Simulating Independence: New Constructions of Condensers, Ramsey Graphs, Dispersers, and Extractors

Preliminary Full Version

Boaz Barak[†] Guy Kindler[†] Ronen Shaltiel[‡] Benny Sudakov[§] Avi Wigderson[†]

November 4, 2004

[†]Institute for Advanced Study, Princeton, NJ. Emails: {boaz,gkindler,avi}@ias.edu
[‡]Department of Computer Science, Haifa University, Israel. Email: ronen@cs.haifa.ac.il
[§]Department of Mathematics, Princeton University, Princeton, NJ. Email: bsudakov@math.princeton.edu

Abstract

A distribution X over binary strings of length n has min-entropy k if every string has probability at most 2^{-k} in X.[1] X is called a δ-source if its rate k/n is at least δ.

We give the following new explicit constructions (namely, poly(n)-time computable functions) of *deterministic* extractors, dispersers and related objects. All work for any fixed rate $\delta > 0$. No previous explicit construction was known for either of these, for any $\delta < 1/2$. The first two constitute major progress to very long-standing open problems.

1. **Bipartite Ramsey** $f_1 : (\{0,1\}^n)^2 \to \{0,1\}$, such that for any two independent δ-sources X_1, X_2 we have $f_1(X_1, X_2) = \{0,1\}$. This implies a new explicit construction of $2N$-vertex bipartite graphs where no induced N^δ by N^δ subgraph is complete or empty.
2. **Multiple source extraction** $f_2 : (\{0,1\}^n)^3 \to \{0,1\}$, such that for any three independent δ-sources X_1, X_2, X_3 we have that $f_2(X_1, X_2, X_3)$ is ($o(1)$-close to being) an unbiased random bit.
3. **Constant seed condenser**[2] $f_3 : \{0,1\}^n \to (\{0,1\}^m)^c$, such that for any δ-source X, one of the c output distributions $f_3(X)_i$, is a 0.9-source over $\{0,1\}^m$. Here c is a constant depending only on δ.
4. **Subspace Ramsey** $f_4 : \{0,1\}^n \to \{0,1\}$, such that for any *affine*-δ-source[3] X we have $f_4(X) = \{0,1\}$.

The constructions are quite involved and use as building blocks other new and known gadgets. But we can point out two important themes which recur in these constructions. One is that gadgets which were designed to work with independent inputs, sometimes perform well enough with correlated, high entropy inputs. The second is using the input to (introspectively) find high entropy regions within itself.

Keywords: Ramsey Graphs, Explicit Constructions, Extractors, Dispersers, Condensers

[1]It is no real loss of generality, and very convenient, to think of X as a uniform distribution on some set of 2^k elements (in which case min-entropy is the same as Shannon entropy).

[2]This result was also independently obtained by Ran Raz.

[3]This is a uniform distribution over an affine subspace of dimension at least δn

Contents

1 Introduction

Randomness extraction is the problem of distilling the entropy present in "weak random sources" into a useful, (nearly) uniform distribution. Its importance and wide applicability to diverse theoretical and practical areas of computer science has motivated a large body of research over the last 20 years.

Much of this research assumes only that the given "weak source" has sufficient (min)-entropy, and that extractors can use an extra, short random "seed" to aid in distilling the randomness. Such a seed (of logarithmic length) is easily seen to be necessary. This research focused on explicitly constructing extractors of small seed (and long output). The survey [SHA] explains most of the recent developments, and the paper [LRVW] contains the state-of-art construction in terms of the ratio between seed length and output length.

However, certain applications (especially in cryptography) cannot "afford" the use of an extra random seed (see e.g. [MP, DS, BST]). To do without it, one must impose extra conditions (beyond entropy content) on the class of sources from which one extracts. This body of work, which includes [vN, BLU, SV, CG, BOL, VAZ, CW, CGH+, TV, MU, KZ, GRS] considers a variety of restrictions, mostly structural, on the given sources, and shows how to extract from them deterministically (without seed).

This paper provides several new explicit constructions of seedless extractors, dispersers and related objects (all to be defined), greatly improving on previous results. We also give several weaker constructions, which are not quite seedless, but only use seeds of *constant size*. These are important as building blocks of the seedless ones, and some are interesting in their own right.

We now turn to describe some of the new constructions, and for each discuss history and related work. For the sake of brevity and clarity, we would skip some of our results, and state others in less than full generality and precision.

1.1 Multiple independent sources

Background. Perhaps the most natural condition allowing seedless extraction is that instead of *one* source with high entropy, we have several independent ones. This model was suggested by Santha and Vazirani [SV], and further studied by Chor and Goldreich [CG] (who introduced the now standard notion of min-entropy).

A function $f : (\{0,1\}^n)^\ell \to \{0,1\}^m$ is called an ℓ-source extractor with entropy requirement k if for every ℓ sources $X_1, \cdots, X_\ell$, each with min-entropy k, the distribution $f(X_1, \cdots, X_\ell)$ (obtained by applying f on an independent sample from each source) is close to uniform on $\{0,1\}^m$.

For this model, the probabilistic method easily gives the best that can be hoped for. Two sources and $\log n$ entropy suffice (and are necessary) for extraction. Moreover, such a function can be computed in time $2^{O(n^2)}$ (while this is a far cry from the explicitness we want (poly(n)-time), it will come as a building block in our constructions). We call such a function `opt` (for Optimal Extractor).

For two sources, the best explicit construction requires in contrast min-entropy $> n/2$. The Hadamard function $\mathtt{Had} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined by $\mathtt{Had}(x, y) = \langle x, y\rangle \pmod 2$ was shown by [CG] to be such an extractor. Moreover, a natural variant $\mathtt{Had}'$ which outputs $m = \Omega(n)$ was shown to be an extractor in [VAZ] under the same conditions (see a simplified and improved analysis in [DEOR]). To date, no one has found an explicit 2-source extractor with sources of rate 1/2 or less[4]. Indeed, until last year no extractor breaking the 1/2 rate barrier was known using

[4]We note that using Weil's estimates the Paley matrix (i.e., for prime p the function $P : \mathbb{F}_p^2 \to \{0,1\}$ where $P(x, y) = 1$ iff $x - y \pmod p$ is a quadratic residue) can be proven to be an extractor even if one source has only

less than $\ell = \mathrm{poly}(n)$ number of sources. Still, the Hadamard function and its extension to long outputs $\mathtt{Had}'$ will be an important building block in our construction.

Last year [BIW] gave a extractor that, for any $\delta > 0$ uses only a constant $\ell = \mathrm{poly}(1/\delta)$ δ-sources.[5] Moreover, on n-bit sources, their extractor outputs n bits and is exponentially close to uniform. The analysis seems to require that the total entropy in all sources is at least the length of one source (and hence that $\ell > 1/\delta$). Still, this too will be a building block in our new constructions in which the number of sources is a constant independent of δ.

New Results. We construct a 3-source extractor which outputs a (nearly) unbiased bit (or any constant number of bits) for every entropy rate $\delta > 0$. We can also increase the output length m to be a constant fraction of the input entropy using 7 independent sources. In both constructions the error (=distance of the output to the uniform distribution) is only sub-constant (about $1/\log\log n$), and as yet we have no idea how to make it exponentially, or even polynomially small. Subsequently to us, Raz [RAZ] used our construction (as well as additional ideas) to obtain a 3-source extractor which outputs $\Omega(n)$ number of bits, and works with one δ-source and two sources of logarithmic entropy.

1.2 Bipartite Ramsey graphs and 2-source dispersers

Ramsey Graphs. The probabilistic method was first used by Erdos to show the existence of *Ramsey graphs*: That is, a 2-coloring of the edges of the complete graph on N vertices such that no induced subgraph of size $K = (2 + o(1))\log N$ is monochromatic. The best known explicit construction of such a coloring by [FW] achieves a much larger value: $K = 2^{\Theta(\sqrt{\log N \log\log N})}$.

Bipartite Ramsey graphs. An even harder variant of this problem is the *bipartite Ramsey problem:* Construct a 2-coloring of the edges of the complete N by N bipartite graph such that no induced K by K subgraph is monochromatic. Setting $N = 2^n$, a coloring is a function $f : (\{0,1\}^n)^2 \to \{0,1\}$. It immediately follows that every 2-source extractor f with entropy-rate δ is a coloring for the bipartite Ramsey problem with $K = N^\delta$ (which is in turn a coloring for the non-bipartite version). Until recently, the best known explicit construction of bipartite Ramsey graphs was that implied by the aforementioned Hadamard 2-source extractor achieving $K > N^{1/2}$. Recently, a slight improvement to $K = N^{1/2}/2^{\sqrt{\log N}}$ (which in our terminology translates to $\delta = 1/2 - 1/\sqrt{n}$) was given by Pudlák and Rödl [PR].[6]

2-source dispersers. An equivalent formulation of this problem is constructing a 1-bit output 2-source *disperser* which is a well-known relaxation of an *extractor*. A 2-source (zero-error) *disperser* of entropy rate δ is a function $\mathtt{disp} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ such that for every 2 independent δ-sources X_1, X_2 we have $\mathtt{disp}(X_1, X_2) = \{0,1\}^m$. In words, every possible output occurs when the inputs range over all possible values in X_1, X_2 (and so only the support matters, not the individual probabilities in the input sources)[7]. Note that when $m = 1$, a disperser is equivalent to a bipartite

about $\log n$ entropy. However the other source still has to have entropy $> n/2$

[5] This extractor was proposed already in [ZUC], but the analysis there relies on an unproven number theoretic assumption.

[6] The construction in that paper is only "weakly explicit" in the sense that the 2-coloring can be found in time polynomial in N. The Hadamard 2-source extractor (as well as all the constructions in this paper) are "strongly explicit" meaning that f is computable in time polynomial in n.

[7] In the extractor literature dispersers usually come with an error parameter ϵ, and then the requirement is that the output of f contains at least $(1-\epsilon)$-fraction of all elements in $\{0,1\}^m$. We use the 0-error definition, which is more relevant to our setting.

Ramsey graph with $K = N^{\delta}$.

New Results. We give an explicit construction of a 2-source disperser with any constant (and even slightly sub-constant) rate $\delta > 0$ (so any $K > N^{\delta}$ in the bipartite Ramsey problem) and any *constant* output length m. Our disperser is strong in the sense that it obtains every output with at least a constant probability, which depends only on δ and on the number of output bits. This construction can therefore be seen as being in between a disperser and an extractor.

1.3 Constant seed (1-source) condensers

Intuitively, a *condenser* is a function whose output distribution is "denser" (has higher entropy rate) than its input distribution. Condensing can be viewed as a weaker form of extraction. Both in the Mathematical sense that an extractor is an ultimate condenser (its output has maximum possible rate), as well as in the practical sense – some constructions of extractors proceed by iterated condensing. Various condensers appear in [RSW, TSUZ, LRVW, CRVW] and other works, mainly as building blocks to constructing extractors and expanders.[8]

It is not hard to see that, like extractors, there are no deterministic condensers. However, unlike extractors, which require logarithmic seed, condensing is possible (for interesting parameters) with only constant length seed. As usual, this was shown via the probabilistic method, and no explicit construction was known. All constructions in the papers above either use a super-constant seed, or use a constant seed without guaranteeing the condensing property.[9]

New Results. We give the first explicit constant seed condenser for linear entropy. More precisely, for every $\delta > 0$ there are integers c, d and a poly(n)-time computable function $\texttt{con} : \{0,1\}^n \rightarrow (\{0,1\}^{n/c})^d$ (i.e. $\texttt{con}$ maps n bit strings into d blocks of length n/c), such that for every δ-source X there is at least one output block $\texttt{con}(X)_i$ (exponentially close to) having entropy rate ≥ 0.9 (here the 0.9 is an arbitrary constant - we can get as close as we want to 1, but all we need is any constant $> 1/2$). The constant seed may be viewed as selecting the output block at random. In the paper we prefer the view of outputting several blocks, and call this construction a *somewhere condenser* (since the concatenation of blocks is condensed somewhere). As we shall see, this condenser is not only interesting in its own right, but it also serves as a basic block in our new constructions. Roughly speaking, it gives us the means to break the 1/2 rate barrier, as it converts an input source of rate *below* that barrier into one (of a few output blocks - something to be dealt with) whose rate is *above* that barrier.

The condenser above is obtained by iterating a constant number of times a *basic* condenser $\texttt{bcon}$ (described in the Section 2.1), which uses only a 2-bit seed (namely has 4 output blocks) and increases the entropy rate by a constant amount.

We note that, independently from us, Ran Raz [RAZ] has obtained a different constant seed condenser construction, which also use the [BIW] paper, but in another way. (In fact, using a new variant of the [LRVW] merger, Raz constructs a condenser with the advantage that *most* of the output blocks are condensed.)

[8] We note that in some of these works the definition is relaxed to allow situations when the output rate is *smaller* than the input rate, as long the the output length is shorter - these gadgets turn out to be useful as well.

[9] The latter are the so called "win-win" condensers whose analysis shows that when they fail to condense, some other good thing must happen.

1.4 A disperser for affine subspaces

Background. There are some other natural restrictions on sources (beyond allowing a few independent ones) that admit deterministic extraction. Two such models are bit-fixing sources [CGH$^+$, BOL, CW, MU, KZ, GRS] and efficiently samplable distributions [TV]. In the bit-fixing model, an (unknown) subset of δn bits (out of the n input bits) is uniformly distributed, and the rest of the bits are fixed to some unknown value.[10] In the "efficiently samplable distributions" model the source is obtained by applying an *efficiently computable* "sampling" function $g : \{0,1\}^{\delta n} \to \{0,1\}^n$ on a uniformly distributed input. Trevisan and Vadhan [TV] give a construction (based on unproven complexity theoretic assumptions) that works for sources with entropy rate $\delta > 1/2$ that are sampled by polynomial sized circuits g. (In fact, in that construction $\delta \geq (1-\gamma)$ for some small constant γ).

In this paper we consider sources which are uniformly distributed over an affine subspace of $\mathrm{GF}(2)^n$. Such sources can be seen as a natural generalization of bit-fixing sources (since any bit-fixing source is in particular a distribution over an affine subspace), or alternatively, as a restriction of the model of [TV] in which the sampling function g is affine (note that every such function can be computed in size n^2).

Strangely, what was known about this model is very similar quantitatively to the two independent source model in the sense one one hand there is a (nonconstructive) "optimal extractor" for entropy $\Theta(\log n)$ but on the other hand an explicit extractor was known only for entropy rate $\delta > 1/2$. (In fact, this explicit extractor is no other than the Hadamard function `Had` mentioned above, applied to two halfs of the sample [BSHRV].)

New Results. We give an explicit zero-error disperser for affine sources of entropy rate δ, that outputs constantly many bits. As in the case of the two-source disperser, the affine-source disperser is strong, obtaining each output string with at least a constant probability. For 1-bit output a Ramsey theoretic interpretation of the result suggests itself – we give a 2-coloring of the Boolean cube F_2^n, in which no affine subspace of linear dimension is monochromatic.

Note that the new results here are quantitatively the same as our 2-source results. The techniques are related as well, but at this point this fact may be surprising - there seem to be little resemblance between the models, and indeed there seem to be no reductions between them in either direction. The similarity in techniques may simply be a byproduct of the fact that we were working on them in parallel, and progress on one suggested related progress on the other (note that at some point we do use the two-source Ramsey construction as a black box in the affine-source disperser, but we could have just as well used a direct construction there). This simply manifests the generality of our techniques, however it would be interesting to find any tighter connections between the two models.

2 Techniques and overview of the main constructions

In the following subsections we describe in some detail some of the extra gadgets we develop, and how to compose them with each other and with known gadgets to get the constructions above. By far the most difficult is the 2-source disperser, and the ideas leading to it take several subsections. Still, reading these will simplify understanding the formal proofs.

[10] In the terminology of the above papers this is the so-called *oblivious* bit-fixing model.

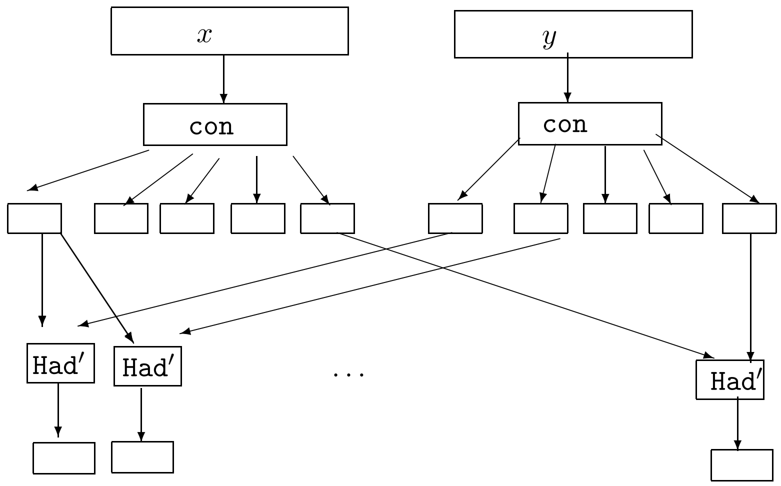

Figure 1: A 2-source somewhere extractor `s_ext`.

2.1 A 2-bit seed condenser

Here we describe our basic condenser `bcon`, which uses only a 2-bit seed to increase the entropy rate of any source by a constant amount. Iterating it a constant number of times on a δ-source allows us to increase to rate (of some output block) above 0.9 (say), thus giving the condenser `con` described in Section 1.3.

Our basic condenser `bcon` will take strings of length n with $n = 3p$ for some prime p.[11] For every $x \in \{0,1\}^n$ let $x = x_1x_2x_3$ its natural partition to three length p blocks. Define $\mathtt{bcon} : \{0,1\}^{3p} \to (\{0,1\}^p)^4$ by $\mathtt{bcon}(x) = x_1, x_2, x_3, x_1 \cdot x_2 + x_3$ (with arithmetic in $GF(2^p)$).

We prove that in X is a δ-source with $\delta < 0.9$, then at least one of the output blocks is a $(\delta + \Omega(\delta^2))$-source.

The proof heavily relies on the main lemma of [BIW], who proved $x_1 \cdot x_2 + x_3$ is condensed *assuming* that the x_i's are *independent*. We certainly *cannot* assume that in our case, as X is a general source. Still, we use that lemma to show that if none of these first 3 blocks is more condensed than the input source, then they are "independent enough" for using that main lemma.

2.2 A 2-source constant-seed/"somewhere" extractor

Our two main deterministic constructions in this paper are a 3-source extractor and a 2-source disperser. For both, an essential building block, is a constant seed 2-source extractor `s_ext` (short for "somewhere extractor") for linear entropy, which we describe next.

What we prove is that for every $\delta > 0$ there are integers c, d and a poly(n)-time computable function $\mathtt{s\_ext} : (\{0,1\}^n)^2 \to (\{0,1\}^{n/c})^d$, such that for every two δ-sources X_1, X_2 there is at least one output block $\mathtt{s\_ext}(X_1, X_2)_i$ which is (exponentially close to) uniform.

Constructing the somewhere extractor `s_ext` is simple, given the condenser `con` of the previous subsection. To compute $\mathtt{s\_ext}(X_1, X_2)$, compute the output blocks of $\mathtt{con}(X_1)$ and $\mathtt{con}(X_2)$. By definition, some output block of each has rate $> .9$. We don't know which, but we can try all pairs!

[11] Note that if n is not of this form than it can be converted to this form by padding the input with $o(n)$ additional bits using standard number theoretic bounds on the density of primes.

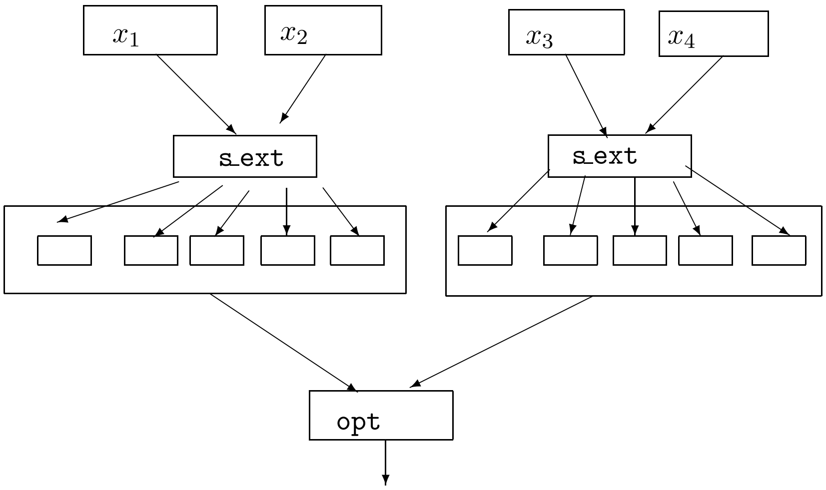

Figure 2: A 4-source extractor `4ext`.

For each pair we compute the Vazirani variant $\texttt{Had}'$ of the Hadamard 2-source extractor for rate $> 1/2$ [Vaz] to obtain a constant number of linear length blocks, one of which is exponentially close to uniform. Formally, if d is the number of output block of `con`, then `s_ext` will produce d^2 blocks, with $\texttt{s\_ext}(X_1, X_2)_{i,j} = \texttt{Had}'(\texttt{con}(X_1)_i, \texttt{con}(X_2)_j)$. This construction is depicted in Figure 1.

To see the power of this gadget, let us first see (intuitively) how to get from it a *deterministic* 4-source extractor for linear entropy. Later we will employ it in several ways to get our 2-source disperser.

2.3 A 4-source extractor (and a 3-source one)

In this subsection we explain how to construct a 4-source extractor `4ext`, and then how to modify it to the promised 3-source extractor `3ext`. These will combine the 2-source somewhere extractor `s_ext` with the nonuniform optimal 2-source extractor `opt`.

Recall that our 2-source *somewhere* extractor `s_ext` produces a constant number (say) d of linear length output blocks, one of which is random. First we note that producing shorter output blocks maintains this property[12].

