



# On the Complexity of Fair Coin Flipping

Iftach Haitner<sup>1</sup>(✉), Nikolaos Makriyannis<sup>1</sup>, and Eran Omri<sup>2</sup>

<sup>1</sup> School of Computer Science, Tel Aviv University, Tel Aviv, Israel  
iftachh@cs.tau.ac.il, n.makriyannis@gmail.com

<sup>2</sup> Department of Computer Science, Ariel University, Ariel, Israel  
omrier@ariel.ac.il

**Abstract.** A two-party coin-flipping protocol is  $\varepsilon$ -fair if no efficient adversary can bias the output of the honest party (who always outputs a bit, even if the other party aborts) by more than  $\varepsilon$ . Cleve [STOC '86] showed that  $r$ -round  $o(1/r)$ -fair coin-flipping protocols do not exist. Awerbuch et al. [Manuscript '85] constructed a  $\Theta(1/\sqrt{r})$ -fair coin-flipping protocol, assuming the existence of one-way functions. Moran et al. [Journal of Cryptology '16] constructed an  $r$ -round coin-flipping protocol that is  $\Theta(1/r)$ -fair (thus matching the aforementioned lower bound of Cleve [STOC '86]), assuming the existence of oblivious transfer.

The above gives rise to the intriguing question of whether oblivious transfer, or more generally “public-key primitives”, is required for an  $o(1/\sqrt{r})$ -fair coin flipping. This question was partially answered by Dachman-Soled et al. [TCC '11] and Dachman-Soled et al. [TCC '14], who showed that *restricted* types of fully black-box reductions cannot establish  $o(1/\sqrt{r})$ -fair coin-flipping protocols from one-way functions. In particular, for constant-round coin-flipping protocols, [10] yields that black-box techniques from one-way functions can only guarantee fairness of order  $1/\sqrt{r}$ .

We make progress towards answering the above question by showing that, for any constant  $r \in \mathbb{N}$ , the existence of an  $1/(c \cdot \sqrt{r})$ -fair,  $r$ -round coin-flipping protocol implies the existence of an infinitely-often key-agreement protocol, where  $c$  denotes some universal constant (independent of  $r$ ). Our reduction is non black-box and makes a novel use of the recent dichotomy for two-party protocols of Haitner et al. [FOCS '18] to facilitate a two-party variant of the attack of Beimel et al. [FOCS '18] on multi-party coin-flipping protocols.

**Keywords:** Coin-flipping · Fairness · Key-agreement

## 1 Introduction

In a two-party coin flipping protocol, introduced by Blum [6], the parties wish to output a common (close to) uniform bit, even though one of the parties

---

I. Haitner and N. Makriyannis—Research supported by ERC starting grant 638121.

I. Haitner and N. Makriyannis—Research supported by the Check Point Institute for Information Security.

E. Omri—Research supported by ISF grant 152/17.

© International Association for Cryptologic Research 2018

A. Beimel and S. Dziembowski (Eds.): TCC 2018, LNCS 11239, pp. 539–562, 2018.

[https://doi.org/10.1007/978-3-030-03807-6\\_20](https://doi.org/10.1007/978-3-030-03807-6_20)

may be corrupted and try to bias the output. Slightly more formally, an  $\varepsilon$ -fair coin flipping protocol should satisfy the following two properties: first, when both parties behave honestly (i.e., follow the prescribed protocol), they both output the *same* uniform bit. Second, in the presence of a corrupted party that may deviate from the protocol arbitrarily, the distribution of the honest party's output may deviate from the uniform distribution (unbiased bit) by at most  $\varepsilon$ . We emphasize that the above notion requires an honest party to *always* output a bit, regardless of what the corrupted party does, and, in particular, it is not allowed to abort if a cheat is detected.<sup>1</sup> Coin flipping is a fundamental primitive with numerous applications, and thus lower bounds on coin flipping protocols yield analogous bounds for many basic cryptographic primitives, including other inputless primitives and secure computation of functions that take input (e.g., XOR).

In his seminal work, Cleve [8] showed that, for *any* efficient two-party  $r$ -round coin flipping protocol, there exists an efficient adversarial strategy that biases the output of the honest party by  $\Theta(1/r)$ . The above lower bound on coin flipping protocols was met for the two-party case by Moran, Naor, and Segev [20] improving over the  $\Theta(n/\sqrt{r})$ -fairness achieved by the majority protocol of Awerbuch, Blum, Chor, Goldwasser, and Micali [2]. The protocol of [20], however, uses oblivious transfer; to be compared with the protocol of [2] that can be based on any one-way function. An intriguing open question is whether oblivious transfer, or more generally “public-key primitives”, is required for an  $o(1/\sqrt{r})$ -fair coin flip. The question was partially answered in the black-box setting by Dachman-Soled, Lindell, Mahmoody, and Malkin [10] and Dachman-Soled, Mahmoody, and Malkin [11], who showed that *restricted* types of fully black-box reductions cannot establish  $o(1/\sqrt{r})$ -bias coin flipping protocols from one-way functions. In particular, for constant-round coin flipping protocols, [10] yields that black-box techniques from one-way functions can only guarantee fairness of order  $1/\sqrt{r}$ .

## 1.1 Our Results

Our main result is that constant-round coin flipping protocols with better bias compared to the majority protocol of [2] imply the existence of infinitely-often key-agreement. We recall that infinitely-often key-agreement protocols satisfy correctness (parties agree on a common bit with overwhelming probability), and, for an infinite number of security parameters, no efficient eavesdropper can deduce the output with probability noticeably far from a random guess.<sup>2</sup>

**Theorem 1.1 (Main result, informal).** *For any (constant)  $r \in \mathbb{N}$ , the existence of an  $1/(c \cdot \sqrt{r})$ -fair,  $r$ -round coin flipping protocol implies the existence*

<sup>1</sup> Such protocols are typically addressed as having *guaranteed output delivery*, or, abusing terminology, as *fair*.

<sup>2</sup> While infinitely-often key-agreement protocols are useless from a cryptographic point of view, constructing such protocols appears to be as hard as obtaining full-blown key agreement.

an infinitely-often key-agreement protocol, for  $c > 0$  being a universal constant (independent of  $r$ ).

As in [8,10,11], our result extends via a simple reduction to general multi-party coin flipping protocols (with more than two-parties) without an honest majority. Our non black-box reduction makes a novel use of the recent dichotomy for two-party protocols of Haitner et al. [12]. Specifically, assuming that io-key-agreement does not exist and applying Haitner et al.’s dichotomy, we show that a two-party variant of the recent multi-party attack of Beimel et al. [3] yields a  $1/(c \cdot \sqrt{r})$ -bias attack.

### 1.2 Our Technique

Let  $\Pi = (A, B)$  be a  $r$ -round two-party coin flipping protocol. We show that the nonexistence of key-agreement protocols yields an efficient  $\Theta(1/\sqrt{r})$ -bias attack on  $\Pi$ . We start by describing the  $1/\sqrt{r}$ -bias *inefficient* attack of Cleve and Impagliazzo [9], and the approach of Beimel et al. [3] towards making this attack efficient. We then explain how to use the recent results by Haitner et al. [12] to obtain an efficient attack (assuming the nonexistence of io-key-agreement protocols).

**Cleve and Impagliazzo’s Inefficient Attack.** We describe the inefficient  $1/\sqrt{r}$ -bias attack due to Cleve and Impagliazzo [9]. Let  $M_1, \dots, M_r$  denote the messages in a random execution of  $\Pi$ , and let  $C$  denote the (without loss of generality) always common output of the parties in a random honest execution of  $\Pi$ . Let  $X_i = \mathbf{E}[C \mid M_{\leq i}]$ . Namely,  $M_{\leq i} = M_1, \dots, M_i$  denotes the partial transcript of  $\Pi$  up to and including round  $i$ , and  $X_i$  is the expected outcome of the parties in  $\Pi$  given  $M_{\leq i}$ . It is easy to see that  $X_0, \dots, X_r$  is a martingale sequence:  $\mathbf{E}[X_i \mid X_0, \dots, X_{i-1}] = X_{i-1}$  for every  $i$ . Since the parties in an honest execution of  $\Pi$  output a uniform bit, it holds that  $X_0 = \Pr[C = 1] = 1/2$  and  $X_r \in \{0, 1\}$ . Cleve and Impagliazzo [9] (see Beimel et al. [3] for an alternative simpler proof) prove that, for such a sequence (omitting absolute values and constant factors),

$$\text{Gap:} \quad \Pr[\exists i \in [r]: X_i - X_{i-1} \geq 1/\sqrt{r}] \geq 1/2. \tag{1}$$

Let the  $i^{\text{th}}$  *backup value* of party  $P$ , denoted  $Z_i^P$ , be the output of party  $P$  if the other party aborts prematurely *after* the  $i^{\text{th}}$  message was sent (recall that the honest party must always output a bit, by definition). In particular,  $Z_r^P$  denotes the final output of  $P$  (if no abort occurred). We claim that without loss of generality for both  $P \in \{A, B\}$  it holds that

Backup values approximate outcome:

$$\Pr[\exists i \in [r]: |X_i - \mathbf{E}[Z_i^P \mid M_{\leq i}]| \geq 1/2\sqrt{r}] \leq 1/4. \tag{2}$$

To see why, assume Eq.(2) does not hold. Then, the (possibly inefficient) adversary controlling  $\bar{P} \in \{A, B\} \setminus P$  that aborts at the end of round  $i$  if

$(-1)^{1-z} \cdot (X_i - \mathbf{E}[Z_i^P \mid M_{\leq i}]) \geq 1/\sqrt{r}$ , for suitable  $z \in \{0, 1\}$ , biases the output of  $\mathbf{P}$  towards  $1 - z$  by  $\Theta(1/\sqrt{r})$ .

Finally, since the coins of the parties are *independent* conditioned on the transcript (a fundamental fact about protocols), if party  $\mathbf{A}$  sends the  $(i + 1)$  message then

$$\text{Independence:} \quad \mathbf{E}[Z_i^B \mid M_{\leq i}] = \mathbf{E}[Z_i^B \mid M_{\leq i+1}]. \tag{3}$$

Combining the above observations yields that without loss of generality:

$$\Pr[\exists i \in [r]: \mathbf{A} \text{ sends the } i^{\text{th}} \text{ message} \wedge X_i - \mathbf{E}[Z_{i-1}^B \mid M_{\leq i}] \geq 1/2\sqrt{r}] \geq 1/8. \tag{4}$$

Equation (4) yields the following (possibly inefficient) attack for a corrupted party  $\mathbf{A}$  biasing  $\mathbf{B}$ 's output towards zero: before sending the  $i^{\text{th}}$  message  $M_i$ , party  $\mathbf{A}$  aborts if  $X_i - \mathbf{E}[Z_{i-1}^B \mid M_{\leq i}] \geq 1/2\sqrt{r}$ . By Eq. (4), this attack biases  $\mathbf{B}$ 's output towards zero by  $\Omega(1/2\sqrt{r})$ .

