Circuit lower bounds and linear codes

Ramamohan Paturi* Pavel Pudlák†

October 1, 2004

Abstract

In 1977 Valiant proposed a graph theoretical method for proving lower bounds on algebraic circuits with gates computing linear functions [13]. He used this method to reduce the problem of proving lower bounds on circuits with linear gates to proving lower bounds on the rigidity of a matrix, a concept that he introduced in that paper. The largest lower bound for an explicitly given matrix is due to J. Friedman, who proved a lower bound on the rigidity of the generator matrices of error correcting codes over finite fields [3]. He showed that the proof can be interpreted as a bound on a certain parameter defined for all linear spaces of finite dimension. In this note we define another parameter which can be used to prove lower bounds on circuits with linear gates. Our parameter may be larger than Friedman's and it seems incomparable with the rigidity, hence it may be easier to prove a lower bound using this concept.

1 Introduction

The problem of proving nontrivial lower bounds on the size of circuits is one of the most fundamental problems in theoretical computer science. Its simplest version—the problem of proving nonlinear lower bounds on the size of logarithmic depth circuits computing an explicitly given function—is still open. It is open not only for boolean circuits, but even for algebraic circuits with gates computing linear functions. This is in spite of the apparent simplicity of such circuits; for example, in case of the two element field the circuits use only the parity gate. In 1977 Valiant found a reduction of the latter problem to an algebraic-combinatorial problem about matrices [13]. He introduced the concept of the rigidity function of a matrix and proved that sufficiently large lower bounds on the rigidity of a matrix imply nonlinear lower bounds on the size of circuits with linear gates computing the linear transformation defined by the matrix. Some other approaches to proving lower bounds on circuits with linear gates were proposed by Morgenstern [10] and Grigoriev [6].

In spite of the considerable amount of work done on this problem, we still lack strong bounds on the rigidity of explicitly defined matrices. The largest lower bound was proved by Friedman in 1990 [3]; he used generator matrices of a good code (the special case with the field being GF_2 is re-proved after Proposition 7). The bound is still far from what is needed for circuit lower bounds. Friedman observed that his proof gives a little more: it gives a lower bound on a natural parameter of linear codes. Later asymptotically the same bounds were proved for infinite fields by Spielman, Stetmann, and Shokrollahi [12] and Lokam [7]. More recently, Alekhnovich studied a hypothesis that is related both to the rigidity of matrices and to linear codes [1].

*University of California, San Diego

†Mathematical Institute, Academy of Sciences of the Czech Republic, Prague. Supported by grants no. EAA1019401 of the Academy of Sciences of the Czech Republic and no. LN0056 of the Ministry of Education of the Czech Republic, pudlak@math.cas.cz

The theory of error correcting (and other) codes is a field with a large body of results and most of these concern linear codes. It is a field that extensively studies relations between algebraic and combinatorial properties of linear spaces. Thus we think that the relations between circuit complexity and codes deserves more attention than it was given so far. Codes with large minimal distances were used in lower bounds for various types of circuits, but the connection with circuits with linear gates seems to be the most promising. Therefore we will propose another reduction in this paper. It is also based on the graph-theoretical transformation of Valiant, but our parameter seems to be incomparable to rigidity and sometimes it is larger than Friedman's. So it may be easier to prove lower bounds on the circuit size using this parameter.

2 Circuits with linear gates

Let F be a field. We consider circuits whose gates are functions of the form $ax + by$, for $a, b \in F$. In particular, the fan-in of all gates is 2. The size of a circuit is the number of gates, the depth is the length of the longest (oriented) path. Such a circuit computes a linear transformation $f : F^n \to F^m$. We shall assume that there are no constant inputs, hence f will always be homogeneous. If $m = 1$, the problem of the circuit size complexity is trivial, thus we always consider $m > 1$. If $n = m$, then there are linear transformations whose complexity is almost quadratic, however, no non-linear lower bound is known for an explicitly defined f. This problem is open even with the additional restriction that the depth of the circuit be $O(\log n)$. The well-known result of Valiant [13] is a reduction of the above problem (with the log-depth restriction) to proving lower bounds on the rigidity of explicitly defined matrices. It is based on the following combinatorial lemma.

Lemma 1 ([13]) *Let r, δ, σ be positive integers with $\delta > 4\sigma$. Let G be a directed acyclic graph with at most $r \log_2 \delta / \log_2(\delta/4\sigma)$ edges and depth at most δ. Then there exists a set of at most r edges such that after removing these edges, G does not contain a path of length σ.*

We shall say that a vector $v \in F^n$ is *s-sparse,* if the number of nonzero elements of v (the weight of v) is at most s. Valiant's result can be stated as follows.

