The locking-decoding frontier for generic dynamics

Jan Florjanczyk

Master of Science

School of Computer Science

McGill University

Montreal,Quebec

2011-06-13

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Master of Science

ABSTRACT

The intuition that the amount of classical correlations between two systems be bounded by their size does not hold true in general for quantum states. In the setting of *information locking*, measurements on a pair of quantum systems that appear to be completely uncorrelated can become maximally correlated with a small increment in the size of one of the systems. A new information locking scheme based on generic unitary channels is presented and a strengthened definition of locking based on a measure of indistinguishability is used. The new definition demonstrates that classical information can be kept arbitrarily low until it can be completely decoded. Unlike previous locking results, non-uniform input messages are allowed and shared entanglement between the pair of quantum systems is considered. Whereas past locking results relied on schemes with an explicit "key" register, this requirement is eliminated in favour of an arbitrary quantum subsystem. Furthermore, past results considered only projective measurements at the receiver. Here locking effects can be shown even in the case where the receiver is armed with the most general type of measurement. The locking effect is found to be generic and finds applications in entropic security and models for black hole evaporation.

ABRÉGÉ

L'intuition que le montant des corrélations classiques entre deux systèmes sont limités par leur taille est incorrect en général pour les états quantiques. En cas de *verrouillage*, des mesures sur une paire de systèmes quantiques qui semblent être totalement décorrélées peuvent devenir corrélées au maximum avec une minuscule augmentation de la taille d'un des systèmes. Une nouvelle forme de verrouillage utilisant des canaux unitaire génériques est introduite et la définition de verrouillage est renforcée a base d'une mesure d'indiscernabilité. La nouvelle définition montre que l'information classique peut être arbitrairement bas jusqu'à ce qu'elle puisse être complètement décodée. Aux contraire des résultats précédents, des messages non-uniforme et l'intrication entre la paire de systèmes sont considérés. Auparavant, il était nécessaire d'avoir un registre explicite pour une "clé", cette nécessité est supprimée en faveure d'un sous-système quantique arbitraire. De plus, les résultats précédent considéraient que les mesures projective mais nous démontrons des effets de verrouillage même dans le cas où le récepteur est armé avec les mesures les plus générales. Nous trouvons l'effet de verrouillage générique et montrons des applications pour la sécurité entropique et pour un modèl d'évaporation des trous noirs.

CONTRIBUTIONS OF AUTHORS

This is a manuscript-based thesis. Sections 2.1, 3, and 4 can be found in [15]. Please note that following convention in the field of quantum information theory, authors were listed alphabetically and we outline their various contributions now.

Section 2.1 was compiled jointly by the candidate, Jan Florjanczyk (JF), and Frédéric Dupuis (FD). The results of Section 3 were produced by JF although Definition 3.1.2 was introduced by FD. The Chernoff bound results of Section 3.5 were initially presented by FD in [14] and adapted for this work by JF. Patrick Hayden (PH) was supervisor to JF for the duration of the candidate's M.Sc. studies and advised on the calculations and analysis leading to the final results here. PH has also contributed comments and reviews during the writing and editing of this thesis. PH and Debbie Leung (DL) have contributed greatly to past results on locking and specifically to locking schemes generated by Haar-random unitaries. DL, PH, and FD found a preliminary result eliminating explicit keys, but one much weaker than presented here.

The remaining Sections of this work (1, 5, and 6) were compiled by the candidate, JF.

ACKNOWLEDGEMENTS

I would like to thank Patrick Hayden, my supervisor, for all of the help, guidance, and instruction over two wonderful years of graduate studies as well as a great deal of support for travel and conferences. I feel very fortunate to have had a mentor that cared so much about the development of his students and I leave his group with an incomparable experience behind me. I would also like to thank Frédéric Dupuis for his patient explanations and thoughtful feedback as well as Debbie Leung for her insights and contributions to the exhausting process of finishing one's first article. I would like to thank Andreas Winter for inviting me to visit and speak about this work in Bristol. I would like to thank Mark Wilde for a great deal of advice and impromptu tutorials in Shannon theory and Pranab Sen for answering my questions patiently and insightfully. I would like to thank Ion Nechita for some fascinating discussions. I would also like to thank CQIL members both past and present: Ivan Savov, Constance Caramanolis, Omar Fawzi, Kamil Bradler, Jürg Wullschleger, Andie Sigler, David Avis, Claude Crépeau, Nicolas Dutil, and Prakash Panangaden. I would finally like to thank my parents and my sister who supported me immensely throughout all of my studies over six years.

Andreas Winter has independently established some locking results for generic unitary transformations. This research was supported by the Canada Research Chairs program, the Perimeter Institute, CIFAR, CFI, FQRNT's INTRIQ, MITACS, NSERC, ORF, ONR through grant N000140811249, QuantumWorks, and the Swiss National Science Foundation through grant no. 200021-119868.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
Introduction

Consider two parties, Alice and Bob, attempting to perform a communication protocol. Alice is attempting to communicate a message m to Bob using at most n bits. Assuming Alice will not waste any bits during their communication, m can take one of 2^n possible values at the beginning of the protocol and Bob can expect Alice to send any of these. She begins by sending the registers of the bitstring, one by one, to Bob. With every received bit, Bob cuts the set of possible messages in half. Thus, when he has received $n-1$ bits, he has only two possible bitstrings left to guess from and could infer Alice's message with probability $\frac{1}{2}$. Throughout this protocol, the information Bob has about Alice's message will increase uniformly and smoothly until he receives all n bits, at which point he can completely infer Alice's message.

This simple example illustrates a proposition we should have no trouble accepting: that the amount of information contained in a physical system is proportional to the size of that system. It is surprising then, that this intuition fails in the setting of quantum information theory. Whereas in the classical setting above Bob has immediate access to the value of the registers he receives, in the quantum setting Bob is required to make a measurement on the quantum registers. Furthermore, via the quantum no-cloning Theorem, Bob cannot make perfect

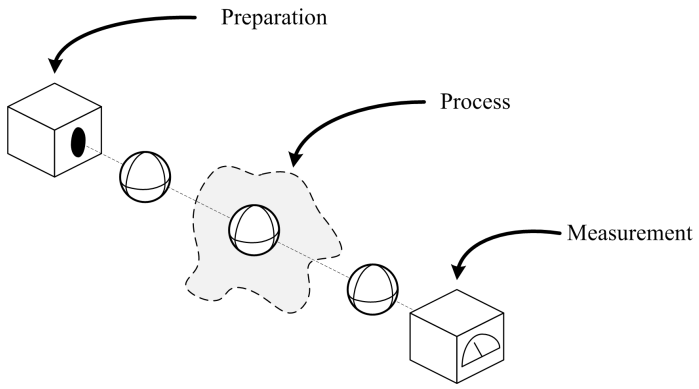

Figure 1–1: The most simple quantum communication protocol for communicating data from a sender to a receiver.

copies of the quantum registers and is restricted to performing the measurement only once.

Note that in most communication tasks, successful communication is defined via a criterion about a classical result (or about the possible classical outcomes of a quantum measurement). For this reason, measurement almost always plays a role in communication tasks. In the study of quantum information theory, we seek to predict the results of a measurement or to describe the best measurement possible for a given communication protocol. A measurement on a quantum state returns a set of possible outcomes with associated probabilities and therefore has a chance of failing to read out the contents of the state. A simple illustration is given in Figure 1–1. First, Alice prepares a quantum state (perhaps imperfectly) after which the state undergoes some process. The process can consist of a transformation due to a channel or a decoding operation performed by Bob at the receiving end. The final step is for Bob to perform a measurement yielding classical data.

In the scenario described above, if Bob were to attempt to guess Alice's message with only $n-1$ registers at his disposal, he would have to perform a measurement on these $n-1$ registers. For a large class of processes known as *information locking schemes*, Bob's measurement on $n-1$ quantum registers fails to read out almost any information about Alice's message but his measurement on n quantum registers succeeds. This large jump in Bob's ability to successfully readout the message in exchange for a very small number of "key" bits of communication is the topic of this work.

Notable improvements are made in this work over previous information locking results and we describe these contributions in a non-technical setting here.

Strengthened definition – Previous results in locking defined Bob's inability to decode Alice's message in terms of the accessible information, the total correlations that Bob can establish with the message via measurements. Our new definition is strictly stronger than this notion. We bound instead, the maximum probability with which Bob can infer whether the result of his measurement yielded a message from a cyphertext that Alice used or if the message came from an independently generated cyphertext. Furthermore, we can make this probability as small as we like. The best result as of the date of this work (see [21]) could yield a value of 3 for the accessible information but our techniques can make this value as low as any given ε for a logarithmic-size key up to corrective factors.

Non-uniform message – Whereas prior results pertained only to equiprobable cyphertexts for Alice, our results allow her to prepare her message according to a non-uniform distribution. By allowing for this larger class of

locking scenarios we capture a more general setting than the cryptographic task of quantum key distribution (QKD) for which locking has been studied [24]. For QKD it is perfectly reasonable to assume that Alice would choose her message uniformly at random but in general communication tasks Alice's cyphertext is not necessarily uniform. We find that the size of the key in the locking scheme increases linearly in the degree to which Alice chooses a non-uniform cyphertext.

Shared entanglement – We generalize our setting even further by allowing Alice and Bob to share an arbitrary pure state. The motivation is that such a state could potentially contain a large amount of entanglement. This is the first study of information locking in a setting with shared entanglement between the sender and the receiver. Allowing for this entanglement resource makes our setup particularly relevant to the Hayden-Preskill black hole evaporation model [22] and gives new evidence that information locking could rescue the long-lived remnant hypothesis for black holes [36].

Generic unitary – Previous locking schemes made use of a random unitary channel which encodes with an explicit register which transformations from a set of unitary evolutions is applied. Our construction, the generic unitary channel, merits a more natural physical motivation. Whereas a random unitary channel is designed with an explicit "key" register, a generic unitary channel applies only one unitary evolution on a larger quantum system. The subsystem that is traced out in the output can be chosen randomly and serves as the key. This is another improvement over the random unitary channel. Eliminating the need for a classical register to encode information about the particular unitary evolution applied

allows us to consider systems such as the Hayden-Preskill model [22] for black hole evaporation, and in general, any closed quantum system whose dynamics are well modeled by Haar random unitaries.

Extension to general measurements – Finally, previous locking results showed that Bob cannot extract information about Alice's message using only projective measurements. Although this is a satisfactory assumption for many results in quantum information theory, we make use of a new technique to extend our results for more general measurements on Bob's system. A Chernoff bound technique originated by [14] and similar to [40] allows us evaluate locking in the case where Bob can perform any measurement, making our results strictly more general than in the past.

The work is divided as follows: In Section 1.2 we introduce the notion of states and define the basic ways in which we quantify the amount of information they contain. Once we establish these notions in the classical setting, we proceed to construct similar ideas in the quantum setting in Section 1.3. As we've hinted above, we must also introduce the measurement formalism in this Section. We review past results in information locking in Section 1.4. Finally, we introduce two important methods of analysis in Sections 1.5.1 and 1.5.2. In Section 2 we construct the communication protocol for which we will analyze the locking effect.

We dedicate Section 3 to proving that without all of the registers available to him, Bob's measurement will fail to reveal Alice's message. We then also prove that Bob can successfully decode Alice's message in Section 4. The calculations found in these Sections are due to the work in [15]. In Section 5 we discuss the

implications of our findings for interesting examples in information theory and physics and we summarize all of our work in Section 6.

1.1 Notation

General

log	Logarithm base 2.
$\mathbb{E}_U[f(U)]$	Expectation value of $f(U)$ over the random variable U.
AB	Composite quantum system whose associated Hilbert space is $\mathcal{H}^A \otimes \mathcal{H}^B$.
$\|A\|$	Dimension of Hilbert space A. However, we will often drop the $\|\ \cdot\ \|$. For example, the dimension of the composite system MCK is denoted by MCK (a scalar value).
$A^{\otimes 2}$	Two identical copies of A the second of which is denoted by $\overline{A}$.
$\|\psi\rangle^A, \|\varphi\rangle^A, \ldots$	Vectors in A.
$\psi^A, \varphi^A, \ldots$	The "unketted" versions denote their associated density matrices: $\psi^A = \|\psi\rangle\langle\psi\|$. Furthermore, if we have defined a state ψ^{AB}, then $\psi^A = \mathrm{Tr}_B[\psi^{AB}]$.
π^A	The maximally mixed state $\frac{\mathbb{I}^A}{\|A\|}$.
$\mathcal{U}(A)$	The unitary group on A.
$\mathrm{Pos}(A)$	The subset of Hermitian operators from A to A consisting of positive semidefinite matrices.
$\mathcal{L}(s, \eta)$	The set of all (s, η)-quasi-measurements, see Definition 2.1.1
$\mathcal{B}(A)$	The set of all bounded positive operators on A.

Operators

$\mathbb{I}^A$	Identity operator on A.
$M^{A\to B}$	Indicates that the operator M is a transformation from states on A to states on B.
$\mathcal{M}^{A\to B}$	Indicates that the superoperator $\mathcal{M}$ is a transformation from operators on A to operators on B. M and $\mathcal{M}$ will be freely identified with their extensions (via tensor product with the identity) to larger systems.
$M \cdot N$	$MNM^\dagger$
$M \leqslant N$	If $M, N \in \mathrm{Herm}(\mathsf{A})$, this means that $N - M \in \mathrm{Pos}(A)$.
$\sqrt{M}$	If $M \in \mathrm{Pos}(A)$ has spectral decomposition $M = \sum_i \lambda_i \lvert\psi_i\rangle\langle\psi_i\rvert$, then $\sqrt{M} = \sum_i \sqrt{\lambda_i}\lvert\psi_i\rangle\langle\psi_i\rvert$.
$\Pi_\pm^A$	Projector onto the symmetric (+) or antisymmetric (−) subspace of $A^{\otimes 2}$.
$\mathrm{op}_{A\to B}(\lvert\psi\rangle^{AB})$	Turns a vector into an operator. See Definition 3.1.2.

Norms and Entropies

$\left\|M^{A\to B}\right\|_1$	$\operatorname{Tr}\sqrt{M^\dagger M}$
$\left\|\|\psi\rangle\right\|_2$	$\sqrt{\lvert\langle\psi\rvert\psi\rangle\rvert}$
$\left\|M^{A\to B}\right\|_2$	$\sqrt{\operatorname{Tr}[M^\dagger M]}$
$\left\|M^{A\to B}\right\|_\infty$	Largest singular value of M, *i.e.* the operator norm of M.
$H_2(A)_\rho$	Renyi 2-entropy of A, defined as $-\log\operatorname{Tr}[\rho^2]$.
$H_{\min}(A)_\rho$	Quantum min-entropy of A, defined as $-\log\min_{\lambda\in\mathbb{R}}\{\lambda : \rho^A \leqslant \lambda\mathbb{I}^A\}$.
$H_{\max}(A)_\rho$	Quantum max-entropy of A, defined as $2\log\operatorname{Tr}\sqrt{\rho^A}$.

1.2 Classical information theory

In this Section we will introduce some of the key ideas in information theory. We will later explain why we refer to the concepts in this Section as "classical" and the concepts in future Sections as "quantum". We will define a *state* as the mathematical object which describes a preparation and the statistics of any possible measurement of that preparation [20]. The most natural representation of a state can be formulated as a *probability space*,

Definition 1.2.1 (Probability space)**.** *We call* $(\mathcal{X}, \mathcal{A}, \Pr)$ *a* probability space *associated with a particular state if* $\mathcal{X}$ *is the alphabet of possible instances of the state,* $\mathcal{A}$ *is a set of subsets of* $\mathcal{X}$*, and* $\Pr$ *is a measure on* $\mathcal{A}$ *normalized so that* $\Pr\{\mathcal{X}\} = 1$.

Note also that if $\mathcal{A}$ is the set of *all* subsets of $\mathcal{X}$ then it is the σ-algebra of $\mathcal{X}$ as defined in Definition A.0.2. We will consider finite alphabets $\mathcal{X}$, i.e.: $|\mathcal{X}| = N$. To each element $x \in \mathcal{X}$ we will associate $p(x) = \Pr(\{x\})$. By the definition of a measure (A.0.3) we know that Pr is additive for disjoint sets and so we can note that $\sum_x p(x) = 1$. We call the collection of outcomes and their probabilities $\{(p(x), x)\}$ a *state*, a *distribution*, or, equivalently, a *random variable*. We note that the set of all states is a convex set.

Definition 1.2.2 (Pure and mixed states)**.** *A* pure state *is an extremal element of the convex set of states, i.e.:* $\Pr(\{x\}) = 1$ *but* $\Pr(A \backslash \{x\}) = 0$ *for any* $A \in \mathcal{A}$*. Any state that is not pure is called a* mixed state*.*

We denote by X the random variable constructed from the alphabet $\mathcal{X}$ and the probability distribution Pr. At this point we have not chosen a particular structure for the set $\mathcal{X}$. A very natural construction would be $\{0,1\}^{\times N}$ (a string of bits), but for the information theoretic quantities defined below, this is not a strict necessity.

Definition 1.2.3 (Shannon entropy)**.** *For the random variable* X *formed from the alphabet* $\mathcal{X} := \{x\}$ *and the probability measure* $p(x)$*, we define the* Shannon entropy *as*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x)$$

where the logarithm is taken to be base 2*.*

The Shannon entropy can be interpreted as the expected value of an "information function" $I(A)$ for each of the events $A \in \mathcal{A}$. We require that such an I be smooth in $\Pr(A)$, dependent exclusively on the probability of A, and additive

under union of disjoint events (i.e.: $I(A \cup A') = I(A) + I(A')$). The only eligible candidates then are $I(A) = -k \log \Pr(A)$ for some fixed k (Theorem 2 of [35]). We choose $k = 1$ and see that if we take the expectation over individual outcomes $\{x\}$ only then

$$\mathbb{E}_{\Pr} I(x) = -\sum_x p(x) \log p(x) = H(X).$$

The appropriate interpretation of the Shannon entropy is the "amount of uncertainty" in a state. Pure states have Shannon entropy 0, and these are states for which we have no uncertainty. On the other extreme lies the maximally mixed state $\{(1/|\mathcal{X}|, x)\}$ which gives the maximum value for the entropy, $H(X) = \log |X|$. Note the flip-side to this interpretation: if a state is very uncertain, then learning its value constitutes gaining a large amount of information.

Where two different states are concerned, two different random variables will share a product alphabet $\mathcal{X} \times \mathcal{Y}$ and one probability measure $\Pr_{\mathcal{X}\times\mathcal{Y}}$ which may or may not factorize into $\Pr_{\mathcal{X}} \times \Pr_{\mathcal{Y}}$. Note that we have implicitly defined the marginal distribution over each of the alphabets $\mathcal{X}$ and $\mathcal{Y}$. The marginal distributions should be taken as

$$\Pr_{\mathcal{X}}(x) = \Pr_{\mathcal{X}\times\mathcal{Y}} \left(\{x\} \times \{\mathcal{Y}\} \right).$$

We define also the uncertainty in two (or more) variables together.

Definition 1.2.4 (Joint entropy)**.** *For the random variables X and Y we define their joint entropy*

$$H(XY) = -\sum_{x\in\mathcal{X}, y\in\mathcal{Y}} p(x, y) \log p(x, y)$$

We also define the amount of uncertainty about one random variable under the assumption that the other is known.

Definition 1.2.5 (Conditional entropy)**.** *For the random variables X and Y we define their conditional entropy*

$$H(Y|X) = -\sum_{x\in\mathcal{X}} p(x)H(Y|X=x)$$

where $H(Y|X=x)$ is defined to be the Shannon entropy for the alphabet $\{x\}\times\mathcal{Y}$ with the probability measure $\Pr(\{x\}\times\mathcal{Y})/\Pr(\{x\})$.

Finally we define one of the most significant quantities for our analysis.

Definition 1.2.6 (Mutual information)**.** *For the random variables X and Y we define the mutual information as*

$$I(X:Y) = H(X) + H(Y) - H(XY)$$

which can equivalently be written as $I(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$ via Bayes' rule.

The mutual information can be interpreted as the total amount of uncertainty that two variables share in common. If two states have high mutual information then they are also highly correlated. As an example, the state for which mutual information is maximized is the maximally correlated state: $p(x,y) = \delta_{xy} 1/\min(|\mathcal{X}|,|\mathcal{Y}|)$.

1.3 Quantum information theory

Our quantum model of information processing will require a new setting. We will need to enhance our definition of state via some physical motivations. The

collection of all quantum states will live in the Hilbert space and inherit its properties. Whereas previously we considered bitstrings and distributions over bitstrings, we will now concern ourselves with quantum states and their distributions. As such, all of the constructions of Section 1 will be named "classical" with respect to the formalism we introduce here.

1.3.1 Hilbert spaces, states, and density operators

In analogy to classical states that take values in an alphabet, *quantum states* will take values in a *Hilbert space* $\mathcal{H}$.