The clear limitation of the above attack is that, assuming one-way functions exist, the value of  $X_i = \mathbf{E}[C \mid M_{\leq i} = (m_1, \dots, m_i)]$  and the value of  $\mathbf{E}[Z_i^P \mid M_{\leq i} = (m_1, \dots, m_i)]$  might *not* be efficiently computable as a function of  $t$ .<sup>3</sup> Facing this difficulty, Beimel et al. [3] considered the martingale sequence  $X_i = \mathbf{E}[C \mid Z_{\leq i}^P]$  (recall that  $Z_i^P$  is the  $i^{\text{th}}$  backup value of  $\mathbf{P}$ ). It follows that, for constant-round protocols, the value of  $X_i$  is only a function of a constant size string, and thus it is efficiently computable ([3] have facilitated this approach for protocols of super-constant round complexity, see Footnote 4). The price of using the alternative sequence  $X_1, \dots, X_r$  is that the independence property (Eq. (3)) might no longer hold. Yet, [3] manage to facilitate the above approach into an efficient  $\tilde{\Omega}(1/\sqrt{r})$ -attack on *multi-party* protocols. In the following, we show how to use the dichotomy of Haitner et al. [12] to facilitate a two-party variant of the attack from [3].

**Nonexistence of Key-Agreement Implies an Efficient Attack.** Let  $U_p$  denote the Bernoulli random variable taking the value 1 with probability  $p$ , and let  $P \stackrel{c}{\approx}_\rho Q$  stand for  $Q$  and  $P$  are  $\rho$ -computationally indistinguishability (i.e., an efficient distinguisher cannot tell  $P$  from  $Q$  with advantage better than  $\rho$ ). We are using two results by Haitner et al. [12]. The first one given below holds for any two-party protocol.

**Theorem 1.2 (Haitner et al. [12]'s forecaster, informal).** *Let  $\Delta = (\mathbf{A}, \mathbf{B})$  be a single-bit output (each party outputs a bit) two-party protocol. Then, for any constant  $\rho > 0$ , there exists a constant output-length poly-time algorithm (forecaster)  $F$  mapping transcripts of  $\Delta$  into (the binary description of) pairs in  $[0, 1] \times [0, 1]$  such that the following holds: let  $(X, Y, T)$  be the parties outputs and transcript in a random execution of  $\Delta$ , then*

<sup>3</sup> For instance, the first two messages might contain commitments to the parties' randomness.

- $(X, T) \stackrel{c}{\approx}_\rho (U_{p^A}, T)_{(p^A, \cdot) \leftarrow F(T)}$ , and
- $(Y, T) \stackrel{c}{\approx}_\rho (U_{p^B}, T)_{(\cdot, p^B) \leftarrow F(T)}$ .

Namely, given the transcript,  $F$  forecasts the output-distribution for each party in a way that is computationally indistinguishable from (the distribution of) the real output.

Consider the  $(r + 1)$ -round protocol  $\tilde{\Pi} = (\tilde{A}, \tilde{B})$ , defined by  $\tilde{A}$  sending a random  $i \in [r]$  to  $\tilde{B}$  as the first message and then the parties interact in a random execution of  $\Pi$  for the first  $i$  rounds. At the end of the execution, the parties output their  $i^{\text{th}}$  backup values  $z_i^A$  and  $z_i^B$  and halt. Let  $F$  be the forecaster for  $\tilde{\Pi}$  guaranteed by Theorem 1.2 for  $\rho = 1/r^2$  (note that  $\rho$  is indeed constant). A simple averaging argument yields that

$$(Z_i^P, M_{\leq i}) \stackrel{c}{\approx}_{1/r} (U_{p^P}, M_{\leq i})_{(p^A, p^B) \leftarrow F(M_{\leq i})} \quad (5)$$

for both  $P \in \{A, B\}$  and every  $i \in [r]$ , letting  $F(m_{\leq i}) = F(i, m_{\leq i})$ . Namely,  $F$  is a good forecaster for the partial transcripts of  $\Pi$ .

Let  $M_1, \dots, M_r$  denote the messages in a random execution of  $\Pi$  and let  $C$  denote the output of the parties in  $\Pi$ . Let  $F_i = (F_i^A, F_i^B) = F(M_{\leq i})$  and let  $X_i = \mathbf{E}[C \mid F_{\leq i}]$ . It is easy to see that  $X_1, \dots, X_r$  is a martingale sequence and that  $X_0 = 1/2$ . We assume without loss of generality that the last message of  $\Pi$  contains the common output. Thus, it follows from Eq. (5) that  $F_r \approx (C, C) \in \{(0, 0), (1, 1)\}$  (otherwise, it will be very easy to distinguish the forecasted outputs from the real ones, given  $M_r$ ). Hence, similarly to Sect. 1.2, it holds that

$$\text{Gap:} \quad \Pr[\exists i \in [r]: X_i - X_{i-1} \geq 1/\sqrt{r}] \geq 1/2. \quad (6)$$

Since  $F_i$  has constant-size support and since  $\Pi$  is constant round, it follows that  $X_i$  is efficiently computable from  $M_{\leq i}$ .<sup>4</sup>

Let  $Z_i^P$  denote the backup value computed by party  $P$  in round  $i$  of a random execution of  $\Pi$ . The indistinguishability of  $F$  yields that  $\mathbf{E}[Z_i^P \mid F_{\leq i}] \approx F_i^P$ . Similarly to Sect. 1.2, unless there is a simple  $1/\sqrt{r}$ -attack, it holds that

Backup values approximate outcome:

$$\Pr[\exists i \in [r]: |X_i - \mathbf{E}[Z_i^P \mid F_{\leq i}]| \geq 1/2\sqrt{r}] \leq 1/4. \quad (7)$$

Thus, for an efficient variant of [9]'s attack, it suffices to show that

$$\text{Independence:} \quad \mathbf{E}[Z_i^P \mid F_{\leq i}] \approx \mathbf{E}[Z_i^P \mid F_{\leq i+1}]. \quad (8)$$

<sup>4</sup> In the spirit of Beimel et al. [3], we could have modified the definition of the  $X_i$ 's to make them efficiently computable even for non constant-round protocols. The idea is to define  $X_i = \mathbf{E}[C \mid F_i, X_{i-1}]$ . While the resulting sequence might not be a martingale, [3] proves that a  $1/\sqrt{r}$ -gap also occurs with constant probability for such a sequence. Unfortunately, we cannot benefit from this improvement, since the results of Haitner et al. [12] only guarantees indistinguishability for constant  $\rho$ , which makes it useful only for attacking constant-round protocols.

for every  $P \in \{A, B\}$  and round  $i$  in which party  $\bar{P} \in \{A, B\} \setminus \{P\}$  sends the  $(i+1)$  message. However, unlike Eq. (3) in Sect. 1.2, Eq. (8) does not hold unconditionally (in fact, assuming oblivious transfer exists, the implied attack must fail for some protocols, yielding that Eq. (8) is false for these protocols). Rather, we relate Eq. (8) to the existence of a key-agreement protocol. Specifically, we show that if Eq. (8) is not true, then there exists a key-agreement protocol.

**Proving that  $F_{i+1}$  and  $Z_i^P$  are Approximately Independent Given  $F_{\leq i}$ .** The next (and last) argument is the most technically challenging part of our proof. At this time, we provide a brief yet meaningful overview of the technique. The full details are provided in the main body (Claim 3.8 in Sect. 3).

We show that assuming nonexistence of io-key-agreement,  $F_{i+1}$  and  $Z_i^P$  are approximately independent given  $F_{\leq i}$ . In more detail, the triple  $(Z_i^P, F_{i+1}, F_{\leq i})$  is  $\rho$ -indistinguishable from  $(Y_1, Y_2, F_{\leq i})$  where  $(Y_1, Y_2)$  is a pair of random variables that are mutually independent given  $F_{\leq i}$ . It would then follow that  $\mathbf{E}[Z_i^P \mid F_{i+1}, F_{\leq i}] \approx \mathbf{E}[Y_1 \mid Y_2, F_{\leq i}] = \mathbf{E}[Y_1 \mid F_{\leq i}] \approx \mathbf{E}[Z_i^P \mid F_{\leq i}]$  as required. To this end, we use a second result by Haitner et al. [12].<sup>5</sup>

**Theorem 1.3 (Haitner et al. [12]’s dichotomy, informal).** Let  $\Delta = (A, B)$  be an efficient single-bit output two-party protocol and assume infinitely-often key-agreement protocol does not exist. Then, for any constant  $\rho > 0$ , there exists a poly-time algorithm (*decorrelator*)  $\text{Dcr}$  mapping transcripts of  $\Delta$  into  $[0, 1] \times [0, 1]$  such that the following holds: let  $(X, Y, T)$  be the parties’ outputs and transcript in a random execution of  $\Delta$ , then

$$(X, Y, T) \stackrel{c}{\approx}_\rho (U_{p^A}, U_{p^B}, T)_{(p^A, p^B) \leftarrow \text{Dcr}(T)}.$$

Namely, assuming io-key-agreement does not exist, the distribution of the parties’ output given the transcript is  $\rho$ -close to the product distribution given by  $\text{Dcr}$ . We assume for simplicity that the theorem holds for *many-bit* output protocols and not merely single bit (we get rid of this assumption in the actual proof).

We define another variant  $\widehat{\Pi}$  of  $\Pi$  that internally uses the forecaster  $F$ , and show that the existence of a decorrelator for  $\widehat{\Pi}$  implies that  $F_{i+1}$  and  $Z_i^P$  are approximately independent given  $F_{\leq i}$ , and Eq. (8) follows. For concreteness, we focus on party  $P = B$ .

Fix  $i$  such that  $A$  sends the  $(i + 1)$  message in  $\Pi$  and define protocol  $\widehat{\Pi} = (\widehat{A}, \widehat{B})$  according to the following specifications: the parties interact just as in  $\Pi$  for the first  $i$  rounds; then  $\widehat{B}$  outputs the  $i^{\text{th}}$  backup value of  $B$  and  $\widehat{A}$  internally computes  $m_{i+1}$  and outputs  $f_{i+1} = F(m_{\leq i+1})$ . By Theorem 1.3 there exists an efficient decorrelator  $\text{Dcr}$  for  $\widehat{\Pi}$  with respect to  $\rho = 1/r$ . That is:

$$(F_{i+1}, Z_i^B, M_{\leq i}) \stackrel{c}{\approx}_{1/r} (U_{p^{\widehat{A}}}, U_{p^{\widehat{B}}}, M_{\leq i})_{(p^{\widehat{A}}, p^{\widehat{B}}) \leftarrow \text{Dcr}(M_{\leq i})}, \tag{9}$$

<sup>5</sup> Assuming the nonexistence of key-agreement protocols, Theorem 1.3 implies Theorem 1.2. Yet, we chose to use both results to make the text more modular.

where now  $p^{\hat{A}}$  describes a non-Boolean distribution, and  $U_{p^{\hat{A}}}$  denotes an independent sample from this distribution.

