A. From uniformity to parameterized identity testing

In this appendix, we explain how the existence of any distributed protocol for uniformity testing implies the existence of one for identity testing with roughly the same parameters, and further even implies one for identity testing in the *massively parameterized* sense[6] ("instance-optimal" in the vocabulary of Valiant and Valiant, who introduced it (Valiant & Valiant, 2017)). These two results will be seen as a straightforward consequence of (Goldreich, 2016), which establishes the former reduction in the standard non-distributed setting; and of (Blais et al., 2017), which implies that massively parameterized identity testing reduces to "worst-case" identity testing. Specifically, we show the following:

Proposition A.1. *Suppose that there exists an ℓ-bit protocol π for testing uniformity of k-ary distributions, with number of players $n(k, \ell, \varepsilon)$ and failure probability $1/3$. Then there exists an ℓ-bit protocol π' for testing identity against a fixed k-ary distribution $\mathbf{q}$ (known to all players), with number of players $n(5k, \ell, \frac{16}{25}\varepsilon)$ and failure probability $1/3$.*

Furthermore, this reduction preserves the setting of randomness (i.e., private-coin protocols are mapped to private-coin protocols).

Proof. We rely on the result of Goldreich (Goldreich, 2016), which describes a randomized mapping $F_{\mathbf{q}} \colon \Delta_{[k]} \to \Delta_{[5k]}$ such that $F_{\mathbf{q}}(\mathbf{q}) = \mathbf{u}_{[5k]}$ and $\mathrm{d}_{\mathrm{TV}}\big(F_{\mathbf{q}}(\mathbf{p}), \mathbf{u}_{[5k]}\big) > \frac{16}{25}\varepsilon$ for any $\mathbf{p} \in \Delta_{[k]}$ ε-far from $\mathbf{q}$.[7] In more detail, this mapping proceeds in two stages: the first allows one to assume, at essentially no cost, that the reference distribution $\mathbf{q}$ is "grained," i.e., such that all probabilities $\mathbf{q}(i)$ are a multiple of $1/m$ for some $m = O(k)$. Then, the second mapping transforms a given m-grained distribution to the uniform distribution on an alphabet of slightly larger cardinality. The resulting $F_{\mathbf{q}}$ is the composition of these two mappings.

Moreover, a crucial property of $F_{\mathbf{q}}$ is that, given the knowledge of $\mathbf{q}$, a sample from $F_{\mathbf{q}}(\mathbf{p})$ can be efficiently simulated from a sample from $\mathbf{p}$; this implies the proposition. □

Remark A.2. The result above crucially assumes that every player has explicit knowledge of the reference distribution $\mathbf{q}$ to be tested against, as this knowledge is necessary for them to simulate a sample from $F_{\mathbf{q}}(\mathbf{p})$ given their sample from the unknown $\mathbf{p}$. If only the referee $\mathcal{R}$ is assumed to know $\mathbf{q}$, then the above reduction does not go through, although one can still rely on any testing scheme based on distributed simulation.

The previous reduction enables a distributed test for any identity testing problem using at most, roughly, as many players as that required for distributed uniformity testing. However, we can expect to use fewer players for specific distributions. Indeed, in the standard, non-distributed setting, Valiant and Valiant in (Valiant & Valiant, 2017) introduced a refined analysis termed the instance-optimal setting and showed that the sample complexity of testing identity to $\mathbf{q}$ is essentially captured by the $2/3$-quasinorm of a sub-function of $\mathbf{q}$ obtained as follows: Assuming without loss of generality $\mathbf{q}_1 \geq \mathbf{q}_2 \geq \ldots \mathbf{q}_k \geq 0$, let $t \in [k]$ be the largest integer that $\sum_{i=t+1}^{k} q_i \geq \varepsilon$, and let $\mathbf{q}_\varepsilon = (\mathbf{q}_2, \ldots, \mathbf{q}_t)$ (i.e., removing the largest element and the "tail" of $\mathbf{q}$). The main result in (Valiant & Valiant, 2017) shows that the sample complexity of testing identity to $\mathbf{q}$ is upper and lower bounded by $\max(\|\mathbf{q}_{\varepsilon/16}\|_{2/3}/\varepsilon^2, 1/\varepsilon)$ and $\max(\|\mathbf{q}_{\varepsilon}\|_{2/3}/\varepsilon^2, 1/\varepsilon)$, respectively.

However, it is not clear if the aforementioned reduction between identity and uniformity of Goldreich preserves this parameterization of sample complexity for identity testing; in particular, the $2/3$-quasinorm characterization does not seem to be amenable to the same type of analysis as that underlying Proposition A.1. Interestingly, a different instance-optimal characterization due to Blais, Canonne, and Gur (Blais et al., 2017) admits such a reduction, enabling us to obtain the analogue of Proposition A.1 for this massively parameterized setting.

To state the result as parameterized by $\mathbf{q}$ (instead of k), we will need the following definition of $\Phi(\mathbf{p}, \gamma)$; see Section 6 of (Blais et al., 2017) for a discussion on basic properties of $\Phi(\mathbf{p}, \gamma)$ and how it relates to notions such as the sparsity of $\mathbf{p}$ and the functional $\|\mathbf{p}_{\gamma}^{-\max}\|$ defined in (Valiant & Valiant, 2017). For $a \in \ell_2(\mathbb{N})$ and $t \in (0, \infty)$, let

$$\kappa_a(t) := \inf_{a'+a''=a} \left(\|a'\|_1 + t\|a''\|_2\right)$$

and, for $\mathbf{p} \in \Delta_{\mathbb{N}}$ and any $\gamma \in (0, 1)$, let

$$\Phi(\mathbf{p}, \gamma) := 2\kappa_{\mathbf{p}}^{-1}(1-\gamma)^2\,. \tag{4}$$

It can be seen that, if $\mathbf{p}$ is supported on at most k elements, $\Phi(\mathbf{p}, \gamma) \leq 2k$ for all $\gamma \in (0, 1)$. We are now in a position to state our general reduction.

