Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 183 (2019)

Randomness Extraction from Somewhat Dependent Sources

Marshall Ball Oded Goldreich Tal Malkin

January 8, 2020

Abstract

We initiate a comprehensive study of the question of randomness extractions from two somewhat dependent sources of defective randomness. Specifically, we present three natural models, which are based on different natural perspectives on the notion of bounded dependency between a pair of distributions. Going from the more restricted model to the less restricted one, our models and main results are as follows.

1. *Bounded dependence as bounded coordination*: Here we consider pairs of distributions that arise from independent random processes that are applied to the outcome of a single global random source, which may be viewed as a mechanism of coordination (which is adversarial from our perspective).

   We show that if the min-entropy of each of the two outcomes is larger than the length of the global source, then extraction is possible (and is, in fact, feasible). We stress that the extractor has no access to the global random source nor to the internal randomness that the two processes use, but rather gets only the two dependent outcomes.

   This model is equivalent to a setting in which the two outcomes are generated by two independent sources, but then each outcome is modified based on limited leakage (equiv., communication) between the two sources. (Here this leakage is measured in terms of the number of bits that were communicated, but in the next model we consider the actual influence of this leakage.)

2. *Bounded dependence as bounded cross influence*: Here we consider pairs of outcomes that are produced by a pair of sources such that each source has bounded (worst-case) influence on the outcome of the other source. We stress that the extractor has no access to the randomness that the two processes use, but rather gets only the two dependent outcomes.

   We show that, while (proper) randomness extraction is impossible in this case, randomness condensing is possible and feasible; specifically, the randomness deficiency of condensing is linear in our measure of cross influence, and this upper-bound is tight. We also discuss various applications of such condensers, including for cryptography, standard randomized algorithms, and sublinear-time algorithms, while pointing out their benefit over using a seeded (single-source) extractor.

3. *Bounded dependence as bounded mutual information*: Due to the average-case nature of this mutual information, here there is a trade-off between the error (or deviation) probability and the randomness deficiency. Loosely speaking, for joint distributions of mutual information t, we can condense with randomness deficiency $O(t/\epsilon)$ and error ϵ, and this trade-off is optimal.

All positive results are obtained by using a standard two-source extractor (or condenser) as a black-box.

ISSN 1433-8092

Contents

1 Introduction

The problem of extracting almost perfect randomness from sources of highly defective randomness is of great theoretical and practical importance, since perfect randomness is essential to cryptography and has numerous applications in algorithmic design, whereas natural sources of randomness are quite defective. In other words, the randomness extraction problem addresses the discrepancy between the perfect randomness that is postulated in various applications and the quite defective randomness that seems available to us in reality.

The foregoing problem has been the focus of much research in the last decades, where research has branched according to how the defective sources of randomness are modelled (see, e.g., [20]). Needless to say, an adequate modelling of such sources is pivotal to such studies. The two main branches, reviewed below, focus on sources of randomness with a worst-case guarantee asserting that no outcome appears with too high a probability. Specifically, the logarithm of the reciprocal of this probability, called *min-entropy*, is a main parameter in these studies.[1]

The two branches differ by the question of whether we have at our disposal a single source of the foregoing type or a few (say, two) *independent* sources of this type. In the first case, a randomness extractor cannot be deterministic; it must use a short random *seed*, where the length of the seed may be logarithmic in the length of the source (and the reciprocal of the desired error probability). Shaltiel's classical survey is focused on this case [19]; see also more recent accounts such as [21, Chap. 6].[2]

In some "off-line" algorithmic applications, seeded extractors provide a good solution to the extraction problem, since one can emulate the use of a short random seed by a deterministic enumeration of all possibilities. This emulation is not possible in "on-line" applications that dominate the areas of cryptography and distributed computing. This reality is the main motivation for the second branch of studies, which considers extraction from several *independent* sources (of defective randomness). Typically, one seeks to minimize the number of independent sources, and it is best to use only two. The rather recent breakthrough result of Chattopadhyay and Zuckerman [5], which provides a two-source extractor for polylogarithmic min-entropy, and its follow-ups (e.g., [14]) belong to this branch.

While some amount of independence between the two sources is definitely necessary for (seedless) extraction, it is desirable to allow as much dependence as possible. In other words, one should seek to relax the postulate of perfectly independent (defective) sources, and study the possibility of extracting randomness from somewhat dependent sources.

Actually, this problem was considered already in the early work of Chor and Goldreich [6], who suggested a simple definition and outlined the possibility and limitation of (proper) extraction under it (see [6, Sec. 3.3]). Their focus was on a small "amount of dependence" (i.e., the sources are almost independence (in a strong sense)), but in that regime their definition (i.e., [6, Def. 10]) does not meet our intuition. On the one hand, it is too rigid, as reflected by the fact that it does not cover natural cases that do allow for good extraction and are covered by our first model (described in Section 1.1.1). On the other hand, their definition led to a negative result regarding extraction

[1] Hence, the postulate that no n-bit long string occurs with probability greater than p takes the form of saying that the distribution has min-entropy at least $\log_2(1/p)$. Note that a perfect n-bit long random string has min-entropy $\log_2(1/2^{-n}) = n$.

[2] The focus of Shaltiel's survey [19] on seeded extractors reflects the focus of research at that period, and specifically the quest for explicit constructions of extractors with optimal parameters such as seed-length, min-entropy bound, output length, and error probability (see, e.g., [19, Sec. 1.4] and [19, Table 1]).

(i.e., [6, Thm. 19]), which seems to have discouraged research in this important direction. (See Section 1.4 for a more detailed account.)

In this work, we initiate a more comprehensive study of the question of randomness extractions from two somewhat dependent sources of defective randomness. Specifically, we present three natural models, which are based on different natural perspectives on the notion of dependency between a pair of distributions. Indeed, the most general model is based on the notion of mutual information, but we believe that the more restricted models are also natural. While the most restricted model (described in Section 1.1.1) allows for (proper) randomness extraction, the other two models (described in Sections 1.1.2 and 1.1.3) do not allow for (proper) extraction, but do allow for randomness condensing. In Section 1.3 we argue that these randomness condensers may be of value for various applications, including in cryptography. But let us discuss the models themselves first.

1.1 Our models

Our three models draw on three different notions of (the amount of) "dependency" between sources. In each model, the "amount of dependence" is specified by a parameter that upper-bounds this amount (as well as by the standard parameter that lower-bounds the min-entropy of the individual sources).

The models are presented in order of generality, going from the more restricted to the less restricted. When the amount of dependency is zero, all models coincide with the standard model of two independent sources. On the other hand, when the amount of dependence reaches the min-entropy of the individual sources, all models include the case of identical sources (which coincides with the single source case). Hence, the amount of dependence that we consider is between these two extreme bounds.

We refrain form taking a categorical position regarding which of the models is "right"; we believe that each of them may be adequate in some settings and less so in others. Indeed, different models may suit different settings, and the fact that the models support different levels of extraction (or condensing) is a good reason to present them all.

1.1.1 The coordinated sources model

A *special case* (equiv., restricted version) of the "coordinated sources" model postulates that the two sources have access to many independent "micro-sources" of randomness that are each extremely defective (i.e., having min-entropy that is too low to be of any use). Furthermore, most of these micro-sources may have no entropy at all. Each source corresponds to a subsequence of the micro-sources, and each such subsequence contains a significant amount of min-entropy, but the subsequences corresponding to the two sources have a small (non-empty) intersection that is not known to us. Of course, if we knew the intersection, then we could have ignored the micro-sources in it when extracting from the two (residual) sources (i.e., from the non-intersecting parts), but the intersection is not known to us (i.e., to the extractor).[3]

More generally, the two sources may be random processes that are each fed by the outcome of some global random process. We have no access to "underlying" global random source nor can we

[3]Likewise, if we knew which of the micro-sources included in one subsequence contain a sufficient amount of min-entropy, then we could partition this subsequence into two good parts and extract from these two auxiliary sources.

access the randomness "added" by each of the two sources (on top of the outcome of the global source). All we get is the outcomes of the two processes (i.e., sources). Specifically, consider the randomized processes A, B and C, each taking a "somewhat random" n-bit long string, denoted r_A, r_B and r_C. Then, we get $A(r_A, C(r_c))$ and $B(r_B, C(r_c))$ only, and we are only guaranteed that their min-entropy is k' bits larger than the length of the outcome of C. Note that we do not require that r_A, r_B and r_B be uniformly distributed, but the min-entropy requirement made regarding the outcomes of A and B does imply that r_A and r_B have min-entropy at least k' each. Here, r_C represents the coordination between the sources.

A different scenario, which is actually equivalent to the above, is that the two sources are initially independent, but become dependent due to (bounded) leakage. Specifically, suppose that each source starts having a somewhat random state (i.e., s_A and s_B), but their state changes in time and may depend also on few bits that are leaked between the states; that is, at each point in time, each source modifies its state based on its current state and bits that are leaked from the current state of the other source. We only see the states at a later time, and after a bounded number of bits were "communicated" (via leakage) between them. Here, this leakage represents the coordination between the sources.

1.1.2 Sources of bounded cross influence

In the bounded "cross influence" model, we envision a setting akin the prior ones, except that here we do not upper-bound the *potential* "influence" of each source on the other – as reflected in the amount of leakage (or communication) – but rather the *actual* influence as embodied in the two outcomes. Specifically, consider randomized processes A and B, each taking a sequence of n random bits, denoted r_A and r_B, and outputting $A(r_A, r_B)$ and $B(r_A, r_B)$; that is, each source is given both r_A and r_B. What we shall bound is the influence of r_B on A's output, and likewise the influence of r_A on B's output.

We say that the influence of r_B on A's output is at most t if for every two values r_A and r_B it holds that $\Pr_r[A(r_A, r) \neq A(r_A, r_B)] \leq 1 - 2^t$. In other words, the value $A(r_A, r_B)$ is maintained, with probability at least 2^{-t}, when r_B is "re-randomized" (i.e., replaced by a random input r). Hence, the influence of r_B on A's output measures the actual effect that r_B has on A's output, regardless of how this effect comes about. The influence of r_A on B's output is defined analogously, and the cross influence of the two sources is defined as the sum of the two (opposite) influences.

To see that the cross influence can be much lower than the amount of coordination, consider the processes $A(r_A, r_B) = r_A$ and $B(r_A, r_B) = (r_B, \mathtt{IP}_2(r_A, r_B))$, where $\mathtt{IP}_2$ denotes the inner product (mod 2) function. These two sources have a single bit of cross influence (i.e., r_B has no influence on A's output, whereas the influence of r_A on B's output is confined to $\mathtt{IP}_2(r_A, r_B)$). It can be shown that the amount of coordination between these sources exceeds $0.499 \cdot |r_A|$ (see Theorems 3.4 and 6.1).[4]

We also comment that, as proved in Proposition 4.2, *if a joint distribution* (X, Y) *has cross*

[4]Indeed, it is well-known that the communication complexity of computing $\mathtt{IP}_2$ on n-bit inputs is $\Omega(n)$, see [6, Thm. 21(ii)], but this does not mean that sampling the distribution $(r_A, (r_B, \mathtt{IP}_2(r_A, r_B)))$ requires $\Omega(n)$ bits of communication (since there may be other ways of sampling this distribution). In other words, the strategies in the coordination protocol need not compute $\mathtt{IP}_2$ on their disjoint inputs; they may sample the desired distribution arbitrarily. Nevertheless, Theorem 6.1 implies that a two-party protocol for sampling the distribution $(r_A, (r_B, \mathtt{IP}_2(r_A, r_B)))$ requires $\Omega(n)$ bits of communication, whereas Theorem 3.4 relates the communication complexity of sampling a joint distribution to the amount of coordination in it.

influence at most t, then $\min_{x,y}\{\ell(x,y)\} \leq t$, *where* $\ell(x,y) = \log_2(\Pr[(X,Y)=(x,y)]/\Pr[X=x]\cdot\Pr[Y=y])$. This means that the min-entropy of (X,Y) is at most t units smaller than the sum of the min-entropies of X and Y. Note that the mutual information of X and Y equals $\mathrm{E}[\ell(X,Y)]$.

1.1.3 Sources of bounded mutual information

The mutual information of the joint distribution (X,Y), denoted $I(X;Y)$, is a well-established measure of the mutual dependence between the two variables. It quantifies the amount of information obtained about one random variable through observing the other random variable; indeed, $I(X;Y) = H(X) + H(Y) - H(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$. It seems that considering this measure in the current context requires no justification. Still, we note that this measure coincides with the *information* communicated between the parties in a generation protocol as considered in the case of coordinated sources.

Recall that the number of bits communicated in such a protocol captures the measure of coordination considered in our first model. But here we consider the *information about one party's random input that is revealed by the communication*. Specifically, consider the trivial protocol in which A generates $(x,y) \leftarrow (X,Y)$ and sends y to B. Then, the information on X communicated to B equals $H(X) - H(X|Y)$, which equals $I(X;Y)$.

1.2 Our results

For the sake of stating our results, we use the following notation, where in all cases n denotes the length of the source's outcome.

$\mathtt{STD}_n(k) =$ the standard two-source model, where each (independent) source has min-entropy k.

$\mathtt{COOR}_n(k,t) =$ the model of *t-coordinated sources*, where each source has min-entropy k. (See Section 1.1.1 and Definition 3.1.)

$\mathtt{CRI}_n(k,t) =$ the model of joint distributions of *cross influence t*, where each source has min-entropy k. (See Section 1.1.2 and Definition 4.1.)

$\mathtt{MI}_n(k,t) =$ the model of joint distributions of *mutual information t*, where each source has min-entropy k. (See Section 1.1.3 and Definition 5.1.)

We first state the fact that, essentially, our models are contained in one another, and that this containment is strict.

Theorem 1.1 (relations between the models):

1. $\mathtt{COOR}_n(k,t)$ is ϵ-close to being in $\mathtt{CRI}_n(k,t+\log_2(1/\epsilon))$: *For every $\epsilon > 0$ and every t, every t-coordinated joint distribution is ϵ-close to some distribution of cross influence $t+\log_2(1/\epsilon)$. On the other hand, there exists a joint distribution* (of min-entropy $n-4$) *that can be generated with at most six bits of cross influence, but is* 0.24*-far from any $(n-O(\log n))$-coordinated distribution.*

2. $\mathtt{CRI}_n(k,t)$ is strictly contained in $\mathtt{MI}_n(k,t)$: *For every t, every joint distribution of cross influence t has t bits of mutual information. On the other hand, for every $\epsilon > 0$, there*

exists a joint distribution (of min-entropy $n - O(1/\epsilon)$) *that that has mutual information at most three, but is* $(\epsilon/2)$*-far from any distribution that can be generated with* $1/\epsilon$ *bits of cross influence.*

The separation between cross influence and mutual information seems to be due to the separation between worst-case and average-case notions of cross influence, as established in Proposition 4.9. Indeed, in retrospect, a worst-case notion is more adequate in the current setting (cf., min-entropy versus entropy).