Theorem 2 *Let r, δ, σ be as above. Suppose a linear transformation defined by an $n \times m$ matrix M can be computed by a circuit with linear gates and with size at most $r \log_2 \delta / (2 \log_2(\delta/4\sigma))$ and depth at most δ. Then the matrix can be decomposed as follows.*

$$M = A + BC, \tag{1}$$

where B is an $n \times r$ matrix, C is an $r \times m$ matrix and the rows of matrices A and C are 2^σ-sparse.

The concept of rigidity was derived by forgetting part of the information contained in equation (1). Namely, one uses only the information that A has at most sn nonzero entries, where $s = 2^\sigma$, and that the matrix BC has rank at most r. The rigidity of a matrix M is the function that expresses the dependence of the sparsity on the rank:

$$R_M(r) =_{df} \min\{R;\ \exists A \text{ a matrix with } R \text{ nonzero entries such that } \operatorname{rank}(M - A) \leq r\}.$$

In order to prove a nonlinear lower bound on log-depth circuits computing the linear transformations defined by a matrix M, we need to prove a bound

$$R_M(\epsilon n) \geq n^{1+\delta},$$

for some constants $\epsilon, \delta > 0$. All the bounds proved for explicit matrices so far give only linear lower bounds on $R_M(\epsilon n)$. For instance, Friedman's lower bound is of the form $R_M(r) = \Omega(\frac{n^2}{r}\log\frac{n}{r})$, which is linear for $r = \epsilon n$.

For the purpose of this paper it is better to consider the function that expresses the dependence of the rank on the sparsity. Furthermore, we shall assume a uniform bound on the sparsity of the rows, as in Theorem 2. So we define:

$$r_M(s) =_{df} \min\{r;\ \exists A \text{ a matrix whose rows are } s\text{-sparse, and } \operatorname{rank}(M-A) = r\}.$$

Suppose that M has fewer rows than columns. We can derive from (1) that the row space of M is contained in the sum of the row space of A and the row space of BC. Since A and C are sparse matrices, we get nontrivial information about the row space of M. We shall use this observation to define two more functions. As they depend only on the row space of the matrix, we shall define them for spaces instead of matrices. The first one is based on the concept of strong rigidity of Friedman [3]. Let $V \subseteq F^n$, $1 \leq s \leq n$.

$$d(s,V) =_{df} \max\{\dim(V \cap U);\ U \subseteq F^n,\ \text{generated by } s\text{-sparse vectors},\ \dim U = \dim V\}.$$

We observe that the equality $\dim U = \dim V$ in the definition can be replaced by $\dim U \leq \dim V$. Indeed, if $\dim U < \dim V$, we can extend an s-sparse basis of U by 1-sparse vectors, so that it has $\dim V$ elements.

Let $\langle M \rangle$ denote the row space of M. The above two functions are related as follows.

Proposition 3 ([3]) *Let M be a full rank matrix such that the number of rows of M is at most the number of columns, let $s \geq 0$. Then*

$$\operatorname{rank} M - d(s,\langle M\rangle) = \dim\langle M\rangle - d(s,\langle M\rangle) \leq r_M(s).$$

Proof. The first equality follows since the rank of a matrix is equal to the dimension of its row space. Let A be a matrix such that $\operatorname{rank}(M-A) = \dim\langle M-A\rangle = r_M(s)$, with s-sparse rows. Since $\dim\langle A\rangle \leq \dim\langle M\rangle$, by definition,

$$\dim(\langle A\rangle \cap \langle M\rangle) \leq d(s,\langle M\rangle).$$

Since $\langle A\rangle + \langle M\rangle = \langle A\rangle + \langle M-A\rangle$, we have

$$\dim(\langle A\rangle + \langle M\rangle) \leq \dim\langle A\rangle + \dim\langle M-A\rangle.$$

From these inequalities and the equality

$$\dim\langle A\rangle + \dim\langle M\rangle = \dim(\langle A\rangle \cap \langle M\rangle) + \dim(\langle A\rangle + \langle M\rangle),$$

we get

$$\dim\langle M\rangle \leq \dim(\langle A\rangle \cap \langle M\rangle) + \dim\langle M-A\rangle \leq d(s,\langle M\rangle) + r_M(s),$$

which is the inequality of the proposition. ■

Thus, for a full rank matrix R, upper bounds on $d(s,\langle M\rangle)$ imply lower bounds on the rigidity of M. Friedman defined M to be (s,t) *strongly rigid,* if $t \geq d(s,\langle M\rangle)$.

We define another function:

$$D(s,V) =_{df} \min\{\dim U;\ V \subseteq U \subseteq F^n,\ U \text{ generated by } s\text{-sparse vectors}\}.$$

The next inequality is also easy.

