Electronic Colloquium on Computational Complexity, Report No. 7 (2009)

Tensor Products of Weakly Smooth Codes are Robust

Eli Ben-Sasson Michael Viderman
Computer Science Department
Technion — Israel Institute of Technology
Haifa, 32000, Israel.
{eli, viderman}@cs.technion.ac.il

January 9, 2009

Abstract

We continue the study of *robust* tensor codes and expand the class of base codes that can be used as a starting point for the construction of locally testable codes via robust two-wise tensor products. In particular, we show that all unique-neighbor expander codes and all locally correctable codes, when tensored with any other good-distance code, are robust and hence can be used to construct locally testable codes. Previous works by [2] required stronger expansion properties to obtain locally testable codes.

Our proofs follow by defining the notion of *weakly smooth* codes that generalize the *smooth* codes of [2]. We show that weakly smooth codes are sufficient for constructing robust tensor codes. Using the weaker definition, we are able to expand the family of base codes to include the aforementioned ones.

1 Introduction

A linear code over a finite field F is a linear subspace $C \subseteq F^n$. A code is *locally testable* if given a word $x \in F^n$ one can verify whether $x \in C$ by reading only a few (randomly chosen) symbols from x. More precisely such a code has a *tester*, which is a randomized algorithm with oracle access to the received word x. The tester reads at most q symbols from x and based on this "local view" decides if $x \in C$ or not. It should accept codewords with probability one, and reject words that are "far" (in Hamming distance) with "noticeable" probability.

Locally Testable Codes (LTCs) were first explicitly studied by Goldreich and Sudan [9] and since then a few constructions of LTCs were suggested (See [8] for an extensive survey of those constructions). All known efficient constructions of LTCs, i.e. that obtain subexponential rate, rely on some form of "composition" of two (or more) codes. One of the simplest ways to compose codes for the construction of LTCs is by use of the tensor product, as suggested by Ben-Sasson and Sudan [1]. They introduced the notion of *robust* LTCs: An LTC is called robust if whenever the received word is far from the code, then with noticeable probability the local view of the tester

[0]Research supported in part by a European Community International Reintegration Grant, an Alon Fellowship, and grants by the Israeli Science Foundation (grant number 679/06) and by the US-Israel Binational Science Foundation (grant number 2006104).

ISSN 1433-8092

is *far* from an accepting local view (see robust definition [2]). It was shown in [1] that a code obtained by tensoring three or more codes (i.e. $C_1 \otimes C_2 \otimes C_3$) is robustly testable when the distances of the codes are big enough, and used this result to construct LTCs. Then they considered the tensor product of two codes. Given two linear codes R, C their tensor product $R \otimes C$ consists of all matrices whose rows are codewords of R and whose columns are codewords of C. If R and C are locally testable, we would like $R \otimes C$ to be locally testable. [1] suggested using the following test for the testing the tensor product $R \otimes C$ and asked whether it is robust:

Test for $R \otimes C$: Pick a random row (or column), accept iff it belongs to R (or C).

Valiant [3] showed a surprising example of two linear codes R and C for which the test above is not robust, by exhibiting a word x that is far from $R \otimes C$ but such that the rows of x are very close to R and the columns of x are very close to C. Additional examples give a codes whose tensor product with itself is not robust [4] and two good codes (with linear rate) whose tensor product is not robust [7].

Despite these examples Dinur et al. showed in [2] that the above test is robust as long as one of the base codes is *smooth*, according to a definition of the term introduced there (see Definition 5). The family of smooth codes includes locally testable codes and certain codes constructed from expander graphs with very good expansion properties. In this work we continue this line of research and enlarge the family of base codes that result in robust tensor codes and do this by working with a weaker definition of smoothness (Definition 4). Using the weaker definition, we still manage to get pretty much the same results as in [2] and do this using the same proof strategy as there. However, our weaker definition allows us to argue — in what we view as the main technical contributions of this paper (Sections 6 and 7) — that a larger family of codes is suitable for forming robust tensor codes. One notable example is that our definition allows us to argue that any expander code with unique-neighbor expansion (i.e., with expansion parameter $\gamma < 1/2$ as per Definition 3) is also weakly smooth, hence robust. We stress that unique-neighbor expansion is the minimal requirement in terms of expansion needed to argue an expander code has good (i.e., constant relative) distance, so our our work shows all "combinatorially good" expander codes[1] are robust. In comparison, the work of [2] required stronger expansion parameters ($\gamma < 1/4$) of the kind needed to ensure an expander code is not merely good in terms of its distance, but can also be decoded in linear time [10].

Another family of codes shown here to be robust under two-wise tensor products is the family of locally correctable codes (LCCs), see Definition 7.

We end this section by pointing out that recently, tensor codes have played a role in the combinatorial construction by Meir [6] of quasilinear length locally testable codes. Better base codes may result in LTCs with improved rate, hence the importance in broadening the class of base codes that can be used to construct robust tensor codes.

Organization of the paper.

In the following section we provide the now-standard definitions regarding robust tensor codes. In Section 3 We formally define weakly smooth codes and state our main results. In Section 4 We

[1] Clearly, there exist non-unique-neighbor expander codes with good distance. However, the distance of these codes cannot be argued merely using the combinatorial structure of the underlying parity check matrix.

prove weakly smooth codes are robust. Section 5 shows the smooth codes of [2] are also weakly smooth. The last two sections prove that unique-neighbor expander codes, and locally correctable codes, respectively, are weakly smooth.

2 Preliminary Definitions

The definitions appearing here are pretty much standard in the literature on tensor-based LTCs.

