



NLTS Hamiltonians and Strongly-Explicit SoS Lower Bounds from Low-Rate Quantum LDPC Codes

Louis Golowich  

Department of Computer Science, University of California at Berkeley, CA, USA

Tali Kaufman 

Department of Computer Science, Bar-Ilan University, Ramat-Gan, Israel

Abstract

Recent constructions of the first asymptotically good quantum LDPC (qLDPC) codes led to two breakthroughs in complexity theory: the NLTS (No Low-Energy Trivial States) theorem (Anshu, Breuckmann, and Nirkhe, STOC'23), and explicit lower bounds against a linear number of levels of the Sum-of-Squares (SoS) hierarchy (Hopkins and Lin, FOCS'22).

In this work, we obtain improvements to both of these results using qLDPC codes of *low rate*:

- Whereas Anshu et al. only obtained NLTS Hamiltonians from qLDPC codes of linear dimension, we show the stronger result that qLDPC codes of arbitrarily small positive dimension yield NLTS Hamiltonians.
- The SoS lower bounds of Hopkins and Lin are only weakly explicit because they require running Gaussian elimination to find a nontrivial codeword, which takes polynomial time. We resolve this shortcoming by introducing a new method of planting a strongly explicit nontrivial codeword in linear-distance qLDPC codes, which in turn yields strongly explicit SoS lower bounds.

Our “planted” qLDPC codes may be of independent interest, as they provide a new way of ensuring a qLDPC code has positive dimension without resorting to parity check counting, and therefore provide more flexibility in the code construction.

2012 ACM Subject Classification Theory of computation → Quantum complexity theory

Keywords and phrases NLTS Hamiltonian, Quantum PCP, Sum-of-squares lower bound, Quantum LDPC code

Digital Object Identifier 10.4230/LIPIcs.ITCS.2024.54

Related Version *Full Version*: <https://arxiv.org/abs/2311.09503>

Funding This work was done in part while the authors were visiting the Simons Institute for the Theory of Computing.

Louis Golowich: Supported by a National Science Foundation Graduate Research Fellowship under Grant No. DGE 2146752, and in part by V. Guruswami’s Simons Investigator award and UC Noyce Initiative Award award.

Acknowledgements We thank Max Hopkins for numerous helpful discussions, and for bringing the problem of strongly explicit SoS lower bounds to our attention. We also thank Venkat Guruswami for helping to improve the exposition.

1 Introduction

Recent breakthrough constructions of asymptotically good quantum LDPC (qLDPC) codes [20, 13, 4] have led to major advances in complexity theory. Specifically, Anshu et al. [1] applied these codes to prove the NLTS theorem, which provides perhaps the most significant progress to date towards the quantum PCP conjecture. Meanwhile, Hopkins and Lin [11] applied the same codes to obtain the first explicit lower bounds against a linear number of



© Louis Golowich and Tali Kaufman;

licensed under Creative Commons License CC-BY 4.0

15th Innovations in Theoretical Computer Science Conference (ITCS 2024).

Editor: Venkatesan Guruswami; Article No. 54; pp. 54:1–54:23

Leibniz International Proceedings in Informatics



LIPIC Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

levels of the Sum-of-Squares semidefinite programming (SoS SDP) hierarchy, which is one of the most powerful algorithmic frameworks for approximating the satisfiability of constraint satisfaction problems (CSPs).

In this paper, we improve upon both of these complexity theoretic results. Along the way, we introduce a new method for ensuring a qLDPC code has positive dimension, which may be of independent interest. Our contributions are therefore threefold:

1. **NLTS Hamiltonians from low-rate codes:** The breakthrough construction of NLTS Hamiltonians of [1] from asymptotically good qLDPC codes relied on both the linear dimension and distance of the codes. A promising approach [19] for further progress towards qPCP is to construct more general NLTS Hamiltonians with additional properties. We make progress in this direction by constructing NLTS Hamiltonians from qLDPC codes of arbitrary positive dimension, thereby removing the linear-dimension requirement in [1]. Our result highlights the usefulness of local Hamiltonians with low-dimensional ground spaces for studying qPCP. Our proof leverages techniques of [6], which conjecturally constructed NLTS Hamiltonians from linear-distance quantum locally testable codes of arbitrary positive dimension (which are not known to exist). However, we obtain the NLTS property without assuming local testability nor linear dimension. Instead, the key ingredient ensuring NLTS Hamiltonians is a small-set expansion property of the qLDPC codes.
2. **Planted quantum LDPC codes:** We show how to plant an explicit nontrivial codeword in a linear-distance qLDPC code, which may have otherwise had rate 0. To the best of our knowledge, this construction yields the first linear-distance qLDPC codes for which nontrivial dimension is established without resorting to parity-check counting. It has been an open question in the literature to develop new such techniques for bounding dimension (see for instance Section 1.1 of [5], and also [2]).
3. **Strongly explicit SoS lower bounds:** We apply our planted qLDPC codes to obtain the first *strongly* explicit family of CSPs that cannot be refuted by a linear number of levels of the SoS hierarchy. This result strengthens the work of [11], which provided the first *weakly* explicit construction of such an SoS lower bound using qLDPC codes. Our improvement stems from the fact that our planted codes have planted codeword given by the all-1s vector, which is strongly explicit.

These results together show new ways to both construct and apply qLDPC codes of low rate. In the remainder of this section, after providing some background on qLDPC codes, we describe each of these results in more depth. We then discuss open questions that arise from our results.

1.1 Background on qLDPC Codes

This section provides some definitions we will need to state our results. The quantum codes we consider in this paper are quantum CSS codes. An n -qudit CSS code $\mathcal{C} = \text{CSS}(C_X, C_Z)$ of alphabet size (i.e. local dimension) q is defined by a pair of classical codes $C_X, C_Z \subseteq \mathbb{F}_q^n$ such that $C_X^\perp \subseteq C_Z$. The associated quantum code is then given by $\mathcal{C} = \text{span}\{\sum_{y' \in C_X^\perp} |y + y'\rangle : y \in C_Z\}$. This code has dimension $k = \dim(C_Z) - \dim(C_X^\perp)$ and distance $d = \min_{y \in (C_Z \setminus C_X^\perp) \cup (C_X \setminus C_Z^\perp)} |y|$, meaning it encodes a k -qudit message into an n -qudit code state, and the message can be recovered from any $n - (d - 1)$ code qudits. We assume $C_X = \ker H_X, C_Z = \ker H_Z$ for associated parity check matrices $H_X \in \mathbb{F}_q^{m_X \times n}, H_Z \in \mathbb{F}_q^{m_Z \times n}$. If every row and column of H_X and H_Z has Hamming weight $\leq \ell$, we say that \mathcal{C} has locality ℓ . A family of qLDPC codes is a family of codes with constant locality ℓ and growing block length n .

It was a longstanding open question to construct linear-distance qLDPC codes. This question was resolved by Panteleev and Kalachev [20], who obtained qLDPC codes of linear distance and linear dimension. Subsequent works [13, 4] provided additional related constructions.

These codes in fact possess¹ the following stronger notion of distance, which guarantees that all low-weight errors have syndromes whose weight is linear in the error weight (as opposed to just having nonzero syndromes). Below, for a code C , we denote $|y|_C = \min_{y' \in C} |y + y'|$.

► **Definition 1** (Small-set (co)boundary expansion; restatement of Definition 14). *Let $C = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$ be a CSS code given by parity check matrices $H_X \in \mathbb{F}_q^{m_X \times n}$ and $H_Z \in \mathbb{F}_q^{m_Z \times n}$. For $c_1, c_2 > 0$, we say that C has (c_1, c_2) -small-set boundary expansion if it holds for every $y \in \mathbb{F}_q^n$ with $|y| \leq c_1 n$ that*

$$\frac{|H_Z y|}{m_Z} \geq c_2 \frac{|y|_{C_X^\perp}}{n}.$$

Similarly, C has (c_1, c_2) -small-set coboundary expansion if it holds for every $y \in \mathbb{F}_q^n$ with $|y| \leq c_1 n$ that

$$\frac{|H_X y|}{m_X} \geq c_2 \frac{|y|_{C_Z^\perp}}{n}.$$

This notion of small-set (co)boundary expansion underlies both the NLTS Hamiltonians of [1] and the SoS lower bounds of [11]. Note that a code with (c_1, c_2) -small set boundary and coboundary expansion by definition has distance $\geq c_1 n$.

1.2 NLTS Hamiltonians from Low-Rate qLDPC Codes

The quantum PCP (qPCP) conjecture, which states that it is QMA-hard to compute a constant-factor approximation to the ground energy of a local Hamiltonian, is a major open question in quantum complexity theory that has remained largely elusive. Perhaps the most significant progress towards this conjecture was the NLTS theorem, which was recently proven by Anshu, Breuckmann, and Nirkhe [1] using an application of asymptotically good qLDPC codes. This result provides a family of local Hamiltonians that have “no low-energy trivial states” (NLTS), where a trivial state is one computed by a constant-depth circuit. The NLTS theorem therefore provides local Hamiltonians exhibiting a weaker form of hardness of approximation than required by qPCP, and is indeed a necessary consequence of the qPCP conjecture under the widely believed assumption that $\text{NP} \neq \text{QMA}$.

Anshu et al. [1] constructed their NLTS Hamiltonians using the asymptotically good quantum Tanner codes of [13]. In particular, their proof of NLTS relied on the codes having both linear distance and dimension. It was an open question whether such linear dimension was necessary for NLTS. This question is motivated by the suggestion [19] that constructing more general families of NLTS Hamiltonians may lead to further progress towards the qPCP conjecture. Furthermore, some earlier partial progress towards NLTS used codes of smaller dimension [6], which again raises the question of whether linear dimension is necessary. Our main result on NLTS resolves this question, as we obtain NLTS Hamiltonians from qLDPC codes of arbitrary positive dimension.

¹ [11] were the first to consider small-set (co)boundary expansion for linear-distance qLDPC codes, and showed that the codes of [13] possess this property. [4] later constructed additional good qLDPC codes for which they proved this expansion property. We explain at the end of Section 3.4 why the decoder of [15, 14] implies that the codes of [20] also possess this expansion property.

NLTS Hamiltonians are formally defined as follows. Recall that a family of Hamiltonians is ℓ -local if every \mathbf{H} in the family can be expressed as a sum of Hamiltonians, each of which act nontrivially on $\leq \ell$ qubits. If $\ell = O(1)$ we say the family is local. We also say that a state ρ is an ϵ -approximate ground state of a Hamiltonian $\mathbf{H} \succeq 0$ if $\text{Tr}(\rho\mathbf{H}) \leq \epsilon$.

► **Definition 2** (NLTS Hamiltonians). *A family of local Hamiltonians $(\mathbf{H}_n)_{n \rightarrow \infty}$ with $0 \preceq \mathbf{H}_n \preceq I$ is NLTS if there exists $\epsilon > 0$ such that the minimum depth of any quantum circuit computing an ϵ -approximate ground state of \mathbf{H}_n approaches ∞ as $n \rightarrow \infty$.*

Recall that for a CSS code $\mathcal{C} = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$, the associated code Hamiltonian is given by

$$\mathbf{H} = \frac{1}{2}(\mathbf{H}_X + \mathbf{H}_Z)$$

for

$$\mathbf{H}_X = \frac{1}{m_X} \sum_{y \in \text{rows}(H_X)} \frac{I - X^y}{2}$$

$$\mathbf{H}_Z = \frac{1}{m_Z} \sum_{y \in \text{rows}(H_Z)} \frac{I - Z^y}{2},$$

where X and Z denote the respective Pauli operators. Thus in particular the ground space of \mathbf{H} is precisely the code space $\mathcal{C} = \text{span}\{\sum_{y' \in C_X^\perp} |y + y'\rangle : y \in C_Z\}$.

