Derandomised Tensor Product Gap Amplification for Quantum Hamiltonians

Thiago Bergamaschi
University of California, Berkeley, CA, USA

Tony Metger
ETH Zurich, Zurich, Switzerland
Simons Institute for the Theory of Computing, Berkeley, CA, USA

Thomas Vidick
Weizmann Institute of Science, Rehovot, Israel
EPFL, Lausanne, Switzerland

Tina Zhang
Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

The quantum PCP conjecture asks whether it is QMA-hard to distinguish between high- and low-energy Hamiltonians even when the gap between "high" and "low" energy is large (constant). A natural proof strategy is gap amplification: start from the fact that high- and low-energy Hamiltonians are hard to distinguish if the gap is small (inverse polynomial) [21] and *amplify* the Hamiltonians to increase the energy gap while preserving hardness. Such a gap amplification procedure is at the heart of Dinur's proof of the *classical* PCP theorem [17]. In this work, following Dinur's model, we introduce a new quantum gap amplification procedure for Hamiltonians which uses random walks on expander graphs to derandomise (subsample the terms of) the *tensor product amplification* of a Hamiltonian. Curiously, our analysis relies on a new technique inspired by quantum de Finetti theorems, which have previously been used to rule out certain approaches to the quantum PCP conjecture [11].

2012 ACM Subject Classification Theory of computation → Quantum complexity theory

Keywords and phrases quantum PCP conjecture, gap amplification, local Hamiltonians, tensor product amplification, expander random walks, quantum de Finetti theorems

Digital Object Identifier 10.4230/LIPIcs.CCC.2026.15

Related Version *Full Version*: `https://arxiv.org/abs/2510.01333` [10]

Funding *Tony Metger*: acknowledges support from SNSF Grant No. 20QU-1_225171 and NCCR SwissMAP.
Thomas Vidick: supported by AFOSR Grant No. FA9550-22-1-0391 and ERC Consolidator Grant VerNisQDevS (101086733).
Tina Zhang: was partially supported by NSF CAREER grant number 2339948.

Acknowledgements We thank Anurag Anshu for useful discussions and suggesting the DL-based construction of a streaming quantum PCP presented in Appendix A. We thank Dorit Aharonov, Niko Breuckmann, Sevag Gharibian, Louis Golowich, Anand Natarajan, Quynh The Nguyen, Nikhil Srivastava, and John Wright for helpful discussions. Part of this work was completed while the authors were visiting the Simons Institute for the Theory of Computing.

1 Introduction

The classical PCP theorem [8] is one of the landmark achievements of classical complexity theory, and the quantum PCP conjecture [2] is one of the major open problems in quantum complexity theory. The classical PCP theorem was proven for the first time using algebraic techniques, and later reproven using elementary combinatorial techniques in the celebrated 2007 paper of Dinur [17]. Both approaches have, so far, resisted quantisation.

The algebraic proof of the PCP theorem seems difficult to quantise because of its reliance on the properties of polynomial codes, which have no clear quantum analogue [2]. The combinatorial proof, which pares the approach down to essentials, has seemed more promising as a guide for quantum PCP. Very broadly speaking, Dinur's proof of the PCP theorem consists of the iterated application of two steps:

▶ **Definition 1.1** (Dinur's template for proving the PCP theorem)**.**
1. *Gap amplification, in which the promise gap of a constraint satisfaction problem family is amplified at the cost of blowing up its alphabet size,*
2. *Alphabet reduction, in which the locality of the constraint satisfaction problem family is reduced back to a constant, and the gap is shown not to shrink too much in the process.*

Applying these two steps iteratively to an NP-complete constraint satisfaction problem (CSP) with inverse polynomial promise gap yields another NP-complete CSP with constant promise gap. Thus, given a purported witness for the amplified CSP, a verifier can check that the witness satisfies the amplified CSP (with constant probability of error) by randomly selecting and checking constantly many clauses, which only requires a logarithmic number of random bits. (The number of bits that the verifier uses in this process is referred to as the *randomness complexity* of the verifier.)

Dinur's classical gap amplification procedure (step (i)) starts from a certain "naïve" gap amplification procedure, which amplifies the promise gap effectively but blows up the randomness complexity of the verifier. To control the randomness complexity of the verifier, Dinur uses expander random walks to derandomise this "naïve" procedure and achieve amplification without paying a steep cost in randomness complexity.

In this work, following Dinur's model, we introduce a new quantum gap amplification procedure for Hamiltonians which uses random walks on expander graphs to derandomise (subsample the terms of) the *tensor product amplification* of a Hamiltonian. The tensor product amplification of a Hamiltonian H (normalized such that $0 \leq H \leq I$) is simply $I - (I - H)^{\otimes t}$. It can be easily shown that tensoring H with itself in this way amplifies the promise gap, but this process blows up the number of terms in the amplified Hamiltonian exponentially in t (and consequently also blows up the amount of randomness required to sample a term at random). By contrast, our derandomised version of tensor product amplification achieves gap amplification without increasing the number of terms in the original Hamiltonian by too much. The following is a formal statement of our gap amplification theorem:

▶ **Theorem 1.2** (Gap amplification)**.** *Let H be a k–local Hamiltonian with m terms acting on n qubits, satisfying the conditions in Definition 1.6. Then, its* derandomised $2t$–fold tensor product amplification $H^{(2t)}$ *(formally defined in Definition 1.9) satisfies:*
1. *$H^{(2t)}$ is $(2t \cdot k)$–local, acts on nt qubits and has $d^{2t} \cdot m$ terms, where d is the degree of some fixed expander graph family.*
2. ***Completeness.*** *The lowest eigenvalue of $H^{(2t)}$ is bounded above by*

$$\lambda_{\min}(H^{(2t)}) \leq 2t \times \lambda_{\min}(H) \tag{1.1}$$

3. ***Soundness.*** *The eigenvalue of $H^{(2t)}$ is bounded below by*

$$\lambda_{\min}(H^{(2t)}) \geq \min\left[\Theta\left(\frac{\log t}{t}\right), \Theta\left(\sqrt{\frac{t}{\log t}} \times \lambda_{\min}(H)\right)\right], \tag{1.2}$$

where the Θ notation obscures constants dependent on the choice of expander graph family and the original Hamiltonian H.

To understand Theorem 1.2, one should consider the case that t is a constant and the completeness parameter of the original Hamiltonian family is exponentially small: in this context, Theorem 1.2 says that we can amplify the additive promise gap of a Hamiltonian family by a constant multiplicative factor, at the cost of increasing both the locality and the number of terms by a constant multiplicative factor. These parameters are comparable, except for one important difference, to those which Dinur achieves with her gap amplification procedure (see [17, Section 6]). We elaborate on this difference more thoroughly in Section 1.1.

In order to prove Theorem 1.2, we devise a new technique inspired by *quantum de Finetti theorems* [27, 28], which we believe could be of independent interest. Curiously, while quantum de Finetti theorems have previously been used to rule out approaches to the quantum PCP conjecture [11], here we use methods inspired by these statements to argue the soundness of our amplification scheme. We give more context and details about our application of de Finetti theorems in the related work Section 1.1 below, and in our technical overview (Section 1.3). We dedicate Section 1.2 to a discussion on potential applications of Theorem 1.2.

1.1 Related work

Quantum gap amplification

In 2008, Aharonov, Arad, Landau and Vazirani proposed a quantum gap amplification procedure that was also based on Dinur's gap amplification [1]. The key ingredient in their work was the *detectability lemma*, which in some sense states that "if the ground state energy of some (non-commuting) Hamiltonian H is at least E, then sequentially measuring all the terms of H on the ground state should detect at least $const \cdot E$ violations (with constant probability)". Although their original motivation lay in the pursuit of the quantum PCP (QPCP) conjecture [2], their work subsequently found many applications to quantum many-body physics and quantum information theory, including area laws [7], k-designs [13], and Gibbs samplers [20].

There are two key differences between the parameters of Dinur's classical amplification procedure and those of the quantum gap amplification procedure proposed by [1]. The first is that (unlike Theorem 1.2) the gap amplification procedure of [1] requires the Hamiltonian family it amplifies to have constant locality to start with.[1] Since [1]'s gap amplification procedure also increases the locality of the Hamiltonian which it amplifies (where the increase in locality is proportional to the increase in promise gap which it achieves, like in Theorem 1.2), this means that [1]'s gap amplification cannot be iterated by itself (without composing it with a locality reduction procedure).

We resolve this issue in this work. Because our gap amplification procedure (Theorem 1.2) can be iterated by itself (albeit only a logarithmic number of times), we can achieve applications that, as far as we know, cannot be achieved with the tools developed by [1]. For example, we are able to get a *locality-gap tradeoff theorem* for QMA-complete Hamiltonians:

▶ **Theorem 1.3** (Locality-gap tradeoff; informal). *Let $\mathcal{H}$ be a k-local Hamiltonian family that is* QMA*-complete with promise gap ϵ.*[2] *Then, given a deterministic construction of $\mathcal{H}$, there is a deterministic construction of a family of Hamiltonians $\mathcal{H}'$ such that $\mathcal{H}'$ has locality $k \cdot 1/\epsilon$ and is* QMA*-complete with some universal constant promise gap c.*

[1] Newer versions of the detectability lemma exist which do not require constant locality of the Hamiltonian to which they are applied (see e.g. [3]); however, we do not know of any detectability lemma that both has this property and is sufficiently strong to complete the proof from [1].

