

On Quantum Obfuscation

Gorjan Alagic¹ Bill Fefferman²

¹ Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark
galagic@gmail.com

² Joint Center for Quantum Information and Computer Science (QuICS), University of Maryland, College Park, MD, USA
wjf@umd.edu

Abstract. Encryption of data is fundamental to secure communication in the modern world. Beyond encryption of data lies *obfuscation*, i.e., encryption of functionality. It is well-known that the most powerful means of obfuscating classical programs, so-called “black-box obfuscation,” is provably impossible [BGI⁺01]. For years since, obfuscation was believed to always be either impossible or useless, depending on the particulars of its formal definition. However, several recent results have yielded candidate schemes that satisfy a definition weaker than black-box, and yet still have numerous applications.

In this work, we initiate the rigorous study of obfuscating programs *via quantum-mechanical means*. We define notions of quantum obfuscation which encompass several natural variants. The input to the obfuscator can describe classical or quantum functionality, and the output can be a circuit description or a quantum state. The obfuscator can also satisfy one of a number of obfuscation conditions: black-box, indistinguishability, and best-possible; the last two conditions come in three variants: perfect, statistical, and computational. We discuss a number of applications, including quantum witness encryption and public-key quantum money. We then prove several impossibility results, extending a number of foundational papers on classical obfuscation to the quantum setting. We prove that quantum black-box obfuscation is impossible in a setting where adversaries can possess more than one output of the obfuscator (possibly even on the same input.) In particular, generic transformation of quantum circuits into black-box-obfuscated quantum circuits is impossible. We also show that statistical indistinguishability obfuscation is impossible, up to an unlikely complexity-theoretic collapse. Our proofs involve a new tool: chosen-ciphertext-secure encryption of quantum data, which was recently shown to be possible provided that quantum-secure one-way functions exist [ABF⁺16].

We emphasize that our results leave open one intriguing possibility: black-box obfuscation of classical or quantum circuits into a single, uncloneable quantum state. This indicates that, in spite of our results, quantum obfuscation may be significantly more powerful than its classical counterpart.

Keywords: Obfuscation, Encryption, Quantum

1 Introduction

The ability to encrypt data is central to modern communications. Basic methods for performing this task with privately-exchanged encryption keys have been known for hundreds of years. More advanced, public-key encryption methods were developed much more recently, beginning with the work of Merkle [Mer78] and Diffie and Hellman [DH76] in the 1970s. These public-key methods have found widespread practical application in virtually all Internet communications. More advanced theoretical methods for encrypting data, such as fully-homomorphic encryption, have only been discovered more recently [Gen09], but show great promise for practical applications.

Arguably the most powerful encryption ability is *obfuscation*; this is the ability to *encrypt functionality*. Obfuscation implies (with some caveats) the ability to perform almost any cryptographic task imaginable, including public-key and fully-homomorphic encryption. Unlike in the case of data encryption, our theoretical understanding of obfuscation is still fairly limited.

To understand obfuscation, it is useful to think about an obvious application: protecting intellectual property in software. In this setting, a software developer wishes to distribute their software to end users. However, the code contains a number of trade secrets which the developer does not want to become public. In order to maintain these secrets, the publisher passes the software through an *obfuscation algorithm* (or obfuscator) prior to publishing. In this application, the obfuscator must be an efficient algorithm that satisfies three core properties:

1. *functional equivalence*: the input/output functionality of the input program does not change;
2. *polynomial slowdown*: if the input program is efficient, then the output program is efficient;
3. *obfuscation*: the code of the output program is “hard to understand.”

The last condition can be formulated rigorously in a number of ways. One possibility is the so-called “virtual black-box” condition, which says that the obfuscated program is no more useful than an impenetrable box which simply accepts inputs and produces outputs. While this condition appears to be too strong in the classical world, there are other formulations (with varying levels of strength and usefulness) which may be achievable. Arguably the most likely of these is computational-indistinguishability obfuscation; promising candidates and numerous applications have been proposed for this case (see, e.g., [GGH⁺13, BGK⁺14, BCC⁺14, BZ14, BR14, GGHW14, HSW14].)

The study of encrypting classical data and classical programs is significantly complicated by the advent of *quantum computation*. One well-known consequence of the presence of quantum computers is that certain data encryption schemes, such as those based on the hardness of factoring or discrete logarithm, are no longer secure. It is conceivable that certain classical obfuscation schemes are also not secure against quantum adversaries. On the other hand, quantum mechanics also appears to enable certain cryptographic tasks (such as information-theoretically secure key exchange) that are impossible classically. It is thus natural to ask what quantum computation means for obfuscation of programs. In particular, we would like to answer the following questions:

- what are the “right” formulations of quantum-mechanical program obfuscation?
- is it possible to quantumly obfuscate classical and/or quantum programs?
- which of the classical results about obfuscation carry over to the quantum setting?
- how do the applications of quantum obfuscation differ from the classical case?

We remark that, in order to address the above questions, we must also properly address the question of encrypting quantum data—a strictly simpler task than encrypting functionality. While information-theoretic encryption of quantum data has been considered before, in this setting we are interested in encryption of quantum data *with computational assumptions*³. Thankfully, this subject has been explored recently, and the basic results we will need are now available [BJ15, ABF⁺16].

Before continuing, we draw attention to the distinction between obfuscating *programs* (e.g., Turing Machines) and obfuscating *circuits*. While these two forms of obfuscation are closely related, there are some important technical differences. In this work, as in most theoretical works on obfuscation, we will focus on obfuscation of circuits. We view the circuit model as more convenient; it is also more standard and convenient in quantum computation.

1.1 Background

We now briefly review the current state of affairs in research on obfuscation and quantum encryption with computational security assumptions. The classical case has been studied significantly. On the other hand, quantum obfuscation has received little to no attention, outside the proposal in [AJJ14] and some related work in copy-protection (e.g., [Aar09].)

³ Note that information-theoretic obfuscation is impossible if the adversary can execute the obfuscated program on all possible inputs; indeed, a computationally unbounded adversary can use this ability to learn everything there is to know about the program.

1.1.1 Classical obfuscation Ad-hoc obfuscation of software has been a fairly common practice for some time. In fact, simply compiling a program can be viewed as a form of obfuscation. The earliest mention of obfuscation in the modern study of theoretical cryptography appears to be in the famous paper of Diffie and Hellman [DH76]. There, it was suggested that public-key cryptosystems might be constructible via obfuscation of private-key schemes; this was viewed as a reasonable possibility because writing code in an obfuscated manner seems relatively easy in practice.

The first major result in classical obfuscation was the 2001 proof by Barak et al. that virtual black-box obfuscation is impossible [BGI⁺01, BGI⁺12]. Their definition is based on the *simulation paradigm*. More precisely, the obfuscation condition (i.e., the third condition in the [previous section](#)) states that any efficient algorithm with access to an obfuscated circuit should be simulable by another efficient algorithm with only oracle (i.e., black-box) access to the original functionality. This definition is very natural in the setting of the aforementioned “software intellectual property protection” application: the end user can only learn that which is learnable by simply running the program. Barak et al. proved that there exist circuit families which are unobfuscatable under this definition. They also showed that some of the most sought-after applications of black-box obfuscation are impossible. For instance, they showed that private-key encryption schemes cannot be transformed to public-key ones by obfuscating the encryption circuits in a generic manner.

The years following the Barak et al. result saw some limited progress in theoretical obfuscation. It was proved possible for some limited forms of functionality [CD08, Wee05], and some additional limits were placed, e.g., on black-box obfuscation with auxiliary input [GK05]. An important step in formulating feasible notions of obfuscation was taken by Goldwasser and Rothblum; they defined *indistinguishability obfuscation* and *best-possible obfuscation* [GR07]. Both of these definitions alter the obfuscation condition, while leaving the functional-equivalence and polynomial-slowdown conditions unchanged. Under indistinguishability, it is required that the obfuscator maps functionally-equivalent circuits to indistinguishable distributions. Under best-possible, the obfuscator maps any circuit to a circuit from which the end user can “learn the least.” Both definitions have a perfect, statistical, and computational variant. Goldwasser and Rothblum proved that the two definitions are equivalent, and that the perfect and statistical versions are impossible (unless the PH collapses) [GR07]. This left one possibility: computational indistinguishability obfuscation. It was widely believed that computational indistinguishability was too weak of a condition to provide any interesting applications.

In 2013, in a breakthrough result, Garg et al. proposed a convincing candidate for computational indistinguishability obfuscation [GGH⁺13]. They proposed an obfuscation scheme for NC1 circuits, based on the presumed hardness of a problem in multilinear maps; they also showed how to use fully-homomorphic encryption (with NC1 decryption circuits) to “bootstrap” their NC1 scheme to obfuscation for all circuits. Around the same time, another breakthrough by Sahai and Waters showed how to use a computational indistinguishability obfuscator to achieve a wide-range of applications, via a new “punctured programs” technique [SW14]. These applications include chosen-ciphertext-secure public-key encryption, injective trapdoor functions, and oblivious transfer. Sahai and Waters suggested that the applications were so wide-ranging that indistinguishability obfuscation might become a “‘central hub’ for cryptography” [SW14]. These two breakthroughs were followed by a flurry of new activity in the area, including several new proposals and applications [BGK⁺14, BCC⁺14, BZ14, BR14, GGHW14, HSW14].

1.1.2 Quantum obfuscation Quantum obfuscation is essentially an unexplored topic, and the present work appears to be the first rigorous treatment of the foundational questions. The question of whether quantum obfuscation is possible was posed as one of Scott Aaronson’s “semi-grand challenges” for quantum computation [Aar05]. Since so little work on quantum obfuscation has appeared, our brief discussion will also mention some results that we believe are related.

In [Aar09], Aaronson proposed two relevant results. The first was a *complexity-theoretic no-cloning theorem*, stating that cloning an unknown, random state by means of a black-box “reflection oracle” requires exponentially many queries. The second theorem stated that an oracle exists relative to which “software copy-protection” is possible. Unfortunately, a full version of [Aar09] with proofs never appeared, although the complexity-theoretic no-cloning theorem was eventually proved in a paper on quantum money [AC12]. In related work, Mosca and Stebila proposed a black-box quantum money scheme, and suggested the possibility of using a quantum circuit obfuscator in place of the black box [MS10].

More recently, Alagic, Jeffery and Jordan proposed obfuscators for both classical (reversible) circuits and quantum circuits, based on ideas from topological quantum computation [AJJ14]. The proposed obfuscator compiles circuits into braids using certain high-dimensional representations of the braid group, and then applies an algorithm for putting braids into normal form. Although it is efficient, this algorithm does not satisfy any of the aforementioned obfuscation definitions; instead, it satisfies perfect indistinguishability for a restricted set of circuit equivalences. The usefulness of such an obfuscator is unclear at this time.

1.1.3 Quantum encryption In order to discuss quantum obfuscation of functionality, we require a more basic primitive: quantum encryption of data. As we will see later, information-theoretic encryption is insufficient for our purposes, and we must thus rely on computational assumptions. In particular, we require (i.) reusability of the key, and (ii.) non-adaptive chosen-ciphertext security. Recent results have shown that this is possible to achieve under the assumption that quantum-resistant one-way functions exist [ABF⁺16]. For completeness, we briefly summarize the relevant results from [ABF⁺16] below, and in more detail in Section 2.

1. **Quantum encryption schemes.** A symmetric-key encryption scheme (for quantum states, with classical keys) consists of three quantum algorithms (key generation, encryption, and decryption) which satisfy correctness: under a fixed key, encryption followed by decryption is equivalent to the identity. Such schemes first appeared in [BJ15].
2. **Chosen-ciphertext security for quantum encryption.** One may also define IND-CCA1 security (or *indistinguishability of ciphertexts under non-adaptive chosen ciphertext attacks*) for these schemes; this formalizes the idea of a “lunchtime attack,” where an adversary has complete access to all aspects of the encryption scheme except the key itself, and is tasked with decrypting a challenge ciphertext later.
3. **An IND-CCA1-secure construction.** If quantum-secure one-way functions (qOWF) exist, then so do IND-CCA1-secure symmetric-key encryption schemes for quantum states. These qOWFs are deterministic classical functions which are easy to compute, but hard to invert for quantum adversaries. The quantum IND-CCA1 scheme is a simple analogue of the standard classical construction using PRFs.

1.2 Summary of results

In this section, we summarize our results and discussions. These are divided by subject, with quantum encryption covered in Section 2, quantum black-box obfuscation in Section 3, and quantum indistinguishability obfuscation in Section 4.

1.2.1 Quantum black-box obfuscation

Definitions. Our main results concern definitions, applications, and (im)possibility of quantum obfuscation in the virtual black-box setting. We will begin by defining the following.

1. **Quantum black-box obfuscator.** This is a polynomial-time quantum algorithm \mathcal{O} which accepts quantum circuits C as input, and produces quantum states $\mathcal{O}(C)$ as output. It preserves

functionality, in the sense that there is a publicly known way to use $\mathcal{O}(C)$ and any input state $|\psi\rangle$ to produce the state $C|\psi\rangle$. It satisfies a black-box condition: for polynomial-time quantum algorithms, possession of $\mathcal{O}(C)$ can be simulated by black-box access to C . This definition is a natural analogue of the classical black-box definition given in [BGI⁺12].

2. **Quantum “two-circuit” black-box obfuscator.** This obfuscator is precisely as above, except the obfuscation condition is strengthened to hold over arbitrary *pairs of circuits* (C_1, C_2) . For us, this definition will be primarily useful because of its role in establishing certain impossibility results.
3. **Information-theoretic quantum black-box obfuscator.** This is a modification of the above definition, in which we posit that *any* adversary with access to $\mathcal{O}(C)$ can be simulated by a *polynomial-time* quantum simulator with black-box access to C . This definition is impossible classically for obvious reasons. It is not a priori clear if the quantum case is impossible: if $\mathcal{O}(C)$ is a single quantum state, even unbounded adversaries are restricted by no-cloning and the (in general irreversible) properties of measurement.

