Probabilistic and Approximate Masking of Quantum Information

Mao-Sheng Li\textsuperscript{1,2} and Kavan Modi\textsuperscript{3,4,*}

\textsuperscript{1}Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China  
\textsuperscript{2}Department of Physics, University of Science and Technology of China, Hefei 230026, China  
\textsuperscript{3}School of Physics & Astronomy, Monash University, Victoria 3800, Australia  
\textsuperscript{4}Institute for Quantum Science and Engineering, and Department of Physics, Southern University of Science and Technology, Shenzhen 518055, China

(Dated: July 8, 2021)

The no-masking theorem states that it is impossible to encode an arbitrary quantum state into the correlations between two subsystems so that no original information about is accessible in the marginal state of either subsystem. In this paper, we generalize this theorem to allow for failure of the protocol. We then bound the performance of a masking protocol when we are allowed a (probabilistic) approximate protocol.

\section{INTRODUCTION}

A quantum cloning machine, that can make perfect copies of an arbitrary quantum state, is forbidden by the structure of quantum mechanics \cite{1, 2}. This simple, yet profound, phenomenon plays a central role in many quantum information tasks, such as quantum teleportation \cite{3, 4} and quantum key distribution \cite{5, 6}. This pioneering result has paved the way for a number of similar no-go theorems, such as the no-broadcasting theorem \cite{7–9}, the impossibility of a bit commitment protocol \cite{10, 11}, the no-deleting theorem \cite{12}, and the no-hiding theorem \cite{13, 14}. On the other hand, the structure of quantum mechanics allows for several potential applications, such quantum key distribution \cite{15, 16} and secret sharing \cite{17–19}. In essence, all these phenomena are consequences of the superposition principle, along with the unitary evolution of quantum mechanics. Importantly, each theorem and task here sheds light on the boundary between the classical and the quantum worlds, which is of fundamental interest in physics.

An important classical protocol that enables secret communication is the one-time pad: here a classical message is combined with a random key (of the same size) so that neither the message nor the key contains the original information. However, the information if perfectly preserved in the correlations between the message and the key \cite{20}. It is natural to consider the quantum analogue of the classical one-time pad. That is, can quantum information be hidden from both subsystems and reside entirely in the correlation? Once again, the answer turns out to be no. Recently, Ref. \cite{21} showed that this is impossible and named such phenomenon as the no-masking theorem. As a by-product, this result implicates a quantum version of the well-known impossibility of bit commitment protocols \cite{10, 11}, i.e., the impossibility of a protocol that allows the commitment of qubits. However, it turns out that, while universal masking is impossible, the set of states that can be masked is non-trivial \cite{21–24}. Moreover, when one allows for some trusted randomness in the protocol, the set of maskable states grows \cite{25, 26}, even to a point where universal masking becomes possible. These non-trivial features of the no-masking theorem has generated significant interest in the recent years \cite{27–32}.

When dealing with information-theoretic protocols, it is important to account for their robustness. That is, while making an exact copy of a quantum state is forbidden, is it also forbidden to make an approximate copy? This question was examined by several authors \cite{33, 34}, who found that it is possible to make imperfect copies of a quantum state. Similarly, others examined if probabilistic cloning of quantum states is possible \cite{35}. Then one natural question is to ask whether approximate or probabilistic quantum masking machines are allowed? If so, then some conditionally secure quantum qubit commitment could also be possible. In this paper, we prove that probabilistic universal masking is impossible and provide bounds on the achievable fidelity for an approximate universal masking machine. Finally, we show that this bound remains intact if we make the approximate masker also probabilistic. We begin by defining the masking protocol.

\section{PRELIMINARIES}

The masking protocols involves two parties, $A$ and $B$. Let $\mathcal{H}_A \otimes \mathcal{H}_B$ be Hilbert spaces of dimension $r \geq 2$ and $s \geq 2$, respectively. We denote the set of all pure states in $\mathcal{H}_X$ as $\mathcal{P}(\mathcal{H}_X)$, while $\mathcal{D}(\mathcal{H}_X)$ denotes all the density matrices of the system $X$, and the set $\mathcal{A} = \{|a_k\rangle_A \in \mathcal{H}_A\}$ is the set of maskable states.

