Semantic Security and Indistinguishability in the Quantum World

June 1, 2016*

Tommaso Gagliardoni[1], Andreas Hülsing[2], and Christian Schaffner[3,4,5]

[1] CASED, Technische Universität Darmstadt, Germany
tommaso@gagliardoni.net
[2] TU Eindhoven, The Netherlands
andreas@huelsing.net
[3] Institute for Logic, Language and Compuation (ILLC), University of Amsterdam, The Netherlands
c.schaffner@uva.nl
[4] Centrum Wiskunde & Informatica (CWI) Amsterdam, The Netherlands
[5] QuSoft, 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 for symmetric-key encryption schemes. We justify our stronger definition by showing its equivalence to novel quantum semantic-security notions that we introduce. Furthermore, we show that our new security definitions cannot be achieved by a large class of ciphers – those which are quasi-preserving the message length. On the other hand, we provide a secure construction based on quantum-resistant pseudorandom permutations; this construction can be used as a generic transformation for turning a large class of encryption schemes into quantum indistinguishable and hence quantum semantically secure ones. Moreover, our construction is the first completely classical encryption scheme shown to be secure against an even stronger notion of indistinguishability, which was previously known to be achievable only by using quantum messages and arbitrary quantum encryption circuits.

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

* An extended abstract of this work appears in the proceedings of CRYPTO 2016. This is the full version.

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,DFNS14,Unr12,Wat09,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. Another possibility occurs when using obfuscation. There are applications where one might want to distribute the obfuscated code of a symmetric-key encryption scheme (with the secret key hardcoded) in order to allow a third party to generate ciphertexts without being able to retrieve the key - think of this as building public-key encryption from symmetric-key encryption using Indistinguishability Obfuscation. Because in these cases an adversary receives the classical code for producing encryptions, he could implement the code on his local quantum computer and query the resulting quantum circuit on a superporition of inputs. Moreover, even in quantum reductions for classical schemes situations could arise where superposition access is needed. A typical example are impossibility results (such as meta-reductions [DFG13]), where giving the adversary additional power often rules out a broader range of secure reductions. Notions covering such settings are often called *quantum-security* notions. In this work we propose new quantum-security notions for encryption schemes.

For encryption, the notion of *semantic security* [GM84,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 for classical encryption schemes 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, in the symmetric-key scenario, 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, see Figure 2 below for an overview.

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. This transformation also works in respect to an even stronger notion of indistinguishability in the quantum world, which was introduced in [BJ15], and previously only known to be achievable in the setting of *computational quantum encryption*, that is, the scenario where all the parties have quantum computing capabilities, and encryption is performed through arbitrary quantum circuits operating on quantum data. Even if this scenario goes in a very different direction from the scope of our work, it is interesting to note that our construction is the first fully classical scheme secure even in respect to such a purely quantum notion of security.

The 'frozen smart-card' example. In order to clarify why quantum security allows the adversary *quantum superposition access* to classical primitives - as opposed to the case of post-quantum security - we give a motivating example. In this mind experiment, we consider a not-so-distant future where the target of an attack is a tiny encryption chip, e.g., integrated into an RFID tag or smart-card. It is reasonable to assume that it will include elements of technology currently researched but undeployed (i.e., extreme miniaturization, optical electronics, etc.) Regardless, the chip we consider is a purely classical device, performing classical encryption (e.g. AES) on classical inputs, and outputting classical outputs.

Consider an adversary equipped with some future technology which subjects the device to a fault-injection environment, by varying the physical parameters (such as temperature, power, speed, etc.) under which the device usually operates. As a figurative example, our 'quantum hacker' could place the chip into an isolation pod, which keeps the device at a very low temperature and shields it from any external electromagnetic or thermal interference. This situation would be analogous to what happens when security researchers perform side channel analysis on cryptographic hardware in nowaday's labs, using techniques such as thermal or electromagnetic manipulation which were previously considered futuristic. There is no guarantee that, under these conditions, the chip does not start to show full or partial quantum behaviour. At this point, the adversary could query the device on a superposition of plaintexts by using, e.g., a laser and an array of beam splitters when feeding signals into the chip via optic fiber.

It is unclear today what a future attacker might be able to achieve using such an attack. As traditionally done in cryptography, we assume the worst-case scenario where the attacker can actually query the target device in superposition. Classical security notions such as IND-CPA do not cover this scenario while our new notion qIND-qCPA does. This setting is an example of what we mean by 'tricking classical parties into quantum behaviour'.

Related work. The idea of considering scenarios where a quantum adversary can force other parties into quantum behaviour has been first considered in [DFNS14]. Attacks exploiting classical encryptions in quantum superposition have been described in [KM10,KM12,KLLNP16,SS16]. 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]. Prior work has considered the security of quantum methods to encrypt classical data in the computational setting [Kos07,XY12]. In concurrent and independent work, Broadbent and Jeffery [BJ15] introduce indistinguishability notions for the public- and secret-key encryption of quantum messages in the context of fully homomorphic quantum computation. We refer to Page 16 for a more detailed description of how their definitions relate to our framework. A more complete overview for these notions, including semantic security for quantum encryption schemes, can be found in another concurrent work [ABF+16].

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. 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 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 indistinguishability of ciphertexts under adaptively chosen plaintext attack (IND-CPA), and semantic security under adaptively chosen plaintext attack (SEM-CPA), which are known to be equivalent (e.g., [Gol04]).

Game-based definitions. 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.

Definition 2.2 (IND-CPA). *A secret-key encryption scheme is said to be* IND-CPA secure *if the success probability of any probabilistic polynomial-time adversary winning the game defined by CPA learning phases and an IND challenge phase is at most negligibly (in n) close to* $1/2$.

The SEM challenge phase: $\mathcal{A}$ sends $\mathcal{C}$ a challenge template (S_m, h_m, f_m) consisting of a poly-sized circuit S_m specifying a distribution over m-bit long plaintexts, an advise function $h_m : \{0,1\}^m \to \{0,1\}^*$, and a target function $f_m : \{0,1\}^m \to \{0,1\}^*$. 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 $f_m(x)$.

In the definition of semantic security it is not required that $\mathcal{A}$'s probability of winning the game is always negligible. Instead, $\mathcal{A}$'s success probability is compared to that of a simulator $\mathcal{S}$ that plays in a *reduced game*: On one hand, $\mathcal{S}$ gets no learning phases. On the other hand, during the challenge phase, $\mathcal{S}$ does not receive the ciphertext but only the output of the advice function. This use of a simulator is what makes the notion hard to work with in proofs as one has to construct a simulator for every possible $\mathcal{A}$ to prove a scheme secure.

Definition 2.3 (SEM-CPA). *A secret-key encryption scheme is said to be* 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 game defined by CPA learning phases and a SEM challenge phase (computed over the coins of $\mathcal{A}$, Gen, and S_m) is negligibly close (in n) 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.

2.2 Previous Notions of Security in the Quantum World

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 $|\varphi\rangle$ belongs to one particular subspace, and we will omit the index n

when the security parameter is implicit, therefore writing just $|\varphi\rangle \in \mathcal{H}$. We will denote pure states with ket notation, e.g., $|\varphi\rangle$, while mixed states will be denoted by lowercase Greek letters, e.g. ρ. We start by defining what we call a *classical description of a quantum state*:

Definition 2.4 (Classical Description). *A* classical description *of a quantum state ρ is a (classical) bitstring describing a quantum circuit S which (takes no input but starts from a fixed initial state $|0\rangle$ and) outputs ρ.*

This definition will be used later in our new notions of security. We deviate here from the traditional meaning of 'classical description' referring to individual numerical entries of the density matrix. The reason is that our definition also covers the cases where those numerical entries are not easily computable, as long as we can give an explicit constructive procedure for that state. Clearly, every pure quantum state $|\varphi\rangle$ has a classical description (given by a description of the quantum circuit which implements the unitary that maps $|0\rangle$ to $|\varphi\rangle$. The classical description of a mixed state ρ_A is given by the circuit which first creates a purification $|\varphi\rangle_{AR}$ of ρ_A and then only outputs the A register. Note that a state admitting a classical description cannot be entangled with any other system.

For encryption, following the approach in [BZ13] and many other works, we define the following:

Definition 2.5 (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:*

$$U_{\mathsf{Enc}_k} : \sum_{x,y} \alpha_{x,y} |x\rangle |y\rangle \mapsto \sum_{x,y} \alpha_{x,y} |x\rangle |y \oplus \mathsf{Enc}_k(x)\rangle \tag{1}$$

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 indistinguishability 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}.