[2] The members of $\mathcal{H}$ also need to satisfy certain technical conditions, such as being a sum of polynomially many projective terms, which we are suppressing in this informal statement. For a formal version, see the full version [10].

In particular, Theorem 1.3 could be applied (for example) to boost the promise gap of a "weak QPCP". Recent progress [4] suggests potential avenues for achieving a so-called *weak quantum PCP*, namely, a QPCP with polylog(n) locality and $1/\text{polylog}(n)$ promise gap instead of constant locality and promise gap. Applying Theorem 1.3 to such an object would immediately yield a PCP with polylog(n) locality and *constant* promise gap.

While our gap amplification procedure can be iterated by itself, like Dinur's, there remains a second important difference between the two. Dinur's gap amplification procedure is *non-locality-increasing*, in the sense that one round of amplification blows up the *alphabet size* of the constraint satisfaction problem but not the number of variables that each constraint involves. All attempted quantisations of Dinur's gap amplification, including ours and [1], are locality-increasing, which is a significant obstacle to applying the kinds of PCP composition theorems that make alphabet reduction possible. [1] gap amplification (and ours) is in some sense a quantum version of classical derandomised *sequential repetition*, and in the classical setting derandomised *parallel repetition* seems necessary to make Dinur's template from Definition 1.1 work: see [19, 18] for a discussion of this issue. As such, producing a quantisation of Dinur's gap amplification procedure which reproduces all of its important characteristics remains an open problem.

Achieving derandomised sequential (locality-increasing) amplification is already non-trivial in the quantum setting, however, and we view one of our primary contributions as that of analysing a natural but so far unstudied locality-increasing derandomised quantum amplification procedure using techniques that have not previously been applied in this context. As far as we know, our techniques are the first to yield iterability. Moreover, the techniques which [1] used to analyse their own amplification procedure have since found widespread application outside of their original context, and we hope the introduction of a new approach for analysing low-randomness locality-increasing quantum gap amplification might be found to have similar utility. We may have particular grounds for hope because the work of [1] can in some ways be viewed as a *black-box reduction* from the quantum problem to the classical one, while our approach more directly interprets certain classical gap amplification strategies in a quantum setting, and as such may more clearly illuminate the differences between the classical and the quantum problems.

Quantum de Finetti theorems

A defining property of quantum entanglement is that it is *monogamous*. That is, a quantum system cannot be very entangled with a large number of other systems. Quantum *de Finetti theorems* attempt to quantify this intuition, and more broadly offer a versatile tool to understand quantum correlations, with applications to quantum information theory [27], cryptography [15, 28], algorithms [11, 9] and complexity theory [22, 12]. Roughly speaking, quantum de Finetti theorems capture the fact that a random k qubit marginal of any $n \gg k$ qubit state should be close to a convex combination of product states, and therefore unentangled.

Most relevant to us are the results of [11], who showed that local Hamiltonians defined on dense (degree d) constraint graphs admit product state approximations, in that there exists a product state whose energy is an additive inverse poly(d) close to the ground state energy.[3] In some sense, their results offer a concrete route to *disproving* the QPCP conjecture: if there

[3] Note that the existence of a product state approximation implies that determining the ground state energy up to said approximation error is in NP, since the description of the product state is a succinct classical witness.

existed a procedure which increased the degree of any local Hamiltonian, without decreasing its ground-state energy (or increasing locality), then deciding the ground state energy of the resulting Hamiltonian up to a constant would be in NP [11, Corollary 11].

Here, we issue two remarks. First, the conclusions of [11] become exponentially weaker if the degree and the locality of the Hamiltonian are allowed to increase simultaneously during amplification [6], which allows us to evade their no-go. Second, despite these "negative" results, de Finetti techniques will nevertheless play a central role in our approach in understanding when quantum correlations can adversarially decrease the ground state energy of our amplified Hamiltonians. A comprehensive discussion is presented in the full version [10].

1.2 Discussion and applications

In this section we discuss a number of potential applications of our gap amplification procedure.

Streaming quantum PCP

Theorem 1.2 can be applied iteratively $1/\log(n)$ times to achieve what we call a "streaming quantum PCP":

▶ **Theorem 1.4** (Streaming quantum PCP)**.** *There exists a deterministic construction of a family $\{H_n\}$ of Hamiltonians on n qubits and an explicit constant c, such that each Hamiltonian H_n is a sum of $\mathrm{poly}(n)$ many terms, each summand is an $O(n)$-fold tensor product of $O(1)$-local projections, and it is QMA-complete to decide whether the ground state energy of $\{H_n\}$ is $\leq \mathsf{negl}(n)$ or $\geq c$.*

Theorem 1.4 can be interpreted as a quantum version of what one would get in the classical setting by iterating Dinur's gap amplification procedure from Definition 1.1 *without* locality or alphabet reduction reduction. Theorem 1.4 can also be viewed as a very preliminary version of quantum PCP. Indeed, the family of Hamiltonians guaranteed by Theorem 1.4 satisfies the conditions of the quantum PCP conjecture – including the condition that there are polynomially many terms in each member Hamiltonian – except for the fact that the member Hamiltonians do not have constant locality, which is of course a substantial omission.

Nonetheless, the terms in each H_n are individually easy to verify: since each term is the $O(n)$-fold tensor product of $O(1)$-local projections, each term can be measured using a constant-depth quantum circuit (and classical post-processing), or by a quantum algorithm that reads a constant number of qubits at a time. There are also only polynomially many such terms in each H_n, which implies that the energy of H_n can be measured on a witness state by a verifier who samples a random term and measures it, and in the process uses only $O(\log n)$ many classical random bits and very limited quantum resources. (We call Theorem 1.4 a "streaming quantum PCP" for this reason.) It also implies that time evolution under H_n can be efficiently simulated (see the full version [10]). Ease of verification is one of the chief motivations for quantum PCP, and ease of simulation is one of the chief motivations for the *sparse* quantum PCP conjecture (see below). As such, Theorem 1.4 can be viewed as progress towards both objectives.

▶ Remark 1.5. There is an alternative way to achieve Theorem 1.4 which does not use our Theorem 1.2 and instead uses improvements to the techniques of [1] due to [3] to prove Theorem 1.4 using "one-shot" (non-iterative) amplification. This result, due to Anurag Anshu, is presented in Appendix A. The main drawback of this alternative approach is that it

does not allow the kind of more "fine-grained" amplification that our Theorem 1.2 provides, where the tunable parameter t simultaneously governs the gap amplification and the locality increase of the input Hamiltonian.

Locality-gap tradeoffs for local Hamiltonians

As mentioned above in Theorem 1.3, Theorem 1.2 can be iteratively applied in order to produce a "locality-gap tradeoff" for local Hamiltonians. More explicitly, Theorem 1.2 implies that, if there is a deterministic construction of a QMA-complete family of Hamiltonians with locality ℓ and promise gap ϵ, then (under mild conditions) there is a deterministic construction of a QMA-complete family of Hamiltonians with locality ℓ/ϵ and gap which is some universal constant c.

In particular, we could use Theorem 1.2 to boost a "weak quantum PCP" statement (cf. [4, Open Problems]) – that is, a family of Hamiltonians with $\operatorname{poly}\log n$ locality and inverse $\operatorname{poly}\log n$ promise gap – into a family of Hamiltonians that has $\operatorname{poly}\log n$ locality and *constant* promise gap.

Quantum games PCP

The quantum games PCP conjecture states that there exists an MIP^* protocol (i.e., a protocol between a classical verifier and multiple efficient, non-communicating but potentially entangled provers) with $\operatorname{poly}\log n$ sized questions and answers, constant completeness-soundness gap, and efficient provers that can decide all of QMA [24]. The motivation for this question lies in the connection between PCPs and succinct MIP (classical multi-prover) protocols with constant gap in the classical world [2]. The connection between quantum PCP and succinct MIP^* protocols is less immediate, although, like in the classical case, quantum games PCP is a necessary consequence of Hamiltonian quantum PCP (but this is considerably harder to show than it is classically). A proof of the quantum games PCP conjecture was claimed by [25], but it has since been established that the proof had errors, and as of now the conjecture remains open.

We believe Theorem 1.4 implies a deterministic construction of an MIP^* protocol with efficient provers, constant completeness-soundness gap and $\log n$ sized *questions* ($\operatorname{poly} n$ sized answers) that decides all of QMA. Such a protocol was not known prior to our work, and it would follow by designing a succinct Pauli braiding test [16, 23] that supports measurements of the terms in our amplified Hamiltonian, which could be made into tensor products of 5-local Clifford projections [14]. Unfortunately, our result does not appear to help with achieving succinct answers, and we leave as an open question whether it is possible to build on these results to prove the quantum games PCP conjecture.

Sparse quantum PCP

The sparse quantum PCP conjecture is similar to the original quantum PCP conjecture except that, instead of requiring that the Hamiltonian is a sum of local terms, we require that the Hamiltonian is expressible as a sparse matrix with efficiently computable entries.

This conjecture is of interest primarily for two reasons:

1. It is, like the NLTS theorem [5], a necessary consequence of quantum PCP, and
2. Sparse Hamiltonians are simulable given only oracle access to their entries, which allows their energy to be measured even in the absence of any term decomposition.

This entails that their ground state energies are easy to verify, and as such a sparse quantum PCP can be viewed as a meaningful stepping stone towards the kind of easy verification that full quantum PCP requires.

