

# Locally Testable Good Codes

Alex Lubotzky

Joint work with:  
Irit Dinur, Shai Evra, Ron Livne and Shahar Mozes



# Error Correcting Codes (ECC)

Let  $\mathbb{F}_q$  =field of order  $q$ , usually  $q = 2$ .

An  $(n, k, d)_q$ -code  $C$  is a subspace of  $\mathbb{F}_q^n$  of dimension  $k$  and distance at least  $d$ , where

$$\begin{aligned} \text{dist}(C) &= \min\{(\text{Hamming}) \text{ dist } (\alpha, \beta) \mid \alpha \neq \beta \in C\} \\ &= \min\{(\text{Hamming}) \text{ weight } (\alpha) \mid 0 \neq \alpha \in C\} \end{aligned}$$

We are interested in the case  $n \rightarrow \infty$ .

A code (or more precisely a family of codes) is **good** (in the sense of Shannon) if  $\exists \varepsilon > 0$  s.t.  $k \geq \varepsilon n$  **and**  $d \geq \varepsilon n$ .

The standard terminology:

$$\text{rate} = \rho(C) = \frac{k}{n} \quad (\text{normalized}) \quad \text{distance} = \delta(C) = \frac{d}{n}$$

and  $C$  is good if  $\rho(C)$  and  $\delta(C)$  are bounded below by some  $\varepsilon$ .

(Say: “**Constant rate and constant distance**”)

- Ex: 1) **Shannon** 40's: Random codes defined by random constraint  
2) many explicit examples

The code(s) is called **LDPC** (= low density parity check) if  $C$  is defined by equations (= linear functionals = constraints) with bounded support.

Equivalently

$$C^\perp = \{\beta \in \mathbb{F}_q^n \mid \langle \alpha, \beta \rangle = 0 \quad \forall \alpha \in C\}$$

is spanned by vectors of bounded support.

$$\langle \alpha, \beta \rangle = \sum_{i=1}^n a_i b_i.$$

Ex: **Gallager**: 60's Random code defined by random equations of bounded support

# Explicit construction of LDPC good codes

A breakthrough by Sipser & Spielman in the 90's was the explicit construction of "Expander codes" which are LDPC good codes

Let  $X = (V, E)$  be an  $r$ -regular graph. Its e.v.'s (the e.v.'s of the adjacency matrix) are in  $[-r, r]$ . The largest is always  $r$  and the second one  $\lambda(X)$  (with  $\lambda(X) < r$  iff  $X$  is connected).

**Def:**  $X$  is called  $\lambda$ -expander for some  $\lambda < r$ ,

$$\text{if } \lambda(X) \leq \lambda$$

Again: we want  $r$  &  $\lambda$  fixed and

$$|X| \rightarrow \infty$$

**Def** :  $X$  is called **Ramanujan graph** if  $\lambda(X) < 2\sqrt{r-1}$

There are explicit constructions of Ramanujan graphs:

For  $r = p + 1$ ,  $p$  prime (**Lubotzky-Phillips-Sarnak**, see also **Margulis**)

$$r = p^e + 1 \quad (\text{Morgenstern})$$

# The S-S construction

$X = (V, E)$ ,  $r$ -regular  $\lambda$ -expander graph on  $m$  vertices  
 $r$  fixed,  $\lambda$  fixed,  $m \rightarrow \infty$

Let  $C_0 \leq \mathbb{F}_2^r$  a subspace (“the small code”).

Label the  $r$  edges around each vertex  $v$  by  $\{1, \dots, r\}$  (no consistency assumption).

Let

$$W = \{f : E \rightarrow \mathbb{F}_2\}$$

so:

$$n := \dim(W) = \frac{mr}{2}$$

and “the large code”

$$C = \{f \in W \mid f|_{\text{link}(v)} \in C_0, \forall v \in V\}$$

i.e. the local view of  $f$  at every vertex is in the small code

## Theorem Sipser-Spielman

If

(a)  $\dim C_0 > \frac{r}{2}$

(b)  $\text{distance}(C_0) > \lambda(X)$

then  $C$  is an LDPC good code.

Sketch of proof:

(1) LDPC - clear !!

(2)

$$\begin{aligned} \dim(C) &\geq n - m(r - \dim C_0) = \\ \frac{mr}{2} - mr + m \dim C_0 &= m(\dim C_0 - \frac{r}{2}) = \\ &= \underbrace{\frac{2}{r}(\dim C_0 - \frac{r}{2})}_{\text{constant.}} n \end{aligned}$$



(3) Let  $0 \neq f \in C$  and let  $F = \text{support}(f)$  and  $G$  the subgraph of  $X$  spanned by  $F$ .

**Lemma (Alon-Chung)** If  $G \subset X$  a subgraph with average degree  $l > \lambda(X)$ , then  $|G| \geq \frac{1}{r}(l - \lambda(X))|X|$ .

Here the degree of every vertex of  $G$  is  $\geq$  i.e., distance  $(C_0) > \lambda(X)$  hence  $|G|$  is linear in  $n = |X|$ .

Now use Ramanujan  $r$ -regular graphs and some fixed code  $C_0 \leq \mathbb{F}_2^r$   
(Even if chosen by random, it is for a fixed(!)  $r$ .)