Let us indeed output only a constant b bits in every block (satisfying $b \geq \log(db)$). Concatenating all output blocks of this $\texttt{s\_ext}(X_1, X_2)$ gives us a distribution (say Z_1) on db bits with min-entropy $\geq b$. If we have 4 sources, we can get another independent such distribution Z_2 from $\texttt{s\_ext}(X_3, X_4)$. But note that these are two independent distributions on a constant number of bits with sufficient min-entropy for (existential) 2-source extraction. Now apply an optimal (non-uniform) 2-source extractors on Z_1, Z_2 to get a uniform bit; as db is only a constant, such an extractor can be found in constant time by brute-force search! To sum up, our 4-source extractor is $\texttt{4ext}((X_1, X_2); (X_3, X_4)) = \texttt{opt}(\texttt{s\_ext}(X_1, X_2), \texttt{s\_ext}(X_3, X_4))$. See Figure 2 for a schematic description of this construction.

Reducing the number of sources to 3 illustrates a simple idea we'll need later. We note that essentially all 2-source constructions mentioned above are "strong". This term, borrowed from the extractor literature, means that the output property is guaranteed for almost every way of

[12] A prefix of a random string is random. This flexibility will prove useful in more ways than one.

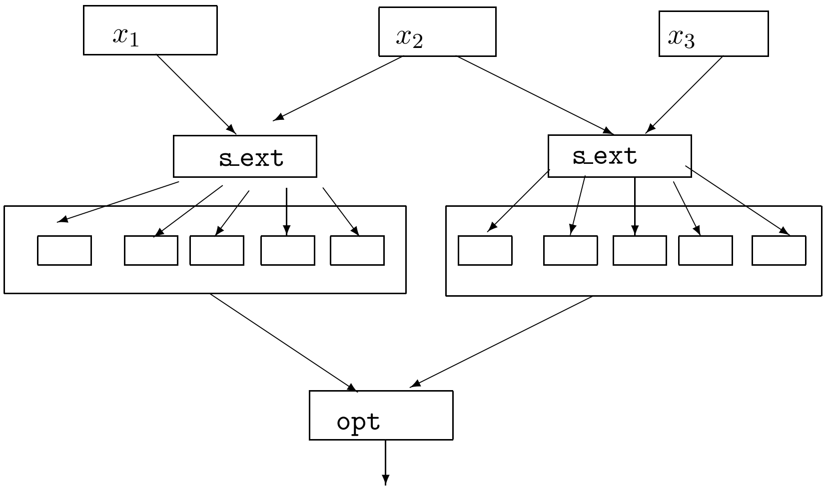

Figure 3: A 3-source extractor 3ext.

fixing the value of *one* of the two input sources. With this in mind, we can *reuse* (say) X_2 in the second somewhere extractor s_ext instead of X_4 to yield a 3-source extractor $\texttt{3ext}((X_1, X_2, X_3) = \texttt{opt}(\texttt{s\_ext}(X_1, X_2), \texttt{s\_ext}(X_3, X_2))$. This construction is depicted in Figure 3. Fixing a random sample x_2 of X_2 will guarantee the output properties of both s_ext's with high probability *and* keep the two inputs to opt independent (since once x_2 is fixed the distributions $\texttt{s\_ext}(X_1, x_2)$ and $\texttt{s\_ext}(X_3, x_2)$ are independent).

2.4 A better 2-source somewhere extractor (with witness)

Our somewhere extractor s_ext produced a constant number of output blocks, one of which is random. It gave us a 4-source (and even a 3-source) extractor, and the intuitive reason for needing the extra independence was that we could not tell which of the output blocks of s_ext was the random one.

To get down to a 2-source (disperser), our general strategy will be to try and "figure out" which of the blocks is random. This requires a "testable" condition on the input, which will point to the right output block. We cannot do that directly with s_ext, so we first develop (using s_ext) a new somewhere extractor $\texttt{s\_ext}'$, for which the random output block is determined by simple entropy conditions on the input sources. The very subtle task of testing these (given only the input sample) is delayed to the next subsection - we now describe $\texttt{s\_ext}'$.

It will be convenient to call now the two independent input δ-sources X and Y. In a nutshell, $\texttt{s\_ext}'$ will consider different partitions of X into $X'X''$ and Y to $Y'Y''$, and for each such partition (X', X'', Y', Y'') apply the 4-source extractor 4ext of the previous subsection to obtain $\texttt{4ext}(X', Y'', Y', X'')$ (see Figure 4).

While these 4 sources are not independent, we can use again the same idea used in decreasing 4 sources into 3 of the previous section to try to analyze this construction. As each of the s_ext components in 4ext are *strong*, we can show that fixing X' and Y' to random samples x' and y' (resp.) will suffice to make the output $\texttt{4ext}((x', Y''), (y', X''))$ close to random, as long as we have the following entropy bounds: the four sources $X', Y', X''|X' = x', Y''|Y' = y'$ all have at least ϵn bits of entropy for some constant $\epsilon > 0$.

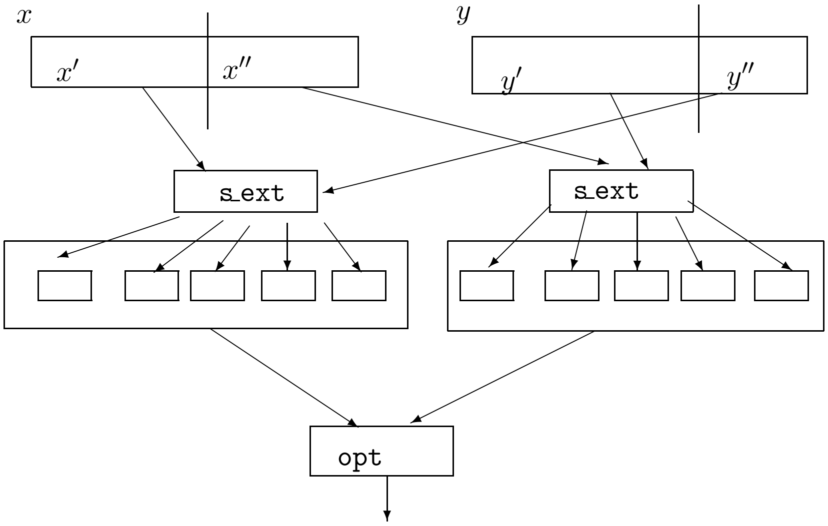

Figure 4: Applying the 4-source extractor to two sources.

Another observation is that a one of a constant number of partitions of X and Y *will* satisfy these entropy conditions. This can be shown using the following reasoning. Specifically, let $t = 4/\delta$ and for $i \in [t]$, denote by X_i the i^{th} block of size n/t in X. That is, $X = X_1, \ldots, X_t$. We let $X_{\leq i}$ denote the concatenation of the first i blocks of X (i.e., $X_{\leq i} = X_1, \ldots, X_i$) and let $X_{>i}$ denote the rest of X (i.e., $X_{>i} = X_{i+1}, \ldots, X_n$). We say that i is a "good" index if $X_{\leq i}$ has min-entropy at least $\delta n/4$ and define i_0 be the *smallest* good i. Since the entropy of $X_{\leq i}$ is at most the min-entropy of $X_{\leq i-1}$ plus the length of X_i we get that it is bounded from above by $2 \cdot (\delta n/4) = \delta n/2$. Combining this with the fact that X has min-entropy at least δn, we get that the min-entropy of $X_{>i_0}$ conditioned on fixing $X_{\leq i_0}$ is at least $\delta n/2$. Therefore if we define $X' = X_{\leq i_0}$ and $X'' = X_{>i_0}$ we get the desired partition. In the same way one can partition Y into Y' and Y'' where $Y' = Y_{\leq j_0}$ and $Y'' = Y_{>j_0}$ for some $j_0 \in [t]$. We caution the reader that the effects of conditioning on min-entropy are actually more subtle than what is reflected by this proof sketch. The actual analysis, which is in Section 7 is thus much more involved than the description here.

Our new somewhere extractor $\mathtt{s\_ext}'$ will apply $\mathtt{4ext}$ to all t^2 possible ways of defining these four sources from the two given ones X and Y. Specifically, we define $\mathtt{s\_ext}'(x, y)_{i,j} = \mathtt{4ext}((x_{\leq i}, y_{>j}), (x_{\leq j}, y_{>i}))$. We know that the output block $\mathtt{s\_ext}'(x, y)_{i_0,j_0}$ corresponding to i_0 for X and j_0 for Y as defined above *will* be uniformly distributed! This "goodness" will be our "testable" entropy condition.

2.5 Finding the entropy: the "Challenge-Response" mechanism

Ideally, we'd like a test that will find the good partitions and point to the correct output block. Namely, we'd like a function $\mathtt{test} : (\{0,1\}^n)^2 \to [t]^2$ such that the output block $\mathtt{s\_ext}'(X, Y)_{\mathtt{test}(X,Y)}$ is uniform. Note that if we could do that, we'd have a 2-source extractor - more than we promised. Our test will only succeed to point out to the random output block with positive probability (thus yielding our strong notion of a disperser).

Before describing the test, we note the obvious subtlety: $\mathtt{test}$ is applied to the *same* sample (x, y) from $X \times Y$ that $\mathtt{s\_ext}'$ is applied to! This single sample is supposed to detect the presence of entropy in some parts of the input sources it is coming from, by looking in the mirror. The fact that this idea can make sense is an important contribution of our paper.

To explain the nature of the test `test`, we oversimplify our problem, and assume we have the following extra gadget to help us.

Temporary unrealistic assumption: We'll assume a deterministic 1-source extractor (called $\mathtt{ext}^*$) for any linear entropy sources. It of course seems very strange to assume an object much stronger than what we're trying to construct (indeed so strong that it doesn't exist). Nonetheless we believe it will be a useful didactic tool. Later, we will examine the properties of $\mathtt{ext}^*$ that we actually needed in order to obtain the function `test` and how these properties can be obtained by more realistic gadgets.

We will test each of the prefixes of input parts (potential choices of the first part of the partition X'), namely $X_{\leq 1}, X_{\leq 2}, \cdots, X_{\leq t}$ of X, for containing entropy. To test in the input part $X_{\leq k}$ has entropy, we apply the (hypothetical) extractor $\mathtt{ext}^*$ to compute a *challenge* $C_k = \mathtt{ext}^*(X_{\leq k})$ of (sufficiently large) constant length c. We will also apply $\mathtt{ext}^*$ on the second source Y to generate a string $R = R_1, R_2, \cdots R_t$, where the length of every *response* R_k is c as well. We note that as Y is a δ source, R is essentially a random string of length tc.

We say that the challenge C_k is met by the *response* R_k simply if $R_k = C_k$. We now define the candidate value of the index i' (this index determines the partition of X to $X'X''$ selected by the test `test`) to be $k+1$ for the largest k for which the challenge was met. We similarly determine the partition of Y by picking j' similarly, reversing the roles of X and Y above. This will give us the pair (i', j') which is the output of the test `test`.

2.6 Analyzing the challenge-response mechanism

Now why does this work, and in what sense? Let us focus on the choice of i' (the argument for j' is analogous). We want to claim that `test` above produces i' which is the smallest good index for X. While this may fail, we can claim that it holds in a constant probability event in $X \times Y$. Essentially, this event will be defined by conditioning parts of X and Y to fix the outcome of some challenge and response strings, resulting in the "correct" choices made by the test `test`, *and* all independence and entropy requirements holding after these fixings. A simple observation, which will be crucial in the analysis below, is that if X is a random variable on $\{0,1\}^n$, and $f : \{0,1\}^n \to \{0,1\}^c$ is a function with a constant output length c, then if we let y be a "typical" output of $f(X)$ and let X' be X conditioned on $f(X) = y$, then then entropy of X' is only a constant smaller than the entropy of X.

Now, let i_0 be the actual smallest good index for X. For all $k = 1, 2, \cdots, i_0 - 1$ (in that order) we condition X_k in a way that fixes the responses $C_k, k < i_0$ to their most popular values (by the observation above, this reduces the entropy of X by a constant only). Now we also fix the values of the responses $R_k, k < i_0$ so that they meet the respective challenges above. Then we fix set R_k for $k \geq i_0$ to an arbitrary value. Again, fixing all the responses reduces the entropy of Y by only a constant.

We now repeat the same process with j_0, the smallest good index for Y. Call the resulting subsources thus generated $\tilde{X}$ and $\tilde{Y}$. Now in the space $\tilde{X} \times \tilde{Y}$ everything we want actually happens.

- $\tilde{X}$ and $\tilde{Y}$ are independent.
- $\tilde{X}$ has constant probability in X (and same for Y), and hence the entropies of $\tilde{X}$ and $\tilde{Y}$ are at least $\delta n - O(1)$.

- This means the the entropy of any subblock of X can also change by at most a constant amount in $\tilde{X}$ (with the same holding for Y). Therefore, $\tilde{X}_{\leq i_0}, \tilde{X}_{>i_0}$ is a "good" partition of $\tilde{X}$, with $\tilde{X}_{\leq i_0}, \tilde{X}_{>i_0}|\tilde{X}_{\leq i_0}$ both having entropy at least $\delta n/4 - O(1)$. Again, the same holds for $\tilde{Y}$.

- By design, in $\tilde{X}$ the test `test` selects $i' = i_0$ with probability very close to one. This happens since we forced all the first $i_0 - 1$ challenges to be met, and all remaining ones will be missed with very high probability $(1 - t/2^c)$, and so by the definition of `test`, it will select $i' = i_0$. Same happens with $j' = j_0$.

To summarize, with constant probability the sample (x, y) from $X \times Y$ actually lands in the event $\tilde{X} \times \tilde{Y}$, in which case we produce a close to uniform output. Otherwise, we have no guarantee. This is precisely the strong disperser promised.

2.7 Removing the unrealistic assumption

We will be very brief here. The main thing to fix of course, is our assumption that we have available (the impossible) one source extractor `ext`*. It will be replaced by our favorite 2-source somewhere extractor `s_ext`. This raises several extra issues to deal with, and we touch on the solutions, ignoring many important details.

- When using `s_ext` instead of `ext`*, what do we use for the second independent input source? The answer is simple - when we used `ext`* on part of X, we use (all of) Y as a second source, and vice versa. This raises as variety of issues regarding independence, and regarding preservation of entropy, which did not arise above. Indeed, when we use `ext`* to compute a *response* (as opposed to a *challenge*) we will use as input only "tiny" (around $\delta^3 n$ size) parts of X and Y so that these parts can be fixed in the analysis without a significant loss in entropy.

- The output of `s_ext` is not uniform, but rather a constant number of blocks one of which is uniform - how does `test` change? The answer again is expected - the challenges and responses comprise the whole output of `s_ext` (of appropriate lengths). However now a response meets a challenge if *one* of the output blocks of the response equals the (the whole) challenge string. Hence, when computing a response we will output a much larger (although still constant size) string than when computing a challenge. This raises several issues regarding the probabilities with which challenges are met, with bearing on entropies lost, which did not arise above.

This is of course by no means a complete description of our 2-source disperser and its analysis. The complete description and analysis can be found in Section 7.

3 Preliminaries and Notations

3.1 Probability Notations.

We will use the following notations for random variables.

Random variables, sources, min-entropy and entropy rate. We will usually deal with random variables which take values over $\{0,1\}^n$. We call such a random variable an *n-bit source*. The *min-entropy* of a random variable X, denoted by $H^\infty(X)$, is defined to equal $\min_x\{-\log_2(\Pr[X = x])\}$, or equivalently $\log_2(1/\max_x\{\Pr[X = x]\})$. We shall usually identify a random variable X

with the distribution it induces (and so one may treat an n-bit source as a vector over $\mathbb{R}^{2^n}$ with positive coordinates that sum up to 1). The *entropy rate* of an n-bit source X, denoted by $\mathsf{r}(X)$, is defined to be $H^\infty(X)/n$.

Flat sources, convex combinations. The support of a random variable X, denoted by $\mathsf{Supp}(X)$, is the set of elements x for which $\Pr[X = x] > 0$. If $\Pr[X = x] = \Pr[X = x']$ for every $x, x' \in \mathsf{Supp}(X)$ we say that X is a flat source. Note that X is a flat source of min-entropy k if and only if X is the uniform distribution over some set A of size 2^k. Let $X^{(1)}, \dots, X^{(k)}$ be random variables, and let $\alpha_1, \dots, \alpha_k$ be non-negative numbers such that $\sum \alpha_i = 1$. We say that an n-bit source Y is the convex combination of $\{X^{(i)}\}_{i\in[k]}$ with coefficients $\{\alpha_i\}_{i\in[k]}$, if Y satisfies $\Pr[Y = y] = \sum_i \alpha_i \Pr[X^{(i)} = y]$ for all y (i.e., in vector notation $Y = \sum_{i=1}^k \alpha_i X^{(i)}$). We sometimes refer to $X^{(i)}$ as a *component* of Y. It is known that every random variable with min entropy $\geq k$ is a convex combination of flat sources each with min-entropy at least k.

Statistical Distance. Let X and Y be random variables taking values in a set Λ. The *statistical distance between X and Y*, denoted by $\mathsf{dist}(X, Y)$ is defined to be $\frac{1}{2}\sum_{x\Lambda} |\Pr[X = x] - \Pr[Y = x]|$. We'll use the following easy to verify fact regarding statistical distance:

Fact 3.1. *Let X and Y be discrete random variables, and let $\epsilon \geq 0$. Then $\mathsf{dist}(X, Y) \leq \epsilon$ if and only if there exist random variables X' and Y' with the same distribution of X and Y respectively, both defined over* the same *probability space Ω, such that $\Pr_{\omega\in_R\Omega}[X'(\omega) \neq Y'(\omega)] \leq \epsilon$.*

Blocks. Let X be an n-bit source, and let us write it as a vector of one-bit variables $(X_1, \dots, X_n)$. For a set $S \subseteq [n]$ of coordinates we define the S-block of X, denoted by X_S, to be $|S|$-bit source $(X_i)_{i\in S}$, obtained from X by only taking its S coordinates. X_S is said to be a block in X, or a sub-block of X. We note that for every n-bit source X and $S \subseteq [n]$ it holds that $H^\infty(X_S) \geq H^\infty(X) - (n - |S|)$. Also, if $\mathsf{dist}(X, Y) < \epsilon$ then $\mathsf{dist}(X_S, Y_S) < \epsilon$ for every $S \subseteq [n]$.

Sub-sources. We will also sometimes use the less-standard notion of *sub-sources*, defined as follows:

Definition 3.2 (Sub-source). Let X and X' be n-bit sources. For a non-negative number $\alpha \leq 1$, we say that X' is a sub-source of X of measure α if $X = \alpha X' + (1 - \alpha)Z$ for some n-bit source Z. We use the notation $X' \subseteq X$ to say that X' is a sub-source of X.

Properties of random variables. We'll also discuss sometimes families or *properties* of random variables. Let $\mathcal{P}$ be a property of sources, we say that a random variable X is within ϵ distance from $\mathcal{P}$, or that X is ϵ-close to having the property $\mathcal{P}$, if there exists a random variable $Y \in \mathcal{P}$ that is within statistical distance at most ϵ from X. We denote by $\mathcal{P}_\epsilon$ the set of random variables that are ϵ-close to $\mathcal{P}$. For example, we say that X is ϵ-close to having min entropy $\geq k$ if there exists a random variable Y with $H^\infty(Y) \geq k$ such that $\mathsf{dist}(X, Y) \leq \epsilon$. In this case we write $H^\infty_\epsilon(X) \geq k$ (and hence $H^\infty_\epsilon(X)$ is defined to be the maximum k such that there exists Y with $H^\infty(Y) = k$ and $\mathsf{dist}(X, Y) \leq \epsilon$).

Blocks and Somewhere-$\mathcal{P}$ sources. We now define the notion of *somewhere-$\mathcal{P}$ sources*, which intuitively means that a source X has the property $\mathcal{P}$ in one of its blocks.

Definition 3.3 (Somewhere$-\mathcal{P}$ sources). Let $\mathcal{P}$ be a property of sources. Let X be an $(n \times \ell)$-bit source, and let us regard it as a concatenation of ℓ consecutive blocks, $X = (X_1, \ldots, X_\ell)$, where each X_i is of length n. A selector I for X is a random variable taking values in $[\ell]$.

If there exists a selector I for X, such that the n-bit source X_I has the property $\mathcal{P}$, we say that X is a somewhere $\mathcal{P}$ source. We use this notion only when the partition of X into sub-blocks is clear from the context.

Proposition 3.4. *Let $\mathcal{P}$ be a property of sources, and let X be an $(n \times \ell)$-bit source. Let us regard it as a concatenation of ℓ consecutive blocks. Then if X is somewhere $\mathcal{P}_\epsilon$, then it is ϵ close to being somewhere $\mathcal{P}$.*

Proof: Write X as a concatenation of blocks $X = (X_1, \ldots, X_\ell)$. Let I be a selector for X such that X_I is ϵ-close to some n-bit source $Y \in \mathcal{P}$. We would like to couple X_I with Y: using Fact 3.1, one may assume without loss of generality that X and Y are defined over the same probability space, and that $\Pr[X_I \neq Y] \leq \epsilon$.

Let us define X' to be the $(n \times \ell)$-bit source

$$X' \stackrel{def}{=} (X'_1, \ldots X'_\ell) \ ,$$

where we let $X'_i = Y$ if $I = i$, and $X'_i = X_i$ otherwise. It is obvious that $\Pr[X \neq X'] \leq \epsilon$, and that $X'_I = Y \in \mathcal{P}$. Therefore X' is a somewhere $\mathcal{P}$ source, and by Fact 3.1, $\mathsf{dist}(X, X') \leq \epsilon$. We thus have that X' is ϵ close to being somewhere $\mathcal{P}$. □

3.2 Extractors and Dispersers

In this section we define some of the objects we will later construct, namely multiple-sample extractors and dispersers. These definitions are essentially taken from [BIW].

