Polynomial Vicinity Circuits and Nonlinear Lower Bounds

Kenneth W. Regan[*]
State University of New York at Buffalo

Abstract

We study families of Boolean circuits with the property that the number of gates at distance t fanning into or out of any given gate in a circuit is bounded above by a polynomial in t of some degree k. We prove that such circuits require size $\Omega(n^{1+1/k}/\log n)$ to compute several natural families of functions, including sorting, finite field arithmetic, and the "rigid linear transformations" of Valiant [26]. Our proof develops a "separator theorem" in the style of Lipton and Tarjan [14] for a new class of graphs, and our methods may have independent graph-theoretic interest.

1. Introduction

Nonlinear lower bounds for natural problems, whether for machine-based or circuit models, have been frustratingly hard to obtain. Indeed, there is currently no example of a function $f : \{0,1\}^n \rightarrow \{0,1\}^n$ whose graph belongs to NP or to E that is known to require Boolean circuits of size more than $4n$.

Faced with this situation, it is natural to seek other conditions on the circuits or machines under which nonlinear lower bounds can be proved. Valiant [26] added the condition that the circuits have logarithmic depth as well as linear size. He defined a class of "highly rigid" linear transformations on $\{0,1\}^n$ (regarded as the n-dimensional vector space over GF(2)), and proved that log-depth circuits require size $\Omega(n \operatorname{loglog} n/\log \operatorname{loglog} n)$ to compute them.

We introduce a different condition that is both limiting and natural. Define a family $C_1, C_2, C_3, \ldots$ of Boolean circuits to have *polynomial vicinity* if there is a polynomial p such that for any gate g in any C_n, and all $t > 0$, the number of gates connected to g by a path of length at most t is at most $p(t)$. For example, circuits whose gates are the nodes of a d-dimensional mesh have vicinity $O(t^d)$. For contrast, a circuit in the form of a full binary tree does not have polynomial vicinity, since $\Theta(2^t)$ gates are within distance t of the root. Indeed, circuits with an output gate that depends on all n input gates must have $n^{\Omega(1)}$ depth if they have polynomial vicinity. Note that the bound $p(t)$ is independent of the size or number of inputs n of the circuits C_n.

The main motivation for this condition is a practically-minded one. Let us regard computing elements, be they solid-state gates or optical nodes or etc., as having some discrete uniform minimum size. Then in 3-dimensional space, the number of elements one can "pack" within distance t of a given element g is $O(t^3)$. If time is measured in units of how long it takes a signal from one element to reach a neighboring one, then only $O(t^3)$ elements can affect the computation of g over the next t time units. Thus cubic vicinity is a property of those networks we can physically build, at least under "physics as we know it." Although real solid-state components are being shrunk to degrees barely imagined ten years ago, the vicinity condition still dominates as $n \rightarrow \infty$, i.e., for the kind of asymptotic bounds that complexity theory is concerned with.

Similar arguments have been made by prominent researchers in related contexts. Schorr [23] argued that the constraints of embeddings in physical space prevent the realization of (poly)logarithmic running times for parallel algorithms; i.e., that PRAM and NC-theory-based running times are too optimistic. This argument was carried further by Vitányi [27], down to fine details of wire thickness and heat concentration and other limits imposed by "the laws of nature." Feldman and Shapiro [8] gave a less-formal version of the above definition of "polynomial vicinity" for a machine model, and presented a 3-dimensional model for which the best possible parallel speedup is n-to-$n^{1/4}$ steps. Kruskal, Rudolph, and Snir [13] advocated the study of such "polynomial speedups" for other reasons.

A formal machine-based definition of "polynomial vicinity" that generalizes Feldman and Shapiro's, and also applies to oracle machines, was introduced in the conference paper [21]. A motivation given there is that whereas unrelativized Turing machines have linear vicinity, all known constructions of oracles A such that $\mathrm{P}^A = \mathrm{NP}^A$ use exponential vicinity. Thus "vicinity" may be the right kind of quantitative concept to get beyond the notorious ob-

[*] Supported in part by National Science Foundation Grant CCR-9409104. Author's current address: Computer Science Department, 226 Bell Hall, Box # 602000, UB North Campus, Buffalo, NY 14260-2000. Email: regan@cs.buffalo.edu

stacles posed by relativization results.

The circuit/graph form of the concept treated here is just as quantitative and couldn't be simpler to state. It also makes no geometrical assumptions about how a graph is represented in finite-dimensional space. Hence our final motivation is to see to what degree lower bounds that have been proved in VLSI or mesh-based models can be carried through in a more-general setting.

The main result of this paper is a *sub-linear graph separator theorem* for polynomial-vicinity (PV) graphs.

Theorem 1.1 (informal statement) *Let $[G_n]$ be a family of graphs of size $s(n)$ and vicinity $O(t^k)$ independent of n. Then for any disjoint vertex subsets A, B of size m in G_n, there is a set S of $O(s(n)\log m/m^{1/k})$ vertices whose removal disconnects almost half the vertices in A from half the vertices in B.*

The set S is the separator. When $m, s(n) = \Theta(n)$, S has sub-linear size $O(n^{1-1/k}\log n)$. We will think of A as the set of input nodes, and B as the set of output nodes, of a circuit computing a function on $\{0,1\}^n$. Valiant [26] observed that circuits computing certain "highly rigid" linear transformations on length-n vectors cannot have separators of size less than n. We extend this observation to functions having a certain "property of randomness" defined by Mansour, Nisan, and Tiwari [17], including sorting, the function $(a, b, c) \mapsto a \cdot b + c$ in finite fields, and any (other) family of universal hash functions. Separators for these functions cannot have size less than $n/2 - 1$. With $m = \Theta(n)$ in these cases, we obtain our main application:

Theorem 1.2 *Circuits of vicinity $O(t^k)$ that sort, or that compute rigid linear transformations, or compute $a \cdot b + c$ in finite fields, or do universal hashing on inputs of size n, must have size $\Omega(n^{k+1}/\log n)$.*

This is a fairly strong nonlinear size lower bound.

