# Symmetric LDPC codes and local testing

Tali Kaufman
MIT
kaufmant@mit.edu

Avi Wigderson
Institute for Advanced Study
avi@ias.edu

April 8, 2014

**Abstract**

Coding theoretic and complexity theoretic considerations naturally lead to the question of generating symmetric, sparse, redundant linear systems. This paper provides a new way of construction with better parameters and new lower bounds.

Low Density Parity Check (*LDPC*) codes are linear codes defined by short constraints (a property essential for *local testing* of a code). Some of the best (theoretically and practically) used codes are LDPC. *Symmetric* codes are those in which all coordinates "look the same", namely there is some transitive group acting on the coordinates which preserves the code. Some of the most commonly used locally testable codes (especially in PCPs and other proof systems), including all "low-degree" codes, are symmetric. Requiring that a symmetric binary code of length $n$ has large (linear or near-linear) distance seems to suggest a "conflict" between 1/rate and density (constraint length). In known constructions, if one is constant then the other is almost worst possible – $n/poly(\log n)$.

Our main positive result simultaneously achieves *symmetric* low density, constant rate codes generated by a *single* constraint. We present an *explicit* construction of a symmetric and transitive binary code of length $n$, near-linear distance $n/(\log\log n)^2$, of constant rate and with constraints of length $(\log n)^4$. The construction is in the spirit of Tanner codes, namely the codewords are indexed by the edges of a sparse regular expander graph. The main novelty is in our construction of a transitive (non Abelian!) group acting on these edges which preserves the code. Our construction is one instantiation of a framework we call *Cayley Codes* developed here, that may be viewed as extending zig-zag product to symmetric codes.

Our main negative result is that the parameters obtained above cannot be significantly improved, as long as the acting group is solvable (like the one we use). More specifically, we show that in constant rate and linear distance codes (aka "good" codes) invariant under solvable groups, the density (length of generating constraints) cannot go down to a constant, and is bounded below by $\log^{(\Omega(\ell))} n$ (an $\Omega(\ell)$ iterated logarithm) if the group has a derived series of length $\ell$. This negative result precludes natural local tests with constantly many queries for such solvable "good" codes.

# 1 Introduction

The work in this paper is partially motivated from several (related) research directions. We first give a very high level description of these, and then proceed to describe our results.

## 1.1 Motivation

**Locally testable codes** Codes in which the proximity to a codeword can be determined by a few coordinate queries have proven a central ingredient in some major results in complexity theory. They appear as low-degree tests in the $IP = PSPACE$, $MIP = NEXP$ and $PCP = NP$ theorems, and indeed the work of [16] (which was later partly derandomized by [8]) elucidates their role as the "combinatorial heart" of PCPs. The quest to simultaneously optimize their coding theoretic parameters and the number of queries used has recently culminated in the combination of [7] and [13] (see also [28]) in a length $n$ binary linear code of linear distance and rate $1/(\log n)^{O(1)}$, testable with a constant number of queries (which are testing linear constraints of constant length). Further improving the rate to a constant is a major open problem. Essential to locally testable codes is having short constraints.

**LDPC codes** Low Density Parity Check codes are precisely linear codes with short constraints. Density is the constraints length. These codes were defined in the seminal work of Gallager [14] in the 60's. Only in the 90's, due to works of [25, 33, 35] and others did LDPC codes start to compete with the algebraic constructions in the coding theory scene. Today these provide some of the best practical and theoretical codes for many noise models, and are extremely efficient to encode and decode. In particular, they can achieve linear distance, constant rate and constant constraint size simultaneously. But their natural potential for local testing was (possibly) devastated by such results as [6], who showed that a general class of LDPC codes, based on expanders, requires a linear number of queries to test, despite having constant-size constraints. We note that possessing short defining constraints is not always an obvious property of a code – e.g., it was only recently discovered in [18] that the sparse dual-BCH codes have such constraints (but unfortunately these codes have a very bad rate).

**Symmetric codes** Many of the classical codes, e.g., Hamming, Reed-Solomon, Hadamard, Reed-Muller, BCH, and some Goppa codes are symmetric, namely there is a transitive group acting on the coordinates which leaves the code invariant. Symmetry is not only elegant mathematically – it often also implies concise representation of the code as well as tools to analyze its quality parameters, like rate and distance. Huge literature is devoted to such codes within coding theory, but even for cyclic codes (those invariant under cyclic shifts) it is still a major open problem if they can have simultaneously constant rate and linear distance. The conjecture is that this is impossible. A major result of Berman from the seventies [9] shows that there are no good cyclic codes of length $n$ where all the prime divisors of $n$ are bounded. Interesting progress on this conjecture was made by Babai Shpilka and Stefankovic [5] that extend Berman's result and relax the conditions on the sizes of the prime divisors of the code length. Moreover [5] show that the conjecture is true if one requires the cyclic code to be defined by constraints of constant length (i.e., to be LDPC). McElice [27] proved (non constructively) that there are asymptotically good *non-linear* codes invariant under the action of very large groups, however these codes are clearly not LDPC.

**Symmetric low-density and locally testable codes** Starting with linearity testing of [10] and the first low-degree tests of [4, 31], nearly all locally testable codes appearing in proof systems *are* symmetric. A theory studying the extent to which symmetry can help (or handicap) local testing was initiated by Kaufman and Sudan [17]. They generalized known examples showing that when the acting group is the affine group (and the coordinates are naturally identified with the elements of the vectors space acted upon), then having short constraints that define the code is not only necessary, but also *sufficient* for local testability. Moreover, in these cases the orbit (under the group action) of a *single* constraint suffices to define the code, and a canonical local test is picking a random constraint from that orbit[1]. Again, the rate of all these codes is poor, and [17] challenge reconciling the apparent conflict between rate and density, possibly for other groups.

**Expanding Cayley graphs** Gallager's construction [14] of LDPC codes was based on sparse random graphs, and Tanner's construction [38] was based on high girth graphs. Sipser and Spielman [35] identified *expansion* as the crucial parameter of graphs which yield codes with good parameters. This was followed up in almost all subsequent works, using expanders to construct codes. This work motivated further explicit constructions of good expanders. As example, we note that the [35] "belief propagation" decoding algorithm for LDPC was simplest if the underlying graph is a *lossless* expander, and subsequently [11] were able to explicitly construct such expanders. *Unfortunately, all codes constructed this way seem far from symmetric.* But expander *graphs* can certainly be symmetric! Indeed, almost all constructions of expander graphs are Cayley graphs, namely the vertices correspond to the elements of a finite group, and edges are prescribed by a fixed generating set of the group. It is evident that such graphs are symmetric, namely the group itself acts transitively on the vertices and preserves the edges. We note importantly that even the zig-zag product construction of expanders [34], which started as a combinatorial alternative to algebraic constructions, was extended to allow iterative probabilistic constructions of Cayley graphs [2, 29] via the semi-direct product of groups. Our codes are partially motivated by making explicit the probabilistic construction of [2, 29]. Attempts to construct codes iteratively exist, with the best example being Meir's, partially explicit construction [28]. However, again, this code is far from symmetric.

Several natural research directions point to the following question: **To what extent can symmetric LDPC codes attain (or even come close to) the coding theory gold standard of linear distance and constant rate?** To fix ideas, let us consider symmetric codes with linear (or even near-linear) distance, and examine the trade-off between density and 1/rate. In all known codes if 1/rate or density is constant then the other is *worst* possible, about $n/poly(\log n)$, the code length! Best density/rate trade-offs for known binary high distance symmetric codes are the following. Reed-Muller codes over binary field (say degree-$d$ polynomials), which are invariant under the affine group, have short constraints ($2^d$-long) but pathetic rate $(\log n)^d/n$. BCH codes, invariant under the cyclic group, have constant rate, but constraints of (worst possible) length $\Omega(n)$. Reed-Muller codes over large fields concatenated with Hadamard achieve density $(\log n)^{1/\epsilon}$ with (1/rate) being $2^{(\log n)^\epsilon}$ [3, 37].[2]

[1] We note that the existence of a *single* constraint that generates a code gives rise to a canonical algorithm for local testing the code. An algorithm that picks a random constraint from the orbit. For codes invariant under the affine group, Kaufman and Sudan have shown that such a canonical algorithm is indeed a valid local tester for the code. This motivates the search for other symmetric codes generated by the orbit(s) of one (or few) generators, with the hope that local testing would be implied.

[2] Note that when this code is mostly used to get constant query complexity, it is modified to make coordinates

Indeed, some believed that the conflict between density and rate in symmetric codes cannot be reconciled. On the other hand, no result precludes the ratio of density/rate from being *best* possible, namely a constant! Our paper addresses both upper and lower bounds on this trade-off.

## 1.2 Our Results and techniques

We first describe our upper bounds and then our lower bounds.

### 1.2.1 Upper bounds

Our main positive result allows simultaneous constant rate and polylogarithmic density, and in particular reduces the upper bound on the ratio density/rate to $poly \log n$! More precisely, Theorem 11 gives an explicit construction of length-$n$ symmetric codes of constant rate and distance $n/(\log \log n)^2$ which is defined by constraints of a length $poly(\log n)$. Moreover, these constraints constitute the orbit of a *single* constraint, under the transitive action of a (non Abelian) group.

In order to prove this result, we develop a framework of "Cayley Codes", which we describe next. They extend Tanner codes in that the coordinates of the code are identified with the edges of a regular expander graph, and constraints are imposed on neighborhoods (namely edges incident on each single vertex) according to a fixed "inner code" $B$. In Cayley codes we naturally insist that the underlying graph is a Cayley graph, namely the vertices are the elements of a group $G$, and a set of generators $S$ of the group determine edges in a natural way. While this a graph is symmetric ($G$ acts transitively on its vertices), there is no such guarantee in general for the code. The problem is to find a group that acts on the *edges* of the graph, and preserves all copies of the internal code. We show that if some group $H$ simultaneously acts transitively on the code $B$ and acts on the group $G$, then the semi-direct product group $G \rtimes H$ acts transitively on the edges. We note that this action is not standard.

We then turn to find an appropriate instantiation of this idea with good parameters. This paragraph is a bit technical and may be skipped at first reading. The group $G$ is chosen to be the hypercube $\mathbb{F}_2^t$, and $S$ a very specific $\epsilon$-biased set [30] in $G$ (so as to make the associated Cayley graph expanding), which can be identified with the elements of a cyclic group $H$ isomorphic to the multiplicative group of of $F_{t^4}^*$. The inner code $B$ is chosen to be a BCH code on $S$ on which the group $H$ acts transitively. The inferior distance and density of the code $B$ are mitigated since its length is only polylogarithmic in the length of the whole code. Now the action of $H$ on $G$ (whose nature we describe in the technical section) allows the construction of the semi-direct product $G \rtimes H$. We now define the action of this group on directed edges of the graph, and prove that all parts fit: this group acts transitively on the Tanner code of the Cayley graph on $G; S$.

