Cloning is as Hard as Learning for Stabilizer States

Nikhil Bansal NIKHIL.BANSAL@WARWICK.AC.UK
Department of Computer Science, University of Warwick, Coventry, UK

Matthias C. Caro MATTHIAS.CARO@WARWICK.AC.UK
Department of Computer Science, University of Warwick, Coventry, UK

Gaurav Mahajan GAURAV.MAHAJAN@YALE.EDU
Department of Computer Science, Yale University, Connecticut, USA
Department of Mathematics and Statistics, Toulouse School of Economics, Toulouse, France

Editors: Steve Hanneke and Tor Lattimore

Abstract

The impossibility of simultaneously cloning non-orthogonal states lies at the foundations of quantum theory. Even when allowing for approximation errors, cloning an arbitrary unknown pure state requires as many initial copies as needed to fully learn the state. Rather than arbitrary unknown states, modern quantum learning theory often considers structured classes of states and exploits such structure to develop learning algorithms that outperform general-state tomography. This raises the question: How do the sample complexities of learning and cloning relate for such structured classes?

We answer this question for an important class of states. Namely, for n-qubit stabilizer states, we show that the optimal sample complexity of cloning is $\Theta(n)$. Thus, also for this structured class of states, cloning is as hard as learning. To prove this result, we use representation-theoretic tools in the recently proposed Abelian State Hidden Subgroup framework and a new structured version of the recently introduced random purification channel to relate stabilizer state cloning to a variant of the sample amplification problem for probability distributions that was recently introduced in classical learning theory. This allows us to obtain our cloning lower bounds by proving new sample amplification lower bounds for classes of distributions with an underlying linear structure. Our results provide a more fine-grained perspective on No-Cloning theorems, opening up connections from foundations to quantum learning theory and quantum cryptography.

Keywords: quantum cloning, quantum learning, stabilizer states, sample amplification

1. Introduction

The *No-Cloning* theorem (Wootters and Zurek, 1982; Dieks, 1982) is a foundational result in quantum theory. It states: There is no *one-fits-all* cloning unitary U such that $U(|\psi\rangle |0\rangle) = |\psi\rangle |\psi\rangle$ holds simultaneously for non-orthogonal pure states $|\psi\rangle$. We also have what one may call an *approximate No-Cloning* theorem (Werner, 1998; Keyl and Werner, 1999): Approximately cloning arbitrary pure states requires as many copies as pure state tomography (Gross et al., 2010; Baldwin et al., 2016; Haah et al., 2016; O'Donnell and Wright, 2016; Pelecanos et al., 2025b). While this tells us cloning general pure states is exponentially costly, imposing sufficiently stringent structure can make cloning easy. As a trivial example, any class of orthogonal states can be cloned (and learned) perfectly from a single copy. Thus, there are two extremes: Approximate cloning and learning are exponentially costly without structure, but even exact cloning and learning become trivial with orthogonality structure.

In many areas of quantum computation and information, and in particular in quantum learning theory (Arunachalam and de Wolf, 2017; Anshu and Arunachalam, 2024), we often consider neither of these extremes. Instead, we focus on classes of states that are structured yet sufficiently rich to exhibit non-trivial quantum behavior. Prominent examples include stabilizer states (Montanaro, 2017), tensor network states (Landon-Cardinal et al., 2010; Cramer et al., 2010; Landau et al., 2015; Arad et al., 2017; Bakshi et al., 2025), quantum example states (Bshouty and Jackson, 1995; Arunachalam and de Wolf, 2017; Caro et al., 2024, 2026), and phase states (Arunachalam et al., 2023). However, whereas questions of learning such structured classes of states are well-studied, questions of (approximate) cloning under structural assumptions remain unexplored. This leads us to our first central question:

Can approximate cloning be easier than learning for structured classes of states?

A computational version of this question was raised in Fefferman et al. (2026), and there is evidence (Nehoran and Zhandry, 2024; Bostanci et al., 2025; Çakan et al., 2026) towards a positive answer. We, however, interpret the question in terms of sample complexity: Are there structured classes of states for which the sample complexity of approximate cloning is strictly smaller than that of learning?

As classical information can be copied, "classical cloning" does not provide any insights useful for the quantum case. However, classical learning theorists have recently introduced the task of *sample amplification* (Axelrod et al., 2020, 2024): Given a dataset of random samples drawn from an unknown distribution from some known class, produce a larger dataset that looks like it consists of true samples. Notably, Axelrod et al. (2020) proved that sample amplification can be strictly easier than learning. When thinking of states loosely as a quantum counterpart of classical probability distributions, we see that the task of sample amplification is conceptually similar to approximate quantum cloning. We posit that, via this analogy, new results in sample amplification can lead to novel insights into quantum cloning.

To this end, we connect sample amplification with a prominent area in computational learning theory, the learnability of Boolean functions classes. Here, given i.i.d. samples of the form $(x, f(x))$, with the input x drawn from some distribution $\mathcal{V}$, and with the unknown function f promised to lie in some known class $\mathcal{F}$, we aim to learn f. If $\mathcal{V}$ is known, this can be viewed as learning the distribution $(\mathcal{V}, f)$ from samples. Hence, from the perspective of computational learning theory, the following question arises naturally:

Can sample amplification for structured classes of distributions $\{(\mathcal{V}, f)\}_{f\in\mathcal{F}}$ be easier than learning?

In this work, we study the above two questions for specific kind of structures. First, we exhibit function classes for which sample amplification is as hard as learning in terms of sample complexity. Concretely, we show that amplifying parities is as hard as learning them, and we provide a coding-theoretic interpretation of this result. Second, we draw on our classical no-go result for amplification to show a no-go for structured quantum cloning: Stabilizer states are no easier to approximately clone than to learn in terms of sample complexity. This serves as an approximate No-Cloning Theorem for stabilizer states, the to our knowledge first No-Cloning Theorems for a practically relevant class of structured quantum states.

1.1. Framework

Our first contribution is to propose structured variants of sample amplification and quantum cloning.

Structured sample amplification. Informally, sample amplification as introduced by Axelrod et al. (2020, 2024) is the following task: Let $\mathcal{D}$ be a known class of distributions. Given t samples drawn i.i.d. from an unknown distribution $D \in \mathcal{D}$, "amplify" them to $t+m$ samples that are indistinguishable from $t+m$ true i.i.d samples drawn from D. Here, it is natural to formalise indistinguishability in terms of small total variation (TV) distance. Thus, a distribution class $\mathcal{D}$ over domain $\mathcal{X}$ admits a $(t, t+m, \epsilon)$ sample amplification scheme if there exists a (randomised) map $T_{\mathcal{D},t,m,\epsilon} : \mathcal{X}^t \to \mathcal{X}^{t+m}$ such that,

$$\sup_{D\in\mathcal{D}} \mathrm{d_{TV}}\left(D^{\otimes t} \circ T^{-1}_{\mathcal{D},t,m,\epsilon}, D^{\otimes t+m}\right) \leq \epsilon\,. \tag{1}$$

As already highlighted in Axelrod et al. (2020), sample amplification can be viewed as a game between an amplifier and a distinguisher. The amplifier tries to ampify t samples from an unknown distribution D to $t+m$ samples. The distinguisher, who knows the distribution D but not the specific samples available to the amplifier, receives a set of $t+m$ samples and aims to distinguish whether it is true i.i.d. data or an output produced by the amplifier. The notion of distance originates as the advantage over random guessing that the distinguisher can achieve. That is, equivalently to the above, a distribution class $\mathcal{D}$ admits a $(t, t+m, \epsilon)$ sample amplification procedure if there exists a (randomised) map $T_{\mathcal{D},t,m,\epsilon} : \mathcal{X}^t \to \mathcal{X}^{t+m}$ such that,

$$\sup_{D\in\mathcal{D}} \sup_{\mathcal{A}} \left| \mathop{\mathbf{Pr}}_{X^{t+m}\leftarrow D^{\otimes t+m}} \left[\mathcal{A}(X^{t+m}) = 1\right] - \mathop{\mathbf{Pr}}_{X^{t}\leftarrow D^{\otimes t}} \left[\mathcal{A}(T_{\mathcal{D},t,m,\epsilon}(X^t)) = 1\right] \right| \leq \epsilon, \tag{2}$$

where the second supremum is over all possible distinguishers.

Axelrod et al. (2020, 2024) studied sample amplification for different classes of discrete and continuous distributions. For discrete distributions over a bounded support of size k, they gave a $(t, t+\Theta(t\epsilon/\sqrt{k}), \epsilon)$ sample amplification scheme whenever $t = \Omega(\sqrt{k}/\epsilon)$. This in particular implies that we can amplify by one sample to constant accuracy already from $\mathcal{O}(\sqrt{k})$ samples, which constitutes a quadratic improvement over the sample complexity of learning this class of distributions to constant accuracy. So, for (unstructured) discrete distributions of bounded support, sample amplification is strictly easier than learning.

While Axelrod et al. (2020, 2024) were motivated by distribution learning, we propose a variant of sample amplification that is inspired by the standard setting of computational probably approximate correct (PAC) learning theory (Valiant, 1984): Given i.i.d. samples $\{(x_i, f(x_i))\}_i$ of some unknown function $f : \{0,1\}^n \to \{0,1\}$ from some known class $\mathcal{F}$, output a hypothesis $\hat{f}$ s.t. $\mathbf{Pr}_x[f(x) \neq \hat{f}(x)]$ is small with high success probability. Often, the inputs $\{x_i\}_i$ are assumed to be i.i.d. uniformly random n-bit strings, so that the learning task can equivalently be formulated as that of learning an unknown distribution from a known class $\mathcal{D}_\mathcal{F} = \{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$ from samples to good approximation in total variation distance. We thus propose the following structured version of sample amplification for distributions described by Boolean functions:

Definition 1 (Sample amplification of functions – Informal) *A class $\mathcal{F}$ of Boolean functions is said to admit a* $(t, t+m, \epsilon)$-sample amplification scheme *if the class $\mathcal{D}_\mathcal{F} = \{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$ of distributions does.*

We will study sample amplification for different Boolean function classes. This will then serve as a tool for understanding approximate quantum cloning of structured classes of states.

Structured quantum cloning. Approximate quantum cloning[1] (Werner, 1998) is the task of mapping t many i.i.d. copies of an n-qubit state to a $((t+m)\cdot n)$-qubit state that is close to $t+m$ i.i.d. copies of the original state. The natural notions of distance to consider are (in)fidelity and trace distance. Werner (1998) studied the problem for pure state cloning with respect to infidelity, so that the figure of merit for a linear completely positive and trace-preserving (CPTP) cloning map $\Lambda : \mathcal{B}((\mathbb{C}^{2^n})^{\otimes t}) \to \mathcal{B}((\mathbb{C}^{2^n})^{\otimes t+m})$ is given as

$$\mathfrak{F}(\Lambda) = \inf_{|\psi\rangle} \operatorname{tr}\left(|\psi\rangle\langle\psi|^{\otimes t+m}\, \Lambda\left(|\psi\rangle\langle\psi|^{\otimes t}\right)\right). \tag{3}$$

Werner (1998) showed that the optimal cloning fidelity for pure state cloning from t copies to $t+m$ copies equals $\frac{d[t]}{d[t+m]}$, where $d[t] = \binom{2^n+t-1}{t}$. This is achieved by projecting $|\psi\rangle^{\otimes t} \otimes I_{2^{nm}}/2^{mn}$ onto the symmetric subspace of $(\mathbb{C}^{2^n})^{\otimes t+m}$. In particular, the optimal achievable fidelity for cloning one extra copy ($m=1$) is $\frac{t+1}{t+2^n}$. Consequently, the sample complexity necessary and sufficient to achieve $1-\epsilon$ fidelity when cloning one extra copy is $t = \Theta\left(2^n/\epsilon\right)$. This matches that of optimal pure state tomography with respect to infidelity (Bruß and Macchiavello, 1999; Gross et al., 2010; Baldwin et al., 2016; Haah et al., 2016; O'Donnell and Wright, 2016; Pelecanos et al., 2025b). While exact cloning is known to become impossible as soon as we allow for any two non-orthogonal states, these bounds for approximate cloning are derived for the case of an arbitrary unknown state; the analysis of approximate cloning in Werner (1998); Keyl and Werner (1999) relies on the representation theory of the *full* unitary group and in particular on Schur-Weyl duality. Thus, it does not carry over to structured classes of quantum states.

Quantum learning theory has recently focused on learning structured classes of states such as stabilizer states (Montanaro, 2017) or states prepared using few non-Clifford gates (Leone et al., 2024; Grewal et al., 2025), matrix-product states (Landon-Cardinal et al., 2010; Cramer et al., 2010; Landau et al., 2015; Arad et al., 2017; Bakshi et al., 2025), states with bounded gate complexity (Zhao et al., 2024), output states of shallow circuits (Huang et al., 2024; Vasconcelos and Huang, 2025; Landau and Liu, 2025), Gibbs states (Anshu et al., 2021; Rouzé and Stilck França, 2024; Haah et al., 2024; Bakshi et al., 2024; Chen et al., 2025a; Bluhm et al., 2025; Arunachalam et al., 2025), and more. This motivates us to connect the modern perspective of quantum learning theory with the foundational question of cloning by proposing a structured version of approximate cloning:

Definition 2 (Cloning of structured states – Informal) *A class $\mathcal{S}$ of quantum states admits a $(t, t+m, \epsilon)$-quantum cloning scheme if there exists a linear CPTP map $\Lambda_{\mathcal{S},t,m,\epsilon} : \mathcal{B}((\mathbb{C}^{2^n})^{\otimes t}) \to \mathcal{B}((\mathbb{C}^{2^n})^{\otimes t+m})$ s.t.*

$$\sup_{\rho\in\mathcal{S}} \mathrm{d}_{\mathrm{TD}}\left(\Lambda_{\mathcal{S},t,m,\epsilon}(\rho^{\otimes t}), \rho^{\otimes t+m}\right) \leq \epsilon\,. \tag{4}$$

We call ϵ the error incurred by the cloning scheme.

We will study structured cloning of stabilizer states. While adapting the reasoning of Werner (1998) to stabilizer state cloning may be achievable via the recently developed representation theory

1. From here on, we often omit the "approximate." We always consider *approximate* cloning unless specified otherwise.

Problem Instance (Classical/Quantum)	Learning (with small constant error)	Sample Amplification/Cloning (with small constant error)
n-bit parities	$\Theta(n)$	$\Omega(n)$ (Corollary 24)
$k_{\mathcal{F}}$-dimensional Boolean function subspace	$\Theta(k_{\mathcal{F}})$	$\Omega(k_{\mathcal{F}})^*$ (Theorem 28) (*depending on properties of the code $C_{\mathcal{F}}$)
n-qubit stabilizer states	$\Theta(n)$ (Montanaro, 2017)	$\Omega(n)$ (Theorem 47)

Table 1: **Overview of our main results:** A comparison between known sample complexity bounds for learning (with small constant error) and our sample complexity lower bounds for amplification/cloning.

of the Clifford group and the Clifford commutant (Gross et al., 2021; Bittel et al., 2025), we take an alternative path that relies on more elementary representation theory and an argument based on linear independence.

1.2. Overview of the Main Results

With the framework established, we can now study the two main questions highlighted above: First, is structured sample amplification easier than learning? And second, is structured quantum cloning easier than learning? In this work, we show that the two questions are related for specific structured classes of functions and states, and for those classes we give negative answers to both questions. Our results on sample amplification and cloning sample complexity lower bounds for different structured classes are highlighted and compared with learning sample complexities in Table 1.

Structured sample amplification. On the classical side, we develop a general lower bound on the optimal achievable error in structured sample amplification for any structured distribution class of the form $\mathcal{D} = \{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$, with $\mathcal{F}$ a known Boolean function class. We then instantiate this bound for parity functions, and more generally for function classes that form a linear space.

Theorem 3 (Structured Sample Amplification Lower Bounds – Informal)

1. ***General lower bound:*** *Let $\mathcal{F}$ be a class of Boolean function on n bits. Let $b_{\mathcal{F}}$ be the teaching dimension (Goldman and Kearns, 1995) of $\mathcal{F}$ (see Section 2 for a definition). Sample amplification for $\mathcal{F}$ from $< b_{\mathcal{F}}$ samples incurs at least an error directly proportional to the probability of exactly learning $\mathcal{F}$ from $b_{\mathcal{F}}$ random samples.*

2. ***Lower bound for parities:*** *Sample amplification for the class $\mathcal{F}_{\text{par}} = \{x \mapsto s\cdot x \mod 2\}_{s\in\mathbb{Z}_2^n}$ of n-bit parity functions from $< n$ samples incurs at least a constant error of 0.14.*

3. ***Lower bounds from coding theory:*** *Let $\mathcal{F}$ be some $k_{\mathcal{F}}$-dimensional linear subspace of all Boolean functions on n bits. Sample amplification for $\mathcal{F}$ from $< k_{\mathcal{F}}$ samples incurs at least an error directly proportional to the probability that a certain corresponding linear code (see Section 2 for a definition) corrects $k_{\mathcal{F}}$ random erasure errors.*

Whereas sample amplification is strictly easier than learning for the class of all Boolean functions as a consequence of results from Axelrod et al. (2020, 2024) (see Section 2.4), Theorem 3, proved in Section 2, shows that structured sample amplification to small constant error is as hard as learning for particular structured classes of functions.

Structured quantum cloning. On the quantum side, we make use of our lower bounds for structured sample amplification to derive optimal cloning lower bounds for different families of states. We first prove a general lower bound on the optimal achievable error for cloning a set of (mixed) states that "hide" unknown symmetry subgroups of a large (known) Abelian group. Here, we say that a state ρ "hides" a subgroup $H \leq G$ under some fixed unitary representation μ if the elements of H leave the state invariant, i.e., $\operatorname{tr}(\rho\mu(h)) = 1, \forall h \in H$, and the elements not in H change the state significantly, i.e., $|\operatorname{tr}(\rho\mu(g))|$ is bounded away from 1 for all $g \in G \setminus H$. We then instantiate this bound for the class of stabilizer states.

Theorem 4 (Structured Quantum Cloning Lower Bounds – Informal)

1. ***General lower bound****: Let G be a known Abelian group. Let $\mathcal{S}_\alpha$ be a class of (mixed) n-qubit states that "hide" unknown symmetry subgroups $H \leq G$ of order α. Let $n_{H^\perp}$ be the number of generators of $H^\perp$, the dual of H (see Section A for definitions). Cloning for S_α from $< n_{H^\perp}$ samples incurs at least an error directly proportional to the probability of exactly learning H from $n_{H^\perp}$ copies of the state.*

2. ***Lower bound for stabilizer states****: Cloning for the class of pure n-qubit stabilizer states from $< \lfloor n/4 \rfloor$ copies incurs at least a constant error of 0.14.*

Approximate cloning is already known to be as hard as learning for the unstructured class of all pure states (Werner, 1998; Keyl and Werner, 1999). Theorem 4, which is proved in Section 3, shows that the same is true for stabilizer states. Thus, imposing such structure does not separate cloning from learning in terms of sample complexity. In proving Theorem 4, we also show the optimality of character POVMs for particular classes of Abelian StateHSPs (Bouland et al., 2025; Hinsche et al., 2026) (with non-isomorphic irreps). This allows us to characterize the sample complexity of solving such problems up to an additive rather than a multiplicative constant, which is needed in our argument. For our proof, we also develop a structured random purification channel, a version of the random purification channel introduced in Tang et al. (2025) that is tailored to the structured classes of states relevant in Abelian StateHSPs. Together, these results allow us to derive sample complexity lower bounds for different Abelian StateHSPs, and thereby cloning lower bounds.

1.3. Techniques

Structured sample amplification lower bound. Sample amplification schemes for discrete distributions (Axelrod et al., 2020, 2024) are based on repeating some observed samples since, by birthday paradox, for a discrete distribution with support size k and for a sample size of $\Omega(\sqrt{k})$, repetitions are not overly suspicious. We show that for specific structured distributions, this approach fails. To understand why, consider a simple example: Take the class of distributions over $\mathbb{Z}_2^{2n}$ that are uniform on some n-dimensional subspace. Learning such a distribution comes down to seeing n linearly independent samples, which is reasonably likely from n samples but impossible from $n-1$ samples. In this sense, n uniformly random samples hold significantly more information about the unknown distribution than $n-1$ samples do. As information cannot be created "for free," amplifying from $n-1$ to n samples with small error becomes impossible. This reasoning can be formalized via the "triangle inequality" (Axelrod et al., 2020, 2024)

$$\epsilon^\star_{SA}(\mathcal{D}, t-1, t) \geq \epsilon^\star_L(\mathcal{D}, t-1) - \epsilon^\star_L(\mathcal{D}, t)\,, \tag{5}$$

where $\epsilon^\star_{SA}(\mathcal{D}, t-1, t)$ is the sample amplification error from $t-1$ samples to t samples for the distribution class $\mathcal{D}$, and $\epsilon^\star_L(\mathcal{D}, t-1)$ and $\epsilon^\star_L(\mathcal{D}, t)$ are the learning errors from $t-1$ and t samples, respectively, for the distribution class $\mathcal{D}$. This triangle inequality arises from the following simple algorithm for learning from $t-1$ samples: First, optimally sample amplify to t samples, then run an optimal learner from t samples. Clearly, this algorithm achieves a learning error of at most $\epsilon^\star_{SA}(\mathcal{D}, t-1, t) + \epsilon^\star_L(\mathcal{D}, t)$.

An obstacle to using Equation (5) in proving sample amplification lower bounds is that sample complexity upper and lower bounds in learning theory are often established only up to *multiplicative* constants. In contrast, to obtain meaningful lower bounds from Equation (5), we require a detailed understanding of the change in learning error when the sample size is increased by an *additive* constant. Using a folklore argument based on the probability of randomly drawn bit strings being linearly independent, we can obtain such understanding for parity learning. Namely, the sample complexity of (exactly) learning n-bit parities with respect to uniformly random inputs (with some high, constant probability $2/3$) is $> n-1$ and $\leq n + \Theta(1)$. We combine this with Equation (5) to prove a constant lower bound on the optimal error $\epsilon^\star_{SA}(\mathcal{D}_{\mathrm{par}}, n-1, n)$ in amplifying from $n-1$ to n samples for parity distributions $\mathcal{D}_{\mathrm{par}} = \{(\mathcal{U}_n, f)\}_{f \in \mathcal{F}_{\mathrm{par}}}$. This shows that non-trivial sample amplification of parity functions requires $\geq n$ samples, which matches the sample complexity of parity learning up to an additive constant. We mention that a similar linear independence idea was considered in Axelrod et al. (2024) while presenting a very different example of a distribution which is as hard to sample amplify as to learn.

To repeat, the simple but crucial observation in the lower bound for parity amplification is as follows: Seeing $n-1$ samples do not suffice to uniquely specify the unknown parity function, and guessing the wrong one incurs non-neglible error. In contrast, when seeing $n + \Theta(1)$ samples, they contain a linearly independent subset of size n with high probability, thus uniquely specifying the unknown parity function and allowing us to learn it with no error. Accordingly, only a small (constant) increase in sample size leads to a sizeable improvement in learning. And as such an improvement cannot come"for free," achieving such an increase in sample size is not possible without incurring a large error. We demonstrate that this linear independence-based reasoning extends beyond parities to general linear spaces of Boolean functions, i.e., to function classes $\mathcal{F}$ that form a linear subspace of the space of all Boolean functions on n bits. More precisely, starting from $\mathcal{F}$ we construct a linear code $C_{\mathcal{F}}$, and we show that the properties of parities that we relied on above—any two distinct parities are far from each other; and we likely obtain a maximal linearly independent set when seeing a sample of size slightly larger than n, the dimension of the space of parity functions—become properties of the code $C_{\mathcal{F}}$ and its dual—relating to the distance of $C_{\mathcal{F}}$ and to the ability of its dual code to correct random erasures.