Anshu et al. [1] showed that for every family of qLDPC codes with linear dimension and constant small-set boundary and coboundary expansion, the associated code Hamiltonians are NLTS. Thus for instance the quantum Tanner codes of [13] yield NLTS code Hamiltonians.

Our result below improves upon this result of [1] by removing the linear dimension requirement.

► **Theorem 3** (NLTS from low-rate codes; informal statement of Corollary 23). *Let $(\mathcal{C}^{(n)})_{n \rightarrow \infty}$ be an infinite family of qLDPC codes over the alphabet \mathbb{F}_2 of block length n and positive dimension that have (c_1, c_2) -small set boundary and coboundary expansion for some constants $c_1, c_2 > 0$. Then the family of associated code Hamiltonians $(\mathbf{H}^{(n)})_{n \rightarrow \infty}$ is NLTS.*

Our proof of Theorem 3 follows the general framework of [6, 1] in showing circuit lower bounds for code Hamiltonians. Specifically, Eldar and Harrow [6] showed that in order to show the code Hamiltonians \mathbf{H} are NLTS, it suffices to show that every distribution obtained by measuring an approximate ground state of \mathbf{H} in either the X or Z basis is well spread. Here a distribution D over \mathbb{F}_2^n is well spread if there exist sets $S_0, S_1 \subseteq \mathbb{F}_2^n$ separated by a linear Hamming distance $\text{dis}(S_0, S_1) \geq \Omega(n)$ such that D assigns constant probability $D(S_0), D(S_1) \geq \Omega(1)$ to both sets.

Both [6, 1] show this well-spreadness property for code Hamiltonians by combining a distance/expansion property of the code with an uncertainty principle. However, the two works different use assumptions on the code as well as different uncertainty principles:

- [6] assumes the code is locally testable and of linear distance, which implies the approximate ground states have a certain linear structure. They then use an uncertainty principle (see Lemma 19) that is able to leverage this linear structure and prove well-spreadness regardless of the code dimension.

- [1] assumes the code has small-set boundary and coboundary expansion, which is weaker than local testability and therefore yields less structure in the approximate ground states. They then use a different uncertainty principle with which they are still able to prove well-spreadness, but only for codes of linear dimension.

Because linear-distance quantum locally testable codes are not known to exist, the NLTS Hamiltonians of [6] remain conjectural.

We prove Theorem 3 by combining these two approaches: we make the weaker assumption that our code has small-set boundary and coboundary expansion, but show that the approximate ground states still have enough linear structure to apply the uncertainty principle in Lemma 19. We then conclude that the code Hamiltonians are NLTS regardless of the code dimension.

At the core of our argument is the application of a “decoding” procedure for approximate ground states of codes with small-set (co)boundary expansion, which is unintuitive in the sense that far-apart approximate ground states may decode to the the same true ground state. However, we are able to show that in some sense, the low-energy space of the code Hamiltonian acts similarly enough to a true code space that the argument still goes through.

1.3 Planted Quantum LDPC Codes

This section presents our result on planting a nontrivial codeword in qLDPC codes.

The recent breakthrough constructions of linear-distance qLDPC codes ([20], followed by [13, 4]) all bound the code dimension by counting parity checks. Specifically, these works use the fact that if $\mathcal{C} = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$ for $H_X \in \mathbb{F}_q^{m_X \times n}, H_Z \in \mathbb{F}_q^{m_Z \times n}$, then \mathcal{C} has dimension $k \geq n - m_X - m_Z$. However, this bound may not be tight if there are redundant parity checks in H_X, H_Z . Indeed, it has been an open question in the coding theory literature to provide new ways of ensuring that LDPC codes have positive dimension; for instance, this question was of central importance in the code constructions of [2, 5].

Our result below makes progress on this question, by showing how to plant a nontrivial codeword in the linear-distance quantum Tanner codes of [13]. In fact, we show that like the codes of [13] our planted codes possess small-set (co)boundary expansion.

► **Theorem 4** (Planted quantum Tanner codes; restatement of Theorem 29). *For every finite field \mathbb{F}_q of characteristic $p \geq 7$, there exist constants $c_1, c_2 > 0$ such that there is a strongly explicit infinite family $(\mathcal{C}^{(n)})_{n \rightarrow \infty}$ of quantum LDPC CSS codes for which every $\mathcal{C}^{(n)} = \text{CSS}(C_X^{(n)}, C_Z^{(n)})$ with $C_X^{(n)}, C_Z^{(n)} \subseteq \mathbb{F}_q^n$ has the following properties:*

1. $\mathcal{C}^{(n)}$ has (c_1, c_2) -small-set boundary and coboundary expansion, and therefore has distance $\geq c_1 n$.
2. The all-1s vector $\mathbf{1} \in \mathbb{F}_q^n$ lies in $C_X^{(n)} \setminus C_Z^{(n)\perp}$ and in $C_Z^{(n)} \setminus C_X^{(n)\perp}$.

While the proof of Theorem 4 that we present here requires the field to have characteristic $p \geq 7$, in follow-up work we have generalized the result to arbitrary fields. The general proof is given in the full version of our paper [8]. The reason that the $p < 7$ case is more difficult is discussed below. Note that for simplicity we only proved our NLTS result (Theorem 3) for a binary alphabet, and thus it only applies to our more general planted quantum Tanner codes described in the full version [8].

Our construction of planted quantum Tanner codes is motivated by a more basic classical analogue. Recall that a classical Tanner code is specified by a Δ -regular graph Γ and an inner code $C_{\text{in}} \subseteq \mathbb{F}_q^\Delta$, where the code components correspond to edges of the graph, and the parity checks impose the constraint that the local view of each vertex is a codeword in C_{in} .

The standard method for ensuring a classical Tanner code C has positive rate is to require C_{in} to have sufficiently large rate $> 1/2$, and then to bound the number of resulting linear constraints on C from the parity checks. However, we may alternatively simply require that C_{in} contain the all-1s vector $\mathbf{1} \in \mathbb{F}_q^\Delta$, so that C then must contain the global all-1s vector $\mathbf{1} \in \mathbb{F}_q^n$. If C contains no other nontrivial codewords, then it is a repetition code, which is typically uninteresting classically.

However, we construct a quantum analogue of this construction, which is more nuanced, and has interesting complexity theoretic implications. Indeed, whereas classically it is easy to achieve linear distance and positive dimension by taking a repetition code, to the best of our knowledge the only known quantum LDPC codes of linear distance and positive dimension are the recent constructions of [20, 13, 4], which can in fact achieve linear dimension.

Recall that a quantum Tanner code $\mathcal{C} = \text{CSS}(C_X, C_Z)$ [13] is constructed by imposing constraints from a *pair* of classical codes $C_A, C_B \subseteq \mathbb{F}_q^\Delta$ on a *square Cayley complex* (V, E, Q) , which is a graph (V, E) with the additional high-dimensional structure of faces, or squares, in Q ; the qudits of the code correspond to the $n = |Q|$ faces in Q .

To prove Theorem 4, we show that if we require the local all-1s vector $\mathbf{1} \in \mathbb{F}_q^\Delta$ to lie in C_A and in C_B^\perp , and q is relatively prime with n , then the global all-1s vector $\mathbf{1} \in \mathbb{F}_q^n$ lies in $C_Z \setminus C_X^\perp$ and $C_X \setminus C_Z^\perp$, so in particular $\mathcal{C} = \text{CSS}(C_X, C_Z)$ has dimension ≥ 1 .

The proof that $\mathbf{1} \in C_A, C_B^\perp$ implies $\mathbf{1} \in C_X, C_Z$ is immediate, as in the classical case. However, we prove that $\mathbf{1} \notin C_X^\perp, C_Z^\perp$ using a parity (or more precisely, arity) mismatch: we argue that C_X^\perp and C_Z^\perp are spanned by vectors whose components sum to $0 \in \mathbb{F}_q$, whereas the components of $\mathbf{1} \in \mathbb{F}_q^n$ do not sum to 0 by the assumption that q, n are relatively prime, so that the characteristic p of \mathbb{F}_q does not divide n . As the strongly explicit square Cayley complexes we use here, which are based on the Cayley expanders of [18], have $n = |Q|$ divisible by 2, 3, 5, we must take $p \geq 7$ in this construction. We show how to remove this limitation in the full version of our paper [8] by using square Cayley complexes based on the Cayley expanders given in Example 3.4 of [17], for which the number of vertices is a power of any desired prime.

We still must show that the resulting planted quantum Tanner codes have good small-set (co)boundary expansion and therefore good distance. By the results of [13], it suffices to show that the inner codes (C_A, C_B) can be chosen to possess a property called *product-expansion* (Definition 8). This property was shown for random inner codes by [12, 4]; we extend the proof of [12] for our case of planted inner codes where $\mathbf{1} \in C_A, C_B^\perp$. As these inner codes are constant-sized as $n \rightarrow \infty$, the randomized construction can be made strongly explicit by a brute force search.

An interesting consequence of our result is that we can construct planted quantum Tanner codes \mathcal{C} of positive dimension $k > 0$ with inner codes C_A, C_B of any desired respective rates $R_A, R_B \in (0, 1)$; for instance, we can take $R_A = R_B$. In contrast, the prior technique of bounding k by counting parity checks only implies that $k \geq -(1 - 2R_A)(1 - 2R_B) \cdot n$, which never gives a meaningful bound when $R_A = R_B$. Thus our construction allows instantiations in new parameter regimes.

We also remark that while we only show how to plant a nontrivial codeword in the qLDPC codes of [13], our techniques also apply to the codes of [20]; to avoid redundancy we do not spell out the details.

1.4 Strongly Explicit SoS Lower Bounds

The Sum-of-Squares semidefinite programming hierarchy is one of the most powerful algorithmic frameworks for approximating the satisfiability of CSPs (see [7] for a survey). However, almost all of the known hard instances (i.e. lower bounds) for this hierarchy are

given by randomized constructions. Hopkins and Lin [11], building on the techniques of Dinur et al. [3], constructed the first explicit unsatisfiable CSPs that cannot be refuted by a linear number of levels of the SoS SDP hierarchy. In contrast, explicit lower bounds prior to their work applied to at best a logarithmic number of levels of the SoS hierarchy.

Hopkins and Lin [11] proved their result by showing that hard instances for SoS can be obtained from a family of qLDPC codes with small-set boundary and coboundary expansion. Explicit such qLDPC codes, such as the quantum Tanner codes of [13], then yield the desired explicit hard CSPs.

► **Remark 5.** The SoS lower bounds of [11] marked the first complexity theoretic application of linear-distance qLDPC codes; the subsequent proof of the NLTS theorem [1] provided a second notable application. Such applications were perhaps surprising given that the construction of asymptotically good qLDPC codes, first obtained by [20] and subsequently extended and modified by [13, 4], was originally motivated in large part by applications to quantum error correction.

However, the explicitness of the CSP construction in [11] was weak in the sense of Definition 6 below. One of the major questions left open by their work was to make this construction strongly explicit [10]. We apply our construction of planted quantum Tanner codes in Theorem 4 to resolve this problem.

► **Definition 6** (Weak vs. strong explicitness). *Let $X = (x_n)_{n \in \mathbb{N}}$ be an infinite family of objects such that each x_n can be represented by a bitstring $x_n \in \{0, 1\}^{a_n}$ of length a_n , where $a_n \rightarrow \infty$ as $n \rightarrow \infty$. We say that X is:*

- **weakly explicit** (or simply “explicit”) if there exist a $\text{poly}(a_n)$ -time algorithm $A(n)$ that outputs x_n
- **strongly explicit** if there exists a $\text{poly}(\log n, \log a_n)$ -time algorithm $A(n, i)$ that outputs the i th bit of x_n for $i \in [a_n]$.