We next turn to our results regarding randomness extraction (and condensing) for each of the three models. Recall that an *extractor with error* ϵ *for a model* $\mathcal{M}$ is a function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ such that for every joint distribution (X,Y) in $\mathcal{M}$, it holds that $F(X,Y)$ is ϵ-close to the uniform distribution over $\{0,1\}^m$.

Theorem 1.2 (extraction for sources of bounded coordination): *If* $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *is an extractor of error* ϵ *for the model* $\mathtt{STD}_n(k)$*, then* F *is an extractor of error* 2ϵ *for the model* $\mathtt{COOR}_n(k + t + \log_2(1/\epsilon), t)$.

We comment that the loss of $t + \log_2(1/\epsilon) - O(1)$ units of min-entropy is unavoidable. But the good news is that modulo this loss, extraction for sources of bounded coordination is possible, let alone via a black-box use of standard extractors, provided that the min-entropy of the individual sources compensates for the amount of coordination. Unfortunately, this good news does not carry over to the other models.

Turning to the other models, we first note that proper extraction is not possible, but condensing is possible. Recall that a *condenser with error* ϵ *and deficiency* d *for a model* $\mathcal{M}$ is a function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ such that for every joint distribution (X,Y) in $\mathcal{M}$, it holds that $F(X,Y)$ is ϵ-close a distribution that has min-entropy at least $m - d$.

Theorem 1.3 (extraction for sources of bounded cross influence):

1. (condensing is possible): *If* $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *is a condenser of error* ϵ *and deficiency* d *for the model* $\mathtt{STD}_n(k)$*, then* F *is a condenser of error* $2^t \cdot \epsilon$ *and deficiency* $d + t$ *for the model* $\mathtt{CRI}_n(k,t)$.

2. (the foregoing is essentially optimal): *There is no condenser with deficiency* $o(t)$ *and error* $2^{-\Omega(t)}$ *for* $\mathtt{CRI}_n(n - O(t), t)$*, when the output length exceeds* t*. Furthermore, there is no extractor with error* $1/4$ *for* $\mathtt{CRI}_n(n-4, 6)$.

Indeed, $d = 0$ is a special case that refers to the case that F is an extractor for $\mathtt{STD}_n(k)$. In general, the fact that the condensing error (in Part 1) grows by a factor of 2^t does not worry us too much, since we can afford to make ϵ small enough to compensate for that (i.e., we can use $\epsilon \leq 2^{-2t}$, provided $k = \Omega(t)$).[5] So the take-home message here is that *condensing is possible at a cost — in terms of deficiency — that equals the amount of cross influence.* As will be articulated in Section 1.3, randomness of bounded deficiency is useful in many setting both in cryptography and in algorithmic design. For mutual information, we obtain a weaker condensing result.

[5] Currently, explicit constructions of standard extractors with error $\epsilon < 1/n$ are not known, but such extractors do exist. Furthermore, explicit constructions of condensers with error ϵ and deficiency $o(\log(1/\epsilon))$ were recently presented in [3].

Theorem 1.4 (extraction for sources of bounded mutual information): *For every $\beta > 0$, the following holds.*

1. (condensing is possible): *If $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is a condenser of error ϵ and deficiency d for the model $\mathtt{STD}_n(k-1)$, then F is a condenser of error $4\beta + 2^{t'} \cdot \epsilon$ and deficiency $d + t'$ for the model $\mathtt{MI}_n(k,t)$, where $t' = (t + O(1))/\beta$.*

2. (the foregoing is essentially optimal): *There is no condenser with deficiency $o(t')$ and error $O(\beta)$ for $\mathtt{MI}_n(n - t', t)$, when the output length exceeds $t' \stackrel{\text{def}}{=} ((t-2)/\beta)$.*

Indeed, this condensing result is inferior to the one for the cross influence model (i.e., Theorem 1.3), since it involves a trade-off between the deficiency and the error parameter. Specifically, in Theorem 1.4 the product of the error and the deficiency is linear in the mutual information bound (i.e., $\beta \cdot t' = \Theta(t)$; indeed, our issue is with the additive error term of β, not with the multiplicative factor of $2^{t'}$ (which can be tolerated as outlined above)). Still, as will be discussed in Section 1.3, some applications can benefit even from the above trade-off.

Yet another definition of bounded dependence. Part 1 of Theorem 1.3 is proved by showing that if (X,Y) can be generated with t bits of cross influence, then, for every x and y,

$$\Pr[(X,Y)=(x,y)] \leq 2^t \cdot \Pr[X=x] \cdot \Pr[Y=y]. \tag{1}$$

(see Proposition 4.2). In other words, the min-entropy of (X,Y) is at most t units smaller than the sum of the min-entropies of X and Y. One may consider this parametrized upper bound as yet another definition of bounded dependence, and note that it implies that (X,Y) has mutual information at most t. Hence, the parameter t in Eq. (1) is sandwiched between cross influence and mutual information. Furthermore, we show approximate converses of both inequalities:

- If (X,Y) satisfies Eq. (1) with parameter t, then it is ϵ-close to a distribution that can be genertated with $t + O(\log(1/\epsilon))$ bits of cross influence (see Proposition 4.13).
- If (X,Y) has mutual information t, it is $O(\beta)$-close to a distribution (X',Y') that satisfies Eq. (1) with respect to the parameter $t' = (t + O(1))/\beta$. (This is the bulk of the proof of Theorem 5.5, which establishes Part 1 of Theorem 1.4.)

We also show that perfect converses do not hold (see Proposition 4.14 and Part 2 of Theorem 1.1, resp.).

1.3 On the usefulness of condensers

Randomness condensers (a.k.a condensers), introduced in [16, 15] (in seeded and seedless versions, respectively), transform highly defective sources of randomness, which are only guaranteed to have some minimal amount of min-entropy, into sources of randomness (that are close to) having a relatively small min-entropy deficiency (see Definition 2.1). This falls short of what is done by randomness extractors, which transform highly defective sources of randomness into almost perfect sources of randomness, but it is highly non-trivial and useful in various ways.

Typically, condensers are viewed as a technical tool; specifically, as an intermediate step towards extractors. However, condensers may be useful by themselves, whenever randomness extraction is

impossible or too expensive. To be more concrete, we distinguish three possible types of uses of randomness: Its uses in cryptography (and other adversary-ridden settings), in standard algorithmic applications, and in sub-linear time computations.

The distinction boils down to the question of whether or not the paradigm of "de-randomizing the seed of a seeded extractor" is applicable. The archetypical case in which this is applicable is when running a randomized decision procedure (or a pseudodeterministic search procedure [9]). In this case, we obtain a single sample from a single source, apply the (seeded) extractor with all possible values of the seed, invoke the original algorithm with its randomness replaced by each of the outcomes, and output the result that is in majority. (Recall that it is impossible to deterministically perform (seedless) randomness extraction from a single defective source, even if the min-entropy of the source is very high [6, Thm. 1]. Furthermore, the seed length of a seeded extractor (for a single source) must be at least logarithmic in the length of the source [17, Thm. 1.9].)

1.3.1 Cryptographic applications

Cryptographic settings are the archetypical case in which the paradigm of "de-randomizing the seed of a seeded extractor" is not applicable. Hence, one may either assume that the small amount of (perfect) randomness required for the seed of an extractor can obtained (independently of the main source of randomness) or assume that we have access to two (or more) independent sources of defective randomness. But what if the two sources are dependent in a manner allowing only for condensing (as in Theorems 1.3 and 1.4)?

As noted by Rao [15] (see articulation in [8, Sec. 5]), it turns out that a level of min-entropy deficiency that is sub-logarithmic in the "security error" can be tolerated in "unpredictability applications" (e.g., unforgeable signatures). Specifically, we say that a cryptographic system (of this type) has a security error of μ if adversaries (with specified resources) can break the system with probability at most μ (see [8, Def. 5.1]). The key observation is that *if the system has a security error of μ when using perfect randomness, then implementing the system with a distribution that is ϵ-close to having deficiency at most d yields a system of security error at most $2^d \cdot \mu + \epsilon$.* An analogous bound holds when a system designed for randomness of deficiency d_0 is used with randomness that has deficiency $d_0 + d$ (see [8, Lem. 5.1]).

The foregoing suggestion is appealing when having access to two sources of sufficiently low cross influence. Specifically, Theorem 1.3 asserts that, when using a standard two-source condenser (or extractor), a cross influence bound of t translates to an added deficiency of t units (while the error gets multipied by a factor of 2^t). Furthermore, in such a case we can use standard two-source condensers with small deficiency rather than standard two-source extractor.

The latter comment is important because the currently known explicit two-source extractors (for low (i.e., lower than 0.4) min-entropy (e.g., [5, 14])) have noticeable error (i.e., error that is polynomially related to the length of the source), which is unacceptable in most cryptographic applications. In contrast, a recent work of Ben-Aroya *et al.* [3] provides an explicit two-source condenser with deficiency that is *sub-logarithmic* in the desired error, which in turn may be set to be negligible. Hence, if an "unpredictability applications" has a security error of μ when using perfect randomness, then implementing it with two sources of cross influence t yields a system of security error at most $2^{2t} \cdot \mu^{1-o(1)}$, since for independent sources a condensing error of μ can be achieved with deficiency $o(\log(1/\mu))$.[6]

[6] Actually, we can get a system of security error at most $2^{t+o(t)} \cdot \mu^{1-o(1)}$ by using a standard condenser with error

1.3.2 Standard algorithmic applications

Somewhat surprisingly, using two-source condensers for dependent sources may also be useful for running standard randomized algorithms. That is, *assuming that we have two somewhat dependent sources of weak randomness at our disposal*, we compare the option of using only one of these sources while employing the paradigm of "de-randomizing the seed of a seeded extractor" to the option of using the two sources with an adequate condenser. We claim that in many cases, the latter option is better.

Recall that the standard suggestion is to run such algorithm by extracting almost perfect randomness from a single defective random source, by using a seeded extractor. Following this suggestion, extraction from an n-bit long source requires a seed of length at least $\log_2 n$, which means that under the foregoing paradigm the original algorithm must be invoked $\Omega(n)$ times (while using all possible seeds with the same n-bit long outcome of the source).

In contrast, using the single outcome extracted from two sources with cross influence at most t yields the same performance provided that the error probability of the algorithm is reduced from $\delta < 1/3$ to $2^{-t} \cdot \delta$. The point is that such an error reduction can be obtained by invoking the original algorithm $O(t)$ times, whereas typically $t \ll n$. The bottom-line is that *if the bound on the cross influence of the two sources is good enough* (i.e., $t \ll n$ as well as t being smaller than (say) half the min-entropy of each source), *then we are better off using two somewhat-dependent sources* (via the condenser) rather than one source (via a seeded extractor). (Recall that Theorem 1.3 asserts that a cross influence bound of t translates to an added deficiency of t units, which means that the error probability can grow by a factor of at most 2^t.)

The foregoing holds even when using two sources that have mutual information at most t (rather than cross influence at most t), where we assume all along that each source has sufficient amount of min-entropy. In this case we use Theorem 1.4 rather than Theorem 1.3, which means that the deficiency bound we obtain is larger. Consequently, the number of invocations grows by a factor of $O(t/\epsilon)$, where ϵ is the desired error. That is, given a randomized algorithm A of error probability δ, and wishing to utilize A when having access to a pair of sources that have mutual information t, while obtaining error probability ϵ, we need to reduce A's error to $2^{-\Omega(t/\epsilon)} \cdot \epsilon$ (rather than to $2^{-t} \cdot \epsilon$, as when using sources of cross influence t). Hence, we shall invoke A for $O(t/\epsilon)$ times (rather than $O(t)$ times), which is feasible only if we are willing to tolerate a noticeable error probability (e.g., $\epsilon = 0.01$ or so).

Finally, note that for search problems that can be solved in probabilistic polynomial-time but are not BPP-search problems (i.e., valid solutions cannot be efficiently recognized (cf. [11])[7]), the paradigm of "de-randomizing the seed of a seeded extractor" is not applicable at all, since listing the solutions that correspond to all possible seeds does not allow to select a valid one. So in this case, if the search problem is solvable with sufficiently high probability (when using perfect randomness), then we can use the "condenser path"; but note that error reduction is also not feasible in this case (when valid solutions cannot be efficiently recognized).

1.3.3 Sublinear-time applications

As noted in [10], the fact that the overhead of the paradigm of "de-randomizing the seed of a seeded extractor" is larger than the randomness complexity of the original algorithm provides

$2^{-t} \cdot \mu$ and deficiency $o(t + \log(1/\mu))$.

[7] So, in particular, they do not have pseudodeterministic algorithms (cf. [9]).

another motivation for reducing the randomness complexity of algorithms. However, in the context of sublinear-time computations, this may not be always possible. Two such cases are sampling and property testing (see analogous discussion in [10, Sec. 3]).

Consider, for example, the task of estimating the average of a function $f : \{0,1\}^n \to [0,1]$. Any reasonable notion of estimation will require randomness complexity $\Omega(n)$. Hence, employing the "seed de-randomization" paradigm will require making $\Omega(n)$ probes to the function. In contrast, when given access to two sources of cross influence t, it suffices to make $O(t)$ probes to obtain a constant factor approximation with probability 0.999. (This corresponds to using a sampler that, when using perfect randomness, obtains such an approximation with probability $1 - 2^{-t-10}$.)

The same considerations apply in the context of property testing (see [12]). In particular, many testers have complexity that only depends on the proximity parameter (and is independent of the size of the tested object), hereafter called strong testers. On the other hand, the randomness complexity of any reasonable testing task is at least logarithmic in the size of the object. Hence, strong testability cannot be achieved when using the paradigm of "de-randomizing the seed of a seeded extractor" but it can be obtained when employing a condenser to a pair of sources of cross influence $O(1)$. Furthermore, some testing tasks may have sublinear query and randomness complexities, but the paradigm of "de-randomizing the seed of a seeded extractor" yields query complexity that is the product of the two, which may be more than linear (i.e., worse than the trivial "tester" that just reads the entire object).[8] In contrast, we can obtain sub-linear complexity when employing a condenser to a pair of sources of cross influence $t = o(n/q)$, where n is the size of the tested object and q is the query complexity of the original tester (which uses perfect randomness).

1.4 Related work

The problem of *dependent sources* of defective randomness was already considered in the early work of Chor and Goldreich [6]. In particular, they suggested a simple definition (i.e., [6, Def. 10]), which allowed for extraction with error proportional to the governing parameter (cf., [6, Lem. 18]), alas this error bound was shown to be essentially tight (i.e., [6, Thm. 19]). Specifically, for $\delta > 0$, the joint distribution (X, Y) was said to be δ-dependent if for every $x, y \in \{0,1\}^n$ it holds that

$$(1+\delta)^{-1} \;\leq\; \frac{\Pr[(X,Y)=(x,y)]}{\Pr[X=x]\cdot\Pr[Y=y]} \;\leq\; 1+\delta. \tag{2}$$

While the formulation intentionally allows to consider also $\delta > 1$, the focus of [6, Sec. 3.3] was on smaller values of δ. They observed that any standard extractor with error ϵ yields an extractor with error $\delta + (1+\delta)\cdot\epsilon$ for δ-dependent sources of the same min-entropy bound (cf., [6, Lem. 18]). On the other hand, they showed that any extractor for the class of δ-dependent sources has error $\Omega(\delta)$, even when the sources have min-entropy $n-2$ (cf., [6, Thm. 19(i)]).