Bell sample amplification lower bound. The superficial similarity between quantum cloning and sample amplification becomes immediate if we consider a restricted version of cloning in which the goal is to generate additional outcomes obtained when performing a fixed measurement on an unknown state. In particular, given the power of Bell sampling in extracting information from stabilizer states (Montanaro, 2017; Gross et al., 2021; Hangleiter and Gullans, 2024; Grewal et al., 2024), one may ask: Given the t outcomes obtained from performing Bell sampling on $2t$ copies of an unknown n-qubit stabilizer state, can we generate an (approximate) additional Bell sampling outcome? Relying on the by now well-developed understanding of Bell sampling, this question is easily seen to be exactly one of sample amplifying a probability distribution uniform over an

unknown n-dimensional subspace of $\mathbb{Z}_2^{2n}$. By the linear independence argument explained above, this sample amplification task requires at least n samples. That is, sample amplification of the outcome distribution of Bell sampling performed on copies of an unknown stabilizer state requires at least n samples.

Optimality of Bell sampling for mixed-phaseless-stabilizer StateHSP instance cloning. The result of the previous paragraph in particular implies that one cannot clone stabilizer states from $< 2n$ copies when using only Bell sampling followed by classical post-processing and state preparation. However, a general stabilizer state cloner may process the initial copies differently, potentially using a global measurement. Thus, to deduce general lower bounds for stabilizer state cloning from the above, we show optimality of Bell sampling measurements for cloning instances of mixed-phaseless-stabilizer StateHSP (See Section A for a definition), that is, for mixed $(2n)$-qubit states that have a well-defined structure and phaseless stabilizer group. Then, we lift the resulting mixed-phaseless-StateHSP instance cloning lower bound to pure stabilizer states via a novel structured version of the recently introduced random purification channel (Tang et al., 2025) that we tailor to mixed-phaseless-stabilizer StateHSP instances.

We take the special instances that are diagonal in the Bell basis. For such states, Bell sampling measurement can be used as a generic pre-processing that does not affect the state, and after observing a Bell sampling outcome, no further information can be extracted from the post-measurement state. This makes Bell sampling followed by classical post-processing optimal for such instances, which then allows us to reduce to a linear independence-based sample amplification lower bounds as above. This idea can be applied to more general mixed-state version of Abelian StateHSP. Namely, we show the optimality of character POVMs in solving mixed Abelian StateHSP when all irreducible representations are non-isomorphic. This allows us to argue that true copies hold strictly more information about the hidden subgroup than cloned copies, leading us to our cloning lower bound for the mixed state case.

From mixed to pure via structured random purification. As mentioned in the previous paragraph, to obtain cloning lower bounds for pure stabilizer states from those for mixed-phaseless-stabilizer StateHSP, we develop a structured version of the random quantum purification channel (Tang et al., 2025). Namely, we construct a channel that maps i.i.d. copies of our mixed-phaseless-stabilizer StateHSP instances to i.i.d. copies of random purifications that are themselves pure stabilizer states[2]. Thus, we can clone mixed-phaseless-stabilizer instances by first applying the structured random purification channel, then applying a pure stabilizer state cloner, and finally tracing out the auxiliary registers. Thereby, the lower bound from the mixed-phaseless-stabilizer StateHSP implies a cloning lower bound also for pure stabilizer states.

1.4. Directions for Future Work

We have studied quantum cloning for particular classes of structured quantum states related to Abelian StateHSP problems, with stabilizer states as a concrete example. For this class, we have shown that the sample complexity of cloning essentially coincides with that of learning. We have achieved this using representation theory and a structured version of random purification to relate

2. We also show that this random purification channel can be viewed as a concrete special case of Walter and Witteveen (2025)'s random purification for general symmetries.

quantum cloning to classical sample amplification. Here, we have established that sample amplification is as hard as learning for parity functions, based on a simple linear independence argument. Finally, we have highlighted some connections between sample amplification and coding theory. Thus, our work raises several natural follow-up questions.

Sample amplification of Reed-Muller codes. As we have given a general recipe for obtaining sample amplification lower bounds for linear spaces of Boolean functions from coding theory, a clear challenge is to instantiate this recipe beyond our example of parities. A natural next example to explore here is the space of low-degree polynomials over $\mathbb{Z}_2$ and the corresponding Reed-Muller codes. Using our arguments, the structured sample amplification error for degree-d polynomials on n bits can be lower bounded by the probability that a Reed-Muller (with suitable parameters) can correct $\binom{n}{\leq d}$ random erasure errors. While the question of robustness to random erasures has been studied extensively for Reed-Muller codes and they are known to achieve capacity for erasure noise channels in a variety of regimes (Abbe et al., 2015; Bhandari et al., 2022), characterising the above probability, to the best of our knowledge, remains open.

Combinatorial dimension for sample amplification. It has been a fruitful endeavour in computational learning theory to characterize learnability of concept classes in different models via combinatorial parameters such as the VC dimension (Vapnik and Chervonenkis, 1971), teaching dimension (Goldman and Kearns, 1995), or Eluder dimension (Li et al., 2022), among many others. Theorem 3 already shows that the teaching dimension plays a key role in lower bounds for sample amplification. However, our results fall short of a full combinatorial characterization of the complexity of sample amplification. Achieving such a characterization may require a new combinatorial dimension and would constitute a significant advance over the current case-by-case understanding of sample amplification.

Cloning hypergraph states. We have shown that stabilizer states are as hard to (approximately) clone as to learn, an approximate No-Cloning theorem for stabilizer states. It is now natural to consider the same question for other structured classes of states. An interesting example to explore here is the class of hypergraph states of order d (Rossi et al., 2013), which also admit a description in terms of stabilizer operators that include Toffoli gates in addition to Paulis. These hypergraph states can alternatively be thought of as phase states

$$|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_x (-1)^{f(x)} |x\rangle \,, \tag{6}$$

with f a degree-d polynomial over $\mathbb{Z}_2$ (Arunachalam et al., 2023). It is an intriguing question whether cloning hypergraph states can be related to sample amplification of low-degree polynomials, and whether such a connection can give rise to lower bounds. Beyond hypergraph states, the study of cloning for classes such as matrix product states, output states of QNC^0 and QAC^0 circuits may be of interest.

Cloning against restricted distinguishers. In the distinguisher-based perspective on cloning, we have not restricted the computational power of the distinguisher. As Axelrod et al. (2020) has highlighted for sample amplification, we may obtain interesting variants of the amplification and cloning tasks by considering restricted distinguishers. Concretely, it is natural to consider the task of cloning structured classes of states against polynomial-time adversaries. This is naturally motivated by quantum cryptography, with potential applications in quantum money (Wiesner, 1983; Ji et al.,

2018; Molina et al., 2013), and related questions have been considered in Fefferman et al. (2026); Nehoran and Zhandry (2024); Bostanci et al. (2025). While our lower bounds for stabilizer state cloning immediately carry over to cloning against polynomial-time adversaries, since our proofs of these bounds use computationally bounded distinguishers, it remains to explore the impact of imposing computational restrictions on the adversary for cloning other classes of quantum states. Specifically, it would be interesting to see whether cloning and learning can be separated for relevant classes of states in this computational model of cloning.

Average-case quantum cloning. Finally, our formulation of quantum cloning as well as our proven lower bounds are in a worst-case picture. In particular, our result on the optimality of Bell sampling for phaseless stabilizer group learning holds for a worst-case notion of learning. In contrast, Werner's optimal cloning map for general pure states is optimal both in the worst case and in the average case (Werner, 1998). We conjecture that our worst-case lower bounds for phaseless stabilizer StateHSP and stabilizer state cloning similarly extend to the average case.

Acknowledgments

We thank Yanlin Chen, Daniel Grier, and Natalie McHugh for insightful discussions.

References

Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. Reed-muller codes for random erasures and errors. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, page 297–306, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450335362. doi: 10.1145/2746539.2746575. URL https://doi.org/10.1145/2746539.2746575.

Anurag Anshu and Srinivasan Arunachalam. A survey on the complexity of learning quantum states. *Nature Reviews Physics*, 6(1):59–69, 2024. URL https://www.nature.com/articles/s42254-023-00662-4.

Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, 17:931–935, 2021. doi: 10.1038/s41567-021-01232-0. URL https://doi.org/10.1038/s41567-021-01232-0.

Itai Arad, Zeph Landau, Umesh Vazirani, and Thomas Vidick. Rigorous rg algorithms and area laws for low energy eigenstates in 1d. *Communications in Mathematical Physics*, 356(1):65–105, August 2017. ISSN 1432-0916. doi: 10.1007/s00220-017-2973-z. URL http://dx.doi.org/10.1007/s00220-017-2973-z.

Srinivasan Arunachalam and Ronald de Wolf. Guest column: A survey of quantum learning theory. *SIGACT News*, 48(2):41–67, June 2017. ISSN 0163-5700. doi: 10.1145/3106700.3106710. URL https://doi.org/10.1145/3106700.3106710.

Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J. Yoder. Optimal Algorithms for Learning Quantum Phase States. In Omar Fawzi and Michael Walter, editors, *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC*

2023), volume 266 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:24, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-283-9. doi: 10.4230/LIPIcs.TQC.2023.3. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.TQC.2023.3.

Srinivasan Arunachalam, Arkopal Dutt, and Francisco Escudero Gutiérrez. Testing and learning structured quantum hamiltonians. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 1263–1270, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715105. doi: 10.1145/3717823.3718289. URL https://doi.org/10.1145/3717823.3718289.

Brian Axelrod, Shivam Garg, Vatsal Sharan, and Gregory Valiant. Sample amplification: increasing dataset size even when learning is impossible. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020. URL https://proceedings.mlr.press/v119/axelrod20a.html.

Brian Axelrod, Shivam Garg, Yanjun Han, Vatsal Sharan, and Gregory Valiant. On the statistical complexity of sample amplification. *The Annals of Statistics*, 52(6):2767–2790, 2024. doi: 10.1214/24-AOS2444. URL https://doi.org/10.1214/24-AOS2444.

Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. Learning quantum hamiltonians at any temperature in polynomial time. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1470–1477, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703836. doi: 10.1145/3618260.3649619. URL https://doi.org/10.1145/3618260.3649619.

Ainesh Bakshi, John Bostanci, William Kretschmer, Zeph Landau, Jerry Li, Allen Liu, Ryan O'Donnell, and Ewin Tang. Learning the closest product state. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 1212–1221. ACM, June 2025. doi: 10.1145/3717823.3718207. URL http://dx.doi.org/10.1145/3717823.3718207.

Charles H. Baldwin, Ivan H. Deutsch, and Amir Kalev. Strictly-complete measurements for bounded-rank quantum-state tomography. *Phys. Rev. A*, 93:052105, May 2016. doi: 10.1103/PhysRevA.93.052105. URL https://link.aps.org/doi/10.1103/PhysRevA.93.052105.

Siddharth Bhandari, Prahladh Harsha, Ramprasad Saptharishi, and Srikanth Srinivasan. Vanishing spaces of random sets and applications to reed-muller codes. In *Proceedings of the 37th Computational Complexity Conference*, CCC '22, Dagstuhl, DEU, 2022. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 9783959772419. doi: 10.4230/LIPIcs.CCC.2022.31. URL https://doi.org/10.4230/LIPIcs.CCC.2022.31.

Lennart Bittel, Jens Eisert, Lorenzo Leone, Antonio A. Mele, and Salvatore F. E. Oliviero. A complete theory of the clifford commutant, 2025. URL https://arxiv.org/abs/2504.12263.

Andreas Bluhm, Matthias C. Caro, Francisco Escudero Gutiérrez, Aadil Oufkir, and Cambyse Rouzé. Certifying and learning quantum ising hamiltonians, 2025. URL https://arxiv.org/abs/2509.10239.

John Bostanci, Barak Nehoran, and Mark Zhandry. A general quantum duality for representations of groups with applications to quantum money, lightning, and fire. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 201–212, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715105. doi: 10.1145/3717823.3718195. URL https://doi.org/10.1145/3717823.3718195.

Adam Bouland, Tudor Giurgică-Tiron, and John Wright. The state hidden subgroup problem and an efficient algorithm for locating unentanglement. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 463–470, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715105. doi: 10.1145/3717823.3718118. URL https://doi.org/10.1145/3717823.3718118.

Dagmar Bruß and Chiara Macchiavello. Optimal state estimation for d-dimensional quantum systems. *Physics Letters A*, 253(5–6):249–251, March 1999. ISSN 0375-9601. doi: 10.1016/s0375-9601(99)00099-7. URL http://dx.doi.org/10.1016/S0375-9601(99)00099-7.

Nader H. Bshouty and Jeffrey C. Jackson. Learning dnf over the uniform distribution using a quantum example oracle. In *Proceedings of the Eighth Annual Conference on Computational Learning Theory*, COLT '95, page 118–127, New York, NY, USA, 1995. Association for Computing Machinery. ISBN 0897917235. doi: 10.1145/225298.225312. URL https://doi.org/10.1145/225298.225312.

Matthias C. Caro, Marcel Hinsche, Marios Ioannou, Alexander Nietner, and Ryan Sweke. Classical Verification of Quantum Learning. In Venkatesan Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 24:1–24:23, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-309-6. doi: 10.4230/LIPIcs.ITCS.2024.24. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2024.24.

Matthias C. Caro, Preksha Naik, and Joseph Slote. Testing Classical Properties from Quantum Data. In Shubhangi Saraf, editor, *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, volume 362 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 34:1–34:26, Dagstuhl, Germany, 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-410-9. doi: 10.4230/LIPIcs.ITCS.2026.34. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2026.34.

Chi-Fang Chen, Anurag Anshu, and Quynh T. Nguyen. Learning quantum gibbs states locally and efficiently. In *2025 IEEE 66th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1458–1464, 2025a. doi: 10.1109/FOCS63196.2025.00076. URL https://doi.org/10.1109/FOCS63196.2025.00076.

Kean Chen, Qisheng Wang, and Zhicheng Zhang. Local test for unitarily invariant properties of bipartite quantum states, 2025b. URL https://arxiv.org/abs/2404.04599.

Kean Chen, Filippo Girardi, Aadil Oufkir, Nengkun Yu, and Zhicheng Zhang. Quantum channel tomography: optimal bounds and a heisenberg-to-classical phase transition, 2026a. URL https://arxiv.org/abs/2604.17369.

Kean Chen, Nengkun Yu, and Zhicheng Zhang. Quantum channel tomography and estimation by local test, 2026b. URL https://arxiv.org/abs/2512.13614.

Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature communications*, 1(1):149, 2010. doi: 10.1038/ncomms1147. URL https://doi.org/10.1038/ncomms1147.

D. Dieks. Communication by epr devices. *Physics Letters A*, 92(6):271–272, 1982. ISSN 0375-9601. doi: https://doi.org/10.1016/0375-9601(82)90084-6. URL https://www.sciencedirect.com/science/article/pii/0375960182900846.

Bill Fefferman, Soumik Ghosh, Makrand Sinha, and Henry Yuen. The Hardness of Learning Quantum Circuits and Its Cryptographic Applications. In Shubhangi Saraf, editor, *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, volume 362 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 56:1–56:21, Dagstuhl, Germany, 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-410-9. doi: 10.4230/LIPIcs.ITCS.2026.56. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2026.56.

Steven R Finch. *Mathematical constants*. Cambridge university press, 2003. URL https://www.cambridge.org/gb/universitypress/subjects/mathematics/recreational-mathematics/mathematical-constants-1?format=PB.

William Fulton and Joe Harris. *Representation theory: a first course*, volume 129. Springer Science & Business Media, 2013. URL https://link.springer.com/book/10.1007/978-1-4612-0979-9.

Filippo Girardi, Francesco Anna Mele, and Ludovico Lami. Random purification channel made simple, 2025a. URL https://arxiv.org/abs/2511.23451.

Filippo Girardi, Francesco Anna Mele, Haimeng Zhao, Marco Fanizza, and Ludovico Lami. Random stinespring superchannel: converting channel queries into dilation isometry queries, 2025b. URL https://arxiv.org/abs/2512.20599.

S.A. Goldman and M.J. Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50(1):20–31, 1995. ISSN 0022-0000. doi: https://doi.org/10.1006/jcss.1995.1003. URL https://www.sciencedirect.com/science/article/pii/S0022000085710033.

Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Improved stabilizer estimation via bell difference sampling. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC '24, page 1352–1363. ACM, June 2024. doi: 10.1145/3618260.3649738. URL http://dx.doi.org/10.1145/3618260.3649738.

Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Efficient learning of quantum states prepared with few non-clifford gates. *Quantum*, 9:1907, November 2025. ISSN 2521-327X. doi: 10.22331/q-2025-11-06-1907. URL http://dx.doi.org/10.22331/q-2025-11-06-1907.

David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010. doi: 10.1103/PhysRevLett.105.150401. URL https://link.aps.org/doi/10.1103/PhysRevLett.105.150401.

David Gross, Sepehr Nezami, and Michael Walter. Schur–weyl duality for the clifford group with applications: Property testing, a robust hudson theorem, and de finetti representations. *Communications in Mathematical Physics*, 385(3):1325–1393, June 2021. ISSN 1432-0916. doi: 10.1007/s00220-021-04118-7. URL http://dx.doi.org/10.1007/s00220-021-04118-7.

Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 913–925, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341325. doi: 10.1145/2897518.2897585. URL https://doi.org/10.1145/2897518.2897585.

Jeongwan Haah, Robin Kothari, and Ewin Tang. Learning quantum hamiltonians from high-temperature gibbs states and real-time evolutions. *Nature Physics*, 20:1027–1031, 2024. doi: 10.1038/s41567-023-02376-x. URL https://doi.org/10.1038/s41567-023-02376-x.

Dominik Hangleiter and Michael J. Gullans. Bell sampling from quantum circuits. *Physical Review Letters*, 133(2), July 2024. ISSN 1079-7114. doi: 10.1103/physrevlett.133.020601. URL http://dx.doi.org/10.1103/PhysRevLett.133.020601.

Marcel Hinsche, Jens Eisert, and Jose Carrasco. Abelian state hidden subgroup problem: Learning stabilizer groups and beyond. *PRX Quantum*, 7(2):020337, 2026. doi: 10.1103/6frk-891j. URL https://doi.org/10.1103/6frk-891j.

Hsin-Yuan Huang, Yunchao Liu, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R. McClean. Learning shallow quantum circuits. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC '24, page 1343–1351. ACM, June 2024. doi: 10.1145/3618260.3649722. URL http://dx.doi.org/10.1145/3618260.3649722.

Zhengfeng Ji, Yi-Kai Liu, and Fang Song. Pseudorandom quantum states. In *Advances in Cryptology – CRYPTO 2018: 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19–23, 2018, Proceedings, Part III*, page 126–152, Berlin, Heidelberg, 2018. Springer-Verlag. ISBN 978-3-319-96877-3. doi: 10.1007/978-3-319-96878-0_5. URL https://doi.org/10.1007/978-3-319-96878-0_5.

R. Jozsa. Quantum factoring, discrete logarithms, and the hidden subgroup problem. *Computing in Science & Engineering*, 3(2):34–43, 2001. doi: 10.1109/5992.909000. URL https://doi.org/10.1109/5992.909000.

M. Keyl and R. F. Werner. Optimal cloning of pure states, testing single clones. *Journal of Mathematical Physics*, 40(7):3283–3299, July 1999. ISSN 1089-7658. doi: 10.1063/1.532887. URL http://dx.doi.org/10.1063/1.532887.

Zeph Landau and Yunchao Liu. Learning quantum states prepared by shallow circuits in polynomial time. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25, page 1828–1838, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400715105. doi: 10.1145/3717823.3718311. URL https://doi.org/10.1145/3717823.3718311.

Zeph Landau, Umesh Vazirani, and Thomas Vidick. A polynomial time algorithm for the ground state of one-dimensional gapped local hamiltonians. *Nature Physics*, 11(7):566–569, 2015. URL https://www.nature.com/articles/nphys3345.

Olivier Landon-Cardinal, Yi-Kai Liu, and David Poulin. Efficient direct tomography for matrix product states, 2010. URL https://arxiv.org/abs/1002.4632.

Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. Learning t-doped stabilizer states. *Quantum*, 8:1361, May 2024. ISSN 2521-327X. doi: 10.22331/q-2024-05-27-1361. URL https://doi.org/10.22331/q-2024-05-27-1361.

Gene Li, Pritish Kamath, Dylan J Foster, and Nati Srebro. Understanding the eluder dimension. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 23737–23750. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/960cfbb846aff424ac20aadce6fa6530-Paper-Conference.pdf.

Antonio Anna Mele. Lecture notes on representation theory for quantum information (qmath masterclass 2025, copenhagen). Lecture Notes, 2025. URL https://antonioannamele.com/documents/QMATH_Basics_of_Representation_Theory_Mele.pdf.

Antonio Anna Mele and Lennart Bittel. Optimal learning of quantum channels in diamond distance, 2025. URL https://arxiv.org/abs/2512.10214.

Francesco Anna Mele, Filippo Girardi, Senrui Chen, Marco Fanizza, and Ludovico Lami. Random purification channel for passive gaussian bosons, 2025. URL https://arxiv.org/abs/2512.16878.

Abel Molina, Thomas Vidick, and John Watrous. Optimal counterfeiting attacks and generalizations for wiesner's quantum money. In Kazuo Iwama, Yasuhito Kawano, and Mio Murao, editors, *Theory of Quantum Computation, Communication, and Cryptography*, pages 45–64, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. ISBN 978-3-642-35656-8. URL https://doi.org/10.1007/978-3-642-35656-8_4.

Ashley Montanaro. Learning stabilizer states by bell sampling, 2017. URL https://arxiv.org/abs/1707.04012.

Barak Nehoran and Mark Zhandry. A Computational Separation Between Quantum No-Cloning and No-Telegraphing. In Venkatesan Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 82:1–82:23, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-309-6. doi: 10.4230/LIPIcs.ITCS.2024.82. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2024.82.

Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010. URL https://www.cambridge.org/highereducation/books/quantum-computation-and-quantum-information/01E10196D0A682A6AEFFEA52D53BE9AE#overview.

Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 899–912, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341325. doi: 10.1145/2897518.2897544. URL https://doi.org/10.1145/2897518.2897544.

Angelos Pelecanos, Jack Spilecki, Ewin Tang, and John Wright. Mixed state tomography reduces to pure state tomography, 2025a. URL https://arxiv.org/abs/2511.15806.

Angelos Pelecanos, Jack Spilecki, and John Wright. The debiased keyl's algorithm: a new unbiased estimator for full state tomography, 2025b. URL https://arxiv.org/abs/2510.07788.

M Rossi, M Huber, D Bruß, and C Macchiavello. Quantum hypergraph states. *New Journal of Physics*, 15(11):113022, November 2013. ISSN 1367-2630. doi: 10.1088/1367-2630/15/11/113022. URL http://dx.doi.org/10.1088/1367-2630/15/11/113022.

Cambyse Rouzé and Daniel Stilck França. Learning quantum many-body systems from a few copies. *Quantum*, 8:1319, April 2024. ISSN 2521-327X. doi: 10.22331/q-2024-04-30-1319. URL https://doi.org/10.22331/q-2024-04-30-1319.

Jean-Pierre Serre et al. *Linear representations of finite groups*, volume 42. Springer, 1977. URL https://link.springer.com/book/10.1007/978-1-4684-9458-7.

P.W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134, 1994. doi: 10.1109/SFCS.1994.365700. URL https://doi.org/10.1109/SFCS.1994.365700.