One might hope that Theorem 1.4 could be useful for sparse QPCP, because our Hamiltonian is a sum of polynomially many relatively structured terms (see below, Definition 1.9). Unfortunately, existing black-box approaches to matrix sparsification appear to be either too randomness-expensive or too computationally inefficient to yield the sparse QPCP conjecture when applied to the terms of our Hamiltonians. We leave it as an interesting open question whether progress can be made in this direction by exploiting the particular tensor-product structure of our Hamiltonians.

1.3 Technical overview

1.3.1 Construction and notation

In this subsection we more formally present our construction of a quantum gap amplification procedure. Our starting point is a Hamiltonian H on n qubits, which can be partitioned into a convex combination of g commuting layers (or colors).

▶ **Definition 1.6** (Layered Hamiltonians). *H is said to be a* layered Hamiltonian *if it admits a decomposition*

$$H = \sum_{\chi\in[g]} w_\chi \cdot H_\chi \,, \qquad \text{where each layer} \qquad H_\chi = \mathop{\mathbb{E}}_{i\in[m_\chi]} \Pi_i^\chi \tag{1.3}$$

is the expectation over m_χ commuting projections, and the weights $\{w_\chi\}_{\chi\in[g]}$ are positive and $\sum_\chi w_\chi = 1$. To quantify imbalance between the layers, we introduce the parameter

$$\omega_{\min} \equiv (\min_\chi w_\chi)^{-1}. \tag{1.4}$$

As we discuss in the full version [10], there are QMA-complete local Hamiltonians which admit such a decomposition with a constant number g of layers (and constant $\omega_{\min}$), so starting from such a decomposition is essentially without loss of generality. Our construction will amplify the individual commuting layers separately. To do so, for each color we impose an expander graph structure on the clauses indexed by $[m_\chi]$.[4]

▶ **Definition 1.7** (Paths of length t in G_χ). *Let G_χ be a d-regular λ-spectral expander graph on m_χ vertices (Definition 2.8). Define a function family $\mathcal{F}_\chi : \{f : [t] \to [m_\chi]\}$ such that there is a one-to-one*[5] *correspondence between member functions $f \in \mathcal{F}_\chi$ and paths of length t in G_χ.*

The amplified Hamiltonian will act on t (tensor) copies of the original n qubit system. For each color χ and path $f \in \mathcal{F}_\chi$, we associate to f a clause Π_f^χ in the amplified Hamiltonian, in the form of a projection onto the intersection of the t clauses on the path. In other words, the clauses in the amplified Hamiltonian are satisfied if all (the "AND") of the clauses on the path f are satisfied:

$$\Pi_f^\chi \equiv \mathbb{I} - \bigotimes_{j\in[t]} \left(\mathbb{I} - \Pi_{f(j)}^\chi\right) \tag{1.5}$$

[4] So long as m_χ is above some constant m_0, there exists a determinstic construction of such a graph, see Lemma 2.9 [26]. The case $m_\chi \le m_0$ is addressed by trivially repeating the clauses in H_χ.

[5] $\forall f \in \mathcal{F}_\chi$, the tuple $(f(1), \ldots, f(t)) \in [m_\chi]^t$ is a path of length t in G_χ, and every such path is represented by some $f \in \mathcal{F}_\chi$.

▶ Remark 1.8. It is easy to check this operator is a projection: it is positive and squares to itself.

The t-fold amplified version of H, which we denote by $H^{(t)}$, can then be expressed as follows:

▶ **Definition 1.9** (Derandomised Tensor Product Amplification)**.** *Let H be layered as in Definition 1.6. We define the t-fold amplification $H^{(t)}$ of H with respect to $\{\mathcal{F}_\chi\}_{\chi\in[g]}$ as*

$$H^{(t)} = \sum_{\chi\in[g]} w_\chi \cdot H_\chi^{(t)}, \quad \textit{where} \quad H_\chi^{(t)} \equiv \mathop{\mathbb{E}}_{f\in\mathcal{F}_\chi} \left(\mathbb{I} - \bigotimes_{j\in[t]} \left(\mathbb{I} - \Pi_{f(j)}^\chi\right)\right). \tag{1.6}$$

If the original H was a k-local Hamiltonian on n qubits and m clauses, then $H^{(t)}$ is $(k\cdot t)$-local Hamiltonian on $n\cdot t$ qubits and $d^t\cdot m$ clauses.

▶ Remark 1.10. Our construction of the amplified Hamiltonian is similar to that of [1], but we use tensor products instead of matrix products.

1.3.2 Main techniques

Here we highlight the main techniques behind the proof of Theorem 1.2.

Background: tensor product amplification

The starting point for our gap amplification procedure is tensor product amplification. The t-fold tensor product amplification of a Hamiltonian H is simply

$$\mathbb{I} - (\mathbb{I} - H)^{\otimes t}. \tag{1.7}$$

If H has a decomposition as a sum of projections, $H = \mathbb{E}_{i\in[m]}\,\Pi_i$, then the tensor product amplification of H can be expanded as

$$\mathop{\mathbb{E}}_{(i_1,\ldots,i_t)\in[m]^t} \left(\mathbb{I} - \bigotimes_{j\in[t]} (\mathbb{I} - \Pi_{i_j})\right). \tag{1.8}$$

If H had m terms, therefore, the t-fold tensor product amplification of H has m^t terms.

Analysing tensor product amplification is straightforward: one simply has to observe that, since the operator $\mathbb{I} - (\mathbb{I} - H)^{\otimes t}$ is diagonal in the basis constituted by tensor products of eigenvectors of H, the ground energy of $\mathbb{I} - (\mathbb{I} - H)^{\otimes t}$ is attained by a tensor product across t registers of ground states of H. If we start with a Hamiltonian family $\{H_n\}_n$ with promise gap of size γ (and completeness sufficiently close to 1), then the gap increases at least to $t\gamma$ after t-fold tensor product amplification.

We would, however, like to achieve a similar amount of gap amplification without increasing the number of terms so steeply. In particular, instead of m^t terms, we would like the t-fold amplification of H to have $c(t)\cdot m$ terms, where $c(t)$ is a function that depends arbitrarily on t but not on m. Then we can set t to be a constant and iterate gap amplification $O(\log n)$ times to get a new Hamiltonian family $\{H'_n\}_n$ such that H'_n only has a multiplicative factor $\mathrm{poly}(n)$ more terms than H_n, but the promise gap of $\{H'_n\}_n$ is also a $\mathrm{poly}(n)$ factor larger than that of $\{H_n\}_n$.

▶ Remark 1.11. If we are content to show QMA-completeness of streaming quantum PCPs under *randomised reductions*, one way that we could reduce the number of terms in Equation (1.8) is to pick some to keep at random and delete the others. More specifically, we could

prove an analogue of our Theorem 1.4 under randomised reductions by randomly subsampling terms of Equation (1.8) using the matrix Chernoff bound: see e.g. [24]. The gap amplification statement that we prove, Theorem 1.2, can be viewed as the "correct" derandomisation of matrix-Chernoff-based subsampling of tensor product amplification. Theorem 1.2 allows us to show Theorem 1.4 under deterministic reductions.

One advantage of derandomisation in view of applications is that Theorem 1.2 is *iterable*, whereas the matrix-Chernoff-based approach is not. That is, Theorem 1.2 holds even when t is a constant, but matrix-Chernoff can only be used to do "one-shot" amplification (to amplify in one go up to constant gap), since it requires the Hamiltonian family being subsampled to have a constant promise gap.

Background: classical sequential repetition and its derandomisation

The classical analogue of tensor product amplification is *sequential repetition*, in which, given a constraint satsifaction problem (CSP) C with m clauses $c_1, \ldots, c_m$, one constructs a new CSP C' whose clauses are the ANDs of clauses from C: that is, every clause in C' is of the form $c_{i_1} \wedge \cdots \wedge c_{i_t}$ for $(i_1, \ldots, i_t) \in [m]^t$. If C is satisfiable, than C' is also satisfiable; and, if γ fraction of clauses in C are unsatisfied by the most satisfying possible assignment to C (the "unsatisfiability" of C is at least γ), then this fraction will be at least $t\gamma$ for C'.

In the classical case, we can also consider *derandomised* sequential repetition, which achieves a similar amplification guarantee but using far fewer clauses. We can achieve derandomised sequential repetition by first constructing C', the full sequential repetition of C, and then *subsampling* the clauses of C' using random walks on expander graphs. More specifically,

▶ **Definition 1.12** (Derandomised sequential repetition)**.** *Given a (classical) CSP C on n bits and m clauses, its t-fold derandomised sequential repetition is a CSP $C^{(t)}$ on $t \cdot n$ bits where the new clauses are the AND of all the clauses that lie along any length t path in some expander graph defined over $[m]$.*

So that this exposition is self-contained, we will sketch how derandomised sequential repetition can be analysed using a particular property of random walks on expander graphs [30, Chapter 4] [17, Section 6]. The only property of random walks which we need is a "set-avoiding" property: For any fixed sets $A, B \subseteq [m]$ of no more than $\delta \cdot m$ in size, the probability that a random walk that starts in A ends up in B after s steps is bounded above by $\delta + \mu^s$, where μ is the second largest eigenvalue of the expander graph. This property follows in a straightforward way from Lemma 2.10.