Definition 3.5 (Multiple-source extractor). A function $\mathtt{ext} : \{0,1\}^{n \times \ell} \to \{0,1\}^m$ is called an *ℓ-source extractor with k min-entropy requirement, n-bit input, m-bit output and ϵ-statistical distance* if for every independent n-bit sources $X^{(1)}, \ldots, X^{(\ell)}$ satisfying $H^\infty(X^{(i)}) \geq k$ for $i = 1, \ldots, \ell$ it holds that

$$\mathsf{dist}\Big(\mathtt{ext}(X^{(1)}, \ldots, X^{(\ell)}), U_m\Big) \leq \epsilon$$

Definition 3.6 (Multiple-source disperser). A function $\mathtt{disp} : \{0,1\}^{n \times \ell} \to \{0,1\}^m$ is called an *ℓ-source disperser with k min-entropy requirement, n-bit input, m-bit output and error parameter* ϵ if for every independent n-bit sources $X^{(1)}, \ldots, X^{(\ell)}$ satisfying $H^\infty(X^{(i)}) \geq k$ for $i = 1, \ldots, \ell$ it holds that

$$|\mathsf{Supp}(\mathtt{disp}(X^{(1)}, \ldots, X^{(\ell)}))| \geq (1 - \epsilon)2^m$$

The disperser is *errorless* if $\epsilon = 0$ (i.e., if $|\mathsf{Supp}(\mathtt{disp}(X^{(1)}, \ldots, X^{(\ell)}))| = 2^m$).

Note that a multiple-source extractor is always a multiple-source disperser with the same parameters.[13] Generally speaking, when constructing extractors our goal is on one hand to use the weakest assumptions on the input and hence we want to minimize the number of samples ℓ and the min-entropy requirement k. On the other hand we want to obtain the strongest guarantees on the output, and thus we want to maximize the output length m and minimize the error ϵ.

We note that one can also define *asymmetric* variants of extractors and dispersers, in which the min-entropy requirements and the input length differ for each of the samples. However, we do not define or use such variants in this paper.

[13] In fact, sometimes it is even a disperser with better parameters. For example, a multiple source extractor with error smaller than 2^{-m} always implies an errorless disperser with all other parameters being equal.

3.3 The BIW Theorem and the BIW Lemma.

In Section 8 we need the following theorem from [BIW].

Theorem 3.7 (The [BIW] extractor). *There exists constants $c_1, c_2 \geq 1$ such that for every constant $\delta > 0$, and every fixed parameter $k \geq K(\delta) \stackrel{def}{=} c_1 \cdot (1/\delta)^{c_2}$, there exists a polynomial-time computable function $\mathtt{ext} : \{0,1\}^{n \cdot k} \to \{0,1\}^n$ with the following property: for every independent random variables $X_1, \ldots, X_k$ over $\{0,1\}^n$, each with min-entropy at least δn, it holds that*

$$\mathsf{dist}(\mathtt{ext}(X_1, \ldots, X_k), U_n) < 2^{-\Omega(n)}$$

We also use the following lemma from [BIW]:[14]

Lemma 3.8 (BIW Lemma). *There exists some constant $\alpha_0 > 0$ such that for every independent random variables A, B, C taking values in the field $\mathrm{GF}(2^p)$ for a prime p, each with entropy rate at least δ, the distribution $A \cdot B + C$ is $2^{-\alpha_0 \delta^2 p}$-close to having entropy rate at least $\min\{\delta + \alpha_0 \delta^2, 0.99\}$.*

When trying to use Lemma 3.8, we will be frequently in the situation when we're trying to map the set $\{0,1\}^n$ in a one-to-one way to the field $\mathrm{GF}(2^p)$ where $p = n(1 + o(1))$. (Note that in this case if X is a δ-source over $\{0,1\}^n$ then its image is a $\delta - o(1)$-source over $\{0,1\}^p$.) To do so, the following two facts will be useful to us:

Fact 3.9. *The exists a large-enough integer n_0 such that for every $n \geq n_0$, there exists a prime p in the segment $[n, n + n^{3/4}]$.*

Fact 3.10. *There exists an algorithm which given a prime p, finds an irreducible p-degree polynomial over $\mathbb{Z}_2$, in time polynomial in p.*

Note that once we have an irreducible polynomial P of degree p over $\mathbb{Z}_2$, then we can do computations over $\mathrm{GF}(2^p)$ in time polynomial in p.

3.4 Affine Sources

Let us begin with the definition of an affine source.

Definition 3.11 (Affine sources). An n-bit source X is said to be affine if it is a flat source, and moreover, $\mathsf{Supp}(X) \subseteq \{0,1\}^n$ is an affine subspace over $\mathbb{Z}_2$.

Note that for an affine source X, the min-entropy of X is equal to the dimension of its support, namely $H^\infty(X) = \dim(\mathsf{Supp}(X))$. The following lemma is helpful in analysis min-entropies of affine sources.

Lemma 3.12. *Let W be a linear space over $\mathbb{Z}_2$, and let $V \subset W$ be a subspace. Suppose that linear maps P_1 and P_2 are a partition of the identity on W, namely $P_1 + P_2 = id$, and $P_1 P_2 = 0$. Then*

$$\dim(P_1(V) \cap V) = \dim(V) - \dim(P_2(V))\,.$$

[14]This lemma is actually more general than stated here. We will use it for fields of the form $\mathrm{GF}(2^p)$, but it holds for any field without large subfields. For *prime* fields [BIW] prove a stronger result: namely, for such fields the factor $\alpha_0\delta^2$ can be replaced with $\alpha_0\delta$.

Proof. Consider a vector $y \in V$, and suppose that $P_2(y) = 0$. Since $P_1 + P_2 = id$, it follows that $P_1(y) = y$, and therefore that $y \in V \cap P_1(V)$. Also if $y \in V \cap P_1(V)$ then $P_2(y) = 0$. Letting $P : V \to W$ denote the restriction of P_2 to V, we therefore have that

$$Ker(P) = V \cap P_1(V)\,. \tag{1}$$

Now, by a well-known identity we have $\dim(Ker(P)) + \dim(\mathsf{Supp}(P)) = \dim(V)$, which by (1) and the definition of P gives $\dim(P_1(V) \cap V) + \dim(P_2(V)) = \dim(V)$, completing the proof. □

The following is a useful corollary of Lemma 3.12.

Corollary 3.13 (Sub-blocks of affine sources). *If X is an affine source, then every block in X is an affine source as well. Moreover, if Y is a block in X and $y \in \mathsf{Supp}(Y)$ then the restriction of X to the event $[Y = y]$ is an affine source with min-entropy $H^\infty(X) - H^\infty(Y)$, and thus $H^\infty(X|Y = y) = H^\infty(X) - H^\infty(Y)$.*

Before proving the corollary, note that it means that if Y is a sub-block of X then for every $y_1, y_2 \in \mathsf{Supp}(Y)$, $H^\infty(X|Y = y_1) = H^\infty(X|Y = y_2)$. We can therefore define the min-entropy of X conditioned on Y by

$$H^\infty(X|Y) \stackrel{def}{=} H^\infty(X|Y = y)\ , \tag{2}$$

where y is any arbitrary element of $\mathsf{Supp}(Y)$. Using Corollary 3.13 we can write

$$H^\infty(X|Y) = H^\infty(X) - H^\infty(Y)\ . \tag{3}$$

Let us now prove the corollary.

Proof. Suppose X is an affine n-bit source, and let $S \subseteq [n]$ be the set such that $Y = X_S$. Let $z \in \mathsf{Supp}(X)$ be any vector which satisfies $x_S = y$, and let $X' \stackrel{def}{=} X + z$. So now $\mathsf{Supp}(X')$ is a linear subspace (not affine) of $\{0,1\}^n$, but since adding a constant vector merely flips some of the coordinates, there is no change in entropy, and therefore it is enough to prove that

$$H^\infty(X'|X'_S = 0) = H^\infty(X') - H^\infty(X'_S)\ .$$

Let $P_1 : \{0,1\}^n \to \{0,1\}^n$ be the linear projection which zeros, for any given x, the S-coordinates on x. Let $P_2 \stackrel{def}{=} id - P_1$ be the projection which zeros the other coordinates. Then by Lemma 3.12,

$$\begin{aligned} H^\infty(X'|X'_S = 0) &= \dim\big(\mathsf{Supp}(X') \cap P_1(\mathsf{Supp}(X'))\big) = \dim\big(\mathsf{Supp}(X')\big) - \dim\big(P_2(\mathsf{Supp}(X'))\big) \\ &= \dim(\mathsf{Supp}(X')) - \dim(\mathsf{Supp}(X'_S)) = H^\infty(X') - H^\infty(X'_S)\ , \end{aligned}$$

and we have the corollary. □

There exists a strong connection between independence of affine sources (viewed as random variables), and independence of linear spaces.

Definition 3.14 (Independent subspaces). It is said that a set $V_1, \ldots, V_k$ of subspaces of some linear space W is *independent*, if

$$\forall\, i\,, \qquad \dim\Big(V_i \cap \big(\textstyle\sum_{j \neq i} V_j\big)\Big) = 0.$$

The following claim is immediate from the definition.

Claim 3.15. *Let $U_1, \ldots, U_k$ and $V_1, \ldots, V_k$ be subspaces of some linear space W, and suppose that $V_1, \ldots, V_k$ are independent and that $U_i \subseteq V_i$ for every i. Then the subspaces $U_1, \ldots, U_k$ are also independent.*

3.4.1 Useful properties of affine sources

Following are some useful properties of affine sources, that we make use of later.

Lemma 3.16. *Let $V_1, \ldots, V_k$ be independent linear subspaces of some linear space W over $\mathbb{Z}_2$, of dimension at most $d_1 + d_2$ each. Also, let V be a subspace of $V_1 + \ldots + V_k$, of dimension at least kd_1. Then there exist subspaces $U_1, \ldots, U_k$ of W, each of dimension at least $d_1 - kd_2$, such that $U_i \subseteq V_i$ for every i, and that $U_1 + \ldots + U_k \subseteq V$.*

Proof. Let $P_1, \ldots, P_k$ be the projections defined on $V_1 + \ldots + V_k$ by

$$P_i(v_1 + \ldots + v_k) \stackrel{def}{=} v_i \quad \text{(where } v_j \in V_j \text{ for every } j\text{)}.$$

This projection is well defined. Denoting $U_i \stackrel{def}{=} P_i(V) \cap V$ for every i, we have that U_i is contained in V_i for every i. Also, since $U_i \subseteq V$, we have that $U_1 + \ldots + U_k \subseteq V$.

It is left to show that the dimension of the U_i's is bounded from below by $d_1 - kd_2$. We show it for U_1, and for the rest of the U_i's it follows similarly. Note that the projection P_1 and the projection $(P_2 + \ldots + P_k)$ are a partition of the identity on $V_1 + \ldots + V_k$. Using Lemma 3.12 and the bounds on the dimensions of V and of $V_1, \ldots, V_k$, we obtain that

$$\begin{aligned}
\dim(U_1) &= \dim(P_1(V) \cap V) = \dim(V) - \dim((P_2 + \ldots + P_k)(V)) \\
&= \dim(V) - \dim(P_2(V)) - \ldots - \dim(P_k(V)) \\
&\geq \dim(V) - \dim(V_2) - \ldots - \dim(V_k) \\
&\geq kd_1 - (k-1)(d_1 + d_2) \geq d_1 - kd_2
\end{aligned}$$

as desired. ☐

Corollary 3.17. *Let $V_1, \ldots, V_k$ be independent linear subspaces of some linear space W over $\mathbb{Z}_2$, of dimension at most $d_1 + d_2$ each. Also, let V be an affine subspace of $V_1 + \ldots + V_k + w$ for some $w \in W$, of dimension at least kd_1. Then there exist subspaces $U_1, \ldots, U_k$ of W, each of dimension at least $d_1 - kd_2$, such that $U_i \subseteq V_i$ for every i, and that V can be partitioned into translations of $U_1 + \ldots + U_k$.*

Proof. Let $v \in V$ be any vector, and apply Lemma 3.16 to the subspaces $V_1 + \ldots + V_k + w - v$, and $V - v$, obtaining $U_1, \ldots, U_k$ with the required dimensions, such that $U_i \subseteq V_i$ for all i, and such that $v \in U_1 + \ldots + U_k + v \subseteq V$. Now note that the space $V - v$ and the space $V_1 + \ldots + V_k + w - v$ are both independent of the choice of $v \in V$. Therefore we may assume without loss of generality that the U_i's are also independent of v, and therefore we have partitioned V into translations of $U_1 + \ldots + U_k$ as needed. ☐

Lemma 3.18. *Let X be an n-bit affine source with $\mathsf{r}(X) \geq \delta$, and let $S_1, \ldots, S_k \subseteq [n]$ be a partition of the coordinates of X into sets of size n/k each. If $\mathsf{r}(X_{S_i}) \leq \delta + \alpha$ for all i, then X is a convex combination of affine sources X' where the blocks $X'_{S_1}, \ldots, X'_{S_k}$ are all independent, and have entropy rate at least $\delta - k\alpha$ each.*

Proof. Let us denote the vector-space $\{0,1\}^n$ over $\mathbb{Z}_2$ by W, and let $V \stackrel{def}{=} \mathsf{Supp}(X)$. For every $i \in [k]$, let $P_i \colon W \to W$ be the projection that given $v \in W$, leaves all coordinates of v whose index is in S_i unchanged, and zeros all other coordinates. Define $V_1, \ldots, V_k$ by taking V_i to be the homogeneous translation of $P_i(V)$, and note that the V_i's are independent subspaces of W. It is easy to verify that for an appropriate $w \in W$,

$$V \subseteq V_1 + \ldots + V_k + w \subseteq W\ .$$

Since the entropy rate of each X_{S_i} is at most $\delta+\alpha$, the dimension of each V_i is at most $(\delta+\alpha)\frac{n}{k}$. Also, since the entropy rate of X is at least δ, we have that the dimension of V is at least δn. We can therefore apply Corollary 3.17, to obtain spaces U_i of dimension at least $(\delta - k\alpha)n/k$ each, such that $U_i \subseteq V_i$, and such that V is partitioned into translations of $U_1 + \ldots + U_k$.

In other words, we have that X is a convex combination of linear sources X', where $\mathsf{Supp}(X')$ is a translation of $U_1 + \ldots + U_k$. It is easy to verify that for such a source X', the sub-blocks $X'_{S_1}, \ldots, X'_{S_k}$ are independent, and have entropy rate at least $\delta - k\alpha$ each. This concludes the proof. $\square$

The following facts concerning affine sources are proven by arguments similar to the ones used in Lemma 3.16, Corollary 3.17 and Lemma 3.18, and are brought without proof.

Lemma 3.19. *Let X be an n-bit affine source, and let $R_0, R_1, \ldots, R_t \subseteq [n]$ be a partition of the set of coordinates. Suppose that $H^\infty(X_{(\cup_{j\geq 1} R_j)}|X_{R_0}) \geq K$. Then for every $\alpha \leq 1$ there exists an $i \in \{1, \ldots, t\}$, for which*

$$H^\infty(X_{R_i}|X_{(\cup_{j<i} R_j)}) \geq \alpha \cdot K \cdot \frac{|R_i|}{|\cup_{j\geq 1} R_j|}$$

(and therefore $\mathsf{r}(X_{R_i}|X_{(\cup_{j<i} R_j)}) \geq \alpha \cdot \mathsf{r}(X_{(\cup_{j\geq 1} R_j)}|X_{R_0})$),

$$\text{and} \qquad H^\infty(X_{(\cup_{j>i} R_j)}|X_{(\cup_{j\leq i} R_j)}) \geq (1-\alpha)K - |R_i| \ .$$

Lemma 3.20. *Let X be an n-bit affine source, and let $R_1, R_2, R_3 \subseteq [n]$ be disjoint sets of coordinates. Then there exists an affine sub-source $X' \subseteq X$ which satisfies*

1. *$H^\infty(X'_{R_1}) = H^\infty(X_{R_1})$, (and thus X'_{R_1} and X_{R_1} have the same distribution).*
2. *$H^\infty(X'_{R_2}|X'_{R_1}) = 0$, (and thus X'_{R_2} is a function of X'_{R_1}).*
3. *$H^\infty(X'_{R_3}|X'_{R_1\cup R_2}) = H^\infty(X_{R_3}|X_{R_1\cup R_2})$,*
   (and thus for every $y \in \mathsf{Supp}(X'_{R_1\cup R_2})$, $(X'_{R_3}|X'_{R_1\cup R_2} = y)$ and $(X_{R_3}|X_{R_1\cup R_2} = y)$ have the same distribution).

Lemma 3.21. *Let X be an n-bit affine source, and let $R_1, R_2 \subseteq [n]$ be disjoint sets of coordinates. Then there exists an affine sub-source $X' \subseteq X$, such that*

1. *X'_{R_1} and X'_{R_2} are independent,*
2. *$H^\infty(X'_{R_1}) \geq H^\infty(X_{R_1}) - |R_2|$,*
3. *$H^\infty(X'_{R_2}) = H^\infty(X_{R_2}|X_{R_1})$.*

4 A Constant-Seed Condenser

Loosely speaking, a *condenser* is a function that maps strings of length n into several strings of shorter length m such that if the input string comes from some distribution with entropy rate δ, then the entropy-rate of one of the shorter strings is higher than δ. A *constant-seed* condenser is a condenser where the number of such shorter strings is a constant. We prefer to take the view that the concatenation of all these blocks is *somewhere-condensed* (as per Definition 3.3). The formal definition of a condenser follows:

Definition 4.1 (Somewhere Condensers). A function $\mathtt{con} : \{0,1\}^n \to \{0,1\}^{m\times\ell}$ is called an ℓ-outputs *somewhere condenser* with expansion factor $\rho = \rho(\mathsf{r}(X))$ and distance $\epsilon = \epsilon(\mathsf{r}(X))$, if for every n-bit source X, $\mathtt{con}(X)$ is somewhere ϵ-close to having rate at least $\min\{\rho\cdot \mathsf{r}(X),\ 0.99\}$.

We will construct a condenser that condenses any δ-source to a source that is somewhere very close to having rate at least 0.99. The heart of this construction is the following *basic condenser* which condenses any δ-source to rate $\delta(1+\Omega(\delta))$, using only four blocks.

Theorem 4.2 (Basic condenser). *There exists a constant $\alpha > 0$ and a polynomial-time computable somewhere condenser* $\mathtt{bcon} : \{0,1\}^n \to \{0,1\}^{(n/3)\times 4}$ *which works for every source X, has expansion factor $\rho = 1+\alpha\cdot\delta$ (where $\delta = \mathsf{r}(X) \gg n^{-1/4}$) and distance $\epsilon = 2^{-\alpha\delta^2 n}$.*

Proof. We start by assuming that $n = 3p$ for some prime p. We later explain what we do in the case that n is not of this form. For $x = (x_1, x_2, x_3) \in \{0,1\}^{3p}$ we define $\mathtt{bcon}(x) = (x_1, x_2, x_3, x_1\cdot x_2 + x_3)$.

Let X be a δ-source over $\{0,1\}^{3p}$. Let $\theta = \alpha_0\delta/1000$, where α_0 is the constant obtained in Lemma 3.8. Note that Lemma 3.8 guarantees that for every three independent $\delta(1-10\theta)$-sources A_1, A_2, A_3 over $\mathrm{GF}(2^p)$, the distribution $A_1\cdot A_2 + A_3$ is $2^{-10\theta\delta p}$-close to having rate at least $\delta(1+10\theta)$. We will show that there is a random variable $I = I(X)$ such that $\mathtt{bcon}(X)_I$ is $2^{-\theta\delta p}$-close to having rate at least $(1+\theta)\delta$.

We assume without loss of generality that X is a flat source, since it is enough to prove the theorem for such sources, and so we can identify X with the subset of elements on which it is supported. For $i \in [4]$ define $H_i \subseteq \mathrm{GF}(2^p)$ to be the set of "heavy" elements of the distribution $\mathtt{bcon}(X)_i$. That is, $H_i = \{y \in \mathrm{GF}(2^p) \mid \Pr[\mathtt{bcon}(X)_i = y] \geq 2^{-(1+\theta)\delta p}\}$. Note that for every $i \in [4]$, $|H_i| \leq 2^{(1+\theta)\delta p}$. If $\Pr[\exists_i \text{ s.t. } \mathtt{bcon}(X)_i \notin H_i] > 1 - 2^{-\theta\delta p}$ then we're done, since we can define for all but an exponentially small fraction of the x's the index $I(x)$ to be the smallest $i \in [4]$ such that $\mathtt{bcon}(x)_i \notin H_i$. Clearly, regardless of how I is defined on these few "bad" x's, we get that $\mathtt{bcon}(X)_I$ is exponentially close to having rate $\geq (1+\theta)\delta$.

Therefore, we may assume that $\Pr[\forall_i \mathtt{bcon}(X)_i \in H_i] \geq 2^{-\theta\delta p}$. Since X is a flat source this just means that

$$\Big|\big\{x = (x_1, x_2, x_3) \in X \mid x \in H_1 \times H_2 \times H_3, x_1\cdot x_2 + x_3 \in H_4\big\}\Big| \geq \frac{2^{3\delta p}}{2^{\theta\delta p}} = 2^{(3-\theta)\delta p} \tag{4}$$

Let X' denote this set, which is a subset of $X \cap H_1 \times H_2 \times H_3$. We observe that Equation 4 implies that $|H_1| \geq 2^{(1-3\theta)\delta p}$, since otherwise we'll have $|X'| \leq |H_1|\cdot|H_2|\cdot|H_3| < 2^{(1-3\theta)\delta p}2^{(1+\theta)\delta p}2^{(1+\theta)\delta p} = 2^{(3-\theta)\delta p}$. Using the same reasoning $|H_2|, |H_3| \geq 2^{(1-3\theta)\delta p}$. We define A_1, A_2, A_3 to be three independent random variables over $\mathrm{GF}(2^p)$ where for $i = 1,2,3$, A_i is the flat distribution over the set H_i. Note that for every $i = 1,2,3$, the random variable A_i is at least a $\delta(1-3\theta)$-source over $\mathrm{GF}(2^p)$ (since $|H_i| \geq 2^{(1-3\theta)\delta p}$). By Equation 4, it holds that

$$\Pr[A_1\cdot A_2 + A_3 \in H_4] = \frac{|X'|}{|H_1|\cdot|H_2|\cdot|H_3|} \geq \frac{2^{(3-\theta)\delta p}}{|H_1|\cdot|H_2|\cdot|H_3|} \geq \frac{2^{(3-\theta)\delta p}}{(2^{(1+\theta)p})^3} = 2^{-4\theta\delta p} \tag{5}$$

However, for every $(\delta+10\theta)$-source Y over $\mathrm{GF}(2^p)$ it holds that

$$\Pr[Y \in H_4] \leq |H_4| 2^{-(1+10\theta)\delta p} \leq 2^{(1+\theta)\delta p}2^{-(1+10\theta)\delta p} = 2^{-9\theta\delta p} \tag{6}$$

Equations 5 and 6 together imply that the statistical distance of $A_1\cdot A_2 + A_3$ to *every* $\delta(1+10\theta)$-source Y is at least $2^{-4\theta\delta p} - 2^{-9\theta\delta p} > 2^{-4\theta\delta p - 1}$, contradicting Lemma 3.8.

