Electronic Colloquium on Computational Complexity, Report No. 41 (1998)

Combinatorics of Monotone Computations *

Stasys Jukna[†‡]

Abstract We consider a general model of monotone circuits, which we call *d-local*. In these circuits we allow as gates: (i) arbitrary monotone Boolean functions whose minterms or maxterms (or both) have length $\leqslant d$, and (ii) arbitrary real-valued non-decreasing functions on $\leqslant d$ variables. Our main result is a general combinatorial lower bounds criterion for such circuits. This resolves a problem, raised by Razborov in 1986, and yields, in a uniform and easy way, non-trivial lower bounds for circuits computing explicit functions even when $d \to \infty$. The proof is relatively simple and direct, and combines the bottlenecks counting method of Haken with the idea of finite limit due to Sipser.

We demonstrate the criterion by super-polynomial lower bounds for explicit Boolean functions, associated with bipartite Paley graphs and partial t-designs. We then derive exponential lower bounds for clique-like graph functions of Tardos. Together with an observation made by Rosenbloom (1997), this implies that (unlike in the Boolean case) the power of monotone real and non-monotone Boolean circuits is *incomparable*. Since we allow real gates, the criterion also implies corresponding lower bounds for the length of cutting planes proof in the propositional calculus.

1 Introduction

The question of determining how much economy the universal *non-monotone* basis of And, Or and Not gates provides over the *monotone* basis with only And and Or gates, has been a long standing open problem in Boolean circuit complexity. The breakthrough in the field was made by Razborov in his seminal paper [20], where the first super-polynomial lower bound of size $n^{\Omega(\log n)}$ for the monotone circuit complexity of the clique function was proved. Shortly after, such (and even exponential) lower bounds were obtained for different Boolean functions [21, 2, 1, 26, 27], including those whose non-monotone circuits are polynomial [21, 26].

After this impressing progress one principal question still remained unclear: is there a tractable lower bounds *criterion* for monotone circuits? Razborov raised this problem as a candidate for a "final chord" in that direction (see [22], Problem 4). The point is that the combinatorial parts of all the above mentioned lower bounds proofs depend heavily

*Revised and improved version of ECCC Report TR96–026. Submitted to: *Combinatorica*.

[†]Supported by a DFG grant Me 1077/10-1.

[‡]University of Trier, Dept. of Theoretical Computer Science, D-54286 Trier, Germany & Institute of Mathematics and Informatics, Vilnius, Lithuania. E-mail: jukna@ti.uni-trier.de.

ISSN 1433-8092

on *specific* properties of concrete Boolean functions, and it was unclear if there are some *common* combinatorial properties of Boolean functions that do actually force their hardness.

In this paper we resolve this problem, and do this in quite general setting. We consider the model of *d-local* circuits where as gates we allow *arbitrary* monotone Boolean functions whose minterms or maxterms (or both) have length $\leqslant d$. For $d = O(1)$ any such circuit can be easily simulated by a circuit with fanin-2 And and Or gates with only polynomial blow-up in size: if the original circuit has ℓ gates then each gate can have at most ℓ^d minterms or maxterms, each of which can be computed with $O(\log d)$ And and Or gates. However, this is no more the case if $d \to \infty$. So (at least directly) previously known methods could not handle the case of growing d. To stress the power of such circuits with growing d, take say, the monotone Boolean function CLIQUE(m, k) which, given a graph on m vertices, accepts it iff this graph contains a k-clique. This function is NP-complete but for $d = \binom{k}{2}$, the whole d-local circuit for it consists of just one gate (computing the function itself).

Our main result is the lower bounds criterion (Theorem 2.1) for monotone d-local circuits. In a somewhat restricted form it states the following. Let $f(x_1, \ldots, x_n)$ be monotone Boolean function all of whose minterms and maxterms have length at least k, for some $1 \leqslant k \leqslant n/d$. If f can be computed by a monotone d-local circuit of size ℓ then, for any $1 \leqslant s, r \leqslant k$ there exists an s-CNF C, an r-DNF D and an s-element set $S \subseteq \{1, \ldots, n\}$ such that $|C| \leqslant \ell(dr)^s$, $|D| \leqslant \ell(ds)^r$ and

$$C \leqslant f \wedge \left(\bigwedge_{i \in S} \overline{x}_i \right) \leqslant D.$$

This, in particular, implies (see Theorem 3.2) that a monotone Boolean function cannot be computed by a small monotone circuit if the sets of its its minterms and maxterms contain partial t-designs with appropriate parameters. Despite its generality, the proof of the criterion is relatively simple and direct. It combines Haken and Cook's *bottlenecks counting* approach [14, 15] (which, as shown in [5], is Razborov's approximation argument in disguise) with Sipser's idea of *finite limit* [24, 25]. A vector x is a k-limit for a set of vectors A if on every subset of k coordinates, x coincides with at least one vector from A. If $f(x) = 0$ and x is a k-limit for the set $f^{-1}(1)$ then x is a hard instance for any circuit computing f since the value $f(x)$ cannot be determined when looking at only k bits of x. The key of the whole argument is one simple observation (see Lemma 4.4) relating limits to transversals of set systems. This correspondence implies that no single gate can make too large progress in classifying such instances. If the function f is such that $f^{-1}(0)$ has many k-limits for $f^{-1}(1)$ (and vice versa) then the progress made by the whole circuit must be large, and hence, there must be many gates.

We then prove that the same criterion holds also for d-local *real circuits*, i.e. for circuits with arbitrary non-decreasing *real-valued* functions of fanin $\leqslant d$ as gates. The proof is a slight modification of that for the Boolean case. This is somewhat surprising because, as observed in [23], for some monotone Boolean functions (so-called, *slice functions*) such circuits (even for $d = 2$) are exponentially more powerful than Boolean circuits over the universal basis with And, Or and Not gates.

In Section 3 we demonstrate how the criterion works in concrete situations. We apply the criterion to Paley-type functions and to monotone functions induced by partial t–(n, k, λ) designs (including the Andreev's "drawing polynomials" function). The Paley-

type function PALEY(q,t) has recently been shown to be hard for monotone *span programs* (see [3, 4, 12]), but its monotone circuit complexity was not known. We show that this function is hard also for monotone d-local circuits (Theorem 3.1).

Then we derive a general lower bound for Boolean functions induced by partial t-designs (Theorem 3.2). These functions are particularly interesting because for them the criterion immediately gives large lower bound if the parameters of the design are good enough. When applied to Andreev's "drawing polynomials" function POLY(q,v), this bound extends the (almost optimal) exponential lower bound of Alon and Boppana [1] for this function to the case of general d-local circuits (Theorem 3.3).

As our last example, we consider *clique-like* functions $\mathrm{T}_\varphi(m,k)$ introduced by Tardos in [26]. Here $\varphi(G)$ is a monotone graph function such that $\omega(G) \leqslant \varphi(G) \leqslant \chi(G)$, where $\omega(G)$ is the clique number and $\chi(G)$ is the chromatic number of G. The function $\mathrm{T}_\varphi(m,k)$ accepts a graph G iff $\varphi(G) \geqslant k$. The conceptual advantage of these functions is that some of them can be computed by a non-monotone Boolean circuit of polynomial size. Using our criterion we prove that any clique-like function $\mathrm{T}_\varphi(m,k)$ requires d-local monotone (Boolean and real) circuits of size exponential in $\Omega\left(k^{1/2}/d\right)$ (Theorem 3.4), thus establishing an exponential gap between non-monotone Boolean and monotone real circuits. Together with the above mentioned result of Rosenbloom [23] this implies that (unlike in Boolean case!) the power of non-monotone Boolean and monotone real circuits is incomparable.

Finally, let us mention that the results in the present paper have also an application to *cutting plane proofs* [10] in the propositional calculus. Cutting plane proofs provide a complete refutation system for unsatisfiable sets of propositional clauses. They efficiently simulate resolution proofs, and in fact are known to provide exponentially shorter proofs on some examples (the pigeonhole clauses). Bonet *et al* [8] and Pudlák [19] reduced the problem to lower bounds for circuits with nondecreasing real functions of fanin 2 as gates. Thus, our general lower bound for such circuits (Theorem 2.1), as well as lower bounds for explicit functions, are also lower bounds for the length of cutting plane proofs.

2 The lower bounds criterion

We will consider monotone circuits of rather general form, so let us recall some standard definitions.

Let Ω be a totally ordered sets containing 0 and 1 with $0 < 1$, and $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. A *circuit* for f over domain Ω is a is a sequence $C = (f_1, \ldots, f_\ell)$ of mappings (called *gates*) $f_i : \{0,1\}^n \to \Omega$ where $f_\ell = f$ and each gate has the form $f_i = \phi(h_1, \ldots, h_m)$ where $\phi : \Omega^m \to \Omega$ is some mapping (called the *operation* of that gate) and each h_j is either a variable or one of the previous gates $f_1, \ldots, f_{i-1}$. The *fanin* of the gate is the number m of gates feeding into it. The gate is *monotone* if ϕ is nondecreasing function (with respect to the order of its domain Ω). The total number t of gates is the *size* of C. The function computed by C is the function f_ℓ computed by the last gate. In this paper we will consider only monotone circuits.

