$$a + a' = b + b' = (c_1 + c_1')a^1 + (c_2 + c_2')a^2 + \cdots + (c_k + c_k')a^k.$$

Since vectors $c$ and $c'$ differ in at least three coordinates, we have on the right-hand side the sum of at least three vectors, say $a^{i_1} + \cdots + a^{i_l}$, with $l \geq 3$. But then in the equation (17.5) we can replace these three (or more) vectors $a^{i_1}, \ldots, a^{i_l}$ by two vectors $a, a'$, which contradicts the minimality of $k$. $\square$

The same argument also implies that no two distinct vectors $c, c' \in C$ can lead to one and the same vector $b \in B$, that is, $c \neq c' \in C$ implies $\sum_i c_i a^i \neq \sum_i c_i' a^i$. This means that $|B| = |C|$.

This, together with Claim 17.15, implies

$$|A| \cdot |C| = |A| \cdot |B| = \sum_{b \in B} |b + A| = |\bigcup_{b \in B} (b + A)| \leq |\text{span}\, A|.$$

Hence, $\log_2 |C| \leq \log_2(1/\alpha)$ which, together with Claim 17.14, yields the desired upper bound (17.4) on $k$. $\square$

## 17.6 Expander codes

If $C \subseteq \{0,1\}^n$ is a linear code with a $k \times n$ generator matrix $G$, then the encoding of messages $w \in \{0,1\}^k$ is very easy: just encode $w$ by the codeword $x = w^\top G$. However, the decoding—that is, given a vector $y \in \{0,1\}^n$ find a codeword $x \in C$ closest to $y$—is in general linear codes a very difficult problem (it is "NP-hard").

We now show how using expander graphs one can construct linear codes for which decoding is almost trivial—it can be done in linear time! Moreover, if the expansion of the graph is good enough then the resulting codes achieve very good rate $(\log_2 |C|)/n$ and minimal distance (both these parameters are then absolute positive constants).

Let $G = (L \cup R, E)$ be a bipartite graph with $|L| = n$, $|R| = m$ and $E \subseteq L \times R$. Each such graph defines a linear code $C \subseteq \{0,1\}^n$ as follows. Associate with each vertex $u \in L$ a boolean variable $x_u$. Given a vector $x \in \{0,1\}^n$, say that a vertex $v \in R$ is *satisfied* by this vector if

$$\sum_{u \in \Gamma(v)} x_u \bmod 2 = 0\,,$$

where $\Gamma(v) = \{u \in L : uv \in E\}$ is the set of all neighbors of $v$ on the left side (see Fig. 17.1). The code defined by the graph $G$ is the set of vectors

$$C = \{x \in \{0,1\}^n : \text{ all vertices in } R \text{ are satisfied by } x\}\,.$$

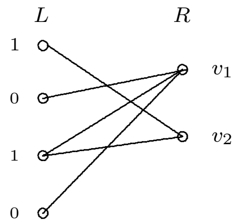


**Fig. 17.1** Vertex $v_2$ is satisfied whereas $v_1$ is not satisfied by the vector $x = (1010)$.

That is, $C$ is just the set of all solutions of $m$ linear equations in $n$ variables. Therefore, $C$ is linear and $|C| \geq 2^{n-m}$.

Let $\mathrm{dist}(C)$ be the minimal Hamming distance between two different vectors in $C$. A graph $G = (L \cup R, E)$ is *left $d$-regular* if each vertex in $L$ has degree $d$. Such a graph is an $(\alpha, c)$-*expander* if every subset $I \subseteq L$ with $|I| \leq \alpha n$ has $|\Gamma(I)| > c|I|$ neighbors on the right side.

**Lemma 17.16.** *If $C \subseteq \{0,1\}^n$ is a code of a left $d$-regular $(\alpha, c)$-expander with $c > d/2$, then*

$$\mathrm{dist}(C) > \alpha n\,.$$

*Proof.* Assume that $\mathrm{dist}(C) \leq \alpha n$. Then $C$ must contain a vector $x$ with at most $\alpha n$ ones. Hence, if we take the set $I = \{u \in L : x_u = 1\}$, then $|I| \leq \mathrm{dist}(C) \leq \alpha n$. Since $G$ is an $(\alpha, d/2)$-expander, this implies $|\Gamma(I)| > d|I|/2$.