But actually there are explicit).

Q.E.D.

## Locally testable codes (LTC)

A code  $C \subseteq \mathbb{F}_2^n$  is a  $(q, \varepsilon)$ -**locally testable code** if there exists a random algorithm  $A$  which for every  $\alpha \in \mathbb{F}_2^n$ ,  $A$  reads only  $q$  random bits of  $\alpha$  and answers YES if  $\alpha \in C$  and answers NO with probability

$$\Pr(A \text{ answers NO}) \geq \varepsilon \cdot \frac{1}{n} \text{distance}(\alpha, C)$$

i.e.  $A$  can decide, by reading only a few bits of  $\alpha$  if  $\alpha$  is in  $C$  or far away from it.

### Example:

Let  $W = \{f : \mathbb{F}_2^m \rightarrow \mathbb{F}_2 \mid f \text{ a function}\}$

so:  $n = \dim W = 2^m$

Let  $C =$  the linear functionals in  $W$

so  $\dim(C) = m = \log_2(n)$ .

Let  $A$  be the test: choose random  $x, y \in \mathbb{F}_2^n$

and answer YES if  $f(x + y) = f(x) + f(y)$

### Theorem (Blum-Luby-Rubinfeld [BLR], 1993)

$\text{Prob}(A \text{ answers NO}) \geq \frac{2}{27} \left(\frac{1}{n} \text{distance}(f, C)\right)$

So  $X$  is  $(3, \frac{2}{27})$ -LTC code.

But far away from being good.

## Other LTC codes:

- Reed-Solomon
- Reed-Muller
- **The tensor code** Let  $C_1 \leq \mathbb{F}_2^{n_1}$  be a “smooth code” and  $C_2 \leq \mathbb{F}_2^{n_2}$ . The tensor code  $C_1 \otimes C_2 \leq \mathbb{F}_2^{n_1} \otimes \mathbb{F}_2^{n_2} = M_{n_1, n_2}(\mathbb{F}_2)$  is the code of the  $n_1 \times n_2$  matrices over  $\mathbb{F}_2$  such that each column is in  $C_1$  and each row in  $C_2$  (so  $n = n_1 n_2$  and  $k = (\dim C_1) \cdot (\dim C_2)$ ).

The algorithm  $A$  chooses a random row or column and check.

If  $C_1$  &  $C_2$  good, so is  $C_1 \otimes C_2$  with  $q = \sqrt{n}$  (say  $n_1 = n_2$ ).

## A long standing problem

Are there good LTC with constant number of queries, i.e.,  $q$  is constant independent of  $n$ ?

## The $C^3$ problem:

constant rate, constant distance and constant number of queries.

## Remark

Random (LDPC) **are not!** Check !

So find a needle in a haystack rather than hay in a haystack, if the needle exists ...

Explicit construction of LTC good codes (with  $q$  constant).

Remark: Panteleev and Kalachev have independently provided a construction of LTC good codes.

**Our construction** Let  $G$  be a finite group with two symmetric sets of generators  $A$  and  $B$ .

Assume:

$$1 \notin A \cup B$$

and

$$\text{(TNC)} \quad \forall a \in A, b \in B, g \in G, g^{-1}ag \neq b$$

i.e., no element of  $A$  is conjugate to an element of  $B$  and for simplicity assume also  $|A| = |B| = r$  and both  $\text{Cay}(G; A)$  and  $\text{Cay}(G; B)$  - the Cayley graphs are Ramanujan,

i.e.

$$\lambda(\text{Cay}(G; A)), \lambda(\text{Cay}(G; B)) < 2\sqrt{r-1}$$

Define the **Left/Right Cayley Complex**  $Cay(A; G; B)$  by letting  $A$  act on  $G$  from the left and  $B$  from the right, giving rise to the following squares:

$$\begin{array}{ccc}
 gb & \xrightarrow{a} & agb \\
 \uparrow b & & \uparrow b \\
 g & \xrightarrow{a} & ag
 \end{array}$$

**Remark** (TNC) ensures that all 4 vertices are different.

There are  $\frac{1}{4}|G| \cdot |A| |B|$  such squares.

(We take  $|A|$  &  $|B|$  fixed and  $|G| \rightarrow \infty$ ).

Let  $C_A \leq \mathbb{F}_2^A$  a smooth code,  $C_B \leq \mathbb{F}_2^B$  and  $C_0 = C_A \otimes C_B$

**Note** The squares containing a vertex  $g$  are in 1 – 1 correspondence with  $A \times B$ .

We can define:

$$\begin{aligned} W &= \{f : \{\text{squares}\} \rightarrow \mathbb{F}_2\} \\ C &= \{f \in W \mid f|_{\text{link}(g)} \in C_0, \forall g \in G\} \\ &= \left\{ f \in W \mid \begin{array}{l} \forall \text{ edge } e \text{ of type } A, f|_{\text{link}(e)} \in C_B \\ \forall \text{ edge of type } B, f|_{\text{link}(e)} \in C_a \end{array} \right\} \end{aligned}$$

### Theorem

Under suitable assumption on  $C_A, C_B$  &  $r$  (large by fixed!)

This is a good LTC code.

The algorithm: Choose a random edge and check only  $r$  bits!