Throughout this paper F is a finite field and C, R are linear codes over F. For $c \in C$ let $\mathrm{supp}(c) = \{i | c_i \neq 0\}$ and $\mathrm{wt}(c) = |\mathrm{supp}(c)|$. We define the *distance* between two words $x, y \in F^n$ to be $d(x, y) = |\{i \mid x_i \neq y_i\}|$ and the relative distance to be $\delta(x, y) = \frac{d(x,y)}{n}$. The distance of a code is denoted $d(C)$ and defined to be the minimal value of $d(x, y)$ for two distinct codewords $x, y \in C$. Similarly, the relative distance of the code is denoted $\delta(C) = \frac{d(C)}{n}$. For $x \in F^n$ and $C \subseteq F^n$, let $\delta_C(x) = min_{y \in C}\{\delta(x, y)\}$ denote the relative distance of x from code C. We let $\dim(C)$ denote the dimension of C. The vector inner product between u_1 and u_2 is denoted by $\langle u_1, u_2 \rangle$. For code C let $C^\perp = \{u \in F^n \mid \forall c \in C : \langle u, c \rangle = 0\}$ be its dual code and let $C^\perp_t = \{u \in C^\perp \mid \mathrm{wt}(u) = t\}$. In similar way we define $C^\perp_{<t} = \{u \in C^\perp \mid \mathrm{wt}(u) < t\}$ and $C^\perp_{\leq t} = \{u \in C^\perp \mid \mathrm{wt}(u) \leq t\}$. For $w \in F^n$ and $S \subseteq [n]$ we let $w|_S = (w_{j_1}, w_{j_2}, ..., w_{j_m})$ when $\{j_1, j_2, ..., j_m\} = S$ be the projection of w on subset S. Similarly, we let $C|_S = \{c|_S \mid c \in C\}$ to denote the projection of code C on subset S.

2.1 Tensor Product of Codes

For $x \in F^m$ and $y \in F^n$ we let $x \otimes y$ denote tensor product of x and y (i.e. the $n \times m$ matrix xy^T). Let $R \subseteq F^m$ and $C \subseteq F^n$ be linear codes. We define the tensor product code $R \otimes C$ to be the linear subspace spanned by words $r \otimes c \in F^{n \times m}$ for $r \in R$ and $c \in C$. Some immediate facts:

- The code $R \otimes C$ consists of all $n \times m$ matrices over F whose rows belong to R and whose columns belong to C.
- $\dim(R \otimes C) = \dim(R) \cdot \dim(C)$
- $\delta(R \otimes C) = \delta(R) \cdot \delta(C)$

Let $M \in F^m \otimes F^n$ and let $\delta(M) = \delta_{R \otimes C}(M)$. Let $\delta^{row}(M) = \delta_{R \otimes F^n}(M)$ denote the distance from the space of matrices whose rows are codewords of R. This is expected distance of a random row in x from R. Similarly let $\delta^{col}(M) = \delta_{F^m \otimes C}(M)$.

2.2 Robust Locally Testable Codes

Definition 1 (Robustness)**.** Let M be a candidate codeword for $R \otimes C$. The *robustness* of M is defined as $\rho(M) = (\delta^{row}(M) + \delta^{col}(M))/2$, i.e., it is the average distance of "local views" of the codeword. The code $R \otimes C$ is *robustly testable* if there exists a constant α such that $\frac{\rho(M)}{\delta(M)} \geq \alpha$ for every M.

The robustness of a Tester T is defined as $\rho^T = min_{M \in R \otimes C} \frac{\rho(M)}{\delta_{R \otimes C}(M)}$.

2.3 Low Density Parity Check (LDPC) Codes

The following definition is the natural generalization of a LDPC codes to fields of size > 2.

Definition 2 (LDPC codes)**.** A check graph $([n], [m], E, F)$ is a bipartite graph $([n], [m], E)$ over F for a code $C \subset F^n$ where each edge $e = (i, j) \in E$ is labeled by some $e_{(i,j)} \neq 0 \in F$ and the following holds (let $N(j)$ denote the neighbors of j in the graph):

$$x \in C \Longleftrightarrow \forall j \in [m] \sum_{i \in N(j)} x_i \cdot e_{(i,j)} = 0,$$

where the sum $\sum_{i \in N(j)} x_i \cdot e_{(i,j)}$ is computed over F.

Clearly, any linear code $C \subseteq F$ has a corresponding check graph $([n], [m], E, F)$. Moreover if $C^{\perp} = span(C^{\perp}_{\leq d})$ then without loss of generality every right hand node $j \in [m]$ has degree at most d.

Definition 3 (Expander graphs)**.** Let $c, d \in N$ and let $\gamma, \delta \in (0, 1)$. Define a (c, d)-regular (γ, δ)-expander to be a bipartite graph (L, R, E, F) with vertex sets L, R such that all vertices in L have degree c, and all vertices in R have degree d; and the additional property that every set of vertices $L' \subset L$, such that $|L'| \leq \delta|L|$, has at least $(1 - \gamma)c|L'|$ neighbors.

We say that a code C is an (c, d, γ, δ)-expander code if it has a check graph that is a (c, d)-regular (γ, δ)-expander. It is well-known that if $\gamma < 1/2$ then the graph has *unique-neighbor* expansion, meaning that for every $L' \subset L$ there exists a set of unique neighbors R' on the right such that each member of R' is a neighbor of a *unique* member of L'. Thus, from here on we refer to (γ, δ)-expanders as *unique-neighbor* expanders. The following well-known proposition (the proof of which is included for the sake of completeness) shows that unique-neighbor expansion of G is sufficient to guarantee the code whose check graph is G has large distance.

Proposition 1. *If C is (c, d, γ, δ)-expander code over F and $\gamma < \frac{1}{2}$, then $\delta(C) \geq \delta$.*

Proof. We prove that every non-zero word in C must have weight more than δn. Indeed let (L, R, E, F) be check graph of C that is a (c, d)-regular (γ, δ)-expander. The proposition follows by examining the unique neighbor structure of the graph. Let $x \in C$ be such that $0 < \mathrm{wt}(x) < \delta n$ and $L' = \mathrm{supp}(x) \subseteq L$. But then L' has at least $(1 - \gamma)c|L'| > \frac{c}{2}|L'|$ neighbors in R. At least one of these sees only one element of L', so the check by this element (corresponding dual word) will give $x_i \cdot e_{(i,j)}$ when $x_i \neq 0, e_{(i,j)} \neq 0$ and thus $x_i \cdot e_{(i,j)} \neq 0$, violating the corresponding constraint and contradicting $x \in C$. □

3 Main Results

Our first main result says that codes obtained by the tensor product of a code with constant relative distance and a unique-neighbor expander code is robust:

Theorem 2 (Unique-Neighbor Expander codes are robust)**.** *Let $R \subseteq F^m$ be a code of distance at least $\delta_R > 0$. Let $C \subseteq F^n$ be a (c, d, γ, δ)-expander code for some $c, d \in N, \delta > 0$, and $0 < \gamma < 1/2$. Then,*

$$\rho^T \geq min\{\frac{0.5\delta \cdot \delta_R}{2d^*}, \frac{\delta_R \cdot 0.25\delta}{2}, 1/8\}.$$

Where $d^* < d^k,\ k = (log_{(0.5+\gamma)}0.05) + 1$.

