in [1], in that they removed a polylogarithmic factor in the bounds.

In this paper we continue (and possibly end) this string of results by proving that $\Gamma = 2$. To be more precise we prove that remaining size is $O(p^2(\log \frac{1}{p})^{3/2}L + p\sqrt{L})$. As discussed above this gives an $\Omega(n^{3-o(1)})$ lower bound for the formula size of the function defined by Andreev.

Our proof is by a sequence of steps. We first analyze the probability of reducing the formula to a single literal. When viewing the situation suitably, this first lemma gives a nice and not very difficult generalization of Khrapchenko's [12, 13] lower bounds for formula size. As an illustration of the power of this lemma we next, without too much difficulty, show how to establish the desired shrinkage when the formula is balanced. The general case is more complicated due to the fact that we need to rely on more dramatic simplifications. Namely, suppose that $\phi = \phi_1 \wedge \phi_2$ and ϕ_1 is much smaller than ϕ_2. Then, from an intuitive point of view, it seems like we are in a good position to prove that we have substantial shrinkage since it seems quite likely that ϕ_1 is reduced to the constant 0 and we can erase all of ϕ_2. The key new point in the main proof is that we have to establish that this actually happens. In the balanced case, we did not need this mechanism. The main theorem is established in two steps. First we prove that the probability that a formula of size at least 2 remains after we have applied a restriction from R_p is small, and then we prove that the expected remaining size is indeed small.

It is curious to note that all except our last and main result are proved even under an arbitrary but "favorable" conditioning, while we are not able to carry this through for the main theorem.

2. Notation. A *de Morgan formula* is a binary tree in which each leaf is labeled by a literal from the set $\{x_1, \ldots, x_n, \bar{x}_1, \ldots, \bar{x}_n\}$ and each internal node v is labeled by an operation which is either $\wedge$ or $\vee$. The *size* of a formula ϕ is defined as the number of leaves and is denoted by $L(\phi)$. The *depth* $D(\phi)$ is the depth of the underlying tree. The *size* and the *depth* of a Boolean function f are, respectively, the minimal size and depth of any de Morgan formula computing f in the natural sense. For convenience we define the size and depth of a constant function to be 0.

A *restriction* is an element of $\{0, 1, *\}^n$. For $p \in [0, 1]$ a random restriction ρ from R_p is chosen by that we set randomly and independently each variable to $*$ with probability p and to $0, 1$ with equal probabilities $\frac{1-p}{2}$. The interpretation of giving the value $*$ to a variable is that it remains a variable, while in the other cases the given constant is substituted as the value of the variable.

All logarithms in this paper are to the base 2. We use the notation x_i^ϵ to denote x_i when $\epsilon = 0$ and $\bar{x}_i$ when $\epsilon = 1$. We also need the concept of a filter.

DEFINITION 2.1. *A set of restrictions Δ is a* filter*, if when $\rho \in \Delta$, and $\rho(x_i) = *$, then for $\epsilon \in \{0, 1\}$ the restriction ρ' obtained by setting $\rho'(x_j) = \rho(x_j)$, for every $j \neq i$ and $\rho'(x_i) = \epsilon$ also belongs to Δ.*

For any event E we will use $Pr[E|\Delta]$ as shorthand for $Pr[E|\rho \in \Delta]$. Note that the intersection of two filters is a filter.

3. Preliminaries. We analyze the expected size of a formula after it has been hit with a restriction from R_p. The variables that are given values are substituted into the formula after which we use the following rules of simplification:

- If one input to a $\vee$-gate is given the value 0 we erase this input and let the other input of this gate take the place of the output of the gate.
- If one input to a $\vee$-gate is given the value 1 we replace the gate by the constant 1.
- If one input to a $\wedge$-gate is given the value 1 we erase this input and let the other input of this gate take the place of the output of the gate.
- If one input to a $\wedge$-gate is given the value 0 we replace the gate by the constant 0.
- If one input of a $\vee$-gate is reduced to the single literal x_i ($\bar{x}_i$) then $x_i = 0$ (1) is substituted in the formula giving the other input to this gate. If possible we do further simplifications in this subformula.
- If one input of a $\wedge$-gate is reduced to the single literal x_i ($\bar{x}_i$) then $x_i = 1$ (0) is substituted in the formula giving the other input to this gate. If possible we do further simplifications in this subformula.

We call the last two rules the *one-variable simplification rules*. All rules preserve the function the formula is computing. Observe that the one-variable simplification rules are needed to get a nontrivial decrease of the size of the formula as can be seen from the pathological case when the original formula consists of an $\vee$ (or $\wedge$) of L copies of a single variable x_i. If these rules did not exist then with probability p the entire formula would remain and we could get an expected remaining size which is pL. Using the above rules we prove that instead we always get an expected remaining size which is at most slightly larger than p^2L.

In order to be able to speak about the reduced formula in an unambiguous way let us be more precise about the order we do the simplification. Suppose that ϕ is a formula and that $\phi = \phi_1 \wedge \phi_2$. We first make the simplification in ϕ_1 and ϕ_2 and then only later the simplifications which are connected with the top gate. This implies that the simplified ϕ will not always consist of a copy of simplified ϕ_1 and a copy of simplified ϕ_2 since the combination might give more simplifications. In particular, this will happen if ϕ_1 is simplified to one variable x_i since then $x_i = 1$ will be substituted in the simplified ϕ_2. Whenever a one-variable simplification rule actually results in a change in the other subformula we say that a one-variable simplification is *active* at the corresponding gate.

We let $\phi\lceil_\rho$ denote formula that results when the above simplifications are done to ϕ. As usual $L(\phi\lceil_\rho)$ denotes the the size of this formula.

It is important to note that simplifications have a certain commutativity property. We say that two restrictions ρ_1 and ρ_2 are *compatible* if they never give two different constant values to the same x_i. In other words, for any x_i the pair $(\rho_1(x_i), \rho_2(x_i))$ is one of the pairs $(*,*), (*,0), (*,1), (0,*), (0,0), (1,*)$ or $(1,1)$. For compatible restrictions we can define the combined restriction $\rho_1 \circ \rho_2$ which in the mentioned 7 cases takes the values $*, 0, 1, 0, 0, 1, 1$ respectively. This combined restriction is the result (on the variable level) of doing first ρ_1 and then doing ρ_2 on the variables given $*$ by ρ_1. Note that the fact that ρ_1 and ρ_2 are compatible makes the combining operator commutative. We need to makes sure that combination acts in the proper way also on formulae.

