

How (not) to Build Quantum PKE in Minicrypt

Longcheng Li* Qian Li† Xingjian Li‡ Qipeng Liu§

Abstract

The seminal work by Impagliazzo and Rudich (STOC’89) demonstrated the impossibility of constructing classical public key encryption (PKE) from one-way functions (OWF) in a black-box manner. Quantum information has the potential to bypass classical limitations, enabling the realization of seemingly impossible tasks such as quantum money, copy protection for software, and commitment without one-way functions. However, the question remains: can quantum PKE (QPKE) be constructed from quantumly secure OWF?

A recent line of work has shown that it is indeed possible to build QPKE from OWF, but with one caveat. These constructions necessitate public keys being quantum and unclonable, diminishing the practicality of such “public” encryption schemes — public keys cannot be authenticated and reused. In this work, we re-examine the possibility of perfect complete QPKE in the quantum random oracle model (QROM), where OWF exists.

Our first main result: QPKE with classical public keys, secret keys and ciphertext, does not exist in the QROM, if the key generation only makes classical queries.

Therefore, a necessary condition for constructing such QPKE from OWF is to have the key generation classically “un-simulatable”. Previous results (Austrin et al. CRYPTO’22) on the impossibility of QPKE from OWF rely on a seemingly strong conjecture. Our work makes a significant step towards a complete and unconditional quantization of Impagliazzo and Rudich’s results.

Our second main result extends to QPKE with quantum public keys.

The second main result: QPKE with quantum public keys, classical secret keys and ciphertext, does not exist in the QROM, if the key generation only makes classical queries and the quantum public key is either pure or “efficiently clonable”.

The result is tight due to these existing QPKEs with quantum public keys, classical secret keys, quantum/classical ciphertext and classical-query key generation require the public key to be mixed instead of pure; or require quantum-query key generation, if the public key is pure. Our result further gives evidence on why existing QPKEs lose reusability.

We also explore other sufficient/necessary conditions to build QPKE from OWF. Along the way, we use a new argument based on conditional mutual information and Markov chain to reprove the classical result; leveraging the analog of quantum conditional mutual information and quantum Markov chain by Fawzi and Renner (Communications in Mathematical Physics), we extend it to the quantum case and prove all our results. We believe the techniques used in the work will find many other usefulness in separations in quantum cryptography/complexity.

*State Key Lab of Processors, Institute of Computing Technology, Chinese Academy of Sciences. Email: lilongcheng22s@ict.ac.cn

†Shenzhen International Center For Industrial And Applied Mathematics, Shenzhen Research Institute of Big Data. Email: liqian.ict@gmail.com

‡Tsinghua University. Email: lxj22@mails.tsinghua.edu.cn

§University of California San Diego. Email: qipengliu0@gmail.com

1 Introduction

Quantum information and computation has the remarkable capability to transform classical impossibility into reality, ranging from breaking classically secure cryptosystems (Shor’s algorithm [Sho99]), realizing classically impossible primitives (quantum money [Wie83], quantum copy-protection [Aar09, ALL⁺21]) to weakening assumptions (quantum key distribution [BB14], oblivious transfer/multi-party computation [BCKM21, GLSV21], commitment [AQY22, MY22b])¹.

In the seminal work by Impagliazzo and Rudich [IR89], they proved that one-way functions (OWFs) were insufficient to imply the existence of public-key encryption (PKE) in a black-box manner. Coined by Impagliazzo [Imp95], the word “Minicrypt” was referred to a world where only one-way functions exist; this word now broadly denotes all cryptographic primitives that are constructible from one-way functions. Thus, their result is now often interpreted as “classical PKE is not in Minicrypt”. Given the increasing instances of quantum making classical impossibility feasible, we explore the following question in this work:

Does quantum PKE (with classical plaintext) exist in Minicrypt?

Upon posing the question, ambiguity arises. A general quantum public-key encryption (QPKE) scheme allows everything to be quantum: its interaction with an OWF, both its public key and secret key, as well as ciphertext. Indeed, many efforts have already been made towards understanding different cases.

Classical Keys, Classical Ciphertext. In the work by Austrin, Chung, Chung, Fu, Lin and Mahmoody [ACC⁺22], they initialized the study on the impossibility of quantum key agreement (QKA) in the quantum random oracle model. QKA is a protocol that Alice and Bob can exchange classical messages in many rounds, quantumly query a random oracle, and eventually agree on a classical key. They show that, under a seemingly strong² assumption called “polynomial compatibility conjecture”, such QKA with perfect completeness does not exist. Since QPKE with both keys and ciphertext being classical implies a two-round QKA, their conditional impossibility extends to this type of QPKE as well.

The result provides evidence on the negative side: such QPKE does not exist under the polynomial compatibility conjecture. However, not only proving or refuting the conjecture is quite challenging, but also the conjecture (or some form of the conjecture) is necessary. To prove QKA is impossible, one needs to design an eavesdropper that observes the whole transcript, interacts with the random oracle, and guesses a classical key. In a general QKA, both Alice and Bob can make quantum queries to the random oracle; the eavesdropper in the attack by Austrin et al. only makes classical queries. It inherently requires a simulation of a quantum-query algorithm using only classical queries (at least for the QKA functionality). Although not directly comparable, such efficient simulation for decision problems [AA09] (Aaronson-Ambainis conjecture) is conjectured to exist but the question still remains open until now. Moreover, [ACC⁺22] showed that if the Aaronson-Ambainis conjecture is false, a classical-query eavesdropper is insufficient to break imperfect complete QKA protocols.

In light of these considerations, our focus in this work centers on the following question:

¹Here, we only cite works that initialized each area.

²It seems strong because it implies that the eavesdropper can attack quantum Alice and Bob with only a polynomial number of classical queries.

Q1. Can we separate QPKE with classical keys, classical ciphertext from Minicrypt, without any conjecture?

Quantum Public Key, Classical Secret Key and Quantum/Classical Ciphertext. The quantum landscape introduces a paradigm shift. It was first realized by Morimae and Yamakawa [MY22a] that some forms of QPKE with quantum public keys and quantum ciphertext might be constructed from OWFs (or even presumably weaker primitives). Subsequent QPKE schemes were later proposed by Coladangelo [Col23], Kitagawa et al. [KMNY23], Malavolta and Walter [MW23] and Barooti et al. [BGH⁺23].

On the surface, it seems to give a good answer: QPKE with quantum public keys exists in Minicrypt, showcasing a notable distinction between the quantum and classical world. However, is this demarcation as unequivocal as it seems? Indeed, all the aforementioned constructions share one limitation — they lose one of the most important properties inherent in all classical PKE — reusability. A classical PKE scheme allows a user who possesses a public key to encrypt any polynomial number of messages, by reusing the classical public key. In contrast, public keys in [MY22a, Col23, KMNY23, MW23, BGH⁺23] are essentially unclonable³ and not reusable. Although some may argue for the practicality of QPKE with an additional public interface for generating quantum public keys, it introduces additional complexities such as the authentication of quantum keys, leading to increased interactions and other potential challenges.

Thus, our second focus is the question:

Q2. Must reusability be sacrificed in constructing QPKE with quantum public keys from OWFs?

1.1 Our Main Results

We make progress towards these two questions. A table discussing and comparing all existing results and our results is provided on the next page (Table 1). Our first main result establishes an impossibility result on QPKE with classical keys and classical ciphertext in the QROM.

Theorem 1.1. *QPKE with classical keys and classical ciphertext does not exist in the QROM, if*

1. *It has perfect completeness.*
2. *The key generation algorithm only makes classical queries to the random oracle.*

Unlike the approaches used in [ACC⁺22], our result does not require any conjecture, as the eavesdropper in our attack makes quantum queries. Perfect completeness is a natural property shared by many PKE schemes⁴

Consequently, if QPKE could be constructed in the QROM, one must have the key generation procedure being classical-query “un-simulatable”; meaning the keys can not be computed by an efficient classical-query algorithm. If we believe the Aaronson-Ambainis conjecture [AA09] (all quantum-query-solvable decision problems can be simulated by classical-query), the theorem suggests that the key generation must be a sampling procedure. Thus, to build QPKE in the QROM, we need to find a search/sampling problem that is only tractable by quantum queries; one such example is the Yamakawa-Zhandry problem [YZ22].

³Although unclonability does not necessarily imply no-reusability, it is an evidence of no-reusability.

⁴Very recently, Mazor [Maz23] proposed the first perfectly complete “Merkle Puzzle”, which was not known for many years.

[ACC⁺22] ruled out perfect complete QPKE with classical keys and classical ciphertext in the QROM, if (i) the encryption only makes classical queries, or (ii) both the key generation and decryption makes classical queries, without using the polynomial compatibility conjecture. Our result immediately improves their result (ii). Additionally, both of their impossibility results apply to the case where the oracle access in encryption and decryption procedure is asymmetric. Our result further completes the picture by making both encryption and decryption symmetric (quantum queries).

Another difference in our work is that our eavesdropper makes quantum queries. The capability of leveraging quantum eavesdroppers and conditioning on quantum events in this work, immediately gives us two strengthened versions (Theorem 1.2 and Theorem 1.3) of Theorem 1.1, which works for quantum public keys and quantum ciphertext. These two extensions provide many interesting discussions on the feasibility and impossibility of building QPKE in Minicrypt. [BGVV23] also discussed the feasibility of QPKE with quantum ciphertext, with a weaker result and based on the polynomial compatibility conjecture. Thus, we believe our framework is versatile and has the potential to completely answer these questions. We elaborate on them now.

Our second main result extends the previous theorem to QPKE with quantum public keys and ciphertexts. This scheme is tight to all existing QPKE with quantum public keys and perfect completeness, see Table 2.

Theorem 1.2 (Subsuming Theorem 1.1). *QPKE with quantum public key, classical secret key, and classical or quantum ciphertext does not exist in the QROM, if*

1. *It has perfect completeness.*
2. *The key generation algorithm only makes classical queries to the random oracle.*
3. *The quantum public key is either pure or “efficiently clonable”.*

At first glance, the theorem may appear unexciting: how can a classical-query key generation produce meaningful quantum public keys? Interestingly, the QPKE constructions in [KMNY23, MW23] have only classical-query key generation procedures. Namely, their public keys are of the form $\frac{1}{\sqrt{2}}(|s_0\rangle + |s_1\rangle)$ of two strings s_0, s_1 , which can be computed using only classical queries.

Let us explain the third condition in our result. In QPKE, one needs to guarantee that having multiple copies of a quantum public key, the IND-CPA security still holds. If the public key generation procedure outputs a pure state, in other words the security holds against copies of the same pure state, we call the public key pure. Otherwise, each copy of a quantum public key is a mixed state (a distribution of pure states) and even holding multiple copies from the same distribution does not mean having the same pure states⁵. In the latter case, we require that there exists a query-efficient cloning procedure that can perfectly duplicate the pure state of the quantum public key.

As the schemes in [KMNY23, MW23] have perfect completeness and classical-query key generation, our result essentially says that their public keys must be mixed and can not be cloned. While the absence of perfect cloning does not imply non-reusability, our result provides crucial insights into the feasibility of building QPKE with quantum public keys and reusability from OWFs. Many intriguing open questions will be discussed further at the end of this section.

Our second theorem also is tight to the QPKE scheme by [Col23, BGH⁺23]. Their scheme has perfect completeness, pure state quantum public keys. Our theorem suggests that their schemes

⁵For example, even if having two copies of the same mixed state $\frac{1}{2}$, they can be two different states, like $|0\rangle|1\rangle$.

must make quantum queries; which is the case in their construction, as they need to query a pseudorandom function (implied by OWFs in a black-box manner) on an equal superposition.

Our next result is on the type of QPKE with classical keys and quantum ciphertext, whose decryption makes no queries to an oracle. The impossibility extends to any oracle model (even quantum oracles), not just the random oracle model.

Theorem 1.3. (*Imperfect/Perfect*) QPKE with classical keys and quantum ciphertext does not exist in any oracle model, if

1. The decryption algorithm makes no queries to the oracle.

Here, an oracle model means an oracle is sampled from a distribution, and the KA is executed under this oracle; both the key generation algorithm, the encryption algorithm and the attacker can have quantum access to the oracle. We remark that in our impossibility result, such QPKE does not need to be perfectly complete. We also note that, our theorem holds in the classical setting; PKE does not exist in any oracle model if the decryption algorithm makes no queries. This does not contradict constructions in the generic group model, as the decryption algorithm is required to make oracle queries.

A recent work by Bouaziz–Ermann, Grilo, Vergnaud and Vu [BGVV23] also discusses a similar separation. They proved that, under the polynomial compatibility conjecture, perfect complete QPKE with classical keys and quantum ciphertext does not exist in the QROM, if decryption makes no queries. Our Theorem 1.3 improves their results in three aspects: we remove the conjecture and the requirement on perfect completeness, and the impossibility works in any oracle model.

Other results. As we progress towards achieving our main results, our techniques also enable us to establish additional impossibility results on QPKE and QKA as well. These results include

- Separation between pseudorandom quantum states and QKA (Remark 5.3);
- Impossibility of Merkle-like QKA (more generally, non-interactive QKA) in *any oracle model* (Theorem 5.2); we emphasize that this result also holds for quantum oracles, e.g., when a black-box unitary is chosen from Haar measure.
- Impossibility of QPKE with a short classical secret key in *any oracle model* (Appendix C).

We refer interested readers to these sections for more details.

1.2 Open Questions and Discussions

Before the overview of our techniques in the next section, we discuss some open questions and directions. Some discussions may become clearer upon reviewing both the overview and the entirety of the paper.

Remove perfect completeness. Our Theorem 1.1 and Theorem 1.2 both rely on the underlying QPKE is perfectly complete. Can we remove this condition and make the impossibility result work for any QPKE?

	[ACC+22]	[ACC+22]	[ACC+22]	Thm 1.1	[BGVV23]	Thm 1.3
perfect complete	✓	✓	✓	✓	✓	
ciphertext	C	C	C	C	Q	Q
Gen	Q	Q	C	C	Q	Q
Enc	Q	C	Q	Q	Q	Q
Dec	Q	Q	C	Q	*	*
conjecture	✓				✓	
oracle	RO	RO	RO	RO	RO	any oracle

Table 1: Comparing impossibility results on classical public keys. ‘Q’ denotes quantum, ‘C’ denotes classical, ‘checkmark’ denotes yes, and ‘*’ denotes no oracle queries. “RO” stands for “(quantum) random oracle”.

	Thm 1.2 (Impossibility)	[KMNY23, MW23]	[Col23, BGH+23]
perfect complete	✓	✓	✓
ciphertext	pure or “efficiently clonable”	mixed and unclonable	pure
Gen	C	C	Q

Table 2: Comparing our Thm 1.2 with existing constructions. [KMNY23, MW23] overcame the impossibility by allowing the ciphertext to be mixed; [Col23, BGH+23] overcame the impossibility by allowing the Gen procedure to make quantum queries.

