Testing Tensor Products of Algebraic Codes

Sumegha Garg ✉ 📧
Department of Computer Science, Rutgers University, Piscataway, NJ, USA

Madhu Sudan ✉ 📧
School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

Gabriel Wu ✉ 📧
Harvard University, Cambridge, MA, USA

— Abstract —

Motivated by recent advances in locally testable codes and quantum LDPCs based on robust testability of tensor product codes, we explore the local testability of tensor products of (an abstraction of) algebraic geometry codes. Such codes are parameterized by, in addition to standard parameters such as block length n and dimension k, their genus g. We show that the tensor product of two algebraic geometry codes is robustly locally testable provided $n = \Omega((k + g)^2)$. Apart from Reed-Solomon codes, this seems to be the first explicit family of two-wise tensor codes of high dual distance that is robustly locally testable by the natural test that measures the expected distance of a random row/column from the underlying code.

1 Introduction

In this work we consider the "robust local testability" of "tensor products" of "algebraic geometry codes". We review each of these notions below before describing our results.

All codes considered in this paper are linear codes over some finite field $\mathbb{F}_q$. Given two codes $R \subseteq \mathbb{F}_q^m$ and $C \subseteq \mathbb{F}_q^n$, their tensor product $R \otimes C \subseteq \mathbb{F}_q^{nm}$ consists of all $n \times m$ matrices M whose rows are codewords of R and whose columns are codewords of C. It is a simple but valuable exercise in linear algebra to note that the dimension of $R \otimes C$ is the product of the dimensions of R and C. It is an arguably simpler exercise to see that the distance of the code $R \otimes C$ is the product of the distance of R and C. (In this paper by distance we mean, either absolute or normalized, Hamming distance. This particular statement holds for either of these terms.) Our interest in tensor product codes comes from their potential testability properties elaborated next.

Given an arbitrary matrix $A \in \mathbb{F}_q^{n \times m}$, the definition of a tensor product defines a natural test to see if $A \in R \otimes C$, namely verify every row is in R and every column is in C. The "robust testability" property explores the robustness of this definition in measuring distances: If the expected distance of a uniformly chosen row of A from R is small, and so is the expected distance of a uniformly chosen column, then is A close to a codeword of $R \otimes C$? A tensor product of codes is said to be robustly testable if the answer is affirmative. (This notion is formulated more precisely and quantitatively in Definition 2.)

The formal study of robustness of tensor product codes was initiated by Ben-Sasson and Sudan [4]. We elaborate on their motivation shortly, but for now mention that most of the general work in this setting, including that in [4] and the work of Viderman [28], considered a variant of the tensor-product testing question raised above. Specifically they consider m-wise tensor products of codes for $m \geq 3$ and showed that these are robust with respect to a slightly more complex "two-wise test" (where the test measures the expected distance of a random two-dimensional projection from the underlying two-wise tensor) as long as the code being tensored has sufficiently large distance.

Robustness of two-wise tensors has been shown for very few classes of codes in the literature so far. Briefly the works have considered the tensor product when the codes R and C are Reed-Solomon codes [24, 1, 23] or when dual distance of C is small [9][1], or when R and C are random. The work of Valiant [26] also gives examples of asymptotically good codes whose tensor product is not robustly testable – suggesting that robustness needs additional properties (other than just rate and distance) in the ingredient codes.

In this work, we consider new classes of codes which are generalizations of Reed-Solomon codes. The codes we study are abstractions of algebraic geometry codes - such codes come as an entire sequence of codes characterized by two main features: (a) they approach the Singleton bound in terms of their distance vs. dimension tradeoff with the additive gap termed the "genus" (a parameter that is derived from the genus of some underlying algebraic curves, but gets a purely coding theoretic interpretation in this abstraction); and (b) the Hadamard product of two codewords of two codes of small dimension is a codeword of a code in the sequence of only slightly larger dimension. (See Definition 1 for a precise formulation.)

The main result in this paper shows that the tensor product of two sequences of AG codes is robustly testable provided their dimension and genus is sufficiently small compared to the length of the code. (See Theorem 4 for a precise statement.) Before describing getting into the specifics we give some context and motivation for the study of testing of tensor product codes.

1.1 Robust Testability of Tensor Product: Background and Motivation

Robust testability of tensor codes was studied in the work of Ben-Sasson and Sudan [4], who raised the question of whether $R \otimes C$ is robustly testable for all codes of sufficiently large relative distance. Their question was inspired by the role of the "bivariate polynomial tester" in the works on PCPs, originating in the works of Babai, Fortnow, Levin and Szegedy [2] and explored systematically in the work of Rubinfeld and Sudan [24]. Seminal results in this space include the work of Arora and Safra [1] who showed that Reed-Solomon codes of inverse polynomial rate have constant robustness, and the work of Polishchuk and Spielman [23], who extended the result of [1] to Reed-Solomon codes of any linear rate bounded away from $1/2$.

---

[1] Dual of a code $C \subseteq \mathbb{F}_q^n$ is defined as the set of vectors orthogonal to C, that is, $C^{\perp} = \{x \in \mathbb{F}_q^n \mid \forall y \in C, \langle x, y \rangle = 0\}$. By dual distance, we refer to the distance of the dual code.

The question raised in [4] was however answered negatively by Valiant [26] (see also [5, 11], who gave codes of relative distance arbitrarily close to 1 that were not robustly testable). These works motivated the search for specific classes of codes, other than Reed-Solomon codes, whose tensor product is robustly testable.