N. J. A. Sloane. A048651, oeis foundation inc. (2026), 1964. URL https://oeis.org/A048651.

Mehdi Soleimanifar and John Wright. *Testing matrix product states*, pages 1679–1701. 2022. doi: 10.1137/1.9781611977073.68. URL https://epubs.siam.org/doi/abs/10.1137/1.9781611977073.68.

Ewin Tang, John Wright, and Mark Zhandry. Conjugate queries can help, 2025. URL https://arxiv.org/abs/2510.07622.

L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, November 1984. ISSN 0001-0782. doi: 10.1145/1968.1972. URL https://doi.org/10.1145/1968.1972.

V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971. doi: 10.1137/1116025. URL https://doi.org/10.1137/1116025.

Francisca Vasconcelos and Hsin-Yuan Huang. Learning shallow quantum circuits with many-qubit gates. In Nika Haghtalab and Ankur Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 5553–5604. PMLR, 30 Jun–04 Jul 2025. URL https://proceedings.mlr.press/v291/vasconcelos25a.html.

Michael Walter and Freek Witteveen. A random purification channel for arbitrary symmetries with applications to fermions and bosons, 2025. URL https://arxiv.org/abs/2512.15690.

R. F. Werner. Optimal cloning of pure states. *Physical Review A*, 58(3):1827–1832, September 1998. ISSN 1094-1622. doi: 10.1103/physreva.58.1827. URL http://dx.doi.org/10.1103/PhysRevA.58.1827.

Stephen Wiesner. Conjugate coding. *SIGACT News*, 15(1):78–88, January 1983. ISSN 0163-5700. doi: 10.1145/1008908.1008920. URL https://doi.org/10.1145/1008908.1008920.

William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886): 802–803, 1982. doi: 10.1038/299802a0. URL https://doi.org/10.1038/299802a0.

John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016. URL https://people.eecs.berkeley.edu/~jswright/papers/thesis.pdf.

Satoshi Yoshida, Ryotaro Niwa, and Mio Murao. Random dilation superchannel, 2025. URL https://arxiv.org/abs/2512.21260.

Haimeng Zhao, Laura Lewis, Ishaan Kannan, Yihui Quek, Hsin-Yuan Huang, and Matthias C. Caro. Learning quantum states and unitaries of bounded gate complexity. *PRX Quantum*, 5: 040306, Oct 2024. doi: 10.1103/PRXQuantum.5.040306. URL https://link.aps.org/doi/10.1103/PRXQuantum.5.040306.

Alper Çakan, Vipul Goyal, and Omri Shmueli. Public-key quantum fire and key-fire from classical oracles, 2026. URL https://arxiv.org/abs/2504.16407.

Appendix A. Notation and Preliminaries

In this section, we introduce the notation and underlying notions used throughout the paper. A familiarity with standard notions as well as Dirac bra-ket notation in quantum information theory is assumed, for this the reader is referred to excellent textbooks such as Nielsen and Chuang (2010).

A.1. Representation Theory

In this subsection, we recall basic notions from representation theory. We refer the reader to the excellent references: Mele (2025); Fulton and Harris (2013); Serre et al. (1977) for a detailed introduction. Throughout, we will use V to denote a finite-dimensional vector space. We use $\mathrm{GL}(V)$ to denote the group of all bijective linear maps from V to itself.

Definition 5 (Representation of finite groups) *Let G be a finite group. Then a tuple (μ, V) with finite-dimensional vector space V and map $\mu : G \to \mathrm{GL}(V)$ is called a* representation *of the group G over the vector space V if μ is a homomorphism. That is,*

$$\mu(gh) = \mu(g)\mu(h) \quad \forall g, h \in G. \tag{7}$$

At times, we will be abusing the terminology: We will not explicitly mention the vector space V and simply call μ a representation, wherever the vector space is clear from the context.

Definition 6 (Sub-representation of finite Groups) *Let G be a finite group and let (μ, V) be a representation of the group G. Furthermore, let $W \subseteq V$ be a subspace such that W is G-invariant, i.e.,*

$$\mu(g)|w\rangle \in W \quad \forall |w\rangle \in W, \forall g \in G\,. \tag{8}$$

Then $(\mu|_W, W)$ is called a sub-representation *of the representation (μ, V).*

Definition 7 (Irreducible Representation) *Let G be a group with a representation (μ, V). Then a sub-representation $(\mu|_W, W)$ is called* irreducible (or, an irrep) *if there are no non-trivial sub-representations of $(\mu|_W, W)$.*

Here, we say that a sub-representation is non-trivial if it is onto a strict but non-trivial (i.e., not equal to $\{0\}$) subspace. It is trivial to see that any sub-representation onto a one-dimensional vector space is irreducible.

Definition 8 (Unitary Representation) *Let G be a finite group and let $(V, \langle\cdot|\cdot\rangle)$ be a complex inner product space. Then, a representation $(\mu, (V, \langle\cdot|\cdot\rangle))$ is called a* unitary representation *with respect to the inner product if,*

$$\langle \mu(g)v|\mu(g)w\rangle = \langle v|w\rangle, \quad \forall v, w \in V \text{ and } \forall g \in G. \tag{9}$$

In other words, a unitary representation is a representation μ such that $\mu(g)$ is unitary w.r.t. the inner product space $(V, \langle\cdot|\cdot\rangle)$ for all $g \in G$. For finite groups, any representation can be treated as a unitary representation with respect to a suitably defined inner product. Thus, it is natural to restrict our focus to unitary representations for finite groups.

The following is the natural notion of equivalence between representations:

Definition 9 (Isomorphic Representations) *Two representations (μ_1, V_1) and (μ_2, V_2) are called* isomorphic *if there exists an invertible linear map $\phi : V_1 \to V_2$ such that*

$$\mu_2(g) = \phi\mu_1(g)\phi^{-1}, \quad \forall g \in G\,. \tag{10}$$

If μ_1 and μ_2 are isormorphic, we write, $\mu_1 \cong \mu_2$.

The following complex-valued map induced by a representation is useful.

Definition 10 (Character) *For a representation μ of a finite group G, its* character *is the function*

$$\chi_\mu : G \to \mathbb{C}\,, \chi_\mu(g) = \mathrm{tr}(\mu(g))\,. \tag{11}$$

Two representations are isomorphic if and only if their characters are same, i.e.,

$$\mu_1 \cong \mu_2 \iff \chi_{\mu_1}(g) = \chi_{\mu_2}(g), \qquad \forall g \in G\,. \tag{12}$$

Lemma 11 (Schur's orthonormality condition) *Let μ and ν be two irreducible representations of a finite group G. Then,*

$$\mathop{\mathbf{E}}_{g\in G}[\overline{\chi_\mu(g)}\chi_\nu(g)] = \delta_{\mu\cong\nu}\,. \tag{13}$$

In this work, we restrict our focus to Abelian groups, i.e., groups G in which the group operation satisfies $gh = hg$ for all $g, h \in G$. For such Abelian groups, it is easy to fully characterize the irreducible representations and characters. To this end, first notice that, if G is a Abelian and if μ is a representation of G, then $\mu(g)$ and $\mu(h)$ commute for all $g, h \in G$, i.e., $[\mu(g), \mu(h)] = 0$. If μ is additionally unitary, then in particular this implies that $\mu(g)$ is unitarily diagonalizable for every $g \in G$, and together with the commutativity this implies that all $\mu(g)$ can be simultaneously diagonalized. It thus makes sense to speak of the eigenvectors of the representation μ.

Lemma 12 *For a unitary representation μ of an Abelian group G, the irreducible sub-representations of μ are exactly the one-dimensional spaces spanned by the eigenvectors of μ, and the corresponding eigenvalues are the corresponding characters.*

In the remainder of the paper, all representations will be unitary representation of an Abelian group over some Hilbert space, unless stated otherwise. We can then write the eigenvectors as $|\lambda, v_\lambda\rangle$, where λ indicates the eigenvalue and where v_λ enumerates the eigenvectors with the same eigenvalue $\chi_\lambda(g), \forall g \in G$. Equivalently, the v_λ enumerates the isomorphic irreducible representations with character χ_λ. The space spanned by isomorphic irreps, $\mathrm{span}\{|\lambda, v_\lambda\rangle\}_{v_\lambda}$, is called *$\lambda$-isotypic component* of the Hilbert space Now, we can define the projective measurement $\{\Pi_\lambda\}_\lambda$ via

$$\Pi_\lambda = \sum_{v_\lambda} |\lambda, v_\lambda\rangle\langle\lambda, v_\lambda|\,, \tag{14}$$

which are projectors on the λ-isotypic subspaces of the Hilbert space. This measurement is called the character POVM of the representation. The following alternative expression for the character POVM elements will be useful:

Fact 13 (Hinsche et al. (2026)) *The character POVM elements can be written as*

$$\Pi_\lambda = \frac{1}{|G|}\sum_{g\in G} \overline{\chi_\lambda(g)}\mu(g). \tag{15}$$

A.2. Abelian State Hidden Subgroup Problem

Symmetries are at the core of physics. It is thus a natural question to ask whether we can learn the symmetries of an unknown quantum state given some form of access to the state. This serves as a generalization of the *Hidden Subgroup Problem (HSP)* (Shor, 1994; Jozsa, 2001), which entails finding the hidden symmetries of a Boolean function from (either classical or quantum) query access. A restricted version of this problem for quantum states, the so-called *Abelian State Hidden Subgroup Problem (Abelian StateHSP)* has recently been formalised (Bouland et al., 2025; Hinsche et al., 2026). Formally, the problem is defined as follows:

Definition 14 (Abelian StateHSP $(G, \mu, \Psi_H^\epsilon)$ Hinsche et al. (2026)) *Let G be an Abelian group and let $H \leq G$ be a subgroup. Further, let $\mu : G \to U(\mathcal{H})$ be a unitary representation of G over some Hilbert space $\mathcal{H}$. Then, we say that a state $\rho \in \mathcal{B}(\mathcal{H})$* ϵ-hides the subgroup $H \leq G$ *(for which we write $\rho \in \Psi_H^\epsilon$) if,*

1. $\forall h \in H$, $\operatorname{tr}(\mu(h)\rho) = 1$, *and*
2. $\forall g \in G \setminus H$, $|\operatorname{tr}(\mu(g)\rho)| \leq 1 - \epsilon$.

The Abelian StateHSP for G is the following problem: Given i.i.d. copies of a state ρ that is promised to ϵ-hide H, identify H. If we are additionally promised that ρ is pure, then we refer to this variant of the problem as pure Abelian StateHSP.

In this formulation, we can also think of the Abelian StateHSP as a problem of discriminating between sets Ψ_H^ϵ of states, each of which ϵ-hides a symmetry subgroup $H \leq G$ of fixed order α. The Abelian StateHSP problem can be solved using $O(\log |G|/\epsilon)$ copies of the state (Hinsche et al., 2026). The main algorithmic step is to perform the character POVM $\{\Pi_\lambda\}_\lambda$ and to then classically post-process the observed measurement outcomes. Next, following Hinsche et al. (2026), we sketch how the Abelian StateHSP algorithm works in a bit more detail.

Note that the outcome probabilities when performing the character POVM are

$$q_\rho(\lambda) = \operatorname{tr}(\rho \Pi_\lambda). \tag{16}$$

For a finite Abelian group G, we can define the dual group $\widehat{G}$ as the group of all character functions over G corresponding to irreducible representations. Then, for a subgroup $H \leq G$, we can define the dual subgroup $H^\perp \leq \widehat{G}$

$$H^\perp = \{\lambda \in \widehat{G} \mid \chi_\lambda(h) = 1, \forall h \in H\}. \tag{17}$$

We also define

$$q_\rho(H^\perp) = \sum_{\lambda \in H^\perp} q_\rho(\lambda) = \operatorname{tr}\left(\rho \sum_{\lambda \in H^\perp} \Pi_\lambda\right). \tag{18}$$

As in Hinsche et al. (2026), the probability can be shown to be equal to

$$q_\rho(H^\perp) = \frac{1}{|H|} \sum_{h \in H} \operatorname{tr}(\mu(h)\rho). \tag{19}$$

For an Abelian StateHSP instance, $\rho \in \Psi_H^\epsilon$, we have

$$q_\rho(H^\perp) = 1. \tag{20}$$

Thus, if we use character POVMs $\{\Pi_\lambda\}_\lambda$, we will always sample a $\lambda \in H^\perp$, and given enough independent samples, we can determine $H^\perp$ and hence its dual H.

A.3. Stabilizer States and Phaseless Stabilizer StateHSP

Let $\mathcal{P}_n$ be the n-qubit Pauli group. The single-qubit Pauli group is generated by X and Z operators, which act as

$$\begin{aligned} X|a\rangle &= |a \oplus 1\rangle, \\ Z|a\rangle &= (-1)^a |a\rangle, \end{aligned} \tag{21}$$

where $a \in \{0, 1\}$. The n-qubit Pauli group consists of tensor products of elements of the single-qubit Pauli group. An n-qubit stabilizer state $|\psi\rangle \in \mathrm{Stab}_n$ is defined to have a stabilizer group $S \subset \mathcal{P}_n$ of order 2^n such that

$$P|\psi\rangle = |\psi\rangle, \forall P \in S. \tag{22}$$

We define the following n-qubit operators

$$V_x = \bigotimes_{i=1}^{n} Z^{a_i} X^{b_i}, \qquad \forall x \in \mathbb{Z}_2^{2n}\,, \tag{23}$$

which are Weyl operators up to an imaginary phase. Moreover, these operators satisfy,

$$V_x V_y = (-1)^{[x,y]} V_y V_x\,. \tag{24}$$

and that,

$$V_x V_y = (-1)^{b\cdot c} V_{x\oplus y}\,. \tag{25}$$

where $x = (a, b)$, $y = (c, d)$ and $[x, y] = a \cdot d + b \cdot c \mod 2$. Now, define the $(2n)$-qubit Bell basis states as,

$$|\Phi_y\rangle = (V_y \otimes I)|\Phi_0\rangle, \qquad \text{with} \qquad |\Phi_0\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\mathbb{Z}_2^{2n}} |x\rangle \otimes |x\rangle\,, \tag{26}$$

for any $y \in \mathbb{Z}_2^{2n}$. It is easy to see that these are orthonormal and forms a basis for the space of $(2n)$-qubit states. Moreover,

$$V_x^{\otimes 2} |\Phi_y\rangle = (-1)^{[x,y]} |\Phi_y\rangle, \qquad \forall x, y \in \mathbb{Z}_2^{2n}\,. \tag{27}$$

A.4. Mixed Phaseless Stabilizer StateHSP

We consider the following unitary representation,

$$\mu : \mathbb{Z}_2^{2n} \to \mathcal{U}(((\mathbb{C}^2)^{\otimes n})^{\otimes 2})\,, x \mapsto V_x^{\otimes 2}. \tag{28}$$

Then, we define the states σ_L as,

$$\sigma_L = \frac{1}{\sqrt{2^n}} \sum_{y\in L^\perp} |\Phi_y\rangle\langle\Phi_y|\,, \tag{29}$$

where L is an n-dimensional subspace of $\mathbb{Z}_2^{2n}$ and $L^\perp$ is its dual subspace defined as,

$$L^\perp = \{y \in \mathbb{Z}_2^{2n} \mid (-1)^{[x,y]} = 1, \forall x \in L\}. \tag{30}$$

Lemma 15 *The states σ_L as defined above have the properties,*

1. $\forall x \in L$, $\mathrm{tr}\left(V_x^{\otimes 2}\sigma_L\right) = 1$, *and*
2. $\forall x \notin L$, $\mathrm{tr}\left(V_x^{\otimes 2}\sigma_L\right) = 0$.

Thus, using Theorem 15, we can define an Abelian StateHSP $(\mathbb{Z}_2^{2n}, \mu, \Psi_L^1)$ where the unitary representation μ is given by Equation (28) and,

$$\Psi_L^1 = \{\sigma_L\}. \tag{31}$$

By Theorem 15, the state σ_L hides the subgroup L of $\mathbb{Z}_2^{2n}$. Now, the irreducible representations (eigenstates) of the unitary representation $V_x^{\otimes 2}$ are just the Bell basis defined in the previous section $\{|\Phi_y\rangle\}_{y\in\mathbb{Z}_2^{2n}}$. Moreover, the character (eigenvalues) are given as,

$$\chi_y(x) = (-1)^{[x,y]}\,. \tag{32}$$

From this, we in particular see that the characters for any two irreps $y \neq \tilde{y}$ satisfy $\chi_y \neq \chi_{\tilde{y}}$. Thus, from the previous discussion, all irreps are non-isomorphic. And, the character POVM is given by,

$$\Pi_y = |\Phi_y\rangle\langle\Phi_y|. \tag{33}$$

We call the Abelian StateHSP $(\mathbb{Z}_2^{2n}, \mu, \Psi_L^1)$ as mixed-phaseless-stabilizer StateHSP because of the similarity of the hidden group to the (phaseless) stabilizer group of the stabilizer states.

A.5. Random Purification

The random purification channel by Tang et al. (2025) allows one to transform t i.i.d. copies of an arbitrary unknown mixed state to t i.i.d. copies of a randomly chosen purification of that state.

Lemma 16 ((Tang et al., 2025), Lemma 2.11) *Let $m, r \in \mathbb{N}$. There is a unitary circuit $C^{(m)}$ such that, such that for all mixed states $\rho \in \mathbb{C}^{d\times d}$ of rank at most r,*

$$C^{(m)}(\rho^{\otimes m}) = \mathop{\mathbf{E}}_{|\rho\rangle} |\rho\rangle\langle\rho|^{\otimes m}\,, \tag{34}$$

where the expectation is over random purifications $|\rho\rangle \in \mathbb{C}^d \otimes \mathbb{C}^r$ of ρ. Additionally, the circuit $C^{(m)}$ can be implemented to accuracy ϵ with $\mathrm{poly}(m, \log d, \log(1/\epsilon))$ gate complexity.

On a very high level, the approach in Tang et al. (2025) is to first perform weak Schur sampling on the m copies of the mixed input state, and to then prepare a specific state conditioned on the outcome observed. The random purification channel has attracted considerable attention in the community recently, evidenced by an alternate construction and proof (Girardi et al., 2025a),an extension to bosonic or fermionic states (Mele et al., 2025; Walter and Witteveen, 2025), extensions to quantum channels (Girardi et al., 2025b; Yoshida et al., 2025), and applications in quantum learning theory (Pelecanos et al., 2025a; Mele and Bittel, 2025). Ideas similar to random purification have been studied before (Soleimanifar and Wright, 2022; Chen et al., 2025a). Specifically, Chen et al. (2025b) studies a notion of local testing, which can be thought of as dual to random purifications,

and which has been employed to study property testing of states (Chen et al., 2025b) and quantum channel tomography (Chen et al., 2026b,a).

We also highlight the more general random purification channel introduced recently for general symmetries (Walter and Witteveen, 2025). We denote the Hilbert space on which our mixed state lives in by $\mathcal{H}$ and the Hilbert space for the purifying register by $\mathcal{H}'$.

Theorem 17 ((Walter and Witteveen, 2025, from Theorem 3.1 and its proof)) *For any $*$-algebra $\mathcal{A}$, there is a quantum channel $\mathcal{P}_{\mathcal{A}} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H} \otimes \mathcal{H}')$ with the following properties:*

1. *Random Purification of symmetric states: For any quantum state $\rho \in \mathcal{S}(\mathcal{H})$ that commutes with $\mathcal{A}$ i.e. $\rho \in \mathcal{A}'$, the commutant of algebra $\mathcal{A}$, we have that*

$$\mathcal{P}_{\mathcal{A}}(\rho) = \int_G (I \otimes g^T)\psi_\rho^{\mathrm{std}}(I \otimes \bar{g})dg \tag{35}$$

   where $G \subseteq U(\mathcal{H})$ is any closed subgroup with $\mathbb{C}G = \mathcal{A}'$, and $\psi_\rho^{\mathrm{std}} = (\sqrt{\rho} \otimes I)|\tilde{\Phi}_0\rangle$, with $|\tilde{\Phi}_0\rangle = \sum_x |x\rangle|x'\rangle \in \mathcal{S}(\mathcal{H} \otimes \mathcal{H}')$ the un-normalized maximally entangled state on $\mathcal{H} \otimes \mathcal{H}'$.

2. *Assume that the Hilbert space decomposes as $\mathcal{H} \cong \bigoplus_{\lambda\in\Lambda} L_\lambda \otimes R_\lambda$ (and $\mathcal{H}' \cong \bigoplus_{\lambda\in\Lambda} L'_\lambda \otimes R'_\lambda$) with finite index set Λ, such that*

$$\mathcal{A} \cong \bigoplus_{\lambda\in\Lambda} \mathcal{B}(L_\lambda) \otimes I_{R_\lambda}, \quad \text{and} \quad \mathcal{A}' \cong \bigoplus_{\lambda\in\Lambda} I_{L_\lambda} \otimes \mathcal{B}(R_\lambda), \tag{36}$$

   (which is always possible for $*$-algebras). Then, the channel $\mathcal{P}_{\mathcal{A}}$ is given by

$$\mathcal{P}_{\mathcal{A}}(\rho) = \bigoplus_{\lambda\in\Lambda} \frac{|\tilde{\Phi}_0\rangle_{L_\lambda L'_\lambda}\langle\tilde{\Phi}_0|_{L_\lambda L'_\lambda}}{\dim L_\lambda} \otimes \mathrm{tr}_{L_\lambda}[P_\lambda \rho P_\lambda] \otimes \frac{I_{R'_\lambda}}{\dim R'_\lambda}, \tag{37}$$

   where $|\tilde{\Phi}_0\rangle_{L_\lambda L'_\lambda}$ is the un-normalised maximally entangled state on the $L_\lambda \otimes L'_\lambda$ component, P_λ be the projector on λ-component $L_\lambda \otimes R_\lambda$ of $\mathcal{H}$.

From Equation (37), we see that to implement the random purification channel $\mathcal{P}_{\mathcal{A}}$, we first measure the given state ρ using projectors $\{P_\lambda\}_{\lambda\in\Lambda}$, obtaining outcome λ and the (un-normalised) post-measurement state $P_\lambda \rho P_\lambda$ on $L_\lambda \otimes R_\lambda$. Then, we discard the state on register L_λ, prepare the maximally entangled state on $L_\lambda \otimes L'_\lambda$ and the maximally mixed state on register R'_λ.

2. Structured Sample Amplification

We now formally define the notion of sample amplification for distributions with a special structure induced by a function class. Namely, we consider the distributions of the form $(\mathcal{V}_n, f)$ over $\mathbb{Z}_2^n \times \mathbb{Z}_2$, where $\mathcal{V}_n$ is some fixed distribution over the n-bit input marginal, and where f is some Boolean function.