Extending clonable keys to reusable keys. When quantum public keys are reusable, that means there exists an encryption procedure Enc' that takes two messages m, m' and only one quantum public key pk , and outputs two valid ciphertexts. Since Enc' is not necessarily honest, it seems difficult to convert such assumptions into the setting of a QKA protocol. An intermediate goal is to have Enc “deterministic”, which means it takes pk and m , outputs the original pk together with a valid ct — this is a weaker but promising step towards a full understanding of reusability. Another direction is to relax the “clonable” condition to some form of “approximate clonable” in our impossibility result.

Make key generation quantum. Our [Theorem 1.1](#) and [Theorem 1.2](#) assume key generation procedure only makes classical queries. The current techniques require addressing some sorts of “heavy queries” made by an encryption procedure. For classical-query key generation, the number of such possible inputs is only polynomial. But it will be less clear how to handle this, when the key generation is fully quantum. Upon resolving this, one can get a full separation between QPKE and OWFs.

The basic idea of our approach is to keep the conditional mutual information (between Alice and Bob, conditioned on quantum Eve) small during the execution of QKA protocols or QPKE schemes. The characterization of quantum states with small CMI provided by Fawzi and Renner [FR15] (see [Theorem 3.11](#)) plays a major role. To achieve a full separation between QPKE and OWF, the argument has to critically utilize the condition that H is a uniform random function rather than drawn from a general oracle distribution. This is because our separation works in both the quantum and classical setting, but in the full classical KA case, KA exists in the generic group model. Thus, we really need to use the structures of a random oracle. In this work, we show it is possible when key generation only makes classical queries but not clear for quantum-query

key generation.

Acknowledgement

Qipeng Liu would like to thank Navid Alamati for bringing up this problem in 2019 while interning at Fujitsu Research of America, and Jan Czajkowski, Takashi Yamakawa and Mark Zhandry for insightful discussions at Princeton during 2019 and 2020. The authors also thank Luowen Qian for pointing out the problem of Merkle-like QKA with Haar-random unitary, which prompted the realization of Theorem 1.3 applicable to even quantum oracles.

Qian Li's work was supported by the National Natural Science Foundation of China Grants (No. 62002229), and Hetao Shenzhen-Hong Kong Science and Technology Innovation Cooperation Zone Project (No. HZQSW-S-KCCYB-2024016).

Longcheng Li was supported in part by the National Natural Science Foundation of China Grants No. 62325210, 62272441, 12204489, 62301531, and the Strategic Priority Research Program of Chinese Academy of Sciences Grant No. XDB28000000.

2 Technical Overview

In this section, we shed light on several key ideas used in the work, especially those on how to attack with quantum queries and how to deal with quantum public keys. Our method for handling quantum ciphertext is similar to that in [BGVV23] by leveraging the Gentle Measurement Lemma.

Translating QPKE to QKA. Although we focus on the separation between QPKE and OWFs (or the quantum random oracle model), we will mostly study two-round QKA protocols. Say, if we have a QPKE scheme in the QROM (Gen, Enc, Dec), we can easily convert it to a two-round QKA:

Definition 2.1 (Informal, Two-round QKA). *We define a two-round QKA as follows.*

- Alice runs Gen to produce pk, sk and sends pk as the first message m_1 to Bob. We call Alice's algorithm at this stage \mathcal{A}_1 .
- Bob upon receiving $m_1 := pk$, it samples a uniformly random key k . It computes $ct \leftarrow \text{Enc}(pk, k)$ and sends $m_2 := ct$ to Alice. We call Bob's algorithm at this stage \mathcal{B} .
- Alice will then run $\text{Dec}(sk, ct)$ to retrieve k . We call Alice's algorithm at this stage \mathcal{A}_2 .

If the QPKE has security against randomly chosen messages, then the underlying QKA is secure. Quantum public keys will make m_1 quantum in the corresponding QKA, quantum ciphertext translates to a quantum m_2 and classical-query Gen makes \mathcal{A}_1 classical-query. Thus, we focus on breaking various types of two-round QKA with perfect completeness.

2.1 Recasting the Classical Idea for Merkle-like KA

The classical proofs [IR89, BM09] shares one common idea: as long as an eavesdropper Eve learns all the queries that are both used in Alice and Bob's computation, Eve can learn the key⁶. Here we re-interpret the idea for a special case: two-round Merkle-like KA. In this type of KA, m_1, m_2 are

⁶[BKS11] also gave a proof for the case of perfect completeness. Their proof works even if not all intersection queries are learned.

generated based on oracle queries and sent simultaneously to the other party (so that m_2 has no dependence on m_1). Furthermore, to recover the shared key, Alice and Bob only need to do local computation on their internal states and the communication, *without making any oracle queries*. This is a strong form of KA, which is not implied by PKE using the aforementioned reduction. We refer to it as Merkle-like because the famous Merkle Puzzles are of this form.

Let us assume for a specific execution of Alice and Bob, the queries made by Alice is the list R_A , her private coins are s_A and similarly R_B, s_B, R_E is the query made by Eve; we define $\text{View}_A = (s_A, R_A)$ as the personal view of Alice, similarly View_B for Bob. A tuple $(\text{View}_A, \text{View}_B, m_1, m_2, H)$ is a possible execution of the KA right after Alice and Bob exchange their messages but have not started working on computing the key. The tuple specifies Alice’s random coins and queries, similarly for Bob, communication m_1, m_2 and the oracle H under which the protocol is executed. We say $(\text{View}_A, \text{View}_B)$ is consistent with a transcript (m_1, m_2) if there exists an oracle H , such that $(\text{View}_A, \text{View}_B, m_1, m_2, H)$ has *strictly positive* probability of appearing in some real execution. Similarly, View_A is consistent with (m_1, m_2) if $(\text{View}_A, m_1, m_2, H)$ has non-zero probability for some oracle H .

For a transcript (m_1, m_2) , it is always easy to find a pair of Alice’s fake $\text{View}'_A = (s'_A, T'_A)$ that is consistent with m_1, m_2 . Since we do not care about the actual computation cost, and only the number of queries matters, we can keep sampling oracles until we see a transcript (m_1, m_2) . Upon receiving View'_A , we hope that $(\text{View}'_A, \text{View}_B)$ has the same distribution as $(\text{View}_A, \text{View}_B)$; if that is true, based on perfect completeness, any non-zero support in $(\text{View}_A, \text{View}_B)$ should provide us with the agreed key.

Unfortunately, this does not hold true. The overlap in inputs between R_A and R_B leads to a correlation in the distribution $(\text{View}_A, \text{View}_B)$, making it impossible to sample independently. This correlation is the resource for Alice and Bob to compute an agreed key, akin to Merkle Puzzles.

The key insight from [IR89, BM09] is to eliminate this correlation, often referred to as “intersection queries”. In their attacks, Eve queries the oracle in a manner such that R_E encompasses all the shared knowledge between Alice and Bob. More precisely, for any execution R_A, R_B consistent with (m_1, m_2) , it holds that $R_A \cap R_B \subseteq R_E$. Consequently, conditioned on R_E and the transcript, the distribution $(\text{View}_A, \text{View}_B)$ becomes close to a product distribution. Eve can then sample View'_A conditioned on m_1, m_2, R_E , and this fake view will yield the correct key.

Finally, [IR89, BM09] demonstrate that as long as Eve queries Alice and Bob’s “heavy queries” (those with a relatively noticeable probability of being queried), with high probability R_E will contain all intersection queries.

2.2 A Quantization Attempt by Austrin et al.

[ACC⁺22] focuses on the fully general QKA, but here we explain their ideas for the Merkle-like protocols. The first challenge arises when attempting to formally describe an execution for both quantum Alice and Bob. As they can make quantum queries, neither the random coins s_A nor the query list R_A can be explicitly delineated. A quantum algorithm can possess randomness that is impossible to be purified as random coins, and can make quantum superposition queries. An execution right after Alice and Bob exchange messages, is represented by $(\text{View}_A, \text{View}_B, m_1, m_2, H)$ where $\text{View}_A := \rho_A, \text{View}_B := \rho_B$ are the internal quantum states of Alice and Bob.

To quantize the strategy of [IR89, BM09], one needs to define “heavy quantum queries” and “quantum intersection queries”, and establish some form of independence between Alice and Bob conditioned on transcripts and Eve’s knowledge. Austrin et al., leveraged the breakthrough

technique (“compressed oracle”) by Zhandry [Zha19], defined “heavy quantum queries”. On a very high level (without introducing Zhandry’s technique), “heavy quantum queries” are classical inputs that have high weights on the oracle, when the oracle is examined under the Fourier basis. Their proposed attack queries all “heavy quantum queries” classically. Under their polynomial compatibility conjecture, they can argue the success of their attacks.

This approach is less ideal in the following aspects. First, even for Merkle-like protocols, the polynomial compatibility conjecture seems necessary. Second, the classical-query Eve has to somehow “simulate” the ability of quantum Alice or quantum Bob. Although such simulation is believed to be true for decision problems [AA09], this kind of simulation in the QKA setting is both unclear to hold and potentially unnecessary. Finally, establishing independence between two quantum states (quantum Alice and Bob) is challenging to define and can be intricate.

2.3 Step Back — Classical and Quantum Proofs Using Markov Chain

Stepping back, let’s reconsider if there are other approaches that are more quantum-friendly: potentially can take advantage of quantum-query Eve. The key insight in [IR89, BM09] is, when conditioned on some query list R_E and (m_1, m_2) , the distribution $(\text{View}_A, \text{View}_B)$ is a product distribution, meaning Alice and Bob are independent. Based on this, Eve can therefore sample View'_A and they together with the real View_B have the same distribution as $(\text{View}_A, \text{View}_B)$.

Our first contribution is to give an alternative view of the classical proofs for Merkle-like KA. We realize that, in the classical proof, when intersection queries $R_A \cap R_B$ always is in R_E , the conditional mutual information (CMI) between R_A, R_B conditioned on E, m_1, m_2 is 0.

$$I(\text{View}_A : \text{View}_B | (m_1, m_2, R_E)) = 0.$$

There are two seemingly classically equivalent consequences when CMI is 0.

- **Perspective 1.** From Pinsker’s inequality, a CMI of 0 immediately implies that View_A and View_B are independent conditioned on (m_1, m_2, R_E) . This further implies that we can sample View'_A accordingly.
- **Perspective 2.** Another perspective is that, $\text{View}_B \rightarrow (m_1, m_2, R_E) \rightarrow \text{View}_A$ forms a Markov chain. Three random variables XYZ form a Markov chain if $p_{xyz} = p_{xy}p_{z|y}$. In our case, it says there exists a way to take (m_1, m_2, R_E) as inputs and sample View'_A such that $(\text{View}'_A, \text{View}_B)$ and $(\text{View}_A, \text{View}_B)$ are identically distributed.

The above discussion explains the attacks for Merkle-like protocols, through the lens of CMI.

From this, we propose one candidate quantum attack for Merkle-like QKA:

Quantum Eve: A Framework

- Eve makes some quantum queries and let $\text{View}_E := \rho_E$ be its internal quantum state. It “somehow” makes sure that the CMI between Alice and Bob, conditioned on Eve and the transcript is small enough.

$$I(\text{View}_A : \text{View}_B | m_1, m_2, \text{View}_E) < \epsilon.$$

- Use “some quantum analogy” of 1 or 2 above to produce a quantum state $\text{View}'_A := \rho_{A'}$ from ρ_E such that,

$$\text{View}'_A \text{View}_B \approx_{\text{poly}(\epsilon)} \text{View}_A \text{View}_B.$$

There are two questions remain to be answered:

- Does there exist a query-efficient Eve’s strategy that always makes CMI small?*
- Does there exist a quantum analogy of step 1 or 2?*

We answer both of the questions affirmatively.

Decreasing CMI. Assume Alice and Bob each makes at most d queries and $H : [2^n] \rightarrow \{0, 1\}$ be any oracle of domain $[2^n]$ and binary range. We show that,

Lemma 2.2 (Informal). *For any standard two-round QKA right after Alice receives m_2 , Eve can run the same Bob t times for some $t \in \{0, 1, \dots, 2dn/\epsilon\}$ and store all t copies of Bob’s internal states, such that the CMI between Alice and Bob conditioned on Eve’s register and m_1, m_2 is at most ϵ .*

A couple of things we clarify here. First, why don’t we set $t := 2dn/\epsilon$ (the largest value)? This is due to the nature of quantum conditional mutual information. Classically, when we condition on more classical queries, the CMI will never increase. However, quantum entropy and mutual information behave unlike their classical counterparts. Still, we are able to show the existence of such small t . The existence of such t will not make our Eve non-uniform, as t is (inefficiently) computable without making any oracle query.⁷

Second, this strategy works for any oracles. Third, the lemma does not distinguish between whether m_1 is classical or quantum. Even if Bob takes a quantum input, as long as we have access to t copies of the same quantum state, we can run Bob with the same pure state t times and thus the lemma still holds. This fact will be useful for the case of quantum public keys.

Finally, the lemma works for any two-round QKA⁸. As Alice and Bob are asymmetric for general protocols, we do not know how to only simulate Alice’s queries and make the CMI small in the general two-round QKA case. We will mention it after finishing the discussion on Merkle-like QKA.

Sampling Fake Alice. When the CMI is small, we need to sample a fake Alice. If the CMI is 0, Hayden et al. [HJPW04] showed an approach that can be viewed as a quantum analogy of Perspective 1. However, their approach only works for the case of CMI being exactly 0 and is not robust.

We realize the second interpretation works much better. The work by Fawzi and Renner showed that

Lemma 2.3 (Approximate Quantum Markov Chain, [FR15]). *Let X, Y, Z be three quantum registers, and ρ_{XYZ} be the state. If $I(X : Z|Y) < \epsilon$, then there exists a channel $\mathcal{T} : Y \rightarrow Y'Z'$ such that*

$$|\rho_{XZ} - \sigma_{XZ'}|_{\text{Tr}} \leq |\rho_{XYZ} - \sigma_{XY'Z'}|_{\text{Tr}} \leq O(\sqrt{\epsilon}),$$

⁷One can also guess a uniform t if we do not require finding the key with probability close to 1.

⁸It is a general lemma that works for any two-classical-message quantum interactive protocol. We focus on its application in QKA in this paper.

where $\sigma_{XY'Z'}$ is the state from applying \mathcal{T} on the Y register of ρ_{XY} . Furthermore, \mathcal{T} is explicitly (and inefficiently) constructible if knowing the state ρ_{XYZ} .

On a high level, [Lemma 2.3](#) states that if the CMI $I(X : Z|Y)$ for a tripartite state ρ_{XYZ} is small enough, then we can apply a local channel on Y to generate a state close to the original state in XY . This directly provides us a way to sample a fake Alice. In our case, X is the view of Bob, Z is the view of Alice and Y is the view of Eve and the transcript. Thus, the whole density matrix of ρ_{XYZ} is known by Eve⁹. By applying the lemma, Eve can sample $Z' := \text{View}'_A$ such that $\text{View}'_A \text{View}_B \approx_{O(\sqrt{\epsilon})} \text{View}_A \text{View}_B$; although \mathcal{T} in this case could be inefficient, it makes no queries. This completes the second step of our Eve.