In the first part of this paper we will consider Boolean circuits, i.e. circuits over the domain $\Omega = \{0,1\}$, and then we will show how the results extend to monotone circuits over

the reals $\Omega = \mathbb{R}$. The arithmetic structure of real numbers will not be used for the lower bounds, one can take any totally ordered set Ω instead of $\mathbb{R}$.

A *minterm* (*maxterm*) of a monotone Boolean function $f(x_1, \ldots, x_n)$ is a minimal set of variables which, if assigned the value 1 (resp., value 0), force the function to take the value 1 (resp., value 0) regardless of the values assigned to the remaining variables. The *degree* of a monotone Boolean function f is the minimal number d such that either all its minterms or all maxterms (or both) have size $\leqslant d$. For example, unbounded fanin And and Or functions have degree 1. The threshold function $T_k^m(x_1, \ldots, x_m)$ (which outputs 1 iff the input has at least k 1's) has degree $d = \min\{k, m-k\}$.

Let C be a monotone circuit over a domain Ω. We say that C is a *d-local Boolean* circuit if $\Omega = \{0,1\}$ and the operations ϕ of its gates are arbitrary monotone Boolean functions of degree at most d. By a *d-local real* circuit we will mean a circuit over the domain $\Omega = \mathbb{R}$, the operations of whose gates are arbitrary nondecreasing real-valued functions $\phi : \mathbb{R}^m \to \mathbb{R}$ on $m \leqslant d$ variables. By a *d-local monotone circuit* we will mean a circuit which is either Boolean or real d-local circuit.

To state (and to use) the criterion, it will be convenient to switch to the set-theoretic language. Namely, we will identify a vector $v \in \{0,1\}^n$ with the set $S_v = \{i \mid v_i = 1\}$ of its non-zero coordinates, so that every Boolean function f becomes a set-theoretic predicate, which accepts a set $S_v \subseteq \{1, \ldots, n\}$ iff $f(v) = 1$.

A *positive* (resp., *negative*) *indicator* is a subset $S \subseteq \{1, \ldots, n\}$ such that $f(A) = 1$ (resp., $f(\overline{A}) = 0$) for some $A \supseteq S$. (Here and throughout, $\overline{A}$ denotes the complement of A.) Note that one set can be both positive and negative indicator. By a *positive* (resp., *negative*) *input* we will mean a positive (negative) indicator S such that $f(S) = 1$ (resp., $f(\overline{S}) = 0$). Put otherwise, a positive (negative) input is a set of variables which, if assigned the value 1 (resp., value 0), force the function to take the value 1 (resp., value 0) regardless of the values assigned to the remaining variables. In particular, minterms and maxterms are the minimal positive and negative inputs, respectively.

Theorem 2.1 [Criterion] *Let f be a monotone Boolean function on n variables. Suppose that f can be computed by a d-local monotone circuit of size ℓ. Then, for every integers $1 \leqslant s, r \leqslant n/d$ there exist a system of $K \leqslant \ell(dr)^s + 1$ s-element sets $S_0, S_1, \ldots, S_K$ and a system of $L \leqslant \ell(ds)^r$ r-element sets $R_1, \ldots, R_L$ such that at least one of the following two conditions holds:*

1. *Every negative input of size at least s contains at least one of the sets $S_1, \ldots, S_K$.*
2. *Every positive input of size at least r either intersects the set S_0 or contains at least one of the sets $R_1, \ldots, R_L$.*

The same holds also with positive inputs replaced by negative inputs and vice versa.

In Theorem 2.1 we measure the "size" of bit sets $S \subseteq [n] \rightleftharpoons \{1, \ldots, n\}$ just as their cardinality $|S|$. Although in most cases this works well (see Section 3), in some situations (especially, when dealing with Boolean functions defined on graphs) taking other measures may lead to better lower bounds.

We say that a mapping $S \mapsto \mu(S) \in \mathbb{Z}_+$ is a *norm* if it is sub-additive: $\mu(S \cup T) \leqslant \mu(S) + \mu(T)$. Given such a norm, the *μ-size* (or just the *size*, if the norm μ is clear from the context) of a set S is the number $\mu(S)$. The *deviation* of μ is the function $D(t) = \max\{|S| : \mu(S) \leqslant t\}$. The *defect* of μ is the maximal length $c = \max\{\mu(\{i\}) : i \in [n]\}$ of a single bit. Since μ is sub-additive, these two characteristics connect the size $\mu(S)$ of a set S with its cardinality: $\mu(S) \leqslant c \cdot |S|$ and $|S| \leqslant D(\mu(S))$. We say that a bit-set T *respects* a norm μ if we cannot add a bit from outside the set T to any of its subsets without increasing their size, i.e. if $\mu(S \cup \{i\}) \geqslant \mu(S) + 1$ for any subset $S \subseteq T$ and any bit $i \notin T$.

For example, if we take the trivial norm $\mu(S) = |S|$, then $c = 1$, $D(t) = t$ and *every* set respects μ. In case of graphs, bits correspond to edges and one can, for example, take $\mu(S)$ to be the number of vertices incident to at least one edge from S. In this case $c = 2$, $D(t) = \binom{t}{2}$ and only cliques will respect such a norm.

In the criterion below we assume that μ_0, μ_1 is an arbitrary pair of norms with deviations D_0, D_1, and defects c_0 and c_1, respectively. The size of a negative (positive) indicator S means here the number $\mu_0(S)$ (resp., $\mu_1(S)$). By a negative (positive) input we mean a negative (positive) input which respects the norm μ_0 (μ_1).

Theorem 2.2 [Criterion Unabridged] *Let $f(X)$ be a monotone Boolean function. Suppose that f can be computed by a d-local monotone circuit of size ℓ. Then, for every integers $1 \leqslant s \leqslant \mu_0(X)/d$ and $1 \leqslant r \leqslant \mu_1(X)/d$ there exist a negative indicator S_0 of size at most s, a system of $K \leqslant \ell \cdot D_1(drc_1)^s$ negative indicators $S_1, \ldots, S_K$ of size at least s and a system of $L \leqslant \ell \cdot D_0(dsc_0)^r$ positive indicators $R_1, \ldots, R_L$ of size at least r such that at least one of the following two conditions holds:*

1. *Every negative input of size at least s contains at least one of the sets $S_1, \ldots, S_K$.*
2. *Every positive input of size at least r either intersects the set S_0 or contains at least one of the sets $R_1, \ldots, R_L$.*

The same holds also with positive inputs replaced by negative inputs and vice versa.

3 Some explicit lower bounds

To motivate the proof of the criterion, let us first show how it works in concrete situations. In this section we apply the criterion to Paley-type functions, to monotone functions induced by partial t-designs and to "clique-like" functions, some of which have non-monotone Boolean circuits of polynomial size.

3.1 Bipartite Paley graphs

Let q be an odd prime power, congruent to 1 modulo 4. A bipartite Paley graph $G = (V_1, V_2, E)$ is defined on the vertex set $V_1 = V_2 = \mathrm{GF}(q)$, where two vertices $x \in V_1$ and $y \in V_2$ are joined by an edge iff $x - y$ is a square in $\mathrm{GF}(q)$. It is known that this graph is $(q-1)/2$-regular, and has the following "uniform neighbourhood" property: for every two

disjoint sets A, B of vertices in the first part V_1 or in the second part V_2, with $|A|+|B| = k$, $k < (\log q)/4$, the number $\gamma(k)$ of vertices (in the opposite part) adjacent to all vertices in A and nonadjacent to every vertex in B is very close to $q/2^k$, namely $|\gamma(k) - q/2^k| \leqslant k\sqrt{q}$. This property was established in [13, 7] (see also [6, Theorem 10 of Chap. 13] for a short proof) in the case of usual (non-bipartite) Paley graphs; the bound in this case is even better: $|\gamma(k) - q/2^k| \leqslant k\sqrt{q}/2 + k/2$. In the bipartite case one must be more careful because now we have two copies of GF(q), and hence, no of the edges (x, x) is present in the graph. Still, with slight modification, the proof carries over also to bipartite case, (with slightly worse bound $k\sqrt{q}/2 + k$, which is still $\leqslant k\sqrt{q}$ for $q \geqslant 5$); an analysis for the bipartite case is given, for example, in [3, 4].

