Semantic Security and Indistinguishability in the Quantum World

April 20, 2015

Tommaso Gagliardoni[1], Andreas Hülsing[2], and Christian Schaffner[3,4]

[1] CASED, Technische Universität Darmstadt, Germany
[2] TU Eindhoven, The Netherlands
[3] Institute for Logic, Language and Compuation (ILLC), University of Amsterdam, The Netherlands
[4] Centrum Wiskunde & Informatica (CWI) Amsterdam, The Netherlands

Abstract. At CRYPTO 2013, Boneh and Zhandry initiated the study of quantum-secure encryption. They proposed first indistinguishability definitions for the quantum world where the actual indistinguishability only holds for classical messages, and they provide arguments why it might be hard to achieve a stronger notion. In this work, we show that stronger notions are achievable, where the indistinguishability holds for quantum superpositions of messages. We investigate exhaustively the possibilities and subtle differences in defining such a quantum indistinguishability notion. We justify our stronger definition by showing their equivalence to novel quantum semantic-security notions that we introduce. Furthermore, we give a generic transformation to turn a big class of encryption schemes into quantum indistinguishable and hence quantum semantically secure ones.

1 Introduction

Quantum computers [NC00] threaten many cryptographic schemes. By using Shor's algorithm [Sho94] and its variants [Wat01], an adversary in possession of a quantum computer can break the security of every scheme based on factorization and discrete logarithms, including RSA, ElGamal, elliptic-curve primitives and many others. Moreover, longer keys and output lengths are required in order to maintain the security of block ciphers and hash functions [Gro96,BHT97]. These difficulties led to the development of *post-quantum cryptography* [BBD09], i.e., classical cryptography resistant against quantum adversaries.

When modeling the security of cryptographic schemes, care must be taken in defining exactly what property one wants to achieve. In classical security models, all parties and communications are classical. When these notions are used to prove *post-quantum* security, one must consider adversaries having access to a quantum computer. This means that, while the communication between the adversary and the user is still classical, the adversary might carry out computations on a quantum computer.

Such post-quantum notions of security turn out to be unsatisfying in certain scenarios. For instance, consider quantum adversaries able to use *quantum superpositions* of messages $\sum_x \alpha_x |x\rangle$ instead of classical messages when communicating with the user, even though the cryptographic primitive is still classical. This kind of scenario is considered, e.g., in [BZ13,DFNS13,Unr12,Wat06,Zha12]. Such a setting might for example occur in a situation where one party using a quantum computer encrypts messages for another party that uses a classical computer and an adversary is able to observe the outcome of the quantum computation before measurement. Other examples are an attacker which is able to trick a classical device into showing quantum behavior, or a classical scheme which is used as subprotocol in a larger quantum protocol. Notions covering such settings are often called *quantum-security* notions. In this work we propose new quantum-security notions for encryptions.

For encryption schemes, the notion of *semantic security* [Gol04] has been traditionally used. This notion models in abstract terms the fact that, without the corresponding decryption key, it is impossible not only to correctly decrypt a ciphertext, but even to recover any non-trivial information about the underlying plaintext. The exact definition of semantic security is cumbersome to work with in security proofs as it is simulation-based. Therefore, the simpler notion of *ciphertext indistinguishability* has been introduced. This notion is given in terms of an interactive game where an adversary has to distinguish the encryptions of two messages of his choice. The advantage of this definition is that it is easier to work with than (but equivalent to) semantic security.

To the best of our knowledge, no quantum semantic-security notions have been proposed so far. For indistinguishability, Boneh and Zhandry introduced indistinguishability notions for quantum-secure encryption under chosen-plaintext attacks in a recent work [BZ13]. They consider a model (IND-qCPA) where a quantum adversary can query the encrypting device in superposition during a learning phase, but is limited to classical communication during the actual challenge phase. However, this approach has the following shortcoming: If we assume that an adversary can get quantum access in a learning phase, it seems unreasonable to assume that he cannot get such access when the actual message of interest is encrypted. Boneh and Zhandry showed that a seemingly natural notion of quantum indistinguishability is unachievable. In order to restore a meaningful definition, they resorted to the compromise of IND-qCPA.

Our contributions. In this paper we achieve two main results. On the one hand, we initiate the study of semantic security in the quantum world, providing new definitions and a thorough discussion about the motivations and difficulties of modeling these notions correctly. This study is concluded by a suitable definition of *quantum semantic security under chosen plaintext attacks (qSEM-qCPA)*. On the other hand, we extend the fundamental work initiated in [BZ13] in finding suitable notions of indistinguishability in the quantum world. We show that the compromise that had to be reached there in order to define an achievable notion instead of a more natural one (i.e., IND-qCPA vs. fqIND-qCPA) can be overcome – although not trivially. We show how various other possible notions

of quantum indistinguishability can be defined. All these security notions span a tree of possibilities which we analyze exhaustively in order to find the most suitable definition of *quantum indistinguishability under chosen plaintext attacks (qIND-qCPA)*. We prove this notion to be achievable, strictly stronger than IND-qCPA, and equivalent to qSEM-qCPA, thereby completing an elegant framework of security notions in the quantum world.

Furthermore, we formally define the notion of a *core function* and *quasi–length-preserving ciphers* – encryption schemes which essentially do not increase the plaintext size, such as stream ciphers and many block ciphers including AES – and we show the impossibility of achieving our new security notion for this kind of schemes. While this impossibility might look worrying from an application perspective, we also present a transformation that turns a block cipher into an encryption scheme fulfilling our notion.

Related work. The idea of considering scenarios where a quantum adversary can force other parties into quantum behaviour has been considered in [DFNS13] where the authors study superposition attacks for multi-party computation, secret sharing, and zero-knowledge. The quantum security of zero-knowledge and zero-knowledge proofs of knowledge has been investigated in [Wat06] and [Unr12]. In [BZ13] the authors also consider the security of signature schemes where the adversary can have quantum access to a signing oracle. Quantum superposition queries have also been investigated relatively to the random oracle model [BDF+11]. Another quantum indistinguishability notion has been suggested (but not further analyzed) by Velema in [Vel13, Def. 5.3].

2 Preliminaries

In this section, we briefly recall the classical security notions for encryption schemes secure against chosen plaintext attacks (CPA). In addition, we revisit the two existing indistinguishability notions for the quantum world. We start by introducing notation we will use throughout the paper.

We say that a function $f : \mathbb{N} \to \mathbb{R}$ is *polynomially bounded* iff there exists a polynomial p and a value $\bar{n} \in \mathbb{N}$ such that: for every $n \geq \bar{n}$ we have that $f(n) \leq p(n)$; in this case we will just write $f = \mathrm{poly}(n)$. We say that a function $\varepsilon : \mathbb{N} \to \mathbb{R}$ is *negligible*, if and only if for every polynomial p, there exists an $n_p \in \mathbb{N}$ such that $\varepsilon(n) \leq \frac{1}{p(n)}$ for every $n \geq n_p$; in this case we will just write $\varepsilon = \mathrm{negl}(n)$. In this work, we focus on secret-key encryption schemes. However, the definitions can be easily extended to the public-key case. In all that follows we use $n \in \mathbb{N}$ as the security parameter.

Definition 2.1 (Secret-key encryption scheme [Gol04]). *A* secret-key encryption scheme *is a triple of probabilistic polynomial-time algorithms* $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *operating on a message space* $\mathcal{M} = \{0,1\}^m$ *(where* $m = poly(n) \in \mathbb{N}$*) that fulfills the following two conditions:*

1. *The key generation algorithm* $\mathsf{Gen}(1^n)$ *on input of security parameter* n *in unary, outputs a bitstring* k*.*

2. *For all k in the range of* $\mathsf{Gen}(1^n)$ *and any message* $x \in \mathcal{M}$*, the algorithms* Enc *(encryption) and* Dec *(decryption) satisfy* $\mathsf{Pr}[\mathsf{Dec}(k, \mathsf{Enc}(k, x)) = x] = 1$*, where the probability is taken over the internal coin tosses of* Enc *and* Dec*.*

We also write $\mathcal{K}$ for the range of $\mathsf{Gen}(1^n)$ (the key space) and $\mathsf{Enc}_k(x)$ for $\mathsf{Enc}(k, x)$.

2.1 Classical Security Notions: IND-CPA and SEM-CPA.

We turn to security notions for encryption schemes. In this work, we will only look at the notions of indistinguishable ciphertexts under adaptively chosen plaintext attacks (IND-CPA), and semantic security under adaptively chosen plaintext attacks (SEM-CPA), which are known to be equivalent (see, e.g., [Gol04]). In general these notions can be defined as a game between a challenger $\mathcal{C}$ and an adversary $\mathcal{A}$. First, $\mathcal{C}$ generates a legitimate key running $k \longleftarrow \mathsf{Gen}(1^n)$ which he uses throughout the game. The game starts with a first learning phase. A challenge phase follows where $\mathcal{A}$ receives a challenge. Afterwards, a second learning phase follows, and finally $\mathcal{A}$ has to output a solution. The learning phases define the type of attack, and the challenge phase the notion captured by the game. We give all our definitions by referring to this game framework and by defining a learning and a challenge phase.

The CPA learning phase: $\mathcal{A}$ is allowed to adaptively ask $\mathcal{C}$ for encryptions of messages of his choice. $\mathcal{C}$ answers the queries using key k. Note that this is equivalent to saying that $\mathcal{A}$ gets oracle access to an encryption oracle that was initialized with key k.

The IND challenge phase: $\mathcal{A}$ defines a challenge template consisting of two equal-length messages x_0, x_1, and sends it to $\mathcal{C}$. The challenger $\mathcal{C}$ samples a random bit $b \xleftarrow{\$} \{0,1\}$ uniformly at random, and replies with the encryption $\mathsf{Enc}_k(x_b)$. $\mathcal{A}$'s goal is to guess b.

The SEM challenge phase: $\mathcal{A}$ sends $\mathcal{C}$ a challenge template consisting of a circuit S_m specifying a distribution over m-bit long plaintexts[5], an advice function $h_m : \{0,1\}^m \to \{0,1\}^{\mathrm{poly}(n)}$ modeling side information, and a target function $f_m : \{0,1\}^m \to \{0,1\}^{\mathrm{poly}(n)}$ for an $m \in \mathbb{N}$ of $\mathcal{A}$'s choice. The challenger $\mathcal{C}$ replies with the pair $(\mathsf{Enc}_k(x), h_m(x))$ where x is sampled according to S_m. $\mathcal{A}$'s challenge is to output $f_m(x)$.

Definition 2.2 (IND-CPA). *A secret-key encryption scheme is called* IND-CPA secure *if the success probability of any probabilistic polynomial-time adversary winning the IND-CPA game is at most negligibly (in n) close to* $1/2$*.*

For the definition of semantic security, $\mathcal{A}$'s success probability is compared to that of a simulator $\mathcal{S}$ that plays in a *reduced game*: $\mathcal{S}$ gets no learning phase and during the challenge phase it only receives $h_m(x)$, not the ciphertext.

[5] Uniform random bits are used as input to the circuit S_m and the outputs are m-bit long plaintexts.

Definition 2.3 (SEM-CPA). *A secret-key encryption scheme is called* SEM-CPA secure *if for any probabilistic polynomial-time adversary $\mathcal{A}$ there exists a probabilistic polynomial-time simulator $\mathcal{S}$ such that the challenge templates produced by $\mathcal{S}$ and $\mathcal{A}$ are identically distributed and the success probability of $\mathcal{A}$ winning the SEM-CPA game is negligibly close to the success probability of $\mathcal{S}$ winning the reduced game.*

Semantic security models what we want an encryption scheme to achieve: An adversary given a ciphertext can learn nothing about the encrypted message which he could not also learn from his knowledge of the message distribution and possibly existing side-information (modeled by h_m). Indistinguishability of ciphertexts is an equivalent technical notion introduced to simplify proofs. A detailed formal definition of both notions can be found in Appendix A.

