Electronic Colloquium on Computational Complexity, Report No. 26 (2006)

Random Selection with an Adversarial Majority

Ronen Gradwohl*
Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science
Rehovot, 76100 Israel
ronen.gradwohl@weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~rgradwoh/

Salil Vadhan†
Division of Engineering & Applied Sciences
Harvard University
33 Oxford Street
Cambridge, MA 02138
salil@eecs.harvard.edu
http://eecs.harvard.edu/~salil/

David Zuckerman‡
Department of Computer Science
University of Texas at Austin
1 University Station C0500
Austin, TX, 78712
diz@cs.utexas.edu
http://www.cs.utexas.edu/~diz/

February 27, 2006

Abstract

We consider the problem of random selection, where p players follow a protocol to jointly select a random element of a universe of size n. However, some of the players may be adversarial and collude to force the output to lie in a small subset of the universe. We describe essentially the first protocols that solve this problem in the presence of a dishonest majority in the full-information model (where the adversary is computationally unbounded and all communication is via broadcast). Our protocols are nearly optimal in several parameters, including the round complexity (as a function of n), the randomness complexity, the communication complexity, and the tradeoffs between the fraction of honest players, the probability that the output lies in a small subset of the universe, and the density of this subset.

Keywords: cryptography, distributed computing, leader election, collective coin-flipping, information-theoretic security, samplers, randomness extractors.

*Research supported by US-Israel Binational Science Foundation Grant 2002246.

†Supported by US-Israel BSF grant 2002246, NSF grants CNS-0430336 and CCR-0133096, and ONR Grant N00014-04-1-0478.

‡Most of this work was done while visiting Harvard University, and was supported in part by a Radcliffe Institute for Advanced Study Fellowship, a John Simon Guggenheim Memorial Foundation Fellowship, a David and Lucile Packard Fellowship for Science and Engineering, and NSF Grant CCR-0310960.

ISSN 1433-8092

1 Introduction

Suppose p players wish to jointly make a random choice from a universe of size n. They follow some protocol, and if all parties play honestly, the output is indeed a uniformly random one. However, some of the players may form a coalition and deviate arbitrarily from the protocol, in an attempt to force some output. The problem of random selection is that of designing a protocol in which the influence of coalitions of dishonest players is somehow limited.

Random selection is a very useful building block for distributed algorithms and cryptographic protocols, because it allows one to first design protocols assuming a public source of randomness, which is often an easier task, and then replace public randomness with the output of a random selection protocol. Of course, for this to work, there must be a good match between the guarantees of the random selection protocol and the requirements of the application at hand. Nevertheless, this general paradigm has been applied successfully numerous times in the past in various settings, e.g., [Yao86, GMW87, GGL98, OVY93, Dam93, DGW94, Oka00, GSV98, Lin01, Bar02, KO04]. This motivates a systematic study of random selection in its own right, like the one we undertake in this paper.

The Setting. The problem of random selection has been widely studied in a variety of settings, which differ in the following respects:

Adversary's Computational Power. In some work on random selection, such as Blum's 'coin-tossing over the telephone' [Blu82], the adversary is assumed to be computationally bounded (e.g., probabilistic polynomial time). Generally, in this setting one utilizes one-way functions and other cryptographic primitives to limit the adversary's ability to cheat, and thus the resulting protocols rely on complexity assumptions. In this paper, we study the *information-theoretic* setting, where the adversary is computationally unbounded (so complexity assumptions are useless).

Communication Model and the Adversary's Information. There is a choice between having point-to-point communication channels, a broadcast channel, or both. In the case of point-to-point communication, one can either assume private channels, as in [BGW88, CCD88], or allow the adversary full access to all communication, as in the *full-information model* of Ben-Or and Linial [BL89]. We allow a broadcast channel and work in the full-information model (so there is no benefit from point-to-point channels).

Number of Players. There has been work specifically studying two-party protocols where one of the players is adversarial; examples in the full-information model include [GGL98, SV05]. Other works study p-player protocols for large p, such as the large body of work on collective coin-flipping (random selection where the universe is of size $n = 2$) and leader election [BL89, Sak89, AN93, CL95, ORV94, BN00, Zuc97, RZ98, Fei99]. In this paper, we focus on the latter setting of *p-player protocols*, but some of our results are significant even for $p = 2$.

To summarize, here we study general multiparty protocols for random selection in the full-information model (with a broadcast channel). This is the first work in this setting to focus on the case that a majority of the players may be dishonest.[1] It may be surprising that protocols exist for

[1] We note that dishonest majorities have been studied extensively in the settings of computationally bounded parties and private channels, both for Byzantine agreement and secure computation, e.g., [GMW87, LPS80, GL02].

this case, as the other two other well-studied problems in this setting, leader election and collective coin-flipping, are provably impossible to solve with an adversarial majority [Sak89].

> **The Goal:** Construct p-player protocols for selecting an element of $[n]$ such that even if a β fraction of players are cheating, the probability that the output lands in any small subset of $[n]$, of density μ, is at most ε.

Particular applications of random selection protocols often have special additional requirements, such as "simulatability." However, all of the existing work on random selection with information-theoretic security, such as [BL89, Sak89, GGL98, NOVY98, AN93, Dam93, DGW94, GSV98, RZ98, Fei99, DHRS04, SV05], seem to include at least some variant of our requirement above. Thus it is of interest to understand this requirement on its own, in particular the tradeoffs between the parameters p, n, β, μ, and ε, as well as the efficiency of protocols meeting the requirement.

As these five parameters vary, we have a very general class of problems, which includes many previously studied problems as special cases (See Section 2.2.). Some natural settings of parameters are n being exponentially large in the security parameter (e.g., choosing a random k-bit string), p being constant or polynomial, β being a constant in $(0, 1)$ (we are particularly interested in $\beta \geq 1/2$), and μ, ε either being constants in $(0, 1)$ or tending to zero.

Regarding protocol efficiency, we focus primarily on information-theoretic measures, such as the communication and round complexities, but we also provide some computationally efficient versions of our protocols.

Our Results. In this paper, we give several protocols for random selection that tolerate an arbitrarily large fraction of cheating players $\beta < 1$. The protocols are nearly optimal in many of the parameters, for example:

- One of our protocols achieves an error probability of $\varepsilon = \tilde{O}(\mu^{1-\beta})$, when the number of players is constant and the density μ of bad outcomes is arbitrary. This comes close to the lower bound of $\varepsilon \geq \mu^{1-\beta}$ proven by Goldreich, Goldwasser, and Linial [GGL98]. For a nonconstant number of players, we can come polynomially close to the lower bound, achieving $\varepsilon = \mu^{\Omega(1-\beta)}$, provided that the fraction β of cheating players is bounded away from 1.

- One of our protocols can handle any density μ of bad outcomes that is smaller than the fraction $\alpha = 1 - \beta$ of honest players while achieving an error probability ε that is bounded away from 1. More generally, we can handle any constants α, μ such that $\lfloor 1/\alpha \rfloor \leq \lceil 1/\mu \rceil - 1$, which is a tight tradeoff by a lower bound of Feige [Fei99].

- In our protocols, the total number of coins tossed by the honest parties is $\log n + o(\log n)$ (when the other parameters are constant), which almost equals the lower bound of $\log n - O(1)$. As the only bits communicated in our protocols are the random coin tosses, the communication complexity is also nearly optimal.

- As a function of n, the round complexity of our protocols is $\log^* n + O(1)$ (when the other parameters are constant). This is within a factor of essentially 2 of the $(1/2 - o(1)) \log^* n$ lower bound proven by Sanghvi and Vadhan [SV05], which applies whenever $\beta \geq 1/2$, and $\mu > 0$ and $\varepsilon < 1$ are constants.

Techniques. Our protocols build upon recent work on round-efficient leader election [RZ98, Fei99] and round-efficient two-party random selection [SV05]. Specifically, the leader election protocols of Russell and Zuckerman [RZ98] and Feige [Fei99] work by iterating a one-round protocol that reduces the task of electing a leader from p players to that of electing from $\text{polylog}(p)$ players. Similarly, the two-party random selection protocol of Sanghvi and Vadhan [SV05] utilizes a one-round protocol that reduces selecting from a universe of size n to selecting from one of size $\text{polylog}(n)$. We combine these approaches, iteratively reducing both the number of players and the universe size in parallel. To do this, we construct new one-round universe reduction protocols that work for many parties (instead of just two, as in [SV05]). We obtain these by establishing a connection between randomness extractors [NZ96] (or, equivalently, randomness-efficient samplers) and universe reduction protocols. Optimizing parameters of the underlying extractors then translates to optimizing parameters of the universe reduction protocols, resulting in the near-optimal bounds we achieve in our final protocols. Our main results, as outlined above, refer to protocols that use optimal extractors, as proven to exist via the probabilistic method, and thus are not explicit or computationally efficient. In Section 6, we also give computationally efficient versions of our protocols, using some of the best known explicit constructions of extractors. Any additional deficiencies in these protocols are due to limitations in the state-of-the-art in constructing extractors, which we view as orthogonal to the issues we study here. Indeed, if the loss turns out to be too much for some application, then that would provide motivation for further research on explicit constructions of extractors.

Organization. Section 2 includes definitions, a more detailed description of previous work and how it relates to this paper, and our results. Section 3 contains the one-round selection protocols that are the final ingredient in our protocols, and in Section 4 we give protocols that reduce the number of players and the size of the universe. In Section 5 we describe how the different pieces fit together to form our final protocols. Finally, our results on explicit protocols, as well as new and known lower bounds and their relation to our results, appear in Sections 6 and 7 respectively.

2 Formal Definitions, Previous Work and Results

2.1 Random Selection Protocols

We now define random selection protocols.

Definition 2.1 (random selection protocol) *A* (p, n)-selection protocol *is specified by p functions (players) $A_1, \ldots, A_p$, a function f, and a number t such that:*

- *At round i, the j'th player outputs (i.e. broadcasts) a message $m_i^{(j)}$, obtained by applying the function A_j to all previous messages sent, namely $\{m_l^{(k)} : k \in [p], l < i\}$, as well as the player's random coins, $r^{(j)}$.*
- *After t rounds, the players output $f(\{m_l^{(k)} : k \in [p], l \in \{1, \ldots, t\}\})$, which is an element of $[n]$.*

In the above description, the protocol cannot require the (honest) players to base their messages in round i on the messages of other players in round i. However, since we do not want to assume simultaneity within a round, we allow dishonest players to base their messages on the outputs of all the other players from the same round (but not from later rounds). That is, we do not have any private communication channels and we consider a "rushing" adversary. This results in the full-information model of [BL89]:

Definition 2.2 (full-information adversary model) *When we say a set $S \subseteq [p]$ of players in a (p, n)-selection protocol is* cheating*, we mean that these players compute their messages $m_i^{(j)}$ using arbitrary functions $(A_j^*)_{j\in S}$ (rather than the A_j's) and these functions A_j^* are applied not only to messages in previous rounds, but also the messages of the honest players in the current round i (i.e., $\{m_i^{(k)} : k \in [p] \setminus S\}$) as well as some shared coin tosses r_S among the dishonest players.*

Given this definition, our notion of security is the following.

Definition 2.3 *A (p, n)*-selection protocol *is called $(\beta, \mu, \varepsilon)$*-resilient *if when at most a β fraction of players are cheating and S is any subset of $[n]$ of density at most μ, the probability that the output lands in S is at most ε. We refer to S as the* target set*.*

We will be interested in the asymptotic behavior of protocols, so when we discuss (p, n)-selection protocols that are $(\beta, \mu, \varepsilon)$-resilient, we are implicitly referring to a family of protocols, one for each value of p, n, β, μ, and ε (or some specified infinite set of tuples $(p, n, \beta, \mu, \varepsilon)$). We are then interested in optimizing a variety of complexity measures:

Definition 2.4 (complexity measures) *The* computation time *of a (p, n)-selection protocol is the maximum total time spent by all (honest) players (to compute their messages using the functions A_j, as well as the function f) in an execution of the protocol. We call a protocol* explicit *if its computation time is* $\mathrm{poly}(\log n, p)$.

The round complexity *is the total number of rounds of the protocol. The* randomness complexity *of a protocol is the maximum total number of random bits used by the players.*[2] *(Typically this maximum is achieved when all players are honest.) The* communication complexity *of a protocol is the maximum total number of bits communicated by the players.*[3]

All our protocols are public-coin, in the sense that the honest players flip their random coins and broadcast the results. Thus, the communication complexity is equal to the randomness complexity. By convention, we assume that if a player sends a message that deviates from the protocol in some syntactically obvious way (e.g., the player outputs more bits than requested), then its message is replaced with some canonical string of the correct form (e.g., the all-zeroes string).

2.2 Previous Work

We now discuss the relationship of the above definitions, specifically of $(\beta, \mu, \varepsilon)$-resilient (p, n)-selection protocols, to existing notions and results in the literature.

[2] Actually, it will be convenient to allow the players to pick elements uniformly at random from $\{1, \ldots, m\}$ where m is determined during the protocol and may not be a power of 2, and in such a case we view this as costing $\log_2 m$ random bits.