Definition 18 (Structured Sample Amplification) *Let $\mathcal{F}$ be a class of Boolean functions on n bits and let $\mathcal{V}_n$ be some distribution over the domain $\mathbb{Z}_2^n$. We say that the class of distributions*

$\mathcal{D} = \{D_f\}_{f\in\mathcal{F}} = \{(\mathcal{V}_n, f)\}_{f\in\mathcal{F}}$ *over the domain* $\mathcal{X} = \mathbb{Z}_2^n \times \mathbb{Z}_2$ admits a $(t, t+m, \epsilon)$-sample amplification scheme *if there exists a stochastic map* $T_{\mathcal{D},t,m,\epsilon} : \mathcal{X}^t \to \mathcal{X}^{t+m}$ *such that,*

$$\sup_{f\in\mathcal{F}} \mathrm{d_{TV}}\left(D_f^{\otimes t} \circ T^{-1}_{\mathcal{D},t,m,\epsilon}, D_f^{\otimes t+m}\right) \leq \epsilon\,, \tag{38}$$

where $\mathrm{d_{TV}}$ *denotes the total variation distance. The* minimax sample amplification error *for the distribution class* $\mathcal{D}$ *is defined by minimising the sample amplification error over all stochastic maps* $T_{SA} : \mathcal{X}^t \to \mathcal{X}^{t+m}$*,*

$$\epsilon^\star_{SA}(\mathcal{D}, t, t+m) := \min_{T_{SA}} \sup_{f\in\mathcal{F}} \mathrm{d_{TV}}\left(D_f^{\otimes t} \circ T^{-1}_{SA}, D_f^{\otimes t+m}\right). \tag{39}$$

This definition of structured sample amplification is simply that of general sample amplification as in Axelrod et al. (2020, 2024) with the additional assumption that the unknown distribution is of the form $(\mathcal{V}_n, f)$, where $\mathcal{V}_n$ is known and where $f \in \mathcal{F}$ for some known function class $\mathcal{F}$. Structured sample amplification can alternatively be thought of as a game between two parties, an amplifier and a distinguisher, where the distinguisher knows the distribution. Then, the minimax sample amplification error is the maximum advantage achievable by an adversarial distinguisher $\mathcal{A}$ in telling the true samples and amplified samples apart:

$$\epsilon^\star_{SA}(\mathcal{D}, t, t+1) = \min_{T_{SA}} \sup_{f\in\mathcal{F}} \max_{\mathcal{A}} \mathrm{Adv}(\mathcal{A}, \mathcal{V}_n, f, T_{SA}, t)\,, \tag{40}$$

where

$$\mathrm{Adv}(\mathcal{A}, f, T_{SA}, t) = \left|\mathop{\mathbf{Pr}}_{Z_{t+1}}[\mathcal{A}^f(Z_{t+1}) = 1] - \mathop{\mathbf{Pr}}_{Z_t}[\mathcal{A}^f(T_{SA}(Z_t)) = 1]\right|. \tag{41}$$

We omit the superscript f from the distinguisher, but it is to be understood that the distinguisher has knowledge of underlying Boolean function (and of the sample amplification procedure).

2.1. General Lower Bound

We denote the random sequence of t elements drawn independently from the distribution $(\mathcal{V}_n, f)$ by $Z_t = (Z_t^{(i)})_{i=1}^t = (X_t^{(i)}, Y_t^{(i)})_{i=1}^t$. A realization of this random variable will be denoted by $z_t = (z_t^{(i)})_{i=1}^t = (x_t^{(i)}, y_t^{(i)})_{i=1}^t$. For any function $f \in \mathcal{F}$, define $b_{\mathcal{F}}(f)$ as the minimum natural number t such that there exists at least one training dataset of size t which uniquely determines f in $\mathcal{F}$. That is,

$$b_{\mathcal{F}}(f) = \min\{t \in \mathbb{N} \mid \exists z_t \text{ s.t. } \forall f' \in \mathcal{F}\setminus\{f\}\ \exists 1 \leq i \leq t : y_t^{(i)} \neq f'(x_t^{(i)})\}\,. \tag{42}$$

Next, let $p^t_{\mathcal{F}}(f)$ be the probability of the random sequence Z_t determining f uniquely, i.e.,

$$p^t_{\mathcal{F}}(f) = \frac{\left|\{x_t \in (\{0,1\}^n)^t \mid z_t = (x_t, f(x_t)) \text{ uniquely determines } f\}\right|}{2^{nt}}\,. \tag{43}$$

Then, define

$$b_{\mathcal{F}} = \max_{f\in\mathcal{F}} b_{\mathcal{F}}(f)\,, \qquad p_{\mathcal{F}} = \min_{f\in\mathcal{F}} p^{b_{\mathcal{F}}}_{\mathcal{F}}(f)\,. \tag{44}$$

We mention that $b_{\mathcal{F}}(f)$ is the minimum size of a teaching sequence for the function f in $\mathcal{F}$. Consequently, $b_{\mathcal{F}}$ equals the teaching dimension for the function class $\mathcal{F}$ (Goldman and Kearns, 1995). Also, for uniform $\mathcal{V}_n$ we can understand $p^t_{\mathcal{F}}(f)$ as the fraction of sequence of length t that are teaching sequences for the function f in $\mathcal{F}$. Finally, we let $n_{\mathcal{F}}(z_t)$ be the number of hypotheses in $\mathcal{F}$ consistent with the sample z_t. That is,

$$n_{\mathcal{F}}(z_t) = |\{f \in \mathcal{F} \mid f(x_t^{(i)} = y_t^{(i)} \; \forall 1 \leq i \leq t\}| \,. \tag{45}$$

With this notation established, we can now state and proof our first main result.

Theorem 19 (Formal Statement of Theorem 3, Point 1: Error Lower Bound) *The minimax sample amplification error for the distribution class $\mathcal{D} = \{(\mathcal{V}_n, f)\}_{f\in\mathcal{F}}$ over the domain $\mathbb{Z}_2^n \times \mathbb{Z}_2$ in amplifying t samples to $t+1$ samples for any $t \leq b_{\mathcal{F}} - 1$ satisfies*

$$\epsilon^{\star}_{SA}(\mathcal{D}, t, t+1) \geq p_{\mathcal{F}}/2, \qquad \forall t \leq b_{\mathcal{F}} - 1. \tag{46}$$

Proof We can focus on the case $t = b_{\mathcal{F}} - 1$ for now due to monotonicity of sample amplification

$$\epsilon^{\star}_{SA}(\mathcal{D}, t_1, t_1+1) \geq \epsilon^{\star}_{SA}(\mathcal{D}, t_2, t_2+1), \qquad \forall t_1 \leq t_2. \tag{47}$$

This can be understood by noting that if there is a sample amplification procedure by one sample from t_1 samples, than this implies sample amplification from any $t_2 \geq t_1$ because we can use sample amplification on first t_1 samples out of t_2 to produce an extra sample. We prove this theorem using a distinguisher-based approach as mentioned in the Definition 18.

We use the simple distinguisher in Algorithm 1.

Algorithm 1 Distinguisher $\mathcal{A}$

Input: t samples, description of f.
Output: Accept or Reject
1. Run a consistent learning algorithm that outputs a uniformly random consistent hypothesis $\hat{f}$ from $\mathcal{F}$ using the t samples.
2. Accept (aka, output 1) if $\hat{f} = f$, else reject (aka, output 0).

Recall that the unknown distribution is of the form $(\mathcal{V}_n, f)$ for some $f \in \mathcal{F}$. Now, we denote by $Z_{b_{\mathcal{F}}(f)}$ the true $b_{\mathcal{F}}(f)$ samples and by $Z^{SA}_{b_{\mathcal{F}}(f)} = T_{SA}(Z_{b_{\mathcal{F}}(f)-1})$ the amplified samples, starting from $Z_{b_{\mathcal{F}}(f)-1}$. Then

$$\mathbf{Pr}\left[\mathcal{A}(Z_{b_{\mathcal{F}}(f)}) = 1\right] = \sum_{z_{b_{\mathcal{F}}(f)}} \mathbf{Pr}\left(z_{b_{\mathcal{F}}(f)}\right) \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)})} \,, \tag{48}$$

where $n_{\mathcal{F}}(Z_{b_{\mathcal{F}}(f)})$ is the number of hypotheses consistent with the data $Z_{b_{\mathcal{F}}(f)}$. Moreover, we mention that sample amplification cannot provide new information useful for learning which we argue in Proposition 20. Thus,

$$\mathbf{Pr}\left[\mathcal{A}(Z^{SA}_{b_{\mathcal{F}}(f)}) = 1\right] \overset{(a)}{\leq} \mathbf{Pr}\left[\mathcal{A}(Z_{b_{\mathcal{F}}(f)-1}) = 1\right] = \sum_{z_{b_{\mathcal{F}}(f)-1}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})} \,. \tag{49}$$

where (a) is proven in Proposition 20 below. Thus, the advantage of the distinguisher satisfies

$$\mathrm{Adv}(\mathcal{A}) \geq \sum_{z_{b_{\mathcal{F}}(f)}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)})} - \sum_{z_{b_{\mathcal{F}}(f)-1}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})}, \tag{50}$$

where we used that the difference is non-negative because $b_{\mathcal{F}}(f)$ samples are at least as good $b_{\mathcal{F}}(f)-1$ samples for a learner that samples randomly from the set of consistent hypotheses. Moreover, denote by $p_f^{b_{\mathcal{F}}(f)}$ (as mentioned earlier), the probability that $Z_{b_{\mathcal{F}}(f)}$ uniquely specifies f, then

$$\mathrm{Adv}(\mathcal{A}) \geq p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f) + \sum_{\substack{z_{b_{\mathcal{F}}(f)} \\ n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)})\geq 2}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)})} - \sum_{z_{b_{\mathcal{F}}(f)-1}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})}. \tag{51}$$

Now, we can group the two sums together as,

$$\mathrm{Adv}(\mathcal{A}) \geq p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f) + \sum_{z_{b_{\mathcal{F}}(f)-1}} \Bigg(\sum_{\substack{z_{b_{\mathcal{F}}(f)} \\ n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)})\geq 2 \\ z_{b_{\mathcal{F}}(f)}|_{b_{\mathcal{F}}(f)-1}=z_{b_{\mathcal{F}}(f)-1}}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)})} - \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})} \Bigg). \tag{52}$$

It is easy to see that $n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)}) \leq n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})$ for $z_{b_{\mathcal{F}}(f)}|_{b_{\mathcal{F}}(f)-1} = z_{b_{\mathcal{F}}(f)-1}$. Thus,

$$\mathrm{Adv}(\mathcal{A}) \geq p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f) + \sum_{z_{b_{\mathcal{F}}(f)-1}} \left(\sum_{\substack{z_{b_{\mathcal{F}}(f)} \\ n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)})\geq 2 \\ z_{b_{\mathcal{F}}(f)}|_{b_{\mathcal{F}}(f)-1}=z_{b_{\mathcal{F}}(f)-1}}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)}\right] - \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \right) \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})}. \tag{53}$$

Now, note that

$$\sum_{\substack{z_{b_{\mathcal{F}}(f)} \\ n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)})\geq 2 \\ z_{b_{\mathcal{F}}(f)}|_{b_{\mathcal{F}}(f)-1}=z_{b_{\mathcal{F}}(f)-1}}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)}\right] = \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \times \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)} \text{ with } z_{b_{\mathcal{F}}(f)}|_{b_{\mathcal{F}}(f)-1} = z_{b_{\mathcal{F}}(f)-1}, n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)}) \geq 2\right]. \tag{54}$$

Sightly abusing notation, we now denote by $q(z_{b_{\mathcal{F}}(f)-1})$ the probability that $z_{b_{\mathcal{F}}(f)}$ with $z_{b_{\mathcal{F}}(f)}|_{b_{\mathcal{F}}(f)-1} = z_{b_{\mathcal{F}}(f)-1}$, uniquely specifies f. It is easy to see that,

$$\sum_{z_{b_{\mathcal{F}}(f)-1}} q(z_{b_{\mathcal{F}}(f)-1}) = p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f). \tag{55}$$

Thus, $q(z_{b_{\mathcal{F}}(f)-1}) \leq p_f$ and hence, the probability that $z_{b_{\mathcal{F}}(f)}$ with first $b_{\mathcal{F}}(f)-1$ samples fixed as $z_{b_{\mathcal{F}}(f)-1}$ does not uniquely specify f is given as $1 - q(z_{b_{\mathcal{F}}(f)-1}) \geq 1 - p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f)$. Thus,

$$\sum_{\substack{z_{b_{\mathcal{F}}(f)} \\ n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)}) \geq 2 \\ z_{b_{\mathcal{F}}(f)} | b_{\mathcal{F}}(f)-1 = z_{b_{\mathcal{F}}(f)-1}}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)}\right] \geq \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \times (1 - p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f)). \tag{56}$$

Hence, we get

$$\begin{aligned} \mathrm{Adv}(\mathcal{A}) &\geq p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f) - \sum_{z_{b_{\mathcal{F}}(f)-1}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \left(1 - 1 + p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f)\right) \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})} \\ &= p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f) \left(1 - \sum_{z_{b_{\mathcal{F}}(f)-1}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})}\right). \end{aligned} \tag{57}$$

Now, note that $n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1}) \geq 2$ by definition of $b_{\mathcal{F}}(f)$, thus

$$\mathrm{Adv}(\mathcal{A}) \geq p_{\mathcal{F}}^{b_{\mathcal{F}}(f)}(f)/2. \tag{58}$$

Thus, the sample amplification error for any unknown distribution $(\mathcal{V}_n, f)$ from $b_{\mathcal{F}}(f)-1$ samples is lower bounded by $p_f^{b_{\mathcal{F}}(f)}/2$. Using Equation (40), we get that for any $f' \in \arg\max_{f \in \mathcal{F}} b_{\mathcal{F}}(f)$,

$$\epsilon^{\star}_{SA}(\mathcal{D}, b_{\mathcal{F}} - 1, b_{\mathcal{F}}) \geq p_{\mathcal{F}}^{b_{\mathcal{F}}}(f')/2 \geq \min_{f \in \mathcal{F}} p_{\mathcal{F}}^{b_{\mathcal{F}}}(f)/2 = p_{\mathcal{F}}/2, \tag{59}$$

where $b_{\mathcal{F}} = \max_{f \in \mathcal{F}} b_{\mathcal{F}}(f)$. To argue the same for any $t < b_{\mathcal{F}} - 1$, we assume that there is a sample amplification algorithm that, given $t' < b_{\mathcal{F}} - 1$ samples, produce $t'+1$ samples with error $\epsilon < p_{\mathcal{F}}/2$. Then this would contradict the case for $t = b_{\mathcal{F}} - 1$, since one could use this sample amplification algorithm on t' samples and append the rest of the samples unchanged to obtain $b_{\mathcal{F}}$ samples with sample amplification error $\epsilon < p_{\mathcal{F}}/2$. Thus,

$$\epsilon^{\star}_{SA}(\mathcal{D}, t, t+1) \geq \frac{p_{\mathcal{F}}}{2}, \quad t \leq b_{\mathcal{F}} - 1. \tag{60}$$

■

Proposition 20 *The probability of acceptance by the distinguisher defined in Algorithm 1 for $Z^{SA}_{b_{\mathcal{F}}(f)} = T_{SA}(Z_{b_{\mathcal{F}}(f)-1})$ is upper bounded as*

$$\mathbf{Pr}\left[\mathcal{A}(Z^{SA}_{b_{\mathcal{F}}(f)}) = 1\right] \leq \mathbf{Pr}\left[\mathcal{A}(Z_{b_{\mathcal{F}}(f)-1}) = 1\right]. \tag{61}$$

Proof Take the realisation $z_{b_{\mathcal{F}}(f)-1}$ of $Z_{b_{\mathcal{F}}(f)-1}$, and write $z^{SA}_{b_{\mathcal{F}}(f)} = T_{SA}(z_{b_{\mathcal{F}}(f)-1})$. The consistent learner in the distinguisher works by choosing a random consistent hypotheses for the given data and the distinguisher compares it to the original function. We argue that for a consistent learner, the amplified sample $z^{SA}_{b_{\mathcal{F}}(f)}$ cannot perform better than on $z_{b_{\mathcal{F}}(f)-1}$. To show this, we compare the number of consistent hypotheses for both cases.

The sample $z_{b_{\mathcal{F}}(f)-1}$ admits $n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})$ consistent hypotheses, none of which is preferred. Thus, if the sample amplifier tries to decrease the number of consistent hypotheses to $n_{\mathcal{F}}(z^{SA}_{b_{\mathcal{F}}(f)})$ by adding one sample, it can do so at best by choosing $n_{\mathcal{F}}(z^{SA}_{b_{\mathcal{F}}(f)})$ hypotheses out of $n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})$ uniformly randomly. Then, the probability that the true hypothesis is among these $n_{\mathcal{F}}(z^{SA}_{b_{\mathcal{F}}(f)})$ hypotheses is

$$\mathbf{Pr}\left[n_{\mathcal{F}}(Z^{SA}_{b_{\mathcal{F}}(f)})\text{random consistent hypotheses contain } f\right] \leq \frac{\binom{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})-1}{n_{\mathcal{F}}(z^{SA}_{b_{\mathcal{F}}(f)})-1}}{\binom{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})}{n_{\mathcal{F}}(z^{SA}_{b_{\mathcal{F}}(f)})}} = \frac{n_{\mathcal{F}}(z^{SA}_{b_{\mathcal{F}}(f)})}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})}. \tag{62}$$

Thus, the probability of success in learning from the amplified sample $z^{SA}_{b_{\mathcal{F}}(f)}$ by outputting a uniformly random consistent hypothesis is

$$\begin{aligned}\mathbf{Pr}\left[\mathcal{A}(Z^{SA}_{b_{\mathcal{F}}(f)}) = 1 \mid Z^{SA}_{b_{\mathcal{F}}(f)} = z^{SA}_{b_{\mathcal{F}}(f)} = T_{SA}(z_{b_{\mathcal{F}}(f)-1})\right] &\leq \frac{n_{\mathcal{F}}(z^{SA}_{b_{\mathcal{F}}(f)})}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})}\frac{1}{n_{\mathcal{F}}(z^{SA}_{b_{\mathcal{F}}(f)})} \\ &= \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})}.\end{aligned} \tag{63}$$

Now, to bound the overall acceptance probability of the distinguisher, we can take the sum over conditioned value above to get

$$\begin{aligned}&\mathbf{Pr}\left[\mathcal{A}(Z^{SA}_{b_{\mathcal{F}}(f)}) = 1\right] \\ &= \sum_{z^{SA}_{b_{\mathcal{F}}(f)}} \mathbf{Pr}\left[z^{SA}_{b_{\mathcal{F}}(f)}\right] \times \mathbf{Pr}\left[\mathcal{A}(Z^{SA}_{b_{\mathcal{F}}(f)}) = 1 \mid Z^{SA}_{b_{\mathcal{F}}(f)} = z^{SA}_{b_{\mathcal{F}}(f)}\right] \\ &\overset{(a)}{=} \sum_{z_{b_{\mathcal{F}}(f)-1}, T_{SA}} \mathbf{Pr}\left[T_{SA}(z_{b_{\mathcal{F}}(f)-1})\right] \times \mathbf{Pr}\left[\mathcal{A}(Z^{SA}_{b_{\mathcal{F}}(f)}) = 1 \mid Z^{SA}_{b_{\mathcal{F}}(f)} = z^{SA}_{b_{\mathcal{F}}(f)} = T_{SA}(z_{b_{\mathcal{F}}(f)-1})\right] \\ &\overset{(b)}{\leq} \sum_{z_{b_{\mathcal{F}}(f)-1}, T_{SA}} \mathbf{Pr}\left[T_{SA}(z_{b_{\mathcal{F}}(f)-1})\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})} \\ &\overset{(c)}{=} \sum_{z_{b_{\mathcal{F}}(f)-1}} \mathbf{Pr}\left[z_{b_{\mathcal{F}}(f)-1}\right] \frac{1}{n_{\mathcal{F}}(z_{b_{\mathcal{F}}(f)-1})} \\ &= \mathbf{Pr}\left[\mathcal{A}(Z_{b_{\mathcal{F}}(f)-1}) = 1\right]\end{aligned} \tag{64}$$

where (a) follows by decomposing the randomness in $z^{SA}_{b_{\mathcal{F}}(f)}$ into the randomness in $z_{b_{\mathcal{F}}(f)-1}$ and the intrinsic randomness of the sample amplifier; (b) follows from Equation (63); and (c) follows by summing over randomness of the amplifier. ■

We now generalize the above bound for sample amplification from any general t samples.

Theorem 21 (General Lower Bound) *Let $d \in \mathbb{N}$. The minimax sample amplification error for the distribution class, $\mathcal{D} = \{(\mathcal{V}_n, f)\}_{f\in\mathcal{F}}$ over the domain $\mathbb{Z}_2^n \times \mathbb{Z}_2$ in amplifying t samples to $t+1$*

samples for any $t \leq d-1$ *satisfies*

$$\epsilon^{\star}_{SA}(\mathcal{D}, t, t+1) \geq \max_{f \in \mathcal{F}} \left[\frac{p^{d}_{\mathcal{F}}(f)}{2} - p^{d-1}_{\mathcal{F}}(f) \right], \quad \forall t \leq d-1. \tag{65}$$

Proof The proof idea is the same as for Theorem 19 and uses the same distinguisher. We again start with the case of sample amplification from $d-1$ samples, and then we generalize to any $t < d-1$. In the case $t = d-1$ case, the advantage of the distinguisher will be

$$\begin{aligned}
\mathrm{Adv}(\mathcal{A}) &\geq \sum_{z_d} \frac{\mathbf{Pr}\left[z_d\right]}{n_{\mathcal{F}}(z_d)} - \sum_{z_{d-1}} \frac{\mathbf{Pr}\left[z_{d-1}\right]}{n_{\mathcal{F}}(z_{d-1})} \\
&\overset{(a)}{=} \sum_{z_{d-1}} \left(\frac{-\mathbf{Pr}\left[z_{d-1}\right]}{n_{\mathcal{F}}(z_{d-1})} + \sum_{\substack{z_d \\ z_d|_{d-1} = z_{d-1}}} \frac{\mathbf{Pr}\left[z_d\right]}{n_{\mathcal{F}}(z_d)} \right) \\
&\overset{(b)}{=} \sum_{\substack{z_{d-1} \\ n_{\mathcal{F}}(z_{d-1})=1}} \left(-\mathbf{Pr}\left[z_{d-1}\right] + \sum_{\substack{z_d \\ z_d|_{d-1} = z_{d-1}}} \mathbf{Pr}\left[z_d\right] \right) \\
&\quad + \sum_{\substack{z_{d-1} \\ n_{\mathcal{F}}(z_{d-1}) \geq 2}} \left(\frac{-\mathbf{Pr}\left[z_{d-1}\right]}{n_{\mathcal{F}}(z_{d-1})} + \sum_{\substack{z_d \\ z_d|_{d-1} = z_{d-1} \\ n_{\mathcal{F}}(z_d) \geq 2}} \frac{\mathbf{Pr}\left[z_d\right]}{n_{\mathcal{F}}(z_d)} \right) + \sum_{\substack{z_d \\ n_{\mathcal{F}}(z_d|_{d-1}) \geq 2 \\ n_{\mathcal{F}}(z_d)=1}} \mathbf{Pr}\left[z_d\right].
\end{aligned} \tag{66}$$

where (a) follows by again grouping using first $b_{\mathcal{F}}(f) - 1$ samples as previously; (b) follows by writing the sum over z_{d-1} in two parts, depending upon the value of $n_{\mathcal{F}}(z_{d-1})$ and then subdividing the case $n_{\mathcal{F}}(z_{d-1}) \geq 2$ according to whether adding an extra sample uniquely specifies f.