2.2 Previous Notions of Security in the Quantum World

Here we briefly recall the results from [BZ13] about quantum indistinguishability notions. We refer to [NC00] for commonly used notation and quantum information-theoretic concepts. Given security parameter n, let $\{\mathcal{H}_n\}_n$ be a family of complex Hilbert spaces such that $\dim \mathcal{H}_n = 2^{\mathrm{poly}(n)}$. We assume that $\mathcal{H}_n$ contains all the subspaces where the message states, the ciphertext states and any auxiliary state live. For the sake of simplicity we will not make a distinction when writing that a state $|\phi\rangle$ belongs to one particular subspace, and we will omit the index n when the security parameter is implicit, therefore writing just $|\phi\rangle \in \mathcal{H}$. We start by defining the following:

Definition 2.4 (Quantum Encryption Oracle ([BZ13])). *Let* Enc *be the encryption algorithm of a secret-key encryption scheme $\mathcal{E}$. We define the* quantum encryption oracle U_{Enc_k} *associated with $\mathcal{E}$ and initialized with key k as (a family of) unitary operators defined by:*

$$U_{\mathsf{Enc}_k} : \sum_{x,y} \alpha_{x,y} |x\rangle |y\rangle \mapsto \sum_{x,y} \alpha_{x,y} |x\rangle |y \oplus \mathsf{Enc}_k(x)\rangle$$

where the same randomness r is used in superposition in all the executions of $\mathsf{Enc}_k(x)$ within one query[6] *– for each new query, a fresh independent r is used.*

The first notion proposed in [BZ13] replaces all classical communication between $\mathcal{A}$ and $\mathcal{C}$ by quantum communication. $\mathcal{A}$ and $\mathcal{C}$ are now quantum circuits operating on quantum states, and sharing a certain number of qubits (the quantum communication register). The definition for the new security game is obtained from Definition 2.2 by changing the learning and challenge phases as follows:

Quantum CPA learning phase (qCPA): $\mathcal{A}$ gets oracle access to U_{Enc_k}.

[6] as shown in [BZ13], this is not restrictive.

Fully quantum IND challenge phase (fqIND): $\mathcal{A}$ prepares the communication register in the state $\sum_{x_0,x_1,y} \alpha_{x_0,x_1,y} |x_0\rangle |x_1\rangle |y\rangle$, consisting of two m-qubit states (the two input-message superpositions) and an ancilla state to store the ciphertext. $\mathcal{C}$ samples a bit $b \xleftarrow{\$} \{0,1\}$ and applies the transformation:

$$\sum_{x_0,x_1,y} \alpha_{x_0,x_1,y} |x_0\rangle |x_1\rangle |y\rangle \mapsto \sum_{x_0,x_1,y} \alpha_{x_0,x_1,y} |x_0\rangle |x_1\rangle |y \oplus \mathsf{Enc}_k(x_b)\rangle .$$

$\mathcal{A}$'s goal is to output b.

The resulting security notion in [BZ13] is called *indistinguishability under fully quantum chosen-message attacks (IND-fqCPA)*. We decided to rename it to *fully quantum indistinguishability under quantum chosen-message attacks (fqIND-qCPA)* in order to fit into our naming scheme: It consists of a quantum CPA learning phase and a fully quantum IND challenge phase.

Definition 2.5 (fqIND-qCPA). *A secret-key encryption scheme is called* fqIND-qCPA secure *if the success probability of any quantum probabilistic polynomial-time adversary winning the game defined by the qCPA learning phase and the fqIND challenge phase above is at most negligibly close (in n) to* $1/2$.

As already observed in [BZ13], this notion is unachievable. The separation by Boneh and Zhandry exploits the entanglement of quantum states, namely the fact that entanglement can be created between plaintext and ciphertext.

Theorem 2.6. *[BZ attack ([BZ13], Theorem 4.2)] No symmetric-key encryption scheme can achieve fqIND-qCPA security.*

Proof. The attack works as follows: The adversary $\mathcal{A}$ chooses as challenge messages the states $|0^m\rangle$ and $H|0^m\rangle$ (where H denotes the m-fold tensor Hadamard transform), i.e. he prepares the register in the state $\sum_x \frac{1}{2^{m/2}} |0^m, x, 0^m\rangle$. When the challenger $\mathcal{C}$ performs the encryption, we can have two cases:

- if $b = 0$, i.e. the first message state is chosen, the state is transformed into

$$\sum_x \frac{1}{2^{m/2}} |0^m, x, \mathsf{Enc}_k(0^m)\rangle = |0^m\rangle \otimes H|0^m\rangle \otimes |\mathsf{Enc}_k(0^m)\rangle ;$$

- if $b = 1$, i.e. the second message state is chosen, the state is transformed into

$$\sum_x \frac{1}{2^{m/2}} |0^m, x, \mathsf{Enc}_k(x)\rangle = |0^m\rangle \otimes \sum_x \frac{1}{2^{m/2}} |x, \mathsf{Enc}_k(x)\rangle .$$

Notice that in the second case we have a fully entangled state between the second and the third register. At this point, $\mathcal{A}$ does the following:

1. measures (traces out) the third register;
2. applies again H to the second register;
3. measures the second register;
4. outputs $b' = 1$ iff the outcome of this last measurement is 0^m, else outputs 0.

In fact, if $b = 0$, then the second register is left untouched: By applying again the Hadamard transformation it will be reset to the state $|0^m\rangle$, and a measurement on this state will yield 0^m with probability 1. If $b = 1$ instead, tracing out one half of a fully entangled state results in a complete mixture in the second register. Applying a Hadamard transform and measuring in the computational basis necessarily gives a fully random outcome, and hence outcome 0^m only with probability $\frac{1}{2^m}$, which is negligible in n, because $m = \text{poly}(n)$. □

This attack implies that the fqIND-qCPA notion is too strong. In order to weaken it, the following notion of indistinguishability under adaptively chosen quantum plaintext attacks was introduced:

Definition 2.7 (IND-qCPA ([BZ13])). *A secret-key encryption scheme is called* IND-qCPA secure *if the success probability of any quantum probabilistic polynomial-time adversary winning the game defined by the qCPA learning phase and the classical IND challenge phase is at most negligibly close (in n) to* $1/2$.

In this definition, the CPA queries are allowed to be quantum, but the challenge query is required to be classical. It has been shown that IND-qCPA is strictly stronger than IND-CPA:

Theorem 2.8. *[([BZ13],Theorem 4.8)] There exists an encryption scheme $\mathcal{E}$ which is IND-CPA secure, but not IND-qCPA secure.*

3 New Notions of Quantum Indistinguishability

IND-qCPA might be viewed as classical indistinguishability (IND) under a quantum chosen plaintext attack (qCPA). The authors in [BZ13] resorted to this definition in order to overcome their impossibility result on one seemingly natural notion of quantum indistinguishability (fqIND-qCPA) which turned out to be too strong. This raises the question whether IND-qCPA is the only possible quantum indistinguishability notion (and hence no classical encryption scheme can achieve indistinguishability of ciphertext superpositions) or if there exists a stronger notion which can be achieved.

In this section we show that by defining fqIND-qCPA, there are many choices which are made implicitly, and that on the other hand there exist other possible quantum indistinguishability notions. We discuss these choices spanning a binary 'security tree' of possible notions. Afterwards, we obtain a small set of candidate notions, eliminating those that are either ill-posed or unachievable because of the BZ attack, described in Theorem 2.6.

In all these notions, we implicitly assume 'quantum CPA learning phases', as in the case of IND-qCPA. However, we limit the discussion in this section to the design of a quantum challenge phase. In the end, we choose a suitable 'qIND-'notion amongst the possible ones we present in this section.

3.1 The 'Security Tree'

To define a general notion of indistinguishability in the quantum world, we have to consider many different distinctions for possible candidate models. For example, can we rule out certain forms of entanglement? How? Does the adversary have complete control over the challenger device? Each of these distinctions leads to a fork in a 'security-model binary tree'. We analyze every 'leaf' of the tree. Some of them lead to unreasonable or ill-posed models, some of them yield unachievable security notions, and others are analyzed in more detail.

Game model: oracle ($\mathcal{O}$) vs. challenger ($\mathcal{C}$). This distinction decides how the game, and especially the challenge phase, are implemented. In the classical world, the following two cases are equivalent but in the quantum world it turns out that they differ. In the *oracle* model, the adversary $\mathcal{A}$ gets oracle access to encryption and challenge oracles, i.e., he plays the game by performing calls to unitary gates $\mathcal{O}_1, \ldots, \mathcal{O}_q$. In this case $\mathcal{A}$ is modeled as a quantum circuit which implements a sequence of unitary gates $U_0, \ldots, U_q$, intertwined by calls to the $\mathcal{O}_i$'s. Given an input state $|\phi\rangle$, the adversary therefore computes the state:

$$U_q \mathcal{O}_q \ldots U_1 \mathcal{O}_1 U_0 \, |\phi\rangle \, .$$

The *structure* of the *oracle* gates $\mathcal{O}_i$ itself is unknown to $\mathcal{A}$, who is only allowed to apply them in a black-box way. The fqIND notion uses this model.

In what we call the *challenger* model instead, the game is played against an *external (quantum) challenger*. Here, $\mathcal{A}$ is a quantum circuit which shares a quantum register (the communication channel) with another quantum circuit $\mathcal{C}$. The main difference is that in this case we can also consider what happens if $\mathcal{C}$ has additional input or output lines out of $\mathcal{A}$'s control. This also covers the case of 'unidirectional' state transmission, i.e., when qubits are sent over a quantum channel to an external entity, and they are not available afterwards until the entity sends them back.

At first glance, the ($\mathcal{O}$) model intuitively represents the scenario where $\mathcal{A}$ has almost complete control of some encryption device, whereas the ($\mathcal{C}$) model is more suited to a 'network' scenario where $\mathcal{A}$ wants to compromise the security of some external target.

Plaintexts: quantum states (Q) vs. classical description (c). In the (Q) model, the two m-qubit plaintexts chosen by $\mathcal{A}$ for the challenge template can be arbitrary superpositions of basis elements and can be potentially entangled with each other and other states. In the (c) model $\mathcal{A}$ is only allowed to choose *classical descriptions* of two m-qubit quantum states, encoded as (poly-sized) binary strings describing, for example, the quantum circuit needed to obtain such states starting from the $|0^m\rangle$ state. In the latter case, $\mathcal{A}$ is only allowed to send classical information to $\mathcal{C}$: the challenger $\mathcal{C}$ will read the states' descriptions and will build one of the two states depending on his challenge bit b.

In classical models, there is no difference between sending a description of a message or the message itself. In the quantum case, there is a big difference between these two cases as the message case allows $\mathcal{A}$ to establish entanglement

of the message(s) with other registers. This is not possible in case of classical descriptions. It might intuitively appear that the (Q) model (considered for the fqIND-qCPA notion) is more natural. However, the (c) scenario models the case where $\mathcal{A}$ is well aware of the message that is encrypted, but the message is not constructed by $\mathcal{A}$ himself. Giving $\mathcal{A}$ the ability to choose the challenge messages for the IND game models the worst case that might happen: $\mathcal{A}$ knows that the ciphertext he receives is the encryption of one out of the two messages that he can distinguish best. This closely reflects the intuition behind the classical IND notions: in that game, the adversary is allowed to send the two messages not because in the real world he would be allowed to do so, but because we want to achieve security even for the worst possible choice of messages. Notice that the (c) model is in fact equivalent to the (Q) model with the additional assumption that the transmitted states are not entangled[7].

Relaying of plaintext states: Yes (Y) vs. No (n). If $\mathcal{C}$ is *not relaying* (n), this means that the two plaintext states chosen by $\mathcal{A}$ will not be 'sent back' to $\mathcal{A}$ (in other words: their registers will not be available anymore to $\mathcal{A}$ after the challenge encryption). In circuit terms, this means that at the beginning of the game, $\mathcal{C}$ will have (one or two) ancilla registers in his internal (private) memory. During the encryption phase, $\mathcal{C}$ will swap these register(s) with the content of the original plaintext register(s), hence transferring their original content outside of $\mathcal{A}$'s control.

If the challenger is relaying (Y) instead, this means that the two plaintext states will be left in the original register (or channel), and may be possibly accessed by $\mathcal{A}$ at any moment. This is the model considered for fqIND.

Again, the (Y) case is more fitting to those cases where $\mathcal{A}$ 'implements locally' the encryption device and has almost full control of it, whereas the (n) case is more appropriate when the game is played against some external entity which is not under $\mathcal{A}$'s control. This is a rather natural assumption, for example, when states are sent over some quantum channel and not returned.