[3] As with randomness complexity, it will be convenient to allow players to send elements of $\{1, \ldots, m\}$, in which case we charge $\log_2 m$ bits of communication.

Two-Party Random Selection. This is the special case where $p = 2$ and $\beta = 1/2$, and attention in previous work has focused on the tradeoff between μ and ε as well as the round complexity. Specifically,

- Goldreich, Goldwasser, and Linial [GGL98] constructed, for every $n = 2^i$, an explicit $(2, n)$-selection protocol that is $(1/2, \mu, O(\sqrt{\mu}))$-resilient for every $\mu > 0$. The protocol takes $2 \log n$ rounds. They also prove that the bound of $\varepsilon = O(\sqrt{\mu})$ is tight (as a special case of a more general result mentioned later).

- Sanghvi and Vadhan [SV05] constructed, for every constant $\delta > 0$ and every n, an explicit $(2, n)$-selection protocol that is $(1/2, \mu, O(\sqrt{\mu + \delta}))$-resilient for every $\mu > 0$. Their protocol takes $\log^* n + O(1)$ rounds. They also prove that $(\log^* n - \log^* \log^* n - O(1))/2$ rounds are necessary for any $(2, n)$-selection protocol that is $(1/2, \mu, \varepsilon)$-resilient for constants $\mu > 0$ and $\varepsilon < 1$.

Collective Coin-Flipping [BL89]. This is the special case when $n = 2$ and $\mu = 1/2$. Attention in the literature has focused on constructing efficient protocols that are $(\beta, 1/2, \varepsilon)$-resilient where β and ε are constants (independent of p), β is as large as possible, and $\varepsilon < 1$. Such a protocol exists for every constant $\beta < 1/2$ [BN00] and can be made explicit [Zuc97]. Conversely, it is impossible to achieve $\beta = 1/2$ and $\varepsilon < 1$ [Sak89]. Efficient constructions of such protocols have been based on leader election (described below).

Leader Election.

Definition 2.5 *A* p-player leader election protocol *is a* (p, p)*-selection protocol. It is* (β, ε)-resilient *if when at most a* β *fraction of players are cheating, the probability that the output is the index of a cheating player is at most* ε*.*

- Every $(\beta, \beta, \varepsilon)$-resilient (p, p)-selection protocol is a (β, ε)-resilient p-player leader election protocol. The converse does not hold because the the former considers *each* subset $S \subset [p]$ of density at most β as a potential target set (i.e., set of 'bad outcomes'), but the latter only considers the subset consisting of the cheating players.

- Nevertheless, a p-player leader election protocol can be used to construct a (p, n)-selection protocol for any n by having the elected leader choose a uniform, random element of $[n]$ as the output. If the election protocol is (β, ε)-resilient, then the resulting selection protocol will be $(\beta, \mu, \varepsilon + (1 - \varepsilon) \cdot \mu)$-resilient for every $\mu \geq 0$.

- By the impossibility result for collective coin-flipping mentioned above [Sak89] and the previous bullet, it is impossible to have an election protocol that is (β, ε)-resilient for $\beta = 1/2$ and $\varepsilon < 1$.

- A long line of work [AN93, CL95, ORV94, Zuc97, RZ98, Fei99] on optimizing the resilience and round complexity for leader election has culminated in the following result of Russell and Zuckerman [RZ98]. For every constant $\beta < 1/2$, there exists an $\varepsilon < 1$ such that for all p, there is an explicit (β, ε)-resilient p-player leader election protocol of round complexity $\log^* p + O(1)$. Consequently, for all constants $\beta < 1/2$ and $\mu > 0$, there is a constant $\varepsilon < 1$ such that for all p and n, there is an explicit $(\beta, \mu, \varepsilon)$-resilient (p, n)-selection protocol.

Multi-Party Random Selection. This is the general problem that encompasses the previous special cases.

- Goldreich, Goldwasser, and Linial [GGL98] constructed, for every $n = 2^i$ and every p, an explicit (p, n)-selection protocol that is $(\beta, \mu, \mu^{1-O(\beta)})$-resilient for all sufficiently small β and every $\mu > 0$. The protocol runs in polylog(n) rounds.
- Russell and Zuckerman [RZ98] constructed an explicit one-round (p, n)-selection protocol that is $(\beta, \mu, \mu \cdot n/n^{\Omega(1-\beta)})$-resilient, when $n \geq p^c$ for a certain constant c.

Notice that all but the last of the above results require that the fraction β of bad players satisfies $\beta \leq 1/2$.[4] For collective coin-flipping and leader election, this is supported by impossibility results showing that $\beta \geq 1/2$ is impossible. For 2-party random selection, it does not make sense to discuss $\beta > 1/2$. The only result which applies to $\beta \geq 1/2$ is the last one (of [RZ98]). However, the resilience $\mu \cdot n/n^{\Omega(1-\beta)}$ is quite weak and only interesting when the density μ of the target set is close to $1/n$.[5] Our work is the first to show strong results for the case $\beta > 1/2$.

2.3 Our Results

In this section, we present our main results. All of our protocols utilize certain kinds of randomness-efficient samplers (equivalently, randomness extractors). Here we present the versions of our results obtained by using optimal samplers, proven to exist via the probabilistic method. We also have explicit (i.e., computationally efficient) versions of our protocols, obtained by using best known explicit constructions of samplers; these are described in Section 6.

The first main result of this paper is the following:

Theorem 2.6 *For all constants $k \in \mathbb{N}$, $k > 0$ and $\delta > 0$, there exists a constant $\varepsilon < 1$ and a (p, n)-selection protocol with the following properties:*

(i) The protocol has $\max(\log^* p, \log^* n) + O(1)$ *rounds.*

(ii) The protocol is $(1 - \alpha, \mu, \varepsilon)$-resilient for $\alpha = 1/(k + 1) + \delta$ and $\mu = 1/k - \delta$.

(iii) The randomness complexity of the protocol is $(\log n)/\alpha + o(\log n) + O(p \log p)$.

The tradeoff between α and μ in the above theorem is optimal up to the slackness parameter δ. This is shown in Corollary 7.5, as a consequence of a lower bound of Feige [Fei99]. Furthermore, the round and randomness complexity are nearly optimal as functions of n, as shown by Corollary 7.2 and Theorem 7.6.

Setting $p = 2$ and $\alpha = 1/2$, we obtain the following two-party protocol:

Corollary 2.7 *For every constant $\delta > 0$, there exists a constant $\varepsilon < 1$ and a $(2, n)$-selection protocol with the following properties:*

(i) The protocol has $\log^* n + O(1)$ *rounds.*

(ii) The protocol is $(1/2, 1/2 - \delta, \varepsilon)$-resilient.

[4] The hidden constant in the protocol of [GGL98] is larger than 2.

[5] The significance of the [RZ98] protocol is that it is one round and only requires n polynomial in p; in fact, there is a trivial protocol with somewhat better parameters when n is exponential in p (Lemma 3.1).

(iii) The randomness complexity of the protocol is $2\log n + o(\log n)$*.*

This protocol improves upon the two-party protocol of [SV05][6] in two ways: first, the randomness complexity is a nearly optimal $2\log n + o(\log n)$, and not $\mathrm{polylog}(n)$. Second, their protocol is $(1/2, \nu, \varepsilon')$-resilient for some small constant ν, and not for the nearly optimal $\frac{1}{2} - \delta$. In other words, their resilience is not optimal in the density of the target set. On the other hand, the error probability ε' of their protocol is smaller than that of ours. However, a special case of our second theorem below gives the parameters of [SV05] with the added benefit of optimal randomness complexity.

Our next two results optimize the error probability ε as a function of the density μ of the target set and fraction β of cheating players. The first achieves a near-optimal tradeoff when the number of players is small (e.g., constant).

Theorem 2.8 *For all* $\mu, \alpha > 0$ *and* $p, n \in \mathbb{N}$ *there exists a* (p, n)*-selection protocol with the following properties:*

(i) The protocol has $t = \log^* n - \log^*(1/\mu) + O(1)$ *rounds.*

(ii) The protocol is $(1-\alpha, \mu, \varepsilon)$*-resilient for*

$$\varepsilon = \mu^\alpha \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu} + (1-\alpha)p\right)^{1-\alpha} \cdot 2^{(1-\alpha)p}.$$

(iii) The randomness complexity is $(\log n + o(\log n) + O(t \log(1/\mu)) + p + \log(1/(1-\alpha)))/\alpha + O(p \log p)$*.*

Note that when the number p of players and the fraction α of honest players are constants, the bound becomes $\varepsilon = \tilde{O}(\mu^\alpha)$, which nearly matches the lower bound of $\varepsilon \geq \mu^\alpha$ proven in [GGL98] (see Theorem 7.3). However, the bound on ε grows exponentially with p. This is removed in the following theorem, albeit at the price of achieving a slightly worse error probability of $\mu^{\Omega(\alpha)}$ (for constrained values of α).

Theorem 2.9 *There is a universal constant* c *such that for all* μ, α *such that* $\alpha \geq \sqrt{c \log\log(1/\mu)/\log(1/\mu)}$ *and all* $p, n \in \mathbb{N}$*, there exists a* (p, n)*-selection protocol with the following properties:*

(i) The protocol has $r = \max\{\log^* p, \log^* n\} - \log^*(1/\mu) + O(1)$ *rounds.*

(ii) The protocol is $(1-\alpha, \mu, \mu^{\Omega(\alpha)})$*-resilient.*

(iii) The randomness complexity is $(\log n + o(\log n) + p)/\alpha + O(p \log p) + \mathrm{poly}(1/\alpha, \log(1/\mu))$*.*

One disadvantage of the above two theorems (as compared to, say, the honest-majority protocols of [GGL98]) is that the protocols require an a priori upper-bound μ on the density of the target set. However, we also benefit from this, in that the round complexity improves as μ tends to zero. In particular, if $\mu \leq 1/\log^{(k)} n$ for some constant k, where $\log^{(k)}$ denotes k iterated logarithms, then the round complexity is *constant*.

[6] Note that in [SV05], the claimed round complexity is $2\log^* n + O(1)$, but this difference from our claim is only a difference of convention: in their model, only one player may communicate in each round, whereas we use the convention of multi-party protocols, in which all players may communicate simultaneously in one round.

3 One-Round Protocols

We start with some simple one-round protocols that will play a role in our later constructions.

Lemma 3.1 *For every $p, \ell \in \mathbb{N}$ and $n = \ell^p$, there is an explicit (p,n)-selection protocol that is $(\beta, \mu, n^\beta \cdot \mu)$-resilient for every $\beta, \mu > 0$.*

Proof: Each player outputs a random element of $[\ell]$, and we take the concatenation of the players' outputs. ■

The above protocol has two main disadvantages. First, the size of the universe $n = \ell^p$ must be at least exponential in the number of players. (We note that Russell and Zuckerman [RZ98] showed how to reduce this requirement to be only polynomial, at the price of a somewhat worse resilience. We will avoid this difficulty in a different manner, by first reducing the number of players.) Second, in terms of resilience, a target set of density μ gets multiplied by a factor that grows polynomially with the universe size (namely, n^β). However, when the number of players is small (e.g., a fixed constant) and the universe is small (e.g., $n = O(1/\mu)$), it can achieve a nearly optimal bound on ε as a function of β and μ (cf., Theorem 7.3).

Lemma 3.2 ([Fei99], Cor. 5) *For every $p, n \in \mathbb{N}$ and $\alpha, \mu \in [0,1]$ such that $\lfloor 1/\alpha \rfloor \le \lceil 1/\mu \rceil - 1$, there exists an $\varepsilon < 1$ and a (p,n)-selection protocol that is $(1-\alpha, \mu, \varepsilon)$-resilient. Specifically, one can take $\varepsilon = 1 - \exp(-\Omega(\alpha \cdot (1-\mu) \cdot np))$.*

Proof: Every player chooses a random set $S \subseteq [n]$ of density at least $1 - \mu$, and the output is the first element of $[n]$ that is contained in every S that was picked by at least an α fraction of players. Such an element exists because there exist at most $\lfloor 1/\alpha \rfloor \le \lceil 1/\mu \rceil - 1$ such sets S, but any $\lceil 1/\mu \rceil - 1$ sets of density at least $1 - \mu$ must have a common intersection. ■

The advantage of the above protocol is that it achieves an optimal tradeoff between α and μ (cf., Theorem 7.5). The main disadvantage is that ε can depend on p and n (this time with exponentially bad dependence), and that it is not sufficiently explicit — even the communication is of length $\Theta(n)$ (rather than $\mathrm{polylog}(n)$).

4 Universe and Player Reduction

The simple 1-round protocols of the previous section behave well when the number of players p and universe size n are small. Thus, as in previous work, our main efforts will be aimed at giving protocols to reduce p and n while approximately preserving the fraction β of bad players and the density μ of the bad set. Roughly speaking, in one round we will reduce p and n to $\mathrm{polylog}(p)$ and $\mathrm{polylog}(n)$, respectively. For this, we consider protocols that select a subset of the universe (or a subset of the players).