Combining the approach of making the CMI arbitrarily small with the quantum Markov chain for sampling a fake Alice, our Eve can attack any Merkle-like QKA in any oracle model. We remark that for a general two-round QKA, [Lemma 2.2](#) and [Lemma 2.3](#) only cover the part up to receiving the last message m_2 . In the next subsection, we will discuss how to handle additional queries made after the last message.

2.4 Ruling out QPKE with Classical-Query Key Generation

The exact same idea from the previous section also applies to the standard two-round QKA as defined in [Definition 2.1](#), particularly when \mathcal{A}_2 (or the decryption algorithm) makes no queries to the oracle. In this scenario, we let View_A and View_B represent the views of Alice and Bob right after m_1 is received by Alice, and she has not yet begun working on producing the key. When we sample $\text{View}'_A \text{View}_B \approx \text{View}_A \text{View}_B$, since \mathcal{A}_2 only applies a local unitary that is independent of the oracle, Eve can perform the same on the fake view and still successfully obtain the key. However, what if \mathcal{A}_2 makes queries?

It is not immediately clear whether our CMI-based method works when \mathcal{A}_1 , \mathcal{A}_2 , and \mathcal{B} all make classical queries only, as in the case of classical KA instead of QKA. There are two attempts, that one might immediately come out.

- Attempt 1: run \mathcal{A}_2 on the fake View'_A using any oracle that is compatible with View'_A ;
- Attempt 2: run \mathcal{A}_2 on the fake View'_A using the real oracle.

Unfortunately, both approaches fail. Consider the following two classical examples for KA.

- Example 1: \mathcal{A}_1 does not query and does not send messages, \mathcal{B} sends a random $m_2 = x$ and both Alice and Bob agree on $H(x)$.
It is easy to see that, even if Eve does not query, $I(\text{View}_A : \text{View}_B | m_2) = 0$ as View_A is empty. Thus, any oracle H' is consistent with a fake View'_A but with overwhelming probability, $H'(x) \neq H(x)$.
- Example 2: \mathcal{A}_1 queries $H(0)$ and stores it as y , but does not send messages to \mathcal{B} ; \mathcal{B} sends a random $m_2 = x$ that is not equal to 0; \mathcal{A}_2 queries $H(0)$ again and aborts if $y \neq H(0)$, otherwise the key will be $H(x)$.
In an honest execution, Alice and Bob will always agree on a key $H(x)$ for some $x \neq 0$ as y is always equal to $H(0)$.

⁹ ρ_{XYZ} in this case, is not the state under a particular oracle, but the mixed state averaged over the distribution of all oracles. Only in this case, Eve knows ρ_{XYZ} .

We claim that $I(\text{View}_A : \text{View}_B | m_2) = 0$ as View_A and View_B have no intersection queries. Let View'_A be the fake view that consists of some y' . A real oracle with overwhelming probability has $H(x) \neq y'$, which fails the second attempt.

Our solution is to combine both Attempt 1 and Attempt 2.

Solution for Simulating \mathcal{A}_2 in the Classical Case. Our solution provides an alternative proof for the classical impossibility "PKE is not in Minicrypt" using our CMI-based framework and our solution for simulating \mathcal{A}_2 below.

When \mathcal{A}_1 , \mathcal{A}_2 , and \mathcal{B} are all classical-query algorithms, we propose the following method to execute a fake $\text{View}'_A = (s'_A, R'_A)$. Here, s'_A represents Alice's random coins, and R'_A represents the query list (a list of input-output pairs).

- Run \mathcal{A}_2 on the fake View'_A using the real oracle, except for every $x \in R'_A$, respond with the corresponding image $y \in R'_A$.

In other words, we adjust the real oracle such that it is consistent with R'_A ; let's denote this modified oracle as H' .

Why it works? Assume we sample View'_A such that $\text{View}'_A \text{View}_B$ has the same distribution as the real views for Alice and Bob. The fake view View'_A together with the real View_B must be reachable under some oracle (not necessarily the real oracle). Therefore, R'_A and R_B must be consistent. We also know that, since View_B is the real view of Bob under the real oracle H , R_B and H must also be consistent.

Thus, changing the oracle to be consistent with R'_A will only alter its behavior on those $x \notin R_B$. That means, under oracle H' , this KA will still have a strictly positive probability to end up with $\text{View}'_A, \text{View}_B, m_1, m_2$. By perfect completeness, when running \mathcal{A}_2 on View'_A with oracle H' , we must obtain the key held by Bob.

Extending to QPKE with Classical-Query Key Generation. For QKA with \mathcal{A}_1 making only classical queries (or QPKE with the key generation making classical queries), we can still run our Eve algorithm such that $I(\text{View}_A : \text{View}_B | (m_0, m_1, E)) < \epsilon$. However, since \mathcal{B} is now quantum, View_B is some quantum state ρ_B and its query list R_B is no longer defined.

We repeat our strategy again: sample View'_A and run \mathcal{A}_2 on View'_A with oracle H' as defined above (mostly the real oracle, but made consistent with R'_A). If Bob has low query weights on R'_A , then our attack still works. If the total query weight of Bob on R'_A is 0, changing the real oracle to H' will not change View_B at all. Similarly, if the weight is small, changing the real oracle to H' will only change View_B by a small amount [BBBV97]. Thus, we can still argue that $\text{View}'_A, \text{View}_B, m_1, m_2$ are reachable under H' and with perfect completeness, we must recover the key.

What if Bob has a large query weight on some $x \in R'_A$? We imagine a hypothetical Bob, who will first produce View_B , but then keep running itself from the beginning multiple times and randomly measure one of its queries. By doing so, the hypothetical Bob's functionality does not change; the advantage is now the hypothetical Bob has a *classical list* L_B that consists of all its queries with high weights, with a high probability.

If we run our Eve with this hypothetical Bob, View'_A must have R'_A consistent with L_B . Thus, changing the oracle to be consistent with R'_A will only change its behavior on those $x \notin L_B$, or in

other words, those x that do not have a large query weight! Then we can use the argument in the previous paragraphs and claim that $\text{View}'_A, \text{View}_B := (\rho_B, L_B), m_1, m_2$ are reachable under H' and with perfect completeness, we can get the key.

Handling Quantum Public Keys Quantum public keys (or quantum m_1) are handled without additional efforts using our CMI-based Eve. This further demonstrates the versatility and power of our new framework; previous approaches based on classical-query Eve do not extend to this case.

When m_1 is quantum, the challenge arises in how Eve can effectively run multiple copies of Bob on quantum input m_1 . In the context of QPKE, the attacker needs access to multiple copies of the quantum m_1 . If m_1 is a pure state, it becomes evident that running the same Bob on the same pure state m_1 is feasible, allowing us to minimize the CMI accordingly. Alternatively, if a query-efficient perfect cloner for m_1 exists, multiple runs of the same Bob become possible.

Given that the subsequent analysis relies solely on the capability to run Bob multiple times on input m_1 , we can extend our conclusion to rule out perfect complete QPKE with quantum public keys and classical-query key generation under the condition that the quantum public keys are either pure or query-efficiently clonable.

3 Preliminaries

We refer reader to [NC10] for more details about quantum computing and quantum information. Below, we mention some backgrounds that are heavily used in this work.

3.1 Distance measures

Let us recall the definition of total variation distance and trace distance.

Definition 3.1 (Total variation distance). *For two probabilistic distributions D_X, D_Y over the same finite domain \mathcal{X} , we define its total variation distance as*

$$TV(D_X, D_Y) = \frac{1}{2} \sum_{x \in \mathcal{X}} |D_X(x) - D_Y(x)|.$$

Definition 3.2 (Trace distance). *For two quantum states ρ, σ , the trace distance between the two states is*

$$TD(\rho, \sigma) = \frac{1}{2} \text{Tr} \left[\sqrt{(\rho - \sigma)^\dagger (\rho - \sigma)} \right] = \sup_{0 \leq \Lambda \leq I} \text{Tr}[\Lambda(\rho - \sigma)].$$

3.2 Quantum Oracle Model and Random Oracle

A quantum oracle algorithm equipped with access to $H: [2^{n\lambda}] \rightarrow [2^{m\lambda}]$ is expressed as a series of unitaries: $U_1, U_H, U_2, U_H, \dots, U_T, U_H, U_{T+1}$. Here, U_i denotes a local unitary acting on the algorithm's internal register. Oracle access to H is defined by a unitary transformation U_H , where $|x, y\rangle$ is transformed to $|x, y + H(x)\rangle$. For an oracle algorithm \mathcal{A} , we would use \mathcal{A}^H to denote the algorithm \mathcal{A} has classical access to the oracle H , and $\mathcal{A}^{|H\rangle}$ to denote \mathcal{A} has quantum access to the oracle.

We would also consider the case when the oracle $H : [2^{n\lambda}] \rightarrow \{0, 1\}$ is sampled from some distribution of oracles \mathcal{H}_λ . We would call some primitive in the quantum random oracle model (QROM) if the distribution \mathcal{H}_λ is uniformly random over all possible oracles.

3.3 Entropy

Definition 3.3 (Von Neumann Entropy). Let $\rho \in \mathbb{C}^{2^n}$ be a quantum state describing a system A. Let $|\phi_1\rangle, |\phi_2\rangle, \dots, |\phi_{2^n}\rangle$ be the eigenbasis of ρ ; ρ is written in this eigenbasis as $\sum_i \eta_i |\phi_i\rangle \langle \phi_i|$. Then its Von Neumann Entropy is denoted by $S(\rho)$ (or $S(A)_\rho$),

$$S(A)_\rho = S(\rho) = - \sum_i \eta_i \log(\eta_i).$$

Given a composite quantum system AB having joint state ρ_{AB} , we define the conditional Von Neumann Entropy as $S(A|B)_\rho$,

$$S(A|B)_\rho = S(AB)_\rho - S(B)_\rho.$$

Below, we often omit ρ in the definition when the quantum state is clear in the context. For example, $S(A)$ and $I(A : B)$ instead of $S(A)_\rho$ and $I(A : B)_\rho$.

Fact 3.4 ([NC10]). Suppose p_k are probabilities, $|k\rangle$ are orthogonal basis of a system A and ρ_k are quantum states for another system B. Then

$$S\left(\sum_k p_k |k\rangle \langle k| \otimes \rho_k\right) = H(p_k) + \sum_k p_k S(\rho_k)$$

where $H(p_k)$ is the Shannon entropy of distribution p_k .

Fact 3.5 ([Wil11]). If ρ_{AB} is a separable state, then $S(A|B) \geq 0$.

Definition 3.6 (Mutual Information). Let ρ be a quantum state describing two joint systems A and B. Then the mutual information between the system A and B is denoted by $I(A : B)$,

$$I(A : B) = S(A) + S(B) - S(AB).$$

Definition 3.7 (Conditional Mutual Information). Let ρ be a quantum state describing three joint systems A, B and C. Then the conditional mutual information $I(A : B|C)$,

$$I(A : B|C) = S(AC) + S(BC) - S(ABC) - S(C).$$

Fact 3.8 (Chain rule). $I(A_1, A_2, \dots, A_t : B | C) = \sum_{i=1}^t I(A_i : B | C, A_1, \dots, A_{i-1})$.

Fact 3.9. Let ABC be a composite quantum system. When a unitary is applied on A, it will not change $S(A)$. Similarly, the local unitary on A will not change $I(A : B)$ or $I(A : B|C)$.

Proof. This directly follows from the definition of $S(A)$ and applying any unitary will not change the spectrum of a density matrix. \square

The strong subadditivity for (conditional) mutual information concludes that both mutual information and conditional mutual information are always non-negative.

Lemma 3.10 (Strong Subadditivity, [AL70]). *Given Hilbert spaces A, B, C ,*

$$S(AC) + S(AB) \geq S(ABC) + S(C).$$

(The conditional form) Given Hilbert spaces A, B, C, D ,

$$S(AC|D) + S(AB|D) \geq S(ABC|D) + S(C|D).$$

3.4 Operational Meaning of Conditional Mutual Information: Approximate Quantum Markov Chain

Fawzi and Renner [FR15] provided a nice characterization of quantum states for which the conditional mutual information is approximately zero. Intuitively, if $I(A : B | E)$ is small, then B can be approximately reconstructed from E .

Theorem 3.11 ([FR15], restate of Lemma 2.3). *For any state ρ_{AEB} over systems AEB , there exists a channel $\mathcal{T} : E \rightarrow E \otimes B'$ such that the trace distance between the reconstructed state $\sigma_{A'E'B'} = \mathcal{T}(\rho_{AE})$ and the original state ρ_{AEB} is at most*

$$\sqrt{\ln 2 \cdot I(A : B|E)_\rho}.$$

3.5 Quantum Key Agreement and Quantum Public Key Encryption

In the following, we provide formal definitions of quantum key agreement (QKA) and quantum public key encryption (QPKE) in the oracle model.

Definition 3.12 (Quantum Key Agreement in the Oracle Model). *Let $\lambda \in \mathbb{Z}_+$ be the security parameter and \mathcal{H}_λ be a distribution of oracles. Let $H \leftarrow \mathcal{H}_\lambda$ be a classical oracle, drawn according to \mathcal{H}_λ .*

A key agreement protocol consists of two parties Alice and Bob, who start with all-zero states and have the ability to apply any quantum operator, get quantum access to H , and send classical messages to each other.

Both Alice and Bob can make at most $\text{poly}(\lambda)$ number of quantum queries to H . Finally, Alice and Bob output classical strings k_A, k_B ¹⁰. We would call the set of classical messages between Alice and Bob, denoted by Π , the transcript of the key exchange protocol.

A key agreement protocol should satisfy both correctness and security.

Definition 3.13 (Correctness). *Let k_A, k_B be the keys outputted in the protocol. Then $\Pr[k_A = k_B] \geq 1/q(\lambda)$ for some polynomial $q(\cdot)$, where the probability is taken over the randomness of Alice and Bob's channels, and the random choice of oracle H .*

Definition 3.14 (Security). *For any eavesdropper Eve that makes at most $\text{poly}(\lambda)$ number of quantum queries to $H \leftarrow \mathcal{H}_\lambda$, eavesdrops classical communication between Alice and Bob and outputs k_E , the probability that $\Pr[k_A = k_E]$ is negligible in λ .*

¹⁰ k_A, k_B can be of any length (even exponential in λ). Our impossibility results apply to protocols with any output key length.

We are interested in whether there exists a protocol that satisfies both correctness and security in the QROM. It is worth noting that the ability of making quantum queries is essential in our setting; when Alice and Bob can only make polynomially many classical queries, secure key agreement does not exist; i.e., there always exists an eavesdropper making polynomially many classical queries and breaking it [IR89, BM09].

We also focus on breaking quantum public key encryption schemes in the oracle model.