For instance, a CSS code $\mathcal{C} = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$ is weakly (resp. strongly) explicit if the matrices H_X, H_Z are weakly (resp. strongly) explicit, meaning that the (i, j) th entry of each matrix can be computed in time $\text{poly}(n, m_X, m_Z)$ (resp. $\text{poly}(\log n, \log m_X, \log m_Z)$).

Similarly, consider a family of CSPs given by ℓ -LIN instances, which are defined by n linear constraints on m variables over a fixed finite field \mathbb{F}_q , such that each linear equation has $\leq \ell = O(1)$ nonzero coefficients. A family of such ℓ -LIN instances is weakly (resp. strongly) explicit if the i th linear equation can be computed in time $\text{poly}(n, m)$ (resp. $\text{poly}(\log n, \log m)$).

Given a qLDPC code $\mathcal{C} = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$ and an arbitrary element $\beta \in C_X \setminus C_Z^\perp$, Hopkins and Lin [11] considered the associated ℓ -LIN instance $\mathcal{I}_{\mathcal{C}, \beta}$ with $m = m_Z$ variables y_1, \dots, y_m and n linear constraints over \mathbb{F}_q given by the system of equations $H_Z^\top y = \beta$ for $y = (y_1, \dots, y_m)$. They showed that if \mathcal{C} has $(\Omega(1), \Omega(1))$ -small-set boundary and coboundary expansion, then at most $1 - \Omega(1)$ fraction of the constraints in $\mathcal{I}_{\mathcal{C}, \beta}$ can be satisfied, but $\mathcal{I}_{\mathcal{C}, \beta}$ is hard to refute for $\Omega(n)$ levels of SoS.

However, even if \mathcal{C} comes from a strongly explicit family of qLDPC codes, the associated ℓ -LIN instance $\mathcal{I}_{\mathcal{C}, \beta}$ is only weakly explicit in general, as one must perform Gaussian elimination to compute some $\beta \in C_X \setminus C_Z^\perp$, which takes $\text{poly}(n, m)$ time.

Because our planted quantum Tanner codes in Theorem 4 by construction have $\mathbf{1} \in C_X \setminus C_Z^\perp$, they resolve this issue, and hence yield the following result.

► **Theorem 7** (Strongly explicit SoS lower bounds; restatement of Corollary 33). *For every prime $p \geq 7$, the ℓ -LIN instances $\mathcal{I}_{C,1}$ for planted quantum Tanner codes \mathcal{C} over the alphabet \mathbb{F}_p provide a family of strongly explicit instances with satisfiability $\leq (1 - \Omega(1))$, such that no instance can be refuted by cn levels of the SoS hierarchy for a sufficiently small constant $c > 0$.*

We remark that [11] actually restricted attention to the binary alphabet $q = 2$ case, though their techniques extend to larger prime alphabets. They also provide a reduction that yields hard ℓ -LIN instances with locality $\ell = 3$. While we only state Theorem 7 here for prime alphabets $q \geq 7$, we generalize to arbitrary prime alphabets in the full version of our paper [8], using the more general planted quantum Tanner codes in [8] mentioned in Section 1.3.

1.5 Open Questions

Our results raise the following open questions:

- Can our construction of NLTS Hamiltonians from low-rate qLDPC codes lead to more progress towards qPCP or hardness of approximation? For instance, perhaps the fact that low-rate codes, which correspond to Hamiltonians with low-dimensional ground spaces, suffice for NLTS will be helpful in constructing Hamiltonians with stronger hardness of approximation guarantees.
- Our results highlight the usefulness of low-rate qLDPC codes, and suggest that for complexity theoretic applications there is often little benefit to having high rate. However, to the best of our knowledge, our planted quantum Tanner codes provide the only known “inherently” low-rate qLDPC codes, and they still have high rate in some parameter regimes. In contrast, there are many interesting classical low-rate LDPC codes such as Hadamard and Reed-Muller codes, which have properties not shared by any high-rate codes. In the quantum case, can similar stronger properties be obtained by allowing for low rate in qLDPC codes?

2 Notation

For a string $y \in \mathbb{F}_q^n$, we denote the Hamming weight by $|y| = |\{i \in [n] : y_i \neq 0\}|$. For subsets $S, T \subseteq \mathbb{F}_q^n$, we denote the Hamming distance by $\text{dis}(S, T) = \min_{s \in S, t \in T} |s - t|$.

Unless explicitly stated otherwise, by a “code” we mean a linear subspace $C \subseteq \mathbb{F}_q^n$. The code C has block length n , dimension $k = \dim_{\mathbb{F}_q}(C)$, and distance $d = \min_{y \in C \setminus \{0\}} |y|$, which can be summarized by saying it is a $[n, k, d]_q$ code. The dual code is $C^\perp = \{x \in \mathbb{F}_q^n : x \cdot y = 0 \forall y \in C\}$.

For codes $C_i \subseteq \mathbb{F}_q^{n_i}$ for $i = 1, 2$, the tensor code $C_1 \otimes C_2 \subseteq \mathbb{F}_q^{n_1 \times n_2}$ consists of all $n_1 \times n_2$ matrices where every column lies in C_1 and every row lies in C_2 . The dual of the tensor code is $(C_1 \otimes C_2)^\perp = C_1^\perp \otimes \mathbb{F}_q^{n_2} + \mathbb{F}_q^{n_1} \otimes C_2^\perp$.

Given a Δ -regular graph Γ with n edges and an inner code $C_{\text{in}} \subseteq \mathbb{F}_q^\Delta$, we denote the associated classical Tanner code by $C = \text{Tan}(\Gamma, C_{\text{in}}) \subseteq \mathbb{F}_q^n$, which is constructed as follows. We associate the set of all edges in Γ with the set $[n]$, and we associate the set of edges incident to each vertex v in Γ with the set $[\Delta]$. Then we define C to be the set of all edge labelings $y \in \mathbb{F}_q^n = \mathbb{F}_q^{E(\Gamma)}$ such that the labels of edges incident to each $v \in \Gamma$ form a codeword in C_{in} .

For a pure quantum state $|\psi\rangle$, we denote the density matrix by $\psi = |\psi\rangle\langle\psi|$. For a set $S \subseteq \mathbb{F}_q^n$, we let $|S\rangle = |S|^{-1/2} \sum_{s \in S} |s\rangle$ denote the uniform superposition over elements of S .

The quantum codes we consider in this paper are CSS codes, which are defined as follows. For classical codes $C_X, C_Z \subseteq \mathbb{F}_q^n$ such that $C_X^\perp \subseteq C_Z$, the associated quantum CSS code $\mathcal{C} = \text{CSS}(C_X, C_Z)$ is defined by $\mathcal{C} = \text{span}\{|y + C_X^\perp\rangle : y \in C_Z\} \subseteq (\mathbb{C}^q)^{\otimes n}$. This code has block length n , dimension $k = \log_q \dim_{\mathbb{C}}(\mathcal{C}) = \dim_{\mathbb{F}_q}(C_Z) - \dim_{\mathbb{F}_q}(C_X^\perp)$, and distance $d = \min_{y \in (C_Z \setminus C_X^\perp) \cup (C_X \setminus C_Z^\perp)} |y|$, which can be summarized by saying that \mathcal{C} is a $[[n, k, d]]_q$ code.

If $C_X = \ker H_X$ and $C_Z = \ker H_Z$ for parity check matrices H_X, H_Z in which each row and column has Hamming weight $\leq \ell$, we say that \mathcal{C} is a CSS code with check weight, or locality, $\leq \ell$. A family of codes with constant locality $\ell = O(1)$ as $n \rightarrow \infty$ is said to be LDPC. The family of codes is (strongly) explicit if the associated families of parity check matrices H_X, H_Z are (strongly) explicit.

3 Review of Quantum Tanner Codes

In this section we review the construction and relevant properties of the asymptotically good quantum LDPC codes of Leverrier and Zémor [13, 14], which are called quantum Tanner codes. Although [13, 14] present the construction over binary alphabets, we consider arbitrary finite field alphabets; all their results and proofs extend to this more general case with just some “+” signs changed to “−” signs for fields of characteristic $\neq 2$.

Recall that a classical Tanner code is constructed by imposing constraints from an inner code on a graph (see Section 2). In contrast, a quantum Tanner code \mathcal{C} is constructed by imposing constraints from *two* inner codes on a higher-dimensional object called a *square Cayley complex*. In particular, $\mathcal{C} = \text{CSS}(C_X, C_Z)$, where both C_X, C_Z are classical Tanner codes on graphs obtained from a square Cayley complex, with distinct inner codes.

3.1 Construction

We now describe the construction of a quantum Tanner code $\mathcal{C} = \text{CSS}(C_X, C_Z)$. We first need to define a square Cayley complex. Recall that for a group G and a subset $A \subseteq G$, the Cayley graph $\text{Cay}(G, A)$ has vertex set G and edge set $\{(g, ag) : g \in G, a \in A\}$. As described below, a square Cayley complex is a sort of 2-dimensional generalization of a Cayley graph.

A square Cayley complex consists a tuple (V, E, Q) of vertices, edges, and faces (or “squares”) that is specified by a group G and two generating sets $A, B \subseteq G$ as follows. We typically take $|A| = |B| = \Delta = O(1)$ as $|G| = \Theta(n) \rightarrow \infty$, and assume that $A = A^{-1}$ and $B = B^{-1}$ are closed under inversion. The complex then has vertex set $V = G \times \{0, 1\}^2$, edge set $E = E_A \sqcup E_B$ for

$$\begin{aligned} E_A &= \{(g, i0), (ag, i1) : g \in G, i \in \{0, 1\}, a \in A\} \\ E_B &= \{(g, 0j), (gb, 1j) : g \in G, j \in \{0, 1\}, b \in B\}, \end{aligned}$$

and face set

$$Q = \{(g, 00), (ag, 01), (gb, 10), (agb, 11) : g \in G, a \in A, b \in B\}.$$

For $i, j \in \{0, 1\}$, let $V_{ij} = G \times (i, j)$. Define bipartite graphs $\Gamma_0 = (V_{00} \sqcup V_{11}, Q)$ and $\Gamma_1 = (V_{01} \sqcup V_{10}, Q)$ whose edges are given by pairs of vertices that form a diagonal in a square in Q ; for instance, Γ_0 has an edge between $v \in V_{00}$ and $v' \in V_{11}$ if v, v' share a face in Q . Observe that both Γ_0 and Γ_1 have a unique edge associated to each square in Q . Furthermore, the Γ_i -edges incident to a vertex v correspond to the squares in Q that contain v ; we let $Q(v)$ denote the set of these squares. But by definition $Q(v)$ consists of an $A \times B$ grid of squares. For instance, for $v = (g, 00) \in V_{00}$ then

$$Q(v) = \{(g, 00), (ag, 01), (gb, 10), (agb, 11) : a \in A, b \in B\}.$$

Therefore given a square Cayley complex (V, E, Q) of degree $\Delta = |A| = |B|$ along with classical codes $C_A \subseteq \mathbb{F}_q^A = \mathbb{F}_q^\Delta$ and $C_B \subseteq \mathbb{F}_q^B = \mathbb{F}_q^\Delta$, we may define a quantum Tanner code $\mathcal{C} = \text{CSS}(C_X, C_Z)$ by

$$\begin{aligned} C_X &= \text{Tan}(\Gamma_0, (C_A \otimes C_B)^\perp) \\ C_Z &= \text{Tan}(\Gamma_1, (C_A^\perp \otimes C_B^\perp)^\perp). \end{aligned}$$

That is, C_X and C_Z are classical Tanner codes on the graphs Γ_0 and Γ_1 respectively, where the inner codes are given by dual tensor codes. Because $E(\Gamma_0) \cong E(\Gamma_1) \cong Q$, both C_X and C_Z are subspaces of \mathbb{F}_q^Q .