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:

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

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

$\mathcal{A}$'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 2.6 (fqIND-qCPA). *A secret-key encryption scheme is said to be* fqIND-qCPA secure *if the success probability of any quantum probabilistic polynomial-time adversary winning the game defined by qCPA learning phases and a fqIND challenge phase 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 2.7 (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 $m = \mathrm{poly}(n)$. □

Theorem 2.7 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 2.8 (IND-qCPA [BZ13]). *A secret-key encryption scheme is said to be* IND-qCPA secure *if the success probability of any quantum probabilistic polynomial-time adversary winning the game defined by qCPA learning phases and a 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, under standard computational assumptions, IND-qCPA is strictly stronger than IND-CPA:

Theorem 2.9 (IND-CPA $\not\Rightarrow$ IND-qCPA [BZ13, Theorem 4.8]). *If classically secure PRFs exist and order-finding in prime groups is classically hard, then 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 from Theorem 2.7. 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 select a suitable 'qIND-'notion amongst all the possible candidate ones.

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[7]. 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, is implemented. In the classical world, the following two cases are equivalent but in the quantum world 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 $|\varphi\rangle$, the adversary therefore computes the state:

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

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. Moreover, $\mathcal{A}$ does not automatically gain access to the inverse (adjoint) of quantum operations performed by $\mathcal{C}$, and $\mathcal{C}$ cannot be 'rewound' by the adversary, which would be far too powerful possibilities. This scenario also covers the case of 'unidirectional' state transmission, i.e., when qubits are sent over a quantum channel to another party, and they are not available afterwards until that party sends them back. Regardless, in security proofs in the ($\mathcal{C}$) model, it is still allowed for an external entity (e.g. a simulator, or a reduction) to rewind the joint circuit composed by adversary and challenger together, if need be. However, we are not aware of any known reduction involving rewinding in this form for encryption schemes in the quantum world.

In order to keep consistency with this choice of the model, when also considering qCPA queries, we implicitly assume the same access mode to the Enc_k oracle as in the qIND game. That is, if we are in the ($\mathcal{O}$) scenario, during the qCPA phase $\mathcal{A}$ has quantum oracle access to Enc_k. In the ($\mathcal{C}$) case, instead, superposition access to Enc_k is provided to $\mathcal{A}$ by an external challenger.

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 (BQP-producible) quantum states and can be entangled with each other and other states. In the (c) model, instead, $\mathcal{A}$ is only allowed to choose *classical descriptions* of two m-qubit quantum states according to Definition 2.4,

[7] We do not rule out that some of them might eventually lead to the same model.

thus being 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 world, there is a big difference between these two cases, as the latter allows $\mathcal{A}$ to establish entanglement of the message(s) with other registers. This is not possible when using 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 best possible choice of messages from the adversary's perspective. Hence, the (c) model is a valid alternative. Will further discuss the difference between these two models later.

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 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. We stress that this distinction in relaying is not trivial: it is not possible for $\mathcal{A}$, in general, to simulate relaying by keeping internal states entangled with the plaintexts. As an example, consider the attack in Theorem 2.7: it is easy to see that this cannot be performed without relaying.

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:

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

This way ensures that the resulting operator is invertible, even if f is not. We call this *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:

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

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 act 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:

$$\sum_{x} \alpha_{x} |x\rangle \mapsto \sum_{x} \alpha_{x} |\mathsf{Enc}_k(x)\rangle . \tag{2}$$

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

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

where the length of the ancilla register is $|y| = |\mathsf{Enc}_k(x)| - |x|$ and $\varphi_{x,0} = \mathsf{Enc}_k(x)$ for every x – i.e., initializing the ancilla y register in the $|0\rangle$ state produces a correct encryption, which is what we expect from an honest quantum executor. One might ask what happens if the ancilla is not initialized to 0, and we leave the general case of arbitrary ancillas manipulation as an interesting open problem, but we stress the fact that this behavior is not considered in the case of honest parties. We call these *type*-(2) *transformations*[8].

Notice that, in general, type-(1) and type-(2) transformations are very different: having quantum oracle access to a type-(2) unitary $U^{(2)}_{\mathsf{Enc}}$ and its adjoint also gives access to the related type-(2) *decryption oracle* $U^{(2)}_{\mathsf{Dec}} : \sum_x \alpha_x |\mathsf{Enc}_k(x)\rangle \mapsto \sum_x \alpha_x |x\rangle$. In fact, notice that $(U^{(2)}_{\mathsf{Enc}})^\dagger = U^{(2)}_{\mathsf{Dec}}$, while the adjoint of a type-(1) encryption operator, $(U^{(1)}_{\mathsf{Enc}})^\dagger$, is generally *not* a type-(1) decryption operator. In particular, type-(2) operators are 'more powerful' in the sense that knowledge of the secret key is required in order to build any efficient quantum circuit implementing them. However, we stress the fact that whenever access to a decryption oracle is allowed, the two models are completely equivalent, because then we can simulate a type-(2) operator by using ancilla qubits and 'uncomputing' the resulting garbage lines (see Figure 1) (as we will see, this will be the case for the challenger in our qIND notion).

3.2 Analysis of the models

By considering these 4 distinctions in the security tree we have $2^4 = 16$ possible candidate models to analyze. We label each of these candidate models by

[8] These are called *minimal quantum oracles* in [KKVB02].

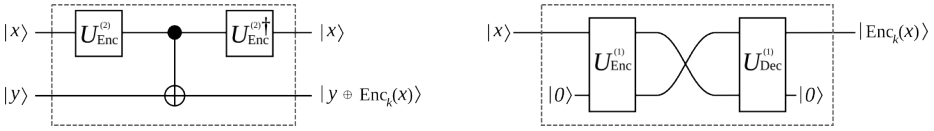

Fig. 1. Equivalence between type-(1) and type-(2) in the case of 1-qubit messages. Left: building a type-(1) encryption oracle by using a type-(2) encryption oracle (and its inverse) as a black-box. Right: building a type-(2) encryption oracle by using type-(1) encryption and decryption oracles as black-boxes.

appending each one of the 4 labels of every tree branch in brackets. Clearly, 16 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 $(\mathcal{O}\boldsymbol{c}\ldots)$. In the $\mathcal{O}$ scenario, the oracle is actually a quantum gate inside $\mathcal{A}$'s quantum circuitry. Therefore $\mathcal{A}$ has the capability of querying the oracle on states which are possibly entangled with other registers kept by $\mathcal{A}$ itself.

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

Leaves of the form $(\ldots\boldsymbol{Y2})$. 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 $\mathcal{A}$).

Excluding these options leaves us with 7 models, but it is easy to see that 3 of them are unachievable because of the attack from Theorem 2.7. This is the case for $(\mathcal{O}QY1)$ (which is exactly fqIND-qCPA), $(\mathcal{C}QY1)$, and $(\mathcal{C}cY1)$. Of the remaining 4, notice that $(\mathcal{C}Qn1)$ and $(\mathcal{C}cn1)$ are equivalent to the IND-qCPA notion from [BZ13]. The reason is that from $\mathcal{A}$'s perspective, a non-relaying $\mathcal{C}$ is indistinguishable from a $\mathcal{C}$ tracing out (measuring) the plaintext register (otherwise $\mathcal{A}$ and $\mathcal{C}$ could communicate faster than light). This measuring operation would make the ciphertext collapse into a single (classical) ciphertext. And since tracing out the challenge register and applying the type-(1) operator $U^{(1)}_{\mathsf{Enc}}$ commute, one can consider (without loss of generality) the case that $\mathcal{A}$ himself first measures the plaintext register, and then initiates a classical IND query with $\mathcal{C}$, therefore recovering a classical definition of IND challenge query[9]. Therefore,

[9] However, we stress that this interpretation is not entirely correct. In fact, one might consider composition scenarios where the IND query is just an intermediate step, and the plaintext and ciphertext registers are reunited at some later step. In such scenarios, not relaying would not be equivalent to measuring. We ignore such considerations in this work, and leave the general case of composable security as an interesting open question.

using any of ($\mathcal{CQ}n1$) or ($\mathcal{C}cn1$) would lead to a weaker notion of quantum indistinguishability. Since we are interested in achieving stronger notions, we will hence consider the more challenging scenarios ($\mathcal{CQ}n2$) and ($\mathcal{C}cn2$).

This argument also leads to the following interesting observation. Ultimately, whether a challenger (or encryption device) performs type-(1) or type-(2) operations depends on its *architecture* which we cannot say anything about - we will focus on the (...2) models in order to be on the 'safe side', as they lead to security notions which are harder to achieve. In order to design a secure encryption device, it is good advice to avoid the possibility that it can be accessed in type-(2) mode. For such a device, it would be sufficient to provide IND-qCPA security, which is weaker and therefore easier to achieve. Clearly, providing guidelines on how to construct encryption devices resilient to type-(2) access lies outside the scope of this work.