Impossibility. We prove three impossibility results, which place several important restrictions on quantum obfuscation. Our impossibility proofs are based on the ideas of Barak et al. [BGI⁺12], with several important quantum adaptations, and a new quantum ingredient: the aforementioned IND-CCA1 quantum encryption scheme.

1. **Two-state black-box obfuscation is impossible.** We prove that there exist families of circuit pairs which can reveal a secret if one is in possession of a circuit description for both of them, but not if one only has black-box access. This impossibility persists even if the obfuscation output is a quantum state, as opposed to a circuit description. Unlike the other results, it is also true even if the obfuscated states are *not reusable*.
2. **If qOWFs exist, then obfuscation with more than one output is impossible.** For this proof, we combine the pairs from the circuit families in the two-circuit impossibility proof in order to build a single unobfuscatable family. The ability to execute obfuscated states from this family *on themselves* is crucial here, and has two requirements: (i.) access to more than one obfuscation, even if the obfuscations are quantum states, and (ii.) secure encryption, which in turn requires the existence of qOWFs. This result applies both to quantum black-box obfuscators (as in the first definition above) and the information-theoretic variant (as in the third definition above.)
3. **Classical algorithms for quantum obfuscation are impossible, unconditionally.** This result follows directly from the previous result and Application 1 below. It can be viewed as an extension of the original Barak et al. impossibility result to the case of quantum functionality and quantum adversaries.

We emphasize that, while our techniques are very similar to Barak et al., our results are not a simple consequence of the fact that classical functions (and in particular, the Barak et al. unobfuscatable functions) are special cases of quantum functions. This “special case” argument fails for obvious reasons when the output of the obfuscator can be a quantum state. As it turns out, it also fails when the output is a quantum circuit; briefly, the reasons are (i.) it is now permitted to specify (even approximate) quantum circuits for classical functions, (ii.) the black-box simulator now has quantum access, which is significantly more powerful than classical (see, e.g., [HR14]) and (iii.) the Barak et al. adversary is insufficient since it can only perform classical gates homomorphically. The last point is where the aforementioned quantum encryption tools become necessary.

Applications. We then move on to discuss potential applications of quantum black-box obfuscators. Some of these applications may still be possible, in some form, in spite of our impossibility results. We view this as a strong indication that quantum obfuscation should be studied further. While some of the applications are analogues of known classical applications (as outlined in

[BGI⁺12],) the last is special to the quantum setting. It is likely that there are other applications which are unique to the quantum setting. We remark that, since black-box obfuscation implies computationally best-possible obfuscation (which, in turn, implies computational indistinguishability obfuscation (iO),) the numerous recently discovered applications of iO may also apply here.

1. **Quantum-secure one-way functions.** We show that, if there exists a classical probabilistic algorithm for quantum obfuscation, then quantum-secure one-way functions exist. The above impossibility result rules this out, but the implication is nonetheless interesting; for one, it enables the proof of the impossibility result itself. We are unable to extend this application to the setting of efficient quantum algorithms for obfuscation. Instead, we give the next application.
2. **Private-key quantum encryption for long messages.** This application uses only the obfuscator, and no other assumptions. We remark that such schemes require the qOWF assumption otherwise.
3. **Public-key quantum encryption from IND-CPA private-key quantum encryption.** In this case, we show (via quantum semantic security [ABF⁺16]) that providing the obfuscated encryption circuit as the public key transforms private-key encryption into public-key encryption. In this case, we require a stronger notion of obfuscation, which is guaranteed *in the presence of auxiliary information*.
4. **Public-key quantum money.** Using circuit obfuscation to produce public-key quantum money was first proposed by Mosca and Stebila [MS10], using a complexity-theoretic no-cloning theorem proposed by Aaronson [Aar09] and proved by Aaronson and Christiano [AC12]. We outline the idea, and adapt it based on the new limitations placed by our results.

We emphasize that all the above applications except quantum money also work for achieving *classical functionality* from a quantum obfuscator; however, depending on the details of the obfuscator and the application, this may require quantum algorithms for encryption and decryption, or even quantum ciphertexts.

1.2.2 Quantum indistinguishability obfuscation We also consider an alternative formulation of obfuscation, motivated by the classical definitions of indistinguishability obfuscation and best-possible obfuscation, as set down by Barak et al. [BGI⁺12] and Goldwasser and Rothblum [GR07]. We establish quantum analogues of the central results in those classical papers. In this setting, rather than comparing the obfuscation of the circuit to that of a black-box, we compare it to obfuscations of other, functionally-equivalent circuits. Starting with the new definitions, our results are as follows.

1. **Quantum indistinguishability obfuscator.** Just as in the black-box definition, this is a polynomial-time quantum algorithm \mathcal{O} which accepts quantum circuits C as input, and produces “functionally-equivalent” quantum states $\mathcal{O}(C)$ as output. The obfuscation condition now states that functionally equivalent circuits are mapped to *indistinguishable* states. Based on the kind of indistinguishability enjoyed by these states, there are three variants: perfect, statistical, and computational.
2. **Quantum best-possible obfuscator.** This is an algorithm precisely as above, except for the obfuscation condition: it now states that $\mathcal{O}(C)$ is the state that “leaks least,” among all states which are “functionally-equivalent” to C . There are again three variants: perfect, statistical, and computational.
3. **Equivalence of definitions.** We prove that each of the three variants of quantum indistinguishability obfuscation is equivalent to the analogous variant of quantum best-possible obfuscation, so long as the obfuscator is efficient.

4. **Impossibility of perfect and statistical indistinguishability obfuscation.** We give a quantum version of the main result of [GR07]: perfect and statistical quantum indistinguishability obfuscation is impossible, unless coQMA is contained in QSZK. An analogous containment in the classical setting (i.e., coMA \subseteq SZK) would imply a collapse of the polynomial-time hierarchy to the second level. Moreover, for the case of obfuscating arbitrary quantum computations (i.e., including measurements and adding/discarding ancillas), we obtain that statistical quantum indistinguishability obfuscation would imply PSPACE \subseteq QSZK.
5. **Application: witness encryption for QMA.** We show that a quantum indistinguishability obfuscator enables witness encryption for QMA. A witness encryption scheme for a language L in QMA encrypts plaintexts x using a particular instance l . The security condition states that, if $l \in L$, then a valid witness w for $l \in L$ allows decryption; on the other hand, if $l \notin L$, then ciphertexts are indistinguishable. While witness encryption has several applications classically [GGSW13], the quantum analogue has not been considered previously.

We remark that, in the classical setting, indistinguishability obfuscation also implies functional encryption [GGH⁺13] and many more applications through the “punctured programs” technique developed by Sahai and Waters [SW14]. We suspect that these results can also be adapted to the quantum setting, but leave them open for now.

1.3 Notation and terminology

In this section, we set down some notation and basic terminology which we will use throughout the rest of the paper.

1.3.1 Classical We will assume that the state space of a classical device can be identified with sets of bitstrings, i.e., $\{0, 1\}^n$ for some positive integer n . The notation $x \in_R \{0, 1\}^n$ will mean that x is an n -bit string selected uniformly at random. The set of all bitstrings (of arbitrary length) will be denoted by $\{0, 1\}^*$. Classical functions will then be maps $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$ from one set of bitstrings to another. We will also sometimes consider function families, written $f : \{0, 1\}^* \rightarrow \{0, 1\}^*$; these can be thought of as a function family $\{f_n\}_{n>0}$ indexed by the input size n .

A classical circuit C is a sequence of local boolean gates which, when composed together, implement some (in general irreversible) function $f_C : \{0, 1\}^n \rightarrow \{0, 1\}^m$. The input size of C is n , the output size is m , and the number of gates is denoted by $|C|$. A probabilistic circuit is also a circuit, but with the input bits divided into two registers: the input register, and the “coin” register. A normal execution of a probabilistic circuit involves initializing the coin register with completely random bits, and inserting the input into the input register. We will frequently discuss **circuit ensembles**; these are infinite families $\{C_n\}_{n>0}$ of circuits, one for each possible input size, so that the size of circuit C_n is bounded by some fixed polynomial function of n . We say that a circuit ensemble is **uniform** if there is a deterministic polynomial-time Turing Machine which, on input 1^n , outputs a classical description of C_n . We will also sometimes make use of **distributions of circuit ensembles**; these are infinite families $\mathcal{C} = \{\mathcal{C}_n\}_{n>0}$ where each \mathcal{C}_n is a finite family of circuits of input size n , along with a probability distribution $P_{\mathcal{C},n}$. For a bitstring x , the notation $\mathcal{C}(x)$ will then denote the probability distribution (on bitstrings) resulting from running a random circuit from the family $\mathcal{C}_{|x|}$, selected according to the distribution $P_{\mathcal{C},|x|}$.

A **deterministic classical algorithm** \mathcal{A} is simply a circuit ensemble. Running \mathcal{A} on an input bitstring x involves selecting the circuit with the appropriate input size, and executing it with input x . If the circuit ensemble is uniform, we will say that \mathcal{A} is efficient; more precisely, it is then a classical deterministic polynomial-time algorithm (**PT** for short.) A **probabilistic algorithm** \mathcal{A}' is an algorithm whose circuits are probabilistic. Running \mathcal{A}' on an input bitstring x involves selecting the circuit with the appropriate input size, initializing its coin register with uniformly random bits, and then executing it with input x . If the circuits of \mathcal{A}' are polynomial-time uniform,

we say that \mathcal{A}' is an efficient, or classical probabilistic polynomial-time algorithm (**PPT** for short.) We will frequently use PPTs to model the most general efficient classical algorithms.

1.3.2 Quantum For our purposes, the space of pure states of a quantum device will be identified with a Hilbert space $\mathcal{H}_n \cong (\mathbb{C}_2)^{\otimes n}$ of a finite number n of qubits. We will identify some fixed orthonormal basis (called the *computational basis*) of \mathcal{H}_n with the corresponding space $\{0, 1\}^n$ of classical states, so that, e.g., $|x\rangle$ for $x \in \{0, 1\}^n$ denotes a basis element of \mathcal{H}_n . The space of density operators (i.e., general quantum states) of n qubits will be denoted $\mathfrak{D}(\mathcal{H}_n)$; this is the set of positive semidefinite, Hermitian trace-one operators in $\text{End}(\mathcal{H}_n)$. A state in $\mathfrak{D}(\mathcal{H}_n)$ can be interpreted as a probabilistic mixture $\sum_j p_j |\varphi_j\rangle\langle\varphi_j|$ of pure states, albeit not in a unique way. We will discuss valid quantum transformations of three types. The first are measurements, which act on a state $|\psi\rangle \in \mathcal{H}_n$ by projecting some or all of the qubits into the computational basis states $\{|0\rangle, |1\rangle\}$. The second are unitary maps, i.e., linear operators $U : \mathcal{H}_n \rightarrow \mathcal{H}_n$ satisfying $U^\dagger U = \mathbb{1}_n$, where $\mathbb{1}_n$ denotes the n -qubit identity operator. The third are CPTP maps, i.e., completely positive trace-preserving maps $\Phi : \mathfrak{D}(\mathcal{H}_n) \rightarrow \mathfrak{D}(\mathcal{H}_m)$. CPTP maps are the most general type of evolution, encompassing unitary maps, measurement, and adding or discarding (or tracing out) qubits. For example, a unitary operator $U \in U(\mathcal{H}_n)$ can be expressed as a CPTP map by writing $\rho \mapsto U\rho U^\dagger$, where $\rho \in \mathfrak{D}(\mathcal{H}_n)$.

A **quantum circuit** C is a sequence of local unitary gates on a fixed number (say n) of qubits; these gates, when composed together, implement some unitary operator $U_C \in U(2^n)$. Definitions of circuit ensembles and distributions over circuit ensembles are defined precisely as in the classical case. A **quantum polynomial time algorithm** (or **QPT** for short) \mathcal{A} is a uniform ensemble of quantum circuits; algorithms can also include measurements and discarding (or tracing-out) of subsystems, so long as these also admit efficient classical descriptions. The input and output size of a quantum algorithm can vary, and will have to be deduced from context. For example, given a QPT \mathcal{A} , the expression $\Pr[\mathcal{A}(|0^n\rangle) = 1]$ will take the value zero unless \mathcal{A} has a specific, labeled output qubit which is measured at the end of the computation.

2 Quantum encryption

In this section, we recall a recently developed encryption scheme for quantum states, with computational assumptions [ABF⁺16]. The scheme satisfies a notion of non-adaptive chosen-ciphertext security, under the assumption that quantum-secure one-way functions exist. We begin with the following definitions.

Definition 1. A *PT-computable function* $f : \{0, 1\}^* \rightarrow \{0, 1\}^*$ is a *quantum-secure one-way function* (qOWF) if for every QPT \mathcal{A} ,

$$\Pr_{x \in_R \{0, 1\}^n} [\mathcal{A}(f(x), 1^n) \in f^{-1}(f(x))] \leq \text{negl}(n),$$

where the probability is taken over $x \in_R \{0, 1\}^n$ as well as the measurements of \mathcal{A} .

Definition 2. A *PT-computable function family* $f_k : \{0, 1\}^n \rightarrow \{0, 1\}^m$ is a *quantum-secure pseudorandom function* (qPRF) if for every QPT \mathcal{A} ,

$$|\Pr_{k \in_R \{0, 1\}^n} [\mathcal{A}^{f_k}(1^n) = 1] - \Pr_{g \in_R \mathcal{F}_{n,m}} [\mathcal{A}^g(1^n) = 1]| \leq \text{negl}(n),$$

where $\mathcal{F}_{n,m}$ denotes the space of all functions from $\{0, 1\}^n$ to $\{0, 1\}^m$.

