Improving the security of device-independent weak coin flipping protocols

Extended abstract

Atul Singh Arora*, Jamie Sikora†, Thomas Van Himbeeck‡

Coin-flipping is the two-party cryptographic primitive where two parties, henceforth called Alice and Bob, wish to flip a coin, but where, to make things interesting, they do not trust each other. This primitive was introduced by Blum [Blu83] who also introduced the first (classical) protocol. In this work, we concentrate on *weak* coin flipping (WCF) protocols where Alice and Bob desire opposite outcomes. Since then, a series of quantum protocols were introduced with successively improved security. Mochon, in his tour de force, finally settled the question about the limits of the security in the quantum regime by proving the *existence* of quantum protocols with security approaching the ideal limit [Moc07]. This was followed by a flurry of results which achieved diverse cryptographic functionality assuming WCF as a black-box, such as strong coin flipping [CK09], bit commitment [CK11], a variant of oblivious transfer [CGS13], leader election [Gan09] and dice rolling [AS], establishing the importance of WCF in the quantum setting. Returning to Mochon, his work was quite technical and based on the notion of *point games*, a concept introduced by Kitaev. Interestingly, his work was never published—only a preprint was available. Subsequently, a sequence of works have continued the study of point games. In particular, the proof of existence was eventually simplified and peer reviewed [ACG+14] and explicit protocols were reported after more than a decade of Mocohn's work [ARW19, ARV].[1] Yet, we note that all of this work is in the *device-dependent* setting where *Alice and Bob trust their quantum devices*. Very little is known in the *device-independent (DI)* setting where a cheating player is allowed to control an honest player's quantum devices, opening up a plethora of new cheating strategies that were not considered in the previously mentioned references.

We introduce some basic concepts to facilitate further discussion. The prefix *weak* in weak coin flipping refers to the situation where Alice and Bob desire opposite outcomes of the coin.[2] When designing weak coin flipping protocols, we desire correctness for honest parties—if both parties are honest, then they agree on the same outcome c—and soundness against one dishonest party—if Alice (Bob) is honest then a cheating Bob (Alice) cannot force the outcome $c = 1$ ($c = 0$). Typically, perfect correctness holds by construction and the goal is to ensure soundness. The latter may be measured in terms of *cheating probabilities* (denoted p_A^* and p_B^*) and *bias* (denoted ϵ)—the maximum probability with which a cheating Alice can force an honest Bob to accept the outcome $c = 0$ ($c = 1$) is defined to be p_A^* (p_B^*) and $\epsilon = \max\{p_A^*, p_B^*\} - 1/2$ captures how much a cheating player can bias the outcome away from uniform.

As hinted at above, we consider these definitions in the *device-independent* setting. To be unambiguous, we emphasise that we study *information theoretic security*—Alice and Bob are only bounded by the laws of quantum mechanics (and not required to be efficient, for instance).

When studying DI protocols, one should first consider whether or not secure classical protocols are known (since these are not affected by the DI assumption). It was proved that every classical WCF protocol

*IQIM, Department of Computing and Mathematical Sciences, California Institute of Technology, USA

†Virginia Polytechnic Institute and State University, USA

‡University of Toronto, Canada; Institute of Quantum Computing, University of Waterloo, Canada

[1] Interestingly, Miller [Mil20] used techniques from Mochon's proof to show that protocols approaching the ideal limit must have an exponentially increasing number of messages. It is an open question to find how small the bias can be made before any such protocol becomes impractical.

[2] To emphasise the distinction, coin-flipping is termed *strong* coin flipping as Alice or Bob could try to bias the coin towards either outcome.

has bias $\epsilon = 1/2$, which is the worst possible value (see [Kit03, HW11]). Thus, it makes sense to study quantum WCF protocols in the DI setting, especially if one with bias $\epsilon < 1/2$ can be found. Remarkably, Silman, Chailloux, Aharon, Kerenidis, Pironio, and Massar presented a protocol [SCA+11], call it Protocol S, ten years ago with $p_A^* = \cos^2(\pi/8) \approx 0.853$ and $p_B^* = 3/4$. We briefly discuss this protocol not only because it is still the state of the art, but also because we build on this result. Their protocol begins with Alice possessing two *boxes*—physical devices that accept classical inputs and yield classical outputs—and Bob possessing one box which are together supposed to contain the GHZ state and measurements.[3] As the protocol proceeds, they, in addition to exchanging classical information, operate these boxes and exchange them.[4] As is, Protocol S has bias $\epsilon \approx 0.353$ but in [SCA+11], Protocol S is composed many times to lower the bias to $\epsilon \leq 0.33664$.

Contribution. In this work, we provide two techniques for lowering the bias of DI WCF and apply them to Protocol S to obtain the first improvement since the work of SCAKPM [SCA+11].