3.3 qIND

At this point we are left with only two candidate notions: ($\mathcal{C}cn2$) and ($\mathcal{CQ}n2$). From now on we will denote them as *'quantum indistinguishability of ciphertexts' (qIND)* and *'general quantum indistinguishability of ciphertexts' (gqIND)* resp., and we summarize the resulting challenge phases as follows.

Quantum IND challenge phase (qIND): $\mathcal{A}$ chooses two quantum states ρ_0, ρ_1 having efficient (poly-sized) classical descriptions, and sends to $\mathcal{C}$ a challenge template consisting of these two classical descriptions according to Definition 2.4. $\mathcal{C}$ samples a bit b and replies to $\mathcal{A}$ with the state obtained by applying the type-(2) operator $U^{(2)}_{\mathsf{Enc}_k}$ as defined in (2) to ρ_b. $\mathcal{A}$'s goal is to output b.

General Quantum IND challenge phase (gqIND): $\mathcal{A}$ chooses two quantum states ρ_0, ρ_1, and sends them to $\mathcal{C}$. $\mathcal{C}$ samples a bit b, discards (traces out) ρ_{1-b}, and replies to $\mathcal{A}$ with the state obtained by applying the type-(2) operator $U^{(2)}_{\mathsf{Enc}_k}$ as defined in (2) to ρ_b. $\mathcal{A}$'s goal is to output b.

Using these challenge phases and the notion of a qCPA learning phase, we define qIND-qCPA and gqIND-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 qCPA learning phases and the qIND challenge phase above is at most negligibly close (in n) to $1/2$.*

Definition 3.2 (gqIND-qCPA). *A secret-key encryption scheme is said to be gqIND-qCPA secure if the success probability of any quantum probabilistic polynomial time adversary winning the game defined by qCPA learning phases and the gqIND challenge phase above is at most negligibly close (in n) to $1/2$.*

Since we mainly consider type-(2) transformations from now on, we will overload notation and also use U_{Enc_k} to denote the type-(2) encryption operator.

Theorem 3.3 (gqIND-qCPA ⇒ qIND-qCPA). *Let $\mathcal{E}$ be a gqIND-qCPA secure symmetric-key encryption scheme. Then $\mathcal{E}$ is also qIND-qCPA secure.*

The reason is that quantum states admitting an efficient classical description (used in qIND) are just a special case of arbitrary quantum plaintext states (used in gqIND). Despite this implication, we will mainly focus on the qIND notion in the following, and we will use the gqIND notion only as a comparison to other existing notions. The main reason for this choice is that in the context of classical encryption schemes resistant to superposition quantum access, we believe that it is important to not lose focus of what the capabilities of a 'reasonable' adversary should be. Namely, recall the following classical IND argument: *allowing the adversary to send plaintexts to the challenger is equivalent to the fact that indistinguishability must hold even for the most favorable case from the adversary's perspective.* Such an argument does *not* hold anymore quantumly. In fact, the (Q) model considered in gqIND presents the following issues:

- it allows entanglement between the adversary and the challenger: $\mathcal{A}$ could prepare a state of the form $\rho_{AB} = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$, sending ρ_A as a plaintext but keeping ρ_B;
- it allows the adversary to create certain non-reproduceable states. For example, consider the state $|\psi\rangle = \sum_{x\in X} \frac{1}{\sqrt{|X|}}|x, h(x)\rangle$, where h is a collision-resistant hash function. $\mathcal{A}$ could measure the second register, obtaining a random outcome y, and knowing therefore that the remaining state is the superposition of the preimages of y, $|\psi_y\rangle = \sum_{x\in X:h(x)=y} \frac{1}{\sqrt{|\{\, x\in X:h(x)=y\,\}|}}|x\rangle$. $\mathcal{A}$ could then use $|\psi_y\rangle$ as a plaintext in the challenge phase, but note that $\mathcal{A}$ cannot reproduce $|\psi_y\rangle$ for a given value y.

Both of the above examples are not reasonable in our scenario. Entanglement between $\mathcal{A}$ and $\mathcal{C}$ represents a sort of 'quantum watermarking' of messages, which goes beyond what a meaningful notion of indistinguishability should achieve. Knowledge of intermediate, unpredictable measurements also renders $\mathcal{A}$ too powerful, because it gives $\mathcal{A}$ access to information not available to $\mathcal{C}$ itself - e.g., in the example above $\mathcal{C}$ would not even know the value of y. As it is $\mathcal{C}$ who prepares the state to be encrypted, it is reasonable to assume that it is $\mathcal{C}$ who should know these intermediate measurements, not $\mathcal{A}$. In the example above, what $\mathcal{A}$ could see instead (provided he knows the circuit generating the state, as we assume in qIND) is that the plaintext is a mixture $\Psi = \sum_y \psi_y$ for all possible values of y.

The possibility offered by gqIND of allowing the adversary to play the IND game with arbitrary states is certainly elegant from a theoretical point of view, but from the perspective of the quantum security of the kind of schemes we are considering, it is too broad in scope. The (c) model used in qIND, on the other hand, inherently provides guidelines and reasonable limitations on what a quantum adversary can or cannot do. Also, qIND is often easier to deal with: notice that in the (c) model, unlike in the (Q) model, $\mathcal{A}$ always receives back an unentangled state from a challenge query. In security reductions, this means that we can more easily simulate the challenger, and that we do not have to take care

of measures of entanglement when analyzing the properties of quantum states - for example, indistinguishability of states can be shown by only resorting to the *trace norm* instead of the more general *diamond norm*.

Furthermore, it is important to notice that all our new results in Section 6 are unaffected by the choice of either qIND or gqIND. Our impossibility result from Theorem 6.3 holds for qIND, and hence also for gqIND because of Theorem 3.3. On the other hand, the security proof of Construction 6.6 (Theorem 6.9) is given for gqIND, and holds therefore also for qIND. In fact, it remains unclear whether a separation between qIND and gqIND can be found at all in the realm of classical encryption schemes. We leave this as an interesting open question.

Finally, we note that the q-IND-CPA-2 indistinguishability notion for secret-key encryption of quantum messages introduced by Broadbent and Jeffery [BJ15, Appendix B] resembles our gqIND notion, and it is in fact equivalent to it in the case that the encryption operation is a symmetric-key classical functionality operating in type-(2) mode.

Theorem 3.4 (gqIND-qCPA $\Leftrightarrow$ q-IND-CPA-2). *Let $\mathcal{E}$ be a symmetric-key encryption scheme. Then $\mathcal{E}$ is gqIND-qCPA secure iff $\mathcal{E}$ is q-IND-CPA-2 secure.*

A proof of the above theorem can be found in Appendix C. A generalization of q-IND-CPA-2 to arbitrary quantum encryption schemes, together with equivalent notions of quantum semantic security, was given and analized in [ABF+16]. All these security notions are given in the context of 'fully quantum encryption', in the sense that the encryption schemes considered in [BJ15] and [ABF+16] are arbitrary quantum circuits acting natively on quantum data, while in this work we consider the quantum security of classical encryption schemes. The fully quantum homomorphic schemes which are shown to be secure in [BJ15], and the other quantum encryption schemes shown to be secure in [ABF+16], do not fall into the category of *classical* encryption schemes which we are studying here. On the other hand, as Theorem 6.9 shows, our Construction 6.6 is the first known example of a classical symmetric-key encryption scheme which is secure even against these kinds of 'fully quantum' security notions.

4 New Notions of Quantum Semantic Security

In this section, we initiate the study of suitable definitions of semantic security in the quantum world. As in the classical case, we are particularly interested in notions that can be proven equivalent to some version of quantum indistinguishability. 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.

Working towards a quantum SEM notion, we restrict our analysis to the SEM challenge phase. For the learning phase, we stick to the 'qCPA learning phase', as in Definition 2.5, where the adversary has access to a quantum encryption oracle. In the end, we give a definition for quantum semantic security under quantum chosen-plaintext attacks (qSEM-qCPA) which we later prove equivalent to qIND-qCPA, thereby adding semantics to our qIND-qCPA notion.

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 game defined by qCPA learning phases and a SEM challenge phase is negligibly close (in n) to the success probability of $\mathcal{S}$ winning the reduced game.*

Spoiler. It is easy to see that the SEM-qCPA notion of semantic security is equivalent to IND-qCPA, see Theorem 5.1.

In Appendix D we discuss what happens if one also allows quantum advice states in this scenario, and why this option would not add anything meaningful.

4.2 Quantum Semantic Security

Here we define *quantum semantic security under chosen-plaintext attacks* (qSEM-qCPA). As in the classical case, we want the definition of semantic security to formally capture what we intuitively understand as a strong security notion. In the quantum case, there are several choices to be made. We start by giving our formal definition of quantum semantic security, and justify our choices afterwards.