Definition 1.3.1 (Hilbert space)**.** *Let $\mathcal{H}$ be a complete inner product vector space such that a quantum state $|\psi\rangle \in \mathcal{H}$ (a "ket" state). We denote by $\langle\phi| \in \mathcal{H}^\dagger$ (a "bra" state) the quantum state belonging to the dual space of $\mathcal{H}$. The dual is taken to be the conjugate transpose and $\mathcal{H}$ and $\mathcal{H}^\dagger$ are related by the inner product,*

$$\langle\phi|\psi\rangle := \langle\phi|\left(|\psi\rangle\right)$$

where ϕ has now assumed its role as a linear functional on $\mathcal{H}$. The distance function

$$d(\psi,\phi) = \sqrt{\left(\langle\phi| - \langle\psi|\right)\left(|\phi\rangle - |\psi\rangle\right)}$$

grants the Hilbert space the properties of complete metric *space as well.*

We will restrict ourselves to finite-dimensional Hilbert spaces in subsequent calculations. However, as we will see later, we will wish to examine a function's asymptotic limit in the dimension of the Hilbert space. Linearity of the Hilbert space is important as it embodies a fundamental behaviour of quantum states, that

of superposition. Unlike variables in our classical construction, the linear combination of two pure quantum states remains a pure quantum state. Completeness of the space is also important as it would be unreasonable to describe a quantum theory where the limit of a Cauchy sequence of states was not itself a possible state within the Hilbert space. Finally, we will only consider quantum states which are unit-normalized, i.e.: $|\phi\rangle$ is a quantum state if it is unit normalized, $|\langle\phi|\phi\rangle| = 1$.

At this point we still have not assumed an underlying field for the Hilbert space. The reader may have already guessed that this choice will be $\mathbb{C}$. If we had described our Hilbert space with the field $\mathbb{R}$, we would encounter a contradiction. Consider a two-dimensional Hilbert space on $\mathbb{R}$. Two possible pure states in this space are the vectors $|(1,0)\rangle$ and $|(0,1)\rangle$. The operator that flips $|(1,0)\rangle$ and $|(0,1)\rangle$ is

$$S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

However, if this operator represents a physical process, it must be possible to implement it continuously. That is, if S takes time Δt to implement, then we could interrupt the process at time $\frac{1}{2}\Delta t$ and expect the resulting state to still be pure. Unfortunately, S does not accept a square root in the space of 2×2 real matrices. $\mathbb{C}$ is algebraically closed meaning that it will contain the eigenvalues of any root of S. Any root of S is thus contained in the space of 2×2 complex matrices. It is useful to remember that in a complex Hilbert space, all pure states are connected and can be continuously rotated into one another. This is not

the case in the space of classical pure states however where the only possible operations between pure states are permutations of the basis.

We can associate any physical system to a Hilbert space. Physical systems can be subdivided however, and we will now establish the notion of subsystems. Recalling that linearity is necessary for superposition of quantum states, the same theory that describes any one system should also describe a composite system. The fashion in which we join two quantum systems will be the tensor product $\otimes$. We denote the composition $|\phi\rangle \otimes |\psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ for two systems A and B. The tensor product is multilinear (linear in all its entries) meaning that quantum theory is extensible.

In analogy to Section 1.2, we can take $\mathcal{X}$ to be the set of all outer products (or projections) associated with each vector in the Hilbert space and $\mathcal{A}$ to be their σ-algebra. As before, our definition of state (now *quantum state*) will be constructed from the probability space although now with any probability space we will also associate a *density operator*.

Definition 1.3.2 (Density operator)**.**

$$\rho^X = \int_{x \in \mathcal{X}} |x\rangle\langle x|^X \mathrm{d}_{\mathrm{Pr}}(x)$$

where $|x\rangle\langle x|$ *is the outer product of the vector corresponding to* x *in* $\mathcal{H}_X$ *and* $\mathcal{H}_X^\dagger$*. For an* n*-dimensional Hilbert space,* $\rho^X \in \mathrm{Herm}(X)$ *and* $\mathrm{Tr}\left[\rho^X\right] = 1$*.*

Most often, the state is defined as a mixture of n pure states, that is, $\Pr(x) = p(x)$ for n values of x and the density operator becomes $\rho = \sum_x p(x)|x\rangle\langle x|$. We

interpret the density operator ρ as the mathematical object which contains all of the information that quantum operations can extract from a quantum state.

As with state vectors, the density operator must be multilinear in subsystems and thus we can also write a composite system as the tensor product of the density operator on two subsystems $\rho^A \otimes \sigma^B$. If, in matrix entry notation, $\left[\rho^A\right]_{(i,j)} = a_{ij}$ and $\left[\sigma^B\right]_{(k,l)} = b_{kl}$ then $\left[\rho^A \otimes \sigma^B\right]_{(m,n)} = [a_{ij}b_{kl}]_{(m,n)}$, where $m = ik$ and $n = jl$. If we are given a state on a composite system but wish only to consider one subsystem, we can recover the marginal state via a partial trace operation.

$$\mathrm{Tr}_B\left[\rho^{AB}\right] = \sum_i \left(\mathbb{I}^A \otimes \langle i|^B\right) \rho^{AB} \left(\mathbb{I}^A \otimes |i\rangle^B\right) = \rho^A$$

for $\{|i\rangle\}_i$ an orthonormal basis in B. We will denote an n-fold tensor product of ρ as $\rho^{\otimes n}$.

Density operators have the uniquely quantum property of purification. Consider that any given density operator ρ^A is positive semidefinite and therefore has a square root factorization $\rho^A = (\sqrt{\rho^A})^\dagger \sqrt{\rho^A}$ (Exercise I.2.2 in [6]). If we take the spectral decomposition $\rho^A = \sum_k \lambda_k |k\rangle\langle k|$ where the $\{|k\rangle\}$ form an orthonormal basis for A, we can construct the state $|\rho\rangle^{AR} = \sum_k \sqrt{\lambda_k}|k\rangle^A \otimes |k\rangle^R$ on the composite system $A \otimes R$. It is easy to check that $\mathrm{Tr}_R[|\rho\rangle\langle\rho|^{AR}] = \rho^A$. Note that the vectors $|k\rangle^R$ can be chosen to be any set of orthonormal vectors in R. Thus, just as all orthonormal bases are related by unitary transformations, all purifications of equivalent dimension are related by unitary transformations.

Unlike pure classical states, some pure quantum states cannot be written as the product of pure states on subsystems. This is another uniquely quantum

property called *entanglement*. Consider the pure state $|\Phi\rangle^{AA'} := 1/n \sum_i |ii\rangle^{AA'}$ where $\dim A = \dim A' = n$. There is no ρ^A and $\sigma^{A'}$ such that $\Phi^{AA'} = \rho^A \otimes \sigma^{A'}$. Furthermore, the complete state purifies the marginal state on either A or A'. In this way we can always interpret a mixed state as the marginal of some purification. We will later show the consequences of entanglement on various quantum information measures.

Finally, we define formally the notion of "process" illustrated in Figure 1–1. We require that any map $\mathcal{E}^A$ from density operators to density operators be completely positive, i.e.: $\mathbb{I}^R \otimes \mathcal{E}^A$ is positive for any ancillary system R of any size. This is a strictly stronger requirement than positivity for $\mathcal{E}$ alone. Furthermore, if $\mathcal{E}$ is to map density operators to density operators, it must preserve the normalization (or trace) of the density operators. We call any completely positive trace-preserving map (CPTP map) from A to B a *quantum channel*,

$$\mathcal{N}^{A\to B}\left(\rho^A\right) = \sigma^B.$$

Note that quantum channels need not be reversible. However, they can be made to be so by the introduction of an environment system E via the Stinespring dilation Theorem.

Theorem 1.3.1 (Stinespring dilation [37])**.** *For any completely positive trace-preserving map $\mathcal{N}^{A\to E} : \mathcal{B}(A) \to \mathcal{B}(B)$, there exists a system E and a unitary transformation $U_{\mathcal{N}}^{A\to BE}$ such that*

$$\mathcal{N}^{A\to B}(\rho) = \mathrm{Tr}_E\left[U_{\mathcal{N}}\rho U_{\mathcal{N}}^\dagger\right].$$

We will often denote the conjugation of a state by a unitary as $U \cdot \rho := U\rho U^\dagger$.

1.3.2 Measurements and distances

We now require a physically meaningful way to extract classical results from a quantum state. In quantum mechanics, observable quantities are traditionally modeled as Hermitian operators M. If M admits a spectral decomposition $\sum_m m|m\rangle\langle m|$ then the results of the observable M for a state ρ are given by m with probability $\langle m|\rho|m\rangle$. However, these do not capture all of the physically realizable measurements and we make use of the following more general formalism.

Definition 1.3.3 (Positive Operator-Valued Measure (POVM))**.** *A countable set of operators $\{E_m\}$ is said to be a* POVM *if $\sum_m E_m = \mathbb{I}$ and $E_m \geqslant 0$ for all m. For a quantum state ρ, we associate with $p(m) = \mathrm{Tr}[E_m\rho]$ the probability of outcome m and note that $\sum_m p(m) = 1$. The post-measurement state for a particular outcome m' is given by*

$$\rho \longrightarrow \frac{\sqrt{E_{m'}}\rho\sqrt{E_{m'}}}{\mathrm{Tr}\left[\sqrt{E_{m'}}\rho\sqrt{E_{m'}}\right]}.$$

Note that in the above, $\sqrt{E_{m'}}$ is not necessarily a projective operator and thus repeated measurement does not necessarily leave the state unchanged. Further, $\sqrt{E_{m'}}$ is in general not uniquely defined, meaning that different measurement procedures can yield different post-measurement states. The most common species of POVM encountered is the projective measurement.

Definition 1.3.4 (Projective measurement)**.** *Consider the POVM $\{E_m\}$ with the additional restriction that $E_i^\dagger E_j = \delta_{ij}E_i$ for all i, j. Such a POVM is called a* projective measurement *as repeated application does not alter the post-measurement quantum state.*

We will also find it useful to assign an operator on density operators (called a superoperator) for any POVM.

Definition 1.3.5 (Measurement superoperator)**.** *We call a completely positive, trace-preserving (CPTP) map* $\mathcal{M} : \mathcal{B}(A) \to \mathcal{B}(X)$ *a* measurement superoperator *if it is of the form* $\mathcal{M}(\rho) = \sum_{i=1}^{N} |i\rangle\langle i|^X \operatorname{Tr}[M_i^A \rho]$*, where* $\{|i\rangle^A : i \in \{1, \ldots, N\}\}$ *is an orthonormal basis for* X*, each* M_i^A *is positive semidefinite, and* $\sum_{i=1}^{N} M_i^A = \mathbb{I}^A$*.*

Often, our task is to distinguish two different quantum states. We've already implicitly equipped the Hilbert space with the Euclidean distance measure (i.e.: the ℓ^2-norm). For the space of density operators, we will assign two distances motivated by norms on the operator space.

1. **The Hilbert-Schmidt distance**, also known as the 2-norm, is defined as

$$\|\rho - \sigma\|_2 = \sqrt{\operatorname{Tr}\left[(\rho - \sigma)^2\right]}$$

   Just as with the ℓ^2-norm, this one is induced by an inner product. The Hilbert-Schmidt inner product for operators is $(A, B) = \operatorname{Tr}[A^\dagger B]$, which one can note for quantum states (which are Hermitian) gives the statement above.

2. **The trace distance**, also known as the 1-norm, is defined as

$$\|\rho - \sigma\|_1 = \operatorname{Tr}\left[|\rho - \sigma|\right]$$

   The absolute value of any operator A is defined as $\sqrt{A^\dagger A}$ which as we note again for Hermitian density operators is simply $\sqrt{A^2}$.

For an operator A on a Hilbert space of dimension n, we have two useful inequalities via [6]

$$\|A\|_1 \leq \sqrt{n}\,\|A\|_2\,,$$
$$\|A\|_2 \leq \sqrt{\|A\|_1\,\|A\|_\infty}.$$

1.3.3 Information Theory and Entropic Quantities

Recall that in Section 1.2 we introduced the Shannon entropy as a measure of uncertainty for a particular state. For a mixed state, the Shannon entropy was a function of probabilities of the ensemble regardless of the actual pure states that composed it. In order to develop a similar idea in quantum information we will have to deal with the fact that the pure states of our Hilbert space are not by default orthogonal (as they are for classical random variables). Fortunately, the density operator language gives a very powerful formula for capturing the amount of uncertainty in a particular quantum state.

Definition 1.3.6 (Von Neumann entropy)**.** *For any quantum state ρ, we define the* Von Neumann entropy *as*

$$H(X)_\rho = -\mathrm{Tr}\left[\rho \log \rho\right]$$

where $\log \rho$ *is taken base* 2*.*

Note that for a diagonalization of $\rho = \sum_k \lambda_k |k\rangle\langle k|$ we have that $H(\rho) = H(\{\lambda_k\})$ the Shannon entropy. In fact, this equality is a special case when the

mixture from which we construct ρ is a mixture of orthogonal pure states. Consider the mixed state taken as the following mixture of probabilities and states:

$$\left\{ \left(\frac{1}{2}, |\psi\rangle^X \right), \left(\frac{1}{2}, (1-\varepsilon)|\psi\rangle + \varepsilon|\psi^\perp\rangle \right) \right\}$$

where we've assumed $\langle\psi|\psi^\perp\rangle = 0$ and a fixed ε. If we naively take the Shannon entropy of this mixture, we calculate that the mixture is maximally uncertain with $H(X) = 1$. However, we note that the density operator can be written as

$$\rho^X = \left(1 - \varepsilon + \frac{1}{2}\varepsilon^2\right)|\psi\rangle\langle\psi| + \left(\frac{1}{2}\varepsilon - \frac{1}{2}\varepsilon^2\right)\left(|\psi\rangle\langle\psi^\perp| + |\psi^\perp\rangle\langle\psi|\right) + \frac{1}{2}\varepsilon^2|\psi^\perp\rangle\langle\psi^\perp|.$$

From which, if we calculate the Von Neumann entropy ignoring ε^2 terms we find $H(X)_\rho \approx \varepsilon$. In the first expression for ρ, the Shannon entropy was near maximal, while in the second it is near 0.

As with the mutual information of two random variables, we can now define the quantum mutual information of a composite quantum state

Definition 1.3.7 (Quantum mutual information)**.** *For a bipartite state ρ^{XY} we define the* quantum mutual information *as*

$$I(X:Y)_\rho = H(X)_\rho + H(Y)_\rho - H(XY)_\rho.$$

We can now note one of the most surprising facts in quantum information theory: the quantum mutual information can be much higher than the classical mutual information. In fact, the maximum value of the $I(X:Y)_\rho$ is twice that of the maximum value for $I(X:Y)$. The maximally correlated state that gave the maximum classical value for mutual information can be written as the quantum

state $\sum_i 1/|X||ii\rangle\langle ii|^{XY}$ (for $|X| = |Y|$). For this state $H(X) = H(Y) = H(XY) = \log|X|$ and $I(X:Y) = \log|X|$. However, in the quantum setting we can write the maximally entangled state $|\Phi\rangle = \sum_i 1/\sqrt{|X|}|ii\rangle^{XY}$ and for this state $\mathrm{Tr}_Y[|\Phi\rangle\langle\Phi|] = \pi^X$ the maximally mixed state (similarily for the marginal state on Y). However, since the state on XY is pure, $H(XY) = 0$ and we have that $I(X:Y) = 2\log|X|$.

We note the following Theorem which gives insight into the relationship between POVMs and the Shannon entropy

Theorem 1.3.2 (Fine-grained POVMs [9])**.** *For any mixture of states $\{p(x), \rho_x^A\}_x$, there exists a POVM M consisting of elements $\{\alpha_y|y\rangle\langle y|^A\}_{y=1}^n$ such that $I(X:Y)$ is maximized and $|A| \leq n \leq |A|^2$.*

Finally, one of the most common quantum systems is the qubit. A qubit is a state on a 2-dimensional Hilbert space with the implicit basis $\{|0\rangle, |1\rangle\}$ (known as the computational basis). Although our results do not require that the quantum states in question be strings of qubits, we will often count the dimension of a Hilbert space in "bits" as this illustrates our statements more intuitively.

1.4 Locking information

We will now introduce another correlation measure to contrast with mutual information. Mutual information, although it does bound the trace distance between two quantum states, does not provide any direct insight into their operational distinguishability. First, consider the following species of quantum state,

Definition 1.4.1 (Classical-quantum state)**.** *A classical-quantum (or "cq") state is any bipartite state of the form*

$$\sigma^{XB} = \sum_x p(x)|x\rangle\langle x|^X \otimes \rho_x^B$$

such that $\{|x\rangle\}_x$ *form an orthogonal basis for* X *and* $\sum_x p(x) = 1$ *with* $p(x) \geq 0$ *for all* x*.*

The following measure, introduced by Fuchs, was defined with the task of distinguishing two quantum states in mind,

Definition 1.4.2 (Accessible information [18])**.** *Let* σ^{XB} *be a cq-state then, the accessible information* $I_{\text{acc}}(X;B)$ *is defined as*

$$I_{\text{acc}}(A;B)_\sigma := \sup_{\mathcal{B}} I(X;Y)_{(\mathbb{I}\otimes\mathcal{B})(\sigma)},$$

where $\mathcal{B}^{B\to Y}$ *is a measurement superoperator, and the supremum is taken over all possible measurement superoperators.*

In other words, the accessible information is the largest possible mutual information between the "classical" part X of the cq-state and a measurement on the "quantum" part B. The accessible information is bounded above by the Holevo quantity for a mixture of states $\{p(x), \rho_x\}$

$$\chi(\rho) = H(X)_\rho - \sum_x p(x) H(X)_{\rho_x}.$$

Applying a measurement superoperator to a quantum state ρ^{AB} yields a cq-state σ^{XB} between the classical measurement results X and the quantum part B. In

this way we can develop a notion similar to accessible information for general quantum states.

Definition 1.4.3 (Classical mutual information for quantum states ([12]))**.** *For any quantum state ρ^{AB}, we define the accessible information as*

$$I_c(A;B)_\rho := \sup_{\mathcal{A}\otimes\mathcal{B}} I(X;Y)_{(\mathcal{A}\otimes\mathcal{B})(\rho)},$$

where the supremum is taken over all possible measurement superators $\mathcal{A}^{A\to X}$ and $\mathcal{B}^{B\to Y}$.

Thus the classical mutual information represents a measure of the bipartite correlations available in a quantum state via local measurement operations. Note that the maximum value that this second definition of accessible information can take is $2\log\min(|A|,|B|)$ (by Theorem 1.3.2).

The accessible information is the correlation measure at the centre of the information locking effect - it is a quantity that can exhibit a drastic jump. We now review three papers which have analyzed schemes for information locking. All of these will have the following general idea in common; Alice will prepare a cq-state state and send the quantum part to Bob through a judiciously chosen channel. Bob will then attempt to infer Alice's classical part of their shared state. If the channel transmits Alice's input to Bob in full, then Bob will be successful with high probability. However, if the channel is built to erase some part of the state, known as the *key*, Bob will be unsuccessful due to low correlation (and low accessible information) with Alice's state.

We will define locking to be the instances where a small key costs Bob greatly in success probability (or equivalently, in correlation with Alice). It is precisely this sharp transition that defies our intuition that the amount of information contained in a state (here the key) should be bounded by the size of the state. It was proved in [11] that $I_c(A;B)_{\rho^{\otimes n}} = nI_c(A;B)_\rho$, implying that the information locking effect is unaffected by multiple copies of the state and thus not an artefact of "one-shot" communication between Alice and Bob. The works reviewed below give results quantitatively relating the key size to the amount of correlation between Alice and Bob.

1.4.1 Past locking results

Locking with one qubit [12] Consider that the channel between Alice (A_1A_2) and Bob (B) establishes the state

$$\rho^{A_1A_2B} = \frac{1}{2d}\sum_{t=0}^{d-1}\sum_{k=0}^{1} |t\rangle\langle t|_{A_1} \otimes |k\rangle\langle k|_{A_2} \otimes \left(U_k|t\rangle\langle t|U_k^\dagger\right)_B \tag{1.4.1}$$

where $\{|t\rangle\}$ and $\{|k\rangle\}$ are respectively d and 2 dimensional states in the computational basis. Also, define $U_0 = \mathbb{I}^B$ and $U_1 = H^{\otimes \log d}$, where H is the Fourier transform (or Hadamard gate) and has the property that $\forall_{i,t}\ |\langle i|U_1|t\rangle| = 1/\sqrt{d}$. Note that we've assumed d is a power of 2 in this scenario. Together, t and k form the bitstring that Alice splits into a "cyphertext" t and a "key" k. It is clear that if Alice measures the register k and then sends it to Bob, Bob can simply undo the appropriate unitary transformation to recover a perfectly correlated register t. In this scenario Alice and Bob establish $\log d + 1$ bits of correlation ($\log d$ from register t and 1 from a shared knowledge of register k).