3.2 Locality and Dimension

We now describe some basic properties of \mathcal{C} . By definition \mathcal{C} has block length $n = |Q|$. Parity checks for C_X are given by tensor codewords in $C_A \otimes C_B$ supported in the neighborhood $Q(v_0)$ of any $v_0 \in V_{00} \sqcup V_{11}$. Similarly, parity checks for C_Z are given by tensor codewords in $C_A^\perp \otimes C_B^\perp$ supported in the neighborhood $Q(v_1)$ of any $v_1 \in V_{01} \sqcup V_{10}$. Because any such $Q(v_0), Q(v_1)$ are either disjoint or intersect in a single row or column, the parity checks for C_X and C_Z are orthogonal, so $C_X^\perp \subseteq C_Z$. Furthermore, as $|Q(v_0)| = |Q(v_1)| = \Delta^2$, the quantum Tanner code \mathcal{C} is LDPC with locality $\Delta^2 = O(1)$ as $n = |Q| \rightarrow \infty$.

Counting parity checks to bound the number of linear constraints on C_X, C_Z implies that \mathcal{C} has dimension $k \geq -(1-2R_A)(1-2R_B) \cdot n$, where $R_A = \dim(C_A)/\Delta$ and $R_B = \dim(C_B)/\Delta$ denote the rate of C_A and C_B respectively.

3.3 Distance

To present the distance bound for quantum Tanner codes, we need the following definition.

► **Definition 8** (Product-expansion). *A pair of codes $C_1, C_2 \subseteq \mathbb{F}_q^n$ is ρ -product-expanding if every $x \in (C_1^\perp \otimes C_2^\perp)^\perp = C_1 \otimes \mathbb{F}_q^n + \mathbb{F}_q^n \otimes C_2$ can be decomposed as $x = c + r$ for some $c \in C_1 \otimes \mathbb{F}_q^n$ and $r \in \mathbb{F}_q^n \otimes C_2$ satisfying*

$$|x| \geq \rho n (|c|_{\text{col}} + |r|_{\text{row}}),$$

where $|c|_{\text{col}}$ denotes the number of nonzero columns in c and $|r|_{\text{row}}$ denotes the number of nonzero rows in r .

It is immediate that product-expansion yields a bound on the distances of the associated codes:

► **Lemma 9** (Well known). *If the pair $C_1, C_2 \subseteq \mathbb{F}_q^n$ is ρ -product expanding, then C_1 and C_2 have distance $\geq \rho n$.*

Proof. Let $x \in C_1 \otimes \mathbb{F}_q^n$ have its first column be a minimum-weight nonzero codeword of C_1 , and have all other columns be 0. Then ρ -product-expansion implies that C_1 has distance $|x| \geq \rho n$. A similar argument holds for C_2 . ◀

The following result bounding the product expansion of random pairs of codes was shown independently by [12] and [4], though only the former explicitly considered non-binary alphabets.

► **Proposition 10** ([12]). *For every fixed $\epsilon > 0$, there exists a constant $\rho = \rho(\epsilon) > 0$ and a function $\delta(n) = \delta(n; \epsilon) \rightarrow 0$ as $n \rightarrow \infty$ such that the following holds. For every pair of integers $k_1, k_2 \in (\epsilon n, (1 - \epsilon)n)$, if $C_i \subseteq \mathbb{F}_q^n$ for $i = 1, 2$ is drawn uniformly at random from the set of linear codes of dimension k_i , then with probability $\geq 1 - \delta(n)$ the pair (C_1, C_2) will be ρ -product-expanding.*

Applying Proposition 10 with a union bound over (C_1, C_2) and (C_1^\perp, C_2^\perp) immediately yields the following corollary.

► **Corollary 11** ([12]). *Defining all variables as in Proposition 10, then with probability $\geq 1 - 2\delta(n)$ both (C_1, C_2) and (C_1^\perp, C_2^\perp) will be ρ -product-expanding.*

The distance bound for quantum Tanner codes will also rely on the Cayley graphs $\text{Cay}(G, A)$ and $\text{Cay}(G, B)$ having sufficiently good expansion. Recall that a Δ -regular graph is *Ramanujan* if the second largest absolute value of an eigenvalue of its (unnormalized) adjacency matrix is $\leq 2\sqrt{\Delta - 1}$. Constructions of Ramanujan Cayley graphs have for instance been given by [16] and [18]; we use the latter construction as it is strongly explicit.

► **Theorem 12** ([18]). *For every prime power $q \geq 3$, there exists a strongly explicit family of $(q + 1)$ -regular Ramanujan Cayley graphs $(\Gamma_m)_{m \in \mathbb{N}}$ with the number of vertices given by*

$$|V(\Gamma_m)| = \begin{cases} q^{2m}(q^{4m} - 1), & q \equiv 0 \pmod{2} \\ q^{2m}(q^{4m} - 1)/2, & q \equiv 1 \pmod{2}. \end{cases}$$

We are now ready to present the distance bound for quantum Tanner codes.

► **Theorem 13** ([13, 14]). *For every fixed $\rho > 0$, the following holds for all sufficiently large Δ . Let \mathcal{C} be a quantum Tanner code for which:*

1. $\text{Cay}(G, A), \text{Cay}(G, B)$ are Ramanujan graphs of degree Δ .
2. $(C_A, C_B), (C_A^\perp, C_B^\perp)$ are ρ -product-expanding.

Then \mathcal{C} has distance $d \geq cn$ for a constant $c > 0$ depending only on ρ, Δ .

Recall that by Lemma 9, Condition 2 in Theorem 13 implies that $C_A, C_B, C_A^\perp, C_B^\perp$ have distance $\geq \rho\Delta$.

Condition 1 in Theorem 13 can be met using the strongly explicit Ramanujan graphs in Theorem 12. If $C_A, C_B \subseteq \mathbb{F}_q^\Delta$ are chosen to be random codes of some fixed rates $0 < R_A, R_B < 1$ for any sufficiently large constant Δ , then Condition 2 is met by Corollary 11. Because Δ is a constant as $n \rightarrow \infty$, we may find C_A, C_B in constant time by a brute force search, so the overall construction of \mathcal{C} is strongly explicit.

3.4 Small-Set (Co)boundary Expansion

For our applications of quantum Tanner codes, we will need a stronger notion than distance, called *small-set (co)boundary expansion*, which was first formally stated in the context of quantum codes by Hopkins and Lin [11]. Below, for a code \mathcal{C} , we denote $|y|_{\mathcal{C}} = \min_{y' \in \mathcal{C}} |y + y'|$.

► **Definition 14** (Small-set (co)boundary expansion). *Let $\mathcal{C} = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$ be a CSS code given by parity check matrices $H_X \in \mathbb{F}_q^{m_X \times n}$ and $H_Z \in \mathbb{F}_q^{m_Z \times n}$. For $c_1, c_2 > 0$, we say that \mathcal{C} has (c_1, c_2) -small-set boundary expansion if it holds for every $y \in \mathbb{F}_q^n$ with $|y| \leq c_1 n$ that*

$$\frac{|H_Z y|}{m_Z} \geq c_2 \frac{|y|_{C_X^\perp}}{n}.$$

Similarly, \mathcal{C} has (c_1, c_2) -small-set coboundary expansion if it holds for every $y \in \mathbb{F}_q^n$ with $|y| \leq c_1 n$ that

$$\frac{|H_X y|}{m_X} \geq c_2 \frac{|y|_{C_Z^\perp}}{n}.$$

Small-set (co)boundary expansion immediately implies a bound on the distance of the code:

► **Lemma 15** (Well known). *If $\mathcal{C} = \text{CSS}(C_X, C_Z)$ of block length n has (c_1, c_2) -small set boundary and coboundary expansion for $c_1, c_2 > 0$, then \mathcal{C} has distance $\geq c_1 n$.*

Proof. For every $y \in \mathbb{F}_q^n \setminus C_X^\perp$ with $|y| \leq c_1 n$, then $|y|_{C_X^\perp} > 0$, so small-set boundary expansion implies that $|H_Z y| \geq c_2 m_Z |y|_{C_X^\perp} / n > 0$ and thus $y \notin C_Z$. An analogous argument shows that $C_X \setminus C_Z^\perp$ has no elements of weight $\leq c_1 n$. ◀

It was originally observed that quantum Tanner codes have small set (co)boundary expansion in [11]. We remark that another proof is given implicitly by the decoder of Leverrier and Zémor [14]. Specifically, there exists a constant $c_1 > 0$ such that for any errors $e_X, e_Z \in \mathbb{F}_q^n$ of sufficiently low weight $|e_X|, |e_Z| \leq c_1 n$, the decoder of [14] takes as input the syndromes $s_X = H_X e_X$ and $s_Z = H_Z e_Z$, and outputs some $e'_X \in e_X + C_Z^\perp$ and $e'_Z \in e_Z + C_X^\perp$ such that $|e'_X| \leq O(|s_X|)$, $|e'_Z| \leq O(|s_Z|)$. It follows that $|e_X|_{C_Z^\perp} \leq |e'_X| \leq O(|s_X|)$ and $|e_Z|_{C_X^\perp} \leq |e'_Z| \leq O(|s_Z|)$, which are precisely the conditions required by small-set coboundary and boundary expansion, respectively. The result below formally summarizes this small-set (co)boundary expansion.

► **Theorem 16** ([11]; also implicit in [14]). *For every fixed $\rho > 0$, the following holds for all sufficiently large Δ . Let \mathcal{C} be a quantum Tanner code that satisfies Conditions 1, 2 in the statement of Theorem 13. Then \mathcal{C} has (c_1, c_2) -small-set boundary and coboundary expansion for constants $c_1, c_2 > 0$ depending only on the values of ρ, Δ .*

Note that Theorem 16 implies Theorem 13 by Lemma 15.

Leverrier and Zémor [15] show how their decoder can also be used to decode the asymptotically good qLDPC codes of Panteleev and Kalachev [20]; again in this case the decoder outputs an error whose weight is linear in the syndrome weight. Therefore a similar result as Theorem 16 holds for the codes of [20] as well.

4 NLTS Hamiltonians from Codes of Arbitrary Dimension

In this section, we show that quantum LDPC codes with linear distance and an appropriate clustering property yield NLTS Hamiltonians, regardless of the code dimension. This result improves upon the prior construction of NLTS Hamiltonians of [1], which required the stronger assumption that the code dimension be linearly large. For simplicity in this section, we restrict attention to binary alphabets, though we expect the results to generalize naturally to qudits for more general alphabet sizes.

4.1 Setup of the Local Hamiltonian

We let $\mathcal{C} = \text{CSS}(C_X, C_Z) = \text{span}\{|y + C_X^\perp\rangle : y \in C_Z\}$ be an $[[n, k, d]]_2$ quantum LDPC CSS code, where all parity checks have weight $\leq \ell$. We assume \mathcal{C} belongs to a family of such codes with constant relative distance $d/n = \Omega(1)$ and constant locality $\ell = O(1)$ as the block length $n \rightarrow \infty$. We will also assume that \mathcal{C} satisfies the clustering property described in

Definition 17 below. The main novel aspect of our proof is that it holds for any nonzero code dimension $k > 0$. In contrast, the prior NLTS proof [1] assumed the rate k/n is constant as $n \rightarrow \infty$.

Recent good qLDPC codes, such as the quantum Tanner codes of [13], satisfy all of the conditions above, and have constant rate. Our planted quantum Tanner codes in Theorem 29 also satisfy these conditions, except that they are defined over non-binary alphabets; we suspect our results on NLTS Hamiltonians below should generalize to this non-binary case, though for simplicity in the presentation we do not pursue this direction.

Denote the parity check matrices of C_X and C_Z by $H_X \in \mathbb{F}_2^{m_X \times n}$ and $H_Z \in \mathbb{F}_2^{m_Z \times n}$ respectively, so that $C_X = \ker H_X$ and $C_Z = \ker H_Z$. By assumption all rows of H_X, H_Z have $\leq \ell$ nonzero entries. Also define $G_X^\epsilon = \{y \in \mathbb{F}_2^n : |H_X y| \leq \epsilon m_Z\}$, and define G_Z^ϵ analogously. We assume \mathcal{C} satisfies the following clustering property for G_X^ϵ and G_Z^ϵ , which is stated as Property 1 in [1]. Below, we denote $|y|_{\mathcal{C}} = \min_{y' \in \mathcal{C}} |y + y'|$.