Since  $F$  and  $Dcr$  both output an estimate of (the expectation of)  $Z_i^B | M_{\leq i}$  in a way that is indistinguishable from the real distribution of  $Z_i^B$  (given  $M_{\leq i}$ ), both algorithms output essentially the same value. Otherwise, the ‘‘accurate’’ algorithm can be used to distinguish the output of the ‘‘inaccurate’’ algorithm from the real output. It follows that

$$(U_{p^{\hat{A}}}, U_{p^{\hat{B}}}, M_{\leq i})_{(p^{\hat{A}}, p^{\hat{B}}) \leftarrow Dcr(M_{\leq i})} \stackrel{c}{\approx}_{1/r} (U_{p^{\hat{A}}}, U_{F_i^B}, M_{\leq i})_{(p^{\hat{A}}, \cdot) \leftarrow Dcr(M_{\leq i})} \tag{10}$$

Using a data-processing argument in combination with Eqs. (9) and (10), we deduce that

$$(F_{i+1}, Z_i^B, F_{\leq i}) \stackrel{c}{\approx}_{1/r} (U_{p^{\hat{A}}}, U_{p^{\hat{B}}}, F_{\leq i})_{(p^{\hat{A}}, p^{\hat{B}}) \leftarrow Dcr(M_{\leq i})} \tag{11}$$

$$\stackrel{c}{\approx}_{1/r} (U_{p^{\hat{A}}}, U_{F_i^B}, F_{\leq i})_{(p^{\hat{A}}, \cdot) \leftarrow Dcr(M_{\leq i})} . \tag{12}$$

Finally, conditioned on  $F_{\leq i}$ , we observe that the pair of random variables  $(U_{p^{\hat{A}}}, U_{F_i^B})_{(p^{\hat{A}}, \cdot) \leftarrow Dcr(M_{\leq i})}$  are mutually independent since  $U_{F_i^B}$  is sampled independently according to  $F_i^B$ , and  $F_i^B$  is fully determined by  $F_{\leq i}$ .

### 1.3 Related Work

We review some of the relevant work on fair coin flipping protocols.

**Necessary Hardness Assumptions.** This line of work examines the minimal assumptions required to achieve an  $o(1/\sqrt{r})$ -bias two-party coin flipping protocols, as done in this paper. The necessity of one-way functions for weaker variants of coin flipping protocol where the honest party is allowed to abort if the other party aborts or deviates from the prescribed protocol, were considered in [5, 13, 17, 18]. More related to our bound is the work of Dachman-Soled et al. [10] who showed that any fully black-box construction of  $O(1/r)$ -bias two-party protocols based on one-way functions (with  $r$ -bit input and output) needs  $\Omega(r/\log r)$  rounds, and the work of Dachman-Soled et al. [11] showed that there is no fully black-box and function *oblivious* construction of  $O(1/r)$ -bias two-party protocols from one-way functions (a protocol is function oblivious if the outcome of protocol is independent of the choice of the one-way function used in the protocol). For the case we are interested in, i.e. constant-round coin flipping protocols, [10] yields that black-box techniques from one-way functions can only guarantee fairness of order  $1/\sqrt{r}$ .

**Lower Bounds.** Cleve [8] proved that, for every  $r$ -round two-party coin flipping protocol, there exists an efficient adversary that can bias the output by  $\Omega(1/r)$ . Cleve and Impagliazzo [9] proved that, for every  $r$ -round two-party coin flipping

protocol, there exists an inefficient fail-stop adversary that biases the output by  $\Omega(1/\sqrt{r})$ . They also showed that a similar attack exists if the parties have access to an ideal commitment scheme. All above bounds extend to the multi-party case (with no honest majority) via a simple reduction. Very recently, Beimel et al. [3] showed that *any*  $r$ -round  $n$ -parties coin flipping with  $n^k > r$ , for some  $k \in \mathbb{N}$ , can be biased by  $1/(\sqrt{r} \cdot (\log r)^k)$ . Ignoring logarithmic factors, this means that if the number of parties is  $r^{\Omega(1)}$ , the majority protocol of [2] is optimal.

**Upper Bounds.** Blum [6] presented a two-party two-round coin flipping protocol with bias  $1/4$ . Awerbuch et al. [2] presented an  $n$ -party  $r$ -round protocol with bias  $O(n/\sqrt{r})$  (the two-party case appears also in Cleve [8]). Moran et al. [19] solved the two-party case by giving a two-party  $r$ -round coin flipping protocol with bias  $O(1/r)$ . Haitner and Tsfadia [14] solved the three-party case up to poly-logarithmic factor by giving a three-party coin flipping protocol with bias  $O(\text{polylog}(r)/r)$ . Buchbinder et al. [7] showed an  $n$ -party  $r$ -round coin flipping protocol with bias  $\tilde{O}(n^{3/2} 2^n / r^{\frac{1}{2} + \frac{1}{2^n - 1 - 2}})$ . In particular, their protocol for four parties has bias  $\tilde{O}(1/r^{2/3})$ , and for  $n = \log \log r$  their protocol has bias smaller than Awerbuch et al. [2].

For the case where less than  $2/3$  of the parties are corrupt, Beimel et al. [4] showed an  $n$ -party  $r$ -round coin flipping protocol with bias  $2^{2^k}/r$ , tolerating up to  $t = (n + k)/2$  corrupt parties. Alon and Omri [1] showed an  $n$ -party  $r$ -round coin flipping protocol with bias  $\tilde{O}(2^{2^n}/r)$ , tolerating up to  $t$  corrupted parties, for constant  $n$  and  $t < 3n/4$ .

## 1.4 Open Questions

We show that constant-round coin flipping protocol with “small” bias (i.e.,  $o(1/\sqrt{r})$ -fair, for  $r$  round protocol) implies io-key-agreement. Whether such a reduction can be extended to protocols with super-constant round complexity remains open. The barrier to extending our results is that the dichotomy result of Haitner et al. [12] only guarantees indistinguishability with constant advantage (as opposed to vanishing or negligible advantage). It is worth mentioning that for protocols of super-constant round complexity, even a black-box separation between optimal (and thus between small bias) coin flipping protocol and one-way functions is not known.

The question of reducing oblivious transfer to optimally-fair coin flip is also open. We recall that all known small bias coin flipping protocols rely on it [7, 15, 20]. It is open whether the techniques of Haitner et al. [12] can provide a similar dichotomy with respect to (io-) oblivious transfer (as opposed to io-key-agreement) allowing for the realization of oblivious transfer from  $o(1/\sqrt{r})$ -fair (constant round) coin flip via the techniques of the present paper.

## Paper Organization

Basic definitions and notation used through the paper are given in Sect. 2. The formal statement and proof of the main theorem are given in Sect. 3.



## 2 Preliminaries

### 2.1 Notation

We use calligraphic letters to denote sets, uppercase for random variables and functions, lowercase for values. For  $a, b \in \mathbb{R}$ , let  $a \pm b$  stand for the interval  $[a - b, a + b]$ . For  $n \in \mathbb{N}$ , let  $[n] = \{1, \dots, n\}$  and  $(n) = \{0, \dots, n\}$ . Let  $\text{poly}$  denote the set of all polynomials, let PPT stand for probabilistic polynomial time and PPTM denote a PPT algorithm (Turing machine). A function  $\nu: \mathbb{N} \rightarrow [0, 1]$  is *negligible*, denoted  $\nu(n) = \text{neg}(n)$ , if  $\nu(n) < 1/p(n)$  for every  $p \in \text{poly}$  and large enough  $n$ . For a sequence  $x_1, \dots, x_r$  and  $i \in [r]$ , let  $x_{\leq i} = x_1, \dots, x_i$  and  $x_{< i} = x_1, \dots, x_{i-1}$ .

Given a distribution, or random variable,  $D$ , we write  $x \leftarrow D$  to indicate that  $x$  is selected according to  $D$ . Given a finite set  $\mathcal{S}$ , let  $s \leftarrow \mathcal{S}$  denote that  $s$  is selected according to the uniform distribution over  $\mathcal{S}$ . The support of  $D$ , denoted  $\text{Supp}(D)$ , be defined as  $\{u \in \mathcal{U} : D(u) > 0\}$ . The *statistical distance* between two distributions  $P$  and  $Q$  over a finite set  $\mathcal{U}$ , denoted as  $\text{SD}(P, Q)$ , is defined as  $\max_{\mathcal{S} \subseteq \mathcal{U}} |P(\mathcal{S}) - Q(\mathcal{S})| = \frac{1}{2} \sum_{u \in \mathcal{U}} |P(u) - Q(u)|$ . Distribution ensembles  $X = \{X_\kappa\}_{\kappa \in \mathbb{N}}$  and  $Y = \{Y_\kappa\}_{\kappa \in \mathbb{N}}$  are  $\delta$ -*computationally indistinguishable in the set*  $\mathcal{K}$ , denoted by  $X \approx_{\mathcal{K}, \delta} Y$ , if for every PPTM  $D$  and sufficiently large  $\kappa \in \mathcal{K}$ :  $|\Pr[D(1^\kappa, X_\kappa) = 1] - \Pr[D(1^\kappa, Y_\kappa) = 1]| \leq \delta$ .

### 2.2 Protocols

Let  $\Pi = (A, B)$  be a two-party protocol. The protocol  $\Pi$  is PPT if the running time of both  $A$  and  $B$  is polynomial in their input length (regardless of the party they interact with). We denote by  $(A(x), B(y))(z)$  a random execution of  $\Pi$  with private inputs  $x$  and  $y$ , and common input  $z$ , and sometimes abuse notation and write  $(A(x), B(y))(z)$  for the parties' output in this execution.

We will focus on no-input two-party single-bit output PPT protocol: the only input of the two PPT parties is the common security parameter given in unary representation. At the end of the execution, each party outputs a single bit. Throughout, we assume without loss of generality that the transcript contains  $1^\kappa$  as the first message. Let  $\Pi = (A, B)$  be such a two-party single-bit output protocol. For  $\kappa \in \mathbb{N}$ , let  $C_\Pi^{A, \kappa}$ ,  $C_\Pi^{B, \kappa}$  and  $T_\Pi^\kappa$  denote the outputs of  $A$ ,  $B$  and the transcript of  $\Pi$ , respectively, in a random execution of  $\Pi(1^\kappa)$ .

**Fair Coin Flipping Protocols.** Since we are concerned with a lower bound, we only give the game-based definition of coin flipping protocols (see [15] for the stronger simulation-based definition).

**Definition 2.1 (Fair coin flipping protocols).** A PPT single-bit output two-party protocol  $\Pi = (A, B)$  is an  $\varepsilon$ -fair coin flipping protocol if the following holds.

**Output delivery:** The honest party always outputs a bit (even if the other party acts dishonestly, or aborts).

**Agreement:** The parties always output the same bit in an honest execution.  
**Uniformity:**  $\Pr \left[ C_{\Pi}^{A,\kappa} = b \right] = 1/2$  (and thus  $\Pr \left[ C_{\Pi}^{B,\kappa} = b \right] = 1/2$ ), for both  $b \in \{0, 1\}$  and all  $\kappa \in \mathbb{N}$ .  
**Fairness:** For any PPT  $A^*$  and  $b \in \{0, 1\}$ , for sufficiently large  $\kappa \in \mathbb{N}$  it holds that  $\Pr \left[ C_{\Pi}^{B,\kappa} = b \right] \leq 1/2 + \varepsilon$ , and the same holds for the output bit of  $A$ .

**Key-Agreement.** We focus on single-bit output key-agreement protocols.

**Definition 2.2 (Key-agreement protocols).** A PPT single-bit output two-party protocol  $\Pi = (A, B)$  is *io-key-agreement*, if there exist an infinite  $\mathcal{K} \subseteq \mathbb{N}$  such that the following hold for  $\kappa$ 's in  $\mathcal{K}$ :

**Agreement.**  $\Pr \left[ C_{\Pi}^{A,\kappa} = C_{\Pi}^{B,\kappa} \right] \geq 1 - \text{neg}(\kappa)$ .  
**Secrecy.**  $\Pr \left[ \text{Eve}(T_{\Pi}^{\kappa}) = C_{\Pi}^{A,\kappa} \right] \leq 1/2 + \text{neg}(\kappa)$ , for every PPT Eve.