These results also show a *size-vicinity tradeoff* that complements known time-space tradeoffs for these functions. Pippenger and Fischer [20] proved that time-t Turing machines can be simulated by Boolean circuits of size $O(t \log t)$. Hence any function computed in quasilinear time *qlin* $= n \cdot (\log n)^{O(1)}$ has circuits of *qlin* size. By our results, however, such circuits for sorting and $a \cdot b + c$ cannot have polynomial vicinity. If our main theorem can be improved in respects discussed in Section 5, it may give a sense in which the older $O(t^2)$-size circuit simulation by Savage [22], which gives quadratic vicinity, is optimal.

The main stem of computer-science interest in graph-separator theorems were two papers by Lipton and Tarjan [14, 15]. They showed that for any planar graph G of size N, and any weighting function wt on the vertex set V (i.e., $\sum_v wt(v) = 1$, $wt(v) \geq 0$ for all v), there is a set $S \subseteq V$ of size less than $2.83n^{1/2}$ whose removal breaks G into two disconnected pieces, each of total weight between $1/3$ and $2/3$. This improved Ungar's theorem [25] giving such an S of size $O(n^{1/2}\log n)$. They also gave an efficient algorithm to find S, and gave applications. The universal quantification over A and B in our Theorem 1.1 is roughly similar to theirs over the wt function, and enables some of the same applications. Note also that we have an "extra" log factor like Ungar's. Whether this can be taken out, and other improvements made, is also discussed in Section 5.

Separator theorems have since been obtained for several other classes of graphs: graphs with a planar representation having $O(n)$ crossings [12], graphs of finite genus [10], graphs of bounded tree-width or with excluded clique minors [2, 3], "d-local" graphs (meaning graphs embedded in d-space so that the ratio of the length of the longest edge in a minimum spanning tree to that of the shortest edge is at most d) [28], and graphs defined by intersections of spheres around points in d-space [18, 19]. The class of graphs of vicinity $O(t^d)$ (for some $d > 0$) is incomparable with each of these. One other connection to note is that our graphs have relatively high (namely, $n^{\Omega(1)}$) diameter, and this property is related both to the second-highest eigenvalue of various matrices associated to the graph and to having small separators in papers by Alon [1], Chung [7], and Spielman and Teng [24]. We do not, however, see how to get our particular result from these connections. Our proof uses just *Menger's Theorem* and a means of obtaining large independent sets in PV bipartite graphs.

2 Low-Vicinity Graphs

Given a subset A of the vertex set V of an undirected graph $G = (V, E)$, let $S(A) = \{ v \in V : (\exists u \in E)(u, v) \in A \}$, and let $N(A) = A \cup S(A)$. Here one calls $S(A)$ the *boundary* of A and $N(A)$ the *neighborhood* of A. Now for $t \geq 2$ inductively define $S^t(A) = S(N^{t-1}(A))$ and $N^t(A) = N(N^{t-1}(A))$. When A has just one vertex v we write $N^t(v)$ for $N^t(\{ v \})$ and so on.

Definition 2.1. (a) A graph G has *vicinity* $f(t)$ if for all vertices v in G, $|N^t(v)| \leq f(t)$.

(b) A family $\mathcal{G}$ of graphs has *vicinity* $f(t)$ if every graph in $\mathcal{G}$ has vicinity at most $f(t)$. The graphs in $\mathcal{G}$ have *polynomial vicinity*, and are *PV* graphs, if they have vicinity $t^{O(1)}$.

Here we mean that $f(t)$ gives an upper bound on the "vicinity function" of G, which could be defined by $v_G(t) = \sup_{v \in V} |N^t(v)|$. This looser usage suffices in this paper. This concept has been studied in only one context that we know, namely where G is a (possibly infinite) *Cayley graph* of a group, and is tied to questions about "polynomial growth" of infinite groups (cf. [16, 4]). We can generalize (b) to bounds $f(t,n)$ that depend on the size or index n of graphs $G_n \in \mathcal{G}$, and will indeed do so in the proof of our main theorem. However, the notion of vicinity is intended to be local and independent of n.

In this paper we focus on vicinity bounds of the form at^k for some fixed (not necessarily integral) constants $a \geq 3$ and $k \geq 1$. The idea is that k is an abstract notion of the dimension of the graph, as exemplified by (rectangular or simplicial or etc.) k-dimensional mesh graphs having vicinity $\Theta(t^k)$. However, the full binary tree is an example of a planar graph of vicinity $\Theta(2^t)$, while not all graphs of quadratic or even linear vicinity are planar. In fact, one can convert any graph H and create a topologically similar graph of nearly linear vicinity by replacing every vertex v in H by a "ring" of degree-3 vertices, and then subdividing every edge with a huge number of degree-2 vertices (exponentially many in case H had exponential vicinity). Thus low vicinity is not a *minor-invariant* property of graphs. We do not know whether graphs of vicinity at^k must have "nice" embeddings in k-dimensional space of the kind studied in [18, 19, 28], and personal communications from some authors of these papers have turned up no reasons why this should be so.

