Symmetric LDPC codes and local testing

Tali Kaufman
MIT
kaufmant@mit.edu

Avi Wigderson
Institute for Advanced Study
avi@ias.edu

March 1, 2010

Abstract

Coding theoretic and complexity theoretic considerations naturally lead to the question of generating symmetric, sparse, redundant linear systems. This paper provides new way of constructions 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 $\log^{(\Omega(\ell))} n$ if the group has a derived series of length ℓ. 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. Following is a very high level description of these.

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 [26]) 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 [22, 30, 32] 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 [19] that the sparse dual-BCH codes have such constraints (but unfortunately this code has 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 [25] 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, 28], 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 [18]. 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 [18] 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 [35] was based on high girth graphs. Sipser and Spielman [32] 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 [32] "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 [31], which started as a combinatorial alternative to algebraic constructions, was extended to allow iterative probabilistic constructions of Cayley graphs [2, 27] via the semi-direct product of groups. Our codes are partially motivated by making explicit the probabilistic construction of [2, 27] Attempts to construct codes iteratively exist, with the best example being Meir's, partially explicit construction [26]. 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, 34].[2]

Indeed, some believed that the conflict between density and rate in symmetric codes cannot be

[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 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.

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.

2 Our Results

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, we provide 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.

Our main negative result 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 $o(\log^{\Omega(\ell)} n)$ if the group has a derived series of length ℓ and n is the code length. This result exclude the possibility of good solvable locally testable codes with few low weight generators.

3 Our Techniques

In order to prove our upper bound, 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 ϵ-biased set 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 $G; S$.

Our lower bound methods extend work of Lubotzky and Weiss [23], 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 ϵ-partitions of the graph to many parts - from which we can deduce information on the distance and rate of the associated Tanner codes.

The proof showing that there are no good solvable codes with few low weight generators has two main parts. First, for a parameter ϵ (later taken to be $o(1)$) we define a new notion that we call an ϵ-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 ϵ-partition if one can remove ϵ fraction of its vertices to make all connected components of relative size at most ϵ. We show that a Schreier graph of a solvable group with $d = o(\log^{\Omega(\ell)} n)$ generators has an ϵ-partition where ϵ is sub-constant. In the second part of the proof we associate codes invariant under groups with Schreier graphs over these groups, and show that if the associated Schreier graph has an ϵ-partition then either the rate or the relative distance of the code is bounded by ϵ.

4 Related work

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 code construction provides such explicit codes generated by *one* orbit!.

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 constant many low-weight constraints.

5 Conclusions and Open Questions

This paper was motivated from 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 can be constructed 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 [26]. 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 $1/(logn)^4$. 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.
- Key to our lower bound is our that Cayley codes of such groups have ϵ-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.

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 applictions*, 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] Holton, D. A. and Sheehan, J. *The Petersen Graph.* Cambridge, England, Cambridge University Press, 1993.

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

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

[20] M. Lackenby, *Large groups, property* (τ) *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] 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.

[23] 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.

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

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

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

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

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

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

[30] 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.

[31] 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.

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

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

[34] 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.

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

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