We sketch how this property would be useful for a classical analysis of derandomised sequential repetition. Recall that $C^{(t)}$ is the derandomised sequential repetition of C, as defined in Definition 1.12. We define a quantity called the *violation number* which will be useful in the analysis:

▶ **Definition 1.13** (Violation number)**.** *The violation number $X_{x,C^{(t)}}$ is a random variable that can be sampled by picking a uniformly random "path clause" $p = c_1 \wedge \cdots \wedge c_t$ from $C^{(t)}$ and evaluating how many out of the t clauses in the AND is violated by x.*

The violation number is a non-negative integer, and $\Pr_p[X_{x,C^{(t)}} > 0]$ precisely captures the fraction of clauses of $C^{(t)}$ which are violated by x. Recall the Second Moment Method, which for any random variable X relates $\Pr[X > 0]$ to the mean and variance of X:

▶ **Fact 1.14** (Second Moment Method). *For any non-negative random variable $X \geq 0$,*

$$\mathbb{P}[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \tag{1.9}$$

As such, if we can get a *lower bound* on the expectation of $X_{x,C^{(t)}}$ and an *upper bound* on the expectation of its square, we will obtain a lower bound on $\Pr[X_{x,C^{(t)}} > 0]$ and by extension a lower bound on the unsatisfiability of $C^{(t)}$.

The expectation of $X_{x,C^{(t)}}$ is easy to control using linearity of expectation. We now sketch how to control the expectation of the square using the "set-avoiding" property of random walks. For any classical assignment $x \in \{0,1\}^{nt}$ to $C^{(t)}$, x can be grouped into blocks $x_1, \cdots, x_t \in \{0,1\}^n$ – which in some sense are assignments to the original C – and one can define a set of clauses $\mathsf{Viol}_{x,i} \subset [m]$ which contain the clauses of C violated by x_i. Then we see that the expectation of the square of the violation number is precisely

$$\sum_{i,j \in [t]} \Pr_{p \leftarrow C^{(t)}}[c_i \in \mathsf{Viol}_{x,i} \text{ and } c_j \in \mathsf{Viol}_{x,j}] \tag{1.10}$$

$$= \sum_{i,j \in [t]} \Pr_{p \leftarrow C^{(t)}}[c_i \in \mathsf{Viol}_{x,i}] \cdot \Pr_{p \leftarrow C^{(t)}}[c_j \in \mathsf{Viol}_{x,j} \mid c_i \in \mathsf{Viol}_{x,i}] \tag{1.11}$$

where the probability is over the sampling of the random "path clause" $p \coloneqq c_1 \wedge \cdots \wedge c_t$. The probability $\Pr_{p \leftarrow C^{(t)}}[c_j \in \mathsf{Viol}_{x,j} \mid c_i \in \mathsf{Viol}_{x,i}]$ is exactly the probability that a random walk which starts in $\mathsf{Viol}_{x,i}$ ends up in $\mathsf{Viol}_{x,j}$ after $|j-i|$ steps, which is the type of quantity controlled by the "set avoiding" property.

▶ Remark 1.15. Since our proof strategy revolves around the second moment of $X_{x,C^{(t)}}$, derandomised sequential repetition would be even easier to analyse if the "path clauses" $p \leftarrow C^{(t)}$ were sampled *pairwise independently* instead of using a random walk. Pairwise independent sampling is too randomness-expensive to yield the parameters we want, so we cannot use it in Definition 1.12; however, this observation is the motivation for the proof strategy that we adopt (see "Technique 1: commuting layers" below).

Difficulties in quantisation

Although this outline is sufficient for classical CSPs, unfortunately, we encounter an obstacle when we try to quantise the argument above. We can, analogously with Definition 1.12, define the *derandomised tensor product amplification* of a Hamiltonian H and call it $H^{(t)}$: this is the motivation for Definition 1.9, in which we define our construction of the amplification of a Hamiltonian. However, $H^{(t)}$ may be non-commuting, so there need not be a valid definition of the violation sets $\mathsf{Viol}_{x,i}$ which we relied on in the classical analysis. Moreover, it may have exclusively entangled ground states, so measuring a clause could collapse the state and affect the probabilities that other clauses are violated.

In this overview we will not attempt to give a complete sketch of our analysis: the reader who wants to understand our entire proof can go directly to the technical sections, which contain further exposition. Here we will focus instead on describing at a high level the two main techniques that we use to overcome the difficulty that the ground states of $H^{(t)}$ may be entangled. Both these techniques rely in a sense on reducing the analysis of an entangled state to that of a convex combination of product states, albeit in different ways.

Technique 1: commuting layers

If all the terms in the original Hamiltonian H *commute*, then we can reduce our analysis to the classical case: we simply imagine measuring the potentially entangled ground state of $H^{(t)}$ in a complete basis that diagonalises all the terms in $H^{\otimes t}$ simultaneously. This collapses the state to a mixture over product states, which we can deal with using the classical argument and convexity. Of course, it is not true that all the terms in H commute, but we can conduct some parts of our analysis by partitioning the terms in H into at most constantly many commuting *layers*, analysing each layer separately, and recombining the layers afterwards (up to some amount of loss).[6]

We use this commuting layers technique only in order to prove that random walk subsampling of terms can be approximately by *pairwise independent* subsampling in our setting. We then develop a new technique – our "miser's de Finetti" technique, which we introduce shortly – in order to prove that pairwise independent sampling is similar enough to fully independent sampling for us even on entangled states. Fully independent sampling is equivalent to (non-derandomised) tensor product amplification, so this statement allows us to get good bounds on the amplification guarantees of our procedure. This last step – i.e. relating the pairwise independent case to the fully independent case – is arguably the hardest part of our analysis.

▶ Remark 1.16. The authors of [1] encounter similar issues, and also resolve them by partitioning the original Hamiltonian into commuting layers. However, they carry through their *entire* analysis using the commuting-layer approach to deal with the entanglement problem, while we use both the commuting-layer approach and also our "miser's de Finetti" technique. Their approach gives rise to constants in the amplification bounds that depend on the locality of terms in H, and in particular become unmanageable when the locality of H is larger than constant; ours, perhaps surprisingly, does not. This is the chief reason we are able to iterate our procedure.

Technique 2: miser's de Finetti

Quantum information theory gives us a class of tools for reducing entangled states to convex combinations of product states, in the form of the *quantum de Finetti theorems* [27, 28]. There are many types of quantum de Finetti theorem, but they all say something like the following: if we take a *symmetric* (permutation-invariant) pure quantum state ρ on t registers and trace out all but $k \ll t$ registers, the mixed state left over on those registers is "close" to a convex combination of product states.

Naïvely, we might try to use a de Finetti theorem to solve our problem as follows. Recall that each term Π_{amp} in the amplified Hamiltonian $H^{(t)}$ (Definition 1.9) is a tensor product $\Pi_{\text{amp}} = (\mathbb{1} - \Pi_1) \otimes \cdots \otimes (\mathbb{1} - \Pi_t)$, where $\Pi_1, \ldots, \Pi_t$ are terms in H, and the constituent terms of Π_{amp} as well as the order in which they appear are determined by a walk. Instead of placing the t terms in Π_{amp} in the order that the walk dictates, we could pick a random permutation π and set $\Pi_{\text{amp}} = (\mathbb{1} - \Pi_{\pi(1)}) \otimes \cdots \otimes (\mathbb{1} - \Pi_{\pi(t)})$ instead – or, if we want our reduction to remain deterministic, we could put all $t!$ possible orderings as terms. This is equivalent to randomly permuting the registers of the state τ to which we apply the terms, so this manoeuvre allows us to assume that τ is permutation symmetric. We trace out all but

[6] Note that, since our amplification procedure (see Definition 1.9) preserves the number of commuting layers, we can iterate our procedure as many times as we like as long as we start with an H that has constantly many layers. We show that such Hamiltonians are QMA-complete in the full version [10].

the last $k \ll t$ registers of τ, leaving us with a state which is (mostly) a convex combination of (approximately) product states. We can then apply the classical argument to these k remaining registers.

Unfortunately, this naïve plan does not work for quantitative reasons. The best possible de Finetti theorem requires $k = \Omega(\log d)$, where d is the dimension of the Hilbert space associated with a single register. This would mean that we get no amplification unless we set $t \geq \text{poly}(n)$, which makes random walk sampling untenably expensive. This would also mean that, if we want a deterministic reduction, we need to have order $n!$ terms in the amplified Hamiltonian. Either of these facts is a sufficient obstacle to achieving an amplified Hamiltonian with a polynomial number of terms.

As such, we cannot hope to rely on a quantum de Finetti theorem as a black box in our setting. However, perhaps surprisingly, we are able to make progress with a technique that is inspired by a particular style of proof for a de Finetti theorem. A strategy for proving a de Finetti theorem goes as follows [31]. The keystone of the strategy is the following fact, which can be made quantitative and then proven using elementary techniques:

▷ Claim 1.17. If ρ is any permutation symmetric state on t registers, and we *project* the last $t - k$ registers onto the state $|\psi\rangle^{\otimes(t-k)}$ for some pure single-register state $|\psi\rangle$, then the first k registers of the residual state must be "close" to the pure state $|\psi\rangle^{\otimes k}$, because the state was permutation symmetric.