Proposition 4 *Let V be a finte-dimensional vector space, let $s \geq 0$. Then*

$$\dim V - d(s, V) \leq D(s, V) - \dim V.$$

Proof. Let U be such that $D(s, V) = \dim U$, $V \subseteq U$ and U is generated by s-sparse vectors. Let $m = \dim V$ and $D = D(s, V)$. Let $u_1, \ldots, u_D$ be a basis of U consisting of s-sparse vectors. Let $U_1 = \langle u_1, \ldots, u_m \rangle$. We have

$$\dim U_1 + \dim V = \dim(U_1 \cap V) + \dim(U_1 + V) \leq \dim(U_1 \cap V) + \dim U.$$

Now apply the equalities $\dim U_1 = \dim V$, $\dim U = D(s, V)$ and the inequality $\dim(U_1 \cap V) \leq d(s, V)$. ■

However, we do not know any nontrivial inequality involving $r_M(s)$ and $D(s, \langle M \rangle)$.

We already know that the rigidity of a matrix M, hence also $d(s, \langle M \rangle)$ can be used to prove a lower bound on the size of circuits computing the transformation defined by M. The function D can be used in the same manner. We state explicitly this corollary of Theorem 2 below. Given a circuit C with n inputs, we shall say that a space $V \subseteq F^n$ is *generated* by the circuit C, if $V = \langle M \rangle$, where M is the matrix of the linear transformation computed by the circuit (in the standard basis).[1]

Corollary 5 *Let r, δ, σ be as above. Suppose a space V can be generated by a circuit of size at most $r \log_2 \delta / (2 \log_2(\delta / 4\sigma))$ and depth at most δ. Then*

$$D(2^\sigma, V) \leq \dim(V) + r.$$

In particular, for all constants $c_1, c_2, \varepsilon > 0$ there exists a constant c_3 such that if V can be generated by a circuit of size at most $c_1 n$ and depth at most $c_2 \log n$, then

$$D(n^\varepsilon, V) \leq \dim(V) + \frac{c_3 n}{\log \log n}.$$

Example. Consider a Vandermonde matrix

$$A = \begin{pmatrix} a_1^0 & a_2^0 & a_3^0 & \ldots & a_n^0 \\ a_1^1 & a_2^1 & a_3^1 & \ldots & a_n^1 \\ a_1^2 & a_2^2 & a_3^2 & \ldots & a_n^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_1^{n-1} & a_2^{n-1} & a_3^{n-1} & \ldots & a_n^{n-1} \end{pmatrix}$$

with $a_1, a_2, \ldots, a_n$ nonzero and pairwise distinct elements of a field F. Special cases of such matrices include Fourier matrices. It has been conjectured that the linear transformation defined by A cannot be computed by linear size and logarithmic depth circuits. The conjecture is open even if one takes $a_1, a_2, \ldots, a_n$ algebraically independent over some subfield. We believe that also the rigidity of this matrix is large (for partial results see [7]), namely that $r_A(s) = \Omega(n)$ if $s = o(n)$. The rows of A generate the whole space F^n, but any circuit which computes A also computes every transformation determined by a submatrix of A. Hence we can apply the corollary above to the space V generated by the first k rows of A. We conjecture that for every $\varepsilon < 1, 0 < \delta < 1/2$ there exists $\eta > 0$ such that if $\delta n \leq k \leq n/2$ and $s \leq n^\varepsilon$, then $D(s, V) \geq (1 - \eta)n$. This conjecture implies that A cannot be computed by linear size and logarithmic depth circuits. Notice that in case $a_1, a_2, \ldots, a_n$ are all nonzero elements of a finite field, the space V is a Reed-Solomon code.

[1] We can also say that V is the space generated by the linear functions computed at the output gates of C, however, it is important to say in which basis these vectors are presented.

3 Some simple bounds

We shall start with an estimate on the maximal value of $D(s,V)$ for spaces of a given dimension k over GF_2. Similar bounds can be proven for other finite fields.

Proposition 6 *Suppose F is GF_2. Let $k, s \leq n$ be arbitrary positive integers. Then there exists a space V with $\dim V = k$ such that*

$$D(s,V) \geq \frac{kn}{k + s\log_2 n}.$$

In particular, if $k \to \infty$ and $s \log n = o(k)$, then $D(s,V) = n - o(n)$.

Proof. We shall use a counting argument. The number of all subspaces of F^n of dimension k is $\begin{bmatrix} n \\ k \end{bmatrix}$. We shall upper bound the number of all sets of s-sparse vectors of size d by n^{sd}. If every space of dimension k is contained in the space spanned by d linearly independent s-sparse vectors, then

$$\begin{bmatrix} d \\ k \end{bmatrix} n^{sd} \geq \begin{bmatrix} n \\ k \end{bmatrix}.$$

Hence,

$$n^{sd} \geq \frac{(2^n-1)\dots(2^n-2^{k-1})}{(2^d-1)\dots(2^d-2^{k-1})} \geq \left(2^{n-d}\right)^k.$$