Quantum SEM (qSEM) challenge phase: $\mathcal{A}$ sends to $\mathcal{C}$ a challenge template consisting of classical decriptions of

- a quantum circuit G_m taking poly (n)-bit classical input and outputting m-qubit plaintext states,
- a quantum circuit h_m taking m-qubit plaintexts as input and outputting poly (n)-qubit advice states,
- a quantum circuit f_m taking m-qubit plaintexts as input and outputting poly (n)-qubit target states.

The challenger $\mathcal{C}$ samples $y \xleftarrow{\$} \{0,1\}^{\mathrm{poly}(n)}$ and computes two copies of the plaintext $\rho_y = G_m(y)$. One is used to compute auxiliary information $h_m(\rho_y)$ and one to compute the ciphertext $U_{\mathsf{Enc}_k}\, \rho_y\, U_{\mathsf{Enc}_k}^{\dagger}$. $\mathcal{C}$ then replies with the pair

$\left(U_{\mathsf{Enc}_k}\,\rho_y\,U^\dagger_{\mathsf{Enc}_k}, h_m(\rho_y)\right)$. $\mathcal{A}$'s goal is to output $f_m(\rho_y)$. We say that $\mathcal{A}$ *wins the qSEM-qCPA game* if no quantum polynomial-time distinguisher can distinguish $\mathcal{A}$'s output from the target state $f_m(\rho_y)$ with non-negligible advantage.

In the reduced game, $\mathcal{S}$ receives no encryption, but only the auxiliary information $h_m(\rho_y)$ from $\mathcal{C}$. Analogously to the above case, $\mathcal{S}$*wins the qSEM-qCPA game* if no quantum polynomial-time distinguisher can distinguish $\mathcal{S}$'s output from the target state $f_m(\rho_y)$ with non-negligible advantage.

Definition 4.2 (qSEM-qCPA). *A secret-key encryption scheme is called qSEM-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 game defined by qCPA learning phases and a qSEM challenge phase is negligibly close (in n) to the success probability of $\mathcal{S}$ winning the reduced game.*

When defining quantum semantic security, we have to deal with several issues: First, we have to define how the plaintext distribution is described. In the classical definition, the distribution is produced by a (classical) circuit G_m running on uniform input bits. We take the same approach here, but let G_m output m-qubit plaintexts.

The second question is how to define the advice function. While the input should be the plaintext quantum state ρ_y, the output could be either quantum or classical. We decided to allow quantum advice as it leads to a more general model and it includes classical outputs as a special case. In order for the challenger to compute both the encryption of the plaintext state ρ_y and the advice state $h_m(\rho_y)$ without violation of the no-cloning theorem, we exploit how we generate the message state. We simply run S_m twice on the same classical randomness y to generate two copies of the plaintext state ρ_y. Another option would have been to allow for entanglement between the plaintext message ρ_y and the advice state $h_m(\rho_y)$. Allowing such entanglement would model side-channel information the attacker could obtain, for instance by learning the content of some internal register of the attacked device. However, the resulting notion would not be equivalent with qIND-qCPA anymore, because in qIND-qCPA, the challenge plaintexts are provided by their classical descriptions and can therefore not be entangled with the attacker.

Third, we have chosen to model the target function f_m in the same way as the advice function h_m, i.e. we allow arbitrary quantum circuits that might output quantum states. The reasoning behind allowing quantum output is again to use the strongest possible, most general model. Allowing quantum output however leads to the problem that, in general, we cannot physically test anymore if an adversary $\mathcal{A}$ outputs exactly the result of the target function $f_m(\rho_y)$. One option would be to require $\mathcal{A}$'s output to be close to $f_m(\rho_y)$ in terms of their trace distance. But two quantum states can be quantum-polynomial-time indistinguishable even if their trace distance is large[10]. Since we are only inter-

[10] Think of two different classical ciphertexts which are encrypted using a quantum-computationally secure encryption scheme. Then, the ciphertext states are orthog-

ested in computational security notions, we solve this problem by requiring QPT indistinguishability as success condition for winning the SEM game.

Spoiler. Our qSEM-qCPA 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

In this section we show relations between our new notions of indistinguishability and semantic security in the quantum world. It is already known [GM84,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 2.

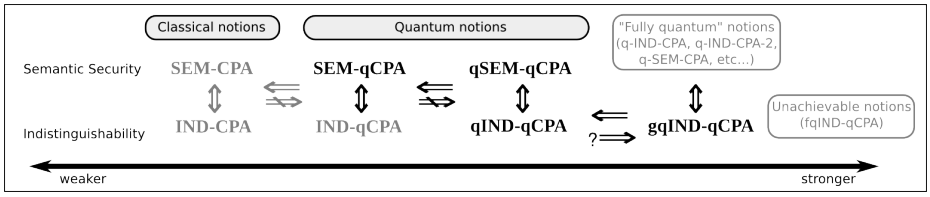

Fig. 2. The relations between notions of indistinguishability and semantic security in the quantum world (previously known results in gray.)

We start by proving equivalence between IND-qCPA and SEM-qCPA.

Theorem 5.1 (IND-qCPA $\Leftrightarrow$ SEM-qCPA). *Let $\mathcal{E}$ be a symmetric-key encryption scheme. Then $\mathcal{E}$ is IND-qCPA secure iff $\mathcal{E}$ 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 Th. 5.4.11]), we recall them as they work as guidelines for the following proofs.

Proposition 5.2 (IND-qCPA $\Rightarrow$ SEM-qCPA).

Proposition 5.3 (SEM-qCPA $\Rightarrow$ IND-qCPA).

Proof (of Proposition 5.2 – Sketch.). The idea of the proof is to hand $\mathcal{A}$'s circuit as non-uniform advice to the simulator $\mathcal{S}$. $\mathcal{S}$ 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

onal (and hence their trace distance is maximal), but they are computationally indistinguishable.

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 qCPA 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 (G_m, h_m, f_m) with G_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 qCPA 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. □

Next, we show equivalence between qIND-qCPA and qSEM-qCPA.

Theorem 5.4 (qIND-qCPA ⇔ qSEM-qCPA). *Let $\mathcal{E}$ be a symmetric-key encryption scheme. Then $\mathcal{E}$ is qIND-qCPA secure iff $\mathcal{E}$ 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, $\mathcal{A}$'s 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 $\mathcal{S}$ simulates $\mathcal{C}$ for $\mathcal{A}$ by generating a new key and answering all of $\mathcal{A}$'s qCPA queries. When it comes to the challenge query, $\mathcal{A}$ produces a qSEM template, which $\mathcal{S}$ forwards to the real $\mathcal{C}$. Then $\mathcal{S}$ forwards $\mathcal{C}$'s reply, plus a bogus encrypted state (e.g., $U_{\mathsf{Enc}_k} \, |1 \dots 1\rangle$), to $\mathcal{A}$. If at this point $\mathcal{A}$ outputs a state φ which can be efficiently distinguished from the correct $f_m(\rho_y)$ computed by the real $\mathcal{C}$, we would have an efficient distinguisher against the qIND-qCPA security of the scheme. Hence, $\mathcal{A}$'s (and therefore also $\mathcal{S}$'s) output must be indistinguishable from $f_m(\rho_y)$ for any QPT distinguisher, which 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 $\mathcal{M}^{\mathcal{A}}$ just forwards everything to the qSEM-qCPA challenger $\mathcal{C}$. When $\mathcal{A}$ performs a challenge qIND query by sending the classical descriptions of two states φ_0 and φ_1, $\mathcal{M}^{\mathcal{A}}$ prepares the qSEM template (G_m, h_m, f_m), with G_m outputing φ_0 for half of the possible y values and φ_1 for the other half, $h_m(\rho_y) = 1^n$, and f_m the identity map $f_m(\rho_y) = \rho_y$. Then $\mathcal{M}^{\mathcal{A}}$ performs a qSEM challenge query with this template. Given challenge ciphertext state $U_{\mathsf{Enc}_k}\,\varphi_b\,U_{\mathsf{Enc}_k}^{\dagger}$ (for $b \in \{0,1\}$), he forwards it as an answer to $\mathcal{A}$'s challenge query. As $\mathcal{A}$ distinguishes $U_{\mathsf{Enc}_k}\,\varphi_0\,U_{\mathsf{Enc}_k}^{\dagger}$ from $U_{\mathsf{Enc}_k}\,\varphi_1\,U_{\mathsf{Enc}_k}^{\dagger}$ with non-negligible success probability, $\mathcal{A}$ returns the correct value of b with non-negligible advantage over guessing. Then $\mathcal{M}^{\mathcal{A}}$, having recorded a copy of the classical descriptions of φ_0 and φ_1, is able to compute the state $f_m(\varphi_b)$ exactly, and consequently win the qSEM-qCPA game with non-negligible advantage. As h_m generates the same advice state $h_m(\rho_y) = 1^n$ independently of the encrypted message, no simulator can do better than guessing the plaintext. 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 (and equivalently qSEM-qCPA) is a strictly stronger notion than IND-qCPA (which is equivalent to 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: $|\varphi_0\rangle = H\,|0\rangle = \frac{1}{\sqrt{2}}\,|0\rangle + \frac{1}{\sqrt{2}}\,|1\rangle$, and $|\varphi_1\rangle = H\,|1\rangle = \frac{1}{\sqrt{2}}\,|0\rangle - \frac{1}{\sqrt{2}}\,|1\rangle$. 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 $|\varphi_0\rangle$ and $|\varphi_1\rangle$ have the same amplitudes on all their components, except for the sign. As these two states are orthogonal, they can be reliably distinguished by the adversary $\mathcal{A}$ who 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 (and equivalently qSEM-qCPA) is impossible to achieve for encryption schemes which do not expand the message (such as stream ciphers and many block ciphers, without considering 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 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 and a part carrying the message-dependent information. This definition 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* $(r, f(k, r, x))$*, where* $r \in \{0,1\}^\tau$ *is independent of the message; 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 Construction 5.3.9 from [Gol04] (where $Enc_k(x)$ is defined as $(r, F_k(r) \oplus x)$ for a PRF F) the core function is $f(k, r, x) = F_k(r) \oplus x$, with $f'(k, r, z) = z \oplus F_k(r)$.