The above theorem extends the result of [2] where a similar result was proved for expanders with the stronger requirement $\gamma < 1/6$. Notice the difference between $\gamma < 1/6$ and unique-neighbor expansion ($\gamma < 1/2$) is qualitative, not merely quantitative. This is because expansion $\gamma < 1/4$ is required to guarantee efficient decoding algorithms, as shown by Sipser and Spielman in [10] whereas $\gamma < 1/2$ is sufficient for claiming the code has large distance, but does not necessarily warrant efficient decoding.

Our next result extends [2] in a different direction by showing that locally correctable codes are also robust. Informally, locally correctable codes allow to recover each entry of a codeword with high probability by reading only a few entries of the codeword even if a large fraction of it is adversely corrupted (see Definition 7).

Theorem 3 (Locally correctable codes are robust). *Let $R \subseteq F^m$ be a code of distance at least $\delta_R > 0$. Let $C \subseteq F^n$ be a (ϵ, δ, q)-locally correctable code with $\epsilon > 0$. Then,*

$$\rho^T \geq min\{\frac{0.5\delta \cdot \delta_R}{2(q+1)}, 1/8\}.$$

To prove both theorems we first define *weakly smooth* codes and prove that the tensor of a weakly smooth code and another code with constant relative distance is robust. Then we show that *smooth* codes are also weakly smooth. Finally we show that all unique-neighbor expander codes (with $\gamma < 1/2$) and all locally correctable codes are weakly smooth, thus obtaining Theorems 2, 3, respectively.

3.1 Weakly Smooth codes

We are coming now to the central definition of the paper, that of a weakly smooth code. This definition allows us to generalize the work of [2] by using pretty much the same proof as there. In particular, in Section 5 we show that every code that is *smooth* according to [2] is also weakly smooth as per Definition 4. Furthermore, using our definition we get robust tensor from a broader family of base codes.

Both the *smooth* codes of [2] and our weakly smooth codes require the code retain large distance even after a portion of its coordinates and constraints have been removed. However there are two subtle differences between the two notions.

1. In the *smooth* codes setting an adversary removes a fraction of *constraints* and then a "Good" player removes a fraction of *indices*. In our Definition 4 both the adversary and the good player remove sets of indices.

2. In the *smooth* codes work with a predefined set of low weight constraints coming from a regular bipartite graph. Our Definition 4 does not assume any graph, nor does it require any regularity of degrees. This slackness and nonregularity will be crucial in arguing that unique-neighbor expanders are weakly smooth.

Definition 4 (Weakly smooth codes). Let $0 \leq \alpha_1 \leq \alpha'_1 < 1,\ 0 < \alpha_2 < 1,\ d^*$ be constants. Code C is $(\alpha_1, \alpha'_1, \alpha_2, d^*)$-*weakly smooth* if $\forall I \subseteq [n],\ |I| < \alpha_1 n$ letting

$$\mathrm{Constr}_{(I)} = \{u \in C^{\perp}_{\leq d^*} \mid \mathrm{supp}(u) \cap I = \emptyset\}$$

and $C' = (\mathrm{Constr}_{(I)})^{\perp}$ there exists $I' \subset [n],\ I \subseteq I'$, $|I'| < \alpha'_1 n$ such that $d(C'|_{[n]\setminus I'}) \geq \alpha_2 n$.

The following is the main technical lemma used to show weakly smooth codes are robust. Its proof follows in the next section.

Lemma 4 (Main Lemma)**.** *Let $R \subseteq F^m$ and $C \subseteq F^n$ be codes of distance δ_R and δ_C. Assume C is $(\alpha_1, \alpha'_1 < \delta_C, \alpha_2, d^*)$-weakly smooth and let $M \in F^m \otimes F^n$. If $\rho(M) < min\{\frac{\alpha_1 \delta_R}{2d^*}, \frac{\delta_R \alpha_2}{2}\}$ then $\delta(M) \leq 8\rho(M)$.*

4 Weakly smooth codes are robust — Proof of Lemma 4

We pretty much follow the proof of the Main Lemma in [2], but attend to the required modifications needed to carry the proof with the weaker requirement of smoothness. The main place where we use the weakly smooth property is the Proposition 6.

Proof of Lemma 4. For row $i \in [n]$, let $r_i \in R$ denote the codeword of R closest to the ith row of M. For column $j \in [m]$, let $c^{(j)} \in C$ denote the codeword of C closest to the jth column of M. Let M_R denote the $n \times m$ matrix whose ith row is r_i, and let M_C denote the matrix whose jth column is $c^{(j)}$. Let $E = M_R - M_C$.

In what follows matrices M_R, M_C and (especially) E will be central objects of attention. We refer to E as the error matrix. Note that $\delta(M, M_R) = \delta^{row}(M)$ and $\delta(M, M_C) = \delta^{col}(M)$ and with some abuse of notation let $\mathrm{wt}(E)$ be the relative weight of E, so

$$\begin{aligned} \mathrm{wt}(E) &= \delta(M_R, M_C) \leq \delta(M, M_R) + \delta(M, M_C) \\ &= \delta^{row}(M) + \delta^{col}(M) = 2\rho(M). \end{aligned} \tag{1}$$

Our proof strategy is to show that the error matrix E is actually very structured. We do this in two steps. First we show that its columns satisfy most constraints of the column code. Then we show that E contains a large submatrix which is all zeroes. Finally using this structure of E we show that M is close to some codeword in $R \otimes C$. The following is from [2, Proposition 4], we give the proof for the sake of completeness.

