Electronic Colloquium on Computational Complexity, Report No. 108 (2025)

Efficient randomized strong 2-source non-malleable extractor for any linear min-entropy

Divesh Aggarwal*, Pranjal Dutta**, Saswata Mukherjee***, Satyajeet Nagargoje†, and Maciej Obremski‡

Abstract. Randomness is a fundamental requirement in cryptographic systems, enabling secure encryption, commitments, and zero-knowledge proofs. However, real-world randomness sources often suffer from weaknesses that adversaries can exploit, leading to significant security vulnerabilities. While deterministic randomness extraction from a single min-entropy source is impossible, two-source extractors provide a robust solution by generating nearly uniform randomness from two independent weak sources. Moreover, cryptographic systems must also be resilient to leakage and tampering attacks, necessitating the development of non-malleable two-source extractors.

In this work, we construct a two-source non-malleable extractor in the Common Reference String (CRS) model, where a random low-degree polynomial is sampled once and made accessible to independent random sources, the distinguisher, and the tamperer. Our extractor requires only linear min-entropy in both sources and doesn't rely on strong computational assumptions, in contrast to prior constructions requiring computational assumptions such as sub-exponential hardness of the Decisional Diffie-Hellman (DDH) problem. Notably, our construction builds upon and relies on the recent breakthrough proof of the polynomial Freiman-Ruzsa conjecture. A connection of the Freiman-Ruzsa conjecture with two-source extractors was considered in prior work [ZBS11,AGMR24], but their construction did not achieve non-malleability.

Our results advance the state of non-malleable cryptographic primitives, with applications in secure storage, leakage-resilient cryptography, and privacy amplification. By eliminating the need for strong computational hardness assumptions, our techniques provide a more foundational and widely applicable method for randomness extraction.

We also show, that the requirements on CRS for our application are so mild that the CRS can be sampled with 2 party computation even when one of the parties is malicious (setting in which establishing unbiased coins is impossible).

Keywords: Randomness Extraction · Tamper-resilient cryptography · Leakage-resilient cryptography.

1 Introduction

Cryptography with weak randomness. Randomness is a fundamental requirement in cryptography. Many core cryptographic primitives, including semantically secure encryption, commitments, and zero-knowledge proofs rely on it. Dodis, Ong, Prabhakaran, and Sahai [DOPS04] established that these primitives cannot be securely implemented using weak randomness sources, even those with high min-entropy, and instead require truly random inputs.

Despite its necessity, obtaining perfect randomness in real-world applications is highly challenging. Numerous cryptographic attacks exploit flaws in randomness generation. For instance, Breitner and Heninger [BH19] demonstrated how weak randomness in key generation allowed them to recover hundreds of Bitcoin private keys. Similar vulnerabilities have been reported in other cryptographic contexts [HDWH12], [BCC+13].

Real-world randomness often originates from physical processes, such as electronic noise or user activity, which contain entropy but rarely produce uniformly random outputs. Additionally, adversaries may gain

* Centre for Quantum Technologies, National University of Singapore. Email: divesh.aggarwal@gmail.com.
** Nanyang Technological University (NTU) Singapore. Email: Email: pranjal.dutta@ntu.edu.sg
*** National University of Singapore. Email: saswatamukherjee607@gmail.com.
† Georgetown University. Email: satyajeetn2012@gmail.com.
‡ National University of Singapore. Email: obremski.math@gmail.com.

ISSN 1433-8092

partial insight into the randomness generation process, further compromising its quality. The difficulty arises not only from the lack of uniformity of the source but also from the uncertainty about the exact distribution.

The best one can hope for is to deterministically extract a nearly perfect random string for direct usage in the desired application. While there are source models which allow for deterministic randomness extraction, such as von Neumann sources [VN51], affine sources [Bou07], and other efficiently generated or recognizable sources [Blu86,SV86,TV00,DGW09,KRVZ06,Dvi12,BGLZ15,CL16], all these models make strong assumptions about the structure of the source.

One can at best assume that they satisfy some minimal property, for example, that none of the outcomes is highly likely. This is the most natural, flexible, and well-studied source model and it is captured by the notion of min-entropy. The situation is complicated even further by the presence of leakage and tampering attacks.

Leakage and Tampering Attacks. In modern cryptographic systems, ensuring security against adversaries who may gain partial knowledge of secret information is critical. This is where leakage-resilient cryptography plays a vital role. Traditional cryptographic models assume that secret keys remain entirely hidden from attackers, but in reality, side-channel attacks, memory leakage, and hardware vulnerabilities can expose portions of the secret state . Leakage-resilient cryptography is designed to withstand such threats by ensuring that security is preserved even when an adversary obtains partial information about the secret. These techniques are particularly crucial in embedded systems, smart cards, and cloud computing, where attackers may exploit unintended information leaks, such as power consumption, timing variations, or electromagnetic emissions. So, even if we have uniform randomness, it is essential for applications that the source remains uniform conditioned on the view of an adversary, who potentially can obtain leakage about the source. For example, this problem was considered for partial key exposure by [CDH$^+$00,DSS01]and later generalized to memory leakage by [Dzi06,DCLW06,AGV09].

Equally important is non-malleable cryptography, which prevents adversaries from tampering with encrypted and signed messages or even worse with the secret keys in a meaningful way. In many real-world scenarios, attackers not only aim to learn secret information but also to modify data in a controlled manner to deceive or manipulate the system. An example of this is a related key attack where the adversary can tamper with the secret key, and observe the outputs of the cryptographic algorithms on those related keys [GLM$^+$04,BDK08,FKOS22]. Non-malleability ensures that any modification/tampering results in an output that is either completely unusable or unrelated to the original message. Together, leakage-resilience and non-malleability strengthen cryptographic foundations, ensuring robust security even in the presence of sophisticated and resourceful adversaries.

Two-source extractors. Sadly, assuming a lower bound on the min-entropy of the source does not allow deterministic extraction of even 1 almost uniformly random bit [CG88]. This holds even in the highly optimistic case where the source is supported on $\{0,1\}^d$ and has min-entropy $d-1$. Fortunately, [CG88] showed that if we are given two independent min-entropy sources, then we can produce uniform randomness via what is called a two-source extractor.

The problem of constructing explicit low-error two-source extractors for low min-entropy sources was an important focus of research in pseudorandomness over more than 30 years, with fundamental connections to combinatorics and many applications in computer science. The first non-trivial explicit construction was given by Chor and Goldreich [CG88], who showed that the inner product function is a low-error two-source extractor for n-bit sources with min-entropy $(1/2+\gamma)n$, where $\gamma > 0$ is an arbitrarily small constant. A standard application of the probabilistic method shows that (inefficient) low-error two-source extractors exist for polylogarithmic min-entropy. Although several attempts were made to improve the construction of [CG88] to allow sources with smaller min-entropy, the major breakthrough results were obtained after almost two decades. Raz [Raz05] gave an explicit low-error two-source extractor where one of the sources must have min-entropy $(1/2+\gamma)n$ for an arbitrarily small constant $\gamma > 0$, while the other source is allowed to have logarithmic min-entropy. In an incomparable result, Bourgain [Bou05] gave an explicit low-error two-source extractor for sources with min-entropy $(1/2-\gamma)n$, where $\gamma > 0$ is a small constant. Recently, an improved analysis by Lewko [Lew19] showed that Bourgain's extractor can handle sources with min-entropy

$4n/9$. In another line of work, Chattopadhyay and Zuckerman [CZ19] succeeded in constructing explicit 1-bit two-source extractors for polylogarithmic min-entropy with polynomially small error (this was quickly improved to larger output length [Li16] and near-logarithmic min-entropy [BADTS17,Coh17,Li17], with the state-of-the-art currently found in [Li23].

The applications for two-source extractors in cryptography goes beyond just extracting randomness from sources with small min-entropy. Two-source extractors play a crucial role in leakage-resilient cryptography by enabling the generation of uniform randomness even when individual sources may be partially compromised. For example, Davi, Dziembowski, and Venturi [DDV10] used two-source extractors to build a leakage-resilient storage scheme in the model where the physical memory may leak some side-channel information.

Non-malleable extractors. In a breakthrough result, Dodis and Wichs [DW09] introduced the notion of seeded non-malleable extractors as a natural tool for applications in tamper-resilient cryptography. Their main goal was towards achieving privacy amplification against active adversaries [MW97] with an optimal number of rounds and small entropy loss. Roughly speaking, the output of a seeded non-malleable extractor with a uniformly random seed and a source X with some min-entropy should look uniformly random to an adversary who can tamper the seed and obtain the output of the non-malleable extractor on a tampered seed. A natural strengthening of both seeded non-malleable extractors, and two-source extractors are two-source *non-malleable* extractors (also known as seedless non-malleable extractors). Two-source non-malleable extractors were introduced by Cheraghchi and Guruswami [CG14][CG17] in the single-tampering setting and by Chattopadhyay, Goyal, and Li [CGL16] in the multi-tampering setting. Roughly speaking, a function $\mathrm{nmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is said to be a non-malleable extractor if the output of the extractor remains close to uniform (in statistical distance), even conditioned on the output of the extractor on an input correlated with the original source. In other words, we require that

$$\mathrm{nmExt}(X,Y), \mathrm{nmExt}(f(X),g(Y)) \approx_\varepsilon U_m, \mathrm{nmExt}(f(X),g(Y)) \ ,$$

where X and Y are independent sources with enough min-entropy, $f, g : \{0,1\}^n \to \{0,1\}^n$ are arbitrary tampering functions such that (f,g) has no fixed points, U_m is uniform over $\{0,1\}^m$ and independent of the rest, and $\approx_\varepsilon$ means the two distributions are ε-close in statistical distance (for small ε). The original motivation for studying efficient two-source non-malleable extractors stems from the fact that they directly yield efficient split-state non-malleable codes [DPW18] (provided the extractor also supports efficient preimage sampling).

The initial constructions of non-malleable codes [DKO13,ADL14] were largely based on the (limited) non-malleability of the inner-product two-source extractor. Later, improved constructions of non-malleable codes in the split-state model can be broadly classified as: those that made use of both the inner-product two-source extractor [ADKO15,AO20,AKO$^+$22] and more sophisticated constructions of two-source non-malleable extractors [CGL16,Li17,Li19,ACO23,Li23] that required alternating extractors. This topic has been extensively explored in the literature; for a more comprehensive list of works on non-malleable codes in the split-state model, see [Li23] and the references therein. Non-malleable codes and two-source non-malleable extractors have since been applied to other areas, including non-malleable secret sharing [GK18a,GK18b,BS19,ADN$^+$19], randomness extraction from adversarial sources [CGGL19], network extraction protocols [GSZ21], non-malleable commitments [GPR16], and privacy amplification [CKOS19,AOR$^+$22].

In particular, in [AOR$^+$22], the authors present an extension of privacy amplification (PA) against active adversaries, where Eve, as an active adversary, is further allowed to *fully corrupt* the internal memory of one of the honest parties, Alice or Bob, before the protocol execution. Their construction required two-source non-malleable extractors where one source has a small entropy rate δ (where δ is a constant close to 0). Such non-malleable two-source extractors were constructed in [ACO23,Li23].

(Non-malleable) Extractors in the CRS Model were introduced by [GKK20] The Common Reference String (CRS) model, introduced by Garg, Kalai, and Khurana [GKK20], provides a computational framework for constructing non-malleable extractors. In this model, a CRS is sampled once and for all, and three adversaries have full access to it: the sampler, which samples independent randomness sources with sufficient

min-entropy; the tamperer, which is allowed to tamper with these source samples; and the distinguisher, which attempts to distinguish the extractor's output from a uniform distribution, given access to outputs on tampered versions. Their work constructed a one-source non-malleable extractor in this model under the DDH assumption, but only handled one-sided tampering. In a follow-up work, [AOR+22] extended this approach to two-source non-malleable extractors, achieving significantly improved parameters while handling both-sided tampering. Their construction remains secure against an unbounded distinguisher under the existence of nearly optimal collision-resistant hash functions. Furthermore, under the quasi-polynomial hardness of the DDH assumption, their extractor requires much lower min-entropy and handles a broader class of tampering functions, significantly improving upon previous constructions.

In [AGMR24] Alrabiah, Goodman, Mosheiff, and Ribeiro explored the properties of random low-degree multivariate polynomials over $\mathbb{F}_2$, they show that such polynomials are good extractors for sumset sources and any small families of sources. Most relevant to our context, they give the construction of a two-source extractor for small min-entropy in the CRS model. Their construction obtains negligible error relevant in cryptographic setting.

1.1 Our Contributions.

Our main result is a two-source non-malleable extractor construction in the CRS model, where a random *low* degree polynomial needs to be sampled once and for all, and then the independent random sources, the distinguisher and the tamperer are allowed arbitrary access to this polynomial. The informal theorem statement is as below.

Theorem 1 (Informal).

1. *For all n, constant $\rho > 0$ there are constants $\mu, \gamma > 0$ so that there is an efficiently sampleable polynomial p and the efficiently computable function $\mathrm{nmExt}_p : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, that depends on p, is with probability $1 - 2^{-\Omega(n^2)}$ over the randomness of p, an $(n, m, \rho n, 2^{-\gamma n})$ 2-source non-malleable extractor, where $m = \mu n$.*
2. *For all n, there are constants $C_0 > 0$ and $\sigma, \nu > 0$ so that there is an efficiently sampleable polynomial p and the efficiently computable function $\mathrm{nmExt}_p : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, that depends on p, is with probability $1 - 2^{-\Omega(n^2)}$ over the randomness of p, an $\left(n, m, \frac{C_0 n}{\log n}, 2^{-\sigma\sqrt{n}}\right)$ 2-source non-malleable extractor, where $m = \nu\sqrt{n}$.*

This should in particular be contrasted with [GKK20,AOR+22] which required very strong computational assumptions in addition to the common reference string, for the construction of a two-source non-malleable extractor. Our construction on the other hand does not require any additional assumption.

Our work is heavily inspired from [ZBS11] that showed a connection between a two-source extractor and the polynomial Freiman-Ruzsa conjecture.

Notice that a non-malleable two-source extractor is also, by default, a two-source extractor. Alrabiah, Goodman, Mosheiff, and Ribeiro [AGMR24] also gave a construction of two-source extractors for small min-entropy in the CRS model, but it doesn't seem easy to extend their construction to give a two-source non-malleable extractor.

It should also be noted that in the absence of CRS our polynomial can even be sampled by two parties that mutually distrust each other. See Section 4, Section 5.

1.2 Technical Overview

In this exposition we focus on extractors with 1 bit output. With a bit of care and handling technical issues, we are able to obtain large output extractors with similar methods.

As a warm-up, let's build two source extractor first. It is well known that if A and B are two subsets of $\{0,1\}^n$ such that $\log|A| + \log|B| > n + 1$ then if we sample $a \leftarrow A$ and $b \leftarrow B$ independently and uniformly at random and evaluate inner product $\langle a, b\rangle$ the output distribution will be statistically close to uniform.

In the language of random variables this translates to $\langle X, Y\rangle$ is statistically close to uniform if X, Y are independent random variables over $\{0,1\}^n$ such that $H_\infty(X)+H_\infty(Y) > n+1$. The equivalence of these two views comes from a standard observation that every random variable with min-entropy k can be decomposed into a convex combination of distributions that are uniform over sets of size 2^k (i.e. flat sources).

Recent progress in additive combinatorics gives us a much stronger statement to work with: if $\dim_{\mathbb{F}_2}(A)+\dim_{\mathbb{F}_2}(B) > n+1$ (where $\dim_{\mathbb{F}_2}$ stands from dimension of the span of the set, see Definition 7), then the inner-product $\langle a, b\rangle$ of independent uniform samples $a \leftarrow A$ and $b \leftarrow B$ will be distributed uniformly at random.

The idea behind two-source extractor is therefore simple – given two random sources X and Y over $\{0,1\}^n$ that fulfil some minimal lower bound on min-entropy $H_\infty(X), H_\infty(Y) > \rho \cdot n$, if we could deterministically encode $X \to \psi(X)$ and $Y \to \psi(Y)$, such that $\psi(\mathrm{Supp}(X))$ has large dimension (specifically, for every random variables with min-entropy at least $\rho \cdot n$ we need $\dim_{\mathbb{F}_2}(\psi(\mathrm{Supp}(X))) > \frac{n+1}{2}$), then we could simply apply inner-product

$$(X, Y) \to \langle \psi(X), \psi(Y)\rangle$$

and know that the output is uniform since

$$\dim_{\mathbb{F}_2}(\psi(\mathrm{Supp}(X))) + \dim_{\mathbb{F}_2}(\psi(\mathrm{Supp}(Y))) \;>\; n+1\ .$$

There are some technical limitations to this approach like the size of the output of ψ has to be linear in the size of the input but for the purpose of this overview, we don't get into those technical details.

So, it suffices to find an appropriate ψ. Let us begin by picking ψ as a random degree $t = O(n)$ polynomial over $\mathbb{F}_{2^n}$, i.e. $\psi : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$. Notice that randomness complexity (or description) of such polynomial is $(t+1)\cdot n$ which is efficient to express and work with.

Now we need to show that for any X random variable with $H_\infty(X) \geq \rho \cdot n$, we have $\dim(\psi(\mathrm{Supp}(X))) > 0.51 \cdot n$ (where 0.51 is a placeholder for "slightly more than half"). Again, it is sufficient to talk about X, a flat distribution over a set of the size $2^{\rho \cdot n}$. This task can be viewed as a *matrix problem*: Find a matrix with rows iterated by all 2^n strings representing all possible values of $x \in \{0,1\}^n$, and let i-th row be filled with $\psi(i) \in \{0,1\}^n$, this means our matrix has 2^n rows and n columns. To show that for any random variable X with $H_\infty(X) > \rho \cdot n$ we have $\dim_{\mathbb{F}_2}(\mathrm{Supp}(X)) > 0.51 \cdot n$, all we need to show is that any $2^{\rho \cdot n}$ rows of our matrix span space of dimension at least $0.51 \cdot n$. We will focus on fixed set of rows and then union-bound over all possible choices of $2^{\rho \cdot n}$ rows.

First, notice that ψ is a random polynomial of degree t, which means that all t rows in our matrix are independent and uniformly random. So clearly probability that the fixed $2^{\rho \cdot n}$ rows of our matrix are spanning a large dimension space is very high, it's as high as the probability that t uniformly random vectors span a high dimensional space. But we have the order of quantifiers wrong, we fixed the choice of rows and said it's very likely they are of high dimension, what we need to show is that every choice of rows gives a high dimensional space. This can be done via union bound over $\binom{2^n}{2^{\rho \cdot n}}$ possible sources. Let's get some optimistic bound on the failure probability i.e. probability that t uniformly random vectors do not span space of dimension at least $0.51 \cdot n$, such probability is clearly smaller than $2^{0.51 \cdot n \cdot t}$, this follows from picking a subspace of dimension $0.49 \cdot n$, each vector falls into this space with probability $2^{0.51 \cdot n}$. Even this extremely optimistic bound on the failure probability $2^{0.51 \cdot n \cdot t}$ can not withstand the union bound $\binom{2^n}{2^{\rho \cdot n}}$.

First idea: Notice that we don't have to union bound over $\binom{2^n}{2^{\rho \cdot n}}$ sources. Instead, it is actually sufficient to bound over all choices of t rows as it is sufficient to show that every t rows span a sufficiently large space, as every source of high entropy contains some t possible values/rows (since $t < 2^{\rho \cdot n}$). But our optimistic failure probability of $2^{0.51 \cdot n}$ still can't withstand a union bound of $\binom{2^n}{t} \approx 2^{n \cdot t}$.