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, Construction 5.3.9 from [Gol04] 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 bitstrings (with 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 unitary type-(2) encryption operation (because it will just 'shuffle' in the basis-state space amplitudes which are exactly the same).

Theorem 6.3 (Impossibility Result). *No quasi–length-preserving secret-key encryption scheme can be qIND 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: $|\varphi_0\rangle = H|0^m\rangle, |\varphi_1\rangle = H|1^m\rangle$, where H is the m-fold tensor Hadamard transformation. Notice that both these states admit efficient classical representations, and are thus allowed in the qIND game.
2. The challenger flips a random bit b and returns $|\psi\rangle = U_{\mathsf{Enc}_k}|\varphi_b\rangle$.
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 $|\varphi_0\rangle$ 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 $b = 1$, $|\varphi_1\rangle = \frac{1}{2^{m/2}}\sum_y (-1)^{y\cdot 1^m}|y\rangle$ where $a\cdot b$ denotes the bitwise inner product between a and b. Hence, $|\varphi_1\rangle$ 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. □

Notice that the above attack also works if $\mathcal{A}$ is allowed to send quantum states to $\mathcal{C}$ directly. Therefore, it also holds for the gqIND notion of quantum indistinguishability described in Section 3. In particular, the above theorem shows that [Gol04, Construction 5.3.9], which in [BZ13] was shown to be IND-qCPA if the used PRF is quantum secure, does not fulfill qIND, nor gqIND.

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 hide the basis bit, and one to hide 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.

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 (see Appendix A for formal definitions).

Construction 6.4. *For security parameter n, let $m = poly(n)$ and $\tau = poly(n)$. Consider an efficient 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. The security is stated in the following theorem.

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

In the next section, we prove the security of a more powerful scheme which includes the above theorem as special case of a single message block.

6.3 Length Extension

Construction 6.4 has the drawback that the message length is upper bounded by the input length of the qPRP (minus the bit length of the randomness). However, like in the case of block ciphers, we can overcome this issue with a *mode of operation*. More specifically, we can handle arbitrary message lengths by splitting the message into m-bit blocks and applying the encryption algorithm of Construction 6.4 independently to each message block (using the same key

but new randomness for each block). This procedure is akin to a 'randomized ECB mode', in the sense that each message block is processed separately, like in the ECB (Electronic Code Book) mode, but in our case the underlying cipher is inherently randomized (since we use fresh randomness for each block), so we can still achieve qCPA security. For simplicity we consider only message lengths which are multiples of m. The construction can be generalized to arbitrary message lengths using standard padding techniques. Moreover, the randomness for every block can be generated efficiently using a random seed and a post-quantum secure PRNG.

Construction 6.6. *For security parameter n, let $m = poly(n)$ and $\tau = poly(n)$. Consider an efficient 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 $\mathcal{M} = \{0,1\}^{\mu m}$ for $\mu \in \mathbb{N}, \mu = poly(n)$, key space $\mathcal{K} = \mathcal{K}_\Pi$, and ciphertext space $\mathcal{C} = \{0,1\}^{\mu(m+\tau)}$. 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 splits x into μ m-bit blocks $x_1, \ldots, x_\mu$. For each block x_i, the encryption algorithm samples a new τ-bit string $r_i \xleftarrow{\$} \{0,1\}^\tau$ uniformly at random, and outputs $y_i = \pi_k(x_i \| r_i)$ ($\|$ denotes string concatenation). The ciphertext is $y = y_1 \| \ldots \| y_\mu$.*

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 splits y into μ $m+\tau$-bit blocks $y_1, \ldots, y_\mu$. Then, it runs $x'_i = (\pi_k^{-1}(y_i))_m$ for each block (where $(s)_m$ refers to taking the first m bits of bit string s). It returns the plaintext $x' = x'_1, \ldots, x'_\mu$.*

The soundness of the construction can be checked easily. For the security, we observe that splitting a μm-qubit plaintext state into μ blocks of m-qubits can introduce entanglement between the blocks. We will address this issue through the following technical lemma.

Lemma 6.7. *Let $\mathcal{E}$ be the quantum channel that takes as input an arbitrary m-qubit state, attaches another τ qubits in state $|0\rangle$, and then applies a permutation picked uniformly at random from $S_{2^{m+\tau}}$ to the computational basis space. Let $\mathcal{T}$ be the constant channel which maps any m-qubit state to the totally mixed state on $m+\tau$ qubits. Then, $\|\mathcal{E} - \mathcal{T}\|_\diamond \leq 2^{-\tau+2}$.*

Proof. In order to consider the fact that the m-qubit input state might be entangled with something else, we have to start with a purification of such a state. Formally, this is a bipartite pure $2m$-qubit state $|\phi\rangle_{XY} = \sum_{x,y} \alpha_{x,y} |x\rangle_X |y\rangle_Y$ whose m-qubit Y register is input into the channel and gets transformed into $id_X \otimes \mathcal{E}(|\phi\rangle\langle\phi|) = \mathrm{tr}_\Pi |\psi\rangle\langle\psi|$ where

$$|\psi\rangle = \sum_{x \in \{0,1\}^m, y \in \{0,1\}^m, \pi \in S_{2^{m+\tau}}} \alpha_{x,y} |x\rangle_X |\pi(y||0)\rangle_C |\pi\rangle_\Pi \ .$$

By definition of the diamond-norm, we have to show that for any $2m$-qubit state ρ, we have that $\|(id \otimes \mathcal{E})(\rho) - (id \otimes \mathcal{T})(\rho)\|_{\mathrm{tr}} \leq 2^{-\tau+2}$. Due to the convexity of the trace distance, we may assume that $\rho = |\phi\rangle\langle\phi|$ is pure with $|\phi\rangle_{XY} = \sum_{x,y} \alpha_{x,y} |x\rangle_X |y\rangle_Y$. Hence, we obtain

$$\begin{aligned}
(id_X \otimes \mathcal{E})(|\phi\rangle\langle\phi|) &= \mathrm{tr}_{\Pi} |\psi\rangle\langle\psi| \\
&= \frac{1}{2^{m+\tau}!} \sum_{x,x',y,y',\pi} \alpha_{x,y}\overline{\alpha_{x',y'}} |x\rangle\langle x'|_X \otimes |\pi(y\|0)\rangle \langle\pi(y'\|0)|_C \\
&= \frac{1}{2^{m+\tau}!} \sum_{x,x',y} \alpha_{x,y}\overline{\alpha_{x',y}} |x\rangle\langle x'|_X \otimes \sum_{\pi} |\pi(y\|0)\rangle \langle\pi(y\|0)|_C \\
&\quad + \frac{1}{2^{m+\tau}!} \sum_{x,x',y\neq y'} \alpha_{x,y}\overline{\alpha_{x',y'}} |x\rangle\langle x'|_X \otimes \sum_{\pi} |\pi(y\|0)\rangle \langle\pi(y'\|0)|_C \\
&= \sum_{x,x',y} \alpha_{x,y}\overline{\alpha_{x',y}} |x\rangle\langle x'|_X \otimes \frac{1}{2^{m+\tau}} \sum_{z} |z\rangle\langle z|_C \\
&\quad + \sum_{x,x',y\neq y'} \alpha_{x,y}\overline{\alpha_{x',y'}} |x\rangle\langle x'|_X \otimes \frac{1}{2^{m+\tau}(2^{m+\tau}-1)} \sum_{z\neq z'} |z\rangle\langle z'|_C \\
&= \mathrm{tr}_Y |\phi\rangle\langle\phi| \otimes \tau_C + \chi_{XC} \\
&= (id_X \otimes \mathcal{T})(|\phi\rangle\langle\phi|) + \chi_{XC} \,,
\end{aligned}$$

where we defined the "difference state"

$$\chi_{XC} := \sum_{x,x',y\neq y'} \alpha_{x,y}\overline{\alpha_{x',y'}} |x\rangle\langle x'|_X \otimes \frac{1}{2^{m+\tau}(2^{m+\tau}-1)} \sum_{z\neq z'} |z\rangle\langle z'|_C \,.$$