We believe that, for small values of δ (i.e., $\delta \approx 0$), the model captured by Eq. (2) is not a satisfactory model of somewhat dependent sources. On the one hand, it is too rigid, as reflected by the fact that it does not cover natural cases that do allow for good extraction and are covered by our first model (described in Section 1.1.1). Consider, for example, a joint distribution (X, Y) such that $X = (X', Z)$ and $Y = (Y', Z)$ where X', Y' and Z are mutually independent (and have each some

[8] E.g., consider the case that both the query and randomness complexities equal a square root of the size of the object.

min-entropy). Then, every pair $(x,y) = ((x',z),(y',z))$ in the support of (X,Y) violates Eq. (2), since the ratio in this case equals $1/\Pr[Z=z]$. Furthermore, for (x',z) and (y',z') (in the support of X and Y, respectively) such that $z \neq z'$, the ratio is not even bounded. However, intuitively, these sources have bounded dependency (i.e., the "dependency" seems $|Z|$), and we should be able to extract good randomness from them, even when the location of the overlapping parts of X and Y are not be known (to the extractor) but $|Z|$ is very small. Indeed, such sources are $|Z|$-coordinated; in fact, they are t-coordinated if 2^t upper-bounds the support of Z (see Definition 3.1).

On the other hand, even for small values of δ, the model captured by Eq. (2) does not allow for (proper) randomness extraction, unless δ is extremely small (i.e., smaller than the desired error of the extractor). However, as mentioned above, extraction is possible in the (incomparable) model of bounded coordination (see Theorem 1.2). In conclusion, the point is that that the model captured by Eq. (2) does not offer a wide class of joint distributions for which proper extraction is possible. Furthermore, while extraction is impossible, condensing may be possible and is possible in this case, since distributions that satisfy Eq. (2) definitely satisfy Eq. (1) with parameter $t = \log_2(1+\delta)$.

Indeed, in retrospect, for large $\delta \gg 1$, Eq. (2) essentially coincides with Eq. (1). On the one hand, as noted above, any joint distribution that satisfies Eq. (2) with parameter δ also satisfies Eq. (1) with parameter $t = \log_2(1+\delta)$. On the other hand, any joint distribution (X,Y) that satisfies Eq. (1) with parameter t also satisfies the upper bound of Eq. (2) with parameter $\delta = 2^t - 1$. Furthermore, while (X,Y) may not satisfy the lower bound of Eq. (2) (for any parameter δ), it is $(2^t-1)^{-1}$-close to satisfy Eq. (2) with parameter $\delta = 2^t - 1$. (More generally, for any $\epsilon \leq (2^t-1)^{-1}$, it holds that (X,Y) is ϵ-close to a distributed that satisfies Eq. (2) with parameter $\delta = 1/\epsilon$.)[9]

Concurrent work. In a concurrent work, Chattopadhyay *et al.* [4] consider the construction of extractors for "adversarial sources" (defined as "somewhat good sources" with "bounded dependency").[10] Specifically, they consider N sources such that at least K of them are good in the sense that they are independent and each contains a considerable amount of min-entropy (i.e., the min-entropy k is $N^{\Omega(1)}$), and each bad source depends on a bounded number of good sources. This seems related to the special case of micro-sounrces discussed in the beginning of Section 1.1.1, where the differences include

- The micro-sources that we consider may each contain a very small amount of min-entropy (e.g., $k = O(1)$). In contrast, in [4] the min-entropy is related to the number of sources (e.g.,

[9] Consider a pair of independent random variables (X',Y') such that $X' \equiv X$ and $Y' \equiv Y$, and let (X'',Y'') equal (X',Y') with probability ϵ and equal (X,Y) otherwise. Then, (X,Y) is ϵ-close to (X'',Y''), which satisfies Eq. (2) with parameter $\delta = 1/\epsilon$. To verify the latter claim, note that

$$\begin{aligned}\Pr[(X'',Y'')=(x,y)] &\geq \epsilon \cdot \Pr[(X',Y')=(x,y)] \\ &= \epsilon \cdot \Pr[X'=x] \cdot \Pr[Y'=y] \\ &= \frac{1}{\delta} \cdot \Pr[X''=x] \cdot \Pr[Y''=y],\end{aligned}$$

and, on the other hand,

$$\begin{aligned}\Pr[(X'',Y'')=(x,y)] &= \epsilon \cdot \Pr[(X',Y')=(x,y)] + (1-\epsilon) \cdot \Pr[(X,Y)=(x,y)] \\ &\leq \epsilon \cdot \Pr[X'=x] \cdot \Pr[Y'=y] + (1-\epsilon) \cdot 2^t \cdot \Pr[X=x] \cdot \Pr[Y=y] \\ &\leq 2^t \cdot \Pr[X''=x] \cdot \Pr[Y''=y],\end{aligned}$$

where the first inequality is due to Eq. (1).

[10] They follow-up on a very recent work of Aggarwal *et al.* [1].

$k = N^{\Omega(1)}$).

- We consider a 2-partition of the micro-sources such that the micro-sources on each side have bounded dependency on the micro-sources of the other wise, where the bound need only be smaller than the total min-entropy on each side. We can allow arbitrary dependency among the micro-sources that reside on the same side as their total minb-entropy is large enough. In contrast, in [4] the bad sources may be coordinated arbitrarily as long as each bad source depends on a bounded number of good sources.

It seems that Chattopadhyay *et al.* [4] envision sources as being controlled by possibly adversarial parties, whereas we try to model sources that are available in nature.

1.5 Organization

In Section 2 we formally define min-entropy and the general notion of an extractor (resp., condenser) for an arbitrary models of joint distributions (i.e., pairs of dependent sources of defective randomness). The core of this work is presented in Sections 3–5, where we study the three models reviewed above. Specifically, the bounded coordination model is studied in Section 3, where it is related to the communication complexity of sampling joint distributions. In Section 4 we study the model of cross influence, and in Section 5 we study joint distributions of bounded mutual information. These two sections are by far the longest and most technically demanding.

In Section 6, we use an idea that underlies the proof of Proposition 4.3 and the results of Section 3 in order to easily derive lower bounds on the communication complexity of sampling. This section is indeed a detour.

2 Preliminaries

We first recall the standard definitions of min-entropy, and (seedless) extractor and condenser for two independent sources. We say that a distribution X is ϵ-close to distributionb Y if the total variation distance between them is at most ϵ (i.e., $\sum_z |\Pr[X=z] - \Pr[Y=z]| \leq 2\epsilon$). Otherwise, we say that X is ϵ-far from Y.

Definition 2.1 (the standard two-source model and extraction from it [6, 15]):[11] *The standard model, denoted* $\mathtt{STD}_n(k)$*, is parameterized by a min-entropy bound, denoted* k*, and the length of the source's outcome, denoted* n*.*

- *An* (n,k)-sources *is a distribution* (or random variable) *over* n*-bit strings having min-entropy at least* k*; that is, a random variables* X *such that* $\Pr[X=x] \leq 2^{-k}$*.*
- *An* extractor with error ϵ *for the model* $\mathtt{STD}_n(k)$ *is a function* $\mathtt{EXT} : \{0,1\}^{n+n} \to \{0,1\}^m$ *such that for every two independent* (n,k)*-sources,* X *and* Y*, it holds that* $\mathtt{EXT}(X,Y)$ *is* ϵ*-close to the uniform distribution over* $\{0,1\}^m$*.*
- *A* condenser with deficiency d and error ϵ *for* $\mathtt{STD}_n(k)$ *is a function* $\mathtt{CND} : \{0,1\}^{n+n} \to \{0,1\}^m$ *such that for every two independent* (n,k)*-sources,* X *and* Y*, it holds that* $\mathtt{CND}(X,Y)$ *is* ϵ*-close to a distribution* (over $\{0,1\}^m$) *that has min-entropy at least* $m-d$*.*

[11] The notions of min-entropy and (n,k)-sources as well as the standard model and extraction from it were introduced in [6]. The notion of a two-source condenser is implicit in [15].

(Indeed, an extractor is a condenser with deficiency 0.)

Whenever we say two-source extractor (resp., condenser), with no farther qualifications, we mean one for the standard model (i.e., for $\mathtt{STD}_n(k)$).

While independent sources can be described each in isolation, as done in Definition 2.1, a pair (or tuple) of dependent sources is described as a joint distribution such that each element of the pair (resp., tuple) represents a source.

In general, a model of joint distributions is merely a set of such distributions. We focus on natural and simply defined models that are specified in terms of few parameters such as min-entropy of individual sources, and a measure of dependency such as bits of *coordination*, bits of *cross influence*, and *mutual information*. The definitions of extractors and condensers extend naturally to these cases.

Definition 2.2 (general definition of extractors and condensers): *Let* $\mathtt{MOD}_n$ *be a model of joint distributions over pairs of* n*-bit long strings.*

- *An* extractor with error ϵ for the model $\mathtt{MOD}_n$ *is a function* $\mathtt{EXT} : \{0,1\}^{n+n} \to \{0,1\}^m$ *such that for every joint distribution* (X,Y) *in* $\mathtt{MOD}_n$ *it holds that* $\mathtt{EXT}(X,Y)$ *is* ϵ*-close to the uniform distribution over* $\{0,1\}^m$*.*
- *A* condenser with deficiency d and error ϵ for $\mathtt{MOD}_n$ *is a function* $\mathtt{CND} : \{0,1\}^{n+n} \to \{0,1\}^m$ *such that for every joint distribution* (X,Y) *in* $\mathtt{MOD}_n$ *it holds that* $\mathtt{CND}(X,Y)$ *is* ϵ*-close to a distribution* (over $\{0,1\}^m$) *that has min-entropy at least* $m-d$*.*

At times we will include condensers in a general discussion of "the (randomness) extraction problem" for a model.

On the computational complexity of extraction. For positive results we will seek explicit constructions, but this is not an issue since we shall use known constructions of extractors (or condensers) for the standard model as black-boxes. Our negative results will rule out certain extractors (or condensers) based on the model's parameters (and the extractor's/condenser's parameters), regardless of explicitness.

On being close to a model (a generic comment). For all models, it holds that if extraction (or condensing) is possible for the model, then extraction (resp., condensing) from any joint distribution that is close to the model (i.e., is close to a distribution in the model) is also possible under almost the same parameters. Specifically, the distance from the model is added to the error of the extractor (resp., condenser). Hence, there is no need to introduce a relaxation of the form of "being close to the model" (in any model).

3 Dependence as coordination

This model is the most restrictive model considered in the current work, and it offers the strongest positive results regarding the extraction of almost perfect randomness.

3.1 Definition

Here we measure the dependence between sources (equiv., the dependence present in a joint distribution) in terms of a "measure of a conditioning" under which these sources are independent. Specifically, we consider a partition of the probability space into a few subspaces and require that in each subspace the conditional distributions are independent. Indeed, the conditioning may be viewed as the result of prior coordination; that is, the joint distribution is based on a global state of coordination such that, given each state of coordination, the sources act independently of one another. Hence, the key parameter is the number of possible states.

An alternative perspective is that the "coordinated component" of the joint distribution is determined *a posteriori*. That is, given a joint distribution, we identify the coordinated states that gave rise to it, such that in each of these states the joint distribution is actually a product distribution. In the following definition, the distribution of these coordinated states is captured by the distribution Z, and, again, the key parameter is the number of possible states.

Definition 3.1 (limited coordination): *A joint distribution (X,Y) is called t-*coordinated *if there exists a related distribution $Z = Z(X,Y)$ such that Z has support size at most 2^t, and, for every z in the support of Z, the conditional distributions $X|_{Z=z}$ and $Y|_{Z=z}$ are independent.*[12] *The corresponding model, denoted $\mathtt{COOR}_n(k,t)$, consists of all t-coordinated joint distribution over $\{0,1\}^{n+n}$ such that each of the n-bit long sources has min-entropy at least k.*

By writing $Z = Z(X,Y)$ we mean to emphasize that Z may be the outcome of a randomized process (not necessarily a function) applied to (X,Y). A natural relaxation allows to discard a small portion (or mass) of the "coordinating distribution" Z (equivalently, to consider only the "effective support size" of Z), but this can be covered by considering a joint distribution that is close to (X,Y).

3.2 Extraction

We observe that any standard two-source extractor works well also for sources that are somewhat coordinated, at the cost of a modest degradation in the parameters (i.e., min-entropy and error).

Theorem 3.2 (extraction for somewhat coordinated sources): *Let $\mathtt{EXT} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ be an extractor with error ϵ for pairs of independent sources each having min-entropy at least k. Then, for every (X,Y) that is t-coordinated such that both X and Y have min-entropy at least $k + t + \log_2(1/\epsilon)$, it holds that $\mathtt{EXT}(X,Y)$ is 2ϵ-close to the uniform distribution over $\{0,1\}^m$. That is, if $\mathtt{EXT}$ is an extractor of error ϵ for the model $\mathtt{STD}_n(k)$, then it is an extractor of error 2ϵ for the model $\mathtt{COOR}_n(k + t + \log_2(1/\epsilon), t)$.*

We note that the loss of $t + \log_2(1/\epsilon) - O(1)$ units of min-entropy is unavoidable; for example, consider the case that $X = (Z, X')$ and $Y = (Z, Y')$, where X' (resp., Y') may depend arbitrarily on Z, which in turn is uniform on a set of 2^t strings of length $n/2$ (e.g., consider the case of $\epsilon = \Omega(1)$).[13]

[12] Indeed, $X|_{Z=z}$ denotes X conditioned on $Z = z$; that is, $\Pr[X|_{Z=z} = x] = \Pr[X = x \mid Z = z]$.

[13] To see what happens even more vividly, notice that if X and Y have min-entropy t, then X' and Y' may have no entropy at all (e.g., we may have $X' \equiv Y' \equiv 0^{n/2}$).