4.1 Definitions

Definition 4.1 *A* $[(p,n) \mapsto n']$-universe reduction protocol *is a p-player protocol whose output is a sequence $(s_1, \ldots, s_{n'})$ of elements of $[n]$. Such a protocol is* $[(\beta, \mu) \overset{\gamma}{\mapsto} \mu']$-resilient *if when at most a β fraction of players are cheating and S is any subset of $[n]$ of density at most μ, the probability* *that at most a μ' fraction of the output sequence is in S is at least γ. It is* explicit *if the players' strategies are computable in time* $\mathrm{poly}(\log n, p)$*, and given the protocol transcript and $i \in [n']$, the i'th element of the output sequence is computable in time* $\mathrm{poly}(\log n, p)$*.*

Notice that a (p, n)-selection protocol is equivalent to a $[(p, n) \mapsto n']$-universe reduction protocol with $n' = 1$, and the former is $(\beta, \mu, \varepsilon)$-resilient if and only if the latter is $[(\beta, \mu) \overset{1-\varepsilon}{\mapsto} 0]$-resilient.

Definition 4.2 *A* $[p \mapsto p']$-player reduction protocol *is a $[(p, p) \mapsto p']$-universe reduction protocol. It is* $[\beta \overset{\gamma}{\mapsto} \beta']$-resilient *if when at most a β fraction of players are cheating, the probability that at most a β' fraction of the output sequence are indices of cheating players is at least γ.*

Definition 4.3 *A* $[(p, n) \mapsto (p', n')]$-universe+player reduction protocol *is a p-player protocol whose output is a sequence $(s_1, \ldots, s_{n'})$ of elements of $[n]$ and a sequence $(t_1, \ldots, t_{p'})$ of elements of $[p]$. Such a protocol is* $[(\beta, \mu) \overset{\gamma}{\mapsto} (\beta', \mu')]$-resilient *if when at most a β fraction of players are cheating and S is any subset of $[n]$ of density at most μ, the probability that at most a β' fraction of the first output sequence are cheating players and at most a μ' fraction of the second output sequence is in S is at least γ. It is* explicit *if the players' strategies are computable in time* $\mathrm{poly}(\log n, p)$*, and given the protocol transcript and $i \in [n']$ (resp., $j \in [p']$), the i'th (resp., j'th) element of the first (resp., second) output sequence is computable in time* $\mathrm{poly}(\log n, p)$*.*

4.2 One-Round Reduction Protocols

In the following one-round protocols, think of $\theta = 1/\mathrm{polylog}(n)$ and $\varepsilon = 1/\mathrm{poly}(n)$.

Theorem 4.4 ([RZ98, Fei99]) *For every $p \in \mathbb{N}$, $\varepsilon > 0$, and $\theta > 0$, there is an explicit, one-round $[p \mapsto p']$-player reduction protocol with*

$$p' = O\left(\frac{1-\beta}{\theta^2} \cdot \log \frac{p}{\varepsilon}\right)$$

that is $[\beta \overset{1-\varepsilon}{\mapsto} \beta + \theta]$-resilient for all $\beta > 0$. Moreover, the randomness complexity is $p \cdot \log(p/p')$.

Proof: Let $\alpha = 1 - \beta$. The protocol used is the Lightest-Bin protocol of [Fei99]. Each player randomly picks one out of b bins, for b to be determined later. The sequence of players output by the protocol consists of the players who picked the bin chosen by the fewest number of players (ties are broken arbitrarily), and arbitrarily adding players to increase the total number to be $p' = \lfloor p/b \rfloor$.

Fix some bin, and suppose players $\{1, \ldots, \alpha p\}$ are the honest players. For each such player i, define a random variable

$$X_i = \begin{cases} 1 & \text{if player } i \text{ chooses the bin, and} \\ 0 & \text{otherwise.} \end{cases}$$

Let $X = \sum_{i=1}^{\alpha p} X_i$. Since $\mathrm{E}[X_i] = 1/b$, $\mathrm{E}[X] = \alpha p/b$. We will bound the probability that fewer than $(\alpha - \theta) \cdot p/b = (1 - \theta/\alpha) \cdot \mathrm{E}[X]$ of the honest players picked the bin, and then take a union bound over all $b \le p$ bins. By a Chernoff Bound,

$$\Pr\left[X \le \left(1 - \frac{\theta}{\alpha}\right) \mathrm{E}[X]\right] \le \exp\left(\frac{1}{2} \cdot \frac{\theta^2}{\alpha^2} \cdot \frac{\alpha p}{b}\right) \le \frac{\varepsilon}{p},$$

if we take $b = \lfloor (1/2) \cdot (\theta^2/\alpha) \cdot p/(\log(p/\varepsilon)) \rfloor$. Thus, with probability at least $1 - \varepsilon$, all $b \le p$ bins have at least $(\alpha - \theta) \cdot p/b \ge (\alpha - \theta) \cdot p'$ honest players, so the output sequence has at least an $\alpha - \theta$ fraction of honest players. We conclude by noting that

$$p' \le \frac{p}{b} = O\left(\frac{1-\beta}{\theta^2} \cdot \log \frac{p}{\varepsilon}\right).$$

■

The starting point for our universe reduction protocol is the simple protocol of Lemma 3.1. That protocol has the property that a β fraction of cheating players cannot make any outcome in $[n]$ appear with probability more than $1/n^{1-\beta}$. (This can be seen by taking $\mu = 1/n$.) Thus the output can be viewed as a source with min-entropy rate at least $1 - \beta$. To get a higher quality output, it is natural to try applying a *randomness extractor*, a function that extracts almost-uniform bits from sources with sufficient min-entropy. However, randomness extractors require an additional random *seed* to do such extraction. Thus we will enumerate over all seeds of the extractor, and the resulting sequence will be the output of our universe reduction protocol. Fortunately, there exist extractors where the number of seeds is only polylogarithmic in n.

Actually, it is more convenient for us to work with an object that is essentially equivalent to extractors, namely (averaging) samplers (cf., [BR94, Zuc97, Gol97]). Samplers are functions that output sample points of a given universe, with the property that the fraction of samples from any particular subset of the universe is roughly equal to the density of that subset. In the following definition, U_r denotes an element of $[r]$ chosen uniformly at random.

Definition 4.5 *A function* $\mathrm{Samp} : [r] \to [n]^t$ *is a* (θ, ε) sampler *if for every set* $S \subseteq [n]$*,*

$$\Pr_{(i_1,\ldots,i_t) \leftarrow \mathrm{Samp}(U_r)} \left[\frac{\#\{j : i_j \in S\}}{t} > \frac{|S|}{n} + \theta \right] \le \varepsilon.$$

We say that Samp *is* explicit *if for every* $x \in [r]$ *and every* $i \in [t]$*, the* i*'th component of* $\mathrm{Samp}(x)$ *can be computed in time* $\mathrm{poly}(\log r, \log n)$*.*

Zuckerman [Zuc97] showed that samplers (as defined above) are essentially equivalent to randomness extractors. We sketch this connection in Appendix B.

Given $p, \ell \in \mathbb{N}$ and a sampler $\mathrm{Samp} : [r] \to [n]^{n'}$ with $r = \ell^p$, we obtain a $[(p, n) \mapsto n']$-universe reduction protocol Π_{Samp} as follows: the players use the protocol of Lemma 3.1 to select an element $x \in [\ell^p]$, and then output the sequence $\mathrm{Samp}(x)$.

Lemma 4.6 *If* Samp *is a* (θ, ε) *averaging sampler, then for every* $\mu, \beta > 0$*,* Π_{Samp} *is* $[(\beta, \mu) \overset{\gamma}{\mapsto} \mu + \theta]$*-resilient for* $\gamma = 1 - r^\beta \cdot \varepsilon$*. Moreover, the randomness complexity is* $\log r$*.*

Proof: Call $x \in [r]$ "bad" if $\#\{j : i_j \in S\}/t > |S|/n + \theta$ when $(i_1, \ldots, i_t) \leftarrow \mathrm{Samp}(x)$, and note that the number of bad x's is at most $\varepsilon \cdot r$ by the properties of the sampler. The players use the protocol of Lemma 3.1 to select an element x from a universe of size r, where the fraction of bad elements is ε. This is a (p, r)-selection protocol that is $(\beta, \varepsilon, r^\beta \cdot \varepsilon)$-resilient, and so the probability of selecting a good x is at least $\gamma = 1 - r^\beta \cdot \varepsilon$. If a good x is selected, then the fraction of bad elements is increased by at most θ. ■

Notice that for this to be useful, we need the error probability ε of the sampler to be smaller than $r^{-\beta}$, and in fact we will be interested in β that are arbitrarily close to 1. Fortunately, we have samplers that achieve this. (This is equivalent to the fact that we have extractors that work for min-entropy rate arbitrarily close to 0.)

Lemma 4.7 (nonconstructive samplers [RT00, Zuc97]) *For every $n \in \mathbb{N}, \theta > 0, \varepsilon > 0$ and $r \geq n/\varepsilon$, there exists a (θ, ε) sampler* $\mathrm{Samp} : [r] \to [n]^t$ *with $t = O(\log(1/\varepsilon)/\theta^2)$.*

It is important to note that the lower bound on r depends linearly on $1/\varepsilon$; this means that we can make the error $\varepsilon \leq r^{-\beta}$ for any $\beta < 1$. Combining the above two lemmas, we have:

Theorem 4.8 (nonconstructive 1-round universe reduction) *For every $p, n \in \mathbb{N}$, $\beta, \varepsilon, \theta > 0$, there exists a 1-round $[(p, n) \mapsto n']$-universe reduction protocol that is $[(\beta, \mu) \overset{1-\varepsilon}{\mapsto} \mu + \theta]$-resilient for every $\mu > 0$, with*

$$n' = O\left(\frac{\log(1/\varepsilon) + (\beta/(1-\beta)) \cdot (\log n + \log(1/\varepsilon)) + \beta \cdot p}{\theta^2}\right).$$

Moreover, the randomness complexity is $p + (\log n + \log(1/\varepsilon))/(1-\beta) + O(1)$.

Proof: First note that without loss of generality, $\theta \geq 1/n$, otherwise we can use the trivial protocol that outputs the entire universe. So now choose $r \in [(cn/(\varepsilon\theta^2))^{1/(1-\beta)}, 2^p \cdot (cn/(\varepsilon\theta^2))^{1/(1-\beta)}]$ such that r is the p'th power of some natural number, and apply Lemma 4.6 with $\varepsilon' = \varepsilon/r^\beta$. ■

Thus, for $p = \mathrm{polylog}(n)$, $\theta = 1/\mathrm{polylog}(n)$, $\varepsilon = 1/\mathrm{poly}(n)$, and $\beta = 1 - 1/\mathrm{polylog}(n)$, we can reduce the universe size from n to $\mathrm{polylog}(n)$. If the number of players is constant, then we can iterate this $\log^* n$ times to reduce the universe size to a constant. However, if the number of players p is large, then the above will not reduce the universe size below βp. Therefore, we will combine this with the player reduction of Theorem 4.4, via the following composition lemma.

Lemma 4.9 *Given a $[(p, n) \mapsto n']$-universe reduction protocol Π and a $[p \mapsto p']$-player reduction protocol Π', we can construct a $[(p, n) \mapsto (p', n')]$-universe+player reduction protocol Γ such that if Π is $[(\beta, \mu) \overset{\gamma}{\mapsto} \mu']$-resilient and Π' is $[\beta \overset{\gamma'}{\mapsto} \beta']$-resilient, then Γ is $[(\beta, \mu) \overset{\gamma\gamma'}{\mapsto} (\beta', \mu')]$-resilient. If Π and Π' are explicit, then so is Γ. The number of rounds (resp. randomness complexity) in Γ is the maximum of (resp., sum of) the number of rounds (resp., randomness complexities) in Π and Π'.*

Proof: Γ consists of applications of protocols Π and Π' in parallel, giving the bound on the round complexity. The honest players use independent random coins for both protocols, so the randomness complexity is the sum of the randomness complexities of the respective protocols. The probability that Γ succeeds is the probability that both Π and Π' succeed, and since they are independent this is just $\gamma\gamma'$. Finally, the computation time of Γ is the sum of the computation times of Π and Π', so that if the latter protocols are explicit, then so is the former. ■

This yields:

Corollary 4.10 (nonconstructive 1-round universe+player reduction) *For every $n, p \in \mathbb{N}$ and $\beta, \theta, \varepsilon > 0$, there exists a 1-round $[(p, n) \mapsto (p', n')]$-universe+player reduction protocol that is $[(\beta, \mu) \overset{1-\varepsilon}{\mapsto} (\beta + \theta, \mu + \theta)]$-resilient for every $\mu > 0$, with*

$$\begin{aligned} n' &= \mathrm{poly}(\log n, \log(1/\varepsilon), 1/\theta, p) \\ p' &= \mathrm{poly}(\log p, \log(1/\varepsilon), 1/\theta) \end{aligned}$$

Moreover, the randomness complexity is $(\log n + O(\log(1/\varepsilon) + \log(1/\theta) + p))/(1 - \beta) + p \log p$.

Notice that if we want to preserve the density μ of the target set up to a constant factor, then we can set $\theta = 1/\mu$ and the above protocol will reduce the universe size to n' depending polynomially on $1/\mu$. However, to optimize one of our results, it will be beneficial to have a universe size that depends almost-linearly on $1/\mu$. To achieve this, we use a variant of our sampler-based protocol that is tailored to a particular value of μ.