Thus

$$sd\log_2 n \geq k(n-d),$$

which proves the proposition. ■

Hence a random space V of dimension $n/2$ has $D(n^\varepsilon, V) = n - o(n)$, whereas every space of dimension $n/2$ which can be generated by linear size log depth circuit has only $D(n^\varepsilon, V) = n/2 + o(n)$. This estimate also shows that $D(s,V) - \dim V$ can be much larger than $\dim V - d(s,V)$. Indeed, let $s(n) = n^\epsilon$, $0 < \epsilon < 1$ be a constant; choose $k(n)$ so that $k(n) = o(n)$ and $s(n)\log n = o(k)$. Then we still have spaces V such that $D(s,V) = n - o(n)$, but $\dim V - d(s,V) \leq \dim V = k(n) = o(n)$. By padding these spaces we can get spaces V such that $\dim V = \Omega(n)$, $D(s,V) - \dim V = \Omega(n)$, and $\dim V - d(s,V) = o(n)$.

We shall present two simple lower bounds based on the distance of a code and the distance of the dual code.

Proposition 7 *Let C be a linear $[n,k,d]$ code (a code with length n, dimension k and minimal distance d), let $s \leq d/2$. Then there exists a $[D(s,C), k, d/s]$ code.*

Proof. Let $D = D(s,C)$. Take D s-sparse vectors which generate a space containing C. Let M be the $D \times n$ matrix composed from these vectors. Let $T : F^D \to F^n$ be the linear transformation defined by $x \mapsto xM$. The fact that the rows of M generate C can be equivalently stated as $C \subseteq T(F^D)$, where $T(F^D)$ denotes the image of the transformation T. Let $C' = T^{-1}(C)$. Then C' is a $[D, k', d']$ code. We claim that $k' \geq k$ and $d' \geq d/s$. The first inequality follows from the fact that C' is a preimage of a space of dimension k. To prove the second inequality, consider an arbitrary nonzero vector $x \in C'$. We know that $xM \in C$. If x has ℓ nonzero elements, then xM has at most ℓs nonzero elements, since the rows of M are s-sparse. Thus $\ell s \geq d$, whence $\ell \geq d/s$.
■

Suppose, for example, that F is the two element field. According to the sphere-packing bound we have

$$2^{k'}\binom{D}{d/2s} \leq 2^D.$$

Hence

$$D \geq k' + \log_2\binom{D}{d/2s} \geq k + \log_2\left(\left(\frac{k}{d/2s}\right)^{d/2s}\right) \geq k + \frac{d}{2s}\log_2(\frac{2sk}{d}).$$

Thus we get the following bound

$$D(s,C) \geq k + \frac{d}{2s}\log_2\left(\frac{2sk}{d}\right). \tag{2}$$

This bound also follows from Friedman's result [3]. Since the bound on D is based on a property that is inherited by subspaces, we can also do the opposite: to derive an upper bound on d from the lower bound on D. Indeed, suppose V is generated by s-sparse vectors, $\dim V = \dim C = k$. Let $C' = C \cap V$, thus C' is a $[n, k', d]$ for some $k' \geq d(s,C)$. Applying Proposition 7 to C' we get

$$k = \dim V \geq D(s,C') \geq k' + \frac{d}{2s}\log_2\left(\frac{2sk}{d}\right).$$

Whence

$$d(s,C) \leq k' \leq k - \frac{d}{2s}\log_2\left(\frac{2sk}{d}\right), \tag{3}$$

which is Friedman's bound.

Using more precise upper bounds on the dimension of codes of a given minimal distance we can get slightly better results, but we only get a better constant multiple for the second term in (2) and (3).

Proposition 8 *Let F be an arbitrary field. Let C be a code of length n whose dual code $C^\perp$ has minimal distance d. Then*

$$D(s,C) \geq d - 1 + \frac{n-d+1}{s}.$$

In particular, if $C^\perp$ is an MDS code (i.e., $d = \dim C + 1$), we get

$$D(s,C) \geq \dim C + \frac{n-d+1}{s}.$$

Proof. Suppose that there are D s-sparse vectors whose span contains C; let N be the $D \times n$ matrix formed by these vectors. Then every $d-1$ columns of N are linearly independent since $C^\perp$ has minimal distance d. Take $\frac{n-d+1}{s}$ rows of N. Since they are s-sparse, there are $d-1$ columns in which these rows have only zeros. The matrix formed by these columns has rank $d-1$, hence it must have $d - 1 + \frac{n-d+1}{s}$ rows. ∎

We do not know the best value of $D(s,C)$ for particular MDS codes such as the Reed-Solomon code. In general, we do not believe that linear rate and linear distance of a code or its dual alone imply a nonlinear bound on log-depth circuits. But it is conceivable that a lower bound proof could be based on codes achieving some extreme parameters.

4 Pseudorandom generators

Our inability to prove lower bounds, even for such restricted computational models as considered here, is frustrating. In a recent paper [1] Alekhnovich suggested that this difficulty can be viewed also positively. Razborov and Rudich [11] showed that if there are pseudorandom generators (with appropriate parameters), then it is not possible to base certain circuit lower bound proofs on simple properties of functions. Alekhnovich's idea is to turn it over and try to design new pseudorandom generators based on the assumption that certain circuit lower bounds are difficult. Our lower bound setting suggests a simple and natural construction that might be a pseudorandom generator.