Definition 3.15 (Quantum Public Key Encryption in the Oracle Model). *Let $\lambda \in \mathbb{Z}_+$ be the security parameter and \mathcal{H}_λ be a distribution over oracles $H: [2^{n_\lambda}] \rightarrow \{0, 1\}$. Let $H \leftarrow \mathcal{H}_\lambda$ be a classical oracle, a QPKE scheme in the oracle model consists of three algorithms (Gen, Enc, Dec), each of which is allowed to make at most $d(\lambda) = \text{poly}(\lambda)$ quantum queries to H :*

- $\text{Gen}^{(H)}(1^\lambda) \rightarrow (\text{pk}, \text{sk})$: The quantum key generation algorithm that generates a pair of classical public key pk and secret key sk .
- $\text{Enc}^{(H)}(\text{pk}, m) \rightarrow \text{ct}$: the quantum encryption algorithm that takes a public key pk , the plaintext m , produces the ciphertext ct .
- $\text{Dec}^{(H)}(\text{sk}, \text{ct}) \rightarrow m'$: the quantum decryption algorithm that takes secret key sk and ciphertext ct and outputs the plaintext m' .

The algorithms should satisfy the following requirements:

Completeness $\Pr \left[\text{Dec}^{(H)}(\text{sk}, \text{Enc}^{(H)}(\text{pk}, m)) = m : \text{Gen}^{(H)}(1^\lambda) \rightarrow (\text{pk}, \text{sk}) \right] \geq 1 - \text{negl}(\lambda)$.

IND-CPA Security For any adversary $\mathcal{E}^{(H)}$ that makes $\text{poly}(\lambda)$ queries, for every two plaintexts $m_0 \neq m_1$ chosen by $\mathcal{E}^{(H)}(\text{pk})$, we have

$$\Pr \left[\mathcal{E}^{(H)}(\text{pk}, \text{Enc}^{(H)}(\text{pk}, m_b)) = b \right] \leq \frac{1}{2} + \text{negl}(\lambda).$$

Especially in Section 6, we will focus on when the key generation algorithm Gen^H is an algorithm with classical access to the oracle H .

It is a folklore result that we can construct a two-round key agreement protocol from a public key encryption scheme as follows:

1. Alice runs $\text{Gen}^{(H)}$ to produce a public key pk and a secret sk , and sends the public key pk as the first message m_1 to Bob. We call Alice's algorithm at this stage \mathcal{A}_1 .
2. Bob upon receiving pk , it samples a uniformly random classical string k as the key. It computes the ciphertext $\text{ct} \leftarrow \text{Enc}^{(H)}(\text{pk}, k)$ and sends $m_2 := \text{ct}$ to Alice. We call Bob's algorithm at this stage \mathcal{B} .
3. Alice will then run $\text{Dec}^{(H)}(\text{sk}, \text{ct})$ to output k' , which is her guess of k . We call Alice's algorithm at this stage \mathcal{A}_2 .

We further notice that the only information required by \mathcal{A}_2 from \mathcal{A}_1 is the secret key sk . Thus if sk is classical, we can assume without loss of generality that the internal state of \mathcal{A}_1 at its termination is a mixed state in the computational basis.

We would also consider the recently proposed QPKE with quantum public key [BGH⁺23, KMNY23, Col23] in Section 6. We would only focus on the variant where the protocol is perfect complete, and key generation algorithms can only make classical queries.

Definition 3.16 (QPKE with quantum public key). Let $\lambda \in \mathbb{Z}_+$ be the security parameter and \mathcal{H}_λ be a distribution over oracles $H: [2^{n_\lambda}] \rightarrow \{0, 1\}$. Let $H \leftarrow \mathcal{H}_\lambda$ be a classical oracle, a QPKE scheme in the oracle model consists of four algorithms (SKGen, PKGen, Enc, Dec), each of which is allowed to make at most $d(\lambda) = \text{poly}(\lambda)$ queries to H :

- $\text{SKGen}^H(1^\lambda) \rightarrow \text{sk}$: The secret key generation algorithm that generates a classical secret key sk .
- $\text{PKGen}^H(\text{sk}) \rightarrow \rho_{\text{pk}}$: The public key generation algorithm that takes the secret key sk and generates a quantum public key ρ_{pk} .
- $\text{Enc}^H(\rho_{\text{pk}}, m) \rightarrow \text{ct}$: the quantum encryption algorithm that takes a public key pk , the plaintext m , produces the (possibly quantum) ciphertext ρ_{ct} .
- $\text{Dec}^H(\text{sk}, \rho_{\text{ct}}) \rightarrow m'$: the quantum decryption algorithm that takes secret key sk and ciphertext ct and outputs the plaintext m' .

The algorithms should satisfy the following requirements:

Perfect Completeness

$$\Pr \left[\text{Dec}^H(\text{sk}, \text{Enc}^H(\rho_{\text{pk}}, m)) = m : \text{SKGen}^H(1^\lambda) \rightarrow \text{sk}, \text{PKGen}^H(\text{sk}) \rightarrow \rho_{\text{pk}} \right] = 1.$$

IND-CPA Security For any adversary \mathcal{E}^H that makes $\text{poly}(\lambda)$ queries, given any polynomial copies of public key $\rho_{\text{pk}}^{\otimes t(\lambda)}$, for every two plaintexts $m_0 \neq m_1$ chosen by $\mathcal{E}(\rho_{\text{pk}}^{\otimes t(\lambda)})$, we have

$$\Pr \left[\mathcal{E}^H(\rho_{\text{pk}}^{\otimes t(\lambda)}, \text{Enc}^H(\rho_{\text{pk}}, m_b)) = b \right] \leq \frac{1}{2} + \text{negl}(\lambda).$$

We call the public key pure if given sk , $\rho_{\text{pk}} \leftarrow \text{PKGen}(\text{sk})$ is a pure state, and call the public key clonable if there is some polynomial query algorithm \mathcal{D}^H that takes $\rho_{\text{pk}} = \sum p_i |\psi_i\rangle \langle \psi_i|$ in its eigenvector decomposition, and generates the state $\rho' = \sum p_i |\psi_i\rangle \langle \psi_i|^{\otimes t(\lambda)}$ for some polynomial $t(\cdot)$.

Remark 3.17. The definition of clonable here might seem odd at first glance, we would give some examples here for further explanations. The public key generation algorithm from [KMNY23, MW23] $\text{PKGen}(\text{sk})$ generates a $\rho_{\text{pk}} = (\text{pk}_r, |\psi_r\rangle)$ according to some private coin r of $\text{PKGen}(\text{sk})$, where pk_r is a classical string, $|\psi_r\rangle$ is a pure quantum state. Thus our cloning algorithm \mathcal{D} can be seen as given a sample $(\text{pk}_r, |\psi_r\rangle)$, it can generate multiple copies of state $|\psi_r\rangle$.

4 Helper Lemmas

In this section, we introduce three helper lemmas. [Lemma 4.2](#) tells how to decrease CMI. [Lemma 4.3](#) and [Lemma 4.4](#) claim that classical communication does not increase CMI. [Lemma 4.6](#) will be used in [Section 6](#).

4.1 Repetition Decreases CMI

Definition 4.1 (Permutation Invariance). Let $A_1, A_2, A_3, \dots, A_t, B$ be $(t+1)$ -partite quantum system. Given the joint state $\rho_{BA_1A_2 \dots A_t}$, we say A_1, \dots, A_t are permutation invariant, if for any permutation π on $[t]$, we have

$$\rho_{BA_1A_2 \dots A_t} = \rho_{BA_{\pi(1)}A_{\pi(2)} \dots A_{\pi(t)}}.$$

Lemma 4.2. *Let $A_1, A_2, A_3, \dots, A_t, B, C$ be $(t + 2)$ -partite quantum system. Suppose the state of the composite system $\rho_{BCA_1A_2\dots A_t}$ is fully separable. If $A_1, A_2, A_3, \dots, A_t$ are permutation invariant, then there is a $0 \leq i \leq t - 1$ such that*

$$I(A_t : B \mid C, A_1, \dots, A_i)_\rho \leq S(B)/t.$$

Proof. By the chain rule of conditional mutual information (see Fact 3.8), we have

$$\sum_{i=1}^t I(A_i : B \mid C, A_1, \dots, A_{i-1}) = I(A_1, \dots, A_t : B \mid C) \quad (1)$$

Besides,

$$I(A_1, \dots, A_t : B \mid C) = S(B \mid C) - S(B \mid C, A_1, \dots, A_t) \leq S(B \mid C) \leq S(B), \quad (2)$$

where the inequalities are by Fact 3.5 and that $I(B : C) = S(B) - S(B \mid C)$ is always non-negative. By (1) and (2), there must exist $i \in [t]$ such that $I(A_i : B \mid C, A_1, \dots, A_{i-1}) \leq S(B)/t$. Finally, by the permutation invariance, we have $I(A_i : B \mid C, A_1, \dots, A_{i-1}) = I(A_t : B \mid C, A_1, \dots, A_{i-1})$. Now we finish the proof. \square

4.2 Classical Communication does not Increase CMI

Lemma 4.3 (Local quantum operation does not increase CMI). *Let ABC be a composite quantum system. After performing a quantum operation M on A , the state of the system becomes $A'B'C'$. Then $I(A' : B'|C') \leq I(A : B|C)$.*

Proof. We introduce another quantum system D , initialized as zero. The quantum operation M can be treated as first performing a unitary transformation U on AD and then discarding D .

$$\begin{aligned} I(A : B|C) &= I(AD : B|C) \\ &= I(U(AD)U^\dagger : B|C) \\ &= I(A'D' : B'|C') \\ &= S(A'D'|C') - S(A'D'B'|C') + S(B'|C'). \end{aligned}$$

By the conditional form of strong sub-additivity (see the condition form of Lemma 3.10), $S(A'D'|C') - S(A'D'B'|C') \geq S(A'|C') - S(A'B'|C')$. Then

$$\begin{aligned} I(A : B|C) &\geq S(A'|C') - S(A'B'|C') + S(B'|C') \\ &= I(A' : B'|C'). \end{aligned} \quad \square$$

Lemma 4.4 (Sending classical message does not increase CMI). *Let ABC be a composite quantum system and $A = (W_A, M_A)$ where W_A is the working register and M_A is the message register containing a classical state. After B and C both obtain a copy of M_A , the system becomes $A'B'C'$. Then $I(A' : B'|C') \leq I(A : B|C)$.*

The following claim will be used.

Claim 4.5 (Copying classical state does not change entropy). *Given a system PQR where Q and R contain identical classical states. Then $S(PQR) = S(PR) = S(PQ)$.*

Proof. The joint state of PQR can be written as

$$\sum_k p_k \rho_P^{(k)} \otimes |k\rangle \langle k|_Q \otimes |k\rangle \langle k|_R$$

where p_k are probabilities, $\rho_P^{(k)}$ are quantum states in P, and $|k\rangle$ are computational basis of Q (also R). Then by [Fact 3.4](#),

$$\begin{aligned} S(\text{PQR}) &= S\left(\sum_k p_k \rho_P^{(k)} \otimes |k\rangle \langle k|_Q \otimes |k\rangle \langle k|_R\right) \\ &= H(p_k) + \sum_k p_k S\left(\rho_P^{(k)}\right) = S\left(\sum_k p_k \rho_P^{(k)} \otimes |k\rangle \langle k|_Q\right) = S(\text{PQ}). \end{aligned}$$

By symmetry, $S(\text{PQR}) = S(\text{PR}) = S(\text{PQ})$. □

Proof of [Lemma 4.4](#). Let $B' = (B, M_B)$, $C' = (C, M_C)$ where M_B, M_C are the message registers which contain a copy of M_A . Then

$$\begin{aligned} I(A' : B'|C') &= I(W_A, M_A : B, M_B|C, M_C) \\ &= S(W_A, C, M_A, M_C) + S(B, C, M_B, M_C) - S(W_A, B, C, M_A, M_B, M_C) - S(C, M_C). \end{aligned}$$

Since M_A, M_B, M_C contain identical classical states, by [Claim 4.5](#), we can remove M_A, M_B from above equation while keeping each term unchanged. Then

$$\begin{aligned} I(A' : B'|C') &= S(W_A, C, M_C) + S(B, C, M_C) - S(W_A, B, C, M_C) - S(C, M_C) \\ &= I(W_A : B|C, M_C) \\ &= I(W_A, M_C : B|C) - I(M_C : B|C). \end{aligned}$$

By the non-negativity of $I(M_C : B|C)$,

$$\begin{aligned} I(A' : B'|C') &\leq I(W_A, M_C : B|C) \\ &= S(W_A, C, M_C) + S(B, C) - S(W_A, B, C, M_C) - S(C). \end{aligned}$$

Since M_A and M_C are identical classical states, M_C can be replaced with M_A by [Claim 4.5](#). Then

$$\begin{aligned} I(A' : B'|C') &\leq S(W_A, C, M_A) + S(B, C) - S(W_A, B, C, M_A) - S(C) \\ &= I(W_A, M_A : B|C) = I(A : B|C). \end{aligned} \quad \square$$

4.3 Other Useful Lemmas

We would also need the following lemma.

Lemma 4.6. *For two classical probabilistic distributions D_X and D_Y over the same domain, if $TV(D_X, D_Y) \leq \epsilon$, we have that*

$$\Pr_{x \leftarrow D_X} [x \notin \text{SUPP}(D_Y)] \leq 2\epsilon.$$

Proof. We use p_x^X, p_x^Y to denote the probability of x drawn from D_X, D_Y respectively.

$$\sum_{x \notin \text{SUPP}(D_Y)} p_x^X \leq \sum_x |p_x^X - p_x^Y| = 2TV(D_X, D_Y) \leq 2\epsilon. \quad \square$$

5 Non-Interactive Quantum Key Agreement

We say a QKA is *non-interactive* if all queries are made before communication. Formally,

Definition 5.1 (Non-interactive Quantum Key Agreement). *A non-interactive key agreement protocol between Alice and Bob consists of the following steps:*

1. Let $\lambda > 0$ be a security parameter and $H \leftarrow \mathcal{H}_\lambda$.
2. Alice and Bob each makes $d_\lambda = \text{poly}(\lambda)$ queries to H .
3. Alice and Bob continue an arbitrary number of rounds of classical communication and local quantum operations, but will never make queries to H .
4. Eventually, Alice and Bob will output k_A, k_B .

Theorem 5.2. *Non-interactive QKA does not exist in any oracle model.*

Proof. Let e_λ denote the number of qubits on which the quantum query unitary U_H acts, i.e., the total length of the input register and the output register. For example, if $H : [2^{n_\lambda}] \rightarrow \{0, 1\}$ is a random function, then $e_\lambda = n_\lambda + 1$. Let ρ_A^0 and ρ_B^0 denote the states of Alice and Bob respectively right after the query algorithm but before any communication.

First, Eve repeatedly runs the same Alice's query algorithm $t = Cd_\lambda e_\lambda$ times (C is some polynomial determined afterwards) and prepares the state ρ_E^0 , which consists of t registers A_1^0, \dots, A_t^0 of ρ_A^0 's copies. Observing that A^0, A_1^0, \dots, A_t^0 are permutation invariant, by [Lemma 4.2](#), we have

$$I(A^0 : B^0 | E^0) \leq \frac{S(B^0)}{t+1} \leq \frac{2d_\lambda e_\lambda}{t+1} \leq \frac{1}{2C}.$$

The second inequality is because we can implement the query unitary U_H by a quantum communication process: suppose there are two parties, namely Bob and Oracle; if Bob wants to apply the unitary U_H , then

1. Bob sends its input register and output register, e_λ qubits in total, to Oracle;
2. Oracle applies U_H on these e_λ qubits and then sends them back to Bob.