In order to conclude, it remains to show that $\|\chi_{XC}\|_{\mathrm{tr}} \leq 2^{-\tau+2}$. For the C-register $\chi_C = \frac{1}{2^{m+\tau}(2^{m+\tau}-1)} \sum_{z\neq z'} |z\rangle\langle z'|_C$, one can verify that the $2^{m+\tau}$ eigenvalues are $(c \cdot (2^{m+\tau}-1), -c, -c, \ldots, -c)$ where $c := \frac{1}{2^{m+\tau}(2^{m+\tau}-1)}$. Hence, the trace norm (which is the sum of the absolute eigenvalues) is exactly $c \cdot 2(2^{m+\tau} - 1) = 2^{-m-\tau+1}$.

For the X-register, we split χ_X into two parts $\chi_X = \xi_X - \xi'_X$ where

$$\begin{aligned}
\xi_X &:= \sum_{x,x'} |x\rangle\langle x'| \sum_{y,y'} \alpha_{x,y}\overline{\alpha_{x',y'}} \,, \\
\xi'_X &:= \sum_{x,x'} |x\rangle\langle x'| \sum_{y} \alpha_{x,y}\overline{\alpha_{x',y}} \,,
\end{aligned}$$

and use the triangle inequality for the trace norm $\|\chi_X\|_{\mathrm{tr}} = \|\xi_X - \xi'_X\|_{\mathrm{tr}} \leq \|\xi_X\|_{\mathrm{tr}} + \|\xi'_X\|_{\mathrm{tr}}$. Observe that $\|\xi_X\|_{\mathrm{tr}} = \|\sum_{x,y} \alpha_{x,y} |x\rangle \sum_{x',y'} \overline{\alpha_{x',y'}} \langle x'|\|_{\mathrm{tr}} = \||s\rangle\langle s|\|_{\mathrm{tr}}$ for the (non-normalized) vector $|s\rangle := \sum_{x,y} \alpha_{x,y} |x\rangle$. Hence, the trace-norm $\|\xi_X\|_{\mathrm{tr}} = |\langle s \mid s\rangle| = \sum_x |\sum_y \alpha_{x,y}|^2 \leq \sum_x \sum_y |\alpha_{x,y}|^2 \cdot 2^m = 2^m$ by the Cauchy-Schwarz inequality and the normalization of the $\alpha_{x,y}$'s. Furthermore, we note that ξ'_X is exactly the reduced density matrix of $|\phi\rangle_{XY}$ after tracing out

the Y register. Hence, ξ'_X is positive semi-definite and its trace norm is equal to its trace which is 1. In summary, we have shown that

$$\begin{aligned}\|\chi_{XC}\|_{\mathrm{tr}} &= \|\chi_X\|_{\mathrm{tr}} \cdot \|\chi_C\|_{\mathrm{tr}} \leq (\|\xi_X - \xi'_X\|_{\mathrm{tr}}) \cdot 2^{-m-\tau+1} \\ &\leq (\|\xi_X\|_{\mathrm{tr}} + \|\xi'_X\|_{\mathrm{tr}}) \cdot 2^{-m-\tau+1} \leq (2^m + 1) \cdot 2^{-m-\tau+1} \leq 2^{-\tau+2}\,.\end{aligned}$$

□

If we consider a slightly different encryption channel $\mathcal{E}^T$ which still maps m qubits to $m+\tau$ qubits but where the permutation π is not picked uniformly from $S_{2^{m+\tau}}$, but instead we are guaranteed that a certain set $T \subset \{0,1\}^{m+\tau}$ of outputs never occurs, we can consider such permutations w.l.o.g. as picked uniformly at random from a smaller set $S_{2^{m+\tau}-|T|}$. In this setting, we are interested in the distance of the encryption operation $\mathcal{E}^T$ from the slightly different constant channel $\mathcal{T}^T$ which maps all inputs to the $(m+\tau)$-qubit state which is completely mixed on the smaller set $\{0,1\}^{m+\tau} \setminus T$. By modifying slightly the proof of Lemma 6.7 we get the following.

Corollary 6.8. *Let $\mathcal{E}^T$ and $\mathcal{T}^T$ be the channels defined above. Then,*

$$\|\mathcal{E}^T - \mathcal{T}^T\|_\diamond \leq \frac{4}{2^\tau - |T|/2^m}\,. \tag{3}$$

We can now prove the security of Construction 6.6. We give the proof for gqIND-qCPA, and then qIND-qCPA follows immediately from Theorem 3.3.

Theorem 6.9 (gqIND-qCPA security of Construction 6.6). *If $\Pi_{m+\tau}$ is a family of quantum-secure pseudorandom permutations (qPRP), then the encryption scheme* $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *defined in Construction 6.6 is gqIND-qCPA secure.*

Proof. We want to show that no QPT distinguisher $\mathcal{D}$ can win the gqIND-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 gqIND-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 are μ new random τ-bit strings that were 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 1 the same

randomness is sampled twice only with negligible probability, 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 $r_1, \ldots, r_\mu$ 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.

When the modified gqIND game 3 starts, $\mathcal{D}$ chooses two different plaintext states and sends them to the challenger, who will then choose one of them and send it back encrypted with fresh randomness $\hat{r}_1, \ldots, \hat{r}_\mu$. Let Q denote the set of $q \cdot \mu = poly(n)$ 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 any $\pi(x\|\hat{r}_i)$ will not take one of these values as $\hat{r}_i$ 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 $|T| = q \cdot \mu \cdot 2^m$, 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 (classical) ciphertexts for every of the $q \cdot \mu$ 'taken' randomnesses, but as we will see, this knowledge will be still insufficient to win the game.)

By construction, the encryption of a μm-qubit (possibly mixed) state σ is performed in μ separate blocks of m qubits each. We are guaranteed that fresh randomness is used in each block, hence it follows from Corollary 6.8 that $\mathsf{Enc}_k(\sigma)$ is negligibly close to the ciphertext state where the first $m+\tau$ qubits are replaced with the completely mixed state (by noting that $|T|/2^m = q \cdot \mu$ is polynomial in n in our case, and hence the right-hand side of (3) is negligible.). Another application of Corollary 6.8 gives negligible closeness to the ciphertext state where the first $2(m+\tau)$ qubits are replaced with the completely mixed state etc. After μ applications of Corollary 6.8, we have shown that $\mathsf{Enc}_k(\sigma)$ is negligibly close to the totally mixed state on $\mu(m+\tau)$ qubits. As this argument can be made for any cleartext state σ, we have shown that from $\mathcal{D}$'s point of view, all encrypted states are negligibly close to the totally mixed state and therefore cannot be distinguished. □

Corollary 6.10 (qIND-qCPA security of Construction 6.6). *If $\Pi_{m+\tau}$ is a family of quantum-secure pseudorandom permutations (qPRP), then the encryption scheme* $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ *defined in Construction 6.6 is qIND-qCPA secure.*

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 leads to 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 CCA1 security in this framework. This topic was also initiated in [BZ13] relative to the IND-qCPA model; it would be interesting to extend the definition of CCA1 security to stronger notions obtained by starting from our qIND-qCPA model.

In Section 3.3 we left open the interesting question on whether it is possible at all to find a separating example between the notions of qIND and gqIND. That is, find a symmetric-key encryption scheme $\mathcal{E}$ which is qIND-secure, but not gqIND-secure. Finding such an example (or provable lack of) would shed further light on the security model we consider.

We have so far not taken into account models where the adversary is allowed to initialize the ancilla qubits used in the encryption operation used by the challenger (i.e. the $|y\rangle$ in $|x, y\rangle \mapsto |x, y \oplus \mathsf{Enc}_k(x)\rangle$). These models lead to the study of *quantum fault attacks*, because they model cases where the adversary is able to 'watermark' or tamper with part of the challenger's internal memory. 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.