Proposition 5. *Let $u \in C_d^{\perp}$ be a constraint of C with* $\mathrm{supp}(u) = \{i_1, ..., i_d\}$*. Let e_i denote the ith row of E. Suppose* $\mathrm{wt}(e_{i_j}) < \delta_R/d$ *for every $j \in [d]$. Then $u^T \cdot E = 0$.*

Proof. Note that $\forall c \in C$: $\langle c, u \rangle = 0$. Let c_i denote the i-th row of the matrix M_C.(Recall that the rows of M_C are not necessarily codewords of any nice code - it is only the columns of M_C that are codewords of C). For every column j, we have $\langle (M_C)_j, u \rangle = 0$ (since the columns of M_C are codewords of C).

Thus we conclude that $u^T \cdot M_C = 0$ as a vector. Clearly, $u^T \cdot M_R \in R$ since each one of the rows of M_R is a codeword of R. But this implies

$$u^T \cdot E = u^T \cdot (M_R - M_C) = u^T \cdot M_R - u^T \cdot M_C = u^T \cdot M_R - 0 \in R$$

Now we use the fact that the e_{i_j}s have small weight for $i_j \in [d]$. This implies that

$$\mathrm{wt}(u^T \cdot E) \leq \mathrm{wt}(u) \cdot (\delta_R/d) < \delta_R.$$

But R is an error-correcting code of the minimum distance δ_R so the only word of weight less than δ_R in it is the zero codeword, yielding $u^T \cdot E = 0$. □

Proposition 6. *There exist subsets $U \subseteq [m]$ and $V \subseteq [n]$ with $|U|/m < \delta_R$ and $|V|/n < \delta_C$ such that letting $\bar{V} = [n] \setminus V$ and $\bar{U} = [m] \setminus U$ we have for all $i \in \bar{V}, j \in \bar{U}$ that $E(i,j) = 0$.*

Proof. Let $V_1 \subseteq [n]$ be the set of indices corresponding to rows of the error matrix E with weight more than δ_R/d^*, i.e.

$$V_1 = \{i \in [n] \mid \mathrm{wt}(e_i) \geq \delta_R/d^*\}.$$

Clearly, $|V_1| < \alpha_1 n$, since $\frac{|V_1|}{n} \cdot \frac{\delta_R}{d^*} \leq \mathrm{wt}(E) \leq 2\rho(M)$ and thus $\frac{|V_1|}{n} \leq \frac{2\rho(M)}{\delta_R/d^*} < \alpha_1$ where the last inequality follows from the assumption $\rho(M) < \frac{\alpha_1 \delta_R}{2d^*}$. Let $\mathrm{Constr}_{(V_1)} = \{u \in C^{\perp}_{\leq d^*} \mid \mathrm{supp}(u) \cap V_1 = \emptyset\}$ and $C' = (\mathrm{Constr}_{(V_1)})^{\perp}$. Proposition 5 implies that $\forall u \in \mathrm{Constr}_{(V_1)}$ we have $u^T \cdot E = 0$, i.e. every column of E, denoted by E_j, satisfies constraint u and thus $E_j \in C'$.

Recall that C is $(\alpha_1, \alpha_1' < \delta_C, \alpha_2, d^*)$-weakly smooth. Associate the set V_1 with I from Definition 4. Following this definition, there exists a set I' (let $V = I'$), $|V| = |I'| < \alpha_1' n$ such that $d(C'_{[n]\setminus I'}) = d(C_{[n]\setminus V}) \geq \alpha_2 n$. We notice that for every column of E, denoted by E_j, we have $(E_j)|_{[n]\setminus I'} \in C_{[n]\setminus V}$. Thus E_j is either zero outside V or has at least $\alpha_2 n$ non-zero elements outside V.

Let U be the set of indices corresponding to the "heavy columns" of E that have $\alpha_2 n$ or more non-zero elements in the rows outside V. We conclude that every column of E that is not zero outside V is located in U. We argue that for each $(i,j) \in \bar{V} \times \bar{U}$ we have $E(i,j) = 0$. This is true since after we remove rows from V all projected nonzero columns have weight at least $\alpha_2 n$ and thus all nonzero columns are located in U. Hence all columns of $\bar{V} \times \bar{U}$ are zero columns.

Clearly, $\frac{|U|}{m} < \delta_R$, since $\frac{|U|}{m} \cdot \alpha_2 \leq \mathrm{wt}(E) \leq 2\rho(M)$ and thus $\frac{|U|}{m} \leq \frac{2\rho(M)}{\alpha_2} < \delta_R$, where the last inequality follows from the assumption $\rho(M) < \frac{\delta_R \alpha_2}{2}$. □

We now use a standard property of tensor products to claim M_R, M_C and M are close to a codeword of $R \otimes C$. Recall that $M \in F^{n \times m}$ and that $\delta(M_C, M_R) \leq 2\rho(M)$. We reproduce the following proof from [2, Proposition 6] for the sake of completeness.

Proposition 7. *Assume there exist sets $U \subseteq [m]$ and $V \subseteq [n]$, $|U|/m < \delta_R$ and $|V|/n < \delta_C$ such that $M_R(i,j) \neq M_C(i,j)$ implies $j \in U$ or $i \in V$. Then $\delta(M) \leq 8\rho(M)$.*

Proof. First we note that there exists a matrix $N \in R \otimes C$ that agrees with M_R and M_C on $\bar{V} \times \bar{U}$ (See [1, Proposition 3]). Recall also that $\delta(M, M_R) = \delta^{row} \leq 2\rho(M)$. So it suffices to show $\delta(M_R, N) \leq 6\rho(M)$.We do so in two steps. First we show that $\delta(M_R, N) \leq 2\rho(M_R)$. We then show that $\rho(M_R) \leq 3\rho(M)$ concluding the proof.