LEMMA 3.1. *Let ρ_1 and ρ_2 be two compatible restrictions then for any ϕ,*

$$(\phi\lceil_{\rho_1})\lceil_{\rho_2} = (\phi\lceil_{\rho_2})\lceil_{\rho_1} = \phi\lceil_{\rho_1 \circ \rho_2}.$$

Proof. Let ρ denote $\rho_1 \circ \rho_2$. Clearly we need only establish $(\phi\lceil_{\rho_1})\lceil_{\rho_2} = \phi\lceil_\rho$ since the other equality follows by symmetry. We proceed by induction over the size of ϕ. When the size is 1 the claim is obvious. For the general case we assume that $\phi = \phi_1 \wedge \phi_2$ (the case $\phi = \phi_1 \vee \phi_2$ being similar) and we have two cases:

1. Some one-variable simplification rule is active at the top gate when defining ϕ_{ρ_1}.
2. This is not the case.

In case 2 there is no problem since there is no interaction between ϕ_1 and ϕ_2 until after both ρ_1 and ρ_2 have been applied. Namely, by induction $\phi_j\lceil_\rho = (\phi_j\lceil_{\rho_1})\lceil_{\rho_2}$ for $j = 1, 2$ and since we do all simplification to the subformulae before we do any simplification associated with the top gate, the result will be the same in both cases.

In case 1, assume that $\phi_1\lceil_{\rho_1} = x_j^\epsilon$. This means that $x_j = \bar{\epsilon}$ is substituted in $\phi_2\lceil_{\rho_1}$. Viewing this as a restriction $\rho^{(j)}$ giving a non-$*$ value to only one variable, ϕ_2 is simplified to $(\phi_2\lceil_{\rho_1})\lceil_{\rho^{(j)}}$. Now we have three possibilities depending on the value of ρ_2 on x_j.

When $\rho_2(x_j) = \epsilon$ then $(\phi_1\lceil_{\rho_1})\lceil_{\rho_2} \equiv 0$ and by induction $\phi_1\lceil_\rho \equiv 0$ and hence $(\phi\lceil_{\rho_1})\lceil_{\rho_2} = \phi\lceil_\rho \equiv 0$.

When $\rho_2(x_j) = \bar{\epsilon}$ then by induction $(\phi_1\lceil_{\rho_1})\lceil_{\rho_2} = \phi_1\lceil_\rho \equiv 1$, and since $\rho^{(j)}$ is compatible with (even a subassignment of) ρ_2 we have

$$(\phi\lceil_{\rho_1})\lceil_{\rho_2} = ((\phi_2\lceil_{\rho_1})\lceil_{\rho^{(j)}})\lceil_{\rho_2} = (\phi_2\lceil_{\rho_1})\lceil_{\rho_2} = \phi_2\lceil_\rho = \phi\lceil_\rho$$

where we again have used the induction hypothesis.

When $\rho_2(x_j) = *$ then since (by induction) $\phi_1\lceil_\rho = (\phi_1\lceil_{\rho_1})\lceil_{\rho_2} = x_j^\epsilon$ when simplifying by ρ, ϕ_2 will be simplified also by the restriction $\rho^{(j)}$ and will be $(\phi_2\lceil_\rho)\lceil_{\rho(j)}$ which by induction is equal to $\phi_2\lceil_{\rho\circ\rho^{(j)}}$. On the other hand when simplifying by ρ_1 and then by ρ_2 ϕ_2 reduces to $((\phi_2\lceil_{\rho_1})\lceil_{\rho^{(j)}})\lceil_{\rho_2}$ which by applying the induction hypothesis twice is equal to $\phi_2\lceil_{\rho_1\circ\rho^{(j)}\circ\rho_2}$ and since $\rho_1 \circ \rho^{(j)} \circ \rho_2 = \rho \circ \rho^{(j)}$ the lemma follows. □

We will need the above lemma in the case of analyzing what happens when we use the one-variable simplification rules. In that case the restriction ρ_2 will just give a non-$*$ value to one variable. Since the lemma is quite simple and natural we will not always mention it explicitly when we use it.

Two examples to keep in mind during the proof are the following:

1. Suppose ϕ computes the parity of m variables and is of size m^2. Then if p is small, the probability that ϕ will depend on exactly one variable is about $pm = p\sqrt{L(\phi)}$ and if p is large, we expect that the remaining formula will compute the parity of around pm variables and thus be of size at least $(pm)^2 = p^2 L(\phi)$.
2. Suppose ϕ is the $\wedge$ of $L/2$ copies of $x_1 \vee x_2$. By our rules of simplification, this will not be simplified if both x_1 and x_2 are given the value $*$ by the restriction. Hence with probability p^2 the entire formula remains and we get expected remaining size of at least $p^2 L$.

4. Reducing to size 1. We start by estimating the probability that a given formula reduces to size one. For notational convenience we set $q = \frac{2p}{1-p}$. This will be useful since we will change the values of restrictions at individual points and if we change a non-$*$ value to $*$, we multiply the probability by q. Since we are interested in the case when p is small, q is essentially $2p$.

LEMMA 4.1. *Let ϕ be a formula of size L and ρ a random restriction in R_p. Let E_δ be the event that ϕ is reduced to the constant δ by ρ for $\delta = 0, 1$. Furthermore let Δ be any filter. Then $Pr[L(\phi\lceil_\rho) = 1|\Delta]$ is bounded by*

$$q\left(LPr[E_0|\Delta]Pr[E_1|\Delta]\right)^{1/2}.$$

Remark: Most of the time the implied bound $q\sqrt{L/4}$ (which follows from $x(1-x) \leq 1/4$) will be sufficient.

Proof. Let ρ be a restriction that satisfies $L(\phi\lceil_\rho) = 1$ and belongs to Δ and suppose that ρ makes ϕ into the literal x_i^ϵ. By definition $\rho(x_i) = *$ and we have two fellow restrictions ρ^δ, $\delta = 0, 1$ where ρ^δ is obtained by changing $\rho(x_i)$ to $xor(\epsilon, \delta)$. ρ^δ contributes to the event E_δ and by the definition of a filter it belongs to Δ. We can hence identify the set of restrictions we are interested in with edges between restrictions that reduce the formula to the constants 0 and 1 respectively and we are on familiar grounds.

Let A be set the of restrictions that satisfy E_0 and belong to Δ and let B be the set of restrictions that satisfy E_1 and belong to Δ. We partition $A \times B$ into rectangles $A_j \times B_j$, where for each j there is some variable x_{i_j} which takes a different value in A_j and in B_j. This was first done in [10] (see also [7]), but we will here need a slight generalization and thus we choose to use the more intuitive framework introduced by Karchmer and Wigderson [4].

In the normal KW-game P_1 gets an input, x, from $f^{-1}(1)$ while P_0 gets an input, y, from $f^{-1}(0)$ and their task is to find a coordinate i such that $x_i \neq y_i$. This is solved by tracing the formula from the output to an input maintaining the property that the two inputs give different values to the gates on the path. This is achieved in the following way. At an $\wedge$-gate, P_0 points to the input of this gate that evaluates to 0 on y. Similarly at an $\vee$-gate P_1 points to the input that evaluates to 1 on x.

We extend this game by giving P_δ a restriction ρ_δ that simplifies the formula to the constant δ. The task is to find an x_i on which the two restrictions take different values (we allow answers where one restriction takes the value $*$ and the other takes the value 0 or 1).