One outstanding open problem is to define CCA2 (adaptive chosen ciphertext attack) security in the quantum world. The problem is that in the CCA2 game the challenger has to ensure that the attacker does not ask for a decryption of the actual challenge ciphertext leading to a trivial break. While this is easily implemented in the classical world, it raises several issues in the quantum world. What does it mean for a ciphertext to be different from the challenge ciphertext? And, more importantly: *How can the challenger check?* There might be several reasonable ways to solve the first issue but, as long as the queries are not classical, we are not aware of any possibility to solve the second issue without disturbing the challenge ciphertext and the query states.

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-qCPA secure scheme into a qIND-qCPA secure one. It is also important to study how our transformation can be applied to modes of operation different from Construction 6.6.

Acknowledgements. The authors would like to thank Ronald de Wolf and Boris Škorić for helpful discussions, the anonymous reviewers for useful comments, Marco Tomamichel for the bibstyle, and the organizers of the Dagstuhl Seminar 15371 "Quantum Cryptanalysis" for providing networking and useful interactions and support. T.G. was supported by the German Federal Ministry of Education and Research (BMBF) within EC SPRIDE and CROSSING. A.H. was supported by the Netherlands Organisation for Scientific Research (NWO) under grant 639.073.005 and the Commission of the European Communities through the Horizon 2020 program under project number 645622 PQCRYPTO. C.S. was supported by a 7th framework EU SIQS grant and a NWO VIDI grant. Part of this work was supported by the COST Action IC1306.

References

ABF+16. G. Alagic, A. Broadbent, B. Fefferman, T. Gagliardoni, C. Schaffner, and M. S. Jules. Computational security of quantum encryption. http://arxiv.org/abs/1602.01441, 2016.

BBD09. D. J. Bernstein, J. Buchmann, and E. Dahmen, editors. *Post-Quantum Cryptography*. Springer Berlin Heidelberg, 2009. DOI: 10.1007/978-3-540-88702-7.

BDF+11. D. Boneh, O. Dagdelen, M. Fischlin, A. Lehmann, C. Schaffner, and M. Zhandry. Random Oracles in a Quantum World. In *ASIACRYPT 2011*, volume 7073 of *LNCS*, pages 41–69, 2011. DOI: 10.1007/978-3-642-25385-0_3.

BHT97. G. Brassard, P. Hoyer, and A. Tapp. Quantum Algorithm for the Collision Problem. arXiv:quant-ph/9705002, 1997. arXiv: quant-ph/9705002.

BJ15. A. Broadbent and S. Jeffery. Quantum homomorphic encryption for circuits of low T-gate complexity. In *CRYPTO 2015*, pages 609–629, 2015. DOI: 10.1007/978-3-662-48000-7_30.

BZ13. D. Boneh and M. Zhandry. Secure Signatures and Chosen Ciphertext Security in a Quantum Computing World. In *CRYPTO 2013*, volume 8043 of *LNCS*, pages 361–379, 2013. DOI: 10.1007/978-3-642-40084-1_21.

DFG13. Ö. Dagdelen, M. Fischlin, and T. Gagliardoni. The fiat-shamir transformation in a quantum world. In *ASIACRYPT 2013, Part II*, pages 62–81, 2013.

DFNS14. I. Damgård, J. Funder, J. B. Nielsen, and L. Salvail. Superposition Attacks on Cryptographic Protocols. In *Information Theoretic Security*, volume 8317 of *LNCS*, pages 142–161, 2014. DOI: 10.1007/978-3-319-04268-8_9.

GM84. S. Goldwasser and S. Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2): 270–299, 1984. DOI: 10.1016/0022-0000(84)90070-9.

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

Gro96. L. K. Grover. A fast quantum mechanical algorithm for database search. In *STOC '96*, pages 212–219. ACM Press, 1996. DOI: 10.1145/237814.237866.

KKVB02. E. Kashefi, A. Kent, V. Vedral, and K. Banaszek. Comparison of quantum oracles. *Physical Review A*, 65(5): 050304, 2002. DOI: 10.1103/PhysRevA.65.050304.

KLLNP16. M. Kaplan, G. Leurent, A. Leverrier, and M. Naya-Plasencia. Breaking symmetric cryptosystems using quantum period finding, 2016. arXiv: 1602.05973. to appear at CRYPTO 2016.

KM10. H. Kuwakado and M. Morii. Quantum distinguisher between the 3-round Feistel cipher and the random permutation. In *ISIT 2010*, pages 2682–2685, 2010. DOI: 10.1109/isit.2010.5513654.

KM12. H. Kuwakado and M. Morii. Security on the quantum-type Even-Mansour cipher. In *ISITA 2012*, pages 312–316, 2012. Online: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6400943.

Kos07. T. Koshiba. Security notions for quantum public-key cryptography. *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, J90-A(5): 367–375, 2007.

LP11. R. Lindner and C. Peikert. Better Key Sizes (and Attacks) for LWE-Based Encryption. In A. Kiayias, editor, *CT-RSA 2011*, volume 6558 of *LNCS*, pages 319–339. Springer, 2011.
DOI: 10.1007/978-3-642-19074-2_21.

McE78. R. J. McEliece. A public-key cryptosystem based on algebraic coding theory. *DSN progress report*, 42(44): 114–116, 1978.
Online: http://www.cs.colorado.edu/~jrblack/class/csci7000/f03/papers/mceliece.pdf.

NC00. M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

Sho94. P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. In *FOCS'94*, pages 124–134. IEEE Comput. Soc. Press, 1994.
DOI: 10.1109/SFCS.1994.365700.

SS16. T. Santoli and C. Schaffner. Using Simon's algorithm to attack symmetric-key cryptographic primitives, 2016.
arXiv: 1603.07856.

Unr12. D. Unruh. Quantum Proofs of Knowledge. In D. Pointcheval and T. Johansson, editors, *EUROCRYPT 2012*, volume 7237 of *LNCS*, pages 135–152. Springer, 2012.
DOI: 10.1007/978-3-642-29011-4_10.

Vel13. M. Velema. Classical Encryption and Authentication under Quantum Attacks. Master's thesis, Master of Logic, University of Amsterdam, 2013.
Online: http://arxiv.org/abs/1307.3753.

Wat01. J. Watrous. Quantum algorithms for solvable groups. In *STOC '01*, pages 60–67. ACM Press, 2001.
DOI: 10.1145/380752.380759.

Wat09. J. Watrous. Zero-Knowledge against Quantum Attacks. *SIAM Journal on Computing*, 39(1): 25–58, 2009.
DOI: 10.1137/060670997.

XY12. C. Xiang and L. Yang. Indistinguishability and semantic security for quantum encryption scheme. *Proc. SPIE*, 8554: 85540G–85540G–8, 2012.
DOI: 10.1117/12.999846.

Zha12. M. Zhandry. How to Construct Quantum Random Functions. In *FOCS 2012*, pages 679–687. IEEE, 2012.
DOI: 10.1109/FOCS.2012.37.

A Formal Definitons

Here we give some formal definitions that we omitted in the main body as they are somewhat standard. We include them for the paper to be self-contained. We begin with detailed formal definitions for SEM-CPA and IND-CPA. Afterwards we define quantum-secure pseudorandom permutations.

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 iff 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*

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

$$< \quad \mathsf{Pr}\left[\begin{array}{ll} v = f_m(x) & \text{where} \\ & ((S_m, h_m, f_m), \sigma) \longleftarrow \mathcal{A}'_1(1^n, z) \\ & x \longleftarrow S_m(U_{poly(n)}) \\ & v \longleftarrow \mathcal{A}'_2(\sigma, 1^{|x|}, h_m(x)) \end{array}\right] + \tfrac{1}{p(n)} \qquad (4)$$

   Recall that (S_m, h_m, f_m) *is a triplet of circuits consisting of a poly-sized circuit* S_m *specifying a distribution over* m*-bit long plaintexts, a circuit computing an advise function* $h_m : \{0,1\}^m \to \{0,1\}^*$*, and a circuit computing a target function* $f_m : \{0,1\}^m \to \{0,1\}^*$*, 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 iff 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

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

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.

Quantum PRP. We now define quantum-secure pseudorandom permutation families. We restrict ourselves to efficient permutation families that have as domain binary strings of a certain length as these are the only ones we are using in this work. Let S_{2^n} be the set of all permutations of n-bit strings.