Alon, Lubotzky and Wigderson [2] provided a randomized construction of high rate high distance codes generated by two orbits. They asked about *explicit* constructions of high rate, high distance codes generated by few orbits (for the group they studied). Our construction provides such explicit codes generated by *one* orbit!

correspond not to the value of the encoded polynomial on a point, but rather its value on an entire line or larger subspace. This has lousy rate, and when derandomized to improve the rate, transitivity of the action is lost.

### 1.2.2 Lower bounds

The second result (Theorem 13) shows that there is no good code invariant under a solvable group with few low-weight generators. In fact we rule out the possibility of such codes even if the support of their generators is $o(\log^{c\ell} n)$ if the group has a derived series of length $\ell$, $n$ is the code length, and $c > 0$ is a consatnt. This result exclude the possibility of good solvable locally testable codes with few low weight generators. Note that the codes we have constructed in the upper bound section are solvable. Our methods extend work of Lubotzky and Weiss [24], who showed a similar lower bound on the number of generators Cayley graphs on these groups to be expanders. The extension is in two directions – we show the same for Schreier graphs, and then extend their argument from finding standard separators to finding $\epsilon$-partitions of the graph to many parts – from which we can deduce information on the distance and rate of the associated Tanner codes.

A work by Babai,Shpilka and Stefankovic [5] showed that there are no good cyclic codes with low weight constraints (with no restriction on the number of generating constraints). Since low weight constraints are a necessary (but not sufficient) condition for testability, they showed that there are no good cyclic locally testable codes. Our work here shows that there are no good solvable locally testable codes with few low weight generating constraints. I.e., we exclude good locally testable codes over larger groups of symmetry but under the assumption of few low weight generating constraints. As far as we know, it could well be the case that a cyclic code whose dual has a low weight basis must have a basis that is generated by a constantly many low-weight constraints.

The proof showing that there are no good solvable codes with few low weight generators has two main parts. First, for a parameter $\epsilon$ (later taken to be $o(1)$) we define a new notion that we call an $\epsilon$-partition of a graph, which extends the notion of a small separator, in that we demand that the separating set splits the graph into *many* pieces. More precisely, a graph has an $\epsilon$-partition (Definition 22) if one can remove $\epsilon$ fraction of its vertices to make all connected components of relative size at most $\epsilon$. We show that a Schreier graph of a solvable group with $d = o(\log^{c\ell} n)$ generators has an $\epsilon$-partition where $\epsilon$ is sub-constant. The proof of this part is by induction on the derived length of the group (Lemma 18) with the Abelian being the base case (Lemma 16). This extends a technique of Lubotzky and Weiss [24] from Cayley to Schrier graphs, and from separators to $\epsilon$-partitions.

In the second part of the proof (Lemma 21) we associate codes invariant under groups with Schreier graphs over these groups (see Definition 26), and show that if the associated Schreier graph has an $\epsilon$-partition then either the rate or the relative distance of the code is bounded by $\epsilon$.

### 1.2.3 Subsequent work

A recent work of Kaufman and Lubotzky [19], building on the framework presented here, solves one of the main open questions arising form our work. Namely, they show an explicit construction of a good symmetric LDPC code, whose constraint space is generated by the orbit of one *constant* weight constraint. Their construction is based on the Ramanujan graphs constructed by Lubotzky-Samuels-Vishne [22, 23] as a special case of Ramanujan complexes. The crucial point is that these graphs are edge transitive and not just vertex transitive as in previous constructions of Ramanujan graphs.

#### 1.2.4 Organization.

In Section 2 we present some general definitions to be used through out the work. In Section 3 we introduce the notion of Cayley codes, and present our main conceptual theorem (Theorem 7) showing sufficient conditions under which symmetric codes with one generator exist. In Section 4 we show our explicit construction of an almost good code with one generator of polylogarithmic weight (Theorem 11). Finally, we show that there are no good Abelian and solvable codes with few low weight generators (Theorem 12, Theorem 13). These lower bounds appear in Section 5.

## 2 Some General Definitions

We start with some basic definitions that are being used throughout this work. The definitions deal with group actions, Abelian and solvable groups, linear codes, Cayley graphs and Schreier graphs.

### 2.1 Group theory definitions

**Definition 1** (Group automorphism)**.** *A* group automorphism *is an isomorphism from a group to itself. The automorphisms of a group $G$ form a group, denoted $Aut(G)$.*

**Definition 2** (The symmetric group)**.** *The* symmetric group *on a set $X$, denoted by $Sym(X)$, is the group of permutations of the set $X$. The composition of two permutations $f, g$ (viewed as bijective functions on $X$) is denoted $f \circ g$.*

**Definition 3** (Action of a group on a group)**.** *An* action *of a group $H$ on a group $G$ is a group homomorphism $\phi : H \to Aut(G)$. In other words, each element $h \in H$ corresponds to an automorphism $\phi_h$ of $G$, where $\phi_{h_1 \cdot h_2} = \phi_{h_1}\phi_{h_2}$. Let $g^h = \phi_h(g)$ denote the action of $h \in H$ on $g \in G$.*

**Definition 4** (Action of a group on a set, transitivity)**.** *An* action *of a group $H$ on a set $X$ is a group homomorphism $\phi : H \to Sym(X)$ that sends each element $h$ to a permutation of the elements of $X$. Let $x^h$ denote the action of $h \in H$ on $x \in X$. That is, an action should satisfy for every $h, h' \in H, x \in X$*

$$x^{hh'} = (x^{h'})^h$$

*The action is called* transitive *if for every $x_1, x_2 \in X$ there exists $h \in H$ such that*

$$x_1^h = x_2$$

**Fact 1.** *If a group $H$ acts on sets $G$, $S$, then $H$ acts also on $G \times S$ in the obvious way.*

**Definition 5** (Orbit)**.** *Suppose a group $H$ acts on a set $X$. The* orbit *of an element $x \in X$ under the action of $H$ is the set*

$$x^H = \{x^h | h \in H\}.$$

**Definition 6** (Semi-direct product group)**.** *Suppose a group $H$ acts on a group $G$. The* semidirect product *$G \rtimes H$ is a group whose elements are pairs $(g, h)$ where $g \in G$ and $h \in H$. The operation of the group is given by*

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 \cdot g_2^{h_1^{-1}}, h_1 \cdot h_2).$$

**Definition 7** (Abelian group)**.** *A finite group $G$ is Abelian if for every $a, b \in G$, $ab = ba$.*

**Definition 8** (Commutator subgroup)**.** *For a group $G$, the* commutator subgroup *of $G$ denoted $[G, G]$ is a subgroup of $G$ generated by all commutators of $G$, where a* commutator *of $g, h \in G$ is $[g, h] = g^{-1}h^{-1}gh$. The identity element of $G$ is the only commutator iff $G$ is Abelian.*

**Definition 9** (Normal subgroup)**.** *For a group $G$, a subgroup $H$ of $G$ is* Normal *if for every $g \in G$, $gHg^{-1} = H$.*

**Definition 10** (Quotient group)**.** *For a group $G$ and a normal subgroup $H$ of $G$, the* quotient group *of $H$ in $G$, written $G/H$ is the set of cosets of $H$ in $G$, with the operation $(Ha)(Hb) = Hab$.*

**Fact 2.** *For a group $G$, the quotient group $G/[G, G]$ is Abelian.*

**Definition 11** (Solvable group)**.** *For a group $G$ denote $G_0 = G$, and $G_{i+1} = [G_i, G_i]$. $G$ is* solvable *of derived length $\ell$ if $G_{\ell-1} \neq 1$ and $G_\ell = 1$, the trivial group of size 1. Also note that by Fact 2 $G_i/G_{i+1}$ is an Abelian group for $i = 0, \cdots \ell - 1$.*

Note that Abelian groups are solvable with derived length 1.

## 2.2 Codes definitions

**Definition 12** (Linear code, length, dimension, rate, distance)**.** *A code $C \subseteq \mathbb{F}^X$ is linear if $C$ forms a linear subspace over $\mathbb{F}$. The orthogonal space to $C$ is denoted as* the dual-code $C^\perp$*. The* length of the code $C$ *is $|X|$. The* dimension of the code $C$ *is the dimension of the subspace that is described by $C$. The* rate of the code $C$ *denoted as $r_C$ is the dimension of the code divided by its length. A* weight of a codeword $c \in C$ *denoted as $w(c)$ is the number of non-zero elements in $c$. The* distance *of $C$ is the minimum weight of a non-zero codeword of $C$, divided by the length of $C$.*

**Definition 13** (Concatenated code)**.** *Given linear codes $C_i \subseteq \mathbb{F}^{X_i}$, $1 \leq i \leq n$, a* concatenated code *$C \subseteq \mathbb{F}^{X_1 \cup \cdots \cup X_n}$ is a linear code such that $c \in C$ iff $c|_{X_i} \in C_i$ for every $i$, where $c|_{X_i}$ denotes the restriction of $c$ to the coordinates $X_i$.*

**Definition 14** (Code invariant under a group)**.** *An action of a group $H$ on a set $X$, induces an action of $H$ on the set $\mathbb{F}^X = \{f : X \to \mathbb{F}\}$ that is defined as follows. For $f \in \mathbb{F}^X$, $h : X \to X$, $f^h = f \circ h \in \mathbb{F}^X$. A code $C \subseteq \mathbb{F}^X$ is said to be $H$-*invariant *if*

$$C = HC \stackrel{\text{def}}{=} \{f^h | f \in C, h \in H\}$$

*$C$ that is $H$-invariant is also $H$-transitive if the action of $H$ on $X$ is transitive (by definition 4).*

**Definition 15** (Short generators, Support of generators)**.** *A linear code $C \subseteq \mathbb{F}^X$ that is $H$-invariant has $(r, w)$-*short generators *if there exist $r$ codewords $c_1, ...c_r \in C^\perp$ each of weight at most $w$, such that the vectors $c_1^H \bigcup \cdots \bigcup c_r^H$ span $C^\perp$. For a generator $c_i \in C^\perp$, let $s(c_i) \subseteq X$ be the coordinates on which $c_i$ has non-zero support. The* support of the given generators *of $C$ is $s(c_1) \bigcup \cdots \bigcup s(c_r) \subseteq X$.*

### 2.3 Graph definitions

**Definition 16** (Cayley graph)**.** *Given a group $G$ and a set of generator $S \subset G$, the* Cayley graph *$Cay(G, S)$ is a graph, whose vertices are labeled by elements of $G$, and there is an edge $(u, v)$ iff there exist $s \in S$, such that $u \cdot s = v$. When set $S$ is symmetric (i.e., $s \in S$ iff $s^{-1} \in S$), then the graph $Cay(G, S)$ is undirected.*

**Definition 17** (Schreier graph)**.** *Given a group $G$ that acts on a set $X$ and a set of generator $S \subseteq G$, the* Schreier graph *$Shr(G, X, S)$ is a graph, whose vertices are labeled by elements of $X$, and there is an edge $(x_1, x_2)$ iff there exist $s \in S$, such that $x_2 = x_1^s$ (recall that $x_1^s$ denotes the action of $s$ on $x_1$).*

**Definition 18** (Edge-transitive graph)**.** *A graph $H = (V, E)$ is* edge-transitive *if there exists a group $G$ that acts on the edges of the graph and preserve the graph, and such that for all pairs of edges $e_1, e_2 \in E$ there exists an element $g \in G$ such that $g(e_1) = e_2$.*

## 3 Cayley Codes

In this section we present our conceptual contribution. We define Cayley codes and show conditions under which these codes are invariant under a group that acts transitively on them. Moreover, we derive conditions, under which these codes are generated by few low weight generators under the action of that group.