One evident property of PV graphs is that they have relatively large diameter, in fact, *min-diameter*. The diameter of a finite graph G is the maximum distance $d(u,v)$ between two vertices u and v. The *min*-diameter equals $\min_{u \in V} \max_{v \in V} d(u,v)$. An n-vertex graph of vicinity at^k has min-diameter at least $n^{1/k}/a$. This already implies that PV graphs are not good *expanders*, but there is a much more striking sense in which PV graphs are the antithesis of expanding graphs. A typical definition for an n-vertex graph to be an expander is that for some $c > 0$ and all $A \subset V$ of size at most (say) $n/2$, $|S(A)| \geq c|A|$. We show that for PV graphs $G = (V,E)$, for all $c > 0$ and $v \in V$, there is a relatively small value of t such that with $A = N^t(v)$, $|S(A)| < c|A|$.

Lemma 2.1 *Let G be a single graph of vicinity at^k, $k \geq 1$. Let c be such that $0 < c \leq \sqrt{2} - 1$, and let $v \in V$. Finally let*

$$b = (2k \log_{1+c} 2) \cdot \log_2(k a_0^{1/k} \log_{1+c} 2), \qquad (1)$$

where $a_0 = \max\{\, a, 2\,\}$. Then for some t, $1 \leq t \leq b$, $|S^t(v)| \leq c|N^{t-1}(v)|$.

Proof. Suppose not; i.e., that for all t, $1 \leq t \leq b$, $|S^t(v)| > c|N^{t-1}(v)|$. Then $|N^b(v)| > (1+c)^b$. Since G has vicinity at^k, this would imply $(1+c)^b < ab^k$. Hence to reach the contradiction that proves the lemma, we need only show that in fact $(1+c)^b \geq ab^k$.

For $k = 1$, we have $b = 2(\log_{1+c} 2)\log_2(a_0 \log_{1+c} 2)$, and need to show that $(1+c)^b \geq ab$. Since $a_0 \geq a$ it suffices to show $(1+c)^b \geq a_0 b$. Then

$$(1+c)^b = 2^{2\log_2(a_0 \log_{1+c} 2)} = (a_0 \log_{1+c} 2)^2$$

and

$$a_0 b = 2a_0(\log_{1+c} 2)\log_2(a_0 \log_{1+c} 2).$$

Upon cancelling $a_0(\log_{1+c} 2)$ from both sides, it suffices to show that

$$(a_0 \log_{1+c} 2) \geq 2\log_2(a_0 \log_{1+c} 2);$$

i.e., that $x \geq 2\log_2 x$ with $x = a_0 \log_{1+c} 2$. This is true so long as $x \geq 4$, and the conditions $a_0 \geq 2$ and $c \leq \sqrt{2} - 1$ bring this about.

For $k > 1$, let $a' = a_0^{1/k}$, and let $c' = (1+c)^{1/k} - 1$, so that $(1+c) = (1+c')^k$. Then

$$\begin{aligned} (1+c)^b \geq a_0 b^k \quad &\Longleftrightarrow \quad (1+c')^{kb} \geq (a'b)^k \\ &\Longleftrightarrow \quad (1+c')^b \geq a'b. \end{aligned}$$

Since $c' \leq \sqrt{2} - 1$ also holds, the desired value of b follows from the case $k = 1$ by substituting a' for a_0 and c' for c in the formula $b = 2(\log_{1+c} 2)\log_2(a_0 \log_{1+c} 2)$. Since $\log_{1+c'} 2 = k \log_{1+c} 2$, this gives (1). □

The main point needed for later results is that in all cases $b = O(\log a)$. We will use this in cases where "a" is not a constant but depends on the size m of certain bipartite graphs, and where m itself may depend on the input-length parameter n. The dependence on c is less important, because c will be fixed. Other tradeoffs between c and a in a bound for b are possible; the general question involves the study of so-called *Lambert functions*, and is left to the interested reader. We emphasize that this counting bound applies concretely to a single graph, not just asymptotically for a family of graphs. We modify the argument to show:

Lemma 2.2 *Let a, c, k and b be as in Lemma 2.1, and let $b' = 2b + 4k\log_{1+c} 2$. Then there exists t, $1 \leq t \leq b'$, such that $|S^t(v)| \leq c|N^{t-1}(v)|$ and $|S^{t+1}(v)| \leq c|N^t(v)|$.*

Proof. If not, then $|N^{b'}(v)| > (1+c)^{b'/2}$. But $|N^{b'}(v)|$ must be at most $ab'^k = (a2^k)(b'/2)^k$. Hence we get the desired conclusion if we show that $(1+c)^{b'/2} \geq (a2^k)(b'/2)^k$. Substituting $a2^k$ for a_0 in (1) tells us that this happens when

$$\begin{aligned} (b'/2) \quad &= \quad (2k \log_{1+c} 2) \cdot \log_2(2k a^{1/k} \log_{1+c} 2) \\ &= \quad b + (2k \log_{1+c} 2). \end{aligned}$$

This gives the result. □

It follows that there is always an *odd* value of $t \leq b'$ such that $|S^{t+1}(v)| \leq c|N^t(v)|$, and this is the consequence of the lemma that we actually use.

3 The Separator Theorem

Let A and B be two disjoint subsets of V in an undirected graph $G = (V, E)$. A subset S of V is said to *separate A from B* in G if S is disjoint from $(A \cup B)$ and the graph G' induced by deleting all vertices in S has no path from a vertex in A to one in B. It will actually be cleaner for use to relax the condition that S be disjoint from A and B, as done by Bollobás in [6]: Call S a "weak separator" for A, B if deleting all vertices in S leaves no path from a vertex in $A \setminus S$ to one in $B \setminus S$. The corresponding form of *Menger's Theorem* that we use is also given in [6]:

Theorem 3.1 *If A, B have no weak separator of size $k-1$, then there are k paths connecting A and B such that no two paths share a vertex.*

The converse also holds (immediately), but this is the direction we use. We now have enough to state and prove our main theorem in full detail.