However, if Bob is left to infer t without the aid of k, he can establish at most $\frac{1}{2}\log d$ bits of correlation. Note that in this particular case $I_c(A;B)_\rho = I_{\text{acc}}(A;B)_\rho$ since the optimal measurement for Alice is the measurement in the computational basis $\{|t\rangle \otimes |k\rangle\}$. If we equip Bob with the fine-grained POVM $\mathcal{B} = \{\alpha_j |\phi_j\rangle\langle\phi_j|\}$ then $I_c(A;B)_\rho$ becomes exactly

$$I_c(A;B)\rho = \max_{\mathcal{B}} \left[\log d + \sum_j \frac{\alpha_j}{d} \left(\frac{1}{2} \sum_{tk} |\langle\phi_j|U_k|t\rangle|^2 \log |\langle\phi_j|U_k|t\rangle|^2 \right) \right]$$

and by an entropic uncertainty relation ([28]) the second of these terms is bounded above by $-\frac{1}{2}\log d$ implying that $I_c(A;B)_\rho \leq \frac{1}{2}\log d$. The locking regime that [12] calculates is illustrated in Figure 1–2. That paper establishes two corners for the accessible information (meaning that the accessible is not necessarily a straight line as illustrated). The corner labeled [DHL+04] indicates that with one qubit register missing, Bob can only establish $\frac{1}{2}\log d$ bits of correlation with Alice.

Locking with random bases [21] The first improvement to the [12] result was made by [21] via the use of random unitary operations. Consider precisely the state ρ described in the first example by [12] but where the register k is allowed to have a larger dimension. The index k then enumerates a set of unitaries acting uniformly on Bob's register. If the unitary matrices are chosen randomly according to the Haar measure (Definition 2.0.3) then Bob fails to infer Alice's register with overwhelmingly high probability. The number of unitary matrices needed to achieve this gives the key-size necessary for locking. In this case the authors achieve $I_c(A;B)_\rho \leq 3$ with a key-size of $k = O(\text{polylog}d)$ although they also

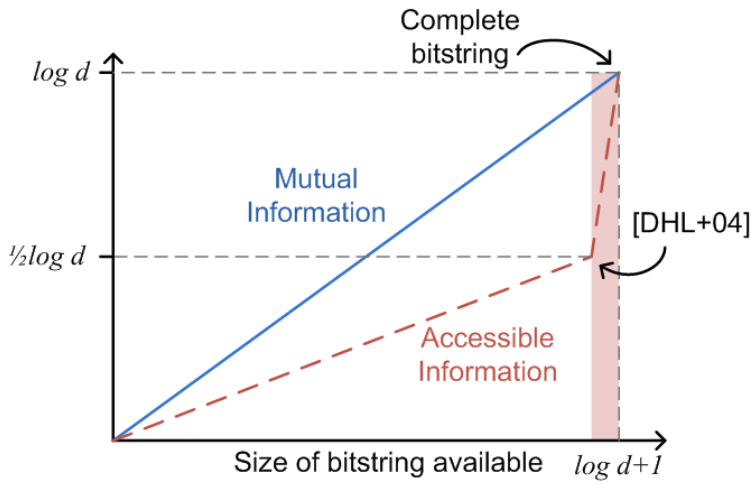

Figure 1–2: The accessible information between Alice and Bob (the dotted red line) is compared to the mutual information (the solid blue line) as a function of the bitstring available to Bob.

show a simple manipulation of their proof that reveals they can also achieve, for a reasonably chosen $\varepsilon > 0$, that $I_c(A;B)_\rho \leq (\log d)\,\varepsilon$ with $k = O(\text{polylog}d - \log \varepsilon)$.

Locking with norm-embeddings [17] Although published almost simultaneously, [17] is actually preceded by the work in this thesis, published as [15]. In this work the authors demonstrate locking as a consequence of *metric uncertainty relations*. A metric uncertainty relation is a statement about the trace distance between a given probability distribution and the uniform distribution. In particular, the authors consider the probability distribution induced by the following coefficients

$$p^A_{U_k|\psi\rangle}(a) = \sum_{b=0}^{d_B-1} \left|\langle a|^A \langle b|^B U_k|\psi\rangle\right|^2 .$$

The intuition is that systems A and B form the cyphertext and key respectively. The collection of t unitaries $\{U_0, ..., U_{t-1}\}$ is said to be a metric uncertainty

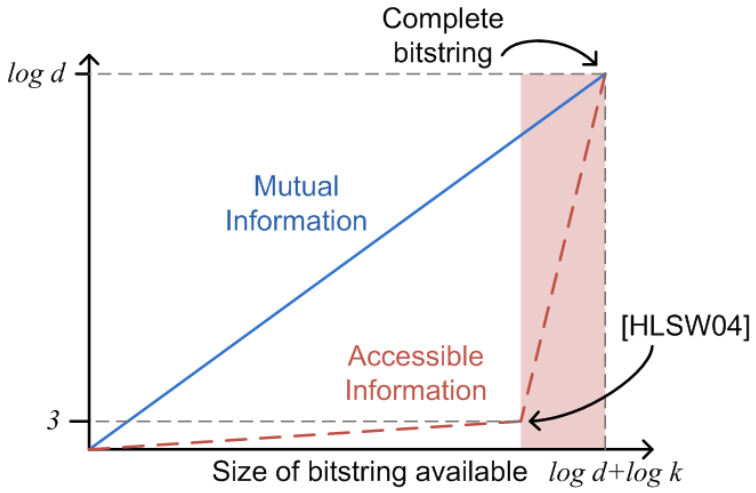

Figure 1–3: The results of [21] gave a new corner specifying the locking region. Although the accessible information is illustrated as a straight line, this is not necessarily the case.

relation if

$$\frac{1}{t}\sum_{k=0}^{t-1}\Delta\left(p^{A}_{U_k|\psi\rangle}, unif([d_A])\right)$$

where $unif$ is the uniform distribution and $\Delta(p,q) = \frac{1}{2}\sum_x |p(x)-q(x)|$, the trace distance between the probability distributions. It is clear that if a metric uncertainty relation holds true for a sufficiently small system B but not for a slightly smaller system, then the trace distance exhibits a locking-type effect. In relation to previous results, a statement about a probability distributions' distance to the uniform distribution directly implies the desired statement about accessible information (see Lemma 2.1.1). In the key/cyphertext language, the authors prove that a locking scheme exists for a key of size $O(\log(1/\varepsilon))$ for a cyphertext of size $\log d - 2\log(18/\varepsilon)$. Note here that the trade-off previously observed between the key size and the size of the cyphertext is eliminated.

1.5 Concentration of measure

The concentration of measure phenomenon is a useful tool for calculating the likelihood of an event vis-a-vis the parameters of a function describing the event. It is precisely the tool used in [21] that permits the authors to claim that the "overwhelming majority of random choices" of a parameter satisfy a desired criterion (in their case, the criterion that a message is locked). The most basic statement of the phenomenon is easily understood graphically and is presented below.

The concentration of measure technique is also often coupled with the use of ε-nets in order to optimize the value of the function in question over a second parameter. The ε-net argument is also easily understood graphically and constitutes the second part of this Section.

1.5.1 Concentration

Consider the 2-dimensional unit-sphere $\mathbb{S}^2$ embedded in 3-dimensional space normalized to have area 1. Any subset of the sphere with area $\frac{1}{2}$ must have a perimeter greater than or equal to the circumference of a greater circle on the sphere. This is because the greater circles form geodesics on the sphere and any subset with a greater circle as its perimeter is a hemisphere cap of the sphere.

We define the r-open-neighbourhood of a set A as

$$A_r = \{x \in \mathbb{S}^2 : d(x, A) < r\}$$

where $d(x, A)$ is the minimum Euclidean distance from a point x to any point in the set A. Visually it is clear that extending an r neighbourhood of a set A with

perimeter greater than the hemisphere cap B increase the total area more than for B_r. This fact is made precise via isoperimetric inequalities in equation (2.5) of [26]

$$\mu(A_r) \geq \mu(B_r), \qquad r > 0$$

where μ is interpreted as the measure of surface area.

Definition 1.5.1 (Concentration function)**.** *For a metric space (X, d) and the probability measure μ we define the* concentration function *as*

$$\alpha_{(X,d,\mu)}(r) = \sup\left\{1 - \mu(A_r) : A \subset X, \mu(A) \geq \frac{1}{2}\right\}$$

where the supremum is taken over all subsets A.

Thus the concentration function gives the worst-case complementary volume to the r-neighbourhood of a hemisphere cap. If we calculate the complementary volume to B_r on $\mathbb{S}^2$ we will find an upper bound on $\alpha_{(\mathbb{S}^2,d,\mu)}(r)$. The calculation is a fairly simple geometric integral and for more general n-dimensional spheres yields the bound in Theorem 2.3 in [26]:

$$\alpha_{(\mathbb{S}^n,d,\mu)}(r) \leq e^{-(n-1)r^2/2}, \qquad r > 0.$$

Note that, although for $\mathbb{S}^2$ the decrease with respect to r is not surprisingly fast, as the dimension of the sphere grows, the concentration increases rapidly in rate (or "strength"). The concentration of $(\mathbb{S}^n, d)$ indeed fits a more general form known as *normal concentration*:

$$\alpha_{(X,d,\mu)}(r) \leq Ce^{-cr^2}, \qquad r > 0$$

where the constants c and C are properties of the metric space and measure. We will later see that another important metric space (namely the unitary group equipped with the Hilbert-Schmidt norm) also has normal concentration.

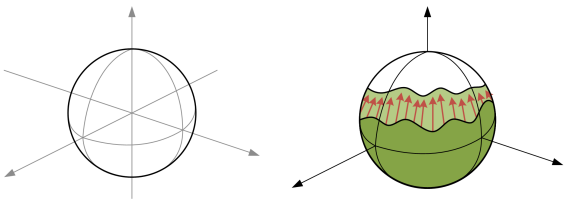

Figure 1–4: A visual representation of the concentration of measure phenomenon for $\mathbb{S}^2$. Any extension to the Southern hemisphere quickly covers the Northern hemisphere.

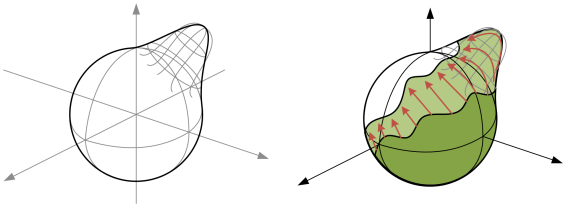

Figure 1–5: A visual representation of concentration over a function on $\mathbb{S}^2$. For a well-behaved function, extensions to any cap still follow a similar behaviour.

Finally, we wish to be able to make, as hinted above, statements about real-valued functions on (X, d, μ). We may, for example, be interested in the space of quantum states and the value of an information theoretic quantity over this space. The result we state below is also sometimes refered to as a statement about the "large deviations" of a function for reasons that will become clear.

Proposition 1.5.1 (Deviation inequality (Proposition 1.3 in [26]))**.** *Let (X, d) be a metric space with probability measure μ and F be a real-valued continuous function on (X, d) with Lipschitz constant θ (see Definition A.0.4) then*

$$\mu\left(\left\{\left|F - \int F \mathrm{d}\mu\right| \geq \varepsilon\right\}\right) \leq \alpha_{(X,d,\mu)}(\varepsilon/\theta)$$

for any ε such that $2\alpha(\varepsilon) < \frac{1}{2}$.

We thus have a bound on the likelihood that a function will deviate from its average value by more than ε.

1.5.2 ε-nets and union bounds

A natural consideration one may have after the results of the previous Section would be the following: given that a function of two parameters is close to uniform with overwhelmingly high probability in one parameter (probability approaching 1 asymptotically in the dimension), what is the probability that the worst case value for the second parameter yields a large deviation? As it turns out, we are already well-equipped to answer this question via our use of the Lipschitz constant θ above.

We have considered so far metric spaces X. Consider however a bounded metric space X (such as, for example, the subspace of normalized quantum states in Hilbert space) then for any fixed ε there exists $N(\varepsilon) > 0$ and $\left\{z_0, z_1, ..., z_{N(\varepsilon)-1}\right\} \subset X$ such that

$$X \subset \bigcup_{k=0}^{N(\varepsilon)-1} \{y : d(y, z_k) < \varepsilon\}.$$

In other words, a bounded metric space can always be covered by a finite number of open ε-balls. The centers of these (i.e.: z_k) are known as the elements of an ε-net. Formally, an ε-net can be defined as

Definition 1.5.2 (ε-net [25])**.** *For any bounded metric space (X, d) there exists an ε-net $\mathcal{J}$ of finite size $N(\varepsilon)$,*

$$\mathcal{J} = \left\{z_0, z_1, ..., z_{N(\varepsilon)-1}\right\}$$

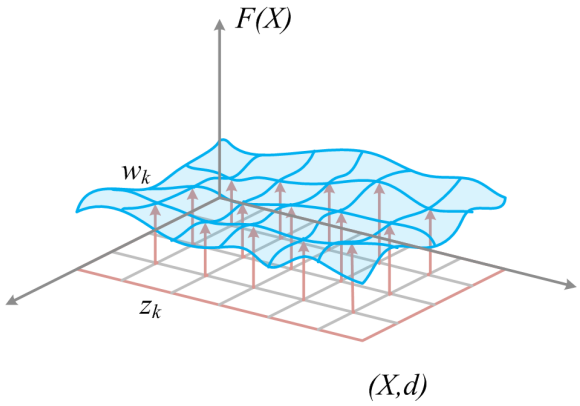

Figure 1–6: An ε-net on the space (X, d) maps to a collection of values of the function F.

such that every element of X *is at most* ε *away from an element of* $\mathcal{J}$ *in the distance* d.

We now examine the behaviour of a function F on the net $\mathcal{J}$. Assign the set $\{w_k = F(z_k)\}$ then, since F is θ-Lipschitz continuous on a bounded metric space, the image of F lies in the interval

$$\{F(x) : x \in X\} \subseteq \left[\min_k \{w_k - \theta\varepsilon\}, \max_k \{w_k + \theta\varepsilon\}\right].$$

A continuous function on a bounded set will achieve its supremum and infimum on the same set. In particular, consider that F achieves its extrema on $x_{\min}$ and $x_{\max}$, then each of these lies in the open ball of an element of $\mathcal{J}$ and we choose those net elements $z_{\min}$ and $z_{\max}$. Finally, since the distance between $x_{\min}$ and $z_{\min}$ is at most ε, $|F(z_{\min}) - F(x_{\min})| < \theta\varepsilon$, and similarly for the maximum. This property is illustrated further in Figure 1–7.

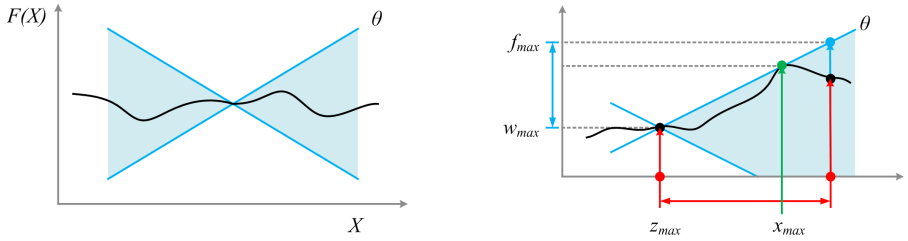

Figure 1–7: One can think of the Lipschitz constant for a 1-dimensional function as the minimum slope for two cones such that for any point on the function, all other points lie within the cones. Thus if a function F is θ-Lipschitz continuous, the maximum value of F is no more than $\theta\varepsilon$ greater than the value at an existing net element.

Recall that at the beginning of this Section we referred to a function of two parameters. As an example, consider a function describing the likelihood of success of some communication protocol between Alice and Bob. The first parameter X may control the first part of the process (e.g.: Alice encoding a message) and due to concentration of measure is successful with exponentially good probability. The second parameter Y may control the second part of the process (e.g.: Bob decoding a message) and is only successful if Bob optimizes over his parameter to reveal some large deviation. We would then like a bound for

$$\mu\left(\left\{\left|\sup_Y F(X,Y) - \int F(X,Y)\mathrm{d}_\mu X\right| \geq \varepsilon\right\}\right).$$

Consider the ε'-net $\mathcal{J} = \left\{y_1, ..., y_{N(\varepsilon')}\right\}$ over inputs Y, then the statement for a fixed y_k is the same as in Proposition 1.5.1

$$\mu\left(\left\{\left|F(X,y_k) - \int F(X,y_k)\mathrm{d}_\mu X\right| \geq \varepsilon\right\}\right) \leq \alpha_{(X,d,\mu)}(\varepsilon/\theta_X)$$

where θ_X is the Lipschitz constant for F with respect to the parameter X. We note now the following fact,

$$\begin{aligned}
&\mu\left(\left\{\left|\max_k F(X,y_k) - \int F(X,y_k)\mathrm{d}_\mu X\right| \geq \varepsilon\right\}\right) \\
&\leq \mu\left(\bigcup_k \left\{\left|F(X,y_k) - \int F(X,y_k)\mathrm{d}_\mu X\right| \geq \varepsilon\right\}\right) \\
&\leq \sum_k \mu\left(\left\{\left|F(X,y_k) - \int F(X,y_k)\mathrm{d}_\mu X\right| \geq \varepsilon\right\}\right)
\end{aligned}$$

which amounts to stating that "the maximum value of F as yielded by net elements is either achieved by the first element, or the second, or the third, or ..." and so on. We can then make the statement about the joint concentration and optimization

$$\mu\left(\left\{\left|\max_k F(X,y_k) - \int F(X,y_k)\mathrm{d}_\mu X\right| \geq \varepsilon\right\}\right) \leq N(\varepsilon')\alpha_{(X,d,\mu)}(\varepsilon/\theta_X).$$

To achieve the desired statement it remains to apply our earlier statement and choose $\varepsilon' = \varepsilon/\theta_Y$,

$$\begin{aligned}
&\mu\left(\left\{\left|\sup_Y F(X,Y) - \int F(X,Y)\mathrm{d}_\mu X\right| \geq \varepsilon\right\}\right) \\
&\leq \mu\left(\left\{\left|\max_k F(X,y_k) - \int F(X,y_k)\mathrm{d}_\mu X\right| \geq 2\varepsilon\right\}\right) \\
&\leq N\left(\frac{\varepsilon}{2\theta_Y}\right)\alpha_{(X,d,\mu)}\left(\frac{\varepsilon}{2\theta_X}\right).
\end{aligned}$$

We see now that the probability of a large deviation in F as optimized against the parameter Y is an asymptotic competition between $N(\cdot)$ which increases as the dimension of Y increases and α which decreases as the dimension of X increases.

In some cases it is interesting to study the question "does there exist a value for X such that a value for Y will yield a large deviation?" In this case, the question reduces to calculating whether the upper bound above is less than 1, i.e.: the set of parameters that yield large deviations has measure strictly greater than 0. We can relax this condition however and require that the set of parameters that yield large deviations has measure at most c, for fixed constant $c < 1$. In this relaxed case, the appropriate interpretation is that we study the question "does there exist a value for X such that none of the possible values for Y yield a large deviation?"

Note that to apply the above ideas we would need to calculate the expectation value of F, the Lipschitz constant θ, the size of the net $N(\varepsilon)$, and the concentration function α.

CHAPTER 2
Generic unitary channels

We will study, as a model of communication between Alice and Bob, the generic unitary channel. This channel takes the form

$$\mathcal{N}^{A\to B}\left(\rho^{A}\right)=\mathrm{Tr}_{E}\left[\left(U^{A\to BE}\right)\rho\left(U^{A\to BE}\right)^{\dagger}\right]$$

where $|A| = |B||E|$ and the unitary matrix U is chosen according to the Haar measure on $U(n)$, the group of unitary matrices. The study of random unitary matrices is well-motivated in physics. Wigner [39] and Dyson [16] developed random matrix theory as a powerful tool for the analysis for nuclear energy levels and the energy levels of complex systems such as one-dimensional Coulomb gas. The study of various random matrix ensembles for analysis of physical systems has been a field of its own since [29]. Amongst these ensembles, the one that models the physics of closed quantum systems is the "circular unitary ensemble" which give rise to the Haar-random unitary matrices defined below,

Definition 2.0.3 (Haar measure)**.** *The* Haar measure $\nu_{U(n)}$ *is the unique measure on* $U(n)$ *that is left-invariant. That is, for any Borel set* $A \subset U(n)$ *and any* $X \in U(n)$*, we have that* $\nu_{U(n)}(A) = \nu_{U(n)}(XA)$*. The Haar measure on* $U(n)$ *is also right-invariant. We also assume that* $\nu_{U(n)}(U(n)) = 1$ *and we denote by* Pr_U *the probability measure induced by the Haar measure.*

If we recall the previous Section, we can note that our observations apply directly to the probability space $(U(n), \|\cdot\|_2, \nu_{U(n)})$ and functions $f : U(n) \to \mathbb{R}$. In fact, the unitary group $U(n)$ exhibits normal concentration ([5]) which yields the following result,

Theorem 2.0.2 (Corollary 4.4.28 in [5])**.** *Let $f : \mathcal{U}(n) \to \mathbb{R}$ be a function with Lipschitz constant θ (see Definition A.0.4; the Lipschitz constant is taken with respect to the Hilbert-Schmidt distance on unitaries). Then,*

$$\Pr_U \{|f(U) - \mathbb{E}_U f| > \varepsilon\} \leqslant \exp\left(-\frac{n\varepsilon^2}{4\theta^2}\right).$$

In the original work [5], the expectation value is defined to be

$$\mathop{\mathbb{E}}_{U} f(\cdot) = \int f(Y\cdot)\mathrm{d}\nu_{SU(n)}(Y).$$

The use of the measure $\mathrm{d}\nu_{SU(n)}(Y)$ (that is, the Haar measure over the special unitary group) is artefact from the proof technique in [5]. We can, however, define the expectation as an integral over $\mathrm{d}\nu_{U(n)}$ in the special case of *class functions* allowing the use of other known results for integrals over the Haar measure. We prove the following Lemma.