The robust testability of tensor product codes has played a central role in many results over the years. The robust testability of Reed-Solomon codes is used ubiquitously in PCP constructions. It also plays a role in the breakthrough result of Ji, Natarajan, Vidick, Wright and Yuen [14] showing MIP*=RE. Indeed this motivated the same authors [15] to explore the quantum testability of tensor products of codes. It also led to the first combinatorial constructions of LTCs of nearly linear rate in the work of Meir [20] and first known constructions of strong LTCs with nearly linear rate in the work of Viderman [27]. The robustness of tensor product codes also plays a central role in the recent breakthroughs of Dinur, Evra, Livne, Lubotzky and Mozes [7] and Panteleev and Kalachev [21] (see also [19]) giving constant query LTCs of linear rate and distance, and quantum LDPCs. We remark that the role of tensor product codes here is essential – these are the only sources of "redundancy" among the local constraints in the constructions of [7, 21, 19] and such redundancies are necessary to get LTCS as shown in [3]. Indeed the only two known sources of redundancy among constraints in LDPCs come from symmetries (see [18]) or the tensor product construction.

These many applications motivate the quest for general tools in the analysis of robustness of tensor product codes. To date the only known works on robust testability of tensor codes are (1) the aforementioned works on Reed-Solomon codes, (2) a work of Dinur, Sudan and Wigderson [9] roughly showing that $R \otimes C$ is robustly testable if C is an LDPC code and (3) recent works, notably by Panteleev and Kalachev [21] and Leverrier and Zémor [19], showing that tensor products of random linear codes (and their duals) are robustly testable[2], with improved parameters in [16, 8]. (We remark that some of the interest in 2-dimensional robust testability is due to an equivalence with a notion called "product expansion" of codes, an equivalence that holds only for two-dimensional tensor products. Both the notion of robustness and product expansion do extend to higher dimensional products but they are no longer equivalent. See [17] and references therein.)

The lack of more general results motivates us to study the testability results for products of Reed-Solomon codes. (We note the other explicit result [9] is already generic, while the robustness in the setting of Reed-Solomon codes is not.) However, the current analyses of robustness in this setting are very algebraic and to make this analysis more generic, one needs an appropriate abstraction of the underlying algebra and we discuss this next.

1.2 Abstracting Algebraic Codes

Over the years coding theorists have proposed nice abstractions of algebraic codes - see for instance [10, 22]. These works abstract a product property that captures the fact that the product of two polynomials of degree d has degree at most $2d$. This corresponds to the fact that the coordinate-wise product of two codewords from a (roughly) d dimensional space is contained in a (roughly) $2d$ dimensional space, which is a highly non-trivial effect.

---

[2] In this work, we prove robust testability of tensor products of an abstraction of algebraic geometry codes. As the dual distance of LDPC codes is a constant by definition, this gives us first explicit family of codes of super-constant dual distance that is robustly locally testable after Reed-Solomon codes (to the best of our knowledge). Explicit constructions for AG codes with super-constant dual distance are known (see Section 2.7 in [13]).

(In contrast for a generic linear code $C \subseteq \mathbb{F}^n$ of dimension d, the smallest linear space that contains all the coordinate-wise products of pairs of codewords from C has dimension $\min\{d^2, n\}$.) This non-trivial product property seen in Reed-Solomon codes when abstracted properly captures most algebraic codes (including Reed-Muller and algebraic geometry codes) nicely and suffices to explain most decoding algorithms for such codes. However other algorithms, such as list-decoding algorithms, use more involved properties (see [12]) that include the ability to capture multiplicities of zeroes and the ability to shift a polynomial without increasing its degree (i.e., $f(x)$ has the same degree as $f(x + a)$).

The current analyses of robust testability of tensor products of Reed-Solomon codes use many aspects of polynomials in addition to the product property. For instance they rely on unique factorization, on the role of resultants in computing greatest common divisors (and the fact that resultants themselves are low-degree polynomials). They use the fact that puncturing of Reed-Solomon codes are Reed-Solomon codes etc. Given all these aspects in the proofs, it is interesting to see how far one can go with more minimal assumptions.

In this work, we use a simple quantitative version of the product property (see Definition 1) which naturally captures algebraic geometry codes, (but not for instance Reed-Muller codes). In particular we avoid use of unique factorization and GCDs, and also avoid explicit use of the puncturing property. This allows us to recover a moderately strong version of the analysis for Reed-Solomon codes: specifically we can analyze codes that have block length at least quadratic in the dimension. Thus our work is not strong enough to imply the result of Polishchuk and Spielman [23] who only require block length linear in the dimension; but is stronger than the previous work of Arora and Safra [1] (and implies their result) who showed that the tensor product of two Reed-Solomon codes of dimension d and length n is robustly testable provided $n = \Omega(d^3)$.

The next section presents our formal definitions and theorem statement.

2   Definitions and Main Result

We use $\mathbb{F}_q$ to denote the finite field on q elements. We use functional notation to describe vectors, so the vector space $\mathbb{F}_q^n$ will be viewed and represented as functions $\mathbb{F}_q^n = \{f : S \to \mathbb{F}_q\}$ for some set S with $|S| = n$. (Often we use $S = [n]$.) Note that with this notation we naturally have the notion of $f + g$ and fg both of which are in $\mathbb{F}_q^n$. For functions $f : S \to \mathbb{F}_q$ and $g : T \to \mathbb{F}_q$, we use $f \otimes g : S \times T \to \mathbb{F}_q$ to denote the function $(f \otimes g)(x, y) = f(x)g(y)$. If $f \in \mathbb{F}_q^m$ and $g \in \mathbb{F}_q^n$, note that we have $f \otimes g \in \mathbb{F}_q^{n \times m} \cong \mathbb{F}_q^{nm}$. Here, we assume that a "row" is indexed by y and varies $x \in [m]$ (vice versa for "column"). For $f, g : S \to \mathbb{F}_q$, we use $\mathrm{dist}(f, g)$ to denote the absolute (non-normalized) Hamming distance between f and g, i.e., $\mathrm{dist}(f, g) = |\{x \in S | f(x) \neq g(x)\}|$ and we use $\delta(f, g) = \frac{1}{|S|} \cdot \mathrm{dist}(f, g)$ to denote the normalized Hamming distance.