It then turns out that, for any permutation symmetric state ρ on t registers, we can write ρ with the last $t - k$ registers traced out as a convex combination of states of the form $(\text{id}_{t-k} \otimes \langle\psi|^{\otimes k})\rho(\text{id}_{t-k} \otimes |\psi\rangle^{\otimes k})$. Combining these two deceptively simple observations yields the de Finetti theorem: if we take a symmetric state ρ on t registers and trace out the final $t - k$ registers, the mixed state left over on the first k registers is "close" to a convex combination of product states $|\psi\rangle^{\otimes k}$.

This argument is at the heart of [31], which gives an astonishingly simple proof of the so-called "exponential de Finetti theorem". However, in order to make Claim 1.17 true quantitatively, t has to be quite large. Intuitively, we can understand this as follows: arguing that the first k registers of a permutation-symmetric state ρ must look similar to $|\psi\rangle^{\otimes k}$ if the last $t - k$ registers are in the state $|\psi\rangle^{\otimes(t-k)}$ is rather like collecting *measurement statistics* about ρ from the last $t - k$ registers and then arguing that, since the state is permutation symmetric, the measurement statistics are representative of the first k registers as well. If t is small, the statistics which were collected from the measured registers are unlikely to be informative about the remaining k registers. In particular, requiring the remaining k registers to be in a state close to $|\psi\rangle^{\otimes k}$ is a stringent condition – we are characterising the leftover state very finely if we find that this condition is true – and one would therefore expect to look at a large number of registers t before one could conclude such a thing with any reasonable probability.

In our situation, we don't actually *need* such a fine characterisation of the state. Indeed, we don't care if the state τ is actually product or not: much weaker guarantees suffice for us. As such, we can potentially get away with collecting far fewer "measurement statistics" about τ than we would have to collect if we wanted to know that τ was product. For example, it is sufficient for us that τ "looks product" with respect to measurements of H, the original Hamiltonian (because the quantity we are trying to bound involves only the trace of H on τ in various registers). As such, we could measure τ in an eigenbasis that diagonalises $H^{\otimes t}$, and then use a *classical* de Finetti theorem on the measurement outcomes. It turns out that this approach also requires an unviably large t (since classical de Finetti theorems also scale unfavourably with the alphabet size of the random variables) – but we can do even better

by realising that we don't need to know the precise eigenvalues of the eigenstates we get in each register after the eigenbasis measurement. Indeed, a very coarse approximation ("is the eigenvalue big or small?") will suffice. This coarse-graining, which allows us to reduce the alphabet size of the measurement outcomes to which we apply de Finetti–inspired techniques, is the chief reason that we can reduce t to something manageable.

This is the idea behind the auxiliary measurement we introduce in Section 5. The binary projective measurements $\{\Pi^{<\alpha}, \mathbb{I} - \Pi^{<\alpha}\}$ (Definition 5.1) act on the same Hilbert space as H, and $\Pi^{<\alpha}$ simply projects onto the energy eigenspaces of H with lower energy than α: therefore, $\Pi^{<\alpha}$ commutes with H. Measuring $\Pi^{<\alpha}$ on a random set of $t/2$ registers allows us to estimate the "energy" E of τ (the estimate we use is the number of $\geq \alpha$ outcomes). We then measure the complementary $t/2$ registers, and use Chernoff-bound-like tools to argue that they are likely to land in eigenspaces with similar energy E. After that, our proof proceeds via a careful case analysis, which hinges on whether E is "high" or "low". In effect, our strategy is to partition the Hilbert space in which τ lives into (constantly many) subspaces, each of which has predictable behaviour with respect to H; and we are able to proceed with the analysis after we project ("pinch") τ into one of these subspaces because the "pinching" measurement commutes with the measurements of H, which are the only measurements that we care about.

1.4 Organization

We begin in section Section 2 by presenting basic definitions of local Hamiltonians and expander graphs, as well as some basic probability facts.

We proceed in Section 3 by formally introducing the violation number measurement, an observable which quantifies the number of violated clauses on a random length t path. In Section 4 we argue the completeness of the amplification scheme (Theorem 1.2, Completeness), by applying the first moment method to the amplified Hamiltonian.

In Section 5 we present the auxiliary energy measurement, a central technical tool which captures the de Finetti ingredient of our proof. In Section 6, we present the proof of soundness of our amplification scheme (Theorem 1.2, Soundness), wherein we apply the second moment method to the violation number measurement.

2 Preliminaries

We dedicate this section to basic facts on the local Hamiltonian problem, probability theory, and expander graphs.

2.1 The Local Hamiltonian problem

The central problem of study in Hamiltonian complexity theory is computing the ground state energy of a local Hamiltonian.

▶ **Definition 2.1** (k-$\mathsf{LH}[\mathsf{a},\mathsf{b}]$)**.** *Let $H = \sum_i^m h_i$ be a k-local Hamiltonian on n qubits and $m = \mathrm{poly}(n)$ terms, where each h_i is a hermitian operator acting on k qubits and is described using $\mathrm{poly}(n)$ bits. Let $0 \leq H \leq \mathbb{I}$.*

Further, let $a < b$ be two real parameters described using $\mathrm{poly}(n)$ bits; $b - a$ is referred to as the "promise gap" of H. The k-local Hamiltonian problem k-$\mathsf{LH}[\mathsf{a},\mathsf{b}]$ then consists of the task of deciding whether the ground state energy $\lambda_{min}(H) \leq a$ or $\lambda_{min}(H) \geq b$, promised that H satisfies one of the two cases.

[21] proved that the local Hamiltonian problem is QMA-Complete.

▶ **Theorem 2.2** ([21]). *The* 5-LH$[2^{-\mathrm{poly(n)}}, 1/\mathrm{poly(n)}]$ *problem is* QMA*-complete.*

The Quantum PCP Conjecture stipulates that the local Hamiltonian problem remains just as hard under a constant promise gap.

▶ **Conjecture 2.3** (The Quantum PCP Conjecture [2]). *There exists a constant locality k as well as $a, b \in [0,1]$ satisfying $b - a = \Omega(1)$ such that the k-*LH$[a, b]$ *problem is* QMA*-complete.*

2.2 Probability

Here we recollect a series of simple inequalities.

▶ **Fact 2.4** (Markov's Inequality). *For any non-negative random variable $X \geq 0$ and real parameter $a > 1$,*

$$\mathbb{P}[X > a \cdot \mathbb{E}[X]] \leq \frac{1}{a} \tag{2.1}$$

▶ **Fact 2.5** (First Moment Method). *For any non-negative random variable $X \geq 0$,*

$$\mathbb{P}[X > 0] \leq \mathbb{E}[X] \tag{2.2}$$

▶ **Fact 2.6** (Second Moment Method). *For any non-negative random variable $X \geq 0$,*

$$\mathbb{P}[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \tag{2.3}$$

▶ **Fact 2.7** ([29], Lemma 6). *Let $Z_1, \cdots Z_{n+k}$ be a collection of binary random variables. Let $S \subset [n+k], |S| = k$ be a uniformly random subset. Then,*

$$\Pr\left[\sum_{i\in S} Z_i \leq k \cdot \delta \text{ and } \sum_{i\in \bar{S}} Z_i \geq n(\delta + \nu)\right] \leq \exp\left[-\frac{2\nu^2 nk^2}{(n+k)(k+1)}\right] \tag{2.4}$$

2.3 Expander graphs

We rely on a series of basic facts of expander graphs.

▶ **Definition 2.8** (Spectral Expansion). *Let $G = (V, E)$ be an undirected d-regular graph, and let A be its adjacency matrix. Let $d = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be its eigenvalues. Then G is said to be a λ-spectral-expander if*

$$\max_{i\neq 1} |\lambda_i| \leq \lambda. \tag{2.5}$$

We often-times will refer to the transition matrix $P = A/d$ of the random walk on G, and let $\mu = \lambda/d$. For our reduction we require an explicit deterministic construction.

▶ **Lemma 2.9** ([26]). *There exists explicit constants $m_0, d \in \mathbb{N}$ and a parameter $d > \lambda > 0$, such that there exists a deterministic polynomial-time constructable family $\{G_m\}_{m\geq m_0}$ of d-regular λ-spectral-expander graphs on m vertices for every $m \geq m_0$.*

See e.g. [17, Lemma 2.1] on how to achieve the condition $\forall m \geq m_0$, from the infinite family of [26].

The constraints on λ entail $\mu < 1$. We will require a short lemma on quadratic forms of the transition matrix P of an expander graph.