Type of unitary transformation: (1) vs. (2). In quantum computing, the 'canonical' way of evaluating a function $f(x)$ in superposition is by using an auxiliary register:

$$\sum_{x,y} |x, y\rangle \mapsto \sum_{x,y} |x, y \oplus f(x)\rangle \,.$$

This way ensures that the resulting operator is invertible, even if f is not. This is what we call *type-*(1) *transformations*: if Enc_k is an encryption mapping m-bit plaintexts to ℓ-bit ciphertexts, the resulting operator in this case will act on $m + \ell$ qubits in the following way:

$$\sum_{x,y} \alpha_{x,y} |x, y\rangle \mapsto \sum_{x,y} \alpha_{x,y} |x, y \oplus \mathsf{Enc}_k(x)\rangle \,,$$

[7] Actually the (c) model is limited to those states having a poly-size classical representation, whereas the (Q) model allows for arbitrary states. But since any plaintext state must eventually be computed by an efficient (i.e., poly-sized) quantum circuit, this restriction is already implied in any computationally meaningful notion.

where the y's are ancillary values. This approach is also used for fqIND.

In our case, though, we do not consider arbitrary functions, but encryptions, which behave as *bijections* on some bit-string spaces (assuming that the randomness is treated as an input.) Therefore, provided that the encryption does not change the size of a message, the following transformation is also invertible:

$$\sum_x \alpha_x |x\rangle \mapsto \sum_x \alpha_x |\mathsf{Enc}_k(x)\rangle .$$

For the more general case of arbitrary message expansion factors, we will consider transformations of the form:

$$\sum_{x,y} \alpha_{x,y} |x, y\rangle \mapsto \sum_{x,y} \alpha_{x,y} |\phi_{x,y}\rangle ,$$

where the length of the ancillary register is $|y| = |\mathsf{Enc}_k(x)| - |x|$ and $\phi_{x,0} = \mathsf{Enc}_k(x)$ for every x – i.e., initializing the ancillary y register in the $|0\rangle$ state produces a correct encryption, which is what we expect from an honest quantum executor. This is what we call *type-*(2) *transformations.*

In quantum computing, it is often implicitly assumed that there is no difference between the two transformation types, but since we possibly consider non-relaying scenarios we will take into account this distinction. Whether we are in the (1) or in the (2) scenario depends on the device architecture, although—as we will see—the two models are equivalent under certain assumptions.

Ancillary y values: initialized by $\mathcal{A}$ (I) or chosen externally (e). In the case that the unitary transformation requires an ancillary state (i.e., $|y\rangle$) for storing the value of the encryption, is this register initialized by $\mathcal{A}$ (I, for 'internal') or by someone else (e, for 'external'), i.e. $\mathcal{C}$? In the latter case we assume that y is set to $|0\rangle$ (because $\mathcal{C}$ executes the protocol honestly), but if $\mathcal{A}$ is allowed to initialize the register, then he might try to inject an ancilla state which is nonzero – potentially entangled with another state he keeps. The (I) case (also considered in fqIND) leads to a very strong model where it is assumed that $\mathcal{A}$ has almost complete control over the encryption device.

3.2 Analysis of the models

By considering these 5 distinctions in the security tree we have 32 possible candidate models to analyze. We label each of these candidate models by appending each one of the 5 labels of every tree branch in brackets. Clearly, 32 different definitions of quantum indistinguishability is too much, but luckily most of these are unreasonable or unachievable. To start with, we can ignore the following:

Leaves of the form ($\mathcal{O}c\ldots$). In the $\mathcal{O}$ scenario, the oracle is actually a quantum gate inside $\mathcal{A}$'s quantum circuitry. Therefore $\mathcal{A}$ has complete faculty of querying the oracle on a superposition of states, possibly entangled with other registers kept by $\mathcal{A}$ itself.

Leaves of the form $(\mathcal{O}Qn\ldots)$. Again, the oracle is here a gate, which has no internal memory to store and keep the plaintext states sent by $\mathcal{A}$.

Leaves of the form $(\mathcal{O}\ldots e)$. Since in this case the transformation is implemented through an oracle (which is a circuit gate), all the ancillary states must be initialized by $\mathcal{A}$.

Leaves of the form $(\ldots Y\ldots 2\ldots)$. Relaying is not taken into account in type-(2) transformations. In these transformations, to some extent, one of the two plaintext registers is *always* relayed (after having been 'transformed' into a ciphertext). If the other plaintext was to be relayed as well, this would immediately compromise indistinguishability (because one of the two states would be modified and the other not, and both of them would be handed over to $\mathcal{A}$).

This leaves us with 13 models, but it is easy to see that 5 of them are unachievable because of the attack from Theorem 2.6. This is the case for $(\mathcal{O}QY1I)$ (which is exactly fqIND-qCPA), $(\mathcal{C}QY1I)$, $(\mathcal{C}QY1e)$, $(\mathcal{C}cY1I)$, and $(\mathcal{C}cY1e)$.

Furthermore, option (I) is too strong for a general notion. The reason is that allowing $\mathcal{A}$ to initialize the ancilla registers of $\mathcal{C}$ models some kind of fault attack. By injecting entangled states in the ancilla register, $\mathcal{A}$ might possibly be able to 'watermark' one of the plaintexts without $\mathcal{C}$ having any possibility of checking. We know that, even in the classical world, fault attacks can compromise indistinguishability (see e.g. [BS03]), and they go well beyond the scope of a general indistinguishability notion. For this reasons, we leave analyzing the (I) cases as an open question for further research.

We are now left with 4 possible candidate models of quantum indistinguishability (qIND) to analyze. Since all of them are of the form $(\mathcal{C}\ldots n\ldots e)$, from now on we will omit the $\mathcal{C}, n, e$ notation. The differences between them are:

- In the two $(Q.)$ models, the encrypting device is 'input reactive', in the sense that it accepts as input external quantum states. In the $(c.)$ models, instead, the device is 'input-autonomous', i.e., it uses its own internal input states, but the adversary is still able to select the most favourable case.
- In the two $(.1)$ models, the (classical version of the) device creates a copy of the (classical) input before encrypting it. When translating this to the quantum behavior it would mean that the device produces the ciphertext state by entangling an ancilla register with the plaintext state. In the $(.2)$ models, instead, the (classical version of the) encrypting device works directly on the input variable, without creating copies.

Spoiler. We show in Section 6.1 that *none* of these notions is achievable for encryption schemes which do not increase the message size (ignoring the randomness). So, for example, classical block ciphers like AES are *not* secure under any of the above IND-notions, and the best one they can achieve is IND-qCPA.

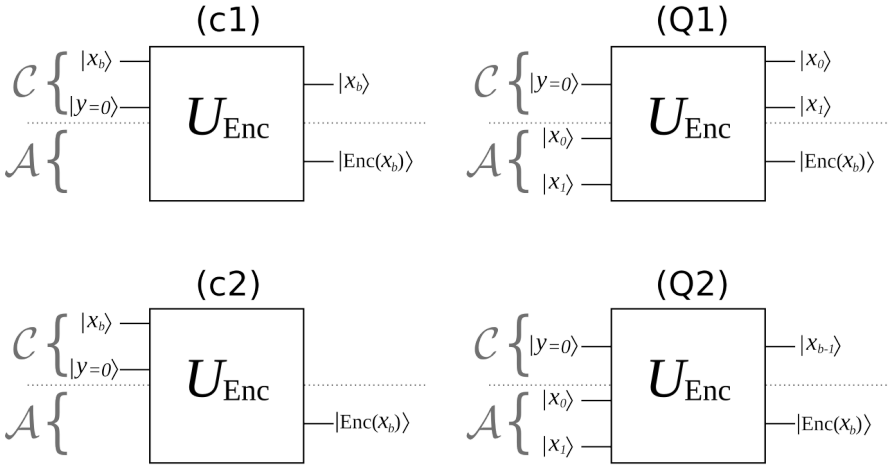

Fig. 1. The four candidate quantum indistinguishability notions. For the (c.) notions, the picture omits the part where the adversary sends classical descriptions of $|x_0\rangle, |x_1\rangle$.

3.3 Our Choice

Our choice for a suitable notion of quantum indistinguishability in the quantum world is ($c2$)IND-qCPA, and from now on we will denote this security notion as *qIND-qCPA*. Before explaining the reasons for this choice, we summarize the resulting qIND challenge phase.

Quantum IND challenge phase (qIND): $\mathcal{A}$ prepares a challenge template, choosing two superpositions of plaintexts $\sum_x \alpha_{0,x} |x\rangle$ and $\sum_x \alpha_{1,x} |x\rangle$. Then he sends $\mathcal{C}$ the challenge template consisting of a classical description of these two states. $\mathcal{C}$ samples a bit b and replies to $\mathcal{A}$ with the state $\sum_x \alpha_{b,x} |\mathsf{Enc}_k(x)\rangle$. $\mathcal{A}$'s goal is to output b.

Using this challenge phase and the notion of a qCPA learning phase, we define qIND-qCPA as follows.

Definition 3.1 (qIND-qCPA). *A secret-key encryption scheme is said to be qIND-qCPA-secure if the success probability of any quantum probabilistic polynomial time adversary winning the game defined by the qCPA learning phase and the qIND challenge phase above is at most negligibly close (in n) to* $1/2$.

Since we consider type-2 transformations, we will sometimes abuse notation and use U_{Enc_k} to denote the type-2 operator relative to Definition 2.4.

The main reason for our choice is that, amongst the 4 models left, this is the only one where, in any case, during a challenge query $\mathcal{A}$ always receives back a *pure state* instead of a *mixed state* (this is easy to check by noting that the ciphertext state received cannot be entangled with anything else). In all the other cases the challenger is left with one or more registers which are not forwarded to $\mathcal{A}$, but might potentially be entangled with the encrypted state $\mathcal{A}$ receives. It

is important to note that this does not automatically lead to attacks. For example, it is not clear how to exploit this entanglement for a BZ-like attack (along the lines of Theorem 2.6), so all these notions *might* potentially be achievable. However, from an application perspective, these remaining models are hard to justify: we would have to assume that the challenger stores the 'unforwarded' registers in an internal, noiseless quantum memory, for an unspecified period of time (until $\mathcal{A}$ terminates the attack at the very least.) In fact, if $\mathcal{C}$ traces out (i.e., measures) these registers, the state $\mathcal{A}$ receives might be disturbed because of the entanglement. It is true that $\mathcal{A}$ has no way of detecting this 'remote measurement' by observing his state only (this would allow $\mathcal{A}$ and $\mathcal{C}$ to communicate faster-than-light), but the effect might become evident in scenarios where the two registers are reunited at a later time—for instance, if the encryption is used in a multi-stage protocol.

An additional criticality about the (Q) model is that allowing $\mathcal{A}$ to directly feed quantum states to $\mathcal{C}$ reminds of the 'quantum watermarking' issue we discussed when we decided to discard the (I) model. There is no way for $\mathcal{C}$ to detect whether $\mathcal{A}$ is feeding states entangled with other external registers or not. This might potentially give $\mathcal{A}$ additional power in a sense which should not be captured by a meaningful indistinguishability notion.

The $(c2)$ model avoids all these difficulties, and reductions to and from other notions, as well as proofs of security, are easier to treat. On the motivational side, we have already made clear that this model is no more restrictive than the other 3. Using type-2 transformations allows us to reduce qIND-qCPA to a reasonable notion of quantum semantic security, which would be problematic otherwise for technical reasons which will become clear in the proof of Theorem 5.4. We remark that the more traditionally used type-1 transformations actually arise from the need of modelling non-invertible functions as unitary (invertible) operators. From this point of view, they are actually a necessary artifice, which we can (and should) avoid in the case of permutations: in that case, we believe that type-2 transformations are actually a more natural choice.

Finally, notice that type-1 and type-2 transformations are equivalent (in the sense that their circuits can be converted to each other) under the assumption that one is also given oracle access to a *decryption oracle* U_{Dec_k}. A type-2 transformation, in fact, provides oracle access to both, encryption (via U_{Enc_k}) and decryption (via $U^{\dagger}_{\mathsf{Enc}_k}$), while the conjugate adjoint (inverse) of a type-1 encryption operator is *not* a type-1 decryption operator in general. However, if we have oracle access to both encryption and decryption type-1 operators, then we can simulate a type-2 operator by using ancilla qubits and 'uncomputing' the resulting garbage lines (see Figure 2). Under this perspective, type-2 transformations are a natural choice also when taking into account *quantum CCA security*, which we leave as an interesting topic to explore.

In any case however, notice that the security of Construction 6.4 likely remains unaffected by the choice of any of these last 4 models, because Theorem 6.5 does not rely on the adversary receiving a pure state or not.