Definition 4.11 *A function* Samp : $[r] \to [n]^t$ *is a* $(\mu, \theta, \varepsilon)$ density-tailored sampler *if for every set $S \subseteq [n]$ with $|S|/n = \mu$,*

$$\Pr_{(i_1,\ldots,i_t) \leftarrow \mathrm{Samp}(U_r)} \left[\frac{\#\{j : i_j \in S\}}{t} > \mu + \theta \right] \leq 1 - \varepsilon.$$

We say that Samp *is* explicit *if for every $x \in [r]$ and every $i \in [t]$, the i'th component of* Samp(x) *can be computed in time* $\mathrm{poly}(\log r, \log n)$.

Density-tailored samplers are essentially equivalent to 'slice extractors,' defined in [RT00]. As in Lemma 4.6, these density-tailored samplers also induce selection protocols.

Lemma 4.12 *If* Samp *is a $(\mu, \theta, \varepsilon)$ density-tailored sampler, then for every $\beta > 0$, Π_{Samp} is $[(\beta, \mu) \overset{\gamma}{\mapsto} \mu + \theta]$-resilient for $\gamma = 1 - r^\beta \cdot \varepsilon$. Moreover the randomness complexity is* $\log r$.

The reason we are interested in these density-tailored samplers is that they exist with slightly better parameters for certain values of μ.

Lemma 4.13 (nonconstructive density-tailored samplers [Vad04]) *There is a universal constant c such that for every $n \in \mathbb{N}, \mu > 0, \theta > 0, \varepsilon > 0$, $t \geq c \cdot \log(1/\varepsilon) \cdot \max\{1/\mu, \mu/\theta^2\}$, and $r \geq c \cdot n \cdot (\mu \log(1/\mu))/(\varepsilon \log(1/\varepsilon))$, there exists a $(\mu, \theta, \varepsilon)$ density-tailored sampler* Samp : $[r] \to [n]^t$.

Note that the number of samples t in these samplers depends linearly on $1/\mu$ (if $\theta = \Omega(\mu)$) and not polynomially as in Lemma 4.7. Combining the above lemma with Lemma 4.6 we get a nonconstructive 1-round universe reduction protocol with different parameters from those of Theorem 4.8:

Theorem 4.14 (nonconstructive, density-tailored 1-round universe reduction) *There is a universal constant c such that for every $p, n \in \mathbb{N}$, $\beta, \mu, \varepsilon, \theta > 0$ and every*

$$n' \geq c \cdot \max\left\{\frac{\mu}{\theta^2}, \frac{1}{\theta}\right\} \cdot \left(\log \frac{1}{\varepsilon} + \frac{\beta}{1 - \beta} \cdot \left(\log n + \log \frac{1}{\beta}\right) + \beta \cdot p\right),$$

there exists a 1-round $[(p, n) \mapsto n']$-universe reduction protocol that is $[(\beta, \mu) \overset{1-\varepsilon}{\mapsto} \mu + \theta]$-resilient. Moreover, the randomness complexity is $p + (\log n + \log(1/\varepsilon) - \log\log(1/\varepsilon) - \log(1/\mu) + \log\log(1/\mu) + \log(1/\beta))/(1 - \beta) + O(1)$.

Proof: We can choose $r \in [r', 2^p \cdot r']$ such that r is the p'th power of some natural number and

$$r' = \left(\frac{cn \cdot \mu \log \frac{1}{\mu}}{\beta \cdot \varepsilon \log \frac{1}{\varepsilon}} \right)^{\frac{1}{1-\beta}}.$$

We then apply Lemma 4.12 with $\varepsilon' = \varepsilon / r^{\beta}$. ■

4.3 Iteration

Now we iterate our 1-round protocols to reduce both the number of players and the size of the universe to a constant. The iteration method is given by the following lemma.

Lemma 4.15 *Given a $[(p, n) \mapsto (p', n')]$-universe+player reduction protocol Π and a $[(p', n') \mapsto (p'', n'')]$-universe+player reduction protocol Π', we can construct a $[(p, n) \mapsto (p'', n'')]$-universe+player reduction protocol $\Pi \circ \Pi'$ such that if Π is $[(\beta, \mu) \stackrel{\gamma}{\mapsto} (\beta', \mu')]$-resilient and Π' is $[(\beta', \mu') \stackrel{\gamma'}{\mapsto} (\beta'', \mu'')]$-resilient, then $\Pi \circ \Pi'$ is $[(\beta, \mu) \stackrel{\gamma\gamma'}{\mapsto} (\beta'', \mu'')]$-resilient. If Π and Π' are explicit, then so is $\Pi \circ \Pi'$, and the number of rounds (resp., randomness complexity) in $\Pi \circ \Pi'$ is the sum of the number of rounds (resp., randomness complexities) in Π and Π'.*

Proof: We construct the protocol $\Pi \circ \Pi'$ as follows: first the initial p players execute protocol Π on the universe of size n, resulting in a collection of p' players and a universe of size n'. Now the selected p' players run protocol Π' on the selected universe of size n'. The output is a collection of p'' players and a universe of size n''.

Now suppose that at most a β fraction of the players are cheating, and S is a subset of $[n]$ of density at most μ. If Π is $[(\beta, \mu) \stackrel{\gamma}{\mapsto} (\beta', \mu')]$-resilient then with probability at least γ, the resulting collection of players contains at most a β' fraction of cheating players, and the resulting universe contains at most a μ' fraction of strings from S. If in addition Π' is $[(\beta', \mu') \stackrel{\gamma'}{\mapsto} (\beta'', \mu'')]$-resilient, then assuming Π was successful (which occurs with probability at least γ), applying Π' on the resulting collections yields the following with probability at least γ': a collection of p'' players of which at most a β'' fraction is cheating, and a universe of size n'' of which at most μ'' belong to S. Thus, with probability at least $\gamma\gamma'$ both protocols are successful, and so $\Pi \circ \Pi'$ is $[(\beta, \mu) \stackrel{\gamma\gamma'}{\mapsto} (\beta'', \mu'')]$-resilient.

The computation time of $\Pi \circ \Pi'$ is the sum of the computation times of Π and Π', and so if the latter two are explicit, then so is the former. Finally, since $\Pi \circ \Pi'$ is the sequential application of Π and Π', both the randomness complexity and the number of rounds are simply the sums of the respective quantities in Π and Π'. ■

Using this composition lemma, we can now construct a many-round protocol that reduces the universe size and the number of players. This protocol will be a main component of our final protocols.

Theorem 4.16 (many-round universe+player reduction) *For every $n, p \in \mathbb{N}$ and every $\beta, \theta, \varepsilon > 0$, there exists a $[(p, n) \mapsto (p', n')]$-universe+player reduction protocol that is $[(\beta, \mu) \stackrel{1-\varepsilon}{\mapsto} (\beta+\theta, \mu+\theta)]$-resilient for every $\mu > 0$, with*

$$\begin{aligned} n' &= \mathrm{poly}(\log(1/\varepsilon), 1/\theta) \\ p' &= \mathrm{poly}(\log(1/\varepsilon), 1/\theta). \end{aligned}$$

Moreover, the number of rounds is $t = \max\{\log^* n, \log^* p\} - \log^* n' + O(1)$ *and the randomness complexity is* $(\log n + o(\log n) + O(t(\log(1/\varepsilon) + \log(1/\theta))))/(1-\beta) + O(p \log p + p/(1-\beta))$.

Proof: We iteratively apply the nonconstructive 1-round universe+player reduction of Corollary 4.10 using the composition of Lemma 4.15. Specifically, we let $p_1 = p$, $n_1 = n$, $\beta_1 = \beta$, $\mu_1 = \mu$ and for $i = 1, \dots, t$ (for t to be determined below), and we let Π_i be a $[(p_i, n_i) \mapsto (p_{i+1}, n_{i+1})]$-universe+player reduction protocol that is $[(\beta_i, \mu) \overset{1-\varepsilon_i}{\mapsto} (\beta_i + \theta_i, \mu_i + \theta_i)]$-resilient obtained from Corollary 4.10 for appropriate choices of the parameters. Specifically, Corollary 4.10 allows us to take

$$\begin{aligned}
\varepsilon_i &= \varepsilon / \max\{\log n_{i-1}, p_{i-1}\} \\
\theta_i &= \theta / \max\{\log \log n_{i-1}, \log p_{i-1}\} \\
\beta_i &= \beta_{i-1} + \theta_{i-1} \\
\mu_i &= \mu_{i-1} + \theta_{i-1} \\
p_i &= \mathrm{poly}(\log p_{i-1}, \log(1/\varepsilon_i), 1/\theta_i) = \mathrm{poly}(\log p_{i-1}, \log \log n_{i-1}, \log(1/\varepsilon), 1/\theta) \\
n_i &= \mathrm{poly}(\log n_{i-1}, \log(1/\varepsilon_i), 1/\theta_i, p_{i-1}) = \mathrm{poly}(p_{i-1}, \log n_{i-1}, \log(1/\varepsilon), 1/\theta).
\end{aligned}$$

We compose these protocols to get $\Pi_1 \circ \dots \circ \Pi_t$, a $[(p, n) \mapsto (p', n')]$-universe+player reduction protocol. We can choose t so that $p' = p_t = \mathrm{poly}(\log(1/\varepsilon), 1/\theta)$, $n' = n_t = \mathrm{poly}(\log(1/\varepsilon), 1/\theta)$, and $t = \max\{\log^* n, \log^* p\} - \log^* n' + O(1)$. (Specifically, this follows by applying Lemma A.1 to the sequences $a_i = p_i$ and $b_i = n_i$ and function $f(x) = ((\log x) \cdot \log(1/\varepsilon) \cdot 1/\theta)^c$.)

By Lemma 4.15, $\Pi_1 \circ \dots \circ \Pi_t$ is $[(\beta, \mu) \overset{1-\varepsilon'}{\mapsto} (\beta + \theta', \mu + \theta')]$-resilient for all $\beta > 0$, for

$$\theta' = \sum_{i=1}^{t} \theta_i = O(\theta_t) \leq \theta,$$

and similarly $\varepsilon' \leq \varepsilon$.

Note that in round i, the randomness complexity is $(\log n_i + O(\log(1/\varepsilon_i) + \log(1/\theta_i)))/(1-\beta_i) + p_i \log p_i + p_i/(1-\beta_i)$. Thus, the total randomness complexity is $(\log n + o(\log n) + O(t(\log(1/\varepsilon) + \log(1/\theta))))/(1-\beta) + O(p \log p + p/(1-\beta))$. ■

5 Putting It Together

Theorem 5.1 (Thm. 2.6, restated) *For all constants* $k \in \mathbb{N}$, $k > 0$ *and* $\delta > 0$, *there exists a constant* $\varepsilon < 1$ *and a* (p, n)*-selection protocol with the following properties:*

(i) The protocol has $\max(\log^* p, \log^* n) + O(1)$ *rounds.*

(ii) The protocol is $(1 - \alpha, \mu, \varepsilon)$*-resilient for* $\alpha = 1/(k+1) + \delta$ *and* $\mu = 1/k - \delta$.

(iii) The randomness complexity of the protocol is $(\log n)/\alpha + o(\log n) + O(p \log p)$.

Proof: The claimed protocol is the composition of two protocols. Let $\alpha = 1/(k+1) + \delta$, and $\mu = 1/k - \delta$. Let Π_1 be the protocol of Theorem 4.16, with $\theta = \delta/2$ and $\varepsilon_1 > 0$ an arbitrary constant. Then Π_1 is a $[(p, n) \mapsto (p', n')]$-universe+player reduction protocol that is $[(1 - \alpha, \mu) \overset{1-\varepsilon_1}{\mapsto} (1 - \alpha +$

$\theta, \mu+\theta)$]-resilient for every $\mu > 0$, where p' and n' are constants. Moreover, the number of rounds is $r = \max\{\log^* n, \log^* p\} + O(1)$ and the randomness complexity is $(\log n)/\alpha + o(\log n) + O(p \log p)$.

With probability at least ε_1, the fraction of good players output by Π_1 is at least $\alpha' = 1/(k+1) + \delta/2$, and the fraction of bad elements of the universe is at most $\mu' = 1/k - \delta/2$. Since $\lfloor 1/\alpha' \rfloor \leq \lceil 1/\mu' \rceil - 1$, we now can apply Π_2, the protocol of Lemma 3.2. Π_2 is a (p', n')-selection protocol that is $(1-\alpha', \mu', \varepsilon_2)$-resilient for some constant $\varepsilon_2 < 1$. Note that Π_2 consists of one round, and its randomness complexity is constant.

Combining Π_1 with Π_2, we get a (p, n)-selection protocol that is $(1-\alpha, \mu, (1-\varepsilon_1) \cdot \varepsilon_2)$-resilient. The number of rounds is $\max\{\log^* n, \log^* p\} + O(1)$, and the randomness complexity is $(\log n)/\alpha + o(\log n) + O(p \log p)$. ■

Next we prove Theorems 2.8 and 2.9 that optimize the relationship between the density μ of the target set and the error probability ε. First, we present a version where we do not reduce the number of players.