We consider linear codes $C \subseteq \mathbb{F}_q^n$. For such a code we use $\dim(C)$ to denote its dimension as a vector space, $\mathrm{dist}(C)$ to denote its (non-normalized) minimum distance (between any two code vectors). For a vector $f \in \mathbb{F}_q^n$ and code $C \subseteq \mathbb{F}_q^n$ we use $\delta(f, C)$ to denote the distance of f to the nearest codeword in C, i.e., $\min_{g \in C}\{\delta(f, g)\}$. For codes $C_1, C_2 \subseteq \mathbb{F}_q^n$, we use $C_1 \star C_2$ to denote its Hadamard product, i.e., $C_1 \star C_2 = \mathrm{span}(\{fg | f \in C_1, g \in C_2\})$. For codes $C_1 \subseteq \mathbb{F}_q^m$ and $C_2 \subseteq \mathbb{F}_q^n$, we let $C_1 \otimes C_2$ denote their tensor product, i.e., $C_1 \otimes C_2 = \mathrm{span}(\{f \otimes g | f \in C_1, g \in C_2\})$. For a matrix $A \in \mathbb{F}_q^{nm}$, it is a simple exercise to see that $A \in C_1 \otimes C_2$ iff every row of A is in C_1 and every column is in C_2 (when C_1 and C_2 are linear codes).

▶ **Definition 1** ((Abstract) Algebraic Geometry Code). *Given integers $0 \leq g \leq n$ and a prime power q, an (q, n, g)-algebraic geometry code sequence is a sequence of linear codes $\mathcal{C} = \{\mathcal{C}(\ell) \subseteq \mathbb{F}_q^n \mid 0 \leq \ell \leq n\}$ with the following requirements:*
(1) For every ℓ, $\dim(\mathcal{C}(\ell)) \geq \ell - g$ and $\mathrm{dist}(\mathcal{C}(\ell)) \geq n - \ell$.
(2) For every ℓ, m, $\mathcal{C}(\ell) \star \mathcal{C}(m) \subseteq \mathcal{C}(\ell + m)$.
We use $\mathcal{AG}(q, n, g)$ to denote the space of all (q, n, g)-algebraic geometry code sequences.

Condition (2) above is called the product property and it is what makes algebraic codes special.

▶ **Definition 2** (Robustness of Tensor Product). *For codes $C_1 \subseteq \mathbb{F}_q^m$ and $C_2 \subseteq \mathbb{F}_q^n$ and $0 \leq \rho \leq 1$ we say that (C_1, C_2) is ρ-robust if for every $F \in \mathbb{F}_q^{n \times m}$ we have*

$$\rho \cdot \delta(F, C_1 \otimes C_2) \leq \frac{1}{2}[\delta(F, C_1 \otimes \mathbb{F}_q^n) + \delta(F, \mathbb{F}_q^m \otimes C_2)].$$

For a linear code C and set $A \subseteq S$, let $C|_A = \{x|_A \mid x \in C\} \subseteq \mathbb{F}_q^{|A|}$ be the projection of C to the coordinates of A. We will use the following proposition about tensor products in our proof.

▶ **Proposition 3.** *If $C_1, C_2 \subseteq \mathbb{F}_q^n$ are linear codes and $A, B \subseteq [n]$, then*

$$C_1|_A \otimes C_2|_B = (C_1 \otimes C_2)|_{A \times B}.$$

Proof. Let $\dim C_1 = k_1$ and $\dim C_2 = k_2$. Then there exist matrices $M_1 \in \mathbb{F}_q^{k_1 \times n}$ and $M_2 \in \mathbb{F}_q^{k_2 \times n}$ such that

$$C_1 = \{xM_1 \mid x \in \mathbb{F}_q^{k_1}\} \quad \text{and} \quad C_2 = \{xM_2 \mid x \in \mathbb{F}_q^{k_2}\},$$

and their tensor product can then be expressed as

$$C_1 \otimes C_2 = \{M_2^T X M_1 \mid X \in \mathbb{F}_q^{k_2 \times k_1}\}.$$

Note that $C_1|_A$ is generated by the restriction of M_1 to the columns of A (i.e. $M_1|_{[k_1] \times A}$). Similarly for $C_2|_B$. Our goal is now to show that

$$\{(M_2^T|_{B \times [k_2]})X(M_1|_{[k_1] \times A}) \mid X \in \mathbb{F}_q^{k_2 \times k_1}\} = \{(M_2^T X M_1)|_{B \times A} \mid X \in \mathbb{F}_q^{k_2 \times k_1}\}.$$

It suffices to show that $\forall X$,

$$(M_2^T|_{B \times [k_2]})X(M_1|_{[k_1] \times A}) = (M_2^T X M_1)|_{B \times A}.$$