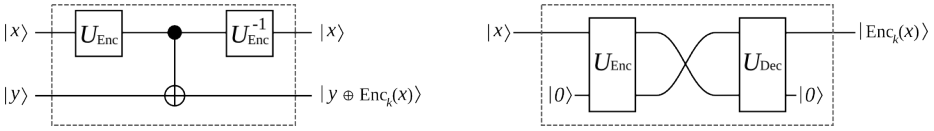

Fig. 2. Equivalence between type-1 and type-2 in the case of 1-qubit messages. Left: building a type-1 encryption oracle by using a type-2 encryption oracle (and its inverse) as a black-box. Right: building a type-2 encryption oracle by using type-1 encryption and decryption oracles as black-boxes.

4 New Notions of Semantic Security in the Quantum World

In this section, we initiate the study of suitable definitions of semantic security in the quantum world. As in the classical case, we will be able to show equivalence to different notions of quantum indistinguishability, which are easier to work with. So these definitions actually describe the *semantics* of the equivalent IND notions. As in the classical case, we present these notions in the non-uniform model of computation.

In what follows, we only look at definitions for a quantum SEM challenge phase. We implicitly assume that the adversary has access to a quantum encryption oracle during the 'qCPA learning phase', as in Definition 2.4. In the end, we give a definition for quantum semantic security under quantum chosen-plaintext attacks (qSEM-qCPA).

4.1 Classical Semantic Security under Quantum CPA

As a first notion of semantic security in the quantum world, we consider what happens if, like in the IND-qCPA notion, we stick to the classical definition but we allow for a quantum chosen-plaintext-attack phase. The definition uses a SEM-qCPA game that is obtained by combining qCPA learning phases with a classical SEM challenge phase as defined in Section 2. As in the classical case, $\mathcal{A}$'s success probability is compared to that of a simulator $\mathcal{S}$ that plays in a reduced game: $\mathcal{S}$ gets no learning phase and during the challenge phase it only receives the advice $h_m(x)$, not the ciphertext.

Definition 4.1. *[SEM-qCPA] A secret-key encryption scheme is called SEM-qCPA-secure if for every quantum polynomial-time machine $\mathcal{A}$, there exists a quantum polynomial-time machine $\mathcal{S}$ such that the challenge templates produced by $\mathcal{S}$ and $\mathcal{A}$ are identically distributed and the success probability of $\mathcal{A}$ winning the SEM-qCPA game is negligibly close (in n) to the success probability of $\mathcal{S}$ winning the reduced game.*

Spoiler. As in the classical case of IND-CPA and SEM-CPA, it is easy to show that this notion of semantic security is equivalent to IND-qCPA. Therefore it is strictly stronger than classical SEM-CPA, see Section 5.

4.2 Semantic Security with Quantum Advice States

Before looking at stronger notions of quantum semantic security, we consider for didactical purposes the case where the messages (and the function to be computed about the message) are still classical, but the auxiliary advice can be a quantum state.

A first approach we try is the following: Let U_{ξ_m} be a unitary (the *advice unitary*) that takes as input a basis element $|x\rangle$ representing a classical m-bit message x as well as (if required) an auxiliary register prepared by $\mathcal{C}$ and computes a quantum advice state $|\xi_m\rangle$. Then we can define the following challenge phase and the corresponding notion.

Quantum-advice SEM challenge phase(qaSEM): $\mathcal{A}$ sends $\mathcal{C}$ a challenge template consisting of: a poly-sized classical circuit S_m specifying a distribution of m-bit plaintexts, a classical description of the advice unitary U_{ξ_m}, and a target function $f_m : \{0,1\}^m \to \{0,1\}^{\mathrm{poly}(n)}$ for an $m \in \mathbb{N}$ of $\mathcal{A}$'s choice. $\mathcal{C}$ replies with the pair $(\mathsf{Enc}_k(x), |\xi_m\rangle)$, where x is sampled according to S_m and $|\xi_m\rangle$ is computed by constructing and evaluating U_{ξ_m}. $\mathcal{A}$'s goal is to output $f_m(x)$.

Definition 4.2. *[qaSEM-qCPA] A secret-key encryption scheme is said to be* quantum advice, classically semantically secure under quantum chosen plaintext attacks *(qaSEM-qCPA) if for every quantum polynomial-time machine $\mathcal{A}$, there exists a quantum polynomial-time machine $\mathcal{S}$ such that the challenge templates produced by $\mathcal{S}$ and $\mathcal{A}$ are identically distributed and the success probability of $\mathcal{A}$ winning the qaSEM-qCPA game is negligibly close (in n) to the success probability of $\mathcal{S}$ winning the reduced game.*

At a first glance it might seem as if qaSEM-qCPA is equivalent to SEM-qCPA as a security notion because having a classical advice function $h(x)$ is just a special case of a quantum advice circuit depending on x. Notice however that as we restrict U_{ξ_m} to be a circuit computing a unitary operator $U|x\rangle$ this notion is meaningless because it is trivially achievable by *any* encryption scheme. The reason is that, in this case, both $\mathcal{A}$ and $\mathcal{S}$ can always apply U^{-1} to $|\xi_m\rangle$ to recover the message – it is like restricting the classical notion to the case where the advice function h is just a permutation *chosen* by $\mathcal{A}$ (resp. $\mathcal{S}$).

To fix this, we have to allow more general quantum circuits U'_{ξ_m} that can provide somehow non-reversible information, for example by applying some partial measurement at the end, or by providing $\mathcal{A}$ (resp. $\mathcal{S}$) only with *some* output qubits, while $\mathcal{C}$ keeps the others. Towards this end let U'_{ξ_m} be an arbitrary quantum circuit (the *advice circuit*) that takes as input a basis element $|x\rangle$ representing a classical m-bit message x, a quantum state ρ_m provided by $\mathcal{A}$ (resp. $\mathcal{S}$) (that includes possibly needed auxiliary registers), and computes a quantum advice state ξ_m. This leads to the following definition:

Ideal quantum advice, classical SEM challenge phase (iqSEM): $\mathcal{A}$ sends $\mathcal{C}$ a challenge template consisting of: a poly-sized classical circuit S_m specifying a distribution of m-bit plaintexts, a classical description of the quantum advice circuit U'_{ξ_m}, a quantum state ρ_m, and a target function $f_m : \{0,1\}^m \to \{0,1\}^{\mathrm{poly}(n)}$

for an $m \in \mathbb{N}$ of $\mathcal{A}$'s choice. $\mathcal{C}$ replies with the pair $(\mathsf{Enc}_k(x), \xi_m)$, where x is sampled according to S_m and ξ_m is computed by constructing and executing U'_{ξ_m}. $\mathcal{A}$'s goal is to output $f_m(x)$.

The iqSEM-qCPA game is defined by qCPA learning phases and a iqSEM challenge phase. This leads to the following definition:

Definition 4.3. *[iqSEM-qCPA] A secret-key encryption scheme is said to be* ideally quantum advice, classically semantically secure under quantum chosen plaintext attacks (iqSEM-qCPA) *if for every quantum polynomial-time machine $\mathcal{A}$, there exists a quantum polynomial-time machine $\mathcal{S}$ such that the challenge templates produced by $\mathcal{S}$ and $\mathcal{A}$ are identically distributed and the success probability of $\mathcal{A}$ winning the iqSEM-qCPA game is negligibly close (in n) to the success probability of $\mathcal{S}$ winning the reduced game.*

This notion turns out to be equivalent to SEM-qCPA (and IND-qCPA). The reason is that having a quantum advice state does not really give any additional power to $\mathcal{A}$ in the case of classical messages and target functions. This can be seen from the reduction between IND-qCPA and SEM-qCPA – see the proofs of Propositions 5.2 and 5.3 below. In one case, the advice state is only used to pass $\mathcal{A}$'s code from the first circuit of $\mathcal{S}$ to the second one (which can also be done with a quantum advice state), in the other case it is set to a constant function.

It seems like introducing arbitrary quantum advice circuits (as opposed to *superpositions of classical advices*) makes little sense as long as the messages are still classical. Consequently, we proceed with our search for a notion of quantum semantic security considering *quantum superpositions* of messages.

4.3 Quantum Semantic Security

When considering semantic security for quantum messages, we mean (as in the qIND-qCPA case), that we can have *superpositions of messages* of the form $|\phi\rangle = \sum_x \alpha_x |x\rangle$. In this case, by *message information function*, we might either mean a *classical function* $f_n : \mathcal{H} \to \{0,1\}^{\mathrm{poly}(n)}$, mapping quantum states to classical bitstrings, or we can mean a *unitary operator* $U_n : \mathcal{H} \to \mathcal{H}$, mapping quantum states to quantum states. However, since extracting any classical information about a quantum state can be modeled as a measurement on that state, we will (at first) restrict to the latter meaning.

In this case, though, we cannot require a QPT machine to *exactly* compute a certain state $U_n |\phi\rangle$ for a particular plaintext state $|\phi\rangle$: it would be sufficient for the machine to output a state $|\psi\rangle$ arbitrarily close to $U_n |\phi\rangle$. We will hence introduce the dependancy on an *accuracy threshold parameter* ε, for some norm $\|.\|$ on $\mathcal{H}$. Notice that in general the dependancy on ε can be 'unloaded' on the security parameter, e.g. $\varepsilon = \frac{1}{n}$.

A first tentative, seemingly natural definition of a quantum SEM challenge phase would be the following:

Fully quantum SEM (fqSEM) challenge phase: $\mathcal{A}$ sends $\mathcal{C}$ a challenge template consisting of: a quantum circuit S_m (including auxiliary quantum registers, if needed) which takes as input classical randomness and outputs an m-qubit quantum message according to a distribution X_m with poly-sized support, an advice circuit U_{ξ_m} (including auxiliary quantum registers, if needed) that takes as input a quantum message state from X_m and outputs a poly-sized quantum state, and a target unitary operator $U_{f,m} : \mathcal{H} \to \mathcal{H}$, for an $m \in \mathbb{N}$ of $\mathcal{A}$'s choice. $\mathcal{C}$ replies with the pair $(U_{\mathsf{Enc}_k}(\rho), U_{\xi_m}(\rho))$, where ρ is sampled from X_m according to $\mathcal{C}$'s randomness. $\mathcal{A}$'s goal is to output a state ψ such that $\|\psi - U_{f,m}(\rho)\| \leq \varepsilon$.

Notice how there are many problems with this definition:

1. The states in X_m can be arbitrary (i.e., possibly entangled). This would lead to problems already discussed in the previous section.
2. It has not been defined which advice operators are allowable ones. For instance, generating entanglement between advice and message state would lead to similar problems as in Definition 4.3.
3. Since the state information operator $U_{f,m}$ is a unitary chosen by $\mathcal{A}$, it can be inverted - and therefore ignored. The definition could thus be reformulated by just requiring $\mathcal{A}$ to output a state arbitrarily close to $|\phi\rangle$ itself, which does not capture the idea of semantic security representing the impossibility of learning *any* interesting (possibly partial) information about the plaintext.

We will solve the first problem as in the qIND-qCPA case: by requiring that the adversary is only allowed to choose classical descriptions of quantum states – the same motivation for this choice holds in this case. In particular, allowable plaintext states are pure states of the form $\sum_x \alpha_x |x\rangle$.

The second problem is also easily solvable by recalling that here we are not modeling security for arbitrary quantum protocols, but instead for classical protocols that are somehow 'forced' to behave quantumly. Namely, since the advice in the classical case is modeled by a (classical) advice function h_m, it makes sense here to restrict ourselves to advice circuits which compute quantum superpositions of values of h_m. This solves the problem at the root, because now the adversary only has to specify a classical function h_m, and then the challenger will compute the advice as a superposition of values of h_m itself in some reversible way. In order to ease the notion we will introduce:

– a unitary operator defined by $U^{\mathcal{A}}_{k,h,m} : \sum_x \alpha_x |x\rangle |0\rangle \mapsto \sum_x \alpha_x |\mathsf{Enc}_k(x), h_m(x)\rangle$ (notice that it is still reversible because Enc_k is an invertible function, the same reasoning we used to justify type-2 transformations in Section 3);
– a density operator $\rho_{h,m} : \sum_x \alpha_x |x\rangle \mapsto \sum_x \|\alpha_x\|^2 |h_m(x)\rangle\langle h_m(x)|$ (notice that ρ is the density operator of $U^{\mathcal{A}}_{k,h,m}$ reduced to the second register).

Finally, concerning issue number 3, recall that the choice of replacing the (classical) message information function with a unitary operator arose from the need of modeling the extraction of arbitrary information from the quantum state. Again in this case, we are more focused on the extraction of superpositions of classical information about the states. Namely, we will also restrict to classical

message functions f_m chosen by the adversary, to be computed in superposition in some reversible way. We will thus define the following unitary operator:

$$U_{f,m} : \sum_x \alpha_x |x\rangle |0\rangle \mapsto \sum_x \alpha_x |x, f_m(x)\rangle .$$

This leads us to the following definition.