Proposition A.3. *Suppose that there exists an ℓ-bit protocol π for testing uniformity of k-ary distributions, with number of players $n(k, \ell, \varepsilon)$ and failure probability $1/3$. Then there exists an ℓ-bit protocol π' for testing identity against a fixed*

[6] Massively parameterized setting, a terminology borrowed from property testing, refers here to the fact that the sample complexity depends not only on a single parameter k but a k-ary distribution $\mathbf{q}$.

[7] In (Goldreich, 2016), Goldreich exhibits a randomized mapping that converts the problem from testing identity over domain of size k with proximity parameter ε to testing uniformity over a domain of size $k' := k/\alpha^2$ with proximity parameter $\varepsilon' := (1-\alpha)^2\varepsilon$, for every fixed choice of $\alpha \in (0, 1)$. This mapping further preserves the success probability of the tester. Since the resulting uniformity testing problem has sample complexity $\Theta\left(\sqrt{k'}/\varepsilon'^2\right)$, the blowup factor $1/(\alpha(1-\alpha)^4)$ is minimized by $\alpha = 1/5$.

distribution $\mathbf{p}$ *(known to all players), with number of players* $O\big(n(\Phi(\mathbf{q}, \frac{\varepsilon}{9}), \ell, \frac{\varepsilon}{18}))\big)$ *and failure probability* $2/5$.

Further, this reduction preserves the setting of randomness (i.e., private-coin protocols are mapped to private-coin protocols).

Proof. This strengthening of Proposition A.1 stems from the algorithm for identity testing given in (Blais et al., 2017), which at a high-level reduces testing identity to $\mathbf{q}$ to three tasks: (i) computing the $(\varepsilon/3)$-effective support[8] of $\mathbf{q}$, $S_{\mathbf{q}}(\varepsilon)$, which can be done easily given explicit knowledge of $\mathbf{q}$; (ii) testing that the unknown distribution $\mathbf{p}$ puts mass at most $\varepsilon/2$ outside of $S_{\mathbf{q}}(\varepsilon)$ (which only requires $O(1/\varepsilon)$ players to be done with a high constant probability, say $1/30$); and (iii) testing identity of $\mathbf{p}$ and $\mathbf{q}$ conditioned on $S_{\mathbf{q}}(\varepsilon)$ with parameter $\varepsilon/18$, which can be done using rejection sampling and Proposition A.1 with $O\big(n(|S_{\mathbf{q}}(\varepsilon)|, \ell, \frac{\varepsilon}{18})\big)$ players and success probability, say $2/3 - 1/30$, where the additional $1/30$ error probability comes from rejection sampling. See Fig. 1 for an illustration.

As shown in Section 7.2 of (Blais et al., 2017), we have $|S_{\mathbf{q}}(\varepsilon)| \le \Phi(\mathbf{q}, \frac{\varepsilon}{9})$, and thereby the claimed result, since it follows that the approach above indeed yields an algorithm which is instance-optimal. Technically, the claimed bound is obtained upon recalling that $n(\Phi(\mathbf{q}, \frac{\varepsilon}{9}), \ell, \frac{\varepsilon}{18})) = \Omega(1/\varepsilon)$ using the trivial lower bound of $\Omega(1/\varepsilon)$ on uniformity testing, so that $n(\Phi(\mathbf{q}, \frac{\varepsilon}{9}), \ell, \frac{\varepsilon}{18})) + O(1/\varepsilon) = O\big(n(\Phi(\mathbf{q}, \frac{\varepsilon}{9}), \ell, \frac{\varepsilon}{18})\big)$. □

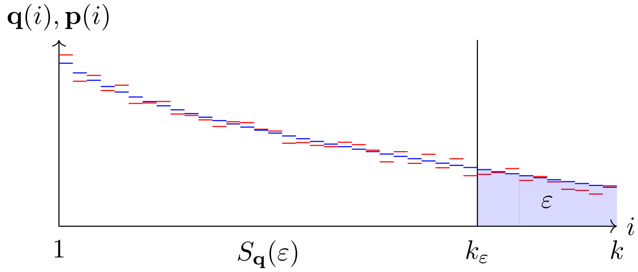

Figure 1. The reference distribution $\mathbf{q}$ (in blue; assumed non-increasing without loss of generality) and the unknown distribution $\mathbf{p}$ (in red). By the reduction above, testing equality of $\mathbf{p}$ to $\mathbf{q}$ is tantamount to (i) determining $S_{\mathbf{q}}(\varepsilon)$, which depends only on $\mathbf{q}$; (ii) testing identity for the conditional distributions of $\mathbf{p}$ and $\mathbf{q}$ given $S_{\mathbf{q}}(\varepsilon)$, and (iii) testing that $\mathbf{p}$ assigns at most $O(\varepsilon)$ probability to the complement of $S_{\mathbf{q}}(\varepsilon)$.

B. Impossibility of perfect simulation when $\ell < \log k$

We begin with a proof of impossibility which shows that any simulation that works for all points in the interior of the $(k-1)$-dimensional probability simplex must fail for a distribution on the boundary. Our main result of this section is the following:

Theorem B.1. *For any* $n \ge 1$, *there exists no* ℓ*-bit public-coin perfect simulation of* k*-ary distributions using* n *players unless* $\ell \ge \log k$.