\textbf{Definition 1} An operation $\mathcal{S}$ is said to mask quantum information contained in states $A$ by mapping them to states $\{|\Psi_k\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B\}$ such that all the marginal states of $|\Psi_k\rangle_{AB}$ are identical, i.e.,

$$\rho_A = \text{Tr}_B(|\Psi_k\rangle_{AB}\langle\Psi_k|) \quad \text{and} \quad \rho_B = \text{Tr}_A(|\Psi_k\rangle_{AB}\langle\Psi_k|),$$

for all $k$. The masker can be modelled by a unitary operator $U_S$ on $AB$, where $B$ is an ancillary system in a fixed state $|b\rangle$:

$$\mathcal{S} : \quad U_S|a_k\rangle_A|b\rangle_B = |\Psi_k\rangle_{AB}.$$  

One can think of the masker to be a triple $\mathcal{M} = (U_S, \rho_A, \rho_B)$. An interesting goal is to determined the maximal set $S$ of states that can be masked by a given masker.
Theorem 1 [21] An arbitrary quantum state in $\mathcal{F}(\mathcal{H}_A)$ cannot be masked.

By linearity the proof also applies for masking states in $\mathcal{D}(\mathcal{H}_A)$. The above theorem assumes that the initial state of $B$ is pure. This is because, as stated in Ref. [21], for more than two parties masking is allowed [18, 19]. Indeed Refs. [25, 26] have shown that if this initial state is allowed to be mixed, i.e., there exists another system $C$ that purifies $B$, then universal masking is possible. The no-masking theorem states that purifying is possible. The no-masking theorem states that if a given input $k$ there is no system $S$ that purifies $B$. Clearly, such a machine will be less powerful than a machine which is such a process. Without loss of generality, we also need to require that $L_{S_p}$ is invertible, i.e., there exists mapping from $|\Psi_k\rangle_{AB} \rightarrow |\alpha_k\rangle_A |b\rangle_B$. Clearly, such a machine will be less powerful than one that is guaranteed to achieve masking. We now show that $L_{S_p}$ also cannot mask all states of $\mathcal{F}(\mathcal{H}_A)$.

Theorem 2 A probabilistic masker, that can mask all the states in $\mathcal{F}(\mathcal{H}_A)$, is impossible.

Proof: We prove this claim by contradiction. Suppose $(L_{S_p}, \rho_A, \rho_B)$ is such a process. Without loss of generality, we can assume $\rho_A = \sum_{j=1}^{n} \lambda_j |j\rangle\langle j|$, with $\lambda_j > 0$. Set

$$L_{S_p}|1\rangle_B = p_1 |\Psi_1\rangle, \quad L_{S_p}|2\rangle_B = p_2 |\Psi_2\rangle. \quad (3)$$

By assumption, $|\Psi_1\rangle, |\Psi_2\rangle$ are linearly independent and can be seen as two different purifications of $\rho_A$. Therefore, they can be written as the following form

$$|\Psi_1\rangle = \sum_{j=1}^{n} \sqrt{\lambda_j} |j\rangle |\mu_j\rangle, \quad |\Psi_2\rangle = \sum_{j=1}^{n} \sqrt{\lambda_j} |j\rangle |\nu_j\rangle, \quad (4)$$

where $\{|\mu_j\rangle\}_{j=1}^{n}$ and $\{|\nu_j\rangle\}_{j=1}^{n}$ are both orthonormal sets.

Now for any pair $u_i, v_i \in \mathbb{C}$, such that $|u_i|^2 + |v_i|^2 = 1$,

$$L_{S_p}((u_i|1\rangle + v_i|2\rangle)|b\rangle) = u_i p_1 |\Psi_1\rangle + v_i p_2 |\Psi_2\rangle$$

we get a nonzero vector as $L_{S_p}$ is invertible. Let $\lambda_i$ be the normalization for this vector. By assumption, the resulting states are purifications of $\rho_A$ and thus can be expressed as

$$|\Psi_i\rangle := \sum_{j=1}^{n} \sqrt{\lambda_j} |j\rangle \frac{u_i |p_1 |\mu_j\rangle + v_i |p_2 |\nu_j\rangle}{\lambda_i}. \quad (6)$$

This means that the set of states $\{|\Psi_i\rangle\}_{i=1, \ldots, n}$ are orthonormal and for all $1 \leq i < k \leq n$, we have

$$p_1 p_2 u_i \overline{v_j} |\mu_j\rangle + p_1 p_2 v_i \overline{v_j} |\mu_j\rangle = 0, \quad \text{here } \overline{v_j} \text{ means the complex conjugation of } v_j.$$