Define PALEY(q, t) to be the function of $n = 2q$ Boolean variables representing the vertices in $V_1 \cup V_2$, which accepts a set of vertices iff this set contains some t-element subset $A \subseteq V_1$ together with the set of its common neighbours $\Gamma(A) \rightleftharpoons \{y \in V_2 \mid (x, y) \in E \text{ for all } x \in A\}$. Thus, minterms of PALEY(q, t) are all the sets of the form $A \cup \Gamma(A)$, with $A \subseteq V_1$, $|A| = t$. To define negative inputs, let $\widehat{\Gamma}(A)$ denote the set of all common non-neighbours, i.e. $\widehat{\Gamma}(A) \rightleftharpoons \{y \in V_2 \mid (x, y) \in E \text{ for no } x \in A\}$. If $t \leqslant (\log q)/6$ and $q \geqslant 5$ then, by the above mentioned uniform neighbourhood property of Paley graphs, we have that $\gamma(2t) > 0$, and hence, for any pair of disjoint t-element sets $A, B \subseteq V_1$ there must be a vertex $y \in V_2$ which is a common neighbour of all the vertices in A and is isolated from all the vertices in B. Thus, $A \cap B = \emptyset$ iff $\Gamma(A) \cap \widehat{\Gamma}(B) \neq \emptyset$. This implies that sets of the form $B \cup \widehat{\Gamma}(B)$ with $B \subseteq V_1$ and $|B| = t$, intersect all the minterms and contain no of them, and hence, are negative inputs for PALEY(q, t).

For $t = \Theta(\log q)$ the minterms of PALEY(q, t) have size at most $t + \gamma(t) \leqslant q$. Hence, the function can be computed by a trivial monotone circuit using $q\binom{q}{t} \leqslant n^{O(\log n)}$ fanin-2 And and Or gates. Is this bound optimal? It was recently proved in [3, 4, 12] that in the case of, so–called, monotone span programs we cannot do better: any such program for PALEY(q, t) with $t = \Theta(\log q)$, requires size $n^{\Omega(\log n)}$. We prove that the same holds also in the case of monotone (Boolean and real) d-local circuits.

Theorem 3.1 *Let $q \equiv 1 (\mathrm{mod} 4)$ be a sufficiently large odd prime power, $\Omega(\log q) = t < (\log q)/8$ and $\log d = o(\log q)$. Then any d-local monotone circuit computing* PALEY(q, t) *has size at least $n^{\Omega(\log n)}$.*

Proof. Let ℓ be the size of a minimal d-local monotone circuit computing PALEY(q, t). According to Theorem 2.1 we have only two possibilities, depending on what of the two items in its conclusion holds. For this proof we take $s = r \rightleftharpoons t$.

Suppose that Item 1 holds. Since sets of the form $B \cup \widehat{\Gamma}(B)$ with $|B| = t$, are negative inputs, each of them must contain at least one of $K \leqslant \ell(dr)^s$ s-element sets $S_1, \ldots, S_K$. Now, if $B_i \cup \widehat{\Gamma}(B_i)$, $i = 1, \ldots, m$ are the negative inputs containing a fixed s-element set S, then $\bigcap_{i=1}^m B_i \supseteq S \cap V_1$ and $\bigcup_{i=1}^m B_i \subseteq \widehat{\Gamma}(S \cap V_2)$. If less than half of the vertices of S lie in V_2 then $|S \cap V_1| \geqslant s/2$, and hence, $m \leqslant \binom{q-s/2}{t-s/2} \leqslant q^{t/2}$. Otherwise, $|S \cap V_2| \geqslant s/2$ and, by the above mentioned universality property of Paley graphs, $m \leqslant \binom{\gamma(s/2)}{t} \leqslant \gamma(t/2)^t \leqslant q^{t(1-\varepsilon)}$ for some absolute constant $0 < \varepsilon \leqslant 1/16$. Thus, in both cases one s-element set S can be contained in at most $q^{t(1-\varepsilon)}$ negative inputs. Since the total number of such inputs is $\binom{q}{t} \geqslant (q/t)^t$, we conclude that in this case $\ell \geqslant K \cdot (dr)^{-s} \geqslant \binom{q}{t} q^{-t(1-\varepsilon)} (dt)^{-t} \geqslant \left(\frac{q^\varepsilon}{dt^2}\right)^t \geqslant n^{\Omega(\log n)}$.

Suppose now that Item 2 of Theorem 2.1 holds. That is, every positive input (i.e. a set of the form $A \cup \Gamma(A)$ with $|A| = t$) either intersects a fixed s-element set S_0 or contains at least one of $L \leqslant \ell(ds)^r$ r-element sets $R_1, \ldots, R_L$. To estimate the number of positive inputs avoiding the set S_0, let $A_i \cup \Gamma(A_i)$, $i = 1, \ldots, m$ be the inputs containing a fixed vertex $x \in S_0$. If $x \in V_1$ then clearly, $m \leqslant \binom{q-1}{t-1} = \frac{t}{q}\binom{q}{t}$. If $x \in V_2$ then $|\bigcup_{i=1}^m A_i|$ cannot exceed the degree $(q-1)/2$, and hence, $m \leqslant \binom{(q-1)/2}{t} = \binom{q-b}{t}$, with $b = (q+1)/2$. Since $\binom{q-b}{t} / \binom{q}{t} \leqslant e^{-(t/q)b}$, we have that in both cases at least $1 - s \cdot e^{-t/2} \geqslant 1 - o(1)$ fraction of all $\binom{q}{t}$ positive inputs avoid the set S_0, and hence, must contain at least one of the sets $R_1, \ldots, R_L$. The same argument as for negative inputs yields that one r-element set $R \subseteq V_1 \cup V_2$ can be contained in at most $\binom{\gamma(r/2)}{t}$ of positive inputs. Hence, in this case we also have that $\ell \geqslant n^{\Omega(\log n)}$ ■

3.2 Partial t–designs

A *partial* t–(n,k,λ) *design* (called also a *covering design*) is a family $\mathcal{D}$ of k-element subsets (called blocks) of an n-element set X such that every t-element set is contained in *at most* λ blocks of $\mathcal{D}$. Every such design $\mathcal{D}$ induces the monotone Boolean function $f_{\mathcal{D}}$ on n variables, which accepts a subset $A \subseteq X$ iff A contains at least one block of $\mathcal{D}$. Say that a partial t–(n,k,λ) design $\mathcal{D}$ is *good* if $2 \leqslant \ln|\mathcal{D}| \leqslant k/(6dt)$ and $k \leqslant (n \ln|\mathcal{D}|)^{1/2}$.

Theorem 3.2 *Let $\mathcal{D}$ be a partial t–(n,k,λ) design. If $\mathcal{D}$ is good then, for every r, $1 \leqslant r \leqslant k$, any d-local monotone circuit computing $f_{\mathcal{D}}$ has size at least* $\min\left\{|\mathcal{D}|/\lambda \cdot (dr)^t,\ \varepsilon \cdot 3^r\right\}$, *where $\varepsilon \geqslant 0.03$.*

Proof. We are going to apply Theorem 2.1 with negative inputs replaced by positive inputs (and vice versa). Let ℓ be the minimum size of a d-local monotone circuit computing $f_{\mathcal{D}}$. Blocks of the design $\mathcal{D}$ are positive inputs for $f_{\mathcal{D}}$. Thus, if the first item of Theorem 2.1 holds, then every block must contain at least one of $K \leqslant \ell(dr)^t$ t-element subsets of X, and hence, in this case $\ell \geqslant |\mathcal{D}|/\lambda \cdot (dr)^t$.