Second idea: We have to boost the probability that each of t rows have a large span. This can be done by picking larger t. However, this will also increase the union-bound penalty and leads nowhere (at least if we want to keep t reasonably small). Let us increase the size of each row then: $\psi : \{0,1\}^n \to \{0,1\}^{10n}$ while maintaining the $t-$wise independence. Now our t random vectors are much longer and span space of

dimension $0.51 \cdot 10 \cdot n$ with much higher probability that can withstand $2^{n \cdot t}$ union-bound penalty. Sampling such function ψ is also easy: pick a random polynomial $p : \mathbb{F}_{2^{10n}} \to \mathbb{F}_{2^{10n}}$ of degree t, and let $\psi(x) = p(0^{9n} \circ x)$. Notice that because of the union bound we get that almost every polynomial p gives us a matrix such that each t rows of this matrix span large dimensional space. Thus we are done:

$$(X, Y) \to \langle p(0^{9n} \circ X), p(0^{9n} \circ Y) \rangle ,$$

is a good extractor for sources with min-entropy[1] at least $\rho \cdot n$.

Building non-malleable extractor: Now that the warm-up is concluded and we have constructed a two-source extractor we can proceed with the construction of a non-malleable two-source extractor. The construction will be basically the same,

$$(X, Y) \to \langle p(0^{9n} \circ X), p(0^{9n} \circ Y) \rangle .$$

It is just that the polynomial p will have to be of slightly higher (but still linear) degree.

To show that $\text{nmExt}(X, Y), \text{nmExt}(f(X), g(Y))$ is close to $U, \text{nmExt}(f(X), g(Y))$ by XOR lemma, all we need to show is that both $\text{nmExt}(X, Y)$ is close to uniform and $\text{nmExt}(X, Y) + \text{nmExt}(f(X), g(Y))$ is also close to uniform.

First statement is already done – that is simply the extractor statement from the warm-up part of this exposition. Let us focus on proving that $\text{nmExt}(X, Y) + \text{nmExt}(f(X), g(Y))$ is uniform. Notice:

$$\begin{aligned}
&\text{nmExt}(X, Y) + \text{nmExt}(f(X), g(Y)) = \\
&= \langle p(0^{9n} \circ X), p(0^{9n} \circ Y) \rangle + \langle p(0^{9n} \circ f(X)), p(0^{9n} \circ g(Y)) \rangle = \\
&= \langle p(0^{9n} \circ X) \circ p(0^{9n} \circ f(X)), p(0^{9n} \circ Y) \circ p(0^{9n} \circ g(Y)) \rangle
\end{aligned}$$

Consider our matrix again, with 2^n rows, with i'th row equal $p(0^{9n} \circ i) \circ p(0^{9n} \circ f(i))$, now each row is $20 \cdot n$ bits long. If we could show again that every t-rows form a high dimension subspace (say, dimension $0.51 \cdot 20 \cdot n$) we would be done. Sadly, this is not quite true. If f is a constant function then the dimension spanned by any t rows is at best $10 \cdot n$ which is not enough.

At this point we identify, that $f-$close to constant is the only barrier standing between us and the proof. But notice that if f is close to constant then $\text{nmExt}(f(X), g(Y))$ is basically a deterministic function of Y, and by a strong extraction property $\text{nmExt}(X, Y) = \langle p(0^{9n} \circ X), p(0^{9n} \circ Y) \rangle$ is close to uniform even given Y, which means $\text{nmExt}(X, Y)$ is close to uniform even given $\text{nmExt}(f(X), g(Y))$. So in the case of close to constant f we are done.

What happens when f is far from constant i.e. almost bijective? Then look at the rows of our matrix $p(0^{9n} \circ i) \circ p(0^{9n} \circ f(i))$, if f is close to bijection than for any large enough set of rows (remember $t << 2^{\rho \cdot n}$ so we have plenty of rows to play with) we can find subset of t rows that $x_1, ..., x_t, f(x_1), ..., f(x_t)$ are all $2t$ *distinct* values, and thus all t vectors $p(0^{9n} \circ x_i) \circ p(0^{9n} \circ f(x_i))$ are uniform and independent. Thus, we can show that every such t row spans large dimension subspace with an overwhelming probability that can withstand a union bound.

Notice that there are plenty of functions that are neither close to constant nor far from constant: function can be constant on some part of domain and bijective on the rest. This is formally handled by splitting domain into subsets: part of domain where f is constant and part of domain where f is not constant. We prove that our extractor is good on each partition(which follows exactly the ideas highlighted above) and then combine it all together.

Two-party setting. To sample the polynomial in two-party setting Alice and Bob can proceed as follows: Alice will come up with polynomial q of degree $10 \cdot 30 \cdot n$, in response Bob will generate a polynomial r of degree $30 \cdot n$, the combined polynomial is degree $10 \cdot 30 \cdot n$ and is equal to $q + r$. Notice that we need $30 \cdot n$-wise

[1] Above we only really required that $t \leq 2^{\rho \cdot n}$, so one can ask a fair question why not require entropy $\log t$? Unfortunately this is a result of our proof approach via additive combinatorics that require us to have entropy linear in n

independent function, and if Bob was honest that is trivially guaranteed by the uniformity of r. If Alice was honest then we can guarantee that the polynomial $q + r$ has been sampled from the source of min-entropy at least $9 \cdot 30 \cdot n^2$ (i.e. the entropy rate of the source was at least ≈ 0.9) which is sufficient by Remark 25. We also note that this can be generalized to multiple parties where each party publishes shorter and shorter polynomials and last one picks polynomial degree $30 \cdot n$. For formal proof one can look at Section 5. This also means that if malicious Bob decides not to publish his polynomial (abort), Alice can safely reset and engage in the selection process again, and again, polynomial number of times, and as long as at least one of reruns terminates Alice is guaranteed that the generated polynomial is good.

This result is impressive because it demonstrates that the common reference string (CRS) required for the extractor is so weak that it can be securely sampled even in the presence of a malicious party using a two-party computation protocol. In contrast, as shown by Cleve [Cle86], it is impossible to sample uniform random coins with guaranteed fairness in the presence of a single malicious participant—one party can always bias the protocol. Therefore, our construction circumvents this barrier by requiring only a highly relaxed form of the CRS, significantly lowering the trust and setup assumptions compared to traditional approaches that demand strong, unbiased randomness. This advances both the practicality and the robustness of protocols in the CRS model.

This protocol can be expanded to larger number of parties where each subsequent party samples smaller and smaller polynomials - albeit this becomes impractical with larger number of parties.

2 Preliminaries

For a positive integer n, we denote $[n] := \{1, \cdots, n\}$. For strings x, y, we denote $x \circ y$ to be the concatenation of the strings x, y. If M be a $n \times n$ matrix with entries from $\mathbb{F}_2$ and S be a set $\{v_1, \cdots, v_t\} \subseteq \mathbb{F}_2^n$, we denote $M \cdot S = \{M.v_1, \cdots, M \cdot v_t\}$. For any set S, we use $\mathbf{X} \sim S$ to denote that $\mathbf{X}$ is a distribution over the set S. And for any positive integer n, we use the notation $\mathcal{U}_n$ to denote the uniform distribution over $\{0,1\}^n$. We define the support and min-entropy of a random variable as follows.

Definition 1 (Support). *For a random variable $\mathbf{X} \sim \{0,1\}^n$, we say support of $\mathbf{X}$,*

$$\mathrm{Supp}(\mathbf{X}) := \big\{x \in \{0,1\}^n : \Pr[\mathbf{X} = x] \neq 0\big\}.$$

Definition 2 (Min-entropy). *For a random source $\mathbf{X} \sim \{0,1\}^n$, we say $\mathbf{X}$ has min-entropy (denote it as $H_\infty(\mathbf{X})$) at least k if, for all $x \in \{0,1\}^n$,*

$$\Pr[\mathbf{X} = x] \leq 2^{-k}.$$

We say $\mathbf{X}$ is an (n,k) source if $\mathbf{X} \sim \{0,1\}^n$ and $H_\infty(\mathbf{X}) \geq k$.

Fact 2. *For any two distributions $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$, we have $H_\infty((\mathbf{X}, \mathbf{Y})) \geq H_\infty(\mathbf{X})$, where $(\mathbf{X}, \mathbf{Y})$ is the joint distribution of $\mathbf{X}, \mathbf{Y}$.*

Flat sources are a special class of (n,k) sources defined as follows.

Definition 3 (Flat sources). *For any random variable $\mathbf{X}$ over $\{0,1\}^n$, we call it a flat-k source if there is $S \subseteq \{0,1\}^n$ so that $|S| = 2^k$ and $\mathbf{X}$ is uniform over S.*

The following lemma gives a relation between min-entropy k sources and flat-k sources.

Lemma 1 ([Vad12, Lemma 6.10]). *Any (n,k)-source $\mathbf{X}$ is a convex combination of flat-k sources, i.e., $\mathbf{X} = \sum_i p_i \mathbf{X}_i$ where for all i, $p_i \geq 0$, $\sum_i p_i = 1$ and each $\mathbf{X}_i$ is a flat-k source.*

A 2-source disperser (with one-bit output) is a deterministic function that takes two entropy sources, and outputs a non-constant bit.

Definition 4 (2-Source Dispersers). *A function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is an (n,k) 2-source disperser if for all sources $\mathbf{X} \sim \{0,1\}^n, \mathbf{Y} \sim \{0,1\}^n$ with $H_\infty(\mathbf{X}) \geq k$, $H_\infty(\mathbf{Y}) \geq k$, $\mathrm{Supp}(f(\mathbf{X}, \mathbf{Y})) = \{0,1\}$.*

The statistical distance is a standard measure for the proximity of two random variables sampled from the same set.

Definition 5 (Statistical Distance). *Given two random variables $A, B \sim \Omega$, we define the statistical distance as*

$$\Delta(A;B) := \frac{1}{2}\sum_{\omega\in\Omega} |\Pr[A=\omega] - \Pr[B=\omega]|.$$

The following inequality states that if $\mathbf{X}, \mathbf{Y}$ are statistically close, then $f(\mathbf{X}), f(\mathbf{Y})$ are also statistically close.

Lemma 2 (Data processing inequality [Vad12, Lemma 6.3]). *For any possibly randomized function $f : \{0,1\}^n \to \{0,1\}^*$, and random sources $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$, $\Delta(f(\mathbf{X}); f(\mathbf{Y})) \le \Delta(\mathbf{X};\mathbf{Y})$.*

A 2-source extractor is a deterministic function that takes as input two entropy sources, and outputs a random variable that is statistically close to uniform.

Definition 6 (2-source extractors). *A function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *is an (n, m, k, ε) 2-source extractor if for all independent sources $\mathbf{X} \sim \{0,1\}^n$, $\mathbf{Y} \sim \{0,1\}^n$ with $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \ge k$ we have*

$$\Delta(\mathrm{Ext}(\mathbf{X},\mathbf{Y}); \mathcal{U}_m) \le \varepsilon\ .$$

The extractor is said to be an (n, m, k, ε) strong 2-source extractor if

$$\Delta(\mathrm{Ext}(\mathbf{X},\mathbf{Y}), \mathbf{Y}; \mathcal{U}_m, \mathbf{Y}) \le \varepsilon\ ,$$

and

$$\Delta(\mathrm{Ext}(\mathbf{X},\mathbf{Y}), \mathbf{X}; \mathcal{U}_m, \mathbf{X}) \le \varepsilon\ .$$

The following theorem says that every 2-source extractor is a strong 2-source extractor with some loss in error.

Lemma 3 ([Rao07, Theorem 5.1]). *If* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *is a (n, m, k, ε) 2-source extractor then* Ext *is a strong (n, m, k', ε') 2-source extractor where $\varepsilon' \le (2^{k-k'} + \varepsilon)2^m$.*

Lemma 4 (Vazirani's XOR lemma [Vaz86]). *$\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_t) \sim \mathbb{F}^t$ be a random variable. If for every $a_1, \ldots, a_t \in \mathbb{F}$, not all zero, $\Delta(\sum_{i=1}^t a_i \mathbf{X}_i; \mathcal{U}_1) \le \varepsilon$. Then, $\Delta(\mathbf{X}; \mathcal{U}_t) \le \varepsilon \cdot |\mathbb{F}|^{(t+2)/2}$.*

The following is a variant of the XOR lemma that includes side information. For a proof, one can look at [ACLV19, Lemma 13].

Lemma 5. *Let $\mathbb{F}$ be a finite field. Given $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_t) \sim \mathbb{F}^t$ and $\mathbf{Y} \sim Y$ for some set Y be random variables. If for every $a_1, \ldots, a_t \in \mathbb{F}$, not all zero, $\Delta((\sum_{i=1}^t a_i \mathbf{X}_i, \mathbf{Y}); (\mathcal{U}_1, \mathbf{Y})) \le \varepsilon$, then, $\Delta((\mathbf{X}, \mathbf{Y}); (\mathcal{U}_t, \mathbf{Y})) \le \sqrt{\varepsilon} \cdot |\mathbb{F}|^{(t+2)/2}$.*

Definition 7. *For a set $A \subseteq \mathbb{F}_2^n$, let $\dim_{\mathbb{F}_2}(A)$ denote the dimension of $\mathrm{Span}_{\mathbb{F}_2}(A)$ over $\mathbb{F}_2$, where*

$$\mathrm{Span}_{\mathbb{F}_2}(A) = \left\{ \sum_{i=1}^n x_i a_i : x_i \in \mathbb{F}_2 \text{ and } a_i \in A \right\}.$$

Definition 8. *For sets $A, B \subseteq \mathbb{F}_2^n$ we define the duality measure of the sets A, B as*

$$D(A,B) = \left| \mathbb{E}_{a\sim A,\, b\sim B}[(-1)^{\langle a,b\rangle}] \right|.$$

For any $\mathbb{F}_2$ subspace $S \subset \mathbb{F}_2^n$, we call $\hat{S}$ is an affine shift of S if,

$$\hat{S} = v + S := \{v + s : s \in S\} \ ,$$

for some fixed $v \in \mathbb{F}_2^n$. Notice that for a pair of set $A, B \subseteq \mathbb{F}_2^n$, we have $D(A, B) = 1$ holds when A is contained in an affine shift of $(\text{Span}_{\mathbb{F}_2} B)^\perp$, where $(\text{Span}_{\mathbb{F}_2} B)^\perp$ denotes the set of all vectors, those are orthogonal to every vectors in $\text{Span}_{\mathbb{F}_2}(B)$, more formally

$$(\text{Span}_{\mathbb{F}_2} B)^\perp := \{y : \textit{ for all } b \in \text{Span}_{\mathbb{F}_2}(B),\ \langle y, b\rangle = 0\}.$$

Definition 9 (Rank of Binary Function). *Rank of a binary function $E : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ is the minimum integer r so that there are functions $h_1, h_2 : \mathbb{F}_2^n \to \mathbb{F}_2^r$ and $E(x, y) = \langle h_1(x), h_2(y)\rangle$.*
Equivalently rank of the function E is same as rank of the $2^n \times 2^n$ matrix over $\mathbb{F}_2$ whose (x, y)th entry is $E(x, y)$.

Also for $\mathbf{X}, \mathbf{Y} \sim \{0, 1\}^n$, we denote the bias of $\langle \mathbf{X}, \mathbf{Y}\rangle$ as

$$\text{bias}(\mathbf{X}, \mathbf{Y}) = \left| \Pr_{x\sim\mathbf{X},\ y\sim\mathbf{Y}}[\langle x, y\rangle = 1] - \frac{1}{2} \right| .$$

Theorem 3 (Polynomial Freiman-Ruzsa Theorem [GGMT25, Theorem 1.2]). *Suppose $A \subset \mathbb{F}_2^n$ such that $|A + A| \leq K|A|$, then A can be covered by atmost $2K^{12}$ translates of a subspace H of $\mathbb{F}_2^n$ of cardinality atmost $|A|$.*

[ZBS11] showed that PFR conjecture (now a theorem as stated above) implies the ADC-exp (Approximate Duality Conjecture-exp) which is now a theorem.

Theorem 4 (ADC-exp [ZBS11, Conjecture 1.2]). *For every pair of constants $0 < \gamma, \beta < 1$ there exists a constant $\zeta > 0$ such that the following holds: Suppose that $A, B \subseteq \mathbb{F}_2^n$ are such that $D(A, B) \geq 2^{-\zeta n}$ and $|A| \geq 2^{\beta n}$, then there exists $A' \subseteq A, B' \subseteq B$ such that $|A'| \geq 2^{-\gamma n}|A|, |B'| \geq 2^{-\gamma n}|B|$, $D(A', B') = 1$.*

In a subsequent work there was another version of ADC has been proved from PFR that says that assuming slightly *larger* duality of given two sets A, B, we can find A', B', subsets of A, B respectively so that they are orthogonal and $|A'|/|A|$ and $|B'|/|B|$ is at least $2^{-cn/\log n}$, where c is some constant. Formally it is stated below.

Theorem 5 (Strong ADC [BSLRZ14, Lemma 1.10]). *For every pair of sets $A, B \subseteq \mathbb{F}_2^n$ which satisfy $D(A, B) \geq 2^{-\sqrt{n}}$, there are subsets $A' \subseteq A$ and $B' \subseteq B$ so that $|A'| \geq 2^{-cn/\log n}|A|$ and $|B'| \geq 2^{-cn/\log n}|B|$ for some absolute constant c and $D(A', B') = 1$*

2.1 Low rank Disperser to Extractor

In this section we show that constructing 2-source dispersers are enough to construct 2-source extractors. A key ingredient in constructing 2-source dispersers is the following elementary lemma whose proof we include for completeness.

Lemma 6 ([ZBS11, Lemma 3.1]). *Let $A, B \subseteq \{0, 1\}^n$ such that $\dim(A) + \dim(B) > n + 1$, then $\langle ., .\rangle : A \times B \to \{0, 1\}$ is a non-constant function.*

Proof. We will argue via contradiction. Throughout the proof when we say span of some set, we mean the span over $\mathbb{F}_2$. Say, $\langle ., .\rangle$ is constant on $A \times B$.

(i) When $\langle A, B\rangle = 0$, this implies that $A \subseteq (\text{Span } B)^{\perp}$. So we have, $\dim(A) \leq n - \dim(B)$, which contradicts the fact that $\dim(A) + \dim(B) > n + 1$.

(ii) When $\langle A, B\rangle = 1$, fix $a \in A$. For any other $a' \in A$, we have $\langle a' - a, b\rangle = 0$ for all $b \in B$. This implies $A - a \subseteq (\text{Span } B)^{\perp}$. Therefore we have $\dim(A - a) + \dim(B) \leq n$, which in turn implies that $\dim(A) + \dim(B) \leq n + 1$, that leads to contradiction.

□

The next lemma shows how low-rank dispersers are also 2-source extractors (Lemma 2.15 from [ZBS11]). We will include the result as well as its proof for self containment.