Lemma 2.0.3. *Given a class function such that $f(X) = f(e^{i\theta}X)$ for any $X \in U(n)$ and any $\theta \in [0, 2\pi)$, we have that*

$$\int f(YX)\mathrm{d}\nu_{SU(n)}(Y) = \int f(Y)\mathrm{d}\nu_{U(n)}(Y).$$

Proof. Consider first that

$$U(n) = \{ST : T \in SU(n), S \in H(n)\} = \{e^{i\theta}T : T \in SU(n), \theta \in [0, 2\pi)\}$$

where $H(n)$ is the the subset of $U(n)$ consisting of scalar multiples of the identity. We have by Lemma 4.4.29 in [5] that

$$\int f(XY)\mathrm{d}\nu_{U(n)}(Y) = \int\int f(XST)\ \mathrm{d}\nu_{SU(n)}(T)\ \mathrm{d}_{H(n)}(S).$$

Since S and T commute and f is a class function, we have that

$$\int f(XY)\mathrm{d}\nu_{U(n)}(Y) = \int f(XT)\mathrm{d}\nu_{SU(n)}(T).$$

Finally, we note that $\mathrm{d}\nu_{U(n)}(Y) = \mathrm{d}\nu_{U(n)}(X^{-1}Y)$ and we recover the desired statement. □

We are now free to use the intuitive definition of the expectation value

$$\mathop{\mathbb{E}}_{U} f := \int f(X)\mathrm{d}\nu_{U(n)}(X) = \int f(U)\mathrm{d}U$$

where we've introduced our simplified notation in this special case.

2.1 Circuit

To end the introduction, we introduce the physical scenario that will occupy us throughout this work. The quantum circuit depicted in Figure 2–1 is our model for any closed quantum system modeled by generic unitary dynamics. The lines running from left to right indicate the subsystems of the various quantum states as labeled. Where two such lines join, the state is described by the composite system. The boxes indicate superoperators, maps of states, or channels. A double line indicates a classical message encoded in a quantum state. The dotted line is used to denote the intermediate state ρ. The classical message M is encoded in N, and the unitary U_{CKE} then mixes it with the E part of the shared entanglement

$|\omega\rangle$. If the information is locked, any joint measurement $\mathcal{M}$ on C and E' will yield a result X that is almost independent of the message. On the other hand, if C is large enough, there will be a joint measurement $\mathcal{M}$ reliably decoding M. The brace represents the classical distributions for which the accessible information is calculated.

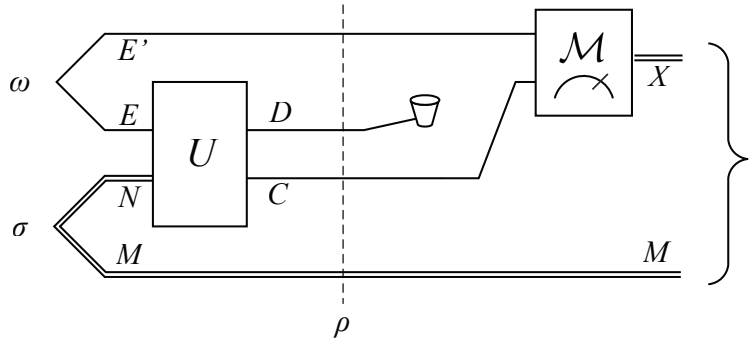

Figure 2–1: A quantum circuit depicting the physical scenario.

Now, let $\{|\psi_m\rangle : 1 \leqslant m \leqslant |M|\}$ be any orthonormal basis for N. The analysis will focus on the properties of the states

$$\sigma^{MN} := \sum_{m=1}^{|M|} p_m |m\rangle\langle m|^M \otimes |\psi_m\rangle\langle\psi_m|^N \quad \text{and} \tag{2.1.1}$$

$$\rho^{MCDE'} := \left(\mathbb{I}^{ME'} \otimes U^{NE\to CD}\right)\left(\sigma^{MN} \otimes \omega^{EE'}\right)\left(\mathbb{I}^{ME'} \otimes U^{NE\to CD}\right)^{\dagger}. \tag{2.1.2}$$

Our objective is to demonstrate that until C is large enough that there exists a measurement on CE' capable of revealing *all* the information about the message M, no measurement will reveal *any* information about the message. This cannot quite be true, of course, so what we will demonstrate is that the jump from no

information to complete information involves enlarging C by a number of qubits logarithmic in the size of the message M and the amount of entanglement E.

In the following example, assume that M is uniformly distributed and that the state $\omega^{EE'}$ is maximally entangled. As a first step, it is necessary to determine how large C needs to be in order for there to exist a measurement on CE' that will reveal the message M. Begin by purifying the state σ to

$$|\sigma\rangle^{RMN} = \frac{1}{\sqrt{|M|}} \sum_{m=1}^{|M|} |m\rangle^R \otimes |m\rangle^M \otimes |\psi_m\rangle^N. \tag{2.1.3}$$

Even more demanding than performing a measurement to reveal m is the task of transmitting the quantum information about RM through U, allowing the decoder, who has access only to CE', to recover a high fidelity copy of the state $|\sigma\rangle^{RMN}$. If U is selected according to the Haar measure, then Theorem IV.1 of [1] implies that there is a quantum operation $\mathcal{D}^{CE'\to N}$ acting only on CE' such that

$$\left\| \mathcal{D}\left(\mathrm{Tr}_D \left[U^{NE\to CD} (\sigma^{RMN} \otimes \omega^{EE'}) (U^{NE\to CD})^\dagger \right] \right) - \sigma^{RMN} \right\|_1 \leq 2\sqrt{\frac{M}{C}}. \tag{2.1.4}$$

Because the trace distance is monotonic under quantum operations, it will not increase by taking the partial trace over R and measuring in the basis $\{|\psi_m\rangle\}$ [30]. If we let $p(m'|m)$ be the probability of getting an outcome $|\psi_{m'}\rangle$ when the message was in fact m, Equation (2.1.4) therefore implies that

$$\frac{1}{M} \sum_m \sum_{m' \neq m} p(m'|m) \leq \sqrt{\frac{M}{C}}. \tag{2.1.5}$$

In words, the probability of the measurement yielding the incorrect outcome, averaged over all messages, is at most $\sqrt{M/C}$, so as soon as C is significantly

larger than M, a measurement on CE' can be found that will reveal the message. Our goal in this thesis will be to demonstrate that until this condition is met, no measurement will reveal any significant information about the message.

We need to introduce the concept of *quasi-measurements* for our analysis. They are, as their name indicates, almost measurements, but differ in three ways: they only contain rank-one elements of equal weight, they have exactly s outcomes, and the sum of all the elements does not necessarily equal the identity, but is instead bounded by $\eta\mathbb{I}$:

Definition 2.1.1 (Quasi-measurement)**.** *We call a superoperator $\mathcal{M}^{A\to B}$ an* (s,η)-quasi-measurement *if it is of the form $\mathcal{M}(\rho) = \frac{|A|}{s}\sum_{i=1}^{s}|i\rangle\langle\chi_i|\rho|\chi_i\rangle\langle i|$ where the $|i\rangle$ index an orthonormal basis for B, and $\frac{|A|}{s}\sum_{i=1}^{s}|\chi_i\rangle\langle\chi_i| \leqslant \eta\mathbb{I}^A$. We call the set of all (s,η)-quasi-measurements on a given system, $\mathcal{L}(s,\eta)$.*

The reason for introducing these, as will soon become apparent, is that they are almost equivalent to POVMs for our purposes while being much easier to handle mathematically. It can easily be seen that projective measurements are simply $(A,1)$-quasi-measurements. Note that we have begun to omit the absolute value $|\cdot|$ to indicate system size.

Definition 2.1.2 (Projective measurement (equivalent to Definition 1.3.4))**.** *We call a superoperator $\mathcal{M}^{CE'\to X}$ a projective measurement if it is a $(CE,1)$-quasi measurement.*

Note that although in our definition of a quasi-measurement we do not require the $|\chi_i\rangle$ to be orthogonal, in the case of a projective measurement they are indeed.

Alternatively, we can write any $\mathcal{M}$ as

$$\mathcal{M}^{CE'\to X}(\rho) = \frac{CE}{n}\sum_{i=1}^{n} |i\rangle\langle i|^X \mathrm{Tr}_{CE'}\left[|\chi_i\rangle\langle\chi_i|\rho\right]$$

We now give the formal, strengthened definition of locking. Because the cyphertext will always be smaller than or equal to the message when locking occurs, certain identifications become possible. In particular, we can assume without loss of generality that $N \cong C \otimes K$ and $D \cong E \otimes K$. Since the analysis will be performed using only C, K and E, we reproduce the illustration of the physical scenario with the identifications made in Figure 2–2.

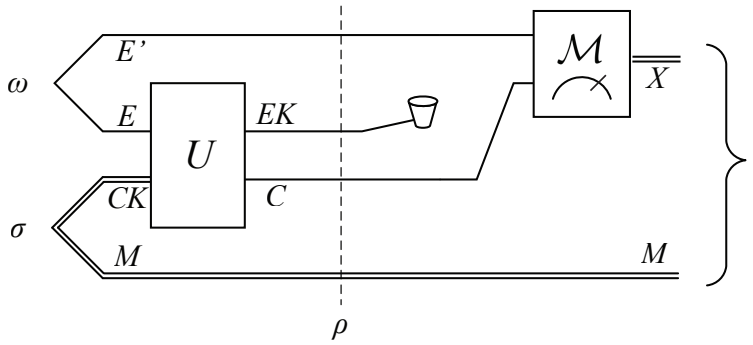

Figure 2–2: A quantum circuit depicting the physical scenario with the locking-specific identifications $N \cong C \otimes K$ and $D \cong E \otimes K$ made.

Definition 2.1.3 (ε-locking scheme)**.** *Let M, C, K, E and E' be quantum systems. Let $\rho^{MCKEE'}$ be a quantum state of the form*

$$\rho^{MCKEE'} = \sum_m p_m U^{CKE}\left(|m\rangle\langle m|^M \otimes |\psi_m\rangle\langle\psi_m|^{CK} \otimes |\omega\rangle\langle\omega|^{EE'}\right) U^{CKE\dagger}, \quad (2.1.6)$$

where the $|\psi_m\rangle$ *are orthogonal and* U^{CKE} *is unitary. Then we call* ρ *an* ε*-locking scheme if for any measurement superoperator* $\mathcal{M}^{CE'\to X}$*, we have that*

$$\left\| \mathcal{M}\left(\rho^{MCE'}\right) - \mathcal{M}\left(\rho^{M} \otimes \rho^{CE'}\right) \right\|_1 \leqslant \varepsilon.$$

Note that this definition of locking is rather different from that used in previous work in the area (see Section 1.4.1). Their definition involved the *accessible information* between the cyphertext and the message. However, this trace-distance definition has a very natural interpretation. It bounds the largest probability for which we can guess, given a message m and the result x of a measurement done on a cyphertext, whether x comes from a valid cyphertext for m or from a cyphertext generated independently of m. In other words, one could almost perfectly reproduce any measurement results made on a valid cyphertext without having access to the cyphertext at all. We can also show that our definition implies the older one:

Lemma 2.1.1. *Let* ξ^{MB} *be a cq-state such that* $\left\|\mathcal{M}(\xi^{MB}) - \xi^{M} \otimes \mathcal{M}(\xi^{B})\right\|_1 \leqslant \varepsilon$ *for all measurement superoperators* $\mathcal{M}^{B\to X}$*. Then,*

$$I_{\mathrm{acc}}(M;B)_{\xi} \leqslant 4\varepsilon \log M + 2\eta(1-\varepsilon) + 2\eta(\varepsilon),$$

where $\eta(x) := -x \log x$ *and* $\eta(0) = 0$*.*

Proof. This is a direct application of the Alicki-Fannes inequality [4]. □

To show that locking has indeed occurred, our calculations will concern the function

$$\begin{aligned} f &: \mathcal{U}(CKE) \times \mathcal{L}(s,\eta) \longrightarrow \mathbb{R} \\ f(U,\mathcal{M}) &= \left\| \mathcal{M}(\rho^{MCE'}) - \mathcal{M}(\rho^{M} \otimes \rho^{CE'}) \right\|_1 . \end{aligned} \tag{2.1.7}$$

where we recall that $\mathcal{L}(s,\eta)$ denotes the space of all (s,η)-quasi-measurements. On occasion, when it does not cause confusion, we will denote by $g_{\mathcal{M}}(U)$ the function $f(U,\mathcal{M})$ with fixed $\mathcal{M}$, and by $h_U(\mathcal{M})$ the function $f(U,\mathcal{M})$ with fixed U. We note that equation (2.1.7) is a measure of independence of the message M from the results of the measurement X.

Four quantities will be particularly useful for quantifying variations from uniform messages and maximal entanglement,

$$\Delta_{M,\infty} := 2^{\log M - H_{\min}(M)_\sigma}, \tag{2.1.8}$$

$$\Delta_{M,2} := 2^{\log M - H_2(M)_\sigma}, \tag{2.1.9}$$

$$\Delta_{E,\infty} := 2^{\log E - H_{\min}(E)_\omega}, \tag{2.1.10}$$

$$\Delta_{E,2} := 2^{\log E - H_2(E)_\omega}. \tag{2.1.11}$$

For a pure classical distribution p_m, $\Delta_{M,\infty} = \Delta_{M,2} = |M|$ and for the uniform distribution $\Delta_{M,\infty} = \Delta_{M,2} = 1$. To give an interpretation of the Δ_E quantities, we can note that for a bipartite state $|\omega\rangle^{EE'}$ with no entanglement, $\Delta_{E,\infty} = \Delta_{E,2} = |E|$. However, if $|\omega\rangle^{EE'}$ is the maximally entangled state, then $\Delta_{E,\infty} = \Delta_{E,2} = 1$, which we call maximal entanglement. The case of a uniformly distributed message and maximal entanglement will give the simplest expressions for minimum key

size. The Δ terms are used in the calculations to provide more general statements relating the entropy of the message and entanglement to the key size.

CHAPTER 3
Locking Results

3.1 Concentration

To be able to use the general concentration of measure Theorem (Theorem 2.0.2) on $g_{\mathcal{M}}(U)$, we must first be able to upper-bound the expectation of $g_{\mathcal{M}}(U)$ with respect to U. The following Lemma does this:

Lemma 3.1.1 (Distinguishability for a fixed measurement)**.** *If $\mathcal{M}^{CE'\to X}$ is an (s,η)-quasi-measurement, then*

$$\mathop{\mathbb{E}}_{U} \left\| \mathcal{M}\left(\rho^{MCE'}\right) - \mathcal{M}\left(\rho^{M}\otimes\rho^{CE'}\right)\right\|_1 \leqslant \frac{2\Delta_{E,\infty}}{\sqrt{KE}}.$$

Proof. We begin by expanding and simplifying the original expression

$$\begin{aligned}
&\mathop{\mathbb{E}}_{U} \left\| \mathcal{M}\left(\rho^{MCE'}\right) - \mathcal{M}\left(\rho^{M}\otimes\rho^{CE'}\right)\right\|_1 && (3.1.1)\\
&= \mathop{\mathbb{E}}_{U} \left\| \mathcal{M}\left(\rho^{MCE'} - \rho^{M}\otimes\rho^{CE'}\right)\right\|_1 && (3.1.2)\\
&\leqslant \mathop{\mathbb{E}}_{U} \sqrt{s\mathrm{Tr}\left[\left((\sigma^M)^{-1/4}\,\mathcal{M}\left(\rho^{MCE'} - \rho^{M}\otimes\rho^{CE'}\right)\,(\sigma^M)^{-1/4}\right)^2\right]}\\
&\leqslant \sqrt{s\mathop{\mathbb{E}}_{U}\mathrm{Tr}\left[\left((\sigma^M)^{-1/4}\,\mathcal{M}\left(\rho^{MCE'} - \rho^{M}\otimes\rho^{CE'}\right)\,(\sigma^M)^{-1/4}\right)^2\right]}. && (3.1.3)
\end{aligned}$$

In the manipulations above, we have used the linearity of the superoperator $\mathcal{M}$ in the first line. In the second line we have used Lemma A.0.10 with $\gamma = \mathbb{I}^X \otimes \sigma^M$, recalling the definition of σ^M in 2.1.1 and noting that $|X| = s$. The third line follows from the concavity of the square root. We will now use a helpful identity

for the trace of an operator squared: $\mathrm{Tr}Z^2 = \mathrm{Tr}(Z \otimes Z)F$, where F is defined as follows.

Definition 3.1.1. *The* swap operator *on $A^{\otimes 2}$, which is written as $A \otimes \overline{A}$, is the unique linear operator F^A satisfying*

$$F^A\left(|\psi\rangle^A|\phi\rangle^{\overline{A}}\right) = |\phi\rangle^A|\psi\rangle^{\overline{A}} \qquad \forall|\psi\rangle, |\phi\rangle.$$

Expressing Equation (3.1.3) using the swap operator gives

$$\begin{aligned}
&\mathop{\mathbb{E}}_U \left\| \mathcal{M}\left(\rho^{MCE'}\right) - \mathcal{M}\left(\rho^M \otimes \rho^{CE'}\right)\right\|_1 && (3.1.4)\\
&\leqslant \sqrt{s\mathrm{Tr}\left[\mathop{\mathbb{E}}_U \left((\sigma^M)^{-1/4}\,\mathcal{M}\left(\rho^{MCE'} - \rho^M \otimes \rho^{CE'}\right)\,(\sigma^M)^{-1/4}\right)^{\otimes 2} F^{XM}\right]} \\
&= \left(\frac{(CE')^2}{s}\sum_{i=1}^{s}\mathrm{Tr}\left[\left(F^M \otimes \left(\chi_i^{CE'}\right)^{\otimes 2}\right)\right.\right. \\
&\qquad \left.\left.\mathop{\mathbb{E}}_U\left[(\sigma^M)^{-1/4}(\rho^{MCE'} - \rho^M \otimes \rho^{CE'})(\sigma^M)^{-1/4}\right]^{\otimes 2}\right]\right)^{1/2} && (3.1.5)
\end{aligned}$$

Equation (3.1.5) follows from the fact that results of the measurement $\mathcal{M}$ are stored in an orthonormal basis of system X. We will proceed by evaluating $\mathop{\mathbb{E}}_U((\sigma^M)^{-1/4}\,(\rho^{MCE'} - \rho^M \otimes \rho^{CE'})\,(\sigma^M)^{-1/4})^{\otimes 2}$, but before continuing we absorb the two $\sigma^{-1/4}$ operators into the operator ρ. That is we define,

$$\begin{aligned}
\tilde{\sigma}^{MCK} &:= \sum_{m=1}^{|M|}\sqrt{p_m}|m\rangle\langle m|^M \otimes |\psi_m\rangle\langle\psi_m|^{CK} \quad \text{and} && (3.1.6)\\
\tilde{\rho}^{MCKEE'} &:= (\sigma^M)^{-1/4}\,\rho^{MCKEE'}\,(\sigma^M)^{-1/4} \\
&= \left(\mathbb{I}^{ME} \otimes U^{CKE}\right) \cdot \left(\tilde{\sigma}^{MCK} \otimes \omega^{EE'}\right). && (3.1.7)
\end{aligned}$$

With these two definitions in hand we can expand $\mathop{\mathbb{E}}_U \left(\tilde{\rho}^{MCE'} - \tilde{\rho}^M \otimes \rho^{CE'}\right)^{\otimes 2}$ as

$$\begin{aligned}
&\mathop{\mathbb{E}}_U \left(\tilde{\rho}^{MCE'} - \tilde{\rho}^M \otimes \rho^{CE'}\right)^{\otimes 2} \qquad (3.1.8)\\
&= \mathop{\mathbb{E}}_U \left(\mathrm{Tr}_{KE}\left[U^{CKE} \cdot \left(\left(\tilde{\sigma}^{MCK} - \tilde{\sigma}^M \otimes \sigma^{CK}\right) \otimes \omega^{EE'}\right)\right]\right)^{\otimes 2}\\
&= \mathrm{Tr}_{KE\overline{KE}}\left[\mathop{\mathbb{E}}_U \left(U^{CKE} \cdot \left(\left(\tilde{\sigma}^{MCK} - \tilde{\sigma}^M \otimes \sigma^{CK}\right) \otimes \omega^{EE'}\right)\right)^{\otimes 2}\right]\\
&= \mathrm{Tr}_{KE\overline{KE}}\left[\int \left(U^{CKE} \otimes U^{\overline{CKE}} \otimes \mathbb{I}^{ME'\overline{ME'}}\right) \cdot \right.\\
&\qquad\qquad \left.\left(\left(\tilde{\sigma}^{MCK} - \tilde{\sigma}^M \otimes \sigma^{CK}\right) \otimes \omega^{EE'}\right)^{\otimes 2} \mathrm{d}U\right]. \qquad (3.1.9)
\end{aligned}$$