Theorem 3.2 *Given fixed $\epsilon > 0$, $k \geq 1$, and $a \geq 2$, we can find $\delta > 0$ such that for any graph $G = (V, E)$ of vicinity at^k, the following holds: For any disjoint $A, B \subseteq V$, with $m = |A| \geq |B|$, there exist $A' \subseteq A$, $B' \subseteq B$, and $S \subseteq V$ such that*

- *S weakly separates A' from B'.*
- *$|A'| \geq (\frac{1}{2} - \epsilon)|A|$,*
- *$|B'| \geq \frac{1}{2}|B|$, and*
- *$|S| \leq \frac{N \log_2 m}{\delta m^{1/k}}$,*

where $N = |V|$.

If $N \log_2 m / \delta m^{1/k} < \epsilon m$, then we can arrange that S is disjoint from A' and B'; i.e., that S separates A' from B'.

The hypothesis of the last sentence will hold asymptotically in cases where $m = \Theta(n)$ and (for sake of contradiction) we suppose that N is not $\Omega(n^{1+1/k})$. The conclusion of the last sentence follows simply by renaming A' to $A' \setminus S$ and B' to $B' \setminus S$ and adjusting δ a little. Hence we can regard Theorem 3.2 as producing a separator in the traditional graph-theoretic sense. Both our proof and the observations in [26, 17] that our applications build on, however, work more naturally for weak separators, and so we refer to weak separators from now on.

Proof. Let $\ell = \delta m^{1/k} / \log_2 m$, where we explain how to choose δ at the end. Now create a bipartite graph Γ with edges from A to B defined by: for u in A and v in B, (u, v) is an edge in Γ if and only if there is a path of length at most ℓ from u to v in G. Let I be any independent set in Γ, and set $A' = A \cap I$, $B' = B \cap I$. Then any path going from A' to B' in G has length greater than ℓ. By Theorem 3.1, if there is no S of size N/ℓ that weakly separates A' from B', then there are N/ℓ vertex-disjoint paths from A' to B' in G. However, the total number of vertices in these paths would be greater than N. This contradiction shows that there does indeed exist a weak separator S of size at most N/ℓ.

Thus all we have to do is construct I so that A' and B' have the desired sizes.

The graph Γ has vicinity bounded by $a\ell^k t^k$. This is because a path of length t in Γ corresponds to a path of length at most ℓt in G. Now let $a_0 = a\ell^k$. Note that a_0 varies with m. The strategy from here on is (1) choose a suitably small constant c—taking $c = \epsilon/(2 + \epsilon)$ will be seen to suffice, (2) calculate the quantity b' in Lemma 2.2 in terms of a_0, c, and k, (3) choose δ (on which ℓ depends) so that $a_0 \ell^k b'^k \leq \epsilon m/2$, and finally (4) show that with these choices, we can build the desired I. Step (3) is possible because

$$b' = O(\log a_0) = O(\log \ell),$$

and so

$$\begin{aligned} a_0 \ell^k b'^k &= [a_0 \delta m / (\log_2 m)^k] \cdot O(\log^k \ell) \\ &= [a_0 \delta m / (\log_2 m)^k] \cdot O(\log^k m) = \delta \cdot O(m). \end{aligned}$$

The constant C inside the "O" depends only on c, k, and a, and we simply choose $\delta = \epsilon/2C$.

A vital fact for our argument is that every vertex-induced subgraph of Γ has the same upper bound $a\ell^k t^k$ on its vicinity, and hence we can use the same estimates in a process that recursively breaks off "pieces" P of Γ: At any step in the process, let v be any vertex in A that does not yet belong to a "piece." Find an odd $t \leq b'$ from Lemma 2.2, and let $P = N^{t+1}(v)$. Now by the vicinity bound and the choice in Step (3), there are at most $\epsilon m/2$ vertices in P. Let $\Gamma' = \Gamma \setminus P$, and continue this process recursively on Γ', until all remaining connected components have size at most $\epsilon m/2$. These remaining connected components are called "leftover pieces."

For a non-leftover piece P, its boundary is $S^{t+1}(v)$, and this is a subset of A. In any event, define the "A-side" of P to be $N^t(v) \cap A$, and the "B-side" to be $N^t(v) \cap B$. The A-side equals $(P \cap A) \setminus S^{t+1}(v)$, and we think of $|S^{t+1}(v)|$ as "lost" when choosing the A-side. The B-side, however, does equal $P \cap B$, so there is no loss from choosing that side. For a leftover piece P, the A-side is simply $P \cap A$ and the B-side is $P \cap B$. Then *any* choice of A-side or B-side from each piece produces an independent set in Γ.

Now order the pieces according to the *ratio* of the cardinality of their A-side to that of their B-side, in non-increasing order of this ratio. Form I by choosing A-sides until the running sum of cardinalities is at least $(1/2 - \epsilon)|A|$, and then for all remaining pieces, choosing the B-sides.

The only thing we have to do now is show that with $B' = I \cap B$, $|B'| \geq |B|/2$.