In the case n is not of the form $n = 3p$ for a prime p, we first pad the input with zeros to length n' for the smallest $n' > n$ such that $n' = 3p$. By Fact 3.9, $n' < n + 3n^{3/4}$. Since the entropy of X

is unchanged by this padding, we get that the rate of X as a random variable over $\{0,1\}^{n'}$ is equal to $\delta\frac{n}{n'} > \frac{\delta}{1+3n^{-1/4}}$. By applying the condenser this rate is going to grow by a multiplicative factor of $1+\theta = 1+\Omega(\delta)$ which (since $n^{-1/4} \ll \delta$) will dwarf this factor of $\frac{1}{1+3n^{-1/4}}$. $\square$

We note that the results of [BIW] are somewhat stronger for a field of *prime* order than for the field GF(2^p) (i.e., a growth factor of $1+\Omega(1)$ instead of $1+\Omega(\delta)$). However, we don't use these stronger results since we are not aware of a deterministic polynomial-time uniform algorithm that on input n outputs a prime of length n bits (or at least a prime with length in the range $[n, n+o(n)]$).[15]

4.1 Composing Condensers

In this section we show that we can compose a condenser with expansion ρ and ℓ blocks together with a condenser with expansion ρ' and ℓ' blocks to obtain a condenser with expansion $\rho \cdot \rho'$ and $\ell \cdot \ell'$ outputs.

Combined with Theorem 4.2 this implies that we get a condenser with arbitrarily large expansion factor (as long as the output is below 0.99 entropy rate) using only a constant number of blocks.

Lemma 4.3 (Condenser composition). *Let:*

1. $\mathtt{con} : \{0,1\}^n \to \{0,1\}^{m\times\ell}$ *be a somewhere condenser with ρ expansion factor and distance ϵ.*
2. $\mathtt{con}' : \{0,1\}^m \to \{0,1\}^{m'\times\ell'}$ *be a somewhere condenser with ρ' expansion factor and distance ϵ', where $\rho\cdot\rho' < 0.99$.*

Define the $\ell\cdot\ell'$-block condenser $\mathtt{con}' \circ \mathtt{con}$ as follows: for $\langle i,i'\rangle \in [\ell]\times[\ell']$ the $\langle i,i'\rangle$-th block of $\mathtt{con}'\circ\mathtt{con}(x)$ is $\mathtt{con}'_{i'}(\mathtt{con}_i(x))$. Then $\mathtt{con}'\circ\mathtt{con} : \{0,1\}^n \to \{0,1\}^{m'\times(\ell\cdot\ell')}$ is a somewhere condenser for $\delta-$sources with $\rho\cdot\rho'$ expansion factor (where $\rho=\rho(\delta)$ and $\rho'=\rho'(\rho\delta)$) and distance $\epsilon+\epsilon'$.

Proof. Let X be some δ-source (where $\rho'\rho\delta < 0.99$, for $\rho=\rho(\delta)$ and $\rho'=\rho'(\rho\delta)$) and consider the random variable $I = I(X)$ over $[\ell]$ such that $Y = \mathtt{con}(X)_I$ is ϵ-close to a $\delta\rho$-source Z (such an I exists since $\mathtt{con}$ is a ρ-condenser). Since $\mathtt{con}'$ is a ρ'-condenser, there exists a random variable I' such that $\mathtt{con}'(Z)_{I'}$ is ϵ'-close to having rate at least $\rho'\rho\delta$, thus if we plug in Y instead of Z we get that $\mathtt{con}'(Y)_{I'} = \mathtt{con}'\circ\mathtt{con}(X)_{\langle I,I'\rangle}$ is $\epsilon+\epsilon'$ close to having rate at least $\rho'\rho\delta$. $\square$

Using composition for every constant $\delta > 0$, we can condense a distribution with rate δ to having entropy rate of 0.99. (In fact, our arguments work even if δ is slightly sub constant.) Lemma 4.3 together with Theorem 4.2 easily imply the following result.

Corollary 4.4. *For every δ there exists $\ell = 2^{\Theta(1/\delta)}$, $m = 2^{-\Theta(1/\delta)}n$ and a $2^{\Theta(1/\delta)}\mathrm{poly}(n)$-time computable somewhere condenser $\mathtt{con} : \{0,1\}^n \to \{0,1\}^{m\times\ell}$ which works for distribution with rate δ and gives output which somewhere $2^{-\Theta(m)}$-close to having rate 0.99.*

We remark than Ran Raz [Raz] had obtained independently a proof of Corollary 4.4 (although not of Theorem 4.2) which also uses the [BIW] construction but in a different way.

[15] Such an algorithm does exist assuming the Extended Riemann Hypothesis.

Obtaining Arbitrarily Small Block Lengths. In some cases, given a condenser $\mathtt{con} : \{0,1\}^n \to \{0,1\}^{m\times\ell}$ with a block size of m, we may want to obtain a condenser blocks of *smaller* size $m' < m$ (where we can always assume without loss of generality that m' divides m). This will be easy to do by simply splitting each m-sized block into several disjoint pieces of size m', thus getting a condenser $\ell \cdot \frac{m}{m'}$ blocks of size m'. By the following proposition, one of these disjoint subblocks will have (roughly) at least the same rate as the original block.

Proposition 4.5. *Let X be a δ-source over $\{0,1\}^m$ and suppose that X is equal to the concatenation of k random variables $X_1, \ldots, X_k$ each of length m'. Let $\epsilon > 0$. Then there exists a random variable I over $[k]$ such that X_I is $2^{-\epsilon m}$-close to having rate at least $\delta - \epsilon$.*

Proof. Assume X is a flat distribution (i.e., X is the uniform distribution over a set of size at least $2^{\delta m}$). For $i \in [k]$, define H_i as the "heavy" elements of X_i, that is $H_i = \{y \in \{0,1\}^{m'} \mid \Pr[X_i = y] \geq 2^{-(\delta-\epsilon)m'}\}$. Then, what we need to prove is that $\Pr[X \in H_1 \times \cdots \times H_k] \leq 2^{-\epsilon n}$. But this immediately follows from the fact that $|H_i| \leq 2^{(\delta-\epsilon)m'}$ and so $|H_1|\cdots|H_k| \leq 2^{(\delta-\epsilon)n}$. □

Using Proposition 4.5 together with Corollary 4.4 we can get the following condenser that condenses a distribution over $\{0,1\}^n$ with entropy rate δ into a distribution over $\{0,1\}^{\beta n}$ that is close to having rate (say) 0.97.

Corollary 4.6 (Final condenser). *For every $\delta > 0$, there exists constant $\nu(\delta) = 2^{-\Theta(1/\delta)}$, such that for every constant $\beta < \nu$ there exist constant $\ell = \ell(\delta, \beta)$ and a polynomial time computable somewhere condenser $\mathtt{con} : \{0,1\}^n \to \{0,1\}^{\beta n \times \ell}$ which works for distributions with rate δ and gives output somewhere $2^{-\Theta(\beta n)}$-close to having rate 0.97.*

5 A 2-sample Somewhere-Extractor

In this section, we construct a 2-sample somewhere-extractor, which is a function that given inputs from two independent n-bit sources with rate at least δ (where δ is an arbitrarily small constant and can be even slightly sub-constant) outputs a somewhere-uniform distribution. We recall that distribution X over $\{0,1\}^{m\times\ell}$ is somewhere-uniform with ℓ blocks if there exists a random variable I over $[\ell]$ (dependent on X) such that $X_I = U_n$. The main facts that we will use regarding somewhere-uniform distributions are summarized in the following claim.

Claim 5.1 (Properties of somewhere-uniform sources).

- *If X is somewhere-uniform distribution over $\{0,1\}^{m\times\ell}$ then it has high min-entropy, i.e., $H^\infty(X) \geq m - \log \ell$. Indeed, note that $H^\infty(X, I) \geq H^\infty(X_I)$ and $H^\infty(X) \geq H^\infty(X, I) - \log |I|$. In fact, we will always consider somewhere-uniform distribution X where the number of blocks ℓ is small enough compared to the block-size m and so $\mathsf{r}(X) \geq 0.99$.*

- *Given a somewhere-uniform distribution X over $\{0,1\}^{m\times\ell}$ and $m' < m$ we can obtain a somewhere-uniform distribution X' over $\{0,1\}^{m'\times\ell}$, by simply truncating each block of X to size m'.*

- *Let X be a distribution over $\{0,1\}^{m\times\ell}$ such that there exists a (dependent) random variable I such that X_I is ϵ-close to uniform. Then, by Proposition 3.4, X is ϵ-close to a somewhere-uniform distribution.*

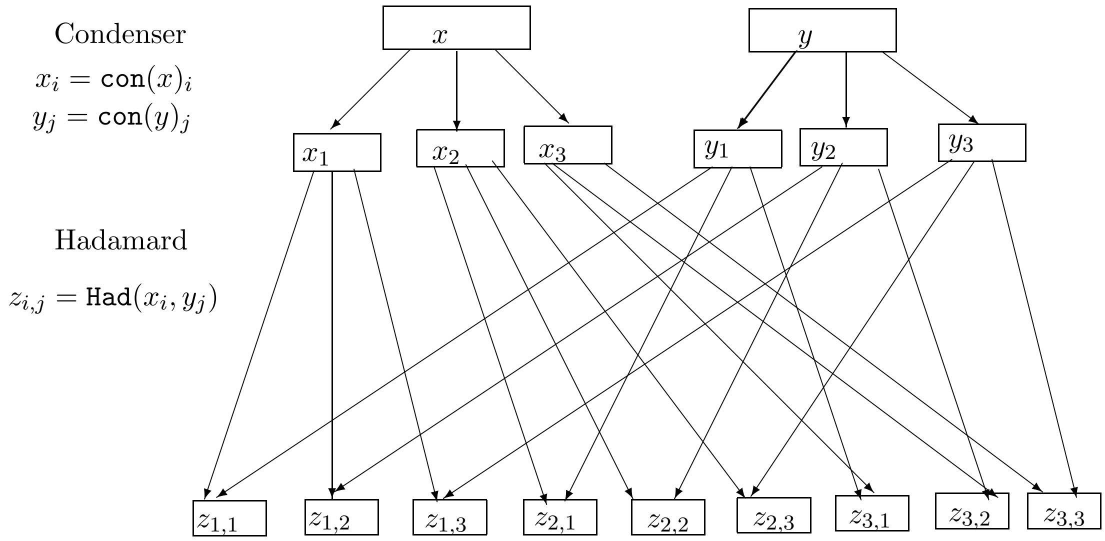

Figure 5: A 2-Sample Somewhere-Extractor

Definition 5.2 (Two-sample somewhere-extractor). We say that

$$\texttt{s\_ext}_m : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^{m \times \ell}$$

is a *somewhere-extractor* with rate requirement δ and distance ϵ if for every two independent distributions X over $\{0,1\}^{n_1}$ and Y over $\{0,1\}^{n_2}$ with rates at least δ, $\texttt{s\_ext}(X,Y)$ is ϵ-close to a somewhere random distribution. The index m indicates the length of the block in the output and we will frequently omit it when its value is clear from the context.

A fixed $y \in \{0,1\}^{n_2}$ is *good* for X if $\texttt{s\_ext}(X,y)$ is ϵ-close to a somewhere random distribution. Similarly, a fixed $x \in \{0,1\}^{n_1}$ is good for Y if $\texttt{s\_ext}(x,Y)$ is ϵ-close to a somewhere random distribution. We say that $\texttt{s\_ext}$ is a *strong* somewhere-extractor with the same parameters as above if for every X, Y as above,

$$\Pr_{y \in_{\mathrm{R}} Y}[y \text{ is good for } X] > 1 - \epsilon$$

and

$$\Pr_{x \in_{\mathrm{R}} X}[x \text{ is good for } Y] > 1 - \epsilon.$$

The main result of this section is the following theorem which claims that for every constant δ there is a somewhere random extractor that works for sources with rate δ and has a constant number of output blocks.

Theorem 5.3 (Somewhere-extractor). *For every fixed $\delta > 0$ and $0 < \gamma \leq 1$, there are constants $\ell, \beta, \eta > 0$ and a polynomial-time computable ℓ-output strong somewhere-extractor $\texttt{s\_ext}_m$: $\{0,1\}^n \times \{0,1\}^{\gamma n} \to \{0,1\}^{m \times \ell}$ that works for all $m \leq \beta n$ and sources with rate δ and has distance $\epsilon \leq 2^{-\eta n}$.*

Note that a special case is $\gamma = 1$ in which the two input sources are of the same length.

The construction. The outline of our construction is sketched in Figure 5. Roughly speaking, the idea behind the construction is to use the condenser derived from Corollary 4.6 to obtain from each source a list of blocks one of whom has entropy rate at least 0.97. We then apply to every possible pair of blocks the previously known construction of a two-sample extractor that works for sources with rate larger than $\frac{1}{2}$ (i.e., for every choice of a block from the first source and a block from the second source we run the two-sample extractor on both blocks and output the results). The number of blocks in the output is the product of the number of output blocks of two condensers. Another tool which we need, in addition to Corollary 4.6, is the following result on two-source extractors.

Theorem 5.4 (Two-source extractors [CG, Vaz, DEOR]). *For every integer n there exists $k = \Omega(n)$ and a polynomial-time two-sample strong extractor* $\mathtt{Had} : \{0,1\}^{n\times 2} \to \{0,1\}^k$ *which works for sources with rate at least* 0.6 *and has distance* $2^{-\Omega(n)}$.

To construct $\mathtt{Had}$ one can take $n \times n$ matrices $A_1, \ldots, A_k$ over $GF(2)$ with the property that sum of every subset of them has high rank, e.g., cyclic shift matrices. Then for any two n-bit vectors x, y the output of the extractor is string $(x \cdot A_1 y, \ldots, x \cdot A_k y)$, where both inner product and matrix-vector multiplications are over $GF(2)$.

Proof of Theorem 5.3. Given the parameters δ and γ let $\nu = \nu(\delta)$ be the constant from Corollary 4.6. Let $n_1 = n$, $n_2 = \gamma n$, $\beta_1 = \gamma\nu$ and $\beta_2 = \nu$. Apply Corollary 4.6 to obtain two condensers $\mathtt{con}_1, \mathtt{con}_2$ using for each $i \in \{1,2\}$ the parameters δ, n_i, β_i. Each condenser $\mathtt{con}_i$ has distance $\epsilon_i = 2^{-\Theta(\beta_i n_i)}$ and outputs ℓ_i blocks of length $n' \stackrel{def}{=} \beta_i n_i = \gamma\nu n$, which is the same for both condensers. Let $\mathtt{Had}_{n'}$ be the two source extractor of Theorem 5.4 with input length n'. Our somewhere random two sample extractor will have $\ell = \ell_1 \cdot \ell_2$ blocks. For every $j = (j_1 - 1)\ell_2 + j_2$ with $j_i \in [\ell_i]$ we define

$$\mathtt{s\_ext}(x_1, x_2)_j = \mathtt{Had}_{n'}(\mathtt{con}_1(x_1)_{j_1}, \mathtt{con}_2(x_2)_{j_2})$$

By definition, the length of every output block is $\Omega(n') = \beta n$ for some constant $\beta > 0$. Let X_1, X_2 be distributions with rate δ over $\{0,1\}^{n_1}$ and $\{0,1\}^{n_2}$ respectively. For $i \in \{1,2\}$, by definition of $\mathtt{con}_i$, there exist random variable I_i such that $\mathtt{con}_i(X_i)_{I_i}$ is ϵ_i close to have rate $0.97 > 0.6$. This implies that $\mathtt{s\_ext}(X_1, X_2)_I$, $I = (I_1 - 1)\ell_2 + I_2$ is ϵ-close to uniform distribution for $\epsilon = \left(\epsilon_1 + \epsilon_2 + 2^{-\Omega(n')}\right) = 2^{-\Omega(n')}$. Therefore the output of the extractor is ϵ-close to a somewhere uniform distribution. Moreover, using the fact that $\mathtt{Had}$ is a strong extractor it is easy to check that $\mathtt{s\_ext}$ also have this property. Since $n' = \Omega(n)$, we conclude there exist a constant $\eta > 0$ (that depends on δ and γ) such that $\epsilon \leq 2^{-\eta n}$. Finally we truncate the output blocks to length m to get $\mathtt{s\_ext}_m$. □

6 A 3-source Extractor

In this section we describe a three source extractor for sources with linear entropy. We prove the following result, which is the first construction of an ℓ-source extractor that works for sources with rate δ such that constant $\ell \cdot \delta$ can be made arbitrarily small.

Theorem 6.1 (3-source extractor). *For every constants $\delta, \epsilon > 0$ and $m \in \mathbb{N}$, and for every sufficiently large integer n there exists a* $\mathrm{poly}(n)$*-time computable* 3*-source extractor* $\mathtt{3ext} : \{0,1\}^{n\times 3} \to \{0,1\}^m$ *with rate requirement δ and distance ϵ.*

The two tools we use in this construction is the somewhere random extractor of Section 5 and an *optimal* two source extractor with very small input size. Such an extractor can be constructed using essentially exhaustive search. The details are given in the next section.

6.1 An optimal 2-source extractor

A probabilistic argument shows the existence of a two source extractor with very low min-entropy requirement. For small input lengths we can efficiently find it by exhaustive search. The next Lemma describes the extractor we use in our construction.

Lemma 6.2 (Optimal Extractor). *For every integers d and $m = \lfloor \log d \rfloor$, there exists a $2^{5d^{14}}$-time computable 2-source extractor* $\texttt{opt} : \{0,1\}^{d\times 2} \to \{0,1\}^m$ *with $6\log d$-entropy requirement and distance $1/d$.*

This result follows from a more general lemma that involves some additional parameters.

Lemma 6.3. *Let $m < k < d$ be integers, and let $\epsilon > 0$. There exist some constant $a > 0$ such that if $k > \log d + 2m + 2\log(1/\epsilon) + a$ then there exists a $2^{5d^2 \cdot 2^{2k}}$-time computable 2-source extractor* $\texttt{opt} : \{0,1\}^{d\times 2} \to \{0,1\}^m$ *with k-entropy requirement and distance ϵ.*

Proof. We first give a probabilistic argument that such an extractor exists, and then find it by brute force. Let $M = 2^m$, $D = 2^d$ and $K = 2^k$. Let A be a randomly chosen $D \times D$ matrix with entries in $\{0,1\}^m$, where the $D^2 m$ bits used as an entries of A are $K^2 m$-wise independent. We show that with positive probability the function $\texttt{opt}(x,y) = A_{x,y}$ is an extractor with the required parameters. We say that R is a *rectangle* if $R \in \binom{[D]}{K} \times \binom{[D]}{K}$. We first show that for every rectangle R, nonzero $v \in \{0,1\}^m$ and $\epsilon > 0$

$$\Pr\left[\Big|\sum_{(i,j)\in R} (-1)^{<A_{ij},v>}\Big| > K^2\epsilon/M\right] < e^{-\Omega(\frac{\epsilon^2 K^2}{M^2})}. \tag{7}$$

For every $(i,j) \in R$ and nonzero $v \in \{0,1\}^m$ define a random variable $B_{ij} = (-1)^{<A_{ij},v>}$ and let $B = \sum_{(i,j)\in R} B_{ij}$. Note that $\mathbb{E}[B] = 0$. Moreover, since every K^2 entries of A are independent, we have the independence of variables B_{ij}. Therefore the Chernoff bound implies that $\Pr\big[|B| > \epsilon K^2/M\big] < e^{-\Omega(\frac{\epsilon^2 K^2}{M^2})}$.

We say that A passes R and v if $\big|\sum_{(i,j)\in R}(-1)^{<A_{ij},v>}\big| \le K^2\epsilon/M$. By a union bound over all rectangles and nonzero vectors we get that the probability that there exist a rectangle R and nonzero v such that A does not pass R and v is bounded by

$$\binom{D}{K}^2 \cdot M \cdot e^{-\Omega(\frac{\epsilon^2 K^2}{M^2})} \le 2^{3Kd - \Omega(\frac{\epsilon^2 K^2}{M^2})}$$

It is easy to verify that there exists some constant $c > 0$ such that the right side of the above inequality is smaller than one when $K \ge \frac{cdM^2}{\epsilon^2}$. This holds for $k > \log d + 2m + 2\log(1/\epsilon) + O(1)$. Thus, there exists a matrix A' which passes all rectangles and all vectors v. Let $\texttt{opt}(x,y) = A'_{x,y}$. We now verify that $\texttt{opt}$ is a 2-sample extractor with k-entropy requirement and distance ϵ. It is sufficient consider only pairs of independent sources where each component is a flat distribution. Each such pair of sources is uniformly distributed over some rectangle R. Fix such an R and let Y denote the output distribution of $\texttt{opt}$ on R. For every nonzero $v \in \{0,1\}^m$ we have that the expected bias of Y in v is $\big|\mathbb{E}[(-1)^{<Y,v>}]\big| \le \epsilon/M$. Thus, the distribution Y is ϵ/M-biased and by

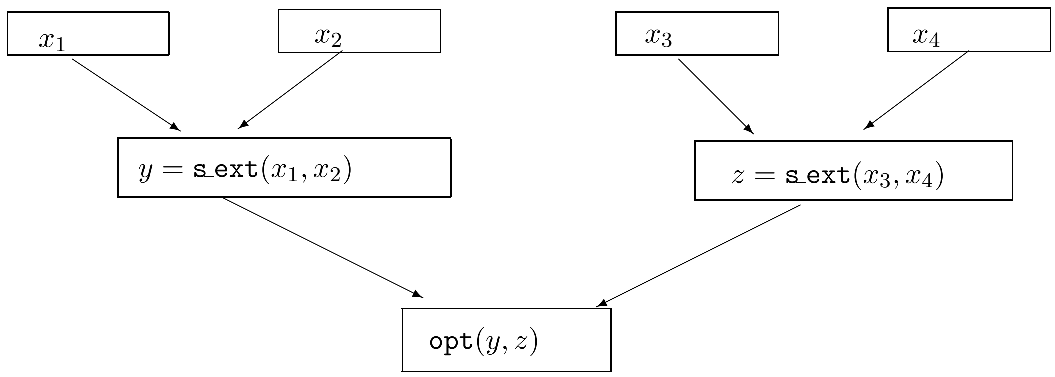

Figure 6: A 4-source extractor

the Vazirani XOR-lemma [Gol], we have that Y is ϵ-close to uniform. This shows that $\mathtt{opt}$ indeed is an extractor.