Particularly, choosing two sets $(u_1, v_1)$ and $(u_2, v_2)$ such that $u_1 \overline{v_1} u_2 \overline{v_2}$ is not in $\mathbb{R}$, we then have

$$u_1 \overline{v_1} |\mu_1\rangle + \overline{v_1} u_2 |\mu_2\rangle = 0, \quad u_1 |\mu_1\rangle + v_1 |\mu_2\rangle = 0.$$  \quad (7)

The condition $u_1 \overline{v_1} u_2 \overline{v_2}$ is not in $\mathbb{R}$ implies that the determinant of the above coefficients is nonzero. Therefore, we can deduce that $|\mu_j\rangle = 0$ for all $j \neq k$. The states corresponding to the images of $L_{S_p}$ can be also viewed as purifications of $\rho_B$ and by Eqs. (4), we have

$$\sum_{j=1}^{n} \lambda_j |\mu_j\rangle |\nu_j\rangle = \rho_B = \sum_{k=1}^{n} \lambda_k |\nu_k\rangle |\mu_k\rangle. \quad (8)$$

Calculating $\text{Tr}(\rho_B |\mu_k\rangle |\mu_k\rangle$ by substituting $\rho_B$ in two ways, we have $\lambda_i = |\langle \mu_i |\mu_k \rangle|^2$, hence $|\langle \mu_i |\mu_k \rangle|^2 = 1$ for $\lambda_i \neq 0$ and $|\langle \mu_i |\nu_k \rangle|^2 = e^{i\theta_l} |\mu_k\rangle$ for some $\theta_l \in \{0, 2\pi\}$ with $l = 1, 2, \ldots, n$.

This turns the coefficient in the numerator of Eq. (6) into $p_1 u_i + p_2 e^{i\theta_j} u_i$. Using the fact that the states in Eq. (6) are also purifications of $\rho_B$, we have equalities: $|p_1 u_i + p_2 e^{i\theta_j} u_i|^2 = |p_1 u_i + p_2 e^{i\theta_j} u_i|^2$ for all $j, k \in \{1, 2, \ldots, n\}$. To see this, we trace out system $A$ of $|\Psi_i\rangle$, which gives us

$$\rho_B = \frac{1}{N^2} \sum_{j=1}^{n} \lambda_j |\mu_j\rangle |\nu_j\rangle.$$  \quad (9)

Comparing this with the $\rho_B$ in Eq. (8), we must have $|p_1 u_i + p_2 e^{i\theta_j} u_i|^2 = N^2$ for all $j, k \in \{1, 2, \ldots, n\}$. This is equivalent to

$$u_1 \overline{v_1} (e^{-i\theta_j} - e^{-i\theta_k}) + \overline{v_1} u_2 (e^{i\theta_j} - e^{i\theta_k}) = 0, \quad u_1 \overline{v_1} (e^{-i\theta_j} - e^{-i\theta_k}) + \overline{v_1} u_2 (e^{i\theta_j} - e^{i\theta_k}) = 0.$$  \quad (10)

The nonzero of determinant $u_1 \overline{v_1} u_2 \overline{v_2}$ gives us $e^{-i\theta_j} - e^{-i\theta_k} = e^{i\theta_j} - e^{i\theta_k} = 0$ for all $1 \leq j, k \leq n$, which simply means $|\Psi_2\rangle = e^{i\theta_j} |\Psi_1\rangle$. This is in contradiction with the assumption of linear independence of $|\Psi_1\rangle$ and $|\Psi_2\rangle$.

Remark: From the above proof, with the assumption that $|\Psi_1\rangle$ and $|\Psi_2\rangle$ are linearly independent and $u_1 \overline{v_1} u_2 \overline{v_2}$ is not in $\mathbb{R}$, we can indeed obtain that four
An approximate universal masker
the information of the original states can only be obtained
marginal states are close to a fiducial state. In this case,
each other if we find it difficult to distinguish between them.
Physically, two states $\rho_A$ and $\rho'_A$ are said to be close to
each other if we find it difficult to distinguish between them.
Consider a masker maps a set of input states to states whose
marginal states are close to a fiducial state. In this case, the
information of the original states can only be obtained
under some threshold hence approximate masking would be
achieved. To formally quantify the distinguishability between
two states we employ the quantum fidelity, $F(P, Q) :=$
$\text{Tr}\sqrt{PQ}\sqrt{P}$, where $P, Q$ are two density matrices [37, 38].
Notice that $0 \leq F(P, Q) \leq 1$ where the greater the fidelity the
more difficult it is to distinguish this two matrices $P$ and $Q$.
With this we define an $\epsilon$-approximate masker as the following.