By subadditivity of entropy, the entropy of Bob can increase by at most $2e_\lambda$ bits through the above quantum communication process. Since B^0 is prepared from a pure state by making d quantum queries, we have that $S(B^0) \leq 2d_\lambda e_\lambda$.

Let ρ_A^f and ρ_B^f denote the states of Alice and Bob respectively right after finishing the communication but before outputting the key. Since classical communication does not increase CMI (see [Lemma 4.4](#)), we have

$$I(A : B | E, \Pi)_{\rho_{ABE}^f} \leq I(A : B | E)_{\rho_{ABE}^0} \leq \frac{1}{2C}.$$

Finally, Eve applies the channel in [Lemma 2.3](#) and obtains a fake view \hat{A}^f of A^f such that the joint state of ρ_{AB}^f is $O(1/\sqrt{C})$ -close to ρ_{AB}^f . By letting $C = O(1/\epsilon^2)$, Eve can use \hat{A}^f to generate k_E such that $\Pr[k_B = k_E]$ is ϵ -close to $\Pr[k_A = k_E]$. \square

Remark 5.3. (a). *Theorem 5.2 directly implies the separation of pseudorandom quantum states (PRS) and QKA. This is because PRS can be prepared by random queries, and without loss of generality, we can assume all random queries are made at the very beginning before any communication.*

(b). In the proof of [Theorem 5.2](#), Eve makes $O(d_\lambda^2 e_\lambda)$ quantum queries. This bound can be improved to $\tilde{O}(d_\lambda^2)$ if Alice and Bob are restricted to making non-adaptive queries (i.e., queries can depend on neither any previous queries nor classical communication) to a random function. The famous Merkle Puzzles and PRS-based protocols are both examples of this non-adaptive case. We refer interested readers to [Appendix B](#) for details.

(c). By almost the same proof, [Theorem 5.2](#) can be extended to include quantum oracles, e.g., Haar random oracles, where the query unitary operator is chosen from some Haar measure.

We can further notice that the non-interactive requirement can be relaxed to one-sided if we are considering two-round key agreement protocols. That is, Bob can make queries after receiving the first message from Alice, while Alice can only make queries before communication. Viewing the protocol in the model of QPKE, we have the following theorem:

Theorem 5.4 (Restate of [Theorem 1.3](#)). *For any QPKE scheme with a classical public key and a classical/quantum ciphertext in any oracle model, if it satisfies the following conditions:*

1. *The key generation algorithm makes at most d quantum queries;*
2. *The encryption algorithm makes at most d quantum queries;*
3. *The decryption algorithm makes no queries.*

There exists an adversary Eve that could break the public key encryption scheme with probability $1 - O(\epsilon)$ by making $O(d^2 e / \epsilon^2)$ number of oracle queries.

Proof. In this case, if Eve generates $O(de/\epsilon^2)$ copies of message m_1 and runs Bob $\mathcal{B}(m_1)$ according to the first message m_1 , we can see that the permutation invariant condition still holds. Consider the stage before Bob sends the second message, similar to [Theorem 5.2](#), we have that

$$I(A : B \mid E) \leq \frac{1}{2C} \leq O(\epsilon^2).$$

Thus by applying the channel in [Theorem 3.11](#) to E , we can obtain a state $\rho_{A'B}$ that is $O(\epsilon)$ close to the state ρ_{AB} . Now Eve receives the message register from Bob and runs the decryption algorithm on A' , by the completeness of the protocol we can obtain that

$$\Pr[k_B = k_E] \geq \Pr[k_A = k_B] - O(\epsilon) \geq 1 - O(\epsilon) - \text{negl}(\lambda).$$

□

Note that our impossibility result can be extended to the Classical Communication One Quantum Message Key Agreement (CC1QM-KA) model from [\[BGVV23\]](#). In their security definition, the quantum message channel is unauthenticated, thus it can be modified by the adversary \mathcal{E} . As long as the completeness of the original key agreement protocol is $1 - \text{negl}(\lambda)$, we can apply the gentle measurement lemma to ensure that $\Pr[k_A = k_B = k_E] \geq 1 - O(\sqrt{\epsilon})$.

6 Public Key Encryption with a Classical Key Generation

In this section, we devote ourselves to proving black-box separation results for public key encryption schemes with a classical key generation process and one-way functions.

Recall the definitions of QPKE schemes from [Section 3.5](#). We will focus on the specific public encryption schemes defined as follows:

Definition 6.1. A public key encryption scheme with a classical key generation process, relative to a random oracle $H \leftarrow \mathcal{H}_\lambda$ consists of the following three bounded-query quantum algorithms:

- $\text{Gen}^H(1^\lambda) \rightarrow (\text{pk}, \text{sk})$: The key generation algorithm that generates a pair of public key pk and secret key sk .
- $\text{Enc}^{|H\rangle}(\text{pk}, m) \rightarrow \text{ct}$: the encryption algorithm that takes a public key pk , the plaintext m , and the randomness r , produces the ciphertext ct .
- $\text{Dec}^{|H\rangle}(\text{sk}, \text{ct}) \rightarrow m'$: the decryption algorithm that takes secret key sk and ciphertext ct and outputs the plaintext m' .

Here we use H to denote the algorithm has classical access to the oracle, and $|H\rangle$ to denote quantum access to the oracle.

The algorithms should satisfy the following requirements:

Perfect Completeness $\Pr \left[\text{Dec}^{|H\rangle}(\text{sk}, \text{Enc}^{|H\rangle}(\text{pk}, m)) = m : \text{Gen}^H(1^\lambda) \rightarrow (\text{pk}, \text{sk}) \right] = 1.$

IND-CPA Security For any QPT adversary $\mathcal{E}^{|H\rangle}$, for every two plaintexts $m_0 \neq m_1$ chosen by $\mathcal{E}^{|H\rangle}(\text{pk})$ we have

$$\Pr \left[\mathcal{E}^{|H\rangle}(\text{pk}, \text{Enc}^{|H\rangle}(\text{pk}, m_b)) = b \right] \leq \frac{1}{2} + \text{negl}(\lambda).$$

To prove a black box separation, we would view the public key encryption scheme as a two-round key agreement protocol as described before, and utilize the previous tools from information theory. From the perspective of a key agreement protocol, it could be viewed as a two-stage Alice: before sending the first message, Alice would first make classical queries to the oracle H ; And after receiving the message from Bob, it would make quantum queries to the oracle, and output the key they agree on. We denote the first stage as \mathcal{A}_1 and the second stage as \mathcal{A}_2 .

We would use the following lemma from [BBBV97].

Lemma 6.2. Consider a quantum algorithm \mathcal{B} that makes d queries to an oracle H . Denote the quantum state immediately after t queries to the oracle as

$$|\psi_t\rangle = \sum_{x,w} \alpha_{x,w,t} |x, w\rangle,$$

where w is the content of the workspace register. Denote the query weight q_x of input x as

$$q_x = \sum_{t=1}^d \sum_w |\alpha_{x,w,t}|^2.$$

For any oracle \tilde{H} , denote $|\phi_d\rangle$ as the final state before measurement obtained by running \mathcal{B} with oracle \tilde{H} , we have that

$$\| |\psi_d\rangle - |\phi_d\rangle \| \leq 2\sqrt{d} \sqrt{\sum_{x: \tilde{H}(x) \neq H(x)} q_x}.$$

To define Eve's algorithm, we would first introduce the following algorithm \mathcal{B}' as the modified algorithm of \mathcal{B} :

1. On receiving message m_1 , first run $\mathcal{B}(m_1)$, and stop before the final output measurement.
2. Repeat the following process $3d^2(\log d + \log(1/\epsilon))$ times: randomly choose $i \leftarrow [d]$, simulate $\mathcal{B}(m_1)$ to its i th query to the oracle, and measure the input register, obtaining output $x \in [2^n]$, and classically query the oracle to obtain $H(x)$.
3. Measure the output register of the first $\mathcal{B}(m_1)$, and obtain key k_B and the second message m_2 .

The first part of the the adversary algorithm \mathcal{E} would be generating $O(dn/\epsilon^2)$ copies of m_1 , and repeating the \mathcal{B}' algorithm on each of these copies without performing step 3. It records all the query input-output pairs $R_E = \{(x_E, H(x_E))\}$ obtained from step 2 of \mathcal{B} .

We denote the heavy weight set W_B of $\mathcal{B}^{(H)}(m_1)$ as $W_B = \{x: q_x \geq \epsilon^2/d^2\}$. We have the following lemma:

Lemma 6.3. *Let In_E be the set of input that is recorded in R_E , we have that*

$$\Pr[W_B \not\subseteq In_E] \leq \epsilon.$$

Proof. For each $x \in W_B$, it would be measured w.p. at least ϵ^2/d^3 at each step 2 \mathcal{B}' performed by \mathcal{E} . Thus the probability it is not measured is bounded by

$$\Pr[x \notin In_E] \leq \left(1 - \frac{\epsilon^2}{d^3}\right)^{3d^3n(\log d + \log(1/\epsilon))/\epsilon^2} \leq \epsilon^3/d^3.$$

Since $\sum_x q_x = d$, we have $|W_B| \leq d^3/\epsilon^2$, thus by a union bound we obtain the desired result. \square

Since step 2 of \mathcal{B}' would not affect the output of the algorithm, we can equivalently think as the key agreement protocol consists of \mathcal{A} and \mathcal{B}' . Note that in this case, \mathcal{E} holds $O(dn/\epsilon^2)$ copies of registers that suffice the permutation invariant condition. Consider the joint state ρ_{ABE} at the stage before \mathcal{B}' performing step 3, by [Lemma 4.2](#) we have that $I(A : B | E) \leq \epsilon^2$. Moreover, applying the channel $\mathcal{T}: E \rightarrow E \otimes A'$ from [Theorem 3.11](#), we can generate a state $\rho_{A'BE}$ such that $TD(\rho_{ABE}, \rho_{A'BE}) \leq \epsilon$.

For the rest of this section, we are only interested in the following registers of A, B, E:

- The classical internal state st_A and oracle query input-output pairs $R_A = \{(x_A, H(x_A))\}$. This assumption can be made since the secret key sk is classical.
- The output register consists of the key k_B and the message m_2 of \mathcal{B} .
- The input-output pairs recorded by Eve $R_E = \{(x_E, H(x_E))\}$.

Without specification, we reuse A, B, E for the beyond registers respectively. We denote the measurement in computational basis on these registers as Π_{ABE} , and the measurement outcome distribution as D_{ABE} . Similarly we can define $\Pi_{A'BE}$ and $D_{A'BE}$.

By perfect completeness, any sample $\text{View}_{ABE} = (st_A, R_A, k_B, m_2, R_E) \leftarrow D_{ABE}$ would be a valid execution. That is to say, there is some oracle H' that is consistent with R_A and R_E , and running \mathcal{A}_2 on H' would always output $k_A = k_B$. It also implies for any $x \in In_A \cap In_E$, the corresponding input-output pair should also be consistent for R_A and R_E .

Since $TD(\rho_{ABE}, \rho_{A'BE}) \leq \epsilon$, we have that $TV(D_{ABE}, D_{A'BE}) \leq \epsilon$ by the operational meaning of trace distance. By [Lemma 4.6](#), we have that

$$\Pr_{\text{View}_{A'BE} \leftarrow D_{A'BE}} [\text{View}_{A'BE} \notin \text{SUPP}(D_{ABE})] \leq 2\epsilon.$$

Now we show that if we reprogram the oracle H to make it consistent with $R_{A'} = \{(x_A, H'(x_A))\}$ from the new sample $\text{View}_{A'BE}$, with high probability the algorithm \mathcal{B} is consistent with \tilde{H} and output (k_B, m_2) . We define the reprogrammed oracle \tilde{H} as follows:

$$\tilde{H}(x) = \begin{cases} H'(x), & x \in In_A; \\ H(x), & \text{else.} \end{cases}$$

We have the following theorem,

Theorem 6.4. *For the reprogrammed oracle \tilde{H} defined as beyond, for any quantum algorithm \mathcal{B} making d queries to the oracle*

$$\Pr_{(k_B, m_2) \leftarrow \mathcal{B}^{|\tilde{H}\rangle}(m_1)} \left[(k_B, m_2) \in \text{SUPP} \left(\mathcal{B}^{|\tilde{H}\rangle}(m_1) \right) \mid \text{View}_{A'BE} \in \text{SUPP}(D_{ABE}) \right] \geq 1 - O(\epsilon),$$

here we slightly abuse the notation $\mathcal{B}^{|\tilde{H}\rangle}(m_1)$ for the classical output distribution of \mathcal{B} .

Proof. Combining [Theorem 3.11](#) and [Lemma 6.3](#) we have that

$$\Pr[\text{View}_{A'BE} \in \text{SUPP}(D_{ABE}) \wedge W_B \subseteq In_E] \geq 1 - \Pr[\text{View}_{A'BE} \notin \text{SUPP}(D_{ABE})] - \Pr[W_B \not\subseteq In_E] \geq 1 - O(\epsilon).$$

Given $\text{View}_{A'BE} \in \text{SUPP}(D_{ABE})$ and $W_B \subseteq In_E$, we apply [Lemma 6.2](#) and obtain that

$$\begin{aligned} TV \left(\mathcal{B}^{|\tilde{H}\rangle}(m_1), \mathcal{B}^{|\tilde{H}\rangle}(m_1) \right) &\leq 4 \| |\psi_d\rangle - |\phi_d\rangle \| \\ &\leq 8\sqrt{d} \sqrt{\sum_{x: \tilde{H}(x) \neq H(x)} q_x} \\ &\leq 8\sqrt{d} \sqrt{d \cdot \frac{\epsilon^2}{d^2}} = O(\epsilon), \end{aligned}$$

where the first inequality comes from [[BBBV97](#), Theorem 3.1], the second inequality is [Lemma 6.2](#), and the third inequality is by that $|\{x: \tilde{H}(x) \neq H(x)\}| \leq |In_A| \leq d$. By [Lemma 4.6](#), we have that

$$\Pr_{(k_B, m_2) \leftarrow \mathcal{B}^{|\tilde{H}\rangle}(m_1)} \left[(k_B, m_2) \in \text{SUPP} \left(\mathcal{B}^{|\tilde{H}\rangle}(m_1) \right) \mid \text{View}_{A'BE} \in \text{SUPP}(D_{ABE}) \wedge W_B \subseteq In_E \right] \geq 1 - O(\epsilon).$$

The final statement can be obtained by a conditional probability formula. □

Now we will prove the first main theorem of the section.