Construction. The parameters are numbers n, k, D, s. The input data are two matrices over GF_2, an arbitrary $k \times D$ matrix A and a $D \times n$ matrix B each of whose rows has exactly s nonzero elements. The output is AB (the matrix product of A and B).

If $kD + D\lceil \log_2 \binom{n}{s} \rceil < kn$, then the function produces more output bits then is the number of bits needed to encode the input data. We conjecture that for parameters of the form stated in Proposition 9 below, if A and B are chosen uniformly, then the output is not computationally distinguishable (by a probabilistic polynomial time algorithm) from the uniform distribution on all $k \times n$ matrices (i.e., it is a pseudorandom generator). The following provides some basic supporting evidence. (Of course, no such statistical properties can be used to derive computational complexity.)

Proposition 9 *Let $0 < \alpha < \beta < 1$ and $0 < \varepsilon < 1/2$. Let $k = \lceil \alpha n \rceil$, $D = \lceil \beta n \rceil$ and s be an odd number, $s = \lceil n^\varepsilon \rceil$. Then*

1. *AB has full rank with probability exponentially tending to 1;*
2. *the distribution of every fixed column of AB is exponentially close to the uniform distribution;*
3. *the distribution of every fixed row of AB is exponentially close to the uniform distribution.*

Proof. 1. The probability that A does not have full rank is bounded by $2^{-D} + 2^{-D+1} + \cdots + 2^{-D+k-1} < 2^{k-D}$. For B we need the following claim.

Claim. Every subspace of GF_2^n of dimension d contains at most $\binom{d+s-1}{s}$ vectors of weight s.

Proof of the Claim. Let S be the set of supports of vectors of weight s of a given space; they are s-element subsets of $[n]$. Construct a sequence $s_1, s_2, \dots, s_e$ of elements of S and a sequence of nonempty subsets $X_1, X_2, \dots, X_e \subseteq [n]$ as follows. $s_1 = X_1$ is an arbitrary element of S. s_{i+1} is an element of S such that s_{i+1} is not a subset of $X_1 \cup \dots \cup X_i$ and such that $s_{i+1} \setminus (X_1 \cup \dots \cup X_i)$ has the minimal cardinality. Then we put $X_{i+1} = s_{i+1} \setminus (X_1 \cup \dots \cup X_i)$. Since the vectors corresponding to $s_1, s_2, \dots, s_e$ are linearly independent, this sequence has to stop with $e \le d$. Then every $s \in S$ is a subset of $X_1 \cup \dots \cup X_e$. More precisely, every $s \in S$ is of the form $t \cup X_{i_1} \cup \dots \cup X_{i_r}$ for some $t \subseteq s_1$ and $1 < i_1 < \dots < i_r \le e$. Clearly, we get the largest possible number of sets of this form if $|X_2| = \dots = |X_e| = 1$ and $e = d$, which is the bound of the Claim.

Now we can bound the probability that B does not have full rank by

$$\frac{\binom{s}{s}}{\binom{n}{s}} + \frac{\binom{s+1}{s}}{\binom{n}{s}} + \dots \frac{\binom{s+k-1}{s}}{\binom{n}{s}} < k\frac{\binom{s+k-1}{s}}{\binom{n}{s}} < k\left(\frac{s+k-1}{n-s+1}\right)^s,$$

which is also exponentially small. Clearly, the product of two full rank matrices of dimensions $k \times D$ and $D \times n$, with $k \le D \le n$, is a full rank matrix too.

2. The probability that a given column in B contains only zeros is

$$(1 - \frac{s}{n})^D = \mathrm{e}^{D \ln(1-\frac{s}{n})} < \mathrm{e}^{-\frac{Ds}{n}} \approx \mathrm{e}^{-\beta n^{\varepsilon}}.$$

If A is random and B is fixed with the i-th column nonzero, then the i-th column of AB has the uniform distribution. Hence if both A and B are random, then the distribution of the i-th column is exponentially close to the uniform distribution.