This follows because:

$$\begin{aligned}
(M_2^T|_{B \times [k_2]})X(M_1|_{[k_1] \times A}) &= (M_2^T X)|_{B \times [k_1]}(M_1|_{[k_1] \times A}) \\
&= (M_2^T X(M_1|_{[k_1] \times A}))|_{B \times A} \\
&= (M_2^T (X M_1)|_{[k_2] \times A}))|_{B \times A} \\
&= ((M_2^T X M_1)|_{[n] \times A})|_{B \times A} \\
&= (M_2^T X M_1)|_{B \times A}. \qquad \blacktriangleleft
\end{aligned}$$

▶ **Theorem 4.** *There exist constants $\rho > 0$ and $c_0 < \infty$ such that for every triple of non-negative integers n, ℓ, g and prime power q satisfying $\ell > \max\{c_0, g\}$ and $n > c_0(\ell + g)^2$ we have the following: If $\mathcal{C}_1, \mathcal{C}_2 \in \mathcal{AG}(q, n, g)$, then $(\mathcal{C}_1(\ell), \mathcal{C}_2(\ell))$ is ρ-robust.*

Theorem 4 is proved at the end of Section 3. Our proof combines elements from the proofs of [1] and [23]. A direct use of the proof from [1] would have resulted in a $n = \Omega((k+g)^3)$ requirement. On the other hand, the proof of [23] uses properties of unique factorization domain, which are no longer true in this general setting. By combining different ingredients we are able to get a bound that is intermediate while still being general.

The fact that $n = \Omega(g^2)$ for the theorem to be useful does limit its applicability. Nevertheless the theorem is not vacuous even given the testability of Reed-Solomon codes and in particular, there exist infinitely many AG codes[3] with $g = o(\sqrt{n})$ and $n \approx q$.

3   Proof of Main Theorem

Let $\mathcal{C}_1, \mathcal{C}_2$ be a pair of (q, n, g)-sequences of algebraic geometry codes. We first show that for a tensor product of codes $\mathcal{C}_1(\ell)$ and $\mathcal{C}_2(\ell)$, if the rows of a matrix are near codewords of $\mathcal{C}_1(\ell)$ and columns are near codewords of $\mathcal{C}_2(\ell)$, then it is also close to a codeword of the tensor product code.

▶ **Theorem 5.** *There exist positive constants ϵ_0, c_0 that make the following true.*

For all $0 < \epsilon < \epsilon_0$, integers n, g, ℓ and q such that $\ell > \max\{c_0, g\}$, and $n > c_0(\ell + g)^2$, and for all (q, n, g)-sequences of AG codes $\mathcal{C}_1, \mathcal{C}_2$: If $R \in \mathcal{C}_1(\ell) \otimes \mathbb{F}_q^n$ and $C \in \mathbb{F}_q^n \otimes \mathcal{C}_2(\ell)$ are such that $\delta(R, C) = \epsilon$, then there exists $Q \in \mathcal{C}_1(\ell) \otimes \mathcal{C}_2(\ell)$ such that

$$\delta(Q, R) + \delta(Q, C) \leq 2\epsilon.$$

Proof. We will prove the theorem for $\epsilon_0 = \frac{1}{100}$ and $c_0 = 15$. Define the following constants:

$$\gamma = 2\sqrt{\epsilon}$$
$$\gamma' = 2\gamma = 4\sqrt{\epsilon}$$
$$L = 2(\ell + g)$$
$$d = \lfloor \sqrt{\epsilon}L \rfloor + g + 2$$

Note that these constants are chosen to satisfy the following inequalities:

$$\gamma < \gamma'(1 - \gamma') \tag{1}$$

$$\left(1 - \frac{\epsilon}{\gamma^2}\right) L > d + \ell \tag{2}$$

$$n(1 - \gamma - \gamma') - dL > L \tag{3}$$

$$(2\sqrt{\epsilon} + \epsilon) \cdot \frac{n}{n - \ell} < 3\sqrt{\epsilon} \tag{4}$$

$$\frac{n - \ell - 3\sqrt{\epsilon}n}{n} > \frac{1}{2} \tag{5}$$

---

[3] For example, for any prime p, there is an elliptic curve over $\mathbb{F}_p$ with $p + 1$ number of rational points (Theorem 14.18 in [6]). One can apply the construction of AG codes from Chapter 2 in [25] to these elliptic curves to get $(p, p, 1)$-sequences of AG codes.

We can quickly check that all of these inequalities hold for our choice of ϵ_0 and c_0. Inequality (1) holds because $1 < 2(1-4\sqrt{\epsilon})$. Inequality (2) holds because the left-hand side can be written as $1.5(\ell+g)$ and the right-hand side can be written as $\lfloor \sqrt{\epsilon}\cdot 2(\ell+g)\rfloor+g+2+\ell$, which is at most $(1+2\sqrt{\epsilon})(\ell+g)+2$. Inequality (3) holds because $(1+d)L < (1+2\sqrt{\epsilon})(\ell+g)\cdot 2(\ell+g) < 6(\ell+g)^2 < 6n/c_0 \le n(1-\gamma-\gamma')$. Inequality (4) holds because $\frac{n}{n-\ell} < \frac{3}{2.1}$ and $2\sqrt{\epsilon}+\epsilon < 2.1\sqrt{\epsilon}$. Inequality (5) holds because $\ell < \frac{1}{15}n < (\frac{1}{2} - 3\sqrt{\epsilon_0})n \implies \ell + 3\sqrt{\epsilon}n < \frac{n}{2}$. Throughout the proof, we will note where we make use of each inequality.

Define an "error" to be a location (x,y) at which $R(x,y) \neq C(x,y)$. The "error fraction" of a row is the number of errors in that row divided by its length (and similarly for columns). Note that the length of a row isn't always n; sometimes we will use notions like "the error fraction of a row among the first L columns".