Suppose now that the second item of Theorem 2.1 holds. Take a random subset $\mathbf{A}$ of X where each element of X is included in $\mathbf{A}$ independently with probability $p \rightleftharpoons (2\ln|\mathcal{D}|)/k$. Let E be the event that the set $\mathbf{A}$ is a negative input for $f_{\mathcal{D}}$, has size at least r and avoids the set S_0. The set $\mathbf{A}$ is *not* a negative input with probability $\mathrm{Prob}\left[f_{\mathcal{D}}(\overline{\mathbf{A}}) = 1\right] \leqslant |\mathcal{D}|\,(1-p)^k \leqslant |\mathcal{D}|\,e^{-pk} \leqslant |\mathcal{D}|^{-1} \leqslant e^{-2}$, by the choice of p. The probability that $\mathbf{A}$ intersects a fixed t-element set S_0 does not exceed $tp \leqslant \frac{1}{3d}$. Finally, the number $|\mathbf{A}|$ is binomially distributed random variable with expectation pn, and hence, $\mathrm{Prob}[|\mathbf{A}| \geqslant pn/2] \geqslant \frac{1}{2}$. Since $\mathcal{D}$ is good, the conditions on the block size k implies that $r \leqslant k \leqslant pn/2$. Therefore, $\mathrm{Prob}\left[\overline{E}\right] \leqslant e^{-2} + \frac{1}{2} + \frac{1}{3} \leqslant 1 - \varepsilon$. By Item 2 of Theorem 2.1, with probability at least ε, the set $\mathbf{A}$ must contain at least one of the r-element sets $R_1, \ldots, R_L$, where $L \leqslant (dt)^r$. Since $\mathrm{Prob}[\mathbf{A} \supseteq R_i] = p^r$ and $\ln|\mathcal{D}| \leqslant k/(6dt)$, we conclude that in this case $\ell \geqslant \varepsilon(dtp)^{-r} = \varepsilon\left(\frac{k}{2dt\ln|\mathcal{D}|}\right)^r \geqslant \varepsilon 3^r$. ■

3.2.1 The design of polynomials

Unlike (usual) t-designs, partial designs with good parameters can be obtained quite easily. As an example, consider the following partial designs $\mathcal{D}_v$. Let q be a prime power and consider the square $X = \mathrm{GF}(q) \times \mathrm{GF}(q)$. If $p(z)$ is a polynomial over $\mathrm{GF}(q)$, then its *graph* is the set of q points $(a, p(a))$ in this square, with $a \in \mathrm{GF}(q)$. Blocks of $\mathcal{D}_v$ are graphs of polynomials over $\mathrm{GF}(q)$ of degree at most $v-1$. For every $1 \leqslant t < v$, this is a partial t–(n,k,λ) design with $n = q^2$, $k = q$ and $\lambda = q^{v-t}$; the number of blocks in this design is $|\mathcal{D}_v| = q^v$.

The corresponding monotone Boolean function $f_{\mathcal{D}_v}$, denoted also $\mathrm{POLY}(q,v)$, was investigated by Andreev [2] who proved that any circuit with fanin-2 And and Or gates computing this function (for appropriate values of v) requires size exponential in $\Omega(n^{1/8-\varepsilon})$. Using Razborov's method of approximations, Alon and Boppana [1] were able to essentially improve this bound until $q^{\Omega(v)}$ for any $v \leqslant (q/\ln q)^{1/2}/2$; for maximal possible v, the bound is exponential in $\Omega(n^{1/4})$. This bound is almost optimal because q^{v+1} is the trivial upper bound for $\mathrm{POLY}(q,v)$ (this function is an Or of q^v monomials, each of length q). Thus, in the case of fanin-2 Boolean gates we have quite sharp bounds for this function. Using Theorem 3.2 we extend this bound to arbitrary monotone d-local (Boolean and real) circuits.

Theorem 3.3 *Let q be a prime power, $q \geqslant 8$. If $1 \leqslant d \leqslant v \leqslant (q/6\ln q)^{1/2}$, then any monotone d-local circuit computing* $\mathrm{POLY}(q,v)$ *has size at least $q^{\Omega(v/d)}$.*

Proof. Take $t \rightleftharpoons \lceil v/d \rceil$ and $r \rightleftharpoons \lceil v \ln q/d \rceil$. Since $\ln|\mathcal{D}_v| = v\ln q \leqslant q/(6v) \leqslant k/(6dt)$, the design $\mathcal{D}_v$ is good (for this value of t). Since $|\mathcal{D}|/\lambda \cdot (dr)^t \geqslant \left(\frac{q}{dr}\right)^t \geqslant q^{\Omega(v/d)}$ and $\varepsilon 3^r \geqslant q^{\Omega(v/d)}$, the desired lower bound follows directly from Theorem 3.2. ■

3.3 Clique-like problems

A graph function $\varphi(G)$ is *clique-like* if $\omega(G) \leqslant \varphi(G) \leqslant \chi(G)$, where $\omega(G)$ is the clique number, i.e. the size of a maximal clique in X, and $\chi(G)$ is the chromatic number. For $2 \leqslant k < m$, let $\mathrm{T}_\varphi(m,k)$ denote the monotone Boolean function of $n = \binom{m}{2}$ boolean variables encoding the edges of a graph on m vertices, which outputs 1 iff $\varphi(G) \geqslant k$. This function is monotone if the underlying graph function φ is such. For $\varphi = \omega$ this is the well-known clique function $\mathrm{CLIQUE}(m,k)$ deciding whether a given graph contains a k-clique.

Although we always have that $\omega(G) \leqslant \chi(G)$, the gap between these two quantities can be quite large: results of Erdös [11] imply that the maximum of $\chi(G)/\omega(G)$ over all m-vertex graphs G has the order $\Theta\left(m/(\log m)^2\right)$. So, at least potentially, the class of clique-like functions $\mathrm{T}_\varphi(m,k)$ is large enough. And indeed, Tardos [26] observed that this class includes not only **NP**-complete problems (like the clique function) but also some problems from **P**. Using Lovász-capacity of graphs, introduced in [18], she defined an explicit clique-like graph function φ, which is monotone and is computable in polynomial time. Hence, the corresponding monotone Boolean functions $\mathrm{T}_\varphi(m,k)$ can be computed by (non-monotone) circuits, with And, Or and Not gates, of polynomial size. On the other hand, the improvement of Razborov's lower bound for the clique function given by Alon and Boppana [1] implies that without Not gates, the function $\mathrm{T}_\varphi(m,k)$ requires size exponential

in $\Omega\left(k^{1/2}\right)$. This has demonstrated that in the Boolean case the gap between monotone and non-monotone complexity is exponential. Super-polynomial gap was previously shown by Razborov in [21] using the perfect matching function.

What about monotone circuits with *real* gates? The question is not trivial because, as we already mentioned in the introduction, there exist monotone Boolean functions, whose non-monotone Boolean circuit size is exponential, and which can be computed by monotone real circuits of polynomial size [23]. It appears that (unlike in the Boolean case) monotone real circuits and non-monotone Boolean circuits are, in fact, *incomparable*: clique-like functions (including the Tardos' function) remain hard even if we allow non-decreasing real-valued functions as gates:

Theorem 3.4 *Let φ be a monotone clique-like function and $3 \leqslant k \leqslant 2m^{1/2}$. Then any d-local monotone circuit computing $\mathrm{T}_\varphi(m,k)$ requires size exponential in $\Omega\left(k^{1/2}/d\right)$.*

Proof. Let ℓ be the minimum size of a d-local monotone circuit computing $\mathrm{T}_\varphi(m,k)$. To apply Theorem 2.2 must first choose positive and negative inputs for $\mathrm{T}_\varphi(m,k)$. With positive inputs the situation is clear: every k-clique G is a positive input because $\varphi(G) \geqslant \omega(G) = k$. To define the negative inputs, we assign each vertex x a color $h(x)$ from the set $\{1,\ldots,k-1\}$, and then put edges between those pairs of vertices with the same color. (Two colorings can lead to the same graph but we consider them as different for counting purposes.) The complement $\overline{G}_h$ of each such graph is a complete l-partite graph, for some $l \leqslant k-1$, and hence, must be rejected by $\mathrm{T}_\varphi(m,k)$ because $\varphi(\overline{G}_h) \leqslant \chi(\overline{G}_h) \leqslant k-1$.

Next, we have to fix a pair of norms μ_0 and μ_1. For positive indicators S, we take $\mu_1(S) \rightleftharpoons v(S)$, where $v(S)$ is the number of vertices incident to at least one edge from S. It is clear that this norm is sub-additive and that every clique respects it. For negative inputs this is no more true, because, for example, $v(G_h \cup \{e\}) = v(G_h)$ if the ends of the edge e belong to different parts of G_h. But in this case, the graph $G_h \cup \{e\}$ has one connected component fewer. This suggests the following norm for negative indicators: take $\mu_0(S) \rightleftharpoons v(S) - \kappa(S)$, where $\kappa(S)$ is the number of connected components in S (such a measure for graphs was already used implicitly in [14] and explicitely in [5]). The sub-additivity of μ_0 can be shown by an easy induction on the number of edges, using the fact that $\mu_0(S \cup \{e\}) = \mu_0(S)$ if the edge e connects two vertices in one connected component of S, and $\mu_0(S \cup \{e\}) = \mu_0(S) + 1 = \mu_0(S) + \mu_0(\{e\})$, otherwise. By the same reason, each negative input $G_h = (V,E)$ respects the norm μ_0, because if $S \subseteq E$ and $e \notin E$, then e cannot connect two vertices in the same connected component of S, and hence, $\mu_0(S \cup \{e\}) = \mu_0(S) + 1$. The defect and the deviation for these norms are: $c_0 = 1$, $c_1 = 2$, $D_0(t) \leqslant t^2$ and $D_1(t) = \binom{t}{2} \leqslant t^2$.

For the rest of the proof we take (with foresee) $s \rightleftharpoons \lceil (m/(2d^2k))^{1/2} \rceil$ and $r \rightleftharpoons \lceil ((k-1)/(8d^2))^{1/2} \rceil$. Our goal is to show that $\ell \geqslant 2^{\Omega(r)}$.