To solve this game both players start by setting $\rho_\delta(x_j) = 1$ for each j such that $\rho_\delta(x_j) = *$. After this they play the standard KW-game. If the path ends at literal x_i^ϵ, then in the extended restrictions $\rho_\delta(x_i) = xor(\delta, \epsilon)$. Note that if $\rho_\delta(x_i) = 0$, then this was the initial value (since we only change values to 1), while if $\rho_\delta(x_i) = 1$ then the initial value was 1 or $*$. In either case we solve the problem.

The extended KW-game creates a partition of $A \times B$ and let $A_j \times B_j$ be the inputs that reach leaf j. Note that the fact that the set of inputs that reach a certain leaf is a product set follows from the fact that each move of the game is determined by one of the players based only on his own input. Let C_j be the set of restrictions ρ that satisfy $L(\phi\lceil_\rho) = 1$ and belong to Δ and such that the pair (ρ^0, ρ^1) reaches leaf j. By necessity the literal that appears at leaf j is the literal to which ρ reduced the formula. Now, note that the probability of C_j is bounded by q times the probability of A_j. This follows since the mapping $\rho \mapsto \rho^0$ gives a one-to-one correspondence of C_j with a subset of A_j and that $Pr(\rho) = qPr(\rho^0)$ for each ρ. We have the same relation between C_j and B_j and hence

$$Pr[L(\phi\lceil_\rho) = 1|\Delta] = \sum_j Pr[C_j|\Delta] \leq \sum_j q\left(Pr[A_j|\Delta]Pr[B_j|\Delta]\right)^{1/2} \leq$$

$$q \left(\sum_j 1 \right)^{1/2} \left(\sum_j Pr[A_j|\Delta]Pr[B_j|\Delta] \right)^{1/2} = q \left(LPr[A|\Delta]Pr[B|\Delta] \right)^{1/2}$$

where we used Cauchy-Schwartz inequality. □

Remark: Please note that the theorem of Khrapchenko is indeed a special case of this lemma. Khrapchenko starts with $A \subseteq f^{-1}(0)$, $B \subseteq f^{-1}(1)$ and C which is a set of edges of the form (a, b) where $a \in A$, $b \in B$ and the hamming distance between a and b is one. As noted above each such edge naturally corresponds to a restriction ρ by setting $\rho(x_i) = a_i$ when $a_i = b_i$ and $\rho(x_{i_0}) = *$ for the unique coordinate i_0 such that $a_{i_0} \neq b_{i_0}$. Abusing notation we can identify the restriction and the edge. Now setting $\Delta = A \cup B \cup C$ we get a filter and since

$$\frac{Pr[C]}{q|C|} = \frac{Pr[A]}{|A|} = \frac{Pr[B]}{|B|}$$

the lemma reduces to Khrapchenko's theorem, i.e. to

$$L \geq \frac{|C|^2}{|A| \cdot |B|}.$$

5. The balanced case. In the general case we cannot hope to have an estimate which depends on the probability of reducing to either constant. The reason is that formulae that describe tautologies do not always reduce to the constant 1.

It remains to take care of the probability that the remaining formula is of size greater than 1.

DEFINITION 5.1. *Let $L^2(\phi)$ be the expected size of the remaining formula where we ignore the result if it is of size 1, i.e. $L^2(\phi) = \sum_{i=2}^{\infty} iPr[L(\phi\lceil_\rho) = i]$. Furthermore, let $L^2(\phi|\Delta)$ be the same quantity conditioned on $\rho \in \Delta$. Here we think of ρ as taken randomly from R_p and thus $L^2(\phi)$ depends on the parameter p. We will, however, suppress this dependence.*

To familiarize the reader with the ideas involved in the proof, we first prove the desired result when the formula is balanced. We will here take the strictest possible definition of balanced, namely that the formula is a complete binary tree and just establish the size of the reduced formula as a function of its original depth.

THEOREM 5.2. *Let ϕ be a formula of depth d and ρ a random restriction in R_p. Let Δ be any filter and assume that $q \leq (d2^{1+d/2})^{-1}$, then*

$$L^2(\phi|\Delta) \leq q^2 d 2^{d-1}.$$

Proof. The proof is by induction over the depth. The base case ($d = 1$) is obvious. Now suppose $\phi = \phi_1 \wedge \phi_2$ (the $\vee$-case is similar) where the depth of each ϕ_i is $d - 1$ and hence the size is bounded by 2^{d-1}. We need to consider the following events:

1. $L(\phi_i\lceil_\rho) \geq 2$, for $i = 1$ or $i = 2$.
2. $L(\phi_1\lceil_\rho) = 1$ and the size of ϕ_2 after the simplification by ρ and the application of the one-variable simplification rule is at least 1.
3. $L(\phi_2\lceil_\rho) = 1$ and the size of ϕ_1 after the simplification by ρ and the application of the one-variable simplification rule is at least 1.

The estimate for $L^2(\phi|\Delta)$ is now

$$2 \cdot q^2(d-1)2^{d-2} + Pr[case\ 2] + Pr[case\ 3].$$

The first term comes from any subformula of size at least two appearing in either of the three cases while the other two terms cover the new contribution in the respective cases.

Let us analyze the probability of case 2. Let p_i^ϵ be the probability that ϕ_1 reduces to x_i^ϵ. We know by Lemma 4.1 that

$$\sum p_i^\epsilon \leq q2^{(d-3)/2}.$$

Now consider the conditional probability that, given that ϕ_1 reduces to x_i^ϵ, ϕ_2 does not reduce to a constant. The condition that ϕ_1 reduces to x_i^ϵ can be written as "$\rho \in \Delta' \wedge \rho(x_i) = *$" for some filter Δ'. The reason for this is that if ρ reduces ϕ_1 to x_i^ϵ, then changing any $\rho(x_j)$, $j \neq i$ from $*$ to a constant the resulting restriction still reduces ϕ_1 to x_i^ϵ. This follows by Lemma 3.1. Thus we should work with the conditioning $\rho \in \Delta \bigcap \Delta' \wedge \rho(x_i) = *$. Now we substitute $x_i = \bar{\epsilon}$ in ϕ_2 and we can forget the variable x_i and just keep the restrictions in $\Delta \bigcap \Delta'$ that satisfy $\rho(x_i) = *$ (as restrictions on the other $n-1$ variables). This yields a filter Δ''. Thus we want to estimate the conditional probability that $\phi_2\lceil_{x_i=\epsilon}$ does not reduce to a constant given that $\rho \in \Delta''$. But now we are in position to apply induction and hence this probability can be estimated by

$$q2^{(d-3)/2} + \frac{1}{2}q^2(d-1)2^{d-2},$$

where the first term is given by Lemma 4.1 and the second term is a bound on $\frac{1}{2}L^2(\phi_2|\Delta'')$. Using $q \leq (d2^{1+d/2})^{-1}$ we get the total bound $q2^{d/2-1}$. This implies that the total probability of case 2 is bounded by

$$\sum p_i^\epsilon q2^{d/2-1} \leq q^2 2^{d-5/2}.$$

the probability of case 3 can be bounded the same way and finally

$$2 \cdot q^2(d-1)2^{d-2} + q^2 2^{d-3/2} \leq q^2 d2^{d-1},$$

and the proof is complete. □

It is not difficult to extend Theorem 5.2 to larger d, but we leave the details to the reader.