Proof. Let $\mathcal{S} = (\pi, \delta)$ be an ℓ-bit perfect simulation for k-ary distributions using n players. Suppose that $\ell < \log k$. We show a contradiction for any such public-coin simulation $\mathcal{S}$. Fix a realization $U = u$ of the public randomness. By the pigeonhole principle we can find a message vector $m = (m_1, \ldots, m_n)$ and distinct elements $x_i, x_i' \in [k]$ for each $i \in [n]$ such that

$$\pi_i(x_i, u) = \pi_i(x_i', u) = m_i.$$

Note that the probability of declaring $\perp$ for a public-coin simulation must be 0 for every k-ary distribution. Therefore, since the message m occurs with a positive probability under a distribution $\mathbf{p}$ with $\mathbf{p}_{x_i} > 0$ for all i, the referee must declare an output $x \in [k]$ with positive probability when it receives m, i.e., there exists $x \in [k]$ such that $\delta_x(m, u) > 0$. Also, since x_i and x_i' are distinct for each i, we can assume without loss of generality that $x_i \neq x$ for each i. Now, consider a distribution $\mathbf{p}$ such that $\mathbf{p}_x = 0$ and $\mathbf{p}_{x_i} > 0$ for each i. For this case, the referee must never declare $\mathbf{p}_x$, i.e., $\Pr\Big[\hat{X} = x\Big] = 0$. In particular, $\Pr\Big[\hat{X} = x \,\Big|\, U = u\Big]$ must be 0, which can only happen if $\Pr[\,M = m \mid U = u\,] = 0$. But since $\mathbf{p}_{x_i} > 0$ for each i,

$$\Pr[\,M = m \mid U = u\,] \ge \prod_{i=1}^{n} \mathbf{p}_{x_i} > 0\,,$$

which is a contradiction. □

Note that the proof above shows, as stated before, that any perfect simulation that works for every $\mathbf{p}$ in the interior of the $(k-1)$-dimensional probability simplex, must fail at one point on the boundary of the simplex. In fact, a much stronger impossibility result holds. We show next that for $k = 3$ and $\ell = 1$, we cannot find a perfect simulation that works in the neighborhood of any point in the interior of the simplex.

Theorem B.2. *For any* $n \ge 1$, *there does not exist any* ℓ*-bit perfect simulation of* 3*-ary distributions unless* $\ell \ge 2$, *even under the promise that the input distribution comes from an open set in the interior of the probability simplex.*

[8] Recall the *ε-effective support* of a distribution $\mathbf{q}$ is the minimal set of elements accounting for at least $1 - \varepsilon$ probability mass of $\mathbf{q}$.

Before we prove the theorem, we show that there is no loss of generality in restricting to *deterministic* protocols, namely protocols where each player uses a deterministic function of its observation to communicate. The high-level argument is relatively simple: By replacing player j by two players j_1, j_2, each with a suitable deterministic strategy, the two 1-bit messages received by the referee will allow him to simulate player j's original randomized mapping.

Lemma B.3. *For $\mathcal{X} = \{0,1,2\}$, suppose there exists a 1-bit perfect simulation $S' = (\pi', \delta')$ with n players. Then, there is a 1-bit perfect simulation $S = (\pi, \delta)$ with $2n$ players such that, for each $j \in [2n]$, the communication π is deterministic, i.e., for each realization u of public randomness*

$$\pi_j(x_j, u) = \pi_j(x), \qquad x \in \mathcal{X}\,.$$

Proof. Consider a mapping $f\colon \{0,1,2\}\times\{0,1\}^* \to \{0,1\}$. We will show that we can find mappings $g_1\colon \{0,1,2\} \to \{0,1\}$, $g_2\colon \{0,1,2\} \to \{0,1\}$, and $h\colon \{0,1\} \times \{0,1\} \times \{0,1\}^* \to \{0,1\}$ such that for every u

$$\Pr[\, f(X,u) = 1\,] = \Pr[\, h(g_1(X_1), g_2(X_2), u) = 1\,], \tag{5}$$

where random variables X_1, X_2, X are independent and identically distributed and take values in $\{0,1,2\}$. We can then use this construction to get our claimed simulation S using $2n$ players as follows: Replace the communication $\pi'_j(x,u)$ from player j with communication $\pi_{2j-1}(x_{2j-1})$ and $\pi_{2j}(x_{2j})$, respectively, from two players $2j-1$ and $2j$, where π_{2j-1} and π_{2j} correspond to mappings g_1 and g_2 above for $f = \pi'_j$. The referee can then emulate the original protocol using the corresponding mapping h and using $h(\pi_{2j-1}(x_{2j-1}), \pi_{2j}(x_{2j}), u)$ in place of communication from player j in the original protocol (recall that, the protocol being known to all parties, the referee knows the mapping $f = \pi'_j$ and thus can implement this strategy). Then, since the probability distribution of the communication does not change, we retain the performance of S', but using only deterministic communication now.

Therefore, it suffices to establish (5). For convenience, denote $\alpha_u := \mathbb{1}_{\{f(0,u)=1\}}$, $\beta_u := \mathbb{1}_{\{f(1,u)=1\}}$, and $\gamma_u := \mathbb{1}_{\{f(2,u)=1\}}$. Assume without loss of generality that $\alpha_u \le \beta_u + \gamma_u$; then, $(\beta_u + \gamma_u - \alpha_u) \in \{0,1\}$. Let $g_i(x) = \mathbb{1}_{\{x=i\}}$ for $i \in \{1,2\}$. Consider the mapping h given by

$$h(0,0,u) = \alpha_u,\ h(1,0,u) = \beta_u,$$
$$h(0,1,u) = \gamma_u,\ h(1,1,u) = (\beta_u + \gamma_u - \alpha_u)\,.$$

Then, for every u,

$$\begin{aligned}
&\Pr[\, h(g_1(X_1), g_2(X_2), u) = 1\,]\\
&= \alpha_u(1-\mathbf{p}_1)(1-\mathbf{p}_2) + \beta_u(1-\mathbf{p}_1)\mathbf{p}_2\\
&\qquad + \gamma_u\mathbf{p}_1(1-\mathbf{p}_2) + (\beta_u + \gamma_u - \alpha_u)\mathbf{p}_1\mathbf{p}_2\\
&= \alpha_u(1-\mathbf{p}_1-\mathbf{p}_2) + \beta_u\mathbf{p}_2 + \gamma_u\mathbf{p}_1\\
&= \Pr[\, f(X,u) = 1\,]\,,
\end{aligned}$$

which completes the proof. □

We now prove Theorem B.2, but in view of our previous observation, we only need to consider deterministic communication.