Suppose the first item of Theorem 2.2 holds. Each negative input G_h consists of t_1 isolated vertices and t_2 mutually disjoint cliques, where $1 \leqslant t_1 + t_2 \leqslant k-1$. Thus, $v(G_h) = m - t_1 \geqslant m - k + 1$ and $\kappa(G_h) = t_2 \leqslant k-1$, which implies that $\mu_0(G_h) \geqslant m - 2k + 2 \geqslant s$. Since the graphs G_h respect the norm μ_0, we have, by Item 1, that each of these graphs must contain at least one of the sets of edges $S_1,\ldots,S_K$, where $\mu_0(S_i) \geqslant s$ and $K \leqslant \ell \cdot D_1(drc_1)^s = \ell \cdot \binom{2dr}{2}^s \leqslant \ell \cdot (2dr)^{2s}$. We have $(k-1)^m$ colourings h, and it

remains to estimate for how many of them, the induced graph G_h can contain a fixed set of edges S, with $\mu_0(S) \geqslant s$. If $V_1, \ldots, V_t$ are the sets of vertices of the connected components of S, then by the definition of the norm μ_0, $|V_1| + \ldots + |V_t| \geqslant s + t$. If $G_h \supseteq S$, then all the vertices in each of the classes V_i must get the same colour. Hence, the number of colourings h, for which $G_h \supseteq S$, does not exceed $(k-1)^t \cdot (k-1)^{m-(s+t)} = (k-1)^{m-s}$. Thus, in this case, $\ell \geqslant (k-1)^s/(2dr)^{2s} = \left(\frac{k-1}{4d^2r^2}\right)^s \geqslant 2^s$, which is at least $2^{\Omega(r)}$, as long as $k \leqslant 2m^{1/2}$.

Suppose now that Item 2 of Theorem 2.2 holds. Positive inputs are k-cliques. At least $\binom{m}{k} - s^2\binom{m-2}{k-2} \geqslant \frac{1}{2}\binom{m}{k}$ of such cliques must avoid a fixed set S_0 of $D_0(s) \leqslant s^2$ edges. By Item 1, each of these k-cliques must contain at least one of $L \leqslant \ell \cdot D_0(drc_0)^r \leqslant \ell \cdot (ds)^{2r}$ r-cliques $R_1, \ldots, R_L$. Since each R_i is contained in $\binom{m-r}{k-r}$ of k-cliques, we conclude that in this case $\ell \geqslant \frac{1}{2}\left(\frac{m}{d^2s^2k}\right)^r \geqslant 2^{\Omega(r)}$. ■

4 Proof of the criterion (Boolean case)

Throughout this and the next section, let $f = f(x_1, \ldots, x_n)$ be an arbitrary (but fixed) monotone Boolean function, and $1 \leqslant s, r \leqslant n/d$ be an arbitrary (but fixed) parameters.

4.1 Witnesses and finite limits

The combinatorial part of our proof is based on the following simple properties of finite limits. Let A be a set of vectors and u be a vector in $\{0,1\}^n$. A *witness* of u *against* A is a subset $S \subseteq \{1, \ldots, n\}$ of coordinates such that every vector $v \in A$ differs from u in at least one coordinate in S. If u does not belong to A then it has at least one witness S against this set. We say that such a witness is *legal* if $S \subseteq I(u)$ where $I(u) \rightleftharpoons \{i \mid u_i = f(u)\}$. If $u \in f^{-1}(\varepsilon)$ and $A \subseteq f^{-1}(\varepsilon \oplus 1)$ then, due to monotonicity of f, the whole set $I(u)$ is a legal witness of u against A. We will be interested in the minimal possible size of its subset with this property.

Definition 4.1 A *k-limit* for a set A is a vector u such that $|S| \geqslant k+1$ for any legal witness S of u against A.

Lemma 4.2 *If $A \subseteq B$ and u is a k-limit for A then u is also a k-limit for B. If $A = A_1 \cup \ldots \cup A_d$ and u is a k-limit for the whole set A then u is a $\lfloor k/d \rfloor$-limit for at least one of the sets $A_1, \ldots, A_d$.*

Proof. The first claim is obvious. For the second claim, observe that if u would have a (legal) witness S_i of size $\lfloor k/d \rfloor$ against A_i, for all $i = 1, \ldots, d$, then $S = S_1 \cup \ldots \cup S_d$ would be a (legal) witness of u against the whole set A, and this witness would have size at most k. ■

Definition 4.3 A set $A \subseteq f^{-1}(\varepsilon)$ is *(a,b)-closed* if there exists a set $\emptyset \neq B \subseteq f^{-1}(\varepsilon \oplus 1)$ such that: (i) *every* vector of A is an a-limit for B, and (ii) *no* vector of B is a b-limit for A.

A family of sets is *a-uniform* if each its member has exactly a elements. For a vector u and a set $S \subseteq \{1,\ldots,n\}$ of its coordinates, we will write $u(S) \equiv \varepsilon$ if $u_i = \varepsilon$ for all $i \in S$.

Lemma 4.4 *If $A \subseteq f^{-1}(\varepsilon)$ is (a,b)-closed then there exists an a-uniform family $\mathcal{F}$ such that $|\mathcal{F}| \leqslant b^a$ and, for every vector $u \in A$, there is an $F \in \mathcal{F}$ for which $u(F) \equiv \varepsilon$.*

For the proof of this lemma we need the following simple property of transversals. By an *a-critical transversal* for a sequence $S_1,\ldots,S_m$ of sets we will mean a set T of size $\geqslant a+1$ for which there is an index $1 \leqslant i \leqslant m$ such that T intersects all the sets $S_1,\ldots,S_i$ but no its subset of size $\leqslant a$ does this.

Lemma 4.5 *Let $S_1,\ldots,S_m$ be a sequence of sets, each of cardinality at most b, and let $\mathcal{T}$ be a family of its a-critical transversals. Then, for every $1 \leqslant k \leqslant a$, there exists a k-uniform family $\mathcal{F}_k$ such that $|\mathcal{F}_k| \leqslant b^k$ and every $T \in \mathcal{T}$ contains at least one $F \in \mathcal{F}_k$.*

Proof. Construct the desired family $\mathcal{F}_k$ by induction on k. For $k=1$ we can take as $\mathcal{F}_1$ the family of all single element sets $\{x\}$ with $x \in S_1$. Suppose now that $\mathcal{F}_{k-1}$ is already constructed. We may assume w.l.o.g. that every member of $\mathcal{F}_{k-1}$ is contained in at least one transversal from $\mathcal{T}$ (if not, we just omit the redundant sets from $\mathcal{F}_{k-1}$). For each set $F \in \mathcal{F}_{k-1}$ choose the first index i such that $F \cap S_i = \emptyset$; such an i exists since $|F| = k-1 < a$ and F is a subset of an a-critical transversal. Put in $\mathcal{F}_k$ all the sets $F \cup \{x\}$ with $x \in S_i$. Each of these sets has $|F \cup \{x\}| = |F|+1 = k$ elements, and $|\mathcal{F}_k| \leqslant |S_i| \cdot |\mathcal{F}_{k-1}| \leqslant b \cdot b^{k-1} = b^k$, as desired.■

Proof of Lemma 4.4. Let $\emptyset \neq B \subseteq f^{-1}(\varepsilon \oplus 1)$ be a set from Definition 4.3. By Item (ii), every vector from B has a legal witness of size $\leqslant b$ against the set A. That is, for every $v \in B$, there must be a set $S = S_v$ of at most b coordinates such that $v(S) \equiv \varepsilon \oplus 1$ but $u(S) \not\equiv \varepsilon \oplus 1$ for all $u \in A$. This, in particular, means that for every $u \in A$, the set $I(u) \rightleftharpoons \{i \mid u_i = \varepsilon\}$ intersects all the sets in the family $\mathcal{S} = \{S_v \mid v \in B\}$. By Item (i), we have that $|I(u)| \geqslant a+1$, because otherwise the set $I(u)$ would be a legal witness of vector u against B, and hence, u could not be an a-limit for B. Thus, sets $I(u)$ with $u \in A$, are a-critical transversals for the family $\mathcal{S}$. By Lemma 4.5, there must be a family $\mathcal{F}$ which consist of at most b^a a-element sets and has the following property: for every $u \in A$ there is an $S \in \mathcal{F}$ such that $I(u) \supseteq S$. Since clearly, $u(S) \equiv \varepsilon$ (because u is constant ε on $I(u)$), $\mathcal{F}$ is the desired family.■

Now we turn to the actual proof of Theorem 2.1 in the case of Boolean gates.