6. The probability of size at least 2 remaining. The reason why things are so easy in the balanced case is that we need not rely on very complicated simplifications. In particular, we did not need the fact that a subformula can kill its brother. This will be needed in general and let us start by:

LEMMA 6.1. *Let ϕ be a formula of size L and ρ a random restriction in R_p. Let Δ be any filter and assume that $q \leq (2\sqrt{L\log L})^{-1}$. Then the probability that $L(\phi\lceil_\rho) \geq 2$ conditioned on $\rho \in \Delta$ is at most $q^2L(\log L)^{1/2}$.*

Proof. We prove the lemma by induction over the size of the formula. It is not hard to see that the lemma is correct when $L \leq 2$. Suppose that $\phi = \phi_1 \wedge \phi_2$ (the $\vee$-case is similar) where $L(\phi_i) = L_i$, $L_1 + L_2 = L$, and $L_1 \leq L_2$. We divide the event in question into the following pieces:

1. $L(\phi_1\lceil_\rho) \geq 2$
2. $L(\phi_1\lceil_\rho) = 1$ and even after the one-variable simplification rule is used the size of the simplified ϕ_2 is at least 1.
3. ϕ_1 is reduced to the constant 1 by ρ and $L(\phi_2\lceil_\rho) \geq 2$.

The probability of the first event is by induction bounded by $q^2 L_1(\log L_1)^{1/2}$. Suppose the probability, conditioned on $\rho \in \Delta$, that ϕ_1 is reduced to the constant 1 is Q. Call the corresponding set of restrictions Δ'. Note that Δ' is a filter. Then using the induction hypothesis with the conditioning $\Delta \bigcap \Delta'$ we get the probability of the third event to be bounded by

$$Qq^2 L_2(\log L_2)^{1/2}.$$

Let us now estimate the probability of the second case. The probability that $L(\phi_1\lceil_\rho) = 1$ is by Lemma 4.1 bounded by $q\left(L_1 Q(1-Q)\right)^{1/2}$. To estimate the conditional probability that, given that $L(\phi_1\lceil_\rho) = 1$, ϕ_2 remains to be of size at least 1 we reason exactly as in the proof of Theorem 5.2. Hence, this probability can be estimated by

$$q\sqrt{L_2/4} + q^2 L_2(\log L_2)^{1/2},$$

where the first term comes from an application of Lemma 4.1 and the second term from the inductive assumption. Thus for the second case we get the total bound

$$q\left(L_1 Q(1-Q)\right)^{1/2} \times \left(q\sqrt{L_2/4} + q^2 L_2(\log L_2)^{1/2}\right) \leq$$

$$q^2\left(L_1 L_2 Q(1-Q)\right)^{1/2} \leq q^2\left(L_1 L_2(1-Q)\right)^{1/2},$$

where we used $q \leq (2\sqrt{L\log L})^{-1} \leq (2\sqrt{L_2 \log L_2})^{-1}$. We want to bound the sum of the estimates for probabilities of cases 2 and 3. Differentiating

$$QL_2(\log L_2)^{1/2} + \left(L_1 L_2(1-Q)\right)^{1/2}$$

with respect to Q yields

$$L_2(\log L_2)^{1/2} - \frac{1}{2}\left(\frac{L_1 L_2}{1-Q}\right)^{1/2}.$$

Thus the derivative is 0 when

$$(1-Q) = \frac{L_1}{4L_2 \log L_2}$$

and this corresponds to a maximum since the second derivative is negative. Using this optimal value for Q we get a total estimate which is

$$q^2 L_1\left(\log L_1\right)^{1/2} + q^2 L_2\left(\log L_2\right)^{1/2} - \frac{q^2}{4} L_1(\log L_2)^{-1/2} + \frac{q^2}{2} L_1(\log L_2)^{-1/2} =$$

$$q^2 L_1\left(\log L_1\right)^{1/2} + q^2 L_2\left(\log L_2\right)^{1/2} + \frac{q^2}{4} L_1(\log L_2)^{-1/2}$$

and we need to prove that this is less than $q^2L(\log L)^{1/2}$ when $L_1 \le L_2$ and $L = L_1 + L_2$. This is only calculus, but let us give the proof.

First note that $\log L_2 \ge \log\lceil L/2\rceil \ge \frac{1}{2}\log L$ when $L \ge 3$ and hence it is sufficient to bound

$$q^2L_1(\log L_1)^{1/2} + q^2L_2(\log L_2)^{1/2} + \frac{q^2}{2}L_1(\log L)^{-1/2}.$$

Now if we set $H(x) = x(\log x)^{1/2}$ it is not difficult to see that $H''(x)$ is positive for $x \ge 2$ and hence the above expression is convex for $L_1 \ge 2$. This means that it is maximized either for $L_1 = 1, 2$ or $L_1 = L_2$. The first two cases correspond to the inequalities

$$(L-1)(\log(L-1))^{1/2} + \frac{1}{2}(\log L)^{-1/2} \le L(\log L)^{1/2}$$

and

$$2 + (L-2)(\log(L-2))^{1/2} + (\log L)^{-1/2} \le L(\log L)^{1/2}$$

which are easily seen to hold for $L \ge 3$ and $L \ge 4$ respectively. To check the required inequality when $L_1 = L_2$ we need to prove that

$$L(\log L/2)^{1/2} + \frac{L}{4}(\log L)^{-1/2} \le L(\log L)^{1/2}.$$

This follows from $\sqrt{x} - \sqrt{x-1} \ge \frac{1}{2\sqrt{x}}$ which is valid for all $x \ge 1$. $\square$

7. Main shrinkage theorem. We are now ready for our main theorem.

THEOREM 7.1. *Let ϕ be a formula of size L and ρ a random restriction in R_p. Then the expected size of $\phi\lceil_\rho$ is bounded by*

$$O\left(p^2(1 + (\log(\min(\frac{1}{p}, L)))^{3/2})L + p\sqrt{L}\right).$$

Crucial to this theorem is the following lemma:

LEMMA 7.2. *Let ϕ be a formula of size L and ρ a random restriction in R_p. If $q \le (2\sqrt{L\log L})^{-1}$, then*

$$L^2(\phi) \le 30q^2L(\log L)^{3/2},$$

while if $\frac{1}{2} \ge q \ge (4\sqrt{L\log L})^{-1}$, then

$$L^2(\phi) \le 200q^2L(\log q^{-1})^{3/2}.$$

First note that Theorem 7.1 follows from Lemma 7.2 together with Lemma 4.1 and thus it is sufficient to prove Lemma 7.2. Also note that Lemma 7.2 and Theorem 7.1 do not consider conditional probabilities. There are two reasons for this. It is not needed and the proof does not seem to allow it.