To evaluate the integral with Lemma A.0.7, we will need to calculate the projections of our operator onto the symmetric and antisymmetric subspaces of $(CKE)^{\otimes 2}$. Since the projectors onto the symmetric and antisymmetric subspaces can be written as $\Pi_\pm = \frac{1}{2}(\mathbb{I} \pm F)$, we can arrive at same results by working with $\mathbb{I}$ and F. We begin with $\mathbb{I}$:

$$\begin{aligned}
&\mathrm{Tr}_{CKE\overline{CKE}}\left[\left(\tilde{\sigma}^{MCK} - \tilde{\sigma}^M \otimes \sigma^{CK}\right)^{\otimes 2} \otimes \left(\omega^{EE'}\right)^{\otimes 2} \mathbb{I}^{CKE\overline{CKE}}\right] \qquad (3.1.10)\\
&= \sum_m \sqrt{p_m}|m\rangle\langle m|^M \otimes \sum_{m'} \sqrt{p_{m'}}|m'\rangle\langle m'|^{\overline{M}} \mathrm{Tr}_{CK}\left[\psi_m - \sum_{m''} p_{m''}\psi_{m''}\right]^2 \otimes \left(\omega^{E'}\right)^{\otimes 2}\\
&= (1-1)^2 \cdot (\tilde{\sigma}^M \otimes \omega^{E'})^{\otimes 2} = 0.
\end{aligned}$$

The projection onto F requires a more subtle calculation,

$$\begin{aligned}
&\mathrm{Tr}_{CKE\overline{CKE}}\left[\left(\tilde{\sigma}^{MCK}-\tilde{\sigma}^{M}\otimes\sigma^{CK}\right)^{\otimes 2}\otimes\left(\omega^{EE'}\right)^{\otimes 2}F^{CKE}\right] \qquad (3.1.11)\\
&= \sum_{m}\sqrt{p_m}|m\rangle\langle m|^{M}\otimes\sum_{m'}\sqrt{p_{m'}}|m'\rangle\langle m'|^{\overline{M}}\\
&\quad\cdot\mathrm{Tr}_{CK}\left[\left(\psi_m-\sum_{m''}p_{m''}\psi_{m''}\right)\left(\psi_{m'}-\sum_{m'''}p_{m'''}\psi_{m'''}\right)\right]\otimes\mathrm{Tr}_{E\overline{E}}\left[\left(\omega^{EE'}\right)^{\otimes 2}F^{E}\right].
\end{aligned}$$

By taking a closer look at Equation (3.1.11) we can make the simplification

$$\begin{aligned}
&\mathrm{Tr}_{CK}\left[\left(\psi_m-\sum_{m''}p_{m''}\psi_{m''}\right)\left(\psi_{m'}-\sum_{m'''}p_{m'''}\psi_{m'''}\right)\right]\\
&= \mathrm{Tr}_{CK}\left[\psi_m\psi_{m'}-\sum_{m''}p_{m''}\psi_{m''}\psi_{m'}-\sum_{m'''}p_{m'''}\psi_m\psi_{m'''}+\sum_{m'',m'''}p_{m''}p_{m'''}\psi_{m''}\psi_{m'''}\right]\\
&= \delta_{mm'}-p_{m'}-p_m+\sum_{m''}p_{m''}^2. \qquad (3.1.12)
\end{aligned}$$

Now we define $\tilde{\sigma}_{\circ}^{M\overline{M}}$ as the quantity evaluated in Equation (3.1.11). Substituting the result of Equation (3.1.12) gives

$$\begin{aligned}
\tilde{\sigma}_{\circ}^{M\overline{M}} &:= \mathrm{Tr}_{CK\overline{CK}}\left[\left(\tilde{\sigma}^{MCK}-\tilde{\sigma}^{M}\otimes\sigma^{CK}\right)^{\otimes 2}F^{CK}\right]\\
&= \left(\sum_{m}p_m\left(|m\rangle\langle m|\right)^{\otimes 2}-\tilde{\sigma}^{M}\otimes(\tilde{\sigma}^{\overline{M}})^3-(\tilde{\sigma}^{M})^3\otimes\tilde{\sigma}^{\overline{M}}+\left(\sum_{m}p_m^2\right)\tilde{\sigma}^{M}\otimes\tilde{\sigma}^{\overline{M}}\right).
\end{aligned}$$

We also define $\Omega^{E'\overline{E'}}$ as the operator acting on system $E'\overline{E'}$ in Equation (3.1.11), or $\Omega^{E'\overline{E'}}=\mathrm{Tr}_{E\overline{E}}[(\omega^{EE'})^{\otimes 2}F^{E}]$. At this point, Lemma A.0.7 can be used to evaluate the integral in Equation (3.1.9). We can make significant simplifications by first expanding the $\alpha_{\pm}$ and then using our result from Equation (3.1.10) to

show that

$$\begin{aligned}\alpha_{\pm} &= \frac{1}{\operatorname{rank}(\Pi_{\pm}^{CKE})}\operatorname{Tr}_{CKE\overline{CKE}}\left[\left(\tilde{\sigma}^{MCK}-\tilde{\sigma}^{M}\otimes\sigma^{CK}\right)^{\otimes 2}\otimes\left(\omega^{EE'}\right)^{\otimes 2}\left(\Pi_{\pm}^{CKE}\otimes\mathbb{I}^{ME'}\right)\right]\\ &= \frac{\pm\left(\tilde{\sigma}_{\circ}^{M\overline{M}}\otimes\Omega^{E'\overline{E'}}\right)}{CKE(CKE\pm 1)},\end{aligned}$$

where the terms $\Pi_{\pm}^{CKE}$ are the projectors onto the symmetric and antisymmetric subspaces of $(CKE)^{\otimes 2}$, that is $\frac{1}{2}(\mathbb{I}^{CKE\overline{CKE}}\pm F^{CKE})$. In particular, because α_+ is proportional to α_-, the integral will have the product form

$$\tilde{\sigma}_{\circ}^{M\overline{M}}\otimes\Omega^{E'\overline{E'}}\otimes\left(\frac{\Pi_{+}^{CKE}}{CKE(CKE+1)}-\frac{\Pi_{-}^{CKE}}{CKE(CKE-1)}\right),$$

so the calculation of the trace in Equation (3.1.5) will factor into a product over the systems $(M)^{\otimes 2}$ and $(CKEE')^{\otimes 2}$. Thus,

$$\begin{aligned}&\operatorname{Tr}_{(MKE)^{\otimes 2}}\left[\left(\operatorname{Tr}_{(CE')^{\otimes 2}}\left[\left(\chi_i^{CE'}\right)^{\otimes 2}\mathop{\mathbb{E}}_{U}\left(\tilde{\rho}^{MCE'}-\tilde{\rho}^{M}\otimes\rho^{CE'}\right)^{\otimes 2}\right]\right)F^{M}\right] \qquad (3.1.13)\\ &= \operatorname{Tr}\left[\tilde{\sigma}_{\circ}^{M\overline{M}}F^{M}\right]\cdot\operatorname{Tr}\left[\left(\chi_i^{CE'}\otimes\mathbb{I}^{KE}\right)^{\otimes 2}\left(\frac{\Pi_{+}^{CKE}\otimes\Omega^{E'\overline{E'}}}{CKE(CKE+1)}-\frac{\Pi_{-}^{CKE}\otimes\Omega^{E'\overline{E'}}}{CKE(CKE-1)}\right)\right].\end{aligned}$$

The first first factor in Equation (3.1.13) can easily be bounded:

$$\begin{aligned}\operatorname{Tr}_{M\overline{M}}\left[\tilde{\sigma}_{\circ}^{M\overline{M}}F^{M}\right] &= \sum_m p_m-\sum_m p_m^{3/2}-\sum_m p_m^{3/2}+\sum_m p_m^2\\ &\leqslant 2\sum_m p_m=2.\end{aligned}$$

To estimate the second factor in Equation (3.1.13) we will need to observe two facts. First, that

$$\operatorname{Tr}\left[(\chi_i^{CE'}\otimes\mathbb{I}^{KE})^{\otimes 2}\,\mathbb{I}^{CKE\overline{CKE}}\otimes\Omega^{E'\overline{E'}}\right]\leqslant(KE)^2\left\|\Omega^{E'\overline{E'}}\right\|_{\infty}, \qquad (3.1.14)$$

which follows from the fact that $\chi_i^{CE'}$ is a rank 1 projector. Second, that

$$\begin{aligned} \mathrm{Tr}\left[(\chi_i^{CE'} \otimes \mathbb{I}^{KE})^{\otimes 2}\, F^{CKE} \otimes \Omega^{E'\overline{E'}}\right] &= KE\, \mathrm{Tr}_{E'\overline{E'}}\left[\left(\mathrm{Tr}_{C\overline{C}}\left[(\chi_i^{CE'})^{\otimes 2} F^C\right]\right)\Omega^{E'\overline{E'}}\right] \\ &\leqslant KE \left\|\Omega^{E'\overline{E'}}\right\|_\infty . \end{aligned} \tag{3.1.15}$$

If we use Equations (3.1.14) and (3.1.15) to estimate the second factor of Equation (3.1.13) we get the bound

$$\begin{aligned} &\mathrm{Tr}\left[\left(\chi_i^{CE'} \otimes \mathbb{I}^{KE}\right)^{\otimes 2}\left(\frac{\Pi_+^{CKE} \otimes \Omega^{E'\overline{E'}}}{CKE(CKE+1)} - \frac{\Pi_-^{CKE} \otimes \Omega^{E'\overline{E'}}}{CKE(CKE-1)}\right)\right] \\ &\leqslant \left(\frac{(KE)^2 + KE}{2CKE(CKE+1)} - \frac{(KE)^2 - KE}{2CKE(CKE-1)}\right) \cdot \left\|\Omega^{E'\overline{E'}}\right\|_\infty \\ &\leqslant \frac{2}{C^2 KE} \cdot \left\|\Omega^{E'\overline{E'}}\right\|_\infty . \end{aligned} \tag{3.1.16}$$

This can be rewritten in a more familiar form using

$$\begin{aligned} \left\|\Omega^{E'\overline{E'}}\right\|_\infty &= \left\|\mathrm{Tr}_{E'\overline{E'}}\left[\left(\omega^{EE'}\right)^{\otimes 2} F^E\right]\right\|_\infty \\ &= \left\|\left(\omega^E\right)^{\otimes 2} F^E\right\|_\infty = \left\|\omega^E\right\|_\infty^2 = 2^{-2H_{\min}(E)_\omega}. \end{aligned}$$

In the above, the third equality follows from the fact that the operator norm is right-invariant under unitary transformations and F is a unitary matrix. Combining the results in Equations (3.1.14) and (3.1.16), as well as the above identity, we obtain an upper bound for the trace distance through Equation

(3.1.5),

$$\begin{aligned}\mathop{\mathbb{E}}_{U}\left\|\mathcal{M}\left(\rho^{MCE'}\right)-\mathcal{M}\left(\rho^{M}\otimes\rho^{CE'}\right)\right\|_1 &\leqslant \sqrt{s\left(\frac{CE'}{s}\right)^2\sum_{i=1}^{s}2\frac{2\cdot 2^{-2H_{\min}(E)_\omega}}{(C)^2KE}}\\ &\leqslant \frac{2\Delta_{E,\infty}}{\sqrt{KE}}.\end{aligned}$$

□

Lemma 3.1.2. *$g_{\mathcal{M}}(U)$, the trace distance to independence for a fixed (s,η)-quasi-measurement, is Lipschitz continuous on the space $(\mathcal{U}(CKE),\|\cdot\|_2)$ with constant $4\eta\sqrt{\Delta_{M,\infty}\,\Delta_{E,\infty}/ME}$.*

Proof. We wish to analyze the behaviour of the trace distance with respect to the unitary matrix defining the channel. Recall the definition of function $g_{\mathcal{M}}(U)$,

$$g_{\mathcal{M}}(U)=\left\|\mathcal{M}\left(\rho^{MCE'}\right)-\mathcal{M}\left(\rho^{M}\otimes\rho^{CE'}\right)\right\|_1.$$

If we denote by ρ_U and ρ_V the states $\mathrm{Tr}_K\left[U\cdot\sigma\right]$ and $\mathrm{Tr}_K\left[V\cdot\sigma\right]$ respectively, we can bound the deviation of $g_{\mathcal{M}}$ using the triangle inequality by

$$\begin{aligned}|g_{\mathcal{M}}(U)-g_{\mathcal{M}}(V)| &\leqslant \left\|\mathcal{M}\left(\rho_U^{MCE'}\right)-\mathcal{M}\left(\rho_V^{MCE'}\right)\right\|_1\\ &\quad+\left\|\mathcal{M}\left(\rho_U^{M}\otimes\rho_U^{CE'}\right)-\mathcal{M}\left(\rho_V^{M}\otimes\rho_V^{CE'}\right)\right\|_1 \qquad (3.1.17)\\ &= \left\|\mathcal{M}\left(\rho_U^{MCE'}-\rho_V^{MCE'}\right)\right\|_1+\left\|\mathcal{M}\left(\sigma^{M}\otimes\left(\rho_U^{CE'}-\rho_V^{CE'}\right)\right)\right\|_1,\end{aligned}$$

where the second line follows from the linearity of the superoperator. We note that for any hermitian operator ζ,

$$\begin{aligned}
\|\mathcal{M}(\zeta)\|_1 &= \left\| \frac{CE'}{s} \sum_{i=1}^{s} |i\rangle\langle\chi_i|\zeta|\chi_i\rangle\langle i| \right\|_1 \\
&= \frac{CE'}{s} \sum_{i=1}^{s} |\langle\chi_i|\zeta|\chi_i\rangle| \leqslant \frac{CE'}{s} \sum_{i=1}^{s} \langle\chi_i||\zeta||\chi_i\rangle \\
&= \frac{CE'}{s} \sum_{i=1}^{s} \mathrm{Tr}\left[\chi_i|\zeta|\right] \leqslant \eta \|\zeta\|_1 ,
\end{aligned}$$

where the last inequality follows from the definition of (s, η)-quasi-measurements. Applying this new fact, our bound in Equation (3.1.17) becomes,

$$\begin{aligned}
|g_{\mathcal{M}}(U) - g_{\mathcal{M}}(V)| &\leqslant \eta \left\| \rho_U^{MCE'} - \rho_V^{MCE'} \right\|_1 \\
&\qquad + \eta \left\| \sigma^M \otimes \left(\rho_U^{CE'} - \rho_V^{CE'} \right) \right\|_1 \qquad (3.1.18) \\
&\leqslant 2\eta \left\| \rho_U^{MCKEE'} - \rho_V^{MCKEE'} \right\|_1 \\
&= 2\eta \|U \cdot (\sigma \otimes \omega) - V \cdot (\sigma \otimes \omega)\|_1 ,
\end{aligned}$$

where the second line follows from monotonicity. We introduce a purification of σ^{MCK} in a new but temporary system N such that $\dim(N) = \dim(M)$. We also recall that ω is pure. This permits us to use Lemma A.0.9 and arrive at the following consequence of Equation (3.1.18),

$$|g_{\mathcal{M}}(U) - g_{\mathcal{M}}(V)| \leqslant 4\eta \left\| \left(U^{CKE} - V^{CKE} \right) \otimes \mathbb{I}_{MNE'} |\sigma\rangle^{MNCK} |\omega\rangle^{EE'} \right\|_2 . \qquad (3.1.19)$$

We now introduce a helpful operation.

Definition 3.1.2 (Vector-operator correspondence)**.** *Endow systems A and B with fixed orthonormal bases $\{|a_i\rangle^A\}_i$ and $\{|b_i\rangle^B\}_i$ respectively, and let $\mathrm{op}_{A\to B} : A \otimes B \to \mathrm{L}(A, B)$, the space of linear transformations from A to B, be defined as*

$$\mathrm{op}_{A\to B}\left(|a_i\rangle|b_j\rangle\right) = |b_j\rangle\langle a_i| \qquad \forall i, j$$

This operation depends on the choice of basis; therefore, whenever it is used, a particular choice of basis is implied. Since this choice will never matter in our calculations, we shall not explicitly define these bases.

Useful properties of the correspondence can be found in [14].

We can think of the operator $(U^{CKE} - V^{CKE}) \otimes \mathbb{I}^{MNE'}$ as bipartite over composite systems MNE' and CKE. Since the 2-norm depends only on the Schmidt coefficients of the states, it will be invariant under the op operation defined in Definition 3.1.2. Our bound from Equation (3.1.19) then becomes,

$$\begin{aligned} |g_{\mathcal{M}}(U) - g_{\mathcal{M}}(V)| &\leqslant 4\eta \left\| \mathrm{op}_{MNE'\to CKE}\left(\left(U^{CKE} - V^{CKE}\right) \otimes \mathbb{I}^{MNE'} |\sigma\rangle^{MNCK} |\omega\rangle^{EE'}\right)\right\|_2 \\ &= 4\eta \left\| (U - V)\, \mathrm{op}_{MNE'\to CKE}\left(|\sigma\rangle|\omega\rangle\right)\right\|_2, \end{aligned}$$

where the second line follows from the fact that $\mathrm{op}_{MNE'\to CKE}$ is linear and commutes with unitary transformations on CKE. We are left with a few easy

steps to bound the Lipschitz constant.

$$\begin{aligned}
|g_{\mathcal{M}}(U) - g_{\mathcal{M}}(V)| &\leqslant 4\eta \left\|U - V\right\|_2 \left\|\mathrm{op}_{MNE'\to CKE}\left(|\sigma\rangle|\omega\rangle\right)\right\|_\infty \\
&= 4\eta \left\|U - V\right\|_2 \sqrt{\left\|\sigma^{CK} \otimes \omega^E\right\|_\infty} \\
&= 4\eta \left\|U - V\right\|_2 \sqrt{\left\|\sum_m p_m |\psi_m\rangle\langle\psi_m|^{CK}\right\|_\infty \left\|\omega^E\right\|_\infty} \\
&= 4\eta \left\|U - V\right\|_2 \sqrt{\max p_m \cdot 2^{-H_{\min}(E)_\omega}} \\
&= 4\eta \left\|U - V\right\|_2 2^{-\frac{1}{2}H_{min}(M)_\sigma} 2^{-\frac{1}{2}H_{\min}(E)_\omega} \\
&= \frac{4\eta\sqrt{\Delta_{M,\infty}\,\Delta_{E,\infty}}}{\sqrt{ME}} \left\|U - V\right\|_2 .
\end{aligned}$$

A proof of the inequality can be found, for example, in [14]. The second line follows from the fact the Schmidt coefficients of $|\sigma\rangle^{MNCK}$ are the square roots of the eigenvalues of σ^{CK}. The last line follows from the definition of $\Delta_{M,\infty}$ and $\Delta_{E,\infty}$. □

3.2 Measurement net

In order to discretize the set of all (s,η)-quasi-measurements, we require a distance measure for the set.