It is easy to see that the first term in the above expression is 0: if z_{d-1} specifies f uniquely, then adding a sample does not change the number of consistent hypotheses, thus we can just take the summation of probability over the last sample to obtain

$$\sum_{\substack{z_d \\ z_d|_{d-1} = z_{d-1} \\ n_{\mathcal{F}}(z_d)=1}} \mathbf{Pr}\left[z_d\right] = \mathbf{Pr}\left[z_{d-1}\right]. \tag{67}$$

Moreover, similarly to the previous case, if $z_d|_{d-1} = z_{d-1}$ then, $n_{\mathcal{F}}(z_d) \leq n_{\mathcal{F}}(z_{d-1})$. Hence, we can again bound the distinguishing advantage as,

$$\begin{aligned}
\mathrm{Adv}(\mathcal{A}) &\geq \sum_{\substack{z_{d-1} \\ n_{\mathcal{F}}(z_{d-1}) \geq 2}} \left(-\mathbf{Pr}\left[z_{d-1}\right] + \sum_{\substack{z_d \\ z_d|_{d-1} = z_{d-1} \\ n_{\mathcal{F}}(z_d) \geq 2}} \mathbf{Pr}\left[z_d\right] \right) \frac{1}{n_{\mathcal{F}}(z_{d-1})} + \sum_{\substack{z_d \\ n_{\mathcal{F}}(z_d|_{d-1}) \geq 2 \\ n_{\mathcal{F}}(z_d)=1}} \mathbf{Pr}\left[z_d\right] \\
&\overset{(a)}{\geq} - \sum_{\substack{z_{d-1} \\ n_{\mathcal{F}}(z_{d-1}) \geq 2}} \mathbf{Pr}\left[z_{d-1}\right] \left(1 - 1 + p^{d}_{\mathcal{F}}(f)\right) \frac{1}{n_{\mathcal{F}}(z_{d-1})} + \sum_{\substack{z_d \\ n_{\mathcal{F}}(z_d|_{d-1}) \geq 2 \\ n_{\mathcal{F}}(z_d)=1}} \mathbf{Pr}\left[z_d\right]
\end{aligned} \tag{68}$$

$$= \sum_{\substack{z_d \\ n_{\mathcal{F}}(z_d|_{d-1})\geq 2 \\ n_{\mathcal{F}}(z_d)=1}} \mathbf{Pr}\left[z_d\right] - p^d_{\mathcal{F}}(f) \sum_{\substack{z_{d-1} \\ n_{\mathcal{F}}(z_{d-1})\geq 2}} \mathbf{Pr}\left[z_{d-1}\right] \frac{1}{n_{\mathcal{F}}(z_{d-1})}.$$

where (a) follows in a similar fashion to the proof of Theorem 19 by noting that

$$\sum_{\substack{z_d \\ z_d|_{d-1}=z_{d-1} \\ n_{\mathcal{F}}(z_d)\geq 2}} \mathbf{Pr}\left[z_d\right] \geq \mathbf{Pr}\left[z_{d-1}\right](1 - p^d_{\mathcal{F}}(f)). \tag{69}$$

Now, note that

$$\sum_{\substack{z_d \\ n_{\mathcal{F}}(z_d|_{d-1})\geq 2 \\ n_{\mathcal{F}}(z_d)=1}} \mathbf{Pr}\left[z_d\right] = p^d_{\mathcal{F}}(f) - p^{d-1}_{\mathcal{F}}(f)\,, \tag{70}$$

because the sum on the left hand side is exactly the probability over d samples such that first $d-1$ samples do not specify the function uniquely but adding an extra sample helps in uniquely specifying the function. Plugging this back in, we get that the advantage of the distinguisher satisfies

$$\begin{aligned} \mathrm{Adv}(\mathcal{A}) &\geq p^d_{\mathcal{F}}(f) \left(1 - \sum_{\substack{z_{d-1} \\ n_{\mathcal{F}}(z_{d-1})\geq 2}} \mathbf{Pr}\left[z_{d-1}\right] \frac{1}{n_{\mathcal{F}}(z_{d-1})}\right) - p^{d-1}_{\mathcal{F}}(f) \\ &\geq \frac{p^d_{\mathcal{F}}(f)}{2} - p^{d-1}_{\mathcal{F}}(f). \end{aligned} \tag{71}$$

Hence, by Equation (40), the minimax sample amplification error satisfies

$$\epsilon^{\star}_{SA}(\mathcal{D}, d-1, d) \geq \max_{f\in\mathcal{F}} \left[\frac{p^d_{\mathcal{F}}(f)}{2} - p^{d-1}_{\mathcal{F}}(f)\right]. \tag{72}$$

To go from $d-1$ to any general $t \leq d-1$, we can argue just like in the proof of Theorem 19. Thus,

$$\epsilon^{\star}_{SA}(\mathcal{D}, t, t+1) \geq \max_{f\in\mathcal{F}} \left[\frac{p^d_{\mathcal{F}}(f)}{2} - p^{d-1}_{\mathcal{F}}(f)\right], \quad \forall t \leq d-1. \tag{73}$$

■

It is easy to see that the above bound generalizes the one provided in Theorem 19 by noticing that for $d = b_{\mathcal{F}}$, $p^{d-1}_{f'} = 0$ for $f' \in \arg\max_{f\in\mathcal{F}} b_{\mathcal{F}}(f)$. Hence,

$$\max_{f\in\mathcal{F}} \left[\frac{p^d_{\mathcal{F}}(f)}{2} - p^{d-1}_{\mathcal{F}}(f)\right] \geq \frac{p_{\mathcal{F}}}{2}. \tag{74}$$

We now apply the above lower bounds for structured sample amplification to the function class of parities and more generally to linear spaces of functions. We will omit the distribution class from the notation $\epsilon^{\star}_{SA}(\mathcal{D}, t, t+1)$ wherever it is clear from the context.

2.2. Parity Lower Bound

We study the structured sample amplification for the n-bit parity distributions $\mathcal{D}_{\text{par}} = \{(\mathcal{U}_n, f_a\}_{a\in\mathbb{Z}_2^n}$, where

$$f_a : \mathbb{Z}_2^n \to \mathbb{Z}_2,\ f_a(x) = a \cdot x\,, \tag{75}$$

and where $\mathcal{U}_n$ is the uniform distribution on $\mathbb{Z}_2^n$. We first study the sample amplification error for more general distributions which are uniformly supported over a n-dimensional subspace $\mathbb{Z}_2^m$. We look at a useful lemma first.

Lemma 22 *Given k uniformly random draws from the vector space $\mathbb{Z}_2^n$, the probability that exactly $1 \leq \alpha \leq k$ of them are linearly independent is given as*

$$\mathbf{Pr}(k,\alpha) = \frac{\prod_{i=0}^{\alpha-1}(2^n - 2^i)(2^k - 2^i)}{2^{nk}\prod_{i=0}^{\alpha-1}(2^\alpha - 2^i)}\,. \tag{76}$$

Thus, the probability of obtaining n linearly independent n-bit strings from n uniformly random draws is

$$\mathbf{Pr}(n,n) = \prod_{i=1}^{n}(1 - 2^{-i})\,. \tag{77}$$

The probability $\mathbf{Pr}(n,n)$ is lower bounded by the constant 0.28 for all n (Sloane, 1964; Finch, 2003). Using the techniques developed and lemma above, we provide the following constant lower bound on minimax sample amplification error for the distribution class $\mathcal{D}_{m,n}$.

Theorem 23 *Let $\mathcal{D}_{m,n}$ with $m \geq n$ denote the uniform distribution supported on an n-dimensional subspace of $\mathbb{Z}_2^m$. Then, the minimax sample amplification error for this class of distributions can be lower bounded as*

$$\epsilon^\star_{SA}(\mathcal{D}_{n,m}, t-1, t) \geq 0.14, \qquad \forall 1 \leq t \leq n\,. \tag{78}$$

Proof The teaching dimension for this class of distributions is clearly n and the teaching sequence is any sequence that contains n linearly independent elements. Thus, $b_{\mathcal{F}} = n$, Using Theorem 19, we conclude that

$$\epsilon^\star_{SA}(\mathcal{D}_{m,n}, t-1, t) \geq p/2, \qquad \forall 1 \leq t \leq n\,, \tag{79}$$

where p is the fraction of samples of size n that are teaching set for any fixed distribution, or equivalently, the probability of learning the distribution given n samples with uniformly random inputs. This is same as probability of obtaining n linearly independent samples in n uniformly random draws, which according to Lemma 22 is given as

$$p = \mathbf{Pr}(n,n) = \prod_{i=1}^{n}(1 - 2^{-i})\,. \tag{80}$$

Thus,

$$\epsilon^\star_{SA}(\mathcal{D}_{m,n}, t-1, t) \geq p/2 = \frac{\prod_{i=1}^{n}(1 - 2^{-i})}{2}, \qquad \forall 1 \leq t \leq n\,, \tag{81}$$

which is bounded from below by the constant 0.14 for any n according to the discussion after Lemma 22. ■

Now, we can view the parity distribution as a uniform distribution supported on a n dimensional subspace of $\mathbb{Z}_2^{n+1}$. Thus, we can obtain a minimax sample amplification error for class of n-bit parity distribution by noticing that $\mathcal{D}_{\text{par}} = \mathcal{D}_{n+1,n}$.

Corollary 24 (Formal Statement of Theorem 3, Point 2: Error Lower Bound for Sample Amplification of Parities) *Let $\mathcal{D}_{\text{par}}$ be the class of n-bit parity distribution as defined above. Then the minimax sample amplification error for this class of distributions can be lower bounded as*

$$\epsilon^{\star}_{SA}(\mathcal{D}_{\text{par}}, t-1, t) \geq 0.14, \qquad \forall 1 \leq t \leq n\,. \tag{82}$$

The above theorem rules out structured sample amplification of n-bit parity distributions to arbitrary small error from $t \leq n-1$ samples. This can be formalised as follows:

Corollary 25 *The class $\mathcal{D}_{\text{par}}$ of n-bit parity distributions does not admit a $(t-1, t, \epsilon)$ sample amplification procedure for any $t \leq n$ and $\epsilon < 0.14$.*

Notice that, since the TV distance between any two distinct parity distributions is a constant independent of n, the number of samples necessary and sufficient for learning an n-bit parity distribution to arbitrarily small error equals that of exact parity learning from uniformly random inputs, which is $\Theta(n)$. Thus, the above corollary shows that the sample complexity for sample amplification of parity distributions with a sufficiently small constant accuracy aymptotically matches that of learning the same class of distributions. In other words: Sample amplification is no easier than learning for parity distributions. (And, clearly, exact parity learning suffices for sample amplification, with an approximation error depending only on the failure probability of the exact learner.)

2.3. Coding Theory Lower Bounds

In this section, we highlight connections between structured sample amplification and coding theory. We first recall that, in principle, for any Boolean function class, we can define an associated linear codes (Definition 27). Then, we show that the minimax sample amplification error for structured distributions $\{(\mathcal{U}_n, f)\}$ with $f \in \mathcal{F}$ can be related to the ability of the corresponding dual code to correct random erasures.

We start with the following relation between the rank of a submatrix obtained by choosing random columns of parity check matrix of a code and the ability of the primal code to correct random erasures:

Lemma 26 (Abbe et al. (2015), Lemma 2.8) *For a parity check matrix H with N columns, take $S \subseteq [N]$, and denote by $\binom{[N]}{s}$ the collection of all s-element subsets of $[N]$. Let $H[S]$ be the restriction of H to columns indexed by elements in S, and let $x[S]$ be the restriction of string x to indices in S. Then, the set of bad s-erasure patterns is given as*

$$\left\{ S \in \binom{[N]}{s} : \exists x, y \in \ker(H), x \neq y, x[S^c] = y[S^c] \right\} = \left\{ S \in \binom{[N]}{s} : \operatorname{rank}(H[S]) < s \right\}. \tag{83}$$

Here, the set of bad s-erasures is the set of erasures at s many locations that cannot be corrected. In particular, the above lemma says that the fraction of bad s-erasures equals the probability that

the submatrix formed by randomly sampling s-columns of the parity check matrix is rank-deficient. Thus, if a code can correct d random erasures with high probability, then any d uniformly sampled columns from its parity check matrix are linearly independent with high probability. As in the parity case above, a high (or at least constant) probability of linear independence will be crucial in proving a sample amplification lower bound.

We first define the special class of codes that we will study in this subsection.

Definition 27 (Linear Function Codes) *Let $\mathcal{F}$ be a linear space of Boolean functions with dimension $k_{\mathcal{F}}$ and with a chosen basis $B_{\mathcal{F}} = \{e_{\mathcal{F},i}\}_{i=1}^{k_{\mathcal{F}}}$. We define the* linear function code $C_{\mathcal{F}}$ *with parameters $[2^n, k_{\mathcal{F}}]$ via the encoding map*

$$\textit{Enc} : \mathbb{Z}_2^{k_{\mathcal{F}}} \to \mathbb{Z}_2^n,\ \textit{Enc}(a) = (f_a(x))_{x\in\mathbb{Z}_2^n}\,, \tag{84}$$

where we define $f_a = \sum_{i=1}^{k_{\mathcal{F}}} a_i e_{\mathcal{F},i}$. That is, a $k_{\mathcal{F}}$-bit string is interpreted as a vector of basis coefficients and is encoded into the vector of evaluations of the corresponding function in $\mathcal{F}$.

The generator matrix $G(C_{\mathcal{F}})$ of such a code is of size $k_{\mathcal{F}} \times 2^n$. Its columns are indexed by elements in $\mathbb{Z}_2^n$, and its rows are indexed by basis functions in $e_{\mathcal{F},i}$. Then, $G(i,j) = e_{\mathcal{F},i}(x_j)$:

	x_0 $\cdots$ $\cdots$ x_{2^n-1}
$e_{\mathcal{F},1}$	
$e_{\mathcal{F},2}$	
$\vdots$	
$e_{\mathcal{F},i}$	$G(i,j) = e_{\mathcal{F},i}(x_j)$
$\vdots$	
$e_{\mathcal{F},k_{\mathcal{F}}}$	

We will denote the parameters of the code as $[2^n, k_{\mathcal{F}}, d]$ with their usual meanings. Moreover, we will denote the dual of the above code as $C_{\mathcal{F}}^{\perp}$, which is given by the kernel of the generator matrix $G(C_{\mathcal{F}})$. A familiar example of linear function codes is the (n, d) Reed-Muller code, where $\mathcal{F}$ is the class of degree-d polynomials over $\mathbb{Z}_2^n$, the basis $B_{\mathcal{F}}$ consists of all monomials of degree at most d, and the dimension is $k_{\mathcal{F}} = \binom{n}{\leq d}$. We now show that a certain resistance to random erasures in the code code $C_{\mathcal{F}}^{\perp}$ implies lower bounds for sample amplification of the distribution class $\mathcal{D} = \{(\mathcal{U}, f)\}_{f\in\mathcal{F}}$.

Theorem 28 (Formal Statement of Theorem 3, Point 3: Sample Amplification Lower Bounds From Coding Theory) *Let $C_{\mathcal{F}}$ be a $[2^n, k_{\mathcal{F}}, d]$ linear function code as defined above, and let $C_{\mathcal{F}}^{\perp}$ be its dual code. Then, the following hold:*

1. *If $C_{\mathcal{F}}^{\perp}$ can correct $k_{\mathcal{F}}$ random erasures with probability p_1 over the erasures, then the minimax sample amplification error for the distribution class $\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$ satisfies*

$$\epsilon_{SA}^{\star}(\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}, t-1, t) \geq p_1/2, \quad \forall 1 \leq t \leq k_{\mathcal{F}}\,. \tag{85}$$

2. *If $C_{\mathcal{F}}^{\perp}$ can correct $k_{\mathcal{F}} - 1$ random erasures with probability p_2 over the erasures, then the minimax sample amplification error for the distribution class $\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$ satisfies*

$$\epsilon_{SA}^{\star}(\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}, t-1, t) \geq p_2 \cdot d/2^{n+1}, \quad \forall 1 \leq t \leq k_{\mathcal{F}}. \tag{86}$$

To prove Theorem 28, we first prove two simple technical lemmas.

Lemma 29 *Let $C_{\mathcal{F}}$ be a $[2^n, k_{\mathcal{F}}, d]$ linear function code. Then, we have*

$$\Pr_{x \sim \mathcal{U}_n}[f(x) = 0] \leq 1 - \frac{d}{2^n}, \quad \forall f \in \mathcal{F}\,, \tag{87}$$

where $\mathcal{U}_n$ is the uniform distribution.

Proof We will denote the functions by $f_a(x) = a \cdot y(x)$, where $a \in \mathbb{Z}_2^{k_{\mathcal{F}}}$ denotes the coefficient vector and $y(x) = e_{\mathcal{F},1}(x) \dots e_{\mathcal{F},k_{\mathcal{F}}}(x)$ denotes the evaluation vector of all basis elements at input x. As the code has distance d by assumption, we have for any $a \neq b \in \mathbb{Z}_2^{k_{\mathcal{F}}}$,

$$\Pr_x[f_a(x) - f_b(x) \neq 0] = \Pr_x[f_a(x) \neq f_b(x)] \geq \frac{d}{2^n}, \tag{88}$$

where implicitly, the probability is over a uniform distribution. Now, notice that $f_a - f_b = f_{a-b} \in \mathcal{F}$ by linearity. Thus, the above implies that $\forall a, b \in \mathbb{Z}_2^{k_{\mathcal{F}}}$,

$$\Pr_x[f_{a-b}(x) \neq 0] \geq \frac{d}{2^n}. \tag{89}$$

Now, since this is true for any arbitrary choice of a and b, we get that,

$$\Pr_x[f_c(x) \neq 0] \geq \frac{d}{2^n}, \quad \forall c \in \mathbb{Z}_2^{k_{\mathcal{F}}}, \tag{90}$$

or,

$$\Pr_x[f_c(x) = 0] = 1 - \Pr_x[f_c(x) \neq 0] \leq 1 - \frac{d}{2^n}\,, \tag{91}$$

as claimed. ■

Lemma 30 *Let $C_{\mathcal{F}}$ be a $[2^n, k_{\mathcal{F}}, d]$ linear function code with generator matrix denoted as G such that,*

$$\mathbf{Pr}\left[S \in \binom{[2^n]}{k_{\mathcal{F}} - 1} : \mathrm{rank(G[S])} = k_{\mathcal{F}} - 1\right] = p\,. \tag{92}$$

Then

$$\mathbf{Pr}\left[S \in \binom{[2^n]}{k_f} : \mathrm{rank(G[S])} = k_{\mathcal{F}}\right] \geq p \cdot \frac{d}{2^n}\,. \tag{93}$$

Proof We define an indicator random variable X_m which takes the value 1 if the first m sampled columns are linearly independent. Then, the probability of interest is $P(X_{k_{\mathcal{F}}} = 1)$ which can be written as,

$$\mathbf{Pr}\left[X_{k_{\mathcal{F}}} = 1\right] = \mathbf{Pr}\left[X_{k_{\mathcal{F}}-1} = 1\right] \times \mathbf{Pr}\left[\text{new linearly independent sample} \mid X_{k_{\mathcal{F}}-1} = 1\right]. \tag{94}$$

Now, notice that given $k_{\mathcal{F}} - 1$ linearly independent samples $\{x_1, \dots x_{k_{\mathcal{F}}-1}\}$, we can define the orthogonal direction to these samples as a with $a \cdot x_i = 0, \forall i$. Now, the next sample is linearly

dependent iff it is orthogonal to a. We can consider function $f_a(x) = a \cdot x$, since $a \in \mathbb{Z}_2^{k_\mathcal{F}}$, and using Theorem 29, we have that

$$\Pr_x[f_a(x) = 0] \leq 1 - \frac{d}{2^n}. \tag{95}$$

Hence, the probability that the new sample will be linearly dependent is at most $1 - d/2^n$. Thus

$$\mathbf{Pr}\left[S \in \binom{[2^n]}{k_f} : \mathrm{rank(G[S])} \geq k_\mathcal{F}\right] \geq p \cdot \frac{d}{2^n} = \mathbf{Pr}\left[X_{k_\mathcal{F}} = 1\right] \geq p \cdot \frac{d}{2^n}. \tag{96}$$

■

We are not well-equipped to prove Theorem 28.

Proof [Proof of Theorem 28] We first prove bullet point 1. Note that if $C_\mathcal{F}^\perp$ can correct $k_\mathcal{F}$ random erasures with probability p_1, then from Lemma 26 we have

$$\Pr_S\left[S \in \binom{[2^n]}{k_\mathcal{F}} : \mathrm{rank}(G[S]) = k_\mathcal{F}\right] = p_1, \tag{97}$$

where G is the parity check matrix of $C_\mathcal{F}^\perp$, i.e., the generator matrix of $C_\mathcal{F}$. So, if we sample columns of G uniformly at random, then with probability p_1, they are linearly independent. Interpreting functions in $\mathcal{F}$ as computing parities of vectors of basis function evaluations, we see that to uniquely identify any function in $\mathcal{F}$, it is necessary and sufficient to see its values on inputs that index $k_\mathcal{F}$ many linearly independent columns of G. Hence, we can reinterpret the above as $b_\mathcal{F} = k_\mathcal{F}$ and $p_\mathcal{F} = p_1$. We can now appeal to Theorem 19 and obtain

$$\epsilon_{SA}^\star(\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}, t-1, t) \geq p_1/2, \qquad \forall 1 \leq t \leq k_\mathcal{F}. \tag{98}$$

For the second part, again using Lemma 26, we first note that if $C_\mathcal{F}^\perp$ can correct $k_\mathcal{F} - 1$ random erasures with probability p_2, then

$$\Pr_S\left[S \in \binom{[2^n]}{k_\mathcal{F}-1} : \mathrm{rank}(G[S]) = k_\mathcal{F} - 1\right] = p_2. \tag{99}$$

Then, using Lemma 30, we get that the probability

$$\Pr_S\left[S \in \binom{[2^n]}{k_\mathcal{F}} : \mathrm{rank}(G[S]) = k_\mathcal{F}\right] \geq p_2 \cdot \frac{d}{2^n}, \tag{100}$$

So, using Theorem 19, we have

$$\epsilon_{SA}^\star(\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}, t-1, t) \geq p_2 \cdot d/2^{n+1}, \qquad \forall 1 \leq t \leq k_\mathcal{F}. \tag{101}$$

■

The above theorem has two consequences, one for coding theory and one for the problem of sample amplification. On one hand, if there exists a linear function code $C_\mathcal{F}$ with parameters $[2^n, k_\mathcal{F}, d]$ with constant fractional distance $d/2^n = \Theta(1)$ which can correct $k_\mathcal{F} - 1$ erasures with very high probability, the corresponding distribution class $\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$ cannot be sample amplified with error less than a constant threshold, starting with less than $k_\mathcal{F}$ samples. On the other hand, if there exists a function class for which $\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$ distribution is easy to sample amplify even for $t \leq k_\mathcal{F}$, say the error being bounded from above by δ which decays with n, then the corresponding code can correct $k_\mathcal{F}$ erasures only with probability at most 2δ.

Corollary 31 *Let $\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$ be a distribution class such that the minimax sample amplification error,*

$$\epsilon^{\star}_{SA}(\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}, k_{\mathcal{F}} - 1, k_{\mathcal{F}}) \leq \delta(n) \tag{102}$$

where δ is some function of n. Then, the linear function code $C_{\mathcal{F}}$ can correct $k_{\mathcal{F}}$ erasure only with probability atmost $2\delta(n)$ or can correct $k_{\mathcal{F}} - 1$ erasures with probability at most $2^{n+1}\delta(n)/d$.

Proof The proof is immediate from Theorem 28. ■

2.4. Sample Amplification for the Class of all Boolean Functions

In the previous sections, we saw examples function classes which structured sample amplification (with small error) is hard as learning. This provides evidence that when enough structure is present, the best strategy is to learn. In this section we provide a simple example of class of distributions based on Boolean functions for which sample amplification is strictly easier than learning in the sample complexity sense.

We consider the distribution class $\mathcal{D} = \{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}_{\text{all}}}$ where $\mathcal{F}_{\text{all}}$ is the class of all Boolean functions over $\mathbb{Z}_2^n$. Note that our lower bound technique fails to provide a meaningful lower bound in this case because the teaching dimension for $\mathcal{F}_{\text{all}}$ is 2^n and because $p_{\mathcal{F}}$ is exponentially small. In fact, we can use the result for sample amplification of any discrete distribution (Axelrod et al., 2020) to sample amplify $\mathcal{F}_{\text{all}}$.