We claim that there must exist a vertex $v_0 \in \Gamma(I)$ with *exactly one* neighbor in $I$, that is, $|\Gamma(v_0) \cap I| = 1$. Indeed, otherwise every vertex $v \in \Gamma(I)$ would have at least two neighbors in $I$. Therefore the number of edges leaving $I$ would be at least $2 \cdot \Gamma(I) > 2 \cdot (d|I|/2) = d|I|$, contradicting the left $d$-regularity of $G$.

Since $x_u = 0$ for all $u \notin I$, this implies that *exactly one* of the bits $x_u$ of $x$ with $u \in \Gamma(v_0)$ is equal to 1. So, $\sum_{u \in \Gamma(v_0)} x_u = 1$, and the vertex $v_0$ cannot be satisfied by the vector $x$, a contradiction with $x \in C$. ☐

By Lemma 17.16, expander codes can correct relatively many errors, up to $\alpha n/2$. Much more important, however, is that the decoding algorithm for such codes is very efficient. The decoding problem is the following one: given a vector $y \in \{0,1\}^n$ of Hamming distance $\leq \alpha n/2$ from some (unknown) codeword $x \in C$, find this codeword $x$. The decoding algorithm for expander codes is amazingly simple:

> *While there exists a variable such that most of its neighbors are not satisfied by the current vector, flip it.*

**Lemma 17.17** (Sipser–Spielman 1996)**.** *If $C$ is a code of a left $d$-regular $(\alpha, c)$-expander with $c > \frac{3}{4}d$, then the algorithm solves the decoding problem in a linear number of steps.*

*Proof.* Let $y \in \{0,1\}^n$ be a vector of Hamming distance $\leq \alpha n/2$ from some (unknown) codeword $x \in C$. Our goal is to find this codeword $x$. Let

$$I = \{u \in L \,:\, y_u \neq x_u\}$$

be the set of errors in $y$. If $I$ is empty, we are done. Otherwise, assume that $|I| \leq \alpha n$. We need this assumption to guarantee the expansion, and we will prove later that this assumption holds throughout the running of the algorithm.

Partition the set $\Gamma(I) = S \cup U$ into the set $S$ of neighbors satisfied by $y$ and the set $U$ of neighbors not satisfied by $y$. Since $c > 3d/4$, we have that

$$|U| + |S| = |\Gamma(I)| > \tfrac{3}{4}d|I|\,. \tag{17.6}$$

Now, count the edges between $I$ and $\Gamma(I)$. At least $|U|$ of these edges must leave $U$. Moreover, at least $2|S|$ of them must leave $S$ because every vertex $v \in S$ must have at least two neighbors in $I$: If $v$ had only one such neighbor, then $y$ would not satisfy the vertex $v$ since $y \neq x$, $x$ satisfies $v$ and $y$ coincides with $x$ outside $I$. Since the total number of edges between $I$ and $\Gamma(I)$ is $d|I|$, this implies $|U| + 2|S| \leq d|I|$. Combining this with (17.6) we get that

$$d|I| - |U| \geq 2|S| > 2\left(\tfrac{3}{4}d|I| - |U|\right)$$

and therefore

$$|U| > \tfrac{1}{2}d|I|\,. \tag{17.7}$$

So, more than $d|I|/2$ neighbors of the $|I|$ vertices in $I$ are unsatisfied. Therefore there is a variable in $I$ that has more than $d/2$ unsatisfied neighbors. We have therefore shown the following claim:

> *If $I \neq \emptyset$ and $|I| \leq \alpha n$ then there is a variable with $> d/2$ unsatisfied neighbors.*

This implies that as long as there are errors *and* $|I| \leq \alpha n$ holds, some variable will be flipped by the algorithm. Since we flip a vertex with more unsatisfied neighbors than satisfied ones, $|U|$ decreases with every step (flipping $x_u$ can only affect the satisfiability of neighbors of $u$). We deduce that if the distance $|I|$ of the actual vector $y$ from $x$ does not exceed $\alpha n/2$ throughout the run of the algorithm, then the algorithm will halt with the codeword $x$ after a linear number of iterations.