Definition 3.2.1 (Metric on the set of (s,η)-quasi-measurements, $\mathcal{L}(s,\eta)$)**.** *Consider $\mathcal{M}$, $\mathcal{N} \in \mathcal{L}(s,\eta)$ defined as*

$$\mathcal{M}(\sigma) = \frac{|CE'|}{s}\sum_{i=1}^{s} |i\rangle\langle\chi_i|\sigma|\chi_i\rangle\langle i|, \qquad \mathcal{N}(\sigma) = \frac{|CE'|}{s}\sum_{i=1}^{s} |i\rangle\langle\nu_i|\sigma|\nu_i\rangle\langle i|.$$

We define the distance between these two elements as

$$d(\mathcal{M},\mathcal{N}) := \sum_{i=1}^{s} \left\|\chi_i - \nu_i\right\|_2 .$$

Now letting $\mathcal{M}$ vary instead of U, we define a new function $h_U(\mathcal{M})$ by

$$h_U(\mathcal{M}) = \left\| \mathcal{M}\left(\rho^{MCE'}\right) - \mathcal{M}\left(\rho^{M} \otimes \rho^{CE'}\right) \right\|_1 .$$

Lemma 3.2.1. *$h_U(\mathcal{M})$ is Lipschitz continuous on the space $(\mathcal{L}(s,\eta), d)$ with constant $\frac{2\sqrt{CE'}}{s}\sqrt{\Delta_{M,2}\Delta_{E,2}}$.*

Proof. As for Lemma 3.1.2, we can use the triangle inequality to rewrite the variation of the trace distance as follows,

$$\begin{aligned}
&|h_U(\mathcal{M}) - h_U(\mathcal{N})| \\
&\leqslant \left\| \mathcal{M}\left(\rho^{MCE'}\right) - \mathcal{N}\left(\rho^{MCE'}\right) \right\|_1 + \left\| \mathcal{M}\left(\rho^{M} \otimes \rho^{CE'}\right) - \mathcal{N}\left(\rho^{M} \otimes \rho^{CE'}\right) \right\|_1 \\
&= \frac{CE'}{s} \sum_{i=1}^{n} \Big(\left\| \mathrm{Tr}_{CE'} \left[\left(\chi_i^{CE'} - \nu_i^{CE'} \right) \rho^{MCE'} \right] \right\|_1 \\
&\qquad\qquad + \left\| \mathrm{Tr}_{CE'} \left[\left(\chi_i^{CE'} - \nu_i^{CE'} \right) \rho^{M} \otimes \rho^{CE'} \right] \right\|_1 \Big) \\
&\leqslant \frac{CE'}{s} \sum_{i=1}^{s} \left\| \left(\chi_i^{CE'} - \nu_i^{CE'} \right) \rho^{MCE'} \right\|_1 \\
&\qquad\qquad + \frac{CE'}{s} \sum_{i=1}^{s} \left\| \left(\chi_i^{CE'} - \nu_i^{CE'} \right) \rho^{M} \otimes \rho^{CE'} \right\|_1 \\
&\leqslant \frac{CE'}{s} \sum_{i=1}^{s} \left\| \chi_i^{CE'} - \nu_i^{CE'} \right\|_2 \left\| \rho^{MCE'} \right\|_2 \\
&\qquad\qquad + \frac{CE'}{s} \sum_{i=1}^{s} \left\| \chi_i^{CE'} - \nu_i^{CE'} \right\|_2 \left\| \rho^{M} \otimes \rho^{CE'} \right\|_2 , \qquad (3.2.1)
\end{aligned}$$

where the last line follows from the operator version of the Cauchy-Schwarz inequality (see Equation (IX.32) in [6]). Consider momentarily the second factor in

the first term in Equation (3.2.1),

$$\begin{aligned}
\left\|\rho^{MCE'}\right\|_2 &= \left\|\mathrm{Tr}_{KE}\left[U_{CKE}\cdot\left(\sigma^{MCK}\otimes\omega^{EE'}\right)\right]\right\|_2 \\
&= \sqrt{\mathrm{Tr}\left[\left(U_{CKE}^{\otimes 2}\cdot\left(\sigma^{MCK}\otimes\omega^{EE'}\right)^{\otimes 2}\right)F^{MCE'}\right]} \\
&= \sqrt{\mathrm{Tr}\left[\left(U_{CKE}^{\otimes 2}\;F^{C}\;U_{CKE}^{\dagger\otimes 2}\right)\left(\sum_m p_m^2(|\psi_m\rangle\langle\psi_m|^{CK})^{\otimes 2}\otimes(\omega^{EE'})^{\otimes 2}\right)F^{E'}\right]} \\
&\leqslant \sqrt{\mathrm{Tr}\left[\sum_m p_m^2(|\psi_m\rangle\langle\psi_m|^{CK})^{\otimes 2}\otimes(\omega^{EE'})^{\otimes 2}F^{E'}\right]} \\
&= \sqrt{\mathrm{Tr}\left[(\omega^{EE'})^{\otimes 2}F^{E'}\right]\sum_m p_m^2} = 2^{-\frac{1}{2}H_2(M)_\sigma-\frac{1}{2}H_2(E)_\omega}. \qquad (3.2.2)
\end{aligned}$$

The third line is true by the cyclic property of the trace. The inequality, however, is true by the following observation: since $F^2=\mathbb{I}$ we know that F has eigenvalues ± 1 and so $F\leq\mathbb{I}$. We can make a similar evaluation for the last factor in Equation (3.2.1),

$$\left\|\rho^{M}\otimes\rho^{CE'}\right\|_2 \leqslant 2^{-\frac{1}{2}H_2(M)_\sigma-\frac{1}{2}H_2(E)_\omega}, \qquad (3.2.3)$$

since this inequality is a just a special case of the calculations leading to Equation (3.2.2). If we apply Equations (3.2.2) and (3.2.3) to Equation (3.2.1), we can extract a very simple bound on the variation of the trace distance

$$\begin{aligned}
|h_U(\mathcal{M})-h_U(\mathcal{N})| &\leqslant \frac{2CE'}{s}2^{-\frac{1}{2}H_2(M)_\sigma-\frac{1}{2}H_2(E)_\omega}\sum_{i=1}^{s}\left\|\chi_i^C-\nu_i^C\right\|_2 \\
&\leqslant \frac{2\sqrt{CE'}}{s}\sqrt{\Delta_{M,2}\Delta_{E,2}}\,d(\mathcal{M},\mathcal{N}),
\end{aligned}$$

where the last line follows from the definition of our metric on $\mathcal{L}(s,\eta)$. We have also ignored a factor of $1/\sqrt{K}$ above when expressing the bound in terms of $\Delta_{M,2}$.

We do this to simplify future calculations and it only gives a slightly less tight bound here. □

Lemma 3.2.2. *Given system A, there exists an ε-net $\mathcal{J}$ over the set $\mathcal{L}(s,\eta)$ of all (s,η)-quasi-measurements on A, such that each element $L \in \mathcal{L}(s,\eta)$ is at most ε-distant from an element of $J \in \mathcal{J}$ with respect to the metric $d(\cdot,\cdot)$. The size of this net can be taken to be*

$$|\mathcal{J}| \leqslant \left(\frac{10s}{\varepsilon}\right)^{2s|A|}.$$

Proof. We begin by consider an ε-net $\mathcal{K}$ over $\mathbb{S}_{2|A|}^{\times s}$ (s-tuples of $2|A|$-dimensional Euclidean unit spheres). First, there exists a ε-net over $\mathbb{S}_{2|A|}$ of size no more than $(5/\varepsilon)^{2|A|}$. (See, for example, Lemma II.4 in [21].) $\mathcal{K}$ can then be constructed by assembling the direct product of all the nets on the individual unit spheres. This produces a new net on the set of s-tuples of $2|A|$-dimensional unit spheres. Recall the distance measure $d(\cdot,\cdot)$ over $\mathcal{L}(s,\eta)$, the set of all (s,η)-quasi-measurements. This metric can be extended to s-tuples. If it is then evaluated for any s-tuple x and its representative in the net y,

$$d(x,y) = \sum_{i=1}^{s} \|\chi_i - \nu_i\|_2 \leq s\varepsilon.$$

Thus the spacing of the net $\mathcal{K}$ over s-tuples is at most $s\varepsilon$ with respect to the desired metric. Consider the following set:

$$\mathcal{K}' := \{y \in \mathcal{K} : \exists\, x \in \mathcal{L}(s,\eta), \|x - y\|_2 \leqslant s\varepsilon\}.$$

This is the set of all elements of the net $\mathcal{K}$ which are close to (s,η)-quasi-measurements. In other words, all (s,η)-quasi-measurements use an element of

$\mathcal{K}'$ as their "representative" in the net. Now, divide $\mathcal{L}(s,\eta)$ into subsets of elements which share the same representation in $\mathcal{K}'$ and construct $\mathcal{J}$ by choosing one $L \in \mathcal{L}$ from each subset. We then have by the triangle inequality that all $L \in \mathcal{L}$ are $2s\varepsilon$ close to their new representative in $\mathcal{J}$. Clearly $|\mathcal{J}| \leqslant |\mathcal{K}|$ since it was constructed from a subset and if we wish to make an ε-net over $\mathcal{L}(s,\eta)$ we need only rescale the ε from above, giving the result. □

3.3 General statement

The Lipschitz constants, expectation value and net size give us all the pieces we need to make the concentration argument. We show that with very high probability, the distinguishability from independence of the joint (potentially unnormalized) distribution of messages and quasi-measurement outcomes is small.

Theorem 3.3.1 (Concentration of probability for distinguishability from independence)**.** *Given the quantum state $\rho^{MCKEE'} = U^{CKE} \cdot (\sigma^{MCK} \otimes \omega^{EE'})$ where U is a random unitary operator chosen according to the Haar measure, σ is as defined in Equation (2.1.1), $E' \cong E$, and $\omega^{EE'}$ is a bipartite pure state, the following bound holds*

$$\begin{aligned}&\Pr_U\left\{\sup_{\mathcal{M}\in\mathcal{L}(s,\eta)}\left\|\mathcal{M}\left(\rho^{MCE'}\right)-\mathcal{M}\left(\rho^{M}\otimes\rho^{CE'}\right)\right\|_1>\varepsilon\right\}\\&\quad\leqslant\exp\left(2sCE\ln\left(\frac{40\sqrt{CE}}{\varepsilon}\sqrt{\Delta_{M,2}\Delta_{E,2}}\right)-\frac{(CKE)^2}{2^8\eta^2\Delta_{M,\infty}\Delta_{E,\infty}}\left(\varepsilon-\frac{4\Delta_{E,\infty}}{\sqrt{KE}}\right)^2\right).\end{aligned}$$

In the above, $\Delta_{M,\infty}$, $\Delta_{M,2}$, $\Delta_{E,2}$ and $\Delta_{E,\infty}$ are as defined in Equations (2.1.8), (2.1.9), (2.1.11) and (2.1.10).

Proof. We apply Theorem 2.0.2 to $g_{\mathcal{M}}$ and consider only one direction of the divergence from the expected value. The exact statement can be written as

$$\Pr_U \left\{ g_{\mathcal{M}}(U) > \varepsilon \right\} \leqslant \exp\left(-\frac{MCKE^2}{64\eta^2 \Delta_{M,\infty}\Delta_{E,\infty}} \left(\varepsilon - \mathop{\mathbb{E}}_U g_{\mathcal{M}} \right)^2 \right). \tag{3.3.1}$$

It is convenient to define

$$f(\mathcal{M}, U) = \left\| \mathcal{M}\left(\rho^{MCE'}\right) - \mathcal{M}\left(\rho^{M} \otimes \rho^{CE'}\right) \right\|_1.$$

Clearly, $g_{\mathcal{M}}$ and h_U are Sections of f and we are interested in bounding

$$\Pr_U \{ \sup_{\mathcal{M}} f(\mathcal{M}, U) > \varepsilon \}.$$

Let

$$\varepsilon' = \frac{s\varepsilon}{2\sqrt{CE\Delta_{M,2}\Delta_{E,2}}},$$

and consider $\mathcal{J}$ an ε'-net over all (s, η)-quasi-measurements $\mathcal{M}$. We found in Lemma 3.2.1 that if two (s, η)-quasi-measurements were ε' apart with respect to the distance measure $d(\cdot, \cdot)$, then for a fixed unitary U, the values of f for each measurement would not differ by more than ε. Thus we can state that the supremum deviation of f is not more than twice the maximum deviation found on measurements in the net,

$$\Pr_U \left\{ \sup_{\mathcal{M}} f(\mathcal{M}, U) > 2\varepsilon \right\} \leqslant \Pr_U \left\{ \max_{\mathcal{M} \in \mathcal{J}} f(\mathcal{M}, U) > \varepsilon \right\}.$$

A union bound argument now bounds the probability of deviation for the maximum measurement by the probability of deviation for a generic measurement,

$$\Pr_U \left\{ \max_{\mathcal{M}\in\mathcal{J}} f(\mathcal{M}, U) > \varepsilon \right\} \leqslant \sum_{\mathcal{M}\in\mathcal{J}} \Pr_U \left\{ g_{\mathcal{M}}(U) > \varepsilon \right\}.$$

Thankfully, we have an explicit bound for the probability of deviation for an arbitrary measurement and we can make a simplification,

$$\begin{aligned}
\Pr_U \left\{ \sup_{\mathcal{M}} f(U, \mathcal{M}) > 2\varepsilon \right\} &\leqslant \sum_{\mathcal{M}\in\mathcal{J}} \exp\left(-\frac{MCKE^2}{64\eta^2 \Delta_{M,\infty}\Delta_{E,\infty}} \left(\varepsilon - \mathop{\mathbb{E}}_U f \right)^2 \right) \\
&\leqslant \left(\frac{20\sqrt{CE\Delta_{M,2}\Delta_{E,2}}}{\varepsilon} \right)^{2sCE} \exp\left(-\frac{MCKE^2}{64\eta^2 \Delta_{M,\infty}\Delta_{E,\infty}} \left(\varepsilon - \mathop{\mathbb{E}}_U f \right)^2 \right) \\
&\leqslant \exp\left(2sCE \ln\left(\frac{20\sqrt{CE\Delta_{M,2}\Delta_{E,2}}}{\varepsilon} \right) - \frac{MCKE^2}{64\eta^2 \Delta_{M,\infty}\Delta_{E,\infty}} \left(\varepsilon - \mathop{\mathbb{E}}_U f \right)^2 \right).
\end{aligned}$$

Substituting in the fact that $CK = M$ yields the desired inequality. □

3.4 Projective measurement

In this Section we will only consider projective measurements, in other words $(s, \eta) = (CE', 1)$. We will also state all of the subsequent Theorems in terms of qubits. For this reason we will identify $C = 2^c$, $K = 2^k$ and $E = E' = 2^e$. This last assumption, namely that E and E' have the same dimension, is crucial for this Section because it restricts the size of the set of measurements sufficiently to allow for a straightforward discretization. The restriction will be lifted when we move on to generalized measurements in the next Section, however.

Our calculations, we will make repeated use of the fact that

$$\log(x + y) \leqslant x + \log(y) \qquad \forall x, y \geqslant 1. \tag{3.4.1}$$

Corollary 3.4.1 (Locking for uniform messages with maximal entanglement)**.** *Consider the locking scheme described in Definition 2.1.3 for a uniform message with maximal entanglement available at the measurement. Choose p and ε such that $\varepsilon > 8\sqrt{1/KE}$ and $p > 2^{-2(CE)^2}$. Then the scheme will be an ε-locking locking scheme except with probability p so long as the measurement superoperators are restricted to projective measurements and*

$$k > 9 + 2\log\frac{1}{\varepsilon} + \frac{1}{2}\log(c+e).$$

Proof. Using Theorem 3.3.1, we ensure that, except with probability p, our state is an ε-locking scheme provided that

$$2(CE)^2 \ln\left(\frac{40\sqrt{CE}}{\varepsilon}\right) - \frac{(CE)^2}{2^8}K^2(\varepsilon')^2 < \ln p,$$

where we've defined for the time being ε' as $\varepsilon - 4/\sqrt{KE}$. A quick rearrangement of the terms reveals that the inequality will be satisfied if

$$\frac{2^9}{(\varepsilon')^2}\ln\left(\frac{40\sqrt{CE}}{\varepsilon}\left(\frac{1}{p}\right)^{1/2(CE)^2}\right) < K^2. \tag{3.4.2}$$

From our choice of p we can easily see $(1/p)^{1/2(CE)^2} < 2$ and from our choice of ε we see that $2^9/(\varepsilon')^2 < 2^{13}/\varepsilon^2$. Thus inequality (3.4.2) is satisfied when

$$\log\left(\frac{2^{13}}{\varepsilon^2}\right) + \log\left(\ln 2\log\left(\frac{80\sqrt{CE}}{\varepsilon}\right)\right) < 2k.$$

Finally, two applications of Equation (3.4.1) reveal that the above is satisfied provided,

$$17 + 2\log\frac{1}{\varepsilon} + \log\log\frac{1}{\varepsilon} + \log(c+e) < 2k.$$

Rearranging the terms we see that the above condition is satisfied provided inequality (3.4.1) is satisfied, and we have completed the proof. □

Corollary 3.4.1, and its extension to arbitrary POVM measurements in Corollary 3.5.4 is a mathematical expression that "generically, information is locked until it can be completely decoded." To arrive at this interpretation, recall from Equation (2.1.4) that to achieve a decoding error of ϵ, the measurement must be supplied with the entanglement through system E' as well as a system C satisfying $c - n > 2\log(1/\epsilon)$. Of course, this condition could never be met if the constraint $n = c + k$ is assumed, but the constraint was only made for convenience to prove the locking results. Using it to re-express Corollary 3.4.1, though, we find that the information about the message is ϵ-locked provided $c = n - k < n - 9 - 2\log(1/\epsilon) - 1/2 \cdot \log(c + e)$. Therefore, regardless of the size of the message or the amount of entanglement, the message goes from being ϵ-locked to being decodable with average probability of error at most ϵ with the transfer of $9 + 4\log(1/\epsilon) + 1/2 \cdot \log(c + e)$ qubits.

At this point, we wish to study the dependence of the minimum key size k on the various entropies of the message M and the entanglement E.

Corollary 3.4.2 (Locking for messages of bounded entropy with imperfect entanglement)**.** *Consider the locking scheme described in Definition 2.1.3 for a message of bounded entropy with entanglement of a bounded fidelity available at the measurement. Choose ε and p satisfying*

$$\varepsilon > \frac{8\Delta_{E,\infty}}{\sqrt{KE}}, \qquad\qquad p > 2^{-2(CE)^2}.$$

Then the scheme will be an ε-locking locking scheme except with probability p so long as the measurement superoperators are restricted to projective measurements and

$$k' + \frac{1}{2}\Big(n - H_{\min}(M)_\sigma\Big) + \frac{1}{2}\Big(e - H_{\min}(E)_\omega\Big) < k, \tag{3.4.3}$$

where we've defined k' as the lower bound given in Corollary 3.4.1, i.e.: $k' = 9 + 2\log(1/\varepsilon) + 1/2 \cdot \log(c+e)$.

Proof. From Theorem 3.3.1, we can ensure ε-locking except with probability p by satisfying

$$2(CE)^2 \ln\left(\frac{40\sqrt{CE}}{\varepsilon}\sqrt{\Delta_{M,2}\Delta_{E,2}}\right) - \frac{(CE)^2}{2^8\Delta_{M,\infty}\Delta_{E,\infty}}K^2(\varepsilon')^2 < \ln p,$$

where we've defined for the time being ε' as $\varepsilon - 4\Delta_{E,\infty}/\sqrt{KE}$. A quick rearrangement of the terms reveals that the inequality can be satisfied if

$$\frac{2^9\Delta_{M,\infty}\Delta_{E,\infty}}{(\varepsilon')^2}\ln\left(\frac{40\sqrt{CE}}{\varepsilon}\sqrt{\Delta_{E,2}\Delta_{M,2}}\left(\frac{1}{p}\right)^{1/2(CE)^2}\right) < K^2, \tag{3.4.4}$$

From our choice of p we can easily see $(1/p)^{1/2(CE)^2} < 2$ and from our choice of ε we see that $2^9/(\varepsilon')^2 < 2^{13}/\varepsilon^2$. Thus the inequality in Equation (3.4.4) is satisfied when

$$13 + 2\log\frac{1}{\varepsilon} + \log(\Delta_{M,\infty}\Delta_{E,\infty}) + \log\left(7 + \log\frac{1}{\varepsilon} + \frac{1}{2}(c+e) + \frac{1}{2}\log(\Delta_{M,2}\Delta_{E,2})\right) < 2k.$$

However, we know that the maximum values of $\Delta_{M,2}$ and $\Delta_{E,2}$ are M and E respectively. Combined with our assumption that $k < c$, we can quickly reduce the

above to,

$$18 + 3\log\frac{1}{\varepsilon} + \log(c+e) + \Big(n - H_{\min}(M)_\sigma\Big) + \Big(e - H_{\min}(E)_\omega\Big) < 2k.$$

Finally, we can identify k' and give the result as desired. □

3.5 General POVM

We will now show that the results of the previous Section hold not only for projective measurements, but also for general POVMs, up to very minor changes in the various constants. The main difficulty at this point is that we cannot use Theorem 3.3.1 directly, since it only gives bounds for (s,η)-quasi-measurements. We must therefore show that a general POVM behaves essentially like an (s,η)-quasi-measurement for the purposes of the Theorem. Our strategy will be probabilistic in nature: we will show that doing a general POVM $\mathcal{M}$ is mathematically equivalent to randomly selecting a measurement constructed from possible sequences of s measurement results obtained from $\mathcal{M}$. With overwhelming probability, the sequence chosen will be an (s,η)-quasi-measurement, and Theorem 3.3.1 will then apply in this case.

We start by proving this last fact, namely that with very high probability, a sequence of s measurement results will be an (s,η)-quasi-measurement, for an appropriately chosen η.

Lemma 3.5.1. *Let $\mathcal{M}^{CE'\to X}$ be any complete measurement superoperator, with $\mathcal{M}(\pi) = \sum_i \alpha_i |i\rangle\langle\chi_i|\pi|\chi_i\rangle\langle i|$, and consider the operator-valued random variable Y which takes the value $|\chi_i\rangle\langle\chi_i|$ with probability $\alpha_i\langle\chi_i|\pi|\chi_i\rangle = \alpha_i/CE'$. Then, s i.i.d.*

copies of Y will fail to be an (s,η)-quasi-measurement with probability at most $2CE'e^{-s(\eta-1)^2/CE'2\ln 2}$.