**Definition 19** (Cayley code)**.** *Given a Cayley Graph $Cay(G, S)$ and a linear code $B \subseteq \mathbb{F}^S$ of length $|S| = d$ define the linear* a Cayley code *$Cay(G, S, B) \subseteq \mathbb{F}^{|G| \cdot |S|/2}$ as follows. Its coordinates are the $|G| \cdot |S|/2$ undirected edges of the graph $Cay(G, S)$, namely the pairs $\{(g, s_i), (gs_i, s_i^{-1})\}$, $g \in G, s_i \in S = \{s_1, \cdots, s_d\}$. The defining linear constraints are the local constraints of $B$ on the edges incident on every vertex, namely*

*$c \in Cay(G, S, B)$ iff for every $g \in G$ the following* vertex consistency *condition holds:*

$$(c_{\{(g,s_1),(gs_1,s_1^{-1})\}}, \cdots, c_{\{(g,s_d),(gs_d,s_d^{-1})\}}) \in B. \tag{1}$$

The following lemmas are concerned with the rate and distance of the Cayley codes.

**Lemma 1** (Rate of Cayley codes)**.** *The code $Cay(G, S, B)$ with $B$ of block length $|S| = d$ and rate $r_B > \frac{1}{2}$, has rate at least $2r_B - 1$.*

*Proof.* The rate of the code follows immediately from the definition. Consider $c \in Cay(G, S, B)$. $c$ is of length $|G||S|/2$.

The codeword $c$ obeys $|G|$ vertex constraints of the form $(c_{\{(g,s_1),(gs_1,s_1^{-1})\}}, \cdots, c_{\{(g,s_d),(gs_d,s_d^{-1})\}}) \in B$, for every $g \in G$. Each such vertex constraint imposes $|S|(1 - r_B)$ constraints of the code $B$. The total number of constraints imposed is $|G||S|(1 - r_B)$. Hence, the rate of the code is $1 - 2(1 - r_B) = 2r_B - 1$. □

**Lemma 2** (Distance of Cayley codes)**.** *Consider the Cayley code $Cay(G, S, B)$ for a linear code $B$ of block length $|S| = d$, and minimum relative distance $\delta$. If the Cayley graph $Cay(G, S)$ is an expander with normalized second largest eigenvalue $\lambda$ then the code $Cay(G, S, B)$ has distance at least $\left(\frac{\delta-\lambda}{1-\lambda}\right)^2$.*

*Proof.* Follows immediately from [35] (Lemma 15). The idea is that the distance of the small code is propagated to the large code due to the expansion of the graph on which the large code is defined. □

In the following we define an action of the semi-direct product group and show that the Cayley graph $Cay(G, S)$ is edge transitive under this action. This is then used for showing that the Cayley codes are invariant under this group that acts transitively on them.

**Definition 20** (Action of the semi-direct product group)**.** *Let $G$ be a group, $S \subseteq G$ a symmetric set. Let $H$ be a group such that $H$ acts on $G$, and $H$ preserves $S$. In such case the semi-direct product $G \rtimes H$ is defined. We define the* action of the semi-direct product $G \rtimes H$ on $G \times S$ *to be as follows for $(g, h) \in G \rtimes H$, $(g', s) \in (G \times S)$.*

$$(g, h)(g', s) = (gg'^{h^{-1}}, s^{h^{-1}})$$

In the following we show that the defined action of the semi direct product group $G \rtimes H$ on $G \times S$ is indeed an action.

**Lemma 3** (The action of the semi-direct product group is valid)**.** *The action of the semi-direct product $G \rtimes H$ on $G \times S$ defined in Definition 20 is indeed an action.*

*Proof.* The defined action of $G \rtimes H$ on $G \times S$ acts on $G$ and on $S$ separately. The separate actions are well defined since $H$ acts on $G$ and $H$ acts on $S$. The combined action of $G \rtimes H$ on $G \times S$ is then well defined by Fact 1, as $H$ acts on $G \times S$ and so is $G$. □

In the following we show conditions under which the Cayley graph $Cay(G, S)$ is edge transitive under the above action of the semi-direct product group $G \rtimes H$.

**Lemma 4** (Conditions for the Cayley graph to be edge-transitive under the action of the semi-direct product group)**.** *Let $G$ be a group, $S \subseteq G$ a symmetric set. Let $H$ be a group such that:*

- *$H$ acts on $G$.*
- *$H$ acts transitively on $S$.*

*Then the graph $Cay(G, S)$ is edge-transitive under the action of the semi-direct product group $G \rtimes H$.*

*Proof.* Since $H$ acts on $G$ and $H$ acts on $S$ the semi-direct product group $G \rtimes H$ is well defined as well as its action on $G \times S$. For proving the edge-transitivity we first need to show that edges

are mapped to edges (edge preservation), and that every edge can be mapped to every other edge (transitivity).

**Edge preservation.** We need to show that undirected edges (which are pairs of anti-directional directed edges) are mapped to undirected edges, namely that for $(g,h) \in G \rtimes H$, $(g',s) \in G \times S$

$$\begin{aligned}(g,h)(g',s) &= (g'^{(g,h)}, s^{(g,h)}) \\ (g,h)(g's,s^{-1}) &= ((g's)^{(g,h)}, (s^{-1})^{(g,h)})\end{aligned}$$

is such that $(g's)^{(g,h)} = g'^{(g,h)} \cdot s^{(g,h)}$, and $(s^{(g,h)})^{-1} = (s^{-1})^{(g,h)}$.

Indeed,

$$\begin{aligned}&(g's)^{(g,h)} = g(g's)^{h^{-1}} = gg'^{h^{-1}}s^{h^{-1}} = g'^{(g,h)} \cdot s^{(g,h)} \ ; \\ &(s^{(g,h)})^{-1} = (s^{h^{-1}})^{-1} = (s^{-1})^{h^{-1}} = (s^{-1})^{(g,h)}\end{aligned}$$

**Transitivity.** We need to show that the action of $G \rtimes H$ on $G \times S$ is transitive. That is, we need to show that for every $(g',s),(g'',s'') \in G \times S$ there exists $(g,h) \in G \rtimes H$ such that $(g,h)(g',s) = (g'',s'')$.

Recall, that the action of $(g,h) \in G \rtimes H$ on $(g',s) \in G \times S$ is of the form $(g,h)(g',s) = (g'^{(g,h)}, s^{(g,h)}) = (gg'^{h^{-1}}, s^{h^{-1}})$.

Pick $h$ such that $s'' = s^{h^{-1}}$ (this can be done since $H$ acts transitively on $S$).

Pick $g$ such that $g'' = gg'^{h^{-1}}$.

Hence the transitivity is obtained. □

In the following we show conditions under which the Cayley code $Cay(G,S,B)$ is invariant under the above action of the semi-direct product group $G \rtimes H$.

**Lemma 5** (Conditions for the Cayley code to be invariant and transitive under the semi-direct product group)**.** *Let $G$ be a group, $S \subseteq G$ a symmetric set. Let $B \subseteq \mathbb{F}^S$ be a linear code. Let $H$ be a group such that:*

- *$H$ acts on $G$.*
- *$H$ acts transitively on $S$.*
- *$B$ is $H$-invariant*

*Then the code $Cay(G,S,B) \subseteq \mathbb{F}^{|G||S|/2}$ is invariant under the action of the semi-direct product group $G \rtimes H$. Moreover, the code $Cay(G,S,B)$ is transitive under this group.*

*Proof.* By Lemma 4 the semi-direct product group $G \rtimes H$ acts transitively on the edges of the Cayley graph $Cay(G,S)$. Assume $|G| = n, |S| = d$. For showing that the code $Cay(G,S,B) \subseteq \mathbb{F}^{|G||S|/2}$ is

invariant under the semi-direct product group $G \rtimes H$ we need to show that for every $(g,h) \in G \rtimes H$, $c = (c_{\{(g_1,s_1)(g_1 s_1, s_1^{-1})\}}, \cdots, c_{\{(g_n,s_d)(g_n s_d, s_d^{-1})\}}) \in Cay(G,S,B)$ we have

$$(g,h) \cdot c = (c_{\{(g_1^{(g,h)}, s_1^{(g,h)})((g_1 s_1)^{(g,h)}, (s_1^{-1})^{(g,h)})\}}, \cdots,$$
$$c_{\{(g_n^{(g,h)}, s_d^{(g,h)}),((g_n s_d)^{(g,h)}, (s_d^{-1})^{(g,h)})\}}) \in Cay(G,S,B)$$

For the last equality we need to show that vertex consistency in (1) is preserved.