For the first part we start by noting that M_R and N agree on every row in $\bar{V}$. This is the case since both rows are codewords of R which may disagree only on entries from the columns of U, but the number of such columns is less that $\delta_R m$. Next we claim that for every column $j \in [m]$ the closest codeword of C to the $M_R(\cdot, j)$, the jth column of M_R, is $N(\cdot, j)$, the jth column of N. This is true since $M_R(i,j) \neq N(i,j)$ implies $i \in V$ and so the number of such i is less than $\delta_C n$. Thus for every j, we have $N(\cdot, j)$ is the (unique) decoding of the jth column of M_R. Averaging over j, we get that $\delta^{col}(M_R) = \delta(M_R, N)$. In turn this yields $\rho(M_R) \geq \delta(M_R)/2 = \delta(M_R, N)/2$. This yields the first of the two desired inequalities.

Now to bound $\rho(M_R)$, note that for any pair of matrices M_1 and M_2 we have $\rho(M_1) \leq \rho(M_2) + \delta(M_1, M_2)$. Indeed it is the case that $\delta^{row}(M_1) \leq \delta^{row}(M_2) + \delta(M_1, M_2)$ and $\delta^{col}(M_1) \leq \delta^{col}(M_2) + \delta(M_1, M_2)$. When the above two arguments are combined it yields $\rho(M_1) \leq \rho(M_2) + \delta(M_1, M_2)$. Applying this inequality to $M_1 = M_R$ and $M_2 = M$ we get $\rho(M_R) \leq \rho(M) + \delta(M_R, M) \leq 3\rho(M)$. This yields the second inequality and thus the proof of the proposition. □

The Main Lemma 4 follows immediately from the two last propositions. □

5 Smooth codes are also weakly so

We now show that our Definition 4 is indeed a generalization of *smooth* codes of Dinur et al. [2]. In what follows F_2 denotes the two-element field and $C(R_0)$ is a code defined by constraints in $R \setminus R_0$ (For further information and definitions see [2].). Recall the definition of smooth code:

Definition 5 (Smooth Codes). A code $C \subseteq F_2^n$ is $(d, \alpha, \beta, \delta)$-smooth if it has a parity check graph $B = (L, R, E)$ where all the right vertices R have degree d, the left vertices have degree $c = d|R|/|L|$, and for every set $R_0 \subseteq R$ such that $|R_0| \leq \alpha|R|$, there exist a set $L_0 \subseteq L$, $|L_0| \leq \beta|L|$ such that the code $C(R_0)|_{[n]\setminus L_0}$ has distance at least δ.

Claim 8. *If $C \subseteq F_2^n$ is a $(d, \alpha, \beta, \delta)$-smooth code then it is $(\alpha_1, \alpha'_1, \alpha_2, d^*)$-weakly smooth with $\alpha_1 = \frac{\alpha}{d}$, $\alpha'_1 = \beta$, $\alpha_2 = \delta$, $d^* = d$.*

Proof. Let R be a set of constraints of degree d and let $I \subseteq [n]$, $|I| < \alpha_1 n = \frac{\alpha n}{d}$ be the index set from Definition 4. Remove all d-constraints that touch at least one index in I. Let R_0 be a set of removed constraints from R. We have left degree $c = \frac{d|R|}{n}$, so, we removed at most $c \cdot \alpha_1 n = d|R|\alpha_1 = \alpha|R|$ constraints. Let $\text{Constr}_{(I)} = \{u \in C_d^{\perp} \mid \text{supp}(u) \cap I = \emptyset\}$ be the set of constraints in $R \setminus R_0$ (low weight dual words). We notice that $C(R_0) = (\text{Constr}_{(I)})^{\perp}$. Let $I' \subseteq [n]$, $|I'| < \beta n = \alpha'_1 n$ be index set from smooth codes definition (Definition 5) that should be thrown out in order to remain with good distance, i.e. $d(C(R_0)|_{[n]\setminus I'}) \geq \delta n = \alpha_2 n$. Clearly $I \subseteq I'$ as otherwise $d(C(R_0)|_{[n]\setminus I'}) = 1$. Thus from the definition of smoothness, letting $C' = (\text{Constr}_{(I)})^{\perp}$ we have $d(C'|_{[n]\setminus I'}) \geq \alpha_2 n$ which proves that C is $(\alpha_1, \alpha'_1, \alpha_2, d^*)$-weakly smooth. □

6 Unique-Neighbor Expander Codes are weakly smooth

As explained in Section 3.1 Dinur et al. [2] showed that expander codes with $\gamma < \frac{1}{6}$ are smooth and thus result in robust tensor product. In this section we show that it is possible to obtain robust tensor codes from expander code with the weaker assumption $\gamma < \frac{1}{2}$. We first define the *gap property* (Definition 6) and prove that it implies weak smoothness. Then we show that unique-neighbor expander codes have the *gap property*.

Definition 6 (Gap property). Code C has a (α, δ, d)-gap property if $\forall J \subseteq [n], |J| < \alpha n$ letting $\text{Constr}_{(J)} = \{u \in C_{\leq d}^{\perp} \mid \text{supp}(u) \cap J = \emptyset\}$ and $C' = (\text{Constr}_{(J)})^{\perp}$ we have that $\forall c \in C'|_{[n]\setminus J}$ either $\text{wt}(c) < 0.1\delta n$ or $\text{wt}(c) > 0.8\delta n$.

Claim 9. *If C has (α, δ, d)-gap property then it is $(\alpha, \alpha + 0.3\delta, 0.5\delta, d)$-weakly smooth.*

Proof. Clearly, C has no codewords of weight between $0.1\delta n$ and $0.8\delta n$. To see this take $J = \emptyset$ and then gap property implies that $\forall w \in F^n$ if $0.1\delta n \leq \text{wt}(w) \leq 0.8\delta n$ then $\langle w, u\rangle \neq 0$ for some $u \in C_{\leq d}^{\perp}$.

Let $S = \{c \in C \mid 0 < \text{wt}(c) < 0.1\delta n\}$ be a set of all non-zero low weight codewords. Let J_S be the union of supports of non-zero low weight words, i.e. $J_S = \bigcup_{c \in S} \text{supp}(c)$ and for any set $A \subseteq C$ let $J_A = \bigcup_{c \in A} \text{supp}(c)$. We show that $|J_S| < 0.3\delta n$.