Lemma 7 ([ZBS11, Lemma 2.15]). *For all constants $\delta, \alpha, t > 0$, there exists a constant ζ such that, every $(n, \delta n)$ 2-source disperser $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ of rank n/t is also a $(n, 1, (\delta + \alpha)n, 2^{-\zeta n})$ 2-source extractor.*

Proof. The following proof is taken almost verbatim from [ZBS11, Lemma 2.15], and is included here for completeness. In the proof we identify $\{0,1\}$ as $\mathbb{F}_2$. As defined before, $E : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ be a 2 source disperser of rank n/t. By Definition 9, there are $f, g : \mathbb{F}_2^n \to \mathbb{F}_2^{n/t}$ so that

$$E(x, y) = \langle f(x), g(y)\rangle \qquad \text{For all } x, y \in \mathbb{F}_2^n \,.$$

Let us define $\tilde{f}, \tilde{g} : \mathbb{F}_2^n \to \mathbb{F}_2^{(2n+n/t)}$, as

$$\tilde{f}(x) = f(x) \circ x \circ 0^n \text{ and } \tilde{g}(y) = g(y) \circ 0^n \circ y$$

Here 0^n is all zero vector of length n. Let $\zeta' > 0$ be the constant from Theorem 4 with constants $\beta = (\alpha + \delta)(1/t + 2)^{-1}$ and $\gamma = \alpha(1/t + 2)^{-1}$. Define $\zeta = \zeta'(2 + 1/t)$.

We proceed by contradiction. Let us assume, there exist two independent distributions $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ with min-entropy at least $(\delta + \alpha)n$ such that

$$\Delta(E(\mathbf{X}, \mathbf{Y}); \mathcal{U}_1) > 2^{-\zeta n}. \tag{1}$$

From Lemma 1 we can assume $\mathbf{X}$ and $\mathbf{Y}$ are flat-$(\delta + \alpha)n$ sources with support A, B respectively. Define $\bar{A} = \{\tilde{f}(x) : x \in A\}$ and $\bar{B} = \{\tilde{g}(y) : y \in B\}$. Note that from definition of $\tilde{f}$ and $\tilde{g}$, we have $|\bar{A}|, |\bar{B}| = 2^{(\delta+\alpha)n} = 2^{\beta(2n+n/t)}$. Note that,

$$\frac{1}{2} D(\bar{A}, \bar{B}) = \frac{1}{2} \left| \mathbb{E}_{x \sim \mathbf{X},\, y \sim \mathbf{Y}} (-1)^{\langle \tilde{f}(x), \tilde{g}(y)\rangle} \right| = \frac{1}{2} \left| \mathbb{E}_{x \sim \mathbf{X},\, y \sim \mathbf{Y}} (-1)^{\langle f(x), g(y)\rangle} \right|.$$

Hence from Equation (1) we have,

$$\frac{1}{2} D(\bar{A}, \bar{B}) = \Delta(E(\mathbf{X}, \mathbf{Y}); \mathcal{U}_1) > 2^{-\zeta n} = 2^{-\zeta'(n/t + 2n)}.$$

Applying Theorem 4, there are $A' \subseteq \bar{A}$ and $B' \subseteq \bar{B}$ so that $D(A', B') = 1$ and

$$|A'| \geq \frac{|\bar{A}|}{2^{\gamma n}} > \frac{2^{(\delta+\alpha)n}}{2^{\gamma(2n+n/t)}} = \frac{2^{(\delta+\alpha)n}}{2^{\alpha n}} = 2^{\delta n} \,.$$

And similarly $|B'| \geq 2^{\delta n}$. As, $\tilde{f}, \tilde{g}$ are injective, we say that $\mathbf{X}'$ and $\mathbf{Y}'$ which are uniformly distributed over $\tilde{f}^{-1}(A'), \tilde{g}^{-1}(B')$ respectively, has min-entropy at least δn but $E(\mathbf{X}', \mathbf{Y}')$ is constant, which is a contradiction. □

The above proof follows from the *weaker* version of ADC. Starting from Theorem 5 we follow the same proof, we get that every two source disperser for the sources of min-entropy $O(n/\log n)$ is also a two source extractor with error $2^{-\sqrt{n}}$ and the same min-entropy requirement.

Lemma 8 (Extractor for lower min-entropic sources from strong ADC). *For any constant $t > 0$ and $\delta > 0$ (δ may not be a constant) we have: Every function $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ which is a $(n, \delta n)$ 2-source disperser of rank n/t, is also a $\left(n, 1, \delta n + \frac{cn}{\log n}, 2^{-\sqrt{n}}\right)$ 2-source extractor. Here c is the absolute constant that we get from Theorem 5.*

Additionally, we will need the following well known fact.

Fact 6. *A polynomial of degree at most d over a field can be uniquely determined by specifying its values at any $d+1$ distinct points. In the case of a finite field GF(q) with q elements (where q is a prime power), this fact provides a way to construct a sequence of q random variables that are $(d+1)$-wise independent. Specifically, if the $d+1$ coefficients of the polynomial are chosen independently and uniformly at random from GF(q), then the polynomial's values at any $d+1$ distinct inputs are also uniformly and independently distributed.*

Throughout the paper we only work with $\mathbb{F}_2$ and some finite extension of it. Hence we state the above more formally for some extension of $\mathbb{F}_2$. Say $\mathcal{H}_{t,r} := \{p \in \mathbb{F}_{2^r}[Z] : \deg(p) \leq t-1\}$. Then, for distinct $x_1, \ldots, x_t$ and $y_1, \ldots, y_t$ from $\mathbb{F}_{2^r}$,

$$\Pr_{p \leftarrow \mathcal{H}_{t,r}}[p(x_1) = y_1 \wedge \cdots \wedge p(x_t) = y_t] = \frac{1}{|\mathbb{F}_{2^r}|^t} = 2^{-rt}.$$

3 Strong two source extractor construction

In this section, we establish that a *random low-degree polynomial* evaluated over a source with sufficient entropy induces a large span. This result serves as a key component in the construction of a strong two-source extractor based on the inner product. Throughout the section, we use $\{0,1\}^r$ and $\mathbb{F}_2^r$ interchangeably for any $r \in \mathbb{N}$. Additionally, note that the additive group of $\mathbb{F}_2^r$ is homomorphic to $\mathbb{F}_{2^r}$. We will use these notions interchangeably and explicitly clarifying any distinctions when necessary.

Throughout this section, for any set $S \subseteq \mathbb{F}_2^n$ and $u \in \mathbb{N}$ with $u > 1$, we define

$$S^{(u)} := \left\{(0^{(u-1)\cdot n} \circ s) : s \in S\right\}.$$

Now we are ready to formally state the large-span property of a randomly chosen low-degree polynomial over a fixed (sufficiently large entropy) source.

Lemma 9. *Let $u > 2$ be an integer and $X \subseteq \mathbb{F}_2^n$, such that $|X| \geq un$. Let $p : \mathbb{F}_{2^{un}} \to \mathbb{F}_{2^{un}}$ be a polynomial of degree un with all the $(un+1)$-many coefficients chosen uniformly and independently from $\mathbb{F}_{2^{un}}$. Define the set $p(X) := \{p(y) : y \in X^{(u)}\}$. Then, for any $d \leq un$,*

$$\Pr_{p \leftarrow \mathcal{H}_{un,un}}[\dim_{\mathbb{F}_2}(p(X^{(u)})) \leq d] \leq 2^{un-(un-d)^2}.$$

Proof. Let $x_1, \ldots, x_{un}$ be kn distinct elements of X, chosen arbitrarily, and let $X' = \{x_1, \ldots, x_{un}\}$. We will prove an upper bound on the probability of the event $\dim_{\mathbb{F}_2}(p(X'^{(u)})) \leq d$. Note that this will trivially imply our desired upper bound.

By definition, $\dim_{\mathbb{F}_2}(p(X'^{(u)})) \leq d$, if $p(X'^{(u)})$ is in the $\mathbb{F}_2$ span of the elements $\{p(y_1^{(u)}), p(y_2^{(u)}), \ldots, p(y_d^{(u)})\}$, for some $y_1, \ldots, y_d \in X'$. For any such $y_1, \ldots, y_d$, the size of the span of $\{p(y_1^{(u)}), \ldots, p(y_d^{(u)})\}$ is at most 2^d, and using $(un+1)$-wise independence of the hash function, the probability that $p(X'^{(u)})$ is contained in the $\mathbb{F}_2$ span of $\{p(y_1^{(u)}), p(y_2^{(u)}), \ldots, p(y_d^{(u)})\}$ is at most

$$\left(\frac{2^d}{2^{un}}\right)^{un-d}.$$

Taking a union bound over all possible choices of $y_1, \ldots, y_d$, we get that

$$\Pr_{p \leftarrow \mathcal{H}_{un,un}}[\dim_{\mathbb{F}_2}(p(X'^{(u)})) \leq d] \leq \binom{un}{d} 2^{-(un-d)^2} \leq 2^{un-(un-d)^2}.$$

□

We now recall the following definitions for a $p \in \mathcal{H}_{10n,10n}$:

$$\begin{aligned}
&\mathrm{Dis}_p : \{0,1\}^n \times \{0,1\}^n \to \{0,1\} \\
&\mathrm{Dis}_p(x,y) \to \langle p(0^{9n} \circ x), p(0^{9n} \circ y) \rangle \\
&\mathrm{Ext}_p : \{0,1\}^n \times \{0,1\}^n \to \{0,1\} \\
&\mathrm{Ext}_p(x,y) \to \langle p(0^{9n} \circ x), p(0^{9n} \circ y) \rangle
\end{aligned}$$

Remark 1. Note that for any $p \in \mathcal{H}_{10n,10n}$, p takes elements from $\mathbb{F}_2^n$ to $\mathbb{F}_2^{10n}$. So, by Definition 9, for all p, we have rank of Dis_p and Ext_p are at most $10n$.

Our claim is that with overwhelming probability over the choice of polynomial $p \in \mathcal{H}_{10n,10n}$, the above object is a 2-source disperser for all logarithmic entropy sources.

Corollary 1 (Disperser of linear rank for log-entropy sources). *Consider any* $\rho \geq \frac{\log 10n}{n}$.

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}}[\mathrm{Dis}_p \textit{ is a } (n, \rho n) \textit{ disperser}] \geq 1 - 2^{-5n^2}.$$

Proof. Consider any $X \subseteq \mathbb{F}_2^n$ with $|X| \geq 10n$. From Lemma 9 with parameters $u = 10$ and $d = 6n$, we get that

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}}[\dim_{\mathbb{F}_2}(p(X^{(10)})) \leq 6n] \leq 2^{-15n^2}.$$

By union bound over all such $X \subseteq \mathbb{F}_2^n$, we get the following,

$$\begin{aligned}
&\Pr_{p \leftarrow \mathcal{H}_{10n,10n}}[\exists\, X \subseteq \mathbb{F}_2^n \text{ of size at least } 10n, s.t.\ \dim_{\mathbb{F}_2}(p(X^{(10)})) \leq 6n] \\
\leq\ &\Pr_{p \leftarrow \mathcal{H}_{10n,10n}}[\exists\, X \subseteq \mathbb{F}_2^n \text{ of size exactly } 10n, s.t.\ \dim_{\mathbb{F}_2}(p(X^{(10)})) \leq 6n] \\
\leq\ &\binom{2^n}{10n} 2^{-15n^2} \leq (2^n)^{10n} 2^{-15n^2} = 2^{-5n^2}.
\end{aligned}$$

This means that with probability at least $1 - 2^{-5n^2}$, for all $X \subseteq \mathbb{F}_2^n$ of size at least $10n$, $\dim_{\mathbb{F}_2}(p(X^{(10)})) > 6n$. Note that, if for some $X, Y \subseteq \mathbb{F}_2^n$ with size at least $10n$, if $\dim_{\mathbb{F}_2}(p(X^{(10)})), \dim_{\mathbb{F}_2}(p(Y^{(10)})) \geq 6n$, then trivially

$$\dim_{\mathbb{F}_2}(p(X^{(10)})) + \dim_{\mathbb{F}_2}(p(Y^{(10)})) > 10n + 1.$$

Combining the above facts along with Lemma 6 it follows that,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}}\left[\begin{array}{l} \forall\, X, Y \subseteq \mathbb{F}_2^n \text{ with size at least } 10n, \\ \langle p(X^{(10)}), p(Y^{(10)}) \rangle \text{ is non-constant} \end{array}\right] \geq 1 - 2^{-5n^2}.$$

Therefore, for all $\rho \geq \frac{\log 10n}{n}$

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}}[\mathrm{Dis}_p \text{ is a 2-source disperser for } \rho n\text{-flat sources}] \geq 1 - 2^{-5n^2}.$$

And combining the above with Lemma 1 we can conclude the proof. □

Combining Lemma 7 and Corollary 1 gives us a *randomized* construction of 2-source extractors for *all* sources having *linear* entropy. More succinctly,

Theorem 7 (Strong extractor of exponential error).

(i) For any constant $\rho > 0$ there exists a constant $\zeta' > 0$ so that the following holds

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} \left[\text{Ext}_p \text{ is a } (n, 1, \rho n, 2^{-\zeta' n}) \text{ strong 2-source extractor} \right] \geq 1 - 2^{-5n^2}.$$

(ii) There is an absolute constant C so that,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} \left[\text{Ext}_p \text{ is a } \left(n, 1, \frac{Cn}{\log n}, 2^{-\sqrt{n}/2} \right) \text{ strong 2-source extractor} \right] \geq 1 - 2^{-5n^2}.$$

Proof. **Proof of (i).** From Remark 1 rank of Ext_p over $\mathbb{F}_2$ is $10n$. By Corollary 1 and Lemma 7 with parameters $\delta = 0.1\rho$, which is strictly greater than $\log(10n)/n$ and $\alpha = 0.8\rho$, we get $\zeta > 0$ so that,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} [\text{Ext}_p \text{ is a } (n, 1, 0.9\rho n, 2^{-\zeta n}) \text{ 2-source extractor}] \geq 1 - 2^{-5n^2}.$$

From Lemma 3 we have,

$$\begin{aligned} &\text{Ext}_p \text{ is } (n, 1, 0.9\rho n, 2^{-\zeta n}) \text{ 2-source extractor} \\ \implies &\text{Ext}_p \text{ is } (n, 1, \rho n, \varepsilon_1) \text{ strong 2-source extractor.} \end{aligned}$$

for $\varepsilon_1 = (2^{-0.1\rho n} + 2^{-\zeta n}) \times 2$. Hence, $\varepsilon_1 \leq 2^{-\zeta' n}$ for some positive constant ζ' and this completes the proof of (i)

Proof of (ii). Using the fact that $(x, y) \mapsto \langle p(0^{9n} \circ x), p(0^{9n} \circ y) \rangle$ has rank $10n$ and Lemma 8 we have,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} \left[\text{Ext}_p \text{ is a } \left(n, 1, \frac{2cn}{\log n}, 2^{-\sqrt{n}} \right) \text{ 2-source extractor} \right] \geq 1 - 2^{-5n^2}.$$

Fix $C = 3c$. Lemma 3 tells us, if Ext_p is a $\left(n, 1, \frac{2cn}{\log n}, 2^{-\sqrt{n}} \right)$ 2-source extractor, it is also a $\left(n, 1, \frac{Cn}{\log n}, \varepsilon_2 \right)$ strong 2-source extractor, where $\varepsilon_2 = \left(2^{-\frac{n}{\log n}} + 2^{-\sqrt{n}} \right) \times 2 \leq 2^{-\sqrt{n}/2}$. From here the theorem follows. □

Remark 2. Note that for any full rank matrix $L \in \mathbb{F}_2^{10n \times 10n}$ we have, $\dim_{\mathbb{F}_2}(p(X^{(10)})) = \dim_{\mathbb{F}_2}(L \cdot p(X^{(10)}))$, for $X \subseteq \mathbb{F}_2^n$. Hence following the line of proof given in Lemma 9 and Corollary 1, we can infer the following,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} \left[\begin{array}{c} \forall\, X, Y \subseteq \mathbb{F}_2^n \text{ with size at least } 10n, \\ \langle p(X^{(10)}), L \cdot p(Y^{(10)}) \rangle \text{ is non-constant} \end{array} \right] \geq 1 - 2^{-5n^2}.$$

If we define a map,

$$\begin{aligned} E_p : \{0,1\}^n \times \{0,1\}^n &\to \{0,1\} \\ (x, y) &\longmapsto \langle p(0^{9n} \circ x), L \cdot p(0^{9n} \circ y) \rangle \end{aligned}$$

Then from proof of part (i) of Theorem 7 we have, for all $\rho > 0$ there is $\zeta' > 0$ so that

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} [E_p \text{ is } (n, 1, \rho n, 2^{-\zeta' n}) \text{ strong 2-source extractor}] \geq 1 - 2^{-5n^2}.$$

And from proof of part (ii) we have, there is an absolute constant C so that,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} \left[E_p \text{ is a } \left(n, 1, \frac{Cn}{\log n}, 2^{-\sqrt{n}/2} \right) \text{ strong 2-source extractor} \right] \geq 1 - 2^{-5n^2}.$$

3.1 Multi-bit output extractors

Previously we have seen a randomized construction of a 2-source extractor with only 1 bit output. In this subsection, we will extend it to the *multi-bit* output extractor, with a *small* increment in the error.

Definition 10. *A set of matrices $L_1, \ldots, L_r \in \mathbb{F}_2^{r \times r}$ are called independent if for all $v_1 \ldots, v_r \in \mathbb{F}_2$ not all zero, $\sum_{j=1}^{r} v_j L_j$ is a full rank matrix.*

It turns out that one can efficiently construct such independent matrices, as shown in [ZBS11].

Lemma 10 (Constructing independent matrices [ZBS11, Section 6.2]). *Consider the field $\mathbb{F}_{2^r}$ and let $v_1, \ldots, v_r \in \mathbb{F}_{2^r}$ be the $\mathbb{F}_2$ basis of $\mathbb{F}_{2^r}$ with $e_j : \mathbb{F}_{2^r} \to \mathbb{F}_{2^r}$ denoting the invertible map $e_j(x) = v_j \cdot x$ for all $j \in [r]$. Let L_i denote the matrices representing the linear transformation e_i, then the matrices $L_1, \ldots, L_r$ are independent.*

We use the above-defined explicit matrices L_i to extend our result in the multi-bit output regime, as follows.

Theorem 8 (Multi-bit output 2-source Extractor). *Define a map $E'_p : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$, for $m \leq 10n$ as follows,*

$$E'_p(x, y) \ := \ (\langle p(0^{9n} \cdot x), L_1 \cdot p(0^{9n} \cdot y)\rangle, \ldots, \langle p(0^{9n} \cdot x), L_m \cdot p(0^{9n} \cdot y)\rangle).$$

where $L_1, \ldots, L_m \in \mathbb{F}_2^{10n \times 10n}$ are explicit matrices from Lemma 10. Then,

(i) For constant $\rho > 0$, say, $\zeta' > 0$ is the constant from Theorem 7. Then there exists constant $\beta > 0$ such that if $m = \zeta' n/8$,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}}[E'_p \text{ is a } (n, m, \rho n, 2^{-\beta n}) \text{ strong 2-source extractor}] \ \geq \ 1 - 2^{-5n^2} \ .$$

(ii) There exists a constant C so that if $m = \sqrt{n}/8$, we have,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} \left[E'_p \text{ is a } \left(n, m, \frac{Cn}{\log n}, 2^{-\sqrt{n}/8} \right) \text{ strong 2-source extractor} \right] \geq 1 - 2^{-5n^2}$$

Proof. Without loss of generality let us assume that our input sources $\mathbf{X}, \mathbf{Y}$ are flat-ρn sources with support X, Y respectively. Note that, for any non-empty set $S \subseteq [m]$,

$$\sum_{i \in S} \langle p(X^{(10)}), L_i \cdot p(Y^{(10)})\rangle = \langle p(X^{(10)}), L \cdot p(Y^{(10)})\rangle.$$

where $L \in \mathbb{F}_2^{10n \times 10n}$ is a full-rank matrix. Define $E_p^S(x, y) := \langle p(0^{9n} \circ x), L \cdot p(0^{9n} \circ y)\rangle$.