If $S \subseteq \{1,\ldots,n\}$ is a positive (negative) input, then the corresponding positive (negative) *input vector* is the vector $v \in \{0,1\}^n$ such that $v_i = 1$ (resp., $v_i = 0$) iff $i \in S$; the *size* of such an input v is the cardinality of S. Hence, the size of a positive input vector is the number of 1's, and the size of a negative input vector is the number of 0's in it. Let $U^0 \subseteq f^{-1}(0)$ be the set of all negative input vectors of size at least s, and $U^1 \subseteq f^{-1}(1)$ be the set of all positive input vectors of size at least r.

Let $C = (f_1,\ldots,f_\ell)$ be a monotone d-local Boolean circuit, and suppose that C computes f, i.e. that $f_\ell = f$. To estimate the size of (i.e. the total number ℓ of gates in)

C, we will follow the "bottlenecks counting" frame suggested by Haken [14]. Every gate makes some "progress" towards separating inputs in $f^{-1}(0)$ from those in $f^{-1}(1)$. The idea now is to send an input vector to the first gate in the circuit for which this input was a "really hard" instance, i.e. at which certain amount of progress in classifying the input is made. The measure of progress is the size of a witness, which intuitively keeps track of how many bits of the input are actually used by the computation at that gate. Dividing an underestimate of the size of the mapped set by an overestimate of how many of vectors can be mapped to one gate, yields the lower bound on the total number of gates.

To capture the progress, made by one gate, let us associate with every gate f_i the set $U_i^0 \times U_i^1 \rightleftharpoons \{(u,v) \in U^0 \times U^1 \mid f_i(u) = 0 \text{ and } f_i(v) = 1\}$ of all those pairs, which the gate separates correctly (just like the function f does). If some vector $u \in U_i^\varepsilon$ is a k-limit for the set of all its neighbours $U_i^{\varepsilon \oplus 1}$, then we can treat u as a "hard instance" for the gate f_i, because this gate correctly separates u from all its neighbours, even though this requires knowledge of more than k bits. Formally, we define the hardness of input vectors by exploring the gates $f_1, \ldots, f_\ell$ one-by-one, as follows.

Initially no input is hard. Suppose we already know what input vectors are hard for the first $i-1$ gates $f_1, \ldots, f_{i-1}$, i.e. that for every $j = 1, \ldots, i-1$ we already know the sets $H_j^\varepsilon \rightleftharpoons \{u \in U_i^\varepsilon \mid \ u \text{ is hard for } f_j\}$, $\varepsilon = 0, 1$. To define what inputs from $U_i \rightleftharpoons U_i^0 \cup U_i^1$ are hard for the i-th gate $f_i = \phi(h_1, \ldots, h_m)$, we consider two cases, depending on what of the terms (minterms or maxterms) of its operation ϕ are short.

All minterms of ϕ have length $\leqslant d$. In this case we first consider the left part U_i^0 and declare an input $u \in U_i^0$ as being *hard* for the i-th gate if this input was hard for no of the previous gates, and is an s-limit for the set of its "easy" neighbours $E_i^1 \rightleftharpoons U_i^1 \setminus \left(H_1^1 \cup \cdots \cup H_{i-1}^1\right)$ Having defined the set H_i^0, we turn to the right part U_i^1. Namely, we say that an input $v \in U_i^1$ is *hard* for the i-th gate if it was hard for no of the previous gates, and is an r-limit for the set $E_i^0 \rightleftharpoons U_i^0 \setminus \left(H_1^0 \cup \cdots \cup H_{i-1}^0 \cup H_i^0\right)$.

Some minterms of ϕ have length $> d$. In this case maxterms must be short, and we define the hardness dually: we start with right part U_i^1, and interchange the parameters s and r.

Let $E^0 \subseteq U^0$ and $E^1 \subseteq U^1$ be the sets of inputs which were hard for *no* of the gates $f_1, \ldots, f_\ell$. Then $U^0 \subseteq E^0 \cup H_1^0 \cup \cdots H_\ell^0$ and $U^1 \subseteq E^1 \cup H_1^1 \cup \cdots H_\ell^1$. By Lemma 4.4, Theorem 2.1 follows directly from the following two claims.

Claim 4.6 *Either $E^0 = \emptyset$ or there exists an s-element subset S_0 of $\{1, \ldots, n\}$ such that $u(S_0) \not\equiv 0$ for all vectors $u \in E^1$.*

Claim 4.7 *For every $i = 1, \ldots, t$ the set H_i^0 is (s, dr)-closed and H_i^1 is (r, ds)-closed*

Proof of Claim 4.6. If $E^0 = \emptyset$, there is nothing to do. Suppose therefore that $E^0 \neq \emptyset$ and take an arbitrary vector $u \in E^0$. By the definition of E^0, this vector u was hard for *no* of the gates, and in particular, was not hard for the last gate f_ℓ. Since our circuit computes the function f we have that $f_\ell = f$, and hence (by the definition of hardness), vector u has a legal witness S_0 of size $|S_0| \leqslant s$ against the set $f^{-1}(1) \setminus \left(H_1^1 \cup \cdots H_{\ell-1}^1\right)$. Since E^1 is a subset of this set, S_0 is a legal witness of u also against E^1. Since $f(u) = 0$, the legality of

S_0 means that $u(S_0) \equiv 0$, and hence, $v(S_0) \not\equiv 0$ for all $u \in E^1$, thus completing the proof of the claim. ■

Proof of Claim 4.7. We will prove this claim only for sets H_i^0 (for sets H_i^1 the argument is dual).

If the i-th gate f_i is one of the variables x_l $(1 \leqslant l \leqslant n)$, then all the pairs (u, v) in the corresponding set $U_i^0 \times U_i^1$ have the property that $u_l = 0$ and $v_l = 1$. Therefore, the one-element set $S = \{l\}$ is a legal witness of all the vectors correctly separated by this gate, and hence, no of these vectors could be hard for it (because $r, s \geqslant 1$). Thus, if the i-th gate is a variable then $H_i^0 = \emptyset$.

Otherwise, this gate has the form $f_i = \phi(h_1, \ldots, h_m)$ where ϕ is a monotone Boolean function, all whose minterms or maxterms (or both) have length at most d. We consider these two cases separately.

Case 1: All minterms of ϕ have length $\leqslant d$. We claim that in this case the set H_i^0 is (s, dr)-closed. Indeed, in this case, every input from H_i^0 is an s-limit for the set $E_i^1 = U_i^1 \setminus \left(H_1^1 \cup \cdots \cup H_{i-1}^1\right)$, by the definition. It remains therefore to show that no input from E_i^1 can be a (dr)-limit for the set H_i^0. Suppose the opposite, and let $v \in E_i^1$ be such a limit for H_i^0. Since $f_i(v) = \phi(h_1(v), \ldots, h_m(v)) = 1$ and the minterms of ϕ have length at most d, there must be a collection of $\leqslant d$ gates $h_{i_1}, \ldots, h_{i_d}$ (feeding into f_i) all of whom output 1 on v, but on every input u from U_i^0 (and hence, from H_i^0), at least one of these gates outputs 0. So, if the input v would be a (dr)-limit for the whole set H_i^0 then, by Lemma 4.2, it would be also an r-limit for the set $A \rightleftharpoons \{u \in H_i^0 \mid f_j(u) = 0\}$, where $f_j \in \{h_{i_1}, \ldots, h_{i_d}\}$ is one of the gates feeding into f_i. Since $A \subseteq H_i^0$ and $j < i$, we have that $A \subseteq E_j^0$ and hence, v would be an r-limit for E_j^0. But this means that this input would already be hard for at least one of the previous gates $f_1, \ldots, f_{i-1}$ and hence, $v \notin E_i^1$, a contradiction.

Case 2: Some minterms of ϕ have length $> d$. We claim that in this case the set H_i^0 is (s, r)-closed. Indeed, in this case we have that $E_i^1 = U_i^1 \setminus \left(H_1^1 \cup \cdots \cup H_i^1\right)$ and $H_i^0 \subseteq E_i^0 = U_i^1 \setminus \left(H_1^0 \cup \cdots \cup H_{i-1}^0\right)$. Now, every input from H_i^0 is an s-limit for E_i^1, by the definition. On the other hand, no input from E_i^1 can be an r-limit for H_i^0, because otherwise it would be an r-limit also for the set E_i^0, and hence, would belong to H_i^1, a contradiction.

This completes the proof of Claim 4.7, and thus, the proof of the Criterion in the case of unbounded fanin Boolean gates. ■

5 Proof of the criterion (real case)

For the case of real-valued gates we use the following "skew version" of closeness.