### 2.3 Martingales

**Definition 2.3 (Martingales).** Let  $X_0, \dots, X_r$  be a sequence of random variables. We say that  $X_0, \dots, X_r$  is a martingale sequence if  $\mathbf{E} [X_{i+1} \mid X_{\leq i} = x_{\leq i}] = x_i$  for every  $i \in [r - 1]$ .

In plain terms, a sequence is a martingale if the expectation of the next point conditioned on the entire history is exactly the last observed point. One way to obtain a martingale sequence is by constructing a *Doob martingale*. Such a sequence is defined by  $X_i = \mathbf{E} [f(Z) \mid Z_{\leq i}]$ , for arbitrary random variables  $Z = (Z_1, \dots, Z_r)$  and a function  $f$  of interest. We will use the following fact proven by [9] (we use the variant as proven in [3]).

**Theorem 2.1.** *Let  $X_0, \dots, X_r$  be a martingale sequence such that  $X_i \in [0, 1]$ , for every  $i \in [r]$ . If  $X_0 = 1/2$  and  $\Pr [X_r \in \{0, 1\}] = 1$ , then*

$$\Pr \left[ \exists i \in [r] \text{ s.t. } |X_i - X_{i-1}| \geq \frac{1}{4\sqrt{r}} \right] \geq \frac{1}{20}.$$

## 3 Fair Coin Flipping to Key-Agreement

In this section, we prove our main result: if there exist constant-round coin flipping protocols which improve over the  $1/\sqrt{r}$ -bias majority protocol of [2], then infinitely-often key-agreement exists as well. Formally, we prove the following theorem.

**Theorem 3.1.** *The following holds for any (constant)  $r \in \mathbb{N}$ : if there exists an  $r$ -round,  $\frac{1}{25600\sqrt{r}}$ -fair two-party coin flipping protocol, see Definition 2.1, then there exists an infinitely-often key-agreement protocol.<sup>6,7</sup>*

<sup>6</sup> Definition 2.1 requires perfect uniformity: the common output in an honest execution is an unbiased bit. The proof given below, however, easily extends to any non-trivial uniformity condition, e.g., the common output equals 1 with probability  $3/4$ .

<sup>7</sup> We remark that we did not optimize the value of the constant.

Before formally proving Theorem 3.1, we briefly recall the outline of the proof as presented in the introduction (we ignore certain constants in this outline). We begin with a good forecaster for the coin flipping protocol  $\Pi$  (which must exist, according to [12]), and we define an efficiently computable conditional expected outcome sequence  $X = (X_0, \dots, X_r)$  for  $\Pi$ , conditioned on the forecaster’s outputs. Then, we show that (1) the  $i^{\text{th}}$  backup value (default output in case the opponent aborts) should be close to  $X_i$ ; otherwise, an efficient attacker can use the forecaster to bias the output of the other party (this attack is applicable regardless of the existence of infinitely-often key-agreement). And (2), since  $X$  is a martingale sequence, “large”  $1/\sqrt{r}$ -gaps are bound to occur in some round, with constant probability. Hence, combining (1) and (2), with constant probability, for some  $i$ , there is a  $1/\sqrt{r}$ -gap between  $X_i$  and the forecasters’ prediction for one party at the preceding round  $i - 1$ . Therefore, unless protocol  $\Pi$  implies io-key-agreement, the aforementioned gap can be exploited to bias that party’s output by  $1/\sqrt{r}$ , by instructing the opponent to abort as soon as the gap is detected. In more detail, the success of the attack requires that (3) the event that a gap occurs is (almost) independent of the backup value of the honest party. It turns out that if  $\Pi$  does not imply io-key-agreement, this third property is guaranteed by the dichotomy theorem of [12]. In summary, if io-key-agreement does not exist, then protocol  $\Pi$  is at best  $1/\sqrt{r}$ -fair.

Moving to the formal proof, fix an  $r$ -round, two-party coin flipping protocol  $\Pi = (A, B)$  (we assume nothing about its fairness parameter for now). We associate the following random variables with a random honest execution of  $\Pi(1^\kappa)$ . Let  $M^\kappa = (M_1^\kappa, \dots, M_r^\kappa)$  denote the messages of the protocol and let  $C^\kappa$  denote the (always) common output of the parties. For  $i \in \{0, \dots, r\}$  and  $P \in \{A, B\}$ , let  $Z_i^{P,\kappa}$  be the “backup” value party  $P$  outputs, if the other party aborts after the  $i^{\text{th}}$  message was sent. In particular,  $Z_r^{A,\kappa} = Z_r^{B,\kappa} = C^\kappa$  and  $\Pr[C^\kappa = 1] = 1/2$ .

**Forecaster for  $\Pi$ .** We are using a forecaster for  $\Pi$ , guaranteed by the following theorem (proof readily follows from Haitner et al. [12, Theorem 3.8]).

**Theorem 3.2 (Haitner et al. [12], existence of forecasters).** *Let  $\Delta$  be a no-input, single-bit output two-party protocol. Then for any constant  $\rho > 0$ , there exists a PPT constant output-length algorithm  $F$  (forecaster) mapping transcripts of  $\Delta$  into (the binary description of) pairs in  $[0, 1] \times [0, 1]$  and an infinite set  $\mathcal{K} \in \mathbb{N}$  such that the following holds: let  $C^{A,\kappa}$ ,  $C^{B,\kappa}$  and  $T^\kappa$  denote the parties’ outputs and protocol transcript, respectively, in a random execution of  $\Delta(1^\kappa)$ . Let  $m(\kappa) \in \text{poly}$  be a bound on the number of coins used by  $F$  on transcripts in  $\text{supp}(T^\kappa)$ , and let  $S^\kappa$  be a uniform string of length  $m(\kappa)$ . Then,*

- $(C^{A,\kappa}, T^\kappa, S^\kappa) \stackrel{c}{\approx}_{\rho, \mathcal{K}} (U_{p^A}, T^\kappa, S^\kappa)_{(p^A, \cdot) = F(T^\kappa; S^\kappa)}$ , and
- $(C^{B,\kappa}, T^\kappa, S^\kappa) \stackrel{c}{\approx}_{\rho, \mathcal{K}} (U_{p^B}, T^\kappa, S^\kappa)_{(\cdot, p^B) = F(T^\kappa; S^\kappa)}$ .

letting  $U_p$  be a Boolean random variable taking the value 1 with probability  $p$ .<sup>8</sup>

Since we require a forecaster for all (intermediate) backup values of  $\Pi$ , we apply Theorem 3.2 with respect to the following variant of protocol  $\Pi$ , which simply stops the execution at a random round.

**Protocol 3.3** ( $\tilde{\Pi} = (\tilde{A}, \tilde{B})$ )

Common input: security parameter  $1^\kappa$ .

Description:

1.  $\tilde{A}$  samples  $i \leftarrow [r]$  and sends it to  $\tilde{B}$ .
2. The parties interact in the first  $i$  rounds of a random execution of  $\Pi(1^\kappa)$ , with  $\tilde{A}$  and  $\tilde{B}$  taking the role of A and B receptively.  
Let  $z_i^A$  and  $z_i^B$  be the  $i^{\text{th}}$  backup values of A and B as computed by the parties in the above execution.
3.  $\tilde{A}$  outputs  $z_i^A$ , and  $\tilde{B}$  outputs  $z_i^B$ .

Let  $\rho = 10^{-6} \cdot r^{-5/2}$ . Let  $\mathcal{K} \subseteq \mathbb{N}$  and a PPT  $F$  be the infinite set and PPT forecaster resulting by applying Theorem 3.2 with respect to protocol  $\tilde{\Pi}$  and  $\rho$ , and let  $S^\kappa$  denote a long enough uniform string to be used by  $F$  on transcripts of  $\tilde{\Pi}(1^\kappa)$ . The following holds with respect to  $\Pi$ .

**Claim 3.4.** For  $I \leftarrow [r]$ , it holds that

- $(Z_I^{A,\kappa}, M_{\leq I}^\kappa, S^\kappa) \stackrel{c}{\approx}_{\rho, \mathcal{K}} (U_{p^A}, M_{\leq I}^\kappa, S^\kappa)_{(p^A, \cdot) = F(M_{\leq I}; S^\kappa)}$ , and
- $(Z_I^{B,\kappa}, M_{\leq I}^\kappa, S^\kappa) \stackrel{c}{\approx}_{\rho, \mathcal{K}} (U_{p^B}, M_{\leq I}^\kappa, S^\kappa)_{(\cdot, p^B) = F(M_{\leq I}; S^\kappa)}$ ,

letting  $F(m_{\leq i}; r) = F(i, m_{\leq i}; r)$ .

*Proof.* Immediate, by Theorem 3.2 and the definition of  $\tilde{\Pi}$ . □

We assume without loss of generality that the common output appears on the last message of  $\Pi$  (otherwise, we can add a final message that contains this value, which does not hurt the security of  $\Pi$ ). Hence, without loss of generality it holds that  $F(m_{\leq r}; \cdot) = (b, b)$ , where  $b$  is the output bit as implied by  $m_{\leq r}$  (otherwise, we can change  $F$  to do so without hurting its forecasting quality).

For  $\kappa \in \mathbb{N}$ , we define the random variables  $F_0^\kappa, \dots, F_r^\kappa$ , by

$$F_i^\kappa = (F_i^{A,\kappa}, F_i^{B,\kappa}) = F(M_{\leq i}; S^\kappa) \tag{13}$$

---

<sup>8</sup> Haitner et al. [12] do not limit the output-length of  $F$ . Nevertheless, by applying [12] with parameter  $\rho/2$  and chopping each of the forecaster’s outputs to the first  $\lceil \log 1/\rho \rceil + 1$  (most significant) bits, yields the desired constant output-length forecaster.

**The Expected Outcome Sequence.** To attack the protocol, it is useful to evaluate at each round the expected outcome of the protocol conditioned on the forecasters' outputs so far. To alleviate notation, we assume that the value of  $\kappa$  is determined by  $|S^\kappa|$ .

**Definition 3.1 (Expected outcome function).** For  $\kappa \in \mathbb{N}$ ,  $i \in [r]$ ,  $f_{\leq i} \in \text{supp}(F_{\leq i}^\kappa)$  and  $s \in \text{Supp}(S^\kappa)$ , let

$$g(f_{\leq i}, s) = \mathbf{E} [C^\kappa \mid F_{\leq i}^\kappa = f_{\leq i}, S^\kappa = s].$$

Namely,  $g(f_{\leq i}, s)$  is the probability that the output of the protocol in a random execution is 1, given that  $F(M_{\leq j}; s) = f_j$  for every  $j \in (i)$  and  $M_1, \dots, M_r$  being the transcript of this execution.

**Expected Outcome Sequence is Approximable.** The following claim, proven in Sect. 3.1, yields that the expected outcome sequence can be approximated efficiently.

**Claim 3.5.** There exists PPTM  $G$  such that

$$\Pr [G(F_{\leq i}^\kappa, S^\kappa) \notin g(F_{\leq i}^\kappa, S^\kappa) \pm \rho] \leq \rho,$$

for every  $\kappa \in \mathbb{N}$  and  $i \in [r]$ .

Algorithm  $G$  approximates the value of  $g$  on input  $(f_{\leq i}, s) \in \text{supp}(F_{\leq i}^\kappa, S^\kappa)$  by running multiple independent instances of protocol  $\Pi(1^\kappa)$  and keeping track of the number of times it encounters  $f_{\leq i}$  and the protocol outputs one. Standard approximation techniques yield that, unless  $f_{\leq i}$  is very unlikely, the output of  $G$  is close to  $g(f_{\leq i}, s)$ . Claim 3.5 follows by carefully choosing the number of iterations for  $G$  and bounding the probability of encountering an unlikely  $f_{\leq i}$ .