Proof. (Of Lemma 7.2) The hard part of the lemma is the case when q is small and we will start by establishing this case. As before we proceed by induction. The base case ($L = 1$) is obvious. Suppose that $\phi = \phi_1 \wedge \phi_2$ (the $\vee$-case is similar) where $L(\phi_i) = L_i$, $L_1 + L_2 = L$ and $L_1 \leq L_2$.

Our basic estimate for $L^2(\phi)$ is $L^2(\phi_1) + L^2(\phi_2)$, which by induction can be bounded by

$$S(L_1, L_2) = 30q^2 L_1 (\log L_1)^{3/2} + 30q^2 L_2 (\log L_2)^{3/2}.$$

We need to revise this estimate and in particular we need to consider the events which contribute either to the event described by the basic estimate ($L(\phi_i\lceil_\rho) \geq 2$ for $i = 1$, or $i = 2$) or to the event we are trying to estimate ($L(\phi\lceil_\rho) \geq 2$).

1. We have $L(\phi_i\lceil_\rho) \geq 2$ for $i = 1$ and $i = 2$.
2. $L(\phi_i\lceil_\rho) = 1$ for $i = 1$ or $i = 2$, and the one-variable simplification rule was active at the top gate.
3. $L(\phi_1\lceil_\rho) = L(\phi_2\lceil_\rho) = 1$ and the one-variable simplification rule was not active at the top gate.
4. $L(\phi_1\lceil_\rho) = 1$ and $L(\phi_2\lceil_\rho) \geq 2$ and the one-variable simplification rule was not active at the top gate.
5. $L(\phi_2\lceil_\rho) = 1$ and $L(\phi_1\lceil_\rho) \geq 2$ and the one-variable simplification rule was not active at the top gate.
6. The function ϕ_1 is reduced to the constant 0 while $L(\phi_2\lceil_\rho) \geq 2$.
7. The function ϕ_1 is reduced to the constant 1 while $L(\phi_2\lceil_\rho) \geq 2$.
8. The function ϕ_2 is reduced to the constant 0 while $L(\phi_1\lceil_\rho) \geq 2$.
9. The function ϕ_2 is reduced to the constant 1 while $L(\phi_1\lceil_\rho) \geq 2$.

Let us first investigate what corrections we need to make to our basic estimate in the various cases.

Case 1 The basic estimate is correct.

Case 2 Suppose that ϕ_1 reduces to x_i^ϵ. If the resulting formula is of size at least 2 then ϕ_2 was of size at least 2 before we did the simplification of substituting $\bar{\epsilon}$ for x_i. This reduced the size of ϕ_2 by at least one. This means that the formula size of ϕ is at most the formula size of ϕ_2 before we did this simplification. But in our basic estimate we have not taken this simplification into account and thus we need not add anything to our basic estimate in this case.

Case 3 In this case we need to add 2 to our basic estimate. We will need some work to estimate the probability of this case.

Case 4 In this case we need to add 1 to our basic estimate. Also the probability of this event needs a little bit of work.

Case 5 In this case we need to add 1 to our basic estimate. From our previous work we can estimate the probability of this event simply by the probability that the remaining size of ϕ_1 is at least 2 and this probability is by Lemma 6.1 bounded by

$$q^2 L_1 (\log L_1)^{1/2}.$$

Cases 6 and 8 In this case we can subtract at least 2 from our original estimate. This follows since in this case we erase a formula of size at least 2 which contributed needlessly to the basic estimate. We will only use case 6.

Cases 7 and 9 The basic estimate is correct.

The above reasoning gives the total bound:

$$S(L_1, L_2) + 2Pr[case\ 3] + Pr[case\ 4] + Pr[case\ 5] - 2Pr[case\ 6].$$

We have already bounded $Pr[case\ 5]$ and for the other probabilities we will establish:

LEMMA 7.3. *If* $q \leq (2\sqrt{L_1 \log L_1})^{-1}$, *then*

$$Pr[case\ 3] \leq 20q^2 L_1 + \frac{1}{2}q^2 L_1 (\log L_1)^{1/2} + \frac{1}{2} Pr[case\ 6]$$

and

LEMMA 7.4. *If* $q \leq (2\sqrt{L_1 \log L_1})^{-1}$, *then*

$$Pr[case\ 4] \leq q^2 L_1 + Pr[case\ 6]$$

Let us just check that this is sufficient to prove Lemma 7.2 in the case when q is small.

$$S(L_1, L_2) + 2Pr[case\ 3] + Pr[case\ 4] + Pr[case\ 5] - 2Pr[case\ 6] \leq$$

$$S(L_1, L_2) + 41q^2 L_1 + q^2 L_1 (\log L_1)^{1/2}$$

We need to prove that this is bounded by $30q^2 L(\log L)^{3/2}$ for all possible L_1. Substituting $L_2 = L - L_1$ and differentiating twice we see that this is a convex function of L_1 when $2 \leq L_1 \leq L/2$. This means that it is maximized either for $L_1 = 1, 2$ or $L_1 = L_2$.

For $L_1 = 1$ we need to establish

$$30(L-1)(\log{(L-1)})^{3/2} + 41 \leq 30L(\log L)^{3/2}$$

which is easily checked to be true for $L = 2$ and $L = 3$. For $L \geq 4$ the inequality follows from

$$L(\log L)^{3/2} \geq (L-1)(\log L)^{3/2} + 2.$$

For $L_1 = 2$ we need to check

$$30(L-2)(\log{(L-2)})^{3/2} + 60 + 82 + 1 \leq 30L(\log L)^{3/2}$$

which for $L \geq 4$ follows from

$$L(\log L)^{3/2} \geq (L-2)(\log L)^{3/2} + 2 \cdot 2^{3/2}$$

and $60 \cdot 2^{3/2} > 143$.

Finally, for $L_1 = L/2$ we need to estimate

$$30L(\log L/2)^{3/2} + L/2(41 + (\log L/2)^{1/2})$$

and using the inequality $x^{3/2} - (x-1)^{3/2} \geq x^{1/2}$, which is valid for any $x \geq 1$, this can be bounded by

$$30L(\log L)^{3/2} - 30L(\log L)^{1/2} + \frac{43}{2}L(\log L)^{1/2} \leq 30L(\log L)^{3/2}$$

and we are done.

Hence we need just establish Lemma 7.3 and Lemma 7.4. The basic principle is to start with a set of restrictions that contribute to the bad case (cases 3 and 4 respectively) and create a set of restrictions that contribute to the good case, namely case 6. In this process there will be some "spills" and hence we need the additive terms. The Lemma 7.4 is by far the easier, and since the basic outline is the same, we start with this lemma.