Proof of (i). From Remark 2, we have constant $\zeta' > 0$ such that

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}}[E_p^S \text{ is } (n, 1, \rho n, 2^{-\zeta' n}) \text{ strong 2-source extractor}] \geq 1 - 2^{-5n^2}.$$

For any $S \subseteq [m]$, if we have $\Delta\Big(\big(\sum_{i \in S} \langle p(\mathbf{X}^{(10)}), L_i \cdot p(\mathbf{Y}^{(10)})\rangle, \mathbf{Y} \big); \big(\mathcal{U}_1, \mathbf{Y} \big) \Big) \leq 2^{-\zeta' n}$, by using Lemma 5, we can deduce that

$$\Delta\Big(\big(\langle p(\mathbf{X}^{(10)}), L_1 \cdot p(\mathbf{Y}^{(10)})\rangle, \ldots, \langle p(\mathbf{X}^{(10)}), L_m \cdot p(\mathbf{Y}^{(10)})\rangle, \mathbf{Y} \big); \big(\mathcal{U}_m, \mathbf{Y} \big) \Big)$$
$$\leq \ \sqrt{2^{-\zeta' n}} \times 2^{(m+2)/2} \ .$$

Here $\mathbf{X}^{(10)}$ and $\mathbf{Y}^{(10)}$ are uniform distributions over $X^{(10)}$ and $Y^{(10)}$. From the choice of our m, $2^{-0.5\zeta' n} \times 2^{(m+2)/2} \leq 2^{-\beta n}$ for some positive constant β and we are done.

Proof of (ii). From Remark 2 we have, there is an absolute constant C, so that for uniformly sampled p from $\mathcal{H}_{10n,10n}$, with probability at least $1 - 2^{-5n^2}$,

$$E_p^S \text{ is a } \left(n, 1, \frac{Cn}{\log n}, 2^{-\sqrt{n}/2}\right) \text{ strong 2-source extractor .}$$

Hence, by the same idea as above, from Lemma 5, we can say,

$$\Pr_{p \leftarrow \mathcal{H}_{10n,10n}} \left[E'_p \text{ is a } \left(n, m, \frac{Cn}{\log n}, \varepsilon'\right) \text{ strong 2-source extractor } \right] \geq 1 - 2^{-5n^2} ,$$

for $\varepsilon' = \sqrt{2^{-\sqrt{n}/2}} \times 2^{(m+2)/2} \leq 2^{-\sqrt{n}/8}$ as, $m \leq \sqrt{n}/8$. □

4 Non-malleable Extractors

For any function $f : \{0,1\}^n \to \{0,1\}^n$, we say that $s \in S$ is a *fixed point* of f, if $f(s) = s$. Let $\mathcal{F}_n$ be the set of all functions from $\{0,1\}^n$ to $\{0,1\}^n$ *without* any fixed points. Now we state the definition of Non-malleable strong 2-source extractors, as defined by Cheraghchi and Guruswami in [CG14].

Definition 11 (Non-malleable strong 2-source extractor [CG14]). *We say* $\text{nmExt} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *is a non-malleable* (n, m, k, ϵ) *strong 2-source extractor if for all functions* $f, g \in \mathcal{F}_n$ *and for all independent sources* $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$*, with* $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq k$*, the following three properties hold:*

1. $\text{nmExt}(\mathbf{X}, \mathbf{Y}), \text{nmExt}(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \text{nmExt}(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y}$
2. $\text{nmExt}(\mathbf{X}, \mathbf{Y}), \text{nmExt}(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \text{nmExt}(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y}$
3. $\text{nmExt}(\mathbf{X}, \mathbf{Y}), \text{nmExt}(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \text{nmExt}(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y}$

Recall the definition of $\mathcal{H}_{t,r}$: it denotes the set of univariate polynomials over the field $\mathbb{F}_{2^r}$ of degree at most $(t-1)$. For a $p \in \mathcal{H}_{30n,10n}$ and $m \leq 10n$, let us define our non-malleable extractor $\text{nmExt}_p^m : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ in the following way:

$$\text{nmExt}_p^m(x, y) \ := \ \left(\langle p(0^{9n} \circ x), L_1 \cdot p(0^{9n} \circ y) \rangle, \ldots, \langle p(0^{9n} \circ x), L_m \cdot p(0^{9n} y) \rangle \right) .$$

In the above equation, $L_1, \ldots, L_m$ are the explicit independent matrices that can be constructed using Lemma 10.

In this section, we will show that (uniformly) sampling a random polynomial of *linear degree*, over an extension field where the degree of extension is *linear*, would yield a strong non-malleable 2-source extractor for all linear entropy sources. Moreover the function obtained by the uniformly chosen polynomial is a strong non-malleable extractor for the sources of min-entropy $O(n/\log n)$, with a slight loss in error. More formally,

Theorem 9 (Efficient strong non-malleable two source extractor).

(i) For every constant $\rho > 0$*, there exist constants* $\mu, \gamma > 0$ *so that with* $m = \mu n$*, we have,*

$$\Pr_{p \leftarrow \mathcal{H}_{30n,10n}} \left[\begin{array}{c} \text{nmExt}_p^m \text{ is a strong two source} \\ (n, m, \rho n, 2^{-\gamma n}) \text{ non-malleable extractor} \end{array} \right] \geq 1 - 2^{-2n^2}.$$

(ii) There are constants C_0 *and* $\sigma, \nu > 0$ *so that when* $m = \nu\sqrt{n}$*,*

$$\Pr_{p \leftarrow \mathcal{H}_{30n,10n}} \left[\text{nmExt}_p^m \text{ is a } \left(n, m, \frac{C_0 n}{\log n}, 2^{-\sigma\sqrt{n}}\right) \text{ strong 2-source non-malleable extractor} \right] \geq 1 - 2^{-2n^2}$$

Remark 3. To prove the above, we will only use the fact that $\mathcal{H}_{30n,10n}$ is a family of $30n$-wise independent hash functions. Instead of polynomials if we take any arbitrary $30n$-wise independent hash family $\mathcal{H}$ over $\mathbb{F}_{2^{10n}}$, for an uniformly random $h \leftarrow \mathcal{H}$ we can prove the same.

To show the above theorem, we will prove Lemma 11 (which would prove the first item of Definition 11) and Lemma 12 (which would prove the second and third items of Definition 11). In particular, for the first item, we will prove the following.

Lemma 11. *Let $f, g \in \mathcal{F}_n$. Then we have:*

(i) For any constant $\rho > 0$, there exist constants $\gamma, \mu > 0$ such that with probability at least $1-2^{-3n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$ the following holds: for all independent sources $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$, with $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq \rho n$, we have

$$\mathrm{nmExt}_p^{\mu n}(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^{\mu n}(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y} \approx_{2^{-\gamma n}} \mathcal{U}_{\mu n}, \mathrm{nmExt}_p^{\mu n}(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y}$$

(ii) There are constants C_0 and $\sigma, \nu > 0$ so that, with probability at least $1 - 2^{-3n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$ the following holds: for all independent sources $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$, with $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq \frac{C_0 n}{\log n}$, we have

$$\mathrm{nmExt}_p^{\nu\sqrt{n}}(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^{\nu\sqrt{n}}(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y} \approx_{2^{-\sigma\sqrt{n}}} \mathcal{U}_{\nu\sqrt{n}}, \mathrm{nmExt}_p^{\nu\sqrt{n}}(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y} .$$

For the second and third items (which are symmetric) i.e. when *exactly* one of the two sources is corrupted, we will prove the following.

Lemma 12. *Let $f, g \in \mathcal{F}_n$. Then,*

(i) For any $\rho > 0$, there are constants $\tilde{\mu}, \tilde{\gamma} > 0$ such that with probability at least $1 - 2 \cdot 2^{-4n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$ the following holds: for all independent sources $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$, with $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq \rho n$, the following two properties hold.

1. $\mathrm{nmExt}_p^m(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^m(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \mathrm{nmExt}_p^m(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y}$
2. $\mathrm{nmExt}_p^m(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^m(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \mathrm{nmExt}_p^m(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y}$

where $m = \tilde{\mu} n$, and $\varepsilon = 2^{-\tilde{\gamma} n}$.

(ii) There are constants C_0 and $\tilde{\sigma}, \tilde{\nu} > 0$ so that, with probability at least $1 - 2 \cdot 2^{-4n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$ the following holds: for all independent sources $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$, with $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq \frac{C_0 n}{\log n}$, we have

1. $\mathrm{nmExt}_p^m(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^m(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \mathrm{nmExt}_p^m(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y}$
2. $\mathrm{nmExt}_p^m(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^m(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \mathrm{nmExt}_p^m(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y}$

where $\varepsilon = 2^{-\tilde{\sigma}\sqrt{n}}$ and $m = \tilde{\nu}\sqrt{n}$

From now on, we will focus on proving Lemma 11, which will be covered in Section 4.1-4.3. We will prove Lemma 12 in Section 4.4, which would follow a similar pattern of proof as Lemma 11.

A sneak-peak into the Partition Lemma. For notational ease, let us denote

$$\begin{aligned} \mathcal{D}_p^m(\mathbf{X}, \mathbf{Y}) &:= \mathcal{U}_m, \mathrm{nmExt}_p^m(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y} \\ \phi_p^m(\mathbf{X}, \mathbf{Y}) &:= \mathrm{nmExt}_p^m(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^m(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y} . \end{aligned}$$

For any set $\mathcal{P} \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$, we denote by $\phi(\mathbf{X}, \mathbf{Y})|_{\mathcal{P}}$, the distribution obtained by restricting $(\mathbf{X}, \mathbf{Y}) \in \mathcal{P}$. We will prove Lemma 11 by conditioning ϕ_p^m on various partitions $\mathcal{P}_1, \ldots, \mathcal{P}_k$ of $\mathcal{P}$, showing that *either* the statistical distance in each case between $\phi_p^m(\mathbf{X}, \mathbf{Y})|_{\mathcal{P}_i}$ and $\mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})|_{\mathcal{P}_i}$ is small, *or* the size of the partition $\mathcal{P}_i$ is small. This is enough to imply that $\phi_p^m(\mathbf{X}, \mathbf{Y})$ is *close* to $\mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})$. A more general statement on the statistical distance between two distributions can be stated as follows.

Lemma 13 (Partition Lemma [AHL16, Lemma 3.3]). *Let ψ_1 and ψ_2 be two functions from $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}^r$. Further, let $\mathcal{P} \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$, and let $\mathbf{A}, \mathbf{B} \sim \{0,1\}^n$ be two distributions such that $\sup(\mathbf{A}, \mathbf{B}) \subseteq \mathcal{P}$. Finally, let $\mathcal{P}_1, \ldots, \mathcal{P}_k$ be disjoint partitions of $\mathcal{P}$ such that for all $i \in [k]$:*

$$\Delta(\psi_1(\mathbf{A}, \mathbf{B})|_{(\mathbf{A},\mathbf{B})\in\mathcal{P}_i}; \psi_2(\mathbf{A}, \mathbf{B})|_{(\mathbf{A},\mathbf{B})\in\mathcal{P}_i}) \;\leq\; \varepsilon_i \,.$$

Then the following holds:

$$\Delta(\psi_1(\mathbf{A}, \mathbf{B}); \psi_2(\mathbf{A}, \mathbf{B})) \;\leq\; \sum_i \varepsilon_i \frac{|\mathcal{P}_i|}{|\mathcal{P}|} \,.$$

Proof strategy using Lemma 13 over flat sources. Let us fix $f, g \in \mathcal{F}_n$ for the rest of the subsections. We will argue that for a uniformly chosen polynomial p and k, there exist ε and m, such that with high probability the following holds:

$$\Delta(\phi_p^m(\mathbf{X}, \mathbf{Y}); \mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})) \;\leq\; \varepsilon \,,$$

for all independent k-flat sources $\mathbf{X}, \mathbf{Y}$ with supports X and Y respectively, and $\mathcal{P} = X \times Y$. In our case k will be either ρn, for any constant ρ or $C_0 n/\log n$, for some fixed constant C_0 and m, ε will be according to the value of k. We will analyze both the cases. Importantly, from Lemma 1, one would then infer the same for any two independent sources $\mathbf{X}, \mathbf{Y}$, where $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq k$.

4.1 Inverse image of the image under f (resp. g) is large

In this subsection, we analyze the case when $f(x)$ has a large pre-image (in X) under f.

When $|X|, |Y|$ is $2^{\rho n}$, for some constant ρ. Formally, define:

$$X_{\text{large}} \;:=\; \left\{x \in X : |f^{-1}(f(x)) \cap X| \;\geq\; 2^{\rho n/2}\right\} .$$

In the next lemma, we will prove that the two distributions $\phi_p^m(\mathbf{X}, \mathbf{Y})$ and $\mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})$, when restricted to $X_{\text{large}} \times Y$, are statistically close.

Lemma 14. *Given any $\rho > 0$, there are constants $\beta', \mu' > 0$ such that with probability at least $1 - 2^{-4n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds: for all $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ flat-ρn sources with supports X and Y respectively, we have*

$$\Delta(\phi_p^m(\mathbf{X}, \mathbf{Y})|_{X_{\text{large}}\times Y}; \mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})|_{X_{\text{large}}\times Y}) \;\leq\; 2^{-\beta' n} \,,$$

where $m = \mu' n$.

To prove the above lemma, we will further partition X_{large} into *disjoint union* of sets $A_{z_1}, A_{z_2}, \ldots, A_{z_\ell}$, with $|A_{z_i}| \geq 2^{\rho n/2}$, where,

$$A_{z_i} \;:=\; \left\{x \in X : f(x) = z_i\right\} \cap X_{\text{large}} \,\forall\, i \in [\ell] \,.$$

We will first show that for suitable m, with high probability, over uniformly random choice of p from $\mathcal{H}_{30n,10n}$, the distribution $\phi_p^m(\mathbf{X}, \mathbf{Y})|_{A_{z_i}\times Y}$ is close to $\mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})|_{A_{z_i}\times Y}$.

Lemma 15. *Given any $\rho > 0$, and $z \in \{z_1, \ldots, z_\ell\}$, there exist constants $\beta', \mu' > 0$ such that with probability at least $1 - 2^{-5n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds: for all independent flat-ρn sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ with supports X and Y respectively, we have*

$$\Delta(\phi_p^m(\mathbf{X}, \mathbf{Y})|_{A_z\times Y}; \mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})|_{A_z\times Y}) \;\leq\; 2^{-\beta' n} \,,$$

where $m = \mu' n$,

Proof. Let $\mathbf{A}_z$ denote the uniform distribution over A_z. Since, by definition $f(A_{z_i}) = z_i$ for all $i \in [\ell]$, we have

$$\begin{aligned}\phi_p^m(\mathbf{X},\mathbf{Y})|_{A_z\times Y} &= \mathrm{nmExt}_p^m(\mathbf{A_z},\mathbf{Y}), \mathrm{nmExt}_p^m(f(\mathbf{A_z}), g(\mathbf{Y})), \mathbf{Y} \\ &= \mathrm{nmExt}_p^m(\mathbf{A_z},\mathbf{Y}), \mathrm{nmExt}_p^m(z, g(\mathbf{Y})), \mathbf{Y}.\end{aligned}$$

Since $|A_z| \geq 2^{\rho n/2}$, using Theorem 8, one can conclude that there exist constants $\beta', \mu' > 0$ such that for uniformly chosen p from $\mathcal{H}_{30n,10n}$, the non-malleable extractor nmExt_p^m is a $(n, \mu' n, \rho n/2, 2^{-\beta' n})$ strong 2-source extractor with probability at least $1 - 2^{-5n^2}$ over uniformly random choice of p from $\mathcal{H}_{30n,10n}$. Further, using Lemma 2, we get

$$\begin{aligned}&\Delta((\mathrm{nmExt}_p^m(A_z,\mathbf{Y}),\mathbf{Y}); (\mathcal{U}_m,\mathbf{Y})) \\ &\leq \Delta((\mathrm{nmExt}_p^m(A_z,\mathbf{Y}),\mathbf{Y}, \mathrm{nmExt}_p^m(z,g(\mathbf{Y}))); (\mathcal{U}_m,\mathbf{Y}, \mathrm{nmExt}_p^m(z,g(\mathbf{Y})))).\end{aligned}$$

Hence, $\Delta(\phi_p^m(\mathbf{X},\mathbf{Y})|_{A_z\times Y}; \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{A_z\times Y}) \leq 2^{-\beta' n}$, for $m = \mu' n$ for all flat-ρn source $\mathbf{X},\mathbf{Y}$. This finishes the proof. □

The above lemma can be directly used for the distributions restricted to $X_{\mathrm{large}} \times Y$, to get the desired result of Lemma 14.

Proof of Lemma 14. From Lemma 15, we know that for all $i \in [\ell]$, with probability at least $1 - \ell \cdot 2^{-5n^2}$, over uniformly chosen p from $\mathcal{H}_{30n,10n}$, for all independent flat-ρn sources $\mathbf{X},\mathbf{Y} \sim \mathbb{F}_2^n$ with supports X and Y respectively the following holds:

$$\Delta(\phi_p^m(\mathbf{X},\mathbf{Y})|_{A_{z_i}\times Y}; \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{A_{z_i}\times Y}) \;\leq\; 2^{-\beta' n}\,.$$

Since trivially $\ell < 2^n$, we have $\ell \cdot 2^{-5n^2} \leq 2^{-4n^2}$. Consider $\mathcal{P} = X_{\mathrm{large}} \times Y$ and its partitions $\mathcal{P}_1, \ldots, \mathcal{P}_\ell$ to be $A_{z_1} \times Y, \ldots, A_{z_\ell} \times Y$. Using Lemma 13, our result follows. □

Remark 4. Similar to X_{large}, we define Y_{large} with respect to g as follows:

$$Y_{\mathrm{large}} \;:=\; \{y \in Y : |g^{-1}(g(y)) \cap Y| \;\geq\; 2^{\rho n/2}\}\,.$$

Let us further partition Y_{large} as $B_{v_1} \sqcup \cdots \sqcup B_{v_p}$, where $\forall\ i \in [p]$,

$$B_{v_i} \;:=\; \{y \in Y : g(y) = v_i\} \cap Y_{\mathrm{large}}\,.$$

For any $v \in \{v_1, \ldots, v_p\}$ and any constants $\rho, \rho' > 0$, there are constants $\alpha', \nu' > 0$ such that taking $m = \nu' n$, with probability at least $1 - 2^{-5n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds:

for all independent flat-ρn sources $\mathbf{X},\mathbf{Y}$ with support X, Y respectively, with $X' \subseteq X$ of size at least $2^{\rho' n}$, we have

$$\big(\mathrm{nmExt}_p^m(\mathbf{X},\mathbf{Y}),\mathbf{Y}\big)|_{X'\times B_v} \;\approx_{2^{-\alpha' n}}\; \big(\mathcal{U}_m,\mathbf{Y}\big)|_{X'\times B_v}\,.$$

Note that $g(B_v) = \{v\}$, and hence if we leak $\mathrm{nmExt}(f(\mathbf{X}), g(\mathbf{Y}))|_{X'\times B_v}$, with a little loss in the parameter we get,

$$\big(\mathrm{nmExt}_p^m(\mathbf{X},\mathbf{Y}),\mathbf{Y}\big)|_{X'\times B_v}, \mathrm{nmExt}_p^m(\mathbf{X},v)|_{X'} \approx_\varepsilon \big(\mathcal{U}_m,\mathbf{Y}\big)|_{X'\times B_v}, \mathrm{nmExt}_p^m(\mathbf{X},v)|_{X'}$$

for $\varepsilon = 2^{-\alpha n}$ and $m = \beta n$, where β, α are positive constants. Therefore we get, for any constant $\rho > 0$ we have positive constants α, β such that with $m = \beta n$ the following holds:

For all independent flat-ρn sources $\mathbf{X},\mathbf{Y} \sim \mathbb{F}_2^n$, with supports X and Y respectively and $X' \subseteq X$ of size at least $2^{\rho' n}$,

$$\phi_p^m(\mathbf{X},\mathbf{Y})|_{(X',Y_{\mathrm{large}})} \;\approx_{2^{-\alpha n}}\; \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{(X',Y_{\mathrm{large}})}\,.$$

When $|X|$ and $|Y|$ is at least $2^{C'n/\log n}$, where C' is any constant bigger than $30C$ and C is the constant from item (ii) of Theorem 8, we analyze in the exact same way we did in the previous case. Define,

$$\hat{X}_{\text{large}} := \left\{x \in X : |f^{-1}(f(x)) \cap X| \geq 2^{C'n/2\log n}\right\}$$

$$\hat{Y}_{\text{large}} := \left\{y \in X : |g^{-1}(g(y)) \cap Y| \geq 2^{C'n/2\log n}\right\} .$$

Also define the partitions $\hat{X}_{\text{large}} = A_{z_1} \sqcup \cdots \sqcup A_{z_\ell}$ and $\hat{Y}_{\text{large}} = B_{v_1} \sqcup \cdots \sqcup B_{v_p}$ as before. If we follow the line of proof exactly as shown in the previous case and instead of (i) if we apply item (ii) of Theorem 8, we have the following lemma.