Classically, one-way functions are the fundamental primitive underpinning encryption. A series of basic results shows that one-way functions can be turned into pseudorandom functions, which can then be used for defining probabilistic encryption schemes. This series of results carries over to the quantum-secure case without much fuss (although some proofs are more involved.) For example, it is known how to construct qPRFs from qOWFs

Theorem 3. [HILL99, Zha12] *If quantum-secure one-way functions exist, then so do quantum-secure pseudorandom functions.*

It is well-known how to encrypt quantum states with information-theoretic security, via the so-called quantum one-time pad. To encrypt a single-qubit state ρ , we choose two classical bits at random, use them to select a random Pauli matrix $P \in \{\mathbb{1}, X, Y, Z\}$, and perform $\rho \mapsto P\rho P^\dagger$. To encrypt an n -qubit quantum state ρ , we select $r \in_R \{0, 1\}^{2n}$ and apply

$$\rho \mapsto P_r \rho P_r^\dagger, \quad (2.1)$$

where P_r denotes the element of the n -qubit Pauli group indexed by r .

A general encryption scheme for quantum states is defined as follows.

Definition 4. *A symmetric-key quantum encryption scheme is a triple of QPTs:*

- (key generation) $\text{KeyGen} : 1^n \mapsto k \in \{0, 1\}^n$;
- (encryption) $\text{Enc}_k : \mathcal{H}_m \longrightarrow \mathcal{D}(\mathcal{H}_c)$;
- (decryption) $\text{Dec}_k : \mathcal{D}(\mathcal{H}_c) \longrightarrow \mathcal{D}(\mathcal{H}_m)$;

where m and c are polynomial functions of n , and the QPTs satisfy $\|\text{Dec}_k \circ \text{Enc}_k - \mathbb{1}_m\|_\diamond \leq \text{negl}(n)$ for all $k \in \text{supp KeyGen}(1^n)$.

Public-key quantum encryption schemes (with classical keys) are defined in an analogous manner [BJ15]. The encryption schemes we will need must produce ciphertexts which are computationally indistinguishable. In some cases, the ciphertexts will need to remain indistinguishable even to adversaries which possess oracle access to the encryption algorithm (and sometimes also even the decryption algorithm.) This security notion is captured by the following definition.

Definition 5. *A symmetric-key quantum encryption scheme is IND-secure if for all QPTs $\mathcal{A}, \mathcal{A}'$,*

$$\left| \Pr[(\mathcal{A}' \circ \text{Enc}_k \otimes \mathbb{1}_s \circ \mathcal{A}) \cdot 1^n = 1] - \Pr[(\mathcal{A}' \circ \Xi_{\text{Enc}_k|0^m\rangle\langle 0^m|} \otimes \mathbb{1}_s \circ \mathcal{A}) \cdot 1^n = 1] \right| \leq \text{negl}(n),$$

where $\Xi_\sigma : \rho \mapsto \sigma$ is the “forgetful” map, and s is a polynomial function of n . If \mathcal{A} and \mathcal{A}' have oracle access to Enc_k , then we say that the scheme is IND-CPA secure. If in addition \mathcal{A}' has oracle access to Dec_k , then we say that the scheme is IND-CCA1 secure.

It turns out that qPRFs can be used to construct simple symmetric-key quantum encryption schemes that satisfy all of the above security conditions. Let $\{f_k : \{0, 1\}^n \rightarrow \{0, 1\}^{2n}\}$ be a qPRF, and consider the following scheme:

- $\text{KeyGen}(1^n)$: output $k \in_R \{0, 1\}^n$;
- $\text{Enc}_k(\rho)$: choose $r \in_R \{0, 1\}^n$; output $|r\rangle\langle r| \otimes P_{f_k(r)} \rho P_{f_k(r)}^\dagger$;
- $\text{Dec}_k(|r\rangle\langle r| \otimes \sigma)$: output $P_{f_k(r)}^\dagger \sigma P_{f_k(r)}$.

In the decryption algorithm, we may assume that the first register is always measured prior to decrypting. Correctness of the scheme is straightforward to check: decrypting with the same key and randomness simply undoes the Pauli operation. IND-CCA1 security was shown in [ABF⁺16]. Combining this with Theorem 3, we have

Theorem 6. *If quantum-secure one-way functions exist, then so do IND-CCA1-secure symmetric-key quantum encryption schemes.*

3 Quantum black-box obfuscation

In this section, we discuss the virtual black-box framework for obfuscating quantum computations. We begin in [Section 3.1](#) with a definition of black-box quantum obfuscator, motivated both by the classical analogue and an intuitive notion of what a “good obfuscator” should achieve. In [Section 3.2](#), we outline several interesting cryptographic consequences that would follow from the existence of such an obfuscator. Finally, in [Section 3.3](#), we prove a few impossibility results which restrict the range of possibilities for the existence of black-box quantum obfuscators. Interestingly, our results leave open some possibilities, which include (restricted versions) of the most interesting applications. Indeed, it is conceivable that quantum obfuscation could be significantly more powerful than its classical counterpart.

3.1 Definitions

Any reasonable notion of obfuscation involves giving the obfuscated circuit $\mathcal{O}(C)$ to an untrusted party. We accept as fundamental the idea that this obfuscated circuit should implement some particular, chosen functionality f_C , and that the object $\mathcal{O}(C)$ allows the untrusted party to execute that functionality. In the black-box formulation of obfuscation, we demand that this is effectively all that the untrusted party will ever be able to do. The rigorous formulation uses the simulation paradigm: anything which can be efficiently learned from the obfuscated circuit, should also be efficiently learnable simply by evaluating f_C some polynomial number of times. This “virtual black-box” notion was first formulated by Barak et al. [[BGI⁺12](#)], and proved impossible to satisfy generically in the classical case.

In the quantum case, there are several complications. First, we are considering the obfuscation of quantum functionalities. This implies that the end user (and hence also any adversary) should be in possession of a quantum computer, and likewise for the simulator. Second, it is conceivable that the obfuscation may not just be another quantum circuit, which is simply a classical state describing a quantum computation. The obfuscator might instead output a quantum state, which is then to be employed by the end user to execute the desired functionality in some well-specified manner. These considerations motivate the following definition.

Definition 7. *A black-box quantum obfuscator is a quantum algorithm \mathcal{O} and a QPT \mathcal{J} such that whenever C is an n -qubit quantum circuit, the output of \mathcal{O} is an m -qubit state $\mathcal{O}(C)$ satisfying*

1. (polynomial expansion) $m = \text{poly}(n)$;
2. (functional equivalence) $\|\mathcal{J}(\mathcal{O}(C) \otimes \rho) - U_C \rho U_C^\dagger\|_{\text{tr}} \leq \text{negl}(n)$ for all $\rho \in \mathfrak{D}(\mathcal{H}_n)$;
3. (virtual black-box) for every QPT \mathcal{A} there exists a QPT \mathcal{S}^{U_C} such that

$$\left| \Pr[\mathcal{A}(\mathcal{O}(C)) = 1] - \Pr[\mathcal{S}^{U_C}(|0^n\rangle) = 1] \right| \leq \text{negl}(n).$$

We emphasize that while the “interpreter” algorithm \mathcal{J} must be polynomial-time, the obfuscator itself need not be. In applications, it will be necessary to make the obfuscator polynomial-time; on the other hand, our impossibility results will hold even for inefficient obfuscators. One could consider variants of [Definition 7](#) where the interpreter algorithm is fixed once and for all, or where $\mathcal{O}(C)$ itself consists of both a quantum “advice state” and a circuit which the end user should execute on the advice state and the desired input. It is straightforward to show that all of these variants are equivalent, in the sense that a black-box quantum obfuscator of each variant exists if and only if the other variants exist. Since we are primarily concerned with possibility vs impossibility, we will stick with the formulation in [Definition 7](#). We also remark that the interpreter is a natural addition to the classical black-box definition when passing to the quantum case. In order for the definition to make sense, there should be *some efficient way* to use $\mathcal{O}(C)$ to implement U_C ; whatever that efficient procedure is, we have here called it an interpreter and denoted it by \mathcal{J} .

We also point out that the no-cloning theorem opens up the possibility of *computationally unbounded adversaries*. In the classical case, such an adversary could simply execute the circuit on every input, and thus learn far more than is possible for a polynomial-time black-box simulator. Quantumly, however, a computationally unbounded adversary is restricted both by the no-cloning theorem and the limitations of measurement. The adversary may not be able to acquire multiple copies of the obfuscated state, and the single state may be partially (or completely) destroyed when measured. It is thus not *a priori* clear that an unbounded adversary could always outmatch a polynomial-time black-box simulator. The appropriate definition is a straightforward modification of [Definition 7](#), where we replace the third condition with the following:

3. (*information-theoretic virtual black-box*) for every quantum adversary \mathcal{A} there exists a QPT \mathcal{S}^{U_C} such that

$$\left| \Pr[\mathcal{A}(\mathcal{O}(C)) = 1] - \Pr[\mathcal{S}^{U_C}(|0^n\rangle) = 1] \right| \leq \text{negl}(n).$$

3.2 Applications of efficient black-box obfuscators

In this section, we motivate the study of quantum black-box obfuscation by giving a few example applications. Unsurprisingly, these applications require that the obfuscation algorithm is itself quantum polynomial-time; strictly speaking, this is not required of [Definition 7](#). Many of these applications are motivated by known classical applications of classical black-box obfuscators. Although our impossibility results will put some restrictions on these applications, they remain interesting. In fact, some of the applications (such as quantum-secure one-way functions) will be used in the impossibility proofs themselves. We point out that, while most of the applications below are written in terms of quantum functionality (e.g., encryption of quantum states), one can just as well consider the weaker case of classical functionality, in this case achieved via quantum means (e.g., via a quantum algorithm for obfuscation.)

3.2.1 Quantum-secure one-way functions. The first application shows that, if there exists a classical algorithm for obfuscating quantum computations, then quantum-secure one-way functions exist. In this section alone, we will assume that functional equivalence property of the black-box obfuscator in [Definition 7](#) is strengthened to be exact (that is, the trace distance is 0). By the results discussed in [Section 2](#), this also implies the existence of quantum-secure pseudorandom generators, quantum-secure pseudorandom functions, and IND-CCA1-secure symmetric-key quantum encryption schemes. We say that a quantum black-box obfuscator is *perfect* if it preserves functionality perfectly.

Proposition 1 *If there exists a classical probabilistic algorithm which is a (perfect) quantum black-box obfuscator, then quantum-secure one-way functions exist.*

Proof. The proof is essentially the same as that of Lemma 3.8 in [\[BGI⁺12\]](#). For all $a \in \{0, 1\}^n$ and $b \in \{0, 1\}$, we define

$$U_{a,b} : |x, y\rangle \mapsto \begin{cases} |a, y \oplus b\rangle & \text{if } x = a; \\ |x, y\rangle & \text{otherwise.} \end{cases}$$

Define a function $f : \{0, 1\}^* \rightarrow \{0, 1\}^*$ by $f(a, b, r) = \mathcal{O}_r(U_{a,b})$ where \mathcal{O} is the obfuscator⁴ as in the hypothesis, and \mathcal{O}_r denotes the same algorithm, but with randomness coins initialized to r . Clearly, inverting f requires computing b from $\mathcal{O}_r(U_{a,b})$. Moreover, with only black-box access

⁴ For simplicity of notation, we omit \mathcal{J} and assume that $f(a, b, r) = \mathcal{O}_r(U_{a,b})$ is in fact a classical circuit for $U_{a,b}$.

to $U_{a,b}$ (for uniformly random a, b) the probability of correctly outputting b in polynomial time is at most $1/2 + \text{negl}(n)$. By the black-box property of \mathcal{O} , we then have

$$\begin{aligned} \Pr_{a,b}[A(f(a, b, r)) = b] &= \Pr_{a,b}[A(\mathcal{O}_r(a, b)) = b] \\ &\leq \Pr_{a,b}[\mathcal{S}^{U_{a,b}}(1^n) = b] + \text{negl}(n) \\ &\leq \frac{1}{2} + \text{negl}(n), \end{aligned}$$

which completes the proof.

We remark that the above proof fails if the obfuscator is a quantum algorithm—even if its output is itself classical. The issue is that one-way functions must be deterministic; while one can turn a classical probabilistic algorithm into a deterministic one by making the coins part of the input, this is not possible quantumly. We leave the problem of constructing cryptographically useful primitives from a fully quantum obfuscator (or even just from a quantum encryption scheme) as an interesting open question.

3.2.2 Private-key encryption for long messages. Let $(\mathcal{O}, \mathcal{J})$ be a quantum black-box obfuscator. We consider an adaptation of the unitary operator $U_{a,b}$ defined above, but now with Pauli group action instead of XOR, and with two registers of, in general, different sizes:

$$V_{r,k} : |x, y\rangle \mapsto \begin{cases} |x, P_r^\dagger y\rangle & \text{if } x = k; \\ |x, y\rangle & \text{otherwise,} \end{cases}$$

Now consider the following scheme for encrypting $l(n)$ -qubit quantum states.

- $\text{KeyGen}(1^n)$: output $k \in_R \{0, 1\}^n$;
- $\text{Enc}_k(\rho)$: choose $r \in_R \{0, 1\}^{2l(n)}$; output $P_r \rho P_r^\dagger \otimes \mathcal{O}(V_{r,k})$;
- $\text{Dec}_k(\sigma \otimes \tau)$: output the second register of $\mathcal{J}(\tau \otimes |k\rangle\langle k| \otimes \sigma)$.