Proof: Let S be the support of Z, and consider

$$S' = \{z \in S : \Pr[Z = z] \geq \epsilon \cdot 2^{-t}\}.$$

Then, $\Pr[Z \in S'] \geq 1 - \epsilon$, since $\Pr[Z \in S \setminus S'] \leq |S| \cdot \epsilon 2^{-t} \leq \epsilon$. Letting $\Delta(.,.)$ denote the total variation distance between distributions (and U_m denote the uniform distribution over $\{0,1\}^m$), we have

$$\begin{aligned}\Delta(\mathtt{EXT}(X,Y),U_m) &\leq \sum_z \Pr[Z=z] \cdot \Delta(\mathtt{EXT}(X|_{Z=z},Y|_{Z=z}),U_m) \\ &\leq \epsilon + \max_{z \in S'}\{\Delta(\mathtt{EXT}(X|_{Z=z},Y|_{Z=z}),U_m)\}.\end{aligned}$$

Next, letting k' denote the min-entropy of X (resp., Y), we observe that, for every $z \in S'$, the min-entropy of $X|_{Z=z}$ (resp., $Y|_{Z=z}$) is at least $k' - (t + \log_2(1/\epsilon)) \geq k$. This holds, because for every x and $z \in S'$, we have

$$\begin{aligned}\Pr[X|_{Z=z}=x] &= \frac{\Pr[X=x \ \& \ Z=z]}{\Pr[Z=z]} \\ &\leq \frac{\Pr[X=x]}{\epsilon \cdot 2^{-t}} \\ &\leq \epsilon^{-1} \cdot 2^t \cdot 2^{-k'},\end{aligned}$$

where the first inequality uses $z \in S'$ and the second inequality uses the min-entropy bound of X. Hence, for every $z \in S'$, we have $\Delta(\mathtt{EXT}(X|_{Z=z},Y|_{Z=z}),U_m) \leq \epsilon$, and the claim follows. ■

3.3 A communication complexity angle

A special case of t-coordinated (joint) distributions arises from t-bit communication protocols for generating joint distributions, with no access to shared randomness. Indeed, such a communication protocol may be viewed as a coordination protocol. Actually, we shall see that that the converse holds too (i.e., every t-coordinated joint distribution can be generated by t-bit communication protocol), which means that the special case is equivalent to the general one.

Definition 3.3 (two-party protocols for generating joint distributions): *We say that a* pair of randomized strategies (A,B) generates the joint distribution (X,Y) *if the interaction of A and B ends with the output pair (X,Y), where A outputs X and B outputs Y. We say that (X,Y)* has communication complexity *t if it can be generated by a pair of randomized strategies that exchange at most t bits of communication* (and have no access to shared randomness).

The parenthetic clarification is crucial; every joint distribution can be generated with zero communication by strategies that have access to shared randomness. We mention that the communication complexity of sampling problems seems to have emerged in [2]; we further study it in Section 6.

Theorem 3.4 (low communication complexity versus low dependence): *A joint distribution has communication complexity t if and only if it is t-coordinated.*

Proof: We first show that having communication complexity t implies being t-coordinated. Let (A, B) be randomized strategies as guaranteed by Definition 3.3, and consider the residual deterministic strategies A' and B' that take random inputs $r \in \Omega$ and $s \in \Omega$, respectively; that is, A selects $r \in \Omega$ uniformly and then acts as $A'(r)$, and ditto for B and B'. Then, a standard communication complexity argument implies that, for every communication transcript $z \in \{0,1\}^t$, the set of input-pairs that yield this transcript is a combinatorial rectangle (i.e., the set of such pairs has the form $I_z \times J_z$ for some $I_z, J_z \subseteq \Omega$). It follows that the corresponding joint output distribution (i.e., $(A'(r), B'(s))$ for (r,s) distributed uniformly in $I_z \times J_z$) is a product distribution; that is, when (r,s) is distributed uniformly in $I_z \times J_z$, the output distribution $(A'(r), B'(s))$ is a Cartesian product of two independent distributions. Recalling that z is distributed over $\{0,1\}^t$, it follows that the output distribution (A, B) is t-coordinated.

Turning to the opposite direction, suppose that (X, Y) is t-coordinated, and let Z be as guaranteed by Definition 3.1. We obtain a t-bit communication protocol for generating (X, Y) by letting A generate Z on its own, send the outcome's index (in the support of Z) to B, and having each party generate the corresponding marginal distribution; that is, if z is the outcome of Z, then A samples $X|_{Z=z}$ and B samples $Y|_{Z=z}$. ■

Digest. The proof of the "low coordination to low communication" direction of Theorem 3.4 uses a uni-directional communication protocol. This protocol make more transparent the fact that *t-coordinated distributions are convex combinations of product distributions*; that is, (X, Y) is t-coordinated if and only if it can be expressed as $\sum_{i \in [2^t]} p_i \cdot (X_i, Y_i)$ for non-negative p_i's (that sum-up to 1), where X_i and Y_i are independent of one another. In other words, the limited dependency in (X, Y) amounts to a joint selection of an $i \in [2^t]$, whereas X_i and Y_i are independent of one another. Hence, the coordination between the two sources is captured by a joint selection of a state, and then each generates an outcome based on this state but independently of the outcome generated by the other process (at the same state). In light of the foregoing, it is natural that the proof of Theorem 3.2 boils down to lower-bounding the min-entropies of the X_i's and Y_i's (for all i's with sufficiently large p_i's).

4 Dependence as cross influence

This model resides in between the two other models considered in the current work. This fact is reflected both by the results that directly compare it to the other two models, and by results obtained for randomness extraction (and condensing) in the various models.

4.1 Definition and a key observation

We view each of the two sources in this model as being generated based mainly on its own private randomness, but also based in a "bounded manner" on the randomness of the other source. Hence, a crucial question is how to measure the influence of (the randomness of) one source on the output of the other source. We measure this influence in terms of the probability that the output of one source changes when the other source's randomness is re-randomized. Specifically, the influence is t' if the probability that the outcome changes is at most $1 - 2^{-t'}$. The sum of the two opposite influences is called the *cross influence*.

We represent the output of the i^{th} source by $G_i(s_1, s_2)$, where s_j is the private randomness of the j^{th} source. That is, G_i is a function describing how the outcome of the i^{th} source is generated, and the generation of joint distribution is described by $G = (G_1, G_2)$. For $s = (s_1, s_2)$, we consider $G(s) = G(s_1, s_2) = (G_1(s_1, s_2), G_2(s_1, s_2))$. For sake of simplicity, we assume that $s_1, s_2 \in \{0,1\}^\ell$, for some $\ell \in \mathbb{N}$ (which may be shorter or longer or equal to $|G_i(s_1, s_2)|$).

Definition 4.1 (generating joint distributions with limited cross influence): *We say that the joint distribution (X, Y) is* generated with at most t bits of cross influence *if there exists a function G such that the following two conditions hold* (for some ℓ)*:*

1. *G generates (X, Y) using a 2ℓ-bit long random seed: That is, $(X, Y) \equiv G(U_{2\ell})$, which means that $\Pr[(X, Y) = (x, y)] = \Pr_{u \in \{0,1\}^{2\ell}}[G(u) = (x, y)]$ holds, for every (x, y).*
2. The influence of each half of the seed on the other part of the outcome is bounded: *Letting G_i denote the i^{th} part of the output of G* (i.e., $G(s) = (G_1(s), G_2(s))$)*, there exists $t_1 \geq 0$ such that, for every $u, v \in \{0,1\}^\ell$, it holds that*

$$\Pr_{v' \in \{0,1\}^\ell}[G_1(u, v') \neq G_1(u, v)] \leq 1 - 2^{-t_1}, \tag{3}$$

$$\Pr_{u' \in \{0,1\}^\ell}[G_2(u', v) \neq G_2(u, v)] \leq 1 - 2^{-(t-t_1)}. \tag{4}$$

Equivalently, for every $u_1, v_1, u_2, v_2 \in \{0,1\}^\ell$, it holds that

$$\Pr_{r \in \{0,1\}^\ell}[G_1(u_1, r) = G_1(u_1, v_1)] \cdot \Pr_{r \in \{0,1\}^\ell}[G_2(r, v_2) = G_2(u_2, v_2)] \geq 2^{-t}. \tag{5}$$

The corresponding model, denoted $\mathtt{CRI}_n(k, t)$, consists of all joint distribution over $\{0,1\}^{n+n}$ that can be generated with at most t bits of cross influence such that each of the n-bit long sources has min-entropy at least k.

Indeed, Eq. (3) (resp., Eq. (4)) captures a worst-case notion of the amount of influence of the second (resp., first) part of the seed on the first (resp., second) part of the outcome.[14] The postulate that the seed of G is a 2ℓ-bit long string is made for simplicity of notation; in general, we may consider a probability space $\Omega_1 \times \Omega_2$ for the two parts of the seed.

A key observation. The following fact plays a key role in our study of joint distributions with low cross influence. It asserts that the min-entropy of such joint distributions is not much smaller than the sum of the min-entropies of their two parts. (Recall that the min-entropy of Z is $\min_z\{\log_2(1/\Pr[Z = z])\}$.)[15]

[14] Indeed, $\mathrm{E}_{u \in \Omega}[\Pr_{v,v' \in \Omega'}[F(u, v) \neq F(u, v')]]$, which equals $\mathrm{E}_{(u,v) \in \Omega \times \Omega'}[\Pr_{v' \in \Omega'}[F(u, v) \neq F(u, v')]]$, is the standard (average-case) notion of the influence of the second argument of F on its outcome. The reason we use the foregoing worst-case notion instead is discussed in Section 4.3.1.

[15] Indeed, Eq. (6) can be written as

$$\log_2(1/\Pr[(X, Y) = (x, y)]) \geq \log_2(1/\Pr[X = x]) + \log_2(1/\Pr[Y = y]) - t.$$

Since this holds for all (x, y)'s, it holds for a pair for which the l.h.s. is minimized (whereas the r.h.s. may be minimized on a different pair).

Proposition 4.2 (low cross influence implies small loss in min-entropy): *Suppose that the joint distribution (X,Y) can be generated with at most t bits of cross influence. Then, for every x and y, it holds that*

$$\Pr[(X,Y)=(x,y)] \leq 2^t \cdot \Pr[X=x] \cdot \Pr[Y=y]. \tag{6}$$

In other words, the min-entropy of (X,Y) is at most t units smaller than the sum of the min-entropies of X and Y. Equivalently, the min-entropy of X conditioned on $Y=y$ is at most t units smaller than the min-entropy of X (equiv., $\Pr[X=x|Y=y] \leq 2^t \cdot \Pr[X=x]$ for every x and y).

Proof: Intuitively, the bound on the cross influence between the two parts of the process $G = (G_1, G_2)$ that generates (X,Y) allows us to re-randomize the second part of the seed fed to G_1 without increasing the probability of any specific outcome by too much. The same holds for re-randomizing the first part of the seed fed to G_2. Once both re-randomizations are performed, the G_i's are fed by independently distributed seeds, and so they produce independently distributed copies of X and Y. Hence, the bound on the cross influence is translated to a bound on the ratio between $\Pr[(X,Y)=(x,y)]$ and $\Pr[X=x] \cdot \Pr[Y=y]$. Details follow.

Let $G = (G_1, G_2)$ be the generator guaranteed by Definition 4.1. Then, for any x and y, it holds that $\Pr_{s\in\{0,1\}^{2\ell}}[G(s)=(x,y)]$ equals

$$\begin{aligned}
&\Pr_{u_1,u_2\in\{0,1\}^\ell}\left[(G_1(u_1,u_2),G_2(u_1,u_2))=(x,y)\right] \\
&= \frac{\Pr_{u_1,u_1',u_2\in\{0,1\}^\ell}\left[(G_1(u_1,u_2),G_2(u_1,u_2))=(x,y)\ \&\ G_2(u_1',u_2)=G_2(u_1,u_2)\right]}{\Pr_{u_1,u_1',u_2\in\{0,1\}^\ell}\left[G_2(u_1',u_2)=G_2(u_1,u_2)\mid (G_1(u_1,u_2),G_2(u_1,u_2))=(x,y)\right]} \\
&= \frac{\Pr_{u_1,u_1',u_2\in\{0,1\}^\ell}\left[(G_1(u_1,u_2),G_2(u_1',u_2))=(x,y)\ \&\ G_2(u_1',u_2)=G_2(u_1,u_2)\right]}{\Pr_{u_1,u_1',u_2\in\{0,1\}^\ell}\left[G_2(u_1',u_2)=G_2(u_1,u_2)\mid (G_1(u_1,u_2),G_2(u_1,u_2))=(x,y)\right]} \\
&\leq \frac{\Pr_{u_1,u_1',u_2\in\{0,1\}^\ell}\left[(G_1(u_1,u_2),G_2(u_1',u_2))=(x,y)\right]}{\min_{u_1,u_2\in\{0,1\}^\ell:(G_1(u_1,u_2),G_2(u_1,u_2))=(x,y)}\left\{\Pr_{u_1'\in\{0,1\}^\ell}[G_2(u_1',u_2)=G_2(u_1,u_2)]\right\}}
\end{aligned}$$

where the first two equalities are merely trivialities that are hard to follow due to the large number of symbols.[16] Note that in the numerator of the last expression G_1 and G_2 obtain seeds that are identical only in half their length, whereas the denominator is related to the influence of the first argument of G_2 on the value of G_2. Applying the same reasoning again (but on the other half), we will obtain independently distributed seeds. Specifically, we have

$$\begin{aligned}
&\Pr_{u_1,u_1',u_2\in\{0,1\}^\ell}\left[(G_1(u_1,u_2),G_2(u_1',u_2))=(x,y)\right] \\
&= \frac{\Pr_{u_1,u_1',u_2,u_2'\in\{0,1\}^\ell}\left[(G_1(u_1,u_2'),G_2(u_1',u_2))=(x,y)\ \&\ G_1(u_1,u_2')=G_1(u_1,u_2)\right]}{\Pr_{u_1,u_1',u_2,u_2'\in\{0,1\}^\ell}\left[G_1(u_1,u_2')=G_1(u_1,u_2)\mid (G_1(u_1,u_2),G_2(u_1',u_2))=(x,y)\right]} \\
&\leq \frac{\Pr_{u_1,u_1',u_2,u_2'\in\{0,1\}^\ell}\left[(G_1(u_1,u_2'),G_2(u_1',u_2))=(x,y)\right]}{\min_{u_1,u_1',u_2\in\{0,1\}^\ell:(G_1(u_1,u_2),G_2(u_1',u_2))=(x,y)}\left\{\Pr_{u_2'\in\{0,1\}^\ell}[G_1(u_1,u_2')=G_1(u_1,u_2)]\right\}}
\end{aligned}$$

Hence,

$$\Pr_{s\in\{0,1\}^{2\ell}}\left[G(s)=(x,y)\right]$$

[16] The first equality merely uses $\Pr[A] = \frac{\Pr[A\ \&\ B]}{\Pr[B\mid A]}$, whereas the second equality uses $\Pr[E(Z)\ \&\ Z=Z'] = \Pr[E(Z')\ \&\ Z=Z']$.

$$
\begin{aligned}
&\leq \frac{\Pr_{u_1,u'_1,u_2,u'_2\in\{0,1\}^\ell}\left[(G_1(u_1,u'_2),G_2(u'_1,u_2))=(x,y)\right]}{\min_{u_1,v_1,u_2,v_2\in\{0,1\}^\ell}\left\{\Pr_{r\in\{0,1\}^\ell}[G_2(r,v_2)=G_2(v_1,v_2)]\cdot\Pr_{s\in\{0,1\}^\ell}[G_1(u_1,s)=G_1(u_1,u_2)]\right\}}\\
&\leq \frac{\Pr_{u_1,u'_1,u_2,u'_2\in\{0,1\}^\ell}\left[(G_1(u_1,u'_2),G_2(u'_1,u_2))=(x,y)\right]}{2^{-t}}\\
&= 2^t\cdot\Pr_{u_1,u'_2\in\{0,1\}^\ell}\left[G_1(u_1,u'_2)=x\right]\cdot\Pr_{u'_1,u_2\in\{0,1\}^\ell}\left[G_2(u'_1,u_2)=y\right]
\end{aligned}
$$

where the second inequality is due to the hypothesis that G generates (X,Y) with at most t bits of cross influence (see Eq. (5)), and the equality is due to the independence of (u_1,u'_2) and (u'_1,u_2). The claim follows. ■

4.2 Extraction and condensing

In Section 4.2.1 we show that while standard extractors may fail to extract almost perfect randomness from joint distributions of very low cross influence, they do condense (in this model) with deficiency linearly related to the cross influence bound. In Section 4.2.2 we present lower bounds on the deficiency of any attempt to condense in the cross influence model, demonstrating that the results of Section 4.2.1 are optimal up to a constant factor.