To find a matrix which passes all rectangles and vectors we go over all possible matrices A in the appropriate sample space. There are efficient constructions of such a space that are of size $(D^2m)^{K^2m}$. For each matrix we check all rectangles and vectors, altogether $\binom{D}{K}^2 \cdot M$ choices. Finally for each such choice we check whether (7) holds in time K^2m. Thus, the total time we need is bounded by

$$(D^2m)^{K^2m} \cdot \binom{D}{K}^2 \cdot M \cdot K^2m \leq 2^{3d^2 \cdot 2^{2k} + d \cdot 2^{k+1} + m + 3k} \leq 2^{5d^2 \cdot 2^{2k}}$$

□

6.2 An Appetizer: A 4-Source Extractor

Before proving Theorem 6.1 we prove a slightly relaxed version of it, which demonstrates the main idea behind the proof. That is, we prove that there exists a 4-source extractor with the same parameters. The construction of this extractor is quite simple.

The construction Given the constants δ, ϵ and m we choose d large enough to satisfy:

- $\frac{1}{d} < \frac{\epsilon}{2}$ (recall that ϵ is the desired statistical distance)
- $\log d \geq m$
- $d/\ell - \log \ell \geq 6 \log d$.

Next set $\gamma = 1$ and apply Theorem 5.3 to get a somewhere random extractor for two equal length sources with entropy rate δ. Let ℓ, β and η be the constants whose existence is guaranteed by the theorem and note that since n is large $\beta n > d/\ell$. We obtain a somewhere random extractor $\mathtt{s\_ext}_{d/\ell}$ that outputs ℓ blocks of length d/ℓ which are $2^{\eta n}$-close to a somewhere random distribution. The total length of the output of $\mathtt{s\_ext}_{d/\ell}$ is d.

Let $\mathtt{opt}$ be the "optimal extractor" from Lemma 6.2 with input length d. As d is a constant it follows that $\mathtt{opt}$ can be computed in constant time. Note that the output length of $\mathtt{opt}$ is $\log d \geq m$,

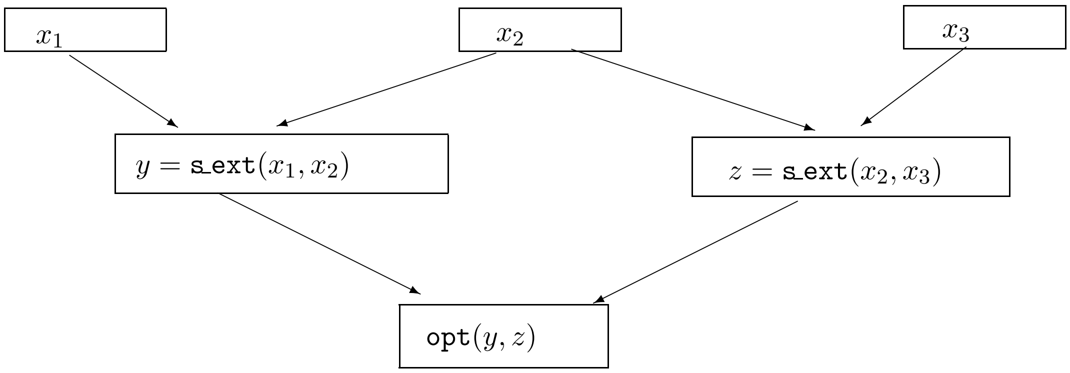

Figure 7: A 3-source extractor

so we can truncate it to length m. Our 4-sample extractor $\mathtt{ext}_4 : \{0,1\}^{n\times 4} \to \{0,1\}^m$ is defined as follows (see also Figure 6):

$$\mathtt{4ext}(x_1, x_2, x_3, x_4) = \mathtt{opt}\Big(\mathtt{s\_ext}_{d/\ell}(x_1, x_2)\ ,\ \mathtt{s\_ext}_{d/\ell}(x_3, x_4)\Big)$$

Proof. Let $X_1, \ldots, X_4$ be independent distributions over $\{0,1\}^n$ all with entropy rate at least δ. We have that $\mathtt{s\_ext}(X_1, X_2)$ is $2^{-\eta n}$-close to a somewhere random distribution with ℓ blocks. Recall that a somewhere random distribution with ℓ blocks of length d/ℓ has min-entropy at least $d/\ell - \log \ell > 6 \log d$. It follows that $\mathtt{s\_ext}(X_1, X_2)$ is $2^{-\eta n}$-close to having min-entropy $6 \log d$ and the same holds for $\mathtt{s\_ext}(X_3, X_4)$. Thus, we apply $\mathtt{opt}$ with two independent distributions which meet its entropy requirement and can conclude that the output is $(2 \cdot 2^{-\eta n} + 1/d)$-close to uniform. We have that $1/d < \epsilon/2$ and therefore for large enough n, $2 \cdot 2^{-\eta n} + 1/d \leq \epsilon$, as required. $\square$

6.3 Down from 4 sources to 3

To construct a 3-source extractors we use the fact that our somewhere random extractor is in fact *strong*. Using this property we are able to show that we can "reuse" one of the sources without harming the input distribution.

The construction: Given the constants δ, ϵ and m we choose the parameters as in the construction for 4-sources. Our 3-source extractor is given by:

$$\mathtt{3ext}(x_1, x_2, x_3) = \mathtt{opt}\Big(\mathtt{s\_ext}_{d/\ell}(x_1, x_2)\ ,\ \mathtt{s\_ext}_{d/\ell}(x_2, x_3)\Big)$$

Proof of Theorem 6.1. Let X_1, X_2, X_3 be independent distributions over $\{0,1\}^n$ with entropy rate at least δ. Recall that $\mathtt{s\_ext}$ is a strong somewhere random extractor, and recall the notion of a "good" input from Definition 5.2. For $i \in \{1,3\}$, let $B_i = \{x : x \text{ is bad for } X_i\}$. We have that $\Pr_{x \in_{\mathrm{R}} X_2}[x \in B_i] \leq 2^{-\eta n}$. Let $B = B_1 \cup B_3$. It follows that if we fix random value of x_2 then with probability $1 - 2 \cdot 2^{-\eta n}$ we obtain that $x_2 \notin B$. For every such fixed $x_2 \notin B$, the distributions $\mathtt{s\_ext}(X_1, x_2)$ and $\mathtt{s\_ext}(x_2, X_3)$ are independent. Furthermore, each one of them is $2^{-\eta n}$-close to a somewhere random distribution with ℓ blocks of length d/ℓ. As in the case of 4 sources, such a distribution meets the entropy requirements of $\mathtt{opt}$ and therefore for all $x_2 \notin B$,

$\texttt{opt}\big(\texttt{s\_ext}(X_1, x_2), \texttt{s\_ext}(x_2, X_3)\big)$ is $(2 \cdot 2^{-\eta n} + 1/d)$-close to uniform distribution. This implies that $\texttt{opt}\big(\texttt{s\_ext}(X_1, X_2), \texttt{s\_ext}(X_2, X_3)\big)$ is $(4 \cdot 2^{-\eta n} + 1/d)$-close to uniform. Once again, we note that $(4 \cdot 2^{-\eta n} + 1/d) \leq \epsilon$ for large enough n. $\square$

7 A 2-Source Disperser

In this section we present one of our main results. For every constant $\delta > 0$ we construct a 2-source disperser which works for sources with rate δ. More precisely we prove the following stronger statement.

Theorem 7.1 (2-source disperser). *For every two constants $\delta > 0$ and $m \in \mathbb{N}$ there exists a polynomial-time computable 2-source errorless disperser $\texttt{disp} : \{0,1\}^{n \times 2} \to \{0,1\}^m$ with rate requirement δ. Moreover, there exists a constant $\vartheta = \vartheta(m, \delta)$, such that for every two sources X, Y with rate at least δ and for every $z \in \{0,1\}^m$ we have*

$$\Pr\big[\texttt{disp}(X, Y) = z\big] \geq \vartheta.$$

Using this result for $m = 1$ one can obtain a new explicit construction of bipartite Ramsey graphs.

Corollary 7.2. *For every constants $\delta > 0$ there there exists a polynomial-time computable 2-edge coloring of the complete bipartite graph $K_{N,N}$ such that every induced subgraph of size N^δ by N^δ has a constant proportion of edges of both colors.*

To prove this corollary, let $V_1 = V_2 = \{0,1\}^n$ be two sets of vertices of size $N = 2^n$. Define a 2-coloring of the edges of the complete bipartite graph $K_{N,N}$ with parts V_1 and V_2 as follows. The color of an edge $(x, y), x \in V_1, y \in V_2$ is $\texttt{disp}(x, y)$. Then the above theorem implies that for every two subsets $A \subseteq V_1$ and $B \subseteq V_2$ of size at least $N^\delta = 2^{\delta n}$ and for every color, the constant proportion of edges between A and B has this color. In particular, this coloring has no monochromatic induced subgraph of size N^δ by N^δ. This is the first known construction of such coloring with constant $\delta < 1/2$.

Given 2-edge coloring $f : N \times N \to \{0,1\}$ of $K_{N,N}$ construct the following 2-edge coloring of complete graph on N vertices. For every $x < y$ the color of the edge (x, y) is $f(x, y)$. Note that if in the original coloring every induced subgraph of $K_{N,N}$ of size M by M had a constant proportion of edges of both colors, then the new coloring has the property that every induced subgraph of K_N of size $2M$ also has constant proportion of edges of both colors. Therefore, for any constant δ we have a polynomial-time computable 2-edge coloring of the complete graph K_N such that every induced subgraph of size $M = N^\delta$ has a constant proportion of edges of both colors. Though the result of Frankl and Wilson [FW] gives explicit construction of Ramsey graphs with better parameters, it can only guarantee that no set of appropriate size is monochromatic. Therefore we obtain a qualitative improvement of Frankl-Wilson construction.

7.1 The Construction

The construction of our disperser uses the same ingredients used in the proof of Theorem 6.1 but in a more complicated way. Let us fix n and δ as in the condition of the theorem, and let x and y be two n-bit strings. We need to compute $\texttt{disp}(x, y)$.

Partitions. Our construction uses various ways to partition the input strings x and y into three parts. We divide $[n]$ into $t \stackrel{def}{=} 10/\delta$ equal parts of length $\delta n/10$. (We assume from now on that t divides n. This is without loss of generality as we can always add dummy bits to inputs x and y to make the requirement hold, at the cost of only slightly decreasing δ.) We use the notation $x[i]$ to denote the i'th block of x. (We will think of $x[1]$ as the rightmost block an of $x[t]$ as the leftmost block). For each $i \in [t]$ we consider a partition of $[n]$ into three segments in the following way. The first segment I_1 consists of the blocks $1, \ldots, i-1$. The second segment I_2 consists of the block i and the third segment I_3 consists of the blocks $i+1, \ldots, t$. (It is possible that some segments are empty.) We use the notation $x^i_j \stackrel{def}{=} x^i_{I_j}$ to denote the restriction of x to the indices in the j'th segment. We often omit i if it is clear from the context. A *partition* I is a pair (i_x, i_y) that is used to partition two strings $x, y \in \{0,1\}^n$. More precisely, given a partition I and strings $(x,y) \in \{0,1\}^{n \times 2}$ we define $x^I_j \stackrel{def}{=} x^{i_x}_j$ and $y^I_j = y^{i_y}_j$. We often omit I if it is clear from the context.

A partial order on partitions. We define a partial order $\prec$ on the partitions in the following way: Given two partitions $I = (i_x, i_y)$ and $I' = (i'_x, i'_y)$ we say that $I \prec I'$ if $i_x \le i'_x$ *and* $i_y \le i'_y$.

Outline. The general operation of our disperser will be as follows:

1. For every partition I we will define a polynomial time procedure $\mathtt{disp}_I(x,y)$ in some way. The final output of the disperser will be $\mathtt{disp}_I(x,y)$ for some partition I that will be chosen as a function of x and y. We next explain how to choose the partition I.
2. For every partition I we will define a polynomial time procedure $\mathtt{test}_I(x,y)$ that outputs a Boolean value. We say that the partition "passes" the test if $\mathtt{test}_I(x,y)$ accepts.
3. The operation of the disperser is as follows: When given the inputs x, y, we go over all partitions I and compute $\mathtt{test}_I(x,y)$. We choose the partition I that is maximal with respect to the partial order $\prec$ amongst the partitions that passed the test. (if there is no unique maximum then we choose arbitrarily between maximal partitions). Finally, we output $\mathtt{disp}_I(x,y)$.

We now describe the operation of the disperser in detail. We will intersperse the description with remarks explaining the intuition. The formal analysis will follow the description.

Preliminaries and parameters Recall that we are given constants $\delta > 0$ and m (where m is the required output length), and that $|x| = |y| = n$. Our construction works as long as n is large enough as a function of δ and m.

By Theorem 5.3 we have a somewhere random extractor for any two sources such that their size ratio is a constant, and for any entropy rate $\delta > 0$. From now on we will abuse the notation and use $\mathtt{s\_ext}$ on input size varying from $\delta^3 n$ to n and entropy requirement varying from $\delta^5 n$ to n. Note that there exist a constant ℓ that is an upper bound on the number of blocks for all such applications. We always assume that the number blocks is ℓ by adding dummy blocks if necessary. Furthermore, we have that the error is at most $2^{-\eta n}$ for some constant $\eta > 0$ for all applications. Thus, we can assume that the error is smaller than any constant for large enough n. Recall that $\mathtt{s\_ext}_b$ denote the somewhere random extractor with output block length b.

We use Lemma 6.2 to obtain a 2-source extractor $\mathtt{opt} : \{0,1\}^{d \times 2} \to \{0,1\}^m$ with a constant d large enough so that $\log d > 10m$ and $\frac{d}{\ell} - \log \ell > 10 \log d$. These requirements are essentially similar to those made in Section 6.2 and guarantee that $\mathtt{opt}$ can be applied on the output of $\mathtt{s\_ext}_{d/\ell}$.

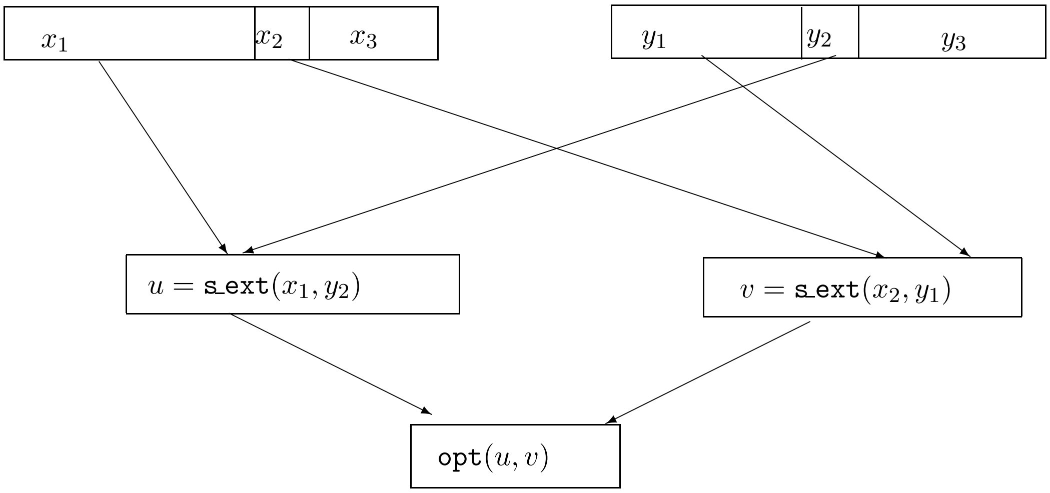

Figure 8: The output $\texttt{disp}_I(x, y)$ with respect to partition I.

The procedure $\texttt{disp}_I(x, y)$. Given a partition I, we define:

$$\texttt{disp}_I(x, y) = \texttt{opt}(\texttt{s\_ext}_{d/\ell}(x_1, y_2), \texttt{s\_ext}_{d/\ell}(y_1, x_2))$$

This procedure (see also Figure 8) is essentially the 4-source extractor of Section 6.2 when applied on x_1, y_2, x_2, y_1. (A technicality is that these inputs have different lengths). Loosely speaking, we will show that in a "correct partition" these 4 sources are "independent enough" so that the construction yields a disperser.

.

Intuition behind this choice. If X is a random variable over $\{0, 1\}^n$, we denote by X_1 the first part of X according to the partition I, and define similarly X_2, X_3, Y_1, Y_2, Y_3. Intuitively, we will later devise the goodness condition to ensure that in a "good" partition both the distributions X_2 and Y_2 will have more than $\delta^2 n \gg \delta^4 n$ entropy in them and that the the distributions X_3 and Y_3 will have essentially zero entropy. Thus, even after conditioning on X_2, X_3, Y_2, Y_3, the distributions X_1 and Y_1 will still have at least $\frac{\delta}{2}n$ entropy in them (since the size of x_2 and y_2 is always at most $\frac{\delta}{10}n$ and hence can take at most $\frac{\delta}{10}n$ entropy). In this case, if we fix X_2 and Y_2 to "typical" values x_2 and y_2 then x_2 will be a good seed, with respect to $\texttt{s\_ext}$, for the distribution $Y_1|Y_2 = y_2$ and y_2 will be a good seed for the distribution $X_1|X_2 = x_2$. Once x_2 and y_2 are fixed these distributions are independent, and hence the outputs of the two somewhere-uniform extractors will both independent and with high entropy. Thus we will get a uniform output when we apply the optimal extractor to them.

The procedure $\texttt{test}_I(x, y)$. Given a partition I, we define $\texttt{test}_I(x, y)$ as follows:

1. Coose k to be large enough constant which we specify later. Let $c_1 = \texttt{s\_ext}_k(x_3, y)$, $c_2 = \texttt{s\_ext}_k(y_3, x)$ and let c be the concatenation of c_1 and c_2. Note that $|c| = 2\ell k$ since the number of output blocks in all somewhere-extractors which we using is ℓ. We refer to c as

the "challenge string" of x, y for partition I. If x_3 is of length zero in the partition I then we let c_1 be some fixed string and similarly we let c_2 be a fixed string if y_3 is of length zero. If I is the full partition (i.e., both x_3 and y_3 are of length zero) then $\mathtt{test}_I(x, y) = 1$ regardless of x and y.

2. For a substring $\hat{x}$ of x which is of length $\delta^3 n$ and whose starting index is an integer multiple of $\delta^3 n$ and a substring $\hat{y}$ of y with the same properties, we let $c_{\hat{x},\hat{y}} = \mathtt{s\_ext}_{2\ell k}(\hat{x}, \hat{y})$. Note that there is at most a constant number of such possible pairs of substrings. We refer to $c_{\hat{x},\hat{y}}$ as a "guess string".

3. If there exist substrings $\tilde{x}, \tilde{y}$ such that c is a sub-block of $c_{\tilde{x},\tilde{y}}$ then we say that the challenge is responded and set $\mathtt{test}_I(x, y) = 1$. Otherwise, $\mathtt{test}_I(x, y) = 0$.

Intuition. X_3 and Y_3 contain very little entropy. Intuitively, if either of X_3 or Y_3 contain a more than $\delta^3 n$ entropy then c should not be predictable using the substrings $\hat{x}$ and $\hat{y}$ which are of length $\delta^3 n$. Note that the requirement that X_3 and Y_3 contain no entropy holds trivially in the case of the full partition (where x_3 and y_3 have length zero). Indeed, in this case we always return 1. However, the partition we will be interested in will be the *maximal* partition under $\prec$ that satisfies the goodness condition, and, as we will show, in this partition X_2 and Y_2 will also have non-negligible entropy (indeed, otherwise they could have been added to X_3 and Y_3).

7.2 The proof

7.2.1 The outline of the proof

We will prove that our disperser works by proving the following claim:

Claim 7.3. *For every X, Y independent random variables over $\{0,1\}^n$ with $H^\infty(X), H^\infty(Y) > \delta n$ there exist independent sub-sources $\tilde{X} \subseteq X$ and $\tilde{Y} \subseteq Y$ and a partition $\tilde{I}$ such that*

1. *$\mathtt{disp}_{\tilde{I}}(\tilde{X}, \tilde{Y})$ is 2^{-2m}-close to uniform.*
2. *The partition $\tilde{I}$ is selected with probability $1 - 2^{-2m}$ on $(x, y) \in_R (\tilde{X}, \tilde{Y})$.*

Theorem 7.1 immediately follows as we have that $\mathtt{disp}_{\tilde{I}}(\tilde{X}, \tilde{Y})$ assigns positive probability to any $z \in \{0,1\}^m$, and this probability can only increase when considering the initial sources X and Y.

While Claim 7.3 suffices to prove Theorem 7.1. We will prove a stronger claim that actually shows that the "good subsources" have large density in the original subsources. The next claim is slightly more technical to state.