**Vertex consistency.** We show that for every $1 \leq i \leq n$:

$$(c_{\{(g_i^{(g,h)}, s_1^{(g,h)})((g_i s_1)^{(g,h)}, (s_1^{-1})^{(g,h)})\}}, \cdots,$$
$$c_{\{(g_i^{(g,h)}, s_d^{(g,h)}),((g_i s_d)^{(g,h)}, (s_d^{-1})^{(g,h)})\}}) \in B.$$

By the assumption that $c \in Cay(G,S,B)$, we know that for every $1 \leq i \leq n$:

$$(c_{\{(g_i,s_1)(g_i s_1, s_1^{-1})\}}, \cdots, c_{\{(g_i,s_d)(g_i s_d, s_d^{-1})\}}) \in B.$$

Hence, similarly, for every $1 \leq i \leq n$:

$$(c_{\{(g_i^{(g,h)}, s_1)(g_i^{(g,h)} s_1, s_1^{-1})\}}, \cdots,$$
$$c_{\{(g_i^{(g,h)}, s_d),(g_i^{(g,h)} s_d, s_d^{-1})\}}) \in B$$

This is since we only got the $G$ coordinate permuted. Since by assumption the code $B$ is $H$-invariant we also know that for every $h^{-1} \in H$ (and every $1 \leq i \leq n$)

$$(c_{\{(g_i^{(g,h)}, s_1^{h^{-1}})(g_i^{(g,h)} s_1^{h^{-1}}, (s_1^{h^{-1}})^{-1})\}}, \cdots,$$
$$c_{\{(g_i^{(g,h)}, s_d^{h^{-1}}),(g_i^{(g,h)} s_d^{h^{-1}}, (s_d^{h^{-1}})^{-1})\}}) \in B.$$

Given the action of the semidirect product group $G \rtimes H$ on the $S$ coordinate in $G \times S$ (where the action on $S$ is performed only by $H$), and where $(g's')^{(g,h)} = g'^{(g,h)} \cdot s'^{(g,h)}$, and $(s'^{(g,h)})^{-1} = (s'^{-1})^{(g,h)}$ we obtain that for every $1 \leq i \leq n$:

$$(c_{\{(g_i^{(g,h)}, s_1^{(g,h)})((g_i s_1)^{(g,h)}, (s_1^{-1})^{(g,h)})\}}, \cdots,$$
$$c_{\{(g_i^{(g,h)}, s_d^{(g,h)}),((g_i s_d)^{(g,h)}, (s_d^{-1})^{(g,h)})\}}) \in B$$

as required.

The transitivity of the code $Cay(G,S,B) \subseteq \mathbb{F}^{|G||S|/2}$ under the action of the semidirect product group $G \rtimes H$ is implied by Lemma 4. □

The following lemma bounds the number of generators of a Cayley code as well as their weight.

**Lemma 6** (The number/weight of generators of a Cayley code)**.** *Let $G$ be a group, $S \subseteq G$ a symmetric set. Let $B \subseteq \mathbb{F}^S$ be a linear code of rate $r_B > \frac{1}{2}$. Let $H$ be a group such that:*

- *$H$ acts on $G$.*
- *$H$ acts transitively on $S$.*
- *$B$ is $H$-invariant*

*Then the code $Cay(G, S, B) \subseteq \mathbb{F}^{|G||S|/2}$ has at most $(1 - r_B)|S|$ generators (under action of the semi-direct product group $G \rtimes H$) of weight at most $|S|$. In particular, if $H$ is the cyclic group the code $Cay(G, S, B)$ has one generator (more generally, if $B$ is generated by one generator with respect to $H$ then the code $Cay(G, S, B)$ has one generator). Recall, that the number of generators of the Cayley code is as the number of generators of the small code.*

*Proof.* By Lemma 4 the code $Cay(G, S, B)$ is invariant under the semi-direct product group $G \rtimes H$. Potential generators are contained in a basis to the code dual to $B$, hence there are at most $(1 - r_B)|S|$ such generators. If $H$ is the cyclic group then the code dual to $B$ is generated by a single vector and its orbit under $H$. As the generators of the Cayley code code $Cay(G, S, B)$ are contained in the generators of the code dual to $B$ (see the definition of a Cayley code), the code $Cay(G, S, B)$ has one generator. □

The following is the main theorem of the of this section. It defines conditions for the existence of codes with high rate, high distance and few generators (under a certain group) of low weight. The proof of the theorem follows immediately from Lemma 4, Lemma 5 and Lemma 6.

**Theorem 7** (Cayley Codes Theorem)**.** *Let $G$ be a group, $S \subseteq G$ a symmetric set. Let $H$ be a group. Let $B \subseteq \mathbb{F}^S$ be a linear code such that:*

- *$Cay(G, S)$ is an expander with second normalized eigenvalue $\lambda$.*
- *Distance of $B$ is $\delta > \lambda$.*
- *Rate of $B$ is greater than $\frac{1}{2}$ ($r_B > \frac{1}{2}$).*
- *$H$ acts on $G$.*
- *$H$ acts transitively on $S$.*
- *$B$ is $H$-invariant with 1-generator.*

*Then the code $Cay(G, S, B) \subseteq \mathbb{F}^{|G||S|/2}$ has constant rate $2r_B - 1$ and distance $[(\delta - \lambda)/(1 - \lambda)]^2$. It is transitive and invariant under the semi-direct product group $G \rtimes H$. The code is generated by one generator of weight bounded by $|S|$.*

# 4 Transitive high rate and high distance code with one low weight generator

In the following we aim at finding groups $G$, $H$, $S \subseteq G$ a symmetric set and a linear code $B$ of length $|S|$ that meets the conditions of the main theorem (Theorem 7). We start by showing that

if we take $G = F_2^t$ then there exists $S \subseteq G$ $\epsilon$-biased set $|S| = t^4 - 1$ and a group $H = F_{t^4}^*$, such that $H$ acts on $G$ and $H$ acts transitively on $S$. Recall that $S$ is $\epsilon$-biased in $G$ is equivalent to saying that the second normalized eigenvalue of $Cay(G, S)$ is at most $\epsilon$.

**Lemma 8** (Choosing $G, H, S$)**.** *Let $t = 2^a$ such that $s|t$ and $4a = s$. Let $G = F_2^t$. Let $q = t^4 = 2^s$. Let $H = F_q^*$ then the following holds.*

- *$H$ acts on $G$.*
- *There exists explicit $g \in F_2^t$ such that the set $S = \{g^H\} \subseteq G$ is an $\epsilon$-biased set with $\epsilon = 1/q^{1/4} \log q$.*

*Proof.* We will consider several related representations of elements in $F_{2^t}$ using its subfield $F_{2^s}$.

Let $d = \frac{t}{s} = \frac{2^{s/4}}{s}$ (recall that $s|t$). Let $e_1, \cdots, e_d$ be a basis for $F_{2^t}$ over $F_{2^s}$ such that for $v \in F_{2^t}$ is written uniquely as $v = v_1 e_1 + v_2 e_2 + \cdots + v_d e_d$ with $v_i \in F_{2^s}$. As elements of the underlying vector spaces, we'll use the vector representation $v = (v_1, v_2, \cdots, v_d)$ where $v_1, \cdots, v_d \in F_{2^s}$. Finally, we will also let $v$ correspond to the univariate degree $d$ polynomial $f_v(x) = v_1 x + v_2 x^2 + \cdots + v_d x^d \in F_{2^s}[x]$.

The action of $H$ on $G$ is simple: for any $h \in H$ and $v \in G$ we define $v^h$ to be

$$(v_1, v_2, \cdots, v_d)^h = (v_1 h, v_2 h^2, \cdots, v_d h^d)$$

with multiplication in $F_{2^s}$ componentwise. This is clearly an action.

To define the $\epsilon$-biased set $S$, let $g = (1, 1, \cdots, 1) = \sum_{i=1}^d e_i \in F_2^t$, and define $S = g^H = \{g^h : h \in H\}$. By definition $H$ acts transitively on $S$ and $|S| = |H| = q - 1$. We now turn to upper bound the bias of this set, namely upper bound $\sum_{h \in H} \quad (g^h)$ for all nontrivial additive characters $\quad$ of $F_{2^t}$.

For this estimate we factor each such character through $F_{2^s}$. Formally, let $\quad_0$ be some fixed nontrivial additive character of $F_{2^s}$. Let $Tr : F_{2^t} \to F_{2^s}$ denote the (linear) trace function. Then all additive characters of $F_{2^t}$ are then obtained as: $\quad(v) = \quad_0(Tr(zv))$ with some $z \in F_{2^t}^*$. Fix any such $z$, and lets call the associated character sum $\Delta_z$. By linearity of the trace, we have

$$\Delta_z = \sum_{h \in H} \quad (g^h) = \sum_{h \in H} \quad ((e_1 + e_2 + \cdots + e_d)^h) = \sum_{h \in H} \quad _0(Tr(z(\sum_{i=1}^d e_i)^h)) \tag{2}$$

$$= \sum_{h \in H} \sum_{i=1}^d \quad _0(Tr(z e_i \cdot h^i)) = \sum_{h \in H} \quad _0(\sum_{i=1}^d (Tr(z e_i)) \cdot h^i) \tag{3}$$

Let us denote by $a_i = Tr(z e_i) \in F_{2^s}$. Since the $e_i$ are a basis and $z \neq 0$, not all $a_i$ can be simultaneously zero. Using the polynomial representation above, we have $\Delta_z = \sum_{h \in H} \quad _0(f_a(h))$, namely a complete character sum of a nonzero degree $d$ polynomial over $F_{2^s}^*$. By Weil bound [39, 12], $|\Delta_z| < d \cdot \sqrt{2^s} = \frac{2^{s/4}}{s} \cdot 2^{s/2} = q/(q^{1/4} \log q)$.

□

**Lemma 9.** *(Bounding the second eigenvalue in $Cay(G,S)$) Let $G$, $S$ defined as in Lemma 8. The Cayley graph $Cay(G,S)$ is an expander with second normalized eigenvalue $\lambda < 2/(q^{1/4}\log q)$. Observe that this graph has $2^t$ vertices. The degree of each vertex is $t^4 - 1 = q - 1$.*

*Proof.* Since $G = F_2^t$ is Abelian the eigenvalues of $Cay(G,S)$ can be expressed in terms of the characters of $F_2^t$. Specifically, if $S$ is an $\epsilon$-biased set then the second normalized eigenvalue of $Cay(G,S)$ is at most $2\epsilon$. The value of $\lambda$ is obtained by the fact that $S$ is an $\epsilon$-biased set, for $\epsilon = 1/(q^{1/4}\log q)$ by Lemma 8. □

In the following, we use $t, q, s$ as they are defined in Lemma 8. To complete the construction of a Cayley code that meets the condition of Theorem 7, we need a linear cyclic code B transitive under the action of the cyclic group $H$. We next show that such a code of length $|S| = q-1$, rate greater than $\frac{1}{2}$ and large distance can be simply taken to be an appropriate BCH code.