The majority of the proof of Theorem 5 is spent on proving the following Lemma.

▶ **Lemma 6.** *There exists a $Q \in \mathcal{C}_1(\ell) \otimes \mathcal{C}_2(\ell)$ such that*

$$\delta(Q, R) \le 2\sqrt{\epsilon}.$$

Proof. First, remove any row or column that has an error fraction greater than ϵ/γ'. This removes less than γ' fraction of the rows and columns (as $\delta(R,C) = \epsilon$). Call the sets of these deleted rows and columns $\mathsf{R}_1, \mathsf{C}_1$, respectively. After this, it is guaranteed that each row and column has at most $n\epsilon/\gamma'$ errors, and its new length is at least $n(1-\gamma')$, so it has an error fraction at most $\epsilon/(\gamma'(1-\gamma')) < \epsilon/\gamma$ by Inequality (1). The total error fraction of the matrix is still at most ϵ because the removed columns and rows each had error fractions above 2ϵ, so overall the removed entries had an error fraction above ϵ.

Next, choose a submatrix $M \subset ([n] \setminus \mathsf{R}_1) \times ([n] \setminus \mathsf{C}_1)$ of size $L \times L$ such that the error fraction within M is at most ϵ (in fact, the number of errors in M is at most $\lfloor L^2\epsilon \rfloor$). Such an M must exist because choosing a random submatrix results in an error fraction at most ϵ in expectation. WLOG, say that M lies in the top left of the matrix (so its rows and columns are indexed by $[L]$).

▷ **Claim 7.** Recall that $d = \lfloor \sqrt{\epsilon}L \rfloor + 2 + g$. There exist vectors $0 \neq E \in \mathcal{C}_1(d) \otimes \mathcal{C}_2(d)$ and $N \in \mathcal{C}_1(d + \ell) \otimes \mathcal{C}_2(d + \ell)$ such that

$$E(x,y)R(x,y) = E(x,y)C(x,y) = N(x,y), \ \ \forall (x,y) \in M.$$

Proof. Consider the projection of $\mathcal{C}_1(d) \otimes \mathcal{C}_2(d)$ onto the set of coordinates $\{(x,y) \in M \mid R(x,y) \neq C(x,y)\}$. Since $\mathcal{C}_1(d)\otimes\mathcal{C}_2(d)$ has dimension at least $(d-g)^2 = (\lfloor \sqrt{\epsilon}L \rfloor + 2)^2 > \epsilon L^2$, but $\{(x,y) \in M \mid R(x,y) \neq C(x,y)\}$ has size at most ϵL^2, the projection has a non-trivial kernel. Choose any nonzero E in this kernel.

We have ensured that $E(x,y) = 0$ whenever R and C disagree on submatrix M, so

$$E(x,y)R(x,y) = E(x,y)C(x,y) \ \ \forall (x,y) \in M.$$

Let's look at the projection of ER on the submatrix M. Every row is an element of $\mathcal{C}_1(d + \ell)|_{[L]}$ and every column is an element of $\mathcal{C}_2(d + \ell)|_{[L]}$ (by product property of AG codes). Thus, $ER|_M$ is an element of $\mathcal{C}_1(d + \ell)|_{[L]} \otimes \mathcal{C}_2(d + \ell)|_{[L]}$. By Proposition 3, any element of $\mathcal{C}_1(d + \ell)|_{[L]} \otimes \mathcal{C}_2(d + \ell)|_{[L]} = (\mathcal{C}_1(d + \ell) \otimes \mathcal{C}_2(d + \ell))|_M$ can be extended to an element of $\mathcal{C}_1(d + \ell) \otimes \mathcal{C}_2(d + \ell)$. Choose any such extension and call it N. So we have

$$E(x,y)R(x,y) = E(x,y)C(x,y) = N(x,y) \ \ \forall (x,y) \in M,$$

as desired. This proves Claim 7.   ◁

Next we extend the relation between E, R, C and N to almost the entire matrix.

▷ **Claim 8.** $E(x,y)R(x,y) = E(x,y)C(x,y) = N(x,y)$ for all (x,y) but γ fraction of the remaining rows and columns. More formally, there exist sets $\mathsf{R}_2 \subset [n] \setminus \mathsf{R}_1$ and $\mathsf{C}_2 \subset [n] \setminus \mathsf{C}_1$ such that the equality holds on all of $([n] \setminus (\mathsf{R}_1 \cup \mathsf{R}_2)) \times ([n] \setminus (\mathsf{C}_1 \cup \mathsf{C}_2))$, and that

$$|\mathsf{R}_2| \leq \gamma(n - |\mathsf{R}_1|) \text{ and } |\mathsf{C}_2| \leq \gamma(n - |\mathsf{C}_1|).$$

Proof. Call a row $y \in [n] \setminus \mathsf{R}_1$ "bad" if

$$\Pr_{x \sim [L]}[R(x,y) \neq C(x,y)] > \frac{\epsilon}{\gamma} \cdot \frac{1}{\gamma}.$$

In other words, a bad row is one that has too many errors in the columns of M. Define a bad column analogously. Let the sets of bad rows and columns be R_2 and C_2, respectively. Call the remaining rows and columns "good" (i.e. the good rows are $[n] \setminus (\mathsf{R}_1 \cup \mathsf{R}_2)$). The fraction $|\mathsf{R}_2|/(n - |\mathsf{R}_1|)$ must be at most γ because, among the first L columns, there is a combined error fraction of at most ϵ/γ (recall that each such column has at most ϵ/γ error fraction). The same holds for the columns.