Let $\alpha_1 = |I \cap A|$, let β_1 be the sum of the sizes of the B-sides of the pieces whose A-sides were chosen, let α_2 similarly sum the A-sides of the pieces whose B-sides were chosen, and let $\beta_2 = |B'|$. Put $M = \alpha_1 + \alpha_2 + \beta_1 + \beta_2$. We claim that $|A| \leq \alpha_1 + \alpha_2 + cM$. This is because at each stage, t is chosen so that $|S^{t+1}(v)| \leq c|N^t(v)|$—hence the total "loss" on the A-side is at most c times the sum of $|N^t(v)|$ over all such stages, which sum in turn is at most M. This proves the claim. Since $\beta_1+\beta_2 = |B| \leq |A|$, we get $M \leq 2(\alpha_1+\alpha_2)+cM$, so $M \leq 2(\alpha_1+\alpha_2)/(1-c)$, and so

$$|A| \leq (\alpha_1 + \alpha_2)(1 + 2c/(1-c)).$$

Now because every piece has size at most $\epsilon|A|/2$ (since $m = |A|$), we obtain:

$$\begin{aligned} \alpha_1 &\leq |A|(\frac{1}{2} - \epsilon) + \epsilon|A|/2 \\ &= |A|(\frac{1-\epsilon}{2}) \\ &\leq \left(\frac{\alpha_1+\alpha_2}{2}\right)((1+2c/(1-c))(1-\epsilon)) \\ &\leq \left(\frac{\alpha_1+\alpha_2}{2}\right) \end{aligned}$$

provided $2c/(1-c) \leq \epsilon$, which we arrange by choosing $c \leq \epsilon/(2+\epsilon)$.

From $\alpha_1 \leq (\alpha_1 + \alpha_2)/2$ it follows directly that $\beta_2 \geq (\beta_1 + \beta_2)/2$, by the scaled ordering of the pieces. This gives $|B'| \geq \frac{1}{2}|B|$, and this completes the entire proof. □

4 Applications

In this section we fix $\epsilon = 1/10$, so that always $|A'| \geq (2/5)|A|$.

Valiant [26] defined a directed acyclic graph G to be an $(f(r), s, t)$-*grate* if there exist disjoint $A, B \subset V$ with $|A| = s$ and $|B| = t$ such that if any vertex set $S \subset V$ of size r is removed from G, then the resulting graph G' still has at least $f(r)$-many pairs $(u, v) \in (A \setminus S) \times (B \setminus S)$ such that there is a path from u to v in G'.

We take $s = t = m$. We will consider functions f of the form $f(r) = (m-r)^2$. Note that if $r = o(m)$, then we can arrange the constants so that $(m-r)^2 > (4/5)m^2$. It follows that in an $f(r)$-grate, there cannot be a set S of r vertices whose deletion separates $2/5$ of the vertices in A from $1/2$ of the vertices in B. With some reasonable abuse of asymptotic notation, we can state:

Lemma 4.1 *A DAG G whose underlying undirected graph has vicinity at^k cannot be an $((m-r)^2, m, m)$-grate (for any given m) unless its size N is $\Omega(m^{1+(1/k)}/\log m)$.*

Proof. The size r of the separator S in Theorem 3.2 becomes $\Omega(m)$ only when $N = \Omega(m^{1+(1/k)}/\log m)$. □

Valiant [26] proved that the graph of any m-input, m-output Boolean circuit computing the linear transformations $x \mapsto Ax$, where A is an $m \times m$ matrix over GF(2) and $x \in \{0,1\}^m$, must be an $(R_A(r), m, m)$-grate. Here $R_A(r)$ is called the *rigidity function* of A, and is defined to be the minimum number of '1' entries in an $m \times m$ matrix B over GF(2) such that the rank of $A + B$ is at most r. (The paper [26] gives definitions for matrices and circuits over arbitrary fields, but the GF(2) case with Boolean circuits suffices for our purposes.) Valiant proved that most $m \times m$ matrices A are "highly rigid," meaning that $R_A(r) = (m-r)^2$ for all r, which is the maximum possible value. Combined with all of the above, we obtain our first nonlinear lower bound result.

Theorem 4.2 *Circuits of vicinity $O(t^k)$ cannot compute linear transformations by highly-rigid $m \times m$ matrices, unless they have size $\Omega(m^{1+1/k}/\log m)$.* □

Note that constants can be supplied to make the bounds concrete for individual m-input circuits rather than asymptotic.

One limitation of this result is that no explicit constructions of families of $m \times m$ matrices A_m (for general m) of rigidity $(m-r)^2$, or even rigidity $\Omega(m^{1+\delta})$ for some fixed $\delta > 0$ and $r = \Theta(m)$, are known. Rigidity $\Omega(m^{1+\delta})$ as above suffices for Valiant's conclusion that log-depth circuits require "just barely superlinear" size $\Omega(m \log\log m / \log\log\log m)$ to compute $A_m x$. Our Theorem 4.2 seems to give superlinear lower bounds on size for PV circuits only for the highest rigidity functions.

However, Valiant's approach ties in readily to ideas and results of Mansour, Nisan, and Tiwari [17]. They define a set E of strings of length m to be a k-*cylinder* if there is a set J of k indices $1 \leq j_1 \leq j_2 \leq \ldots \leq j_k \leq m$ and a string v of length k such that $E = \{ x \in \{0,1\}^m : (\forall r, 1 \leq r \leq k)\, x_{j_r} = v_r \}$. (We abbreviate this condition by writing $E = \{ x : x_J = v \}$.) Then they define:

Definition 4.1 ([17]). A function $f : \{0,1\}^\ell \to \{0,1\}^m$ has the *property of randomness with parameters* (n, α, β) if for all $k \leq m$, every n-cylinder $D \subseteq \{0,1\}^\ell$, and every k-cylinder $E \subseteq \{0,1\}^m$,

$$\Pr_{x \in D}[f(x) \in E] \leq 2^\beta / 2^{\alpha k}. \qquad (2)$$

To state this definition another way, let I be the set of n indices and $u = u_1 \ldots u_n$ the fixed input values that define the n-cylinder D, and let J and $v = v_1 \ldots v_k$ similarly define E for the output values. Then (2) can be rewritten as the conditional probability

$$\Pr_{x \in \{0,1\}^\ell}[f(x)_J = v \mid x_I = u] \leq 2^{-\alpha k + \beta}.$$