Theorem 1 (informal)**.** *There exist device-independent weak coin flipping protocols with bias, ϵ, approaching* 0.31486*. Assuming a continuity conjecture (see the manuscript), the bias can be lowered to $\epsilon \approx 0.29104$.*

We now discuss the key ideas that go into the proof of our main theorem, above. Protocol S was, in fact, a strong coin flipping protocol and we begin by turning it into a weak coin flipping protocol—Protocol W—in a routine manner. Intuitively, since weak coin flipping has the notion of a "winner" (if $c = 0$ Alice wins and if $c = 1$ Bob wins) we have the party who does not win, conduct an additional test.

Our first technique is to add a pre-processing step to Protocol W which *self-tests* the boxes shared by Alice and Bob at the start of the protocol. Our second technique is to compose and analyse the resulting protocols in a new way,[5] which we call *abort-phobic* composition.

First technique: Self-testing. Since we are in the DI setting, in Protocol S and it's WCF variant, Protocol W, a cheating party may control what measurement is performed in the boxes of the other party and how the state of the boxes is correlated to its own quantum memory. However, we employ the concept of self-testing to prevent Bob (or Alice) from applying such a strategy. Intuitively, self testing is a powerful property which allows one to, just from certain input-output behaviours of given devices (satisfying minimal assumptions), conclude uniquely which quantum states and measurements constitute the devices (up to relabelling). The GHZ state which was used in Protocols S and W can be self-tested. Clearly, this property has the potential to improve their security.[6]

We define two variants of Protocol W: Protocol P, where Alice self-tests Bob before executing Protocol W, and Protocol Q, where Bob self-tests Alice instead. The procedure is simple: Alice and Bob start with n triples of boxes and, for instance when Alice self-tests, Alice asks Bob to send all but one randomly selected triple and tests if the GHZ test passes for these. If so, the remaining triple is used for the actual protocol. If n is chosen large enough, a dishonest Bob is unable to significantly tamper with the boxes. Indeed, this step already allows us to reduce the cheating probabilities.

Lemma 2 (Informal)**.** *For Protocol P, i.e. where Alice self tests Bob, the cheating probabilities, in the limit of large n, are*

$$p_A^* = \cos^2(\pi/8) \approx 0.85355 \quad \text{and} \quad p_B^* \approx 0.6667. \tag{1}$$

[3] A GHZ state is a widely used tripartite entangled quantum state with interesting non-local properties.

[4] Any protocol described using boxes is readily converted into one where Alice and Bob communicate over an insecure quantum channel.

[5] The composition in [SCA+11] may also be seen as "abort-phobic" but their analysis doesn't rely on the "abort" probability; their bound essentially neglects the abort event.

[6] In [SCA+11], it was noted that self-testing doesn't help improve the security of Protocol S. Alternatively stated, Protocol S has the curious property that its device dependent variant has the same security as it (the device dependent variant).

For comparison, recall that for Protocol S (it turns out, also for Protocol W), $p_A^* = \cos^2(\pi/8)$ and $p_B^* = 3/4$. We prove this lemma in two stages. In the *first* stage, we assume perfect self-testing—the self-testing step results in exactly specifying (up to a relabelling) the state and measurements governing Alice's boxes. It is known that for device-dependent protocols, where Alice and Bob trust their devices, the cheating probabilities can be cast as values of semidefinite programs (SDPs) [Kit03, Moc07]. Thus one can express Bob's cheating probabilities as an SDP. Its numerical evaluation yields the quoted value. Analysis for Alice's cheating probability is unchanged from Protocol W. In the *second* stage, we take n to be finite and first show how self-testing results apply in this cryptographic setting—where based on $n-1$ tests, we conclude the properties of the as-yet untouched triple of boxes—and then use this characterisation to write the cheating probability for Bob as an optimisation problem which converges to the SDP above for large n. These techniques may be of independent interest. For Protocol Q we establish an analogous result under a continuity conjecture.

Second technique: abort-phobic composition. For a protocol with cheating probabilities p_B^* and p_A^*, we say that it has *polarity* towards Alice (Bob) when it satisfies $p_A^* > p_B^*$ ($p_B^* > p_A^*$). Given a polarised protocol $\mathcal{R}$, we may switch the roles of Alice and Bob to flip the polarity—explicitly, we write $\mathcal{R}_A$ to be the version of the protocol with $p_A^* > p_B^*$ and $\mathcal{R}_B$ to be the version with $p_B^* > p_A^*$. The *standard composition* then, is simply that Alice and Bob perform $\mathcal{R}$ first. If Alice (Bob) wins, they perform $\mathcal{R}_A$ ($\mathcal{R}_B$) to determine the outcome c. Intuitively, the winner of the first round gets to run the protocol in a polarity which favours them. If $p_A^* > p_B^*$, then for the composed protocol, Alice's cheating is

$$(p_A^*)^2 + (1-p_A^*)p_B^* < p_A^* \tag{2}$$

while Bob's cheating probability is

$$p_B^* p_A^* + (1-p_B^*)p_B^* < p_A^*. \tag{3}$$

This does indeed reduce the bias since the maximum cheating probability is now smaller.