Proof. (Of Lemma 7.4) Let C be the set of restrictions such that ϕ_1 reduces to exactly one variable or its negation and such that the reduced ϕ_2 does not contain this variable. Let A be the set of restrictions that is formed by setting the variable that remains in ϕ_1 in such a way to make ϕ_1 reduce to the constant 0 and let B be the corresponding set that makes ϕ_1 reduce to 1. Each element in C corresponds to an edge between A and B and we can (as in the proof of Lemma 4.1) let this define a path in ϕ_1. Thus each leaf in ϕ_1 corresponds to a set $A_j \times B_j$ which reaches this leaf and a subset C_j of C such that for any $\rho \in C_j$, its neighbors belong to A_j and B_j respectively. The sets $A_j \times B_j$ form a partition of $A \times B$. Suppose furthermore that the literal at leaf j of ϕ_1 is $x_{d_j}^{\epsilon_j}$. Note that this implies that if $\rho \in C_j$ then ρ simplifies ϕ_1 to $x_{d_j}^{\epsilon_j}$.

Let q_j be the conditional probability that, when ρ is chosen uniformly from C_j, $L(\phi_2\lceil_\rho) \geq 2$. The probability of case 4 is then given by

$$\sum_j Pr[C_j]q_j.$$

If we take any restriction ρ contributing to this event and change the value of ρ at x_{d_j} to ϵ_j then we get a restriction ρ' contributing to case 6. This follows since x_{d_j} does not appear in the reduced ϕ_2. The set of restrictions created at leaf j will be of total probability $q^{-1}Pr[C_j]q_j$ and we seem to be in good shape. However the same restriction ρ' might be created at many leaves and hence we would be over counting if we would just sum these probabilities for various j. However, note that ρ' belongs to A and if it is created at leaf j then it belongs to A_j. Now, since $A_j \times B_j$ form a partition of $A \times B$ we have for any $\rho \in A$

$$\sum_{j|\rho \in A_j} Pr[B_j] = Pr[B] \leq 1.$$

This means that if we multiply the total probability of restrictions created at leaf j by $Pr[B_j]$ we avoid over counting. Thus the sure contribution to the probability of case 6 is

$$\sum_j q^{-1}Pr[C_j]Pr[B_j]q_j.$$

We need to compare this to $\sum_j Pr[C_j]q_j$. For the j for which $Pr[B_j] \geq q$ the term in the sum for case 6 is bigger than the corresponding term for the case 4, while for other j, we use that $Pr[C_j] \leq qPr[B_j] \leq q^2$ and thus summing over those j gives a contribution of at most $q^2 L_1$. We have proved Lemma 7.4. □

Next we turn to Lemma 7.3. This will be more complicated, mainly because the restrictions contributing to case 6 are more difficult to construct.

Proof. (Of Lemma 7.3) Let A_j, B_j and C_j be as in the previous proof. For fixed j let r_j be the conditional probability that $L(\phi_2\lceil_\rho) = 1$ given that $\rho \in C_j$. We divide the leaves into two cases depending on whether $r_j \leq 20qPr[B_j]^{-1}$. If we restrict the summation to those j that satisfy this inequality then

$$\sum_j Pr[C_j]r_j \leq \sum_j qPr[B_j]20qPr[B_j]^{-1} \leq 20q^2 L_1$$

and this gives the first term in the right hand side of Lemma 7.3. We thus concentrate on the case when $r_j \geq 20qPr[B_j]^{-1}$.

Let $A^{2,j}$ and $B^{2,j}$ be the subsets of C_j which reduce ϕ_2 to 0 and 1 respectively. Let ρ be a restriction that belongs to C_j and contributes to case 3. Assume that ρ reduces ϕ_2 to x_d^ϵ. We can obtain a restriction ρ' in $A^{2,j}$ (if $\epsilon = 1$) or $B^{2,j}$ (if $\epsilon = 0$) by setting $\rho'(x_k) = \rho(x_k)$ for $k \neq d$ and $\rho'(x_d) = 1$. To see that $\rho' \in C_j$ note that before we play the KW-game we give all variables given the value $*$ by ρ the value 1. Thus the executions on ρ and ρ' are identical and thus $\rho' \in C_j$. Also, clearly, ρ' forces ϕ_2 to $\bar{\epsilon}$. Now, suppose that

$$\sum_{\rho|\rho' \in A^{2,j}} Pr[\rho] \geq \sum_{\rho|\rho' \in B^{2,j}} Pr[\rho] \tag{2}$$

(the other case being symmetric). Suppose $A^{2,j}$ consists of the restrictions $\rho_1 \ldots \rho_k$. For ρ_i we define a set of *critical variables* and x_k is in this set if

- $\rho_i(x_k) = 1$.
- Creating the restriction ρ_i' by setting $\rho_i'(x_l) = \rho_i(x_l)$ for every $l \neq k$ while $\rho_i'(x_k) = *$ creates a restriction in C and $\phi_2\lceil_{\rho_i'}$ reduces to $\bar{x}_k$.

Note that, as observed above, $\rho' \in C$ in fact implies that $\rho' \in C_j$ since $\rho \in C_j$ and we go from ρ' to ρ by changing the value on a variable from $*$ to 1.

Suppose there are s_i critical variables for ρ_i. By definition, each restriction contributing to the conditional probability r_j gives one critical variable for one $\rho_i \in A^{2,j}$ exactly when ϕ_2 is reduced to a negated variable and otherwise it gives no contribution. By (2) the first case happens in at least half the cases and hence we have

$$r_j = \alpha Pr[C_j]^{-1} \sum_i q s_i Pr[\rho_i],$$

for some α satisfying $1 \leq \alpha \leq 2$. We now create a set of restriction in the following way. We obtain $\binom{s_i}{2}$ new restrictions from ρ_i by choosing two critical variables for ρ_i and for each such choice (s,t) create a restriction $\rho_i^{(s,t)}$ by setting $\rho_i^{(s,t)}(x_k) = \rho_i(x_k)$ for every $k \notin \{s,t\}$ while $\rho_i^{(s,t)}(x_s) = \rho_i^{(s,t)}(x_t) = *$. This way we get a set of restrictions of total probability $q^2\binom{s_i}{2}Pr[\rho_i]$.

Let us relate the total probability of these constructed restrictions to r_j. Note that

$$\begin{aligned}
r_j^2 &= \left(\alpha Pr[C_j]^{-1} \sum_i (q s_i Pr[\rho_i])\right)^2 \\
&\leq \alpha^2 \left(Pr[C_j]^{-1} \sum_i (q^2 s_i^2 Pr[\rho_i])\right) \cdot \left(Pr[C_j]^{-1} \sum_i Pr[\rho_i]\right) \\
&= \alpha^2 Pr[C_j]^{-1} Pr[A^{2,j}|C_j] \sum_i q^2 s_i^2 Pr[\rho_i],
\end{aligned}$$

where the inequality comes from Cauchy-Schwartz inequality. Now

$$\sum_i q^2 \binom{s_i}{2} Pr[\rho_i] = \sum_i \frac{1}{2} q^2 s_i^2 Pr[\rho_i] - \sum_i \frac{1}{2} q^2 s_i Pr[\rho_i]$$

$$\geq \frac{r_j^2 Pr[C_j]}{2\alpha^2 Pr[A^{2,j}|C_j]} - \frac{q}{2\alpha} r_j Pr[C_j]$$
$$\geq (\frac{1}{2\alpha^2} - \frac{1}{40\alpha}) r_j^2 Pr[C_j] \geq \frac{1}{10} r_j^2 Pr[C_j],$$

since $\alpha \leq 2$ and $r_j \geq 20q$.