3. The distribution of a fixed row is given as follows. Take random $\mathbf{h}_1, \cdots, \mathbf{h}_D \in \{0,1\}$ and random vectors $\mathbf{u}_1, \cdots, \mathbf{u}_D \in GF^n$ each having exactly s ones. Then the distribution is

$$\sum_{i=1}^{D} \mathbf{h}_i \mathbf{u}_i.$$

We can view it as the result of the Markov process in which we start with the zero vector, and at each step we do nothing with probability $1/2$, or add a random vector with exactly s ones with probability $1/2$. Put differently, it is the result of D steps of a random walk on the n-dimensional Boolean cube which starts at the zero vector, and at each step with probability $1/2$ we do not move and with probability $1/2$ we move to a vertex in Hamming distance s, choosing such a vertex with uniform probability. Thus we need only to estimate the size of the second largest eigenvalue of the matrix of this process. The eigenvectors of this matrix are the same as the eigenvectors of the graph G_s on $\{0,1\}^n$ in which two vertices are connected iff their Hamming distance is exactly s. This graph is a Cayley graph on the additive group of the vector space GF_2^n, hence the eigenvectors are the characters of this group. It is well-known that they are

$$\chi_a(x) = (-1)^{x^\top a}, \text{ for } a \in GF_2^n.$$

We shall first estimate the eigenvalues of G_s. To compute the eigenvalue associated with χ_a, it suffices to consider the vertex $\bar{0}$ and its neighbors. The eigenvalue is

$$(\chi_a(\bar{0}))^{-1} \sum_{|x|=s} \chi_a(x) = (-1)^{x^\top a}.$$

The largest eigenvalue (associated with $\chi_{\bar{0}}$) is the degree of the graph $\binom{n}{s}$. The second largest eigenvalue is at most $(1 - \frac{\gamma s}{n})\binom{n}{s}$, for a constant $\gamma > 0$. (We believe that it is precisely $(1 - \frac{s}{n})\binom{n}{s}$ and it is associated with all χ_a such that a contains exactly one 1, but the weaker statement is all that we need.) This follows from the lemma below.

Lemma 10 *Let s be an odd number and $s = o(n)$ and $\emptyset \ne X \subseteq [n]$. Then the probability that the intersection of X with a random subset S of size s is odd is at least $\frac{\gamma s}{n}$ for a constant $\gamma > 0$ and provided that n is sufficiently large.*

Proof. Let us *fix* an S, $|S| = s$, s an odd number, and for $0 < k \le n/2$, let X be *a random set* of size k. Think of X as the result of the random process of choosing distinct elements $x_1, \cdots, x_n \in [n]$. Consider $|S \cap \{x_1, \cdots, x_{k-1}\}|$ and $|([n] \setminus S) \cap \{x_1, \cdots, x_{k-1}\}|$. The distribution of these random variables are sharply concentrated around the values $\frac{s(k-1)}{n}$, respectively $\frac{(n-s)(k-1)}{n}$. Hence the probability that $|S \setminus \{x_1, \cdots, x_{k-1}\}| \geq s/3$ and $|([n] \setminus S) \setminus \{x_1, \cdots, x_{k-1}\}| \geq n/3$ is bigger than some constant $\delta > 0$. Suppose this event happens. If $|S \cap \{x_1, \cdots, x_{k-1}\}|$ is odd, then the probability that $|S \cap X|$ is odd is at least $1/3$. Otherwise the probability is at least $s/(3n)$. Thus the probability that $|S \cap X|$ is odd is at least $\delta s/(3n)$.

If $k > n/2$ think of X as the results of the random process of choosing distinct elements in its complement and then argue in the same way.

■

The matrix of the Markov process is

$$\frac{1}{2}\binom{n}{s}^{-1} A + \frac{1}{2} I,$$

where A is the adjacency matrix of G_s and I is the $2^n \times 2^n$ identity matrix. Hence the second largest eigenvalue of the Markov process is $1 - \frac{s}{2n}$. Thus the distance from the uniform distribution is bounded by $c^{D \cdot s/(2n)} \approx c^{\beta n^{\epsilon}/2}$, for some constant $c < 1$. ■

For sake of simplicity, we are using vectors with exactly s ones here instead of the vectors with at most s ones used before. We think that the difference is not essential (except that now we have to talk about s being odd). Let us see what is the connection to lower bounds on the size of circuits. The conjecture about the generator can be restated as follows (we assume the same parameters as in Proposition 9).

> *A random k-dimensional subspace of a D-dimensional space generated by vectors with s ones is not computationally distinguishable from a random space of dimension k. We assume that the spaces are given by randomly chosen bases.*

Thus if we had a simple test that would distinguish the outputs of the generator from random spaces, then, probably, we could use this test to prove a lower bound.

Notice also that the spaces are described very compactly, so we may not be able to test properties such as the minimal distance. Hence the conjecture is more likely than if the spaces were given by the lists of all vectors. Therefore, it is also possible that the generator is a pseudorandom generator and still there exists a "natural" lower bound proof.

5 Good codes and explicit expanders

In this section, we prove a connection between good codes with *highly* sparse parity check matrices and expander graphs. In the context of parity check matrices, we will informally apply the term "highly sparse" for matrices with bounded number of non-zero entries in each row and column. Codes with highly sparse parity check matrices have been studied in coding theory [8]. One of their advantages is that such codes can be efficiently decoded at least to some fraction of the minimal distance. We shall show that with suitable parameters highly sparse parity check matrices of good codes can be used to construct expanders.