Theorem 6.5. *For any public key encryption scheme in the random oracle model, if it satisfies the following conditions:*

1. *Perfect completeness;*
2. *The key generation algorithm makes at most d classical queries to the oracle;*
3. *The encryption algorithm makes at most d quantum queries to the oracle;*
4. *The decryption algorithm makes at most D quantum queries to the oracle.*

There exists an adversary Eve that could break the public key encryption scheme w.p. $1 - O(\epsilon)$ by making $O(d^4 n (\log d + \log(1/\epsilon)) / \epsilon^2 + D)$ queries to the oracle.

Proof. We have already described most of the adversary algorithm \mathcal{E} as we prove the beyond lemmata and theorems. We summarize the algorithm \mathcal{E} as follows: it first generates $O(dn/\epsilon^2)$ copies of m_1 , and repeats the \mathcal{B}' algorithm on each of these copies without performing step 3. It applies the channel \mathcal{T} to the register E (including the full workspace) and generates a state $\rho_{A'BE}$. Applying $\Pi_{A'BE}$ to $\rho_{A'BE}$ to obtain a sample $\text{View}_{A'BE} = (st_{A'}, R_{A'}, k_B, m_2, R_E)$. It runs the algorithm $\mathcal{A}_2(m_2)$ on register A' with the reprogrammed oracle \tilde{H} , and obtaining the final output $k_E = k_{A'}$.

Now we argue the correctness of the algorithm. Assume that before performing $\Pi_{A'BE}$, we first perform Π_{ABE} on the state ρ_{ABE} . This assumption can be made since $\Pi_{A'BE}$ and Π_{ABE} commute. By perfect completeness of the protocol, we can see that the result View_{ABE} is a valid internal state, compatible with oracle H . Now we discuss the measurement result of $\Pi_{A'BE}$. Note that in this case, the (k_B, m_2, R_E) in $\text{View}_{A'BE}$ is the same as in View_{ABE} .

Consider the case that $\text{View}_{A'BE} \in \text{SUPP}(D_{ABE})$, i.e. $\text{View}_{A'BE}$ is a valid execution for some oracle H' , we now argue that actually $\text{View}_{A'BE}$ is also a valid execution for oracle \tilde{H} with high probability. From the perspective of A' , the result $st_{A'}, R_{A'}$ is compatible with \tilde{H} , and will output message m_1 . From the perspective of $\mathcal{B}^{(H)}(m_1)$, before the measurement on the output register, it can be viewed as a distribution of key-message pairs $\{(k_B, m_2)\}$. Now we imagine we are running the algorithm $\mathcal{B}^{(\tilde{H})}(m_1)$ instead. By [Theorem 6.4](#), we can see that with probability $1 - O(\epsilon)$, $(k_B, m_2) \leftarrow \mathcal{B}^{(H)}(m_1)$ would also be in the support of $\mathcal{B}^{(\tilde{H})}(m_1)$.

Together they imply that $\text{View}_{A'BE}$ would be a valid execution for \mathcal{A}, \mathcal{B} under oracle \tilde{H} when \mathcal{B} outputs key k_B . Thus if we run $\mathcal{A}_2(m_2)$ with the reprogrammed oracle \tilde{H} , by perfect completeness, it would also output $k'_A = k_B$.

For each run of \mathcal{B}' , it would simulate \mathcal{B} for $O(d^2(\log(d) + \log(1/\epsilon)))$ times, and each time it makes at most d queries. For the adversary algorithm \mathcal{E} , it would run \mathcal{B}' for $O(dn/\epsilon^2)$ times, giving us the query complexity beyond. \square

Reviewing our proof beyond, we notice that m_1 may not necessarily be a classical message. As long as $|m_1\rangle$ is a pure state, and we can obtain any polynomial copies of $|m_1\rangle$, the analysis in our proof exactly applies. Recall the IND-CPA Security from [Definition 3.16](#), if the public key is pure, the adversary algorithm \mathcal{E} can obtain polynomial many copies of $|\text{pk}\rangle$. For any message $m_0 \neq m_1$, we can construct a key agreement for one bit by setting the second message as the ciphertext ct_0/ct_1 respectively. Thus by running our \mathcal{E} beyond, we can break the IND-CPA security game with advantage $1 - O(\epsilon)$.

Now we show that our attack would still apply when the ciphertext is also quantum. The proof would be similar to the proof beyond. By running the same algorithm \mathcal{E} we obtained the state $\rho_{A'BE}$, we would perform the measurement $\Pi_{A'BE}$, but for the B part we do not include the ciphertext ρ_{m_2} register. We can define \tilde{H} as beyond, and prove the following statements given $\text{View}_{A'BE}$ is a valid execution:

1. From the perspective of A' , the result $st_{A'}, R_{A'}$ is compatible with \tilde{H} , and will output message $|m_1\rangle$. By definition, $R_{A'}$ is compatible with \tilde{H} . Since B is not affected by channel \mathcal{T} , by uncomputing \mathcal{B} on state $\rho_{A'B}$, we can see that \mathcal{B} would also receive $|m_1\rangle$ from A' . From the observation, we can see that after performing the uncomputation, $TD(\rho_A, \rho_{A'}) = TD(\rho_{A'B}, \rho_{AB}) \leq \epsilon$, thus $st_{A'}, R_{A'}$ is compatible with output $|m_1\rangle$ under oracle \tilde{H} w.p. $1 - \epsilon$.
2. From the perspective of $\mathcal{B}^{(H)}(|m_1\rangle)$, the state before the measurement $\Pi_{A'BE}$ can be written as $\rho_B = \sum_{k_B} p_{k_B} |k_B\rangle \langle k_B| \otimes \rho_{k_B}$. Similarly, we denote the state of $\mathcal{B}^{(\tilde{H})}(|m_1\rangle)$ as $\sigma_B = \sum_{k_B} p'_{k_B} |k_B\rangle \langle k_B| \otimes \sigma_{k_B}$. In the proof of [Theorem 6.4](#), when we apply [Lemma 6.2](#), we can

obtain that $TD(\rho_B, \sigma_B) \leq O(\epsilon)$. Now we consider the classical part of ρ_B and σ_B . Since partial trace will not increase the trace distance, we have for distribution $D_B = \{p_{k_B}\}$ and $D'_B = \{p'_{k_B}\}$, $TV(D_B, D'_B) \leq \epsilon$. Thus for state $\rho'_{AB} = \sum_{k_B} p'_{k_B} |k_B\rangle \langle k_B| \otimes \rho_{k_B}$, $TD(\rho_B, \rho'_{AB}) \leq \epsilon$, and by triangular inequality, $TD(\rho'_B, \sigma_B) \leq 2\epsilon$. Since $TD(\rho'_B, \sigma_B) = \mathbb{E}[TD(\rho_{k_B}, \sigma_{k_B})]$, using Markov inequality, we can see that

$$\Pr_{(sk, k_B) \leftarrow D_{A'B}} [TD(\rho_{k_B}, \sigma_{k_B}) \leq C\epsilon] \geq 1 - \frac{2}{C}.$$

If we take $C = 1/\sqrt{\epsilon}$, we can obtain that w.p. $1 - O(\sqrt{\epsilon})$, $TD(\rho_{k_B}, \sigma_{k_B}) \leq \sqrt{\epsilon}$.

We can see that $|m_1\rangle, \sigma_{k_B}$ would be a valid transcript for \mathcal{A}, \mathcal{B} under oracle \tilde{H} . Thus when ρ_{k_B} and σ_{k_B} is $O(\sqrt{\epsilon})$ close, $\mathcal{A}^{\tilde{H}}(sk, \rho_{k_B})$ will output $k_{A'} = k_B$ with probability $1 - O(\sqrt{\epsilon})$. Thus we obtain the following theorem:

For the clonable public key case $\rho_{m_1} = \sum_i p_i |\psi_i\rangle \langle \psi_i|$, we observe that the output of $\mathcal{B}(\rho_{m_1})$ would be the convex combination of $\mathcal{B}(|\psi_i\rangle)$. Thus by the perfect completeness property of the protocol, each $|\psi_i\rangle$ would also be a valid public key. Further observing that the clonable case is a convex combination of pure public keys, we have the following theorem:

Theorem 6.6. *For any QPKE scheme with quantum public key in the random oracle model, if it satisfies the following conditions:*

1. *Perfect completeness;*
2. *The public key ρ_{pk} is pure or clonable.*
3. *The key generation algorithm makes at most d classical queries to the oracle;*
4. *The encryption algorithm makes at most d quantum queries to the oracle;*
5. *The decryption algorithm makes at most D quantum queries to the oracle.*

There exists an adversary Eve that could break the public key encryption scheme w.p. $1 - O(\sqrt{\epsilon})$ by making $O(d^4 n (\log d + \log(1/\epsilon))/\epsilon^2 + D)$ queries to the oracle.

This theorem gives a tight characterization of multiple existing QPKE schemes. In [BGH⁺23] and [Col23], they both provided a QPKE scheme with a pure quantum public key, but their public key generation algorithms need to make quantum queries. Our result shows that the quantum query is necessary for their key agreement scheme. In [KMNY23, MW23], they provided another QPKE scheme where the quantum public key is mixed, but the key generation algorithm can only make classical queries. Our result shows that their key must be mixed and unclonable in a strong sense.

References

- [AA09] Scott Aaronson and Andris Ambainis. “The need for structure in quantum speedups”. In: *arXiv preprint arXiv:0911.0996* (2009) (cit. on pp. 2, 3, 9).
- [Aar09] Scott Aaronson. “Quantum copy-protection and quantum money”. In: *2009 24th Annual IEEE Conference on Computational Complexity*. IEEE, 2009, pp. 229–242 (cit. on p. 2).

- [ACC⁺22] Per Austrin, Hao Chung, Kai-Min Chung, Shiuan Fu, Yao-Ting Lin, and Mohammad Mahmoody. “On the impossibility of key agreements from quantum random oracles”. In: *Annual International Cryptology Conference*. Springer. 2022, pp. 165–194 (cit. on pp. 2–4, 6, 8).
- [AL70] Huzihiro Araki and Elliott H Lieb. “Entropy inequalities”. In: *Communications in Mathematical Physics* 18.2 (1970), pp. 160–170 (cit. on p. 15).
- [ALL⁺21] Scott Aaronson, Jiahui Liu, Qipeng Liu, Mark Zhandry, and Ruizhe Zhang. “New approaches for quantum copy-protection”. In: *Annual International Cryptology Conference*. Springer. 2021, pp. 526–555 (cit. on p. 2).
- [AQY22] Prabhanjan Ananth, Luowen Qian, and Henry Yuen. “Cryptography from pseudo-random quantum states”. In: *Annual International Cryptology Conference*. Springer. 2022, pp. 208–236 (cit. on p. 2).
- [BB14] Charles H. Bennett and Gilles Brassard. “Quantum cryptography: Public key distribution and coin tossing”. In: *Theoretical Computer Science* 560 (2014), pp. 7–11. DOI: [10.1016/j.tcs.2014.05.025](https://doi.org/10.1016/j.tcs.2014.05.025) (cit. on p. 2).
- [BBBV97] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. “Strengths and Weaknesses of Quantum Computing”. In: *SIAM J. Comput.* 26.5 (1997), pp. 1510–1523 (cit. on pp. 12, 22, 24).
- [BCKM21] James Bartusek, Andrea Coladangelo, Dakshita Khurana, and Fermi Ma. “On the round complexity of secure quantum computation”. In: *Advances in Cryptology—CRYPTO 2021: 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16–20, 2021, Proceedings, Part I 41*. Springer. 2021, pp. 406–435 (cit. on p. 2).
- [BGH⁺23] Khashayar Barooti, Alex B. Grilo, Loïc Huguenin-Dumittan, Giulio Malavolta, Or Sattath, Quoc-Huy Vu, and Michael Walter. *Public-Key Encryption with Quantum Keys*. 2023. arXiv: [2306.07698](https://arxiv.org/abs/2306.07698) [quant-ph] (cit. on pp. 3, 4, 6, 16, 26).
- [BGVV23] Samuel Bouaziz–Ermann, Alex B Grilo, Damien Vergnaud, and Quoc-Huy Vu. “Towards the Impossibility of Quantum Public Key Encryption with Classical Keys from One-Way Functions”. In: *arXiv preprint arXiv:2311.03512* (2023) (cit. on pp. 4–7, 21).
- [BKS⁺11] Zvika Brakerski, Jonathan Katz, Gil Segev, and Arkady Yerukhimovich. “Limits on the power of zero-knowledge proofs in cryptographic constructions”. In: *Theory of Cryptography: 8th Theory of Cryptography Conference, TCC 2011, Providence, RI, USA, March 28–30, 2011. Proceedings 8*. Springer. 2011, pp. 559–578 (cit. on p. 7).
- [BM09] Boaz Barak and Mohammad Mahmoody-Ghidary. “Merkle puzzles are optimal—an $O(n^2)$ -query attack on any key exchange from a random oracle”. In: *Annual International Cryptology Conference*. Springer. 2009, pp. 374–390 (cit. on pp. 7–9, 16).
- [Col23] Andrea Coladangelo. “Quantum trapdoor functions from classical one-way functions”. In: *arXiv preprint arXiv:2302.12821* (2023) (cit. on pp. 3, 4, 6, 16, 26).
- [FR15] Omar Fawzi and Renato Renner. “Quantum conditional mutual information and approximate Markov chains”. In: *Communications in Mathematical Physics* 340.2 (2015), pp. 575–611 (cit. on pp. 6, 10, 15).

- [GLSV21] Alex B Grilo, Huijia Lin, Fang Song, and Vinod Vaikuntanathan. “Oblivious transfer is in MiniQCrypt”. In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2021, pp. 531–561 (cit. on p. 2).
- [HJPW04] Patrick Hayden, Richard Jozsa, Denes Petz, and Andreas Winter. “Structure of states which satisfy strong subadditivity of quantum entropy with equality”. In: *Communications in mathematical physics* 246 (2004), pp. 359–374 (cit. on p. 10).
- [Imp95] Russell Impagliazzo. “A personal view of average-case complexity”. In: *Proceedings of Structure in Complexity Theory. Tenth Annual IEEE Conference*. IEEE. 1995, pp. 134–147 (cit. on p. 2).
- [IR89] Russell Impagliazzo and Steven Rudich. “Limits on the provable consequences of one-way permutations”. In: *Proceedings of the twenty-first annual ACM symposium on Theory of computing*. 1989, pp. 44–61 (cit. on pp. 2, 7–9, 16).
- [KMNY23] Fuyuki Kitagawa, Tomoyuki Morimae, Ryo Nishimaki, and Takashi Yamakawa. “Quantum public-key encryption with tamper-resilient public keys from one-way functions”. In: *arXiv preprint arXiv:2304.01800* (2023) (cit. on pp. 3, 4, 6, 16, 17, 26).
- [Maz23] Noam Mazon. “Key-agreement with perfect completeness from random oracles”. In: *Cryptology ePrint Archive* (2023) (cit. on p. 3).
- [Mül92] JW Müller. “Parity moments for a Poisson variable”. In: *BIPM WPN-235* (1992) (cit. on p. 32).
- [MW23] Giulio Malavolta and Michael Walter. *Robust Quantum Public-Key Encryption with Applications to Quantum Key Distribution*. Cryptology ePrint Archive, Paper 2023/500. <https://eprint.iacr.org/2023/500>. 2023. URL: <https://eprint.iacr.org/2023/500> (cit. on pp. 3, 4, 6, 17, 26).
- [MY22a] Tomoyuki Morimae and Takashi Yamakawa. “One-wayness in quantum cryptography”. In: *arXiv preprint arXiv:2210.03394* (2022) (cit. on p. 3).
- [MY22b] Tomoyuki Morimae and Takashi Yamakawa. “Quantum commitments and signatures without one-way functions”. In: *Annual International Cryptology Conference*. Springer. 2022, pp. 269–295 (cit. on p. 2).
- [NC10] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: [10.1017/CBO9780511976667](https://doi.org/10.1017/CBO9780511976667) (cit. on pp. 13, 14).
- [Sho99] Peter W Shor. “Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer”. In: *SIAM review* 41.2 (1999), pp. 303–332 (cit. on p. 2).
- [Wie83] Stephen Wiesner. “Conjugate coding”. In: *ACM Sigact News* 15.1 (1983), pp. 78–88 (cit. on p. 2).
- [Wil11] Mark M Wilde. “From classical to quantum Shannon theory”. In: *arXiv preprint arXiv:1106.1445* (2011) (cit. on p. 14).
- [YZ22] Takashi Yamakawa and Mark Zhandry. “Verifiable Quantum Advantage without Structure”. In: *arXiv preprint arXiv:2204.02063* (2022) (cit. on p. 3).