Theorem 5.2 (Thm. 2.8, restated) *For all $\mu, \alpha > 0$, $p, n \in \mathbb{N}$ there exists a (p, n)-selection protocol with the following properties:*

(i) The protocol has $r = \log^* n - \log^*(1/\mu) + O(1)$ rounds.

(ii) The protocol is $(1-\alpha, \mu, \varepsilon)$-resilient for

$$\varepsilon = \mu^\alpha \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu} + \beta p \right)^\beta \cdot 2^{\beta p}.$$

(iii) The randomness complexity is $(\log n + o(\log n) + O(t \log(1/\mu)) + p + \log(1/\beta))/\alpha + O(p \log p)$.

To prove this theorem, we use the protocol of Lemma 3.1 as the final one-round protocol. In order for this to work well, we need to reduce the size of the universe to $n' \approx 1/\mu$. Lemma 3.1 also requires that the universe size n' is larger than the number of players (indeed, n' must be some natural number to the power of p), which results in the bad dependence of ε on p above. However, when the number p of players and the fraction α of honest players are constant, the bound becomes $\varepsilon = \tilde{O}(\mu^\alpha)$, which nearly matches the lower bound of $\varepsilon \geq \mu^\alpha$ proven in [GGL98] (see Theorem 7.3).

Proof: In the proof we often write $\alpha = 1 - \beta$ to denote the fraction of honest players. We can assume without loss of generality that $n \geq 1/\mu$, otherwise we can trivially output an arbitrary element of $[n]$. Our aim will be to reduce the size of the universe to $n_2 \approx 1/\mu$. This is done in two steps:

1. First, we reduce the size of the universe to $n_1 = \text{poly}(1/\mu)$ using the protocol of Theorem 4.16 just as a $[(p, n_1) \mapsto n_2]$-universe reduction protocol (ignoring the player reduction). The parameters of the protocol will be $\varepsilon_1 = \mu$ and $\theta_1 = \mu$, so the protocol is $[(\beta, \mu) \overset{1-\varepsilon_1}{\mapsto} \mu + \theta_1]$-resilient for every $\mu > 0$, with

   $$n_1 \quad = \quad \text{poly}(1/\mu).$$

   The number of rounds of Π_1 is $r = \log^* n - \log^*(1/\mu) + O(1)$, and its randomness complexity is $(\log n + o(\log n) + O(t \log(1/\mu) + p))/\alpha + O(p \log p)$.

2. Let $\mu_1 = \mu + \theta_1 = 2\mu$. We now reduce the size of the universe from n_1 to n_2 using the protocol of Theorem 4.14: a 1-round $[(p, n_1) \mapsto n_2]$-universe reduction protocol that is $[(\beta, \mu_1) \overset{1-\varepsilon_2}{\mapsto} \mu_1 + \theta_2]$-resilient with $\varepsilon_2 = \mu^\alpha$ and $\theta_2 = \mu_1 = 2\mu$. Theorem 4.14 allows us to choose any $n_2 \geq v$, where

$$\begin{aligned} v &= O\left(\frac{1}{\mu_1} \cdot \left(\log \frac{1}{\varepsilon_2} + \frac{\beta}{\alpha} \cdot \left(\log n_1 + \log \frac{1}{\beta}\right) + \beta \cdot p\right)\right) \\ &= O\left(\frac{1}{\mu} \cdot \left(\alpha \log \frac{1}{\mu} + \frac{\beta}{\alpha} \cdot \left(\log \frac{1}{\mu} + \log \frac{1}{\beta}\right) + \beta \cdot p\right)\right) \\ &= O\left(\frac{1}{\mu} \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu} + \beta p\right)\right). \end{aligned}$$

We choose $n_2 \in [v, 2^p \cdot v]$ to be the smallest p'th power of an integer greater than or equal to v. The randomness complexity is $O(\log(1/\mu) + \log(1/\beta))/\alpha$.

Composing the above two protocols as in Lemma 4.15, we are in the following situation: There is a universe of size n_2, out of which a $\mu_2 = \mu + \theta_1 + \theta_2 = O(\mu)$ fraction of the elements are in the target set. Now, since we chose n_2 to be an integer to the power of p, we can use the protocol of Lemma 3.1. This is a (p, n_2)-selection protocol that is $(\beta, \mu_2, \varepsilon_3)$-resilient. Note that

$$\begin{aligned} \varepsilon_3 &= n_2^\beta \cdot \mu_2 \\ &\leq (v \cdot 2^p)^\beta \cdot O(\mu) \\ &= O\left(\frac{1}{\mu} \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu} + \beta p\right) \cdot 2^p\right)^\beta \cdot O(\mu) \\ &= O\left(\mu^\alpha \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu} + \beta p\right)^\beta \cdot 2^{\beta p}\right). \end{aligned}$$

The randomness complexity of this last sub-protocol is $\log n_2 \leq O(p + \log(1/\alpha) + \log(1/\mu))$.

Putting all three parts together, we get a (p, n)-selection protocol that is $(\beta, \mu, \varepsilon)$-resilient for $\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 = O(\varepsilon_3)$. The number of rounds of the protocol is $t = \log^* n - \log^*(1/\mu) + O(1)$, and the randomness complexity is $(\log n + o(\log n) + O(t \log(1/\mu)) + p + \log(1/\beta))/\alpha + O(p \log p)$. ∎

Now we eliminate the bad dependence on p by combining the above with a player reduction, at the price of a slightly worse error probability of $\varepsilon = \mu^{\Omega(\alpha)}$.

Theorem 5.3 (Thm. 2.9, restated) *There is a universal constant c such that for all μ, α such that $\alpha \geq \sqrt{c \log\log(1/\mu)/\log(1/\mu)}$ and all $p, n \in \mathbb{N}$, there exists a (p, n)-selection protocol with the following properties:*

(i) The protocol has $r = \max\{\log^* p, \log^* n\} - \log^*(1/\mu) + O(1)$ rounds.

(ii) The protocol is $(1 - \alpha, \mu, \varepsilon)$-resilient for $\varepsilon = \mu^{\Omega(\alpha)}$.

(iii) The randomness complexity is $(\log n + o(\log n) + p)/\alpha + O(p \log p) + \text{poly}(1/\alpha, \log(1/\mu))$.

Proof: In what follows, we use α and β interchangeably, with $\alpha = 1 - \beta$. The protocol is essentially the same protocol as in Theorem 2.8, except that we precede it with an appropriate player reduction.

1. For the first stage, we use the protocol of Theorem 4.16, used only as a $[p \mapsto p_1]$-player reduction protocol (ignoring the universe reduction). We use parameters $\varepsilon_1 = \mu$ and $\theta_1 = \alpha/3$, so the protocol is $[\beta \overset{1-\varepsilon_1}{\mapsto} \beta + \theta_1]$-resilient for every $\mu > 0$, with

$$p_1 = \mathrm{poly}(1/\alpha, \log(1/\mu)).$$

   The number of rounds of Π_1 is $t = \log^* p - \log^*(1/\mu) + O(1)$, and its randomness complexity is $(\log n + o(\log n) + O(t(\log(1/\mu) + \log(1/\alpha))) + p)/(1-\beta) + O(p \log p)$.

2. Let $\beta_2 = \beta + \theta_1$, $\alpha_2 = 1 - \beta_2 = 2\alpha/3$. Π_2 is the protocol of Theorem 4.4, a $[p_1 \mapsto p_2]$-player reduction protocol that is $[\beta_2 \overset{1-\varepsilon_2}{\mapsto} \beta_2 + \theta_2]$-resilient with $\varepsilon_2 = \mu^{\alpha/c}$ and $\theta_2 = \alpha/3$ for a constant c to be determined later. The resulting number of players is

$$\begin{aligned} p_2 &= O\left(\frac{\alpha_2}{\theta_2^2}\left(\log\frac{1}{\varepsilon_2} + \log p_1\right)\right) \\ &= O\left(\frac{1}{\alpha}\left(\frac{\alpha}{c}\log\frac{1}{\mu} + \log\frac{1}{\alpha} + \log\log\frac{1}{\mu}\right)\right) \\ &= \log\frac{1}{\mu} + O\left(\frac{1}{\alpha}\log\frac{1}{\alpha} + \frac{1}{\alpha}\log\log\frac{1}{\mu}\right), \end{aligned}$$

   for a sufficiently large choice of the constant c. The randomness complexity is $p_1 \log p_1 = \mathrm{poly}(1/\alpha, \log(1/\mu))$.

3. Now we do a universe reduction in the same way as Steps 1 and 2 of the proof of Theorem 2.8, except that we run it using only the p_2 players selected above. Recall that these steps reduce the universe size to n_2, where n_2 is the smallest number that is an integer to the power of p_2 greater than or equal to

$$\begin{aligned} v &= O\left(\frac{1}{\mu}\cdot\left(\frac{1}{\alpha}\cdot\log\frac{1}{\mu} + \beta p_2\right)\right) \\ &= \frac{1}{\mu}\cdot\mathrm{poly}\left(\frac{1}{\alpha}, \log\frac{1}{\mu}\right). \end{aligned}$$

   Now we observe that

$$2^{p_2} = \frac{1}{\mu}\cdot\left(\frac{1}{\alpha}\cdot\log\frac{1}{\mu}\right)^{O(1/\alpha)} \geq v,$$

   so in fact $n_2 = 2^{p_2}$. This leaves us with a universe of size n_2, out of which a $\mu_2 = \mu + \theta_1 + \theta_2 = O(\mu)$ fraction of the elements are in the target set.

4. The protocol now concludes in the same way as the proof of Theorem 2.9, using the protocol

of Lemma 3.1. This is a (p_2, n_2)-selection protocol that is $(\beta_3, \mu_2, \varepsilon_3)$-resilient, where

$$\begin{aligned}
\varepsilon_3 &= n_2^{\beta_3} \cdot \mu_2 \\
&\leq (2^{p_2})^{\beta} \cdot O(\mu) \\
&= \frac{1}{\mu^{\beta_3}} \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu}\right)^{O(\beta_3/\alpha)} \cdot O(\mu) \\
&= \mu^{1-\beta_3} \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu}\right)^{O(1/\alpha)} \\
&= \mu^{\alpha/3} \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu}\right)^{O(1/\alpha)}.
\end{aligned}$$

The randomness complexity of this sub-protocol is $\log n_2 = p_2$.

Putting everything together, we get a (p, n)-selection protocol that is $(\beta, \mu, \varepsilon)$-resilient for

$$\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3 = \mu^{\Omega(\alpha)} \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu}\right)^{O(1/\alpha)} = \mu^{\Omega(\alpha)},$$

when $\alpha \geq \sqrt{c \log\log(1/\mu)/\log(1/\mu)}$ for a large enough constant c.

Finally, in order to save on the round and randomness complexity, note that the player reduction in Step 1 can be done in parallel with the universe reduction in Step 3. So this yields a protocol with $t = \max\{\log^* p, \log^* n\} - \log^*(1/\mu) + O(1)$ rounds, where the randomness complexity is $(\log n + o(\log n) + p)/\alpha + O(p \log p) + \mathrm{poly}(1/\alpha, \log(1/\mu))$.

■

6 Explicit Protocols

We now give explicit versions of the above results. The first is the explicit version of Theorem 2.6.

Theorem 6.1 *For all constants $k \in \mathbb{N}$, $k > 0$, $\gamma > 0$ and $\delta > 0$, there exists a constant $\varepsilon < 1$ and an explicit (p, n)-selection protocol with the following properties:*

(i) The protocol has $\max(\log^* p, \log^* n) + O(1)$ rounds.

(ii) The protocol is $(1 - \alpha, \mu, \varepsilon)$-resilient for $\alpha = 1/(k+1) + \delta$ and $\mu = 1/k - \delta$.

(iii) The randomness complexity of the protocol is $(\log n)^{1+\gamma}/\alpha + O(p \log p)$.

Apart from its explicitness, note that the randomness complexity of the above theorem is now $(\log n)^{1+\gamma}$ for an arbitrarily small constant γ, rather than $(1 + o(1)) \log n$. Intuitively, this occurs because the explicit sampler we use (based on an extractor of [RRV99]) only has randomness complexity within a polynomial factor of $(1 + \gamma)$ of optimal. It is possible to remedy this and obtain a randomness complexity of $(1 + o(1)) \log n$ by using other samplers (e.g. based on the extractors of [RRV02]) for the first few rounds of universe reduction, but this creates some messy constraints on the other parameters, so we omit a formal statement.

We also give explicit versions of Theorem 2.8 and Theorem 2.9.

Theorem 6.2 *For all $\mu, \alpha > 0$, $p, n \in \mathbb{N}$, and constant $\gamma > 0$ there exists a (p,n)-selection protocol with the following properties:*

(i) The protocol has $r = \log^* n - \log^*(1/\mu) + O(1)$ rounds.

(ii) The protocol is $(1-\alpha, \mu, \varepsilon)$-resilient for

$$\varepsilon = \mu^{\alpha} \cdot \left(\frac{1}{\alpha} \cdot \log \frac{1}{\mu} + (1-\alpha)\, p\right)^{1-\alpha} \cdot 2^{(1-\alpha)p}.$$

(iii) The randomness complexity is $((\log n)^{1+\gamma} + O(t \log(1/\mu)) + p + \log(1/(1-\alpha)))/\alpha + O(p \log p)$.