**Definition 3** An operation $S_\epsilon$ is said to $\epsilon$-approximate mask quantum information contained in states of set $A$ by mapping
them to states $\{|\Psi_k\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B\}$ such that all marginal states of $|\Psi_k\rangle_{AB}$ approximately satisfy Eq. (1), i.e., they are only $\epsilon$ distinguishable from each other:

$$F(\rho_A |k, \rho'_A |k') \geq 1 - \epsilon, \quad F(\rho_B |k, \rho'_B |k') \geq 1 - \epsilon,$$

for all $k, k'$. Here $\rho_A |k$ and $\rho_B |k$ are the marginals of $|\Psi_k\rangle_{AB}$.

An $\epsilon$-approximate masker can be designed by replacing $U_S$ in
Eq. (2) with another unitary $U_S$, and demanding conditions
(11). Quantum fidelity is closely related to the trace norm $\| \cdot \|_1$ by the inequality (see Ref. [38], p161):

$$F(P, Q) \leq \sqrt{1 - \frac{1}{4} \| P - Q \|_1^2}.$$  

If the approximate conditions of Ineq. (11) holds, the above
inequality implies that

$$\| \rho_{X |k} - \rho_{X |k'} \|_1 \leq 2\sqrt{\epsilon} \quad \text{with} \quad X \in \{A, B\}.$$  

We now bound how well $\epsilon$-approximate masking is possible.

**Theorem 3** An approximate universal masker $S_\epsilon$, that can
mask all states in $\mathcal{F}(\mathcal{H}_A)$ with fidelity greater than $1 - \epsilon$,
can only do so for $\epsilon \geq \frac{1}{\sqrt{2}} \sqrt{1 - \min_{r, \sigma} \frac{3r}{2}}$, where $r = \dim \mathcal{H}_A, s = \dim \mathcal{H}_B$.

**Proof:** Set $t := \min\{r, s\}$. Let $\{|j\rangle_A | 1 \leq j \leq r\}$ and
$\{|k\rangle_A | 1 \leq k \leq s\}$ be orthonormal basis of $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. Suppose that $U_{S_\epsilon} |1\rangle_A |B = |\Psi\rangle, U_{S_\epsilon} |2\rangle_A |B = |\Phi\rangle$. The matrices representation of $|\Psi\rangle$ and $|\Phi\rangle$ are denoted by two matrices $M = (m_{jk}), N = (n_{jk}) \in \text{Mat}_{r \times s}(\mathbb{C})$ respectively, i.e., $|\Psi\rangle = \sum_{j=1}^r \sum_{k=1}^s m_{jk} |j\rangle_A |k\rangle_B, |\Phi\rangle = \sum_{j=1}^r \sum_{k=1}^s n_{jk} |j\rangle_A |k\rangle_B$. Then the partial traces of $|\Psi\rangle\langle\Psi|$ and $|\Phi\rangle\langle\Phi|$ can be calculated as follows:

$$\text{Tr}_A (|\Psi\rangle\langle\Psi|) = MM^\dagger, \quad \text{Tr}_A (|\Phi\rangle\langle\Phi|) = NN^\dagger;$$  

$$\text{Tr}_B (|\Psi\rangle\langle\Psi|) = MM^\dagger, \quad \text{Tr}_B (|\Phi\rangle\langle\Phi|) = NN^\dagger$$

The definition of $\epsilon$-approximate quantum masking yields

$$\| MM^\dagger - NN^\dagger \|_1 \leq 2\sqrt{2}\epsilon, \quad \| MM^\dagger - NN^\dagger \|_1 \leq 2\sqrt{2}\epsilon.$$  

Moreover, considering another general state $|\Omega\rangle = u|\Psi\rangle + v|\Phi\rangle \in \mathcal{F}(\mathcal{H}_A \otimes \mathcal{H}_B)$ for all $|u|^2 + |v|^2 = 1$, we may find that the matrix corresponding to $|\Omega\rangle$ is just $uM + vN$. Therefore, $\text{Tr}_B (|\Omega\rangle\langle\Omega|) = (uM + vN)(uM + vN)^\dagger$. By the definition of $\epsilon$-approximate quantum masking, we should also have $\| \text{Tr}_B (|\Omega\rangle\langle\Omega|) - \text{Tr}_B (|\Psi\rangle\langle\Psi|) \|_1 \leq 2\sqrt{2}\epsilon$. Writing it into the
matrix form we deduce the following inequality

$$\| u^2 (MM^\dagger - NN^\dagger) + u\sigma MN^\dagger + v\sigma NM^\dagger \|_1 \leq 2\sqrt{2}\epsilon.$$  