► **Definition 17** (Clustering property [1]). *For constants $c_1, c_2, \epsilon_0 > 0$, we say that $\mathcal{C} = \text{CSS}(C_X, C_Z)$ exhibits (c_1, c_2, ϵ_0) -clustering if for all $0 < \epsilon < \epsilon_0$, the following hold:*

1. *Every $y \in G_X^\epsilon$ satisfies either $|y|_{C_Z^\perp} \leq c_1 \epsilon n$ or $|y|_{C_Z^\perp} \geq c_2 n$.*
2. *Every $y \in G_Z^\epsilon$ satisfies either $|y|_{C_X^\perp} \leq c_1 \epsilon n$ or $|y|_{C_X^\perp} \geq c_2 n$.*

This clustering property follows from small-set (co)boundary expansion (Definition 14), as is shown below.

► **Lemma 18.** *If \mathcal{C} has (c'_1, c'_2) -small-set boundary and coboundary expansion, then \mathcal{C} has (c_1, c_2, ϵ_0) -clustering for $c_1 = 1/c'_2$, $c_2 = c'_1$, $\epsilon_0 = 1$.*

Proof. Assume that $y \in G_Z^\epsilon$ satisfies $|y|_{C_X^\perp} \leq c_2 n = c'_1 n$. Let y' be the minimum-weight element of $y + C_X^\perp$, so that $H_Z y' = H_Z y$ and $|y'| = |y|_{C_X^\perp}$. Then the small-set boundary expansion implies that $|H_Z y'|/m_Z \geq c'_2 \cdot |y'|/n$, so $|y|_{C_X^\perp} = |y'| \leq (n/c'_2 m_Z) |H_Z y'| \leq \epsilon n/c'_2 = c_1 \epsilon n$. Thus we have shown the desired clustering for G_Z^ϵ ; an analogous argument applies to G_X^ϵ . ◀

For the remainder of Section 4, we assume that \mathcal{C} satisfies (c_1, c_2, ϵ_0) -clustering for some constants $c_1, c_2, \epsilon_0 > 0$ as $n \rightarrow \infty$.

Following [1], we define our ℓ -local Hamiltonian \mathbf{H} to be the code Hamiltonian

$$\mathbf{H} = \frac{1}{2}(\mathbf{H}_X + \mathbf{H}_Z)$$

for

$$\mathbf{H}_X = \frac{1}{m_X} \sum_{y \in \text{rows}(H_X)} \frac{I - X^y}{2}$$

$$\mathbf{H}_Z = \frac{1}{m_Z} \sum_{y \in \text{rows}(H_Z)} \frac{I - Z^y}{2}.$$

Thus in particular the ground space of \mathbf{H} is precisely the code space $\mathcal{C} = \text{span}\{|y + C_X^\perp\rangle : y \in C_Z\}$

While our general proof will follow that of [1], our use of an uncertainty principle instead follows the earlier work of [6]. Below we state the uncertainty principle we will use, which appears implicitly in [6]; we also provide a proof in the full version of our paper [8].

► **Lemma 19** (Uncertainty principle [9, 6]). *Let A, B be Hermitian observables with $AB+BA = 0$ and $A^2 = B^2 = I$. Then for every (possibly mixed) state ρ , at least one of the inequalities $|\text{Tr}(A\rho)| \leq 1/2 + 1/2\sqrt{2}$ or $|\text{Tr}(B\rho)| \leq 1/2 + 1/2\sqrt{2}$ holds.*

We follow prior works such as [6, 1] in establishing circuit lower bounds for approximate ground states of \mathbf{H} by showing that the measurement distributions of these states are well-spread, in the following sense.

► **Definition 20.** *For $\mu, \delta > 0$, A probability distribution D over \mathbb{F}_2^n is (μ, δ) -spread if there exist $S_0, S_1 \subseteq \mathbb{F}_2^n$ such that $D(S_0) \geq \mu$, $D(S_1) \geq \mu$, and $\text{dis}(S_0, S_1) \geq \delta n$.*

We specifically use following result, which appears as Fact 4 in [1] but is similar to an earlier result of [6]. Below, for an n -qubit state ψ , we let D_X^ψ and D_Z^ψ denote the distributions over \mathbb{F}_2^n obtained by measuring ψ in the X and Z bases respectively.

► **Lemma 21** (Circuit lower bound [6, 1]). *Let ψ be a (possibly mixed) quantum state on n qubits such that the Z -measurement distribution D_Z^ψ is (μ, δ) -spread. Then any circuit (on $\geq n$ qubits) that constructs ψ must have depth at least*

$$\frac{1}{3} \log \left(\frac{\delta^2 n}{400 \log(1/\mu)} \right).$$

4.2 Statement of Main Result on NLTS Hamiltonians

In this section, we state our main technical result, which implies that the code Hamiltonian \mathbf{H} for a CSS code with linear distance that exhibits the clustering property is NLTS, that is, its approximate ground states cannot be constructed by constant-depth circuits. Crucially, we only assume that the dimension of the code is positive.

Specifically, our main technical result below shows that the measurement distribution of every approximate ground state of \mathbf{H} is well-spread in either the X or Z basis.

► **Theorem 22.** *Let \mathbf{H} be the code Hamiltonian for a $[[n, k, d]]_2$ CSS code $\mathcal{C} = \text{CSS}(C_X, C_Z)$ of positive dimension $k > 0$ that exhibits (c_1, c_2, ϵ_0) -clustering. For any*

$$\epsilon < \frac{1}{1000} \cdot \min \left\{ \frac{\epsilon_0}{2}, \frac{c_2}{4c_1}, \frac{d}{2c_1 n} \right\}, \quad (1)$$

let ρ be an ϵ -approximate ground state of \mathbf{H} , so that $\text{Tr}(\mathbf{H}\rho) \leq \epsilon$. Then at least one of D_X^ρ or D_Z^ρ is (μ, δ) -spread for $\mu = .02$ and $\delta = c_2$.

Lemma 21 immediately yields the following corollary.

► **Corollary 23** (\mathbf{H} is NLTS). *Define ϵ, μ, δ , and \mathbf{H} as in Theorem 22. Then no ϵ -approximate ground state of \mathbf{H} can be constructed by a circuit of depth less than*

$$\frac{1}{3} \log \left(\frac{\delta^2 n}{400 \log(1/\mu)} \right) + 1.$$

Our proof of Theorem 22 is similar to [6] in that we combine an uncertainty principle with a decoding procedure to obtain uncertainty for approximate ground states. We furthermore use the clustering property of \mathcal{C} similarly to [1]. As such, the key novel aspect of our proof is the use of a “decoding” procedure that handles clusters of approximate ground states which do not correspond to any true codeword.

4.3 Proof of Well-Spreadness for Approximate Ground States

In this section, we prove Theorem 22. Throughout this section, we maintain the notation in the statement of Theorem 22, so that $\mathcal{C} = \text{CSS}(C_X, C_Z)$ is a $[[n, k, d]]_2$ CSS code exhibiting (c_1, c_2, ϵ_0) -clustering, ρ is an ϵ -approximate ground state of \mathbf{H} for ϵ as in (1), and $\mu = .02$, $\delta = c_2$.

4.3.1 Reducing to Well-Spreadness of Pure States with Small Syndrome

We will first show that D_X^ρ and D_Z^ρ are mostly supported inside $G_X^{O(\epsilon)}$ and $G_Z^{O(\epsilon)}$ respectively, so that up to a small loss in parameters we may assume they are entirely supported inside these sets. We will also show that it suffices to consider pure states $\psi' = |\psi'\rangle\langle\psi'|$, rather than arbitrary mixed states ρ .

Formally, we may decompose our Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ into orthogonal subspaces as

$$\mathcal{H} = \bigoplus_{e_X + C_X \in \mathbb{F}_2^n / C_X, e_Z + C_Z \in \mathbb{F}_2^n / C_Z} X^{e_Z} Z^{e_X} \mathcal{C},$$

where the choices of coset representatives in the above sum does not matter because by definition $X^{e_Z} Z^{e_X} \mathcal{C} = \mathcal{C}$ for $c_X \in C_X$, $c_Z \in C_Z$. Observe furthermore that each subspace $X^{e_Z} Z^{e_X} \mathcal{C}$ is by definition an eigenspace of the code Hamiltonian \mathbf{H} with eigenvalue $|H_X e_X|/2m_X + |H_Z e_Z|/2m_Z$.

Set

$$\epsilon' = 1000\epsilon,$$

and let

$$\mathcal{C}^{\leq \epsilon'} = \bigoplus_{e_X + C_X : |H_X e_X| \leq \epsilon' m_X, e_Z + C_Z : |H_Z e_Z| \leq \epsilon' m_Z} X^{e_Z} Z^{e_X} \mathcal{C}.$$

Therefore $\mathcal{C}^{\leq \epsilon'}$ is the span of some of the eigenspaces of energy $\leq \epsilon'$, and contains all of the eigenspaces of energy $\leq \epsilon'/2$. Let $\Pi_{\mathcal{C}^{\leq \epsilon'}}$ denote projection onto this subspace. Note that by definition, every $|\psi'\rangle \in \mathcal{C}^{\leq \epsilon'}$ has $\text{supp}(D_X^{\psi'}) \subseteq G_X^{\epsilon'}$ and $\text{supp}(D_Z^{\psi'}) \subseteq G_Z^{\epsilon'}$.

We now reduce the task of proving Theorem 22 to the following proposition. Below, recall that we carry the definitions of \mathbf{H} , ρ , ϵ , μ , δ from Theorem 22.

► **Proposition 24.** *There exist sets $S_X^0, S_X^1, S_Z^0, S_Z^1 \subseteq \mathbb{F}_2^n$ such that $\text{dis}(S_X^0, S_X^1), \text{dis}(S_Z^0, S_Z^1) \geq \delta n$, and such that for every pure state $|\psi'\rangle \in \mathcal{C}^{\leq \epsilon'}$, either*

$$D_X^{\psi'}(S_X^0), D_X^{\psi'}(S_X^1) \geq \mu' \quad \text{or} \quad D_Z^{\psi'}(S_Z^0), D_Z^{\psi'}(S_Z^1) \geq \mu', \quad (2)$$

where $\mu' = 1/4 - 1/4\sqrt{2}$.

In the full version of our paper [8], we show how to prove Theorem 22 assuming Proposition 24; this proof uses relatively standard techniques, though it is slightly tedious. We will subsequently prove the proposition, which contains the key ideas for our result.

4.3.2 Decoding Clusters of Small-Syndrome States

To prove Proposition 24, we begin by using the clustering property of \mathcal{C} to partition $G_X^{\epsilon'}$ and $G_Z^{\epsilon'}$ into clusters, for which we will subsequently choose representative elements that we will use to define a decoding map for states $|\psi'\rangle$ with small syndrome. Below, we first analyze the clustering of $G_Z^{\epsilon'}$; the case of $G_X^{\epsilon'}$ will be exactly analogous.

We consider clusters defined similarly as in [1]. However, to obtain our improvement over [1], we will leverage an additional linear structure in the set of clusters (Property 2 in Lemma 2 below), which ultimately allows us to use the uncertainty principle in Lemma 19.

Given $y \in G_Z^{\epsilon'}$, define a cluster $Y_Z^y \subseteq G_Z^{\epsilon'}$ by

$$Y_Z^y = \{y' \in G_Z^{\epsilon'} : |y + y'|_{C_X^\perp} \leq 2c_1\epsilon'n\}.$$

The following lemma follows directly from our definitions.