Now suppose C is a circuit computing f on inputs of length ℓ, and suppose S is a set of nodes in C that (weakly) separates all but n input nodes of C from some k output nodes. Then the size s of S must be at least $\alpha k - \beta$. Otherwise, let I be the indices of the leftover n input nodes, and let us (arbitrarily!) fix values $u = u_1, \ldots, u_n$ for those nodes, giving us D. Conditioned on $x_I = u$ (i.e. $x \in D$), the values in the k output nodes then depend only on the s values induced by x on the gates in S. It follows that some pattern $v = v_1, \ldots, v_k$ in those output nodes occurs with probability at least $1/2^s$ over $x \in D$—indeed, this is so of every pattern that occurs with nonzero probability. If $s < \alpha k - \beta$, then this probability is too large.

Theorem 4.3 *If f as in Definition 4.1 has $l, m = \Theta(n)$ and has the property of randomness with parameters (n, α, β), with α constant and $\beta = o(n)$, then circuits C of vicinity at^k that compute f must have size $\Omega(n^{1+(1/k)}/\log n)$.*

Proof. Let N be the size of C. Start by choosing A_1 to be the set of input nodes of C, and B_1 to be the set of output nodes. Theorem 3.2 then gives a subset A_1' of A of size at least $(2/5)\ell$, a subset B_1' of B of size at least $m/2$, and a set S_1 of size $O(N \log n / n^{1/k})$ that weakly separates A_1' from B_1' in (the underlying *undirected* graph of) C. Now take $A_2 = A_1 \setminus A_1'$ and $B_2 = B_1'$, and re-apply the theorem to get a separator S_2 of similar size. Continue until some stage r at which $A_r \setminus A_r'$ has size at most n. Then r is a constant depending only on ℓ, and the set $S = S_1 \cup S_2 \cup \ldots S_r$ weakly separates $A_1' \cup A_2' \cup \ldots A_r'$ from B_r. Since r is constant, the size s of S is $O(N \log n / n^{1/k})$. It follows that $s \geq \alpha|B_r| - \beta$. Since $|B_r| = \Theta(n)$, we obtain $N \log n / n^{1/k} = \Omega(\alpha n - \beta)$, so $N = \Omega(n^{1+(1/k)}/\log n)$ as stated in the theorem. □

Mansour, Nisan, and Tiwari showed that all families of *universal$_2$ hash functions* $h : \{0,1\}^n \to \{0,1\}^n$ are such that the function $f(h,x) = h(x)$ has the property of randomness with parameters $(n, 1/2, 1)$. One example is the family $H = \{\, h_{a,b} : a, b \in \mathrm{GF}(2^n) \,\}$, where $h_{a,b}(x) = ax + b$ for $x \in \mathrm{GF}(2^n)$. Then the function $f(h,x)$ is just $ax + b$, with input length $\ell = 3n$ and output length $m = n$. Another example in [17] is defined in terms of "string convolutions": Given binary strings $x = x_1 \ldots x_n$ and $y = y_1 \ldots y_r$, where $r \geq n$, define $x \circ y$ to be the string z of length $r - n + 1$ such that for all i, $1 \leq i \leq r - n + 1$, bit z_i equals the GF(2) inner product of x with $y_i \ldots y_{i+n-1}$. Then the family $H = \{\, h_{y,w} : y \in \{0,1\}^{2n-1}, w \in \{0,1\}^n \,\}$ with $h_{y,w}(x) = (x \circ y) + w$ is a universal$_2$ hash family. Here $+$ is the same as bit-wise exclusive-or, and since this has trivial circuits, the separator-size lower bound applies to circuits computing $x \circ y$. That sorting functions have similar parameters α and β is also mentioned in [17].

Corollary 4.4 *For all $k \geq 1$, circuits of vicinity $O(t^k)$ that compute sorting, $ax + b$ in finite fields, or string convolutions require size $\Omega(n^{1+1/k}/\log n)$.*

It follows that no sorting networks of size $n(\log n)^{O(1)}$ can have polynomial vicinity.

5 Discussion: Stronger Results?

Let us fix attention momentarily on graphs of quadratic vicinity, and draw some useful comparisons with the Lipton-Tarjan theorem (LT) for planar graphs. Recall that these two classes of graphs are incomparable, although their practical motivations are similar. Let $G = (V, E)$ be an undirected graph with N nodes that belongs to one, respectively the other, class.

LT finds a set S_0 of size $O(N^{1/2})$ whose removal breaks G into two "halves" A_0 and B_0, each of size at least $N/3$. (The resulting A_0 and B_0 need not themselves be connected, so long as they are disconnected from each other in $G \setminus S_0$.) Thus we have a partition of V into A_0, B_0, S_0 such that S_0 separates A_0 from B_0 in G. Using Theorem 3.2, we obtain:

Theorem 5.1 *Graphs G of quadratic vicinity and size N can be partitioned into A_1, B_1, S_1 such that A_1 and B_1 have size at least $N(1/4 - \epsilon)$ (for any desired fixed $\epsilon > 0$), S_1 has size $O(N^{1/2} \log N)$, and S_1 separates A_1 from B_1 in G.*

More generally, if the graphs G have vicinity $O(t^k)$, then S_1 has size $O(N^{1-(1/k)} \log N)$, and A_1 and B_1 still have size at least $N(1/4 - \epsilon)$.