To check correctness, we apply the functionality-preserving property of the obfuscator. A decryption of a valid encryption with the same key yields

$$\text{Dec}_k(\text{Enc}_k(\rho)) = \text{Tr}_1 [\mathcal{J}(\mathcal{O}(V_{r,k}) \otimes |k\rangle\langle k| \otimes P_r \rho P_r^\dagger)] = \text{Tr}_1 [U_{r,k}(|k\rangle\langle k| \otimes P_r \rho P_r^\dagger) V_{r,k}^\dagger] = \rho$$

as desired. We argue briefly that the scheme satisfies semantic security, which was defined and shown equivalent to IND in [ABF⁺16]. Let \mathcal{A} be a SEM adversary; we are charged with describing a simulator \mathcal{A}' . By the black-box property, \mathcal{A} on input $\text{Enc}_k(\rho)$ can be simulated by an adversary \mathcal{S} who receives $P_r \rho P_r^\dagger$ and only queries $V_{r,k}$ as a black box. We note that, since r is selected uniformly at random, and the plaintext and the obfuscated circuit are independent, we do not require the obfuscator to be secure in the presence of auxiliary information. Indeed, we can rewrite $P_r \rho P_r^\dagger = P_{r'} \rho' P_{r'}^\dagger$ for any other r' (and some ρ') and this does not change the execution of \mathcal{A} or \mathcal{S} . By a Grover lower bound, \mathcal{S} will succeed at distinguishing its oracle from the identity with at most negligible probability. Given this, the input of \mathcal{S} simply appears maximally mixed. We can thus simply define the simulator \mathcal{A}' to execute \mathcal{A} on $(\mathbb{1}/d) \otimes \mathcal{O}(\mathbb{1})$, regardless of the plaintext.

3.2.3 Public-key encryption from private-key encryption. For this application, we will need a stronger form of black-box obfuscation. Namely, the obfuscator should be secure in the presence of auxiliary information (see, e.g., Definition 3 in [GK05]). This amounts to requiring that the black-box condition now states

$$|\Pr[\mathcal{A}(\mathcal{O}(C) \otimes \rho) = \pi(C, \rho)] - \Pr[\mathcal{S}^{U_C}(|0^n\rangle \otimes \rho) = \pi(C, \rho)]| \leq \text{negl}(n),$$

where π is any predicate, and the auxiliary information ρ is allowed to depend on C . It's conceivable that the construction is possible even without auxiliary information; we leave this as an open problem.

Let $\Pi = (\text{KeyGen}, \text{Enc}, \text{Dec})$ be a symmetric-key quantum encryption scheme, with classical keys. We define a public-key quantum encryption scheme $\Pi' = (\text{KeyGen}', \text{Enc}', \text{Dec}')$ so that key generation outputs the key of Π_S as the private key, and $\mathcal{O}(\text{Enc}_k)$ as the public key. We will assume that the obfuscator outputs circuits, so that the public key is classical. One could also consider the case where public keys are quantum; more care is then required in setting down the definition of public-key schemes.

We argue that, if Π satisfies SEM-CPA (see [ABF⁺16]) then Π' satisfies SEM (for public-key schemes.) A Π' SEM adversary \mathcal{A} receives the public key $\mathcal{O}(\text{Enc}_k)$ and a state

$$\sigma_{AB} := [(\text{Enc}_k)_A \otimes \mathbb{1}_B] \rho_{AB}$$

which contains an encryption of the plaintext in register A , and a register B of side information about the plaintext. Assume there exists such an \mathcal{A} , for which no simulator \mathcal{S} (receiving only the public key and register B) exists. By black-box obfuscation with auxiliary information, there exists a QPT $\mathcal{A}_0^{\text{Enc}_k}$ which simulates⁵ \mathcal{A} on input σ_{AB} and interacts with the public key in a black-box manner. Note that \mathcal{A}_0 is a valid (but perhaps not successful) SEM-CPA adversary against Π : by functional equivalence of \mathcal{O} , the CPA oracle for Π is equivalent to the oracle \mathcal{A}_0 expects. We claim that \mathcal{A}_0 (with the same plaintext, side information, and objective predicate as \mathcal{A}) has no simulator. For a contradiction, suppose \mathcal{S}_0 was such a simulator. We can then build a simulator \mathcal{S} for \mathcal{A} simply by running \mathcal{S}_0 and answering its CPA queries using the public key $\mathcal{O}(\text{Enc}_k)$. We conclude that Π' satisfies SEM, and thus also IND.

3.2.4 Public-key quantum money

Quantum money. The idea of “quantum money” first arose in work by Wiesner [Wie83]. The core idea is simple: use a quantum state for representing currency in such a way that the no-cloning theorem of quantum mechanics prevents counterfeiting. These ideas were refined and developed further in several works [Aar09, AC12, BBBW83, FGH⁺12, MS10]; some of these works also included explicit proposals based on various hardness assumptions.

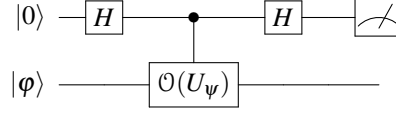
Informally, a *quantum money scheme* consists of two algorithms: Mint, which produces quantum states, and Verify, which accepts an input state and then either accepts or rejects. If the different states produced by Mint are distinguishable, then we refer to them as *bills*; if they are indistinguishable, then we call them *tokens* (if Verify consumes them) or *coins* (if Verify does not consume them.) In all quantum money schemes, we imagine an authority (typically called the bank) which runs Mint repeatedly to produce money; in addition, the Verify algorithm should accept only on states produced by the bank. Depending on the particular scheme, this might only be true if Verify is executed by the bank (private-key money), or it might be true for any party (public-key money.)

In this language, Wiesner's original idea [Wie83] was for a private-key scheme for bills, which is as follows. Each execution of Mint produces two random classical bitstrings $r, s \in \{0, 1\}^{2n}$ as well as an n -qubit quantum state $|\psi_r\rangle$, with each qubit initialized in one of the states $|0\rangle, |1\rangle, |+\rangle, |-\rangle$, as determined by the bits of r . The bank records the pair (r, s) in a secret table, and publishes $(s, |\psi_r\rangle)$. The bank verifies by using s to look up the correct r in the table, and then performing the measurements in the correct basis and checking the results against r .

Public-key money from circuit obfuscation. While private-key money schemes are relatively straightforward to construct, public-key proposals appear to be much more difficult, and require computational assumptions. Here we show that an obfuscator can sometimes be used to turn

⁵ Here one must check that it is sufficient to consider SEM adversaries which output predicates.

private-key money schemes to public-key ones. The use of an obfuscator to create a particular quantum money scheme was considered by Mosca and Stebila [MS10]. Their scheme (in our language) is as follows. Each execution of Mint produces a Haar-random n -qubit quantum state $|\psi\rangle$, together with the obfuscation $\mathcal{O}(U_\psi)$ of a circuit⁶ for $U_\psi = \mathbb{1} - 2|\psi\rangle\langle\psi|$. The bill consists of the pair $(\mathcal{O}(U_\psi), |\psi\rangle)$. Verify($|\phi\rangle$) consists of executing the following:



and accepting iff the measurement returns 1. It's easy to check that the above succeeds only on valid states; moreover, in that case, the state $|\psi\rangle$ is output in the second register, so that verification can be repeated. To show resistance of the above scheme to counterfeiting, one can use Aaronson's Complexity-Theoretic No-Cloning Theorem [Aar09], which states that cloning the state $|\psi\rangle$ while in possession of oracle access to $|U_\psi\rangle$ requires $\Omega(2^{n/2})$ queries. The first published proof of this theorem (as well as its first appearance in the form required here) was in [AC12].

Unfortunately, we will later show that obfuscation of quantum circuits in the form required by Mosca and Stebila is impossible. What remains possible is a setting in which both $|\psi\rangle$ and $\mathcal{O}(U_\psi)$ are quantum states, and another circuit (which is publicly known and independent of $|\psi\rangle$) is used for verification. Moreover, as we will also show, any black-box obfuscation scheme which outputs states that can be efficiently cloned is also impossible. We thus conjecture the following.

Conjecture 8. If quantum black-box obfuscators exist, then so do public-key quantum money schemes.

If the relevant obfuscation is a consumable state, then this would result in a token scheme. If it can be reused to perform verification repeatedly⁷, then the result would be a bills scheme. We remark that, in any case, all of the public-key money states discussed above should be authenticated by the bank; otherwise a merchant would only know that he was handed *some* pair (state, circuit) where the circuit executed on the state outputs “accept”—a clearly inadequate state of affairs.

3.3 Impossibility results

3.3.1 Impossibility of two-circuit obfuscation Barak et. al. [BGI⁺12] showed that black-box obfuscation is impossible by constructing an explicit circuit family that cannot be black-box obfuscated. We begin with a similar result in the quantum setting. We show that quantum black-box obfuscation is impossible in any setting where the adversary can gain access to two outputs of the obfuscator on *different* inputs. We formalize this notion by defining a “black-box two-circuit obfuscator,” defined just as in Definition 7 but with the following strengthening of the virtual black-box condition:

3. (two-circuit virtual black-box) for every pair of quantum circuits C_1 and C_2 and every quantum adversary \mathcal{A} there exists a quantum simulator $\mathcal{S}^{U_{C_1}, U_{C_2}}$ and a negligible ϵ_2 such that

$$\left| \Pr[\mathcal{A}(\mathcal{O}(C_1) \otimes \mathcal{O}(C_2)) = 1] - \Pr[\mathcal{S}^{U_{C_1}, U_{C_2}}(|0\rangle^{\otimes |C_1| + |C_2|}) = 1] \right| \leq \epsilon_2(n, \min\{|C_1|, |C_2|\}).$$

⁶ For most $|\psi\rangle$, the circuit U_ψ will not have polynomial length. However, as pointed out by [Aar09], one can instead select $|\psi\rangle$ from an approximate t -design without a significant loss in security.

⁷ For example, if successful verification also outputs another state which is sufficiently close to the original state.

We now show that there exists a family of circuits which is unobfuscatable under the above definition. We emphasize that our result holds even when the outputs of the obfuscator are quantum states, and even if these states are *single-use only*, i.e., if the interpreter \mathcal{J} irrevocably destroys the obfuscated state during use.

We first define a *circuit-pair family* to be an ensemble of distributions over pairs of circuits. More precisely, if \mathcal{C} is a circuit-pair family, then there exists a Turing machine M which, on input a positive integer parameter n (in unary), outputs a classical description of a pair of circuits (C_n, D_n) drawn at random from some distribution \mathcal{C}_n on pairs of $\text{poly}(n)$ -size circuits. If M is polynomial-time, then we say that \mathcal{C} is a *poly-time circuit-pair family*.

We also define a *state-pair family* analogously. If \mathcal{C}' is a state-pair family, then there exists a (not necessarily polynomial-time) quantum algorithm which, on input n in unary, outputs a pair of density operators (ρ_n, σ_n) drawn at random from some distribution \mathcal{C}'_n on quantum states on $\text{poly}(n)$ -many qubits. Given a circuit-pair family \mathcal{C} and a state-pair family \mathcal{C}' , we say that \mathcal{C}' is an obfuscation of \mathcal{C} if there exists a computable map $\mathcal{C} \rightarrow \mathcal{C}'$ assigning to each circuit a corresponding state, in a manner that satisfies the two-circuit obfuscation definition above.

With these definitions, we can now state our first impossibility result.

Theorem 9. *There exists a poly-time quantum circuit-pair family \mathcal{C} such that no state-pair family is an obfuscation of \mathcal{C} .*

Proof. Let $(\mathcal{O}, \mathcal{J})$ be a black-box quantum two-circuit obfuscator. The poly-time quantum circuit-pair family \mathcal{C} consists of quantum circuits for implementing the following pairs of unitary operators. Each pair is parameterized by an input size n , as well as bitstrings a, b chosen uniformly at random from $\{0, 1\}^n$.

$$U_{a,b} : |x, y\rangle \mapsto \begin{cases} |x, y \oplus b\rangle & \text{if } x = a; \\ |x, y\rangle & \text{otherwise.} \end{cases} \quad (3.1)$$

$$V_{a,b} : |C, z\rangle \mapsto \begin{cases} |C, z \oplus 1\rangle & \text{if } C(a) = b; \\ |C, z\rangle & \text{otherwise.} \end{cases} \quad (3.2)$$

The registers indexed by x and y are of size n . The register indexed by C accepts a circuit description (under some fixed encoding), and needs to be able to handle inputs of size $|\mathcal{O}(C_{a,b})|$ (i.e. of size equal to the number of qubits in the state $\mathcal{O}(C_{a,b})$). Here $C_{a,b}$ is a fixed, explicit $\text{poly}(n)$ -size circuit for $U_{a,b}$. The second register of $V_{a,b}$ has size one.

Note that both of these unitaries can be implemented by efficient quantum circuits. We choose some particular set of such circuits, and henceforth denote them by $C_{a,b}$ and $D_{a,b}$, respectively. The idea for the proof is as follows. Consider an adversary \mathcal{A} which is ignorant of the randomly selected a and b , and consider two scenarios: in the first, \mathcal{A} is given access to *any* pair of circuits that implement $U_{a,b}$ and $V_{a,b}$; in the second, \mathcal{A} only has oracle access to $U_{a,b}$ and $V_{a,b}$. The point is that, in the first case, \mathcal{A} can execute $V_{a,b}$ on a circuit for $U_{a,b}$; provided that the latter is not too long, \mathcal{A} will achieve something that is impossible to do with only black-box access. Specifically, it is only in the first case that \mathcal{A} will be able to tell if the first circuit/oracle implements $U_{a,b}$, or if it has surreptitiously been replaced by the identity operator!