An (n, m, k)-*concentrator* is a bipartite graph $E \subseteq I \times O$ such that $|I| = n, |O| = m$ and for every subset $X \subseteq I$ of size k, there exists a matching from X into O with k edges (by Hall's theorem this is equivalent to the condition that every subset $X \subseteq I$ of size at most k has at least $|X|$ neighbors in O); to exclude trivial cases, we assume $m < n$. We shall call I inputs and O outputs.

An (n,α)-*expander* is a symmetric graph on n vertices such that every subset X of vertices of size at most $n/2$ has at least $\alpha|X|$ neighbors. Usually we assume that we have an infinite family of such graphs with the *expansion ratio* α being a constant greater than 1. The most important families of expanders are those which have also a constant bound on the degree of vertices. If an expander is a regular graph, then there is a gap, which is a positive constant, between the two largest absolute values of eigenvalues of the graph; this is often taken as the definition of expanders (see [2]). The main idea of the connection between expanders and parity check matrices is in the following two lemmas.

Proposition 11 *Let A be an $m \times n$ parity check matrix of a linear code with the code length n and minimal distance d. Further assume that each row and column of A has at most s 1's. Let $E \subseteq I \times O$, $|I| = n, |O| = m$ be the bipartite graph whose edges correspond to the nonzero entries of the matrix. Then E is an $(n, m, d-1)$-concentrator with the degrees of the inputs and outputs bounded by s.*

Proof. For $\ell \leq d-1$, every ℓ columns of the parity check matrix are linearly independent. Hence the submatrix determined by these columns must have at least ℓ independent, hence nonzero, rows. ∎

Proposition 12 *Given an $(n, m, n/2)$-concentrator E with the degrees bounded by s. Then one can explicitly construct*

1. *an (n,α)-expander, with $\alpha \geq 1 + \lceil n/(n-m) \rceil^{-1}$ in which the degrees of outputs are at most $s + 1 + \max(s, \lceil n/(n-m) \rceil)$;*
2. *an (n,α)-expander, with $\alpha \geq 1 + \lceil n/(n-m) \rceil^{-1}$ in which the degrees of outputs are at most $4s$.*

Proof. 1. Given an (n, m, k)-concentrator E with inputs I and outputs O, add to it a set O' of additional outputs, $|O'| = n - m$, and connect every input vertex to some vertex in O' so that at most $\lceil n/(n-m) \rceil$ vertices in I are connected to the same vertex in O'. One can easily check that in the resulting bipartite graph every subset $X \subseteq I$ such that $|X| \leq n/2$ has at least $\alpha|X| \geq 1 + \lceil n/(n-m) \rceil^{-1}$ neighbors in $O \cup O'$. Thus if we take an arbitrary matching between I and $O \cup O'$ and identify the pairs of vertices we get an expander with the parameters stated in the proposition.

2. In the second construction, we also add $n - m$ vertices O'. Suppose w.l.o.g. that $O \cup O' = \{1, 2, \ldots, n\}$. Make another copy of E by shifting its output vertices O by $n-m$. Thus the expander consists of two partially overlapping copies of the concentrator. Let us prove the lower bound on the expansion ratio. Let $X \subseteq I$, $|X| \leq k$. Let Y, resp. Y' be the neighbors of X in the original copy of E, resp. in the shifted copy. Since $|Y| \geq |X|$, we need only to estimate $\ell = |Y' \setminus Y|$. Notice that if we shift O by $n - m$ $\lceil n/(n-m) \rceil$-times, we get disjoint copies. After every shift we get at most ℓ new neighbors of X and the neighbors in the last copy are disjoint with Y. Hence $\ell \geq |X|/(\lceil n/(n-m) \rceil)$. The rest of the proof is the same as in 1. ∎

Since there are no good binary codes with the minimal distance $\geq n/2$, the above estimates are not sufficient for constructing expanders from binary codes. However, the bound in Proposition 11 can be improved if we slightly weaken the condition on concentrators. Then one can get, at least theoretically, an expander also from binary codes.

Proposition 13 *Let M be an $m \times n$ parity check matrix of a linear code $[n, n-m, d]$ over a finite field with q elements, and suppose $k < d/(1-q^{-1})$. Then among every k columns of M there are at least*

$$k - \log_q \frac{d}{d-(1-q^{-1})k}$$

independent columns. Consequently, in the associated bipartite graph of nonzero elements every set of k inputs has at least so many neighbors.

Proof. Let k columns be given, let M' be the corresponding submatrix of M. Then M' is a parity check matrix of a code $[k, \ell, d]$. According to the Plotkin bound (see eg. [14]),

$$\ell \leq \log_q \frac{d}{d-(1-q^{-1})k}.$$

The matrix M' has $k-\ell$ independent rows, from which we infer that it has $k-\ell \geq k-\log_q \frac{d}{d-(1-q^{-1})k}$ independent columns. ∎

In particular, if we take the two-element field and $k = 2d-1$, then every set of k elements of input vertices has at least $k - \log_2 d - 1$ neighbors. For large k the additive term $-\log_2 d - 1$ is negligible, thus if $2d > n/2$ we can apply Propositions 11 and 12 with a minor modification.