[Zha19] Mark Zhandry. “How to record quantum queries, and applications to quantum indistinguishability”. In: *Annual International Cryptology Conference*. Springer. 2019, pp. 239–268 (cit. on pp. 9, 29).

A Additional Preliminaries

A.1 Compressed Oracle

The analysis of [Appendix B](#) relies on the compressed oracle technique of Zhandry [Zha19]. Here, We briefly introduce the ideas we needed. Zhandry shows that the standard quantum random oracle is perfectly indistinguishable from the purified random oracle by any unbound quantum adversary. For a standard random oracle $H : [2^n] \rightarrow \{0, 1\}$, the initial state of the purified random oracle is $\sum_{H \in \{0,1\}^{2^n}} |H\rangle$. The oracle access to $|H\rangle$ is defined by a unitary transformation U_H where

$$U_H |x, y\rangle |H\rangle := (-1)^{y \cdot H(x)} |x, y\rangle |H\rangle.$$

Given a vector $S \in \{0, 1\}^{2^n}$, define $|\hat{D}\rangle := \sum_{H \in \{0,1\}^{2^n}} (-1)^{\langle H, D \rangle} |H\rangle$ where $\langle \cdot, \cdot \rangle$ denotes inner product. We call $\{|\hat{D}\rangle\}$ the Fourier basis of the oracle space. Then we have following lemma.

Lemma A.1. *Let \mathcal{A} be a quantum algorithm that makes d queries to the random oracle, the final state of \mathcal{A} can be written as $\sum_{|D| \leq d} \alpha_D |\psi_D\rangle |\hat{D}\rangle$ where $|D|$ denotes the number of non-zero entries in D .*

Proof. Prove by induction on d . When $d = 0$, the initial joint state of working register A, the query register Q and oracle register H can be written as $|\psi\rangle_{AQ} \sum_{H \in \{0,1\}^{2^n}} |H\rangle_H = |\psi\rangle_{AQ} |\hat{0}\rangle_H$. Supposing the statement holds for d , then the current state can be written as $\sum_{u,x,y,|D| \leq d} \alpha_{u,x,y,D} |u\rangle_A |x, y\rangle_Q |\hat{D}\rangle_H$. After making one more query, the state becomes

$$\begin{aligned} \sum_{u,x,y,|D| \leq d} \alpha_{u,x,y,D} |u\rangle_A U_H |x, y\rangle_Q |\hat{D}\rangle_H &= \sum_{u,x,y,|D| \leq d} \alpha_{u,x,y,D} |u\rangle_A U_H |x, y\rangle_Q \sum_H (-1)^{\langle H, D \rangle} |H\rangle_H \\ &= \sum_{u,x,y,|D| \leq d} \alpha_{u,x,y,D} |u\rangle_A |x, y\rangle_Q \sum_H (-1)^{\langle H, D \rangle + y \cdot H(x)} |H\rangle_H \\ &= \sum_{u,x,y,|D| \leq d} \alpha_{u,x,y,D} |u\rangle_A |x, y\rangle_Q \sum_H (-1)^{\langle H, D \oplus (x, y) \rangle} |H\rangle_H \\ &= \sum_{u,x,y,|D| \leq d} \alpha_{u,x,y,D} |u\rangle_A U_H |x, y\rangle_Q |\widehat{D \oplus (x, y)}\rangle_H \end{aligned}$$

where $|D \oplus (x, y)| \leq d + 1$. □

B Non-Adaptive Quantum Key Agreement

Non-adaptive key agreement protocol is a special case of non-interactive protocol in [Section 5](#). It restricts Alice and Bob by only allowing them to make non-adaptive queries, i.e., queries can depend on neither any previous queries nor classical communication. The famous Merkle Puzzles and PRS-based protocols are both examples of this non-adaptive case. Formally,

Definition B.1 (Non-adaptive Quantum Key Agreement). A QKA protocol between Alice and Bob with access to a random oracle H is called non-adaptive if (i) the protocol is non-interactive (see [Definition 5.1](#)), (ii) Alice's (Bob's) query algorithm consists of the following steps:

1. Prepare a quantum state $\sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_d$ where σ_i is the state of the input and output registers of the i -th query.
2. Make the d queries to H in parallel.

We show that for this special case, the number of queries Eve needed can be improved from $O(d^2e)$ to $\tilde{O}(d^2)$, which is independent of the input size of the random oracle. We remark that unlike [Theorem 5.2](#), this result only works for random oracle.

Theorem B.2. Given a non-adaptive QKA protocol between Alice and Bob with access to a random oracle H , let k_A, k_B be key of Alice and Bob respectively. If $\Pr[k_A = k_B] \geq \delta$, then there exists an eavesdropper Eve who outputs her key k_E with $O\left(\frac{d^2}{\epsilon^2} \log^2 \frac{d}{\epsilon}\right)$ queries such that $\Pr[k_E = k_A] \geq \delta - \epsilon$.

The rest of this section is a proof of [Theorem B.2](#). Similar to non-interactive case, our goal is to show that Eve can somehow repeat Alice's query algorithm to make CMI sufficiently small so that she can recover the key.

B.1 Recasting It as a Classical Random Walk Problem

Since queries are non-adaptive, we assume Alice and Bob use the same query algorithm denoted by \mathcal{Q} without loss of generality. Let $H : [2^n] \rightarrow \{0, 1\}$ be the random oracle. By [Lemma A.1](#), the state after applying \mathcal{Q} once can be written as

$$\sum_{D \in \{0,1\}^{2^n} : |D| \leq d} \alpha_D |\psi_D\rangle |\hat{D}\rangle$$

where $|\psi_D\rangle$ are states in the working register and $|\hat{D}\rangle$ are the Fourier basis states in the oracle register (see [Appendix A.1](#)). The following lemma describes properties of the states $|\psi_D\rangle$, which follows directly from [Lemma A.1](#).

Lemma B.3. (a) If the query is non-adaptive, then $|\psi_D\rangle$ and $|\psi_{D'}\rangle$ are orthogonal if $D \neq D'$.
(b) If we repeat \mathcal{Q} for t times, then the joint state of working register and oracle register is

$$\sum_{D_1, D_2, \dots, D_t} \alpha_{D_1} \alpha_{D_2} \cdots \alpha_{D_t} |\psi_{D_1}\rangle |\psi_{D_2}\rangle \cdots |\psi_{D_t}\rangle |\widehat{\bigoplus_{i=1}^t D_i}\rangle.$$

Let \mathcal{D} denote the distribution on $\{0, 1\}^{2^n}$ where $\Pr_{D \sim \mathcal{D}}[D = x] = |\alpha_x|^2$. Define random variable $D^t := D_1 \oplus \cdots \oplus D_t$ where D_1, D_2, \dots, D_t are i.i.d. samples drawn from \mathcal{D} . We remark that D^t can be viewed as a random walk with t steps in the hypercube $\{0, 1\}^{2^n}$ where each step is sampled by distribution \mathcal{D} .

We first consider the same attack strategy as in [Section 5](#) where Eve repeats \mathcal{Q} for sufficient large times. We have the following lemma.

Lemma B.4. Let A^0 and B^0 denote the state of Alice and Bob right after applying \mathcal{Q} respectively, E^0 be the state of Eve repeating \mathcal{Q} for t times. Then $I(A^0 : B^0 | E^0) = 2S(D^{t+1}) - S(D^t) - S(D^{t+2})$.

Proof. By the definition of CMI, we have

$$I(A^0 : B^0 | E^0) = S(A^0 E^0) + S(B^0 E^0) - S(E^0) - S(A^0 B^0 E^0).$$

Then we compute each term separately. Since Alice and Eve together run \mathcal{Q} for $t + 1$ times, by [Lemma B.3\(b\)](#) the joint state of Alice A, Eve E and Oracle H can be viewed as pure state

$$|\phi\rangle_{\text{AEH}} = \sum_{D_1, D_2, \dots, D_{t+1}} \alpha_{D_1} \alpha_{D_2} \cdots \alpha_{D_{t+1}} |\psi_{D_1}\rangle_{\text{A}} (|\psi_{D_2}\rangle \cdots |\psi_{D_{t+1}}\rangle)_{\text{E}} |\widehat{\bigoplus_{i=1}^{t+1} D_i}\rangle_{\text{H}}.$$

Let $\sigma_{\text{AEH}} = |\phi\rangle_{\text{AEH}} \langle\phi|_{\text{AEH}}$, $\sigma_{\text{AE}} = \text{Tr}_{\text{H}}(\sigma_{\text{AEH}})$, $\sigma_{\text{H}} = \text{Tr}_{\text{AE}}(\sigma_{\text{AEH}})$. By [Lemma B.3\(a\)](#), $|\psi_{D_i}\rangle$ are orthogonal states. Then we have

$$\sigma_{\text{H}} = \text{Tr}_{\text{AE}}(\sigma_{\text{AEH}}) = \sum_{D_1, D_2, \dots, D_{t+1}} |\alpha_{D_1} \alpha_{D_2} \cdots \alpha_{D_{t+1}}|^2 |\widehat{\bigoplus_{i=1}^{t+1} D_i}\rangle \langle\widehat{\bigoplus_{i=1}^{t+1} D_i}|.$$

Since σ_{AEH} is a pure state, we have $S(\sigma_{\text{AE}}) = S(\sigma_{\text{H}})$. Then $S(A^0 E^0) = S(\sigma_{\text{AE}}) = S(\sigma_{\text{H}}) = S(D^{t+1})$. Similarly, we have $S(B^0 E^0) = S(D^{t+1})$ and $S(E^0) = S(D^t)$ and $S(A^0 B^0 E^0) = S(D^{t+2})$. Thus $I(A^0 : B^0 | E^0) = 2S(D^{t+1}) - S(D^t) - S(D^{t+2})$. \square

Thus the quantum CMI can be expressed using the entropy of classical variable D^t . If we can show that $2S(D^{t+1}) - S(D^t) - S(D^{t+2})$ is small, then we can conclude [Theorem B.2](#). Unfortunately, $S(D^t)$ is difficult to compute, so we slightly modify Eve's strategy using Poissonization trick. We first define Poissonized version of Alice:

1. For each $i \in [d]$, Alice samples an independent Poisson variable P_i with parameter $\log \mu d$ where $\mu \geq 2$, and then repeats the i -th query for P_i times.
2. If Alice repeats each query at least once, then she will execute the original protocol and output the key as before. Otherwise, she aborts and output key 0.

Poissonized version of Bob is defined similarly. We remark that Poissonized Alice and Bob will make expected $O(d \log d)$ queries. W.p. $(1 - e^{-\log \mu d})^{2d} \geq 1 - 1/\mu$, Poissonized Alice and Bob will repeat each query at least once and then agree on the same key as before. Thus if Eve can recover Poissonized Alice and Bob's key, she can also recover the real key w.p. at least $1 - 1/\mu$. We have the following theorem.

Theorem B.5. *Given a positive integer t , let \tilde{A}^0 and \tilde{B}^0 be the state of Poissonized Alice and Bob right after querying, \tilde{E}^0 be t copies of \tilde{A}^0 . Then $I(\tilde{A}^0 : \tilde{B}^0 | \tilde{E}^0) = O(d \log \mu d / t)$.*

Before proving the above theorem, we first use it to conclude [Theorem B.2](#).

Proof of Theorem B.2. Let \tilde{A}^0 and \tilde{B}^0 be the state of Poissonized Alice and Bob right after querying, \tilde{A}^f and \tilde{B}^f be the states of Poissonized Alice and Bob respectively right after finishing the communication but before outputting the key, Π denote the communications. Let Eve's attack strategy be as follows:

1. Eve repeats \tilde{A}^0 for $O\left(\frac{d \log \mu d}{\epsilon^2}\right)$ times and obtains state \tilde{E}^0 .
2. Eve applies the channel in [Lemma 2.3](#) w.r.t. system $\tilde{A}^f \tilde{E}^0 \tilde{B}^f$ and obtains a fake view $\tilde{A}^{f'}$ of \tilde{A}^f . Then use $\tilde{A}^{f'}$ to output key k_E .

By [Theorem B.5](#), we have $I(\tilde{\mathbf{A}}^0 : \tilde{\mathbf{B}}^0 | \tilde{\mathbf{E}}^0) = O(\epsilon^2)$. Since classical communication does not increase CMI by [Lemma 4.4](#) and [Lemma 4.3](#), we have $I(\tilde{\mathbf{A}}^f : \tilde{\mathbf{B}}^f | \tilde{\mathbf{E}}^0, \Pi) \leq I(\tilde{\mathbf{A}}^0 : \tilde{\mathbf{B}}^0 | \tilde{\mathbf{E}}^0) = O(\epsilon^2)$. Then by [Theorem 3.11](#), $\tilde{\mathbf{A}}^f \tilde{\mathbf{B}}^f$ is $O(\epsilon)$ -close to $\tilde{\mathbf{A}}^f \tilde{\mathbf{B}}^f$. Let \tilde{k}_A be the key of Poissonized Alice. Then $\Pr[k_E = \tilde{k}_A] \geq \delta - O(\epsilon)$. Since w.p. $1 - 1/\mu$ the Poissonized Alice and Bob will be the same as real Alice and Bob, we have $\Pr[k_E = k_A] \geq \delta - O(\epsilon) - 1/\mu$. By setting $\mu = 2/\epsilon$ and adjusting constants, we have $\Pr[k_E = k_A] \geq \delta - \epsilon$ and Eve needs $O\left(d \log \mu d \cdot \frac{d \log \mu d}{\epsilon^2}\right) = O\left(\frac{d^2}{\epsilon^2} \log^2 \frac{d}{\epsilon}\right)$ queries. \square

B.2 Proof of [Theorem B.5](#)

First observe that because non-adaptive queries do not depend on each other, the distribution \mathcal{D} can be viewed as the sum of d independent distributions $\mathcal{D}[1], \mathcal{D}[2], \dots, \mathcal{D}[d]$, all of which are on $\{0, 1\}^{2^n}$ with Hamming weight 1. That is, $\mathcal{D} = \mathcal{D}[1] \oplus \mathcal{D}[2] \oplus \dots \oplus \mathcal{D}[d]$. Given a positive integer t , define a random variable \tilde{D}^t as follows:

1. For each $\ell \in [d]$, first sample \tilde{t}_ℓ from $\text{Pois}(t \log \mu d)$ independently and then draw i.i.d. samples $D_1[\ell], D_2[\ell], \dots, D_{\tilde{t}_\ell}[\ell]$ from $\mathcal{D}[\ell]$.
2. Let $\tilde{D}^t = \bigoplus_{\ell=1}^d \bigoplus_{i=1}^{\tilde{t}_\ell} D_i[\ell]$.