Mixed-State quantum SEM (mqSEM) challenge phase: $\mathcal{A}$ sends $\mathcal{C}$ a challenge template consisting of a classical circuit S_m generating a distribution X_m (with poly-sized support) of classical descriptions of m-qubit quantum message states in $\mathcal{H}$, an advice function $h_m : \{0,1\}^m \to \{0,1\}^{\mathrm{poly}(n)}$, and a target function $f_m : \{0,1\}^m \to \{0,1\}^{\mathrm{poly}(n)}$ for an $m \in \mathbb{N}$ of $\mathcal{A}$'s choice. $\mathcal{C}$ replies with $U^{\mathcal{A}}_{k,h,m} |\phi\rangle$, where $|\phi\rangle$ is sampled according to X_m. $\mathcal{A}$'s goal is to output $|\psi\rangle$ such that $\||\psi\rangle - U_{f,m} |\phi\rangle\| \leq \varepsilon$.

In the reduced game corresponding to mqSEM, the simulator, instead of getting $U^{\mathcal{A}}_{k,h,m} |\phi\rangle$, has only access to the reduced state $\rho_{h,m}$ (i.e., he has only access to the advice register, as opposed to the adversary who has access to all the qubits).

Notice that there are still a couple of issues with this definition. First of all, when outputting the message function $U_{f,m}$, we do not really care whether $\mathcal{A}$ or $\mathcal{S}$ produce a state accurate enough on the first register (which basically encodes the plaintexts). In fact, we have already noticed that the message function f_m might represent (classically) partial or incomplete information about the message. It is thus natural to require that the only state $\mathcal{A}$ and $\mathcal{S}$ have to output correctly (close enough in terms of some space norm) is the state which represents the partial knowledge about f_m, namely, the reduced density operator defined by $\rho_{f,m} : \sum_x \alpha_x |x\rangle \mapsto \sum_x \|\alpha_x\|^2 |f_m(x)\rangle\langle f_m(x)|$.

The second issue is more subtle. Recall that in the classical SEM notion, we have an adversary who is provided with both the ciphertext *and* the advice, and a simulator who is *only* provided with the advice – but the two advice functions are the same. In the above definition, instead, we have a problem. We are faced with a simulator which is provided with a (mixed) advice state, and an adversary which is provided with a pure state instead, from which he could extract (by partially tracing out) *either* the same advice state as the simulator *or* a mixed state encoding the ciphertext. The situation is clearly asymmetrical in respect to the classical case, and presents problems when dealing with reductions to the qIND-qCPA notion.

Luckily, we are dealing with *classical descriptions* of quantum states. We can therefore ask that the challenger prepares *two copies* of the selected state and produces both: a (pure) state encoding the ciphertext, and another (mixed) reduced state encoding the advice. The choice of a pure state for the ciphertext instead of another mixed state is because it will simplify a reduction to qIND-qCPA: remember in fact that in the (c2) notion we have chosen in Section 3 we use the type-2 transformation $U_{\mathsf{Enc}_k} : \sum_x \alpha_x |x\rangle \mapsto \sum_x \alpha_x |\mathsf{Enc}_k(x)\rangle$. Furthermore, we specify as a norm the *trace norm* [NC00]. This finally leads to the definition of our quantum SEM challenge phase.

Quantum SEM (qSEM) challenge phase: $\mathcal{A}$ sends $\mathcal{C}$ a challenge template consisting of a classical circuit S_m generating a distribution X_m (with poly-sized support) of classical descriptions of m-qubit message states in $\mathcal{H}$,an advice function $h_m : \{0,1\}^m \to \{0,1\}^{\mathrm{poly}(n)}$, and a target function $f_m : \{0,1\}^m \to \{0,1\}^{\mathrm{poly}(n)}$ for an $m \in \mathbb{N}$ of $\mathcal{A}$'s choice. $\mathcal{C}$ replies with the pair $(U_{\mathsf{Enc}_k} |\phi\rangle, \rho_{h,m}(|\phi\rangle))$, where two copies of the same $|\phi\rangle$ (sampled once from X_m) are used. $\mathcal{A}$'s goal is to output $|\psi\rangle$ such that $\| |\psi\rangle - U_{f,m} |\phi\rangle \|_{\mathrm{tr}} \leq \varepsilon$.

In the reduced game, $\mathcal{S}$ only receives the second part of $\rho_{h,m}(|\phi\rangle)$ of $\mathcal{C}$'s response. This finally allows us to define our notion of quantum semantic security:

Definition 4.4. *[qSEM-qCPA] A secret-key encryption scheme is called qSEM-qCPA-secure if for every quantum polynomial-time machine $\mathcal{A}$ and for any $\varepsilon \leq \frac{1}{n}$, there exists a quantum polynomial-time machine $\mathcal{S}$ such that the challenge templates produced by $\mathcal{S}$ and $\mathcal{A}$ are identically distributed and the success probability of $\mathcal{A}$ winning the qSEM-qCPA game is negligibly close (in n) to the success probability of $\mathcal{S}$ winning the reduced game.*

Spoiler. This notion of semantic security is equivalent to qIND-qCPA, and unachievable for those schemes which leave the size of the message unchanged (like most block ciphers), see Section 6.1.

5 Relations

We show the relations between our new notions of indistinguishability and semantic security in the quantum world. It is already known [Gol04] that classically, IND-CPA and semantic security are equivalent. Our goal is to show a similar equivalence for our new notions, plus to show a hierarchy of equivalent security notions. Our results are summarized in Figure 3.

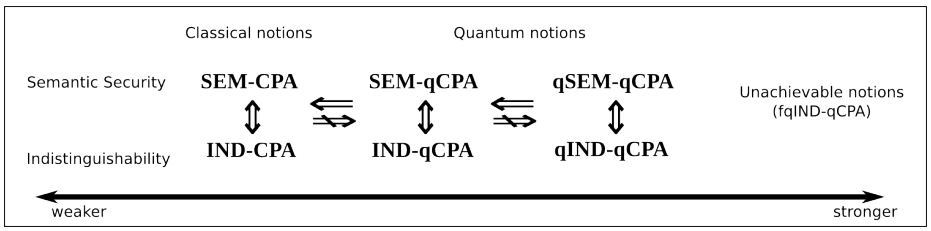

Fig. 3. The relations between notions of indistinguishability and semantic security.

We start by completing the analysis of classical security notions combined with quantum chosen plaintext attacks. Namely, we show that SEM-qCPA and IND-qCPA are equivalent, which implies that SEM-qCPA is strictly stronger than SEM-CPA, completing the basic rectangle of relations.

Theorem 5.1. *[IND-qCPA ⇔ SEM-qCPA] Let $\mathcal{E}$ be a symmetric-key encryption scheme. Then $\mathcal{E}$ is IND-qCPA secure if and only if $\mathcal{E}$ is SEM-qCPA secure.*

We split the proof of Theorem 5.1 into two propositions – one per direction. They closely follow the proofs for the classical case (see [Gol04, Proof of Theorem 5.4.11]), we recall them as they work as a guideline for the following proofs.

Proposition 5.2. *[IND-qCPA ⇒ SEM-qCPA.]*

Proposition 5.3. *[SEM-qCPA ⇒ IND-qCPA]*

Proof (of Proposition 5.2 – Sketch.). The idea of the proof is to hand the simulator $\mathcal{S}$ as non-uniform advice $\mathcal{A}$'s circuit. $\mathcal{S}$ runs $\mathcal{A}$'s circuit and impersonates the challenger $\mathcal{C}$ by generating a new key and answering all of $\mathcal{A}$'s queries using this key. When it comes to the challenge query, $\mathcal{S}$ encrypts the $1 \dots 1$ string of the same length as the original message. It follows from the indistinguishability of encryptions that the adversary's success probability in this game must be negligibly close to its success probability in the real semantic security game, which concludes the proof. The only difference in the -qCPA case is that $\mathcal{A}$ and $\mathcal{S}$ are quantum circuits, and that $\mathcal{S}$ has to emulate the quantum encryption oracle instead of a classical one. □

Proof (of Proposition 5.3). We recall here the full proof as it is short. Assume there exists an efficient distinguisher $\mathcal{A}$ against the IND-qCPA security of $\mathcal{E}$. Then we show how to construct an oracle machine $\mathcal{M}^{\mathcal{A}}$ that has access to $\mathcal{A}$ and breaks the SEM-qCPA security of the scheme. $\mathcal{M}^{\mathcal{A}}$ runs $\mathcal{A}$, emulating the quantum encryption oracle by simply forwarding all the CPA queries to its own oracle. As $\mathcal{A}$ executes an IND challenge query on m-bit messages $(x_0, x_1), \mathcal{M}^{\mathcal{A}}$ produces the SEM template (S_m, h_m, f_m) with S_m describing the uniform distribution over $\{ x_0, x_1 \}, h_m = 1^n$ (or any other function such that $h_m(x_0) = h_m(x_1)$), and f_m a function that fulfills $f_m(x_0) = 0$ and $f_m(x_1) = 1$ (i.e., the distinguishing function). Then $\mathcal{M}^{\mathcal{A}}$ performs a SEM challenge query with this template, and given challenge ciphertext c, uses it to answer $\mathcal{A}$'s query. If, at that point, $\mathcal{A}$ performs more CPA queries, $\mathcal{M}^{\mathcal{A}}$ answers again by forwarding all these queries to its own oracle. Finally, $\mathcal{M}^{\mathcal{A}}$ outputs $\mathcal{A}$'s output. As $\mathcal{A}$ distinguishes encryptions of x_0 and x_1 with non-negligible success probability, $\mathcal{A}$ will return the correct value of f_m with recognizably higher probability than guessing. As h_m is independent of the encrypted message, no simulator can do better than guessing. Hence, $\mathcal{M}^{\mathcal{A}}$ has a non-negligible advantage to output the right value of f_m. □

Theorem 5.4. *[qIND-qCPA ⇔ qSEM-qCPA] Let $\mathcal{E}$ be a symmetric-key encryption scheme. Then $\mathcal{E}$ is qIND-qCPA secure if and only if $\mathcal{E}$ is qSEM-qCPA secure.*

Again, we split the proof of Theorem 5.4 into two propositions.

Proposition 5.5. *[qIND-qCPA ⇒ qSEM-qCPA.]*

Proposition 5.6. *[qSEM-qCPA ⇒ qIND-qCPA]*

Proof (of Proposition 5.5 – Sketch.). The proof follows that of Proposition 5.2, with some careful observations. Since $\mathcal{A}$ is a QPT adversary against the qSEM-qCPA game, its circuit has a short classical representation ξ. So $\mathcal{S}$ gets ξ as non-uniform advice and hence can implement and run $\mathcal{A}$. The simulator $\mathcal{S}$ will simulate $\mathcal{C}$ for $\mathcal{A}$ by generating a new key and answering all of $\mathcal{A}$'s qCPA queries. When it comes to the challenge query, $\mathcal{A}$ will produce a qSEM template, which $\mathcal{S}$ will forward to the real $\mathcal{C}$. Then $\mathcal{S}$ will forward $\mathcal{C}$'s reply, plus a bogus encrypted state (e.g., $|\mathsf{Enc}_k(1\dots1)\rangle$), to $\mathcal{A}$. If at this point $\mathcal{A}$ outputs a state $|\psi\rangle$ which can be reliably distinguished from the correct $U_{f,m}|\phi\rangle$ computed by the real $\mathcal{C}$, we would have an efficient distinguisher against the qSEM-qCPA security of the scheme. Since the probability of distinguishing two quantum states is upper bounded by their trace distance, this concludes the proof. □

Proof (of Proposition 5.6). This is also similar to the proof of Proposition 5.3. Given an efficient distinguisher $\mathcal{A}$ for the qIND-qCPA game, our adversary for the qSEM-qCPA game is an oracle machine $\mathcal{M}^{\mathcal{A}}$ running $\mathcal{A}$ and acting as follows. Concerning $\mathcal{A}$'s qCPA queries, as usual $\mathcal{M}^{\mathcal{A}}$ will just forward everything to the qSEM-qCPA challenger $\mathcal{C}$. When $\mathcal{A}$ performs a challenge qIND query by sending the description of two states $|\phi_0\rangle$ and $|\phi_1\rangle$, $\mathcal{M}^{\mathcal{A}}$ will prepare the qSEM template (S_m, h_m, f_m), with S_m describing the uniform distribution over states $\{\ |\phi_0\rangle,\ |\phi_1\rangle\ \}$, $h_m = 1^n$ (or any other function such that $\rho_{h,m}(|\phi_0\rangle) = \rho_{h,m}(|\phi_1\rangle)$), and f_m the identity function $f_m(x) = x$. Then $\mathcal{M}^{\mathcal{A}}$ performs a qSEM challenge query with this template, and given challenge ciphertext state $U_{\mathsf{Enc}_k}|\phi_b\rangle$, he forwards it as an answer to $\mathcal{A}$'s challenge query. As $\mathcal{A}$ distinguishes $U_{\mathsf{Enc}_k}|\phi_0\rangle$ and $U_{\mathsf{Enc}_k}|\phi_1\rangle$ with non-negligible success probability, $\mathcal{A}$ will return the correct value of b with recognizably higher probability than guessing. Then $\mathcal{M}^{\mathcal{A}}$, having recorded a copy of the classical descriptions of $|\phi_0\rangle$ and $|\phi_1\rangle$, will be able to compute exactly the reduced state $U_{f,m}|\phi_b\rangle$, and win the qSEM-qCPA game for any arbitrarily small ε. As h_m generates the same advice state $\rho_{h,m}$ independently of the encrypted message, no simulator can do better than guessing. This concludes the proof. □

Finally, we show the separation result between the two classes of security we have identified (we show it between IND-qCPA and qIND-qCPA). This shows that qIND-qCPA (eq., qSEM-qCPA) is a strictly stronger notion than IND-qCPA (eq., SEM-qCPA).