Claim 7.4. *For every $\delta > 0$ and integer m there exists a constant $\nu = \nu(\delta, m) > 0$ such that for every independent random variables X, Y over $\{0,1\}^n$ with $H^\infty(X), H^\infty(Y) > \delta n$ there exist independent subsources $\underline{X} \subseteq X$ and $\underline{Y} \subseteq Y$ such that $\underline{X}$ and $\underline{Y}$ have density at least ν in their respective sources, and furthermore $\underline{X}$ and $\underline{Y}$ are both convex combinations, such that for every pair of components $\tilde{X}$ of $\underline{X}$ and $\tilde{Y}$ of $\underline{Y}$ There exists a partition $\tilde{I}$ such that*

1. *$\mathtt{disp}_{\tilde{I}}(\tilde{X}, \tilde{Y})$ is 2^{-2m}-close to uniform.*
2. *The partition $\tilde{I}$ is selected with probability $1 - 2^{-2m}$ on $(x, y) \in_R (\tilde{X}, \tilde{Y})$.*

Note that Claim 7.3 immediately follows from Claim 7.4, by considering some fixed pair of components $\tilde{X}$ and $\tilde{Y}$. We stress that Claim 7.4 gives a stronger notion of a disperser. The definition of dispersers allow that in the output distribution there is one heavy element z while the probability $\{0,1\}^m \setminus \{z\}$ can be very small. In contrast, Claim 7.4 gives that there exists a constant $\vartheta = \vartheta(m, \delta) > 0$ such that for every $z \in \{0,1\}^m$, the probability assigned to z by the output of the disperser is at least ϑ.

7.2.2 Properties of subsources

In Claim 7.4 we explained what we expect to get from the subsources $\tilde{X}, \tilde{Y}$. We now list properties of subsources that will suffice to prove the claim.

1. $\texttt{test}_{\tilde{I}}(\tilde{X}, \tilde{Y})$ accepts with probability one.
2. $H^\infty(\tilde{X}_2) \geq \frac{\delta^2 n}{100}$, $H^\infty(\tilde{Y}_2) \geq \frac{\delta^2 n}{100}$.
3. $H^\infty(\tilde{X}_1) \geq \frac{\delta n}{100}$, $H^\infty(\tilde{Y}_1) \geq \frac{\delta n}{100}$.
4. With probability $1 - 2^{-2m}$, $\texttt{test}_{I'}(\tilde{X}, \tilde{Y})$ rejects all partition I' such that $I' \not\prec \tilde{I}$.

The second item in Claim 7.4 easily follows from the first and last items above. We now show that the first one also follows. This is essentially similar to the 4-source extractor of Section 6.2.

Claim 7.5. *Let $\tilde{X}$ and $\tilde{Y}$, and $\tilde{I}$ be a partition such that the properties above are fulfilled, more precisely, such that*

1. $H^\infty(\tilde{X}_2) \geq \frac{\delta^2 n}{100}$.
2. $H^\infty(\tilde{X}_1) \geq \frac{\delta n}{100}$.
3. $H^\infty(\tilde{Y}_2) \geq \frac{\delta^2 n}{100}$.
4. $H^\infty(\tilde{Y}_1) \geq \frac{\delta n}{100}$.

Then $\texttt{disp}_{\tilde{I}}(\tilde{X}, \tilde{Y})$ *is 2^{-2m}-close to uniform.*

We prove Claim 7.5, in Section 7.2.5. The remainder of the section is devoted to proving the existence of subsources with the properties listed above. In section 7.2.3 we state some technical lemmas we need in the proof. We then give the proof in Section 7.2.4.

7.2.3 Technical preliminaries

A partitioning lemma We use a special notation for partitions that is more suitable for this section. Given a string x of length n, consider the partition of x into t blocks of length n/t and let x_i denote the i-th block, $x_{<i}$ denote the concatenation of the first $i-1$ blocks, $x_{>i}$ denote the concatenation of blocks $i+1, .., n$ and $x_{\geq i}$ denote the concatenation of blocks $i, .., n$.

Lemma 7.6. *Let Z be a distribution over $\{0,1\}^n$ with min-entropy $k = \delta n$, and let $\alpha > 0$ and let t be the number of blocks. There exists a subsource $Z' \subseteq Z$ with density at least $1 - \alpha$ such that Z' is a convex combination of components such that each component X of Z' satisfies that The density of X in Z' is at least $\frac{\alpha}{4t}$ and in addition every X has an index i such that for every $x \in \mathsf{Supp}(x)$:*

- $H^{\infty}(X_i|X_{<i}=x_{<i}) > \frac{\delta n}{4t} - \log t - \log(1/\alpha) - 2$
- $H^{\infty}(X^{>i}|X_{\leq i}=x_{\leq i}) > \frac{3\delta n}{4} - \frac{n}{t} - \log t - \log(1/\alpha) - 2$

Proof. Let $p_i(x) = \Pr[Z_i = x_i|Z_{<i} = x_{<i}]$. We note that for every x, the product of the p_i's is smaller than $2^{-\delta n}$. Therefore, there must exist one i in which $p_i(x) \leq 2^{-\delta n/t}$. Let $v = \delta n/4t$ and let

$$A_i = \{x|i \text{ is the first index such that } p_i(x) \text{ is smaller than } 2^{-v}\}$$

The sets A_i are a partition of the support of Z. We say that A_i is "sufficiently large" if $\Pr[Z \in A_i] \geq \alpha/2t$ and call i a "good index". For a "sufficiently large" A_i we say than an element $x \in A_i$ is "too small" if $p_i(x) < \frac{\alpha}{4t}2^{-(n/t)}$. For every "sufficiently large" set A_i,

$$\Pr[Z \text{ is too small for } A_i] \leq \frac{\alpha}{4t}2^{-(n/t)} \cdot 2^{n/t} \leq \frac{\alpha}{4t}$$

We now define BAD to be the union of all sets A_i that are not sufficiently large as well as all elements x that are "too small". Note that $\Pr[Z \in BAD] \leq t \cdot (\alpha/2t) + t \cdot (\alpha/4t) \leq \alpha$. For the good indices i's we let B_i contain all the elements in A_i that are not too small. It follows that $\Pr[Z \in B_i] \geq \alpha/4t$. Fix some good index i and let X to denote the distribution $Z|B_i$.

An important observation is that X depends only on the first i blocks. More formally, there exist a set A of strings of length in/t such that $x \in B_i$ if and only if $x \in \mathsf{Supp}(Z)$ and $x_{\leq i} \in A$.

For the first item, note that for every $x \in B_i$,

$$\Pr[X_i = x_i|X_{<i} = x_{<i}] = \Pr[Z_i = x_i|Z \in X, Z_{<i} = x_{<i}]$$

$$\leq \frac{\Pr[Z_i = x_i|Z_{<i} = x_{<i}]}{\Pr[Z \in B_i]} \leq p_i(x) \cdot (4t/\alpha) \leq 2^{-(v - \log t - \log(1/\alpha) - 2)}$$

For the second item, note that for every $x \in B_i$

$$\Pr[X_{>i} = x_{>i}|X_{\leq i} = x_{\leq i}] = \Pr[Z_{>i} = x_{>i}|Z \in B_i, Z_{\leq i} = x_{\leq i}]$$

$$\leq \frac{\Pr[Z_{>i} = x_{>i}, Z_{\leq i} = x_{\leq i}]}{\Pr[Z \in B_i, Z_{\leq i} = x_{\leq i}]} = \frac{\Pr[Z = x]}{\Pr[Z \in B_i, Z_{\leq i} = x_{\leq i}]}$$

$$\leq \frac{2^{-\delta n}}{\Pr[Z \in B_i, Z_{\leq i} = x_{\leq i}]}$$

Thus, it is sufficient to show that $\Pr[Z \in B_i, Z_{\leq i} = x_{\leq i}]$ is large. We use the fact that for $x \in B_i$, the event above is equivalent to $\{Z_{\leq i} = x_{\leq i}\}$. This is because whether or not $Z \in B_i$ depends only on the first i blocks of Z. Thus,

$$\Pr[Z \in B_i, Z_{\leq i} = x_{\leq i}] = \Pr[Z_{\leq i} = x_{\leq i}]$$

$$= \Pr[Z_{<i} = x_{<i}]\Pr[Z_i = x_i|Z_{<i} = x_{<i}] > (2^{-v})^{i-1} \cdot \frac{\alpha}{4t}2^{-(n/t)}$$

Thus, overall the conditional min-entropy is at least $\delta n - (i-1)v - (n/t) - \log t - \log(1/\alpha) - 2$. Note that $(i-1)v < tv \leq \delta n/4$ and therefore the conditional min-entropy is at least $\frac{3}{4} \cdot \delta n - (n/t) - \log t - \log(1/\alpha) - 2$. $\square$

We now deduce the following corollary from Lemma 7.6 which is stated using the notation of the previous section.

Corollary 7.7. *Let $\delta > 0$. For large enough n let X be a distribution over $\{0,1\}^n$ with $H^\infty(X) \geq \delta n$. There exists a subsource $\underline{X} \subseteq X$ with density $1/2$ in X such that $\underline{X}$ is a convex combination of distributions such that each component X' of X has an index i_x such that when partitioning $[n]$ to 3 segments according to i_x:*

- *X'_3 is a constant distribution.*
- *$H^\infty(X'_2) \geq \frac{\delta^2 n}{50}$*
- *$H^\infty(X'_1) > \frac{\delta n}{2}$*

Proof. The Corollary follows by using $t = 10/\delta$ in the Lemma above. Each subsource X we consider all values x_3 to X_3. X can be expressed as a convex combination of distributions X' where for each x_3, $X' = X|X_3 = x_3$. Note that these components satisfy the properties of the Lemma. $\square$

We also need the following corollary (that is stated for general t).

Corollary 7.8. *Let $\delta > 0$ and t be some constants. For large enough n let X be a distribution over $\{0,1\}^n$ with min-entropy $k = \delta n$. There exists a subsource $X' \subseteq X$ with density $1/8t$ and an index i such that $H^\infty(X'[i]) \geq \frac{\delta n}{5t}$.*

Subsources and min-entropy We also need the following Lemmas.

Lemma 7.9. *Suppose that $H^\infty(X) \geq k$, and let E be an event such that $\Pr[X \in E] \geq p$ and let X' be the subsource $X|X \in E$. Then, $H^\infty(X') \geq k - \log p$.*

Lemma 7.10. *Let X_1, X_2 be random variables such that $H^\infty(X_1) \geq k$, $|X_2| = r$, then with probability $1 - 2^{-r}$ over choosing $x_2 \in_R X_2$,*

$$H^\infty(X_1|X_2 = x_2) \geq k - 2r$$

7.2.4 Getting the subsources

We will gradually find smaller and smaller subsources until we reach the desired subsource.

Obtaining the partition. We first apply Corollary 7.7 on the sources X. We conclude that there exist a subsource $X^1 \subseteq X$ of density $1/2$. Furthermore, this subsource is a convex combination of subsources, such that for each component X' of X^1 there exists an index i'_x, such that when partitioning $[n]$ to three parts according to i'_x:

- X'_3 is a constant distribution.
- $H^\infty(X'_2) \geq \frac{\delta n}{5t} \geq \frac{\delta^2 n}{50}$
- $H^\infty(X'_1) \geq \frac{\delta n}{2}$

We do the same process on Y to obtain a subsource Y^1 with components Y' and indices i'_y. For every pair of components X', Y' we associate the partition $I = (i'_x, i'_y)$. We now consider a fixed pair of components X', Y' and their associated partition.

Fixing the challenge. Recall that $c_1 = \mathtt{s\_ext}_k(x_3, y)$ and $c_2 = \mathtt{s\_ext}_k(y_3, x)$. Our next goal is to restrict our attention to subsources in which c_1 and c_2 are fixed. We start with c_2. We have that Y_3' is fixed and therefore the value of c_2 given by $\mathtt{s\_ext}(Y_3', X')$ depends only on X'. There is a fixing c_2' such that $\Pr[\mathtt{s\_ext}(Y_3', X') = c_2'] \geq 2^{-|c_2|}$ which is a constant. Let E denote the event in the probability above. We define $X'' \subseteq X'$ by $X'' = X|E$. As the probability of E is large we have that the when restricting our attention from X' to X'' the first and second block lose only a few bits of entropy.

We repeat the same process and fix c_1 to some value c_1' by restricting our attention to a subsource Y'' of Y'. Thus, in X'', Y'' the value of $c = c_1 \circ c_2$ is fixed to some value c'. We have that

- $H^\infty(X_2'') \geq \frac{\delta^2 n}{60}$
- $H^\infty(X_1'') \geq \frac{\delta n}{4}$
- On the subsources X'', Y'' the value of the "challenge string" is fixed.

Finding a short substring $\hat{x}$ with sufficient randomness We now apply Lemma 7.8 on the distribution X'' dividing it to substrings of length $\delta^3 n$. There are $t' \stackrel{def}{=} 1/\delta^3$ such blocks. This distribution has min-entropy at least $\delta n/4$. By Corollary 7.8 we conclude that there exists a subsource X''' of X'' of constant density $(1/8t')$ and an index j such that if we denote the j'th block in X''' by $\widehat{X'''}$ then $H^\infty(\widehat{X'''}) \geq \delta^4 n/20$. As we have restricted X'' to a subsource with constant probability, we have by Lemma 7.9 that the entropy of X_2''', X_3''' only reduced by a constant compared to X_2'', X_3'''.

We repeat the same process for Y'' to generate a subsource Y''' with the same properties. Thus, we have that:

- $H^\infty(X_2''') > \frac{\delta^2 n}{70}$
- $H^\infty(X_1''') \geq \frac{\delta n}{8}$
- In the subsource X''', Y''' the value of the "challenge string" c is fixed.
- There is a substring $\widehat{X'''}$ of X_2''' of length $\delta^3 n$, such that $H^\infty(\widehat{X'''}) \geq \frac{\delta^4 n}{20}$.

Responding to the challenge. Recall that for every two short substrings $\hat{x}, \hat{y}$ of x and y the construction computes a "guess string" $c_{\hat{x},\hat{y}} = \mathtt{s\_ext}_{2\ell k}(\hat{x}, \hat{y})$ and checks whether it "meets the challenge" in the sense that it has the "challenge string" c as a sub-block.

We now further restrict the subsources so that the guess string meets the challenge. We have that both the substrings $\widehat{X'''}$ and $\widehat{Y'''}$ have constant min-entropy rate. It follows, that when applying $\mathtt{s\_ext}_{2\ell k}(\widehat{X'''}, \widehat{Y'''})$ we obtain a distribution $\mathcal{R}$ that is $2^{-\eta n}$-close to a somewhere random distribution. This distribution has blocks of length $2\ell k$ that is a constant. We have that there exists a random variable J over $[\ell]$ (that depends on $\mathcal{R}$) such that $\mathcal{R}_J$ is $2^{-\eta n}$ close to uniform. Thus, $\mathcal{R}_J$ equals the challenge c' with probability at least $2^{-|c'|} - 2^{-\eta n}$ which is larger than some positive constant. Thus, with positive probability the challenge c' is responded. We now consider all possible fixings of $\hat{x}$ of $\widehat{X'''}$ and $\hat{y}$ of $\widehat{Y'''}$ such that

- $\mathtt{s\_ext}_{2\ell k}(\hat{x}, \hat{y})$ contains the challenge c' as a sub-block.
- $\Pr[\widehat{X'''} = \hat{x}] > 2^{-2\delta^3 n}$ and the same requirement for the y's.

We notice that these pairs occur with constant probability under the choice of $\widehat{X'''}, \widehat{Y'''}$. For each such pair we define the subsources $\tilde{X}$ and $\tilde{Y}$ of X''' and Y''' as follows: $\tilde{X} = X'''|\widehat{X'''} = \hat{x}$. We define $\tilde{Y}$ analogously. We define the $\tilde{I}$ to be the partition associated with X' and Y' that we've been using all along. It follows that $\mathtt{test}_{\tilde{I}}(\tilde{X}, \tilde{Y})$ accepts with probability one, as the challenge is always responded. Furthermore, the event $E = \{\widehat{X'''} = \hat{x}\}$ occurs with probability at least $2^{-2\delta^3 n}$. Therefore, by Lemma 7.9, when moving from X''' to $\tilde{X}$, the entropy of the the first and second block decreases by at most $2\delta^3 n$ bits. We thus, meet the requirements on the entropy of the blocks stated in Section 7.2.2. That is we have that:

- $H^\infty(\tilde{X}_2) \geq \frac{\delta^2 n}{100}$.
- $H^\infty(\tilde{X}_1) \geq \frac{\delta n}{100}$.
- $\mathtt{test}_{\tilde{I}}(\tilde{X}, \tilde{Y})$ accepts with probability one.

We now let $\underline{X}$ be the "union" of all sources $\tilde{X}$. More precisely, $\underline{X}$ is the convex combinations of sources $\tilde{X}$ for all choices of $\hat{x}, \hat{y}$ and all choices of components X' of X^1 with the appropriate weights. It follows that $\underline{X}$ has some constant density in X.

Avoiding "incorrect" partitions. Fix some subsources $\tilde{X}, \tilde{Y}$ and the partition associated with them $\tilde{I}$. Let I' be a partition such that $I' \not\leq \tilde{I}$. We want to show that $\mathtt{test}_{I'}(\tilde{X}, \tilde{Y})$ rejects with probability close to 1. Let $\tilde{I} = (\tilde{i}_x, \tilde{i}_y)$ and $I' = (i'_x, i'_y)$. As $I' \not\leq \tilde{I}$ we assume without loss of generality that $i'_x > \tilde{i}_x$ (the proof similar in the case $i'_y > \tilde{i}_y$). We know that $\tilde{X}_2^{\tilde{I}}$ contains randomness. Note that, this block is contained in $\tilde{X}_3^{I'}$, and therefore

$$H^\infty(\tilde{X}_3^{I'}) \geq \delta^2 n/100$$

We now show that for every two substring of length $\delta^3 n$ of $\tilde{X}, \tilde{Y}$ which we will denote by $\widehat{X}$ and $\widehat{Y}$ respectively, the guess $c_{\widehat{X},\widehat{Y}}$ fails to respond to the challenge with high probability.

By Lemma 7.10 applied on the variables $\tilde{X}_3^{I'}, \widehat{X}$, we have that with probability $1 - 2^{-\delta^3 n}$ over choosing $\hat{x} \in_{\mathrm{R}} \widehat{X}$,

$$H^\infty(\tilde{X}_3^{I'}|\widehat{X} = \hat{x}) \geq \delta^2 n/100 - 2\delta^3 n \geq \delta^2 n/200$$

Note that the same holds for the y part. For every such fixed values $\hat{x}, \hat{y}$ we consider the distribution

$$P = (\tilde{X}_3^{I'}, \tilde{Y})|\widehat{X} = \hat{x}, \widehat{Y} = \hat{y}$$

It follows that the two blocks in this distributions are independent and have high min-entropy. Therefore. when applying $\mathtt{s\_ext}_k(P)$ to compute c_1, the output is $2^{-\eta m}$ close to a somewhere random distribution. Nevertheless, in P the "guess string" depends only on the two substrings and is therefore fixed. It follows that the probability that the "guess string" responds to the challenge is at most $3\ell 2^{-k}$ for large enough n. Recall that we are still free to choose the constant k to make this quantity as small as we like. There are a constant number of choices for partitions I', and a constant number of possible choices for substrings. We choose k to be large enough so that by a union bound all of the partitions I' and substrings, all challenges are not responded simultaneously with probability $1 - 2^{-2m}$. This proves the last item in the list on Section 7.2.2.

7.2.5 Proof of Claim 7.5.

Recall that $\mathtt{disp}_I(x,y) = \mathtt{opt}(\mathtt{s\_ext}_{d/\ell}(x_1,y_2), \mathtt{s\_ext}_{d/\ell}(y_1,x_2))$. Our goal is to show that opt is applied on two independent distributions with sufficient entropy requirement. This will be achieved for most fixings of $\tilde{X}_2$ and $\tilde{Y}_2$.

We first apply Lemma 7.10 on the sources $\tilde{X}_1, \tilde{X}_2$, we conclude that with probability $1-2^{-\delta^2 n/100}$ over choosing $x_2 \in_{\mathrm{R}} \tilde{X}_2$,

$$H^{\infty}(\tilde{X}_1|\tilde{X}_2 = x_2) \geq \delta n/100 - 2\delta^2 n/100 > \delta n/200$$

We say that x_2 is "useful" if the equation above holds for x_2. We define a the set of "useful" y_2's in an analogous way.

Recall that we apply a strong somewhere random extractor on x_1, y_2. We say that a useful y_2 is "good" with respect to x_2 if y_2 is a good seed for the distribution $X_1|X_2 = x_2$. For every value of x_2 the distribution has a constant rate of randomness. We conclude that for every x_2,

$$\Pr[Y_2 \text{ is good with respect to } x_2] \geq 1 - 2^{-\eta n}$$

which implies that

$$\Pr[Y_2 \text{ is good with respect to } X_2] \geq 1 - 2^{-\eta n}$$

We deduce analogously that

$$\Pr[X_2 \text{ is good with respect to } Y_2] \geq 1 - 2^{-\eta n}$$

Thus, with probability $1 - 2 \cdot 2^{\eta n}$ over choosing $(x_2, y_2) \in_{\mathrm{R}} (\tilde{X}_2, \tilde{Y}_2)$ the pair that is chosen is a "good match", meaning that each one is good with respect to the other.

For such a fixed pair (x_2, y_2) we define $\tilde{X'}_1 = \tilde{X}_1|\tilde{X}_2 = x_2$ and $\tilde{Y'}_1 = \tilde{Y}_1|\tilde{Y}_2 = y_2$. In words, these are the subsources of $\tilde{X}_1$ and $\tilde{Y}_1$ relative to fixed x_2, y_2.

We have seen that $\tilde{X'}_1$ has a constant rate of entropy. As y_2 is a good seed for $\tilde{X'}_1$ we have that $\mathtt{s\_ext}_{d/\ell}(\tilde{X'}_1, y_2)$ is a somewhere random distribution with ℓ blocks of length d/ℓ. Recall that such a distribution has min-entropy $d/\ell - \log \ell$ and we required that this quantity is larger than $10 \log d$. The same reasoning gives that $\mathtt{s\_ext}_{d/\ell}(\tilde{Y'}_1, x_2)$ has entropy larger than $10 \log d$. Thus, these distributions meet the entropy requirements of opt. We conclude that when we apply opt on these distributions we obtain a distribution that is $1/d$-close to uniform. Recall that we have required that $\log d \geq 10m$ which gives $1/d \leq 2^{-10m}$.