**Lemma 10.** *(A linear code $B$ invariant under the cyclic group $F_q^*$ with good rate and distance) There exists a linear cyclic code $B$ of length $q-1$ (i.e., invariant under the cyclic group $H = F_q^*$) that has rate greater than $1/2$ and normalized distance greater than $1/c\log q$ for some constant $c > 1$.*

*Proof.* By Corollary 10 in [36] the BCH code is a linear cyclic code of length $N = q-1$ that can be chosen to obey the following relation $N - k = (d/2)\log N$, where $N$ is the length of the code, $k$ is the dimension of the code, $d$ is the distance (actual distance, not normalized). Thus if we pick $k$ to be greater than $\frac{N}{2}$ , we obtain $d > N/c\log N$ for some constant $c > 1$. I.e., a normalized distance greater than $1/c\log q$ for some constant $c > 1$. □

**Theorem 11** (Almost good code with one generator of polylogrithmic weight)**.** *Let $G$, $H$, $S$, $B$ be as above. Let $M = 2^t$. The Cayley code $Cay(G,S,B)$ has length $n = M \times (logM)^4/2$, constant rate, distance greater than $\frac{1}{c'(\log\log n)^2}$ for some constant $c' > 1$. The code is transitive and invariant under the semi direct product group $G \rtimes H$, and it is generated by one generator of weight at most $(\log n)^4$.*

*Proof.* Consider $G, H, S, B$ from above. By Lemma 9, the Cayley graph $Cay(G,S)$ is an expander with second largest eigenvalue $\lambda < 2/q^{1/4}\log q$. By Lemma 10 the linear code $B$ is a cyclic code that has distance $\delta = 1/c\log q > \lambda$, moreover $r_B > \frac{1}{2}$. By definition of $H, G, S, B$, $H$ acts on $G$, $H$ acts transitively on $S$ and $B$ is $H$-invariant with 1-generator. Hence all the conditions of Theorem 7 are met and the code has constant rate, distance

$$\left(\frac{\delta-\lambda}{1-\lambda}\right)^2 = \left(\frac{1/c\log q - 2/q^{1/4}\log q}{1 - 2/q^{1/4}\log q}\right)^2$$
$$\geq \frac{1}{c'\log^2 q} = \frac{1}{c'(\log\log M)^2}$$

Moreover, the code is transitive and invariant under the semi direct product group $G \rtimes H$, and it is generated by one generator of weight at most $(\log M)^4$. □

# 5 There are no good Abelian and solvable codes with few low weight generators

In the following we show that good codes with few low weight generators cannot be obtained from invariance under Abelian and solvable groups.

The main theorems of this section are the following.

**Theorem 12** (No good Abelian codes with generators of low support)**.** *An $n$-length linear code transitive and invariant under an Abelian group, defined by the orbits of constraints that are supported on $s \leq (\log n)^{1/4}$ coordinates, has either rate or relative distance bounded by $\epsilon$, where $\epsilon = n^{-1/6d^2}$ with $d = s^2$. Moreover, if the rate of the code exceeds $\epsilon$, the code is a concatenation of codes of length at most $\epsilon n$ each.*

Before stating the theorem for the solvable case we need the following definition.

**Definition 21.** *For a function $f(x)$ the $k$-iterate of $f$, denoted $f^{(k)}(x) = f(f(....f(x)))$ where the last has $k$-repetitions.*

The lower bound theorem for the solvable case is the following.

**Theorem 13** (No good solvable codes with generators of low support)**.** *An $n$-length linear code transitive and invariant under a solvable group of derived length $\ell$, defined by the orbits of constraints that are supported on $s \leq (\log n)^{(2l)}$ coordinates, has either rate or relative distance bounded by $\epsilon$, where $\epsilon = d/(\log n)^{(l)}$ with $d = s^2$. Moreover, if the rate of the code exceeds $\epsilon$, the code is a concatenation of codes of length at most $\epsilon n$ each.*

The theorems imply the following corollary.

**Corollary 1.** *For a linear code invariant and transitive under an Abelian/solvable group with $s$, $\epsilon$ as above. Let $\delta = \sqrt{\epsilon}$. If the rate of the code is at least $\delta$, then the code must have $1/\delta$ codewords with pairwise disjoint support.*

*Proof of Corollary 1.* If the rate of the code is at least $\delta = \sqrt{\epsilon}$ then the theorems imply that the code is a concatenation of codes of length at most $\epsilon n$ each. Every such concatenated code contributes $\epsilon$ to the total rate of the code. Hence the code is obtained by a concatenation of at least $1/\delta = 1/\sqrt{\epsilon}$ linear codes. Each such concatenated code contributes at least one codeword of disjoint support from the other codewords, so the proof of the corollary is established. □

The proofs of the theorems have two main parts. First we define a new notion that we call an $\epsilon$-partition of a graph (Definition 22), and show that a Schreier graph of an Abelian/solvable group with $d$ generators has an $\epsilon$-partition (where $d$ and $\epsilon$ are as defined as in the theorems). The proof of this part is by induction on the derived length of the group (Lemma 18) with the Abelian being the base case (Lemma 16).

In the second part (Lemma 21) we associate codes invariant under groups with Schreier graphs over these groups (see Definition 26), and show that if the associated Schreier graph has an $\epsilon$-partition then either the rate or the relative distance of the code is bounded by $\epsilon$.

Theorem 12 follows immediately from Lemma 21 and Lemma 16 below. Similarly, the proof of Theorem 13 follows immediately from Lemma 21 and Lemma 18.

We start by introducing the notion of an $\epsilon$-partition of a graph that plays a major role in the proofs.

**Definition 22** ($\epsilon$-Partition of a Graph)**.** *A graph $H$ on vertices $X$ has an* $\epsilon$-partition *if $X$ can be partition into $X_0, X_1, \cdots, X_t$ (i.e., $X = X_0 \cup X_1 \cup, \cdots, \cup X_t$, where $X_i$'s are disjoint), each of size at most $\epsilon|X|$, such that there are no edges between $X_i, X_j$ for all distinct $i, j > 0$ (i.e., the removal of $X_0$ partitions the graph to many small connected components).*

We stress that the notion of $\epsilon$-partition for sub-constant $\epsilon$ is stronger than non-expansion. E.g., take a disjoint union of two identical Cayley expanders, the obtained graph is transitive, non-expanding and has no $\epsilon$-partition.

In the following we study $\epsilon$-partitions of non-expanding Schreier graphs of Abelian and solvable groups. An interesting open question is which other groups obey the condition that a non-expanding Schreier graphs of these groups have $\epsilon$-partition for sub-constant $\epsilon$.

### 5.1 $\epsilon$-Partitions of non-expanding Abelian Schreier graphs

In the following we study non-expanding Schreier graphs of Abelian groups and show that they have $\epsilon$-partition for sub-constant $\epsilon$. The proofs of this subsection use the followings pair of tools. The first tool is a decomposition theorem for Abelian groups (Theorem 14). The second tool is a reduction from Schreier graphs to Cayley graphs (Claim 15).

**Theorem 14** (Decomposition Theorem for Abelian groups)**.** *A finite Abelian group $G$ can be obtained by a direct product of constant many cyclic groups. I.e., $G = C_1 \times C_2 \times, \cdots, \times C_\ell$ where $C_i$'s are cyclic groups.*

**Claim 15** (A reduction from Schreier to Cayley graphs )**.** *Let $G$ be an Abelian group and $S \subseteq G$ a (symmetric) set. Let $X$ be a set such that $G$ acts on $X$ transitively. Then the Schreier graph $Sch(G, X, S)$ is isomorphic to a Cayley graph $Cay(G', S')$ where $G'$ is Abelian, $|G'| = |X|$ and $|S'| = |S|$.*

*Proof.* By the Decomposition Theorem for Abelian groups $G = C_1 \times C_2 \times \cdots \times C_\ell$ where $C_i$ are cyclic groups. Since $G$ acts on $X$ transitively we have for every $x \in X$, $X = \{x^g | g \in G\}$. Fix the stabilizer $H$ of a point $x \in X$. I.e., $H = \{g \in G | x^g = x\}$. The stabilizer $H$ is a subgroup of $G$, hence $H$ is also Abelian and the group $G' = G/H$ is well defined. Also, $G'$ acts on $X$ and the natural homomorphism $G \to G'$ maps the generators $S$ to a set $S' \subseteq G'$. So we obtain $|G'| = |X|$ and by the transitivity of the action of $G$ on $X$ we can label $X$ with the elements of $G'$ so the Schreier graph $Sch(G, X, S)$ is isomorphic to a Cayley graph $Cay(G', S')$. □

We now turn to show that non-expanding Schreier graphs of Abelian groups have $\epsilon$-partition for sub-constant $\epsilon$.

**Lemma 16.** *Let $G$ be an Abelian group and $X$ be a set such that $G$ acts on $X$ and $|X| = n$. Let $S \subseteq G$ be a (symmetric) set of size $d \leq (\log n)^{1/2}$. Then, the Schreier graph $Sch(G, X, S)$ has an $\epsilon$-partition, with $\epsilon \leq n^{-\frac{1}{6d^2}}$.*

*Proof.* Assume first that the action of $G$ on $X$ is transitive. By Claim 15 above, the graph $Sch(G, X, S)$ is isomorphic to the Cayley graph $Cay(G', S')$ where $G'$ is Abelian and $|S'| = |S|$. Now by Claim 17 below, the Cayley graph $Cay(G', S')$ (and hence also the Schreier graph $Sch(G, X, S)$) has an $\epsilon$-partition with $\epsilon \leq n^{-\frac{1}{4d^2}}$.