**Forecasted Backup Values are Close to Expected Outcome Sequence.**

The following claim bounds the probability that the expected outcome sequence and the forecaster's outputs deviate by more than  $1/8\sqrt{r}$ . The proof is given in Sect. 3.2.

**Claim 3.6.** Assuming  $\Pi$  is  $\frac{1}{6400\sqrt{r}}$ -fair, then

$$\Pr [\exists i \in [r] \text{ s.t. } |g(F_{\leq i}^\kappa, S^\kappa) - F_i^{\mathbf{P}, \kappa}| \geq 1/8\sqrt{r}] < 1/100$$

for both  $\mathbf{P} \in \{\mathbf{A}, \mathbf{B}\}$  and large enough  $\kappa \in \mathcal{K}$ .

Loosely speaking, Claim 3.6 states that the expected output sequence and the forecaster's outputs are close for a fair protocol. If not, then either of the following attackers  $\mathbf{P}_0^*$ ,  $\mathbf{P}_1^*$  can bias the output of party  $\mathbf{P}$ : for fixed randomness  $s \in \text{supp}(S^\kappa)$ , attacker  $\mathbf{P}_z^*$  computes  $f_i = F(m_{\leq i}, s)$  for partial transcript  $m_{\leq i}$  at round  $i \in [r]$ , and aborts as soon as  $(-1)^{1-z}(G(f_{\leq i}^\kappa, s) - f_i) \geq 1/8\sqrt{r} - \rho$ . The desired bias is guaranteed by the accuracy of the forecaster (Claim 3.4), the accuracy of algorithm  $G$  (Claim 3.5) and the presumed frequency of occurrence of a suitable gap. The details of the proof are given in Sect. 3.2.

**Expected Outcome Sequence has Large Gap.** Similarly to [9], the success of our attack depends on the occurrence of large gaps in the expected outcome sequence. The latter is guaranteed by [9] and [3], since the expected outcome sequence is a suitable martingale.

**Claim 3.7.** For every  $\kappa \in \mathbb{N}$ , it holds that

$$\Pr [\exists i \in [r]: |g(F_{\leq i}^\kappa, S^\kappa) - g(F_{\leq i-1}^\kappa, S^\kappa)| \geq 1/4\sqrt{r}] > 1/20.$$

*Proof.* Consider the sequence of random variables  $G_0^\kappa, \dots, G_r^\kappa$  defined by  $G_i^\kappa = g(F_{\leq i}^\kappa, S^\kappa)$ . Observe that this is a Doob (and hence, strong) martingale sequence, with respect to the random variables  $Z_0 = S^\kappa$  and  $Z_i = F_i^\kappa$  for  $i \in [r]$ , and the function  $f(S^\kappa, F_{\leq r}^\kappa) = g(F_{\leq r}^\kappa, S^\kappa) = F_r^\kappa[0]$  (i.e., the function that outputs the actual output of the protocol, as implied by  $F_r^\kappa$ ). Clearly,  $G_0^\kappa = 1/2$  and  $G_r^\kappa \in \{0, 1\}$  (recall that we assume that  $F(M_{\leq r}; \cdot) = (b, b)$ , where  $b$  is the output bit as implied by  $M_{\leq r}$ ). Thus, the proof follows by Theorem 2.1.  $\square$

**Independence of Attack Decision.** Claim 3.4 immediately yields that the expected values of  $F_i$  and  $Z_i^P$  are close, for both  $P \in \{A, B\}$  and every  $i \in [r]$ . Assuming io-key-agreement does not exist, the following claim essentially states that  $F_i$  and  $Z_i^P$  remain close in expectation, even if we condition on some event that depends on the other party’s next message. This observation will allow us to show that, when a large gap in the expected outcome is observed by one of the parties, the (expected value of the) backup value of the other party still lags behind. The following claim captures the core of the novel idea in our attack, and its proof is the most technical aspect towards proving our main result.

**Claim 3.8 (Independence of attack decision).** Let  $D$  be a single-bit output PPTM. For  $\kappa \in \mathbb{N}$  and  $P \in \{A, B\}$ , let  $E_1^{P,\kappa}, \dots, E_r^{P,\kappa}$  be the sequence of random variables defined by  $E_i^{P,\kappa} = D(F_{\leq i}^\kappa, S^\kappa)$  if  $P$  sends the  $i^{\text{th}}$  message in  $\Pi(1^\kappa)$ , and  $E_i^{P,\kappa} = 0$  otherwise.

Assume io-key-agreement protocols do not exist. Then, for any  $P \in \{A, B\}$  and infinite subset  $\mathcal{K}' \subseteq \mathcal{K}$ , there exists an infinite set  $\mathcal{K}'' \subseteq \mathcal{K}'$  such that

$$\mathbf{E} \left[ E_{i+1}^{P,\kappa} \cdot (Z_i^{\bar{P},\kappa} - F_i^{\bar{P},\kappa}) \right] \in \pm 4r\rho$$

for every  $\kappa \in \mathcal{K}''$  and  $i \in (r - 1)$ , where  $\bar{P} \in \{A, B\} \setminus \{P\}$ .

Since  $\mathbf{E} \left[ E_{i+1}^{P,\kappa} \cdot (Z_i^{\bar{P},\kappa} - F_i^{\bar{P},\kappa}) \right] = \mathbf{E} \left[ E_{i+1}^{P,\kappa} \cdot \mathbf{E} \left[ Z_i^{\bar{P},\kappa} - F_i^{\bar{P},\kappa} \mid E_{i+1}^{P,\kappa} = 1 \right] \right]$ , Claim 3.8 yields that the expected values of  $F_i$  and  $Z_i^P$  remain close, even when conditioning on a likely-enough-event over the next message of  $P$ .

The proof of Claim 3.8 is given in Sect. 3.3. In essence, we use the recent dichotomy of Haitner et al. [12] to show that if io-key-agreement does not exist, then the values of  $E_{i+1}^{P,\kappa}$  and  $Z_i^{\bar{P},\kappa}$  conditioned on  $M_{\leq i}$  (which determines the value of  $F_i^{\bar{P},\kappa}$ ), are (computationally) close to be in a product distribution.

**Putting Everything Together.** Equipped with the above observations, we prove Theorem 3.1.

*Proof of Theorem 3.1.* Let  $\Pi$  be an  $\varepsilon = \frac{1}{25600\sqrt{r}}$ -fair coin flipping protocol. By Claims 3.6 and 3.7, we can assume without loss of generality that there exists an infinite subset  $\mathcal{K}' \subseteq \mathcal{K}$  such that

$$\Pr \left[ \exists i \in [r]: \text{A sends } i^{\text{th}} \text{ message in } \Pi(1^\kappa) \wedge g(F_{\leq i}^\kappa, S^\kappa) - F_{i-1}^{\text{B},\kappa} \geq \frac{1}{8\sqrt{r}} \right] \geq \frac{1}{80} - \frac{1}{100} = \frac{1}{400} \tag{14}$$

We define the following PPT fail-stop attacker  $\text{A}^*$  taking the role of  $\text{A}$  in  $\Pi$ . We will show below that assuming io-key-agreement do not exist, algorithm  $\text{A}^*$  succeeds in biasing the output of  $\text{B}$  towards zero by  $\varepsilon$  for all  $\kappa \in \mathcal{K}''$ , contradicting the presumed fairness of  $\Pi$ . In the following, let  $\text{G}$  be the PPTM guaranteed to exist by Claim 3.5.

**Algorithm 3.9** ( $\text{A}^*$ )

*Input:* security parameter  $1^\kappa$ .

*Description:*

1. Sample  $s \leftarrow S^\kappa$  and start a random execution of  $\text{A}(1^\kappa)$ .
2. Upon receiving the  $(i - 1)$  message  $m_{i-1}$ , do
  - (a) Forward  $m_{i-1}$  to  $\text{A}$ , and let  $m_i$  be the next message sent by  $\text{A}$ .
  - (b) Compute  $f_i = (f_i^{\text{A}}, f_i^{\text{B}}) = \text{F}(m_{\leq i}, s)$ .
  - (c) Compute  $\tilde{g}_i = \text{G}(f_{\leq i}, s)$ .
  - (d) If  $\tilde{g}_i \geq f_{i-1}^{\text{B}} + 1/16\sqrt{r}$ , abort (without sending further messages). Otherwise, send  $m_i$  to  $\text{B}$  and proceed to the next round.

It is clear that  $\text{A}^*$  is a PPTM. We conclude the proof showing that assuming io-key-agreement do not exist,  $\text{B}$ 's output when interacting with  $\text{A}^*$  is biased towards zero by at least  $\varepsilon$ .

The following random variables are defined with respect to a a random execution of  $(\text{A}^*, \text{B})(1^\kappa)$ . Let  $S^\kappa$  and  $F^\kappa = (F_1^\kappa, \dots, F_r^\kappa)$  denote the values of  $s$  and  $f_1, \dots, f_r$  sampled by  $\text{A}^*$ . Let  $Z^{\text{B},\kappa} = (Z_1^{\text{B},\kappa}, \dots, Z_r^{\text{B},\kappa})$  denote the backup values computed by  $\text{B}$ . For  $i \in [r]$ , let  $E_i^\kappa$  be the event that  $\text{A}^*$  decides to abort in round  $i$ . Finally, let  $J^\kappa$  be the index  $i$  with  $E_i^\kappa = 1$ , setting it to  $r + 1$  if no such index exist. Below, if we do not quantify over  $\kappa$ , it means that the statement holds for any  $\kappa \in \mathbb{N}$ .

By Claim 3.5 and Eq. (14),

$$\Pr [J^\kappa \neq r + 1] > \frac{1}{400} - \rho \geq \frac{1}{800} \tag{15}$$

for every  $\kappa \in \mathcal{K}'$ . Where since the events  $E_i^\kappa$  and  $E_j^\kappa$  for  $i \neq j$  are disjoint,

$$\begin{aligned} \mathbf{E} \left[ Z_{J^{\kappa-1}}^{\mathbf{B},\kappa} - F_{J^{\kappa-1}}^{\mathbf{B},\kappa} \right] &= \mathbf{E} \left[ \sum_{i=1}^{r+1} E_i^\kappa \cdot (Z_{i-1}^{\mathbf{B},\kappa} - F_{i-1}^{\mathbf{B},\kappa}) \right] \\ &= \sum_{i=1}^{r+1} \mathbf{E} \left[ E_i^\kappa \cdot (Z_{i-1}^{\mathbf{B},\kappa} - F_{i-1}^{\mathbf{B},\kappa}) \right] \\ &= \sum_{i=1}^r \mathbf{E} \left[ E_i^\kappa \cdot (Z_{i-1}^{\mathbf{B},\kappa} - F_{i-1}^{\mathbf{B},\kappa}) \right]. \end{aligned} \tag{16}$$