(iv) The protocol is explicit given appropriate samplers of size

$$s = 2^p \cdot \text{poly}\left(\frac{1}{\mu}, \log^{(3)} n\right)^{\frac{1}{\alpha}},$$

which can be obtained probabilistically in time $O(s)$ and deterministically in time $2^{O(s)}$.

Theorem 6.3 *There is a universal constant c such that for all μ, α such that $\alpha \geq \sqrt{c \log\log(1/\mu)/\log(1/\mu)}$, $p, n \in \mathbb{N}$ and constant $\gamma > 0$, there exists a (p,n)-selection protocol with the following properties:*

(i) The protocol has $r = \max\{\log^* p, \log^* n\} - \log^*(1/\mu) + O(1)$ rounds.

(ii) The protocol is $(1-\alpha, \mu, \mu^{\Omega(\alpha)})$-resilient.

(iii) The randomness complexity is $((\log n)^{1+\gamma} + p)/\alpha + O(p \log p) + \text{poly}(1/\alpha, \log(1/\mu))$.

(iv) The protocol is explicit given appropriate samplers of size

$$s = 2^{\text{poly}\left(\frac{1}{\mu}\right) \cdot \log^{(4)} n},$$

which can be obtained probabilistically in time $O(s)$ and deterministically in time $2^{O(s)}$.

Note that the protocols are explicit whenever $s = O(\log\log n)$ (in particular, when μ and α are constants).

6.1 Explicit Reduction Protocols

In order to prove the above theorems, we wish to make the protocol given in Theorem 4.16 explicit. Note that the player reductions are already constructive, and the problem is that there are no known optimal sampler constructions. After a few rounds of the universe reduction, however, we can already use the nonconstructive samplers: because the universe is small (i.e., $\log^{(3)} n$), we can exhaustively search all possible nonconstructive samplers until we find the optimal one (in time polylog(n)).

For the first few rounds of the universe reduction, however, we need to use constructive samplers. The explicit construction we will use is the sampler equivalent of an extractor construction of Raz, Reingold and Vadhan [RRV99] (where the equivalence of extractors and samplers is given by [Zuc97]).

Lemma 6.4 (constructive samplers ([RRV99])) *For every constant* $1 > \delta > 0$ *and* $n \in \mathbb{N}, \varepsilon > 0, \theta > \exp(-l/(\log^* l)^{\log^* l})$ *for* $l = (\log n)^{1+\delta} + \log(1/\varepsilon)$*, and* $r \geq 2^{(\log n)^{1+\delta}}/\varepsilon$*, there exists a* (θ, ε) *sampler* Samp : $[r] \to [n]^t$ *with*

$$t = \text{poly}\left(\frac{1}{\theta}\right) \cdot \left(\log n + \log \frac{1}{\varepsilon}\right)^{O\left(1 + \frac{\log\log\frac{1}{\varepsilon}}{\log\log n}\right)}.$$

As before, this sampler yields a 1-round universe reduction protocol. If we use this sampler in Lemma 4.6, and fix some parameters, we get the following constructive protocol.

Theorem 6.5 (constructive 1-round universe reduction) *For every constant* $1 > \delta > 0$*,* $p, n \in \mathbb{N}$ *with* $p = \text{polylog}(n)$*,* $\beta, \varepsilon > 1/n^{\text{polylog}(n)}, \theta > 1/n^{\text{polylog}(n)}$*, there exists an explicit 1-round* $[(p, n) \mapsto n']$*-universe reduction protocol that is* $[(\beta, \mu) \overset{1-\varepsilon}{\mapsto} \mu + \theta]$*-resilient for every* $\mu > 0$*, with*

$$n' = \text{poly}\left(\log n, \frac{1}{\theta}\right).$$

Moreover, the randomness complexity is $(\log n)^{1+\delta}/(1-\beta) + p + O(\log(1/\varepsilon))/(1-\beta)$.

We now wish to iterate this constructive 1-round universe reduction protocol, along with an appropriate player-reduction protocol, in order to obtain a constructive equivalent of Theorem 4.16. However, the expressions in the constructive many-round universe+player reduction get rather messy, so we instead give two different variants with parameters tailored to our needs. The first will be used in the protocol of Theorem 6.1, and the second will be used in the protocols of Theorems 6.2 and 6.3.

Theorem 6.6 (constructive many-round universe+player reduction for constant parameters) *For every* $n, p \in \mathbb{N}$ *and constants* $\beta, \varepsilon > 0, \theta$ *and* $0 < \delta < 1$*, there exists an explicit* $[(p, n) \mapsto (p', n')]$*-universe+player reduction protocol that is* $[(\beta, \mu) \overset{1-\varepsilon}{\mapsto} (\beta+\theta, \mu+\theta)]$*-resilient for every constant* $\mu > 0$*, with*

$$\begin{aligned} n' &= \text{poly}(\log(1/\varepsilon), 1/\theta) \\ p' &= \text{poly}(\log(1/\varepsilon), 1/\theta). \end{aligned}$$

Moreover, the number of rounds is $t = \max\{\log^* n, \log^* p\} - \log^* n' + O(1)$ *and the randomness complexity is* $(\log n)^{1+\delta}/(1-\beta) + O(p \log p) + O(p + t\log(1/\varepsilon) + t\log(1/\theta))/(1-\beta)$.

Proof: The protocol proceeds in stages. First, we iteratively apply the player reduction protocol of Theorem 4.4 to reduce the number of players to a constant $p' = \text{poly}(\log(1/\varepsilon'), 1/\theta')$, where $\varepsilon' = \varepsilon/4$ and $\theta' = \theta/4$. The iteration is done as in Theorem 4.16, except that we disregard the universe reduction. The fraction of bad players increases from β to at most $\beta' = \beta + \theta/2$.

Second, we reduce the size of the universe to

$$n' = \text{poly}\left(\frac{1}{\theta}, \log^{(3)} n\right) = \text{poly}\left(\log^{(3)} n\right).$$

This is done by iteratively applying the protocol of Theorem 6.5 three times, each time with errors $\varepsilon' = \varepsilon/12$ and $\theta' = \theta/6$. Since μ, θ, ε, and β are all constants, the bound on n' follows.

Note that we are now in the following situation, with probability at least $1 - \varepsilon/2$: there are a constant p' players, out of which at most a $\beta' = \beta + \theta/2$ fraction are adversarial. The universe is of size n', and the fractional size of the bad set is at most $\mu' = \mu + \theta/2$.

We now continue as in the proof of Theorem 4.16, with $\varepsilon' = \varepsilon/2$ and $\theta' = \theta/2$. Recall that in the proof of Theorem 4.16, we used non-constructive samplers. For this protocol, we use exhaustive search to find these optimal samplers of the appropriate size. We need a sampler of size

$$s = 2^{p'} \cdot \text{poly}\left(\log^{(3)} n\right) = \text{poly}(\log^{(3)} n),$$

which can be obtained probabilistically in time $O(s)$ and deterministically in time $2^{O(s)} = o(\log n)$ (which is efficient). Thus, the protocol is explicit. In order to get the desired round complexity, note that we can do the player reduction and the universe reduction in parallel, as in Theorem 4.16. ■

The following theorem will be used in the protocols of Theorems 6.2 and 6.3. We note that the range of parameters for which we prove this theorem and some of the constraints obtained are not optimal, as our emphases here are clarity and readability.

Theorem 6.7 (constructive many-round universe+player reduction for μ parameters) *For every $n, p \in \mathbb{N}$, constant $0 < \delta < 1$, every β, and with $\varepsilon = \theta = \mu$, there exists a $[(p, n) \mapsto (p', n')]$-universe+player reduction protocol that is $[(\beta, \mu) \overset{1-\varepsilon}{\mapsto} (\beta + \theta, \mu + \theta)]$-resilient with*

$$\begin{aligned} n' &= \text{poly}(1/\mu) \\ p' &= \text{poly}(1/\mu). \end{aligned}$$

Moreover, the number of rounds is $t = \max\{\log^* n, \log^* p\} - \log^* n' + O(1)$ and the randomness complexity is $(\log n)^{1+\delta}/(1-\beta) + O(p \log p) + O(p + t\log(1/\varepsilon) + t\log(1/\theta))/(1-\beta)$. Finally, the protocol is explicit given an appropriate sampler of size

$$s = \min\left\{2^p, 2^{\text{poly}\left(\frac{1}{\mu}\right)}\right\} \cdot \text{poly}\left(\frac{1}{\mu}, \log^{(3)} n\right)^{\frac{1}{1-\beta}}.$$

This sampler can be obtained probabilistically in time $O(s)$ and deterministically in time $2^{O(s)}$.

Proof: The protocol proceeds as in Theorem 6.6. If $p > \text{poly}(1/\mu)$,, iteratively apply the player reduction protocol of Theorem 4.4 to reduce the number of players to $p' = \text{poly}(\log(1/\varepsilon'), 1/\theta') = \text{poly}(1/\mu)$, where $\varepsilon' = \mu/4$ and $\theta' = \mu/4$. The fraction of bad players increases from β to at most $\beta' = \beta + \mu/2$. Otherwise, let $p' = p$. We then reduce the size of the universe to

$$n' = \text{poly}\left(\frac{1}{\mu}, \log^{(3)} n\right).$$

Again, this is done by iteratively applying the one-round universe+player reduction three times, as in the proof of Theorem 4.16. Instead of using the protocol of Theorem 4.8, we use the protocol of Theorem 6.5. At each one of the three iterations, we first check that the constraints on p', μ, and the current size of the universe are satisfied. If this is not the case (and we can not apply the universe reduction), we do not do anything. But note that this occurs when μ is very small, and so $\mathrm{poly}(1/\mu)$ is the dominating term in our expression for n'. Thus, the universe is still reduced to size n'.

As in the previous proof, we are now in the following situation, with probability at least $1-\mu/2$: there are p' players, out of which at most a $\beta' = \beta + \mu/2$ fraction are adversarial. The universe is of size n', and the fractional size of the bad set is at most $\mu' = 3\mu/2$.

We now continue as in the proof of Theorem 4.16, with $\varepsilon' = \varepsilon/2$ and $\theta' = \theta/2$. Recall that in the proof of Theorem 4.16, we used non-constructive samplers. For this protocol, we use exhaustive search to find these optimal samplers of the appropriate size. We need a sampler of size

$$s = \min\left\{2^p, 2^{\mathrm{poly}\left(\frac{1}{\mu}\right)}\right\} \cdot \mathrm{poly}\left(\frac{1}{\mu}, \log^{(3)} n\right)^{\frac{1}{1-\beta}},$$

which can be obtained probabilistically in time $O(s)$ and deterministically in time $2^{O(s)}$. ■

6.2 Putting It Together

The proof of Theorem 6.1 follows exactly the same proof as that of Theorem 2.6, except that it uses the protocol of Theorem 6.6 rather than that of Theorem 4.16.

The explicit versions of Theorem 2.8 and Theorem 2.9 are not as immediate as that of Theorem 2.6. Both of the former protocols rely critically on a density-tailored sampler that attains optimal parameters, and it is not known how to explicitly construct such samplers. Thus, in both protocols we have to exhaustively search over all samplers of the correct size to find the optimal one. Hopefully, by the time the sampler is needed, the universe size and number of players are small enough that such an exhaustive search is affordable.

We now prove Theorem 6.2.

Proof: This is the same proof as that of Theorem 2.8, with two differences. Instead of using the protocol of Theorem 4.16 in step 1., use the protocol of Theorem 6.7. Second, in step 2., exhaustively search over all possible density-tailored samplers for a universe of size $\mathrm{poly}(1/\mu)$ to find the optimal. Note that with our parameters, the construction times for the protocol Theorem 6.7 is larger than the exhaustive search over density-tailored samplers, and so the total construction time (up to constant factors) is the former. ■

The proof of Theorem 6.3 is the same as that of Theorem 2.9, with the same differences as in the proof of Theorem 6.2.

7 Lower Bounds

In this section we state known lower bounds for different parameters of random selection, and how they relate to our protocols.

7.1 Round Complexity

Theorem 7.1 ([SV05]) *For any $(2, n)$-selection protocol that is $(1/2, \mu, \varepsilon)$-resilient for constants $\mu > 0$ and $\varepsilon < 1$, the round complexity is at least* $(\log^* n - \log^* \log^* n - O(1))/2$.

A corollary of this theorem for the case of many parties is the following:

Corollary 7.2 *For any (p, n)-selection protocol that is $(\beta, \mu, \varepsilon)$-resilient with $\beta \geq 1/2$ and for constants $\mu > 0$ and $\varepsilon < 1$, the randomness complexity is at least* $(\log^* n - \log^* \log^* n - O(1))/2$.