Theorem 5.7. *[IND-qCPA $\not\Rightarrow$ qIND-qCPA] There exists a symmetric-key encryption scheme $\mathcal{E}$ which is IND-qCPA secure but not qIND-qCPA secure.*

Proof (of Theorem 5.7). The scheme we use as a counterexample is the one from [Gol04](Construction 5.3.9). It has been proven in [BZ13] that this scheme is IND-qCPA secure if the used PRF is post-quantum secure. We exhibit a distinguisher $\mathcal{A}$ which breaks the qIND-qCPA security of this scheme with high probability. For ease of notation we restrict to the case of single-bit messages 0 and 1. $\mathcal{A}$ will simply choose as challenge states: $|\phi_0\rangle = H|0\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$, and $|\phi_1\rangle = H|1\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle$. When the challenger $\mathcal{C}$ applies the type-2

transformation to either of these two states, it is easy to see that in any case the state is left unchanged. This is because U_{Enc_k} just applies a permutation in the space of the basis elements, but $|\phi_0\rangle$ and $|\phi_1\rangle$ have the same amplitudes on all their components, except for the sign. As these two states are orthogonal, they can be reliably distinguished by $\mathcal{A}$, which can then win the qIND-qCPA game with probability 1. □

The above proof can be generalized to message states of arbitrary length, as our impossibility result in Section 6.1 shows.

6 Impossibility and Achievability Results

In this section we show that qIND-qCPA (equivalently, qSEM-qCPA) is impossible to achieve for encryption schemes which do not expand the message (such as stream ciphers and many block ciphers, ignoring the randomness part in the ciphertext). Therefore, for a scheme to be secure according to this new definition, it is necessary (but not sufficient) to increase the message size during the encryption. Interestingly, such an increase happens in most public-key post-quantum encryption schemes, like for example LWE based schemes [LP11] or the McEliece scheme [McE78].

Then we propose a construction of a qIND-qCPA–secure symmetric-key encryption scheme. Our construction works for any (quantum-secure) pseudorandom permutation (PRP). Given that block ciphers are usually modelled as PRPs, it seems reasonable to assume that we can obtain a secure scheme when using block ciphers with sufficiently large key and block size. Hence, our construction can be used to patch existing schemes, or as a guideline in the design of quantum-secure encryption schemes from block ciphers.

6.1 Impossibility Result

First we have to formally define what it means for a cipher to expand or keep constant the message size by defining the *core function* of a (secret-key) encryption scheme. Intuitively, the definition splits the ciphertext into the randomness, a part carrying the message, and an auxiliary part (e.g., a MAC as used in generic constructions of CCA-secure schemes). This covers most encryption schemes in the literature.

Definition 6.1. *[Core function] Let* $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *be a secret-key encryption scheme. We call the function* $f : \mathcal{K} \times \{0,1\}^\tau \times \mathcal{M} \to \mathcal{Y}$ *the* core function *of the encryption scheme if, for some* $\tau \in \mathbb{N}$*:*

- *for all* $k \in \mathcal{K}$ *and* $x \in \mathcal{M}$, $\mathsf{Enc}_k(x)$ *can be written as* $(r, f(k,r,x), g(k,r,x))$*, where* $r \in \{0,1\}^\tau$ *is independent of the message and* g *can be an arbitrary function; and*
- *there exists a function* f' *such that for all* $k \in \mathcal{K}, r \in \{0,1\}^\tau, x \in \mathcal{M}$*, we have:* $f'(k,r,f(k,x,r)) = x$*.*

For example, in case of the scheme from [BZ13, Construction 4.9] (see Appendix B) the core function is $f(k,r,x) = F_{k_1}(r) \oplus x$ where $k = (k_1, k_2)$ and $g(k,r,x) = H_{k_2}(r\|x)$.

Definition 6.2 (Quasi–length-preserving encryption). *We call a secret-key encryption scheme with core function f* quasi–length-preserving *if*

$$\forall x \in \mathcal{M}, r \in \{0,1\}^\tau, k \in \mathcal{K} \Rightarrow |f(k,x,r)| = |x|,$$

i.e., if the output of the core function has the same bit length as the message.

Continuing the above example, [BZ13, Construction 4.9] is quasi–length-preserving.

The crucial observation is the following: For a quasi–length-preserving encryption scheme, the space of possible input and (core function) output bit-strings (in respect to plaintext and ciphertext) coincide, therefore these ciphers act as permutations on this space. This means that, if we start with an input state which is a superposition of *all* the possible basis states, all of them with the *same* amplitude, this state will be unchanged by the encryption unitary type-2 operation (because it will just 'shuffle' in the basis-state space amplitudes which are exactly the same).

Theorem 6.3. *[Impossibility result for quasi–length-preserving schemes.] No quasi–length-preserving secret-key encryption scheme can be qIND-qCPA secure.*

Proof. Let $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be a quasi–length-preserving scheme. We show an attack that is a generalization of the distinguishing attack in Theorem 5.7.

1. for m-bit message strings, the distinguisher $\mathcal{D}$ sets the two plaintext states for the qIND- game to be: $|\phi_0\rangle = H|0^m\rangle, |\phi_1\rangle = H|1^m\rangle$, where H is the m-fold tensor Hadamard transformation.
2. The challenger flips a random bit b and returns $|\psi\rangle = U_{\mathsf{Enc}_k}|\phi_b\rangle$.
3. $\mathcal{D}$ applies H to the core-function part of the ciphertext $|\psi\rangle$ and measures it in the computational basis. $\mathcal{D}$ outputs 0 if and only if the outcome is 0^m, and outputs 1 otherwise.

As already observed, applying U_{Enc_k} to $H|0^m\rangle$ leaves the state untouched: since the encryption oracle merely performs a permutation in the basis space, and since $|\phi_0\rangle$ is a superposition of every basis element with the same amplitude, it follows that whenever b is equal to 0, the ciphertext state will be unchanged. In this case, after applying the self-inverse transformation H again, $\mathcal{D}$ obtains measurement outcome 0^m with probability 1. On the other hand, if $b = 1$, $|\phi_1\rangle = \frac{1}{2^{m/2}} \sum_y (-1)^{y \cdot 1^m} |y\rangle$ where $a \cdot b$ denotes the bitwise inner product between a and b. Hence, $|\phi_1\rangle$ is a superposition of every basis element where (depending on the parity of y) half of the elements have a positive amplitude and the other half have a negative one, but all of them will be equal in absolute value. Applying $U_{\mathsf{Enc},k}$ to this state, results in $\frac{1}{2^{m/2}} \sum_y (-1)^{y \cdot 1^m} |\mathsf{Enc}_k(y)\rangle$. After re-applying H, the amplitude of the basis state $|0^m\rangle$ becomes $\sum_y (-1)^{y \cdot 1^m + \mathsf{Enc}_k(y) \cdot 0^m}$ which is easily calculated to be 0. Hence, the above attack gives $\mathcal{D}$ a way of perfectly distinguishing between encryptions of the two plaintext states. □

This attack is a consequence of the well-known fact that, in order to perfectly (information-theoretically) encrypt a single quantum bit, *two* bits of classical information are needed: one to mask the basis permutation, and one to mask the permutation on the phase (i.e. the signs of the amplitudes). The fact that we are restricted to quantum operations of the form U_{Enc_k} - that is, quantum instantiations of classical encryptions - means that we cannot afford to hide the phase as well, and this restriction allows for an easy distinguishing procedure.

It is easy to see that this attack also works for the other notions of quantum indistinguishability described in Section 3. In particular, the above theorem shows that [Gol04, Construction 5.3.9], which is IND-qCPA if the used PRF is quantum secure, and [BZ13, Construction 4.9] which is even IND-qCCA1-secure, do not fulfill any of these indistinguishability notions. For convenience of the reader, we recap these constructions in Appendix B.

6.2 Secure Construction

Here we propose a construction of a qIND-qCPA secure symmetric-key encryption scheme from any family of quantum-secure pseudorandom permutations.

Construction 6.4. *For security parameter n, let $m = poly(n)$ and $\tau = poly(n)$. Consider a family of permutations $\Pi_{m+\tau} = (\mathcal{I}, \Pi, \Pi^{-1})$ with key space $\mathcal{K}_\Pi$ that operates on bit strings of length $m+\tau$, and consider a plaintext message space $\mathcal{M} = \{0,1\}^m$, key space $\mathcal{K} = \mathcal{K}_\Pi$, and ciphertext space $\mathcal{C} = \{0,1\}^{m+\tau}$. The construction is given by the following algorithms:*

Key generation algorithm $k \longleftarrow \mathsf{Gen}(1^n)$**:** *on input of security parameter n, the key generation algorithm runs $k \longleftarrow \mathcal{I}(1^{m+\tau})$ and returns secret key k.*

Encryption algorithm $y \longleftarrow \mathsf{Enc}_k(x)$**:** *on input of message $x \in \mathcal{M}$ and key $k \in \mathcal{K}$, the encryption algorithm samples a τ-bit string $r \xleftarrow{\$} \{0,1\}^\tau$ uniformly at random, and outputs $y = \pi_k(x\|r)$ ($\|$ denotes string concatenation).*

Decryption algorithm $x \longleftarrow \mathsf{Dec}_k(y)$**:** *on input of ciphertext $y \in \mathcal{C}$ and key $k \in \mathcal{K}$, the decryption algorithm first runs $x' = \pi_k^{-1}(y)$, and then returns the first m bits of x'.*

The soundness of the construction can be easily checked.

Theorem 6.5. *[qIND-qCPA security of Construction 6.4] If $\Pi_{m+\tau}$ is a family of quantum-resistant pseudorandom permutations (QPRP), then the encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ defined in Construction 6.4 is qIND-qCPA secure.*

Proof. We want to show that no QPT distinguisher $\mathcal{D}$ can win the qIND-qCPA game with probability substantially better than guessing. We first transform the game through a short game-hopping sequence into an indistinguishable game for which we can bound the success probability of any such $\mathcal{D}$.

Game 0. This is the original qIND-qCPA game.

Game 1. This is like Game 0, but instead of using a permutation drawn from the QPRP family $\Pi_{m+\tau}$, a random permutation $\pi \in S_{2^{m+\tau}}$ is chosen from the

set of all permutations over $\{0,1\}^{m+\tau}$. The difference in the success probability of $\mathcal{D}$ winning one or the other of these two games is negligible. Otherwise, we could use $\mathcal{D}$ to distinguish a random permutation drawn from $\Pi_{m+\tau}$ from one drawn from $S_{2^{m+\tau}}$. This would contradict the assumption that $\Pi_{m+\tau}$ is a QPRP.

Game 2. This is like Game 1, but $\mathcal{D}$ is guaranteed that the randomness used for each encryption query is a new random value that was not used before. In other words, the challenger keeps track of all random values used so far and excludes those when sampling a new randomness. Since in Game 0, the same randomness is sampled twice only with negligible probability, it is clear that the probability of winning these two games differs by at most a negligible amount.