Except for codes constructed from expanders, we do not know of any families of explicit good codes with sparse parity check matrices. It is, however, possible that some algebraic-geometric codes have this property. Construction of these codes are known for finite fields GF_{q^2}. Such a code C is determined by a curve, n places of degree on the curve, and a divisor G whose support is disjoint with the places. If the degree of the divisor satisfies $2g-2 < degG < n$, where g is the genus of the curve, then the code has parameters $[n, k, d]$ satisfying

$$k = degG + 1 - g, \quad \text{and} \quad d \geq n - degG.$$

The dual code $C^{\perp}$ is also an algebraic-geometric code on the same curve and the same places and its parameters $[n, k', d']$ satisfy

$$k' = n - k = n + g - 1 - degG, \quad \text{and} \quad d' \geq degG - 2g + 2.$$

Explicit families of curves have been constructed such that $n/g \to q-1$. If we choose $degG = 2g-2$, then the lower bound on d' becomes trivial, while

$$k = (\frac{1}{q-1} + o(1))n \quad \text{and} \quad d \geq (1 - \frac{2}{q-1} - o(1))n.$$

Thus if $q \geq 5$ we get a family of good codes with $d \geq (\frac{1}{2} - o(1))n$, which can be used to construct a family of expanders. The fact that we do not have any nontrivial lower bound on the minimal distance of the dual code, gives some hope that the parity check matrices of these codes may be sparse. If that were true, we would get sparse expanders from these codes. Garcia and Stichtenoth found very explicit constructions of such codes [4, 5]. Unfortunately, these codes are still not sufficiently well understood, in particular, even the minimal distance has not been computed precisely. Thus it is an open problem whether or not one can get good codes with sparse parity check matrices in this way.

For another relation between good codes and expanders see [9].

6 Conclusions

The most natural question suggested by this approach is the following. Can codes with sparse parity check matrices achieve the same asymptotic tradeoff between the rate and the minimal distance as general codes? If one could prove that the answer is no with sparsity being n^{ε}, and if one could construct explicitly codes that have better tradeoff than sparse ones, then one would solve the old problem of Valiant. It would also be interesting to solve this problem for constant sparsity. This would only give a lower bound for a weaker type of circuits (series parallel circuits), but it would be even more interesting from the point of view of coding theory.

References

[1] M. Alekhnovich, More on average case vs. approximation complexity. *Proc. of the 44rd IEEE Symposium on Foundations of Computer Science,* 2003, pp.298-307.

[2] N. Alon, J. Spencer, and P. Erdös, The Probabilistic Method. John Wiley & Sons, Inc. 1992.

[3] J. Friedman, A note on matrix rigidity. *Combinatorica,* 13(2), 1993, pp.235-239.

[4] A. Garcia and H. Stichtenoth, A tower of Artin–Schierer extensions of function fields attaining the Drinfeld–Vlăduţ bound. *Invent. Math.* 121, 1995, pp.211-222.

[5] A. Garcia and H. Stichtenoth, On the asymptotic behavior of some towers of function fields over finite fields. *J. of Number Theory* 61(2), 1996, pp.248-273.

[6] D.Yu. Grigoriev, Using the notion of separability and independence for proving lower bounds on circuit complexity. *Notes of the Leningrad branch of the Steklov Institute* 60, 1976, pp.38-48.

[7] S.V. Lokam, On the rigidity of Vandermonde matrices. *Theoretical Computer Science* 237(1-2), 2000, pp.477-483.

[8] D.J.C. MacKay, Good error-correcting codes based on very sparse matrices. *IEEE Transaction on Information Theory* 45(2), 1999, pp.399-401.

[9] R. Meshulam and A. Wigderson, Expanders from symmetric codes. Proc. ACM Symposium on Theory of Computing, 2002, pp.669-677.

[10] M. Morgenstern, Note on a lower bound on the linear complexity of the Fast Fourier Transform. *J. of ACM* 20(2), 1973, pp. 305-306.

[11] A.A. Razborov, and S. Rudich, Natural proofs. In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing*, 1994, pp. 204-213.

[12] D.A. Spielman, V. Stetmann, and M.A. Shokrollahi, A remark on matrix rigidity. *Information Processing Letters,* 64(6), 1997, pp.283-285.

[13] L.G. Valiant, Graph–theoretic arguments in low–level complexity, In *Proceedings of the 6th Symposium on Mathematical Foundations of Computer Science*, Springer–Verlag, Lecture Notes in Computer Science, vol. 53, 1977, pp. 162-176.

[14] J.H. van Lint, Introduction to Coding Theory. Springer-Verlag 1992.