By triangle inequality, $\| u\sigma MN^\dagger + v\sigma NM^\dagger \|_1$ is less than

$$|u|^2 \| MM^\dagger - NN^\dagger \|_1 + \| u\|^2 = \| MM^\dagger - NN^\dagger \|_1 + \| v\|^2$$

which is less than $2\sqrt{2}(1 + |u|^2)\epsilon$. Therefore,

$$\| u\sigma MN^\dagger + v\sigma NM^\dagger \|_1 \leq 2\sqrt{2}(1 + |u|^2)\epsilon.$$  

Choosing $(u, v)$ as the following pairs $(u_1, v_1) = (\sqrt{\frac{1}{\sqrt{2}}} - \frac{1}{\sqrt{2}})$
and $(u_2, v_2) = (\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}})$ the above inequality becomes

$$\| MM^\dagger - NN^\dagger \|_1 \leq 3\sqrt{2}\epsilon.$$  

Using the triangle inequality of trace norm, we obtain $\| MN^\dagger \|_1 \leq 6\sqrt{2}\epsilon$, and $\| NN^\dagger \|_1 \leq 6\sqrt{2}\epsilon$. Using the compatible condition of norm, we obtain an upper bound of the modulo of $\text{Tr}(MN^\dagger MN^\dagger)$.

That is, $\| \text{Tr}(MN^\dagger MN^\dagger) \|_1 \leq \| MN^\dagger \|_1 \| NN^\dagger \|_1$, and the last term is less than $72\epsilon^2$.

On the other hand, $\text{Tr}(MN^\dagger MN^\dagger) = \text{Tr}(MM^\dagger N)$.
We will give a lower bound of $\| \text{Tr}(MM^\dagger N) \|$. Set $L := N^\dagger M^\dagger M \in \text{Mat}_{r \times s}(\mathbb{C})$. Substituting $N^\dagger M^\dagger = MM^\dagger + L$ into the above trace, we obtain

$$\| \text{Tr}(MM^\dagger N) \| = \frac{1}{t} \| \text{Tr}(MM^\dagger) \| - \| \text{Tr}(MM^\dagger L) \|.$$  

We claim that $\| \text{Tr}(MM^\dagger L) \| \leq 2\sqrt{2}\epsilon$, which we prove below. Using this claim and the fact that $\text{Tr}(MM^\dagger) = 1$, we have

$$\| \text{Tr}(MM^\dagger N) \| \geq \frac{1}{t} - \| \text{Tr}(MM^\dagger L) \| \geq \frac{1}{t} - 2\sqrt{2}\epsilon.$$  

Hence we obtain the inequality which we need to prove
\[ 72\epsilon^2 \geq |\text{Tr}(MN^\dagger NM^\dagger)| = |\text{Tr}(M^\dagger M N)| \geq \frac{1}{\ell} - 2\sqrt{2}\epsilon. \]

Using the quadratic formula we find the lower bound on \( \epsilon \) given in the statement of the theorem.

Now we focus on proving the claim stated above. First, the second inequality of (14) implies that \( \|L\|_2 \leq \|L\|_1 \leq 2\sqrt{2}\epsilon \). There exists some unitary matrix \( U \in U(s) \) such that \( UM^1MU^\dagger \) is diagonal and the diagonals are \( x_1, x_2, \ldots, x_s \). Let \( (y_1, y_2, \ldots, y_s) \) be the diagonal elements of \( ULU^\dagger \). As \( \|ULU^\dagger\|_2 = \|L\|_2 \), we have
\[ |\text{Tr}(M^\dagger ML)| = |\text{Tr}(UM^1MU^\dagger ULU^\dagger)| \leq \sum_{j=1}^{s} x_j |y_j| \leq \sum_{j=1}^{s} x_j \|L\|_2 = \|L\|_2 \leq 2\sqrt{2}\epsilon. \]

**Remark:** In the case \( \epsilon = 0 \) for the above theorem, we can also arrive at the conclusion of the no-masking theorem.

Bužek and Hillery showed that a universal approximate cloning machine can clone an arbitrary unknown qubit state with the surprisingly high fidelity of \( \sqrt{5/6} \approx 0.913 \) [33]. Our bound here places restrictions on the performance of a universal masker in terms of the fidelity between the marginals of different masked states. It is interesting to compare our bound with that of Bužek and Hillery for the case of a qubit. In this case, the fidelity of the local states cannot greater than \( 0.6411 \), which quantifies how much of the input information is revealed by the marginals.