Game 3. This is like Game 2 except that the answer to each query asked by $\mathcal{D}$ also contains the randomness r used by the challenger for answering that query. Clearly, $\mathcal{D}$'s probability of winning this game is at least the probability of winning Game 2.

We now show that $\mathcal{D}$'s probability of success in winning Game 3 is negligible in the security parameter. Let $|\varphi\rangle = \sum_{x\in\{0,1\}^m} \alpha_x |x\rangle$ be a m-qubit state. For a τ-bit string $r \in \{0,1\}^\tau$ and a permutation $\pi \in S_{2^{m+\tau}}$, define $|\mathsf{Enc}_{r,\pi}(|\varphi\rangle)\rangle = \sum_{x\in\{0,1\}^m} \alpha_x |\pi(x\|r)\rangle$.

When the modified qIND game starts, $\mathcal{D}$ chooses two different superpositions of messages and sends them to the challenger, who will then choose one of them and send it back encrypted with a fresh randomness $\hat{r}$. Let Q denote the set of $q = poly(n)$ query values used during the previous qCPA-phase. We have to consider that, from this phase, $\mathcal{D}$ knows a set $T \subset \{0,1\}^{m+\tau}$ of 'taken' outputs, i.e. he knows that $\pi(m\|\hat{r})$ will not take one of these values as $\hat{r}$ has not been used before. So, from the adversary's point of view, π is a permutation randomly chosen from S', the set of those permutations over $\{0,1\}^{m+\tau}$ that fix these $|T|$ values. In order to simplify the proof, we will consider a very conservative bound where $|T| = q \cdot 2^m$, and the size of S' is $|S'| = (2^{m+\tau} - |T|)!$ (notice that this bound is very conservative because it assumes that the adversary learns 2^m different ciphertexts for every of the q 'taken' randomnesses, but as we will see, this knowledge will be still insufficient to win the game.)

As $\mathcal{D}$ receives an unknown pure state picked at random from some set, this state is a mixture of every possible state in the set from his point of view. We are interested in the resulting mixture when $\hat{r} \leftarrow_\$ \{0,1\}^\tau$ is known, and $\pi \leftarrow_\$ S'$ is picked uniformly at random:

$$\begin{aligned}\rho_\varphi &= \frac{1}{(2^{m+\tau} - |T|)!} \sum_{\pi\in S'} |\mathsf{Enc}_{\hat{r},\pi}(|\varphi\rangle)\rangle\langle\mathsf{Enc}_{\hat{r},\pi}(|\varphi\rangle)| \qquad (1)\\ &= \frac{1}{(2^{m+\tau} - |T|)!} \sum_{\pi\in S'} \sum_{x,y\in\{0,1\}^m} \alpha_x\overline{\alpha}_y \ |\pi(x\|\hat{r})\rangle \langle\pi(y\|\hat{r})| \,.\end{aligned}$$

For $i \in [0, 1, \ldots, 2^{m+\tau} - 1]$, the ith diagonal entry of ρ_φ is

$$\begin{aligned}
\langle i|\rho_\varphi|i\rangle &= \frac{1}{(2^{m+\tau} - |T|)!} \sum_{\pi \in S'} \sum_{x,y \in \{0,1\}^m} \alpha_x \overline{\alpha}_y \ \langle i|\pi(x\|\hat{r})\rangle \langle \pi(y\|\hat{r})|i\rangle \\
&= \frac{1}{(2^{m+\tau} - |T|)!} \sum_{\pi \in S'} \sum_{x \in \{0,1\}^m} |\alpha_x|^2 \ \langle i|\pi(x\|\hat{r})\rangle \langle \pi(x\|\hat{r})|i\rangle \qquad (2) \\
&= \frac{1}{(2^{m+\tau} - |T|)!} \sum_{x \in \{0,1\}^m} |\alpha_x|^2 \sum_{\substack{\pi \in S' \\ \pi(x\|\hat{r}) = i}} \langle i|\pi(x\|\hat{r})\rangle \langle \pi(x\|\hat{r})|i\rangle \qquad (3)
\end{aligned}$$

where (2) follows from the fact that $x = y$ must hold if both $\pi(x\|\hat{r})$ and $\pi(y\|\hat{r})$ have to be equal to i as π is a permutation. (3) is changing the order of summation and again requiring that $\pi(x\|\hat{r}) = i$. If $i \in T$, the ith diagonal entry of ρ_φ vanishes, because no permutation in S' maps to i. Otherwise, if $i \notin T$,

$$\begin{aligned}
\langle i|\rho_\varphi|i\rangle &= \frac{(2^{m+\tau} - |T| - 1)!}{(2^{m+\tau} - |T|)!} \sum_{x \in \{0,1\}^m} |\alpha_x|^2 \qquad (4) \\
&= \frac{1}{2^{m+\tau} - |T|} \ ,
\end{aligned}$$

(4) follows from the fact that there are exactly $(2^{m+\tau} - |T| - 1)!$ permutations that fix $|T| + 1$ input/output pairs. Hence, all but $|T|$ diagonal elements of ρ_φ are equal to $\frac{1}{2^{m+\tau} - |T|}$ and the others are 0.

Let us perform a similar calculation to determine the off-diagonal element $(\rho_\varphi)_{ij}$ for $i \neq j$:

$$\begin{aligned}
\langle i|\rho_\varphi|j\rangle &= \frac{1}{(2^{m+\tau} - |T|)!} \sum_{\pi \in S'} \sum_{x,y \in \{0,1\}^m} \alpha_x \overline{\alpha}_y \ \langle i|\pi(x\|\hat{r})\rangle \langle \pi(y\|\hat{r})|j\rangle \\
&= \frac{1}{(2^{m+\tau} - |T|)!} \sum_{\pi \in S'} \sum_{x \in \{0,1\}^m} \sum_{y \neq x} \alpha_x \overline{\alpha}_y \ \langle i|\pi(x\|\hat{r})\rangle \langle \pi(y\|\hat{r})|j\rangle \qquad (5) \\
&= \frac{1}{(2^{m+\tau} - |T|)!} \sum_{x \in \{0,1\}^m} \alpha_x \sum_{y \neq x} \overline{\alpha_y} \sum_{\substack{\pi \in S' \\ \pi(x\|\hat{r}) = i \\ \pi(y\|\hat{r}) = j}} \langle i|\pi(x\|\hat{r})\rangle \langle \pi(y\|\hat{r})|j\rangle \qquad (6)
\end{aligned}$$

where (5) follows from the fact that $x \neq y$ must hold if both $\pi(x\|\hat{r}) = i$ and $\pi(y\|\hat{r}) = j$ have to hold and $i \neq j$. (6) is changing the order of summation and again requiring that $\pi(x\|\hat{r}) = i$ and $\pi(y\|\hat{r}) = j$. If either $i \in T$ or $j \in T$ (or both), then the i, jth entry of ρ_φ vanishes, as no permutation maps to an element in T. Otherwise, if both i and j are not in T,

$$\begin{aligned}
\langle i|\rho_\varphi|j\rangle &= \frac{(2^{m+\tau} - |T| - 2)!}{(2^{m+\tau} - |T|)!} \sum_{x \in \{0,1\}^m} \alpha_x \sum_{y \neq x} \overline{\alpha_y} \qquad (7) \\
&= \frac{1}{(2^{m+\tau} - |T|) \cdot (2^{m+\tau} - |T| - 1)} \sum_{x \in \{0,1\}^m} \alpha_x \sum_{y \neq x} \overline{\alpha_y} \ ,
\end{aligned}$$

where (7) follows from the fact that there are exactly $(2^{m+\tau} - |T| - 2)!$ permutations that fix $|T| + 2$ different input/output pairs. It follows that all non-zero off-diagonal elements of ρ_φ are equal.

Let us define $a := \sum_{x\in\{0,1\}^m} \alpha_x \sum_{y\neq x} \overline{\alpha_y}$ and derive an upper bound on a using that $\sum_x |\alpha_x|^2 = 1$:

$$\begin{aligned} a &= \sum_{x\in\{0,1\}^m} \alpha_x \left(\left(\sum_{y\neq x} \overline{\alpha_y} \right) + \overline{\alpha_x} - \overline{\alpha_x} \right) \\ &= \sum_x \alpha_x \sum_y \overline{\alpha_y} - \sum_x \alpha_x \overline{\alpha_x} \\ &= \left| \sum_x \alpha_x \right|^2 - 1 \\ &\leq \sum_x |\alpha_x|^2 \cdot 2^m - 1 = 2^m - 1 \ , \end{aligned} \tag{8}$$

where the inequality is Cauchy-Schwarz. Note that the upper bound is achieved for instance for uniform amplitudes $\alpha_x = \frac{1}{2^{m/2}}$. On the other hand, we can also conclude from Equation (8) that $a \geq -1$ which is achieved for $\alpha_x = (-1)^{x\cdot 1^m}$.

When investigating how well one can distinguish between the encryption of two different quantum states $|\varphi\rangle = \sum_{x\in\{0,1\}^m} \alpha_x |x\rangle$ and $|\psi\rangle = \sum_{x\in\{0,1\}^m} \beta_x |x\rangle$, we have to consider the trace distance between the resulting density matrices [FvdG99]

$$\|\rho_\varphi - \rho_\psi\|_{\mathrm{tr}} = \frac{1}{2} \operatorname{tr} |\rho_\varphi - \rho_\psi| \, ,$$

where $|A| := \sqrt{A^\dagger A}$ is the positive square root of $A^\dagger A$. Hence, we have to sum the absolute eigenvalues of $\rho_\varphi - \rho_\psi$ [NC00]. From the structure of the ρ matrices derived above, we know that $\rho_\varphi - \rho_\psi$ is a $2^{m+\tau} \times 2^{m+\tau}$ matrix with 0's on the diagonal and

$$c := \frac{1}{(2^{m+\tau} - |T|) \cdot (2^{m+\tau} - |T| - 1)} \left(\sum_{x\in\{0,1\}^m} \alpha_x \sum_{y\neq x} \overline{\alpha_y} - \sum_{x\in\{0,1\}^m} \beta_x \sum_{y\neq x} \overline{\beta_y} \right)$$

as off-diagonal elements, whenever $i \notin T$ and $j \notin T$ and zero otherwise.

It is easy to verify that the spectrum of such matrices is

$$\{(2^{m+\tau} - |T| - 1) \cdot c, \overbrace{0, \ldots, 0}^{|T|}, \overbrace{-c, \ldots, -c}^{2^{m+\tau}-|T|-1}\} \ ,$$

because the eigenvectors are:

- the vector that has 0 entries at every position with indices in T and 1 entries everywhere else, with associated eigenvalue $(2^{m+\tau} - |T| - 1) \cdot c$,
- vectors of the form $(0, \ldots, 0, 1, 0, \ldots, 0)^T$, where the single 1 entry spaces all the positions with indices in T, with associated eigenvalues 0, and

– vectors of the form $(0,\ldots,0,-1,0,\ldots,0,1,0,\ldots,0)^T$, where the -1 entry is at the first position with index not in T and the 1 entry of the ith such vector is at the $(i+1)$th position with index not in T, for $i \in \{1,\ldots,2^{m+\tau}-|T|-1\}$, with associated eigenvalues $-c$.

Therefore, the sum of the absolute eigenvalues is

$$\begin{aligned}\frac{1}{2}\operatorname{tr}|\rho_\varphi - \rho_\psi| &= \frac{1}{2}\cdot 2\cdot(2^{m+\tau}-|T|-1)\cdot|c| \\ &= \frac{2^{m+\tau}-|T|-1}{(2^{m+\tau}-|T|)\cdot(2^{m+\tau}-|T|-1)}\left|\sum_{x\in\{0,1\}^m}\alpha_x\sum_{y\neq x}\overline{\alpha_y} - \sum_{x\in\{0,1\}^m}\beta_x\sum_{y\neq x}\overline{\beta_y}\right| \\ &\leq \frac{1}{2^{m+\tau}-|T|}2^m = 2^{-(\tau-1)}\ ,\end{aligned}$$

where the inequality follows from the upper and lower bounds on the off-diagonal elements derived above. We conclude that the trace distance between two encryptions of arbitrary quantum states is negligible in the security parameter and hence, they cannot be distinguished except with negligible probability. □

7 Conclusions and Further Directions

We believe that many of the current security notions used in different areas of cryptography are unsatisfying in case quantum computers become reality. In this respect, our work contributes to a better understanding of which properties are important for the long-term security of modern cryptographic primitives. Our work opens many interesting follow-up questions.