The last inequality holds since the protocol’s output appears in the last message, by assumption, and thus without loss of generality  $Z_r^{\mathbf{B},\kappa} = F_r^{\mathbf{B},\kappa}$ . Consider the single-bit output PPTM  $\mathbf{D}$  defined as follows: on input  $(f_{\leq i}, s)$  where  $f_{\leq i}$  is a sequence of pairs of values, i.e.,  $f_{\leq i} = (f_1^{\mathbf{A}}, f_1^{\mathbf{B}}), \dots, (f_i^{\mathbf{A}}, f_i^{\mathbf{B}})$ , it outputs 1 if  $\mathbf{G}(f_{\leq i}, s) - f_{i-1}^{\mathbf{B}} \geq 1/16\sqrt{r}$ , and  $\mathbf{G}(f_{\leq j}, s) - f_{j-1}^{\mathbf{B}} < 1/16\sqrt{r}$  for all  $j < i$ . Otherwise, it outputs zero. Observe that  $E_i^\kappa$  is the indicator of the event  $\mathbf{A}$  sends the  $i^{\text{th}}$  message in  $\Pi(1^\kappa)$  and  $\mathbf{D}(F_{\leq i}^\kappa, S^\kappa) = 1$ , for any fixing of  $(F^\kappa, S^\kappa, Z^{\mathbf{B},\kappa})$ . Thus, assuming io-key-agreement protocols do not exist, Claim 3.8 yields that that there exists an infinite set  $\mathcal{K}'' \subset \mathcal{K}'$  such that

$$\mathbf{E} \left[ E_{i+1}^\kappa \cdot (Z_i^{\mathbf{B},\kappa} - F_i^{\mathbf{B},\kappa}) \right] \in \pm 4r\rho \tag{17}$$

for every  $\kappa \in \mathcal{K}''$  and  $i \in [r-1]$ . Putting together Eqs. (16) and (17), we conclude that, for every  $\kappa \in \mathcal{K}''$ ,

$$\mathbf{E} \left[ Z_{J^{\kappa-1}}^{\mathbf{B},\kappa} - F_{J^{\kappa-1}}^{\mathbf{B},\kappa} \right] \in \pm 4r^2\rho. \tag{18}$$

Recall that our goal is to show that  $\mathbf{E} \left[ Z_{J^{\kappa-1}}^{\mathbf{B},\kappa} \right]$  is significantly smaller than  $1/2$ . We do so by showing that it is significantly smaller than  $\mathbf{E} [g(F_{\leq J^\kappa}^\kappa, S^\kappa)]$  which equals  $1/2$ , since, by tower law (total expectation),

$$\mathbf{E} [g(F_{\leq J^\kappa}^\kappa, S^\kappa)] = \mathbf{E} [C^\kappa] = 1/2. \tag{19}$$

Finally, let  $G_i$  be the value of  $\mathbf{G}(F_{\leq i}, S^\kappa)$  computed by  $\mathbf{A}^*$  in the execution of  $(\mathbf{A}^*, \mathbf{B})(1^\kappa)$  considered above, letting  $G_{r+1} = g(F_{\leq r+1}^\kappa, S^\kappa)$ . Claim 3.5 yields that

$$\mathbf{E} [g(F_{\leq J^\kappa}^\kappa, S^\kappa) - G_{J^\kappa}] \leq 2r\rho. \tag{20}$$



Putting all the above observations together, we conclude that, for every  $\kappa \in \mathcal{K}''$ ,

$$\begin{aligned} & \mathbf{E} \left[ Z_{J^{\kappa}-1}^{\mathbf{B},\kappa} \right] \\ &= \mathbf{E} \left[ g(F_{\leq J^{\kappa}}^{\kappa}, S^{\kappa}) \right] - \mathbf{E} \left[ G_{J^{\kappa}} - F_{J^{\kappa}-1}^{\mathbf{B},\kappa} \right] \\ & \quad + \mathbf{E} \left[ Z_{J^{\kappa}-1}^{\mathbf{B},\kappa} - F_{J^{\kappa}-1}^{\mathbf{B},\kappa} \right] - \mathbf{E} \left[ g(F_{\leq J^{\kappa}}^{\kappa}, S^{\kappa}) - G_{J^{\kappa}} \right] \\ &\leq \frac{1}{2} - \mathbf{E} \left[ G_{J^{\kappa}} - F_{J^{\kappa}-1}^{\mathbf{B},\kappa} \mid J^{\kappa} \neq r + 1 \right] \cdot \Pr [J^{\kappa} \neq r + 1] + 4r^2\rho + 2r\rho \\ &\leq \frac{1}{2} - (1/16\sqrt{r}) \cdot (1/800) + 4r^2\rho + 2r\rho \\ &< \frac{1}{2} - \frac{1}{25600\sqrt{r}}. \end{aligned}$$

The first inequality holds by Eqs. (18) to (20). The second inequality holds by the definition of  $J^{\kappa}$  and Eq. (15). The last inequality holds by our choice of  $\rho$ .  $\square$

### 3.1 Approximating the Expected Outcome Sequence

In this section we prove Claim 3.5, restated below.

**Claim 3.10 (Claim 3.5, restated).** There exists PPTM  $G$  such that

$$\Pr [G(F_{\leq i}^{\kappa}, S^{\kappa}) \notin g(F_{\leq i}^{\kappa}, S^{\kappa}) \pm \rho] \leq \rho,$$

for every  $\kappa \in \mathbb{N}$  and  $i \in [r]$ .

The proof of Claim 3.10 is straightforward. Since there are only constant number of rounds and  $F$  has constant output-length, when fixing the randomness of  $F$ , the domain of  $G$  has constant size. Hence, the value of  $g$  can be approximated well via sampling. Details below.

Let  $c$  be a bound on the number of possible outputs of  $F$  (recall that  $F$  has constant output-length). We are using the following implementation for  $G$ . In the following, let  $\bar{F}((m_1, \dots, m_i); s) = (F(m_1; s), \dots, (F(m_i; s)))$  (i.e.,  $\bar{F}(M_{\leq i}; S^{\kappa}) = F_{\leq i}$ ).

**Algorithm 3.11** ( $G$ )

*Parameters:*  $v = \left\lceil \frac{1}{2} \cdot \left(\frac{2c^r}{\rho}\right)^4 \cdot \ln\left(\frac{8}{\rho}\right) \right\rceil$ .

*Input:*  $f_{\leq i} \in \text{supp}(F_{\leq i}^{\kappa})$  and  $s \in \text{Supp}(S^{\kappa})$ .

*Description:*

1. Sample  $v$  transcripts  $\{m^j, c^j\}_{j \in [v]}$  by taking the (full) transcripts and outputs of  $v$  independent executions of  $\Pi(1^{\kappa})$ .
2. For every  $j \in [v]$  let  $f_i^j = \bar{F}(m_{\leq i}^j; s)$ .
3. Let  $q = \left| \left\{ j \in [v] : f_{\leq i}^j = f_{\leq i} \right\} \right|$  and  $p = \left| \left\{ j \in [v] : f_{\leq i}^j = f_{\leq i} \wedge c^j = 1 \right\} \right|$ .
4. Set  $\tilde{g} = p/q$ . (Set  $\tilde{g} = 0$  if  $q = p = 0$ .)
5. Output  $\tilde{g}$ .

*Remark 3.1 (A more efficient approximator.).* The running time of algorithm G above is exponential in  $r$ . While this does not pose a problem for our purposes here, since  $r$  is constant, it might leave the impression that our approach cannot be extended to protocols with super-constant round complexity. So it is worth mentioning that the running time of G can be reduced to be polynomial in  $r$ , by using the augmented weak martingale paradigm of Beimel et al. [3]. Unfortunately, we currently cannot benefit from this improvement, since the result of [12] only guarantees indistinguishability for constant  $\rho$ , which makes it useful only for attacking constant-round protocols.

We prove Claim 3.10 by showing that the above algorithm approximates  $g$  well.

*Proof of Claim 3.10.* To prove the quality of G in approximating  $g$ , it suffices to prove the claim for every every  $\kappa \in \mathbb{N}$ ,  $i \in [r]$  and fixed  $s \in \text{supp}(S^\kappa)$ . That is

$$\Pr \left[ \left| g(\bar{F}(M_{\leq i}, s), s) - G(\bar{F}(M_{\leq i}, s), s) \right| \geq \rho \right] \leq \rho, \tag{21}$$

where the probability is also taken over the random coins of G.

Fix  $\kappa \in \mathbb{N}$  and omit it from the notation, and fix  $i \in [r]$  and  $s \in S^\kappa$ . Let  $\mathcal{D}_i = \{f_{\leq i} : \Pr [\bar{F}(M_{\leq i}, s) = f_{\leq i}] \geq \rho/2c^r\}$ . By Hoeffding’s inequality [16], for every  $f_{\leq i} \in \mathcal{D}$ , it holds that

$$\begin{aligned} \Pr \left[ \left| g(f_{\leq i}, s) - G(f_{\leq i}, s) \right| \geq \rho \right] &\leq 4 \cdot \exp \left( -2 \cdot v \cdot (\rho/2c^r)^4 \right) \\ &\leq 4 \cdot \exp \left( -\frac{v\rho^4}{8c^{4r}} \right) \\ &\leq \rho/2. \end{aligned} \tag{22}$$

It follows that

$$\begin{aligned} \Pr \left[ \left| g(\bar{F}(M_{\leq i}, s), s) - G(\bar{F}(M_{\leq i}, s), s) \right| \geq \rho \right] &\leq \Pr \left[ (\bar{F}(M_{\leq j}, s) \notin \mathcal{D}) + \rho/2 \right] \\ &\leq |\text{Supp}(\bar{F}(M_{\leq j}, s))| \cdot \rho/2c^r + \rho/2 \\ &\leq c^r \cdot \rho/2c^r + \rho/2 = \rho. \end{aligned}$$

□

### 3.2 Forecasted Backup Values Are Close to Expected Outcome Sequence

In this section, we prove Claim 3.6 (restated below).

**Claim 3.12 (Claim 3.6, restated).** Assuming  $\Pi$  is  $\frac{1}{6400\sqrt{r}}$ -fair, then

$$\Pr \left[ \exists i \in [r] \text{ s.t. } \left| g(F_{\leq i}^\kappa, S^\kappa) - F_i^{\mathbb{P}, \kappa} \right| \geq 1/8\sqrt{r} \right] < 1/100$$

for both  $\mathbb{P} \in \{\mathbb{A}, \mathbb{B}\}$  and large enough  $\kappa \in \mathcal{K}$ .

*Proof.* Assume the claim does not hold for  $\mathbf{P} = \mathbf{B}$  and infinitely many security parameters  $\mathcal{K}$  (the case  $\mathbf{P} = \mathbf{A}$  is proven analogously). That is, for all  $\kappa \in \mathcal{K}$  and without loss of generality, it holds that

$$\Pr \left[ \exists i \in [r] \text{ s.t. } g(F_{\leq i}^\kappa, S^\kappa) - F_i^{\mathbf{B}, \kappa} \geq \frac{1}{8\sqrt{r}} \right] \geq \frac{1}{200}. \tag{23}$$

Consider the following PPT fail-stop attacker  $\mathbf{A}^*$  taking the role of  $\mathbf{A}$  in  $\Pi$  to bias the output of  $\mathbf{B}$  towards zeros.

**Algorithm 3.13** ( $\mathbf{A}^*$ )

*Input:* security parameter  $1^\kappa$ .

*Description:*

1. Samples  $s \leftarrow S^\kappa$  and start a random execution of  $\mathbf{A}(1^\kappa)$ .
2. For  $i = 1 \dots r$ :

*After sending (or receiving) the prescribed message  $m_i$ :*

- (a) Let  $f_i = \mathbf{F}(m_{\leq i}; s)$  and  $\mu_i = \mathbf{G}(f_{\leq i}, s) - f_i$ .
- (b) Abort if  $\mu_i \geq \frac{1}{8\sqrt{r}} - \rho$  (without sending further messages).