Assume the contrary, i.e. $|J_S| \geq \delta \cdot 0.3n$. Then there exists $S' \subseteq S$, such that $0.2\delta n < |J_{S'}| < 0.3\delta n$. To see this remove low weight words one by one from S, each time decreasing S at most by $0.1\delta n$.

Consider a random linear combination of codewords from S'. The expected weight of the above is more than $0.1\delta n$ but can not exceed $0.3\delta n$, thus there exists such a linear combination of low weight codewords that produces a codeword with weight more than $0.1\delta n$ but less than $0.3\delta n$. Contradiction.

Thus for the rest of the proof we assume $|J_S| < 0.3\delta n$. We are ready to show that C is $(\alpha, \alpha + 0.3\delta n, 0.5\delta n, d)$-weakly smooth. Let $I \subset [n]$, $|I| < \alpha n$ be arbitrarily chosen set. Let $\mathrm{Constr}_{(I)} = \{u \in C^{\perp}_{\leq d} \mid \mathrm{supp}(u) \cap I = \emptyset\}$ and $C' = (\mathrm{Constr}_{(I)})^{\perp}$.

¿From the definition of the gap property and from the above it follows that $\forall c \in C'|_{[n]\setminus I}$ either $\mathrm{wt}(c) < 0.1\delta n$ and thus $\mathrm{supp}(c) \subseteq J_S$ or $\mathrm{wt}(c) > 0.8\delta n$.

Let $I' = J_S \cup I$ and then $|I'| \leq |J_S| + |I| < \alpha n + 0.3\delta n$. We claim that $d(C'|_{[n]\setminus(I\cup J_S)}) = d(C'|_{[n]\setminus(I')}) \geq 0.5\delta n$. To see this assume $c' \in C'|_{[n]\setminus I}$, $c'' = c'|_{[n]\setminus(I\cup J_S)}$, $c'' \in C'|_{[n]\setminus(I\cup J_S)}$ such that $0 < \mathrm{wt}(c'') < 0.5\delta n$ but then $0 < \mathrm{wt}(c'') \leq \mathrm{wt}(c') \leq |J_S| + \mathrm{wt}(c'') < 0.8\delta n$ and thus c' is a low weight word, i.e. $\mathrm{supp}(c') \subseteq J_S$. Hence $c'' = c'|_{[n]\setminus(I\cup J_S)} = 0$, contradicting $\mathrm{wt}(c'') > 0$. □

Proposition 10. *Let C be a linear code over F. If $u_1 \in C^{\perp}_{<f}$ and $u_2 \in C^{\perp}_{<g}$ and $i \in \mathrm{supp}(u_1) \cap \mathrm{supp}(u_2)$ then exists $u_3 \in C^{\perp}_{<f+g}$ such that $\mathrm{supp}(u_3) \subseteq (\mathrm{supp}(u_1) \cup \mathrm{supp}(u_2)) \setminus \{i\}$.*

Proof. Let $a \in F$ be ith entry in u_1 and $b \in F$ be ith entry in u_2. Then $u_3 = a^{-1}u_1 + b^{-1}u_2 \in C^{\perp}_{<f+g}$ has desired properties. □

Claim 11. *Let C be a (c, d, γ, δ)-expander code over F with constant $\gamma < \frac{1}{2}$. Let $w \in F^n$ with $0 < \mathrm{wt}(w) < \delta n$ with $I = \mathrm{supp}(w)$. Then at least a 0.95-fraction of indices $i \in I$ have $u_i \in C^{\perp}_{<d^*}$ where $d^* < d^k$, $k = (log_{0.5+\gamma}(0.05)) + 1$ such that $\mathrm{supp}(u_i) \cap I = \{i\}$.*

Proof. Fix set I with $|I| < \delta n$. Let (L, R, E) be a check graph of C that is a (c, d)-regular (γ, δ)-expander. The claim follows from examining the unique neighbor structure of the graph. We prove this by induction on $j = 1...k$ and show set constructions I_j satisfying

- $I_1 = I$, $I_{j+1} \subset I_j$
- $|I_{j+1}| \leq (\frac{1}{2} + \gamma)|I_j|$
- $\forall i \in I_j \setminus I_{j+1}$ exists $u_i \in C^{\perp}_{\leq d^j}$ with $\mathrm{supp}(u_i) \cap I = \{i\}$

We then conclude $(\frac{1}{2}+\gamma)^k < 0.05$ and thus from the induction follows that $I_k \subset I$, $|I_k| < 0.05 \cdot |I|$ and $\forall i \in I \setminus I_k$ exists $u_i \in C^{\perp}_{<d^k}$ with $\mathrm{supp}(u_i) \cap I = \{i\}$. And the the proof of the claim is completed.

For the base case let $I_1 = I$. Since C is an expander and $|I_1| \leq \delta n$, I_1 has at least $(1-\gamma)c|I_1| = (\frac{c}{2} + (0.5-\gamma)c)|I_1|$ neighbors in R. Each index $i \in I_1$ is asked by c constraints in R. And thus the number of neighbors that ask at least 2 indices from I_1 is bounded from above by $(\frac{c}{2})|I_1|$. Hence there are at least $((\frac{1}{2} - \gamma)c)|I_1|$ unique neighbors in R. Since a single index can not have more than c unique neighbors in R, the number of indices in I_1 having unique neighbor is at least $(\frac{1}{2} - \gamma)|I_1|$. I.e. at least $(\frac{1}{2} - \gamma)$-fraction of all indices in I_1 have a unique neighbor with support $d = d^1$. Let $I_2 \subset I_1$ be subset of all indices $i \in I_1$ that have no unique neighbor of weight at most d^1. We constructed set I_1, I_2 such that

- $I_1 = I$, $I_2 \subset I_1$
- $|I_2| \leq (\frac{1}{2} + \gamma)|I_1|$
- $\forall i \in I_1 \setminus I_2$ exists $u_i \in C^{\perp}_{\leq d^1}$ with $\mathrm{supp}(u_i) \cap I = \{i\}$

And this completes the base case.