Now we prove that $E(x,y) \cdot C(x,y) = N(x,y)$ on all good columns. Fix any good column $\hat{x}$. For any row $y_0 \in [L]$, since row vectors $E(x, y_0) \cdot R(x, y_0)$ and $N(x, y_0)$ belong to $\mathcal{C}_1(d + \ell)$ and agree on L points (i.e. the columns in $[L]$), they must be identical because the distance of the code $\mathcal{C}_1(d + \ell)$ is at least $n - d - \ell$ which is greater than $n - L$ (by Inequality (2)). So $E(\hat{x}, y_0) \cdot R(\hat{x}, y_0) = N(\hat{x}, y_0)$. Since $\hat{x}$ is good, there are at least $L\left(1 - \epsilon/\gamma^2\right)$ values of $y_0 \in [L]$ such that

$$E(\hat{x}, y_0) \cdot C(\hat{x}, y_0) = E(\hat{x}, y_0) \cdot R(\hat{x}, y_0) = N(\hat{x}, y_0).$$

Now, since

$$L\left(1 - \frac{\epsilon}{\gamma^2}\right) > d + \ell,$$

by Inequality (2), this tells us that the column vectors $E(\hat{x}, y) \cdot C(\hat{x}, y)$ and $N(\hat{x}, y)$ (which belong in $\mathcal{C}_2(d + \ell)$) are identical. Thus, $E(x,y) \cdot C(x,y) = N(x,y)$ on all good columns. The same argument can be applied on the rows to show that $E(x,y) \cdot R(x,y) = N(x,y)$ on all good rows. On the intersection of all good columns and rows, we have

$$E(x,y) \cdot R(x,y) = N(x,y) = E(x,y) \cdot C(x,y).$$

This proves Claim 8.   ◁

Now, we have the necessary claims to prove Lemma 6. Our first step is to find a submatrix of size $L \times L$ upon which $E(x,y)$ is never 0. To do this, consider the good submatrix $G = ([n] \setminus (\mathsf{R}_1 \cup \mathsf{R}_2)) \times ([n] \setminus (\mathsf{C}_1 \cup \mathsf{C}_2))$ upon which $E(x,y) \cdot R(x,y) = E(x,y) \cdot C(x,y)$ from Claim 8. Note that G has at least $n(1 - \gamma' - \gamma)$ rows and columns. Choose L columns $\{x_1, \ldots, x_L\} \subset [n] \setminus (\mathsf{C}_1 \cup \mathsf{C}_2)$ upon which E is not identically 0 (these must exist: simply consider a row upon which E is not identically 0, then choose among the $n(1 - \gamma' - \gamma) - d > L$ columns in $[n] \setminus (\mathsf{C}_1 \cup \mathsf{C}_2)$ in which the row takes a nonzero value).

In each of these columns, E is 0 at most d times (this follows by the distance of the code; the all-zero vector is always a codeword of a linear code). Thus, there must be $n(1 - \gamma' - \gamma) - dL$ rows in $[n] \setminus (\mathsf{R}_1 \cup \mathsf{R}_2)$ upon which E is never 0 in the columns $\{x_1, \ldots, x_L\}$. But since $n(1 - \gamma' - \gamma) - dL > L$ by Inequality (3), we can find an $L \times L$ submatrix of G upon which E is never 0. Call this submatrix M'. Since $E(x,y) \cdot R(x,y) = E(x,y) \cdot C(x,y)$ on $M' \subset G$, we know that $R(x,y) = C(x,y)$ on M'.

Permute the matrix so that the rows and columns of M' are indexed by $[L]$. Every row of M' is an element of $\mathcal{C}_1(\ell)|_{[L]}$ and every column of M' is an element of $\mathcal{C}_2(\ell)|_{[L]}$. Thus, by Proposition 3, the submatrix M' is an element of $\mathcal{C}_1(\ell)|_{[L]} \otimes \mathcal{C}_1(\ell)|_{[L]} = (\mathcal{C}_1(\ell) \otimes \mathcal{C}_2(\ell))|_{M'}$ and can be extended to some $Q \in \mathcal{C}_1(\ell) \otimes \mathcal{C}_2(\ell)$ on the entire matrix, such that $R(x,y) = C(x,y) = Q(x,y)$ on M'.

For this last step, we can bring back all of the rows and columns $\mathsf{R}_1, \mathsf{C}_1, \mathsf{R}_2, \mathsf{C}_2$ that were previously ignored. Let $\mathsf{R}_3 \subset [n]$ be the set of all rows that have an error fraction above ϵ/γ^2 among the first L columns. Since each column in $[L] \subset [n] \setminus \mathsf{C}_1$ has an overall error fraction at most ϵ/γ, we know that $|\mathsf{R}_3| < \gamma n$.

We will show that R is identical to Q on all rows not in R_3. Consider any row $\hat{y} \in [n] \setminus \mathsf{R}_3$. For any column $x_0 \in [L]$, we know that the column vectors $Q(x_0, y)$ and $C(x_0, y)$ belong to $\mathcal{C}_2(\ell)$ and are identical (using distance of the code), so $Q(x_0, \hat{y}) = C(x_0, \hat{y})$. Thus, for any x_0 such that $R(x_0, \hat{y}) = C(x_0, \hat{y})$, we have $Q(x_0, \hat{y}) = R(x_0, \hat{y})$ as well. Since we're assuming that there is an error fraction of at most ϵ/γ^2 among the first L columns of row $\hat{y}$, row vectors $Q(x, \hat{y})$ and $R(x, \hat{y})$ (in $\mathcal{C}_1(\ell)$) are equal in at least $(1 - \epsilon/\gamma^2)L$ places. Thus, $Q(x, \hat{y})$ is identical to $R(x, \hat{y})$ (using Inequality (2) and the fact that distance of code $\mathcal{C}_1(\ell)$ is at least $n - \ell$).