Definition 5.1 A set $A = \{u_1, \ldots, u_m\} \subseteq f^{-1}(\varepsilon)$ is *weakly (a, b)-closed* if there exists a sequence of sets $\emptyset \neq B_1 \subseteq \ldots \subseteq B_m \subseteq f^{-1}(\varepsilon \oplus 1)$ such that, for every $k = 1, \ldots, m$,

(i) input u_k is an a-limit for B_k, and

(ii) *no* input from B_k is a b-limit for the set $A_k = \{u_k, \ldots, u_m\}$.

Lemma 5.2 *If $A \subseteq f^{-1}(\varepsilon)$ is weakly (a,b)-closed then there exist an a-uniform family $\mathcal{F}$ such that $|\mathcal{F}| \leqslant b^a$ and for every vector $u \in A$ there is an $F \in \mathcal{F}$ such that $u(F) \equiv \varepsilon$.*

Proof. Similar to that of Lemma 4.4. By (ii), every input from B_k has a legal witness of length at most b against the set $A_k \rightleftharpoons \{u_k, \ldots, u_m\}$. That is, for every input $v \in B_k$ there is a subset of bits $S_{k,v} \subseteq I(v) \rightleftharpoons \{i \mid v_i = \varepsilon \oplus 1\}$ such that $|S_{k,v}| \leqslant b$ and every input $u \in A_k$ takes the value $v_i \oplus 1 = \varepsilon$ on at least one bit $i \in S_{k,v}$. This, in particular, means that for every $u \in A_k$, the set $I(u)$ intersects all the sets in the sequence $\mathcal{S}_k = \{S_{k,v} : v \in B_i\}$ (with sets $S_{k,v}$ arranged in arbitrary order). Now, for each $j = 1, \ldots, m$ the input u_j belongs to all the sets $A_1, \ldots, A_j$, and hence, the set $I(u_j)$ must intersect all the sets in the sequence $\mathcal{S}^j = \{\mathcal{S}_1, \ldots, \mathcal{S}_j\}$. On the other hand, by (i), no a-element subset of $I(u_j)$ can do this, since any such subset would be a legal witness of u_j against B_j. Thus, for every $j = 1, \ldots, m$, the set $I(u_j)$ is an a-critical transversal for the sequence $\mathcal{S}^j$, and hence, is such a transversal for the whole sequence $\mathcal{S}^m$. By Lemma 4.5, there must be a family $\mathcal{F}$ which consist of at most b^a a-element sets such that, for every $u \in A$, the set $I(u)$ contains at least one member of $\mathcal{F}$. Since u is constant ε on the whole set $I(u)$, $\mathcal{F}$ is the desired family. ■

Now we turn to the actual proof of Theorem 2.1 in the case of real gates.

Let $C = (f_1, \ldots, f_\ell)$ be a circuit with arbitrary non-decreasing real-valued functions as gates, and suppose that C computes f, i.e. that $f_\ell = f$. Let $U^0 \subseteq f^{-1}(0)$ be the set of all negative input vectors of size at least s, and $U^1 \subseteq f^{-1}(1)$ be the set of all positive input vectors of size at least r. To capture the progress made by one gate, we associate with every gate f_i the bipartite graph $G_i \rightleftharpoons \left\{(u,v) \in U^0 \times U^1 \mid f_i(u) < f_i(v)\right\}$. We define the hardness of input vectors by exploring the gates $f_1, \ldots, f_\ell$ one-by-one, as follows. For a vector $u \in U^0 \cup U^1$, let $G_i(u)$ denote the set of all its neighbours in the i-th graph G_i.

Initially no input is hard. Suppose we already know what input vectors are hard for the first $i-1$ gates $f_1, \ldots, f_{i-1}$, i.e. that for every $j = 1, \ldots, i-1$ we already know the sets $H_j^\varepsilon \rightleftharpoons \{u \in U_i^\varepsilon \mid \ u \text{ is hard for } f_j\}$, $\varepsilon = 0, 1$. We say that an input vector $u \in U_i^\varepsilon$ is *hard* for the i-th gate f_i is u was hard for no of the previous gates and is a k-limit for the set

$$G_i^*(u) \rightleftharpoons G_i(u) \setminus \left(H_1^{\varepsilon\oplus 1} \cup \ldots \cup H_{i-1}^{\varepsilon\oplus 1}\right)$$

of all its "easy" neighbours in the i-th graph G_i. If $E^0 \subseteq U^0$ and $E^1 \subseteq U^1$ denote the sets of inputs which were hard for *no* of the gates $f_1, \ldots, f_\ell$, Then $U^0 \subseteq E^0 \cup H_1^0 \cup \cdots H_\ell^0$ and $U^1 \subseteq E^1 \cup H_1^1 \cup \cdots H_\ell^1$.

Since Claim 4.6 holds also in this case, it remains, by Lemma 5.2, to prove the analogy of Claim 4.7.

Claim 5.3 *For every $i = 1, \ldots, t$ the set H_i^0 is weakly (s, dr)-closed and H_i^1 is weakly (r, ds)-closed*

To prove this, we will employ the following specific properties of graphs $G_1, \ldots, G_\ell$.

Property 5.4 *If $f_i = \phi(f_{i_1}, \ldots, f_{i_d})$ then $G_i \subseteq G_{i_1} \cup \ldots \cup G_{i_d}$.*

Indeed, if some edge (u,v) appears in *no* of the graphs $G_{i_1},\ldots,G_{i_d}$ then $f_{i_j}(u) \geqslant f_{i_j}(v)$ for all $j = 1,\ldots,d$. Since ϕ is nondecreasing, this implies that $f_i(u) \geqslant f_i(v)$, i.e. that $(u,v) \notin G_i$, as desired.

Property 5.5 *For every $i = 1,\ldots,t$ and $\varepsilon \in \{0,1\}$, it is possible to order the vectors $u_1,\ldots,u_p$ from U^ε so that $G_i(u_1) \subseteq G_i(u_2) \subseteq \ldots \subseteq G_i(u_p)$.*

To see this, arrange the inputs $U^0 = \{u_1,\ldots,u_p\}$ and $U^1 = \{v_1,\ldots,v_p\}$ so that $f_i(u_1) \geqslant f_i(u_2) \geqslant \ldots \geqslant f_i(u_p)$ and $f_i(v_1) \leqslant f_i(v_2) \leqslant \ldots \leqslant f_i(v_p)$.

Proof of Claim 5.3. We will prove the claim only for sets H_i^0 (for sets H_i^1 the argument is dual). If the i-th gate f_i is one of the variables x_l $(1 \leqslant l \leqslant n)$, then $H_i^0 = \emptyset$ as before. So, assume that $f_i = \phi(f_{i_1},\ldots,f_{i_d})$. By Property 5.5 we can order the vectors $u_1,\ldots,u_m$ from H_i^0 so that $G_i^*(u_1) \subseteq G_i^*(u_2) \subseteq \ldots \subseteq G_i^*(u_m)$. We are going to show that both the items (i) and (ii) of Definition 5.1 hold with with $a \rightleftharpoons s$, $b \rightleftharpoons dr$, $A \rightleftharpoons H_i^0$ and $B_k \rightleftharpoons G_i^*(u_k)$ for $k = 1,\ldots,u_m$.

Item (i) holds by the definition of H_i^0. To verify the second item (ii), suppose the opposite that some input $v \in B_k$ is a (dr)-limit for the set $A_k \rightleftharpoons \{u_k,u_{k+1},\ldots,u_m\}$. Since $B_k = G_i^*(u_k)$, the vector v is an easy neighbour of u_k, and since $G_i^*(u_k) \subseteq \ldots \subseteq G_i^*(u_m)$, v is also an easy neighbour of all the vectors in A_k. Thus, A_k is a subset of $G_i^*(v)$, and if v would be a (dr)-limit for the set A_k, it would be also such a limit for $G_i^*(v)$. But by Property 5.4, $G_i^*(v) \subseteq \bigcup_{j=1}^d G_{i_j}^*(v)$, and by Lemma 4.2, vector v should be an r-limit for at least one of the sets $G_{i_j}^*(v)$, which is impossible since then v would be already hard for some previous gate.

This completes the proof of Claim 4.7, and thus, the proof of Theorem 2.1 in the case of real gates. ■

6 Proof sketch for Theorem 2.2

The proof of Theorem 2.2 is similar to that of Theorem 2.1 using more general notion of finite limit (depending this time on the norm).

Definition 6.1 Let μ be a norm and c be its defect. A *k-limit* for a set A *under* μ is an input u such that $\mu(S) > k \cdot c$ for any legal witness S of u against A.

Since norms are sub-additive, Lemma 4.2 remains true also for this notion of finite limit. Thus, the only place, where the possible deviation between the norm $\mu(S)$ and the cardinality $|S|$ needs more care, is the lemma about transversals (Lemma 4.5). Using the estimates $\mu(S)/c \leqslant |S| \leqslant D(\mu(S))$ connecting the norm of sets with their cardinality, one can easily modify the proof of this lemma to the case of arbitrary norms.