▶ **Lemma 2.10.** *Let $P = A/d$ be the transition matrix of a d-regular $(d \cdot \mu)$-spectral-expander graph on m vertices. Then, for all vectors $a, b \in [0,1]^{\times m}$ and integer t:*

$$a^T P^t b \leq \frac{1}{m}\|a\|_1 \cdot \|b\|_1 + \mu^t\big(\|a\|_1 + \|b\|_1\big) \tag{2.6}$$

Proof. Let us diagonalize $P = \sum_k \mu_k \gamma_k \gamma_k^T$, where $\mu_1 = 1$, $|\mu_{k>1}| \leq \mu$, and the eigenvectors γ_k form an orthonormal basis.

$$a^T P^t b = \frac{1}{m}\|a\|_1 \cdot \|b\|_1 + \sum_k \mu_k^t (a \cdot \gamma_k)(b \cdot \gamma_k) \tag{2.7}$$

$$\leq \frac{1}{m}\|a\|_1 \cdot \|b\|_1 + \mu^t \sum_k |a \cdot \gamma_k||b \cdot \gamma_k| \tag{2.8}$$

$$\leq \frac{1}{m}\|a\|_1 \cdot \|b\|_1 + \mu^t \left(\sum_k |a \cdot \gamma_k|^2\right)^{1/2} \left(\sum_k |b \cdot \gamma_k|^2\right)^{1/2} \tag{2.9}$$

$$\leq \frac{1}{m}\|a\|_1 \cdot \|b\|_1 + \mu^t \|a\|_2 \cdot \|b\|_2 \tag{2.10}$$

Where, in sequence, we used $|\mu_k| \leq \mu$, the Cauchy-Schwartz inequality, and the orthonormality of the γ_k. Now,

$$\|a\|_2 \cdot \|b\|_2 \leq \|a\|_2^2 + \|b\|_2^2 \leq \|a\|_1 + \|b\|_1 \tag{2.11}$$

by the AM-GM inequality and since each entry of a, b is ≤ 1. ◀

3 The Main Quantity of Interest: The Violation Number Measurement

We recollect that we consider families of layered Hamiltonians as in Definition 1.6. Following Definition 1.9, to define the amplified Hamiltonian, we amplify the layers separately:

$$H = \sum_{\chi \in [g]} w_\chi H_\chi \overset{\text{amplify}}{\longrightarrow} H^{(2t)} := \sum_{\chi \in [g]} w_\chi H_\chi^{(2t)} \tag{3.1}$$

Let the number of projective terms in H_χ be m_χ.

For any given color χ and function $f \in \mathcal{F}_\chi$ (the paths of length $2t$, Definition 1.7), it is useful to associate a measurement N_f^χ. N_f^χ counts the number of projections of color χ that are violated when the projectors specified by $f \in \mathcal{F}_\chi$ are measured, i.e. the projector $\Pi_{f(j)}$ is measured on the jth register for all $j \in [2t]$.

▶ **Definition 3.1** (Violation Number Measurement)**.** *Fix a color $\chi \in [g]$. Given a state ρ on $2t \cdot n$ qubits and a function $f : [2t] \to [m_\chi]$:*

1. *Divide the $2t \cdot n$ qubits into $2t$ blocks of n qubits, and measure each block as follows: On the jth block, perform the measurement $\{\mathbb{I} - \Pi_{f(j)}^\chi, \Pi_{f(j)}^\chi\}$ with outcome $b_j \in \{0,1\}$.*
2. *Output $\sum_j b_j$.*

The observable associated with this measurement is denoted

$$N_f^\chi \equiv \sum_{j \in [2t]} (\Pi_{f(j)}^\chi)_{\text{reg}[j]} \otimes \mathbb{I}_{[2t]\setminus j} \ . \tag{3.2}$$

The notation $(\Pi_{f(j)})_{\text{reg}[j]}$ means $\Pi_{f(j)}$ acting on register j. Intuitively, N_f^χ chooses $2t$ projectors from the Hamiltonian H_χ according to the given function f and then measures these projectors in parallel to count how many of them are violated on the state ρ.

▶ Remark 3.2. N_f^χ is a Hermitian operator with eigenvalues $\{0, 1, \ldots, 2t\}$ whose eigenspaces are the projectors onto the different outcomes of the measurement procedure above. However, as written Equation (3.2) is not the eigendecomposition of N_f^χ.

We also define another collection of measurements, this one indexed by χ (color), f (function) and also $S \subseteq [2t]$ (subset). Measuring $N_{f,S}^\chi$ will correspond to measuring the projectors specified by f only in the registers whose indices are inside S, and counting how many violations result.

▶ **Definition 3.3** (Subset Violation Number). *Fix a color $\chi \in [g]$. Given a state ρ on $2t \cdot n$ qubits, a function $f : [2t] \to [m_\chi]$ and a subset $S \subseteq [2t]$:*

1. *Divide the $2t \cdot n$ qubits into $2t$ blocks of n qubits, and for $j \in S$ measure $\{\mathbb{I} - \Pi_{f(j)}^\chi, \Pi_{f(j)}^\chi\}$ on the jth block with outcome $b_j \in \{0, 1\}$.*
2. *Output $\sum_{j \in S} b_j$.*

We denote the associated observable as $N_{f,S}^\chi$ in the natural way. We will be interested in the probability that measuring N_f^χ (or $N_{f,S}^\chi$) on some state ρ does *not* yield outcome 0. We will write this probability as

$$\Pr_\rho\left[N_f^\chi > 0\right].$$

The projector onto the 0-eigenspace of N_f^χ is $\bigotimes_{j \in [2t]} (\mathbb{1} - \Pi_{f(j)}^\chi)$, so we have

$$\Pr_\rho\left[N_f^\chi > 0\right] = 1 - \mathrm{Tr}\left[\bigotimes_{j \in [2t]} (\mathbb{I} - \Pi_{f(j)}^\chi)\rho\right]. \tag{3.3}$$

The following lemmas give the relationship between the energy of $H_\chi^{(2t)}$ and the observables N_f^χ and $N_{f,S}^\chi$.

▶ **Lemma 3.4.** *Consider $H_\chi^{(2t)}$ and N_f^χ as defined above. Then, for any state ρ:*

$$\mathrm{Tr}\left[H_\chi^{(2t)}\rho\right] = \mathop{\mathbb{E}}_{f \in \mathcal{F}_\chi} \Pr_\rho\left[N_f^\chi > 0\right]. \tag{3.4}$$

Proof. This follows by comparing the definition of $H_\chi^{(2t)}$ and Equation (3.3). ◀

▶ **Lemma 3.5.** *Consider $H_\chi^{(2t)}$, N_f^χ and $N_{f,S}^\chi$ as defined above. Then, for any function f, state ρ, and subset S,*

$$\mathrm{Tr}\left[N_f^\chi \rho\right] \geq \mathrm{Tr}\left[N_{f,S}^\chi \rho\right], \quad \Pr_\rho\left[N_f^\chi > 0\right] \geq \Pr_\rho\left[N_{f,S}^\chi > 0\right] \tag{3.5}$$

Proof. The bits b_j that are summed in the measurement procedure for $N_{f,S}^\chi$ are a subset of the bits summed in the measurement procedure for N_f^χ. ◀

4 On the Completeness of the Amplified Hamiltonian

In this section, we present an upper bound on the ground state energy of the amplified Hamiltonian. The upper bound says that if the ground state energy of the original Hamiltonian was very low, then the ground state energy of the amplified Hamiltonian remains fairly low.

▶ **Proposition 4.1.** *Fix a layered Hamiltonian H as in Definition 1.6, and the collection of function families $\{\mathcal{F}_\chi\}$ from Definition 1.7. For any $t \in \mathbb{N}$, write $H^{(2t)} = \sum_\chi w_\chi H_\chi^{(2t)}$ to denote the derandomized $2t$-fold tensor product amplification of H (Definition 1.9). The lowest eigenvalue of $H^{(2t)}$ is bounded from above by*

$$\lambda_{\min}(H^{(2t)}) \leq 2t \cdot \lambda_{\min}(H)\,.$$

Proof. Let ρ be the ground state of the original Hamiltonian $H = \sum_\chi w_\chi H_\chi$. Let us consider the Hamiltonian terms of each color one at a time. Note that, since we are proving an upper bound on the lowest eigenvalue of the amplified Hamiltonian, it suffices to show that the value is upper bounded on some specific state. Consider the state $\rho^{\otimes 2t}$. From Lemma 3.4,

$$\mathrm{Tr}[H_\chi^{(2t)}\rho^{\otimes 2t}] = \mathbb{E}_{f\in\mathcal{F}_\chi} \Pr_{\rho^{\otimes 2t}}\left[N_f^\chi > 0\right] \tag{4.1}$$

From the first moment method (Fact 2.5), we have

$$\Pr_{\rho^{\otimes 2t}}\left[N_f^\chi > 0\right] \leq \mathrm{Tr}\left[N_f^\chi \rho^{\otimes 2t}\right] \tag{4.2}$$

$$\Rightarrow \quad \mathbb{E}_{f\in\mathcal{F}_\chi} \Pr_{\rho^{\otimes 2t}}\left[N_f^\chi > 0\right] \leq \sum_{i\in[2t]} \mathrm{Tr}\left[(H_\chi)_{\mathrm{reg}[i]}\rho^{\otimes 2t}\right] = 2t \cdot \mathrm{Tr}[H_\chi\rho] \tag{4.3}$$

By linearity, we conclude

$$\lambda_{min}(H^{(2t)}) \leq 2t \cdot \sum_\chi w_\chi \mathrm{Tr}[H_\chi\rho] = 2t \cdot \lambda_{min}(H) \tag{4.4}$$

◀

5 A Technical Tool: The Auxiliary Energy Measurement

To simplify notation, for $j \in [2t]$, let us denote by $H_{\mathrm{reg}[j]}$ the application of H on the j-th register.

$$H_{\mathrm{reg}[j]} = \underbrace{\mathbb{I}\otimes\cdots\otimes\mathbb{I}}_{j-1} \otimes H \otimes \underbrace{\mathbb{I}\otimes\cdots\otimes\mathbb{I}}_{2t-j} \tag{5.1}$$

In this section, we introduce a measurement which we refer to as the "auxiliary energy measurement". This measurement is never performed in a real measurement of the amplified Hamiltonian $H^{(2t)}$; it exists purely in the analysis. Loosely speaking, we think of this measurement as a *diagnostic*: we select some random subset of registers on which to perform it, and its outcomes on those "auxiliary" registers will help us determine how to proceed in the analysis of the quantity we care about, namely $N_{f,S}^\chi$, which is defined on the remaining (non-auxiliary) registers. The fact that we select the subset of registers randomly will mean that the prover cannot adversarially bias the outcome of the auxiliary measurement in order to mislead us about the non-auxiliary registers.