*Otherwise, proceed to the next round.*

In the following, we fix a large enough  $\kappa \in \mathcal{K}$  such that Eq. (23) holds, and we omit it from the notation when the context is clear. We show that algorithm  $\mathbf{A}^*$  biases the output of  $\mathbf{B}$  towards zero by at least  $1/(6400\sqrt{r})$ .

We associate the following random variables with a random execution of  $(\mathbf{A}^*, \mathbf{B})$ . Let  $J$  denote the index where the adversary aborted, i.e., the smallest  $j$  such that  $\mathbf{G}(F_{\leq j}, S) - F_j^{\mathbf{B}} \geq \frac{1}{8\sqrt{r}} - \rho$ , or  $J = r$  if no abort occurred. The following expectations are taken over  $(F_{\leq i}, S)$  and the random coins of  $\mathbf{G}$ . We bound  $\mathbf{E}[Z_J^{\mathbf{B}}]$ , i.e. the expected output of the honest party.

$$\begin{aligned} \mathbf{E}[Z_J^{\mathbf{B}}] &= \mathbf{E}[Z_J^{\mathbf{B}}] + \mathbf{E}[g(F_{\leq J}, S)] - \mathbf{E}[g(F_{\leq J}, S)] + \mathbf{E}[\mathbf{G}(F_{\leq J}, S) - F_J^{\mathbf{B}}] - \mathbf{E}[\mathbf{G}(F_{\leq J}, S) - F_J^{\mathbf{B}}] \\ &= \mathbf{E}[g(F_{\leq J}, S)] - \mathbf{E}[\mathbf{G}(F_{\leq J}, S) - F_J^{\mathbf{B}}] + \mathbf{E}[\mathbf{G}(F_{\leq J}, S) - g(F_{\leq J}, S)] + \mathbf{E}[Z_J^{\mathbf{B}} - F_J^{\mathbf{B}}] \\ &= \frac{1}{2} - \mathbf{E}[\mathbf{G}(F_{\leq J}, S) - F_J^{\mathbf{B}}] + \mathbf{E}[\mathbf{G}(F_{\leq J}, S) - g(F_{\leq J}, S)] + \mathbf{E}[Z_J^{\mathbf{B}} - F_J^{\mathbf{B}}]. \end{aligned} \tag{24}$$

The last equation follows from  $\mathbf{E}[g(F_{\leq J}, S)] = \mathbf{E}[C]$  and thus  $\mathbf{E}[g(F_{\leq J}, S)] = \frac{1}{2}$  (for a more detailed argument see Eq. (19) and preceding text). We bound each of the terms above separately. First, observe that

$$\begin{aligned}
 & \Pr [J \neq r] \\
 & \geq \Pr [(\forall i \in [r]: |G(F_{\leq i}, S) - g(F_{\leq i}, S)| \leq \rho) \\
 & \quad \wedge \left( \exists j \in [r]: g(F_{\leq j}, S) - F_j^{\mathbf{B}} \geq \frac{1}{8\sqrt{r}} \right)] \\
 & \geq \Pr \left[ \exists j \in [r]: g(F_{\leq j}, S) - F_j \geq \frac{1}{8\sqrt{r}} \right] \\
 & \quad - \Pr [\exists i \in [r]: |G(F_{\leq i}, S) - g(F_{\leq i}, S)| > \rho] \\
 & \geq \frac{1}{200} - \rho \\
 & \geq \frac{1}{400}. \tag{25}
 \end{aligned}$$

The penultimate inequality is by Eq. (24) and Claim 3.5. It follows that

$$\begin{aligned}
 \mathbf{E} [g(F_{\leq J}, S) - F_J^{\mathbf{B}}] &= \Pr [J \neq r] \cdot \mathbf{E} [g(F_{\leq J}, S) - F_J^{\mathbf{B}} \mid J \neq r] \\
 &\geq \frac{1}{400} \cdot \left( \frac{1}{8\sqrt{r}} - \rho \right) - \mathbf{E} [G(F_{\leq J}, S) - g(F_{\leq J}, S)] \\
 &\geq \frac{1}{400} \cdot \frac{1}{8\sqrt{r}} - 3\rho. \tag{26}
 \end{aligned}$$

The penultimate inequality is by Claim 3.5. Finally, since we were taking  $\kappa$  large enough, Claim 3.4 and a data-processing argument yields that

$$\mathbf{E} [Z_J^{\mathbf{B}} - F_J^{\mathbf{B}}] \leq r\rho \tag{27}$$

We conclude that  $\mathbf{E} [g(F_{\leq J}, S) - F_J^{\mathbf{B}}] \geq \frac{1}{400} \cdot \frac{1}{8\sqrt{r}} - (r + 3)\rho > 1/(6400\sqrt{r})$ , in contradiction to the assumed fairness of  $\Pi$ .  $\square$

### 3.3 Independence of Attack Decision

In this section, we prove Claim 3.8 (restated below).

**Claim 3.14 (Claim 3.8, restated).** Let  $D$  be a single-bit output PPTM. For  $\kappa \in \mathbb{N}$  and  $P \in \{A, B\}$ , let  $E_1^{P, \kappa}, \dots, E_r^{P, \kappa}$  be the sequence of random variables defined by  $E_i^{P, \kappa} = D(F_{\leq i}^{\kappa}, S^{\kappa})$  if  $P$  sends the  $i^{\text{th}}$  message in  $\Pi(1^{\kappa})$ , and  $E_i^{P, \kappa} = 0$  otherwise.

Assume io-key-agreement protocols do not exist. Then, for any  $P \in \{A, B\}$  and infinite subset  $\mathcal{K}' \subseteq \mathcal{K}$ , there exists an infinite set  $\mathcal{K}'' \subseteq \mathcal{K}'$  such that

$$\mathbf{E} \left[ E_{i+1}^{P, \kappa} \cdot (Z_i^{\bar{P}, \kappa} - F_i^{\bar{P}, \kappa}) \right] \in \pm 4r\rho$$

for every  $\kappa \in \mathcal{K}''$  and  $i \in (r - 1)$ , where  $\bar{P} \in \{A, B\} \setminus \{P\}$ .

We prove for  $P = A$ . Consider the following variant of  $\Pi$  in which the party playing  $A$  is outputting  $E_i^A$  and the party playing  $B$  is outputting its backup value.

**Protocol 3.15** ( $\widehat{\Pi} = (\widehat{A}, \widehat{B})$ )

Common input: security parameter  $1^\kappa$ .

Description:

1. Party  $\widehat{A}$  samples  $i \leftarrow [r]$  and  $s \leftarrow S^\kappa$ , and sends them to  $\widehat{B}$ .
2. The parties interact in the first  $i - 1$  rounds of a random execution of  $\Pi(1^\kappa)$ , with  $\widehat{A}$  and  $\widehat{B}$  taking the role of  $A$  and  $B$  respectively.  
 Let  $m_1, \dots, m_{i-1}$  be the messages, and let  $z_{i-1}^B$  be the  $(i - 1)$  backup output of  $B$  in the above execution.
3.  $\widehat{A}$  sets the value of  $e_i^A$  as follows:  
 If  $A$  sends the  $i - 1$  message above, then it sets  $e_i^A = 0$ .  
 Otherwise, it
  - (a) Continues the above execution of  $\Pi$  to compute its next message  $m_i$ .
  - (b) Computes  $f_i = F(m_{\leq i}, s)$ .
  - (c) Let  $e_i^A = D(f_{\leq i}, s)$ .
- (4)  $\widehat{A}$  outputs  $e_i^A$  and  $B$  outputs  $z_{i-1}^B$ .

We apply the the following dichotomy result of Haitner et al. [12] on the above protocol.

**Theorem 3.16 (Haitner et al.[12], Theorem 3.18, dichotomy of two-party protocols).** *Let  $\Delta$  be an efficient single-bit output two-party protocol. Assume io-key-agreement protocol do not exist, then for any constant  $\rho > 0$  and infinite subset  $\mathcal{K} \subseteq \mathbb{N}$ , there exists a PPT algorithm  $Dcr$  (decorrelator) mapping transcripts of  $\Delta$  into (the binary description of) pairs in  $[0, 1] \times [0, 1]$  and an infinite set  $\mathcal{K}' \in \mathbb{N}$ , such that the following holds: let  $C^{A,\kappa}$ ,  $C^{B,\kappa}$  and  $T^\kappa$  denote the parties' output and protocol transcript in a random execution of  $\Delta(1^\kappa)$ . Let  $m(\kappa) \in \text{poly}$  be a bound on the number of coins used by  $Dcr$  on transcripts in  $\text{supp}(T^\kappa)$ , and let  $S^\kappa$  be a uniform string of length  $m(\kappa)$ . Then*

$$(C^{A,\kappa}, C^{B,\kappa}, T^\kappa, S^\kappa) \stackrel{c}{\approx}_{\rho, \mathcal{K}'} (U_{p^A}, U_{p^A}, T^\kappa, S^\kappa)_{(p^A, p^B) = Dcr(T^\kappa; S^\kappa)}$$

letting  $U_p$  be a Boolean random variable taking the value 1 with probability  $p$ .

*Proof of Claim 3.14.* Assume io-key-agreement does not exists, and let  $\mathcal{K}'' \subseteq \mathcal{K}'$  and a PPT  $Dcr$  be the infinite set and PPT decorrelator resulting by applying Theorem 3.16 with respect to protocol  $\widehat{\Pi}$  and  $\rho$ . Let  $\widehat{S}^\kappa$  denote a long enough uniform string to be used by  $Dcr$  on transcripts of  $\widehat{\Pi}(1^\kappa)$ . Then for  $I \leftarrow (r - 1)$ , letting  $Dcr(m_{\leq i}, s; \widehat{s}) = Dcr(i, s, m_{\leq i}; \widehat{s})$ , it holds that

$$(E_{I+1}^{A,\kappa}, Z_I^{B,\kappa}, M_{\leq i}^\kappa, S^\kappa, \widehat{S}^\kappa) \stackrel{c}{\approx}_{\rho, \mathcal{K}''} (U_{p^A}, U_{p^B}, M_{\leq I}^\kappa, S^\kappa, \widehat{S}^\kappa)_{(p^A, p^B) = Dcr(M_{\leq I}, S^\kappa; \widehat{S}^\kappa)}. \tag{28}$$

For  $i \in [r]$ , let  $W_i^\kappa = (W_i^{A,\kappa}, W_i^{B,\kappa}) = Dcr(M_{\leq i}, S^\kappa; \widehat{S}^\kappa)$ . The proof of Claim 3.17 follows by the following three observations, proven below, that hold for large enough  $\kappa \in \mathcal{K}''$ .