As the output is close to random for almost all fixings x_2, y_2 we conclude that $\mathtt{disp}_{\tilde{I}}(\tilde{X}, \tilde{Y})$ is within distance $2^{-10m} + 2 \cdot 2^{-\eta n} \leq 2^{-2m}$ from uniform as required.

8 Somewhere extractor for affine sources

In this section we describe a somewhere extractor that given an affine n-bit source X with entropy rate δ, produces a constant (depending on δ) number of blocks of linear length in n, such that one of the blocks is exponentially close to being uniform.

Definition 8.1 (Affine somewhere extractors). A function $\mathtt{as\_ext} : \{0,1\}^n \to \{0,1\}^{m \times \ell}$ is called an ℓ-outputs *affine somewhere extractor* with rate requirement δ and error ϵ, if for every affine n-bit source X with entropy rate at least δ, one of the output blocks of $\mathtt{as\_ext}(X)$ is ϵ-close to being uniformly distributed over $\{0,1\}^m$.

Note that the above definition requires the output of an affine somewhere extractor to satisfy a stronger condition than simply being ϵ-close to somewhere uniform, since the uniform output block is fixed. We can generate affine somewhere extractor because of the nice properties of affine subspaces.

Theorem 8.2. *There exist positive constants $c_1 > 1$ and c_2, such that for any constant parameter $\delta > 0$ the following holds. Taking*

$$\ell = \ell(\delta) = \exp\left(O\left(\frac{\ln(1/\delta)}{\delta^{c_2+1}}\right)\right) ,$$

and for every m satisfying

$$m \leq m(\delta) = n \cdot \exp\left(-O\left(\frac{\ln(1/\delta)}{\delta^{c_2+1}}\right)\right) ,$$

there exists a polynomial-time computable ℓ-outputs affine somewhere extractor $\mathtt{as\_ext}_{(m,\delta)}$ with rate requirement δ, which outputs blocks of length m, and whose error is at most $2^{-c_1 m}$.

Overview of the construction. Let X be an affine source with entropy rate δ, and let $S_1, \ldots, S_{c_1 \cdot (1/\delta)^{c_2}}$ be an arbitrary partition of its coordinates into sets of equal size. The main idea behind the construction of $\mathtt{as\_ext}_{(m,\delta)}$ (much like in the condenser presented in Section 4), is to consider the cases where entropy is particularly dense within one of the blocks, and the case where it is "evenly spread" among all blocks.

- **Almost independent blocks.** If the entropy rate in each block X_{S_j} is exactly δ for all j's, then X is actually a concatenation of independent sources. We prove that this is roughly the case even when $\max_j \{\mathsf{r}(X_{S_j})\}$ is slightly above δ, namely if it is bounded by $\delta + \alpha(\delta)$ for some function α. In this case we can use the [BIW] extractor to get the entropy out of X. One of the blocks in the output of $\mathtt{as\_ext}_{(m,\delta)}$), will therefore be obtained by applying the [BIW] extractor to the sub-blocks in X.
- **A concentrated block.** If X does not fall in the previous category, then there must be some S_j block where the entropy rate is larger than $\delta + \alpha(\delta)$. Then it makes sense to apply $\mathtt{as\_ext}_{(m,\delta+\alpha(\delta))}$ recursively to that block. Since we do not know in advance which is the "dense" block, we concatenate the output generated in case 1 with the outputs of $\mathtt{as\_ext}_{(m,\delta+\alpha(\delta))}$, applied to each of the blocks S_j in X. This recursive construction ends when $\delta + \alpha(\delta) > 1$, in which case X must be in case 1.

Let us now give the formal proof of Theorem 8.2.

8.1 Construction of a somewhere extractor for an affine source

We start with formal construction of $\mathtt{as\_ext}_{(m,\delta)}$. The output $\mathtt{as\_ext}_{(m,\delta)}(X)$ consists of a concatenation of ℓ blocks of length m, that we will call *output blocks* to distinguish them from sub-blocks of X. Set

$$\alpha(\delta) = \frac{\delta^{c_2+1}}{4c_1c_2} , \tag{8}$$

and take $k \stackrel{def}{=} 2K(\delta)$, where c_1, c_2 and $K(\delta)$ are as in Theorem 3.7.

The initial partition. We partition X into sub-blocks, by taking $S_1, \ldots, S_k$ to be an arbitrary partition of $[n]$ (namely the set of coordinates of X) into k subsets of equal size.

The output blocks. The first output block in $\texttt{as\_ext}_{(m,\delta)}$ is obtained by applying the [BIW]-extractor to $X_{S_1}, \ldots, X_{S_k}$ with entropy rate parameter $\delta - k\alpha(\delta)$ (we prove in Claim 8.4 that there are enough blocks to apply the extractor for the designated rate). The obtained block may contain more than m bits (but not less than m bits, by Claim 8.5 below), in which case we take only the first m bits of the output.

Now if $\delta + \alpha(\delta) > 1$, then we are done constructing $\texttt{as\_ext}_{(m,\delta)}$. Otherwise, the rest of the blocks are obtained by applying $\texttt{as\_ext}_{(m,\delta+\alpha(\delta))}$ recursively to each of the sub-blocks $X_{S_1}, \ldots, X_{S_k}$. The construction of $\texttt{as\_ext}_{m,\delta}$ therefore forms a tree of sub-partitions and output blocks, where the root contains the initial partition and the first output block, and each offspring contains a sub-partition of one of the blocks in the parent partition, and the output block obtained by applying the [BIW]-extractor to it.

8.2 Correctness of the Construction

To prove that the construction above is correct, we first show that it is well defined, namely that the recursion ends, that the [BIW]-extractor is always applied with an appropriate number of blocks for the designated rate parameters, and that the length of the block generated by each application of the [BIW]-extractor is at least m. We then verify that the number of generated blocks is as stated, and finally conclude by showing that one of the output blocks is close to random.

Let us start by showing that the recursion completes at some point.

Claim 8.3. *The depth of the recursion is bounded by* $4c_1c_2 \cdot \delta^{-(c_2+1)}$.

Proof. Note that $\alpha(\delta)$ is an increasing function in δ, and that δ increases by $\alpha(\delta)$ each time the depth of the recursion increases. Since the recursion ends when $\delta + \alpha(\delta)$ reaches 1, it follows that the depth of the recursion is bounded by

$$\frac{1-\delta}{\alpha(\delta)} \leq \frac{4c_1c_2}{\delta^{c_2+1}} \ .$$

□

The next claim shows that we partition the source into sufficiently many blocks for the application of the [BIW] extractor, with respect to the specified rate. Due to the recursive nature of the construction, it is enough to verify that for the first application of the extractor. Since the rate parameter used in this application is $\delta - k\alpha(\delta)$, it is enough to prove the following.

Claim 8.4. *Let c_1, c_2, and K be as in Theorem 3.7, and let $k = 2K(\delta)$ be the number of sub-blocks in the initial partition of X. Then*

$$k \geq K(\delta - k\alpha(\delta)) \ .$$

Proof. Note that

$$\delta - k\alpha(\delta) = \delta - 2 \cdot c_1 \cdot (1/\delta)^{c_2} \cdot \frac{\delta^{c_2+1}}{4c_1c_2} = \delta \cdot \left(1 - \frac{1}{2c_2}\right)$$

Hence

$$K(\delta - k\alpha(\delta)) = K(\delta) \cdot \left(1 - \frac{1}{2c_2}\right)^{-c_2} \leq K(\delta) \cdot (\tfrac{1}{2})^{-1} = k \ ,$$

where the last inequality follows from the easy-to-verify inequality

$$\forall\,(0 < t \leq 1)\,,(c > 1)\,, \qquad (1-t)^c \geq 1 - ct$$

□

The next claim shows that we indeed get at least m bits in each block.

Claim 8.5. *During the construction of* $\mathtt{as\_ext}_{(m,\delta)}$*, each application of the [BIW]-extractor generates at least* m *bits.*

Proof. Since the number of bits generated by the [BIW]-extractor is the same as the size of the blocks to which it is applied, we need to verify that the block sizes are always larger than m. Indeed, we initially partition X into $k = 2c_1 \cdot \delta^{-c_2}$ blocks, and since δ increases with the depth of the recursion, at each level of the recursion we further break the blocks into no more than k sub-blocks. Since the depth of the recursion is bounded by $4c_1c_2 \cdot \delta^{-(c_2+1)}$ (according to the above claim), we have that the size of the smallest block ever generated is

$$n/k^{O(\delta^{-(c_2+1)})}\ .$$

Substituting the value of k yields that the size of all blocks that are generated is bigger than m, which implies the claim. □

It is now left to prove the main property of $\mathtt{as\_ext}_{(m,\delta)}$, namely that one of the blocks in it is uniform. As mentioned in the overview of the construction, we will distinguish between two cases.

Two partition types. Let $T_1, \ldots, T_{k(\delta')}$ be a sub-partition obtained at some node of the tree-of-partitions, where the corresponding output block is obtained by applying the [BIW]-extractor to it with parameter δ'. We call the partition *well-balanced*, if for all i the entropy rate of X_{S_i} is at most $\delta' + \alpha(\delta')$. If the partition is not well balanced then there must exist a block X_{S_i} with entropy rate more than $\delta' + \alpha(\delta')$. This is called a *concentrated block*.

Let us first deal with the case where one of the sub-partitions is well balanced, and then show that there must exist a well-balanced partition.

Lemma 8.6. *If the any of the sub-partitions of* X *applied during the construction is well balanced, then its corresponding output-block is of distance at most* $2^{-\Omega(n)}$ *from uniform.*

Proof. Let $T_1 \ldots, T_{k(\delta')}$ be a well-balanced sub-partition of some block X_T (where δ' is the parameter with which the [BIW]-extractor is applied for that sub-partition). Using Lemma 3.18 we have that X_T is a convex combination of sources X', for which the sub-blocks $X'_{S_1}, \ldots X'_{S_k}$ are independent and have entropy rate at least $\delta' - k(\delta') \cdot \alpha(\delta')$ each. Together with Claim 8.4 and Theorem 3.7, this completes the proof. □

To show the correctness of the construction, it now only remains to show that one of the sub-partitions is well-balanced.

Lemma 8.7. *At least one of the sub-partitions applied by* $\mathtt{as\_ext}_{(m,\delta)}$ *is well-balanced.*

Proof. We prove the lemma by induction on the tree of sub-partitions. If the initial partition is well-balanced then we are obviously done. If it is not, then X has entropy rate at least $\delta_1 \stackrel{def}{=} \delta+\alpha(\delta)$ on one of its sub-blocks. Denote the coordinates of this sub-block by T_1. Then by the definition of the construction, at some point it applies $\texttt{as\_ext}_{(m,\delta_1)}$ to a sub-partition of X_{T_1}, so if this partition is well-balanced, we are done again. If it is not well-balanced, we get a subset $T_2 \subseteq T_1$ where of X_{T_2} has entropy rate at least $\delta_2 \stackrel{def}{=} \delta_1 + \alpha(\delta_1)$, to which the construction eventually applies $\texttt{as\_ext}_{(m,\delta_2)}$ ...

Continually following the concentrated blocks in this manner, if we eventually encounter a well-balanced partition then we are done. Indeed, this must be the case: Claim 8.3 implies that at some depth t of the tree of sub-partitions, the construction applies $\texttt{as\_ext}_{(m,\delta_t)}$ to sub-blocks of X, where $\delta_t + \alpha(\delta_t) > 1$. If at that point we do not already have a balanced partition, then we must have a concentrated sub-block X_{T_t} of X whose entropy rate is at least δ_t. The partition of X_t *must* be well-balanced, since obviously the entropy rate of each of its sub-blocks is bounded by 1 (and therefore by $\delta_t + \alpha(\delta_t)$). □

9 Affine-source dispersers

In this section we show a zero error disperser for affine sources whose entropy rate is δ, where δ is any positive constant. Our disperser is strong in the sense that it obtains every output with at least some constant probability.

Theorem 9.1 (Affine-source disperser). *Let $\delta > 0$ be some constant, and m_0 be a fixed integer. There exists a polynomial-time computable function* $\texttt{a-disp} : \{0,1\}^n \to \{0,1\}^{m_0}$ *such that for every affine n-bit source X with $\mathsf{r}(X) \geq \delta$,* $\texttt{a-disp}(\mathsf{Supp}(X)) = \{0,1\}^{m_0}$. *Moreover, there exists a constant $d = d(m,\delta)$, such that for every $z \in \{0,1\}^m$,*

$$\Pr_{x_1 \in_R X_1, x_2 \in_R X_2}[\texttt{disp}(x_1,x_2) = z] \geq d(m,\delta) \ .$$

In the following, we define the affine-source disperser, but we do not prove its strongness property. This will be proven in the final version of our paper.

9.1 The disperser

Let X be an n-bit affine source with entropy rate δ. The idea of the affine disperser is to set a partition $\mathcal{S} = \{S_1, \ldots, S_5\}$ of the coordinates of X into five consecutive blocks, and then to find an affine sub-source X' of X, where X'_{S_2} and X'_{S_4} are both independent, and have sufficient length and entropy rate to apply the two-source disperser to them. Of course, we will have to choose among many possible block-partitions of X, and we will use the challenge mechanism to do that. Note that while X is an affine source, X' need not necessarily be one.

Candidate partitions. The partitions we consider of coordinates of X come from the following family. By a *segment*, we refer to a set of consecutive coordinates $S \subseteq [n]$. A *candidate partition* $\mathcal{S}$ of X is a partition of $[n]$ into 5 consecutive segments $S_1, \ldots, S_5$ (namely if $i > j$ and $a \in S_i$ and $b \in S_j$ then $a > b$), with the following properties.

- $|S_1|$ is a multiple of $\delta n/4$.
- $|S_2| = \delta n/4$.

- $|S_3|$ is a multiple of $\delta^2 n/32$.
- $|S_4| = \delta^2 n/32$.
- S_5 is the remaining set of coordinates.

Ordering partitions. Given a string x, our disperser will choose one of the candidate partitions by applying some sort of test to each partition, declaring a partition valid if it passed the test. It then picks, among the valid partitions, the largest partition with respect to the following order. We say that one candidate partition $\mathcal{S} = \{S_1, \ldots, S_5\}$ is larger than another, $\mathcal{T} = \{T_1, \ldots, T_5\}$, if either $|S_1| > |T_1|$, or if $|S_1| = |T_1|$ and $|S_3| > |T_3|$.

9.1.1 The construction of `a-disp`.

Let $\mathrm{SRL}_{(m,\delta)}$ denote the affine somewhere extractor as in the previous section, and let $\ell(\delta)$ denote the number of output blocks that are generated by the affine somewhere extractor, for rate requirement δ. Also, let $c(m,\delta)$ be as stated in Theorem 7.1, and set

$$m_2 \stackrel{def}{=} \log_2\left(\frac{4000 \cdot \ell(\delta/2)}{\delta^{13} \cdot c(m_0, \delta/20)}\right) , \quad \text{and} \quad m_1 \stackrel{def}{=} m_2 \cdot \ell(\delta^3/100) + \log_2\left(\frac{4000 \cdot \ell(\delta/8)}{\delta^{13} \cdot c(m_0, \delta/20)}\right) .$$

The condenser tests each partition $\mathcal{S}$, by first computing the following:

- The first so called *challenge*,
$$\mathtt{ch}_1(\mathcal{S}, X) \stackrel{def}{=} \mathrm{SRL}_{(m_1, \delta^2/50)}(X_{S_1}) ,$$
- We partition S_2 into small consecutive segments of length $\delta^{10} n$, and for each such segment $S_{2,i}$ we compute a *guess*
$$u_i(\mathcal{S}, X) \stackrel{def}{=} \mathrm{SRL}_{(\ell(\delta^2/50)m_1, \delta/8)}(X_{S_{2,i}}) .$$
- The second challenge,
$$\mathtt{ch}_2(\mathcal{S}, X) = \mathrm{SRL}_{(m_2, \delta^3/100)}(X_{S_3}) ,$$
- We partition S_4 into small parts of length $\delta^{10} n$, and for each such part $S_{4,j}$ we compute the second guess
$$v_j(\mathcal{S}, X) = \mathrm{SRL}_{(\ell(\delta^3/100)m_2, \delta/2)}(X_{S_{4,j}}) .$$

Valid partitions. For a given value x of X, we say a candidate partition $\mathcal{S}$ is *valid* with respect to x (namely it "passes the test"), if there exists an i such that $\mathtt{ch}_1(\mathcal{S}, x)$ appears as one of the output blocks in $u_i(\mathcal{S}, x)$ (in this case we say that u_i guesses $\mathtt{ch}_1$), and if there exists a j such that $\mathtt{ch}_2(\mathcal{S}, x)$ is one of the output blocks in $v_j(\mathcal{S}, x)$.

The output of `a-disp`. To obtain the final output of the disperser we let $\mathcal{S} = \mathcal{S}(x) = \{S_1, \ldots, S_5\}$ denote the largest valid candidate partition (according to the order defined above). If there is no valid candidate partition we choose an arbitrary one. The final output is taken to be

$$\mathtt{a\text{-}disp}(X) = \mathtt{disp}(x_{S_2}, (x_{S_4}, \bar{0})) , \tag{9}$$

where `disp` is the 2-source disperser for rate parameter $\delta^2/20$ and with m_0 output size, and where $(x_{S_4}, \bar{0})$ denotes the padding of x_{S_4} by zeros, so that it has the same length as x_{S_2}.

Some intuition. For a $\mathcal{S} = \{S_1, \ldots, S_5\}$, we consider the fact that the challenge $\texttt{ch}_1$ is guessed as an evidence that the S_2-block of the source X has enough min-entropy, and if $\texttt{ch}_2$ is guessed, it is viewed as indicating that there is some min-entropy in the S_4-block of X as well. In the remaining part of this section we show that this intuition is correct, at least for some sub-source $X' \subseteq X$.

Particularly, we show a partition $\mathcal{S}$, and a sub-source X' of X, such that X'_{S_2} and X'_{S_4} are independent and have high-enough entropy rate, and such that $\mathcal{S}$ is always valid with respect to X'. We will also show that for this sub-source X', the probability that there exists a valid partition larger than $\mathcal{S}$ is so small that we can think of $\mathcal{S}$ as always being the maximal valid partition with respect to X'. In that case the disperser $\texttt{disp}$ will be always applied to the S_2 and S_4 blocks of X', whose entropy rate is high enough for $\texttt{disp}$ to produce an output with full support.

9.1.2 The dream sub-source

Following the intuition, we start by defining the properties of our dream sub-source X' and the according good partition.

Good partition for a sub-source. We say that a candidate partition $\mathcal{S} = \{S_1, \ldots, S_5\}$ is good with respect to a sub-source X' of X, if $\mathcal{S}$ is valid with respect to all possible values of X', and in addition X' satisfies the following:

1. X'_{S_1} is constant.
2. $H^\infty(X'_{S_2}) \geq \delta^2 n/40$.
3. $H^\infty(X'_{S_3}|X'_{S_2}) = 0$, (and therefore the S_3-block of X' is a function of its S_2-block).
4. X'_{S_2} and X'_{S_4} are independent.
5. $H^\infty(X'_{S_4}) \geq \delta^3 n/80$.
6. The probability in X' that $\mathcal{S}$ is not the maximal valid partition is bounded by $c(m_0, \delta^2/20)/2$.

We will show that there exists a sub-source X' of X and a candidate partition which is good with respect to X'. This will suffice to obtain Theorem 9.1.

Lemma 9.2. *There exists a sub-source X' of X and a candidate partition $\mathcal{S}$ which is good with respect to X'.*

Let us show why the Lemma 9.2 implies Theorem 9.1 before we prove it.

Claim 9.3. *Lemma 9.2 implies Theorem 9.1.*

Proof. We need to show that $\mathsf{Supp}(\texttt{a-disp}(X)) = \{0,1\}^{m_0}$. Let X' and $\mathcal{S}$ be the sub-source of X and the good partition obtained from Lemma 9.2 respectively. By property 2 of a good partition, we have that $\mathsf{r}(X'_{S_2}) \geq \delta/10 \gg \delta^2/20$. Using property 5 we have that $\mathsf{r}((X'_{S_4}, \bar{0})) \geq \delta^2/20$ as well, where $(X'_{S_4}, \bar{0})$ is the padding of X'_{S_4} by zeros so it becomes the same size as X'_{S_2}, namely $\delta n/4$. In addition, property 4 implies that those two blocks are independent.

We can therefore apply Theorem 7.1 to the two blocks, obtaining that for every $z \in \{0,1\}^{m_0}$,

$$\Pr\left[\texttt{disp}(X'_{S_2},\, (X'_{S_4}, \bar{0})\,) = z\right] \geq c(m_0, \delta^2/20)\ ,$$

where $\texttt{disp}$ is the disperser as in (9). It follows, however, from property 6 that

$$\Pr\left[\texttt{a-disp}(X') \neq \texttt{disp}(X'_{S_2},\, (X'_{S_4}, \bar{0})\,)\right] \leq c(m_0, \delta^2/20)/2\ ,$$

and therefore $\texttt{a-disp}(X')$ obtains every output $z \in \{0,1\}^{m_0}$ with probability at least $c(m_0, \delta^2/16)/2$. This completes the proof since obviously $\mathsf{Supp}(X) \supseteq \mathsf{Supp}(X')$. $\square$

9.1.3 Proof of Lemma 9.2

In the following, we find the good partition hand in hand with the sub-source X' of X. We gradually restrict X into X', taking affine sub-sources of X at first, and on the last step restricting to a non-affine sub-source. Note that we use, at different points of the proof, several lemmas from Section 3 concerning properties of linear sources. We use notation of the form R_i to denote parameters with which the lemmas are applied, and thus the meaning of the sets R_i may change from one paragraph to another.