Let μ be a norm and c be its defect. An *a-critical transversal* for a sequence of sets $S_1,\ldots,S_m$ under the norm μ is a set T, which respects μ and for which there is an index i such that T intersects all the sets $S_1,\ldots,S_i$ but no its subset $T' \subseteq T$ with $\mu(T') \leqslant a \cdot c$, does this.

Lemma 6.2 *Let $S_1, \ldots, S_m$ be a sequence of sets, each of cardinality at most b, and let $\mathcal{T}$ be a family of its transversals which are a-critical under some norm μ. Then, for every $1 \leqslant k \leqslant a$, there exists a k-uniform family $\mathcal{F}_k$ such that:* (i) $|\mathcal{F}_k| \leqslant b^k$, (ii) $k \leqslant \mu(F) \leqslant k \cdot c$ *for all $F \in \mathcal{F}_k$, and* (iii) *every set from $\mathcal{T}$ contains at least one set $F \in \mathcal{F}_k$.*

Proof. As in the proof of Lemma 4.5, we will construct the desired family $\mathcal{F}_k$ by induction on k. For $k = 1$ we can choose the first set S_i such that $\mu(\{x\}) \neq 0$ for all $x \in S_i$, and take as $\mathcal{F}_1$ the family of all one element sets $\{x\}$ with $x \in S_i$. This family has at most $|S_i| \leqslant b$ sets, each of which has size (under μ) at most c, as desired. Suppose now that the family $\mathcal{F}_{k-1}$ is already constructed. For a set of bits F, let $\mathrm{ext}(F)$ denote the set of all transversals in $\mathcal{T}$ containing F. We can assume w.l.o.g. that $\mathrm{ext}(F) \neq \emptyset$ for every set F in $\mathcal{F}_{k-1}$ (if not, just remove such sets). We construct the family $\mathcal{F}_k$ by applying the following procedure to the family $\mathcal{F}_{k-1}$.

Take a set F in $\mathcal{F}_{k-1}$ and choose the first index i such that $F \cap S_i = \emptyset$ but $T \cap S_i \neq \emptyset$ for all $T \in \mathrm{ext}(F)$; such an i exists since $\mu(F) \leqslant (k-1)c < ac$ and F is a subset of an a-critical under μ transversal. There are two possibilities: either there is some bit $x \in S_i$ for which $\mu(F \cup \{x\}) = \mu(F)$, or not. In the first case replace the set F in $\mathcal{F}_{k-1}$ by $F \cup \{x\}$. Since $\mu(F \cup \{x\}) = \mu(F)$ and all the transversals in $\mathrm{ext}(F)$ respect the norm μ, we have that $\mathrm{ext}(F \cup \{x\}) = \mathrm{ext}(F)$. Hence, no transversal gets lost during this step, and we can repeat the procedure with the new family. In the second case include in $\mathcal{F}_k$ all the sets $F \cup \{x\}$ with $x \in S_i$, remove F from $\mathcal{F}_{k-1}$ and repeat the procedure with this smaller family $\mathcal{F}_{k-1} \setminus \{F\}$. Since $\mu(F \cup \{x\}) \leqslant \mu(F) + c \leqslant (k-1)c + c \leqslant ac$, no a-critical transversal gets lost also during this step, since every transversal containing F, must contain at least one of the sets $F \cup \{x\}$ with $x \in S_i$. Moreover, we have that $\mu(F \cup \{x\}) \geqslant \mu(F) + 1 \geqslant (k-1) + 1 = k$, as desired. Since every set in $\mathcal{F}_{k-1}$ produces at most $|S_i| \leqslant b$ new sets, the resulting family $\mathcal{F}_k$ will have at most $b \cdot |\mathcal{F}_{k-1}| \leqslant b^k$ sets, and we are done. ■

With this lemma instead of Lemma 4.5, the rest of the proof is the same as in the case of norms $\mu_0(S) = \mu_1(S) = |S|$.

Acknowledgements

I thank Anna Gál, Armin Haken and Avi Wigderson for interesting discussions on the topic of this paper. My special thanks to Sasha Razborov for his attention to my (partially successful, cf. [16]) attempts to derive similar criterion ten years ago, and to Mike Sipser for telling me the notion of finite limits while visiting Vilnius in 1991.

References

[1] N. Alon and R. Boppana, The monotone circuit complexity of Boolean functions, *Combinatorica,* 7:1 (1987), 1–22.

[2] A. E. Andreev, On a method for obtaining lower bounds for the complexity of individual monotone functions, *Doklady Akademii Nauk SSSR,* 282:5 (1985), pp. 1033-1037. English translation in: *Soviet Mathematics Doklady,* 31 (1985), 530–534.

[3] L. Babai, A. Gál, J. Kollár, L. Rónyai, T. Szabó and A. Wigderson, Extremal bipartite graphs and superpolynomial lower bounds for monotone span programs. In: *Proc. 28th ACM STOC* (1996), 603–611.

[4] L. Babai, A. Gál, and A. Wigderson, Superpolynomial lower bounds for monotone span programs, to appear in: *Combinatorica*.

[5] C. Berg, S. Ulfberg, Symmetric approximation argument for monotone lower bounds without sunflowers, to appear in: *Computational Complexity*.

[6] B. Bollobás, *Random graphs*, Academic Press, New-York, 1985.

[7] B. Bollobás and A. Thomason, Graphs which contain all small graphs, *European J. Combin.* **2** (1981), 13–15.

[8] M. Bonet, T. Pitassi, and R. Raz, Lower bounds for cutting planes proofs with small coefficients. In: *Proc. Twenty-seventh Ann. ACM Symp. Theor. Comput.*, (1995), 575–584.

[9] R. B. Boppana and M. Sipser, The complexity of finite functions. In *Handbook of Theoretical Computer Science*, Vol. A, *Algorithms and Complexity*, J. van Leeuwen, Ed., MIT Press (1990), 757–804.

[10] W. Cook, C. R. Coullard, and Gy. Turán, On the complexity of cutting plane proofs. *Disc. Appl. Math.*, (1987), 25–38.

[11] P. Erdös, Some remarks on chromatic graphs, *Colloquium Mathematicum* **16** (1967), 253–256.

[12] A. Gál, A characterization of span program size and improved lower bounds for monotone span programs. In: *Proc. 30th ACM STOC* (1998), 429–437.

[13] R.L. Graham and J. Spencer, A constructive solution to a tournament problem, *Canad. Math. Bull.*, **14** (1971), 45–48.

[14] A. Haken, Counting Bottlenecks to Show Monotone P$\neq$NP, In *Proc. of the 36th Ann. IEEE Symp. Found. Comput. Sci.*, (1995), 36–40.

[15] A. Haken and S. Cook, An exponential lower bound for the size of monotone real circuits. Submitted to: *J. Comput. System Sci.*, 1996.

[16] S. Jukna, Monotone circuits and local computations. In: *Proc. of 31th Conf. of Lithuanian Math. Society* (Kaunas, June 28-29, 1990), 100–101.

[17] S. Jukna, Finite limits and monotone computations: the lower bounds criterion. In: *Proc. of the 12th Ann. IEEE Conf. on Comput. Complexity*, (1997), 302-313.

[18] L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. on Information Theory*, **25** (1979), 1–7.

[19] P. Pudlák Lower bounds for resolution and cutting planes proofs and monotone computations, *Journal of Symbolic Logic*, **62**:3 (1997), 981–998.

[20] A. A. Razborov, Lower bounds on the monotone complexity of some Boolean functions, *Doklady Akademii Nauk SSSR,* 281:4 (1985), pp. 798-801. English translation in: *Soviet Mathematics Doklady,* 31 (1985), 354–357.

[21] A. A. Razborov, A lower bound on the monotone network complexity of the logical permanent, *Matematicheskie Zametki*, 37:6 (1985) pp. 887-990 (in Russian); English translation in: Math. Notes Acad. of Sci. USSR, 37:6 (1985), 485–493.

[22] A. A. Razborov, Lower bounds on the monotone complexity of Boolean functions. In: *Proc. of Int. Congress of Mathematicians* (Berkeley, California, USA,1986), 1987, 1478–1487.

[23] A. Rosenbloom, Monotone circuits are more powerful than monotone boolean circuits, *Inform. Process. Letters*, **61**:3 (1997), 161–164.

[24] M. Sipser, A topological view of some problems in complexity theory. In: *Colloquia Mathematica Societatis János Bolyai* **44** (1985), 387–391.

[25] M. Sipser, *Personal communication*, 1991.

[26] É. Tardos, The gap between monotone and non-monotone circuit complexity is exponential, *Combinatorica*, 7:4 (1987), 141–142

[27] A. C. Yao, Circuits and local computations. In: *Proc. 21th Ann. ACM Symp. Theor. Comput.* (1989), 186–196.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'