Things are somewhat complicated by the fact that the obfuscator outputs states instead of circuits. We will need to enable \mathcal{A} to execute these states on one another. It will thus be necessary to replace $D_{a,b}$ with a related circuit $D'_{a,b}$. Roughly speaking, this circuit will check if its input, when interpreted as a quantum advice state to the algorithm \mathcal{J} , maps the input a to the output b . A precise description follows. First, $D'_{a,b}$ will have three registers: an input register of m qubits, a work register of $2n$ qubits, and an output register of 1 qubit initialized in the $|0\rangle$ state. When given as input a quantum state ρ on m qubits, it will initialize the first n bits of the work register to $|a\rangle$, then execute the appropriate unitary circuit of \mathcal{J} on $\rho \otimes |a\rangle$. Finally, if the output register of the latter computation contains $|b\rangle$, $D'_{a,b}$ will flip the contents of the output register. We remark

that, by a simple counting argument over circuits, this occurs for only an exponentially small fraction of possible input states ρ .

Recall that the $2n$ -qubit identity operator is denoted by $\mathbb{1}_{2n}$, and is implemented by the obvious circuit which we will denote by I_{2n} . We observe that, for every QPT algorithm \mathcal{S} there exists a polynomial t and a negligible ε_1 so that:

$$\left| \Pr[\mathcal{S}^{U_{a,b}, D'_{a,b}}(|0\rangle^{\otimes t(n)}) = 1] - \Pr[\mathcal{S}^{I_{2n}, D'_{a,b}}(|0\rangle^{\otimes t(n)}) = 1] \right| \leq \varepsilon_1(n). \quad (3.3)$$

Here the probability is taken over the uniformly random choice of a and b as well as all of the measurement outcomes of \mathcal{S} . The above is an easy corollary of the tightness of the Grover bound for unstructured quantum search [BBBV97]. Indeed, given the definitions of $U_{a,b}$ and $D'_{a,b}$, it's clear that with only polynomial queries and no knowledge of a or b , \mathcal{S} is faced precisely with unstructured search for an exponentially small “marked space.” This marked space is only encountered if \mathcal{S} correctly guesses a , or correctly guesses an obfuscation of a circuit that maps a to b .

Now consider the QPT algorithm \mathcal{A} that, given as input the obfuscated states $\mathcal{O}(C)$ and $\mathcal{O}(D)$, simply executes the quantum algorithm \mathcal{J} on their tensor product, accepting if and only if the outcome is 1. Notice that this succeeds with constant probability $\alpha > 0$ if C is functionally equivalent to $C_{a,b}$ and D is functionally equivalent to $D'_{a,b}$. On the other hand, this same algorithm \mathcal{A} accepts with at most negligible probability when C is functionally equivalent to I_{2n} (and D is still functionally equivalent to $D'_{a,b}$); indeed, this only happens if $a = b$. Thus there exists a negligible function ε_2 so that:

$$\left| \Pr[\mathcal{A}(\mathcal{O}(D'_{a,b}), \mathcal{O}(I_{2n})) = 1] - \Pr[\mathcal{A}(\mathcal{O}(D'_{a,b}) \otimes \mathcal{O}(C_{a,b})) = 1] \right| \geq \alpha - \varepsilon_2(n). \quad (3.4)$$

To complete the proof, we explicitly define the poly-time circuit-pair family \mathcal{C} . The distribution \mathcal{C}_n is generated by choosing a, b uniformly at random from $\{0, 1\}^n$, and then choosing a bit $r \in \{0, 1\}$ at random; if $r = 0$, we output the circuit pair $(C_{a,b}, D'_{a,b})$, and otherwise we output $(I_{2n}, D'_{a,b})$. For this distribution, equations (3.3) and (3.4) together show that no state-pair family is an obfuscation of \mathcal{C} .

3.3.2 Generalizing the impossibility result Our goal in this section is to extend the two-circuit impossibility proof from the prior section to the case of obfuscating a single circuit. For our impossibility proof, we require an additional condition on the obfuscator: that each of its outputs is reusable a polynomial number of times. This is a natural condition which is automatically satisfied by classical obfuscators (as well as quantum obfuscators with classical outputs), since their outputs can be perfectly copied.

Definition 10. A *reusable-black-box quantum obfuscator* is a quantum algorithm \mathcal{O} and a QPT \mathcal{J} such that whenever C is an n -qubit quantum circuit, $\mathcal{O}(C)$ is an m -qubit quantum state satisfying

1. (polynomial slowdown) $m = \text{poly}(n, |C|)$;
2. (functional equivalence) $\|\mathcal{J}(\mathcal{O}(C) \otimes \cdot) - C \cdot C^\dagger\|_\diamond \leq \text{negl}(n, |C|)$;
3. (reusability) after execution of \mathcal{J} , an output register contains a state which satisfies (2.);
4. (virtual black-box) for every QPT adversary \mathcal{A} there exists a QPT simulator \mathcal{S}^{U_C} such that:

$$\left| \Pr[\mathcal{A}(\rho_{(i)}) = 1] - \Pr[\mathcal{S}^{U_C}(|0\rangle^{\otimes |C|}) = 1] \right| \leq \text{negl}(n, |C|).$$

We remark that reusability can be achieved in any number of ways: by providing a state which partially survives uses by the interpreter \mathcal{J} , by providing sufficiently many copies, or by providing a means of cloning the state. We prove impossibility of the above definition in any setting where the adversary receives two copies of the obfuscator output, even on identical inputs. This is

automatically satisfied if the obfuscator provides multiple copies in order to satisfy reusability, or if the state is (even approximately) cloneable. The key new obstacle is to prove impossibility even though the functionality for both copies is *the same*.

To state the result, we define (in analogy to circuit-pair families and state-pair families) a *circuit family* to be an ensemble of distributions over circuits, and a *state family* to be an ensemble of distributions over states. A state family \mathcal{C}' is said to be an obfuscation of a circuit family \mathcal{C} if there exists a computable map $\mathcal{C} \rightarrow \mathcal{C}'$ assigning to each circuit a corresponding state, in a manner that satisfies [Definition 10](#). With these definitions, we will prove the following theorem.

Theorem 11. *If quantum-secure one-way functions exist, then there exists a quantum circuit family \mathcal{C} such that no state family is a reusable-black-box quantum obfuscation of \mathcal{C} .*

Since the full proof of [Theorem 11](#) is somewhat lengthy and involved, we will first prove a simpler case, showing that quantum circuits cannot be obfuscated into quantum circuits, under any of the definitions considered so far. This corollary (stated below as [Theorem 12](#)) is arguably the most direct quantum generalization of the impossibility result of [\[BGI⁺01\]](#). Once we have proved it, we will explain in detail how the proof should be adapted in order to achieve [Theorem 11](#).

Theorem 12. *If quantum-secure one-way functions exist, then there exists a quantum circuit family \mathcal{C} such that no quantum circuit family is a black-box obfuscation of \mathcal{C} .*

Proof. Let \mathcal{O} be a black-box quantum obfuscator satisfying [Definition 7](#), such that its outputs are classical bitstrings. Since these states are used to describe an efficiently implementable quantum computation, we can assume that these bitstrings are in fact quantum circuits under some particular encoding.

To construct the unobfuscatable circuit family, we will need a notion of combining the functionality of two quantum circuits into one.

Definition 13. *The **combined quantum circuit** of a finite collection $\{C_1, C_2, \dots, C_k\}$ of n -qubit quantum circuits is the circuit that has two registers (a control register of $\log k$ qubits, and an input register of n qubits) and, controlled on the value of the first register, applies the respective quantum circuit to the input register.*

Notice that if each circuit C_i in the collection is polynomial size, and k is bounded by a polynomial in n , then the associated combined quantum circuit is also of polynomial size. We will denote the operation of combining circuits with $\#$. For example, the combined circuit of two circuits C_1 and C_2 is denoted $C_1\#C_2$.

Now recall the two circuits $C_{a,b}$ and $D_{a,b}$ from [Section 3.3.1](#), as well as the circuit I_{2n} , which simply implements the identity operator on $2n$ qubits. Consider the combined quantum circuits $C_{a,b}\#D_{a,b}$ and $I_{2n}\#D_{a,b}$, sampled by selecting a and b uniformly at random from $\{0, 1\}^n$. We again choose $C = C_{a,b}$ or $C = I_{2n}$, each with probability $1/2$, and ask the adversary to determine which is the case. Using the same reasoning as in the proof of [Theorem 9](#) from [Section 3.3.1](#), these combined quantum circuit distributions are indistinguishable from the perspective of any QPT simulator that is ignorant of a and b , and is given only black-box access to $U_{C\#D_{a,b}}$. On the other hand, unlike in the prior proof, it is not immediately apparent how to distinguish the two possibilities given a circuit description of $\mathcal{O}(C\#D_{a,b})$. Still, the idea is simple. Suppose we have two copies of the circuit. We can hard-wire the control register of one copy to implement $D_{a,b}$, and hard-wire the control register of the other copy to implement C . If we then run the first copy on the second, the result should be equivalent to implementing $D_{a,b}$ on input C , which will determine the nature of C and conclude the proof just as in [Theorem 9](#).

Unfortunately, this idea does not work as stated, because the two circuit copies have the same size. Since they are functionally nontrivial, their input size is much smaller than their description, making it impossible to run one on the other. The core difficulty is that $D_{a,b}$ cannot be made large enough to universally execute circuits of size $|\mathcal{O}(C\#D_{a,b})| > p(|D_{a,b}|)$ where p is the running

time of \mathcal{O} . This issue arises in the classical proof as well, and is resolved as follows. First, note that we could have $D_{a,b}$ simply provide a and b , thus offloading the gate-by-gate execution of C to the algorithm \mathcal{A} in the black-box definition. Unfortunately, this would also provide the simulator \mathcal{S} with a and b , enabling it to simulate \mathcal{A} . The resolution is to have $D_{a,b}$ provide *encryptions of a and b* , as well as a *quantum fully-homomorphic encryption (QFHE) oracle* for homomorphically applying the gates of C . We emphasize that the functionality and security of the QFHE oracle crucially depends on the obfuscation property; in particular, an actual QFHE scheme is not required for the proof.

Concretely, we prove the following Lemma, which is a quantum analogue of Lemma 3.6 (including Claim 3.6.1) from [BGI⁺12].

Lemma 14. *If quantum-secure one-way functions exist, then for each $n \in \mathbb{N}$ and $a, b \in \{0, 1\}^n$, there exists a distribution $\mathcal{D}_{a,b}$ over circuits such that:*

1. *There exists a PPT algorithm that, given $n \in \mathbb{N}$ and $a, b \in \{0, 1\}^n$, samples from $\mathcal{D}_{a,b}$;*
2. *There is a QPT algorithm \mathcal{A} so that $C|a\rangle|0^n\rangle = |a\rangle|b\rangle$ implies $\mathcal{A}^{U_D}(C, 1^n) = a$; this holds for all $n \in \mathbb{N}$, all $a, b \in \{0, 1\}^n$, any $D \in \text{supp}(\mathcal{D}_{a,b})$, and every n -qubit circuit C ;*
3. *For any QPT \mathcal{S} , $\Pr[\mathcal{S}^{U_D}(1^n) = a] \leq \text{negl}(n)$, where the probability is over $a, b \in \{0, 1\}^n$, $D \sim \mathcal{D}_{a,b}$, and the measurements of \mathcal{S} .*

Proof. The distribution $\mathcal{D}_{a,b}$ will be sampled by choosing $k, r \in_R \{0, 1\}^{2n}$. The bitstring k is to serve as a private key for the IND-CCA1-secure symmetric-key quantum encryption scheme from Theorem ?? . Each circuit $D \in \text{supp}(\mathcal{D}_{a,b})$ will be a combination (again via #) of the following three circuits.

1. $E_{K,a}$; this outputs $\text{Enc}_k(|a\rangle)$, executed with randomness r .
2. $\text{Hom}_K(G, \rho)$; on input a gate description G and a state ρ , it outputs $[\text{Enc}_k \circ G \circ \text{Dec}_k](\rho)$.
3. $B_{k,a,b}$; on input ρ , it outputs $|a\rangle$ if $\text{Dec}_k(\rho) = |b\rangle$ and $|0^n\rangle$ otherwise.

We remark that the Hom oracle requires randomness in order to re-encrypt the state. This is handled in the usual way: we expand the input in some register via a (quantum-secure) pseudo-random function; these exist by the assumption of quantum-secure one-way functions and Theorem 3.

Clearly, given a and b , $\mathcal{D}_{a,b}$ can be sampled efficiently by choosing k uniformly at random and outputting the combined quantum circuit $D_{k,a,b} := E_{k,a} \# \text{Hom}_k \# B_{k,a,b}$. This establishes Property 1 from the Lemma. For Property 2, let \mathcal{A} be the algorithm that, on input a quantum circuit C , (i.) uses the first two circuits comprising $D_{k,a,b}$ to homomorphically simulate C gate-by-gate on a , and then (ii.) plugs the final state into $B_{k,a,b}$.

To complete the proof of Lemma 14, we must verify Property 3, i.e., that no QPT simulator algorithm that has black-box access to each of the three algorithms comprising $D_{k,a,b}$ can discover a with non-negligible probability. This amounts to showing that

$$\left| \Pr[\mathcal{S}^{\text{Hom}_k, \text{Enc}_k}(\text{Enc}_k(|0\rangle)) = 1] - \Pr[\mathcal{S}^{\text{Hom}_k, \text{Enc}_k}(\text{Enc}_k(|a\rangle)) = 1] \right| \leq \text{negl}(n), \quad (3.5)$$

where the probabilities are over $k \in_R \{0, 1\}^n$ and the measurement outcomes of \mathcal{S} . We proceed by contradiction, and assume that there's a QPT \mathcal{S} that violates the claim.