4.2.1 Standard extractors fail as such but do condense

It would have been great if Theorem 3.2 could have been extended to joint distributions of low cross influence. Unfortunately, this is not the case. We first observe that standard two-source extractors may fail miserably even when the amount of cross influence is very small (e.g., one bit) and the min-entropy is very high.

Proposition 4.3 (failure of a standard extractor when applied to sources of very low cross influence): *There exists a function* $\mathtt{EXT}:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ *such that the following two condition hold:*

1. $\mathtt{EXT}$ extracts from pairs of independent sources: *The function* $\mathtt{EXT}$ *is an extractor with error* $\exp(-\Omega(n))$ *for pairs of independent sources that have min-entropy at least* $0.51n$ *each.*
2. $\mathtt{EXT}$ fails on some pairs of low cross influence: *There exists a joint distribution* (X,Y) *that can be generated with at most one bit of cross influence such that both* X *and* Y *have min-entropy at least* $n-1$*, but* $\mathtt{EXT}(X,Y)\equiv 0$.

We stress that Proposition 4.3 does *not* say that one cannot extract from any pair of sources having low cross influence and high min-entropy; it only says that one cannot use an arbitrary standard two-source extractor (for independent sources) towards this end. Indeed, the begging question is whether there are other ways. A negative answer to this question appears as Proposition 4.6 (albeit Proposition 4.6 asserts only a joint distribution (X,Y) with six bits of cross influence such that $|\Pr[\mathtt{EXT}(X,Y)=0]-0.5|\geq 0.25$).

Proof: We use the inner-product (mod 2) function, denoted $\mathtt{IP}_2$, in the role of the standard two-source extractor, $\mathtt{EXT}$, and recall that Part 1 is proved in [6, Thm. 9]. Towards proving Part 2, we take the joint distribution (X,Y) such that $X=(X',1)$ and $Y=(Y',\mathtt{IP}_2(X',Y'))$, where (X',Y') is uniformly distributed in $\{0,1\}^{n-1}\times\{0,1\}^{n-1}$. On the one hand, X and Y have each

min-entropy at least $n-1$, the randomness used to generate Y (i.e., Y') has no influence on X, and the randomness used to generate X (i.e., X') has at most one bit of influence on Y (i.e., for every $(x',y') \in \{0,1\}^{n-1} \times \{0,1\}^{n-1}$, it holds that $\Pr_{r\in\{0,1\}^{n-1}}[(y', \mathtt{IP}_2(r,y')) = (y', \mathtt{IP}_2(x',y'))] \geq 1/2$).[17] On the other hand,

$$\mathtt{EXT}(X,Y) = \mathtt{IP}_2(X,Y) \equiv \sum_{i\in[n]} X_i Y_i \equiv \mathtt{IP}_2(X',Y') + 1 \cdot \mathtt{IP}_2(X',Y') \equiv 0 \pmod 2$$

and the claim follows. ■

Digest. The joint distribution used in the proof of Proposition 4.3 seems to represent a "very mild form of dependence" between sources, and one would want to extract randomness in such a case. While Proposition 4.3 only asserts that this cannot be done in a straightforward manner (i.e., by using an arbitrary standard extractor), Proposition 4.6 asserts that extraction from joint distribution of low cross influence is impossible in general.

In light of this fact, we lower our goals and aim for only condensing such joint distributions. This less ambitious task turns out to be achievable. Specifically, we show that any standard extractor constitutes a good condenser for joint distributions with low cross influence, where the deficiency of the condenser is upper-bounded by the cross influence of the sources. Actually, a stronger statement holds.

Theorem 4.4 (condensers for sources with low cross influence): *Let* $\mathtt{CND} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *be a condenser with error* ϵ *and min-entropy deficiency* d *for pairs of independent sources that have min-entropy at least* k *each; that is, for every such pair* (X,Y) *it holds that* $\mathtt{CND}(X,Y)$ *is* ϵ*-close to a distribution that has min-entropy at least* $m-d$*. Then, for every* (X,Y) *that is generated with at most* t *bits of cross influence such that both* X *and* Y *have min-entropy at least* k*, it holds that* $\mathtt{CND}(X,Y)$ *is* $2^t \cdot \epsilon$*-close to a distribution that has min-entropy at least* $m-d-t$*. That is,* CND *is a condenser with error* $2^t \cdot \epsilon$ *and deficiency* $d+t$ *for sources of cross influences* t *and min-entropy at least* k*.*

That is, if CND is a condenser of error ϵ and deficiency d for the model $\mathtt{STD}_n(k)$, then it is a condenser of error $2^t \cdot \epsilon$ and deficiency $d+t$ for the model $\mathtt{CRI}_n(k,t)$.

Proof: By Proposition 4.2, we have for every $z \in \{0,1\}^m$,

$$\begin{aligned}
\Pr[\mathtt{CND}(X,Y)=z] &= \sum_{x,y:\mathtt{CND}(x,y)=z} \Pr[(X,Y)=(x,y)] \\
&\leq \sum_{x,y:\mathtt{CND}(x,y)=z} 2^t \cdot \Pr[X=x] \cdot \Pr[Y=y] \\
&= 2^t \cdot \Pr[\mathtt{CND}(X',Y')=z],
\end{aligned}$$

where X' and Y' are two independent random variables such that $X' \equiv X$ and $Y' \equiv Y$. Since X' and Y' are two independent sources each of min-entropy at least k, it follows that $\mathtt{CND}(X',Y')$ is ϵ-close to a distribution with min-entropy at least $m-d$, denoted W'. Note that if $\mathtt{CND}(X',Y')$ itself had min-entropy at least $m-d$ (i.e., $\mathtt{CND}(X',Y') \equiv W'$), then it would follow that $\mathtt{CND}(X,Y)$

[17] Indeed, $\Pr_r[(y', \mathtt{IP}_2(r,y')) = (y', \mathtt{IP}_2(x',y'))]$ equals $1/2$ if $y \neq 0^{n-1}$, and equals 1 otherwise.

has min-entropy at least $m-d-t$. But, in general, $\mathtt{CND}(X',Y')$ is only ϵ-close to W', and the rest of the proof is devoted to showing that in this case $\mathtt{CND}(X,Y)$ is $2^t\cdot\epsilon$-close to a distribution having min-entropy at least $m-d-t$.

Intuitively, z's that are assigned less or equal weight under $\mathtt{CND}(X',Y')$ in comparison to their weight under W' (i.e., $\Pr[\mathtt{CND}(X',Y')=z]\leq\Pr[W'=z]$) pose no problem since their weight under $\mathtt{CND}(X,Y)$ is at most $2^t\cdot\Pr[\mathtt{CND}(X',Y')=z]\leq 2^t\cdot\Pr[W'=z]\leq 2^t\cdot 2^{-(m-d)}=2^{-(m-d-t)}$. The other z's do pose a problem, but their total weight under $\mathtt{CND}(X,Y)$ is larger by at most an additive term of $2^t\cdot\epsilon$ than what it would have been if $\Pr[\mathtt{CND}(X',Y')=z]=\Pr[W'=z]$.

To formalize the foregoing intuition, we consider the pointwise difference between the distributions $\mathtt{CND}(X',Y')$ and W'; that is, we let

$$\delta(z)\stackrel{\rm def}{=}\Pr[\mathtt{CND}(X',Y')=z]-\Pr[W'=z],$$

and recall that $\sum_{z:\delta(z)>0}\delta(z)\leq\epsilon$. Now, for every z, it holds that

$$\begin{aligned}\Pr[\mathtt{CND}(X,Y)=z] &\leq 2^t\cdot\Pr[\mathtt{CND}(X',Y')=z]\\ &= 2^t\cdot(\Pr[W'=z]+\delta(z))\\ &\leq 2^t\cdot 2^{-(m-d)}+2^t\cdot\delta(z),\end{aligned}$$

where the last inequality uses the min-entropy bound of W'. Hence, $\Pr[\mathtt{CND}(X,Y)=z]\leq 2^{-m(m-d-t)}$ for every z such that $\delta(z)\leq 0$, whereas for the remaining z's (i.e., z such that $\delta(z)>0$) we only have $\Pr[\mathtt{CND}(X,Y)=z]\leq 2^{-m(m-d-t)}+2^t\cdot\delta(z)$.

Recalling that $\sum_{z:\delta(z)>0}\delta(z)\leq\epsilon$, we present a distribution W of min-entropy at least $m-(d+t)$ that is $2^t\cdot\epsilon$-close to $\mathtt{CND}(X,Y)$ by moving probability mass that is assigned to z's that violate the probability bound (i.e., $\Pr[\mathtt{CND}(X,Y)=z]>2^{d+t-m}$) to other z's. Indeed, W is obtained by moving probability mass from z's that have an excess of $\Pr[\mathtt{CND}(X,Y)=z]-2^{d+t-m}>0$ to the other z's, without violating the probability bound on the latter, where the moved mass is fully accounted for in $2^t\cdot\delta(z)$. Specifically, we reduce the probability mass of z only if $\Pr[\mathtt{CND}(X,Y)=z]>2^{d+t-m}$, and in this case $\Pr[\mathtt{CND}(X,Y)=z]-2^{d+t-m}\leq 2^t\cdot\delta(z)$. We can move this mass to z's that satisfy $\delta(z)<0$ without assigning any of these z's more than weight 2^{d+t-m}, because the weight assigned to each of these z's by $\mathtt{CND}(X,Y)$ is at most $2^{d+t-m}+2^t\cdot\delta(z)$ (where $\delta(z)<0$).[18] ■

On the optimality of Theorem 4.4. It turns out that the min-entropy loss suffered by a standard condenser when applied to a joint distribution with t bits of cross influence is inevitable. This holds even if the condenser is actually an extractor. (Proposition 4.7 will provide a far more general result that refers to any condensing method, but with slightly less tight parameters.)

Proposition 4.5 (Proposition 4.3, generalized): *There exists a function* $\mathtt{EXT}:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^{0.04n}$ *such that the following two condition hold:*

1. $\mathtt{EXT}$ extracts from pairs of independent sources: *The function* $\mathtt{EXT}$ *is an extractor with error* $\exp(-\Omega(n))$ *for pairs of independent sources that have min-entropy at least* $0.6n$ *each.*

2. When applied to a joint distribution with cross influece t, the output of $\mathtt{EXT}$ may have a min-entropy deficiency of t: *For every* $t\leq 0.04n$*, there exists a joint distribution* (X,Y) *that can*

[18] This uses $\sum_{z:\delta(z)>0}\delta(z)=-\sum_{z:\delta(z)<0}\delta(z)$.

be generated with at most t bits of cross influence such that both X and Y have min-entropy at least $0.9n$, but the first t bits of $\mathtt{EXT}(X,Y)$ are identically zero.

The choice of the various constants in Proposition 4.5 is quite arbitrary: we can replace 0.6 by any constant $\kappa > 0.5$, replace 0.04 by any constant $\mu > 0$ smaller than $\kappa - 0.5$, and replace 0.9 by any constant smaller than $1 - 2 \cdot \mu$.

Proof: We generalized the proof of Proposition 4.3, while using one of the (inner-product (mod 2) based) extractors presented in [7]. Specifically, we use the multi-shifts inner-product (mod 2) extractor, denoted $\mathtt{EXT}^{(\mathtt{mip})} : \{0,1\}^{n+n} \to \{0,1\}^m$, whose i^{th} output bit, denoted $\mathtt{EXT}_i^{(\mathtt{mip})}(x,y)$, equals $\sum_{j\in[n-i+1]} x_j y_{(i-1)+j} \bmod 2$. Recall that $\mathtt{EXT}^{(\mathtt{mip})} : \{0,1\}^{n+n} \to \{0,1\}^m$ has error at most $2^{-(k-0.5n-m)}$ for pairs of independent sources that have min-entropy at least k each. This establishes Part 1.