Proof. Y fulfills all the conditions for the operator Chernoff bound (Lemma A.0.8) to apply, with $\mathbb{E}Y = \pi^{CE'}$. This yields

$$\Pr\left\{\frac{1}{s}\sum_{j=1}^{s} Y_j \not\leqslant \eta\pi\right\} \leqslant 2CE'e^{-s(\eta-1)^2/CE'2\ln 2},$$

and the probability on the left is an upper bound on the probability that the s-tuple $Y_1,\ldots,Y_s$ is not an (s,η)-quasi-measurement. □

We now use this to show that the best general POVM cannot do much better than the best (s,η)-quasi-measurement:

Lemma 3.5.2. *It is true that*

$$\begin{aligned}&\sup_{\mathcal{M}} \left\|\mathcal{M}\left(\rho^{MCE'} - \rho^M \otimes \rho^{CE'}\right)\right\|_1 \\ &\quad\leqslant \max_{\mathcal{M}'\in\mathcal{L}(s,\eta)} \left\|\mathcal{M}'\left(\rho^{MCE'} - \rho^M \otimes \rho^{CE'}\right)\right\|_1 + 4(CE')^2 e^{-s(\eta-1)^2/(CE'(2\ln 2))}, \quad (3.5.1)\end{aligned}$$

where the supremum on the left-hand side is taken over all measurement superoperators.

Proof. Let $\mathcal{M}^{CE'\to X}$ be any complete measurement superoperator of the form $\mathcal{M}(\sigma) = \sum_i \alpha_i |i\rangle\langle\chi_i|\sigma|\chi_i\rangle\langle i|$, and define Y to be the operator-valued random variable which takes value χ_i with probability α_i/CE'. Let Q be the event that $Y_1,\ldots,Y_n$ is an (s,η)-quasi-measurement, where the Y_i are i.i.d. with the same

distribution as Y.

$$\begin{aligned}
&\left\|\mathcal{M}\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right\|_1 \\
&= \sum_i \alpha_i \left\|\mathrm{Tr}_{CE'}\left[\chi_i\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right]\right\|_1 \\
&= CE'\mathbb{E}_Y \left\|\mathrm{Tr}_{CE'}\left[Y\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right]\right\|_1 \\
&= \frac{CE'}{s}\mathbb{E}_{Y_1,\ldots,Y_s}\sum_i^s \left\|\mathrm{Tr}_{CE'}\left[Y_i\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right]\right\|_1 .
\end{aligned}$$

At this point we separate the expression into two terms, one for the event Q and another for its complement.

$$\begin{aligned}
&\left\|\mathcal{M}\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right\|_1 \\
&= \frac{CE'}{s}\Pr\{Q\}\mathbb{E}\left[\sum_{i=1}^s \left\|\mathrm{Tr}_{CE'}\left[Y_i\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right]\right\|_1 \middle| Q\right] \\
&\qquad + \frac{CE'}{s}\Pr\{\bar{Q}\}\mathbb{E}\left[\sum_{i=1}^s \left\|\mathrm{Tr}_{CE'}\left[Y_i\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right]\right\|_1 \middle| \bar{Q}\right] \\
&\leqslant \max_{\mathcal{M}'\in\mathcal{L}(s,\eta)} \left\|\mathcal{M}'\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right\|_1 \Pr\{Q\} + 2CE'\Pr\{\bar{Q}\} \\
&\leqslant \max_{\mathcal{M}'\in\mathcal{L}(s,\eta)} \left\|\mathcal{M}'\left(\rho^{MCE'}-\rho^{M}\otimes\rho^{CE'}\right)\right\|_1 + 4(CE')^2 e^{-s(\eta-1)^2/CE'2\ln 2}.
\end{aligned}$$

In the above, the sum of trace distances given Q was interpreted as executing an (s,η)-quasi-measurement described by $Y_1,\ldots,Y_s$, and the same sum given $\bar{Q}$ was simply bounded by 2η (there are s terms in the sum, each of which cannot exceed 2). In the last step, we have bounded $\Pr\{\bar{Q}\}$ using Lemma 3.5.1 and made use of the fact that we can assume without loss of generality that $|E| = |E'|$.

Finally, a non-complete measurement superoperator can always be decomposed into a complete one by splitting the POVM elements of rank greater than 1; this process always increases the trace distance. □

What we have achieved with the above statement is to show that the decoupling distance for a general measurement superoperator is very close to the decoupling distance of an (s, η)-quasi-measurement. All that is now left to do is to use Theorem 3.3.1 to bound the supremum over (s, η)-quasi-measurements, and we get the main Theorem of this Section:

Theorem 3.5.3 (Locking Theorem for general measurements)**.** *Given the quantum state $\rho^{MCKEE'} = U^{CKE} \cdot (\sigma^{MCK} \otimes \omega^{EE'})$ where U is a random unitary operator chosen according to the Haar measure, σ is as defined in Equation (2.1.1) and $\omega^{EE'}$ a bipartite pure state, then*

$$\begin{aligned}
&\Pr_U \left\{ \sup_{\mathcal{M}} \left\| \mathcal{M}\left(\rho^{MCE'}\right) - \mathcal{M}\left(\rho^{M} \otimes \rho^{CE'}\right) \right\|_1 > \varepsilon \right\} \\
&\leqslant \exp\left(9(CE)^2 \ln(CE) \ln\left(\frac{40\sqrt{CE}}{\varepsilon} \sqrt{\Delta_{M,2}\Delta_{E,2}} \right) - \frac{(CKE)^2}{2^{10}\Delta_{M,\infty}\Delta_{E,\infty}} \left(\varepsilon - \frac{8\Delta_{E,\infty}}{\sqrt{KE}} \right)^2 \right).
\end{aligned}$$

In the above, $\Delta_{M,\infty}$, $\Delta_{M,2}$, $\Delta_{E,2}$ and $\Delta_{E,\infty}$ are as defined in Equations (2.1.8), (2.1.9), (2.1.11) and (2.1.10).

Proof. We may assume without loss of generality that $|E'| \leq |E|$. If not, let E'' be the range of $\rho^{E'} = \omega^{E'}$. Because ω is pure, $|E''| = \operatorname{rank} \omega^{E'} \leq |E|$. Let V be the isometric embedding $E'' \hookrightarrow E'$ and $\rho^{MCE''}$ the projection of ρ to MCE''. Then for any POVM measurement superoperator $\mathcal{M}^{CE' \to X}$,

$$\mathcal{M}(\rho^{MCE'}) = \mathcal{M}(V \rho^{MCE''} V^\dagger)$$

so measuring $\mathcal{M}$ or $\mathcal{M} \circ (V \cdot V^\dagger)$ will yield exactly the same measurement statistics. But the latter is a POVM on CE'' and E'' satisfies the desired dimension bound.

Substituting the results of Lemma 3.5.2 into those of Theorem 3.3.1, we get the following:

$$\Pr_U \left\{ \sup_{\mathcal{M}} \left\| \mathcal{M} \left(\rho^{MCE'} - \rho^M \otimes \rho^{CE'} \right) \right\|_1 \geqslant \varepsilon \right\} \leqslant \exp \left(2sCE \ln \left(\frac{40\sqrt{CE}}{\varepsilon} \sqrt{\Delta_{M,2} \Delta_{E,2}} \right) \right.$$
$$\left. - \frac{(CKE)^2}{2^8 \eta^2 \Delta_{M,\infty} \Delta_{E,\infty}} \left(\varepsilon - 4(CE)^2 e^{-s(\eta-1)^2/(CE(2\ln 2))} - \frac{4\Delta_{E,\infty}}{\sqrt{KE}} \right)^2 \right). \quad (3.5.2)$$

We now choose $\eta = 2$ and $s = (6 \ln 2) CE \ln CE$ and note that this immediately implies

$$2(CE)^2 e^{-s(\eta-1)^2/CE2\ln 2} = \frac{2}{CE}.$$

We absorb this factor into our "offset" for the ε factor,

$$\left(\varepsilon - 4(CE)^2 e^{-s(\eta-1)^2/CE2\ln 2} - \frac{4\Delta_{E,\infty}}{\sqrt{KE}} \right)^2 \geqslant \left(\varepsilon - \frac{8\Delta_{E,\infty}}{\sqrt{KE}} \right)^2.$$

Substituting the choices for s and η into Equation 3.5.2 reveals the desired result. □

We now wish to express, in qubits, a lower bound for the key size for a given probability p and a given ε. The relevant variables are $M = 2^n$, $C = 2^c$, $K = 2^k$, and $E = 2^e$. Unlike in the previous Section, it is unnecessary to make any assumptions about the dimension of E'.

Corollary 3.5.4 (Locking against POVMs for a uniform message with maximal entanglement)**.** *Consider the locking scheme described in Definition 2.1.3 for a uniform message and maximal entanglement available at the measurement. Choose*

p and ϵ such that $\varepsilon > 16\sqrt{1/KE}$ and $p > 2^{-9(CE)^2}$. Then the scheme will be an ε-locking locking scheme except with probability p so long as

$$11 + 2\log\frac{1}{\varepsilon} + \log(c+e) < k.$$

Proof. From Theorem 3.5.3 we can ensure ε-locking except with probability p given

$$9\ln(CE)\ln\left(\frac{40\sqrt{CE}}{\varepsilon}\right) + \frac{1}{9(CE')^2}\ln\frac{1}{p} < \frac{K^2(\varepsilon')^2}{2^{10}},$$

where we've defined for the time being ε' as $\varepsilon - 8/\sqrt{KE}$. We now make use of our lower bound for p as well as the assumption that $\ln(CE) \geqslant 1$ to show that the above can satisfied provided

$$9\ln(CE)\ln\left(\frac{80\sqrt{CE}}{\varepsilon}\right) < \frac{K^2(\varepsilon')^2}{2^{10}}.$$

Solving the above equation for k and applying the condition on ε reveals that the bound can be satisfied by the statement in the Lemma. □

Corollary 3.5.5 (Locking against POVMs for messages of bounded entropy with imperfect entanglement)**.** *Consider the locking scheme described in Definition 2.1.3 for a uniform message and maximal entanglement available at the measurement. Choose p and ϵ such that*

$$\varepsilon > \frac{16\Delta_{E,\infty}}{\sqrt{KE}}, \qquad\qquad p > 2^{-9(CE)^2}.$$

Then the scheme will be an ε-locking locking scheme except with probability p so long as

$$k' + \frac{1}{2}\Big(n - H_{\min}(M)_\sigma\Big) + \frac{1}{2}\Big(e - H_{\min}(E)_\omega\Big) < k, \tag{3.5.3}$$

where we've defined k' as the lower bound given in Corollary 3.5.4, i.e.: $k' = 11 + 2\log(1/\varepsilon) + \log(c+e)$.

Proof. From Theorem 3.3.1, we can ensure ε-locking with probability p by satisfying, From Theorem 3.5.3 we can ensure ε-locking with probability p given

$$9\ln(CE)\ln\left(\frac{40\sqrt{CE}\sqrt{\Delta_{M,2}\Delta_{E,2}}}{\varepsilon}\right) + \frac{1}{9(CE')^2}\ln\frac{1}{p} < \frac{K^2(\varepsilon')^2}{2^{10}\Delta_{M,\infty}\Delta_{E,\infty}},$$

where we've defined for the time being ε' as $\varepsilon - 8/\sqrt{KE}$. We now make use of our lower bound for p as well as the assumption that $\ln(CE) \geqslant 1$ to show that the above can satisfied provided

$$9\ln(CE)\ln\left(\frac{80\sqrt{CE}\sqrt{\Delta_{M,2}\Delta_{E,2}}}{\varepsilon}\right) < \frac{K^2(\varepsilon')^2}{2^{10}\Delta_{M,\infty}\Delta_{E,\infty}}.$$

Next, we use our definition for ε' and our bound for ε and we solve for k to find that the bound is satisfied provided

$$21 + 3\log\frac{1}{\varepsilon} + 2\log(c+e) + \log(\Delta_{M,\infty}\Delta_{E,\infty}) < 2k.$$

Finally, we can identify k' and give the result as desired. □

The lower bound requirement on ε in Corollary 3.5.5 limits the corollary's range of applicability to situations in which $H_{\min}(E)_\omega$ is not too small. Specifically,

the requirement can be rewritten in light of (3.5.3) as

$$2\log(c+e) + (n - H_{\min}(M)_\sigma) + 3H_{\min}(E)_\omega > e + \text{const.}$$

So, at least when the message is uniform, the requirement is roughly that $H_{\min}(E)_\sigma > e/3$. We suspect that this requirement can be eliminated but leave it as an open problem to find a way to do so.

CHAPTER 4
Decoding Results

The previous Sections have shown that, under certain conditions, no classical information is recoverable by the receiver. Here we aim to show that, in many regimes, these results are essentially optimal. We do this by showing that if we make the key only very slightly smaller, then with overwhelming probability, the classical message will be decodable with a negligible error probability. In fact we prove even more: in this regime where the information is decodable, the decoder can even decode a *purification* of the classical message. In other words, in this generic scenario where U is chosen with no preferred basis, either all *classical* information is locked away, or we can decode *quantum* information. This is formalized in the next Theorem.

In order to study decodability, we must discard the identifications made in Figure 2–2 to study locking and return to the original scenario described by Figure 2–1. Whereas k was previously the number of qubits in system K, there is no system K in Figure 2–2. Instead, we define $k = n - c$, which is consistent with its earlier definition. Now, however, it might be the case that k is negative since decoding could require the cyphertext to be longer than the message.

The following Theorem generalizes the discussion of Section 2.1 to nonuniform messages and imperfect entanglement.

Theorem 4.0.6. *If U is chosen according to the Haar measure, then the information in the scheme described in Figure 2–1 is such that there exists a decoding CPTP map $\mathcal{D}^{CE'\to N}$ such that*

$$\left\|\mathcal{D}\left(\mathrm{Tr}_D\left[U^{NE\to CD}\left(\sigma^{RMN}\otimes\omega^{E'E}\right)(U^{NE\to CD})^\dagger\right]\right)-\sigma^{RMN}\right\|_1\leqslant\varepsilon$$

asymptotically almost surely, where σ^{RMN} is a purification of σ^{MN}, as long as

$$k\leqslant\frac{1}{2}\Big(n-H_{\max}(M)_\sigma\Big)-\frac{1}{2}\Big(e-H_2(E)_\omega\Big)-2\log(1/\varepsilon)-4$$

Proof. Using Theorem 3.7 from [14], we get that

$$\mathbb{E}_U\left\|\mathrm{Tr}_C\left[U^{NE\to CD}\left(\sigma^{RMN}\otimes\omega^{E}\right)(U^{NE\to CD})^\dagger\right]-\sigma^{RM}\otimes\rho^D\right\|_1 \leqslant 2^{\frac{1}{2}H_{\max}(M)_\sigma-\frac{1}{2}H_2(E)_\omega}\sqrt{\frac{D}{C}}.$$

It can also be shown that the value of this trace distance will asymptotically almost surely not exceed twice this bound. Under this condition, we have that:

$$\left\|\mathrm{Tr}_C\left[U^{NE\to CD}\left(\sigma^{RMN}\otimes\omega^{E}\right)(U^{NE\to CD})^\dagger\right]-\sigma^{RM}\otimes\rho^D\right\|_1 \leqslant 2\times 2^{\frac{1}{2}H_{\max}(M)_\sigma-\frac{1}{2}H_2(E)_\omega}\sqrt{\frac{D}{C}}.$$

Uhlmann's Theorem then implies the existence of a partial isometry $V^{CE'\to NG}$ and of a purification of ρ^D on system G that we call θ^{DG} such that

$$\left\|VU\left(\sigma^{RMN}\otimes\omega^{E'E}\right)U^\dagger V^\dagger-\sigma^{RMN}\otimes\theta^{DG}\right\|_1\leqslant 4\left(2^{H_{\max}(M)_\sigma-H_2(E)_\omega}\frac{D}{C}\right)^{1/4}.$$

Defining $\mathcal{D}^{CE'\to N}$ as $\mathcal{D}(\xi) = \mathrm{Tr}_G\left[V\xi V^\dagger\right]$ and tracing out system D, we get that

$$\left\|\mathcal{D}\left(\mathrm{Tr}_D\left[U^{NE\to CD}\left(\sigma^{RMN}\otimes\omega^{E'E}\right)(U^{NE\to CD})^\dagger\right]\right) - \sigma^{RMN}\right\|_1 \leqslant 4\left(2^{H_{\max}(M)_\sigma - H_2(E)_\omega}\frac{D}{C}\right)^{1/4}.$$

Now, to satisfy the Theorem statement, we need to ensure that

$$4\left(2^{H_{\max}(M)_\sigma - H_2(E)_\omega}\frac{D}{C}\right)^{1/4} \leqslant \varepsilon.$$

Taking logarithms on both sides and using the fact that $\log D = k + e$, we get that

$$2 + \frac{1}{4}\left[H_{\max}(M)_\sigma - H_2(E)_\omega + e + k - c\right] \leqslant \log\varepsilon.$$

Substituting in the fact that $c = n - k$, we arrive at the statement of the Theorem. □

CHAPTER 5
Discussion

The results of this work establish that the information locking effect is a generic effect and a stronger effect than was previously known. We describe the width of the regime between locking and decoding (in number of qubits) as a function of various setup parameters. We call this quantity the "frontier width". Outside of this frontier, the receiver has either asymptotically low probability of decoding the message or asymptotically low probability of failing to do so. Unlike previous results, we consider arbitrarily small values for the accessible information thanks to our strictly stronger Definition 2.1.3. We also do away with schemes based on an explicit key register. Whereas those schemes relied on the availability of unbiased bases to hide the information from the receiver, we achieve the same result by tracing out any arbitrary small "key" quantum subsystem. Our results are strengthened further by extending our analysis to generalized POVMs at the receiver, non-uniform input messages, and the availability of a pure state containing potentially large amounts of entanglement between the sender and the receiver.

Information locking in a generic unitary channel may appear reminiscent of a strong converse to a channel capacity problem. Any quantum channel has a classical capacity defined as the supremum over all achievable rates of the channel in the limit of many simultaneous channel uses. The rate of a channel, in turn, is

the ratio of successfully communicated classical bits in n uses of the channel. The strong converse Theorem states that any attempt to transmit above the channel capacity will result in the decoding error probability approaching one as $n \to \infty$. In our setting, the analog of the strong converse would be a matching lower bound to Equation (2.1.5) of the form

$$1 - \epsilon < \frac{1}{M} \sum_{m} \sum_{m' \neq m} p(m'|m) \tag{5.0.1}$$

whenever $C < M$, indicating the the probability of incorrectly decoding the message is at least $1 - \epsilon$. What we prove here is much stronger. Equation (5.0.1) doesn't rule out the possibility of being able to pin the message down to some relatively small set. More generally, it doesn't imply a small mutual information between the message and the measurement outcome. Information locking does imply these stronger statements.

Our results are general enough to easily recover the setting of [21] and reproduce the state in equation (1.4.1) by the following steps: first, we remove the entanglement state ω. It is not hard to see that all of our calculations hold for $|E| = |E'| = 1$ and with support on the one-dimensional system EE', ω would manifest as nothing more than a global phase in the quantum circuit which we can ignore. Next, we can rewrite the states $|m\rangle\langle m|^M = |c\rangle\langle c|^{M_1} \otimes |k\rangle\langle k|^{M_2}$ and $|\psi_m\rangle\langle\psi_m|^{CK} = |c\rangle\langle c|^C \otimes |k\rangle\langle k|^K$ and the unitary operator $U^{CK} = \sum_{i,j} U^C_{(i,j)} \otimes |i\rangle\langle j|^K$, where $U^C_{(i,j)}$ is defined to be the matrix U^{CK} restricted to the (i, j) block in the

basis of K.

$$\begin{aligned}
\rho^{MCK} &= \sum_m |m\rangle\langle m|^M \otimes \left(\left(U^\dagger\right)^{CK} |\psi_m^{CK}\rangle\langle\psi_m^{CK}| U^{CK} \right) \\
&= \sum_{c,k} |ck\rangle\langle ck|^{M_1 M_2} \otimes \left(\sum_{i,j} |j\rangle\langle i|^K \otimes \left(U_{(i,j)}^\dagger\right)^C \right) |ck\rangle\langle ck|^{CK} \\
&\qquad\qquad \left(\sum_{i,j} |s\rangle\langle t|^K \otimes \left(U_{(s,t)}\right)^C \right) \\
&= \sum_{c,k} |c\rangle\langle c|^{M_1} \otimes |k\rangle\langle k|^{M_2} \otimes \sum_{j,t} |j\rangle\langle t|^K \otimes \left(U_{(k,j)}^\dagger\right)^C |c\rangle\langle c|^C \left(U_{(k,t)}\right)^C
\end{aligned}$$

If we now make the identifications $M_1 \cong A_1$, $M_2 \otimes K \cong A_2$ and $C \cong B$ we can recover almost exactly the state $\rho^{A_1 A_2 B}$ in equation (1.4.1). The only step required is that Alice either dephase the state on A_2 completely, or that she send that state to Bob through a completely dephasing channel,

$$\mathcal{N}(\rho) = \sum_i |i\rangle\langle i|\rho|i\rangle\langle i|,$$

where the $\{|i\rangle\}_i$ form an orthonormal basis. By using a completely dephasing channel, we recover the classical key that was used in [21].