Assume correctness until $j-1$ and let us prove for j. Consider I_j, $|I_j| \leq |I_1| \leq \delta n$. By the unique neighbor expansion at least $(\frac{1}{2} - \gamma)$-fraction of indices $i \in I_j$ have bounded unique neighbor, i.e. $u_i \in C^{\perp}_d$ such that $\mathrm{supp}(u_i) \cap I_j = \{i\}$. Let $I_{j+1} \subset I_j$ be indices $i \in I_j$ that have no bounded unique neighbor and thus $|I_{j+1}| \leq (\frac{1}{2} + \gamma)|I_j|$.

Fix $i \in I_j \setminus I_{j+1}$ arbitrarily. There exists $u_i \in C^{\perp}_d$ such that $\mathrm{supp}(u_i) \cap I_j = \{i\}$. Every index $l \in \mathrm{supp}(u_i)$, $l \neq i$ is located either in $[n] \setminus I_1$ or in $I_1 \setminus I_j$. We handle all $l \in I_1 \setminus I_j$ using linear combination according to Proposition 10 to obtain a constraint $u'_i \in C^{\perp}_{\leq d^j}$ such that $\mathrm{supp}(u'_i) \cap I = \{i\}$. This is possible since every $l \in I_1 \setminus I_j$ is located in some I_f for $1 \leq f < j$ and thus from induction assumption has $u_l \in C^{\perp}_{\leq d^{j-1}}$ such that $\mathrm{supp}(u_l) \cap I = \{l\}$. Since $\mathrm{wt}(u_i) \leq d$ we obtain $u'_i \in C^{\perp}_{\leq d^{j-1} \cdot d} = C^{\perp}_{\leq d^j}$ such that $\mathrm{supp}(u'_i) \cap I = \{i\}$. So we showed

- $I_{j+1} \subset I_j$
- $|I_{j+1}| \leq (\frac{1}{2} + \gamma)|I_j|$
- $\forall i \in I_j \setminus I_{j+1}$ exists $u_i \in C^{\perp}_{\leq d^j}$ with $\mathrm{supp}(u_i) \cap I = \{i\}$

This yields the induction and the claim. □

Corollary 12. *If C is (c, d, γ, δ) expander code with $\gamma < \frac{1}{2}$ then C has $(0.5\delta, 0.5\delta, d^*)$ gap property where $d^* < d^k$, $k = (log_{(0.5+\gamma)} 0.05) + 1$.*

Proof. Let $J \subset [n]$, $|J| < 0.5\delta$ be arbitrarily chosen. Let $\mathrm{Constr}_{(J)} = \{u \in C^{\perp}_{<d^k} \mid \mathrm{supp}(u) \cap J = \emptyset\}$ and $C' = (\mathrm{Constr}_{(J)})^{\perp}$. Assume by contradiction, there exists $w \in C'_{[n]\setminus J}$ such that $0 < 0.1 \cdot (0.5\delta)n \leq \mathrm{wt}(w) \leq 0.8 \cdot (0.5\delta)n$. And thus there is no $u \in \mathrm{Constr}_{(J)}$ such that $|\mathrm{supp}(u) \cap \mathrm{supp}(w)| = 1$.

Let $I = J \cup \mathrm{supp}(w)$, $|I| \leq |J| + \mathrm{wt}(w) < 0.5\delta n + 0.4\delta n < \delta n$. We notice that $\mathrm{supp}(w) \cap J = \emptyset$ and $|\mathrm{supp}(w)| > 0.05 \cdot |I|$. Thus Claim 11 implies that there exists $u \in C^{\perp}_{<d^k}$ such that $|\mathrm{supp}(u) \cap \mathrm{supp}(w)| = 1$ and $|\mathrm{supp}(u) \cap I| = |\mathrm{supp}(u) \cap \mathrm{supp}(w)| = 1$. Thus $u \in \mathrm{Constr}_{(J)}$ such that $|\mathrm{supp}(u) \cap \mathrm{supp}(w)| = 1$. Contradiction. □

Claim 13. *If C is (c, d, γ, δ) expander code with $\gamma < \frac{1}{2}$ then C is $(0.5\delta, 0.65\delta, 0.25\delta, d^*)$-weakly smooth where $d^* < d^k$, $k = (log_{(0.5+\gamma)} 0.05) + 1$.*

Proof. Follows immediately from Corollary 12 and Claim 9. Corollary 12 implies that C has $(0.5\delta, 0.5\delta, d^*)$ gap property where $d^* < d^k$, $k = (log_{(0.5+\gamma)} 0.05) + 1$. Claim 9 implies that C is $(0.5\delta, 0.5\delta + 0.15\delta, 0.25\delta, d^*)$-weakly smooth . □

Proof of Theorem 2. Let $R \subseteq F^m$ and $C \subseteq F^n$ be codes of distance δ_R and δ_C. Let $M \in F^m \otimes F^n$. Claim 13 implies that C is $(0.5\delta, 0.65\delta, 0.25\delta, d^*)$-weakly smooth where $d^* < d^k$, $k = (log_{(0.5+\gamma)} 0.05) + 1$. Main Lemma implies that if $\rho(M) < min\{\frac{(0.5\delta) \cdot \delta_R}{2d^*}, \frac{\delta_R \cdot (0.25\delta)}{2}\}$ then $\delta(M) \leq 8\rho(M)$. □

7 Locally correctable codes are weakly smooth

Definition 7 (Locally Correctable Code). A $[n, k, d]_{|F|}$ code C is called (q, ϵ, δ) locally correctable code if there exists a randomized decoder (D) that reads at most q entries and the following holds: $\forall c \in C$, $\forall i \in [n]$ and $\forall \hat{c} \in F^n$ such that $d(c, \hat{c}) \leq \delta n$ we have

$$Pr[D^{\hat{c}}[i] = c_i] \geq \frac{1}{|F|} + \epsilon,$$

i.e. with probability at least $\frac{1}{|F|} + \epsilon$ entry c_i will be recovered correct.