Towards proving Part 2, we let $m = 0.04n$ and define functions $f_1, ..., f_t : \{0,1\}^{n-(2m-1)} \times \{0,1\}^{n-m} \to \{0,1\}$ such that for every $i \in [t]$ it holds that $f_i(x',y') \stackrel{\text{def}}{=} \mathtt{EXT}_i^{(\mathtt{mip})}(x'0^{2m-1}, y'0^m)$, and denote their concatenation by $f(x',y') = (f_1(x',y'), ..., f_t(x',y'))$. We consider the joint distribution (X,Y) such that $X = (X'0^{m-1}, 10^{m-1})$ and $Y = (Y', f(X',Y')0^{m-t})$, where (X',Y') is uniformly distributed in $\{0,1\}^{n-(2m-1)} \times \{0,1\}^{n-m}$. On the one hand, X and Y have each min-entropy at least $n - (2m-1) > 0.9n$, the randomness used to generate Y (i.e., Y') has no influence on X, and the randomness used to generate X (i.e., X') influences at most t bits of Y (i.e., for every $(x',y') \in \{0,1\}^{n-(2m-1)} \times \{0,1\}^{n-m}$, it holds that $\Pr_{r\in\{0,1\}^{n-(2m-1)}}[(y', f(r,y')0^{m-t}) = (y', f(x',y')0^{m-t})] \geq 2^{-t}$).[19] On the other hand, we shall show that the t-bit prefix of $\mathtt{EXT}^{(\mathtt{mip})}(X,Y)$ always equals 0^t. First, observe that for every $x = (x'0^{m-1}, x'') \in \{0,1\}^{n-m} \times \{0,1\}^m$ and $y = (y', y'') \in \{0,1\}^{n-m} \times \{0,1\}^m$, and every $i \in [m]$, it holds that

$$\begin{aligned}
\mathtt{EXT}_i^{(\mathtt{mip})}(x'0^{m-1}x'', y'y'') &= \sum_{j\in[|x'|]} x_j y_{(i-1)+j} + \sum_{j\in[|x'|+m,n-i+1]} x_j y_{(i-1)+j} \\
&= \mathtt{EXT}_i^{(\mathtt{mip})}(x'0^{(m-1)+|x''|}, y'0^{|y''|}) + \mathtt{EXT}_i^{(\mathtt{mip})}(0^{|x'|+m-1}x'', 0^{|y'|}y''),
\end{aligned}$$

since $\sum_{j\in[|x'|+1,|x'|+m-1]} x_j y_{(i-1)+j}$ is identically zero (because $x_j = 0$ for any $j \in [|x'|+1, |x'|+m-1]$). Hence (recalling that $X = (X', 0^{m-1}, 10^{m-1})$ and $Y = (Y', f(X',Y')0^{m-t})$), for every $i \in [t]$, we have

$$\begin{aligned}
\mathtt{EXT}_i^{(\mathtt{mip})}(X,Y) &= \mathtt{EXT}_i^{(\mathtt{mip})}(X'0^{(m-1)+m}, Y'0^m) \\
&\quad + \mathtt{EXT}_i^{(\mathtt{mip})}(0^{n-(2m-1)+(m-1)}10^{m-1}, 0^{n-m}f(X',Y')0^{m-t}).
\end{aligned} \tag{7}$$

Now, using the definition of $\mathtt{EXT}_i^{(\mathtt{mip})}$, and denoting the j^{th} bit of α by $\mathtt{bit}_j(\alpha)$, we have

$$\begin{aligned}
&\mathtt{EXT}_i^{(\mathtt{mip})}(0^{n-m}10^{m-1}, 0^{n-m}f(X',Y')0^{m-t}) \\
&\quad = \sum_{j\in[n-i+1]} \mathtt{bit}_j(0^{n-m}10^{m-1}) \cdot \mathtt{bit}_{(i-1)+j}(0^{n-m}f(X',Y')0^{m-t}) \\
&\quad = \mathtt{bit}_{(i-1)+(n-m+1)}(0^{n-m}f(X',Y')0^{m-t})
\end{aligned}$$

[19] To lower-bound $\Pr_r[f(r,y')) = f(x',y')]$, we observe that, for every $y' \in \{0,1\}^{n-m}$ and $x' \in \{0,1\}^{n-(2m-1)}$, the set $\{r \in \{0,1\}^{|x'|} : f(r,y') = f(x',y')\}$ is a linear subspace of dimention at least $|x'| - t$.

which equals $f_i(X',Y') = \mathtt{EXT}_i^{(\mathtt{mip})}(X'0^{2m-1}, Y'0^m)$. Hence,

$$\mathtt{EXT}_i^{(\mathtt{mip})}(0^{n-m}10^{m-1}, 0^{n-m}f(X',Y')0^{m-t}) = \mathtt{EXT}_i^{(\mathtt{mip})}(X'0^{2m-1}, Y'0^m). \quad (8)$$

Combining Eq. (7) with Eq. (8), it follows that $\mathtt{EXT}_i^{(\mathtt{mip})}(X,Y) = 0$ (for every $i \in [t]$), and the Part 2 follows. ■

4.2.2 General imposibility of extraction and limits to condensing

Recall that Proposition 4.3 asserted that some standard extractors may fail to extract from a joint distribution generated with a single bit of cross influence, whereas Proposition 4.5 asserted that the deficiency bound of the generic condenser of Theorem 4.4 is optimal. In both cases, this was shown by tailoring a specific joint distribution (of low cross influence) to a specific standard extractor. This raises the question of whether some other standard extractors can do better, let alone whether one can get lower deficiency by using a function that is not a standard extractor (which is unlikely and yet *a priori* possible). We answer these questions negatively, modulo a small constant factor. (Due to the non-tightness of the following two results, they do not fully superseed Propositions 4.3 and 4.5.)

The general impossibility of extraction. In contrast to Proposition 4.3, which only asserts the failure of some standard two-source extractors in the context of low cross influence, the following result asserts the failure of any function to extract almost-perfect randomness in this context. In fact, we show that one cannot even extract a single bit that is not significantly biased.

Proposition 4.6 (general impossibility of extraction from joint distribution with low cross influence): *For any $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, there exists a joint distribution (X,Y) having at most six bits of cross influence such that both X and Y have min-entropy at least $n-4$, but $\Pr[F(X,Y)=\sigma] \geq 3/4$ for some $\sigma \in \{0,1\}$* (equiv., the bias of $F(X,Y)$ is at least $1/2$).[20]

Note that Proposition 4.6 only asserts that F fails to extract a bit that has *biased smaller than* $1/2$, when fed with a joint distribution having six bits of cross influence (and very high min-entropy). Recall that Theorem 4.4 asserts that in such a case it is possible to output a 7-bit long string such that no outcome appears with probability higher than 0.51. (In general, Theorem 4.4 asserts that condensing with deficiency that equals the cross influence bound is possible, and in Proposition 4.7 we show that this is optimal up to a constant factor.)

Proof: Suppose that F is more likely to evaluate to σ; that is, $F^{-1}(\sigma) = \{(x,y) : F(x,y)=\sigma\}$ has size at least 2^{2n-1}. Then, one may be tempted to define (X,Y) to be uniform on $F^{-1}(\sigma)$, but it is not clear how to generate this distribution using few bits of cross influence. The first idea that comes to mind is to use the generator G that is given a sequence of random pairs in $\{0,1\}^{n+n}$ and outputs the first pair in $F^{-1}(\sigma)$. That is, for some parameter $m = \ell/n$, let $G((u_1,u_2,...,u_m),(v_1,v_2,...,v_m)) = (u_i,v_i)$ if $F(u_i,v_i) = \sigma$ and $F(u_j,v_j) \neq \sigma$ for every $j \in [i-1]$.

Ignoring the issue of what happens if none of the pairs is in $F^{-1}(\sigma)$, observe that G does not have low cross influence (in the worst case). Specifically, suppose that $G((u_1,u_2,...,u_m),(v_1,v_2,...,v_m)) = (u_i,v_i)$. Then, the probability that a re-randomization of all u_j's (resp., all v_j's) yields a sequence

[20] The bias of a Boolean random variable ζ is $|\Pr[\zeta=0] - \Pr[\zeta=1]|$.

that generates $(\cdot, v_i)$ (resp., $(u_i, \cdot)$) is at most $2^{-(i-1)} + 2^{-n}$, where the first term bounds the probability that the i^{th} pair in the re-randomized seed is the first pair in $F^{-1}(\sigma)$. This suggest to modify the generation process by setting $m = 2$ and having the generator output the second pair (regardless of whether it is in $F^{-1}(\sigma)$ or not) unless the first pair is in $F^{-1}(\sigma)$.

Another issue that we will need to address is the possibility that for some u's (resp., v's) there may be too few or too many v's (resp., u's) that yield a pair in $F^{-1}(\sigma)$. This is a problem, since when we re-randomize v (resp., u), while keeping u (resp., v) intact, we hope to get a pair that maintains the value of $F(u, v)$. We start the actual proof by addressing this issue.

We first show that one may focus on the case that for at least $7/8$ of the x's (resp., y's) it holds that $\Pr_{y\in\{0,1\}^n}[F(x,y)=\sigma] \geq 1/4$ (resp., $\Pr_{x\in\{0,1\}^n}[F(x,y)=\sigma] \geq 1/4$) for both $\sigma \in \{0,1\}$. If this is not the case, then F fails even as a standard two-source extractor for min-entropy $n-4$ (and error $1/4$). Specifically, suppose that for a set B of at least $2^n/16$ of the x's it holds that $\Pr_y[F(x,y)=1] < 1/4$. Then, defining X to be uniform on B and Y as uniform on $\{0,1\}^n$ (and independent of X), we get $\Pr[F(X,Y)=0] > 3/4$. The same holds for $\Pr_y[F(x,y)=0] < 1/4$, and ditto for the y's. Having established the forgoing claim, we define

$$\begin{aligned} R &\stackrel{\text{def}}{=} \{x \in \{0,1\}^n : 0.25 \leq \Pr_{y\in\{0,1\}^n}[F(x,y)=1] \leq 0.75\} \\ S &\stackrel{\text{def}}{=} \{y \in \{0,1\}^n : 0.25 \leq \Pr_{x\in\{0,1\}^n}[F(x,y)=1] \leq 0.75\}, \end{aligned}$$

and note that $|R|, |S| \geq \frac{7}{8} \cdot 2^n$. Furthermore, for every $x \in R$ it holds that $\Pr_{y\in S}[F(x,y)=1] \in [1/8, 7/8]$ (resp., for every $y \in S$ it holds that $\Pr_{x\in R}[F(x,y)=1] \in [1/8, 7/8]$).[21]

Now, let $\sigma \in \{0,1\}$ be such that $p \stackrel{\text{def}}{=} \Pr_{(x,y)\in R\times S}[F(x,y)=\sigma] \geq 1/2$, and consider the joint distribution (X,Y) defined by a generator $G = (G_1, G_2)$ that, on input $((r_1,r_2),(s_1,s_2)) \in R^2 \times S^2$, outputs (r_1, s_1) if $F(r_1, s_1) = \sigma$ and outputs (r_2, s_2) otherwise; that is, $G_1((r_1,r_2),(s_1,s_2)) = r_1$ if $F(r_1,s_1)=\sigma$ and equals r_2 otherwise.

Then, $\Pr[F(X,Y)=\sigma] = p + (1-p)\cdot p \geq 3/4$, where the first (resp., second) term corresponds to the case that $F(r_1,s_1)=\sigma$ (resp., $F(r_1,s_1)\neq\sigma$). Next, note that the min-entropy of X is at least $n-2$, since

$$\begin{aligned} \Pr[X=x] &= \Pr_{(r_1,s_1)\in R\times S}[F(r_1,s_1)=\sigma \ \&\ r_1=x] + \Pr_{(r_1,r_2,s_1)\in R^2\times S}[F(r_1,s_1)\neq\sigma \ \&\ r_2=x] \\ &\leq \Pr_{r_1\in R}[r_1=x] + \Pr_{r_2\in R}[r_2=x], \end{aligned}$$

which equals $2/|R| < 2^{-(n-2)}$, since $|R| > 2^{n-1}$. Ditto for Y. Lastly, we get to the crux of the argument, which is showing that the cross influence of (X,Y) is upper-bounded by six. Consider an arbitrary input to G, denoted $((r_1,r_2),(s_1,s_2)) \in R^2 \times S^2$, and the following two cases.

Case 1: $F(r_1,s_1)=\sigma$**.** In this case $G_1((r_1,r_2),(s_1,s_2))=r_1$, and so

$$\begin{aligned} \Pr_{(s_1',s_2')\in S^2}[G_1((r_1,r_2),(s_1',s_2'))=G_1((r_1,r_2),(s_1,s_2))] &= \Pr_{s_1',s_2'\in S}[G_1((r_1,r_2),(s_1',s_2'))=r_1] \\ &\geq \Pr_{s_1'\in S}[F(r_1,s_1')=\sigma] \\ &\geq 1/8. \end{aligned}$$

[21] This uses $\Pr_{y\in S}[F(x,y)=1] \geq \Pr_{y\in\{0,1\}^n}[F(x,y)=1] - \Pr_{y\in\{0,1\}^n}[y\notin S]$ and $\Pr_{y\in S}[F(x,y)=1] \leq \Pr_{y\in\{0,1\}^n}[F(x,y)=1]/\Pr_{y\in\{0,1\}^n}[y\in S]$.

Case 2: $F(r_1,s_1)\neq\sigma$**.** In this case $G_1((r_1,r_2),(s_1,s_2))=r_2$, and so

$$\begin{aligned}\Pr_{(s'_1,s'_2)\in S^2}[G_1((r_1,r_2),(s'_1,s'_2))=G_1((r_1,r_2),(s_1,s_2))] &= \Pr_{s'_1,s'_2\in S}[G_1((r_1,r_2),(s'_1,s'_2))=r_2]\\ &\geq \Pr_{s'_1\in S}[F(r_1,s'_1)\neq\sigma]\\ &\geq 1/8.\end{aligned}$$

Hence, the influence of the second part of the seed on G_1 is at most three units. The same holds for G_2, and the claim follows. ■

A general limit to condensing. Generalizing the ideas that underlie the proof of Proposition 4.6, we show that joint distributions of low cross influence cannot be condensed much better than the result of Theorem 4.4. Specifically, we show that the deficiency of any potential condenser of such distributions is at least linear in the cross influence.

Proposition 4.7 (lower-bounding the deficiency of condensers as a function of the cross influence of the sources): *For any $F:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^m$ and every $t\in[8,m]$, there exists a joint distribution (X,Y) having at most t bits of cross influence such that both X and Y have min-entropy at least $n-0.5t+2$, but there exists a set H of density at most $2^{-0.5t+4}$ such that $\Pr[F(X,Y)\in H]\geq 2^{-0.25t-2}$.*

Hence, H witnesses the fact that $F(X,Y)$ has min-entropy at most

$$\begin{aligned}\min_{h\in H}\{\log_2(1/\Pr[F(X,Y)=h])\} &\leq \log_2(|H|/\Pr[F(X,Y)\in H])\\ &\leq (m-0.5t+4)-(-0.25t-2)\\ &= m-(0.25t-6),\end{aligned}$$

which implies that it has deficiency at least $0.25t-6$. Furthermore, any distribution W that is $2^{-0.25\cdot t-3}$-close to $F(X,Y)$ satisfies $\Pr[W\in H]\geq 2^{-0.25t-3}$, which implies that the min-entropy of W is smaller than $m-(0.25t-7)$. It follows that *there is no condenser with deficiency $0.25\cdot t-8$ and error $2^{-0.25\cdot t-3}$ for sources of cross influence t that each have min-entropy at least $n-0.5t+2$.* This means that the deficiency obtained by Theorem 4.4 is optimal up to a constant factor (of four), since combining it with standard extractors (of error 2^{-2t}) yields condensers with deficiency t and error 2^{-t} for sources of cross influence t (and min-entropy $m+4t+O(1)$).[22]

Proof: While the proof follows the high-level struction of the proof of Proposition 4.6, the details are much more complex in the current case. As in the proof of Proposition 4.6, things would have been simpler if the function F was "typical" in the sense that $\Pr_y[F(x,y)=v]=\Theta(2^{-m})$ for every v and x (resp., $\Pr_x[F(x,y)=v]=\Theta(2^{-m})$ for every v and y).[23] In such a case, for $t'=0.5\cdot t-O(1)$, we would have selected an arbitrary set $H\subset\{0,1\}^m$ of density $2^{-t'}$, and defined (X,Y) by considering

[22] Specifically, there exist extractors $\texttt{EXT}:\{0,1\}^{n+n}\to\{0,1\}^m$ of error $\epsilon>0$ for independent sources of min-entropy $m+2\log_2(1/\epsilon)+\log_2(9n)$ (see [6, Thm. 7(i)]). Explicit constructions of condensers with deficiency $o(\log(1/\epsilon)$ and error $\epsilon>0$ for independent sources of min-entropy $\min(m+O(\log(1/\epsilon)),\mathrm{poly}(\log(n/\epsilon)))$ are also known (see [3]), and combining Theorem 4.4 with them yields condensers with deficiency $t+o(t)$ and error 2^{-t} for sources of cross influence t (and min-entropy $\min(m+O(t),\mathrm{poly}(t\log n))$).