If the action of $G$ on $X$ is not transitive then consider partition of $X$ induced by the action of $G$. I.e., $X = X_1 \cup X_2 \cup \cdots \cup X_k$ where the action of $G$ on each $X_i$ is transitive. In this case the graph $Sch(G, X, S) = Sch(G, X_1, S) \cup \cdots \cup Sch(G, X_k, S)$, where $G$ is transitive on each $X_i$. We can assume that each $X_i$ is such that $|X_i| \geq n^{1-\frac{1}{4d^2}} \geq n^{\frac{3}{4}}$ since otherwise the relevant connected component is already of the required size. By the previous discussion the graph $Sch(G, X_i, S)$ has an $\epsilon$-partition with $\epsilon \leq n^{-\frac{3}{4 \cdot 4d^2}} \leq n^{-\frac{1}{6d^2}}$. This implies that the Schreier graph $Sch(G, X, S)$ has an $\epsilon$-partition with $\epsilon \leq n^{-\frac{1}{6d^2}}$. □

**Claim 17.** *Let $G$ be an Abelian group of size $n$, and $S \subseteq G$ a (symmetric) set of size $d \leq (\log n)^{1/2}$. Then, the Cayley graph $Cay(G, S)$ has an $\epsilon$-partition, with $\epsilon \leq n^{-\frac{1}{4d^2}}$.*

*Proof.* By the Decomposition Theorem for Abelian groups every Abelian group is direct product of cyclic groups. Moreover, we can assume that there are at most $d$ of these, otherwise $S$ does not even generate $G$, and we can work on each connected component separately. Thus, we may assume that one of these cyclic groups is $H = Z_m$, with $m > n^{1/d}$, and $G = H \times K$. We will partition $H$ only by removing from it at most $\epsilon m$ vertices $D$, and leaving all components $C_i$ with size at most $\epsilon m$ each. This induces in $G$ a set of vertices $D \times K$ to remove of size of at most $\epsilon n$. After the removal the components of $G$ will be $C_i \times K$, each with at most $\epsilon n$ vertices.

Let $T$ be the projection of $S$ on $H$, namely the $H$ coordinates of each element of $S$ in its $H \times K$ representation. $T$ may be a multiset, but has size at most $d$, and consider the Cayley graph $Cay(H, T)$.

Let $T = \{h_1, h_2, ..., h_d\}$ where each $h_i$ is an integer in $[-m/2, m/2]$, which we use to represent $Z_m$.

Let $t$ be such that $t^{2d} = m/2$. Thus, $t = (m/2)^{1/2d}$.

$$1/t = 1/(m/2)^{1/2d} < 1/n^{1/4d^2} = 1/2^{(\log n)/4d^2}$$

**First case.** Assume all $|h_i| < m/t^2$. We break $Z_m$ to $t^2$ intervals of length $m/t^2$ each. Then the partition to disjoint pieces is clear: go around the circle $Z_m$, and remove every t'th interval. Thus, in total we remove $m/t$ vertices and get disconnected pieces of size at most $m/t$ each. Hence, we obtain an $\epsilon$-partition with $\epsilon \leq 1/2^{(\log n)/4d^2} = n^{-\frac{1}{4d^2}}$.

**Second case.** If the first case does not apply we show that the Cayley graph $Cay(H, T)$ is isomorphic to the first case, so we could be done by the proof of the first case once we order the vertices right.

Denote $I(x)$ the name (in $[t^2]$) of the interval that $x$ (an element of $Z_m$) belongs to, in the intervals of length $m/t^2$ defined above. Now consider the map $f : Z_m \to (t^2)^d$ defined by $f(a) = (I(a \cdot h_1), I(a \cdot h_2), ..., I(a \cdot h_d))$. Since $t^{2d} = m/2 < m$ we have a collision $f(a) = f(b)$, which means that

for $c = b - a$ we have $|c \cdot h_i| < m/t^2$ for all $i$. In other words, if we order the elements according to $c \cdot 1, c \cdot 2, ..., c \cdot m$, then we are back to the first case.

□

## 5.2 $\epsilon$-Partitions of non-expanding solvable Schreier graphs

In the following we study non-expanding Schreier graphs of solvable groups and show that they have $\epsilon$-partition for sub-constant $\epsilon$.

**Lemma 18.** *Let $G$ be a solvable group of derived length $\ell$, and $X$ be a set $|X| = n$, such that $G$ acts on $X$. Let $S \subseteq G$ be a (symmetric) set of size $d \leq [(\log n)^{1/4}]^{(\ell)}$. Then, the Schreier graph $Sch(G, X, S)$ has an $\epsilon$-partition, with $\epsilon \leq \frac{1}{2^{\frac{([(\log n)^{1/4}]^{(\ell)})^2}{6d^2}}}$.*

Before moving to the proof of the lemma we describe the proof strategy. Given a Schreier graph $Sch(G, X, S)$ where $G$ is solvable of derived length $\ell$, acts on $X$, with a set of generators $S \subseteq G$ of size $d$, we define a subgroup $H$ of G, and two simpler Schreier graphs $Sch(G/H, X/H, S')$ and $Sch(H, X_1, S_1)$. The first graph is Abelian and the second has a shorter derived series. We first show that either $Sch(G/H, X/H, S')$ or $Sch(H, X_1, S_1)$ has an $\epsilon$-partition. We then show (Claim 20) that if either of the graphs $Sch(G/H, X/H, S')$ or $Sch(H, X_1, S_1)$ has an $\epsilon$-partition then so does $Sch(G, X, S)$. This follows closely the structure of proof in [24], which we need to extend in two ways: from Cayley to Schreier graphs, and from non-expansion to $\epsilon$-partition. This reduction/induction increases the number of generators exponentially.

### 5.2.1 Definitions of simpler related Schreier graphs

**Definition 23** (The Schreier graph $Sch(G/H, X/H, S')$)**.** *Let $H = [G, G]$ and recall that $G/H$ is Abelian, and $H$ is normal.*

*For $x \in X$, let $x^H = \{x^{g_1} | g_1 \in H\}$, i.e., $x^H$ is the orbit of $x \in X$ under $H$. Let $X/H$ be the set of different orbits. Note that the orbits $X/H$ are a partition of $X$. The cosets of $X/H$ are simply $x^{Ha}$ with $a$'s being coset representatives of $H$ in $G$.*

*Thus, the action of $G/H$ on $X/H$ is induced by the action of $G$ on $X$, and it can be defined as $(Ha)(x^H) = (x^a)^H = x^{Ha}$ (by the normality of $H$).*

*Hence, we build a Schreier graph on the orbits using the induced action on them of $G/H$. Namely, $Sch(G/H, X/H, S')$ is the Schreier graph on the $X/H$ orbits where $S' = Im(S)$ under the natural map from $G \to G/H$.*

In order to define the Schreier graph $Sch(H, X_1, S_1)$ we need to define a map $f$ with its properties first.

**Definition 24** (The map $f$)**.** *Let $H = [G, G]$ and $n = [G : H]$. Let $T = \{t_1, \cdots t_n\}$ be the right coset representatives of $H$ ($T = T^{-1}$), and denote by $f : G \to T$ the map from $G$ to $T$, such that $f(g) = t_i$ iff $g$ is in the coset $Ht_i$.*

**Claim 19.** *The map $f : G \to T$ has the following properties*

1. $g[f(g)]^{-1} \in H$
2. *For every* $g, h \in G$, $f(f(g)f(h)) = f(gh)$.
3. *For every* $g \in G$, $[f(g)]^{-1} = f(g^{-1})$.
4. *For every* $z_j, z_{j'} \in H$, $t_i, t_{i'} \in T$ *and* $s \in G$, *if* $(z_j)t_i s = (z_{j'})t_{i'}$ *then* $z_j t_i s[f(t_i s)]^{-1} = z_{j'}$.

*Proof.* In the following we prove the four different properties of $f$.

**Proof of 1.** For $g \in G$, $g = g_1 t$ where $g_1 \in H, t \in T$, we have that $f(g) = t$. Hence $g[f(g)]^{-1} = gt^{-1} = g_1 t t^{-1} = g_1 \in H$.

**Proof of 2.** Let $g, h \in G$ be as follows. $g = g_1 t_i, h = h_1 t_j$ where $g_1, h_1 \in H$, $t_i, t_j \in T$.

$$f(gh) = f(g_1 t_i h_1 t_j) = f(g_1(t_i h_1)t_j) = f(g_1(h_1' t_i)t_j)$$
$$\text{for some } h_1' \in H \text{ since } H \text{ is normal.}$$

Thus,

$$\begin{aligned} f(gh) =& f((g_1 h_1')t_i t_j) = f(t_i t_j) = f(f(g_1 t_i)f(h_1 t_j)) \\ =& f(f(g)f(h)) \end{aligned}$$

**Proof of 3.** By the normality of $H$, there exists $g' \in H$ such that $g'^{-1} t_i^{-1} = t_i^{-1} g_1^{-1}$.

$$\begin{aligned} [f(g)]^{-1} =& [f(g_1 t_i)]^{-1} = t_i^{-1} = f(g'^{-1} t_i^{-1}) \\ =& f(t_i^{-1} g_1^{-1}) = f(g^{-1}) \end{aligned}$$

**Proof of 4.** Note first the following facts that follow easily from the previous properties of $f$.

- If $s \in G$ is such that $(z_j)t_i s = (z_{j'})t_{i'}$ then $f(s) = f(t_i^{-1} t_{i'})$. This is since $f(t_i s) = f(t_{i'})$, so $f(f([t_i]^{-1})f(t_i s)) = f(f([t_i]^{-1})f(t_{i'}))$. Hence $f([t_i]^{-1} t_i s) = f(s) = f([t_i]^{-1} t_{i'})$.
- For every $t_i, t_{i'}$, $f(t_i) = t_i$, $f(t_{i'}) = f(t_i f({t_i}^{-1} t_{i'}))$.

Assume that for $s \in G$, $(z_j)t_i s = (z_{j'})t_{i'}$.

$$\begin{aligned} (z_j)t_i s =& (z_{j'})t_{i'} = (z_{j'})f(t_{i'}) = (z_{j'})f(t_i f(t_i^{-1} t_{i'})) \\ =& (z_{j'})f(t_i f(s)) = (z_{j'})f(t_i s) \end{aligned}$$

Hence, $(z_j)t_i s[f(t_i s)]^{-1} = (z_{j'})$. □

We now move to the definition of the Schreier graph $Sch(H, X_1, S_1)$.

**Definition 25** (The Schreier graph $Sch(H, X_1, S_1)$)**.** *Let $H = [G, G]$. For $x \in X$, let $X_1 = x^H$. Given $G$ and $X$, let $G'$ to be the smallest subgroup of $G$ containing $H$ such that $G'/H$ acts transitively on $X/H$. The coset representatives of $H$ in this $G'$ are denoted $T_X = \{t_i \in T | x^{Ht_i} \in X/H\}$ where $T_X \subseteq T$ ($T$ is as defined in the definition of the map $f$). W.l.o.g we can assume $G' = G$. Hence, $T_X = T$ and $|T_X| = |X/H| = |G/H| = |T|$. The set of generators $S_1 \subseteq H$ are of the following form. $t_i s[f(t_i s)]^{-1}$, $t_i \in T$, $s \in S$. Hence, $|S_1| \leq |S||X/H|$.*

### 5.2.2 Proof of Lemma 18

The proof is by induction on $\ell$. For $\ell = 1$, $G$ is Abelian so the lemma follows immediately from Lemma 16. Assume the lemma holds for $\ell - 1$. We next prove that it holds for $\ell$. Let $H = [G, G]$. Recall that $G/H$ is Abelian. Let $g(n) = \frac{[(\log n)^{1/4}]^{(\ell)}}{d}$.