Abort-phobic composition. WCF protocols are usually viewed as having two outcomes indicating which player wins as a player that gets detected cheating immediately loses.[7] For "one-shot" applications, this makes sense. However, even in the standard composition, it is evident that Bob (Alice) should not really accept to continue onto the second protocol if he catches Alice (Bob) cheating in the first round. Denote this event by $c = \perp$, an *abort*. We therefore define *abort-phobic compositions* as before, where Alice and Bob use a WCF protocol to determine who gets to choose the polarity in the subsequent round, except that if either party aborts (detects maleficence), the composite protocol is aborted. As an illustration, suppose Bob adopts a cheating strategy which has a probability v_B of him winning ($c = 1$), a probability v_A of him losing ($c = 0$), and a probability $v_\perp$ of Alice catching him cheating. Then his cheating probability for abort-phobic composition is now

$$v_B \cdot p_A^* + v_A \cdot p_B^* + v_\perp \cdot 0. \tag{4}$$

This quantity may be a strict improvement if $v_\perp > 0$ when $v_B = p_B^*$. Optimising over *cheat vectors*, $v_B, v_A, v_\perp$ in the DI setting in general is difficult but the self-testing step allows for an SDP characterisation for Alice (Bob) in Protocol P (Q). Using techniques reminiscent of dynamic programming, one can apply the analysis even if abort-phobic compositions are repeated many times.

When applied to Protocol P (where Alice self-tests), this yields protocols which converge onto a bias of $\epsilon \approx 0.31486$ proving the first part of the main result. For the second part, we use Protocol P in the final round, and Protocol Q (where Bob self-tests) until then, to obtain protocols whose bias approaches $\epsilon \approx 0.29104$ (assuming the aforementioned continuity conjecture for Protocol Q).

[7] We implicitly assume perfect correctness in doing this.

References

[ACG+14] Dorit Aharonov, André Chailloux, Maor Ganz, Iordanis Kerenidis, and Loïck Magnin, *A simpler proof of existence of quantum weak coin flipping with arbitrarily small bias*, SIAM Journal on Computing **45** (2014), no. 3, 633–679.

[ARV] Atul Singh Arora, Jérémie Roland, and Chrysoula Vlachou, *Analytic quantum weak coin flipping protocols with arbitrarily small bias*, pp. 919–938.

[ARW19] Atul Singh Arora, Jérémie Roland, and Stephan Weis, *Quantum weak coin flipping*, Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing - STOC 2019, ACM Press, 2019.

[AS] N. Aharon and J. Silman, *Quantum dice rolling: A multi-outcome generalization of quantum coin flipping*.

[Blu83] Manuel Blum, *Coin flipping by telephone a protocol for solving impossible problems*, SIGACT News **15** (1983), no. 1, 23–27.

[CGS13] André Chailloux, Gus Gutoski, and Jamie Sikora, *Optimal bounds for semi-honest quantum oblivious transfer*.

[CK09] André Chailloux and Iordanis Kerenidis, *Optimal Quantum Strong Coin Flipping*, 50th FOCS, 2009, pp. 527–533.

[CK11] ———, *Optimal Bounds for Quantum Bit Commitment*, 52nd FOCS, 2011, pp. 354–362.

[Gan09] Maor Ganz, *Quantum Leader Election*.

[HW11] Esther Hänggi and Jürg Wullschleger, *Tight bounds for classical and quantum coin flipping*, Theory of Cryptography (Berlin, Heidelberg) (Yuval Ishai, ed.), Springer Berlin Heidelberg, 2011, pp. 468–485.

[Kit03] Alexei Kitaev, *Quantum coin flipping*, Talk at the 6th workshop on Quantum Information Processing, 2003.

[Mil20] Carl A. Miller, *The impossibility of efficient quantum weak coin flipping*, Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (New York, NY, USA), STOC 2020, Association for Computing Machinery, 2020, pp. 916–929.

[Moc07] Carlos Mochon, *Quantum weak coin flipping with arbitrarily small bias*, arXiv:0711.4114 (2007).

[SCA+11] J. Silman, A. Chailloux, N. Aharon, I. Kerenidis, S. Pironio, and S. Massar, *Fully distrustful quantum bit commitment and coin flipping*, Physical Review Letters **106** (2011), no. 22.