Remark: Note that the constructed restrictions need not be in C_j. The reason is that there is no control when you change a variable from being fixed to being $*$. In particular, if we were trying to estimate a conditional expectation we would be in deep trouble, since it need not be the case that these recently constructed restrictions satisfy the condition.

Let us look more closely at these obtained restrictions. They give the value $*$ to the variable x_{d_j} since the restriction we started with belonged to C_j. They also give the value $*$ to the two special variables x_s and x_t.

We now change the value at x_{d_j} to ϵ_j in an attempt to force ϕ_1 to 0. Note that this attempt might not always be successful since once x_s and x_t become unassigned ϕ_1 might also depend on those variables (as well as others). We leave this problem for the time being. Let us analyze the set of restrictions created in this way.

At leaf j we have this way created a set of restrictions of total probability at least $q^{-1}\frac{1}{10}Pr[C_j]r_j^2$. However, the same restriction might appear many times and we need to adjust for this fact. Take any restriction ρ created from $\rho_i \in A^{2,j}$. First note that ρ determines the identity of the two special variables x_s and x_t. These are namely the only variables x_k given the value $*$ by ρ with the property that setting $\rho(x_k) = 1$ makes ϕ_2 depend on only one variable. This follows since we recreate a restriction from C_j with the additional property that x_{d_j} is set, but since we are considering cases when ϕ_2 was independent of x_{d_j}, setting a value to x_{d_j} does not matter. To complete the characterization of x_s and x_t, note that after setting any other variable x_k to any value it is still true that ϕ_2 depends on both x_s and x_t.

Let x_s be the variable with lower index of the variables x_s and x_t which we just have identified. Consider the restriction ρ' obtained by setting $\rho'(x_s) = 1$ while $\rho'(x_k) = \rho(x_k)$ for every $x_k \neq x_s$. We claim that ρ' belongs to A_j. Remember that $A_j \times B_j$ was the set of inputs reaching leaf j when playing the KW-game on the formula ϕ_1. To see this claim let ρ'' be obtained by setting $\rho''(x_k) = \rho'(x_k)$ for $x_k \neq x_{d_j}$ while $\rho''(x_{d_j}) = *$. By the conditions for x_t being a critical variable for ρ_i, $\rho'' \in C_j$ and hence $\rho' \in A_j$.

Thus, starting with ρ we have created a unique restriction ρ' such that whenever ρ is created at leaf j then $\rho' \in A_j$. Thus, reasoning as in the proof of Lemma 7.4, if we multiply the probability of the restrictions produced at leaf j by $Pr[B_j]$, then we avoid making an overestimate. This means that we have created a set of restrictions of total probability at least

$$\sum_j \frac{1}{10} q^{-1} Pr[C_j] r_j^2 Pr[B_j].$$

The created restrictions are of two types, either they reduce the formula ϕ_1 to 0 or not. In the former case they contribute to case 6 (since ϕ_2 depends on x_s and x_t), and we have to estimate the probability of the latter case. We claim that in this case the reduced ϕ_1 contains both the variables x_s and x_t. This follows, since setting $\rho(x_s) = 1$ or $\rho(x_t) = 1$ simplifies ϕ_1 to 0 which in its turn is basically the fact

$\rho' \in A_j$ established above. By Lemma 6.1 it follows that the probability of this case is bounded by $q^2 L_1 (\log L_1)^{1/2}$. Summing up we have

$$\begin{aligned} Pr[case\ 3] &= \sum_{j=1}^{L_1} Pr[C_j] r_j = \sum_{j | r_j\ small} Pr[C_j] r_j + \sum_{j | r_j\ large} Pr[C_j] r_j \\ &\leq 20 q^2 L_1 + \frac{1}{20} \sum_j q^{-1} Pr[C_j] r_j^2 Pr[B_j] \\ &\leq 20 q^2 L_1 + \frac{1}{2} \left(Pr[case\ 6] + q^2 L_1 (\log L_1)^{1/2} \right), \end{aligned}$$

where the first inequality uses the bound for r_j and the last inequality is based on the above reasoning. The proof of Lemma 7.3 is complete. $\square$

All that remains is to complete the proof of Lemma 7.2 when $q \geq (4\sqrt{L \log L})^{-1}$. To simplify the calculations we will in this case prove the slightly stronger bound

$$L^2(\phi) \leq 200 q^2 L (\log q^{-1})^{3/2} - 2.$$

First note that when $(4\sqrt{L \log L})^{-1} \leq q \leq (2\sqrt{L \log L})^{-1}$ the second bound follows from the first bound since

$$\begin{aligned} 200 q^2 L (\log q^{-1})^{3/2} &\geq 200 q^2 L (1/2 \log L)^{3/2} \geq 62 q^2 L (\log L)^{3/2} \\ &\geq 30 q^2 L (\log L)^{3/2} + 32 (4\sqrt{L \log L})^{-2} L (\log L)^{3/2} \\ &\geq 30 q^2 L (\log L)^{3/2} + 2. \end{aligned}$$

It remains to establish the second bound when $q \geq (2\sqrt{L \log L})^{-1}$ and we do this by induction over L. Assume that $\phi = \phi_1 \wedge \phi_2$, (the $\vee$-case being similar) where $L(\phi_i) = L_i$, $L_1 + L_2 = L$, and $L_1 \leq L_2$. This implies that we can always use the second bound when bounding $L^2(\phi_2)$. We have two cases depending on whether $q \leq (4\sqrt{L_1 \log L_1})^{-1}$. If $q \geq (4\sqrt{L_1 \log L_1})^{-1}$, then using the induction hypothesis and $L^2(\phi) \leq L^2(\phi_1) + L^2(\phi_2) + 2$ the result follows immediately.