**Claim 3.17.**  $\mathbf{E} \left[ E_{I+1}^{A,\kappa} \cdot Z_I^{B,\kappa} - W_I^{A,\kappa} \cdot W_I^{B,\kappa} \right] \in \pm\rho.$

**Claim 3.18.**  $\mathbf{E} \left[ W_I^{A,\kappa} \cdot F_I^{B,\kappa} - E_{I+1}^{A,\kappa} \cdot F_I^{B,\kappa} \right] \in \pm\rho.$

**Claim 3.19.**  $\mathbf{E} \left[ W_I^{A,\kappa} \cdot W_I^{B,\kappa} - W_I^{A,\kappa} \cdot F_I^{B,\kappa} \right] \in \pm 2\rho.$

We conclude that  $\mathbf{E} \left[ E_{I+1}^{P,\kappa} \cdot Z_I^{\bar{P},\kappa} - E_{I+1}^{P,\kappa} \cdot F_I^{\bar{P},\kappa} \right] \in \pm 4\rho,$  and thus

$\mathbf{E} \left[ E_{i+1}^{P,\kappa} \cdot Z_i^{\bar{P},\kappa} - E_{i+1}^{P,\kappa} \cdot F_i^{\bar{P},\kappa} \right] \in \pm 4r\rho$  for every  $i \in (r - 1).$  □

**Proving Claim 3.17.**

*Proof of Claim 3.17.* Consider algorithm D that on input  $(z^A, z^B, \cdot),$  outputs (the product)  $z^A z^B.$  By definition,

1.  $\Pr \left[ D(U_{W_I^{A,\kappa}}, U_{W_I^{B,\kappa}}, M_{\leq I}^\kappa, S^\kappa) = 1 \right] = \mathbf{E} \left[ U_{W_I^{A,\kappa}} \cdot U_{W_I^{B,\kappa}} \right] = \mathbf{E} \left[ W_I^{A,\kappa} \cdot W_I^{B,\kappa} \right],$
2.  $\Pr \left[ D(E_{I+1}^{A,\kappa}, Z_I^{B,\kappa}, M_{\leq I}^\kappa, S^\kappa) = 1 \right] = \mathbf{E} \left[ E_{I+1}^{A,\kappa} \cdot Z_I^{B,\kappa} \right].$

Hence, the proof follows by Eq. (28). □

**Proving Claim 3.18.**

*Proof of Claim 3.18.* Consider the algorithm D that on input  $(z^A, z^B, (m_{\leq I}, s)):$  (1) computes  $(\cdot, f^B) = F(m_{\leq I}; s),$  (2) samples  $u \leftarrow U_{f^B},$  and (3) outputs  $z^A \cdot u.$  By definition,

1.  $\Pr \left[ D(U_{W_I^{A,\kappa}}, U_{W_I^{B,\kappa}}, M_{\leq I}^\kappa, S^\kappa) = 1 \right] = \mathbf{E} \left[ U_{W_I^{A,\kappa}} \cdot U_{F_I^{B,\kappa}} \right] = \mathbf{E} \left[ W_I^{A,\kappa} \cdot F_I^{B,\kappa} \right],$
2.  $\Pr \left[ D(E_{I+1}^{A,\kappa}, Z_I^{B,\kappa}, M_{\leq I}^\kappa, S^\kappa) = 1 \right] = \mathbf{E} \left[ E_{I+1}^{A,\kappa} \cdot U_{F_I^{B,\kappa}} \right] = \mathbf{E} \left[ E_{I+1}^{A,\kappa} \cdot F_I^{B,\kappa} \right].$

Hence, also in this case the proof follows by Eq. (28). □

**Proving Claim 3.19.**

*Proof of Claim 3.19.* Since  $|W_I^{A,\kappa}| \leq 1,$  it suffices to prove  $\mathbf{E} \left[ |W_I^{B,\kappa} - F_I^{B,\kappa}| \right] \leq 2\rho.$  We show that if  $\mathbf{E} \left[ |W_I^{B,\kappa} - F_I^{B,\kappa}| \right] > 2\rho,$  then there exists a distinguisher with advantage greater than  $\rho$  for either the real outputs of  $\widehat{H}$  and the emulated outputs of Dcr, or, the real outputs of  $\widehat{H}$  and the emulated outputs of F, in contradiction with the assumed properties of Dcr and F.

Consider algorithm D that on input  $(z^A, z^B, m_{\leq i}, s)$  acts as follows: (1) samples  $\widehat{s} \leftarrow \widehat{S}^\kappa,$  (2) computes  $(\cdot, f^B) = F(m_{\leq i}; s)$  and  $(\cdot, w^B) = \text{Dcr}(m_{\leq i}, s; \widehat{s}),$  (3) outputs  $z^B$  if  $w^B \geq f^B,$  and  $1 - z^B$  otherwise. We compute the difference in

probability that  $D$  outputs 1 given a sample from  $\text{Dcr}(M_{\leq I}^\kappa)$  or a sample from  $F(M_{\leq I}^\kappa)$  (we omit the superscript  $\kappa$  and subscript  $I$  below to reduce clutter)

$$\begin{aligned}
 & \Pr \left[ D(U_{W_I^{A,\kappa}}, U_{W_I^{B,\kappa}}, M_{\leq I}^\kappa, S^\kappa) = 1 \right] - \Pr \left[ D(U_{F_I^{A,\kappa}}, U_{F_I^{B,\kappa}}, M_{\leq I}^\kappa, S^\kappa) = 1 \right] \\
 &= \mathbf{E} \left[ U_{W^B} \mid W^B \geq F^B \right] \cdot \Pr \left[ W^B \geq F^B \right] \\
 &\quad + \mathbf{E} \left[ 1 - U_{W^B} \mid W^B < F^B \right] \cdot \Pr \left[ W^B < F^B \right] \\
 &\quad - \mathbf{E} \left[ U_{F^B} \mid W^B \geq F^B \right] \cdot \Pr \left[ W^B \geq F^B \right] \\
 &\quad - \mathbf{E} \left[ 1 - U_{F^B} \mid W^B < F^B \right] \cdot \Pr \left[ W^B < F^B \right] \\
 &= \mathbf{E} \left[ W^B \mid W^B \geq F^B \right] \cdot \Pr \left[ W^B \geq F^B \right] \\
 &\quad - \mathbf{E} \left[ W^B \mid W^B < F^B \right] \Pr \left[ W^B < F^B \right] \\
 &\quad - \mathbf{E} \left[ F^B \mid W^B \geq F^B \right] \cdot \Pr \left[ W^B \geq F^B \right] \\
 &\quad + \mathbf{E} \left[ F^B \mid W^B < F^B \right] \cdot \Pr \left[ W^B < F^B \right] \\
 &= \mathbf{E} \left[ W^B - F^B \mid W^B \geq F^B \right] \cdot \Pr \left[ W^B \geq F^B \right] \\
 &\quad + \mathbf{E} \left[ -W^B + F^B \mid W^B < F^B \right] \Pr \left[ W^B < F^B \right] \\
 &= \mathbf{E} \left[ |W^B - F^B| \right] \\
 &> 2\rho.
 \end{aligned}$$

An averaging argument yields that either  $D$  is a distinguisher for the tuples  $(U_{F_I^{A,\kappa}}, U_{F_I^{B,\kappa}}, M_{\leq I}^\kappa, S^\kappa)$  and  $(Z_I^{A,\kappa}, Z_I^{B,\kappa}, M_{\leq I}^\kappa, S^\kappa)$  with advantage greater than  $\rho$ , in contradiction with Claim 3.4, or, algorithm  $D$  is a distinguisher for the tuples  $(U_{W_I^{A,\kappa}}, U_{W_I^{B,\kappa}}, M_{\leq I}^\kappa, S^\kappa)$  and  $(E_I^{A,\kappa}, Z_I^{B,\kappa}, M_{\leq I}^\kappa, S^\kappa)$  with advantage greater than  $\rho$ , in contradiction with Eq. (28).  $\square$

## References

1. Alon, B., Omri, E.: Almost-optimally fair multiparty coin-tossing with nearly three-quarters malicious. In: Hirt, M., Smith, A. (eds.) TCC 2016. LNCS, vol. 9985, pp. 307–335. Springer, Heidelberg (2016). [https://doi.org/10.1007/978-3-662-53641-4\\_13](https://doi.org/10.1007/978-3-662-53641-4_13)
2. Awerbuch, B., Blum, M., Chor, B., Goldwasser, S., Micali, S.: How to implement Bracha’s  $O(\log n)$  byzantine agreement algorithm (1985). Unpublished manuscript
3. Beimel, A., Haitner, I., Makriyannis, N., Omri, E.: Tighter bounds on multi-party coin flipping via augmented weak martingales and differentially private sampling. In: Proceedings of the 59th Annual Symposium on Foundations of Computer Science (FOCS) (2018)
4. Beimel, A., Omri, E., Orlov, I.: Protocols for multiparty coin toss with a dishonest majority. *J. Cryptol.* **28**(3), 551–600 (2015)
5. Berman, I., Haitner, I., Tentes, A.: Coin flipping of any constant bias implies one-way functions. *J. ACM* **65**(3), 14 (2018)
6. Blum, M.: How to exchange (secret) keys. *ACM Trans. Comput. Syst.* **1**, 175–193 (1983)

7. Buchbinder, N., Haitner, I., Levi, N., Tsfadia, E.: Fair coin flipping: tighter analysis and the many-party case. In: Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 2580–2600 (2017)
8. Cleve, R.: Limits on the security of coin flips when half the processors are faulty. In: Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC), pp. 364–369 (1986)
9. Cleve, R., Impagliazzo, R.: Martingales, collective coin flipping and discrete control processes (extended abstract) (1993). <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.1797>
10. Dachman-Soled, D., Lindell, Y., Mahmoody, M., Malkin, T.: On the black-box complexity of optimally-fair coin tossing. In: Ishai, Y. (ed.) TCC 2011. LNCS, vol. 6597, pp. 450–467. Springer, Heidelberg (2011). [https://doi.org/10.1007/978-3-642-19571-6\\_27](https://doi.org/10.1007/978-3-642-19571-6_27)
11. Dachman-Soled, D., Mahmoody, M., Malkin, T.: Can optimally-fair coin tossing be based on one-way functions? In: Lindell, Y. (ed.) TCC 2014. LNCS, vol. 8349, pp. 217–239. Springer, Heidelberg (2014). [https://doi.org/10.1007/978-3-642-54242-8\\_10](https://doi.org/10.1007/978-3-642-54242-8_10)
12. Haitner, I., Nissim, K., Omri, E., Shaltiel, R., Silbak, J.: Computational two-party correlation. In: Proceedings of the 59th Annual Symposium on Foundations of Computer Science (FOCS) (2018)
13. Haitner, I., Omri, E.: Coin flipping with constant bias implies one-way functions. *SIAM J. Comput.* **43**(2), 389–409 (2014)
14. Haitner, I., Tsfadia, E.: An almost-optimally fair three-party coin-flipping protocol. In: Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC), pp. 817–836 (2014)
15. Haitner, I., Tsfadia, E.: An almost-optimally fair three-party coin-flipping protocol. *SIAM J. Comput.* **46**(2), 479–542 (2017)
16. Hoeffding, W.: Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13–30 (1963)
17. Impagliazzo, R., Luby, M.: One-way functions are essential for complexity based cryptography. In: Proceedings of the 30th Annual Symposium on Foundations of Computer Science (FOCS), pp. 230–235 (1989)
18. Maji, H.K., Prabhakaran, M., Sahai, A.: On the computational complexity of coin flipping. In: Proceedings of the 51st Annual Symposium on Foundations of Computer Science (FOCS), pp. 613–622 (2010)
19. Moran, T., Naor, M., Segev, G.: An optimally fair coin toss. In: Reingold, O. (ed.) TCC 2009. LNCS, vol. 5444, pp. 1–18. Springer, Heidelberg (2009). [https://doi.org/10.1007/978-3-642-00457-5\\_1](https://doi.org/10.1007/978-3-642-00457-5_1)
20. Moran, T., Naor, M., Segev, G.: An optimally fair coin toss. *J. Cryptol.* **29**(3), 491–513 (2016)