There are many other directions to investigate, once the basic framework of 'indistinguishability versus semantic security' presented in this work is completed. A natural direction is to look at quantum CCA security in this framework. This topic was also initiated in [BZ13] relative to the IND-qCPA model; it is intriguing to extend the definition of CCA security to stronger notions obtained by starting from our qIND-qCPA model.

With respect to qIND-qCPA, we have left as an open problem a detailed study of three other possible notions, namely the models (Q1), (Q2), and (c1). It is interesting to check whether they are equivalent to qIND-qCPA; if not, whether they are achievable, and their relation to other notions of semantic security. We have also decided not to take into account (I) models: these models lead to the study of *quantum fault attacks*. Moreover, we have not considered superpositions of keys or randomness: these lead to a quantum study of *weak-key* and *bad-randomness* models. The authors of this paper are not aware of any results in these directions.

With respect to semantic security, it is also possible to weaken qSEM-qCPA by restricting the messages to be quantum, but the advice function to be classical. All the semantic security notions can be also studied in the *uniform model*.

Our secure construction shows how to turn block ciphers into qIND-qCPA secure schemes. An interesting research question is whether there exists a general patch transforming an IND-CPA secure scheme into a qIND-qCPA secure one. It is important to study how our transformation can be applied to general modes of operation.

Finally, although much different in scope, it would also be possible to study *fully quantum encryption*, i.e., encryption schemes for protecting quantum information, meant to be run on quantum computers, where all the data and parties involved behave fully quantum, and the encryption and decryption operations are arbitrary unitaries.

Acknowledgements: The authors would like to thank the anonymous reviewers for useful comments. Tommaso Gagliardoni was supported by the German Federal Ministry of Education and Research (BMBF) within EC SPRIDE. Andreas Hülsing was supported by the Netherlands Organisation for Scientific Research (NWO) under grant 639.073.005. Christian Schaffner was supported by a 7th framework EU SIQS grant. Part of this work was supported by the COST Action IC1306.

References

BBD09. Daniel J. Bernstein, Johannes Buchmann, and Erik Dahmen. *Post-Quantum Cryptography*. Springer, 2009.

BDF+11. Dan Boneh, Özgür Dagdelen, Marc Fischlin, Anja Lehmann, Christian Schaffner, and Mark Zhandry. Random Oracles in a Quantum World. In *Asiacrypt 2011*, number 7073 in LNCS, pages 41–69. Springer, 2011.

BHT97. Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum algorithm for the collision problem. arXiv:quant-ph/9705002, 1997.

BS03. Johannes Blömer and Jean-Pierre Seifert. Fault Based Cryptanalysis of the Advanced Encryption Standard (AES). In Rebecca N. Wright, editor, *Financial Cryptography*, volume 2742 of *LNCS*, pages 162–181. Springer, 2003.

BZ13. Dan Boneh and Mark Zhandry. Secure Signatures and Chosen Ciphertext Security in a Post-Quantum World. Cryptology ePrint Archive, Report 2013/088, 2013. An extended abstract appeared in Ran Canetti and Juan A. Garay, editors, *Crypto 2013*, volume 8043 of *LNCS*, pages 361–379. Springer, 2013.

DFNS13. Ivan Damgård, Jakob Funder, Jesper Buus Nielsen, and Louis Salvail. Superposition Attacks on Cryptographic Protocols. In Carles Padró, editor, *ICITS 2013*, volume 8317 of *LNCS*, pages 142–161. Springer, 2013.

FvdG99. C.A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *Information Theory, IEEE Transactions on*, 45(4):1216–1227, May 1999.

Gol04. Oded Goldreich. *Foundations of Cryptography: Volume 2, Basic Applications*. Cambridge University Press, Cambridge, UK, 2004.

Gro96. Lov K. Grover. A Fast Quantum Mechanical Algorithm for Database Search. In *STOC '96*, pages 212–219. ACM, 1996.

LP11. Richard Lindner and Chris Peikert. Better Key Sizes (and Attacks) for LWE-Based Encryption. In Aggelos Kiayias, editor, *CT-RSA 2011*, volume 6558 of *LNCS*, pages 319–339. Springer, 2011.

McE78. Robert J. McEliece. A public-key cryptosystem based on algebraic coding theory. *DSN progress report*, 42(44):114–116, 1978.

NC00. Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

Sho94. Peter W. Shor. Algorithms for Quantum Computation: Discrete Logarithms and Factoring. In *FOCS 1994*, pages 124–134. IEEE Computer Society Press, 1994.

Unr12. Dominique Unruh. Quantum proofs of knowledge. In David Pointcheval and Thomas Johansson, editors, *Eurocrypt 2012*, volume 7237 of *LNCS*, pages 135–152. Springer, 2012.

Vel13. Maria Velema. Classical encryption and authentication under quantum attacks. Master's thesis, Master of Logic, University of Amsterdam, 2013. http://arxiv.org/abs/1307.3753.

Wat01. John Watrous. Quantum Algorithms for Solvable Groups. In *STOC '01*, pages 60–67. ACM, 2001.

Wat06. John Watrous. Zero-knowledge Against Quantum Attacks. In *STOC '06*, pages 296–305. ACM, 2006.

Zha12. Mark Zhandry. How to Construct Quantum Random Functions. In *FOCS 2012*, pages 679–687. IEEE, 2012.

A Formal Definitons of SEM-CPA and IND-CPA

The following definitions are more precise than the ones we use in the main text. They are included here for reference and were taken from Goldreich ([Gol04]).

Definition A.1 (Sem-CPA). *A secret-key encryption scheme,* $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$*, is said to be semantically secure under chosen plaintext attacks if for every pair of probabilistic polynomial-time oracle machines* $\mathcal{A}_1$ *and* $\mathcal{A}_2$*, there exists a pair of probabilistic polynomial-time algorithms* $\mathcal{A}'_1$ *and* $\mathcal{A}'_2$ *such that the following two conditions hold:*

1. *For every positive polynomial* $p(\cdot)$*, and all sufficiently large* n *and* $z \in \{0,1\}^{poly(n)}$ *it holds that*

$$\mathsf{Pr}\left[\begin{array}{l} v = f_m(x) \text{ where} \\ \qquad k \longleftarrow \mathsf{Gen}(1^n) \\ \qquad ((S_m, h_m, f_m), \sigma) \longleftarrow \mathcal{A}_1^{\mathsf{Enc}_k}(1^n, z) \\ \qquad c \longleftarrow (\mathsf{Enc}_k(x), h_m(x)), \text{where } x \longleftarrow S_m(U_{poly(n)}) \\ \qquad v \longleftarrow \mathcal{A}_2^{\mathsf{Enc}_k}(\sigma, c) \end{array}\right]$$

$$< \quad \mathsf{Pr}\left[\begin{array}{l} v = f_m(x) \text{ where} \\ \qquad ((S_m, h_m, f_m), \sigma) \longleftarrow \mathcal{A}'_1(1^n, z) \\ \qquad x \longleftarrow S_m(U_{poly(n)}) \\ \qquad v \longleftarrow \mathcal{A}'_2(\sigma, 1^{|x|}, h_m(x)) \end{array}\right] + \tfrac{1}{p(n)} \qquad (9)$$

Recall that (S_m, h_m, f_m) *is a triplet of circuits produced as in Step 3 of the foregoing description, and that* x *is a sample from the distribution induced by* S_m*.*

2. *For every n and z, the first elements (i.e., the (S_m, h_m, f_m) part) in the random variables $\mathcal{A}'_1(1^n, z)$ and $A_1^{\mathsf{Enc}_{\mathsf{Gen}(1^n)}}(1^n, z)$ are identically distributed.*

Definition A.2 (IND-CPA). *A secret-key encryption scheme, $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$, is said to have indistinguishable encryptions under chosen plaintext attacks if for every pair of probabilistic polynomial-time oracle machines, $\mathcal{A}_1$ and $\mathcal{A}_2$, for every positive polynomial $p(\cdot)$, and all sufficiently large n and $z \in \{0,1\}^{poly(n)}$ it holds that*

$$\left| p_{n,z}^{(1)} - p_{n,z}^{(2)} \right| < \frac{1}{p(n)}$$

where

$$p_{n,z}^{(i)} \stackrel{def}{=} \mathsf{Pr}\left[\begin{array}{l} v = 1 \text{ where} \\ \quad k \longleftarrow \mathsf{Gen}(1^n) \\ \quad ((x_1, x_2), \sigma) \longleftarrow \mathcal{A}_1^{\mathsf{Enc}_k}(1^n, z) \\ \quad c \longleftarrow \mathsf{Enc}_k(x_i) \\ \quad v \longleftarrow \mathcal{A}_2^{\mathsf{Enc}_k}(\sigma, c) \end{array}\right]$$

where $|x_1| = |x_2|$.

Please note that there are no restrictions regarding $\mathcal{A}$'s oracle queries, i.e. $\mathcal{A}_1$ as well as $\mathcal{A}_2$ are allowed to ask for encryptions of x_1 and x_2.

B Example Encryption Schemes

In this section we recall two generic constructions of classically secure encryption schemes that we use as examples throughout our work. The first construction is Construction 5.3.9 from [Gol04] which achieves IND-CPA security starting from a pseudorandom function family. Afterwards, we recall construction 4.9 from [BZ13] which extends the first construction with a PRF-MAC to achieve IND-CCA1 security.

Construction B.1 ([Gol04],Construction 5.3.9). *Let $n \in \mathbb{N}$ be the security parameter, $\tau, m \in poly(n)$, $\mathcal{F} = \{F_k : \{0,1\}^\tau \to \{0,1\}^m \mid k \in \mathcal{K}\}$ be a pseudorandom function family with key space $\mathcal{K}$. Then the following triple of algorithms form a symmetric-key encryption scheme with message space $\{0,1\}^m$:*

$\mathsf{Gen}(1^n)$: *On input of the security parameter, returns a uniformly random key $k \xleftarrow{\$} \mathcal{K}$ for the PRF $\mathcal{F}$ as secret key.*

$\mathsf{Enc}(x, k)$: *On input of message x and key k returns cipher text $c = (r, c')$ where randomness $r \xleftarrow{\$} \{0,1\}^\tau$ is a uniformly random τ bit string and c' is computed as*

$$c' \longleftarrow F_k(r) \oplus x.$$

$\mathsf{Dec}(c, k)$: *On input of cipher text $c = (r, c')$ and key k returns plain text*

$$x \longleftarrow c' \oplus F_k(r).$$

The construction above only achieves CPA security. A known way to achieve CCA security is to add a MAC such that an adversary is unable to produce valid cipher texts himself. Construction B.2 below uses a PRF MAC for this purpose.

Construction B.2 ([BZ13],Construction 4.9). *Let $n \in \mathbb{N}$ be the security parameter, $\tau, m \in poly(n)$, $\mathcal{F} = \{F_k : \{0,1\}^\tau \to \{0,1\}^m \mid k \in \mathcal{K}_\mathcal{F}\}$ and $\mathcal{H} = \{H_k : \{0,1\}^{(\tau+m)} \to \{0,1\}^n \mid k \in \mathcal{K}_\mathcal{H}\}$ be a pseudorandom function family with key space $\mathcal{K}_\mathcal{F}$ and $\mathcal{K}_\mathcal{H}$, respectively. Then the following triple of algorithms form a symmetric-key encryption scheme with message space $\{0,1\}^m$:*

$\mathsf{Gen}(1^n)$**:** *On input of the security parameter, returns two uniformly random keys $k_1 \xleftarrow{\$} \mathcal{K}_\mathcal{F}$ and $k_2 \xleftarrow{\$} \mathcal{K}_\mathcal{H}$ for the PRFs as secret key $k = (k_1, k_2)$.*

$\mathsf{Enc}(x, k)$**:** *On input of message x and key $k = (k_1, k_2)$ returns cipher text $c = (r, c_1, c_2)$ where randomness $r \xleftarrow{\$} \{0,1\}^\tau$ is a uniformly random τ bit string and c_1, c_2 are computed as*

$$c_1 \longleftarrow F_{k_1}(r) \oplus x$$
$$c_2 \longleftarrow H_{k_2}(r \| m).$$

$\mathsf{Dec}(c, k)$**:** *On input of cipher text $c = (r, c_1, c_2)$ and key $k = (k_1, k_2)$ returns*

$$m \longleftarrow c_1 \oplus F_{k_1}(r), \text{ if } c_2 = H_{k_2}(r \| m)$$
$$\perp, \text{ otherwise.}$$