Proof. Take A to be any set of $N/2$ nodes and $B = V \setminus A$, Then Theorem 3.2 produces A' and B', each of size at least $N(1/4 - \epsilon)$, and an S_1 of size $O(N^{1/2} \log N)$ that separates them. Here, as remarked in Theorem 3.2, we can arrange that S_1 is disjoint from A' and B', so it is a true separator. Then we can add the remaining nodes from $G \setminus S_1$ to A' and B' to obtain a partition A_1, B_1, S_1 of V such that S_1 separates A_1 from B_1. □

The extra $\log N$ factor in the size of S and the $N/4$ rather than $N/3$ represent a slight slippage in bounds compared to the LT theorem. For general weighting functions $wt : V \to [0,1]$, things become somewhat more problematic. Still, these bounds are good enough for some of the applications in [15] and elsewhere. Our first open question is: Can the bounds in Theorem 5.1, and those in the main Theorem 3.2, be improved?

The second open question we raise is whether the theorem can be improved in cases where the size m of the chosen subsets A and B is little-oh of the size N of the graph, as happens in Theorems 4.3 and 4.4. Let us ignore factors of $\log m$ or $\log N$ in this paragraph. The upper bound $N/m^{1/k}$ on the size of S in Theorem 3.2 is unusual insofar as it scales upward as m goes downward. One might think that smaller m should make the subsets A and B easier to separate, hence that the upper bound should go lower or at least stay the same, but what happens in the dynamics of the proof is that the smaller m leaves less choice in finding the subsets A' and B' that are actually separated. If the bound were $N^{1-(1/k)}$ *as in Theorem 5.1*, then Corollary 4.4 would hold with a size lower bound of $\Omega(n^{1+1/(k-1)})$ in place of $\Omega(n^{1+(1/k)})$. The change from "$1/k$" to "$1/(k-1)$" is significant: in the case of quadratic vicinity (i.e., $k = 2$), it would yield a quadratic lower bound on the circuit size for sorting and $ax + b$ and $x \circ y$. Since these functions are computable in linear time *cum* log factors, this lower bound is met by the circuits of Savage [22] (see also [11] or [5]), which have quadratic size and quadratic vicinity. Hence this stronger version of Corollary 4.4 would give a sense in which Savage's construction is best possible, and would tighten the "size-vicinity tradeoff" implied by our results.

It is notable that the *only* use of the PV condition on G in Theorem 3.2 is to get an analogous condition on the bipartite graph Γ. Moreover, its only use on Γ is to ensure that each individual "piece" in the proof is small and has an even smaller boundary. Perhaps some tangibly weaker condition than PV can support the same theorem, with a similar proof. It is true that the graphs Γ obtained from G are special: one can show that under various notions of a "random" bipartite graph of size m and degree $a\ell^k = m^{\Omega(1)}$, such graphs do not have independent sets I with a constant proportion of nodes in each partition. However, it is also true that when $m = o(N)$, as in the last paragraph, going from G to Γ "throws away" most of the graph! This above all leads us to suspect that there is some wastage in our main proof that can be exploited for stronger results when m is relatively small.

Note that we could also get this improvement to Corollary 4.4 if we could choose the initial sets A_1 and B_1 in the proof of Theorem 4.3 to have size $\Omega(N)$ rather than $O(m)$. The "property of randomness" on which all this is based, however, restricts attention to the input and output nodes. Hence a third question is whether the results in [17] can themselves be extended to yield conclusions about the distribution of Boolean values at other levels of circuits computing these functions besides the inputs and outputs. This seems more problematic than improving Corollary 4.4 directly, however.

A fifth matter is that all of our work has been based on *un*directed graphs. We do not see any immediate improvement that would follow from the hypotheses that G is directed, and that A is a set of sources and B a set of sinks. However, let us also note that the circuits of [22] are *leveled*. It seems that this last condition should be exploitable for stronger bounds, but we have not (yet) achieved this.

Finally, we ask whether our proof itself can be simplified, or whether our theorem follows from some combination of results about other known classes of graphs. We have given some counter-indication on the latter in Sections 1 and 2. We suspect there is a connection via eigenvalues and graph diameters drawing on the results of [1] and [7], but thus far it seems both far from "simple" and to give different bounds, at least in the analogous situation of [24]. One possibility for a simpler proof leads to our final question: In graphs G of vicinity $O(t^k)$, for any vertex subsets A, B as in the statement of Theorem 3.2, does there always exist a vertex v such that one of the "shells" $S^t(v)$ fulfills the conclusions of the theorem? This is plausible because the radius of the graph from v is $\Omega(N^{1/k})$, and hence some $S^t(v)$ has size $\Omega(N^{1-1/k})$. The question is whether at least one of these large $S^t(v)$ shells must separate a large chunk of A from a large chunk of B.

6 Conclusions

We have defined a condition on circuits that is both theoretically natural and practically significant, and have shown nonlinear lower bounds under this condition. The lower bounds are substantial: $\Omega(n^{4/3})$ for cubic-vicinity graphs, $\Omega(n^{3/2})$ for quadratic vicinity (ignoring log factors).

Results of this kind have been obtained for specific kinds of "well-behaved mesh" networks or other VLSI circuits. The special interest in our results is that they abstract away from many model-specific details of the circuits, focusing on the information-theoretic nature of the problem of how many data bits can be accessed within a span of t time units. The graphs defined by our condition seem to be incomparable with other classes of graphs that have been used for the former results.