Corollary 32 *Let $\mathcal{F}$ be the class of all Boolean functions $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$. Then, the class of distributions $\{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}}$ admits a $(t, t+1, \epsilon)$ sample amplification scheme with $t = \Theta\left(\frac{\sqrt{2^n}}{\epsilon}\right)$.*

Proof As the class of distributions $\mathcal{D} = \{(\mathcal{U}_n, f)\}_{f\in\mathcal{F}_{\text{all}}}$ consists of distributions with support size 2^n, this immediately follows from (Axelrod et al., 2020, Theorem 1). ■

In particular, the complexity of sample amplification for $\mathcal{F}_{\text{all}}$ is quadratically better than the complexity of learning the same class, which by coupon collector is $\widetilde{\Theta}(2^n)$.

3. Structured Cloning

We can define structured quantum cloning analogously as structured sample amplification.

Definition 33 (Structured Quantum Cloning) *Let $\mathcal{S}$ be a class of n-qubit states. Then, $\mathcal{S}$ is said to admit $(t, t+m, \epsilon)$-quantum cloning scheme if there exists a CPTP map $\Lambda_{S,t,m,\epsilon} : \mathcal{B}((\mathbb{C}^{2^n})^{\otimes t}) \to \mathcal{B}((\mathbb{C}^{2^n})^{\otimes t+m})$ such that,*

$$\sup_{\rho\in\mathcal{S}} \mathrm{d}_{\mathrm{TD}}\left(\Lambda_{\mathcal{S},t,m,\epsilon}(\rho^{\otimes t}), \rho^{\otimes t+m}\right) \leq \epsilon, \tag{103}$$

and the optimal cloning error for the class of states $\mathcal{S}$ is defined by minimisung the cloning error over all CPTP maps,

$$\epsilon^{\star}_{\mathrm{Cl}}(\mathcal{S}, t, t+m) := \min_{\Lambda} \sup_{\rho\in\mathcal{S}} \mathrm{d}_{\mathrm{TD}}\left(\Lambda(\rho^{\otimes t}), \rho^{\otimes t+m}\right). \tag{104}$$

We consider classes of states that have symmetries. Let G be an Abelian group, and let Ψ_H^ϵ be the set of states that ϵ-hides the subgroup $H \leq G$. Then we consider the cloning of $\mathcal{S}_\alpha = \bigcup_{H,|H|=\alpha} \Psi_H^\epsilon$, i.e., states that are promised to hide an unknown hidden subgroup of fixed order. Here, we do not make any assumptions about the purity of the states. Thus, $\mathcal{S}_\alpha$ also contains mixed states, which is relevant for our lower bound proof below. As mentioned earlier, our cloning lower bound technique is based on learning error and requires sample complexity lower bounds for learning that are tight up to additive constants. We show that for $\mathcal{S}_\alpha$, the sample complexity of learning can be determined up to additive constants by considering the task of learning the hidden subgroup. To argue this, we identify the optimal measurement for the Abelian state hidden subgroup problem in the next section.

3.1. Optimality of Character POVMs for a Class of Abelian StateHSPs

We show that in the worst case sense, character POVMs are optimal up to classical post-processing for solving a class of Abelian StateHSP. We show this optimality for Abelian StateHSPs with non-isomorphic irreducible representations. To show this, we use the following result (Wright, 2016).

Proposition 34 (Wright (2016), Proposition 1.2.7) *Suppose ρ is block diagonal with blocks corresponding to the (known) orthogonal projectors $\{\Pi_i\}_i$. Then,*

- *One can without loss of generality pre-process any quantum algorithm that acts on ρ by measuring $\{\Pi_i\}_i$ on ρ.*
- *If ρ is a multiple of the identity in each block, then $\{\Pi_i\}_i$ is an optimal measurement for extracting information from ρ.*

The proof of this result is simple. It is easy to see that, assuming a block diagonal structure, we can always perform the projective measurement onto the blocks without losing any information. Moreover, if the state is proportional to the maximally mixed state within each block then we cannot get any more information after identifying the block.

Let us now sketch how we will use Theorem 34 before presenting the detailed proof. For an Abelian StateHSP $(G, \mu, \Psi_H^\epsilon)$ and for any $\rho \in \Psi_H^\epsilon$, we consider the state

$$\sigma = \frac{1}{|G|} \sum_{g \in G} \mu(g) \rho \mu(g)^\dagger \,. \tag{105}$$

We will argue that this state is block-diagonal with respect to the character POVM $\{\Pi_\lambda\}_\lambda$. Then, using the non-isomorphic property of irreducible representations, we will argue that each block is one-dimensional and hence the above state is also maximally mixed in each block. We can combine this with Proposition 34 to obtain the optimality of character POVMs on σ. However, this optimality result is limited, it only applies to algorithms that process single copies of σ at a time, but it does not account for general multi-copy measurements. To show this optimality among this more general class of algorithms, we consider $\sigma^{\otimes t}$, which is block diagonal with respect to tensor powers of single-copy character POVMs. As each block is still one-dimensional, this will lead to the optimality of single-copy character POVMs even when multi-copy measurements are allowed. Intuitively, the optimality can be understood by noting that allowing for multiple i.i.d. copies of

σ is equivalent to allowing non-i.i.d. states from Ψ_H^ϵ, and in this scenario it is easy to see that single-copy character POVMs are optimal.

The rest of this subsection makes the above reasoning precise to establish the following result:

Theorem 35 (Optimality of Character POVMs for Abelian StateHSP) *Let $(G, \mu, \Psi_H^\epsilon)$ be an Abelian StateHSP problem such that the unitary representation μ has non-isomorphic irreducible representations. Then, the character POVM $\{\Pi_\lambda\}_\lambda$ is the optimal procedure to solve the Abelian StateHSP in the worst-case sense.*

Proof Let t be the number of copies provided. We assume that we are allowed any arbitrary t-copy measurements on instances of Ψ_H^ϵ, i.e., we are allowed to make measurements on $\rho^{\otimes t}$ with $\rho \in \Psi_H^\epsilon$. Now, for any $\rho \in \Psi_H^\epsilon$, define

$$\sigma = \frac{1}{|G|} \sum_{g \in G} \mu(g) \rho \mu(g)^\dagger. \tag{106}$$

Claim 36 *Let $\rho \in \Psi_H^\epsilon$, for an underlying Abelian Group G with the unitary representation μ and let*

$$\sigma = \frac{1}{|G|} \sum_{g \in G} \mu(g) \rho \mu(g)^\dagger, \tag{107}$$

then,

1. $\sigma \in \Psi_H^\epsilon$,
2. $[\sigma, \mu(g)] = 0, \forall g \in G$.

Proof See Section 4. ■

As a consequence of bullet point 1 in Claim 36, σ is also a valid instance of the Abelian StateHSP. So, for the purposes of a worst-case analysis, we can assume that we are given $\sigma^{\otimes t}$. As a consequence of bullet point 2 of Claim 36, σ has the same eigenvectors as $\mu(g)$, i.e., σ is diagonal in the basis $\{|\lambda_i, v_{\lambda_i}\rangle = |\lambda_i\rangle\}_i$, where we can drop the v_{λ_i} because the irreducible representations are by assumption non-isomorphic. Now, we define the t-copy POVM $\{\bigotimes_i^t \Pi_{\lambda_i}\}$, with character POVM elements $\Pi_{\lambda_i} = |\lambda_i\rangle\langle\lambda_i|$, which because of the assumption on non-isomorphic irreducible representations are rank-one projections. Note that:

1. $\sigma^{\otimes t}$ is diagonal with respect to the basis projected on by the POVM elements $\{\bigotimes_{i=1}^t \Pi_{\lambda_i}\}$.
2. As $\sigma^{\otimes t}$ is block-diagonal with blocks of size 1, it is proportional to the identity in each block.

Hence, by Theorem 34, the POVM $\{\bigotimes_{i=1}^t \Pi_{\lambda_i}\}$ is optimal to solve the Abelian StateHSP in the worst case sense. ■

In the case of the phaseless stabilizer StateHSP, we have non-isomorphic irreducible representations. This yields the optimality of Bell sampling for phaseless stabilizer StateHSP.

3.2. Structured Random Purification

The random purification channel allows to transform t i.i.d. copies of a mixed states into t i.i.d. copies of a random purification of the mixed state (Tang et al., 2025; Girardi et al., 2025a) .Below, we provide a structured version of the random purification map for a class of Abelian StateHSP problems with non-isomorphic irreducible representations. Namely, we give a CPTP map that transforms i.i.d. copies of any mixed state that is diagonal with respect to the character POVMs to i.i.d. copies of a random purification.

Theorem 37 (Structured Random Purification) *Let $\sigma \in \Psi_H^\epsilon$ be a mixed state instance of an Abelian StateHSP $(G, \mu, \Psi_H^\epsilon)$ with non-isomorphic irreducible representations. Assume that $\sigma = \sum_\lambda a_\lambda |\lambda\rangle\langle\lambda|$ is diagonal with respect to the character POVM $\{\Pi_\lambda = |\lambda\rangle\langle\lambda|\}_\lambda$. Then, for any t, there is a CPTP map C^t such that*

$$C^t(\sigma^{\otimes t}) = \mathop{\mathbf{E}}_g |\sigma_g\rangle\langle\sigma_g|^{\otimes t}\,, \tag{108}$$

*where $|\sigma_g\rangle = (I \otimes \mu(g))|\sigma\rangle$ with $|\sigma\rangle = \sum_\lambda \sqrt{a_\lambda}|\lambda\rangle \otimes |\lambda\rangle$ is a purification of σ for any $g \in G$. We call $|\sigma_g\rangle$ the g-*purification *of σ.*

Theorem 37 will allow us to map copies of the worst case instance for Abelian StateHSP constructed in the proof of Theorem 35 to copies of a random purification. This will enable us to infer pure Abelian StateHSP lower bounds from the mixed Abelian StateHSP lower bound of Theorem 37.

Proof We begin by rewriting our target state:

$$\mathop{\mathbf{E}}_g |\sigma_g\rangle\langle\sigma_g|^{\otimes t} \tag{109}$$

$$= \mathop{\mathbf{E}}_g \sum_{\substack{\lambda_1\dots\lambda_t \\ \eta_1,\dots\eta_t}} \sqrt{a_{\lambda_1}\dots a_{\lambda_t}}\sqrt{a_{\eta_1}\dots a_{\eta_t}}|\lambda_1\dots\lambda_t\rangle\langle\eta_1\dots\eta_t| \otimes \mu(g)^{\otimes t}|\lambda_1\dots\lambda_t\rangle\langle\eta_1\dots\eta_t|(\mu(g)^\dagger)^{\otimes t} \tag{110}$$

$$= \sum_{\substack{\lambda_1\dots\lambda_t \\ \eta_1,\dots\eta_t}} \sqrt{a_{\lambda_1}\dots a_{\lambda_t}}\sqrt{a_{\eta_1}\dots a_{\eta_t}}|\lambda_1\dots\lambda_t\rangle\langle\eta_1\dots\eta_t| \otimes \mathop{\mathbf{E}}_g \mu(g)^{\otimes t}|\lambda_1\dots\lambda_t\rangle\langle\eta_1\dots\eta_t|(\mu(g)^\dagger)^{\otimes t} \tag{111}$$

$$= \sum_{\bar\lambda,\bar\eta} \sqrt{a_{\bar\lambda}a_{\bar\eta}}|\bar\lambda\rangle\langle\bar\eta| \otimes \mathop{\mathbf{E}}_g \mu(g)^{\otimes t}|\bar\lambda\rangle\langle\bar\eta|(\mu(g)^\dagger)^{\otimes t}\,. \tag{112}$$

Now, observe that

$$\mathop{\mathbf{E}}_g \mu(g)^{\otimes t}|\bar\lambda\rangle\langle\bar\eta|(\mu(g)^\dagger)^{\otimes t} = \delta_{\bar\lambda\cong\bar\eta}|\bar\lambda\rangle\langle\bar\eta|\,. \tag{113}$$

This can be seen by noticing that $\mu(g)^{\otimes t}|\bar\lambda\rangle = \chi_{\bar\lambda}(g)|\bar\lambda\rangle$ and thus we can use Schur's orthonormality condition (Lemma 11) to obtain

$$\mathop{\mathbf{E}}_g \mu(g)^{\otimes t}|\bar\lambda\rangle\langle\bar\eta|\mu(g)^{\otimes t,\dagger} = \mathop{\mathbf{E}}_g \chi_{\bar\lambda}(g)\overline{\chi_{\bar\eta}(g)}|\bar\lambda\rangle\langle\bar\eta| = \delta_{\bar\lambda\cong\bar\eta}|\bar\lambda\rangle\langle\bar\eta|\,. \tag{114}$$

Thus, we obtain

$$\mathop{\mathbf{E}}_g |\sigma_g\rangle\langle\sigma_g|^{\otimes t} = \sum_{\substack{\bar\lambda,\bar\eta \\ \bar\lambda\cong\bar\eta}} \sqrt{a_{\bar\lambda}a_{\bar\eta}}|\bar\lambda\rangle\langle\bar\eta| \otimes |\bar\lambda\rangle\langle\bar\eta|\,. \tag{115}$$

The character of the representation $\bar{\lambda} = \lambda_1 \dots \lambda_t$ is $\chi_{\bar{\lambda}} = \chi_{\lambda_1} \dots \chi_{\lambda_t}$. Hence, because we assumed the irreducible representations of μ to be non-isomorphic, and using that two representations are isomorphic if and only if they have the same characters, we see that $\bar{\lambda} \cong \bar{\eta}$ holds if and only if $\bar{\lambda}$ and $\bar{\eta}$ are equal up to a permutation, i.e., $\bar{\eta} = \pi\bar{\lambda}$ for some $\pi \in S_t$. In this case, clearly $a_{\bar{\lambda}} = a_{\bar{\eta}}$. Hence, we can write

$$\mathop{\mathbf{E}}_{g} |\sigma_g\rangle\langle\sigma_g|^{\otimes t} = \sum_{\{\bar{\lambda}\}} f(\{\bar{\lambda}\}) \frac{a_{\{\bar{\lambda}\}}}{f(\{\bar{\lambda}\})} \sum_{\substack{\bar{\eta},\bar{\eta}' \\ \bar{\eta}\cong\bar{\lambda} \\ \bar{\eta}'\cong\bar{\lambda}}} |\bar{\eta}\bar{\eta}\rangle\langle\bar{\eta}'\bar{\eta}'| = \sum_{\{\bar{\lambda}\}} f(\{\bar{\lambda}\}) a_{\{\bar{\lambda}\}} \rho_{\{\bar{\lambda}\}}\,, \tag{116}$$

where we have used $\{\bar{\lambda}\}$ to denote a representative for each equivalence class of isomorphic tensor product irreps, where $f(\{\bar{\lambda}\})$ denotes the number of distinct permutations of the tuple $\bar{\lambda}$, and where we have defined the state

$$\rho_{\{\bar{\lambda}\}} = \frac{1}{f(\{\bar{\lambda}\})} \sum_{\substack{\bar{\eta},\bar{\eta}' \\ \bar{\eta}\cong\bar{\lambda} \\ \bar{\eta}'\cong\bar{\lambda}}} |\bar{\eta}\bar{\eta}\rangle\langle\bar{\eta}'\bar{\eta}'|\,. \tag{117}$$

We define the POVM $\{\Pi_{\{\bar{\lambda}\}}\}_{\{\bar{\lambda}\}}$, where

$$\Pi_{\{\bar{\lambda}\}} = \sum_{\substack{\bar{\eta} \\ \bar{\eta}\cong\bar{\lambda}}} |\bar{\eta}\rangle\langle\bar{\eta}|. \tag{118}$$

With this, we can describe our structured random purification procedure

Algorithm 2 Structured Random Purification Channel C^t

Input: $\sigma^{\otimes t}$, where σ is a mixed state instance of the Abelian StateHSP $(G, \mu, \Psi_H^\epsilon)$
Assumption: The Abelian StateHSP $(G, \mu, \Psi_H^\epsilon)$ has non-isomorphic irreducible representations, and σ is block diagonal with respect to these irreducible representations.
Output: $\mathbf{E}_g |\sigma_g\rangle\langle\sigma_g|^{\otimes t}$
1. Measure $\sigma^{\otimes t}$ in the basis $\{\Pi_{\{\bar{\lambda}\}}\}_{\{\bar{\lambda}\}}$.
2. Then, based on the observed outcome $\{\bar{\lambda}\}$, prepare the state $\rho_{\{\bar{\lambda}\}}$.

We now verify that indeed $C^t(\sigma^{\otimes t}) = \mathbf{E}_g |\sigma_g\rangle\langle\sigma_g|^{\otimes t}$. First, notice that the measurement outcome in first step is $\{\bar{\lambda}\}$ with probability $a_{\{\bar{\lambda}\}} f(\{\bar{\lambda}\})$ because

$$\begin{aligned}
\operatorname{tr}\left(\Pi_{\{\bar{\lambda}\}} \sigma^{\otimes t}\right) &= \operatorname{tr}\left(\Pi_{\{\bar{\lambda}\}} \sum_{\bar{\eta}} a_{\bar{\eta}} |\bar{\eta}\rangle\langle\bar{\eta}|\right) \\
&= \operatorname{tr}\left(\sum_{\bar{\eta}_i\cong\bar{\lambda}} \sum_{\bar{\eta}'} a_{\bar{\eta}'} \delta_{\bar{\eta}\bar{\eta}'} |\bar{\eta}'\rangle\langle\bar{\eta}|\right) \\
&= \sum_{\bar{\eta}\cong\bar{\lambda}} a_{\bar{\eta}} \\
&= f(\{\bar{\lambda}\}) a_{\{\bar{\lambda}\}}\,.
\end{aligned} \tag{119}$$

Thus, when applying our structured random purification procedure to $\sigma^{\otimes t}$, in Step 1 we observe the outcome $\{\bar{\lambda}\}$ with probability $f(\{\bar{\lambda}\})a_{\{\bar{\lambda}\}}$, and then in Step 2 prepare the state $\rho_{\{\bar{\lambda}\}}$. Thus, we have shown

$$C^t(\sigma^{\otimes t}) = \sum_{\{\bar{\lambda}\}} f(\{\bar{\lambda}\})a_{\{\bar{\lambda}\}}\rho_{\{\bar{\lambda}\}} \overset{(a)}{=} \mathop{\mathbf{E}}_{g} |\sigma_g\rangle\langle\sigma_g|^{\otimes t}, \tag{120}$$

where (a) used Equation (116). ■

Remark 38 *Our structured random purification channel can also be obtained from Walter and Witteveen (2025)'s random purification channel for general symmetries. Namely, for a particular choice of block diagonal states and corresponding symmetries, the random purification channel of Walter and Witteveen (2025) gives exactly our structured random purification channel. For any Abelian StateHSP $(G, \mu, \Psi_H^{\epsilon})$ and any t (copies), we take the algebras,*

$$\mathcal{A} \cong \bigoplus_{\{\bar{\lambda}\}\in\Lambda_t} b_{\{\bar{\lambda}\}} I_{\{\bar{\lambda}\}} \quad \text{and} \quad \mathcal{A}' \cong \bigoplus_{\{\bar{\lambda}\}\in\Lambda_t} \sigma_{\{\bar{\lambda}\}}, \tag{121}$$

where $\bar{\lambda} = \lambda_1 \dots \lambda_t$ with each λ_i being the irreducible representation of the unitary representation μ, $\{\bar{\lambda}\}$ being the equivalence class of $\bar{\lambda}$ under any t-permutations, and $b_{\{\bar{\lambda}\}}$ being any complex numbers. Essentially, we are looking at states that are block diagonal with each block characterised by $\{\bar{\lambda}\}$ and basis states $\{|\pi\bar{\lambda}\rangle\}_{\pi\in S_t}$. Then, t copies of the diagonal state $\sigma = \sum_\lambda a_\lambda |\lambda\rangle\langle\lambda|$ satisfy $\sigma^{\otimes t} \in \mathcal{A}'$. Thus, using Theorem 17, we get that there is a channel $\mathcal{P}_{\mathcal{A}}$ such that

$$\mathcal{P}_{\mathcal{A}}(\sigma^{\otimes t}) = \bigoplus_{\{\bar{\lambda}\}\in\Lambda_t} \frac{|\tilde{\Phi}_0\rangle_{L_{\{\bar{\lambda}\}}L'_{\{\bar{\lambda}\}}}\langle\tilde{\Phi}_0|_{L_{\{\bar{\lambda}\}}L'_{\{\bar{\lambda}\}}}}{\dim L_{\{\bar{\lambda}\}}} \otimes \mathrm{tr}_{L_{\{\bar{\lambda}\}}}\left[P_{\{\bar{\lambda}\}}\sigma^{\otimes t}P_{\{\bar{\lambda}\}}\right] \otimes \frac{I_{R'_{\{\bar{\lambda}\}}}}{\dim R'_{\{\bar{\lambda}\}}}. \tag{122}$$

We now recognise that $P_{\{\bar{\lambda}\}}$ is the projection on block $\{\bar{\lambda}\}$ and thus corresponds to $\Pi_{\{\bar{\lambda}\}}$ in Equation (118), $\dim L_{\{\bar{\lambda}\}} = f(\{\bar{\lambda}\})$ is the number of distinct permutations of $\bar{\lambda}$, and

$$\frac{|\tilde{\Phi}_0\rangle_{L_{\{\bar{\lambda}\}}L'_{\{\bar{\lambda}\}}}\langle\tilde{\Phi}_0|_{L_{\{\bar{\lambda}\}}L'_{\{\bar{\lambda}\}}}}{\dim L_{\{\bar{\lambda}\}}} = \frac{1}{f(\{\bar{\lambda}\})} \sum_{\substack{\bar{\eta},\bar{\eta}' \\ \bar{\eta}\cong\bar{\lambda} \\ \bar{\eta}'\cong\bar{\lambda}}} |\bar{\eta}\bar{\eta}\rangle\langle\bar{\eta}'\bar{\eta}'| = \rho_{\{\bar{\lambda}\}}. \tag{123}$$

Moreover, $R_{\{\bar{\lambda}\}}$ and $R'_{\{\bar{\lambda}\}}$ are just one-dimensional spaces. Thus, using Equation (119) and Equation (123), we get that

$$\mathcal{P}_{\mathcal{A}}(\sigma^{\otimes t}) = \bigoplus_{\{\bar{\lambda}\}\in\Lambda_t} f(\{\bar{\lambda}\})a_{\{\bar{\lambda}\}}\rho_{\{\bar{\lambda}\}}. \tag{124}$$

Comparing it with Equation (120), we get that

$$\mathcal{P}_{\mathcal{A}}(\sigma^{\otimes t}) = \mathop{\mathbf{E}}_{g} |\sigma_g\rangle\langle\sigma_g|^{\otimes t} = C^t(\sigma^{\otimes t}). \tag{125}$$

Therefore, the Algorithm 2 gives exactly the channel output $\mathcal{P}_{\mathcal{A}}(\sigma^{\otimes t})$. This completes the remark.

We now define the states

$$\sigma_H = \frac{1}{|H^\perp|} \sum_{\lambda \in H^\perp} |\lambda\rangle\langle\lambda| \,. \tag{126}$$

where $H \leq G$ is a subgroup of G with the dual $H^\perp$. It is easy to see that these states hide the subgroup H with $\epsilon = 1$.