[23] Actually, it would have sufficed to let $F'(x,y)$ denote the t'-long prefix of $F(x,y)$ and assume that F' is typical (i.e., $\Pr_y[F'(x,y)=v]=\Theta(2^{-t'})$ etc).

the generator that, on input $((r_0, r_1, ..., r_{2^{t'}}), (s_0, s_1, ..., s_{2^{t'}}))$, outputs (r_i, s_i) if $i \in [2^{t'}]$ is the first index satisfying $F(r_i, s_i) \in H$ and (r_0, s_0) otherwise (i.e., if such an i does not exists). Intuitively, in this case, one can show that $\Pr[F(X,Y) \in H] = \Omega(1)$, whereas X (resp., Y) has min-entropy $n - t' - O(1)$ and (X,Y) has $2t' + O(1) = t$ bits of cross influence (since the influence of each part of the seed on the opposite output is at most $t' + O(1)$, where this claim relies on the simplifying assumption).

Getting rid of the simplifying assumption leads to various complications; in particular, this forces us to use $2^{t'/2}$ indices (i.e., i's) rather than $2^{t'}$, which means that we only prove that $\Pr[F(X,Y) \in H] = \Omega(2^{-t'/2})$. Under this setting, we are able to show that the corresponding distribution has $2t' + O(1)$ bits of cross influence. (In light of the foregoing, in the actual proof, we set $t' = (t - O(1))/4$; actually, we shall replace $2^{-t'}$ by ρ as representing the density of the set H.)

The actual proof: preliminaries. For $\rho = \Theta(2^{-t/2})$ (i.e., $\rho = 2^{-0.5 \cdot t + 4}$), let $H \subset \{0,1\}^m$ be a set of $\rho \cdot 2^m$ elements such that $\Pr_{x,y \in \{0,1\}^n}[F(x,y) \in H] \geq \rho$. We first show that one may focus on the case that for at least $1 - 0.25 \cdot \rho$ of the x's (resp., y's) it holds that $\Pr_{y \in \{0,1\}^n}[F(x,y) \in H] \leq \rho^{1/2}$ (resp., $\Pr_{x \in \{0,1\}^n}[F(x,y) \in H] \leq \rho^{1/2}$). If this is not the case, then F has deficiency at least $\log_2(\rho^{1/2}/\rho) = \Omega(t/4)$ even as a standard two-source condenser for min-entropy $n - \log_2(4/\rho) = n - 0.5t - O(1)$. Specifically, suppose that for a set B of at least $\rho \cdot 2^{n-2}$ of the x's it holds that $\Pr_{y \in \{0,1\}^n}[F(x,y) \in H] > \rho^{1/2}$. Then, defining X to be uniform on B and Y as uniform on $\{0,1\}^n$ (and independent of X), we get $\Pr[F(X,Y) \in H] > \rho^{1/2}$, whereas $\Pr_{r \in \{0,1\}^m}[r \in H] = \rho$. Hence, $F(X,Y)$ has min-entropy at most $\log_2(|H|/\rho^{1/2}) = m - \log_2(1/\rho^{1/2})$, which means deficiency at least $\log_2(1/\rho^{1/2})$. In this case the proposition follows for these independent sources (which have min-entropy at least $n - \log_2(4/\rho) = n - 0.5 \cdot t - O(1)$).

Having established that at least $1 - 0.25 \cdot \rho$ of the x's (resp., y's) satisfy $\Pr_y[F(x,y) \in H] \leq \rho^{1/2}$ (resp., $\Pr_x[F(x,y) \in H] \leq \rho^{1/2}$), we focus on these typical x's (resp., y's) and define the corresponding sets

$$\begin{aligned} R &\stackrel{\text{def}}{=} \{x \in \{0,1\}^n : \Pr_{y \in \{0,1\}^n}[F(x,y) \in H] \leq \rho^{1/2}\} \\ S &\stackrel{\text{def}}{=} \{y \in \{0,1\}^n : \Pr_{x \in \{0,1\}^n}[F(x,y) \in H] \leq \rho^{1/2}\}. \end{aligned}$$

Note that $|R| \cdot |S| \geq ((1 - 0.25 \cdot \rho) \cdot 2^n)^2 > (1 - 0.5 \cdot \rho) \cdot 2^{2n}$. Furthermore, for every $x \in R$ it holds that $p_x^{(0)} \stackrel{\text{def}}{=} \Pr_{y \in S}[F(x,y) \in H] < 2 \cdot \rho^{1/2}$, since $\Pr_{y \in \{0,1\}^n}[F(x,y) \in H] \leq \rho^{1/2}$ whereas $\Pr_{y \in \{0,1\}^n}[y \in S] > 1/2$. Likewise, for every $y \in S$ it holds that $q_y^{(0)} \stackrel{\text{def}}{=} \Pr_{x \in R}[F(x,y) \in H] < 2 \cdot \rho^{1/2}$. Lastly, we define $p^{(0)} \stackrel{\text{def}}{=} \mathrm{E}_{x \in R}[p_x^{(0)}] = \mathrm{E}_{y \in S}[q_y^{(0)}]$, and note that $p^{(0)} = \Pr_{(x,y) \in R \times S}[F(x,y) \in H] \geq \rho/2$, since $\Pr_{(x,y) \in \{0,1\}^{n+n}}[F(x,y) \in H] \geq \rho$ whereas $\Pr_{(x,y) \in \{0,1\}^{n+n}}[(x,y) \notin R \times S] \leq \rho/2$.

Recalling that $p_x^{(0)} < 2 \cdot \rho^{1/2}$ for every $x \in R$, note that if $p_x^{(0)} = \Omega(\rho)$ was also true (and ditto for the $q_y^{(0)}$'s), then we could proceed as outlined above, since all $p_x^{(0)}$'s would be approximately equal up to a factor of $O(\rho^{1/2})$, which we can afford to pay. Furthermore, x's satisfying $p_x^{(0)} = 0$ would cause no harm. Hence, we employ the following iterative process with the aim of obtaining corresponding $p_x^{(\cdot)}$'s (and $q_y^{(\cdot)}$'s) such that for every $x \in R$ either $p_x^{(\cdot)} = 0$ or $p_x^{(\cdot)} = \Omega(\rho)$ (and ditto for the $q_y^{(\cdot)}$'s). The $p_x^{(\cdot)}$'s (and $q_y^{(\cdot)}$'s) are modified by giving-up on x's (resp., y's) that violate the foregoing condition, and setting these probabilities to 0, which means that we don't take these x's (resp., y's) into account when considering the probability of hitting H.

To streamline the presentation, we define $\chi^{(0)}(x,y) = \texttt{true}$ if $F(x,y) \in H$ and $\chi^{(0)}(x,y) = \texttt{false}$ otherwise, and note that $p_x^{(0)} = \Pr_{y\in S}[\chi^{(0)}(x,y)]$ (resp., $q_y^{(0)} = \Pr_{x\in R}[\chi^{(0)}(x,y)]$). In the i^{th} iteration, we shall define $\chi^{(i)}$ such that $\chi^{(i)}(x,y)$ always implies $\chi^{(i-1)}(x,y)$, and define $p_x^{(i)} = \Pr_{y\in S}[\chi^{(i)}(x,y)]$ (resp., $q_y^{(i)} = \Pr_{x\in R}[\chi^{(i)}(x,y)]$). Note that $\chi^{(i)}(x,y) = \texttt{false}$ if either $p_x^{(i)} = 0$ or $q_y^{(i)} = 0$.

The iterative process. For $i \geq 1$, we stop before the i^{th} iteration if for every $x \in R$ and $y \in S$ it holds that $p_x^{(i-1)}, q_y^{(i-1)} \notin (0, \rho/8)$, which means that these probabilities are either zero or lower-bounded by $\rho/8$. Otherwise, we proceed as follows:

- If there exists an $x \in R$ such that $p_x^{(i-1)} \in (0, \rho/8)$, then we pick such an x, denote it $x^{(i)}$, and say that the i^{th} iteration is of the first type. Otherwise, we pick $y^{(i)} \in S$ such that $q_{y^{(i)}}^{(i-1)} \in (0, \rho/8)$, and say that the i^{th} iteration is of the second type.

- Assuming that the i^{th} iteration is of the first (resp., second) type, for every $(x,y) \in R \times S$, we define $\chi^{(i)}(x,y) = \texttt{false}$ if $x = x^{(i)}$ (resp., if $y = y^{(i)}$) and $\chi^{(i)}(x,y) = \chi^{(i-1)}(x,y)$ otherwise. (Intuitively, we give-up on $x^{(i)}$ (resp., $y^{(i)}$) when lower-bounding the probability of hitting H; this will be useful when upper-bounding the cross influence.)

- Next, we define $p_x^{(i)} \stackrel{\text{def}}{=} \Pr_{y\in S}[\chi^{(i)}(x,y)]$ for every x and $q_y^{(i)} \stackrel{\text{def}}{=} \Pr_{x\in R}[\chi^{(i)}(x,y)]$ for every y. Note that $p_x^{(i)} \leq p_x^{(i-1)} \leq p_x^{(0)} < 2 \cdot \rho^{1/2}$ and $q_y^{(i)} \leq q_y^{(i-1)} < 2 \cdot \rho^{1/2}$. Furthermore, assuming that the i^{th} iteration is of the first type, for every $x \in R$, it holds that $p_x^{(i)} \in \{0, p_x^{(i-1)}\}$, depending on whether or not $x = x^{(i)}$. (If the i^{th} iteration is of the second type, then for every $y \in S$ it holds that $q_y^{(i)} \in \{0, q_y^{(i-1)}\}$, depending on whether or not $y = y^{(i)}$.)

- Lastly, we define $p^{(i)} \stackrel{\text{def}}{=} \mathrm{E}_{x\in R}[p_x^{(i)}] = \mathrm{E}_{y\in S}[q_y^{(i)}]$, and observe that $p^{(i)} > p^{(i-1)} - \frac{\rho/8}{|R|}$ if the i^{th} iteration is of the first type (and $p^{(i)} > p^{(i-1)} - \frac{\rho/8}{|S|}$ if the i^{th} iteration is of the second type), since we modified a single term in the expectation.[24]

After $i \leq |R| + |S|$ iterations, the process stops, and we have $p^{(i)} > p^{(0)} - 2 \cdot \rho/8 \geq 0.5 \cdot \rho - 0.25 \cdot \rho = 0.25 \cdot \rho$. At this point we define $\chi \stackrel{\text{def}}{=} \chi^{(i)}$, $p \stackrel{\text{def}}{=} p^{(i)}$, and $p_x \stackrel{\text{def}}{=} p_x^{(i)} \notin (0, \rho/8)$ for each $x \in R$. Likewise, $q_y \stackrel{\text{def}}{=} p_x^{(i)} \notin (0, \rho/8)$ for each $y \in S$. Also recall that $p_x, q_y < 2 \cdot \rho^{1/2}$. Hence, each p_x (resp., q_y) is either zero or in $[0.125 \cdot \rho, 2 \cdot \rho^{1/2}]$.

Defining the joint distribution (X,Y). For $T = 0.25 \cdot \rho^{-1/2}$, the joint distribution (X,Y) is defined by its generator, denoted $G = (G_1, G_2)$. On input $(\overline{r}, \overline{s}) \in R^{T+1} \times S^{T+1}$ such that $\overline{r} = (r_0, r_1, ..., r_T)$ and $\overline{s} = (s_0, s_1, ..., s_T)$, the generator outputs (r_i, s_i) if $i \in [T]$ is the first index such that $\chi(r_i, s_i)$ holds, and outputs (r_0, s_0) otherwise (i.e., if no pair (r_i, s_i) satisfies χ). That is, for each $i \in [T]$, consider the indicator $\chi_i = \chi_i((r_0, r_1, ..., r_T), (s_0, s_1, ..., s_T))$ that evaluates to $\texttt{true}$ if and only if $\chi(r_i, s_i)$ holds and $\neg\chi(r_j, s_j)$ holds for every $j \in [i-1]$. Then, $G_1((r_0, r_1, ..., r_T), (s_0, s_1, ..., s_T)) = r_i$ if $\chi_i((r_0, r_1, ..., r_T), (s_0, s_1, ..., s_T))$ holds, and equals r_0 otherwise (i.e., if $\neg \bigvee_{i\in[T]} \chi(\overline{r}, \overline{s})$). We now turn to the analysis of (X,Y).

[24] In the first case, we consider $\sum_{x\in R} p_x^{(i)}$ and note that it equals $\sum_{x\in R\setminus\{x^{(i)}\}} p_x^{(i-1)}$, whereas $p_{x^{(i)}}^{(i-1)} < \rho/8$. Otherwise, we use $\sum_{y\in S} q_y^{(i)} = \sum_{y\in S\setminus\{y^{(i)}\}} q_y^{(i-1)}$.

Lower-bounding the min-entropy of X and Y. The min-entropy of X is at least $\log_2(1/(\rho^{1/2}\cdot 2^{-n})) = n - 0.25\cdot t - O(1)$, since for every x it holds that

$$\begin{aligned}\Pr[X=x] &= \sum_{i\in[T]} \Pr_{(\overline{r},\overline{s})\in R^{T+1}\times S^{T+1}}[\chi_i(\overline{r},\overline{s}) \ \ \& \ \ r_i=x] \\ &\quad + \Pr_{(\overline{r},\overline{s})\in R^{T+1}\times S^{T+1}}[(\forall i\in[2^{t'}])\,\neg\chi(\overline{r},\overline{s}) \ \ \& \ \ r_0=x] \\ &\leq \sum_{i\in[T]} \Pr_{r_i\in R}[r_i=x] \ \ + \Pr_{r_0\in R}[r_0=x],\end{aligned}$$

which equals $(T+1)\cdot|R|^{-1} < \rho^{1/2}\cdot 2^{-n}$, since $T+1 < 0.5\cdot\rho^{-1/2}$ and $|R| > 2^{n-1}$. Ditto for Y.