Lemma 16. *1. There exists a constant C' (as defined before) so that for uniformly chosen p from $\mathcal{H}_{30n,10n}$, with probability at least $1 - 2^{-4n^2}$ we have the following: For all $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ flat $C'n/\log n$-sources with supports X and Y respectively,*

$$\Delta\left(\phi_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{large}}\times Y}; \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{large}}\times Y}\right) \leq 2^{-\sqrt{n}/8}$$

where $m = \sqrt{n}/8$.

2. There exists a constant C' (as defined before) so that for any $\tilde{C}$ satisfying $C < \tilde{C} < C'$ and uniformly chosen p from $\mathcal{H}_{30n,10n}$, the following holds with probability at least $1 - 2^{-4n^2}$: For all $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ flat $C'n/\log n$-sources with supports X, Y respectively, and for any $X' \subseteq X$ of size $2^{\tilde{C}n/\log n}$,

$$\Delta\left(\phi_p^m(\mathbf{X},\mathbf{Y})|_{X'\times Y_{\text{large}}}; \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{X'\times Y_{\text{large}}}\right) \leq 2^{-\sqrt{n}/8}$$

where $m = \sqrt{n}/8$.

4.2 Inverse image of the image under f (resp. g) is small

Till now we have taken care of the case when $\phi_p^m(\mathbf{X},\mathbf{Y})$ and $\mathcal{D}_p^m(\mathbf{X},\mathbf{Y})$ are restricted to the partition $X_{\text{large}} \times Y$. Now let us focus on the subsets of X and Y which contain those elements whose images have *small* pre-image under f and g respectively. When $|X| = |Y| = 2^{\rho n}$, define,

$$X_{\text{small}} := \{x \in X : 0 < |f^{-1}(f(x)) \cap X| < 2^{\rho n/2}\} \subseteq X,$$
$$Y_{\text{small}} := \{y \in Y : 0 < |g^{-1}(g(y)) \cap Y| < 2^{\rho n/2}\} \subseteq Y .$$

And when $|X| = |Y| = 2^{\hat{C}n/\log n}$ where $\hat{C} \geq 30C$,

$$\hat{X}_{\text{small}} := \left\{x \in X : 0 < |f^{-1}(f(x)) \cap X| < 2^{\hat{C}n/2\log n}\right\} \subseteq X,$$
$$\hat{Y}_{\text{small}} := \left\{y \in Y : 0 < |g^{-1}(g(y)) \cap Y| < 2^{\hat{C}n/2\log n}\right\} \subseteq Y .$$

We will analyze the first case in detail. The analysis of the second case will be almost same to the first one. We will make remarks in suitable places to mention the formal statement and the small changes that we will need to prove the second case.

When $|X|$ and $|Y|$ are exponential (linear min-entropy), i.e. $|X| = |Y| = 2^{\rho n}$. In this case, our main goal is to prove the following:

Lemma 17. *Fix $f, g \in \mathcal{F}_n$. For $X, Y \subseteq \mathbb{F}_2^n$ with $|X| = |Y| = 2^{\rho n}$ and X_{small}, Y_{small}, defined as before, consider the following property:*

$$|X_{\text{small}} \times Y_{\text{small}}| \geq 2^{1.7\rho n} , \tag{2}$$

where $\rho > 0$ is some constant. Then, there are positive constants β'', μ'' such that for $m = \mu'' n$, with probability at least $1 - 2^{-4n^2}$ over uniformly random choice of p from $\mathcal{H}_{30n,10n}$ the following holds:

For all $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ such that $|X| = |Y| = 2^{\rho n}$ and satisfies property 2,

$$\Delta\Big(\phi_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{small}}\times Y_{\text{small}}}; \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{small}}\times Y_{\text{small}}}\Big) \leq 2^{-\beta'' n} .$$

In order to prove the above lemma, we start with a small technical claim that will be used later.

Lemma 18. *Let $X, Y, X_{\text{small}}, Y_{\text{small}}$ be defined as above with $|X_{\text{small}}|, |Y_{\text{small}}| \geq 2^{0.7\rho n}$. Then, the following two properties hold.*

1. *There exists a subset $X^{''} \subseteq X_{\text{small}}$, with $|X^{''}| \geq 2^{0.1\rho n}$, such that all elements in X'' and $f(y)$, for all $y \in X^{''}$ are* distinct. *That is, $f|_{X''}$ is injective, and for any $y \in X^{''}$, $y \notin f(X^{''})$.*
2. *There exists a subset $Y^{''} \subseteq Y_{\text{small}}$, with $|Y^{''}| \geq 2^{0.1\rho n}$, such that all elements in Y'' and $g(z)$ for all $z \in Y^{''}$ are* distinct. *That is, $g|_{Y''}$ is injective, and for any $z \in Y^{''}$, $z \notin f(Y^{''})$.*

Proof. Consider the following algorithm (see Algorithm 1).

Algorithm 1 Finding a small subset $X^{''}$

1: $X^{''} \leftarrow \emptyset$
2: **while** $X_{\text{small}} \neq \emptyset$ **do**
3: Pick x from X_{small}
4: $X^{''} \leftarrow X^{''} \cup \{x\}$
5: $S \leftarrow \{x' \in X_{\text{small}} : f(x') = f(x)\} \cup \{y \in X_{\text{small}} : f(y) = x\} \cup \{f(x)\}$
6: $X_{\text{small}} \leftarrow X_{\text{small}} \backslash S$
7: Return $X^{''}$

In an iteration of the Algorithm 1 (Line 5), if we insert x inside $X^{''}$, then by the step specified, we remove $f^{-1}(f(x))$, $f^{-1}(x) \cap X_{\text{small}}$ and $f(x)$. Therefore, $X^{''}$ satisfies the required property. Further, by assumption, we have

$$|f^{-1}(f(x))| \ < \ 2^{\rho n/2} \quad \text{and} \quad |f^{-1}(x) \cap X_{\text{small}}| < 2^{\rho n/2} \ .$$

Hence, in each iteration, at most $2^{0.5\rho n+1} + 1$ many elements are removed from X_{small}. On the other hand, by assumption, $|X_{\text{small}}| \geq 2^{0.7\rho n}$. Therefore,

$$|X^{''}| \times (2^{0.5\rho n+1} + 1) \ \geq \ 2^{0.7\rho n} \implies |X^{''}| \ \geq \ 2^{0.1\rho n} \ .$$

This proves our claim for $X^{''}$. Similarly, we can find $Y^{''} \subseteq Y_{\text{small}}$ with the desired properties. □

Remark 5. When $|X| = |Y| = 2^{\hat{C}n/\log n}$ for $\hat{C} \geq 30C$ and $|\hat{X}_{\text{small}}|, |\hat{Y}_{\text{small}}| \geq 2^{0.7\hat{C}n/\log n}$, in the same way as above we can show there are $S_1 \subseteq \hat{X}_{\text{small}}$ and $S_2 \subseteq \hat{Y}_{\text{small}}$ so that $|S_1|, |S_2| \geq 2^{0.1\hat{C}n/\log n}$ and they satisfy same property as X'' and Y'' do.

We now follow the proof strategy previously employed in Section 3 for the randomized construction of strong 2-source extractors. First, we define the following.

Definition 12. *For any set $X \subseteq \mathbb{F}_2^n$, a function $f \in \mathcal{F}_n$ and a positive integer $u \geq 2$, and a polynomial $p \in \mathcal{H}_{3un,un}$, define*

$$p\left(X_f^{(u)}\right) \ := \ \left\{p(0^{(u-1)n}x) \circ p(0^{(u-1)n}f(x)) : x \in X \ \right\} ,$$

where

$$X_f^{(u)} \ := \ \{(0^{(u-1)n} \circ x) \circ (0^{(u-1)n} \circ f(x)) : x \in X\} \ .$$

30-feet above proof overview in two lines. We will show that $\dim((X'')_f^{(10)})$ and $\dim((Y'')_g^{(10)})$ are large, which in turn will imply that $\dim((X_{\text{small}})_f^{(10)})$ and $\dim((Y_{\text{small}})_g^{(10)})$ are also large. To show this we will prove a more general statement that all the sets which satisfy the property given in Lemma 18, have large dimension. For that we need the following formal definition.

Property P_f. For a set $A \subseteq \{0,1\}^n$ and $f \in \mathcal{F}_n$, we say A satisfies property P_f if

$$f|_A \text{ is injective and } a \notin f(A) \ \forall\ a \in A. \tag{3}$$

Theorem 10. *Let $f \in \mathcal{F}_n$. Let $A \subseteq \mathbb{F}_2^n$, such that $|A| \geq 2un$, for a positive integer $u \geq 2$. Further, assume that A satisfies property P_f. Then for any $d \leq 1.4un$, the following holds.*

$$\Pr_{p \leftarrow \mathcal{H}_{3un,un}}[\dim_{\mathbb{F}_2}(p(A_f^{(u)})) \ \leq\ d] \ \leq\ 2^{2un-(2un-d)^2}\ .$$

Proof. We follow the exact proof structure of Lemma 9. Let $A' := \{x_1, \ldots, x_{2un}\}$, where x_i, for $i \in [2un]$ are arbitrarily chosen *distinct* elements from A. Let $B := \{x_1, \ldots, x_d\}$. We will prove an upper bound on the probability that $\dim_{\mathbb{F}_2}(p(A_f'^{(u)})) \leq d$, which will readily imply the desired upper bound. For some $x \in A$, define:

$$p_{f,x}^u \ :=\ p(0^{(u-1)n}x) \circ p(0^{(u-1)n}f(x))\ .$$

Pick any $\mathbf{t} = (t_1, \ldots, t_d) \in \mathbb{F}_{2^{kn}}^d$, and $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{F}_{2^{cn}}^d$. For $\mathbf{a} = (a_1, \ldots, a_d) \in \mathbb{F}_2^d$, define the following events:

1. $\mathcal{E}_{\mathbf{t}}^{\mathbf{a}} := \left(\bigwedge_{i=1}^d p(0^{(u-1)n}x_i) = t_i\right) \bigwedge \left(p(0^{(u-1)n}z) = \sum_{i=1}^d a_i t_i\right)$.
2. $\Gamma_{\mathbf{z}}^{\mathbf{a}} \ :=\ \left(\bigwedge_{i=1}^d p(0^{(u-1)n}f(x_i)) = k_i\right) \ \bigwedge \ \left(p(0^{(u-1)n}f(z)) \ =\ \sum_{i=1}^d a_i k_i\right)$.

Pick any $z \in A' \backslash B$. Then,

$$\begin{aligned}
\Pr_{p \leftarrow \mathcal{H}_{3un,un}}[p_{f,z}^u \in \text{Span}_{\mathbb{F}_2}(p_{f,x_1}^u, \ldots, p_{f,x_d}^u)] \ &\leq\ \sum_{\mathbf{a}\ \in\ \mathbb{F}_2^d} \left(\sum_{\mathbf{t},\mathbf{k}\ \in\ \mathbb{F}_{2^{un}}^d} \Pr_{p \leftarrow \mathcal{H}_{3un,un}} \left[\mathcal{E}_{\mathbf{t}}^{\mathbf{a}} \wedge \Gamma_{\mathbf{k}}^{\mathbf{a}}\right] \right) \\
&=\ \sum_{\mathbf{a}\ \in\ \mathbb{F}_2^d} \left(\sum_{\mathbf{t},\mathbf{k}\ \in\ \mathbb{F}_{2^{un}}^d} \frac{1}{2^{2un(d+1)}} \right) \\
&=\ \frac{2^{2und+d}}{2^{2un(d+1)}} \ =\ \frac{1}{2^{2un-d}}\ .
\end{aligned}$$

Since A satisfies property 3, all of $x_1, \ldots, x_d$ and $f(x_1), \ldots, f(x_d)$ must be distinct. Since $2d \leq 3un$, we can use the property of $(2d)$-wise hash family in the first equality above.

Further, the probability that $p_{f,z}^u$ is in $\text{Span}_{\mathbb{F}_2}(B_f^{(u)})$, for all $z \in A' \backslash B$, is $2^{-(2un-d)^2}$, since p is chosen from a $(3un)$-wise independent hash family.

Finally, we need to take a union bound on all such $B \subseteq A'$ which in turn will give the final probability bound of $2^{2un-(2un-d)^2}$, as desired.

□

A similar calculation as above shows the following; we state it without proving in detail.

Lemma 19. *Let $f \in \mathcal{F}_n$. Let $A \subseteq \mathbb{F}_2^n$, such that $|A| \geq 50n$. Further, assume that A satisfies property P_f. Then, the following holds.*

$$\Pr_{p \leftarrow \mathcal{H}_{300n,10n}}[\dim_{\mathbb{F}_2}(p(A_f^{(10)})) \ \leq\ 10n] \ \leq\ 2^{50n-400n^2}\ .$$

By applying the union bound over all set of size $50n$, we get the following corollary.

Corollary 2. *Let $f \in \mathcal{F}_n$. Then, with probability at most 2^{-349n^2} over the choice of uniformly random p from $\mathcal{H}_{300n,10n}$ we have,*

$$\exists\, A \subseteq \mathbb{F}_2^n \text{ of size at least } 50n \text{ that satisfies } P_f \,, \dim_{\mathbb{F}_2}(p(A_f^{(10)})) \;\leq\; 10n \,.$$

Following the (exact) same line of proof that has been provided for the above, it is easy to see that in fact picking a polynomial of degree $300n$, sampled from a source of min-entropy rate 0.9 *still works*!

In the next remark, we argue that all the above probability calculations hold if one multiplies the space by a full-rank matrix.

Remark 6. Since $\mathrm{Span}_{\mathbb{F}_2}(p(A_f^{(u)})) \subseteq \mathbb{F}_2^{2un}$, for any full-rank matrix $M \in \mathbb{F}_2^{2un \times 2un}$, we have

$$M \cdot \mathrm{Span}_{\mathbb{F}_2}(p(A_f^{(u)})) \;=\; \mathrm{Span}_{\mathbb{F}_2}(M \cdot p(A_f^{(u)})) \;=\; \mathrm{Span}_{\mathbb{F}_2}(p(A_f^{(u)})).$$

Hence, for given A, f, c, d same as in Theorem 10 and any full-rank matrix $M \in \mathbb{F}_2^{2un \times 2un}$, we also have,

$$\Pr_{p \leftarrow \mathcal{H}_{3un,un}}[\dim_{\mathbb{F}_2}(M \cdot p(A_f^{(u)})) \;\leq\; d] \;\leq\; 2^{2un-(2un-d)^2} \,.$$

Lemma 20. *Let $f \in \mathcal{F}_n$ and let $\rho > 0$ be any constant. Further let M be any $20n \times 20n$ full-rank matrix over $\mathbb{F}_2$. Then with probability at least $1 - 2^{-5n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds: for all $X \subseteq \mathbb{F}_2^n$ of size $2^{\rho n}$ with $|X_{\mathrm{small}}| \geq 2^{0.7\rho n}$ we have,*

$$\dim_{\mathbb{F}_2}\left(M \cdot p\big((X_{\mathrm{small}})_f^{(10)}\big)\right) \;\geq\; 12n \,.$$

Proof. Substitute $u = 10$ and $d = 12n$ in Theorem 10 to get that with probability at most 2^{-50n^2}, over the uniformly random choice of p from $\mathcal{H}_{30n,10n}$, we have $\dim_{\mathbb{F}_2}(p(A_f^{(10)})) \leq 12n$. By applying the union bound over all $A \subseteq \mathbb{F}_2^n$ of size $20n$, we obtain the following.

With probability at most 2^{-5n^2}, where the probability is over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, there exists $A \subseteq \mathbb{F}_2^n$, of size at least $20n$, where A satisfies property P_f such that

$$\dim_{\mathbb{F}_2}\left(M \cdot p(A_f^{(10)})\right) \;\leq\; 12n \,.$$

Recall that, we are only interested in $X \subseteq \mathbb{F}_2^n$ such that $|X| = \rho n$ and $|X_{\mathrm{small}}| \geq 2^{0.7\rho n}$. By Lemma 18, we know that there exists $A \subseteq X_{\mathrm{small}}$ with $|A| \geq 20n$, such that A satisfies property P_f. Therefore, the following holds:

$$\exists X \subseteq \mathbb{F}_2^n, \text{ of size } 2^{\rho n}, \text{ where } |X_{\mathrm{small}}| \geq 2^{0.7\rho n}, \text{ such that}$$

$$\dim_{\mathbb{F}_2}\left(p\big((X_{\mathrm{small}})_f^{(10)}\big)\right) \;\leq\; 12n \,,$$

with probability at most 2^{-5n^2}, where the probability is over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$. Since M is a full rank matrix, our lemma follows. □

Now we will formally argue that the inner-product of the image of $(X_{\mathrm{small}})_f^{(10)}$ and image of $(Y_{\mathrm{small}})_g^{(10)}$ under p is *non-constant*. From there, we show that a random polynomial can be used to construct a strong extractor on the sources $(X_{\mathrm{small}})_f^{(10)}$ and $(Y_{\mathrm{small}})_f^{(10)}$.

By $(\mathbf{X}_{\mathrm{small}})_f^{(10)}$ and $(\mathbf{Y}_{\mathrm{small}})_g^{(10)}$, we denote uniform distributions over $(X_{\mathrm{small}})_f^{(10)}$ and $(Y_{\mathrm{small}})_g^{(10)}$ respectively.