First, we replace the responses to all of \mathcal{S} 's queries to the Hom_k oracle with encryptions of $|0^n\rangle$, and deploy a hybrid argument to show that the success probability of \mathcal{S} is not significantly affected. To this end, consider the following hybrids of the computation of \mathcal{S} on input $\text{Enc}_k(|a\rangle)$: in the i -th hybrid, the first i oracle queries of \mathcal{S} are answered using Hom_k , and the rest are answered with $\text{Enc}_k(|0^n\rangle)$. Notice that any gap in distinguishing between the i and $i+1$ st hybrid must be due to the $i+1$ st query \mathcal{S} makes to Hom_k . We can use this to create a CCA1 adversary \mathcal{T} which breaks the encryption scheme, as follows. The QPT \mathcal{T} simulates \mathcal{S} and replies to its first i oracle queries by means of \mathcal{T} 's Enc_k and Dec_k oracles. Upon receiving the challenge ciphertext,

\mathcal{T} passes it to \mathcal{S} as the response to its $i + 1$ st oracle query. Finally, \mathcal{T} answers the remaining oracle queries of \mathcal{S} with $\text{Enc}_K(|0^n\rangle)$. We conclude that, if \mathcal{S} violated (3.5), then \mathcal{T} succeeds with non-negligible probability. This establishes (by contradiction) that we can replace the oracle queries of \mathcal{S} with $\text{Enc}_K(|0\rangle)$. With this replacement, \mathcal{S} can distinguish an encryption of $|0^n\rangle$ from an encryption of $|a\rangle$, when given access to only an encryption oracle, which again contradicts IND-CCA1 security of the scheme.

Now we are ready to describe the unobfuscatable family of quantum circuits and complete the proof of [Theorem 12](#). First, for a fixed $a, b \in \{0, 1\}^n$ we let $\mathcal{D}_{a,b}$ be the distribution over circuits constructed in [Lemma 14](#). Then consider the following two distributions over circuits:

1. \mathcal{F}_n : Choose $a, b \in_R \{0, 1\}^n$, sample a circuit $D_{a,b}$ from $\mathcal{D}_{a,b}$ and output $C_{a,b} \# D_{a,b}$
2. \mathcal{G}_n : Choose $a, b \in_R \{0, 1\}^n$, sample a circuit $D_{a,b}$ from $\mathcal{D}_{a,b}$ and output $I_{2n} \# D_{a,b}$

By Property 2 of [Lemma 14](#) there exists an algorithm \mathcal{A} that, on input $\mathcal{O}(C_0)$, accepts if C_0 was sampled from \mathcal{F}_n and rejects if C was sampled from \mathcal{G}_n . Thus there exists a constant α and a negligible function ε_1 so that:

$$\left| \Pr[\mathcal{A}(\mathcal{O}(\mathcal{F}_n)) = 1] - \Pr[\mathcal{A}(\mathcal{O}(\mathcal{G}_n)) = 1] \right| \geq \alpha - \varepsilon_1(n).$$

On the other hand, by Property 3 of [Lemma 14](#), we know that for every QPT \mathcal{S} there exists some negligible function ε_2 so that:

$$\left| \Pr[\mathcal{S}^{\mathcal{F}_n}(|0\rangle^{\otimes n}) = 1] - \Pr[\mathcal{S}^{\mathcal{G}_n}(|0\rangle^{\otimes n}) = 1] \right| \leq \varepsilon_2(n).$$

We conclude that the circuit family formed by taking the union of \mathcal{F}_n with \mathcal{G}_n (and assigning them each equal probability) is an unobfuscatable circuit family.

We now return to the proof of [Theorem 11](#), and show how to extend the above proof to the case where the obfuscator outputs reusable quantum states.

Proof. (of [Theorem 11](#)) Our proof will still use the same distribution $\mathcal{D}_{a,b}$ over circuits, which were provided by [Lemma 14](#) and described above, but with some slight modifications. The goal will still be to take two copies of $\mathcal{O}(C_0)$ for any circuit C_0 sampled from that distribution, and give an algorithm \mathcal{A} that can “execute one copy on the other.” This will enable us to distinguish if C_0 is from the distribution \mathcal{F}_n , or the distribution \mathcal{G}_n (just as above), a task which is impossible with only black-box access.

However, executing one copy of $\mathcal{O}(C_0)$ on another is now somewhat more complicated, due to the fact that we no longer have explicit circuit descriptions in hand, and must instead use the interpreter \mathcal{J} (with some register initialized to $\mathcal{O}(C_0)$) whenever we want to run C_0 . To do this, we will need to describe a new distribution $\mathcal{D}'_{a,b}$ of circuits, closely related to the distribution $\mathcal{D}_{a,b}$ from [Lemma 14](#).

Attempt 1. To warm up, a first attempt at describing \mathcal{A} and the modified circuits $D'_{a,b}$ from the distribution $\mathcal{D}'_{a,b}$ is as follows. First, we simply define $D'_{a,b}$ to be a composition of circuits which simply output both a and b . Set $C_0 = C \# D'_{a,b}$, and suppose that inputting $|0\rangle$ in the first register executes the first circuit in the combination, while $|1\rangle$ executes the second circuit in the combination. The algorithm \mathcal{A} is in possession of two copies of $\mathcal{O}(C_0)$. It performs:

1. run $\mathcal{J}(\mathcal{O}(C_0) \otimes |1\rangle|0^n\rangle)$ to retrieve $|a\rangle|b\rangle$ (by functional equivalence of \mathcal{O});
2. run $\mathcal{J}(\mathcal{O}(C_0) \otimes |0\rangle|a\rangle)$ (now using the other copy of $\mathcal{O}(C_0)$);
3. compare the result to $|b\rangle$.

This does exactly what we want, except of course that the black-box simulator \mathcal{S} will also be able to retrieve a and perform this experiment. Our valiant attempt failed.

Attempt 2. Undeterred, we now try a more sophisticated approach, returning to the idea of encryption and homomorphic execution. We now ask that (as before) $\mathcal{D}'_{a,b}$ outputs an encryption $\text{Enc}(a)$ of a (when given the flag A), implements a Hom oracle (when given the flag H), and checks if the input is $\text{Enc}(b)$ (when given the flag B). We again set $C_0 = C\#D'_{a,b}$ and give \mathcal{A} two copies of $\mathcal{O}(C_0)$; for convenience we denote them $\mathcal{O}(C_0)$ and $\mathcal{O}(C_0)'$. Now, \mathcal{A} performs:

1. run $\mathcal{J}(\mathcal{O}(C_0) \otimes [|1A\rangle \otimes |0^n\rangle])$ to retrieve $\text{Enc}(a)$;
2. run $\mathcal{J}(\mathcal{O}(C_0) \otimes [|1H\rangle \otimes \mathcal{O}(C_0)' \otimes |0\rangle \otimes \text{Enc}(a)])$ to “homomorphically run C ”;
3. run $\mathcal{J}(\mathcal{O}(C_0) \otimes [|1B\rangle \otimes \rho])$ where ρ is the output of the previous step; output the result.

The first and last step are largely self-explanatory: we start with the encryption of a , and check at the end that we have the encryption of b . What happened in the second step? We tried to homomorphically evaluate⁸ C on $\text{Enc}(a)$. By functional equivalence of \mathcal{O} , we executed the first copy of C_0 on input $|1H\rangle \otimes \mathcal{O}(C_0)' \otimes \text{Enc}(a)$; this specifies that $D'_{a,b}$ should run the Hom oracle with input $\mathcal{O}(C_0)' \otimes \text{Enc}(a)$. To make this sensible, we can redefine Hom to accept two registers, and homomorphically evaluate the appropriate circuit of \mathcal{J} ; the result is that, whenever Hom is called on $\mathcal{O}(C) \otimes \text{Enc}(z)$ for a circuit C and state z , the output is $\text{Enc}(C(z))$.

This attempt looks like it succeeds, but there is a disastrous flaw: Hom must now accept inputs with at least as many qubits as $\mathcal{O}(C_0)$, which is significantly bigger than the circuit description allowed for Hom itself (since it is a part of $D'_{a,b}$ and hence also of C_0).

Attempt 3. In the final attempt, we will repair the flaw of Attempt 2. The key is to again offload some of the execution, but this time from the Hom oracle to the main algorithm \mathcal{A} . More precisely, we will expand step (2) in Attempt 2, and execute it gate-by-gate. In this iteration, Hom is back to its original version, and is used only to apply two-qubit gates. It accepts n input qubits, decrypts them, applies the desired gate (as specified in another register), and then re-encrypts. In addition, will also expand $D'_{a,b}$ to provide an Enc oracle (when given the flag E); we can do this for free, by equation (3.5) in Lemma 14. The final algorithm \mathcal{A} will proceed as follows. Here we have let m denote the number of qubits of $\mathcal{O}(C_0)'$, and we let J_m be the circuit of \mathcal{J} for executing m -qubit obfuscated states.

1. run $\mathcal{J}(\mathcal{O}(C_0) \otimes [|1E\rangle \otimes \mathcal{O}(C_0)'_{(k)}])$ for all $k \in [m]$, to encrypt all qubits of $\mathcal{O}(C_0)'$;
2. run $\mathcal{J}(\mathcal{O}(C_0) \otimes [|1A\rangle \otimes |0^n\rangle])$ to retrieve $\text{Enc}(a)$;
3. let $j = 0$ and let $\sigma_j := \text{Enc}(a)$;
4. let G_j be the j th gate in the description of J_m ;
 - (a) let s, t be the qubits G_j acts on; assume s is a qubit of $\mathcal{O}(C_0)'$ and t is a qubit of σ_j ;⁹
 - (b) set $\sigma_{j+1} = \mathcal{J}(\mathcal{O}(C_0) \otimes [|1H\rangle \otimes |G_j, s, t\rangle \otimes \text{Enc}(\mathcal{O}(C_0)'_{(s)}) \otimes \sigma_j])$;¹⁰
 - (c) if $j = |J_m|$, continue; otherwise increment j by 1 and go to step 4.
5. run $\mathcal{J}(\mathcal{O}(C_0) \otimes [|1B\rangle \otimes \sigma_j])$ and output the result.

A few remarks are in order. First, the reusability of the state $\mathcal{O}(C)$ was crucial in our ability to repeatedly execute the Hom oracle in step 4.(b). Second, one checks that by the functional equivalence of the obfuscator and the definition of the Hom oracle, if $C_0 = C_{a,b}\#D'_{a,b}$ then the state σ_j when step 5. is reached will indeed be $\text{Enc}(b)$. Third, note that the “compactness” issue of Hom from Attempt 2 has been resolved, and the input to C_0 in step 4.(b) is now of size n (plus a constant.)

Finally, despite all of the difficulties in defining the algorithm \mathcal{A} appropriately, the hardness of the corresponding search problem for the black-box simulator \mathcal{S} is essentially unchanged from

⁸ Selecting C was done by passing the bit flag $|0\rangle$ into the control register.

⁹ This assumption is only made for simplicity of the algorithm description; the other possibilities are similar.

¹⁰ The notation $\mathcal{O}(C_0)'_{(s)}$ is meant to indicate that only the s -th qubit of that state is to be placed in the input register.

the proof of [Theorem 12](#). The only difference is that $D'_{a,b}$ now also provides an encryption oracle; the encryption scheme we selected is certainly secure in this setting.

To finish the proof, we again build a circuit family by choosing $C_{a,b}\#D'_{a,b}$ or $I_{2n}\#D'_{a,b}$, each with equal probability, for random a and b . By the above arguments, this circuit family is unobfuscatable. This concludes the proof of [Theorem 11](#).

4 Quantum indistinguishability obfuscation

In this section, we analyze a different definition of quantum obfuscation, motivated by classical definitions established by Goldwasser and Rothblum [[GR07](#)]. As opposed to the black-box approach, these definitions assess the quality of an obfuscation in relative terms, e.g., as compared to other functionally-equivalent circuits (or, in the quantum case, states).

4.1 CPTP circuits and ensembles of states

For our discussions on indistinguishability obfuscation, we modify the notations of [Section 1.3](#) slightly, as follows.

First, henceforth we take the point of view that any quantum circuit C can include unitary gates *as well as measurement gates and instructions for discarding qubits*. In keeping with this view, we overload notation so that C denotes both the circuit itself (i.e., a classical description of a set of wires and gates) as well as the CPTP map that the circuit implements.

Second, we allow our ensembles to be indexed by infinite sets of strings rather than the natural numbers. As before, we will denote ensembles by italicized capital letters (and optionally place the indexing set in the subscript.) Elements of the ensemble will be denoted by the corresponding non-italicized capital letter (optionally, with the index in the subscript.) So, for example, we may write $\mathcal{C}_S := \{C_s : s \in S\}$. Just as in the single-circuit case, we will overload this notation for circuit ensembles and use it to also refer to the corresponding family of CPTP maps. We will need one new definition in this context: given two circuit ensembles \mathcal{C}_S and \mathcal{D}_S , we say that \mathcal{C}_S is **functionally equivalent** to \mathcal{D}_S (denoted $\mathcal{C}_S \cong \mathcal{D}_S$) if the circuits themselves are functionally equivalent, i.e., if $\|C_s - D_s\|_\diamond \leq \text{negl}(|s|)$ for all $s \in S$.

We will also now need to handle infinite collections of states. We thus define a **state ensemble** to be an infinite collection $\{\rho_x : x \in X\}$ of density operators indexed by some set $X \subset \{0, 1\}^*$, such that $\rho_x \in \mathcal{D}(\mathcal{H}_{p(|x|)})$ where p is bounded by some fixed polynomial function. We remark that a circuit ensemble is a special case of a state ensemble, where each state is a classical string describing wires, gates, and so on. A **uniform state ensemble** will be a state ensemble $\{\rho_x : x \in X\}$ together with a uniform circuit ensemble $\{C_x : x \in X\}$ such that $\rho_x = C_x|0^m\rangle$ for appropriately chosen m . We remark that a QPT is defined by choosing a uniform circuit ensemble, and that the set of possible outputs of a QPT are a uniform state ensemble. In particular, if \mathcal{S} is a state ensemble and \mathcal{A} is a QPT, then (ignoring some uninteresting bookkeeping) we may write $\mathcal{A}(\mathcal{S})$ to denote the state ensemble that results from running \mathcal{A} on inputs from \mathcal{S} . Note that if \mathcal{S} is uniform, then $\mathcal{A}(\mathcal{S})$ is also uniform.