► **Lemma 25.** *The clusters Y_Z^y for $y \in G_Z^{\epsilon'}$ form a partition of $G_Z^{\epsilon'}$ satisfying the following properties:*

1. *Every pair of distinct clusters $Y_Z^y \neq Y_Z^{y'}$ satisfies $\text{dis}(Y_Z^y, Y_Z^{y'}) \geq c_2n$.*
2. *For $c \in C_Z$, then $Y_Z^{y+c} = Y_Z^y + c$, and in particular $Y_Z^{y+c} = Y_Z^y$ if and only if $c \in C_X^\perp$.*

Proof. We first show that the clusters form a partition of $G_Z^{\epsilon'}$. Fix some $y \in G_Z^{\epsilon'}$. Then for every $y' \in Y_Z^y$, it follows that every $y'' \in Y_Z^{y'}$ has $|y'' + y|_{C_X^\perp} \leq |y'' + y'|_{C_X^\perp} + |y' + y|_{C_X^\perp} \leq 4c_1\epsilon'n$. But by assumption (see the statement of Theorem 22) $2\epsilon' < \epsilon_0$ and $4c_1\epsilon' < c_2$, so because $y'' + y \in G_Z^{2\epsilon'}$, the clustering property implies that $|y'' + y|_{C_X^\perp} \leq 2c_1\epsilon'n$, so that $y'' \in Y_Z^y$. Thus we have shown that every $y' \in Y_Z^y$ has $Y_Z^{y'} \subseteq Y_Z^y$, and by the same reasoning $Y_Z^y \subseteq Y_Z^{y'}$, so $Y_Z^{y'} = Y_Z^y$. Thus every pair of clusters is either equal or disjoint, so the clusters Y_Z^y form a partition of $G_Z^{\epsilon'}$.

Now every pair of distinct clusters $Y_Z^y \neq Y_Z^{y'}$ satisfies $\text{dis}(Y_Z^y, Y_Z^{y'}) \geq c_2n$, as if this distance was $< c_2n$, the clustering property would imply that it is $\leq 2c_1\epsilon'n$, which then implies that $Y_Z^y = Y_Z^{y'}$.

It remains to show Property 2 in the lemma statement. For every $y \in G_Z^{\epsilon'}$ and $c \in C_Z$, by definition $H_Z(y+c) = H_Zy$ and thus Y_Z^{y+c} is also a cluster in $G_Z^{\epsilon'}$, which is isomorphic to Y_Z^y under the isomorphism $y' \mapsto y' + c$; that is, $Y_Z^{y+c} = Y_Z^y + c$. If $c \in C_X^\perp$, then $|y + (y+c)|_{C_X^\perp} = 0$ so that $y + c \in Y_Z^y$ and therefore $Y_Z^{y+c} = Y_Z^y$. Meanwhile, if $c \in C_Z \setminus C_X^\perp$, then $Y_Z^{y+c} \neq Y_Z^y$, as otherwise it would follow that $|y + (y+c)|_{C_X^\perp} = |c|_{C_X^\perp} \leq 2c_1\epsilon'n$. But by assumption (see Theorem 22) C has distance $d > 2c_1\epsilon'n$, so $|c|_{C_X^\perp} > 2c_1\epsilon'n$. ◀

Lemma 25 implies that Y_Z^y has distinct translates by all $c + C_X^\perp \in C_Z/C_X^\perp$, where all representatives of a given coset of C_X^\perp yield the same translate. We denote the collection of these translates for a given cluster Y_Z^y by

$$\mathcal{Y}_Z^{Y_Z^y} = \{Y_Z^{y+c} : c \in C_Z\}.$$

For each such collection $\mathcal{Y}_Z = \mathcal{Y}_Z^{Y_Z^y}$ of clusters, we fix an arbitrary representative $Y_Z(\mathcal{Y}_Z) \in \mathcal{Y}_Z$.

Now for a given syndrome $s = H_Zy \in \mathbb{F}_2^{m_Z}$ of some $y \in G_Z^{\epsilon'}$, so that $|s| \leq \epsilon'm_Z$, then the set of bit strings with syndrome s is precisely the coset $y + C_Z$. By Lemma 25, $(y + C_Z) \cap Y_Z(\mathcal{Y}_Z^{Y_Z^y})$ is a coset of C_X^\perp , and is in particular therefore nonempty. Thus we may associate to s an arbitrary representative $e_Z(s) \in (y + C_Z) \cap Y_Z(\mathcal{Y}_Z^{Y_Z^y})$.

We now let Dec_Z be a unitary acting on $n + m_Z$ qubits with the following “decoding” property: for every $y \in G_Z^{\epsilon'}$, it holds that

$$\text{Dec}_Z |y\rangle \otimes |0\rangle = |y + e_Z(H_Zy)\rangle \otimes |H_Zy\rangle.$$

We let Dec_Z^1 be the channel acting on n qubits that simply applies Dec_Z and traces out the syndrome register. Formally, for every $y \in G_Z^{\epsilon'}$, then

$$\text{Dec}_Z^1(|y\rangle \langle y|) = |y + e_Z(H_Zy)\rangle \langle y + e_Z(H_Zy)|.$$

Equivalently, Dec_Z^1 is the channel that performs a Z -syndrome measurement on its input $|y\rangle$, and then adds $e_Z(s)$ to the post-measurement state, where $s = H_Z y$ was the measurement outcome.

The channel Dec_Z^1 performs a weak form of decoding in the following sense. Recall that an ordinary decoder in the Z basis for a CSS code maps all bit strings near a given codeword $c \in C_Z$ to some element of the coset $c + C_X^\perp$. In contrast, as shown below, the key property of our decoding channel Dec_Z^1 is that it sends all bit strings in a given cluster Y_Z^y to elements of the same coset $c + C_X^\perp \in C_Z/C_X^\perp$, though this coset may be far away from the cluster Y_Z^y .

► **Lemma 26.** *For every cluster Y_Z^y and every pair of elements $y, y' \in Y_Z^y$, then $y' + e_Z(H_Z y') \in y + e_Z(H_Z y) + C_X^\perp$.*

Proof. By definition $e_Z(H_Z y) \in y + C_Z$ and $e_Z(H_Z y') \in y' + C_Z$ both lie in the cluster $Y_Z(\mathcal{Y}_Z^{Y_Z^y})$. Therefore by Lemma 25, both $e_Z(H_Z y')$ and $y' + (y + e_Z(H_Z y))$ belong to both $y' + C_Z$ and $Y_Z(\mathcal{Y}_Z^{Y_Z^y})$, and thus $e_Z(H_Z y') \in y' + (y + e_Z(H_Z y)) + C_X^\perp$, or equivalently, $y' + e_Z(H_Z y') \in y + e_Z(H_Z y) + C_X^\perp$. ◀

To conclude this section, we extend all of the clustering terminology and results above for $G_Z^{\epsilon'}$ to their analogues for $G_X^{\epsilon'}$. Specifically, we similarly obtain a partition of $G_X^{\epsilon'}$ into clusters Y_X^y for $y \in G_X^{\epsilon'}$. We again conclude that each cluster Y_X^y in $G_X^{\epsilon'}$ has a set $\mathcal{Y}_X^{Y_X^y}$ of distinct translates by all $c + C_Z^\perp \in C_X/C_Z^\perp$. We fix arbitrary representative clusters $Y_X(\mathcal{Y}_X) \in \mathcal{Y}_X$, and assign to each syndrome $s = H_X y$ for $y \in G_X^{\epsilon'}$ an element $e_X(s) \in (y + C_X) \cap Y_X(\mathcal{Y}_X^{Y_X^y})$. We then obtain an X decoding unitary Dec_X and channel Dec_X^1 , which are defined analogously to their Z analogues, except the syndrome measurement and error correction steps are performed in the X basis instead of the Z basis. Observe that Dec_X^1 and Dec_Z^1 commute, so we can define $\text{Dec}^1 = \text{Dec}_X^1 \text{Dec}_Z^1 = \text{Dec}_Z^1 \text{Dec}_X^1$.

4.3.3 Applying Decoding to Prove Well-Spreadness

We now complete the proof of Proposition 24, which as shown above in turn implies Theorem 22, by applying the uncertainty principle in Lemma 19 to the decodings of small-syndrome states for \mathbf{H} .

Proof of Proposition 24. Because \mathcal{C} has dimension $k > 0$, the space $C_Z/C_X^\perp = (C_X/C_Z^\perp)^\perp$ is nonzero, so there exist $\bar{c}_X \in C_X \setminus C_Z^\perp$, $\bar{c}_Z \in C_Z \setminus C_X^\perp$ such that $\bar{c}_X \cdot \bar{c}_Z = 1$. Fix an arbitrary pair of such elements \bar{c}_X, \bar{c}_Z , so that $\bar{X} := X^{\bar{c}_Z}$ and $\bar{Z} := Z^{\bar{c}_X}$ are anticommuting logical operators for the code \mathcal{C} .

For $b = 0, 1$, define

$$S_X^b = \{y \in G_X^{\epsilon'} : \bar{c}_Z \cdot (y + e_X(H_X y)) = b\}$$

$$S_Z^b = \{y \in G_Z^{\epsilon'} : \bar{c}_X \cdot (y + e_Z(H_Z y)) = b\}.$$

Then by Lemma 26, for a given cluster Y_Z^y in $G_Z^{\epsilon'}$, all $y' \in Y_Z^y$ have $y' + e_Z(H_Z y')$ lying in the same coset $y + e_Z(H_Z y) + C_X^\perp$, and thus all $y' \in Y_Z^y$ have the same value of $\bar{c}_X \cdot (y' + e_Z(H_Z y')) = \bar{c}_X \cdot (y + e_Z(H_Z y))$. Therefore all $y' \in Y_Z^y$ lie in the same set S_Z^b , where $b = \bar{c}_X \cdot (y + e_Z(H_Z y))$. It follows from Lemma 25 that $\text{dis}(S_Z^0, S_Z^1) \geq c_2 n = \delta n$. Analogous reasoning implies that $\text{dis}(S_X^0, S_X^1) \geq c_2 n = \delta n$.