Proof: This is a simple reduction from Theorem 7.1. Given any (p, n)-selection protocol Π that is $(\beta, \mu, \varepsilon)$-resilient with $\beta \geq 1/2$, construct a $(1/2, \mu, \varepsilon)$-resilient $(2, n)$-selection protocol by having each of the two players simulate $p/2$ players in the protocol Π. ■

The round complexity of our protocols is $\max\{\log^* n, \log^* p\} + O(1)$. For $p \leq n$, this is optimal up to a factor of 2 (for the case $\beta \geq 1/2$), by the above corollary. It is not known whether $\log^* p$ rounds are necessary. Indeed, for $\beta < 1/2$, ruling out even 1-round protocols for collective coin-flipping (i.e., random selection when $n = 2$) or leader election is a long-standing open problem. (It is known that $\log^* p$ rounds are necessary in case the players send only one bit per round [RSZ02].)

7.2 Error Probability

Theorem 7.3 ([GGL98]) *For any (p, n)-selection protocol that is $(1 - \alpha, \mu, \varepsilon)$-resilient, $\varepsilon \geq \mu^\alpha$.*

Our Theorem 2.8 achieves a nearly matching bound of $\varepsilon = \tilde{O}(\mu^\alpha)$ in case the number of players is constant, $\alpha = \Omega(1)$ and $\mu = o(1)$. For a nonconstant number of players, Theorem 2.9 achieves $\varepsilon = \mu^{\Omega(\alpha)}$, which is tight up to a constant factor in the exponent.

7.3 μ versus α

Theorem 7.4 ([Fei99], Thm. 4) *Suppose that $\alpha, \mu > 0$ satisfy $\lfloor 1/\alpha \rfloor > \lceil 1/\mu \rceil - 1$. Then in any (p, n)-selection protocol, there exists a set $H \subseteq [p]$ of at least αp players and a set $T \subseteq [n]$ of density at least $1 - \mu$ such that no matter what deterministic strategies the players in H play, the players in $[p] \setminus H$ have a strategy to force the outcome to land outside of T.*

Proof: Suppose that $\lfloor 1/\alpha \rfloor > \lceil 1/\mu \rceil - 1$. Assume towards a contradiction that there is some (p, n)-selection protocol Π such that for all $H \subseteq [p]$ of at least αp players and for all $T \subseteq [n]$ of density at least $1 - \mu$, the players in H have a deterministic strategy to force the outcome into T.

Since the subsets H have density at least αp, there exist $h = \lfloor 1/\alpha \rfloor$ such subsets that are mutually disjoint. Call them $H_1, \ldots, H_h$. Since the sets T have density at least $1 - \mu$, there exist $t = \lceil 1/\mu \rceil$ such subsets that have no common intersection, say $T_1, \ldots, T_t$.

Π guarantees that for every $i \in \{1, \ldots, t\}$, there exists a strategy for the players in H_i to force the output into T_i. This is impossible, however, since there are no elements in the intersection of the T_i's. ■

We use this to deduce that the tradeoff achieved in our Theorem 2.6 between the fraction α of honest players and the density μ of the target set is nearly optimal.

Corollary 7.5 *For any (p, n)-selection protocol that is $(1 - \alpha, \mu, \varepsilon)$-resilient with $\varepsilon < 1$, $\lfloor 1/\alpha \rfloor \leq \lceil 1/\mu \rceil - 1$.*

Proof: Let Π be a (p, n)-selection protocol that is $(1 - \alpha, \mu, \varepsilon)$-resilient with $\varepsilon < 1$. Suppose, for sake of contradiction, that $\lfloor 1/\alpha \rfloor > \lceil 1/\mu \rceil - 1$, and let H and T be the set of players and target set guaranteed by Theorem 7.4. Now we view Π as inducing a game between two players A and B, where A controls the players in H and B controls the players in $[p] \setminus H$, and A wins if the outcome is in T and B wins if the outcome is in $[n] \setminus T$. A basic result in game theory [NM44] says that in every finite, full-information two-player game such as this, one of the players has a winning strategy, i.e., a strategy that wins regardless of how the other player plays. Theorem 7.4 says that A does not have a winning strategy in this game. Thus, B must have a winning strategy. That is, if the players outside H cheat, they can force the outcome to land in $\overline{T}$, regardless of how the players in H play (in particular, if they play honestly). Since there are at most $(1 - \alpha)p$ players outside H, and $\overline{T}$ has density at most μ, this contradicts the fact that Π is $(1 - \alpha, \mu, \varepsilon)$-resilient with $\varepsilon < 1$. ■

7.4 Randomness and Communication Complexity

Theorem 7.6 *For any (p, n)-selection protocol that is $(1 - \alpha', \mu, \varepsilon)$-resilient for $\varepsilon < 1$, the randomness and communication complexities are at least*

$$\max\left\{(1 - \alpha')p, \frac{1 - \varepsilon}{\alpha} \log \frac{\mu n}{\varepsilon}\right\},$$

where $\alpha = \lfloor \alpha' p \rfloor / p$.

To prove this theorem, we first need a definition:

Definition 7.7 (entropy) *The* entropy *of a discrete distribution X, written as $H(X)$, is defined as*

$$H(X) = -\sum_{x \in supp(X)} \Pr[X = x] \cdot \log \Pr[X = x].$$

We now prove Theorem 7.6.

Proof: We first show that a randomness complexity of $(1 - \varepsilon)/(\alpha) \log(\mu n/\varepsilon)$ is necessary. Suppose not, and there exists some protocol Π that is $(1 - \alpha', \mu, \varepsilon)$-resilient for $\varepsilon < 1$, but with randomness complexity $r < (1 - \varepsilon)/(\alpha) \log(\mu n/\varepsilon)$. Each random choice in the protocol can be viewed as a random variable, so let C_i be a random variable denoting the i'th random choice. We wish to calculate the entropy provided by the random choices, $H(C_1, C_2, \ldots)$. Note that the entropy is maximal when all choices are made uniformly at random, which is precisely the randomness complexity. Hence, $H(C_1, C_2, \ldots) \leq r$.

What we would like to say now is that if we pick a $1 - \alpha$ fraction of the players and fix their random choices, then the entropy of the choices made by the other players is at most αr. This is not necessarily true, however, as the number of choices made by the players may also be a random variable. However, we argue that on expectation the entropy is at most αr.

By the chain rule for entropy, $H(C_1, C_2, \ldots) = H(C_1) + H(C_2|C_1) + H(C_3|C_1, C_2) + \ldots \leq r$. Let Q be a subset of the players of size αp. Let C_Q be the set of choices made by players in Q, and note that even for fixed Q, C_Q is a random variable whose value is set by choices made during the protocol. In particular, for every i, choices $C_1, \ldots, C_i$ decide whether or not $C_{i+1} \in C_Q$. Define the random variables

$$J_i^Q = \begin{cases} H(C_i|C_1,\ldots,C_{i-1}) & \text{if } C_i \in C_Q \text{ given } C_1,\ldots,C_{i-1}, \\ 0 & \text{otherwise.} \end{cases}$$

for every i and Q, and

$$J_Q(C_1, C_2, \ldots) = \sum_i J_i^Q.$$

J_Q is a random variable whose value is the entropy provided by the random choices made by players in the set Q, where $\mathrm{E}[J_Q(C_1, C_2, \ldots)] \le \alpha r$. By averaging, there exists some set Q^* such that $J_{Q^*}(C_1, C_2, \ldots) \le \alpha r$.

We now describe an adversary who forces the total entropy of all choices in a run of the protocol to be at most αr. First, the adversary corrupts the players not in Q^*. Then, in the i'th choice made in the protocol, one of two scenarios are possible:

1. The i'th choice may belong to a player in Q^*. That is, $C_i \in C_{Q^*}$. In this case, an honest player makes the choice.

2. Conversely, the i'th choice may belong to a player not in Q^*, a corrupted player. Suppose the $i-1$ choices made thus far have values $C_j = c_j$, and that the amount of entropy contributed is $h = J_1^{Q^*} + \ldots + J_{i-1}^{Q^*}$. Then by averaging, there exists some c_i such that $h + \sum_{j>i}(J_j^{Q^*}|C_i = c_i) \le \alpha r$. So in this case, the adversary (deterministically) fixes his choice C_j to c_j.

With this adversary, the total amount of entropy going into the protocol is at most αr. How much entropy is output by the protocol? Recall that we assumed that Π is a (p, n)-selection protocol that is $(1-\alpha', \mu, \varepsilon)$-resilient for $\varepsilon < 1$, and let X be the output distribution of Π. We need the following claim:

Claim 7.8

$$H(X) \ge (1-\varepsilon)\log\frac{\mu n}{\varepsilon}.$$

Proof: Let $S \subset [n]$, $|S| = \mu n$ have maximal weight over all sets of size μn. Note that the resilience of Π guarantees that $\Pr[X \in S] \le \varepsilon$. Then for all $x \notin S$, $\Pr[X = x] \le \varepsilon/(\mu n)$. This is due to the fact that there exists some $y \in S$ such that $\Pr[X = y] \le \varepsilon/(\mu n)$ (since the total probability is ε, and there are μn elements). Thus, if there were some $x \notin S$ such that $\Pr[X = x] > \varepsilon/(\mu n)$, we could define $S' = S \cup \{x\} \setminus \{y\}$. But $|S'| = \mu n$ and $\Pr[X \in S'] > \Pr[X \in S]$, contradicting the maximality of S. Thus, for all $x \notin S$, $\Pr[X = x] \le \varepsilon/(\mu n)$.

We can now compute the entropy of X.

$$\begin{aligned} H(X) &= -\sum_{x \in supp(X)} \Pr[X = x]\log\Pr[X = x] \\ &\ge -\sum_{x \notin S} \Pr[X = x]\log\Pr[X = x] \\ &\ge \log\frac{\mu n}{\varepsilon}\sum_{x \notin S}\Pr[X = x] \\ &\ge (1-\varepsilon)\log\frac{\mu n}{\varepsilon}. \end{aligned}$$

■

It must be the case that $H(C_1, C_2, \ldots) \geq H(X)$, and so

$$r \geq \frac{1-\varepsilon}{\alpha} \log \frac{\mu n}{\varepsilon}.$$

Now we show that a randomness complexity of $(1-\alpha')p$ is necessary. Again suppose this is not the case, and there exists some protocol Π that is $(1-\alpha', \mu, \varepsilon)$-resilient for $\varepsilon < 1$, but with randomness complexity $r < (1-\alpha)p$. Consider some run of the protocol Π, and suppose that each random choice C_j was set to c_j. Since there are at most r C_j's, the number of players that actually made a choice in the run of the protocol is at most r. Now consider the adversary that corrupts these r players, and sets their random choices to c_j. Running the protocol with this adversary yields a fixed output, since there are no random choices made. Clearly, such a protocol can not be $(r/p, \mu, \varepsilon)$-resilient for $\varepsilon < 1$, but the resilience guarantees that it is $(1-\alpha', \mu, \varepsilon)$-resilient. Thus, $r \geq (1-\alpha')p$.

Finally, note that the communication complexity must be at least as large as the minimal randomness complexity. This is because the output of the protocol is a function of the messages communicated, not of the randomness, and the entropy of the messages communicated is at most the total length of the messages. Thus, for any protocol, its randomness is at most the length of the messages and at least the minimal randomness complexity. ∎

The randomness complexity of the protocol given in Theorem 2.6 is $(\log n/\alpha) + o(\log n) + O(p \log p)$. In terms of the dependence on n, this is optimal up to lower-order terms. The dependence on p is within a logarithmic factor. It would be interesting to improve the randomness complexity to p, or to show that $p \log p$ is necessary.

Acknowledgements

We would like to thank Omer Reingold for many useful discussions.

References

[AN93] Noga Alon and Moni Naor. Coin-flipping games immune against linear-sized coalitions. *SIAM Journal of Computing*, 22(2):403-417, April 1993.

[Bar02] Boaz Barak. Constant-round coin-tossing with a man in the middle or realizing the shared random string model . In *43rd FOCS*, 2002.

[Blu82] Manuel Blum. Coin flipping by telephone. In *IEEE Spring COMPCOM*, 1982.

[BG89] Donald Beaver and Shafi Goldwasser. Multiparty computation with faulty majority. In *Advances in Cryptology – CRYPTO 89*, volume 435 of *Lecture Notes in Computer Science*, pages 589-590. IACR, Springer-Verlag, August 1989.

[BGW88] Michael Ben-Or, Shafi Goldwasser and Avi Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation. In 20th STOC, pages 1-10. ACM Press, 1988.

[BL89] Michael Ben-Or and Nati Linial. Collective coin fliping. In *Advances in Computing Research*, volume 5: Randomness and Computation, JAI Press, Greenwich, CT, 1989, 91-115.

[BN00] Ravi B. Boppana and Babu O. Narayanan. Perfect-information leader election with optimal resilience. *SIAM Journal of Computing.* 29(4): 1304-1320 (2000).

[BR94] Mihir Bellare and John Rompel. Randomness-efficient oblivious sampling. In *35th FOCS*, 1994.

[CCD88] David Chaum, Claude Cropeau, and Ivan Damgard. Multiparty unconditionally secure protocols. In *20th STOC*, pages 11-19. ACM Press, 1988.

[CL95] Jason Cooper and Nathan Linial. Fast perfect-information leader-election protocols with linear immunity. *Combinatorica*, 15:319-332, 1995.

[Dam93] Ivan B. Damgård. Interactive hashing can simplify zero-knowledge protocol design without computational assumptions (extended abstract). In *CRYPTO*, 1993.