**Probabilistic \( \epsilon \)-approximate masking.**— Before we make our conclusions, we look at the case where approximate masking is also allowed to be probabilistic. The next theorem is stronger statement contains the last theorem as a limiting case. Here, we bound the fidelity of an approximate masker if we allow for a probability of failure.

**Definition 4** An invertible operation \( L_{\mathcal{S}_x} \) is said to be probabilistic \( \epsilon \)-approximate masking quantum information contained in states \( \mathcal{A} = \{ |\alpha_k\rangle_A \in \mathcal{H}_A \} \) by mapping them to \( \{ |\beta_k\rangle_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B \} \) such that the marginal states of \( |\beta_k\rangle_{AB} \) cannot be distinguished with each other in the sense of \( \epsilon \), i.e., they satisfy the approximate bounds in (11).

**Theorem 4** An \( \epsilon \)-approximate-probabilistic masker \( L_{\mathcal{S}_x} \) that can mask all states in \( \mathcal{P}(\mathcal{H}_A) \) with fidelity greater than \( 1 - \epsilon \), can only do so for \( \epsilon \geq \frac{\pi/2}{\sqrt{2}} \left(-1 + \sqrt{1 + \min_{r,s}(\epsilon)}\right)\).

**Proof:** Since \( L_{\mathcal{S}_x} \) is invertible, \( \dim_{\mathcal{C}}(L_{\mathcal{S}_x}(\mathcal{H}_A \otimes |b\rangle)) = \dim_{\mathcal{C}}(\mathcal{H}_A) = r \geq 2 \). Hence there exist \( |\alpha_1\rangle, |\alpha_2\rangle \in \mathcal{P}(\mathcal{H}_A) \) such that \( L_{\mathcal{S}_x}|\alpha_1\rangle_A |b\rangle_B = p_2 |\Psi_j\rangle_j \), \( j = 1, 2 \) with \( p_1, p_2 > 0 \) and \( \langle \Psi_1 |\Psi_2\rangle = 0 \). The matrix representation of \( |\Psi_j\rangle \) is denoted by the matrix \( M_j = (m_{ij}^{(j)}) \in \text{Mat}_{r \times s}(\mathcal{C}) \). That is, \( |\Psi_j\rangle = \sum_{k=1}^{r} \sum_{l=1}^{s} m_{kl}^{(j)} |k\rangle_A |l\rangle_B \). Then we have the following partial traces
\[ \text{Tr}_A(|\Psi_j\rangle \langle \Psi_j|) = M_j^\dagger M_j, \quad \text{Tr}_B(|\Psi_j\rangle \langle \Psi_j|) = M_j M_j^\dagger. \]

The definition of \( \epsilon \)-approximate probabilistic quantum masking yields
\[ \|M_1 M_1^\dagger - M_2 M_2^\dagger\|_1 \leq 2\sqrt{2}\epsilon. \] (21)

We define \( S(\theta) := \{(x, y) \in \mathbb{R}^2 | x|a_1\rangle + ye^{i\theta}|a_2\rangle \in \mathcal{P}(\mathcal{H}_A) \} \) for all \( \theta \in [0, 2\pi] \). For all \( (x, y) \in S(\theta) \), we have
\[ L_{\mathcal{S}_x}((x|a_1\rangle + ye^{i\theta}|a_2\rangle) |b\rangle_B) = \sqrt{(xp_1)^2 + (yp_2)^2} |\Omega_{x,y,\theta}\rangle, \]
where \( |\Omega_{x,y,\theta}\rangle = (xp_1 |\Psi_1\rangle + yp_2 e^{i\theta} |\Psi_2\rangle) / \sqrt{(xp_1)^2 + (yp_2)^2} \) is a normalized state, where we have use the orthonormality of \( |\Psi_1\rangle \) and \( |\Psi_2\rangle \) which plays a significant role in the proof. Lemma 1, given below, states that for any \( \theta \in [0, 2\pi] \) we can find some \( (x, y) \in S(\theta) \) such that \( xp_1 = yp_2 \). For such parameters, \( |\Omega_{x,y,\theta}\rangle = (|\Psi_1\rangle + e^{i\theta} |\Psi_2\rangle) / \sqrt{2} \). The matrix representation of such state is \( (M_1 + e^{i\theta} M_2) / \sqrt{2} \). The partial trace \( \text{Tr}_B(|\Omega_{x,y,\theta}\rangle \langle \Omega_{x,y,\theta}|) \) can be computed as
\[ \frac{1}{2} (M_1 M_1^\dagger + M_2 M_2^\dagger + e^{-i\theta} M_1 M_1^\dagger + e^{i\theta} M_2 M_2^\dagger). \] (22)