Proof of Theorem B.2. Suppose by contradiction that there exists such a 1-bit perfect simulation protocol $S = (\pi, \delta)$ for n players on $\mathcal{X} = \{0,1,2\}$ such that $\pi(x,u) = \pi(x)$. Assume that this protocol is correct for all distributions $\mathbf{p}$ in the neighborhood of some $\mathbf{p}^*$ in the interior of the simplex. Consider a partition the players into three sets $\mathcal{S}_0$, $\mathcal{S}_1$, and $\mathcal{S}_2$, with

$$\mathcal{S}_i := \{\, j \in [n] : \ \pi_j(i) = 1 \,\}, \qquad i \in \mathcal{X}\,.$$

Note that for deterministic communication the message M is independent of public randomness U. Then, by the definition of perfect simulation, it must be the case that

$$\begin{aligned}
\mathbf{p}_x &= \mathbb{E}_U \sum_{m\in\{0,1\}^n} \delta_x(m,U) \Pr[\, M = m \mid U\,] && (6)\\
&= \mathbb{E}_U \sum_m \delta_x(m,U) \Pr[\, M = m\,]\\
&= \sum_m \mathbb{E}_U[\delta_x(m,U)] \Pr[\, M = m\,] && (7)
\end{aligned}$$

for every $x \in \mathcal{X}$, which with our notation of $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2$ can be re-expressed as

$$\begin{aligned}
&\mathbf{p}_x\\
&= \sum_{m\in\{0,1\}^n} \mathbb{E}_U[\delta_x(m,U)] \prod_{i=0}^{2} \prod_{j\in\mathcal{S}_i} (m_j\mathbf{p}_i + (1-m_j)(1-\mathbf{p}_i))\\
&= \sum_{m\in\{0,1\}^n} \mathbb{E}_U[\delta_x(m,U)] \prod_{i=0}^{2} \prod_{j\in\mathcal{S}_i} (1 - m_j + (2m_j - 1)\mathbf{p}_i)\,,
\end{aligned}$$

for every $x \in \mathcal{X}$. But since the right-side above is a polynomial in $(\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2)$, it can only be zero in an open set in the interior if it is identically zero. In particular, the constant term must be zero:

$$\begin{aligned}
0 &= \sum_{m\in\{0,1\}^n} \mathbb{E}_U[\delta_x(m,U)] \prod_{i=0}^{2} \prod_{j\in\mathcal{S}_i} (1 - m_j)\\
&= \sum_{m\in\{0,1\}^n} \mathbb{E}_U[\delta_x(m,U)] \prod_{j=1}^{n} (1 - m_j)\,.
\end{aligned}$$

Noting that every summand is non-negative, this implies that for all $x \in \mathcal{X}$ and $m \in \{0,1\}^n$, $\mathbb{E}_U[\delta_x(m,U)] \prod_{j=1}^{n}(1 - m_j) = 0$. In particular, for the all-zero message $\mathbf{0}^n$, we get $\mathbb{E}_U[\delta_x(\mathbf{0}^n, U)] = 0$ for all $x \in \mathcal{X}$, so that again by non-negativity we must have $\delta_x(\mathbf{0}^n, u) = 0$ for all $x \in \mathcal{X}$

and randomness u. But the message $\mathbf{0}^n$ will happen with probability

$$\begin{aligned}\Pr[\, M = \mathbf{0}^n \,] &= \prod_{i=0}^{2} \prod_{j\in\mathcal{S}_i} (1-\mathbf{p}_i) \\ &= (1-\mathbf{p}_0)^{|\mathcal{S}_0|}(1-\mathbf{p}_1)^{|\mathcal{S}_1|}(1-\mathbf{p}_2)^{|\mathcal{S}_2|} > 0,\end{aligned}$$

where the inequality holds since $\mathbf{p}$ lies in the interior of the simplex. Therefore, for the output $\hat{X}$ of the referee we have

$$\begin{aligned}\Pr\Big[\, \hat{X} \neq \bot \,\Big] &= \sum_{m} \sum_{x\in\mathcal{X}} \mathbb{E}_U[\delta_x(m,U)] \cdot \Pr[\, M = m \,] \\ &= \sum_{m\neq \mathbf{0}^n} \Pr[\, M = m \,] \sum_{x\in\mathcal{X}} \mathbb{E}_U[\delta_x(m,U)] \\ &\leq \sum_{m\neq \mathbf{0}^n} \Pr[\, M = \mathbf{0}^n \,] \\ &= 1 - \Pr[\, M = \mathbf{0}^n \,] < 1\,,\end{aligned}$$

contradicting the fact that π is a perfect simulation protocol. □

Remark B.4. It is unclear how to extend the proof of Theorem B.2 arbitrary k, ℓ. In particular, the proof of Lemma B.3 does not extend to the general case. A plausible proof-strategy is a black-box application of the $k = 3$, $\ell = 1$ result to obtain the general result using a direct-sum-type argument.

We close this section by noting that perfect simulation is impossible even when the communication from each player is allowed to depend on that from the previous ones. Specifically, we show that availability of such an interactivity can at most bring an exponential improvement in the number of players.

Lemma B.5. *For every $n \geq 1$, if there exists an interactive public-coin ℓ-bit perfect simulation of k-ary distributions with n players, then there exists a public-coin ℓ-bit perfect simulation of k-ary distributions with $2^{\ell n+1}$ players that uses only SMP.*

Proof. Consider an interactive communication protocol π for distributed simulation with n players and ℓ bits of communication per player. We can view the overall protocol as a (2^ℓ)-ary tree of depth n where player j is assigned all the nodes at depth j. An execution of the protocol is a path from the root to the leaf of the tree. Suppose the protocol starting at the root has reached a node at depth j, then the next node at depth $j+1$ is determined by the communication from player j. Thus, this protocol can be simulated non-interactively using at most $((2^\ell)^n - 1)/(2^\ell - 1) < 2^{\ell n+1}$ players, where players $(2^{j-1}+1)$ to 2^j send all messages correspond to nodes at depth j in the tree. Then, the referee receiving all the messages can output the leaf by following the path from root to the leaf. □

Corollary B.6. *Theorems B.1 and B.2 extend to interactive protocols as well.*

C. Distributed Simulation with one bit

Proof of Theorem 4.1. To help the reader build heuristics for the proof, we describe the protocol and analyze its performance in steps. We begin by describing the basic idea and building blocks; we then build upon it to obtain a full-fledged protocol, but with potentially unbounded expected number of players used. Finally, we describe a simple modification which yields our desired bound for expected number of player's accessed.