▶ **Definition 5.1** (The Auxiliary Energy Measurement)**.** *Given any state ρ on $n \cdot 2t$ qubits, a subset $S \subseteq [2t]$, and a threshold parameter $\alpha \in \mathbb{R}$:*

1. *For $j \in [2t]$ define $\Pi_{\mathrm{reg}[j]}^{\geq\alpha}$ as the projection onto the direct sum of all eigenspaces of $H_{\mathrm{reg}[j]}$ with associated eigenvalue larger than α.*
2. *For all $j \in \overline{S}$, measure $\{\mathbb{I} - \Pi_{\mathrm{reg}[j]}^{\geq\alpha}, \Pi_{\mathrm{reg}[j]}^{\geq\alpha}\}$ and call the outcome $c_j \in \{0,1\}$.*
3. *Output $c = (c_j)_{j\in\overline{S}}$.*

We will write the probability of receiving an outcome c in this measurement as $\Pr_\rho[c]$, where it will be clear from context that we are referring to the auxiliary energy measurement and what the set S and threshold α are. We will denote by $\rho_{|S,\alpha,c}$ the post-measurement state after receiving outcome c in the auxiliary measurement with set S and threshold α. Note that in the measurement procedure, we measure all registers *not* in S, i.e. $\rho_{|S,\alpha,c}$ is a post-measurement state whose registers in S have been left untouched.

Our analysis will hinge on a careful case division, where we classify the possible outcomes of the auxiliary energy measurement into two categories, "high-energy" and "low-energy", and decide what to do based on whether we got a high– or low-energy outcome. For this purpose, let $r \in [t]$ be an integer parameter (to be picked soon). For any subset $S \subseteq [2t]$, we define

$$U_{\bar{S}}^r \equiv \left\{ c \in \{0,1\}^{\bar{S}} : |c| \geq 4r \right\}^7 \tag{5.2}$$

The following lemma captures the intuition that our auxiliary measurement is a good diagnostic because we pick the subset on which to perform it randomly. Suppose we were to extend the auxiliary energy measurement to **all** the registers, i.e., suppose we were to perform it in S as well as in $\bar{S}$. We expect that, since S is picked randomly, it is likely that, when we condition on receiving a high energy outcome in $\bar{S}$, we will also receive a high energy outcome in S with decent probability.

▶ **Lemma 5.2.** *For each $j \in [2t]$, let the binary random variable C_j denote the outcome of the measurement $\{\mathbb{1} - \Pi_{\mathrm{reg}[j]}^{\geq\alpha}, \Pi_{\mathrm{reg}[j]}^{\geq\alpha}\}$ on any $2t \times n$ qubit state ρ. Then, for every $r \in [t]$,*

$$\mathbb{E}_{|S|=t} \Pr_\rho \Big[\sum_{j \in S} C_j < 2r \;\; \textit{and} \;\; \sum_{i \in \overline{S}} C_i \geq 4r \Big] \leq e^{-\frac{2r^2}{t}} \tag{5.3}$$

Proof. Here we remark that the $\{C_j\}$ correspond to the outcomes of a sequence of commuting measurements, so their joint distribution is well-defined. The proof then follows immediately from Fact 2.7. ◀

As we later discuss, this is the central "de Finetti" statement we require in our approach.

▶ Remark 5.3. So long as one picks the same set S in both, the subset violation number measurement and the auxiliary energy measurement commute. We have therefore the following decomposition for any state ρ, color χ, threshold α, subset S and function f:

$$\Pr_\rho \Big[N_{f,S}^\chi > 0 \Big] = \sum_{c \in \{0,1\}^{\overline{S}}} \Pr_\rho[c] \Pr_{\rho_{|S,\alpha,c}} \Big[N_{f,S}^\chi > 0 \Big] \, . \tag{5.4}$$

Our final lower bound on the ground energy of the amplified Hamiltonian, introduced in Section 6, is based solely on the behavior of $N_{f,S}^\chi$, which we manage to control using the auxiliary measurements. We are "sacrificing" the auxiliary registers, in the sense that we do not count violations on them except as a diagnostic tool. We can do this because the outcome of $N_{f,S}^\chi$ always lower bounds that of N_f^χ (Lemma 3.5).

---

[7] We also denote $\overline{U}_{\bar{S}}^r \subset \{0,1\}^{\bar{S}}$ as the complement set. Whenever implicit, the superscript r is omitted for conciseness.

6 On the Soundness of the Amplified Hamiltonian

In this section, we prove a lower bound on the ground energy of the amplified Hamiltonian: we show that one cycle of amplification boosts the ground energy of the original Hamiltonian by at least a constant multiplicative factor which depends on t, i.e. the number of "copies" of the original Hamiltonian. Together with Proposition 4.1, this shows that our amplification procedure amplifies the promise gap of any Hamiltonian with sufficiently good completeness.

The key challenge is that the ground state of the amplified Hamiltonian may not be a product state, and therefore may be entangled in a way that causes its energy to be lower than what a purely classical analysis of a similar gap amplification procedure, such as that of [17], would lead us to expect. Nevertheless, by combining ingredients from [17]'s proof of classical gap amplification with techniques inspired by the proof of the exponential de Finetti theorem [27, 28, 31], we still are able to show amplification of the ground state energy:

▶ **Theorem 6.1** (Simplified). *Fix a layered Hamiltonian H as in Definition 1.6, and the collection of function families $\{\mathcal{F}_\chi\}$ from Definition 1.7. For any $t \in \mathbb{N}$, write $H^{(2t)} = \sum_\chi w_\chi H_\chi^{(2t)}$ to denote the derandomized $2t$-fold tensor product amplification of H (Definition 1.9). The lowest eigenvalue of $H^{(2t)}$ is bounded from below by*

$$\lambda_{\min}(H^{(2t)}) \geq \min\left\{\Theta\Big(\frac{\log t}{t}\Big)\ ,\ \Theta\Big(\lambda_{\min}(H)\cdot\frac{t^{1/2}}{(\log t)^{1/2}}\Big)\right\},$$

where the big-O hides only constants that do not depend on t.

Due to space constraints, the proof of the theorem is available only in the arXiv version [10].

References

1 Dorit Aharonov, Itai Arad, Zeph Landau, and Umesh Vazirani. The detectability lemma and quantum gap amplification. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, pages 417–426, New York, NY, USA, 2009. Association for Computing Machinery. doi:10.1145/1536414.1536472.

2 Dorit Aharonov, Itai Arad, and Thomas Vidick. The quantum PCP conjecture, 2013. arXiv:1309.7495.

3 Anurag Anshu, Itai Arad, and Thomas Vidick. Simple proof of the detectability lemma and spectral gap amplification. *Physical Review B*, 93(20), May 2016. doi:10.1103/physrevb.93.205142.

4 Anurag Anshu, N. P. Breuckmann, and Quynh T. Nguyen. Circuit-to-hamiltonian from tensor networks and fault tolerance. *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, 2023. URL: https://api.semanticscholar.org/CorpusID:262935180.

5 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 '23, pages 1090–1096. ACM, June 2023. doi:10.1145/3564246.3585114.

6 Anurag Anshu and Chinmay Nirkhe. Limitations on brandão-harrow limitation for 4-local hamiltonians, 2022. URL: https://anuraganshu.seas.harvard.edu/sites/g/files/omnuum6156/files/anshu/files/bh_4local.pdf.

7 Itai Arad, Zeph Landau, and Umesh Vazirani. Improved one-dimensional area law for frustration-free systems. *Phys. Rev. B*, 85:195145, May 2012. doi:10.1103/PhysRevB.85.195145.

8 S. Arora, C. Lund, R. Motwani, M. Sudan, and Mario Szegedy. Proof verification and hardness of approximation problems. In *Proceedings., 33rd Annual Symposium on Foundations of Computer Science*, pages 14–23, 1992. doi:10.1109/SFCS.1992.267823.

9 Thiago Bergamaschi. Improved product-state approximation algorithms for quantum local hamiltonians. In *International Colloquium on Automata, Languages and Programming*, 2023. doi:10.4230/LIPIcs.ICALP.2023.20.

10 Thiago Bergamaschi, Tony Metger, Thomas Vidick, and Tina Zhang. Derandomised tensor product gap amplification for quantum hamiltonians. *arXiv preprint arXiv:2510.01333*, 2025. doi:10.48550/arXiv.2510.01333.

11 Fernando G. S. L. Brandão and Aram Wettroth Harrow. Product-state approximations to quantum ground states. In *Symposium on the Theory of Computing*, 2013. URL: https://api.semanticscholar.org/CorpusID:9421795.

12 Fernando GSL Brandao and Aram W Harrow. Quantum de finetti theorems under local measurements with applications. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 861–870, 2013.

13 Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346(2):397–434, August 2016. doi:10.1007/s00220-016-2706-8.

14 Anne Broadbent, Zhengfeng Ji, Fang Song, and John Watrous. Zero-knowledge proof systems for qma. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 31–40. IEEE, 2016. doi:10.1109/FOCS.2016.13.