Let $|X/H| = k$.

**Case one:** $2^{(d^2 g^2(n))} < k$**.** I.e., $d < \frac{(\log k)^{1/2}}{g(n)}$, then again by Lemma 16, the Schreier graph $Sch(G/H, X/H, S')$ has an $\epsilon$-partition, with

$$\epsilon \leq \frac{1}{2^{\frac{\log k}{6d^2}}} \leq \frac{1}{2^{\frac{g^2(n)}{6}}} = \frac{1}{2^{\frac{([(\log n)^{1/4}]^{(\ell)})^2}{6d^2}}}.$$

**Case two:** $2^{(d^2 g^2(n))} \geq k$**.** Here the Schreier graph $Sch(H, X_1, S_1)$ is such that $n' = |X_1| \geq \frac{n}{2^{(d^2 g^2(n))}} \geq \frac{n}{[(\log n)^{1/4}]^{(\ell-1)}}$ and it has a set of generators $S_1$, where

$$\begin{aligned} |S_1| &\leq kd \leq d2^{(d^2 g^2(n))} \leq 2^{(d^4 g^2(n))} \\ &\leq ([(\log n)^{1/4}]^{(\ell-1)})^{1/g^2(n)} \leq \frac{[(\log n')^{1/4}]^{(\ell-1)}}{g(n)}. \end{aligned}$$

Thus, according to the induction hypothesis applied to $Sch(H, X_1, S_1)$, the graph $Sch(H, X_1, S_1)$ has an $\epsilon$-partition, with

$$\epsilon \leq \frac{1}{2^{\frac{([(\log n')^{1/4}]^{(\ell-1)})^2}{6|S_1|^2}}} \leq \frac{1}{2^{\frac{g^2(n)}{6}}} = \frac{1}{2^{\frac{([(\log n)^{1/4}]^{(\ell)})^2}{6d^2}}}.$$

We complete the proof of the lemma by showing in the next claim (Claim 20) that if either of the graphs $Sch(G/H, X/H, S')$ or $Sch(H, X_1, S_1)$ has an $\epsilon$-partition then so does $Sch(G, X, S)$.

**Claim 20.** *If either of the graphs $Sch(G/H, X/H, S')$ or $Sch(H, X_1, S_1)$ has an $\epsilon$-partition then so does $Sch(G, X, S)$.*

*Proof.* If the graph $Sch(G/H, X/H, S')$ has an $\epsilon$-partition, then by removing at most $\epsilon|X/H|$ cosets a partition of $|X/H|$ into sets $Z_1, Z_2, \cdots, Z_m \subseteq X/H$ is obtained, where each set is of size at most $\epsilon|X/H|$ with no edges from $S'$ between pieces. Let $T_{Z_i}$ be the coset representatives of the cosets in

$Z_i$. An $\epsilon$-partition of $Sch(G, X, S)$ is immediately implied by the partition of $Sch(G/H, X/H, S')$ by removing at most $\epsilon|X/H|$ whole cosets of size $|x^H|$ each from $Sch(G, X, S)$. Namely, such a partition is $HT_{Z_1}, \cdots HT_{Z_m}$ where each part is of size at most $\epsilon|X/H| \cdot |H| = \epsilon|X|$ and there are no edges from $S$ between the parts.

If the graph $Sch(H, X_1, S_1)$ has an $\epsilon$-partition, recall the definitions of the map $f : G \to T$ and the coset representatives $T$.

Assume that the $\epsilon$-partition of the graph $Sch(H, X_1, S_1)$ partition the vertex set $X_1$ into sets $Z_1, Z_2, \cdots, Z_m$ each of size at most $\epsilon|X_1|$ with no edges from $S_1$ between pieces. The $\epsilon$-partition of $Sch(H, X_1, S_1)$ implies an $\epsilon$-partition of $Sch(Ht_i, X_1t_i, S_1)$ , $t_i \in T$ into sets $Z_1t_i, Z_2t_i, \cdots, Z_mt_i$.

Now consider the partition of $Sch(G, X, S)$ into pieces of the form $\{Z_jt_1, \cdots, Z_jt_{|T|}\}$, $(t_i \in T)$. We next show that this is indeed a partition. I.e., there is no edge $s \in S$ that connects *any* $(z_j)t_i$ and $(z_{j'})t_{i'}$. This follows from the forth property proved in Claim 19 showing that for every $z_j, z_{j'} \in H$, $t_i, t_{i'} \in T$ and $s \in S$, if $(z_j)t_is = (z_{j'})t_{i'}$ then $z_jt_is[f(t_is)]^{-1} = z_{j'}$, i.e., there exists $s_1 = t_is[f(t_is)]^{-1} \in S_1$, such that $z_js_1 = z_{j'}$. The forth property, proved in Claim 19, implies that if there is an edge between parts defined by the above partition of $Sch(G, X, S)$, then there is an edge between parts in the above partition of $Sch(H, X_1, S_1)$. However, we assumed that the partition of $Sch(H, X_1, S_1)$ is valid and hence so is the partition of $Sch(G, X, S)$.

□

### 5.3 A Coding Result and the proofs of Theorems 12 and 13

In the following we consider codes invariant under groups and associate with them Schreier graphs which we call the Schreier-graph of the code (see following Definition 26). We show that if the Schreier-graph of the code has an $\epsilon$-partition then either the rate or the relative distance of the code are bounded by $\epsilon$ (Lemma 21). The $\epsilon$-partitions of Abelian and solvable Schreier graphs (Lemma 16 and Lemma 18) combined with this coding theory result (Lemma 21) immediately imply the proofs of Theorem 12 and Theorem 13.

**Definition 26** (Schreier graph of a code)**.** *Let $C \subseteq \mathbb{F}^X$ be a linear code invariant and transitive under a group $G$, where $C$ is defined by the orbits of constraints that are supported on a set $W \subseteq X$ of size $d$ (i.e., $|W| = d$). The* Schreier graph of the code *$C$ is the graph $Sch(G, X, S)$ where $S$ is defined as follows. Pick $x \in X$, Let $S' \subseteq G$ be such that $W = \{x^g | g \in S'\}$. $S = \{g_i(g_j)^{-1} | g_i, g_j \in S'\}$, i.e., $|S| = d^2$ is such that for each generating constraint of $C$, there is a clique in the graph between its coordinates.*

**Lemma 21.** *Let $C \subseteq \mathbb{F}^X$ be a linear code invariant and transitive under a group $G$. Assume $C$ is defined by the orbits of constraints that are supported on $d$ coordinates of $C$. If the Schreier graph of the code $Sch(G, X, S)$ (recall $|S| = d^2$) has an $\epsilon$-partition then either the rate or the relative distance of $C$ are bounded by $\epsilon$. Moreover, if the rate of the code exceeds $\epsilon$, the code is a concatenation of codes of length at most $\epsilon|X|$ each.*

*Proof.* If the rate of $C$ is bounded by $\epsilon$ then we are done. It remains to show that if the rate of $C$ is larger than $\epsilon$ then the distance of $C$ is at most $\epsilon|X|$.

By the definition of a Schreier graph of a code, The Schreier graph of the code $C$, $Sch(G, X, S)$, contains a clique between the coordinates of each generating constraint of the code $C$. Since the graph $Sch(G, X, S)$ has an $\epsilon$-partition, there exists a partition of $X = X_0 \cup X_1 \cup \cdots \cup X_q$ into disjoint sets $X_i$ each of size at most $\epsilon|X|$, such that there are no edges between $X_i, X_j$ for $i \neq j > 0$. The fact that the graph $Sch(G, X, S)$ contains a clique between the coordinates of each generating constraint of the code $C$, and the $\epsilon$-partition of $Sch(G, X, S)$ imply that each generating constraint of the code $C$ has support completely contained in $X_i \cup X_0$ for some $i$.

Consider constraints of the form $\{x = 0 | x \in X_0\}$. There are at most $\epsilon|X|$ such constraints. Since we assume that the rate of $C$ is larger than $\epsilon$, there exists a sub-code $B \subseteq C$, $B \neq 0$, that satisfies all of the generating constraints of $C$ as well as the constraints $\{x = 0 | x \in X_0\}$. Using the fact that each generating constraint of $C$ has support completely contained in $X_i \cup X_0$ for some $i$, we obtain that each generating constraint of $B$ has support completely contained in one $X_i$, $i \geq 0$.

So $B \subseteq C \subseteq \mathbb{F}^X$ is a non-zero linear code, for which there exists a partition of $X = X_0 \cup X_1 \cup \cdots \cup X_q$ into subsets $X_i$ each of size at most $\epsilon|X|$, and a set of constraints that generate $B$, each of which has support completely contained in one $X_i$, hence by Claim 22 (below), $B$ (and hence $C$) has distance bounded by $\epsilon|X|$. □

**Claim 22.** *Let $C \subseteq \mathbb{F}^X$ be a linear code. If there exists a partition of $X = X_0 \cup X_1 \cup \cdots \cup X_q$ into subsets $X_i$ each of size at most $\epsilon|X|$, and a set of constraints that span $C^\perp$ each of which has support completely contained in one $X_i$, then $C$ is a concatenation of disjoint codes $C_i \subseteq \mathbb{F}^{X_i}$ of length at most $\epsilon|X|$ each, and in particular, $C$ has distance bounded by $\epsilon|X|$.*

*Proof.* To see that $C$ is a concatenation of codes we need to show $c \in C$ iff $c_i \in C_i$ for every $i$ (where $c_i$ is the restriction of $c$ to the coordinates of $X_i$). If $c \in C$ then it obeys the constraints of the code $C$. Since the constraints that define the code $C$ can be partitioned into disjoint sets, where each set defines the code $C_i$, we have that if $c \in C$ then $c_i \in C_i$ for every $i$. Next we show that if $c \in \mathbb{F}^X$ is such that $c_i \in C_i$ for every $i$, then $c \in C$. Since the constraints that define $C$ can be partitioned into constraints that define $C_i$ for every $i$, and since we know that $c_i$ obeys all the constraints of $C_i$ we have that $c$ obeys all the defining constraints of $C$ and hence $c \in C$. To see that the distance of $C$ is bounded by $\epsilon|X|$ note that if $C$ is not empty then it contains a non zero codeword $c \in C$, where $c = c_0||c_1|| \cdots ||c_q$; $c_i \in C_i$. Since $c$ is non zero there exists some non-zero $c_i$ (assume w.l.o.g that $i = 0$). As $C$ is a concatenation of codes $C_i$'s, $c \in C$ implies that $c' = c_0||0||0 \cdots 0||0$ is also in $C$ and it has weight bounded by $|X_1| \leq \epsilon|X|$, hence $C$ has distance bounded by $\epsilon|X|$. □

# 6 Conclusions and Open Questions

This paper was motivated by the construction of locally-testable codes of good coding-theoretic parameters. As is well known, Goldreich and Sudan [16] showed how to obtain such codes from PCPs with related parameters, and good parameters are achieved by combining the PCPs of Dinur [13] with Ben-Sasson and Sudan [7]. Specifically, they achieve linear binary codes of length $n$ with linear distance, rate $1/(\log n)^c$ and constant-size queries. These codes are completely explicit.