Determining S_1 and S_2. We use Lemma 3.19 to choose S_1 and S_2. We would like to choose the block S_2 so that X has entropy on it conditioned on its S_1-block. For this purpose, we partition $[n]$ into $t = 4/\delta$ consecutive segments $R_1 \ldots, R_t$, of length $\delta n/4$ each, and set $R_0 = \emptyset$. Applying Lemma 3.19 with parameter $\alpha = 1/4$ we obtain that there exists an i such that

$$H^\infty(X_{R_i}|X_{(\cup_{j<i}R_j)}) \geq \delta^2 n/16 \ ,$$

$$\text{and moreover,} \qquad H^\infty\left(X_{(\cup_{j>i}R_j)}|X_{(\cup_{j\leq i}R_j)}\right) \geq 3\delta n/4 - \delta n/4 = \delta n/2 \ .$$

We now set $S_1 \stackrel{def}{=} \cup_{j<i}R_j$, and $S_2 = R_i$. Hence

$$H^\infty(X_{S_2}|X_{S_1}) \geq \delta^2 n/16 \ , \tag{10}$$

$$\text{and} \qquad H^\infty\left(X_{[n]\setminus(S_1\cup S_2)}|X_{(S_1\cup S_2)}\right) \geq \delta n/2 \ . \tag{11}$$

Fixing X_{S_1}. We would now like to use Lemma 3.20 to find a sub-source $X^{(1)}$ of X such that $X^{(1)}_{S_1}$ is constant, and that Equations (10) and (11) still hold for X'. For this purpose we set $R_1 \stackrel{def}{=} \emptyset$, $R_2 \stackrel{def}{=} S_1$, and $R_3 \stackrel{def}{=} [n] \setminus S_1$, and let $X^{(1)}$ be the sub-source obtained by applying Lemma 3.20. Hence $X^{(1)}_{S_1}$ is indeed constant, and it follows from the remark on section 3 of Lemma 3.20, that the bounds in (10) and (11) hold also for $X^{(1)}$.

Choosing S_3, S_4 and S_5. We use Lemma 3.19 again to choose S_3 and S_4, such that the S_4-block of $X^{(1)}$ has entropy conditioned on its $S_1 \cup S_2 \cup S_3$-block. We will later restrict the source so that its S_4-block and its S_2-block become independent. We denote $R_0 \stackrel{def}{=} S_1 \cup S_2$, and partition $[n] \setminus (S_1 \cup S_2)$ into consecutive segments of length $\delta^2 n/32$ each, $R_1, \ldots, R_t$. Applying Lemma 3.19 with parameter $\alpha = 1$ we have that for some $i > 0$,

$$H^\infty(X^{(1)}_{R_i}|X^{(1)}_{(\cup_{j<i}R_j)}) \geq \delta^3 n/64 \ .$$

We set $S_3 \stackrel{def}{=} R_1 \cup \ldots \cup R_{i-1}$, and $S_4 \stackrel{def}{=} R_i$, and thus obtain

$$H^\infty(X^{(1)}_{S_4}|X^{(1)}_{S_1\cup S_2\cup S_3}) \geq \delta^3 n/64 \tag{12}$$

Taking $S_5 \stackrel{def}{=} [n] \setminus (S_1 \cup \ldots \cup S_4)$, we now have the complete partition $\mathcal{S} = (S_1, \ldots, S_5)$.

Restricting $X^{(1)}$. We now restrict $X^{(1)}$ using Lemma 3.20, to get a source for which property 3 of good partitions holds. We set the parameters by taking $R_1 \stackrel{def}{=} S_1 \cup S_2$, $R_2 \stackrel{def}{=} S_3$, and $R_3 \stackrel{def}{=} S_4$, and let $X^{(2)} \subseteq X^{(1)}$ be the sub-source obtained from Lemma 3.20 for these parameters. $X^{(2)}$ now satisfies

$$H^\infty(X^{(2)}_{S_3}|X^{(2)}_{S_1\cup S_2}) = H^\infty(X^{(2)}_{S_3}|X^{(2)}_{S_2}) = 0 \ ,$$

and it also satisfies the entropy bounds stated for $X^{(1)}$, namely the S_1-block of $X^{(2)}$ is constant, the min-entropy of its S_2-block is at least $\delta^2 n/16$, and (12) holds for it. Since the S_3-block of $X^{(2)}$ is now determined by the value of it S_2-block we obtain from the latter statement that

$$H^\infty(X^{(2)}_{S_4} | X^{(2)}_{S_2}) \geq \delta^3 n/64 \tag{13}$$

Making the blocks independent. We now further restrict $X^{(2)}$ to make its S_2-block and its S_4-block independent. We set $R_1 \stackrel{def}{=} S_2$ and $R_2 \stackrel{def}{=} S_4$, and let $X^{(3)}$ be the sub-source of $X^{(2)}$ obtained by applying Lemma 3.21 to $X^{(2)}$ with these parameters. We thus have that $X^{(3)}_{S_2}$ and $X^{(3)}_{S_4}$ are independent, that

$$H^\infty(X^{(3)}_{S_2}) \geq H^\infty(X^{(2)}_{S_2}) - |S_4| \geq \delta^2 n/16 - \delta^2 n/32 = \delta^2 n/32 \ , \tag{14}$$

and that $H^\infty(X^{(3)}_{S_4}) \geq \delta^3 n/64$.

Finding a good u_i. Since the S_1-block of $X^{(3)}$ is constant, then so is $\mathtt{ch}_1(\mathcal{S}, X^{(3)})$. We would like to slightly restrict $X^{(3)}$ to a sub-source $X^{(4)}$ such that for some fixed i, $u_i(\mathcal{S}, X^{(4)})$ always guesses it. We first find, using Lemma 3.19, an i for which the entropy rate of $X^{(3)}_{S,i}$ is high-enough for u_i to be able to guess $\mathtt{ch}_1$.

We set $R_0 \stackrel{def}{=} \emptyset$, and define $R_1, \ldots, R_t$ by $R_i \stackrel{def}{=} S_{2,i}$, where $S_{2,i}$ is as in the definition of u_i. Then by (14) we have

$$\mathsf{r}(X^{(3)}_{S_2}) = \mathsf{r}(X^{(3)}_{(\cup_{j \geq 1} R_j)} | X^{(3)}_{R_0}) \geq \delta/8 \ ,$$

and applying Lemma 3.19 to the source $X^{(3)}_{S_2}$ with these R_i parameters and with $\alpha = 1$, we obtain an index i for which

$$\mathsf{r}(X^{(3)}_{S_{2,i}}) = \mathsf{r}(X^{(3)}_{R_i}) \geq \mathsf{r}(X^{(3)}_{R_i} | X^{(3)}_{(\cup_{j<i} R_j)}) \geq \delta/8$$

Guessing $\mathtt{ch}_1$. The rate of $X^{(3)}_{S_{2,i}}$ satisfies the requirement of the affine somewhere extractor applied by u_i, and therefore one of the output blocks of $u_i(\mathcal{S}, X^{(3)})$ has full support, namely obtains every possible value in $\{0,1\}^{\ell(\delta^2/50)m_1}$. This implies that there exists at least one value z in the support of $X^{(3)}_{S_{2,i}}$ for which u_i guesses $\mathtt{ch}_1$. We take $X^{(4)} \subseteq X^{(3)}$ to be the sub-source where the value of the $S_{2,i}$-block is fixed to be z, namely $X^{(4)} \stackrel{def}{=} (X^{(3)} | X^{(3)}_{S_{2,i}} = z)$.

Since $H^\infty(X^{(3)}_{S_{2,i}}) \leq |S_{2,i}| = \delta^{10} n$, we have from Corollary 3.13 that the min-entropy of the S_2-block in $X^{(4)}$ is smaller by at most $\delta^{10} n$ from the min-entropy of the same block in $X^{(3)}$. We can therefore write $H^\infty(X^{(4)}_{S_2}) \geq \delta^2 n/35$. Also, the independence between the S_2-block and the S_4-block is not affected by the restriction, nor is the min-entropy of the S_4-block.

Fixing a v_j guess. We now restrict $X^{(4)}$ further, in order to make sure that $\mathtt{ch}_2$ is always guessed. This step is somewhat trickier than what we did for $\mathtt{ch}_1$ because $\mathtt{ch}_2(\mathcal{S}, X^{(4)})$ is not a fixed string. We therefore find for $\mathtt{ch}_2(\mathcal{S}, X^{(4)})$ a string w, which it produces with probability at least $2^{-\ell(\delta^3/100)m_2}$ (such a string must exist, because $\mathtt{ch}_2$ produces strings of length $\ell(\delta^3/100)m_2$). We now find an index j such that the $S_{4,j}$-block in $X^{(4)}$ has entropy rate at least $\delta/2$ – this is done similarly to how we found a good block $S_{2,i}$, hence we omit the details.

Now since $X^{(4)}_{S_{4,j}}$ satisfies the rate requirement of the affine somewhere extractor applied by v_j, there must be an element $z \in \mathsf{Supp}(X^{(4)}_{S_{4,j}})$ for which one of the output strings of v_j is w. We let $X^{(5)} \stackrel{def}{=} (X^{(4)} | X^{(4)}_{S_{4,j}} = z)$, and use Corollary 3.13 to obtain that the min-entropy of the S_4-block of

$X^{(5)}$ is still greater than, say, $\delta^3 n/80$. Note that the min-entropy of the S_2-block is not affected by the restriction.

Fixing $\mathtt{ch}_2$. Our next step is a restriction of $X^{(5)}$ which fixes the value of $\mathtt{ch}_2$ to the string w, so that v_j always guesses it (note that the restriction to $X^{(5)}$ does not change the probability of getting w as the output of $\mathtt{ch}_2$, since it was obtained by restricting the S_4-block, which is independent of the S_3-block). The next step is the only restriction which may produce a non-affine sub-source, and is obtained by taking $X' \stackrel{def}{=} (X^{(5)}|\mathtt{ch}_2(\mathcal{S}, X^{(5)}) = w)$.

The min-entropy of the S_4-block is not affected by this restriction. Moreover, since this is a restriction of $X^{(5)}$ to an event of probability at least $2^{-\ell(\delta^3/100)m_2}$, the min-entropy of its S_2-block is reduced by at most $\ell(\delta^3/100)m_2$, and therefore $H^\infty(X'_{S_2}) \geq \delta^2 n/40$.

9.1.4 $\mathcal{S}$ is good

Now that we have a sub-source X' and the partition $\mathcal{S}$, let us verify that $\mathcal{S}$ is good with respect to X', going over the definition of good partitions and verifying each of the required properties.

It is clear that $\mathcal{S}$ is always valid, since we made sure that $u_i(\mathcal{S}, X')$ guesses $\mathtt{ch}_1(\mathcal{S}, X')$ and that $v_j(\mathcal{S}, X')$ guesses $\mathtt{ch}_2(\mathcal{S}, X')$. Also, we already stated that X'_{S_1} is constant, that X'_{S_3} is a function of X'_{S_2}, and that the S_2 and S_4 blocks of X' are independent. We also showed that the min-entropy of the S_2-block of X' is at least $\delta^2/40$ and that the min-entropy of its S_4-block is at least $\delta^3 n/80$.

This takes care of all the properties of a good partition except for property 6, which we state and prove in the following claim.

Claim 9.4. *The probability that $\mathcal{S}$ is not the maximal valid partition is bounded by $c(m_0, \delta/20)/2$.*

Proof. Let $\mathcal{T} = \{T_1, \ldots, T_5\}$ be a candidate partition that is bigger than $\mathcal{S}$. Since the number of such partitions is easily bounded by $1000/\delta^3$, to complete the claim it is enough to show that the probability that $\mathcal{T}$ is valid is bounded by $c(m_0, \delta/20) \cdot \delta^3/2000$.

We consider two cases. The easier case is where $T_1 = S_1$, $T_2 = S_2$, and T_3 contains $S_3 \cup S_4$. In this case the entropy of $H^\infty(X'_{T_3}|X'_{T_1\cup T_2}) \geq H^\infty(X'_{S_4}) \geq \delta^3 n/80$. For every j and for every $z \in \mathsf{Supp}(X'_{T_1\cup T_2\cup T_{4,j}})$, let us estimate

$$\Pr\big[v_j(\mathcal{T}, X') \text{ guesses } \mathtt{ch}_1(\mathcal{T}, X') \mid X'_{T_1\cup T_2\cup T_{4,j}} = z\big] = \Pr[v_j(\mathcal{T}, X^{(j,z)}) \text{ guesses } \mathtt{ch}_1(\mathcal{T}, X^{(j,z)})] ,$$

where we have used $X^{(j,z)} \stackrel{def}{=} (X'|X'_{T_1\cup T_2\cup T_{4,j}} = z)$.

Note that by Corollary 3.13,

$$\begin{aligned} H^\infty(X^{(j,z)}_{T_3}) &= H^\infty(X'_{T_3}|X'_{T_1\cup T_2\cup T_{4,j}}) = H^\infty(X'_{T_3\cup T_{4,j}}|X'_{T_1\cup T_2\cup T_{4,j}}) \\ &= H^\infty(X'_{T_3\cup T_{4,j}}|X'_{T_1\cup T_2}) - H^\infty(X'_{T_{4,j}}|X'_{T_1\cup T_2}) \geq H^\infty(X'_{T_3}|X'_{T_1\cup T_2}) - H^\infty(X'_{T_{4,j}}) \\ &\geq \delta^3 n/80 - |T_{4,j}| \geq \delta^3 n/100 , \end{aligned}$$

and hence the T_3-block of $X^{(j,z)}$ satisfies the rate requirement of the affine somewhere extractor applied to it by $\mathtt{ch}_2$. It follows that the "almost uniform" output block of $\mathtt{ch}_2$ (which is of length m_2) obtains each of its values with probability at most 2^{1-m_2}.

Since in order for $v_j(\mathcal{T}, X^{(j,z)})$ to guess $\mathtt{ch}_2$ this block must appear as a sub-string of one of the $\ell(\delta/2)$ output blocks of $v_j(\mathcal{T}, X^{(j,z)})$ (which are all fixed), it follows that the probability that $v_j(\mathcal{T}, X^{(j,z)})$ guesses $\mathtt{ch}_2(\mathcal{T}, X^{(j,z)})$ is bounded by

$$2^{(1-m_2)} \cdot \ell(\delta/2) \leq c(m_0, \delta/20) \cdot \delta^{13}/2000 .$$

This bound holds when $X'_{T_1 \cup T_2 \cup T_{4,j}}$ is fixed to any string z, and therefore it holds in general that

$$\Pr[v_j(\mathcal{T}, X') \text{ guesses } \mathtt{ch}_1(\mathcal{T}, X')] \leq c(m_0, \delta/20) \cdot \delta^{13}/2000 \ .$$

Applying the union bound over all j's (there are less than $1/\delta^{10}$ many), we have

$$\Pr[v_j(\mathcal{T}, X') \text{ guesses } \mathtt{ch}_1(\mathcal{T}, X') \text{ for any } j\ ] \leq c(m_0, \delta/20) \cdot \delta^{3}/2000 \ .$$

We thus have the desired bound on the probability that $\mathcal{T}$ is valid.

Now let us consider the second case, where $S_1 \cup S_2 \subseteq T_1$. This case is a bit trickier, and we start by analysing the probability that $\mathtt{ch}_1(\mathcal{T}, X^{(5)})$ is guessed by $u_i(\mathcal{T}, X^{(5)})$ for some fixed i (the source $X^{(5)}$ has the advantage of being affine). As in the first case, we pick a string $z \in \mathsf{Supp}(X^{(5)}_{T_{2,i}})$, and define $X^{(i,z)} = (X^{(5)} | X^{(5)}_{T_{2,i}} = z)$. In a way similar to the analysis of the first case, we use the fact that $S_2 \subseteq T_1$ and Corollary 3.13, and obtain that $\mathsf{r}(X^{(i,z)}_{T_1}) \geq \delta^2/50$.

The T_1-block of $X^{(i,z)}$ therefore satisfies the rate requirement of the affine somewhere extractor applied to it by $\mathtt{ch}_1(\mathcal{T}, X^{(i,z)})$, and hence $\mathtt{ch}_1(\mathcal{T}, X^{(i,z)})$ has an "almost uniform" output block, which obtains each of its values with probability at most 2^{1-m_1}. Since $u_i(\mathcal{T}, X^{(i,z)})$ is a fixed string containing $\ell(\delta/8)$ blocks, the probability that it guesses $\mathtt{ch}_1(\mathcal{T}, X^{(i,z)})$ is bounded by $2^{1-m_1} \cdot \ell(\delta/8)$. Since the same analysis holds for every z, we obtain

$$\Pr[u_i(\mathcal{T}, X^{(5)}) \text{ guesses } \mathtt{ch}_1(\mathcal{T}, X^{(5)})] \leq 2^{1-m_1} \cdot \ell(\delta/8) \ .$$

Summing up over all i's we get

$$\Pr[u_i(\mathcal{T}, X^{(5)}) \text{ guesses } \mathtt{ch}_1(\mathcal{T}, X^{(5)}) \text{ for any } i\ ] \leq 2^{1-m_1} \cdot \ell(\delta/8)/\delta^{10} \ . \tag{15}$$

Now X' is a sub-source of $X^{(5)}$ of measure at least $2^{-\ell(\delta^3/100)m_2}$, and thus the probability of any event for $X^{(5)}$ can increase by a factor of at most $2^{\ell(\delta^3/100)m_2}$ when restricted to X'. Hence (15) implies

$$\begin{aligned}\Pr[u_i(\mathcal{T}, X') \text{ guesses } \mathtt{ch}_1(\mathcal{T}, X') \text{ for any } i\ ] &\leq 2^{\ell(\delta^3/100)m_2} \cdot 2^{1-m_1} \cdot \ell(\delta/8)/\delta^{10} \\ &\leq c(m_0, \delta/20) \cdot \delta^3/2000 \ ,\end{aligned}$$

which completes the analysis of the second case. □

References

[BIW] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting Randomness from Few Independent Sources. In *Proc. 45th FOCS*. IEEE, 2004.

[BST] B. Barak, R. Shaltiel, and E. Tromer. True Random Number Generators Secure in a Changing Environment. In *Workshop on Cryptographic Hardware and Embedded Systems (CHES)*, pages 166–180, 2003. LNCS no. 2779.

[BOL] M. Ben-Or and N. Linial. Collective Coin Flipping, Robust Voting Schemes and Minima of Banzhaf Values. In *Proc. 26th FOCS*, pages 408–416. IEEE, 1985.

[BSHRV] E. Ben-Sasson, S. Hoory, E. Rosenman, and S. Vadhan. Personal Communication, 2001.

[Blu] M. Blum. Independent Unbiased Coin Flips From a Correlated Biased Source: A Finite State Markov Chain. In *Proc. 25th FOCS*, pages 425–433. IEEE, 1984.

[CRVW] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proc. 34th STOC*, pages 659–668. ACM, 2002.

[CG] B. Chor and O. Goldreich. Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. In *Proc. 26th FOCS*, pages 429–442. IEEE, 1985.

[CGH+] B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, and R. Smolensky. The Bit Extraction Problem of t-Resilient Functions (Preliminary Version). In *Proc. 26th FOCS*, pages 396–407. IEEE, 1985.

[CW] A. Cohen and A. Wigderson. Dispersers, Deterministic Amplification, and Weak Random Sources. In *Proc. 30th FOCS*, pages 14–19. IEEE, 1989.

[DEOR] Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz. Improved Randomness Extraction from Two Independent Sources. In *Proc. of 8th RANDOM*, 2004.

[DS] Y. Dodis and J. Spencer. On the (non)Universality of the One-Time Pad. In *Proc. 43rd FOCS*, pages 376–388. IEEE, 2002.

[FW] P. Frankl and R. M. Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1(4):357–368, 1981.

[GRS] A. Gabizon, R. Raz, and R. Shaltiel. Deterministic Extractors for Bit-Fixing Sources by Obtaining an Independent Seed. In *Proc. 45th FOCS*. IEEE, 2004.

[Gol] O. Goldreich. Three XOR-Lemmas – An Exposition. ECCC Report TR95-056, 1995.

[KZ] Kamp and Zuckerman. Deterministic Extractors for Bit-Fixing Sources and Exposure-Resilient Cryptography. In *Proc. 44th FOCS*. IEEE, 2003.

[LRVW] C.-J. Lu, O. Reingold, S. Vadhan, and A. Wigderson. Extractors: optimal up to constant factors. In *Proc. 35th STOC*, pages 602–611. ACM, 2003.

[MP] J. L. McInnes and B. Pinkas. On the Impossibility of Private Key Cryptography with Weakly Random Keys. In *Crypto '90*, pages 421–436, 1990. LNCS No. 537.

[MU] Mossel and Umans. On the Complexity of Approximating the VC Dimension. *J. Comput. Syst. Sci.*, 65, 2002.

[PR] P. Pudlák and V. Rödl. Pseudorandom sets and explicit constructions of Ramsey graphs, 2004. Submitted for publication.

[Raz] R. Raz. Extractors with Weak Random Seeds, 2004. Submitted for publication.

[RSW] O. Reingold, R. Shaltiel, and A. Wigderson. Extracting randomness via repeated condensing. In *Proc. 41st FOCS*, pages 22–31. IEEE, 2000.

[SV] M. Santha and U. V. Vazirani. Generating Quasi-Random Sequences from Slightly-Random Sources. In *Proc. 25th FOCS*, pages 434–440. IEEE, 1984.

[Sha] R. Shaltiel. Recent developments in extractors. *Bulletin of the European Association for Theoretical Computer Science*, 2002. Available from `http://www.wisodm.weizmann.ac.il/~ronens`.

[TSUZ] A. Ta-Shma, C. Umans, and Z. Zuckerman. Loss-less condensers, unbalanced expanders, and extractors. In *Proc. 42nd FOCS*, pages 143–152. IEEE, 2001.

[TV] L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *Proc. 41st FOCS*, pages 32–42. IEEE, 2000.

[Vaz] U. Vazirani. Strong Communication Complexity or Generating Quasi-Random Sequences from Two Communicating Semi-Random Sources. *Combinatorica*, 7, 1987. Preliminary version in STOC' 85.

[vN] J. von Neumann. Various Techniques Used in Connection with Random Digits. *Applied Math Series*, 12:36–38, 1951.

[Zuc] D. Zuckerman. General Weak Random Sources. In *Proc. 31st FOCS*, pages 534–543. IEEE, 1990.