Next, we consider three different notions of distinguishability for state ensembles, in order of decreasing power. We will write $\mathcal{R} \approx \mathcal{S}$ to indicate that \mathcal{R} and \mathcal{S} are indistinguishable state ensembles; the type of indistinguishability should be clear from context. Let $\mathcal{R} = \{\rho_x : x \in X\}$ and $\mathcal{S} = \{\sigma_x : x \in X\}$ be two state ensembles. We say that \mathcal{R} and \mathcal{S} are **perfectly indistinguishable** if $\rho_x = \sigma_x$ for every $x \in X$. While this is an unnatural notion of indistinguishability for general quantum states, it is more reasonable (and is easy to test for individual cases) if ρ_x and σ_x happen to all be classical. A weaker notion is **statistical indistinguishability**, which demands that $\|\rho_x - \sigma_x\|_{\text{tr}} \leq \text{poly}(|x|)$ for all but finitely many $x \in X$. In the weakest notion we will consider, we say that \mathcal{R} and \mathcal{S} are **computationally indistinguishable** if for every QPT \mathcal{A} ,

$$|\Pr[\mathcal{A}(\rho_x) = 1] - \Pr[\mathcal{A}(\sigma_x) = 1]| \leq \text{negl}(|x|)$$

for all but finitely many x ; here the probabilities are taken over the coins and measurements of \mathcal{A} . One may also consider a non-uniform version of the above definition, in which \mathcal{A} also ranges over non-uniform circuit families, and is allowed access to an auxiliary state ensemble $\{\xi_x : x \in X\}$. All of our results hold (with appropriate adjustments) for both the uniform and the non-uniform setting; we will focus on the uniform setting for convenience.

4.2 Definitions: indistinguishability, best-possible

We begin with the notion of a quantum indistinguishability obfuscator. As before, the interpreter and the obfuscated state must be efficient, while the obfuscation algorithm itself might not be. We also assume that all “interpreter” algorithms \mathcal{J} have two registers: an advice register (where the obfuscated state is to be inserted), and an input register (where the input is to be inserted.) It will thus be convenient to write, e.g., \mathcal{J}_ρ for the CPTP map family defined by the circuits of \mathcal{J} with the advice register initialized to the state ρ .

Definition 15. A *quantum translator* is a quantum algorithm \mathcal{O} and a QPT \mathcal{J} such that whenever \mathcal{C} is a circuit ensemble and $C \in \mathcal{C}$ is an n -qubit circuit,

1. (polynomial slowdown) $\mathcal{O}(C)$ has at most $\text{poly}(n)$ qubits;
2. (functional equivalence) $\|\mathcal{J}_{\mathcal{O}(C)} - C\|_\diamond \leq \text{negl}(n)$;

Definition 16. A *quantum statistical (resp., computational) indistinguishability obfuscator* is a quantum translator $(\mathcal{O}, \mathcal{J})$ such that whenever \mathcal{C}_S and \mathcal{D}_S are functionally-equivalent circuit ensembles with $|C_s| = |D_s|$ for all $s \in S$, then $\mathcal{O}(\mathcal{C}_S)$ and $\mathcal{O}(\mathcal{D}_S)$ are statistically (resp., computationally) indistinguishable.

Note that an obfuscator which simply outputs circuit descriptions is included as a special case of the above; in that case, $\mathcal{O}(C)$ is always a quantum circuit, and \mathcal{J} is a universal circuit which executes $\mathcal{O}(C)$ on the state given in the input register. One may also define a quantum perfect-indistinguishability obfuscator, where the obfuscated states $\mathcal{O}(C_1)$ and $\mathcal{O}(C_2)$ are identical. We remark that the condition of equal length can be relaxed to any fixed polynomial (e.g., $|C_1|$ can be of length at most $|C_2|^2$.)

The notion of indistinguishability obfuscation may not seem intuitive at first. To begin to see why it is a useful definition, we now show that it is equivalent to a semantic definition of obfuscation. To achieve this, we first need to properly define functional equivalence for state ensembles. If we fix a quantum translator $(\mathcal{O}, \mathcal{J})$, then the QPT \mathcal{J} defines a way to implement functionality via states. We then say that two state ensembles \mathcal{R}_X and \mathcal{S}_X are functionally equivalent if

$$\|\mathcal{J}_{\rho_x} - \mathcal{J}_{\sigma_x}\|_\diamond \leq \text{negl}(|x|)$$

for all but finitely many $x \in X$, $\rho_x \in \mathcal{R}_X$ and $\sigma_x \in \mathcal{S}_X$. In this case, we will write $\mathcal{R}_X \cong_{\mathcal{J}} \mathcal{S}_X$. This allows us to compare the relative strengths of one obfuscated ensemble versus another, as follows.

Definition 17. A *quantum statistical (resp., computational) best-possible obfuscator* is a quantum translator $(\mathcal{O}, \mathcal{J})$ such that for any QPT \mathcal{A} there exists a QPT \mathcal{S} with the following property: for every circuit ensemble \mathcal{C}_X and any uniform state ensemble \mathcal{R}_X which is functionally-equivalent to $\mathcal{O}(\mathcal{C}_X)$ and has same-size states¹¹, we have that $\mathcal{A}(\mathcal{O}(\mathcal{C}_X))$ and $\mathcal{S}(\mathcal{R}_X)$ are statistically (resp., computationally) indistinguishable.

This definition captures the relative “leakage” of the obfuscated state ensemble: among all functionally-equivalent state ensembles (i.e., potential obfuscations), the best-possible obfuscation is the ensemble that leaks the least. From this point of view, we think of \mathcal{A} as a “learner” which tries to learn something from the obfuscated ensemble, and \mathcal{S} as a “simulator” which can learn the same thing as \mathcal{A} , but from any other functionally-equivalent ensemble.

¹¹ meaning that, for each $x \in X$, the corresponding states in the two ensembles have the same number of qubits.

Proposition 18. *Let $(\mathcal{O}, \mathcal{J})$ be a polynomial-time quantum translator. Then $(\mathcal{O}, \mathcal{J})$ is a quantum statistical (resp., computational) best-possible obfuscator if and only if it is a quantum statistical (resp., computational) indistinguishability obfuscator.*

Proof. First, let $(\mathcal{O}, \mathcal{J})$ be a best-possible obfuscator. Set \mathcal{A} to be the trivial learner which simply implements the identity operator, and let \mathcal{S} be the corresponding simulator. Let \mathcal{C} and \mathcal{D} be two functionally-equivalent circuit ensembles with same-size circuits. Note that their obfuscations are functionally-equivalent (i.e., $\mathcal{O}(\mathcal{C}) \cong_{\mathcal{J}} \mathcal{O}(\mathcal{D})$) same-size ensembles. By two applications of the best-possible property (one on the left, and one on the right), we get

$$\mathcal{O}(\mathcal{C}) = \mathcal{A}(\mathcal{O}(\mathcal{C})) \approx \mathcal{S}(\mathcal{O}(\mathcal{D})) \approx \mathcal{A}(\mathcal{O}(\mathcal{D})) = \mathcal{O}(\mathcal{D}),$$

where \approx denotes the appropriate form of indistinguishability (statistical or computational.) It follows that $(\mathcal{O}, \mathcal{J})$ is an indistinguishability obfuscator.

For the other direction, let $(\mathcal{O}, \mathcal{J})$ be an indistinguishability obfuscator. Given a learner \mathcal{A} , define a simulator \mathcal{S} as follows. Let $(\mathcal{C}, \mathcal{R})$ be a (circuit, ensemble) pair as in the best-possible definition. Since \mathcal{R} is a uniform state ensemble, there is a corresponding circuit ensemble \mathcal{D} for preparing it. Given a circuit $D \in \mathcal{D}$ and the corresponding circuit $J \in \mathcal{J}$ of the interpreter, we can then build a circuit $D \circ J$ which is functionally equivalent to the corresponding circuit $C \in \mathcal{C}$ (by the definition of a quantum translator.) Let $\mathcal{D} \circ \mathcal{J}$ be the corresponding circuit family, and define $\mathcal{S}(\mathcal{R}) = \mathcal{S}(\mathcal{O}(\mathcal{D} \circ \mathcal{J}))$. Now, by the indistinguishability property, $\mathcal{O}(\mathcal{C}) \approx \mathcal{O}(\mathcal{D} \circ \mathcal{J})$, from which it follows that

$$\mathcal{A}(\mathcal{O}(\mathcal{C})) \approx \mathcal{A}(\mathcal{O}(\mathcal{D} \circ \mathcal{J})) = \mathcal{S}(\mathcal{R}),$$

as desired.

We remark that the forward implication did not require the obfuscator to be efficient.

4.3 Impossibility of statistical obfuscation

In this section, we show that efficient perfect or statistical indistinguishability obfuscation is impossible. We begin by recalling the following computational problems about distinguishability of circuit ensembles and state ensembles, and some relevant complexity-theoretic results.

Problem 19. Circuit distinguishability.

Input: two quantum circuits C and D ; parameter $\varepsilon > 0$.

Promise: $\|C - D\|_{\diamond}$ is greater than $2 - \varepsilon$ or less than ε .

Output: YES in the former case and NO in the latter.

Problem 20. Quantum State Distinguishability.

Input: m -qubit quantum circuits C_0 and C_1 , positive integer $k \leq m$ and parameters a, b with $a < b^2$.

Promise: let $\rho_i = \text{Tr}_{(k+1, m)}[C_i | 0^m \rangle \langle 0^m | C_i^\dagger]$; then $\|\rho_0 - \rho_1\|_{tr}$ is greater than b or less than a .

Output: YES in the former case and NO in the latter.

Theorem 21. [RW05] *The problem Circuit distinguishability is QIP-complete for every $\varepsilon > 0$.*

Theorem 22. [Wat02] *The problem Quantum State Distinguishability is QSZK-complete.*

The following theorem is a straightforward matter of assembling the above results together with the definition of indistinguishability obfuscation.

Theorem 23. *If there exists a polynomial-time quantum statistical indistinguishability obfuscator, then PSPACE is contained in QSZK.*

Proof. We will actually show $\text{QIP} \subset \text{BQP}^{\text{QSZK}}$; since BQP is contained in QSZK and $\text{QIP} = \text{PSPACE}$ [JJUW10], the result will follow. We will solve Circuit distinguishability using a subroutine that solves Quantum State Distinguishability. Given the classical descriptions of C and D , we run the obfuscator \mathcal{O} on both to get states $\mathcal{O}(C)$ and $\mathcal{O}(D)$. Now we apply the subroutine, and output its result.

Note that, if C and D are elements of a pair of functionally-equivalent circuit ensembles, then $\mathcal{O}(C)$ and $\mathcal{O}(D)$ will be indistinguishable and the subroutine will output YES. On the other hand, if C and D are far from being functionally equivalent, then there exists some input on which they differ significantly. It follows that $\mathcal{O}(C)$ and $\mathcal{O}(D)$ must be distinguishable; if they were not, then $\mathcal{J}_{\mathcal{O}(C)}$ and $\mathcal{J}_{\mathcal{O}(D)}$ would be functionally equivalent, contradicting the translator conditions in the definition of the obfuscator.

In addition, we can prove an impossibility result for the case of statistical obfuscators which can only obfuscate unitary circuits. We first recall the Identity Check problem. Given an m -qubit state ρ and indices $l, k \geq 0$, we let $\text{Tr}_{l,k}[\rho]$ denote the result of tracing out qubits l through k of ρ . We adopt the convention that nothing is traced out (i.e., $\text{Tr}_{l,k}[\rho] = \rho$) if $l > m$.

Problem 24. Identity Check.

Input: an n -qubit unitary quantum circuit C and parameters a, b so that $b - a \geq 1/\text{poly}(n)$.

Promise: $\min_{\alpha} \|C - e^{i\alpha} I\|$ is less than a or greater than b .

Output: YES in the former case and NO in the latter.

Theorem 25. [JWB05] *The problem Identity Check is coQMA-complete.*

Theorem 26. *If there exists a polynomial-time quantum statistical indistinguishability obfuscator for unitary circuits, then coQMA is contained in QSZK.*

Proof. We will actually show $\text{coQMA} \subset \text{BQP}^{\text{QSZK}}$; since BQP is contained in QSZK, the result will follow. Let a and b satisfy $b - a = 1/\text{poly}(n)$. We will solve Identity Check using a subroutine that solves Quantum State Distinguishability.

Let C be the input, i.e., a classical description of an n -qubit quantum circuit. Create an identity circuit D with an equal number of inputs as C , and of equal length to C . Let O_C be a circuit that initializes a register with the classical state $|C\rangle$ containing the classical description of C , and applies the circuit of \mathcal{O} which corresponds to the input length $|C|$. Likewise, let O_D be a circuit that initializes a register with the classical state $|D\rangle$ containing the classical description of D , and applies the circuit of \mathcal{O} which corresponds to the input length $|D| = |C|$. Note that, after tracing out ancillas, the outputs of these circuits are given by

$$\text{Tr}_{\text{anc.}}[O_C|0\rangle\langle 0|O_C^\dagger] = \mathcal{O}(C) \quad \text{and} \quad \text{Tr}_{\text{anc.}}[O_D|0\rangle\langle 0|O_D^\dagger] = \mathcal{O}(D).$$

Now apply the subroutine for solving quantum state distinguishability to the pair (O_C, O_D) . If it says “close”, we output YES; otherwise we output NO. Let’s show that this has solved (a, b) -identity-check. Note that the states $\mathcal{O}(C)$ and $\mathcal{O}(D)$ must have the same number of qubits, and denote that number by m .