This work also suggests other lines of research, most notably on size-vicinity tradeoffs for circuits, and their possible relation to other tradeoffs in complexity theory. One idea is to see if our lower bounds can be met by upper bounds—e.g., can linear-time Turing machines be simulated by cubic-vicinity circuits of $\Omega(n^{4/3})$, or at least $\Omega(n^{3/2})$, size? In this regard we note results by Fortnow [9] in these proceedings: SAT cannot be accepted by uniform circuits of log depth and $n(\log n)^{O(1)}$ size; and for Turing machines, SAT must lie outside either nondeterministic log space or deterministic $n(\log n)^{O(1)}$ time. Is there any connection? One weakness of the technique in Section 4, as with [26, 17], is that it really applies only for circuits computing *functions*, not those recognizing *languages*. Can we define and find non-separation properties for languages?

Acknowledgments I specially thank Shang-Hua Teng and Richard Lipton for extensive discussions of graph separators and their applications. Tao Jiang and Satya Lokam also gave helpful information about graphs and Valiant's papers. Jin-Yi Cai and Lijie Shi suggested simplifications for Lemma 2.1, some of which have been adopted here. I am also grateful to students in theory seminars at UB as a sounding board for my preliminary attempts at the results in this paper. Lokam, Lipton, the anonymous referees of Complexity'97, and D. Sivamukar also gave some assistance in preparing the present draft of this paper.

References

[1] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6:83–96, 1986.

[2] N. Alon, P. Seymour, and R. Thomas. A separator theorem for graphs with an excluded minor and its applications. In *Proc. 22nd Annual ACM Symposium on the Theory of Computing*, pages 293–299, 1990.

[3] N. Alon, P. Seymour, and R. Thomas. A separator theorem for non-planar graphs. *J. Amer. Math. Soc.*, 3, 1990.

[4] L. Babai. The growth rate of vertex-transitive planar graphs. In *Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 564–573, New Orleans, Louisiana, 5–7 Jan. 1997.

[5] J. Balcázar, J. Díaz, and J. Gabarró. *Structural Complexity I*. Springer Verlag, 1988.

[6] B. Bollobás. *Extremal Graph Theory*. Academic Press, 1976.

[7] F. Chung. Diameters and eigenvalues. *J. Amer. Math. Soc.*, 2:187–196, 1989.

[8] Y. Feldman and E. Shapiro. Spatial machines: A more-realistic approach to parallel computation. *Comm. Assn. Comp. Mach.*, 35:60–73, October 1992.

[9] L. Fortnow. Nondeterministic polynomial time versus nondeterministic logarithmic space. In *Proceedings, 12th IEEE Conference on Computational Complexity (formerly Structure in Complexity Theory), Ulm, Germany, June*, 1997. To appear.

[10] J. Gilbert, J. Hutchinson, and R. Tarjan. A separation theorem for graphs of bounded genus. *J. Alg*, 5:391–407, 1984.

[11] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison–Wesley, Reading, MA, 1979.

[12] H. B. Hunt, S. S. Ravi, and R. E. Stearns. Separators, graph homomorphisms and chromatic polynomials (extended abstract). manuscript, 1988.

[13] C. Kruskal, L. Rudolph, and M. Snir. A complexity theory of efficient parallel algorithms. *Theor. Comp. Sci.*, 71:95–132, 1990.

[14] R. Lipton and R. Tarjan. A separator theorem for planar graphs. *SIAM J. Appl. Math.*, 36:177–189, 1979.

[15] R. Lipton and R. Tarjan. Applications of a planar separator theorem. *SIAM J. Comput.*, 9:615–627, 1980.

[16] H. Macpherson. Growth rates in infinite graphs and permutation groups. *Proc. London Math. Soc.*, 51:285–294, 1985.

[17] Y. Mansour, N. Nisan, and P. Tiwari. The computational complexity of universal hashing. *Theor. Comp. Sci.*, 107:121–133, 1993.

[18] G. Miller, S.-H. Teng, W. Thurston, and S. Vavasis. Finite element meshes and geometric separators, 1996. To appear in *SIAM J. Sci. Comp.*

[19] G. Miller, S.-H. Teng, W. Thurston, and S. Vavasis. Separators for sphere-packings and nearest-neighbor graphs, 1996. Submitted toJ. Assn. Comp. Mach.

[20] N. Pippenger and M. Fischer. Relations among complexity measures. *J. Assn. Comp. Mach.*, 26:361–381, 1979.

[21] K. Regan. On superlinear lower bounds in complexity theory. In *Proc. 10th Annual IEEE Conference on Structure in Complexity Theory*, pages 50–64, 1995.

[22] J. Savage. Computational work and time on finite machines. *J. Assn. Comp. Mach.*, 19:660–674, 1972.

[23] A. Schorr. Physical parallel devices are not much faster than sequential ones. *Inf. Proc. Lett.*, 17:103–106, 1983.

[24] D. Spielman and S.-H. Teng. Spectral partitioning works: Planar graphs and finite element meshes. In *Proc. 37th Annual IEEE Symposium on Foundations of Computer Science*, pages 96–105, 1996.

[25] P. Ungar. A theorem on planar graphs. *J. London Math. Soc.*, 26:256–262, 1951.

[26] L. Valiant. Graph-theoretic arguments in low-level complexity. In *Proc. 2nd International Symposium on Mathematical Foundations of Computer Science*, volume 53 of *Lect. Notes in Comp. Sci.*, pages 162–176. Springer Verlag, 1977.

[27] P. Vitányi. Locality, communication, and interconnect length in multicomputers. *SIAM J. Comput.*, 17:659–672, 1988.

[28] L. Wang and T. Jiang. An approximation scheme for some Steiner tree problems in the plane, 1996.