It remains to be shown that (2) holds for every $|\psi'\rangle \in \mathcal{C}^{\epsilon'}$. By Lemma 19, either $|\text{Tr}(\bar{X} \text{Dec}^1(\psi'))| \leq 1/2 + 1/2\sqrt{2}$ or $|\text{Tr}(\bar{Z} \text{Dec}^1(\psi'))| \leq 1/2 + 1/2\sqrt{2}$. Assume the latter; the proof for the former is analogous. Now because \bar{Z} by definition commutes with

Dec_X , the distribution from measuring \bar{Z} on $\text{Dec}^1(\psi') = \text{Dec}_X^1 \text{Dec}_Z^1(\psi')$, or equivalently on $\text{Dec}_X(\text{Dec}_Z^1(\psi' \otimes |0\rangle\langle 0|)) \text{Dec}_X^\dagger$, is the same as the distribution from measuring \bar{Z} on $\text{Dec}_Z^1(\psi')$. Therefore $|\text{Tr}(\bar{Z} \text{Dec}_Z^1(\psi'))| \leq 1/2 + 1/2\sqrt{2}$. But by definition if we expand $|\psi'\rangle = \sum_{y \in G_Z^{\epsilon'}} \psi'_y |y\rangle$, then it follows that

$$\begin{aligned}
\frac{1}{2} + \frac{1}{2\sqrt{2}} &\geq \text{Tr}(\bar{Z} \text{Dec}_Z^1(\psi')) \\
&= \text{Tr}\left((\bar{Z} \otimes I) \text{Dec}_Z(|\psi'\rangle\langle\psi'| \otimes |0\rangle\langle 0|) \text{Dec}_Z^\dagger\right) \\
&= \left(\langle\psi'| \otimes \langle 0| \text{Dec}_Z^\dagger\right) \left((\bar{Z} \otimes I) \text{Dec}_Z |\psi'\rangle \otimes |0\rangle\right) \\
&= \left(\sum_{y \in G_Z^{\epsilon'}} \langle y + e_Z(H_Z y) | \otimes \langle H_Z y | (\psi'_y)^\dagger\right) \\
&\quad \cdot \left(\sum_{y \in G_Z^{\epsilon'}} \psi'_y (-1)^{\bar{e}_X \cdot (y + e_Z(H_Z y))} |y + e_Z(H_Z y)\rangle \otimes |H_Z y\rangle\right) \\
&= \sum_{y \in G_Z^{\epsilon'}} (-1)^{\bar{e}_X \cdot (y + e_Z(H_Z y))} |\psi'_y|^2 \\
&= \left| \sum_{y \in S_Z^0} |\psi'_y|^2 - \sum_{y \in S_Z^1} |\psi'_y|^2 \right| \\
&= |D_Z^{\psi'}(S_Z^0) - D_Z^{\psi'}(S_Z^1)|.
\end{aligned}$$

Then because $D_Z^{\psi'}$ is supported inside $G_Z^{\epsilon'} = S_Z^0 \sqcup S_Z^1$ by the definition of ψ' , it follows that $D_Z^{\psi'}(S_Z^0) + D_Z^{\psi'}(S_Z^1) = 1$, so we must have

$$D_Z^{\psi'}(S_Z^0), D_Z^{\psi'}(S_Z^1) \geq \frac{1}{4} - \frac{1}{4\sqrt{2}} = \mu',$$

as desired. ◀

5 Planting Codewords in QLDPC Codes

In this section, we show how to plant a nontrivial codeword in the quantum Tanner codes of [13], thereby ensuring the code has positive dimension regardless of other parameters in the instantiation. For instance, when the inner codes C_A, C_B are chosen to be of rate $1/2$ in the quantum Tanner code construction, the only prior method for bounding dimension, namely by counting parity checks, fails to ensure the dimension of the global code is positive (see Section 3). However, our planted construction of quantum Tanner codes has positive dimension regardless of the rates of the inner codes, and thus provides a new way to ensure positive dimension, that works in previously unfeasible parameter regimes. We remark that a similar technique also works for the codes of [20], though we do not present the details to avoid redundancy.

Using the strongly explicit nature of the planted codeword, we apply our construction to improve upon the explicit SoS lower bounds of [11] to obtain *strongly* explicit SoS lower bounds.

5.1 Intuition: Planted Classical Tanner Codes

In this section, we present the simpler case of how to plant a codeword in a classical Tanner code, which motivates our construction in the quantum case.

Recall that a classical Tanner code $C = \text{Tan}(\Gamma, C_{\text{in}}) \subseteq \mathbb{F}_q^n$ is constructed from a Δ -regular graph Γ with n edges and an inner code $C_{\text{in}} \subseteq \mathbb{F}_q^\Delta$ as follows. We associate the set of all edges in Γ with the set $[n]$, and we associate the set of edges incident to each vertex $v \in \Gamma$ with the set $[\Delta]$. Then we define C to be the set of all edge labelings $y \in \mathbb{F}_q^n = \mathbb{F}_q^{E(\Gamma)}$ such that the labels of edges incident to each $v \in \Gamma$ form a codeword in C_{in} .

The standard method for ensuring that the rate R of C is positive (and in fact linear in n) is to require that C_{in} be a linear code of rate $R_{\text{in}} > 1/2$, so that by counting linear constraints it follows that $R \geq 1 - 2(1 - R_{\text{in}})$.

However, if we only care about ensuring that $R > 0$, we may instead simply require that C_{in} contains the all-1s vector $\mathbf{1} \in \mathbb{F}_q^\Delta$, as then by definition the global all-1s vector $\mathbf{1} \in \mathbb{F}_q^n$ must lie in C . If the resulting “planted” classical code has no other nontrivial codewords, it is simply a repetition code, which is typically uninteresting classically.

However, below we construct a quantum analogue of such planted codes, which are more difficult to construct than their classical counterparts, and yield interesting complexity theoretic applications regardless of their rate. For instance, because the planted codeword is trivial to describe and therefore strongly explicit, we improve the explicit SoS lower bounds of [11] to be strongly explicit. Furthermore, recall that in Corollary 23 of Section 4, we showed that qLDPC codes of arbitrarily small rate also yield NLTS Hamiltonians.

5.2 Construction of Planted Quantum Tanner Codes

In this section, we present our construction of planted quantum Tanner codes. This construction can be viewed as a quantum analogue of the planted classical Tanner codes described in Section 5.1. The quantum case requires significantly more care, as described below.

The following proposition presents our paradigm for planting a nontrivial codeword in a quantum Tanner code

► **Proposition 27.** *Let C be a quantum Tanner code as defined in Section 3.1 such that the following hold:*

1. *The all-1s vector $\mathbf{1} \in \mathbb{F}_q^\Delta$ lies in C_A and in C_B^\perp .*
 2. *$n = |Q| = |G||A||B|$ is relatively prime with q .*
- Then the all-1s vector $\mathbf{1} \in \mathbb{F}_q^Q$ lies in $C_Z \setminus C_X^\perp$ and in $C_X \setminus C_Z^\perp$.*

Proof. Because $\mathbf{1} \in C_A$, the components in every given codeword of C_A^\perp sum to 0. Therefore every codeword in $C_A^\perp \otimes C_B^\perp$, and thus also in C_Z^\perp , has components summing to 0, as C_Z^\perp is by definition spanned by codewords in $C_A^\perp \otimes C_B^\perp$ supported in neighborhoods of vertices in the square Cayley complex. Thus as the components of $\mathbf{1} \in \mathbb{F}_q^Q$ sum to $n \neq 0$ in \mathbb{F}_q because n is relatively prime with q , it follows that $\mathbf{1} \notin C_Z^\perp$.

However, as $\mathbf{1} \in \mathbb{F}_q^\Delta$ lies in C_B^\perp , it follows that $\mathbf{1} \in \mathbb{F}_q^{\Delta \times \Delta}$ lies in $(C_A \otimes C_B)^\perp$, and thus $\mathbf{1} \in \mathbb{F}_q^Q$ lies in $C_X = \text{Tan}(\Gamma_0, (C_A \otimes C_B)^\perp)$.

Thus we have shown that $\mathbf{1} \in C_X \setminus C_Z^\perp$. Analogous reasoning shows that $\mathbf{1} \in C_Z \setminus C_X^\perp$. ◀

We will instantiate the construction in Proposition 27 by choosing C_A and C_B^\perp at random from the set of codes of some constant rate that contain $\mathbf{1}$. The following result, which we prove in the full version of our paper [8], shows that such random “planted” codes are still product-expanding, thereby providing a planted analogue of Corollary 11.

► **Proposition 28.** *For every fixed $\epsilon > 0$, there exists a constant $\rho = \rho(\epsilon) > 0$ and a function $\delta(n) = \delta(n; \epsilon) \rightarrow 0$ as $n \rightarrow \infty$ such that the following holds. For every pair of integers $k_1, k_2 \in (\epsilon n, (1 - \epsilon)n)$, if $C_i \subseteq \mathbb{F}_q^n$ for $i = 1, 2$ is drawn uniformly at random from the set of linear codes of dimension k_i that contain $\mathbf{1} \in \mathbb{F}_q^n$, then with probability $\geq 1 - \delta(n)$ both (C_1, C_2^\perp) and (C_1^\perp, C_2) will be ρ -product-expanding.*

Combining the results above, we immediately obtain the following strongly explicit construction of quantum Tanner codes with a planted all-1s vector.

► **Theorem 29** (Planted quantum Tanner codes). *For every finite field \mathbb{F}_q of characteristic $p \geq 7$, there exist constants $c_1, c_2 > 0$ such that there is a strongly explicit infinite family $(\mathcal{C}^{(n)})_{n \rightarrow \infty}$ of quantum LDPC CSS codes for which every $\mathcal{C}^{(n)} = \text{CSS}(C_X^{(n)}, C_Z^{(n)})$ with $C_X^{(n)}, C_Z^{(n)} \subseteq \mathbb{F}_q^n$ has the following properties:*

1. $\mathcal{C}^{(n)}$ has (c_1, c_2) -small-set boundary and coboundary expansion (and therefore has distance $\geq c_1 n$ by Lemma 15).

2. The all-1s vector $\mathbf{1} \in \mathbb{F}_q^n$ lies in $C_X^{(n)} \setminus C_Z^{(n)\perp}$ and in $C_Z^{(n)} \setminus C_X^{(n)\perp}$.

In particular, for a sufficiently large constant Δ and a sufficiently small constant $\rho > 0$, such a family $(\mathcal{C}^{(n)})_{n \rightarrow \infty}$ is given by quantum Tanner codes, where we choose $\text{Cay}(G, A) = \text{Cay}(G, B)$ from a strongly explicit family of Ramanujan graphs given by Theorem 12, and the inner codes $C_A, C_B \subseteq \mathbb{F}_q^\Delta$ are found by a brute force search to ensure that $\mathbf{1} \in C_A, C_B^\perp$ and that $(C_A, C_B), (C_A^\perp, C_B^\perp)$ are ρ -product expanding.

Proof. By Proposition 28, if $\rho > 0$ is sufficiently small and $\Delta > 0$ is sufficiently large then we can find codes $C_A, C_B \subseteq \mathbb{F}_q^\Delta$ satisfying the criteria in the theorem statement, namely that $\mathbf{1} \in C_A, C_B^\perp$ and that $(C_A, C_B), (C_A^\perp, C_B^\perp)$ are ρ -product expanding. Furthermore, because $p \geq 7$ is a prime, we can specifically choose $\Delta = p' + 1$ for some sufficiently large prime p' such that $p'^4 \not\equiv 0, 1 \pmod{p}$, and then choose $\text{Cay}(G, A) = \text{Cay}(G, B)$ from the strongly explicit family of Ramanujan graphs $(\Gamma_m)_{m \in p\mathbb{N}}$ of degree $\Delta = p' + 1$ in Theorem 12. Here we restrict to graphs Γ_m for $m \in p\mathbb{N}$ to ensure that $p'^{2m}(p'^{4m} - 1) \equiv p'^2(p'^4 - 1) \not\equiv 0 \pmod{p}$, so that $|V(\Gamma_m)| \not\equiv 0 \pmod{p}$. Furthermore $p'^4 \not\equiv 0, 1 \pmod{p}$ implies that $\Delta = p' + 1 \not\equiv 0 \pmod{p}$, so $n = |Q| = |V(\Gamma_m)|\Delta^2 \not\equiv 0 \pmod{p}$, and thus n and q are relatively prime. Thus Proposition 27 implies that $\mathbf{1} \in \mathbb{F}_q^n$ lies in $C_X^{(n)} \setminus C_Z^{(n)\perp}$ and in $C_Z^{(n)} \setminus C_X^{(n)\perp}$, while Theorem 16 implies that $\mathcal{C}^{(n)}$ has (c_1, c_2) -small-set boundary and coboundary expansion for sufficiently small constants $c_1, c_2 > 0$.

Because the Ramanujan graphs in Theorem 12 are strongly explicit, and the inner codes C_A, C_B have constant size because Δ is constant as $n \rightarrow \infty$, the parity check matrices H_X, H_Z for C_X, C_Z respectively are strongly explicit, which by definition means that \mathcal{C} is strongly explicit. ◀

In Theorem 29, we may choose C_A, C_B to have any fixed rates $0 < R_A, R_B < 1$ for sufficiently large Δ . Because \mathcal{C} has rate $R \geq -(1 - 2R_A)(1 - 2R_B)$ (see Section 3.2), it follows that we can in fact ensure that the codes in Theorem 29 have any desired constant rate $0 < R < 1$.