- **completeness.** In this case, the obfuscated states satisfy $\|\mathcal{O}(C) - \mathcal{O}(D)\|_{\text{tr}} \leq \alpha$. By the definition of the induced trace norm, this implies that $\|\mathcal{J}_{\mathcal{O}(C)}^n - \mathcal{J}_{\mathcal{O}(D)}^n\|_{\diamond} \leq \alpha$. By functional equivalence for C and D and the triangle inequality, it follows that $\|U_C - U_D\|_{\diamond} = \|U_C - I\|_{\diamond} \leq \alpha$, as desired.
- **soundness.** In this case, the obfuscated states satisfy $\|\mathcal{O}(C) - \mathcal{O}(D)\|_{\text{tr}} \geq \beta$. We claim that this implies $\|U_C - U_D\|_{\diamond} > b$. Suppose this is not the case, i.e., that these operators are in fact close; then by the indistinguishability property, it would follow that $\mathcal{O}(C)$ and $\mathcal{O}(D)$ are close as well, a contradiction.

The above amounts to a BQP^{QSZK} protocol for a coQMA-hard problem, thus placing coQMA in QSZK.

4.4 Application: quantum witness encryption

We now give an interesting application of the surviving case of quantum indistinguishability obfuscation, i.e., the computational variant.

The classical idea of witness encryption was first studied in [GGSW13]; its connection to indistinguishability obfuscation was first considered in [GGH⁺13]. In the quantum case, we set up the problem as follows. Suppose Alice wishes to encrypt a quantum state ρ , but not to a particular key or for a particular person; instead, the encryption is tied to a challenge question, and anyone that can answer the question correctly can decrypt the plaintext. The question will be of a particularly quantum nature: a correct answer will be a quantum state, e.g., the ground state of some Hamiltonian.

More formally, we consider a QMA language L , and would like to enable Alice to encrypt her state ρ using a particular problem instance x . If x is a “yes” instance, then we’d like to allow Bob, who holds a witness state, to be able to decrypt Alice’s message. On the other hand, if x is a no instance, then we demand that no QPT can distinguish between encryptions of any two quantum states on the same number of qubits. Interestingly, the definition says nothing about the case where x is a yes instance but a witness is not known. While this may seem counterintuitive, the classical primitive has a number of natural applications (e.g., public-key encryption and identity encryption, see [GGSW13]). It is likely that the quantum primitive has the same or similar applications, but we will not explore that question here.

We now show that witness encryption for QMA is possible, assuming the existence of a quantum computational-indistinguishability obfuscator. It is well-known (see, e.g., [GN13]) that QMA contains languages L which have a poly-time uniform circuit family \mathcal{C}_L with completeness $1 - 2^{-\Omega(n)}$ and soundness $2^{-\Omega(n)}$. Given an instance x of such a language L , and a quantum state ρ , consider the quantum circuit $Q_{x,\rho}(\sigma)$, which runs the appropriate circuit from \mathcal{C}_L and outputs ρ on accept, and $|0^n\rangle$ otherwise. We claim the computational-indistinguishability obfuscation $\mathcal{O}(Q_{x,\rho})$ is a valid witness encryption for L and ρ . Correctness of decryption is clear, since the ability to provide a valid witness for L allows Bob to use $\mathcal{O}(Q_{x,\rho})$ to obtain Alice’s state ρ . Of course, if x is not in L , then no witness will suffice; more precisely, $\|Q_{x,\rho_1} - Q_{x,\rho_2}\| \leq 2^{-\Omega(n)}$ for any two quantum states ρ_1, ρ_2 on the same number of qubits. By the indistinguishability condition of Definition 16, the obfuscation of Q_{x,ρ_1} will be computationally indistinguishable from the obfuscation of Q_{x,ρ_2} .

5 Conclusion

5.1 Open problems

The central remaining open problem is whether quantum-mechanical means of obfuscation are possible, within the restricted framework placed by our results. In the classical setting, it is known that black-box obfuscation and statistical obfuscation are impossible; on the other hand, it is generally believed that computational-indistinguishability obfuscation can be achieved. In the quantum setting, on the other hand, much less is known. Although our results place some significant restrictions, several possibilities remain. For example, it is conceivable that *information-theoretically secure black-box obfuscation* of a program is possible using quantum-mechanical means. Moreover, such a construction may even enable the user to use the obfuscated state to evaluate the program a polynomial number of times. Such an obfuscator would be tremendously powerful, and would be another example of quantum supremacy in the world of cryptography. It is difficult to imagine how such an obfuscation would function for arbitrary quantum programs (due to restrictions placed by the monogamy of entanglement); it is perhaps more reasonable to imagine it for classical functions. In that case, one could imagine using the state to perform the computation, then copying the (classical) output of the program, and restoring the obfuscated state by undoing the unitary circuit.

In the indistinguishability setting, constructions for obfuscating quantum programs are not known. It is conceivable that classical ideas (such as those in [GGH⁺13]) can be translated to the quantum setting. However, a key ingredient in [GGH⁺13] is fully-homomorphic encryption, which is yet to be achieved quantumly—although significant progress in this direction was recently made by Broadbent and Jeffery [BJ15].

In terms of applications, our results raise another interesting question: what applications can we envision for obfuscators whose outputs are not cloneable or reproducible? Many of the standard applications are restricted in the case. On the other hand, one may be able to design applications where the inability to copy the obfuscator’s output is a desired feature. In fact, one can imagine quantum obfuscators where the number of executions is highly limited in a manner analogous to one-time programs (see, e.g., [BGS13]); this could also be of use in cryptographic settings.

Bibliography

- [Aar05] Scott Aaronson. Ten semi-grand challenges for quantum computing theory, July 2005. Retrieved 02/16.
- [Aar09] Scott Aaronson. Quantum copy-protection and quantum money. In *Computational Complexity, 2009. CCC'09. 24th Annual IEEE Conference on*, pages 229–242. IEEE, 2009.
- [ABF⁺16] Gorjan Alagic, Anne Broadbent, Bill Fefferman, Tommaso Gagliardoni, Christian Schaffner, and Michael St. Jules. Computational security of quantum encryption. 2016.
- [AC12] Scott Aaronson and Paul Christiano. Quantum money from hidden subspaces. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 41–60. ACM, 2012.
- [AJJ14] Gorjan Alagic, Stacey Jeffery, and Stephen Jordan. Circuit obfuscation using braids. In *Proceedings of TQC 2014*, volume 27, pages 141–160, 2014.
- [BBBV97] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh V. Vazirani. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, 26(5):1510–1523, 1997.
- [BBBW83] Charles H. Bennett, Gilles Brassard, Seth Breidbart, and Stephen Wiesner. Quantum cryptography, or unforgeable subway tokens. In David Chaum, Ronald L. Rivest, and Alan T. Sherman, editors, *Advances in Cryptology*, pages 267–275. Springer US, 1983.
- [BCC⁺14] Nir Bitansky, Ran Canetti, Henry Cohn, Shafi Goldwasser, Yael Tauman Kalai, Omer Paneth, and Alon Rosen. The impossibility of obfuscation with auxiliary input or a universal simulator. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology CRYPTO 2014*, volume 8617 of *Lecture Notes in Computer Science*, pages 71–89. Springer Berlin Heidelberg, 2014.
- [BGI⁺01] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. In *Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology, CRYPTO '01*, pages 1–18, London, UK, UK, 2001. Springer-Verlag.
- [BGI⁺12] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. *J. ACM*, 59(2):6:1–6:48, May 2012.
- [BGK⁺14] Boaz Barak, Sanjam Garg, Yael Tauman Kalai, Omer Paneth, and Amit Sahai. Protecting obfuscation against algebraic attacks. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology EUROCRYPT 2014*, volume 8441 of *Lecture Notes in Computer Science*, pages 221–238. Springer Berlin Heidelberg, 2014.
- [BGS13] Anne Broadbent, Gus Gutoski, and Douglas Stebila. *Advances in Cryptology – CRYPTO 2013: 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18–22, 2013. Proceedings, Part II*, chapter Quantum One-Time Programs, pages 344–360. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
- [BJ15] Anne Broadbent and Stacey Jeffery. *Advances in Cryptology – CRYPTO 2015: 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16–20, 2015. Proceedings, Part II*, chapter Quantum Homomorphic Encryption for Circuits of Low T-gate Complexity, pages 609–629. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.
- [BR14] Zvika Brakerski and Guy N. Rothblum. Virtual black-box obfuscation for all circuits via generic graded encoding. In Yehuda Lindell, editor, *Theory of Cryptography*,

- volume 8349 of *Lecture Notes in Computer Science*, pages 1–25. Springer Berlin Heidelberg, 2014.
- [BZ14] Dan Boneh and Mark Zhandry. Multiparty key exchange, efficient traitor tracing, and more from indistinguishability obfuscation. In JuanA. Garay and Rosario Gennaro, editors, *Advances in Cryptology CRYPTO 2014*, volume 8616 of *Lecture Notes in Computer Science*, pages 480–499. Springer Berlin Heidelberg, 2014.
- [CD08] Ran Canetti and Ronny Ramzi Dakdouk. Obfuscating point functions with multibit output. In Nigel Smart, editor, *Advances in Cryptology EUROCRYPT 2008*, volume 4965 of *Lecture Notes in Computer Science*, pages 489–508. Springer Berlin Heidelberg, 2008.
- [DH76] Whitfield Diffie and Martin E. Hellman. New directions in cryptography. *IEEE transactions on Information Theory*, 22(6):644–654, 1976.
- [FGH⁺12] Edward Farhi, David Gosset, Avinandan Hassidim, Andrew Lutomirski, and Peter Shor. Quantum money from knots. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, ITCS '12*, pages 276–289, New York, NY, USA, 2012. ACM.
- [Gen09] Craig Gentry. *A Fully Homomorphic Encryption Scheme*. PhD thesis, Stanford, CA, USA, 2009. AAI3382729.
- [GGH⁺13] S. Garg, C. Gentry, S. Halevi, M. Raykova, A. Sahai, and B. Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 40–49, Oct 2013.
- [GGHW14] Sanjam Garg, Craig Gentry, Shai Halevi, and Daniel Wichs. On the implausibility of differing-inputs obfuscation and extractable witness encryption with auxiliary input. In JuanA. Garay and Rosario Gennaro, editors, *Advances in Cryptology CRYPTO 2014*, volume 8616 of *Lecture Notes in Computer Science*, pages 518–535. Springer Berlin Heidelberg, 2014.
- [GGSW13] Sanjam Garg, Craig Gentry, Amit Sahai, and Brent Waters. Witness encryption and its applications. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing, STOC '13*, pages 467–476, New York, NY, USA, 2013. ACM.
- [GK05] S. Goldwasser and Y.T. Kalai. On the impossibility of obfuscation with auxiliary input. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 553–562, Oct 2005.
- [GN13] David Gosset and Daniel Nagaj. Quantum 3-sat is qma1-complete. In *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 756–765. IEEE Computer Society, 2013.
- [GR07] Shafi Goldwasser and Guy N. Rothblum. On best-possible obfuscation. In Salil P. Vadhan, editor, *Theory of Cryptography*, volume 4392 of *Lecture Notes in Computer Science*, pages 194–213. Springer Berlin Heidelberg, 2007.
- [HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudo-random generator from any one-way function. *SIAM J. Comput.*, 28:1364–1396, March 1999.
- [HR14] Aram W. Harrow and David J. Rosenbaum. Uselessness for an oracle model with internal randomness. *Quantum Info. Comput.*, 14(7&8):608–624, May 2014.
- [HSW14] Susan Hohenberger, Amit Sahai, and Brent Waters. Replacing a random oracle: Full domain hash from indistinguishability obfuscation. In PhongQ. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology EUROCRYPT 2014*, volume 8441 of *Lecture Notes in Computer Science*, pages 201–220. Springer Berlin Heidelberg, 2014.
- [JJUW10] Rahul Jain, Zhengfeng Ji, Sarvagya Upadhyay, and John Watrous. QIP = PSPACE. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 573–582, 2010.

- [JWB05] Dominik Janzing, Pawel Wocjan, and Thomas Beth. Non-identity check is qma-complete. In *International Journal of Quantum Information*, 2005.
- [Mer78] Ralph C. Merkle. Secure communications over insecure channels. *Commun. ACM*, 21(4):294–299, April 1978.
- [MS10] Michele Mosca and Douglas Stebila. Quantum coins. *Error-Correcting Codes, Finite Geometries and Cryptography*, 523:35–47, 2010.
- [RW05] B. Rosgen and J. Watrous. On the hardness of distinguishing mixed-state quantum computations. In *Computational Complexity, 2005. Proceedings. Twentieth Annual IEEE Conference on*, pages 344–354, June 2005.
- [SW14] Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: Deniable encryption, and more. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 475–484, New York, NY, USA, 2014. ACM.
- [Wat02] John Watrous. Limits on the power of quantum statistical zero-knowledge. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*, page 459. IEEE Computer Society, 2002.
- [Wee05] Hoeteck Wee. On obfuscating point functions. In *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing*, STOC '05, pages 523–532, New York, NY, USA, 2005. ACM.
- [Wie83] Stephen Wiesner. Conjugate coding. *ACM Sigact News*, 15(1):78–88, 1983.
- [Zha12] Mark Zhandry. How to Construct Quantum Random Functions. In *FOCS 2012*, pages 679–687. IEEE, 2012.