The scheme, base version. Consider a protocol with $2k$ players where the 1-bit communication from players $(2i-1)$ and $(2i)$ just indicates if their observation is i or not, namely $\pi_{2i-1}(x) = \pi_{2i}(x) = \mathbb{1}_{\{x=i\}}$.

On receiving these $2k$ bits, the referee $\mathcal{R}$ acts as follows:

- if exactly one of the bits $M_1, M_3, \ldots, M_{2k-1}$ is equal to one, say the bit M_{2i-1}, and the corresponding bit M_{2i} is zero, then the referee outputs $\hat{X} = i$;
- otherwise, it outputs $\bot$.

In the above, the probability $\rho_\mathbf{p}$ that some $i \in [k]$ is declared as the output (and not $\bot$) is

$$\rho_\mathbf{p} := \sum_{i=1}^{k} (1-\mathbf{p}_i)\cdot \mathbf{p}_i \prod_{j\neq i} (1-\mathbf{p}_j) = \prod_{j=1}^{k}(1-\mathbf{p}_j),$$

so that

$$\begin{aligned}\rho_\mathbf{p} &= \exp\sum_{j=1}^{k} \ln(1-\mathbf{p}_j) = \exp\left(-\sum_{t=1}^{\infty} \frac{\|\mathbf{p}\|_t^t}{t}\right) \\ &\geq \exp\left(-\left(1+\sum_{t=2}^{\infty}\frac{\|\mathbf{p}\|_2^t}{t}\right)\right) = \frac{1-\|\mathbf{p}\|_2}{e^{1-\|\mathbf{p}\|_2}}\end{aligned}$$

which is bounded away from 0 as long as $\mathbf{p}$ is far from being a point mass (i.e., $\|\mathbf{p}\|_2$ is not too close to 1).

Further, for any fixed $i \in [k]$, the probability that $\mathcal{R}$ outputs i is

$$\mathbf{p}_i \cdot \prod_{j=1}^{k} (1-\mathbf{p}_j) = \mathbf{p}_i \rho_\mathbf{p} \propto \mathbf{p}_i\,.$$

The scheme, medium version. The (almost) full protocol proceeds as follows. Divide the countably infinitely many players into successive, disjoint batches of $2k$ players each, and apply the base scheme to each of these runs. Execute the base scheme to each of the batch, one at a time and moving to the next batch only when the current batch declares a $\bot$; else declare the output of the batch as $\hat{X}$.

It is straightforward to verify that the distribution of the output $\hat{X}$ is exactly $\mathbf{p}$, and moreover that on expectation $1/\rho_\mathbf{p}$

runs are considered before a sample is output. Therefore, the expected number of players accessed (i.e., bits considered by the referee) satisfies

$$\frac{2k}{\rho_{\mathbf{p}}} \leq 2k \cdot \frac{e^{1-\|\mathbf{p}\|_2}}{1-\|\mathbf{p}\|_2}\,. \qquad (8)$$

The scheme, final version. The protocol described above can have the expected number of players blowing to infinity when $\mathbf{p}$ has ℓ_2 norm close to one. To circumvent this difficulty, we modify the protocol as follows: Consider the distribution $\mathbf{q}$ on $[2k]$ defined by

$$\mathbf{q}_{2i} = \mathbf{q}_{2i-1} = \frac{\mathbf{p}_i}{2}, \qquad i \in [k]\,.$$

Clearly, $\|\mathbf{q}\|_2 = \|\mathbf{p}\|_2/\sqrt{2} \leq 1/\sqrt{2}$, and therefore by (8) the expected number of players required to simulate $\mathbf{q}$ using our previous protocol is at most

$$4k \cdot \frac{e^{1-1/\sqrt{2}}}{1-1/\sqrt{2}} \leq 20k.$$

But we can simulate a sample from $\mathbf{p}$ using a sample from $\mathbf{q}$ simply by mapping $(2i-1)$ and $2i$ to i. The only thing remaining now is to simulate samples from $\mathbf{q}$ using samples from $\mathbf{p}$. This, too, is easy. Every 2 players in a batch that declare 1 on observing symbols $(2i-1)$ and $(2i)$ from $\mathbf{q}$ declare 1 when they see i from $\mathbf{p}$. The referee then simply flips each of this 1 to 0, thereby simulating the communication corresponding to samples from $\mathbf{q}$. In summary, we modified the original protocol for $\mathbf{p}$ by replacing each player with two identical copies and modifying the referee to flip 1 received from these players to 0 independently with probability $1/2$; the output is declared in a batch only when there is exactly one 1 in the modified messages, in which case the output is the element assigned to the player that sent 1. Thus, we have a simulation for k-ary distributions that uses at most $20k$ players, completing the proof of the theorem. □