Corollary 3. *Fix any $f, g \in \mathcal{F}_n$. Let M be any $20n \times 20n$ full-rank matrix over $\mathbb{F}_2$. Then for any constant $\rho > 0$, there is another constant $\beta_1 > 0$ such that with probability at least $\geq 1-2^{-4.5n^2}$, over uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds: For $X, Y \subseteq \mathbb{F}_2^n$ such that $|X| = |Y| = 2^{\rho n}$ and $|X_{\text{small}}|, |Y_{\text{small}}| \geq 2^{0.7\rho n}$, we have,*

$$\Delta\Big(\langle p((\mathbf{X}_{\text{small}})_f^{(10)}), M \cdot p(\mathbf{Y}_{\text{small}})_g^{(10)})\rangle, \mathbf{Y}_{\text{small}} \; ; \; \mathcal{U}_1, \mathbf{Y}_{\text{small}}\Big) \; \leq \; 2^{-\beta_1 n} \; .$$

Proof. By Lemma 20, we know that with probability at least $1 - 2^{-5n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds: for all $X \subseteq \mathbb{F}_2^n$ of size $2^{\rho n}$ with $|X_{\text{small}}| \geq 2^{0.7\rho n}$,

$$\dim_{\mathbb{F}_2}\left(M \cdot p\big((X_{\text{small}})_f^{(10)}\big)\right) \; \geq \; 12n \; .$$

And the similar will hold for all $Y \subseteq \mathbb{F}_2^n$ with $|Y| = 2^{\rho n}$ and $|Y_{\text{small}}| \geq 2^{0.7\rho n}$, with probability at least $1 - 2 \cdot 2^{5n^2}$ if p is chosen uniformly from $\mathcal{H}_{30n,10n}$. If both the dimensions: $\dim_{\mathbb{F}_2}(p((X_{\text{small}})_f^{(10)}))$ and $\dim_{\mathbb{F}_2}(M \cdot p((Y_{\text{small}})_g^{(10)}))$ are at least $12n$, then trivially their sum is *strictly* larger than $20n + 1$.

Therefore, applying Lemma 6, one concludes that with probability at least $1-2\cdot 2^{-5n^2}$, over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds:

For all $X, Y \subseteq \mathbb{F}_2^n$, each of size $2^{\rho n}$, where both $|X_{\text{small}}|$, and $|Y_{\text{small}}|$ are at least $2^{0.7\rho n}$, we have

$$\langle p((X_{\text{small}})_f^{(10)}), p((Y_{\text{small}})_g^{(10)})\rangle \;\text{ is non-constant.}$$

Note that, both $|(X_{\text{small}})_f^{(10)}|, |(Y_{\text{small}})_g^{(10)}| \geq 2^{0.7\rho n}$. Finally, we can replace ρ by 0.7ρ in Theorem 7, and get our desired result. □

From the proof idea of Lemma 20 and Corollary 3 and using Remark 5, item (ii) of Theorem 7 we can have the following lemma, which we are stating without formal proof.

Lemma 21. *Fix any $f, g \in \mathcal{F}_n$ and M be any $20n \times 20n$ full rank matrix over $\mathbb{F}_2$. For any $\hat{C} \geq 30C$ where C is from Theorem 8, with probability at least $1 - 2^{-4.5n^2}$ over uniformly random choice of p from $\mathcal{H}_{30n,10n}$ we have: For all $X, Y \subseteq \mathbb{F}_2^n$ so that $|X| = |Y| = 2^{\hat{C}n/\log n}$ and $|\hat{X}_{\text{small}}|, |\hat{Y}_{\text{small}}| \geq 2^{0.7\hat{C}n/\log n}$,*

$$\Delta\Big(\langle p((\hat{\mathbf{X}}_{\text{small}})_f^{(10)}), M \cdot p(\hat{\mathbf{Y}}_{\text{small}})_g^{(10)})\rangle, \hat{\mathbf{Y}}_{\text{small}} \; ; \; \mathcal{U}_1, \hat{\mathbf{Y}}_{\text{small}}\Big) \; \leq \; 2^{-\sqrt{n}/8} \; .$$

Here $\hat{\mathbf{X}}_{\text{small}}$ and $\hat{\mathbf{Y}}_{\text{small}}$ are uniform distributions over $\hat{X}_{\text{small}}$ and $\hat{Y}_{\text{small}}$ respectively.

As, $\hat{C} \geq 30C$, we have $0.1 \times 0.7 \times \hat{C} > C$, we can impose item (ii) of Theorem 8.

Now we are ready to prove Lemma 17 (main lemma of this subsection). For completeness, we restate the lemma here.

Lemma 22. *Fix $f, g \in \mathcal{F}_n$. For $X, Y \subseteq \mathbb{F}_2^n$ with $|X| = |Y| = 2^{\rho n}$ and X_{small}, Y_{small} are as defined before, consider the following property,*

$$|X_{\text{small}} \times Y_{\text{small}}| \geq 2^{1.7\rho n} \; , \tag{4}$$

where $\rho > 0$ is some constant. Then, there are positive constants β'', μ'' such that for $m = \mu'' n$, with probability at least $1 - 2^{-4n^2}$ over uniformly random choice of p from $\mathcal{H}_{30n,10n}$ the following holds: For all $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ such that $|X| = |Y| = 2^{\rho n}$ and satisfies Equation (4),

$$\Delta\Big(\phi_p^m(\mathbf{X}, \mathbf{Y})|_{X_{\text{small}} \times Y_{\text{small}}}; \mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})|_{X_{\text{small}} \times Y_{\text{small}}}\Big) \leq 2^{-\beta'' n} \; .$$

Proof. If for $X, Y \subseteq \mathbb{F}_2^n$, both of size ρn, we have $2^{1.7\rho n} \leq |X_{\text{small}} \times Y_{\text{small}}| \leq |X_{\text{small}}| \times 2^{\rho n}$, we can easily conclude that

$$|X_{\text{small}}| \ \geq \ 2^{0.7\rho n}, \text{ and } |Y_{\text{small}}| \ \geq \ 2^{0.7\rho n} \ .$$

For this proof, we introduce the following notations:

We use $\mathbf{X}_{\text{small}}$ and $(\mathbf{X}_{\text{small}})^{(10)}$ to denote uniform distribution over X_{small} and $(X_{\text{small}})^{(10)}$ respectively. And similarly, we use notations $\mathbf{Y}_{\text{small}}$ and $(\mathbf{Y}_{\text{small}})^{(10)}$ to denote uniform distribution over Y_{small} and $(Y_{\text{small}})^{(10)}$ respectively.

We will show a more general statement that there exists a constant $\beta'' > 0$ (see Equation (5)), such that with probability at least $1 - 2^{-4n^2}$, over uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds:

For all $X, Y \subseteq \mathbb{F}_2^n$ where $|X| = |Y| = 2^{\rho n}$ such that they satisfy Equation (4), we have

$$\Delta\Big(\phi_p^m(\mathbf{X}, \mathbf{Y})|_{X_{\text{small}} \times \mathbf{Y}_{\text{small}}}; (\mathcal{U}_m, \mathcal{U}_m, \mathbf{Y}_{\text{small}})\Big) \ \leq \ 2^{-\beta'' n} \ . \tag{5}$$

To show the above, we will apply Vazirani's XOR lemma (Lemma 5). We need to show that for any $a_1, \ldots, a_m$, $b_1, \ldots, b_m \in \mathbb{F}_2$, where not all of them are zero, with probability at least $1-2^{-4n^2}$, over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, we have:

For all $X, Y \subseteq \mathbb{F}_2^n$ both of size $2^{\rho n}$ satisfying *Equation* (4),

$$\begin{aligned}
&\Delta\Bigg(\Big(\sum_{i=1}^{m} a_i \langle p((\mathbf{X}_{\text{small}})^{(10)}), L_i \cdot p((\mathbf{Y}_{\text{small}})^{(10)}) \rangle \\
&+ \sum_{i=1}^{m} b_i \langle p((f(\mathbf{X}_{\text{small}}))^{(10)}), L_i \cdot p((g(\mathbf{Y}_{\text{small}}))^{(10)}) \rangle, \mathbf{Y}_{\text{small}} \Big); \Big(\mathcal{U}_1, \mathbf{Y}_{\text{small}} \Big) \Bigg) \\
&\leq 2^{-\beta n} \ ,
\end{aligned}$$

for some $\beta > 0$. Fix arbitrary $a_1, \cdots, a_n, b_1, \cdots, b_n \in \mathbb{F}_2$, where not all of them are zero. Observe that the above distribution of the sum of the output bits of nmExt_p^m reduces to the following distribution

$$\langle p((\mathbf{X}_{\text{small}})^{(10)}), L' \cdot p((\mathbf{Y}_{\text{small}})^{(10)}) \rangle \ + \ \langle p((f(\mathbf{X}_{\text{small}}))^{(10)}), L'' \cdot p((g(\mathbf{Y}_{\text{small}}))^{(10)}) \rangle \ ,$$

where $L' := \sum_{i \in [m]} a_i L_i$ and $L^{''} := \sum_{i \in [m]} b_i L_i$.

We will denote the above distribution by $\psi_{L',L''}(\mathbf{X}_{\text{small}}, \mathbf{Y}_{\text{small}})$, for the rest of the subsection.

Since, by assumption, not all of $a_1, \cdots, a_n, b_1, \cdots, b_n$ are zero, and since L_i are independent matrices, we know that *at least* one of L' or $L^{''}$ must be full-rank. In particular, either both are nonzero full-rank matrices, or one is 0, while the other remains a full-rank matrix. We will go over all the 3 cases below.

Case 1: Both $L^{'}, L^{''}$ are full-rank.

Proof of Case 1. In this case, we have $\psi_{L',L''}(\mathbf{X}_{\text{small}}, \mathbf{Y}_{\text{small}})$

$$\begin{aligned}
&= \langle p(0^{9n} || \mathbf{X}_{\text{small}}), L' \cdot p(0^{9n} || \mathbf{Y}_{\text{small}}) \rangle \ + \ \langle p(0^{9n} || f(\mathbf{X}_{\text{small}})), L^{''} \cdot p(0^{(9n} || g(\mathbf{Y}_{\text{small}})) \rangle \\
&= \langle p(0^{9n} || \mathbf{X}_{\text{small}}) \circ p(0^{9n} || f(\mathbf{X}_{\text{small}})), M \cdot p(0^{9n} || \mathbf{Y}_{\text{small}}) \circ p(0^{9n} || g(\mathbf{Y}_{\text{small}})) \rangle \ ,
\end{aligned}$$

where

$$M \ := \ \begin{pmatrix} L' & \mathbf{0} \\ \mathbf{0} & L'' \end{pmatrix}_{20n \times 20n} \ .$$

Note that M is full-rank since both L' and $L^{''}$ are full-rank matrices (by assumption).

Therefore from Corollary 3 we can conclude, with probability at least $1-2^{-4.5n^2}$ over uniformly random choice of p from $\mathcal{H}_{30n,10n}$, we have the following: For all X, Y such that both of size ρn and satisfy *Equation* (4),

$$\Delta\Big(\psi_{L',L''}(\mathbf{X}_{\text{small}}, \mathbf{Y}_{\text{small}}), \mathbf{Y}_{\text{small}}\ ;\ \mathcal{U}_1, \mathbf{Y}_{\text{small}}\Big)\ \leq\ 2^{-\beta_1 n}\ .$$

Case 2: L' is full-rank and $L^{''} = \mathbf{0}$.

Proof of Case 2. In this case,

$$\psi_{L',L''}(\mathbf{X}_{\text{small}}, \mathbf{Y}_{\text{small}})\ =\ \langle p((\mathbf{X}_{\text{small}})^{(10)}), L' \cdot p((\mathbf{Y}_{\text{small}})^{(10)})\rangle\ .$$

Therefore, if we start with the min-entropy of the input sources $0.7\rho n$, and use Theorem 7, we get a constant $\beta_2 > 0$ such that, with probability at least $1 - 2^{-5n^2}$ over uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds:

for all $X, Y \subseteq \mathbb{F}_2^n$ such that $|X| = |Y| = 2^{\rho n}$ and satisfy Equation (4), we have

$$\Delta\Big(\langle p((\mathbf{X}_{\text{small}})^{(10)}), p((\mathbf{Y}_{\text{small}})^{(10)}), \mathbf{Y}_{\text{small}}\ ;\ \mathcal{U}_1, \mathbf{Y}_{\text{small}}\Big)\ \leq\ 2^{-\beta_2 n}\ ,$$

which is what we wanted.

Case 3: $L' = \mathbf{0}$ and $L^{''}$ is full-rank.

Proof of Case 3. In this case,

$$\psi_{L',L''}(\mathbf{X}_{\text{small}}, \mathbf{Y}_{\text{small}})\ =\ \langle p((f(\mathbf{X}_{\text{small}})^{(10)}), L'' \cdot p((g(\mathbf{Y}_{\text{small}})^{(10)})\rangle\ .$$

We consider our random sources as $f(\mathbf{X}_{\text{small}})$ and $g(\mathbf{Y}_{\text{small}})$. Let $X, Y \subseteq \mathbb{F}_2^n$, each of size $2^{\rho n}$ which satisfy Equation (4). Using Lemma 18, we get that there exists $X^{''} \subseteq X_{\text{small}}$, and $Y^{''} \subseteq Y_{\text{small}}$, each of size at least $2^{0.1\rho n}$, such that both $f|_{X''}$ and $g|_{Y''}$ are injective.

If we replace A by $f(X^{''})$, and $c = 10$, $d = 6n$ in Lemma 9, we get that for $X^{''}$, the following holds:

$$\Pr_{p \leftarrow \mathcal{H}_{30n,10n}}\left[\dim_{\mathbb{F}_2}\left(p((f(X^{''})^{(10)})\right) \leq 6n\right]\ \leq\ 2^{-15n^2}\ .$$

Since both $|f(X_{\text{small}})|$ and $|g(Y_{\text{small}})|$ are at least $2^{0.1\rho n}$, from Remark 2, there exist a constant $\beta_3 > 0$ such that, with probability at least $1 - 2^{-4.5n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$, the following holds:

for all $X, Y \subseteq \mathbb{F}_2^n$ each of size $2^{\rho n}$ and $|X_{\text{small}}|$ and $|Y_{\text{small}}|$ at least $2^{0.7\rho n}$, we have

$$\Delta\Big(\langle p((f(\mathbf{X}_{\text{small}}))^{(10)}), L^{''} \cdot p((g(\mathbf{Y}_{\text{small}}))^{(10)})\rangle, \mathbf{Y}_{\text{small}}\ ;\ \mathcal{U}_1, \mathbf{Y}_{\text{small}}\Big)\ \leq\ 2^{-\beta_3 n}\ .$$

This finishes the proof of Case 3.

Finishing off the proof. Therefore, case $1-3$, combined with XOR lemma (Lemma 5), yields that with probability at least $\geq 1 - 2^{-4n^2}$, over the uniformly random choice of p from $\mathcal{H}_{30n,10n}$, the following holds:

$$\begin{aligned}&\text{For all } X, Y \subseteq \mathbb{F}_2^n \text{ each of size } 2^{\rho n} \text{ and satisfying 2}\\ &\Delta_1\big(\phi_p^m(\mathbf{X}, \mathbf{Y}))|_{X_{\text{small}} \times Y_{\text{small}}}; (\mathcal{U}_m, \mathcal{U}_m, \mathbf{Y}_{\text{small}})\big) \leq 2^{-\beta n/2} \cdot 2^m\ ,\end{aligned}$$

where $\beta'' := \min\{\beta_1, \beta_2, \beta_3\}$ and take $m = \mu'' n = \beta n/4$. This finishes the proof of our lemma. □

When $|X| = |Y| = 2^{\hat{C} n/\log n}$, again by the similar proof idea as above, from Lemma 21, item (ii) of Theorem 7 and Theorem 8 we have,

Lemma 23. *There is a constant $\hat{C} \geq 30C$ and $\hat{\sigma}, \hat{\nu} > 0$ so that with probability at least $1 - 2^{-4n^2}$ over the uniformly choice of p from $\mathcal{H}_{30n,10n}$ we have: For all $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$ with supports X, Y respectively so that $|X| = |Y| = 2^{\hat{C} n/\log n}$ and $|\hat{X}_{\text{small}} \times \hat{Y}_{\text{small}}| \geq 2^{1.7\hat{C} n/\log n}$,*

$$\Delta\left(\phi_p^m(\mathbf{X}, \mathbf{Y})|_{X_{\text{small}} \times Y_{\text{small}}}; \mathcal{D}_p^m(\mathbf{X}, \mathbf{Y})|_{X_{\text{small}} \times Y_{\text{small}}}\right) \leq 2^{-\hat{\sigma}\sqrt{n}}$$

where $m = \hat{\nu}\sqrt{n}$.

4.3 Putting it all together: Proof of Lemma 11

Till now, we have shown that our nmExt_p^m *works* with overwhelming probability, over the uniformly random choice of p, on disjoint partitions $\{\mathcal{P}_i\}_i$ of $X \times Y \subseteq \mathbb{F}_2^n$. That is :

$$\text{for all } i,\ \Delta(\phi_p^m(\mathbf{X},\mathbf{Y})|_{\mathcal{P}_i}; \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{\mathcal{P}_i}) \ \leq\ \varepsilon_i\ .$$

where ε_i's can be replaced by $2^{-O(n)}$ when input sources have entropy $O(n)$ and by $2^{-O(\sqrt{n})}$ when input sources have entropy $O(n/\log n)$ respectively. Using Lemma 13, we will argue that for any two independent sources $\mathbf{X}$ and $\mathbf{Y}$ each having *sufficiently large* support, our non-malleable extractor still works.

Proof of Lemma 11.
(i) When $X, Y \subseteq \mathbb{F}_2^n$, each of size $2^{\rho n}$, recall the definitions of the partition of X and Y with respect to f, g respectively.

$$\begin{aligned} X_{\text{large}} &= \left\{x \in X : |f^{-1}(f(x)) \cap X| \geq 2^{\rho n/2}\right\}, \\ X_{\text{small}} &= \left\{x \in X : 0 < |f^{-1}(f(x)) \cap X| < 2^{\rho n/2}\right\}. \end{aligned}$$

Similarly for Y, we had defined the following:

$$\begin{aligned} Y_{\text{large}} &= \left\{y \in Y : |g^{-1}(g(y)) \cap Y| \geq 2^{\rho n/2}\right\}, \\ Y_{\text{small}} &= \left\{y \in Y : 0 < |g^{-1}(g(y)) \cap Y| < 2^{\rho n/2}\right\}. \end{aligned}$$

Note that, $X = X_{\text{large}} \sqcup X_{\text{small}}$ and $Y = Y_{\text{large}} \sqcup Y_{\text{small}}$. We consider two cases: (1) when $|X_{\text{small}}| = 2^{\Omega(n)}$, and (2) when $|X_{\text{small}}| = 2^{o(n)}$. We solve both cases one by one.