Then we have the following lemma, which can be proved similarly as [Lemma B.4](#).

Lemma B.6. *Given a positive integer t , let $\tilde{\mathbf{E}}^0$ be t copies of $\tilde{\mathbf{A}}^0$. Then*

$$I(\tilde{\mathbf{A}}^0 : \tilde{\mathbf{B}}^0 | \tilde{\mathbf{E}}^0) = 2S(\tilde{D}^{t+1}) - S(\tilde{D}^t) - S(\tilde{D}^{t+2}).$$

Thus [Theorem B.5](#) follows directly from the following theorem.

Theorem B.7. *Given an integer $t > 1$, $2S(\tilde{D}^t) - S(\tilde{D}^{t-1}) - S(\tilde{D}^{t+1}) = O(d \log \mu d / t)$.*

Proof. Let $p_{\ell,i} = \Pr(D[\ell] = \mathbf{e}_i)$ and $p_{\ell,0} = \Pr(D[\ell] = 0^{2^n})$ for $\ell \in [d]$ where \mathbf{e}_i denotes the i -th standard basis vector in $\{0, 1\}^{2^n}$. The crucial observation is that \tilde{D}^t can also be generated by the following process:

- For each $\ell \in [d]$ and $i \in [2^n]$, sample $\tilde{t}_{\ell,i}$ from $\text{Pois}(p_{\ell,i} t \log \mu d)$ independently.
- Let $\tilde{D}^t = \sum_{\ell \in [d]} \sum_{i \in [2^n]} \tilde{t}_{\ell,i} \mathbf{e}_i \bmod 2$.

Note that $\{\tilde{t}_{1,1}, \dots, \tilde{t}_{d,2^n}\}$ are mutually independent. Then all of the 2^n bits of \tilde{D}^t are mutually independent, so

$$S(\tilde{D}^t) = \sum_{i=1}^{2^n} S(\text{the } i\text{-th bit of } \tilde{D}^t) = \sum_{i=1}^{2^n} S\left(\text{parity}\left(\sum_{\ell \in [d]} \tilde{t}_{\ell,i}\right)\right).$$

Note that $\sum_{\ell \in [d]} \tilde{t}_{\ell,i} \sim \text{Pois}\left(t \log \mu d \cdot \sum_{\ell \in [d]} p_{\ell,i}\right)$. According to Formula (6) in [[Mül92](#)],

$$\Pr\left[\text{parity}\left(\sum_{\ell \in [d]} \tilde{t}_{\ell,i}\right) = 1\right] = \frac{1 - e^{-2\mathbb{E}[\sum_{\ell \in [d]} \tilde{t}_{\ell,i}]}}{2} = \frac{1 - e^{-2t \log \mu d \cdot \sum_{\ell \in [d]} p_{\ell,i}}}{2}.$$

Then we have

$$2S(\tilde{D}^t) - S(\tilde{D}^{t-1}) - S(\tilde{D}^{t+1}) = \sum_{i=1}^{2^n} f\left(\log \mu d \cdot \sum_{\ell \in [d]} p_{\ell,i}\right)$$

where

$$f(p) := 2H\left(\frac{1 - e^{-2tp}}{2}\right) - H\left(\frac{1 - e^{-2(t-1)p}}{2}\right) - H\left(\frac{1 - e^{-2(t+1)p}}{2}\right),$$

$$H(p) := -p \log p - (1-p) \log(1-p).$$

By Lemma B.8 and the fact that $\sum_{\ell,i} p_{\ell,i} \leq d$, we conclude that

$$2S(\tilde{D}^t) - S(\tilde{D}^{t-1}) - S(\tilde{D}^{t+1}) = O\left(\left(e^{-t}d + \sum_{\ell,i} p_{\ell,i}/t\right) \log \mu d\right) = O(d \log \mu d/t).$$

□

Lemma B.8. For $p \in [0, 1]$, $f(p) = O(p/t)$. For $p > 1$, $f(p) = O(e^{-t})$.

Proof. $f(p) = 0$ when $p = 0$. When $p \in (0, 1]$, since $H(p)$ is concave,

$$\begin{aligned} f(p) &= \left[H\left(\frac{1}{2} - \frac{e^{-2tp}}{2}\right) - H\left(\frac{1}{2} - \frac{e^{-2(t-1)p}}{2}\right) \right] - \left[H\left(\frac{1}{2} - \frac{e^{-2(t+1)p}}{2}\right) - H\left(\frac{1}{2} - \frac{e^{-2tp}}{2}\right) \right] \\ &\leq H'\left(\frac{1}{2} - \frac{e^{-2(t-1)p}}{2}\right) \left(\frac{e^{-2(t-1)p}}{2} - \frac{e^{-2tp}}{2}\right) - H'\left(\frac{1}{2} - \frac{e^{-2(t+1)p}}{2}\right) \left(\frac{e^{-2tp}}{2} - \frac{e^{-2(t+1)p}}{2}\right) \\ &= \frac{1}{\ln 2} (e^{2p} - 1) e^{-2p(t+1)} \left(e^{2p} \tanh^{-1}\left(e^{-2p(t-1)}\right) - \tanh^{-1}\left(e^{-2p(t+1)}\right) \right). \end{aligned}$$

For $p \in (0, 1]$, $(e^{2p} - 1) \leq e^{2p} = O(p)$. Then $f(p) \leq \frac{e^{2p}}{\ln 2} \cdot q^{t+1} \left(\frac{1}{q} \tanh^{-1}(q^{t-1}) - \tanh^{-1}(q^{t+1})\right)$ where $q = e^{-2p} \in [e^{-2}, 1)$. By Lemma B.9, $f(p) = O(p/t)$.

Note that $H(1/2 + x) = 1 + O(x^2)$ when x is small. Then for $p > 1$,

$$\begin{aligned} f(p) &= H\left(\frac{1 - e^{-2tp}}{2}\right) - H\left(\frac{1 - e^{-2(t-1)p}}{2}\right) - H\left(\frac{1 - e^{-2(t+1)p}}{2}\right) \\ &= [2 + O(e^{-4pt})] - [1 + O(e^{-4p(t-1)})] - [1 + O(e^{-4p(t+1)})] = O(e^{-t}). \end{aligned}$$

□

Lemma B.9. Given $t > 1$, $q \in [0, 1)$, then $q^{t+1} \left(\frac{1}{q} \tanh^{-1}(q^{t-1}) - \tanh^{-1}(q^{t+1})\right) = O(1/t)$.

Proof. Split the left-hand side of the equation into two parts:

$$[q^t (\tanh^{-1}(q^{t-1}) - \tanh^{-1}(q^{t+1}))] + [(1-q)q^t \tanh^{-1}(q^{t+1})].$$

By the fact that $\tanh^{-1}(x)$ is convex for $x \in [0, 1)$, the first part

$$\begin{aligned}
q^t (\tanh^{-1}(q^{t-1}) - \tanh^{-1}(q^{t+1})) &\leq q^t (q^{t-1} - q^{t+1}) \tanh^{-1'}(q^{t-1}) \\
&= q^{2t-1} \frac{1 - q^2}{1 - q^{2(t-1)}} \\
&= \frac{q^{2t-1}}{1 + q^2 + q^4 + \dots + q^{2(t-2)}} \\
&\leq \frac{q^{2t-1}}{(t-1)q^{2(t-2)}} = \frac{q^3}{t-1} \leq \frac{1}{t-1}.
\end{aligned}$$

By the inequality $\tanh^{-1}(x) \leq \frac{1}{2} \left(\frac{1+x}{1-x} - 1 \right)$ for $x \in [0, 1)$, the second part

$$\begin{aligned}
(1-q)q^t \tanh^{-1}(q^{t+1}) &\leq (1-q)q^t \cdot \frac{1}{2} \left(\frac{1+q^{t+1}}{1-q^{t+1}} - 1 \right) \\
&= q^{2t+1} \frac{1-q}{1-q^{t+1}} \\
&= \frac{q^{2t+1}}{1+q+q^2+\dots+q^t} \\
&\leq \frac{q^{2t+1}}{(t+1)q^t} = \frac{q^{t+1}}{t+1} \leq \frac{1}{t+1}.
\end{aligned}$$

Thus $q^{t+1} \left(\frac{1}{q} \tanh^{-1}(q^{t-1}) - \tanh^{-1}(q^{t+1}) \right) \leq \frac{1}{t-1} + \frac{1}{t+1} = O(1/t)$ for $q \in [0, 1)$. \square

C Public Key Encryption with a Short Classical Secret Key

In this section, we consider a special case of QPKE where the secret key is classical string of logarithmic length. We show that such kind of QPKE does not exist in any classical oracle model. In the following, we use the same notations as in [Definition 6.1](#) except that the key generation $\text{Gen}^{|H\rangle}$ is also quantum. Formally,

Theorem C.1. *Given a QPKE scheme $(\text{Gen}^{|H\rangle}, \text{Enc}^{|H\rangle}, \text{Dec}^{|H\rangle})$ with classical keys in the oracle model where $H \leftarrow \mathcal{O}_\lambda$ is any oracle whose quantum query unitary U_H acts on e_λ qubits, if the scheme satisfies the following conditions:*

1. *the secret key sk is classical string of $O(\log \lambda)$ length;*
2. *it is δ -complete i.e., $\Pr_{r,H}[\text{Dec}^{|H\rangle}(\text{sk}, \text{Enc}^{|H\rangle}(\text{pk}, m)) = m : \text{Gen}^{|H\rangle}(1^\lambda) \rightarrow (\text{pk}, \text{sk})] \geq \delta$;*
3. *it makes at most d_λ queries to H at each stage,*

then there exists an adversary Eve who outputs her guess m_E of m such that $\Pr[m_E = m] \geq \delta - \epsilon$ by making $O(d_\lambda^2 e_\lambda \text{poly}(\lambda)/\epsilon^2)$ queries.

Proof. Eve only needs to break the two-round QKA between Alice and Bob consisting of the following steps:

1. Alice runs $\text{Gen}^{|H\rangle}$ to produce pk, sk where sk is a classical string of $O(\log \lambda)$ length, and then sends pk as the first message π_A to Bob. We call Alice's algorithm at this stage \mathcal{A}_1 .

2. Upon receiving $\pi_A := \text{pk}$, Bob samples a uniformly random key k_B , computes $\text{ct} \leftarrow \text{Enc}^{(H)}(\text{pk}, k_B)$ and sends $\pi_B := \text{ct}$ to Alice. We call Bob's algorithm at this stage \mathcal{B} .
3. Upon receiving $\pi_B := \text{ct}$, Alice outputs her key $k_A := \text{Dec}^{(H)}(\text{sk}, \text{ct})$. We call Alice's algorithm at this stage \mathcal{A}_2 .

Let B denote the state of Bob after running \mathcal{B} , A denote the state of Alice after running \mathcal{A}_2 , and π denote (π_A, π_B) . Eve's strategy is as follows:

1. Run \mathcal{B} for t times and obtain state B_1, B_2, \dots, B_t .
2. For each $i \in \{0, 1\}^{O(\log \lambda)}$, Run \mathcal{A}_2 with $\text{sk} = i$ for t times and obtain state $A_1^{(i)}, A_2^{(i)}, \dots, A_t^{(i)}$.
3. Finally, apply the recovery channel in [Theorem 3.11](#) to generate a fake view \hat{B} of Bob B and then use \hat{B} to compute key k_E .

By setting $t = 2 \ln 2 \cdot e_\lambda d_\lambda / \epsilon^2$, we have $A\hat{B}$ is $O(\epsilon)$ -close to AB by [Lemma C.3](#) and [Lemma 2.3](#). Thus we have $\Pr[k_E = k] \geq \delta - \epsilon$ and Eve needs to make $t \cdot O(d_\lambda \text{poly}(\lambda)) = O(d_\lambda^2 e_\lambda \text{poly}(\lambda) / \epsilon^2)$ queries. \square

Remark C.2. Similar to the generalization from [Theorem 1.1](#) to [Theorem 1.2](#), [Theorem C.1](#) can be generalized to the case where the public key pk is quantum, but pk needs to be either pure or "efficiently clonable".

Lemma C.3. *There exists $q, q_1, q_2, \dots, q_{\text{poly}(\lambda)} \in [t]$ such that*

$$I(\text{sk}, A : B|E, \pi) \leq 2e_\lambda d_\lambda / t$$

where $E = (B_1, B_2, \dots, B_q, A_1^{(1)}, A_2^{(1)}, \dots, A_{q_1}^{(1)}, A_1^{(2)}, A_2^{(2)}, \dots, A_{q_2}^{(2)}, \dots, A_1^{(\text{poly}(\lambda))}, \dots, A_{q_{\text{poly}(\lambda)}}^{(\text{poly}(\lambda))})$.

Proof. By chain rule and the fact that sk is classical,

$$\begin{aligned} I(\text{sk}, A : B|E, \pi) &= I(\text{sk} : B|E, \pi) + I(A : B|E, \pi, \text{sk}) \\ &= I(\text{sk} : B|E, \pi) + \mathbb{E}_s I(A : B|E, \pi, \text{sk} = s). \end{aligned}$$

For the first part, observe that states B, B_1, \dots, B_t are permutation invariant w.r.t. sk, π and the rest of Eve. Then by [Lemma 4.2](#), there exists $q \in [t]$ such that $I(\text{sk} : B|\pi, \dots, B_1, B_2, \dots, B_q) \leq S(B)/(t+1) \leq 2e_\lambda d_\lambda / t$. For the second part, observe that conditioned on $\text{sk} = s$, states $A, A_1^{(s)}, \dots, A_t^{(s)}$ are permutation invariant w.r.t. π and the rest of Eve. Then by [Lemma 4.2](#), there exists $q_s \in [t]$ such that $I(A : B|\text{sk} = s, \pi, \dots, A_1^{(s)}, A_2^{(s)}, \dots, A_{q_s}^{(s)}) \leq S(A|\text{sk} = s)/(t+1) \leq 2e_\lambda d_\lambda / t$. Thus we have $I(\text{sk}, A : B|E, \pi) \leq 2e_\lambda d_\lambda / t$. \square