It is natural in physics to measure the "correlation" between two quantum physical systems using the correlation between the outcomes of measurements on those two systems. Information locking suggests, however, that measurements can be distressingly bad ways to detect correlation, grossly underestimating the total correlation between the system. It is important to distinguish however, between open and closed quantum systems. Although closed quantum systems (those that

do not interact with their environment) are modelled by unitary evolution, open quantum systems are not and the generic unitary channel model does not apply.

If trying to build a locking scheme, one runs up against a seemingly insurmountable obstacle: Haar-random unitaries are not easily simulated by quantum computers. The unitary U_{CKE}, if implemented using random two-qubit gates, would require exponentially many such gates before its distribution reached the required uniform Haar measure. There is hope however: the follow-up article [17] demonstrates a locking scheme achievable in a quantum circuit of depth only slightly superlinear in the number of qubits.

Black hole evaporation is inconsistent with the reversibility of quantum mechanics in that a black hole radiates quantum states that are close to maximally mixed (i.e.: "thermal") even when it is formed from a pure state. In an effort to reconcile these opposing statements, one could claim that the inside of the black hole contains a purification to the otherwise mixed emitted radiation. However, near the end of the black hole's evaporation, one would require that the black hole remnant hold the purification of a rather large radiated quantum system. The large number of remnant species required to describe purifications of this size is unfortunately inconsistent with low energy physics [2, 8]. Oppenheim and Smolin suggested that the problem could be rescued were the black hole to lock the information about its contents [36]. In this way, the number of possible remnants is kept small (i.e.: the size of a key) yet the emitted radiation reveals nothing about the state inside the black hole.

The Smolin-Oppenheim locking result used a state similar to equation (1.4.1). In both [36] and [21] the state which locks information has an explicit key register. Our method, however, avoids this ad hoc distinction and any small subsystem can be used as a key. This is particularly useful in the black hole setting as one could not expect reasonable black hole dynamics from which an explicit key would emerge.

Our results for generic unitary channels and ε-locking schemes are perfectly compatible with (and in fact, inspired by) the unitary evaporation process for black holes described in [22]. Here, Hayden and Preskill, show that for a black hole that is highly entangled with prior emitted radiation, a message sent into the black hole would be decodable in subsequent emitted radiation after only a short time. The time scales obtained were such that the message was decodable immediately after the black hole had sufficiently "scrambled" the message with internal degrees of freedom. Our extension to this and other papers concerning generic unitary dynamics in black holes [34, 27, 7] is that the information is not decodable until moments before it can *all* be obtained. The conclusion depends, of course, on whether the generic unitary transformation is a good model of the evaporation process.

The quantum discord is a measure of correlations defined by Ollivier and Zurek as follows in [31]

$$D(A;B)_{\rho,\mathcal{M}\otimes\mathcal{N}} = I(A;B)_\rho - I(X;Y)_{\mathcal{M}\otimes\mathcal{N}}$$

where $\mathcal{M}$ and $\mathcal{N}$ are POVMs on systems A and B with results X and Y respectively. The discord is then the measure between the mutual information, the "total correlations", and the accessible information, the "classical correlations". In [38], the authors claim that quantum discord is a robust measure of quantum correlations as it decays exponentially for a two-qubit decoherence model whereas measures of entanglement decay in finite time. We show, however, that for large quantum systems, quantum discord can decay to an arbitrarily low value at the cost of a relatively low key size. This puts into doubt the previous claims of robustness for quantum discord as a good measure of quantum correlations.

Comparing Corollary 3.5.5 and Theorem 4.0.6 reveals that the difference between being ε-locked and being able to decode quantum information to within ε is determined by a frontier width of at most

$$\frac{1}{2}\left[H_{\max}(M)_\sigma - H_{\min}(M)_\sigma\right] + \left[e - H_{\min}(E)_\omega\right] + \log(c+e) + 4\log(1/\varepsilon) + 15$$

qubits, where the inequality $H_2 \geq H_{\min}$ has been used to simplify the expression.

In other words, if we consider the case of maximal entanglement, then the frontier width between locking and decodability can only be as wide as the difference between the min- and max-entropy of the message modulo logarithmic terms. One should note that this gap is real, and not only an artifact of our proof technique. To see this, consider an n-bit message distributed such that with probability $\frac{1}{2}$, the first bit is uniform and the rest of the string is always zero, and with probability $\frac{1}{2}$ the whole string is uniform. The max-entropy of such a message is n, but the min-entropy is tiny. Now, to be able to decode, one must be able

to decode the entire string in the "worst-case" scenario where the whole string is uniform, so the max-entropy is relevant in this case. But in the locking case, we must be able to lock in the worst-case scenario of only one bit being random, so the min-entropy is the relevant quantity here.

The effect of non-maximal entanglement on the frontier width is not entirely clear however. There is a fairly large gap between our locking and decodability results here, but the locking side is almost certainly not tight in general. For instance, we can easily set up the system in such a way that there is a part of E' that is clearly useless, but our proof technique forces us to take this part into account, which artificially hurts our bound. This extreme case can be ruled out by restricting E' to the support of $\omega^{E'}$, but we are confident that similar gains could be found in the general case. A max-min gap in the entanglement would lend itself to a nicer interpretation of its contribution to the frontier width. In [23], the authors prove that the max-min entropy gap is the minimum amount of classical communication required to transform an arbitrary bipartite pure state into a maximally entangled state. If the optimal decoding procedure for Bob makes use of a maximally entangled state then this extra cost in the frontier width is associated with Bob transforming $\omega^{EE'}$ to $(\Phi^+)^{EE'}$. Note that the factor of $\frac{1}{2}$ can then be interpreted as superdense coding for this communication.

In general, although imperfect entanglement widens the frontier, the presence of perfect entanglement (i.e.: the maximally entangled state) helps to reduce the parameters p and ε greatly. Remember that these two govern the likelihood with which a locking scheme fails and the likelihood with which Bob might guess the

cyphertext, respectively. Thus our claim that the accessible information can be made arbitrarily small relies heavily on the presence of entanglement. This is because to achieve a given low value for ε at a fixed cyphertext size, one needs to include a large amount of perfect entanglement in the scheme.

In [13], Dodis and Smith generalized the definition of *entropic security* from an earlier result [33]. By assuming a lower bound on the min-entropy of Alice's message, *entropic security* was shown to imply the, much stronger, semantic security defined in [19]. Using the notion of quantum conditional min-entropy introduced by Renner [32], Desrosiers and Dupuis were able to show that entropic security was equivalent to a definition of entropic indistinguishability,

Definition 5.0.1 (Entropic Indistinguishability [10])**.** *An encryption system $\mathcal{E}$ is (t, ε)-indistinguishable if there exists a state $\Omega^{A'}$ such that for all states ρ^{AE} such that $H_{\min}(A|E)_\rho \geq t$ we have that:*

$$\left\| \mathcal{E}(\rho^{AE}) - \Omega^{A'} \otimes \rho^{E} \right\|_1 < \varepsilon$$

This definition, roughly equivalent to the definition of an ε-secure key in [24], is the proper characterization of entropic security. In [24], the authors considered a quantum key distribution setting and showed that n bits of key secured via an accessible information criterion could be learned by an adversary with access to $n - 1$ bits of the key. In this work we've shown a much more drastic violation. We've demonstrated that a key of size n that is ε-secure with respect to classical correlations (i.e.: accessible information), can be learned in as little as

$\log(n/\varepsilon)$ bits, reinforcing the need for entropic security over the use of accessible information.

CHAPTER 6
Conclusion

This work has defined information locking with a stronger definition than previously used and in a more generic setting. The effect of equipping the sender and receiver with shared entanglement is studied. Messages in non-uniform distributions are studied as well. The previous necessity for a distinct key register is eliminated in place of any arbitrary subsystem of sufficient size. All of these parameters contribute to a calculation of the frontier width; the middle ground between locking and decoding. Our results show that this middle ground is indeed very small, logarithmic in the size of the cyphertext and linear in the max-min entropy gap of the message. The results find potential application in black hole physics, as well as the study of the quantum discord.

The entanglement gap in the frontier width is likely to be improved to a max-min gap. Although the method for this would be via arguments about measurements on the support of the entanglement state, even further improvements might be possible via the application of measurement compression results such as [40].

References

[1] Anura Abeyesinghe, Igor Devetak, Patrick Hayden, and Andreas Winter. The mother of all protocols: Restructuring quantum information's family tree. *Proceedings of the Royal Society A*, (465):2537–2563, 2009. quant-ph/0606225.

[2] Y. Aharonov, A. Casher, and S. Nussinov. The unitarity puzzle and Planck mass stable particles. *Physics Letters B*, 191:51–55, 1987.

[3] Rudolf Ahlswede and Andreas Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002. quant-ph/0012127.

[4] Robert Alicki and Mark Fannes. Continuity of quantum mutual information. *Journal of Physics A: Mathematical and General*, 37(5):L55–L57, 2004. quant-ph/0312081.

[5] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2009. http://www.wisdom.weizmann.ac.il/ zeitouni/cupbook.pdf.

[6] Rajendra Bhatia. *Matrix Analysis*. Springer-Verlag, 1996.

[7] Samuel L. Braunstein, Hans-Jürgen Sommers, and Karol Zyczkowski. Entangled black holes as ciphers of hidden information. arXiv:0907.0739, 2009.

[8] R. D. Carlitz and R. S. Willey. Lifetime of a black hole. *Physical Review D*, 36:2336–2341, 1987.

[9] E. Davies. Information and quantum measurement. *IEEE Transactions on Information Theory*, 24:596–599, 1978. 10.1109/TIT.1978.1055941.

[10] Simon Pierre Desrosiers and Frédéric Dupuis. Quantum entropic security and approximate quantum encryption. *IEEE Transactions on Information Theory*. to appear. arXiv:0707.0691.

[11] David DiVincenzo, Debbie Leung, and Barbara Terhal. Quantum data hiding. *IEEE Trans. Inf. Theory*, 48(3):580–598, 2002. quant-ph/0103098.

[12] David P. DiVincenzo, Michał Horodecki, Debbie W. Leung, John A. Smolin, and Barbara M. Terhal. Locking classical correlation in quantum state. *Phys. Rev. Lett.*, (92, 067902), 2004. quant-ph/0303088.

[13] Yevgeniy Dodis and Adam Smith. Entropic security and the encryption of high entropy messages. Cryptology ePrint Archive, Report 2004/219, 2004.

[14] Frédéric Dupuis. *The decoupling approach to quantum information theory.* PhD thesis, Université de Montréal, 2009. arXiv:1004.1641.

[15] Frédéric Dupuis, Jan Florjanczyk, Patrick Hayden, and Debbie Leung. Locking classical information. 2010. arxiv:1011.1612.

[16] Freeman J. Dyson. Statistical theory of the energy levels of complex systems. i. *Journal of Mathematical Physics*, 3(1):140–156, 1962.

[17] Omar Fawzi, Patrick Hayden, and Pranab Sen. From low-distortion embeddings to metric uncertainty relations and information locking. 2010. arxiv:1010.3007v3.

[18] Christopher A. Fuchs. *Distinguishability and accessible information in quantum theory.* PhD thesis, University of New Mexico, 1996. quant-ph/9601020.

[19] Shafi Goldwasser and Silvio Micali. Probabilistic encryption & how to play mental poker keeping secret all partial information. In *STOC '82: Proceedings of the fourteenth annual ACM Symposium on Theory of computing*, pages 365–377, New York, NY, USA, 1982. ACM Press.

[20] Lucien Hardy. Quantum theory from five reasonable axioms. 2001. arxiv:quant-ph/0101012.

[21] Patrick Hayden, Debbie Leung, Peter Shor, and Andreas Winter. Randomizing quantum states: Constructions and applications. *Comm. Math. Phys.*, 250(2):371–391, 2004. quant-ph/0307104.

[22] Patrick Hayden and John Preskill. Black holes as mirrors: quantum information in random subsystems. *Journal of High Energy Physics*, 07(09):120, 2007.

[23] Patrick Hayden and Andreas Winter. Communication cost of entanglement transformations. *Physical Review A*, 67, 2003.

[24] Robert König, Renato Renner, Andor Bariska, and Ueli Maurer. Locking of accessible information and implications for the security of quantum cryptography. *Phys. Rev. Lett.*, 98(140502), 2007. quant-ph/0512021.

[25] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Dover Publications, February 1999.

[26] Michel Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.

[27] Seth Lloyd. Almost certain escape from black holes in final state projection models. *Physical Review Letters*, 96:061302, 2006.

[28] Hans Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Phys. Rev. Lett.*, 60(12):1103–1106, Mar 1988.

[29] M.L. Mehta. *Random matrices*. Pure and applied mathematics. Elsevier/Academic Press, 2004.

[30] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, New York, NY, USA, 2000.

[31] Harold Ollivier and Wojciech H. Zurek. Quantum discord: a measure of the quantumness of correlations. *Physical Review Letters*, 88:017901, 2001.

[32] Renato Renner. *Security of quantum key distribution*. PhD thesis, ETH Zurich, 2005. quant-ph/0512258.

[33] Alexander Russell and Hong Wang. How to fool an unbounded adversary with a short key. In *EUROCRYPT '02: Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques*, pages 133–148, London, UK, 2002. Springer-Verlag.

[34] Y. Sekino and L. Susskind. Fast scramblers. *Journal of High Energy Physics*, 10:65–+, 2008. arxiv:0808.2096.

[35] Claude Elwood Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[36] John Smolin and Jonathan Oppenheim. Locking information in black holes. *Physical Review Letters*, 96(8):081302–+, 2006.

[37] W. Forrest Stinespring. Positive functions on c*-algebras. *Proceedings of the American Mathematical Society*, 6(2):pp. 211–216, 1955.

[38] T. Werlang, S. Souza, F. F. Fanchini, and C. J. Villas Boas. Robustness of quantum discord to sudden death. *Physical Review A*, 80:024103, 2009.

[39] Eugene P. Wigner. On the statistical distribution of the widths and spacings of nuclear resonance levels. *Mathematical Proceedings of the Cambridge Philosophical Society*, 47(04):790–798, 1951.

[40] A. Winter and S. Massar. Compression of quantum measurement operations. Phys. Rev. A 64, 012311 (2001), 2000.

APPENDIX A
Appendix

Definition A.0.2 (σ-algebra)**.** *Consider the set $\mathcal{X}$. We call $\mathcal{A}$ the* σ-algebra *of $\mathcal{X}$ if it is the collection of subsets of $\mathcal{X}$ such that: $\mathcal{A}$ is non-empty, closed under countable union, and closed under countable intersection.*

Definition A.0.3 (Measure)**.** *Consider the set $\mathcal{X}$ with the σ-algebra $\mathcal{A}$. We call a function $\mu : \mathcal{A} \to \mathbb{R}$ a* measure *if it is positive, takes the value 0 on the empty set and is countably additive as follows:*

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu\left(A_i\right)$$

for countably many disjoint sets A_i.

Definition A.0.4 (Lipschitz constant)**.** *Let $f : \mathfrak{X} \to \mathfrak{Y}$ be a function from the metric space $(\mathfrak{X}, d_{\mathfrak{X}})$ to the metric space $(\mathfrak{Y}, d_{\mathfrak{Y}})$. Then, the Lipschitz constant of f is defined as*

$$\sup_{x_1, x_2 \in \mathfrak{X}} \frac{d_{\mathfrak{Y}}(f(x_1), f(x_2))}{d_{\mathfrak{X}}(x_1, x_2)}.$$

If the above quantity is not bounded, the constant is not defined.

Lemma A.0.7 (Lemma IV.3 in [1])**.** *For any matrix $X^{A\overline{A}R}$ and for $\mathrm{d}U$ the Haar measure over unitaries, we have the following property:*

$$\int_U \left(U_A \otimes U_{\overline{A}} \otimes \mathbb{I}^R\right) X^{A\overline{A}R} \left(U_A^\dagger \otimes U_{\overline{A}}^\dagger \otimes \mathbb{I}^R\right) \mathrm{d}U = \alpha_+\left(X\right) \otimes \Pi_+^A + \alpha_-\left(X\right) \otimes \Pi_-^A$$

where

$$\alpha_{\pm}(X) = \frac{\mathrm{Tr}_{A\overline{A}}\left[X(\Pi_{\pm}^{A} \otimes \mathbb{I}^{R})\right]}{\mathrm{rank}\left(\Pi_{\pm}^{A}\right)} \quad \Pi_{\pm}^{A} = \frac{1}{2}\left(\mathbb{I}^{A\overline{A}} \pm F_{\overline{A}}^{A}\right) \quad \mathrm{rank}\left(\Pi_{\pm}^{A}\right) = \frac{|A|(|A| \pm 1)}{2}.$$

Lemma A.0.8 (Operator Chernoff bound [3])**.** *Let $X_1, \ldots, X_M$ be i.i.d. random variables taking values in the operators* $\mathrm{Pos}(\mathsf{A})$*, with $0 \leqslant X_j \leqslant \mathbb{I}$, with $A = \mathbb{E}X_j \geqslant \alpha\mathbb{I}$, and let $0 < \eta \leqslant 1/2$. Then*

$$\Pr\left\{\frac{1}{M}\sum_{j=1}^{M} X_j \notin (1+\eta)A\right\} \leqslant 2|A| \exp\left(-M\frac{\alpha\eta^2}{2\ln 2}\right). \tag{A.0.1}$$

Lemma A.0.9 (Trace distance versus Euclidean norm for pure states (See, e.g. [30].))**.** *Consider any two quantum states $|\varphi\rangle, |\tilde{\varphi}\rangle$ with associated density operators $\varphi, \tilde{\varphi}$ respectively. We can relate the* 1*-norm distance between the operators to the* 2*-norm distance of the states as follows,*

$$\|\varphi - \tilde{\varphi}\|_1 \leq 2\,\||\varphi\rangle - |\tilde{\varphi}\rangle\|_2\,.$$

Lemma A.0.10 (A bound for the 1-norm in terms of conditional entropy [32, 14])**.** *Let $\rho \in \mathrm{L}(A)$ be any Hermitian operator and let $\gamma \in \mathrm{Pos}(A)$ be a positive definite operator. Then,*

$$\|\rho\|_1 \leqslant \sqrt{\mathrm{Tr}\left[\gamma\right]\mathrm{Tr}\left[\left(\gamma^{-1/4}\rho\gamma^{-1/4}\right)^2\right]}.$$

Proof.

$$\begin{aligned}
\|\rho\|_1 &= \max_{U\in\mathcal{U}(A)} |\mathrm{Tr}\left[U\rho\right]| \\
&= \max_{U\in\mathcal{U}(A)} \left|\mathrm{Tr}\left[\left(\gamma^{1/4}U\gamma^{1/4}\right)\left(\gamma^{-1/4}\,\rho\,\gamma^{-1/4}\right)\right]\right| \\
&\leqslant \max_{U\in\mathcal{U}(A)} \sqrt{\mathrm{Tr}\left[\left(\gamma^{1/4}U\gamma^{1/4}\right)\left(\gamma^{1/4}U^\dagger\gamma^{1/4}\right)\right]\mathrm{Tr}\left[\gamma^{-1/4}\,\rho\,\gamma^{-1/2}\,\rho^\dagger\,\gamma^{-1/4}\right]} \\
&= \sqrt{\max_{U\in\mathcal{U}(A)} \mathrm{Tr}\left[\gamma^{1/2}U\gamma^{1/2}U^\dagger\right]\mathrm{Tr}\left[\gamma^{-1/4}\,\rho\,\gamma^{-1/2}\,\rho^\dagger\,\gamma^{-1/4}\right]} \\
&= \sqrt{\mathrm{Tr}\left[\gamma\right]\mathrm{Tr}\left[\gamma^{-1/4}\,\rho\,\gamma^{-1/2}\,\rho^\dagger\,\gamma^{-1/4}\right]},
\end{aligned}$$

where the first equality is an application of Lemma I.6 in [14] and the inequality results from an application of Cauchy-Schwarz, and the maximizations are over all unitaries on A. The last equality follows from

$$\begin{aligned}
\max_{U\in\mathcal{U}(A)} \mathrm{Tr}\left[\gamma^{1/2}U\gamma^{1/2}U^\dagger\right] &\leqslant \max_{U\in\mathcal{U}(A)} \sqrt{\mathrm{Tr}\left[\gamma\right]\mathrm{Tr}\left[U\gamma^{1/2}U^\dagger U\gamma^{1/2}U^\dagger\right]} \\
&= \mathrm{Tr}\left[\gamma\right] \\
&\leqslant \max_{U\in\mathcal{U}(A)} \mathrm{Tr}\left[\gamma^{1/2}U\gamma^{1/2}U^\dagger\right].
\end{aligned}$$

□

Lemma A.0.11 (Transpose trick)**.** *Given any positive operator M and the maximally entangled bipartite state $|\Phi^+\rangle$,*

$$(M^A\otimes I^B)|\Phi^+\rangle\left(I^A\otimes(M^T)^B\right)|\Phi^+\rangle.$$