1. **Case I $\left(|X_{\text{small}}| \geq 2^{\rho' n}$ for $\rho \geq \rho' > 0\right)$:** Consider the partition,

   $$X \times Y \ =\ \left(X_{\text{large}} \times Y\right) \sqcup \left(X_{\text{small}} \times Y_{\text{large}}\right) \sqcup \left(X_{\text{small}} \times Y_{\text{small}}\right).$$

   Since $|X_{\text{small}}| \geq 2^{\rho' n}$ for some constant $\rho' > 0$, using Remark 4, we know that there are constants $\nu, \alpha > 0$ such that with probability at least $1 - 2^{-4n^2}$, over uniformly random choice of p from $\mathcal{H}_{30n,10n}$, the following holds:

   for all independent flat-ρn sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$, with supports respectively X and Y, and further with, $|X_{\text{small}}| \geq 2^{\rho' n}$, we have

   $$\Delta(\phi_p^{\nu n}(\mathbf{X},\mathbf{Y})|_{X_{\text{small}} \times Y_{\text{large}}}; \mathcal{D}_p^{\nu n}(\mathbf{X},\mathbf{Y})|_{X_{\text{small}} \times Y_{\text{large}}}) \ \leq\ 2^{-\alpha n}\ .$$

   We further divide this case into two sub-cases, as follows.

   (i) **Sub-case I $\left(|X_{\text{small}} \times Y_{\text{small}}| \leq 2^{1.7\rho n}\right)$:**
   For a constant $\rho > 0$, we know there are constants $\beta', \mu' > 0$, from Lemma 14. Further, let α, ν be the positive constants as mentioned above. Finally, consider $\mu_1 := \min\{\nu, \mu'\}$ and $m := \mu_1 n$.

   Using these parameters, along with Lemma 13, we get that with probability at least $1 - 3 \cdot 2^{-4n^2}$, for all independent flat-ρn sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$, with support X and Y respectively, where $|X_{\text{small}}| \geq 2^{\rho' n}$ and $|X_{\text{small}} \times Y_{\text{small}}| \leq 2^{1.7\rho n}$, the following holds:

   $$\begin{aligned} &\Delta\left(\phi_p^m(\mathbf{X},\mathbf{Y}), \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})\right) \\ &\leq\ 2^{-\beta' n} \cdot \frac{|X_{\text{large}} \times Y|}{|X \times Y|} \ +\ 2^{-\alpha n} \cdot \frac{|X_{\text{small}} \times Y_{\text{large}}|}{|X \times Y|} \ +\ \frac{|X_{\text{small}} \times Y_{\text{small}}|}{|X \times Y|} \\ &\leq\ 2^{-\beta' n} \ +\ 2^{-\alpha n} \ +\ 2^{-0.3\rho n} \ \leq\ 2^{-\hat{\gamma}_1 n}\ , \end{aligned}$$

   where $\hat{\gamma}_1 := \frac{1}{2} \cdot \min\{\beta',\ \alpha,\ 0.3\rho\}$.

(ii) **Sub-case II** $\big(|X_{\mathbf{small}} \times Y_{\mathbf{small}}| \geq 2^{1.7\rho n}\big)$**:**
For a constant $\rho > 0$, we know there are positive constants $\beta'', \mu'' > 0$, from Lemma 17. Take $\mu_2 := \min\{\mu', \mu'', \nu\}$ and take $m := \mu_2 n$.

Then, with probability at least $1-2^{-4n^2}$ for uniformly random choice of p from $\mathcal{H}_{30n,10n}$, the following holds:

For all independent flat-ρn sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$, with support X and Y respectively, where $|X_{\text{small}}| \geq 2^{\rho' n}$ and $|X_{\text{small}} \times Y_{\text{small}}| \geq 2^{1.7\rho n}$, we have

$$\Delta\big(\phi_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{small}}\times Y_{\text{small}}}, \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{small}}\times Y_{\text{small}}}\big) \;\leq\; 2^{-\beta'' n}\,.$$

Hence, with probability at least $1 - 3\cdot 2^{-4n^2}$ we have,

$$\begin{aligned}
&\Delta\big(\phi_p^m(\mathbf{X},\mathbf{Y}), \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})\big)\\
&\leq\; 2^{-\beta' n}\cdot\frac{|X_{\text{large}}\times Y|}{|X\times Y|} \;+\; 2^{-\alpha n}\cdot\frac{|X_{\text{small}}\times Y_{\text{large}}|}{|X\times Y|} \;+\; 2^{-\beta'' n}\cdot\frac{|X_{\text{small}}\times Y_{\text{small}}|}{|X\times Y|}\\
&\leq\; 2^{-\hat{\gamma}_2 n}\,,
\end{aligned}$$

where $\hat{\gamma}_2 := \min\{\beta',\ \beta'',\ \alpha\}$.

2. **Case II** $\big(|X_{\mathbf{small}}| < 2^{\lambda n}$ **for all constant** $\lambda > 0\big)$**:**
Consider the partition

$$X\times Y \;=\; \big(X_{\text{large}}\ \times Y\big)\sqcup\big(X_{\text{small}}\times Y\big)\,.$$

For a constant $\rho > 0$, we know there are constants $\mu', \beta' > 0$, from Lemma 14. Therefore, with probability at least $1 - 3\cdot 2^{-4n^2}$, over uniformly random choice of p from $\mathcal{H}_{30n,10n}$, the following holds:

For independent flat-ρn sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$, with support X and Y respectively, where $|X_{\text{small}}| < 2^{\lambda n}$ for all constant $\lambda > 0$, we have

$$\begin{aligned}
&\Delta(\phi_p^m(\mathbf{X},\mathbf{Y}), \mathcal{D}_p^m(\mathbf{X},\mathbf{Y}))\\
&\leq\; 2^{-\beta' n}\cdot\frac{|X_{\text{large}}\times Y|}{|X\times Y|} \;+\; 2^{-0.5\rho n} \qquad \big(\text{since } |X_{\text{small}}| \leq 2^{0.5\rho n}\ \big)\\
&\leq 2^{-\beta' n} + 2^{-\rho n/2} \;\leq\; 2^{-\hat{\gamma}_3 n}\,,
\end{aligned}$$

where $\hat{\gamma}_3 := \frac{1}{2}\cdot\min\{\beta', \frac{\rho}{2}\}$.

This finishes the proof of case II.

Finishing off. From our analysis above, we know that for any constant $\rho > 0$, there exist positive constants $\gamma := \min\{\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3\}$, and $\mu := \min\{\mu_1, \mu_2, \mu'\}$, such that taking $m = \mu n$, with probability at least $1 - 2^{-3n^2}$ over the uniformly random choice of p from $\mathcal{H}_{30n,10n}$, the following holds:

For all independent flat-ρn sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$, we have

$$\begin{aligned}
&\Delta\Big((\text{nmExt}_p^m(\mathbf{X},\mathbf{Y}), \text{nmExt}_p^m(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y}); (\mathcal{U}_m, \text{nmExt}_p^m(f(\mathbf{X}), g(\mathbf{Y})), \mathbf{Y})\Big)\\
&\leq\; 2^{-\gamma n}\,.
\end{aligned}$$

The above statement combined with Lemma 1 yields the desired statement.

(ii) When $|X| = |Y| = 2^{\hat{C}n/\log n}$ for $\hat{C} \geq 30C$, we give the brief proof sketch as the idea is quite similar to the previous case. Let us again define:

$$\begin{aligned}
\hat{X}_{\text{large}} \;&=\; \big\{x \in X : |f^{-1}(f(x)) \cap X| \geq 2^{\hat{C}n/2\log n}\big\}\,,\\
\hat{X}_{\text{small}} \;&=\; \big\{x \in X : 0 < |f^{-1}(f(x)) \cap X| < 2^{\hat{C}n/2\log n}\big\}\,.
\end{aligned}$$

And similarly for Y,

$$\hat{Y}_{\text{large}} = \left\{y \in Y : |g^{-1}(g(y)) \cap Y| \geq 2^{\hat{C}n/2\log n}\right\},$$
$$\hat{Y}_{\text{small}} = \left\{y \in Y : 0 < |g^{-1}(g(y)) \cap Y| < 2^{\hat{C}n/2\log n}\right\}.$$

Here also, we break this into the following cases,

1. **When $|\hat{X}_{\text{small}}| \geq 2^{\hat{C}n/2\log n}$.** Under this case, there are two sub-cases as before:

   (i) $|\hat{X}_{\text{small}} \times \hat{Y}_{\text{small}}| \geq 2^{1.7\hat{C}n/\log n}$
   (ii) $|\hat{X}_{\text{small}} \times \hat{Y}_{\text{small}}| < 2^{1.7\hat{C}n/\log n}$.

   Consider the partition of $X \times Y$ as follows,

   $$X \times Y = \hat{X}_{\text{large}} \times Y \sqcup \hat{X}_{\text{small}} \times \hat{Y}_{\text{large}} \sqcup \hat{X}_{\text{small}} \times \hat{Y}_{\text{small}}.$$

   When sub-case (i) occurs, by Lemma 16, Lemma 23 and the partition lemma Lemma 13 we can conclude that

   $$\Delta\left(\phi_p^m(\mathbf{X},\mathbf{Y}); \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})\right) \leq 2^{-\sigma\sqrt{n}}$$

   where $m = \nu\sqrt{n}$ for some constants ν, σ.
   When we are in sub-case (ii), again by Lemma 16 we know,

   $$\phi_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{large}}\times Y} \approx_{2^{-\tilde{\sigma}\sqrt{n}}} \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{large}}\times Y}$$
   $$\phi_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{small}}\times Y_{\text{large}}} \approx_{2^{-\tilde{\sigma}\sqrt{n}}} \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})|_{X_{\text{small}}\times Y_{\text{large}}}$$

   where $m = \tilde{\nu}\sqrt{n}$ and $\tilde{\sigma}, \tilde{\nu}$ are fixed constants. As, $|\hat{X}_{\text{small}} \times \hat{Y}_{\text{small}}| \leq 2^{-0.3\hat{C}n/\log n}$ which is at most $2^{-\tilde{\sigma}\sqrt{n}}$, by Lemma 13 we have $\Delta\left(\phi_p^m(\mathbf{X},\mathbf{Y}); \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})\right) \leq 2^{-\sigma\sqrt{n}}$ where $m = \nu\sqrt{n}$ and ν, σ are constants.

2. **When $|\hat{X}_{\text{small}}| < 2^{\hat{C}n/2\log n}$,** we make the partition

   $$X \times Y = \hat{X}_{\text{large}} \times Y \sqcup \hat{X}_{\text{small}} \times Y.$$

   As, $\frac{|\hat{X}_{\text{small}} \times Y|}{|X \times Y|} \leq 2^{-0.5\hat{C}n/\log n}$, by item (i) of Lemma 16 and Lemma 13 we can again conclude that,

   $$\Delta\left(\phi_p^m(\mathbf{X},\mathbf{Y}); \mathcal{D}_p^m(\mathbf{X},\mathbf{Y})\right) \leq 2^{-\sigma\sqrt{n}}$$

   for $m = \nu\sqrt{n}$ and we achieve our desired result.

□

4.4 Exactly one of the two sources is tampered: Proof of Lemma 12

After finishing the proof of Lemma 11, it remains to prove that when *exactly* one of the two sources is tampered (i.e. second and third items of Definition 11), then also a *random* polynomial *works* with overwhelming probability, proving Lemma 12. For the sake of completeness, we restate the lemma below.

Lemma 24. *Let $f, g \in \mathcal{F}_n$. Then,*

(i) For any $\rho > 0$, there are constants $\tilde{\mu}, \tilde{\gamma} > 0$ such that with probability at least $1 - 2 \cdot 2^{-4n^2}$ over the uniformly random choice of $p \in \mathcal{H}_{30n,10n}$ the following holds: for all independent sources $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$, with $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq \rho n$, the following two properties hold.

   1. $\text{nmExt}_p^m(\mathbf{X},\mathbf{Y}), \text{nmExt}_p^m(f(\mathbf{X}),\mathbf{Y}), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \text{nmExt}_p^m(f(\mathbf{X}),\mathbf{Y}), \mathbf{Y}$

2. $\mathrm{nmExt}_p^m(\mathbf{X},\mathbf{Y}), \mathrm{nmExt}_p^m(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \mathrm{nmExt}_p^m(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y}$

where $m = \tilde{\mu} n$, *and* $\varepsilon = 2^{-\tilde{\gamma} n}$.

(ii) *There are constants* C_0 *and* $\tilde{\sigma}, \tilde{\nu} > 0$ *so that, with probability at least* $1 - 2 \cdot 2^{-4n^2}$ *over the uniformly random choice of* $p \in \mathcal{H}_{30n,10n}$ *the following holds: for all independent sources* $\mathbf{X}, \mathbf{Y} \sim \{0,1\}^n$, *with* $H_\infty(\mathbf{X}), H_\infty(\mathbf{Y}) \geq \frac{C_0 n}{\log n}$, *we have*

1. $\mathrm{nmExt}_p^m(\mathbf{X},\mathbf{Y}), \mathrm{nmExt}_p^m(f(\mathbf{X}),\mathbf{Y}), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \mathrm{nmExt}_p^m(f(\mathbf{X}),\mathbf{Y}), \mathbf{Y}$
2. $\mathrm{nmExt}_p^m(\mathbf{X},\mathbf{Y}), \mathrm{nmExt}_p^m(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \mathrm{nmExt}_p^m(\mathbf{X}, g(\mathbf{Y})), \mathbf{Y}$

where $\varepsilon = 2^{-\tilde{\sigma}\sqrt{n}}$ *and* $m = \tilde{\nu}\sqrt{n}$

Proof sketch. We will only sketch the proof of the first statement of item (i) since we have proved similar arguments before. The second statement can be similarly proved as well. And also by the similar argument using Lemma 8 and Theorem 8 item (ii), one can prove the case when input sources have min-entropy at least $C_0 n/\log n$ for some fixed constant C_0.

Without loss of generality, we assume $\mathbf{X}, \mathbf{Y}$ to be flat-ρn sources with supports X and Y respectively from Lemma 1. Observe that in this case $Y = Y_{\text{small}}$, with respect to the identity function. Consider the following partition,

$$X \times Y \;=\; (X_{\text{large}} \times Y) \;\sqcup\; (X_{\text{small}} \times Y) \;.$$

The subsets X_{large} and X_{small} are already defined before in Section 4.3. For now, let us restrict our focus on the case when $|X_{\text{small}}| \geq 2^{0.7\rho n}$.

Consider the matrix $\tilde{A} \in \mathbb{F}_2^{|Y| \times 20n}$, where each row is indexed by an element of Y, and for some $y \in Y$, the corresponding row is $p(0^{9n} \circ y) \circ p(0^{9n} \circ y)$. By the exact similar proof strategy of Theorem 10, we can prove that

$$\Pr_{p \leftarrow \mathcal{H}_{30n,10n}} [\dim_{\mathbb{F}_2}(\tilde{A}) \geq 12n] \;\geq\; 1 - 2^{-50n^2} \;.$$

Define $p(Y^{(10)} \circ Y^{(10)}) := \{p(0^{9n} \circ y) \circ p(0^{9n} \circ y) : y \in Y\}$. For any full-rank matrix $M \in \mathbb{F}_2^{20n \times 20n}$, taking a union bound over all such Y of size $2^{\rho n}$, we conclude that with probability at least $1 - 2^{-5n^2}$, over uniformly random choice over p from $\mathcal{H}_{30n,10n}$, the following holds: $\forall\, Y \subseteq \mathbb{F}_2^n$, of size $2^{\rho n}$, we have

$$\dim_{\mathbb{F}_2} \left(M \cdot p(Y^{(10)} \circ Y^{(10)}) \right) \;\geq\; 12n \;.$$

By Lemma 20, we get that with probability at least $1 - 2^{-5n^2}$, over uniformly random choice of p from $\mathcal{H}_{30n,10n}$, the following holds:

$\forall X \subseteq \mathbb{F}_2^n$, with $|X| = 2^{\rho n}$ and $|X_{\text{small}}| \geq 2^{0.7\rho n}$, we have

$$\dim_{\mathbb{F}_2} \left(p((X_{\text{small}})_f^{(10)} \right) \;\geq\; 12n \;.$$

Combining the above two equations, we have that with probability at least $1 - 2 \cdot 2^{-5n^2}$, for all X, Y of size $2^{\rho n}$ with $|X_{\text{small}}| \geq 2^{0.7\rho n}$,

$$\dim_{\mathbb{F}_2}(p(X_{\text{small}})_f^{(10)}) + \dim_{\mathbb{F}_2}(p(Y^{(10)} \circ Y^{(10)}) > 20n + 1$$

So, following the line of proof similar to Corollary 3 and Lemma 17, we can conclude that there are constants $\mu_0, \beta_0 > 0$, depending on $\rho > 0$, such that with probability at least $1 - 2^{-5n^2}$, over uniformly chosen p from $\mathcal{H}_{30n,10n}$, the following holds:

$$\forall X, Y \subseteq \mathbb{F}_2^n \text{ with } |X| = |Y| = 2^{\rho n} \text{ and } |X_{\text{small}}| \geq 2^{0.7\rho n}, \text{we have}$$

$$\mathrm{nmExt}_p^m(\mathbf{X}_{\mathrm{small}}, \mathbf{Y}), \mathrm{nmExt}_p^m(f(\mathbf{X}_{\mathrm{small}}), \mathbf{Y}), \mathbf{Y} \approx_\varepsilon \mathcal{U}_m, \mathrm{nmExt}_p^m(f(\mathbf{X}_{\mathrm{small}}), \mathbf{Y}), \mathbf{Y},$$

where $\varepsilon = 2^{-\beta_0 n}$ and $m = \mu_0 n$ and further, $\mathbf{X}_{\mathrm{small}}$ is the uniform distribution over X_{small} for all X.

Finally, we can conclude that for $\tilde{\mu} = \min\{\mu', \mu_0\}$ and $m = \tilde{\mu} n$, with probability at least $1 - 2^{-3n^2}$, over uniformly random choice of p from $\mathcal{H}_{30n,10n}$ the following holds:

For all independent flat-ρn sources $\mathbf{X}, \mathbf{Y} \sim \mathbb{F}_2^n$, with the supports X and Y respectively, when

1. **Case I.** ($|X_{\mathrm{small}}| \leq 2^{0.7\rho n}$):

$$\begin{aligned}
&\Delta(\mathrm{nmExt}_p^m(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^m(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y} \; ; \; \mathcal{U}_m, \mathrm{nmExt}_p^m(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y}) \\
&\leq \; 2^{-\beta' n} \frac{|X_{\mathrm{large}} \times Y|}{|X \times Y|} + \frac{|X_{\mathrm{small}} \times Y|}{|X \times Y|} \\
&\leq \; 2^{-\beta' n} + 2^{-0.3\rho n} \; \leq \; 2^{-\tilde{\gamma}_1 n} \; ,
\end{aligned}$$

   where $\tilde{\gamma}_1 := \frac{1}{2} \cdot \min\{\beta', 0.3\rho\}$,
2. **Case II.** ($|X_{\mathrm{small}}| \geq 2^{0.7\rho n}$):

$$\begin{aligned}
&\Delta(\mathrm{nmExt}_p^m(\mathbf{X}, \mathbf{Y}), \mathrm{nmExt}_p^m(f(\mathbf{X}, \mathbf{Y}), \mathbf{Y}; \mathcal{U}_m, \mathrm{nmExt}_p^m(f(\mathbf{X}), \mathbf{Y}), \mathbf{Y}) \\
&\leq \; 2^{-\beta' n} \cdot \frac{|X_{\mathrm{large}} \times Y|}{|X \times Y|} + 2^{-\beta_0 n} \cdot \frac{|X_{\mathrm{small}} \times Y|}{|X \times Y|} \\
&\leq \; 2^{-\beta' n} + 2^{-\beta_0 n} \; \leq \; 2^{-\tilde{\gamma}_2 n} \; ,
\end{aligned}$$

   where $\tilde{\gamma}_2 := \frac{1}{2} \cdot \min\{\beta', \beta_0\}$.