To take care of the other case, notice that our estimates for the corrections to the basic estimates depended only on L_1. This means that in this case we get the total bound

$$L^2(\phi_1) + L^2(\phi_2) + 41 q^2 L_1 + q^2 L_1 (\log L_1)^{1/2}$$

and using the induction hypothesis (the first case for ϕ_1 and the second for ϕ_2) we can bound this by

$$\begin{aligned} & 30 q^2 L_1 (\log L_1)^{3/2} + (200 q^2 L_2 (\log q^{-1})^{3/2} - 2) + q^2 L_1 (41 + (\log L_1)^{1/2}) \\ \leq\ & 60 q^2 L_1 (\log q^{-1})^{3/2} + (200 q^2 L_2 (\log q^{-1})^{3/2} - 2) + 43 q^2 L_1 (\log q^{-1})^{1/2}) \\ \leq\ & 200 q^2 L (\log q^{-1})^{3/2} - 2 \end{aligned}$$

and the proof is complete. $\square$

8. Application to formula size lower bounds. As mentioned in the introduction, it is well known that shrinkage results can be used to derive lower bounds on formula size. Let us just briefly recall the function which seems to be the most appropriate for this purpose. The input bits are of two types. For notational simplicity assume that we have $2n$ input bits and that $\log n$ is an integer that divides n. The first n bits define a Boolean function H on $\log n$ bits. The other n bits are divided into $\log n$ groups of $n/\log n$ bits each. If the parity of the variables in group i is y_i then the output is $H(y)$. We call this function A as it was first defined by Andreev [9].

THEOREM 8.1. *The function A requires formulae of size*

$$\Omega\left(\frac{n^3}{(\log n)^{7/2}(\log\log n)^3}\right).$$

Proof. Assume that we have a formula of size S which describes A. We know ([8], Chap 4, Theorem 3.1) that there is a function of $\log n$ variables which requires a formula size which is

$$\Omega\left(\frac{n}{\log\log n}\right).$$

We fix the first set of values to describe such a function. This might decrease the size of the formula, but it is not clear by how much and hence we just note that the resulting formula is of size at most S. Apply an R_p-restriction with $p = \frac{2\log n \log\log n}{n}$ on the remaining formula. By our main theorem the resulting formula will be of expected size at most $O(Sn^{-2}(\log n)^{7/2}(\log\log n)^2 + 1)$. The probability that all variables in a particular group are fixed is bounded by

$$(1-p)^{\frac{n}{\log n}} \le e^{-\frac{pn}{\log n}} \le (\log n)^{-2}.$$

Since there are only $\log n$ groups, with probability $1 - o(1)$ there remains at least one live variable in each group. Now since a positive random variable is at most twice its expected with probability at least $1/2$, it follows that there is a positive probability that we have at most twice the expected remaining size and some live variable in each group. It follows that

$$O\left(Sn^{-2}(\log n)^{7/2}(\log\log n)^2\right) \ge \Omega\left(\frac{n}{\log\log n}\right).$$

and the proof is complete. $\square$

We might not that there are indeed formulas for the function A of size $O(n^3(\log n)^{-1})$ and hence our bounds are close to optimal.

9. Conclusions. As we see it there remain two interesting questions in shrinking:

- **What is the shrinkage exponent for monotone formulae?** In some sense we have established that it is 2, namely one of the two examples given in the introduction is monotone and shrinks only by a factor p^2. This is the example of $L/2$ copies of $x_1 \wedge x_2$. This is not a natural example and if it is the only one, we are asking the wrong question. We can get around it by using 2-variable and 3-variable simplification rules. We could also ask a slightly different question, namely what is the minimal α such that for arbitrary small p there is a monotone formula of size $O(p^{-\alpha})$ that is trivialized by a restriction from R_p with probability at most $1/2$?
  Apart from its inherent interest, a successful answer to this question would in most cases (depending on the exact form of the answer) lead to an $\omega(n^2)$ lower bound for monotone formulae for the majority function.

- **Are these annoying log factors really needed?** This is really of minor importance. If they were indeed needed it would be surprising.

Acknowledgment. I am most grateful to Sasha Razborov and V.N. Lebedev for catching gaps in the proofs in previous versions of this paper. Sasha Razborov was also together with Andy Yao very helpful in giving ideas in the initial stages of this research. Finally, I am most grateful to two anonymous referees for a careful reading of the manuscript.

REFERENCES

[1] M. Dubiner, U. Zwick, How do read-once formulae shrink? *Combinatorics, Probability & Computing* 3, 455–469 (1994).

[2] M. Furst, J. B. Saxe, and M. Sipser. Parity, circuits and the polynomial time hierarchy. *Math. Syst. Theory*, 17:13–27, 1984.

[3] J. Håstad, A. Razborov, and A. Yao On the Shrinkage Exponent for Read-Once Formulae *Theoretical Computer Science*, Vol 141, 269-282, 1995.

[4] M. Karchmer and A. Wigderson Monotone circuits for connectivity require super-logarithmic depth *SIAM J. on Discrete Mathematics* Vol 3, 1990, pp 255-265.

[5] N. Nisan and R. Impagliazzo. The effect of random restrictions on formulae size. *Random Structures and Algorithms*, Vol. 4, No. 2, 1993, pp 121-134.

[6] M. S. Paterson and U. Zwick. Shrinkage of de Morgan formulae under restriction. *Random Structures and Algorithms*, Vol. 4, No. 2, 1993, pp 135-150.

[7] A. Razborov. Applications of matrix methods to the theory of lower bounds in computational complexity. *Combinatorica*, 10(1):pp 81–93, 1990.

[8] I. Wegener, The Complexity of Boolean function, 1987, John Wiley & Sons Ltd.

Russian references

[9] А. Е. Андреев. О методе получения более чем квадратичных нижних оценок для сложности π-схем. *Вестник МГУ, сер. матем и механ.*, т. 42, в. 1, 1987, стр. 70-73. (A.E. Andreev, On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes, *Moscow Univ. Math. Bull* 42(1)(1987), 63-66).

[10] К.Л. Рычков. Модификация метода В. М. Храпченко и применение ее к оценкам сложности π-схем для кодобых функций– в сб "Методы дискртного анализа б торрии графов и схем" вып. 42 Новосибирск, 1985, стр. 91-98. (K.L. Rychkov,, A modification of Khrapchenko's method and its application to bounding the complexity of Π-networks for coding functions. In: Methods of discrete analysis in the theory of graphs and circuits, Novosibirsk, 1985, pages 91-98.)

[11] Б. А. Субботовская. О реализации линейных функций формулами в базисе $\&, \vee, -$. *ДАН СССР*, т. 136, в. 3, 1961, стр. 553-555. (B.A.Subbotovskaya, Realizations of linear functions by formulas using $+, *, -$, *Soviet Mathematics Doklady* 2(1961), 110-112).

[12] В. М. Храпченко. О сложности реализации линейной функции в классеП-схем. *Матем. заметки*, т. 9, в. 1, 1971, стр. 35-40. (V.M. Khrapchenko, Complexity of the realization of a linear function in the class of π-circuits, *Math. Notes Acad. Sciences USSR* 9(1971), 21-23).

[13] В. М. Храпченко. Об одном методе получения нижних оценок сложности П-схем. *Матем. заметки*, т. 10, в. 1, 1971, стр. 83-92. (V.M. Khrapchenko, A method of determining lower bounds for the complexity of Π-schemes, *Math. Notes Acad. Sciences USSR* 10(1971), 474-479).