Definition A.3 (Efficient Permutation Family). *Let $n \in \mathbb{N}$, we call a family of permutations $\Pi_n = \{\pi_k : \{0,1\}^n \to \{0,1\}^n\} \subset S_{2^n}$ with key space $\mathcal{K}_\Pi$ and domain $\{0,1\}^n$ efficient if there exists a triple of probabilistic polynomial-time algorithms $(\mathcal{I}, \Pi, \Pi^{-1})$ such that:*

1. *The initialization algorithm $\mathcal{I}(1^n)$ takes as input the parameter n and outputs a random function key $k \stackrel{\$}{\longleftarrow} \mathcal{K}_\Pi$ from the key space.*
2. *The function Π takes as input a function key k and a domain element x and outputs $\pi_k(x)$.*
3. *The function Π^{-1} takes as input a function key k and a domain element x and outputs $\pi_k^{-1}(x)$.*

We sometimes abuse notation and write π instead of π_k and $\pi \stackrel{\$}{\longleftarrow} \Pi_n$ for the process of running $\mathcal{I}(1^n)$. A quantum-secure pseudorandom permutation family (qPRP) is an efficient permutation family that achieves the pseudorandomness property in presence of a quantum adversary that can query the permutation π with superpositions of domain elements x. It is defined as follows:

Definition A.4 (Quantum PRP). *An efficient permutation family Π_n is said to be a* quantum-secure pseudorandom permutation family *if for every quantum polynomial-time oracle machine $\mathcal{A}$, it holds that*

$$\left|\mathsf{Pr}_{\pi \stackrel{\$}{\longleftarrow} \Pi_n}\left[\mathcal{A}^{|\pi\rangle}(1^n) = 1\right] - \mathsf{Pr}_{\pi \stackrel{\$}{\longleftarrow} S_{2^n}}\left[\mathcal{A}^{|\pi\rangle}(1^n) = 1\right]\right| \leq negl(n) \ ,$$

where the superscript $|\cdot\rangle$ denotes oracle access in superposition.

Note that the permutations are chosen by the game. Hence, keys are classical.

A permutation family Π_n is called a *strong quantum PRP*, if a random member of Π_n is computationally indistinguishable from a uniform permutation even if the attacker $\mathcal{A}$ can query (in superposition) both the permutation π and the inverse permutation π^{-1}. Notice that the construction in Theorem 6.5 does not

require *strong* quantum PRPs. The reason is that, even if we are considering type-(2) transformations (which could be used to compute π^{-1}), these transformations are implemented by the challenger, because we are in the $(\mathcal{C})$ model. And since we only consider CPA scenarios here, and not CCA, the adversary is never granted access to the decryption oracle. Hence, π^{-1} is not needed by the reduction.

B Example Encryption Scheme

In this section we recall Construction 5.3.9 from [Gol04] which achieves IND-CPA security starting from a pseudorandom function family.

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*

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

C Proof of Theorem 3.4

In this section we explain how the q-IND-CPA-2 indistinguishability notion for secret-key encryption of quantum messages introduced by Broadbent and Jeffery [BJ15, Appendix B] is equivalent to our gqIND-qCPA notion in the case that the encryption operation is a symmetric-key classical functionality operating in type-(2) mode. In [BJ15], the authors study the definition of quantum indistinguishability relative to the case of *quantum fully homomorphic encryption*. The general definition of *quantum symmetric-key encryption scheme* has been formalized in [ABF+16] in the following way.

Definition C.1 (Quantum Symmetric-Key Encryption Scheme). *A* quantum symmetric-key encryption scheme (or qSKE) *is a triple of quantum circuit families of polynomial depth:*

1. *(key generation)* $Q.\mathsf{Gen} : 1^n \mapsto k \in \mathcal{K}$
2. *(encryption)* $Q.\mathsf{Enc} : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$
3. *(decryption)* $Q.\mathsf{Dec} : \mathcal{K} \times \mathcal{Y} \to \mathcal{X}$

such that $\|Q.\mathsf{Dec} \circ Q.\mathsf{Enc} - \mathbb{I}_{\mathcal{X}}\|_{\diamond} \leq negl(n)$ *for all* $k \in Supp\,(Q.\mathsf{Gen}(1^n))$*, where* $\mathcal{K}$ *is the* key space*,* $\mathcal{X}$ *is the* plainstate space*,* $\mathcal{Y}$ *is the* cipherstate space*,* $\mathbb{I}$ *is the identity operator, and* $Q.\mathsf{Dec}, Q.\mathsf{Enc}$ *must be intended acting with the same (classical) key* k*.*

Then the authors of [ABF+16] define a notion of *quantum indistinguishability for quantum symmetric-key encryption schemes* (which they call IND, but which we relabel here as q-IND-qse for ease of reading) as follows.

Definition C.2 (q-IND-qse). *A qSKE* $(Q.\mathsf{Gen}, Q.\mathsf{Enc}, Q.\mathsf{Dec})$ *has* indistinguishable encryptions *(or is* q-IND-qse secure*) if for every QPT adversary* $\mathcal{A} = (\mathcal{M}, \mathcal{D})$ *we have:*

$$|\Pr\,[\,\mathcal{D}_{Q.\mathsf{Enc}}(\rho_{ME}) = 1\,] - \Pr\,[\,\mathcal{D}_{Q.\mathsf{Enc}}(|0\rangle\langle 0|_M \otimes \rho_E) = 1\,]| \leq negl(n)$$

where $\rho_{ME} \leftarrow \mathcal{M}$*,* $\rho_E = \mathrm{tr}_M(\rho_{ME})$*,* $\mathcal{D}_{Q.\mathsf{Enc}} = \mathcal{D} \circ (\mathsf{Enc}_k \otimes \mathbb{I}_E)$ *and the probabilities are taken over* $k \leftarrow Q.\mathsf{Gen}(1^n)$ *and the internal randomness of* Enc*,* $\mathcal{M}$*, and* $\mathcal{D}$*.*

Basically, the above definition states that for any QPT adversary $\mathcal{A}$, it must be hard to distinguish an encryption of any state ρ_M from an encryption of $|0\rangle\langle 0|_M$ (where ρ_E is auxiliary information carried between the two parts $\mathcal{M}$ and $\mathcal{D}$ of $\mathcal{A}$). Once we add a quantum CPA phase ($\mathcal{M}$ and $\mathcal{D}$ are given oracle access to Enc_k), Definition C.2 translates to the notion of q-IND-CPA from [BJ15]. And, also in [BJ15, Theorem B.2], this notion q-IND-CPA has been shown to be equivalent to another notion, q-IND-CPA-2, which considers the case where in the above game there are two messages chosen by the adversary, ρ^0 and ρ^1, instead of a single state ρ and the fixed $|0\rangle\langle 0|$ state. In other words, the q-IND-CPA-2 game can then be summarized as follows.

Definition C.3 (q-IND-CPA-2). *A qSKE* $(Q.\mathsf{Gen}, Q.\mathsf{Enc}, Q.\mathsf{Dec})$ *is* q-IND-CPA-2 secure*) if any QPT adversary* $\mathcal{A}$ *having oracle access to* $Q.\mathsf{Enc}_k$ *has probability at most negligibly better than guessing of winning the following game:*

1. $\mathcal{A}$ *generates two plaintext state messages* $\rho^0, \rho^1 \in \mathcal{X}$ *and sends them to the challenger* $\mathcal{C}$*;*
2. $\mathcal{C}$ *flips a random bit* $b \xleftarrow{\$} \{0,1\}$*;*
3. $\mathcal{C}$ *traces out (discards)* ρ^{1-b}*;*
4. $\mathcal{C}$ *encrypts* ρ^b *to* $\varphi \leftarrow Q.\mathsf{Enc}_k(\rho^b)$*;*
5. $\mathcal{A}$ *receives back* φ *from* $\mathcal{C}$*;*
6. $\mathcal{A}$ *outputs a bit* b'*, and wins the game iff* $b = b'$*.*

Finally, notice that Definition C.3 is equivalent to Definition 3.2 when the encryption algorithm $Q.\mathsf{Enc}$ is actually a type-(2) unitary operator U_{Enc} of a classical simmetric-key encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$. This concludes the proof of Theorem 3.3. □

D Semantic Security with Quantum Advice States

In Section 4.1 we left open the question of what happens if the messages (and the function to be computed about the message) are still classical, but the auxiliary advice can be a quantum state. Here we discuss this scenario.

A possible first approach 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 over m-bit plaintexts x, 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} on $|x\rangle$. $\mathcal{A}$'s goal is to output $f_m(x)$. Again, $\mathcal{S}$ plays in the reduced game and learns only $|\xi_m\rangle$.

Definition D.1 (qaSEM-qCPA). *A secret-key encryption scheme is said to be qaSEM-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 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 problem, we have to allow more general quantum circuits U'_{ξ_m} that can somehow provide 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 over 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 D.2 (iqSEM-qCPA). *A secret-key encryption scheme is said to be iqSEM-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 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. 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*) is not meaningful as long as the messages are still classical. Consequently, we proceed in Section 4.2 with our search for a notion of quantum semantic security considering quantum message states.