15 Matthias Christandl, Robert König, and Renato Renner. Postselection technique for quantum channels with applications to quantum cryptography. *Physical review letters*, 102 2:020504, 2008. URL: https://api.semanticscholar.org/CorpusID:10062534.

16 Mikael de la Salle. Spectral gap and stability for groups and non-local games, 2022. arXiv:2204.07084.

17 Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12–es, June 2007. doi:10.1145/1236457.1236459.

18 Irit Dinur and Or Meir. Derandomized parallel repetition via structured pcps, 2014. arXiv:1002.1606.

19 Irit Dinur and Omer Reingold. Assignment testers: Towards a combinatorial proof of the pcp theorem. *SIAM Journal on Computing*, 36(4):975–1024, 2006. doi:10.1137/S0097539705446962.

20 Michael J. Kastoryano and Fernando G. S. L. Brandao. Quantum gibbs samplers: the commuting case, 2016. arXiv:1409.3435.

21 Alexei Y. Kitaev, A. H. Shen, and Mikhail N. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate studies in mathematics*. American Mathematical Society, 2002. URL: https://bookstore.ams.org/gsm-47/.

22 Cécilia Lancien and Andreas Winter. Flexible constrained de finetti reductions and applications. *Journal of Mathematical Physics*, 58(9), September 2017. doi:10.1063/1.5003633.

23 Tony Metger, Anand Natarajan, and Tina Zhang. Succinct arguments for qma from standard assumptions via compiled nonlocal games, 2024. doi:10.48550/arXiv.2404.19754.

24 Anand Natarajan and Chinmay Nirkhe. The status of the quantum PCP conjecture (games version), 2024. doi:10.48550/arXiv.2403.13084.

25 Anand Natarajan and Thomas Vidick. Retracted: Two-player entangled games are np-hard. In *33rd Computational Complexity Conference, CCC 2018, San Diego, CA, USA, June 22-24, 2018*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.CCC.2018.20.

26 Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 3–13, 2000. URL: https://api.semanticscholar.org/CorpusID:420651.

27 Renato Renner. Symmetry of large physical systems implies independence of subsystems. *Nature Physics*, 3(9):645–649, 2007.

28 Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.

29 Marco Tomamichel and Anthony Leverrier. A largely self-contained and complete security proof for quantum key distribution. *Quantum*, 1:14, 2017. doi:10.22331/Q-2017-07-14-14.

30 Salil P. Vadhan. Pseudorandomness. *Foundations and TrendsÂ® in Theoretical Computer Science*, 7:1–336, 2012. doi:10.1561/0400000010.

31 Thomas Vidick and Henry Yuen. A simple proof of Renner's exponential de Finetti theorem. *arXiv preprint*, 2016. arXiv:1608.04814.

A Amplification from the Detectability Lemma

In this section we present an alternative "one-shot" construction of a non-local Hamiltonian with constant gap, based on the detectability lemma and its converse.[8] Roughly speaking, for any integer t, we will define an amplification scheme which maps local Hamiltonians on n qubits and m clauses of (sufficiently small) inverse-polynomial promise gap γ, to a new Hamiltonian on $t \cdot n$ qubits and $O(1)$ clauses, with promise gap $\Theta(\gamma \cdot m \cdot t)$. While these results are conceptually similar to our derandomization scheme in Theorem 6.1, the new Hamiltonian will be $O(n \cdot t)$ local (i.e. fully global, see Theorem A.3 below). We refer the reader back to Section 1.2 for a discussion.

We begin with a brief background on the detectability lemma and its converse.

▶ **Lemma A.1** (Detectability Lemma, e.g. [3, Lemma 2]). *Let $H = \mathbb{E}_{i\in[m]}\Pi_i$, where each Π_i is a projection that commutes with all except at most d of the other projections Π_j. Then, if $\lambda_{\mathsf{min}}(H) \geq \gamma$, we have for every state $|\psi\rangle$:*

$$\Big\| \prod_{i\in[m]} (\mathbb{I} - \Pi_i)|\psi\rangle \Big\|^2 \leq (1 + m\gamma/d^2)^{-1} \,. \tag{A.1}$$

▶ **Lemma A.2** (The Converse to the DL, e.g. [3, Lemma 4]). *Let $A_1, \cdots A_g$ be projections. Then, if $|\psi\rangle$ is a state satisfying for some $\epsilon \in [0,1]$,*

$$\Big\| \prod_{\chi\in[g]} A_i|\psi\rangle \Big\|^2 \leq 1 - \epsilon \,, \tag{A.2}$$

then there exists a $\chi \in [g]$ such that $\|A_\chi|\psi\rangle\|^2 \leq 1 - \frac{\epsilon}{4g}$.

For simplicity of presentation, in this section we fix our attention to local Hamiltonians which are expectations over projections $H = \frac{1}{m}\sum_i \Pi_i$; where further we assume each projection does not commute with at most $(g-1)$ other projections. This implies that the local terms of H can be efficiently (and equitably) partitioned into g commuting layers/colors (Definition 1.6). We express the terms in each layer $\chi \in [g]$ as:

$$H^\chi = \mathbb{E}_{i\in[m_\chi]}\Pi_i^\chi \tag{A.3}$$

and associate a "weight" $w_\chi = m_\chi/m = \Omega(g^{-1})$ to be the number of clauses labeled the color χ. For any integer t, we consider the amplified Hamiltonian defined by tensor products of the DL operators applied to each layer:

$$H^{(t)} = \mathbb{I} - \mathbb{E}_\chi \bigotimes_j^t \Big(\prod_{i\in[m_\chi]} (\mathbb{I} - \Pi_i^\chi) \Big) \,, \tag{A.4}$$

[8] We thank Anurag Anshu for pointing us to this construction.

where as before, $\mathbb{E}_\chi$ indicates a sample χ from the distribution $(w_1, w_2, \cdots w_g)$ over $[g]$. The amplified Hamiltonian is Hermitian because the projections appearing in the product $\prod_{i\in[m_\chi]}(\mathbb{I} - \Pi_i^\chi)$ all commute, since they belong to the same layer.

▶ **Theorem A.3** (Amplification from the DL). *Assume that the minimum eigenvalue of H is $\lambda_{\mathsf{min}}(H) \geq \gamma$. Then, the minimum eigenvalue of the amplified Hamiltonian $H^{(t)}$ from Equation A.4 satisfies*

$$\lambda_{\mathsf{min}}(H^{(t)}) \geq \min\left(\Theta\big(g^{-2}\big), \quad \Omega\left(\frac{t\cdot m\cdot\gamma}{g^4}\right)\right). \tag{A.5}$$

Proof. From Lemma A.1, we have that for any state $|\psi\rangle \in \mathbb{C}^{2^n}$ on a single copy of the system,

$$\left\| \prod_{\chi\in[g]} \prod_{i\in[m_\chi]} \left(\mathbb{I} - \Pi_i^\chi\right)|\psi\rangle \right\|^2 \leq (1 + m\gamma/g^2)^{-1}\,.$$

This can be amplified by taking tensor products since the operator norm is sub-multiplicative. That is, for any $|\psi\rangle \in (\mathbb{C}^{2^n})^{\otimes t}$ on t copies of the original system:

$$\left\| \bigotimes_{j=1}^{t} \prod_{\chi\in[g]} \prod_{i\in[m_\chi]} \left(\mathbb{I} - \Pi_i^\chi\right)_{\mathsf{reg}[j]} |\psi\rangle \right\|^2 \leq (1 + m\gamma/g^2)^{-t}\,.$$

Next, we observe that the order of the tensor product can be exchanged with that of the product over colors. As a consequence, we can introduce the operators $A_\chi = \otimes_j^t \prod_{i\in[m_\chi]} \left(\mathbb{I} - \Pi_i^\chi\right)_{\mathsf{reg}[j]}$, and re-express the operator above as

$$\bigotimes_{j=1}^{t} \prod_{\chi\in[g]} \prod_{i\in[m_\chi]} \left(\mathbb{I} - \Pi_i^\chi\right)_{\mathsf{reg}[j]} = \prod_{\chi\in[g]} A_\chi\,.$$

Each A_χ is a product of *commuting* projections; hence, A_χ itself is also a projection. We can therefore apply Lemma A.2, which implies that there must exist one layer labeled by index χ with

$$\left\| \bigotimes_{j}^{t} \prod_{i\in[m_\chi]} \left(\mathbb{I} - \Pi_i^\chi\right)_{\mathsf{reg}[j]} |\psi\rangle \right\|^2 \leq 1 - \frac{1}{4g}\left(1 - \frac{1}{(1 + m\gamma/g^2)^t}\right).$$

To conclude, we obtain that the energy of $|\psi\rangle$ under the amplified Hamiltonian is at least

$$\begin{aligned}\langle\psi|H^{(t)}|\psi\rangle &= 1 - \mathbb{E}_\chi\langle\psi|\bigotimes_{j}^{t} \prod_{i\in[m_\chi]} (\mathbb{I} - \Pi_i^\chi)|\psi\rangle \\ &\geq \frac{\min_\chi w_\chi}{4g}\cdot\left(1 - \frac{1}{(1 + m\gamma/g^2)^t}\right).\end{aligned}$$

The claimed result then follows by considering the regimes of $m\gamma t/g^2 \leq 0.1$ and ≥ 0.1 separately. ◀