Finally, this means that $Q(x,y) = R(x,y)$ on all points in $([n] \setminus \mathsf{R}_3) \times [n]$. Since R_3 contains at most γ fraction of all rows, we know $\delta(Q, R) \leq \gamma = 2\sqrt{\epsilon}$. This proves Lemma 6.   ◄

Now we prove Theorem 5. If two $\mathcal{C}_1(\ell)$ row vectors are distinct, they must disagree on at least $n - \ell$ points (which follows from the distance of the code). Thus, the fraction of rows upon which R and Q disagree anywhere is at most $\frac{n}{n-\ell} 2\sqrt{\epsilon} < 3\sqrt{\epsilon}$ by Inequality (4). Similarly, the fraction of errors between C and Q is at most $2\sqrt{\epsilon} + \epsilon$, so the fraction of columns upon which C and Q disagree is at most $\frac{n}{n-\ell}(2\sqrt{\epsilon} + \epsilon) < 3\sqrt{\epsilon}$, again by Inequality (4). Permute the matrix so that these rows and columns are contiguous in the bottom right. Label the four regions of the matrix $A_{11}, A_{12}, A_{21}, A_{22}$:

$$
\begin{array}{|c|c|}
\hline
 & \\
A_{11} & A_{12} \\
 & \\
\hline
A_{21} & A_{22} \\
\hline
\end{array}
$$

where $A_{11} \cup A_{12}$ is the region where $R(x,y) = Q(x,y)$ and $A_{11} \cup A_{21}$ is the region where $C(x,y) = Q(x,y)$. The submatrix A_{22} has size at most $3\sqrt{\epsilon}n$ in each dimension. Now, notice that each row of A_{21} has at least $n - \ell - 3\sqrt{\epsilon}n$ disagreements between R and $Q = C$. Thus,

$$
\frac{\text{dist}_{A_{21}}(R,C)}{|A_{21} \cup A_{22}|} \geq \frac{n - \ell - 3\sqrt{\epsilon}n}{n} > \frac{1}{2}
$$

by Inequality (5), where $\text{dist}_M(R,C)$ is the number of disagreements between R and C on submatrix M. Applying the same logic to the columns, we get

$$
\frac{\text{dist}_{A_{12}}(R,C)}{|A_{12} \cup A_{22}|} > \frac{1}{2}.
$$

But now, since

$$\operatorname*{dist}_{A_{12}}(R, C) + \operatorname*{dist}_{A_{21}}(R, C) \le \operatorname{dist}(R, C) = \epsilon n^2,$$

we can combine these inequalities to get

$$|A_{12} \cup A_{22}| + |A_{21} \cup A_{22}| < 2\epsilon n^2.$$

Finally, since $\operatorname{dist}(Q, R) \le |A_{12} \cup A_{22}|$ and $\operatorname{dist}(Q, C) \le |A_{21} \cup A_{22}|$, we conclude that

$$\delta(Q, R) + \delta(Q, C) \le 2\epsilon.$$

This finishes the proof of Theorem 5. ◀

Proof of Theorem 4. We prove the theorem for $\rho = \epsilon_0/2$. The constants c_0 and ϵ_0 are defined in Theorem 5. For any $F \in \mathbb{F}_q^{n \times n}$, let $R \in \mathcal{C}_1(\ell) \otimes \mathbb{F}_q^n$ and $C \in \mathbb{F}_q^n \otimes \mathcal{C}_2(\ell)$ be the closest vectors to F in their respective codes. Let $\epsilon = \delta(R, C)$.

If $\epsilon \ge \epsilon_0$, then $\rho = \epsilon_0/2$ trivially works because

$$\frac{\epsilon_0}{2} \cdot \delta(F, \mathcal{C}_1(\ell) \otimes \mathcal{C}_2(\ell)) \le \frac{\epsilon_0}{2} \le \frac{1}{2}\delta(R, C) \le \frac{1}{2}[\delta(F, R) + \delta(F, C)].$$

Otherwise, if $\epsilon < \epsilon_0$, we apply Theorem 5 to find a $Q \in \mathcal{C}_1(\ell) \otimes \mathcal{C}_2(\ell)$ such that $\delta(Q, R) + \delta(Q, C) \le 2\epsilon$. Then we know

$$\begin{aligned}
\delta(F, Q) &\le \min(\delta(F, R) + \delta(R, Q), \delta(F, C) + \delta(C, Q)) \\
&\le \frac{1}{2}\left(\delta(F, R) + \delta(F, C) + \delta(R, Q) + \delta(C, Q)\right) \\
&\le \frac{1}{2}\left(\delta(F, R) + \delta(F, C) + 2\epsilon\right) \\
&\le \frac{3}{2}\left(\delta(F, R) + \delta(F, C)\right) \\
\frac{1}{3}\delta(F, Q) &\le \frac{1}{2}\left[\delta(F, R) + \delta(F, C)\right].
\end{aligned}$$

Since $\rho < 1/3$, the theorem is proven. ◀

References

1   Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998. `doi:10.1145/273865.273901`.

2   László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In Cris Koutsougeras and Jeffrey Scott Vitter, editors, *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 21–31. ACM, 1991. `doi:10.1145/103418.103428`.