D. Distributed Simulation for any ℓ

Proof of Theorem 1.2. For simplicity, assume that $2^\ell - 1$ divides k. We can then extend the previous protocol by considering a partition of domain into $m = k/(2^\ell - 1)$ parts and assigning one part of size $2^\ell - 1$ each to a player. Each player then sends the all-zero sequence of length ℓ when it does not see an element from its assigned set, or indicates the precise element from its assigned set that it observed. For each batch, the referee, too, proceeds as before and declares an output if exactly one player in the batch sends a 1 – the declared output is the element indicated by the player that sent a 1; else it moves to the next batch. To bound the number of players, consider the analysis of the base protocol. The probability that an output is declared for a batch (a $\perp$ is not declared in the base protocol) is given by

$$\begin{aligned}\rho_{\mathbf{p}} &:= \sum_{i=1}^{m}(1-\mathbf{p}(S_i)) \cdot \sum_{\ell\in S_i}\mathbf{p}_\ell \prod_{j\neq i}(1-\mathbf{p}(S_j)) \\ &= \prod_{j=1}^{m}(1-\mathbf{p}(S_j)) \cdot \sum_{i=1}^{m}\sum_{\ell\in S_i}\mathbf{p}_\ell \\ &= \prod_{j=1}^{m}(1-\mathbf{p}(S_j))\,,\end{aligned}$$

where $\{S_1,\dots,S_m\}$ denotes the partition used. Then, writing $\mathbf{p}^{(S)}$ for the distribution on $[m]$ given by $\mathbf{p}^{(S)}(j) = \mathbf{p}(S_j)$, by proceeding as in the $\ell = 1$ case we obtain

$$\rho_{\mathbf{p}} \geq \frac{1-\|\mathbf{p}^{(S)}\|_2}{e^{1-\|\mathbf{p}^{(S)}\|_2}}\,.$$

Once again, this quantity may be unbounded and we circumvent this difficulty by replacing each player with two players that behave identically and flipping their communicated 1's to 0's randomly at the referee; the output is declared in a batch only when there is exactly one 1 in the modified messages, in which case the output is the element indicated by the player that sent 1. The analysis can be completed exactly in the manner of the $\ell = 1$ case proof by noticing that the protocol is tantamount to simulating $\mathbf{q}$ with $\|\mathbf{q}^{(S)}\|_2 \leq 1/\sqrt{2}$ and accesses messages from at most $20m$ players in expectation. □

E. Proof of Theorem 6.3

In this appendix, we prove Theorem 6.3, stating that taking a random balanced partition of the domain in $L \geq 2$ parts preserves the ℓ_2 distance between distributions with constant probability. Note that the special case of $L = 2$ was proven in (Acharya et al., 2018a). In fact, the proof for general L is similar to the proof in (Acharya et al., 2018a), but requires some additional work. We provide a self-contained proof here for easy reference.

We begin by recall the Paley–Zigmund inequality, a key tool we shall rely upon.

Theorem E.1 (Paley–Zygmund)**.** *Suppose U is a non-negative random variable with finite variance. Then, for every $\theta \in [0,1]$,*

$$\Pr[\,U > \theta\mathbb{E}[U]\,] \geq (1-\theta)^2\frac{\mathbb{E}[U]^2}{\mathbb{E}[U^2]}\,.$$

We will prove a more general version of Theorem 6.3, showing that the ℓ_2 distance to any fixed distribution $\mathbf{q} \in \Delta_{[k]}$ is preserved with a constant probability.[9] Let random variables $X_1,\dots,X_k$ be as in Theorem 6.3; in particular, each

[9] For this application, one should read the theorem statement with $\delta := \mathbf{p} - \mathbf{q}$.

X_i is distributed uniformly on $[L]$ and for every $r \in [L]$, $\sum_{i=1}^{k} \mathbb{1}_{\{X_i=r\}} = \frac{k}{L}$.

Theorem E.2. *Suppose $2 \leq L < k$ is an integer dividing k, and fix $\delta \in \mathbb{R}^k$ such that $\sum_{i\in[k]} \delta_i = 0$. For random variables $X_1, ..., X_k$ above, let $Z = (Z_1, \dots, Z_L) \in \mathbb{R}^L$ with*

$$Z_r := \sum_{i=1}^{k} \delta_i \mathbb{1}_{\{X_i=r\}}, \qquad r \in [L]\,.$$

Then, there exists a constant $c > 0$ such that

$$\Pr\left[\, \|Z\|_2 > \frac{1}{2} \cdot \|\delta\|_2 \right] \geq c.$$

Proof of Theorem E.2. As in Theorem 14 of (Acharya et al., 2018a), the gist of the proof is to consider a suitable non-negative random variable (namely, $\|Z\|_2^2$) and bound its expectation and second moment in order to apply the Paley–Zygmund inequality to argue about anticoncentration around the mean. The difficulty, however, lies in the fact that bounding the moments of $\|Z\|_2$ involves handling the products of correlated L-valued random variables X_i's, which is technical even for the case $L = 2$ considered in (Acharya et al., 2018a). For ease of presentation, we have divided the proof into smaller results.

Lemma E.3 (Each part has the right expectation)**.** *For every $r \in [L]$,*

$$\mathbb{E}[Z_r] = 0\,.$$

Proof. By linearity of expectation,

$$\mathbb{E}[Z_r] = \sum_{i=1}^{k} \delta_i \mathbb{E}\big[\mathbb{1}_{\{X_i=r\}}\big] = \frac{1}{L}\sum_{i=1}^{k} \delta_i = 0.$$

□

Lemma E.4 (The ℓ_2^2 distance to uniform of the flattening has the right expectation)**.** *For every $r \in [L]$,*

$$\begin{aligned}&\operatorname{Var} Z_r = \mathbb{E}\big[Z_r^2\big]\\ &= \frac{1}{L}\|\delta\|_2^2\left(1 - \frac{1}{L} + \frac{L-1}{L(k-1)}\right) \geq \frac{1}{2L}\|\delta\|_2^2\,.\end{aligned}$$

In particular, the expected squared ℓ_2 norm of Z is

$$\mathbb{E}\Big[\|Z\|_2^2\Big] = \mathbb{E}\left[\sum_{r=1}^{L} Z_r^2\right] \geq \frac{1}{2}\|\delta\|_2^2\,.$$