Lemma 39 *For σ_H states defined above,*

1. *if $g \in H$, $\mathrm{tr}\left(\mu(g)\sigma_H\right) = 1$, and*
2. *if $g \in G \setminus H$, $\mathrm{tr}\left(\mu(g)\sigma_H\right) = 0$.*

Proof The proof is provided in Section 4. ■

Next, we show that all g-purifications corresponding to σ_H as defined above for any H are instances of a suitably constructed Abelian StateHSP. To this end, we consider the Abelian group $G \times G$, and we consider the unitary representation

$$\begin{aligned} &\mu' : G \times G \to \mathrm{GL}(V \otimes V \otimes V \otimes V)\,, \\ &\mu'(g,h) = \mu(g) \otimes \mu(h)^\dagger \otimes \mu(g)^\dagger \otimes \mu(h). \end{aligned} \tag{127}$$

Lemma 40 *The two-copy g-purifications $|\sigma_{H,g}\rangle^{\otimes 2}$, where*

$$|\sigma_{H,g}\rangle = (I \otimes \mu(g)) \left(\frac{1}{\sqrt{|H^\perp|}} \sum_{\lambda \in H^\perp} |\lambda\rangle \otimes |\lambda\rangle \right) \tag{128}$$

(as defined in Theorem 37) purifies the mixed state σ_H for some H, are instances of the Abelian StateHSP $(G \times G, \mu', \Psi^{\epsilon'}_{H'})$ for a suitable $\epsilon' > 0$, where the hidden subgroup H' is given by

$$H' = \{(g_1, g_2) \mid \chi_{\lambda_1}(g_1)\overline{\chi_{\lambda_2}(g_1)\chi_{\lambda_1}(g_2)}\chi_{\lambda_2}(g_2) = 1, \forall \lambda_1, \lambda_2 \in H^\perp\}\,, \tag{129}$$

and with

$$\Psi^{\epsilon'}_{H'} = \bigcup_{g \in G} \{|\sigma_{H,g}\rangle\}\,. \tag{130}$$

Proof We note that $\mu'(g_1, g_2)$ acting on some g-purification gives

$$\begin{aligned} &\mu'(g_1,g_2)|\sigma_{H,g}\rangle^{\otimes 2} \\ &\quad = (\mu(g_1) \otimes \mu(g_2)^\dagger |\sigma_{H,g}\rangle) \otimes (\mu(g_1)^\dagger \otimes \mu(g_2)|\sigma_{H,g}\rangle) \\ &\quad \overset{(a)}{=} ((I \otimes \mu(g))(\mu(g_1) \otimes \mu(g_2)^\dagger)|\sigma_H\rangle) \otimes ((I \otimes \mu(g))(\mu(g_1)^\dagger \otimes \mu(g_2))|\sigma_H\rangle) \\ &\quad = (I \otimes \mu(g))^{\otimes 2} \left(\sum_{\lambda_1} \frac{1}{\sqrt{|H^\perp|}} \mu(g_1)|\lambda_1\rangle \otimes \mu(g_2)^\dagger|\lambda_1\rangle \otimes \sum_{\lambda_2} \frac{1}{\sqrt{|H^\perp|}} \mu(g_1)^\dagger|\lambda_2\rangle \otimes \mu(g_2)|\lambda_2\rangle \right) \\ &\quad \overset{(b)}{=} (I \otimes \mu(g))^{\otimes 2} \left(\sum_{\lambda_1,\lambda_2} \frac{1}{|H^\perp|} \chi_{\lambda_1}(g_1)|\lambda_1\rangle \otimes \overline{\chi_{\lambda_1}(g_2)}|\lambda_1\rangle \otimes \overline{\chi_{\lambda_2}(g_1)}|\lambda_2\rangle \otimes \chi_{\lambda_2}(g_2)|\lambda_2\rangle \right), \end{aligned} \tag{131}$$

where (a) follows from Abelianity of G, and where (b) follows from noting that the $|\lambda\rangle$ are eigenvectors of the representation $\mu(g)$ with the eigenvalues given by characters. From this, if $(g_1, g_2) \in H'$, we see directly from the definition of H' that

$$\mu'(g_1, g_2)|\sigma_{H,g}\rangle^{\otimes 2} = |\sigma_{H,g}\rangle^{\otimes 2}, \qquad \forall g, g_1, g_2 \in G. \tag{132}$$

In contrast, if $(g_1, g_2) \notin H'$, we get

$$\begin{aligned} \left|\langle\sigma_{H,g}|^{\otimes 2}\mu'(g_1, g_2)|\sigma_{H,g}\rangle^{\otimes 2}\right| &= \left|\langle\sigma_H|^{\otimes 2}\mu'(g_1, g_2)|\sigma_H\rangle^{\otimes 2}\right| \\ &= \left|\sum_{\lambda_1,\lambda_2} \frac{1}{|H^\perp|^2}\chi_{\lambda_1}(g_1)\overline{\chi_{\lambda_2}(g_1)\chi_{\lambda_1}(g_2)}\chi_{\lambda_2}(g_2)\right| \\ &\overset{(a)}{<} 1. \end{aligned} \tag{133}$$

Here, (a) follows by Cauchy-Schwarz inequality $\langle v, w\rangle \leq \|v\| \cdot \|w\|$ being strict unless v and w are linearly dependent. In our case, the vectors are $v = (\chi_{\lambda_1}(g_1)\overline{\chi_{\lambda_2}(g_1)})_{\lambda_1,\lambda_2\in H^\perp}$ and $w = (\chi_{\lambda_1}(g_2)\overline{\chi_{\lambda_2}(g_2)})_{\lambda_1,\lambda_2\in H^\perp}$. They have norm 1 because $|\chi_\lambda(g)| = 1$ for all g and λ, and they are not linearly dependent if $(g_1, g_2) \notin H'$ because, by definition of H', there exists at least one pair $(\lambda_1, \lambda_2) \in H^\perp$ such that

$$\chi_{\lambda_1}(g_1)\overline{\chi_{\lambda_2}(g_1)\chi_{\lambda_1}(g_2)}\chi_{\lambda_2}(g_2) \neq 1. \tag{134}$$

We can thus define

$$\epsilon'_H = 1 - \max_{g_1,g_2\in(G\times G)\setminus H'} \left|\langle\sigma_H|^{\otimes 2}\mu'(g_1, g_2)|\sigma_H\rangle^{\otimes 2}\right| > 0\,, \tag{135}$$

and (since we consider only finitely many H) also

$$\epsilon' = \min_H \epsilon'_H > 0\,, \tag{136}$$

so that for any arbitrary H, the g-purifications are instances of the Abelian StateHSP $(G\times G, \mu', \Psi^{\epsilon'}_{H'})$. ■

3.3. Cloning Lower Bounds for Abelian StateHSPs

We are now well-equipped to prove cloning lower bounds for particular classes of mixed and pure Abelian StateHSP. Our approach is similar to the one that led to the lower bound for structured sample amplification (Theorem 19): Solving an Abelian StateHSP problem is equivalent to collecting independent generators of the dual group $H^\perp$ of the hidden subgroup H. By Theorem 35, for an Abelian StateHSP (assuming non-isomorphic irreps), the optimal way to collect these generators is via character POVM measurements. Hence, $n_{H^\perp} - 1$ copies do not suffice to obtain $n^\perp_H$ independent generators of $H^\perp$. However, if we had a quantum cloner for that generates an extra copy from initially $n_{H^\perp} - 1$ copies of an arbitrary instance of the Abelian StateHSP, then we could measure the character POVM on the extra copy to obtain a new independent generator with not-too-small probability. Thus no such cloner can exist.

To formalise the above argument, we consider the set of states

$$\mathcal{S}_\alpha = \bigcup_{H\leq G:|H|=\alpha} \Psi_H^\epsilon\,, \tag{137}$$

for some fixed α. That is, we consider the (mixed and pure) state instances of an Abelian StateHSP with Abelian group G, representation μ, and some hidden subgroup $H \leq G$ of order α. For a given state ρ with hidden subgroup $H \leq G$, we refer to $n_{H^\perp}$ as the number of independent generators of H, and we let p_ρ be the success probability of learning H given $n_{H^\perp}$ copies of the state ρ. Also, we define

$$t_{G,\alpha} = \max_{H\leq G:|H|=\alpha} n_{H^\perp}. \tag{138}$$

Moreover, we take the states σ_H as defined in Equation (126). Then, we define

$$\mathcal{S}_\alpha^\sigma = \bigcup_{H\leq G,|H|=\alpha} \{\sigma_H\}. \tag{139}$$

It is easy to see that $\mathcal{S}_\alpha^\sigma \subseteq \mathcal{S}_\alpha$.

Theorem 41 (Formal Statement of Theorem 4, Point 1: Error Lower Bounds for Cloning of Abelian StateHSP) *Let G be an Abelian group, let μ be a representation of G with non-isomorphic irreps. Then, for any $t \leq t_{G,\alpha}$, the optimal error in cloning from $t-1$ copies to t copies for the class of states $\mathcal{S}_\alpha$ is lower bounded as*

$$\epsilon^\star_{\mathrm{Cl}}(\mathcal{S}_\alpha, t-1, t) \geq \epsilon^\star_{\mathrm{Cl}}(\mathcal{S}_\alpha^\sigma, t-1, t) \geq \min_{\sigma_H\in\mathcal{S}_\alpha^\sigma} p_{\sigma_H}/2\,, \qquad \forall t \leq t_{G,\alpha}\,, \tag{140}$$

where the maximum is over all states σ_H defined in Equation (126) for $H \leq G$ with $|H| = \alpha$.

Proof Let Λ be any quantum cloning map for the states $\mathcal{S}_\alpha$. We first consider cloning from $t-1 = t_{G,\alpha}-1$ copies, which we can later generalize to any $t-1 \leq t_{G,\alpha}-1$ copies. Notice that $\mathcal{S}_\alpha^\sigma \subseteq \mathcal{S}_\alpha$, thus

$$\epsilon^\star_{\mathrm{Cl}}(\mathcal{S}_\alpha, t-1, t) \geq \epsilon^\star_{\mathrm{Cl}}(\mathcal{S}_\alpha^\sigma, t-1, t). \tag{141}$$

Then, given a $\sigma_H \in \mathcal{S}_\alpha^\sigma$, we compare

1. true copies, $\sigma_H^{\otimes t_{G,\alpha}}$, and
2. cloned copies, $\Lambda(\sigma_H^{\otimes(t_{G,\alpha}-1)})$.

We define two probability distribution over $(H^\perp)^{\times t_{G,\alpha}}$ as

$$q_{\Lambda(\sigma_H^{\otimes(t_{G,\alpha}-1)})}(\lambda_1,\ldots,\lambda_{t_{G,\alpha}}) = \mathrm{Tr}\left(\Lambda(\sigma_H^{\otimes(t_{G,\alpha}-1)}) \bigotimes_{i=1}^{t_{G,\alpha}} \Pi_{\lambda_i}\right), \tag{142}$$

and

$$q_{\sigma_H^{\otimes t_{G,\alpha}}}(\lambda_1,\ldots,\lambda_{t_{G,\alpha}}) = \mathrm{Tr}\left(\sigma_H^{\otimes t_{G,\alpha}} \bigotimes_{i=1}^{t_{G,\alpha}} \Pi_{\lambda_i}\right), \tag{143}$$

where $\{\Pi_\lambda\}_\lambda$ is the character POVM. We now again use a distinguisher based-approach, noting that the optimal cloning error can be rewritten as

$$\epsilon^\star_{\mathrm{Cl}}(\mathcal{S}^\sigma_\alpha, t_{G,\alpha}-1, t_{G,\alpha}) = \min_{\Lambda} \sup_{\rho\in\mathcal{S}^\sigma_\alpha} \max_{\mathcal{A}} \mathrm{Adv}(\mathcal{A}, H, \Lambda, t_{G,\alpha})\,. \tag{144}$$

We consider the following distinguisher:

Algorithm 3 Distinguisher $\mathcal{A}$

Input: An $n\cdot t$ qubit state $\sigma_H^{\otimes t}$, classical description of H
Output: Accept or Reject
1. Measure the character POVM $\{\Pi_\lambda\}_\lambda$ on every copy of σ_H.
2. Run a consistent learning algorithm that outputs one of the consistent subgroups $\hat{H}$ uniformly at random using the t measurement samples from the first step.
2. Accept if $\hat{H}=H$, else Reject.

For the distinguisher defined above, the distinguishing advantage is

$$\mathrm{Adv}(\mathcal{A}) = \left| \underset{q\ \sigma_H^{\otimes t_{G,\alpha}}}{\mathbf{Pr}} \left[\mathcal{A}\left(\lambda_1,\dots\lambda_{t_{G,\alpha}}\right)=1\right] - \underset{q\ \Lambda(\sigma_H^{\otimes(t_{G,\alpha}-1)})}{\mathbf{Pr}} \left[\mathcal{A}\left(\lambda_1,\dots\lambda_{t_{G,\alpha}}\right)=1\right] \right|, \tag{145}$$

where the λ_i are sampled from the respective distributions. We argue that since the single copy character POVMs are optimal for learning H correctly, the probability for specifying H correctly from cloned $t_{G,\alpha}$ copies can be upper-bounded by the probability for specifying the hidden subgroup from $t_{G,\alpha}-1$ copies analogous to Proposition 20:

$$\underset{q\ \Lambda(\sigma_H^{\otimes t_{G,\alpha}-1})}{\mathbf{Pr}} \left[\mathcal{A}\left(\lambda_1,\dots\lambda_{t_{G,\alpha}}\right)=1\right] \leq \underset{q\ \sigma_H^{\otimes(t_{G,\alpha}-1)}}{\mathbf{Pr}} \left[\mathcal{A}\left(\lambda_1,\dots\lambda_{t_{G,\alpha}}\right)=1\right]\,. \tag{146}$$

Then, the advantage of the distinguisher in distinguishing the true and cloned copies is lower bounded as

$$\mathrm{Adv}(\mathcal{A}) \geq \underset{q\ \sigma_H^{\otimes t_{G,\alpha}}}{\mathbf{Pr}} \left[\mathcal{A}\left(\lambda_1,\dots\lambda_{t_{G,\alpha}}\right)=1\right] - \underset{q\ \sigma_H^{\otimes(t_{G,\alpha}-1)}}{\mathbf{Pr}} \left[\mathcal{A}\left(\lambda_1,\dots\lambda_{t_{G,\alpha}}\right)=1\right]. \tag{147}$$

Notice that the above difference on the right hand side can alternatively be thought of as the difference in probability of specifying a function f uniquely from $t_{G,\alpha}$ samples with that of from $t_{G,\alpha}-1$ samples. Thus, we can use similar calculation as used in the proof of Theorem 19 to lower bound this by $p_{\sigma_H}/2$. Using Equation (144), we get that for any $H'\in \arg\max_{H\leq G,|H|=\alpha} n_{H^\perp}$,

$$\epsilon^\star_{\mathrm{Cl}}(\mathcal{S}^\sigma_\alpha, t_{G,\alpha}-1, t_{G,\alpha}) \geq p_{\sigma_H}/2 \geq \min_{\sigma_H\in\mathcal{S}^\sigma_\alpha} p_{\sigma_H}/2\,, \tag{148}$$

where p_{σ_H} is the probability of learning the hidden subgroup H from $t_{G,\alpha}$ copies of the state. This gives

$$\epsilon^\star_{\mathrm{Cl}}(\mathcal{S}_\alpha, t_{G,\alpha}-1, t_{G,\alpha}) \geq \epsilon^\star_{\mathrm{Cl}}(\mathcal{S}^\sigma_\alpha, t_{G,\alpha}-1, t_{G,\alpha}) \geq \min_{\sigma_H\in\mathcal{S}^\sigma_\alpha} p_{\sigma_H}/2. \tag{149}$$

We can then extend this to a lower bound for cloning from $t-1\leq t_{G,\alpha}-1$ to t copies via Lemma 42 below. ■

Lemma 42 *If a family of states $\mathcal{S}$ does not admit a $(k-1,k,\delta)$-quantum cloning scheme for some ϵ, then it also does not admit a $(t-1,t,\epsilon)$-quantum cloning scheme for any $t \leq k$.*

Proof We prove this by contradiction. Assume that Λ is a $(t-1,t,\epsilon)$ cloner for the family of states $\mathcal{S}$ with error ϵ for some $t \leq k$. This means that, for any $\rho \in \mathcal{S}$,

$$\mathrm{d_{TD}}(\Lambda(\rho^{\otimes t-1}), \rho^{\otimes t}) \leq \epsilon. \tag{150}$$

Define Λ' as $\Lambda' = \Lambda \otimes I^{\otimes k-t}$. Then,

$$\begin{aligned} \mathrm{d_{TD}}\left(\Lambda'(\rho^{\otimes k-1}), \rho^{\otimes k}\right) &= \mathrm{d_{TD}}\left(\Lambda(\rho^{\otimes t-1}) \otimes \rho^{\otimes k-t}, \rho^{\otimes t} \otimes \rho^{\otimes k-t}\right) \\ &\overset{(a)}{=} \mathrm{d_{TD}}(\Lambda(\rho^{\otimes t-1}), \rho^{\otimes t}) \leq \epsilon\,. \end{aligned} \tag{151}$$

Thus, Λ' is a $(k-1,k,\epsilon)$-cloner for the family of states $\mathcal{S}$, contradicting our assumption that no such cloner exists. ■

The proof of Theorem 41 works for mixed state Abelian StateHSP, more precisely when we allow $\mathcal{S}_\alpha$ to contain the mixed states $\mathcal{S}_\alpha^\sigma$. We can lift the cloning lower bound to the pure Abelian StateHSP $(G \times G, \mu', \Psi_{H'}^{\epsilon'})$ as defined in the previous section. Here, we consider the set of states

$$\mathcal{S}'_\alpha = \bigcup_{H'} \Psi_{H'}^{\epsilon'}\,, \tag{152}$$

where the union is over all subgroups of $G \times G$ of the form H' as in Equation (129) for some subgroup $H \leq G$ of order α.

Corollary 43 *Let G be an Abelian group, let μ be a representation of G with non-isomorphic irreps. Then, for any $t \leq t_{G,\alpha}$, the optimal error in cloning from $t-1$ copies to t copies for the class of states $\mathcal{S}'_\alpha$ is lower bounded as*

$$\epsilon^\star_{\mathrm{Cl}}(\mathcal{S}'_\alpha, t-1, t) \geq \min_{\sigma_H \in \mathcal{S}_\alpha^\sigma} p_{\sigma_H}/2\,, \qquad \forall t \leq t_{G,\alpha}\,. \tag{153}$$

The intuition for obtaining this lower bound from the lower bound for cloning $\mathcal{S}_\alpha$ is easy to understand: One way of cloning a mixed state from $\mathcal{S}_\alpha^\sigma$ is to first apply the structured random purification channel, yielding a pure state from $\mathcal{S}'_\alpha$, then to apply a cloner for $\mathcal{S}'_\alpha$, and finally to trace out the auxiliary systems.

Proof We prove this by contradiction. Suppose there exists a $(t-1,t,\epsilon)$-quantum cloning scheme Λ for $\mathcal{S}'_\alpha$ for some $t \leq t_{G,\alpha}$ and $\epsilon < \max_{\sigma_H \in \mathcal{S}_\alpha^\sigma} p_{\sigma_H}/2$. Then, for the g-purifications of any states in $\mathcal{S}_\alpha^\sigma$, the cloner satisfies

$$\mathrm{d_{TD}}(\Lambda(|\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t-1}), |\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t}) \leq \epsilon, \qquad \forall g, H\,. \tag{154}$$

Define the map $\Lambda' = \mathrm{tr}_{PR}\,\Lambda \circ C^{t-1}$, where tr_{PR} is the partial trace over the purifying registers, and where C^{t-1} is the structured random purification channel for $t-1$ copies from Theorem 37. Then,

$$\mathrm{d_{TD}}\left(\Lambda'(\sigma_H^{\otimes t-1}), \sigma_H^{\otimes t}\right) \leq \mathrm{d_{TD}}\left(\Lambda \circ C^{t-1}(\sigma_H^{\otimes t-1}), \mathop{\mathbf{E}}_g |\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t}\right)\,, \tag{155}$$

since the partial trace is a CPTP map and thus cannot increase trace distance. We can further upper bound the trace distance as

$$\begin{aligned}\mathrm{d_{TD}}\left(\Lambda\circ C^{t-1}(\sigma_H^{\otimes t-1}),\mathop{\mathbf{E}}_g|\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t}\right) &= \mathrm{d_{TD}}\left(\Lambda(\mathop{\mathbf{E}}_g|\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t-1}),\mathop{\mathbf{E}}_g|\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t}\right)\\ &\overset{(a)}{=} \mathrm{d_{TD}}\left(\mathop{\mathbf{E}}_g\Lambda(|\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t-1}),\mathop{\mathbf{E}}_g|\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t}\right)\\ &\overset{(b)}{\leq} \mathop{\mathbf{E}}_g\left(\Lambda(|\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t-1}),|\sigma_{H,g}\rangle\langle\sigma_{H,g}|^{\otimes t}\right)\leq\epsilon\,,\end{aligned} \tag{156}$$

where (a) follows from linearity of Λ, (b) follows by triangle inequality, and the final inequality is Equation (154). Thus, Λ' is a $(t-1,t,\epsilon)$-cloner for states in $\mathcal{S}_\alpha^\sigma$, which contradicts the cloning error lower bound in Theorem 41. ■

The above results provide a lower bound on sample complexity for cloning the states in $\mathcal{S}_\alpha$ and $\mathcal{S}'_\alpha$ with arbitrarily small error.

Corollary 44 *Let G be an Abelian group, let μ be a representation of G with non-isomorphic irreps. Then, the sets of states $\mathcal{S}_\alpha$ and $\mathcal{S}'_\alpha$ defined above do not admit $(t,t+1,\epsilon)$-quantum cloning schemes for $t\leq t_{G,\alpha}$ and $\epsilon<p_\sigma/2$.*

3.4. Cloning Lower Bound for Stabilizer States

We now turn our attention from the general setting of cloning Abelian StateHSP instances to the concrete question of cloning stabilizer states. We consider the following kind of $(2n)$-qubit mixed states,

$$\sigma_L=\frac{1}{\sqrt{2^n}}\sum_{y\in L^\perp}|\Phi_y\rangle\langle\Phi_y|, \tag{157}$$

where $|\Phi_x\rangle$ are Bell states as defined earlier, L as some n-dimensional subspace of $\mathbb{Z}_2^{2n}$. Then, as mentioned previously, we consider the mixed phaseless stabilizer StateHSP as $(\mathbb{Z}_2^{2n},\mu,\Psi_L^1)$, where Ψ_L^1 is the singleton set

$$\Psi_L^1=\{\sigma_L\}, \tag{158}$$

and with the unitary representation $x\mapsto V_x^{\otimes 2}$, with V_x as defined in Equation (23). Then, we consider the class states $\mathcal{S}_n$ defined as

$$\mathcal{S}_n=\bigcup_{L\leq\mathbb{Z}_2^{2n},|L|=2^n}\Psi_L^1=\{\sigma_L\}_{L\leq\mathbb{Z}_2^{2n},|L|=2^n}. \tag{159}$$

Note that

$$t_{G,n}=n\,. \tag{160}$$

Since the states σ_L are block-diagonal with respect to irreps (Bell basis), and since the representation has non-isomorphic irreps, measuring in the Bell basis is optimal from Theorem 35. As a

consequence of this optimality, the probability of learning the hidden subspace L and hence the hidden subgroup is given by Lemma 22 as,

$$p_\sigma = \prod_{i=1}^{n}(1-2^{-i}) \geq 0.28. \tag{161}$$

The following corollary is now immediate from Theorem 41 when noting that $\mathcal{S}_n = \mathcal{S}_n^\sigma$ in this case.

Corollary 45 *The set of states $\mathcal{S}_n$ does not admit a $(t-1,t,\epsilon)$-quantum cloning scheme for $t \leq n$ and $\epsilon \leq 0.14$.*

We now show our cloning lower bound for stabilizer states. To obtain this from Theorem 45, we prove that for states in $\mathcal{S}_n$, the g-purifications are $(4n)$-qubit stabilizer states. This allows us to "lift" the cloning lower bound for $\mathcal{S}_n$ to a cloning lower bound for $(4n)$-qubit stabilizer states.