Without loss of generality we assume that given $\hat{c} \in F^n$ the "correction" of entry i (obtaining c_i) is done by choosing random $u \in S \subseteq C^{\perp}_{\leq q+1}$ such that $i \in \mathrm{supp}(u)$. Formally, assume the ith entry of u is u_i, let $u^{proj} = u|_{[n]\setminus\{i\}}$, $\hat{c}^{proj} = \hat{c}|_{[n]\setminus\{i\}}$ and then c_i is recovered by $D^{\hat{c}}[i] = \frac{\langle u^{proj}, \hat{c}^{proj}\rangle}{u_i}$, notice that $u_i \neq 0$.

The next claim holds for every $\epsilon > 0$ which can be arbitrarily close to 0 (e.g. $o(1)$) whereas usually locally correctable codes are defined with $\epsilon = \Omega(1)$.

Claim 14. *If C is (ϵ, δ, q)-locally correctable code with $\epsilon > 0$ then it is $(0.5\delta, 0.5\delta, 0.5\delta, q + 1)$-weakly smooth and its relative distance is at least δ.*

Proof. We first show that $\forall I \subseteq [n]$, $|I| \leq \delta n$ and $\forall i \in I$ we have $u_i \in C^{\perp}_{\leq q+1}$ with $\mathrm{supp}(u_i) \cap I = \{i\}$. Assume the contrary and fix $I \subseteq [n]$, $|I| \leq \delta n$ and $i \in I$. So, for all $u_i \in C^{\perp}_{\leq q+1}$ with $i \in \mathrm{supp}(u_i) \cap I$ we have $|\mathrm{supp}(u_i) \cap I| \geq 2$. Consider an adversary that takes $c \in C$ and sets c_j to random element from F for all $j \in I$, obtaining $\hat{c}$. Clearly, c_i will be recovered with probability at most $\frac{1}{|F|}$ since for every $u^{(i)} \in C^{\perp}_{\leq q+1}$ such that $i \in \mathrm{supp}(u^{(i)})$ the inner product $\langle (u^{(i)})|_{[n]\setminus\{i\}}, c|_{[n]\setminus\{i\}}\rangle$ will produce a uniformly random value in F.

We next show that $d(C) \geq \delta n$. To see this assume $c \in C$ such that $0 < \mathrm{wt}(c) < \delta n$. Let $I = \mathrm{supp}(c)$, $|I| < \delta n$ and $i \in I$. There exists $u \in C^{\perp}_{\leq q+1}$ with $\mathrm{supp}(u) \cap \mathrm{supp}(c) = \{i\}$ and thus $\langle u, w\rangle \neq 0$ implies $c \notin C$.

We finally show the weak smoothness of C. Let $I \subset [n]$, $|I| < 0.5\delta n$ be the adversary chosen set and let $I' = I$. Let $\mathrm{Constr}_{(I)} = \{u \in C^{\perp}_{\leq q+1} \mid \mathrm{supp}(u) \cap I = \emptyset\}$ and $C' = (\mathrm{Constr}_{(I)})^{\perp}$. We claim that $d(C'|_{[n]\setminus I}) \geq 0.5\delta n$. This is true, since otherwise we have $c' \in C'$, $c'_{[n]\setminus I} \in C'|_{[n]\setminus I}$ such that $0 < \mathrm{wt}(c'_{[n]\setminus I}) < 0.5\delta n$. But then $0 < \mathrm{wt}(c') < 0.5\delta n + |I| \leq \delta n$ and thus exists $u \in \mathrm{Constr}_{(I)}$ such that $|\mathrm{supp}(u) \cap \mathrm{supp}(c')| = 1$ which implies $\langle u, c'\rangle \neq 0$ and $c' \notin C'$. Contradiction. So, C is $(0.5\delta, 0.5\delta, 0.5\delta, q + 1)$-weakly smooth. □

of Theorem 3. Let $R \subseteq F^m$ and $C \subseteq F^n$ be linear codes such that $\delta(R) \geq \delta_R$. Let $M \in F^m \otimes F^n$. Claim 14 implies that C is $(0.5\delta, 0.5\delta, 0.5\delta, q + 1)$-weakly smooth and $\delta(C) \geq \delta$. The Main Lemma 4 implies that if $\rho(M) < \min\{\frac{(0.5\delta)\cdot\delta_R}{2(q+1)}, \frac{\delta_R\cdot(0.5\delta)}{2}\} = \frac{(0.5\delta)\cdot\delta_R}{2(q+1)}$ then $\delta(M) \leq 8\rho(M)$. □

Acknowledgements.

We thank Madhu Sudan for helpful discussions.

References

[1] E. Ben-Sasson and M. Sudan.: Robust locally testable codes and products of codes, APPROX-RANDOM (2004), pp. 286-297 (See ECCC TR04-046, 2004).

[2] I. Dinur, M. Sudan and A. Wigderson.: Robust local testability of tensor products of LDPC codes, APPROX-RANDOM (2006), pp. 304-315

[3] P. Valiant.: The tensor product of two codes is not necessarily robustly testable, APPROX-RANDOM (2005), pp. 472-481

[4] D. Copersmith and A. Rudra.: On the robust testability of tensor products of codes, ECCC TR05-104,(2005)

[5] O. Meir.: On the rectangle method in proofs of robustness of tensor products, ECCC TR07,(2007)

[6] O. Meir.: Combinatorial Construction of Locally Testable Codes. M.Sc. Thesis, Weizmann Institute of Science, (2007)

[7] O. Goldreich and O. Meir.: The tensor product of two good codes is not necessarily robustly testable, ECCC TR07, (2007)

[8] O. Goldreich.: Short locally testable codes and proofs (survey), ECCC TR05-014,(2005)

[9] O. Goldreich and M. Sudan.: Locally testable codes and PCPs of almost linear length, FOCS (2002), pp. 13-22 (See ECCC TR02-050, 2002)

[10] D. Spielman.: Linear-time encodable and decodable error-correcting codes, IEEE Trans. on Information Theory, (1996), pp. 1723-1731

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/