Proof. For a fixed $r \in [L]$, using the definition of Z, the fact that $\sum_{i=1}^{k} \mathbb{1}_{\{X_i=r\}} = \frac{k}{L}$, and Lemma E.3, we get that

$$\begin{aligned}&\operatorname{Var}[Z_r]\\ &= \mathbb{E}\big[Z_r^2\big]\\ &= \mathbb{E}\left[\left(\sum_{i=1}^{k} \delta_i \mathbb{1}_{\{X_i=r\}}\right)^2\right]\\ &= \sum_{1\leq i,j\leq k} \delta_i\delta_j \mathbb{E}\big[\mathbb{1}_{\{X_i=r\}}\mathbb{1}_{\{X_j=r\}}\big]\\ &= \sum_{i=1}^{k} \delta_i^2 \mathbb{E}\big[\mathbb{1}_{\{X_i=r\}}\big] + 2\sum_{1\leq i<j\leq k} \delta_i\delta_j \mathbb{E}\big[\mathbb{1}_{\{X_i=r\}}\mathbb{1}_{\{X_j=r\}}\big]\,.\end{aligned}$$

Since the X_i's – while not independent – are identically distributed, it is enough by symmetry to compute $\mathbb{E}\big[\mathbb{1}_{\{X_k=r\}}\big]$ and $\mathbb{E}\big[\mathbb{1}_{\{X_{k-1}=r\}}\mathbb{1}_{\{X_k=r\}}\big]$. The former is $1/L$; for the latter, note that

$$\mathbb{E}\big[\mathbb{1}_{\{X_{k-1}=r\}}\mathbb{1}_{\{X_k=r\}}\big] \tag{9}$$

$$= \mathbb{E}\big[\mathbb{E}\big[\,\mathbb{1}_{\{X_{k-1}=r\}}\mathbb{1}_{\{X_k=r\}} \mid \mathbb{1}_{\{X_k=r\}}\,\big]\big] \tag{10}$$

$$= \frac{1}{L}\Pr[\,X_{k-1} = r \mid X_k = r\,]$$

$$= \frac{1}{L}\Pr\left[X_{k-1} = r \,\middle|\, \sum_{i=1}^{k-1} \mathbb{1}_{\{X_i=r\}} = \frac{k}{L} - 1\right] \tag{11}$$

$$= \frac{1}{L^2}\cdot\frac{k-L}{k-1}, \tag{12}$$

where the final identity uses symmetry once again, along with the observation that

$$\sum_{i=1}^{k-1} \mathbb{E}\left[\mathbb{1}_{\{X_i=r\}} \,\middle|\, \sum_{j=1}^{k-1} \mathbb{1}_{\{X_j=r\}} = \frac{k}{L} - 1\right] = \frac{k}{L} - 1.$$

Putting it together, we get the result as follows:

$$\begin{aligned}\operatorname{Var}[Z_r] &= \frac{1}{L}\sum_{i=1}^{k}\delta_i^2 + \frac{1}{L^2}\cdot\frac{k-L}{k-1}\cdot 2\sum_{1\leq i<j\leq k}\delta_i\delta_j\\ &= \frac{1}{L}\|\delta\|_2^2 - \frac{1}{L^2}\left(1 - \frac{L-1}{k-1}\right)\|\delta\|_2^2\\ &= \frac{1}{L}\|\delta\|_2^2\left(1 - \frac{1}{L} + \frac{L-1}{L(k-1)}\right).\end{aligned}$$

□

Lemma E.5 (The ℓ_2^2 distance to uniform of the flattening has the required second moment)**.** *There exists an absolute constant $C > 0$ such that*

$$\mathbb{E}\Big[\|Z\|_2^4\Big] \leq C\|\delta\|_2^4\,.$$

Proof of Lemma E.5. Expanding the square, we have

$$\mathbb{E}\Big[\|Z\|_2^4\Big] = \mathbb{E}\left[\left(\sum_{r=1}^{L} Z_r^2\right)^2\right] = \sum_{r=1}^{L}\mathbb{E}\big[Z_r^4\big] + 2\sum_{r<r'}\mathbb{E}\big[Z_r^2 Z_{r'}^2\big] \tag{13}$$

We will bound both terms separately. For the first term, we note that using Equation(21) of (Acharya et al., 2018a) with $\mathbb{1}_{\{X_i=r\}}$ in the role of X_i there, each term $\mathbb{E}\left[Z_r^4\right]$ is bounded above by $19\|\delta\|_2^4/L$ whereby

$$\sum_{r=1}^{L}\mathbb{E}\left[Z_r^4\right] \leq 19\|\delta\|_2^4. \tag{14}$$

However, we need additional work to handle the second term comprising roughly L^2 summands. In particular, to complete the proof we show that each summand in the second term is less than a constant factor times $\|\delta\|_2^4/L^2$.

Claim E.6. *There exists an absolute constant $C' > 0$ such that*

$$\sum_{r<r'}\mathbb{E}\left[Z_r^2 Z_{r'}^2\right] \leq C'\|\delta\|_2^4\,.$$

Proof. Fix any $r \neq r'$. As before, we expand

$$\begin{aligned}
&\mathbb{E}\left[Z_r^2 Z_{r'}^2\right]\\
&= \mathbb{E}\left[\left(\sum_{i=1}^{k}\delta_i\mathbb{1}_{\{X_i=r\}}\right)^2\left(\sum_{i=1}^{k}\delta_i\mathbb{1}_{\{X_i=r'\}}\right)^2\right]\\
&= \sum_{1\leq a,b,c,d\leq k}\delta_a\delta_b\delta_c\delta_d\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r\}}\mathbb{1}_{\{X_c=r'\}}\mathbb{1}_{\{X_d=r'\}}\right]\,.
\end{aligned}$$

Using symmetry once again, note that the term $\mathbb{E}\left[\tilde{X}_a\tilde{X}_b\tilde{X}_c\tilde{X}_d\right]$ depends only on the number of distinct elements in the multiset $\{a,b,c,d\}$, namely the cardinality $|\{a,b,c,d\}|$. The key observation here is that if $\{a,b\}\cap\{c,d\}\neq\emptyset$, then $\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r\}}\mathbb{1}_{\{X_c=r'\}}\mathbb{1}_{\{X_d=r'\}} = 0$. This will be crucial as it implies that the expected value can only be non-zero if $|\{a,b,c,d\}|\geq 2$, yielding a $1/L^2$ dependence for the leading term in place of $1/L$.