To show that $|I|$ can never exceed $\alpha n$, recall that $|I| \leq \alpha n/2$, and hence,

$$|U| \leq |\Gamma(I)| \leq \tfrac{1}{2}\alpha dn \tag{17.8}$$

hold in the beginning. Moreover, $|U|$ decreases after each iteration. Hence, if at some step we had that $|I| > \alpha n$, then (17.7) would imply $|U| > \alpha dn/2$, contradicting (17.8). ☐

In general, every linear code $C \subseteq \{0,1\}^n$ is defined by its *parity-check* matrix $H$ such that $x \in C$ iff $Hx = \mathbf{0}$. Note that, if $C$ is a code defined by a bipartite graph $G$, then $H$ is just the transpose of the adjacency matrix of $G$. If $G$ is left $d$-regular, then every row of $H$ has exactly $d$ ones. If $G$ is an $(\alpha, c)$-expander, then every subset $I$ of $|I| \leq \alpha n$ columns of $H$ has ones in at least $c|I|$ rows. The decoding algorithm above is, given a vector $y \in \{0,1\}^n$ such that $Hy \neq \mathbf{0}$, to flip its $i$-th bit provided that vector $H(y \oplus e_i)$ has fewer ones than vector $Hy$.

## 17.7 Expansion of random graphs

Explicit constructions of bipartite left $d$-regular $(\alpha, c)$-expanders with $\alpha = \Omega(1)$ and $c > 3d/4$ are known. These constructions are however too involved to be presented here. Instead of that, we will show that *random* bipartite left-regular graphs have good expansion properties.

Let $d \geq 3$ be a constant. We construct a random bipartite left $d$-regular $n \times n$ graph $G_{n,d} = (L \cup R, E)$ as follows: For each vertex $u \in L$ choose its $d$ neighbors independently at random, each with the same probability $1/n$. The graph obtained may have multi-edges, that is, some pairs of vertices may be joined by several edges.

**Theorem 17.18.** *For every constant $d \geq 3$, there is a constant $\alpha > 0$ such that for all sufficiently large $n$, the graph $G_{n,d}$ is an $(\alpha, d-2)$ expander with probability at least $1/2$.*

*Proof.* Set (with foresight) $\alpha := 1/(\mathrm{e}^3 d^4)$. Fix any $s \leq \alpha n$, and take any set $S \subseteq L$ of size $|S| = s$. We want to upper bound the probability that $S$ does not expand by $d-2$. This means that the $ds$ neighbors (including multiplicities) of the vertices in $S$ hit fewer than $(d-2)s$ distinct vertices on the right side, that is, some $2s$ of these $ds$ neighbors land on previously picked vertices. Each neighbor lands on a previously picked vertex with probability at most $ds/n$, so

$$\Pr\left[S \text{ does not expand by } (d-2)\right] \leq \binom{ds}{2s}\left(\frac{ds}{n}\right)^{2s}.$$

By the union bound, the probability that at least one subset $S$ of size $s$ does not expand by $(d-2)$ is at most

$$\binom{n}{s}\binom{ds}{2s}\left(\frac{ds}{n}\right)^{2s} \leq \left(\frac{\mathrm{e}n}{s}\right)^{s}\left(\frac{\mathrm{e}ds}{2s}\right)^{2s}\left(\frac{ds}{n}\right)^{2s} \leq \left(\frac{\mathrm{e}^3 d^4}{4n}\right)^{s} \leq \left(\frac{1}{4}\right)^{s},$$

by the choice of $\alpha$. Thus, the probability that some set $S$ of size $|S| \leq \alpha n$ does not expand by $(d-2)$ does not exceed