Lemma 46 *Suppose $\sigma = \sigma_L$ as in Equation (157) for some n-dimensional subspace L. Then, the g-purifications $|\sigma_g\rangle$ of σ (defined using Theorem 37) are $(4n)$-qubit stabilizer states.*

Proof Let L be an n-dimensional subspace of $\mathbb{Z}_2^{2n}$. As $\sigma = \sigma_L$ is diagonal in the Bell basis and supported uniformly on 2^n elements, the corresponding base-purification takes the form

$$|\sigma\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in L^\perp} |\Phi_x\rangle \otimes |\Phi_x\rangle \,. \tag{162}$$

We first show that this base purification $|\sigma\rangle$ is a $(4n)$-qubit stabilizer state. From the discussion in Section A.3, we can obtain,

$$(V_x \otimes I)|\Phi_y\rangle = (-1)^{b\cdot c}|\Phi_{x\oplus y}\rangle. \tag{163}$$

where $x = (a,b)$ and $y = (c,d)$. Thus, on the level of two copies,

$$(V_x \otimes I)^{\otimes 2}|\Phi_y\rangle^{\otimes 2} = |\Phi_{x\oplus y}\rangle^{\otimes 2}. \tag{164}$$

Now, we make a choice of n basis vectors for $L^\perp$ as $\{u_i\}_i^n$, and of n basis vectors of L, $\{h_i\}_i^n$. We then extend the basis $\{h_i\}_i^n$ to a basis for $\mathbb{Z}_2^{2n}$, $\{h_i\}_i^n \cup \{s_i\}_i^n$. Then, we define four sets

$$U = \{V_u \otimes I \otimes V_u \otimes I \mid u \in \{u_i\}_i^n\}\,, \tag{165}$$

$$W_1 = \{V_h \otimes V_h \otimes I^{\otimes 2} \mid h \in \{h_i\}_i^n\}\,, \tag{166}$$

$$W_2 = \{I^{\otimes 2} \otimes V_h \otimes V_h \mid h \in \{h_i\}_i^n\}\,, \tag{167}$$

and

$$S = \{V_s^{\otimes 4} \mid s \in \{s_i\}_i^n\}\,. \tag{168}$$

Then, notice that for any $V \in U$ described by $u' \in \{u_i\}_{i=1}^n$,

$$\begin{aligned} V|\sigma\rangle &= (V_{u'} \otimes I \otimes V_{u'} \otimes I)|\sigma\rangle \\ &= \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} (V_{u'} \otimes I)|\Phi_u\rangle \otimes (V_{u'} \otimes I)|\Phi_u\rangle \\ &\overset{(a)}{=} \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} |\Phi_{u\oplus u'}\rangle \otimes |\Phi_{u\oplus u'}\rangle \\ &\overset{(b)}{=} \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} |\Phi_u\rangle \otimes |\Phi_u\rangle = |\sigma\rangle\,, \end{aligned} \tag{169}$$

where for (a) follows from Equation (164), and where (b) is a simple reindexing of the summation by noting that $L^\perp$ is closed under addition. Thus, elements of U stabilize $|\sigma\rangle$.

Next, any $V \in W_1 \cup W_2$ can be written as $V = V_{h_i}^{\otimes 2} \otimes V_{h_j}^{\otimes 2}$ for $h_i, h_j \in \{h_k\}_{k=1}^n$, and thus

$$\begin{aligned} V|\sigma\rangle &= (V_{h_i}^{\otimes 2} \otimes V_{h_j}^{\otimes 2})|\sigma\rangle \\ &= \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} V_{h_i}^{\otimes 2}|\Phi_u\rangle \otimes V_{h_j}^{\otimes 2}|\Phi_u\rangle \\ &\overset{(c)}{=} \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} \chi_u(h_i)|\Phi_u\rangle \otimes \chi_u(h_j)|\Phi_u\rangle \\ &\overset{(d)}{=} \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} |\Phi_u\rangle \otimes |\Phi_u\rangle = |\sigma\rangle\,, \end{aligned} \tag{170}$$

where (c) follows by noting that $V_{h_i}^{\otimes 2} = \mu(h_i)$ is the underlying unitary representation and thus $V_{h_i}^{\otimes 2}|\Phi_u\rangle = \chi_u(h_i)|\Phi_u\rangle$ since $|\Phi_u\rangle$ are one-dimensional irreps of the representation (eigenstates) with character (eigenvalue) given by $\chi_u(h_i) = (-1)^{[h_1,u]}$, and where (d) follows by noting that for $h \in L$, we have $\chi_u(h) = 1$ for any $u \in L^\perp$ by assumption.

Finally, take $V \in S$ described by $s \in \{s_i\}_{i=1}^n$. Then

$$\begin{aligned} V|\sigma\rangle &= V_s^{\otimes 4}|\sigma\rangle \\ &= \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} V_s^{\otimes 2}|\Phi_u\rangle \otimes V_s^{\otimes 2}|\Phi_u\rangle \\ &\overset{(e)}{=} \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} \chi_u(s)|\Phi_u\rangle \otimes \chi_u(s)|\Phi_u\rangle \\ &\overset{(f)}{=} \frac{1}{\sqrt{2^n}} \sum_{u \in L^\perp} |\Phi_u\rangle \otimes |\Phi_u\rangle = |\sigma\rangle\,, \end{aligned} \tag{171}$$

where (e) follows similarly to (c) above, and (f) follows by noting that $\chi_u(s) = (-1)^{[s,u]}$ and thus the square of the character equals 1.

These calculations show that the sets U, W_1, W_2, S all stabilize $|\sigma\rangle$. Moreover, each of these sets has cardinality n. Hence, it remains to show that they are also independent, then we have $4n$ independent stabilizers for the state $|\sigma\rangle$, thus making it a $(4n)$-qubit stabilizer state. Firstly, note that these operators all commute pairwise. To see this, note that the operators V_x either commute or anti-commute. As two anti-commuting operators cannot simultaneously stabilize the same state, by the above calculations we conclude that the operators from U, W_1, W_2, S commute. Now, to show independence, we need to show that any non-trivial multiplication of elements of $A = U \cup W_1 \cup W_2 \cup S$ does not equal the identity. So, we consider a product

$$\begin{aligned} &\prod_{\substack{i \\ u_i \in \{u_i\}_i}} ((V_{u_i} \otimes I)^{\otimes 2})^{a_i} \prod_{\substack{j \\ h_j \in \{h_i\}_i}} (V_{h_j}^{\otimes 2} \otimes I^{\otimes 2})^{b_j} \prod_{\substack{k \\ h_k \in \{h_i\}_i}} (I^{\otimes 2} \otimes V_{h_k}^{\otimes 2})^{c_k} \prod_{\substack{l \\ s_l \in \{s_i\}_i}} V_{s_l}^{\otimes 4, d_l} \\ &\overset{(g)}{=} \pm (V_{\sum_i a_i u_i} \otimes I)^{\otimes 2} (V_{\sum_j h_j b_j}^{\otimes 2} \otimes V_{\sum_k h_k c_k}^{\otimes 2})(V_{\sum_l d_l s_l})^{\otimes 4} \\ &\overset{(h)}{=} \pm \left(V_{\sum_i a_i u_i + \sum_j h_j b_j + \sum_l d_l s_l} \otimes V_{\sum_j h_j b_j + \sum_l d_l s_l} \otimes V_{\sum_i a_i u_i + \sum_k h_k c_k + \sum_l d_l s_l} \otimes V_{\sum_k h_k c_k + \sum_l d_l s_l} \right)\,, \end{aligned} \tag{172}$$

where (g) and (h) follows by using the fact that $V_x V_y = V_{x\oplus y}$ up to a negative sign. Now, for the above term to equal the identity, it has to equal the identity on each factor of the tensor product. For it to be identity on the second register,

$$\sum_j b_j h_j + \sum_l d_l s_l = 0. \tag{173}$$

As $\{h_j\}_{j=1}^n \cup \{s_l\}_{l=1}^n$ is a basis and thus linearly independent by assumption, the above equation can only hold for $b_j = 0, d_l = 0, \forall j, l$. Then, for the fourth register to be identity, it is immediate that $c_k = 0, \forall k$ due to linear independence of $\{h_j\}_j$. Finally, we see that for the first register to be identity as well, $a_i = 0, \forall i$ due to linear independence of u_i. Thus, the only product of elements in A that yields the identity is the trivial product. We have thus shown the desired independence.

The above arguments show that elements in A are $4n$ stabilizers for the state $|\sigma\rangle$ and are independent, thus they generate the stabilizer group for $|\sigma\rangle$. So, $|\sigma\rangle$ is a $(4n)$-qubit stabilizer state. To argue that each of g-purification,

$$|\sigma_g\rangle = (I \otimes V_g^{\otimes 2})|\sigma\rangle \tag{174}$$

is also an $(4n)$-qubit stabilizer state, we simply define their stabilizer group using the stabilizer group of the state $|\sigma\rangle$: We take the stabilizer generators of state $|\sigma\rangle$ and then conjugate by the Pauli $(I^{\otimes 2} \otimes V_g^{\otimes 2})$. Thus, all g-purifications $|\sigma_g\rangle$ of σ are $(4n)$-qubit stabilizer state. ■

The above tells us that we can use our structured random purification to map any σ_L to a random purification, which is a $(4n)$-qubit stabilizer state. Thus, we can define a cloner for $\mathcal{S}_n$ by first using structured random purification, then applying a cloner for pure $(4n)$-qubit stabilizer states, and finally tracing out the environment registers. This allows us to carry the cloning lower bound from $\mathcal{S}_n$ over to pure $(4n)$-qubit stabilizer states. Then, by noting that a $(4n)$-qubit stabilizer states can be trivially be embedded into $(4n+1)$-qubit stabilizer states (by fixing the last qubit to $|0\rangle$), we can lift the lower bounds to $(4n+1)$-qubit stabilizer states. This allows us to set up the following result for cloning of any n-qubit stabilizer states.

Theorem 47 (Formal Statement of Theorem 4, Point 2: Cloning Lower Bound for Stabilizer States) *Let* Stab_n *be the set of pure* n*-qubit stabilizer states. Then* Stab_n *does not admit a* $(t-1, t, \epsilon)$*-quantum cloning scheme for*

$$t \leq \lfloor n/4 \rfloor \quad \textit{and} \quad \epsilon \leq 0.14. \tag{175}$$

To prove this, we first present some useful lemmas.

Lemma 48 *Let* Stab_{4n} *be the set of pure* $(4n)$*-qubit stabilizer states for* $n \geq 1$*. Then,* Stab_{4n} *does not admit an* $(n-1, n, \epsilon)$ *quantum cloning scheme with* $\epsilon \leq 0.14$*.*

Proof The proof is identical to that of Theorem 43 and follows by lifting the cloning lower bound of $\mathcal{S}_n = \{\sigma_L\}_{L \leq \mathbb{Z}_2^{2n}, |L| = 2^n}$ from Theorem 45 to the random purification states, which we have shown to be stabilizer states in Theorem 46. For completeness, the proof is provided in Section 4. ■

So far, we discussed stabilizer states with this peculiar $(4n)$-qubit number, which is an artifact of the proof technique that we used. A simple observation allows us to extend the result to any number of qubits.

Lemma 49 *Let* Stab_n *be the class of* n*-qubit stabilizer states. Suppose that* Stab_n *does not admit a* $(t-1,t,\epsilon)$ *quantum cloning scheme, then the class of states* Stab_{n+1} *(the* $(n+1)$*-qubit stabilizer states) does not admit a* $(t-1,t,\epsilon)$ *quantum cloning scheme.*

Proof The proof is provided in Section 4 and follows by noting that for a $(4n)$-qubit stabilizer state $|\psi\rangle$, $|\psi\rangle\otimes|0\rangle$ is a $(4n+1)$-qubit stabilizer state and thus any cloning scheme for $(4n+1)$-qubit stabilizer would be a cloning scheme for $(4n)$-qubit stabilizer states. ■

We are now well-equipped to provide the proof of Theorem 47.

Proof [Proof of Theorem 47] To show this, we start with Lemma 48, which says that there is no $(t-1,t,\epsilon)$ cloner for n-qubit stabilizer states with $n=4a$ with integer $a\geq 1$ and with $t=n/4=a$, Next, we use Lemma 42 to lift this result to any $t\leq n/4$. Finally, we lift this using Lemma 49 to stabilizer states on a number of qubits that is not an integer multiple of 4. We use the notation $\nexists\Lambda(\mathcal{S},t,t+1,\epsilon)$ to mean that $\mathcal{S}$ does not admit a $(t,t+1,\epsilon)$-quantum cloning scheme. Take $\epsilon\leq 0.14$, then

$$\begin{aligned}\text{Lemma 48}\Longrightarrow\nexists\Lambda(\mathrm{Stab}_n,n/4-1,n/4,\epsilon)&\xrightarrow{\text{Lemma 42}}\nexists\Lambda(\mathrm{Stab}_n,t-1,t,\epsilon),\quad\forall t\leq n/4\\&\xrightarrow{\text{Lemma 49}}\nexists\Lambda(\mathrm{Stab}_{n+1},t-1,t,\epsilon),\forall t\leq n/4=\left\lfloor\frac{n+1}{4}\right\rfloor.\end{aligned}\tag{176}$$

Similarly, we can argue about $(n+2)$ and $(n+3)$-qubit stabilizer states. When we arrive at $(n+4)$-qubit states, we can again use Lemma 48 since it is also a multiple of 4. ■

Thus, we have shown that cloning pure n-qubit stabilizer states to a small constant accuracy requires linearly-in-n many copies. This matches the sample complexity of exactly learning the same class of states up to constant factors; stabilizer state cloning is no easier than stabilizer state learning.

4. Deferred Proofs

Proof [Proof of Theorem 36] Notice that, for any $g'\in G$,

$$\begin{aligned}\mathrm{tr}\left(\mu(g')\sigma\right)&=\mathrm{tr}\left(\mu(g')\frac{1}{|G|}\sum_{g\in G}\mu(g)\rho\mu(g)^\dagger\right)\\&=\frac{1}{|G|}\sum_{g\in G}\mathrm{tr}\left(\mu(g')\mu(g)\rho\mu(g)^\dagger\right)\\&\overset{(a)}{=}\frac{1}{|G|}\sum_{g\in G}\mathrm{tr}\left(\mu(g)\mu(g')\rho\mu(g)^\dagger\right)\\&\overset{(b)}{=}\frac{1}{|G|}\sum_{g\in G}\mathrm{tr}\left(\mu(g')\rho\right)=\mathrm{tr}\left(\mu(g')\rho\right),\end{aligned}\tag{177}$$

where (a) follows because the group is Abelian, and (b) follows from cyclic property of trace. Hence, $\rho\in\Psi_H^\epsilon$ if and only if $\sigma\in\Psi_H^\epsilon$.

Now, to show the second part, take any $g' \in G$, then,

$$\begin{aligned}\mu(g')\sigma &= \frac{1}{|G|}\sum_{g\in G}\mu(g')\mu(g)\rho\mu(g)^\dagger \\ &= \frac{1}{|G|}\sum_{g\in G}\mu(g'g)\rho\mu(g)^\dagger \\ &\overset{(c)}{=} \frac{1}{|G|}\sum_{g''\in G}\mu(g'')\rho\mu(g''(g')^{-1})^\dagger \\ &\overset{(d)}{=} \frac{1}{|G|}\sum_{g''\in G}\mu(g'')\rho\mu(g'')^\dagger\mu(g') = \sigma\mu(g')\,,\end{aligned} \tag{178}$$

where (c) follows by redefining $g'' = g'g = gg'$, and (d) follows by Abelianity and noticing that $\mu((g')^{-1})^\dagger = \mu(g')$. ■

Proof [Proof of Theorem 39] To show this, we use the property of characters that,

$$\sum_{\lambda\in H^\perp}\chi_\lambda(g) = \begin{cases}|H^\perp| & \text{if } g\in H,\\ 0 & \text{otherwise.}\end{cases} \tag{179}$$

Now,

$$\begin{aligned}\operatorname{tr}\left(\mu(g)\sigma_H\right) &= \operatorname{tr}\left(\frac{1}{|H^\perp|}\sum_{\lambda\in H^\perp}\mu(g)|\lambda\rangle\langle\lambda|\right) \\ &= \frac{1}{|H^\perp|}\sum_{\lambda\in H^\perp}\chi_\lambda(g) = 1\end{aligned} \tag{180}$$

Thus, using the above equations,

$$\operatorname{tr}\left(\mu(g)\sigma_H\right) = \begin{cases}1 & \text{if } g\in H,\\ 0 & \text{otherwise}\,.\end{cases} \tag{181}$$

■

Proof [Proof of Theorem 15] As mentioned previously,

$$V_x^{\otimes 2}|\Phi_y\rangle = (-1)^{[x,y]}|\Phi_y\rangle, \tag{182}$$

where $[x,y] = a\cdot d + b\cdot c \mod 2$ with $x=(a,b)$ and $y=(c,d)$. Now, by definition,

$$L^\perp = \{y\in\mathbb{Z}_2^{2n} \mid (-1)^{[x,y]} = 1, \forall x\in L\}. \tag{183}$$

Thus, for any $x\in L$,

$$\begin{aligned}\operatorname{tr}\left(V_x^{\otimes 2}\sigma\right) &= \frac{1}{2^n}\sum_{y\in L^\perp}\operatorname{tr}\left(V_x^{\otimes 2}|\Phi_y\rangle\langle\Phi_y|\right) \\ &= \frac{1}{2^n}\sum_{y\in L^\perp}(-1)^{[x,y]} = 1.\end{aligned} \tag{184}$$

Now, for any $x \notin L$, there exists $y \in L^\perp$ such that,

$$(-1)^{[x,y]} = -1 \Longrightarrow [x,y] \neq 0. \tag{185}$$

If we choose a basis $\{l_i\}_{i=1}^n$ for $L^\perp$, then $y = \sum_i a_i l_i$ for some $a_i \in \{0,1\}, \forall i$. Then,

$$\sum_i a_i [x, l_i] \neq 0. \tag{186}$$

So, there exists at least one l_i such that $[x, l_i] \neq 0$. Then,

$$\begin{aligned} \operatorname{tr}\left(V_x^{\otimes 2}\sigma\right) &= \frac{1}{2^n}\sum_{y\in L^\perp}(-1)^{[x,y]} \\ &= \frac{1}{2^n}\sum_{a_1\ldots a_n}(-1)^{\sum_i a_i[x,l_i]} \\ &= \frac{1}{2^n}\prod_{i=1}^{n}\left(1+(-1)^{[x,l_i]}\right). \end{aligned} \tag{187}$$

Since, there exists atleast one l_i such that $[x, l_i] \neq 0$, above quantity is 0. ■

Proof [Proof of Theorem 48] We prove this by contradiction. Assume that there exist a $(n-1, n, \epsilon)$ cloner Λ for pure $(4n)$-qubit stabilizer states with error ϵ. Then, we consider the mixed states $\{\sigma_L\}_L$

$$\sigma_L = \frac{1}{2^n}\sum_{u\in L^\perp}|\Phi_u\rangle\langle\Phi_u|, \tag{188}$$

where L is some n dimensional subspace of $\mathbb{Z}_2^{2n}$ and their g-purifications $|\sigma_{L,g}\rangle$. As g-purifications are pure $(4n)$-qubit stabilizer states,

$$\mathrm{d_{TD}}(\Lambda(|\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n-1}), |\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n}) \leq \epsilon, \qquad \forall g, L. \tag{189}$$

Define the map $\Lambda' = \operatorname{tr}_{PR}\Lambda \circ C^{n-1}$, where tr_{PR} is the partial trace on purifying registers, and where C^{n-1} is the structured random purification channel for $n-1$ copies from Theorem 37. Then,

$$\mathrm{d_{TD}}\left(\Lambda'(\sigma^{\otimes n-1}), \sigma^{\otimes n}\right) \leq \mathrm{d_{TD}}(\Lambda \circ C^{n-1}(\sigma^{\otimes n-1}), \mathop{\mathbf{E}}_g |\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n}), \tag{190}$$

since the partial trace is a CPTP map and thus cannot increase trace distance. We can further upper bound the trace distance as

$$\begin{aligned} \mathrm{d_{TD}}\left(\Lambda \circ C^{n-1}(\sigma^{\otimes n-1}), \mathop{\mathbf{E}}_g |\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n}\right) &= \mathrm{d_{TD}}\left(\Lambda(\mathop{\mathbf{E}}_g |\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n-1}), \mathop{\mathbf{E}}_g |\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n}\right) \\ &\overset{(a)}{=} \mathrm{d_{TD}}\left(\mathop{\mathbf{E}}_g \Lambda(|\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n-1}), \mathop{\mathbf{E}}_g |\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n}\right) \\ &\overset{(b)}{\leq} \mathop{\mathbf{E}}_g \left(\Lambda(|\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n-1}), |\sigma_{L,g}\rangle\langle\sigma_{L,g}|^{\otimes n}\right) \leq \epsilon\,, \end{aligned} \tag{191}$$

where (a) follows from linearity of Λ and (b) follows by triangle inequality. Thus, Λ' is an $(n-1, n, \epsilon)$-cloner for $\{\sigma_L\}_L$. If $\epsilon < 0.14$, this contradicts Theorem 45. ■

Proof [Proof of Theorem 49] We show this by using contradiction. Assume that there exist a $(t-1, t, \epsilon)$ quantum cloner Λ for Stab_{n+1}. Then, we simply define a new cloner Λ' for n-qubit stabilizer states as

$$\Lambda' = \mathrm{tr}_1 \Lambda \circ V^{\otimes t-1}, \tag{192}$$

where V is an isometry from 2^n dimensions to 2^{n+1} dimensions such that for any n-qubit stabilizer state $|\psi\rangle$, $V|\psi\rangle = |\psi\rangle \otimes |0\rangle$ which is $(n+1)$-qubit stabilizer state from Claim 50, and tr_1 is partial trace on all copies of this single qubit register. Now, notice that for any n-qubit stabilizer state $|\psi\rangle$,

$$\mathrm{d_{TD}}\left(\Lambda'(|\psi\rangle^{\otimes t-1}), |\psi\rangle^{\otimes t}\right) \leq \mathrm{d_{TD}}\left(\Lambda \circ V^{\otimes t-1}(|\psi\rangle^{\otimes t-1}), |\psi 0\rangle^{\otimes t}\right), \tag{193}$$

which follows again by using the contraction property of trace distance under CPTP maps. Now,

$$\mathrm{d_{TD}}\left(\Lambda \circ V^{\otimes t-1}(|\psi\rangle^{\otimes t-1}), |\psi 0\rangle^{\otimes t}\right) = \mathrm{d_{TD}}\left(\Lambda(|\psi 0\rangle^{\otimes t-1}), |\psi 0\rangle^{\otimes t}\right) \leq \epsilon. \tag{194}$$

This gives a $(t-1, t, \epsilon)$ cloner for n-qubit stabilizer states by using the $(t-1, t, \epsilon)$ cloner for $n+1$ qubit stabilizer states, but this contradicts the assumption. ■

Claim 50 *For any n-qubit stabilizer state $|\psi\rangle$, $|\psi 0\rangle$ is a $(n+1)$-qubit stabilizer state.*

Proof To show this, we create stabilizer generators of $|\psi 0\rangle$. Let $g_1, \ldots g_m$ be stabilizer generators of $|\psi\rangle$, then we take the set,

$$A = \{g_1 \otimes I, \ldots g_m \otimes I, I^{\otimes m} \otimes Z\}. \tag{195}$$

Notice that the elements of A stabilizes $|\psi 0\rangle$. Moreover, they can be shown to be independent by leveraging the independence of g_i, and they also commute. Hence, the above $n+1$ elements would be stabilizer generators for $|\psi 0\rangle$, and so it is an $(n+1)$-qubit stabilizer state. ■