[DGW94] Ivan B. Damgård, Oded Goldreich, and Ave Wigderson. Hashing functions can simplify zero-knowledge protocol design (too). TR RS-94-39. BRICS, 1994.

[DHRS04] Yan Zong Ding, Danny Harnik, Alon Rosen, and Ronen Shaltiel. Constant-round oblivious transfer in the bounded storage model. In *1st TCC*, 2004.

[Fei99] Uriel Feige. Noncryptographic selection protocols. Proceedings of *40th Annual Symposium on Foundations of Computer Science*, pages 142-152, 1999. IEEE.

[GGL98] Oded Goldreich, Shafi Goldwasser, Nathan Linial. Fault-tolerant computation in the full information model. *SIAM J. Computing* 27(2), 1998. IEEE.

[Gol97] Oded Goldreich. A sample of samplers - a computational perspective on sampling (survey). In Electronic Colloquium on Computational Complexity (ECCC) (20), volume 4, 1997.

[GL90] Shafi Goldwasser and Leonid A. Levin. Fair computation of general functions in presence of immoral majority. In *Advances in Cryptology – CRYPTO 90*, volume 537 of *Lecture Notes in Computer Science*, pages 77-93. Spriger-Verlag, August 1990.

[GL02] Shafi Goldwasser and Yehuda Lindell. Secure computation without agreement. In *DISC* 2002: 17-32

[GMW87] Oded Goldreich, Shafi Micali and Ave Wigderson. How to play ANY mental game. In Proceedings of the nineteenth annual ACM conference on Theory of computing, pages 218-229. ACM Press, 1987.

[GSV98] Oded Goldreich, Amit Sahai, and Salil Vadhan. Honest-verifier statistical zero-knowledge equals general statistical zero-knowledge. In *30th STOC*, 1998.

[KO04] Jonathan Katz and Rafail Ostrovsky. Round-optimal secure two-party computation. In *CRYPTO*, 2004.

[KOS03] Jonathan Katz, Rafail Ostrovsky and Adam Smith: Round efficiency of multi-party computation with a dishonest majority. EUROCRYPT 2003: 578-595.

[Lin01] Yehuda Lindell. Parallel coin-tossing and constant-round secure two-party computation. In *CRYPTO*, 2001.

[LPS80] Marshall Pease, Robert Shostak, and Leslie Lamport. Reaching agreement in the presence of faults. Journal of the ACM, 27(2):228–234, April 1980.

[LRVW03] C. Lu, O. Reingold, S. Vadhan, and A. Wigderson. Extractors: Optimal up to constant factors. In *Proceedings of 35th ACM STOC*, 2003.

[NM44] John von Neumann and Oskar Morgenstern. Theory of games and economic behavior. In Princeton University Press, 1944.

[NOVY98] Moni Naor, Rafail Ostrovsky, Ramarathnam Venkatesan, and Moti Yung. Perfect zero-knowledge arguments for NP can be based on general complexity assumptions. *J. Cryptology* 11, 1998.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 52(1): 43–52, 1996.

[Oka00] Tatsuaki Okamoto. On relationships between statistical zero-knowledge proofs. *Journal of Computer and System Sciences*, 60(1): 47–108, 2000.

[ORV94] Rafail Ostrovsky, Sridhar Rajagopalan, and Umesh Vazirani. Simple and efficient leader election in the full information model. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, pages 234-242, Montréal, Québec, Canada, 23-25 May 1994.

[OVY93] Rafail Ostrovsky, Ramarathnam Venkatesan. and Moti Yung. Interactive hashing simplifies zero-knowledge protocol design. In *Proceedings of Eurocrypt*, 1993.

[RRV02] Ran Raz, Omer Reingold, and Salil Vadhan. Extracting all the randomness and reducing the error in Trevisan's extractors. *Journal of Computer and System Sciences*, 65(1):97–128, August 2002.

[RRV99] Ran Raz, Omer Reingold, and Salil Vadhan. Error Reduction for Extractors. In *40th FOCS*, 1999.

[RSZ02] Alexander Russell, Michael Saks and David Zuckerman. Lower bounds for leader election and collective coin- flipping in the perfect information model. *SIAM Journal on Computing*, 31:1645–1662, 2002.

[RT00] J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13(1):2–24 (electronic), 2000.

[RZ98] Alexander Russell and David Zuckerman. Perfect-information leader election in $\log^* n+O(1)$ rounds. *Journal of Computer and System Sciences*, 63:612–626, 2001.

[Sak89] Michael Saks. A robust noncryptographic protocol for collective coin flipping. *SIAM Journal on Discrete Mathematics,* 2(2):240-244, May 1989.

[SV05] Saurabh Sanghvi and Salil Vadhan. The round complexity of two-party random selection. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, Baltimore, MD, May 2005. ACM.

[Vad04] Salil Vadhan. Constructing locally computable extractors and cryptosystems in the bounded-storage model. *Journal of Cryptology*, Vol. 17, No. 1, pages 43-77, Winter 2004. Special Issue on the Bounded-Storage Model, edited by Oded Goldreich

[WZ99] Avi Wigderson and David Zuckerman, Expanders that beat the eigenvalue bound: explicit construction and applications. Combinatorica, 19 (1999) 125–138. 17

[Yao86] Andrew Yao. How to generate and exchange secrets. In *Proc. 27th FOCS*, 1986.

[Zuc97] David Zuckerman. Randomness-optimal oblivious sampling. Random Structures and Algorithms, 11(4): 345-367 (1997).

A Miscellaneous

Lemma A.1 *Suppose* $a_1, a_2, \ldots$ *and* $b_1, b_2, \ldots$ *are sequences of real numbers and* f *is a monotone non-increasing function such that* $a_i \leq \max\{f(a_{i-1}), f(f(b_{i-1}))\}$ *and* $b_i \leq \max\{a_{i-1}, f(b_{i-1})\}$ *for all* i*. Then* $a_t, b_{t+1} \leq \max\{f^{(t-1)}(a_1), f^{(t-1)}(b_1)\}$ *for all* $t \geq 2$*.*

Proof: The lemma is proved by induction on t. For the case $t = 2$, we are given that

$$a_2 \leq \max\{f(a_1), f(f(b_1))\} \leq \max\{f(a_1), f(b_1)\}$$

by the monotonicity of f. Also,

$$b_3 \leq \max\{a_2, f(b_2)\} \leq \max\{\max\{f(a_1), f(b_1)\}, f(a_1), f(f(b_1))\} \leq \max\{f(a_1), f(b_1)\}.$$

Now assume the lemma holds for all $2 \leq i \leq t-1$. We will prove the lemma for $i = t$.

$$\begin{aligned} a_t &\leq \max\{f(a_{t-1}), f(f(b_{t-1}))\} \\ &\leq \max\{f(f^{(t-2)}(a_1)), f(f^{(t-2)}(b_1)), f(f(f^{(t-3)}(a_1))), f(f(f^{(t-3)}(b_1)))\} \\ &= \max\{f^{(t-1)}(a_1), f^{(t-1)}(b_1)\}. \end{aligned}$$

In addition,

$$\begin{aligned} b_t + 1 &\leq \max\{a_t, f(b_t)\} \\ &\leq \max\{f^{(t-1)}(a_1), f^{(t-1)}(b_1), f(f^{(t-2)}(a_1)), f(f^{(t-1)}(b_1))\} \\ &= \max\{f^{(t-1)}(a_1), f^{(t-1)}(b_1)\} \end{aligned}$$

as claimed. ∎

B Connection Between Extractors and Samplers

In this section we sketch the connection between extractors and samplers, as shown by Zuckerman [Zuc97]. First, we need a couple of definitions.

Definition B.1 (statistical difference) *For two distributions X and Y over some finite domain, denote the* statistical difference *between them by $\Delta(X,Y)$, where:*

$$\Delta(X,Y) = \frac{1}{2} \sum_{i \in supp(X \cup Y)} |\Pr[X=i] - \Pr[Y=i]|.$$

*X and Y are ε-*close *if $\Delta(X,Y) \leq \varepsilon$.*

We also need a measure of the randomness of a distribution.

Definition B.2 (min-entropy) *The* min-entropy *of a distribution X, denoted by $H_\infty(X)$, is defined as*

$$H_\infty(X) = \min_{i \in supp(X)} \log \frac{1}{\Pr[X=i]}.$$

Definition B.3 (extractor) $\mathrm{Ext} : \{0,1\}^\ell \times \{0,1\}^d \mapsto \{0,1\}^m$ *is a (k,ε)-extractor if, for any distribution X with $H_\infty(X) \geq k$, when choosing x according to X and r uniformly at random from $\{0,1\}^d$, the distribution of* $\mathrm{Ext}(x,r)$ *is ε-close to uniform.*

The connection between extractors and samplers is given by two lemmas:

Lemma B.4 *Let* $\mathrm{Ext} : \{0,1\}^\ell \times \{0,1\}^d \mapsto \{0,1\}^m$ *be a (k,ε)-extractor. Then* Ext *is also a $(2^{k-\ell}\varepsilon)$-sampler* $\mathrm{Samp} : [R] \to [N]^T$ *for $R = 2^\ell$, $N = 2^m$, and $T = 2^d$.*

Proof: Suppose we are given a (k,ε)-extractor $\mathrm{Ext} : \{0,1\}^\ell \times \{0,1\}^d \mapsto \{0,1\}^m$. We can view this as a sampler $\mathrm{Samp} : [R] \to [N]^T$ for $R = 2^\ell$, $N = 2^m$, and $T = 2^d$ as follows: on input $x \in [R]$, $\mathrm{Samp}(x) = \{s | s = \mathrm{Ext}(x,i), i \in \{0,1\}^d\}$.

Now suppose that Samp is not an $(\varepsilon, 2^{k-\ell})$-sampler. This means that there exists some $S \subseteq [N]$ such that

$$\Pr_{(i_1,\ldots,i_t) \leftarrow \mathrm{Samp}(U_R)} \left[\frac{\#\{j : i_j \in S\}}{T} > \frac{|S|}{N} + \varepsilon \right] > \frac{2^k}{2^\ell}.$$

Thus, there exists some set of 2^k x's for which

$$\frac{\#\{s | s = \mathrm{Ext}(x,i), i \in \{0,1\}^d\}}{T} > \frac{|S|}{N} + \varepsilon.$$

Denote by X the uniform distribution over these x's, and note that $H_\infty(X) = k$. However, we claim that $\mathrm{Ext}(X, U_{\{0,1\}^d})$ can not be ε-close to uniform. Consider the following statistical test: $f(y) = 1$ if $y \in S$, and $f(y) = 0$ otherwise. Then

$$\left| \Pr\left[f(U_{\{0,1\}^m}) = 1\right] - \Pr\left[\mathrm{Ext}(X, U_{\{0,1\}^d})\right] = 1 \right| > \varepsilon,$$

contradicting the assumption that Ext is a (k,ε)-extractor. ■

Lemma B.5 *Let* $\mathrm{Samp} : [R] \to [N]^T$ *be a $(\varepsilon, 2^{k-\ell})$-sampler, with R, N, and T as above. Then* Samp *is also a $(k + \log(1/\varepsilon), 2\varepsilon)$-extractor* $\mathrm{Ext} : \{0,1\}^\ell \times \{0,1\}^d \mapsto \{0,1\}^m$.

Proof: Given a statistical test $T \subseteq \{0,1\}^m$, let

$$B_T \stackrel{\text{def}}{=} \left\{ x \in \{0,1\}^\ell : \left| \Pr_{U_{\{0,1\}^d}}[\text{Ext}(x, U_{\{0,1\}^d}) \in T] - \frac{|T|}{n} \right| > \varepsilon \right\}.$$

Since Samp is a $(2^k/2^\ell, \varepsilon)$-sampler we know that $\Pr_{U_R}[U_R \in B_T] \leq 2^{k-\ell}$, so $|B_T| \leq 2^k$.

Let X be a $(k + \log(1/\varepsilon))$ source. We have:

$$\Pr_X[X \in B_T] \leq 2^{-(k+\log(1/\varepsilon))} \cdot |B_T| \leq \varepsilon$$

$$\Rightarrow \Pr_{X,U_{\{0,1\}^d}}\left[\text{Ext}(X, U_{\{0,1\}^d}) \in T\right] \leq \Pr_X[X \in B_T] \cdot 1 + \Pr_X[X \notin B_T] \cdot \left(\frac{|T|}{N} + \varepsilon\right) \leq \frac{|T|}{N} + 2\varepsilon$$

$$\Rightarrow \Pr_{X,U_{\{0,1\}^d}}\left[\text{Ext}\left(X, U_{\{0,1\}^d}\right) \in T\right] - \frac{|T|}{N} \leq 2\varepsilon.$$

Since T was an arbitrary statistical test, we can also apply the above inequality to $\overline{T}$ to get $|\Pr_{X,U_{\{0,1\}^d}}[\text{Ext}(X, U_{\{0,1\}^d}) \in T] - |T|/N| \leq 2\varepsilon$. Since T was an arbitrary subset of $\{0,1\}^m$, we have that Ext is a $(k + \log(1/\varepsilon), 2\varepsilon)$-extractor. ■

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'