Removing the PCP machinery and obtaining such codes (and even better ones) directly is a basic question, motivated at length in the paper of Meir [28]. He succeeds only partially, in that his

construction that is partly probabilistic. Moreover, the construction cleverly retains "proofs of membership" in the code, as part of the code, which make it resemble Dinur's PCP construction.

We take a completely different approach. As all locally-testable codes must be LDPC codes (since low query complexity means low density in the parity check matrix), and moreover many locally-testable codes are symmetric (have a transitive group acting on them), we ask first if the above coding theoretic parameters can be attained by codes that are simultaneously symmetric and low-density. We give the first such construction. Our codes are linear binary codes of length $n$ with near linear distance $n/(\log\log n)^2$, constant rate and both density bounded by $1/(\log n)^4$. The group acting transitively is non-Abelian. All previously known symmetric codes with such (or even weaker) distance had either density or (1/rate) close to $n$, and groups in all cases are Abelian.

There are several open questions that arise from this work.

- Cayley codes and local testing. Are the Cayley codes we construct actually locally testable? We tend to think that they are not, in which case would be the *first* example of a symmetric LDPC code which is not locally testable. As we offer a general framework of Cayley codes, possibly other choices of components in this framework can lead to to locally-testable codes.
- Improving the parameters. Can one get the ultimate – symmetric, constant density *good* codes (namely with linear distance and constant rate)? Our lower bounds imply that for such a result the acting group must be "more noncommutative" than the one we use, namely it cannot be solvable with a constant-length derived series. As we have discussed in the introduction, this question was recently solved affirmatively by a work of Kaufman and Lubotzky [19].
- Key to our lower bound is our that Cayley codes of such groups have $\epsilon$-partition, a property which implies in particular that such codes must have two *disjoint* codewords. Interestingly, the question of proving the latter property for similar codes comes up naturally in the work of Lackenby [20, 21] on 3-dimensional manifolds. Specifically, he asks if linear codes symmetric under the action of $p$-groups (which are solvable, but can have constant degree Cayley graphs), which have constant rate, density and normalized distance, must have two codewords with disjoint support. Our lower-bound techniques fails for such groups.

## Acknowledgements

We wish to thank Madhu Sudan, Peter Sarnak, and Jean Bourgain for very helpful conversions, and to Or Meir for helpful comments on an earlier version of this manuscript. The first author was supported in part by NSF Awards CCF-0514167 and NSF-0729011.

## References

[1] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn and Dana Ron, *Testing Low Degree Polynomials Over GF(2)*, Proceedings of 7th International Workshop on Randomization and Computation,(RANDOM), Lecture Notes in Computer Science 2764, 188-199, 2003. Also, IEEE Transactions on Information Theory, Vol. 51(11), 4032-4039, 2005.

[2] Noga Alon, Alex Lubotzky and Avi Wigderson *Semi Direct product in groups and zig-zag product in graphs: connections and applications*, Proceedings of the 42nd Annual Symposium on the Foundations of Computer Science (FOCS), 630-637, 2001.

[3] Sanjeev Arora and Madhu Sudan. *Improved low degree testing and its applications.* Combinatorica, 23(3): 365-426, 2003.

[4] L. Babai and L. Fortnow and C. Lund, *Non-Deterministic Exponential Time has Two-Prover Interactive Protocols*, Computational Complexity, volume 1, number 1, 3–40, 1991.

[5] László Babai and Amir Shpilka and Daniel Stefankovic, *Locally testable cyclic codes*,IEEE Transactions on Information Theory, Vol 51, No 8, pp. 2849–2858. 2005.

[6] Eli Ben-Sasson, Prahladh Harsha and Sofya Raskhodnikova, *Some 3CNF Properties are Hard to Test*, SIAM Journal on Computing, volume 35, issue 1, pages 1-21, 2005.

[7] Eli Ben-Sasson, Madhu Sudan, *Simple PCPs with poly-log rate and query complexity*, STOC 2005: 266-275.

[8] E. Ben-Sasson, M. Sudan, S. Vadhan, A. Wigderson. *Randomness-efficient Low Degree Tests and Short PCPs via Epsilon-Biased Sets* 35th Annual ACM Symposium, STOC 2003, pp. 612-621, 2003.

[9] S. D. Berman. *Semisimple Cyclic and Abelian Codes.* Cybernetics 3 (1967), 2130.

[10] Blum, M., Luby, M., Rubinfeld, R., *Self-Testing/Correcting with Applications to Numerical Problems*, In J. Comp. Sys. Sci. Vol. 47, No. 3, December 1993.

[11] M. Capalbo, O. Reingold, S. Vadhan, A. Wigderson, *Randomness Conductors and Constant-Degree Expansion Beyond the Degree /2 Barrier*, Proceedings of the 34th STOC, pp. 659-668, 2002.

[12] L. Carlitz and S. Uchiyama, *Bounds for exponential sums*, Duke Math. J., 24:37-41, 1957.

[13] Irit Dinur,*The PCP theorem by gap amplification*, J. ACM 54(3): 12 (2007).

[14] R. G. Gallager, *Low density parity check codes*, MIT Press, Cambridge, MA, 1963.

[15] Elena Grigorescu, Tali Kaufman and Madhu Sudan, *Succinct Representation of Codes with Applications to Testing*, manuscript.

[16] Oded Goldreich, Madhu Sudan, *Locally testable codes and PCPs of almost-linear length*, J. ACM 53(4): 558-655 (2006).

[17] Tali Kaufman and Madhu Sudan, *Algebraic Property Testing: The Role of Invariance*, Proceedings of the 40th ACM Symposium on Theory of Computing (STOC), 2008.

[18] Tali Kaufman, Simon Litsyn, *Almost Orthogonal Linear Codes are Locally Testable*, FOCS 2005: 317-326.

[19] Tali Kaufman, Alexander Lubotzky, *Edge transitive ramanujan graphs and symmetric LDPC good codes*, STOC 2012: 359-366.

[20] M. Lackenby, *Large groups, property* $(\tau)$ *and the homology growth of subgroups.* Math. Proc. Cambridge Philos. Soc. 146 (2009), no. 3, 625–648.

[21] M. Lackenby, *Covering spaces of 3-orbifolds.* Duke Math. J. 136 (2007), no. 1, 181–203.

[22] A. Lubotzky, B. Samuels, and U. Vishne. *Ramanujan complexes of type* $\tilde{A}_d$*. Israel Journal of Mathematics*, 149(1):267-299, 2005.

[23] A. Lubotzky, B. Samuels, and U. Vishne. *Explicit constructions of ramanujan complexes of type* $\tilde{A}_d$*.* Eur. J. Comb., 26(6):965-993, 2005.

[24] A. Lubotzky, B. Weiss, *Groups and expanders*, In *Expanding Graphs* (e. J. Friedman), DIMACS Ser. Discrete Math. Theoret. Compt. Sci. 10pp. 95-109, Amer. Math. Soc., Prividence, RI 1993.

[25] Michael G. Luby, Michael Mitzenmacher, M. Amin Shokrollahi and Daniel A. Spielman, *Improved Low-Density Parity-Check Codes Using Irregular Graphs* IEEE Transactions on Information Theory, 47(2), pp. 585-598. 2001.

[26] F. J. MacWilliams and N. J. A. Sloan, *The Theory of Error Correcting Codes*, North Holland, Amsterdam, 1977.

[27] Robert J. McElice. *On the Symmetry of Good Nonlinear Codes.* IEEE Trans. Inform. Theory IT-16 (1970), pp. 609611.

[28] Or Meir, *Combinatorial Construction of Locally Testable Codes*, proceedings of STOC 2008, pages 285-294.

[29] R. Meshulam, A. Wigderson, *Expanders in Group Algebras*, Combinatorica, vol. 24, no. 4, 659–680, 2004.

[30] J. Naor and M. Naor, *Small-Bias Probability Spaces: Efficient Constructions and Applications*, SIAM J. Comput., 22(4), 838-856, 1993.

[31] Ronitt Rubinfeld and Madhu Sudan, *Robust characterizations of polynomials with applications to program testing*, SIAM Journal on Computing, 25(2):252-271, April 1996.

[32] E. Rozenman, A. Shalev, A. Wigderson, *A new family of Cayley expanders (?)*, 36th Annual ACM Symposium, STOC 2004, pp. 445-454, 2004.

[33] T. Richardson and R. Urbanke, *The Capacity of Low-Density Parity Check Codes under Message-Passing Decoding*, IEEE Transactions on Information Theory, 47(2):599-618, 2001.

[34] O. Reingold, S. Vadhan, A. Wigderson, *Entropy Waves, the Zig-Zag Graph Product, and New Constant-Degree Expanders*, Annals of Mathematics, vol. 155, no.1, pp. 157-187, 2002.

[35] Michael Sipser and Daniel A. Spielman, *Expander codes*, IEEE Transactions on Information Theory, Vol 42, No 6, pp. 1710-1722. 1996.

[36] Madhu Sudan Lecture notes http://people.csail.mit.edu/madhu/FT01/scribe/bch.ps.

[37] Madhu Sudan, Luca Trevisan, and Salil Vadhan, *Pseudorandom generators without the XOR Lemma*, Journal of Computer and System Sciences, 62(2): 236–266, March 2001.

[38] Robert M. Tanner, *A recursive approach to low complexity codes*, IEEE Transactions on Information Theory, 27(5):533-547, 1981.

[39] A. Weil, *Sur les courbes algebriques et les varietes qui s'en deduisent*, Actualities Sci. et Ind. no. 1041. Hermann, Paris, 1948.