$$\mathbb{E}\left[Z_r^2 Z_{r'}^2\right] \tag{15}$$

$$\begin{aligned}
&= \sum_{|\{a,b,c,d\}|=2}\delta_a^2\delta_b^2\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r'\}}\right]\\
&+ \sum_{|\{a,b,c,d\}|=3}\delta_a^2\delta_b\delta_c\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r'\}}\mathbb{1}_{\{X_c=r'\}}\right]\\
&+ \sum_{|\{a,b,c,d\}|=3}\delta_a\delta_b\delta_c^2\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r\}}\mathbb{1}_{\{X_c=r'\}}\right]\\
&+ \sum_{|\{a,b,c,d\}|=4}\delta_a\delta_b\delta_c\delta_d\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r\}}\mathbb{1}_{\{X_c=r'\}}\mathbb{1}_{\{X_d=r'\}}\right]\,.
\end{aligned} \tag{16}$$

The first term, which we will show dominates, is bounded as

$$\begin{aligned}
&\sum_{|\{a,b,c,d\}|=2}\delta_a^2\delta_b^2\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r'\}}\right]\\
&= \mathbb{E}\left[\mathbb{1}_{\{X_{k-1}=r\}}\mathbb{1}_{\{X_k=r'\}}\right]\|\delta\|_2^4 \leq \frac{2}{L^2}\|\delta\|_2^4
\end{aligned}$$

where the inequality uses

$$\mathbb{E}\left[\mathbb{1}_{\{X_{k-1}=r\}}\mathbb{1}_{\{X_k=r'\}}\right] = \frac{1}{L^2}\cdot\frac{k}{k-1} \leq \frac{2}{L^2},$$

which in turn is obtained in the manner of (12).

For the second and the third terms, noting that

$$\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r'\}}\mathbb{1}_{\{X_c=r'\}}\right] = \left|\delta_a^2\delta_b\delta_c\right|\cdot\frac{1}{L^3}\frac{k(k-L)}{(k-1)(k-2)},$$

and that

$$\sum_{|\{a,b,c,d\}|=3}\delta_a^2\delta_b\delta_c = \sum_{1\leq a,b,c\leq k}\delta_a^2\delta_b\delta_c - \sum_{a\neq b}\delta_a^2\delta_b^2 - 2\sum_{a\neq b}\delta_a^3\delta_b$$

with $\sum_{1\leq a,b,c\leq k}\delta_a^2\delta_b\delta_c = \left(\sum_{a=1}^k\delta_a^2\right)\left(\sum_{a=1}^k\delta_a\right)^2 = 0$, $\sum_{a\neq b}\delta_a^2\delta_b^2 \leq \sum_{1\leq a,b\leq k}\delta_a^2\delta_b^2 = \|\delta\|_2^4$, and $\sum_{a\neq b}\delta_a^3\left|\delta_b\right| \leq \sum_{1\leq a,b\leq k}\delta_a^3\left|\delta_b\right| \leq \|\delta\|_\infty\|\delta\|_3^3 \leq \|\delta\|_2^4$, we get

$$\begin{aligned}
&-\frac{6}{L^3}\|\delta\|_2^4\\
&\leq \sum_{|\{a,b,c,d\}|=3}\delta_a^2\delta_b\delta_c\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r'\}}\mathbb{1}_{\{X_c=r'\}}\right]\\
&\leq \frac{6}{L^3}\|\delta\|_2^4\,.
\end{aligned}$$

Finally, as $\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r\}}\mathbb{1}_{\{X_c=r'\}}\mathbb{1}_{\{X_d=r'\}}\right] = \frac{1}{L^4}\frac{k^2(k-L)^2}{(k-1)(k-2)(k-3)(k-4)} \leq \frac{10}{L^4}$, similar manipulations yield

$$\begin{aligned}
&-\frac{\alpha}{L^4}\|\delta\|_2^4\\
&\leq \sum_{|\{a,b,c,d\}|=4}\delta_a\delta_b\delta_c\delta_d\mathbb{E}\left[\mathbb{1}_{\{X_a=r\}}\mathbb{1}_{\{X_b=r\}}\mathbb{1}_{\{X_c=r'\}}\mathbb{1}_{\{X_d=r'\}}\right]\\
&\leq \frac{\alpha}{L^4}\|\delta\|_2^4
\end{aligned}$$

for some absolute constant $\alpha > 0$. Gathering all this in (16), we get that there exists some absolute constant $C' > 0$ such that

$$\sum_{r<r'}\mathbb{E}\left[Z_r^2 Z_{r'}^2\right] \leq C'\sum_{r<r'}\frac{1}{L^2}\|\delta\|_2^4 \leq \frac{C'}{2}\|\delta\|_2^4\,.$$

□

The lemma follows by combining the previous claim with (14). □

We are now ready to establish Theorem 6.3. By Lemmas E.4 to E.5, we have $\mathbb{E}\left[\|Z\|_2^2\right] \geq \frac{1}{2}\|\delta\|_2^2$ and $\mathbb{E}\left[\|Z\|_2^4\right] \leq C\|\delta\|_2^4$, for some absolute constant $C > 0$. Therefore, by the Payley–Zygmund inequality (Theorem E.1) applied to $\|Z\|_2^2$ for $\theta = 1/2$,

$$\begin{aligned}\Pr\left[\|Z\|_2^2 > \frac{1}{4}\|\delta\|_2^2\right] &\geq \Pr\left[\|Z\|_2^2 > \frac{1}{2}\mathbb{E}\left[\|Z\|_2^2\right]\right] \\ &\geq \frac{1}{4}\frac{\mathbb{E}\left[\|Z\|_2^2\right]^2}{\mathbb{E}\left[\|Z\|_2^4\right]} \geq \frac{1}{16C}\,.\end{aligned}$$

This concludes the proof. □