The definition of \( \epsilon \)-approximate probabilistic quantum masking yields \( \|\text{Tr}_B(|\Omega_{x,y,\theta}\rangle \langle \Omega_{x,y,\theta}|) - M_1 M_1^\dagger\|_1 \leq 2\sqrt{2}\epsilon \). That is,
\[ \|M_2 M_2^\dagger - M_1 M_1^\dagger + e^{-i\theta} M_1 M_1^\dagger + e^{i\theta} M_2 M_2^\dagger\|_1 \leq 4\sqrt{2}\epsilon. \] (23)

Then by Ineq. (21) and the triangle inequality, one has \( \|e^{-i\theta} M_1 M_1^\dagger \|_1 \leq 6\sqrt{2}\epsilon \). As this holds for all \( \theta \in [0, 2\pi] \), particularly, we have
\[ \|M_1 M_1^\dagger + e^{-i\theta} M_2 M_2^\dagger\|_1 \leq 6\sqrt{2}\epsilon. \] (24)
Therefore, by using triangle inequality again, we deduce \( \|M_1M_2^\dagger M_2M_1^\dagger\|_1 \leq 6\sqrt{2}\epsilon \). Using the compatible condition of norm, we obtain an upper bound of the modulo of \( \text{Tr}(M_1M_2^\dagger M_2M_1^\dagger) \). That is,
\[
\|M_1M_2^\dagger M_2M_1^\dagger\|_1 \leq \|M_1M_2^\dagger\|_1 \|M_2M_1^\dagger\|_1 \leq 72\epsilon^2. \tag{25}
\]
Then we can take similar argument as the proof of Theorem 3 to complete the proof and therefore we omit it here.

\section{Conclusions}

In this paper, we have considered three generalisations of the no-masking theorem. First, by replacing the unitary operation with an invertible linear operation we obtain the impossibility of probabilistic universal masking. While giving up certainty does not help for masking, we find that giving up the precision does. We derive a necessary condition of the precision \( \epsilon \) to make the approximate masking possible. At last, we combine the previous process together and consider the \( \epsilon \)-approximate-probabilistic masker. These results can be seen as the robustness of the no-masking theorem. On the other hand, our results and methods open up new research questions, namely, finding optimal approximate maskers and testing the tightness of our fidelity bounds. The implications of our results the subsequent open questions have important applications quantum secret sharing, data hiding, and other quantum information protocols that require storing of information in composite quantum systems. Some important examples include the studies of interacting prover problems \cite{39}, quantum error correction methods, and out of time order correlators, which are important for the black hole physics and the black hole information paradox.

\section{Appendix A: Appendix}

\textbf{Lemma 1} Let \( |\alpha_1\rangle, |\alpha_2\rangle \in \mathcal{F}(\mathcal{H}_A) \) and \( 0 \leq \langle \alpha_1 | \alpha_2 \rangle < 1 \).

For any positive real numbers \( p_1, p_2 \in \mathbb{R}_+ \) and \( \theta \in [0, 2\pi] \), there exists \( x, y \in \mathbb{R} \) such that \( x|\alpha_1\rangle + ye^{i\theta}|\alpha_2\rangle \in \mathcal{F}(\mathcal{H}_A) \) and \( px = p_2y \).

\textbf{Proof.} Denote \( r := \langle \alpha_1 | \alpha_2 \rangle \). The vector \( x|\alpha_1\rangle + ye^{i\theta}|\alpha_2\rangle \in \mathcal{F}(\mathcal{H}_A) \) if and only if
\[
x^2 + 2r \cos \theta xy + y^2 = 1. \tag{A1}
\]

Fixed \( r \) and \( \theta \), the above function defines an elliptic curve whose center is \((0, 0)\). In fact, we can write it into the following normal form:
\[
\frac{(x + r \cos \theta y)^2}{1} + \frac{y^2}{(1 - r^2 \cos^2 \theta)^{-1}} = 1. \tag{A2}
\]

Obviously, the curve defined by Eq. (A1) has incidents with the line defined by
\[
p_1x = p_2y. \tag{A3}
\]
Equivalently, there is some common solution of the Eqs. (A1) and (A3) as shown in Fig. 1. This completes the proof. \( \blacksquare \)

\textbf{Acknowledgements.} We thank J. Fitzsimons for bringing our attention to this problem. M.-S. L. thanks Xiangjing Liu for helpful discussions. K.M. received travel support from the Australian Academy of Technology and Engineering via the 2018 Australia China Young Scientists Exchange Program and the 2019 Next Step Initiative. M.-S. L. would like also to thank the National Natural Science Foundation of China (Grants No. 11875160 and No. 11871295) for partial financial support.

\begin{thebibliography}{99}

[1] W. K. Wootters and W. H. Zurek. A Single quantum cannot be cloned. Nature (London) \textbf{299}, 802 (1982).