This finishes the proof of the first statement as desired. □

5 Selecting polynomial via two party computation

We have seen that if a uniformly sample a polynomial of degree $O(n)$ and coefficients from a *small* extension of $\mathbb{F}_2$ (degree of extension $O(n)$), with high probability we have: For any constant $\rho > 0$ there are constants $\mu, \gamma > 0$ so that,

$$\mathrm{nmExt}_p^m \text{ is a } (n, m, k, \varepsilon) \text{ strong 2 source non-malleable extractor .}$$

where when $k = \rho n$ for some constant $\rho > 0$, we have $\varepsilon \leq 2^{-\gamma n}$ and $m = \mu n$ for some appropriate constants $\mu, \gamma > 0$. And when $k = C_0 n / \log n$ for some fixed constant C_0, we have $\varepsilon \leq 2^{-\sigma n}$ and $m = \nu\sqrt{n}$, for constants $\nu, \sigma > 0$.

In this section we will show that even without access to uniformly random coin tosses, one can come up with a *good* polynomial via two party computation. We will define an explicit two party protocol, that outputs a low degree polynomial p so that even when one of the parties are corrupted, nmExt_p retains its properties. Recall the definition of $\mathcal{H}_{t,cn}$

$$\mathcal{H}_{t,cn} := \{p \in \mathbb{F}_{2^{cn}}[Z] : \deg(p) \leq t\} \; .$$

Let us start by proving a technical claim that tells that instead of choosing the polynomial uniformly if one samples it according to a distribution of min-entropy rate at least 0.9, with high probability it will be a *good* polynomial.

Lemma 25. *Let* $\mathbf{X}$ *be any distribution on* $\mathcal{H}_{300n,10n}$ *so that* $\frac{H_\infty(\mathbf{X})}{\log |\mathcal{H}_{300n,10n}|} \geq 0.9$ *(i.e. min entropy rate of* $\mathbf{X}$ *is at least* 0.9*). Then:*

1. *For all constant $\rho > 0$, there exist constants $\mu, \gamma > 0$ so that the following holds.*

$$\Pr_{p \sim \mathbf{X}} \left[\text{nmExt}_p^{\mu n} \text{ is a } (n, \mu n, \rho n, 2^{-\gamma n}) \text{ strong 2-source non-malleable extractor } \right] \geq 1 - 2^{-48n^2}$$

2. *There are constants $C_0 > 0$ and $\nu, \sigma > 0$ so that,*

$$\Pr_{p \sim \mathbf{X}} \left[\text{nmExt}_p^{\mu n} \text{ is a } \left(n, \nu\sqrt{n}, \frac{C_0 n}{\log n}, 2^{-\sigma\sqrt{n}} \right) \text{ strong 2-source non-malleable extractor } \right] \geq 1 - 2^{-48n^2}$$

Proof. We will only prove (1) as by Theorem 9, the second item immediately follows. Fix $\rho > 0$ and $\mu, \gamma > 0$ are the constants that we get from Theorem 9, item (i). Let us define the following set

$$\text{Bad} := \left\{ p \in \mathcal{H}_{300,10n} : \text{nmExt}_p^{\mu n} \text{ is not a } (n, \mu n, \rho n, 2^{-\gamma n}) \text{ strong 2-source non-malleable extractor} \right\}.$$

From Lemma 19 and Corollary 2 and following the line of proof same as Theorem 9 we can say,

$$\Pr_{p \leftarrow \mathcal{H}_{300n,10n}} [p \in \text{Bad}] \leq 2^{-348n^2}$$

That implies $|\text{Bad}| \leq |\mathcal{H}_{300n,10n}| \times 2^{-348n^2} \leq 2^{2652n^2}$. Now consider any distribution $\mathbf{X}$ over $\mathcal{H}_{300n,10n}$ with min-entropy rate at least 0.9. By definition, for any $p \in \mathcal{H}_{300n,10n}$, we have $\Pr[\mathbf{X} = p] \leq 2^{-2700n^2}$. Therefore finally we have,

$$\Pr_{p \sim \mathbf{X}} [p \in \text{Bad}] \leq 2^{-48n^2}.$$

This completes our proof. □

Now we are ready to describe the two party protocol.

Definition 13 (Two party protocol for selecting the polynomial). *The protocol is as follows:*

(i) ***Alice:*** *Uniformly samples a polynomial q from $\mathcal{H}_{300n,10n}$ and broadcasts.*
(ii) ***Bob:*** *Uniformly samples a polynomial r from $\mathcal{H}_{30n,10n}$ and broadcasts.*
(iii) Both of them agree on the final polynomial $p(Z) := q(Z) + r(Z)$ and output.

Lemma 26. *Say $\mathbf{Y}$ is the output distribution of the protocol 13. Even if either Alice or Bob is corrupted: If for $\rho > 0$ we have $\mu, \gamma > 0$ are the constants from item (i) and $C_0, \sigma, \nu > 0$ are the constants from item (ii) of Theorem 9*

1.
$$\Pr_{p \sim \mathbf{Y}} \left[\text{nmExt}_p^{\mu n} \text{ is a } (n, \mu n, \rho n, 2^{-\gamma n}) \text{ strong 2-source non-malleable extractor } \right] \geq 1 - 2^{-2n^2}$$

2.
$$\Pr_{p \sim \mathbf{Y}} \left[\text{nmExt}_p^{\nu\sqrt{n}} \text{ is a } \left(n, \nu\sqrt{n}, \frac{C_0 n}{\log n}, 2^{-\sigma\sqrt{n}} \right) \text{ strong 2-source non-malleable extractor } \right] \geq 1 - 2^{-2n^2}$$

Proof. Note that from Theorem 9 and Lemma 25 it is enough to prove the item 1.
When both of Alice and Bob are honest, $\mathbf{Y}$ is uniform distribution over $\mathcal{H}_{300n,10n}$. Hence, the lemma follows immediately from Theorem 9.

When Alice is honest and Bob is corrupted, Alice picks polynomial from uniform distribution over $\mathcal{H}_{300n,10n}$. We denote it as $\mathcal{U}'$. Bob picks a polynomial from some arbitrary distribution $\mathbf{X}'$ over $\mathcal{H}_{30n,10n}$. Note that $\mathcal{U}'$ can be written as $(\mathcal{U}^{(1)}, \mathcal{U}^{(2)})$ where,

- $\mathcal{U}^{(1)}$ denotes the uniform distribution from which coefficients of monomials of degree $\leq 30n + 1$ is picked
- $\mathcal{U}^{(2)}$ denotes the uniform distribution from which coefficients of remaining monomials are picked.

Therefore, $\mathbf{Y} = \mathcal{U}' + \mathbf{X}' = (\mathcal{U}^{(1)}, \mathcal{U}^{(2)} + \mathbf{X}')$. As, $H_\infty(\mathcal{U}^{(1)}) \geq 270n^2$, i.e. min-entropy rate of $\mathcal{U}^{(1)}$ is at least 0.9, together Theorem 2 and Lemma 25 imply the lemma.

When Bob is honest and Alice is corrupted. As Alice has to start the communication, trivially the final polynomial p will be $30n$-wise independent. From Remark 3 we know that if we pick p uniformly from $30n$-wise independent hash family, $\mathrm{nmExt}_p^{\mu n}$ is a $(n, \mu n, \rho n, 2^{-\gamma n})$ strong 2-source non-malleable extractor with probability at least $1 - 2^{-3n^2}$. Hence, this concludes our proof. □

6 Acknowledgements

This work was supported by the NRF investigatorship grant (NRF-NRFI09-0005), the NSF CAREER award (grant CCF-2338730) and the MOE Tier 2 grant titled "Breaking the Box-On Security of Cryptographic Devices." Additionally, this work was carried out while Pranjal Dutta was a Research Fellow at NUS and Satyajeet Nagargoje was hosted by Divesh Aggarwal at the Centre for Quantum Technologies, National University of Singapore. Satyajeet Nagargoje would also like to thank Zeyong Li for the insightful discussions during the visit.

References

ACLV19. Divesh Aggarwal, Kai-Min Chung, Han-Hsuan Lin, and Thomas Vidick. A quantum-proof non-malleable extractor: With application to privacy amplification against active quantum adversaries. In *Advances in Cryptology,EUROCRYPT*, 2019.

ACO23. Divesh Aggarwal, Eldon Chung, and Maciej Obremski. Extractors: Low entropy requirements colliding with non-malleability. In *Annual International Cryptology Conference*, pages 580–610. Springer, 2023.

ADKO15. Divesh Aggarwal, Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Leakage-resilient non-malleable codes. In *Theory of Cryptography Conference, TCC*, 2015.

ADL14. Divesh Aggarwal, Yevgeniy Dodis, and Shachar Lovett. Non-malleable codes from additive combinatorics. In *Proceedings of the 46th annual ACM Symposium on Theory of Computing, STOC*, 2014.

ADN+19. Divesh Aggarwal, Ivan Damgård, Jesper Buus Nielsen, Maciej Obremski, Erick Purwanto, Joao Ribeiro, and Mark Simkin. Stronger leakage-resilient and non-malleable secret sharing schemes for general access structures. In *Advances in Cryptology, CRYPTO*. Springer, 2019.

AGMR24. Omar Alrabiah, Jesse Goodman, Jonathan Mosheiff, and João Ribeiro. Low-degree polynomials are good extractors. *arXiv preprint arXiv:2405.10297*, 2024.

AGV09. Adi Akavia, Shafi Goldwasser, and Vinod Vaikuntanathan. Simultaneous hardcore bits and cryptography against memory attacks. In *Theory of Cryptography Conference, TCC*, 2009.

AHL16. Divesh Aggarwal, Kaave Hosseini, and Shachar Lovett. Affine-malleable extractors, spectrum doubling, and application to privacy amplification. In *International Symposium on Information Theory, ISIT*, 2016.

AKO+22. Divesh Aggarwal, Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, Maciej Obremski, and Sruthi Sekar. Rate one-third non-malleable codes. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1364–1377, 2022.

AO20. Divesh Aggarwal and Maciej Obremski. A constant rate non-malleable code in the split-state model. In *FOCS*, 2020.

AOR+22. Divesh Aggarwal, Maciej Obremski, João Ribeiro, Mark Simkin, and Luisa Siniscalchi. Privacy amplification with tamperable memory via non-malleable two-source extractors. *IEEE Transactions on Information Theory*, 68(8):5475–5495, 2022.

BADTS17. Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. An efficient reduction from two-source to non-malleable extractors: achieving near-logarithmic min-entropy. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, 2017.

BCC+13. Daniel J Bernstein, Yun-An Chang, Chen-Mou Cheng, Li-Ping Chou, Nadia Heninger, Tanja Lange, and Nicko Van Someren. Factoring rsa keys from certified smart cards: Coppersmith in the wild. In *Advances in Cryptology-ASIACRYPT 2013*, 2013.

BDK08. Eli Biham, Orr Dunkelman, and Nathan Keller. A unified approach to related-key attacks. In *Fast Software Encryption: FSE 2008*, 2008.

BGLZ15. Abhishek Bhowmick, Ariel Gabizon, Thái Hoang Lê, and David Zuckerman. Deterministic extractors for additive sources. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS*, 2015.

BH19. Joachim Breitner and Nadia Heninger. Biased nonce sense: Lattice attacks against weak ecdsa signatures in cryptocurrencies. In *Financial Cryptography and Data Security: 23rd International Conference, FC*, 2019.

Blu86. Manuel Blum. Independent unbiased coin flips from a correlated biased source—a finite state markov chain. *Combinatorica*, 1986.

Bou05. Jean Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 2005.

Bou07. Jean Bourgain. On the construction of affine extractors. *GAFA Geometric And Functional Analysis*, 17(1):33–57, 2007.

BS19. Saikrishna Badrinarayanan and Akshayaram Srinivasan. Revisiting non-malleable secret sharing. In *Advances in Cryptology – EUROCRYPT*, 2019.

BSLRZ14. Eli Ben-Sasson, Shachar Lovett, and Noga Ron-Zewi. An additive combinatorics approach relating rank to communication complexity. *J. ACM*, 61(4), July 2014.

CDH+00. Ran Canetti, Yevgeniy Dodis, Shai Halevi, Eyal Kushilevitz, and Amit Sahai. Exposure-resilient functions and all-or-nothing transforms. In *Advances in Cryptology – EUROCRYPT*, 2000.

CG88. Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 1988.

CG14. Mahdi Cheraghchi and Venkatesan Guruswami. Non-malleable coding against bit-wise and split-state tampering. In Yehuda Lindell, editor, *Theory of Cryptography*, pages 440–464, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

CG17. Mahdi Cheraghchi and Venkatesan Guruswami. Non-malleable coding against bit-wise and split-state tampering. *J. Cryptology*, 30(1):191–241, January 2017.

CGGL19. Eshan Chattopadhyay, Jesse Goodman, Vipul Goyal, and Xin Li. Extractors for adversarial sources via extremal hypergraphs. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC*, 2019.

CGL16. Eshan Chattopadhyay, Vipul Goyal, and Xin Li. Non-malleable extractors and codes, with their many tampered extensions. In *Proceedings of the 34th annual ACM Symposium on Theory of Computing, STOC*, 2016.

CKOS19. Eshan Chattopadhyay, Bhavana Kanukurthi, Sai Lakshmi Bhavana Obbattu, and Sruthi Sekar. Privacy amplification from non-malleable codes. In *International Conference on Cryptology in India*, pages 318–337. Springer, 2019.

CL16. Eshan Chattopadhyay and Xin Li. Explicit non-malleable extractors, multi-source extractors, and almost optimal privacy amplification protocols. In *FOCS*, 2016.

Cle86. R Cleve. Limits on the security of coin flips when half the processors are faulty. In *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing*, page 364–369, New York, NY, USA, 1986. Association for Computing Machinery.

Coh17. Gil Cohen. Towards optimal two-source extractors and ramsey graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, 2017.

CZ19. Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the 48th annual ACM Symposium on Theory of Computing, STOC*, 2019.

DCLW06. Giovanni Di Crescenzo, Richard Lipton, and Shabsi Walfish. Perfectly secure password protocols in the bounded retrieval model. In *Theory of Cryptography Conference*, pages 225–244. Springer, 2006.

DDV10. Francesco Davì, Stefan Dziembowski, and Daniele Venturi. Leakage-resilient storage. In *Security and Cryptography for Networks: 7th International Conference, SCN 2010, Amalfi, Italy, September 13-15, 2010. Proceedings 7*, pages 121–137. Springer, 2010.

DGW09. Zeev Dvir, Ariel Gabizon, and Avi Wigderson. Extractors and rank extractors for polynomial sources. *Computational Complexity*, 18:1–58, 2009.

DKO13. Stefan Dziembowski, Tomasz Kazana, and Maciej Obremski. Non-malleable codes from two-source extractors. In *Advances in Cryptology, CRYPTO*, 2013.

DOPS04. Yevgeniy Dodis, Shien Jin Ong, Manoj Prabhakaran, and Amit Sahai. On the (im) possibility of cryptography with imperfect randomness. In *FOCS*, 2004.

DPW18. Stefan Dziembowski, Krzysztof Pietrzak, and Daniel Wichs. Non-malleable codes. *Journal of the ACM (JACM)*, 65(4):1–32, 2018.

DSS01. Yevgeniy Dodis, Amit Sahai, and Adam Smith. On perfect and adaptive security in exposure-resilient cryptography. In *Advances in Cryptology, EUROCRYPT*, 2001.

Dvi12. Zeev Dvir. Extractors for varieties. *Computational complexity*, 21:515–572, 2012.

DW09. Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the 31st annual ACM Symposium on Theory of Computing, STOC*, 2009.

Dzi06. Stefan Dziembowski. Intrusion-resilience via the bounded-storage model. In *Theory of Cryptography Conference, TCC*, 2006.

FKOS22. Sebastian Faust, Juliane Krämer, Maximilian Orlt, and Patrick Struck. On the related-key attack security of authenticated encryption schemes. In *International Conference on Security and Cryptography for Networks*, 2022.

GGMT25. W. T. Gowers, Ben Green, Freddie Manners, and Terence Tao. On a conjecture of marton. *Annals of Mathematics*, 2025. To appear.

GK18a. Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, 2018.

GK18b. Vipul Goyal and Ashutosh Kumar. Non-malleable secret sharing for general access structures. In *Advances in Cryptology,CRYPTO*, 2018.

GKK20. Ankit Garg, Yael Tauman Kalai, and Dakshita Khurana. Low error efficient computational extractors in the crs model. In *Advances in Cryptology–EUROCRYPT 2020: 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10–14, 2020, Proceedings, Part I 39*, pages 373–402. Springer, 2020.

GLM+04. Rosario Gennaro, Anna Lysyanskaya, Tal Malkin, Silvio Micali, and Tal Rabin. Algorithmic tamper-proof (atp) security: Theoretical foundations for security against hardware tampering. In *TCC*, 2004.

GPR16. Vipul Goyal, Omkant Pandey, and Silas Richelson. Textbook non-malleable commitments. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 1128–1141, 2016.

GSZ21. Vipul Goyal, Akshayaram Srinivasan, and Chenzhi Zhu. Multi-source non-malleable extractors and applications. In *Advances in Cryptology, EUROCRYPT*, 2021.

HDWH12. Nadia Heninger, Zakir Durumeric, Eric Wustrow, and J Alex Halderman. Mining your ps and qs: Detection of widespread weak keys in network devices. In *21st USENIX Security Symposium (USENIX Security 12)*, pages 205–220, 2012.

KRVZ06. Jesse Kamp, Anup Rao, Salil Vadhan, and David Zuckerman. Deterministic extractors for small-space sources. In *Proceedings of the 38th annual ACM Symposium on Theory of Computing, STOC*, 2006.

Lew19. Mark Lewko. An explicit two-source extractor with min-entropy rate near. *Mathematika*, 65(4):950–957, 2019.

Li16. Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *FOCS*, 2016.

Li17. Xin Li. Improved non-malleable extractors, non-malleable codes and independent source extractors. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, 2017.

Li19. Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. In *CCC*, 2019.

Li23. Xin Li. Two source extractors for asymptotically optimal entropy, and (many) more. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1271–1281. IEEE, 2023.

MW97. Ueli Maurer and Stefan Wolf. Privacy amplification secure against active adversaries. In *Advances in Cryptology, CRYPTO*, 1997.

Rao07. Anup Rao. An exposition of bourgain's 2-source extractor. *Electronic Colloquium on Computational Complexity*, 2007.

Raz05. Ran Raz. Extractors with weak frandom seeds. In *Proceedings of the 37th annual ACM Symposium on Theory of Computing, STOC*, 2005.

SV86. Miklos Santha and Umesh V Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 1986.

TV00. Luca Trevisan and Salil Vadhan. Extracting randomness from samplable distributions. In *FOCS*, 2000.

Vad12. Salil P. Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 2012.

Vaz86. Umesh Virkumar Vazirani. *Randomness, adversaries and computation (random polynomial time)*. PhD thesis, 1986. AAI8718194.

VN51. John Von Neumann. Various techniques used in connection with random digits. *Applied Math Ser*, 12(36-38), 1951.

ZBS11. Noga Zewi and Eli Ben-Sasson. From affine to two-source extractors via approximate duality. In *Proceedings of the 43rd annual ACM Symposium on Theory of Computing, STOC*, 2011.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il