However, our construction alternatively allows us to obtain quantum Tanner codes of positive dimension for R_A, R_B in previously impossible parameter regimes. Because Theorem 29 ensures that $\mathbf{1} \in C_Z \setminus C_X^\perp$, it follows that \mathcal{C} always has dimension $\dim(\mathcal{C}) = \dim(C_Z) - \dim(C_X^\perp) \geq 1$, even when we choose R_A, R_B to be constants for which the bound $R \geq -(1 - 2R_A)(1 - 2R_B)$ is meaningless. For instance, we can choose $R_A = R_B$, or take both $R_A, R_B < 1/2$, in which case counting parity checks fails to show that the resulting quantum Tanner code \mathcal{C} has positive dimension. Nevertheless the planted all-1s vector ensures that even in this case \mathcal{C} must have dimension ≥ 1 .

As mentioned Section 1.3, while Theorem 29 provides planted quantum Tanner codes over the alphabet given by any field \mathbb{F}_q of characteristic $p \geq 7$, we generalize this result to arbitrary finite fields in the full version [8]. The main idea in this generalization is to replace the Cayley expanders in [18] with those in Example 3.4 of [17], as the number of vertices in the former (but not the latter) construction is always a multiple of 2, 3, and 5.

5.3 Application to Strongly Explicit Sum-of-Squares Lower Bounds

In this section, we describe how we use our planted quantum Tanner codes to obtain *strongly* explicit lower bounds against a linear number of levels of the SoS hierarchy, thereby improving upon the weakly explicit SoS lower bounds of Hopkins and Lin [11].

Hopkins and Lin [11] show that quantum LDPC codes with small-set boundary and coboundary expansion yield CSPs that are hard for a linear number of levels of the Sum-of-Squares SDP hierarchy. Specifically, the CSPs they use are instances of ℓ -LIN over \mathbb{F}_2 (or equivalently, ℓ -XOR) defined as follows. Below, we let the *locality* of a CSS code $\mathcal{C} = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$ refer to the maximum Hamming weight of any row or column of H_X or H_Z . The qLDPC codes we consider by definition have locality $\ell = O(1)$ as $n \rightarrow \infty$.

► **Definition 30** (ℓ -LIN instance from qLDPC codes [11]). *Let $\mathcal{C} = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$ be a CSS code of locality ℓ . Also fix any $\beta \in C_X \setminus C_Z^\perp$. Then define the associated ℓ -LIN instance $\mathcal{I}_{\mathcal{C},\beta}$ to have $m = m_Z$ variables $y_1, \dots, y_m \in \mathbb{F}_q$ and n linear constraints over \mathbb{F}_q given by the system of equations $H_Z^\top y = \beta$, where $y = (y_1, \dots, y_m)$.*

[11] instantiates this definition with quantum Tanner codes. Although quantum Tanner codes are strongly explicit, meaning that the matrices H_X, H_Z are strongly explicit, any ℓ -LIN instance $\mathcal{I}_{\mathcal{C},\beta}$ from these codes requires a description of some $\beta \in C_X \setminus C_Z^\perp$. Previously, the only known method for finding such a codeword was via Gaussian elimination, which runs in $\text{poly}(n)$ time, and thus only yields a (weakly) explicit construction of β and of $\mathcal{I}_{\mathcal{C},\beta}$.

In contrast, our planted quantum Tanner codes in Theorem 29 are guaranteed to have the all-1s vector $\mathbf{1} \in C_X \setminus C_Z^\perp$, which is by definition strongly explicit. As such, we immediately obtain the following.

► **Lemma 31.** *If \mathcal{C} is chosen from a family of planted quantum Tanner codes from Theorem 29 and $\beta = \mathbf{1}$, then $\mathcal{I}_{\mathcal{C},\beta}$ gives a family of strongly explicit ℓ -LIN instances for a constant $\ell = O(1)$.*

Formally, [11] obtain their SoS lower bounds by showing the following result, which they applied to quantum Tanner codes. Below, recall that an ℓ -LIN instance is μ -satisfiable if there exists an assignment of the variables satisfying $\geq \mu$ -fraction of the linear constraints. We refer to [11] and the references within for background on the SoS SDP hierarchy.

► **Theorem 32** ([11]). *Let $\mathcal{C} = \text{CSS}(C_X = \ker H_X, C_Z = \ker H_Z)$ be a quantum LDPC code of locality ℓ with (c_1, c_2) -small-set boundary and coboundary expansion. Then for every $\beta \in C_X \setminus C_Z^\perp$, the ℓ -LIN instance $\mathcal{I}_{\mathcal{C},\beta}$ with $m = m_Z$ variables and n constraints satisfies the following:*

1. *Soundness: $\mathcal{I}_{\mathcal{C},\beta}$ is at most $(1 - c_1)$ -satisfiable.*
2. *Completeness: $\mathcal{I}_{\mathcal{C},\beta}$ cannot be refuted by $c_1 c_2 m / 4\ell$ levels of the SoS hierarchy.*

Although Hopkins and Lin [11] only showed Theorem 32 for the binary alphabet \mathbb{F}_2 , their same proof extends to arbitrary fields \mathbb{F}_p for prime p . Specifically, their proof uses small-set (co)boundary expansion to establish a bound on *refutation complexity*, which was then shown to imply an SoS bound for the binary alphabet \mathbb{F}_2 by Schoenebeck [21], and for prime-sized alphabets \mathbb{F}_p by Tulsiani [22].

Thus as described above, [11] obtained (weakly) explicit, but not strongly explicit, lower bounds against $\Omega(n)$ levels of SoS by taking \mathcal{C} to be a quantum Tanner code in Theorem 32. Meanwhile, applying our planted quantum Tanner codes in Theorem 29 with Lemma 31, we immediately obtain the following corollary to Theorem 32.

► **Corollary 33** (Strongly explicit SoS lower bounds). *For every prime $p \geq 7$, the ℓ -LIN instances $\mathcal{I}_{\mathcal{C},1}$ for planted quantum Tanner codes \mathcal{C} over the alphabet \mathbb{F}_p provide a family of strongly explicit instances with satisfiability $\leq (1 - \Omega(1))$, such that no instance can be refuted by cn levels of the SoS hierarchy for a sufficiently small constant $c > 0$.*

A generalization of the above result to arbitrary primes p is presented in the full version [8]. [11] also showed a reduction that used their ℓ -XOR (i.e. ℓ -LIN over \mathbb{F}_2) SoS lower bounds to obtain 3-XOR SoS lower bounds. We suspect a similar reduction should work in our case of ℓ -LIN over \mathbb{F}_p , but for conciseness we will not spell out the details.

References

- 1 Anurag Anshu, Nikolas P. Breuckmann, and Chinmay Nirkhe. NLTS Hamiltonians from Good Quantum Codes. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023*, pages 1090–1096, New York, NY, USA, June 2023. Association for Computing Machinery. doi:10.1145/3564246.3585114.
- 2 Yotam Dikstein, Irit Dinur, Prahladh Harsha, and Noga Ron-Zewi. Locally testable codes via high-dimensional expanders. *arXiv:2005.01045 [cs]*, May 2020. arXiv:2005.01045.
- 3 Irit Dinur, Yuval Filmus, Prahladh Harsha, and Madhur Tulsiani. Explicit SoS Lower Bounds from High-Dimensional Expanders. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 38:1–38:16, Dagstuhl, Germany, 2021. Schloss Dagstuhl — Leibniz-Zentrum für Informatik. ISSN: 1868-8969. doi:10.4230/LIPIcs.ITCS.2021.38.
- 4 Irit Dinur, Min-Hsiu Hsieh, Ting-Chun Lin, and Thomas Vidick. Good Quantum LDPC Codes with Linear Time Decoders. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023*, pages 905–918, New York, NY, USA, June 2023. Association for Computing Machinery. doi:10.1145/3564246.3585101.
- 5 Irit Dinur, Siqi Liu, and Rachel Zhang. New Codes on High Dimensional Expanders, August 2023. ISSN: 1433-8092. URL: <https://eccc.weizmann.ac.il/report/2023/127/>.
- 6 Lior Eldar and Aram W. Harrow. Local Hamiltonians Whose Ground States are Hard to Approximate. *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 427–438, October 2017. arXiv: 1510.02082. doi:10.1109/FOCS.2017.46.
- 7 Noah Fleming, Pravesh Kothari, and Toniann Pitassi. Semialgebraic Proofs and Efficient Algorithm Design. *Foundations and Trends® in Theoretical Computer Science*, 14(1-2):1–221, December 2019. Publisher: Now Publishers, Inc. doi:10.1561/04000000086.
- 8 Louis Golowich and Tali Kaufman. NLTS Hamiltonians and Strongly-Explicit SoS Lower Bounds from Low-Rate Quantum LDPC Codes, November 2023. arXiv:2311.09503 [quant-ph]. doi:10.48550/arXiv.2311.09503.
- 9 Holger F. Hofmann and Shigeki Takeuchi. Violation of local uncertainty relations as a signature of entanglement. *Physical Review A*, 68(3):032103, September 2003. Publisher: American Physical Society. doi:10.1103/PhysRevA.68.032103.
- 10 Max Hopkins. Personal Communication, 2023.
- 11 Max Hopkins and Ting-Chun Lin. Explicit Lower Bounds Against Omega(n)-Rounds of Sum-of-Squares. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 662–673. IEEE Computer Society, October 2022. doi:10.1109/FOCS54457.2022.00069.
- 12 Gleb Kalachev and Pavel Panteleev. Two-sided Robustly Testable Codes, August 2023. arXiv:2206.09973 [cs, math]. doi:10.48550/arXiv.2206.09973.

- 13 Anthony Leverrier and Gilles Zémor. Quantum Tanner codes. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 872–883. IEEE Computer Society, October 2022. doi:10.1109/FOCS54457.2022.00117.
- 14 Anthony Leverrier and Gilles Zémor. Decoding Quantum Tanner Codes. *IEEE Transactions on Information Theory*, 69(8):5100–5115, August 2023. Conference Name: IEEE Transactions on Information Theory. doi:10.1109/TIT.2023.3267945.
- 15 Anthony Leverrier and Gilles Zémor. Efficient decoding up to a constant fraction of the code length for asymptotically good quantum codes. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Proceedings, pages 1216–1244. Society for Industrial and Applied Mathematics, January 2023. doi:10.1137/1.9781611977554.ch45.
- 16 A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, September 1988. doi:10.1007/BF02126799.
- 17 A. Lubotzky and B. Weiss. Groups and Expanders. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 10:95–109, 1993. URL: <http://ma.huji.ac.il/~alexlub/PAPERS/groups%20and%20expanders/groupsAndExpanders.pdf>.
- 18 M. Morgenstern. Existence and Explicit Constructions of $q + 1$ Regular Ramanujan Graphs for Every Prime Power q . *Journal of combinatorial theory. Series B*, 62(1):44–62, 1994. Place: SAN DIEGO Publisher: Elsevier Inc. doi:10.1006/jctb.1994.1054.
- 19 Chinmay Nirkhe. Making the Leap to Quantum PCPs, July 2023. URL: <https://simons.berkeley.edu/talks/chinmay-nirkhe-ibm-2023-07-12>.
- 20 Pavel Panteleev and Gleb Kalachev. Asymptotically good Quantum and locally testable classical LDPC codes. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, pages 375–388, New York, NY, USA, June 2022. Association for Computing Machinery. doi:10.1145/3519935.3520017.
- 21 Grant Schoenebeck. Linear Level Lasserre Lower Bounds for Certain k -CSPs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 593–602, October 2008. ISSN: 0272-5428. doi:10.1109/FOCS.2008.74.
- 22 Madhur Tulsiani. CSP gaps and reductions in the lasserre hierarchy. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, STOC '09, pages 303–312, New York, NY, USA, May 2009. Association for Computing Machinery. doi:10.1145/1536414.1536457.