3   Eli Ben-Sasson, Venkatesan Guruswami, Tali Kaufman, Madhu Sudan, and Michael Viderman. Locally testable codes require redundant testers. *SIAM J. Comput.*, 39(7):3230–3247, 2010. `doi:10.1137/090779875`.

4   Eli Ben-Sasson and Madhu Sudan. Robust locally testable codes and products of codes. *Random Struct. Algorithms*, 28(4):387–402, 2006. `doi:10.1002/rsa.20120`.

5   Don Coppersmith and Atri Rudra. On the robust testability of product of codes. *Electron. Colloquium Comput. Complex.*, TR05-104, 2005. URL: `https://eccc.weizmann.ac.il/eccc-reports/2005/TR05-104/index.html`.

6   David A Cox. *Primes of the Form x2+ ny2: Fermat, Class Field Theory, and Complex Multiplication. with Solutions*, volume 387. American Mathematical Soc., 2022.

7 Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 357–374. ACM, 2022. `doi:10.1145/3519935.3520024`.

8 Irit Dinur, Min-Hsiu Hsieh, Ting-Chun Lin, and Thomas Vidick. Good quantum LDPC codes with linear time decoders. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 905–918. ACM, 2023. `doi:10.1145/3564246.3585101`.

9 Irit Dinur, Madhu Sudan, and Avi Wigderson. Robust local testability of tensor products of LDPC codes. In Josep Díaz, Klaus Jansen, José D. P. Rolim, and Uri Zwick, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2006 and 10th International Workshop on Randomization and Computation, RANDOM 2006, Barcelona, Spain, August 28-30 2006, Proceedings*, volume 4110 of *Lecture Notes in Computer Science*, pages 304–315. Springer, 2006. `doi:10.1007/11830924_29`.

10 Iwan M. Duursma and Ralf Kötter. Error-locating pairs for cyclic codes. *IEEE Trans. Inf. Theory*, 40(4):1108–1121, 1994. `doi:10.1109/18.335964`.

11 Oded Goldreich and Or Meir. The tensor product of two good codes is not necessarily robustly testable. *Inf. Process. Lett.*, 112(8-9):351–355, 2012. `doi:10.1016/j.ipl.2012.01.007`.

12 Venkatesan Guruswami and Madhu Sudan. On representations of algebraic-geometry codes. *IEEE Trans. Inf. Theory*, 47(4):1610–1613, 2001. `doi:10.1109/18.923745`.

13 Tom Høholdt, Jacobus H Van Lint, and Ruud Pellikaan. Algebraic geometry codes. *Handbook of coding theory*, 1(Part 1):871–961, 1998.

14 Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Mip*=re. *CoRR*, abs/2001.04383, 2020. `doi:10.48550/arXiv.2001.04383`.

15 Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. Quantum soundness of testing tensor codes. *CoRR*, abs/2111.08131, 2021. `doi:10.48550/arXiv.2111.08131`.

16 Gleb Kalachev and Pavel Panteleev. Two-sided robustly testable codes. *CoRR*, abs/2206.09973, 2022. `doi:10.48550/arXiv.2206.09973`.

17 Gleb Kalachev and Pavel Panteleev. Maximally extendable product codes are good coboundary expanders. *CoRR*, abs/2501.01411, 2025. `doi:10.48550/arXiv.2501.01411`.

18 Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 403–412. ACM, 2008. `doi:10.1145/1374376.1374434`.

19 Anthony Leverrier and Gilles Zémor. Quantum tanner codes. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 872–883. IEEE, 2022. `doi:10.1109/FOCS54457.2022.00117`.

20 Or Meir. Combinatorial construction of locally testable codes. *SIAM J. Comput.*, 39(2):491–544, 2009. `doi:10.1137/080729967`.

21 Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 375–388. ACM, 2022. `doi:10.1145/3519935.3520017`.

22 Ruud Pellikaan. On decoding by error location and dependent sets of error positions. *Discret. Math.*, 106-107:369–381, 1992. `doi:10.1016/0012-365X(92)90567-Y`.

23 Alexander Polishchuk and Daniel A. Spielman. Nearly-linear size holographic proofs. In Frank Thomson Leighton and Michael T. Goodrich, editors, *Proceedings of the Twenty-Sixth Annual ACM Symposium on Theory of Computing, 23-25 May 1994, Montréal, Québec, Canada*, pages 194–203. ACM, 1994. `doi:10.1145/195058.195132`.

24 Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996. `doi:10.1137/S0097539793255151`.

25 Henning Stichtenoth. *Algebraic function fields and codes*, volume 254 of *Universitext*. Springer, 1993.

26 Paul Valiant. The tensor product of two codes is not necessarily robustly testable. In Chandra Chekuri, Klaus Jansen, José D. P. Rolim, and Luca Trevisan, editors, *Approximation, Randomization and Combinatorial Optimization, Algorithms and Techniques, 8th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2005 and 9th InternationalWorkshop on Randomization and Computation, RANDOM 2005, Berkeley, CA, USA, August 22-24, 2005, Proceedings*, volume 3624 of *Lecture Notes in Computer Science*, pages 472–481. Springer, 2005. `doi:10.1007/11538462_40`.

27 Michael Viderman. Strong ltcs with inverse poly-log rate and constant soundness. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 330–339. IEEE Computer Society, 2013. `doi:10.1109/FOCS.2013.43`.

28 Michael Viderman. A combination of testability and decodability by tensor products. *Random Struct. Algorithms*, 46(3):572–598, 2015. `doi:10.1002/rsa.20498`.