[2] D. Dieks. Communication by EPR devices. Phys. Lett. A \textbf{92}, 271 (1982).

[3] C. H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. Phys. Rev. Lett. \textbf{70}, 1895 (1993).

[4] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter and A. Zeilinger. Experimental quantum teleportation. Nature \textbf{390}, 575-579 (1997).

[5] C. H. Bennett. Quantum cryptography using any two nonorthogonal states. Phys. Rev. Lett. \textbf{68}, 3121 (1992).

[6] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden. Quantum cryptography. Rev. Mod. Phys. \textbf{74}, 145 (2002).

[7] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher. Noncommuting Mixed States Cannot Be Broadcast. Phys. Rev. Lett. \textbf{76}, 2818 (1996).

[8] M. Piani, P. Horodecki, and R. Horodecki. No-Local-Broadcasting Theorem for Multipartite Quantum Correlations. Phys. Rev. Lett. \textbf{100}, 090502 (2008).

[9] A. Kalt and I. Hen. No-Broadcasting Theorem and Its Classical Counterpart. Phys. Rev. Lett. \textbf{100}, 210502 (2008).

[10] H. K. Lo and H. F. Chau. Is quantum bit commitment really possible? Phys. Rev. Lett. \textbf{78}, 3410 (1997).

[11] D. Mayers. Unconditionally Secure Quantum Bit Commitment is Impossible. Phys. Rev. Lett. \textbf{78}, 3414 (1997).

[12] A. K. Pati and S. L. Braunstein. Possibility of deleting an unknown quantum state. Nature (London) \textbf{404}, 164 (2000).

[13] S. L. Braunstein and A. K. Pati. Quantum Information Cannot Be Completely Hidden in Correlations: Implications for the Black-Hole Information Paradox. Phys. Rev. Lett. \textbf{98}, 080502 (2007).

[14] J. R. Smal, A. K. Pati, and A. Kumar. Experimental Test of the Quantum No-Hiding Theorem. Phys. Rev. Lett. \textbf{106}, 080401 (2011).

[15] C. H. Bennett and G. Brassard, Quantum Cryptography: Public Key Distribution and Coin Tossing.
[16] A. K. Ekert. Quantum cryptography based on Bell's theorem. Phys. Rev. Lett. 67, 661 (1991).
[17] D. P. DiVincenzo, D. W. Leung, B. M. Terhal. Quantum Data Hiding. IEEE Trans. Inf Theory Vol. 48, No. 3, 580-599 (2002)
[18] M. Hillery, V. Bužek, and A. Berthiaume. Quantum secret sharing. Phys. Rev. A 59, 1829 (1999).
[19] R. Cleve, D. Gottesman, and H.-K. Lo. How to Share a Quantum Secret? Phys. Rev. Lett. 83, 648 (1999).
[20] C. E. Shannon, Bell Syst. Tech. J. 28, 656 (1949).
[21] K. Modi, A. K. Pati, A. Sen(De), and U. Sen. Masking quantum information is impossible. Phys. Rev. Lett. 120, 230501 (2018).
[22] M.-S. Li and Y.-L. Wang. Masking quantum information in multipartite scenario. Phys. Rev. A 98, 062306 (2018).
[23] X.-B. Liang, B. Li, and S.-M. Fei. Complete characterization of qubit masking. Phys. Rev. A 100 030304(2019).
[24] F. Ding, X. Hu. Masking Quantum Information on Hyperdisks. arXiv:1909.11256 (2019).
[25] S. H. Lie and H. Jeong. Randomness cost of masking quantum information and the information conservation law. arXiv:1903.12304v1 (2019).
[26] S. H. Lie and H. Jeong. Randomness cost of masking quantum information and the information conservation law. arXiv:1903.12304v1 (2019).
[27] T. Guha, M. Atimuddin, P. Parashar. Some no-go results in Quantum Thermodynamics. arXiv:1909.13503 (2019).