Lower-bounding $\Pr[F(X,Y)\in H]$. We lower-bound the probability that $F(X,Y)\in H$ as follows.

$$\begin{aligned}\Pr[F(X,Y)\in H] &\geq \sum_{i\in[T]} \Pr_{(\overline{r},\overline{s})\in R^{T+1}\times S^{T+1}}[\chi_i(\overline{r},\overline{s})] \\ &= \sum_{i\in[T]} \Pr_{(r_i,s_i)\in R\times S}[\chi(r_i,s_i)]\cdot \prod_{j\in[i-1]} \Pr_{(r_j,s_j)\in R\times S}[\neg\chi(r_j,s_j)] \\ &= \sum_{i\in[T]} p\cdot(1-p)^{i-1} \\ &= 1-(1-p)^T \\ &> T\cdot p.\end{aligned}$$

Recalling that $p\geq\rho/4$ and $T = 0.25\cdot\rho^{-1/2}$, we get $\Pr[F(X,Y)\in H]\geq 0.25\rho^{-1/2}\cdot\rho/4 = \rho^{1/2}/16$.

Upper-bounding the cross influence of (X,Y). Finally, we get to the analysis we were preparing for all along. We show that the cross influence of (X,Y) is upper-bounded by $2\log(8/\rho)$. For an arbitrary input $(\overline{r},\overline{s})$ to G, we consider the following two cases.

Case 1: $\chi_i(\overline{r},\overline{s})$ ***holds for some*** $i\in[T]$**.** If $\chi_i(\overline{r},\overline{s})$ holds, then $\chi(r_i,s_i)$ holds for that i, and in particular $p_{r_i}\geq\rho/8$ holds. In that case, using the fact that $p_x < 2\cdot\rho^{1/2}$ holds (for every $x\in R$), we have

$$\begin{aligned}\Pr_{\overline{u}=(u_0,u_1,\ldots,u_T)\in S^{T+1}}[G_1(\overline{r},\overline{u})=G_1(\overline{r},\overline{s})] &= \Pr_{\overline{u}=(u_0,u_1,\ldots,u_T)\in S^{T+1}}[G_1(\overline{r},\overline{u})=r_i] \\ &\geq \Pr_{\overline{u}=(u_0,u_1,\ldots,u_T)\in S^{T+1}}[\chi_i(\overline{r},\overline{u})] \\ &= \Pr_{u_i\in S}[\chi(r_i,u_i)]\cdot\prod_{j\in[i-1]}\Pr_{u_j\in S}[\neg\chi(r_j,u_j)] \\ &\geq p_{r_i}\cdot(1-\max_{x\in R}\{p_x\})^{i-1} \\ &> \frac{\rho}{8}\cdot(1-2\cdot\rho^{1/2})^T \\ &> \rho/16.\end{aligned}$$

Case 2: $\neg\chi_i(\overline{r},\overline{s})$ ***holds for every*** $i\in[T]$**.** In that case, $\neg\chi(r_i,s_i)$ holds for every i. Using, again, the fact that $p_x < 2\cdot\rho^{1/2}$ holds (for every $x\in R$), we have

$$\begin{aligned}
\Pr_{\overline{u}=(u_0,u_1,...,u_{2^{t'}})\in S^{T+1}}[G_1(\overline{r},\overline{u})=G_1(\overline{r},\overline{s})] &= \Pr_{\overline{u}=(u_0,u_1,...,u_T)\in S^{T+1}}[G_1(\overline{r},\overline{u})=r_0] \\
&\geq \Pr_{\overline{u}=(u_0,u_1,...,u_T)\in S^{T+1}}[(\forall i\in[T])\,\neg\chi(r_i,u_i)] \\
&= \prod_{i\in[T]} \Pr_{u_i\in S}[\neg\chi(r_i,u_i)] \\
&\geq (1-\max_{x\in R}\{p_x\})^T \\
&> (1-2\cdot\rho^{1/2})^T \\
&> 1/2.
\end{aligned}$$

Hence, the influence of the second part on G_1 is at most $\log_2(16/\rho) = 0.5\cdot t$ units, provided $\rho = 2^{-0.5\cdot t+4}$. The same holds for G_2 (using the fact that $q_y \geq 2\cdot\rho^{1/2}$ (for every $y\in S$)).

Conclusion. Using $\rho = 2^{-0.5\cdot t+4}$, we have shown that $\Pr[F(X,Y)\in H] \geq \rho^{1/2}/16 = 2^{-0.25\cdot t-2}$, whereas H has density ρ, and (X,Y) can be generated with t bits of cross influence, and X (resp., Y) has min-entropy $n - 0.5\log_2(1/\rho) = n - 0.25t + 2$. The claim follows. ■

4.3 Other issues

In Section 4.3.1 we undertake a brief study of the possibility of defining cross influence on the average, rather than in the worst-case (as done in Definition 4.1). The bottom-line is that, in retrospect, we believe that the worst-case definition is the 'right' choice for the current context. In Section 4.3.2 we show that cross influence is (essentially) much more expressive than the notion of coordination (treated in Section 3). In Section 4.3.3 we show that an approximate version of the converse of Proposition 4.2 does hold, while a perfect converse does not hold.

4.3.1 On an average-case notion of cross influence

A somewhat unsatisfying aspect of the definition of the amount of cross influence (i.e., Definition 4.1) is that, in some cases, this amount may be larger than the length of the substring of the output that is affected by the opposite part of the seed. Consider, for example, the joint distribution $(X,Y) = (X,Y'f(X))$, where X and Y' are independent of one another. Then, we would expect that the "influence" of X on Y to be at most $|f(X)|$ bits, but $\Pr_r[(y',f(r))=(y',f(x))] \geq 2^{-|f(x)|}$ does not necessarily hold for all x and y' (i.e., it might be the case that for some x it holds that $\Pr_r[f(r)=f(x)] \ll 2^{-|f(x)|}$).

This phenomenon would not have occurred if cross influence was defined as (minus the logarithm of) the expected value of the probability of agreement (i.e., $\mathrm{E}_x[\Pr_r[f(r)=f(x)]]$), since this expectation would equal the collision probability of $f(X)$ (which is always lower-bounded by $2^{-|f(x)|}$). Unfortunately, using this average-case measure (rather than our worst-case measure) would have hinder the proof of Eq. (6). Actually, this is not an artifact of our proof, but rather a reflection of reality (see Proposition 4.10). But let us first spell-out the definition that we have in mind.

Definition 4.8 (an average-case version of cross influence): *We say that the joint distribution (X,Y) is* generated with at most t bits of cross influence on the average *if there exists a function $G=(G_1,G_2)$ that generates (X,Y), as in Definition 4.1, and satisfies the following condition:*

There exists $t_1 \geq 0$ such that,

$$\mathrm{E}_{u,v\in\{0,1\}^\ell}\left[\mathrm{Pr}_{v'\in\{0,1\}^\ell}[G_1(u,v')\neq G_1(u,v)]\right] \leq 1-2^{-t_1}, \quad (9)$$

$$\mathrm{E}_{u,v\in\{0,1\}^\ell}\left[\mathrm{Pr}_{u'\in\{0,1\}^\ell}[G_2(u',v)\neq G_2(u,v)]\right] \leq 1-2^{-(t-t_1)}, \quad (10)$$

where, as in Definition 4.1, G_i denotes the i^{th} part of the output of G (i.e., $G(s) = (G_1(s), G_2(s))$).

(That is, the influence of each half of the seed on the opposite part of the outcome is bounded *on the average* (rather than on the worst case).)

Indeed, Eq. (9) (resp., Eq. (10)) captures an average-case notion of the amount of influence of the second (resp., first) part of the seed on the first (resp., second) part of the outcome. It coincides with the standard (average-case) notion of the influence of part of the argument of a function F on its outcome, which is typically written as $\mathrm{E}_{u\in\Omega}\left[\mathrm{Pr}_{v,v'\in\Omega'}[F(u,v)\neq F(u,v')]\right]$, which in turn equals $\mathrm{E}_{(u,v)\in\Omega\times\Omega'}\left[\mathrm{Pr}_{v'\in\Omega'}[F(u,v)\neq F(u,v')]\right]$. We first show that the average-case definition of cross influence may be much lower than the worst-case definition (used in Definition 4.1).

Proposition 4.9 (average-case vs worst-case cross influence): *There exists a joint distribution (X,Y) over $\{0,1\}^{n+n}$ that can be generated with a single bit of cross influence* on the average*, but cannot be generated with $n-1$ bits of cross influence* (on the worst-case)*.*

Proof: Consider the generator G that, on input $((b,r),s)$, outputs (s,s) if $b=1$ and (r,s) otherwise; that is, $G_2((b,r),s) = s$ whereas $G_1((1,r),s) = s$ and $G_1((0,r),s) = r$. Obviously, the first part of the seed has no influence on G_2. We upper-bound the average influence of the second part of the seed on G_1 by observing that

$$\begin{aligned}
&\mathrm{E}_{(b,r,s)\in\{0,1\}^{1+n+n}}\left[\mathrm{Pr}_{s'\in\{0,1\}^n}[G_1((b,r),s')\neq G_1((b,r),s)]\right]\\
&\quad= \frac{1}{2}\cdot\mathrm{E}_{(r,s)\in\{0,1\}^{n+n}}\left[\mathrm{Pr}_{s'\in\{0,1\}^n}[G_1((1,r),s')\neq G_1((1,r),s)]\right]\\
&\qquad+\frac{1}{2}\cdot\mathrm{E}_{(r,s)\in\{0,1\}^{n+n}}\left[\mathrm{Pr}_{s'\in\{0,1\}^n}[G_1((0,r),s')\neq G_1((0,r),s)]\right]\\
&\quad= \frac{1}{2}\cdot\mathrm{E}_{(r,s)\in\{0,1\}^{n+n}}\left[\mathrm{Pr}_{s'\in\{0,1\}^n}[s'\neq s]\right]+\frac{1}{2}\cdot\mathrm{E}_{(r,s)\in\{0,1\}^{n+n}}\left[\mathrm{Pr}_{s'\in\{0,1\}^n}[r\neq r]\right]\\
&\quad= \frac{1}{2}\cdot\left(1-2^{-n}\right)+\frac{1}{2}\cdot 0,
\end{aligned}$$

which is smaller than $1/2$. Hence, G generates a distribution (X,Y) with a single bit of cross influence *on the average*. Next, we lower-bound the cross influence (in the worst-case sense) of (X,Y) by using Proposition 4.2 and showing that (X,Y) has much lower min-entropy than the sum of the min-entropies of X and Y. Specifically, the min-entropy of X is n, since for every $x\in\{0,1\}^n$ it holds that

$$\begin{aligned}
&\mathrm{Pr}_{(b,r,s)\in\{0,1\}^{1+n+n}}[G_1((b,r),s')=x]\\
&\quad= \mathrm{Pr}_{(b,r,s)\in\{0,1\}^{1+n+n}}[b=0\ \&\ r=x]+\mathrm{Pr}_{(b,r,s)\in\{0,1\}^{1+n+n}}[b=1\ \&\ s=x]\\
&\quad= 0.5\cdot 2^{-n}+0.5\cdot 2^{-n} = 2^{-n}.
\end{aligned}$$

As for the min-entropy of Y, it is even easier to see that it equals n, since $G_2((b,r),s)=s$ always holds. Lastly, we upper-bound the min-entropy of (X,Y) by observing that for every $z\in\{0,1\}^n$ it holds that

$$\begin{aligned}\Pr_{(b,r,s)\in\{0,1\}^{1+n+n}}[G((b,r),s)=(z,z)] &> \Pr_{(b,r,s)\in\{0,1\}^{1+n+n}}[b=1 \ \&\ G((b,r),s)=(z,z)]\\ &= \Pr_{(b,r,s)\in\{0,1\}^{1+n+n}}[b=1 \ \&\ s=z],\end{aligned}$$

which equals 2^{-n-1}. Hence,

$$\Pr[(X,Y)=(z,z)] \ > \ 2^{-n-1} \ = \ 2^{-n+1}\cdot\Pr[X=z]\cdot\Pr[Y=z],$$

which by (a counter-positive of) Proposition 4.2 implies that (X,Y) cannot be generated with $n-1$ bits of cross influence (on the worst-case). ■

Failure of condensing. The proof of Proposition 4.9 provides a strong indication that joint distributions with low cross influence *on the average* cannot be significantly condensed. The reason is that such joint distributions may assign probability half to pairs of identical elements, which means that condensing such joint distributions is not easier than condensing a single source without a seed.

Proposition 4.10 (average-case cross influence cannot be condensed): *For every $F:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^m$, there exists a joint distribution (X,Y) over $\{0,1\}^{n+n}$ that can be generated with a single bit of cross influence* on the average *such that both X and Y have min-entropy at least $n-m$ but $F(X,Y)$ may assume some value with probability at least $1/2$.*

In other words, F has deficiency at least $m-1$ for sources of min-entropy $n-m$ with average cross influence 1. (Note that any distribution of min-entropy $n-m$ over $\{0,1\}^n$ has deficiency m, and so condensing to that level is trivial.)

Proof: We use the construction presented in the proof of Proposition 4.9, except that we let s and r be distributed independently according to arbitrary sources of min-entropy $n-m$. Note that it still holds that $\Pr[X=Y]\geq 1/2$, which means that a condenser for (X,Y) yields a (seedless) condenser for X itself, which is impossible. Specifically, consider $F'(x)=F(x,x)$ and note that $\Pr[F'(X)=v]\leq 2\cdot\Pr[F(X,Y)=v]$ holds for every v (equiv., $\log_2(1/\Pr[F'(X)=v])\geq\log_2(1/\Pr[F(X,Y)=v])-1$). But if X is only guranateed to have min-entropy $n-m$, then $F'(X)$ can be a constant; specifically, let v be such that $\Pr_{r\in\{0,1\}^n}[F(r)=v]\geq 2^{-m}$, and consider X that is uniform over $\{r:F'(r)=v\}$. It follows that $\Pr[F(X,Y)=v]\geq 1/2$. ■

4.3.2 Relation to the coordination model

It seems quite tempting to believe that if a joint distribution is t-coordinated then it can be generated with t bits of cross influence. We are only able to show that this is approximately true in the sense that such a distribution is close to one that can be generated with approximately t bits of cross influence. The (small) gap seems related to the issue discussed in Section 4.3.1.

Proposition 4.11 (low coordination implies low cross influence): *For every $\epsilon>0$, every joint distribution that is t-coordinated is ϵ-close to a distribution having at most $t+\log_2(1/\epsilon)$ bits of cross influence.*

Proof: Let (X,Y) be a joint distribution that is t-coordinated, and let Z be the related distribution guaranteed by Definition 3.1. Suppose (for simplicity)[25] that $Z = f(U_\ell)$, $X|_{Z=z} = g_z(U_\ell)$ and $Y|_{Z=z} = h_z(U_\ell)$, where U_ℓ denotes the uniform distribution on $\{0,1\}^\ell$. Consider the generator $G = (G_1, G_2)$ such that $G_1((s,r_1),r_2) = g_{f(s)}(r_1)$ and $G_2((s,r_1),r_2) = h_{f(s)}(r_2)$. Then, r_2 has no influence on G_1, whereas the influence of (s,r_1) on G_2 is upper-bounded as follows: For every (s,r_1) and r_2, it holds that

$$\begin{aligned}\Pr_{s',r_1'}[G_2((s',r_1'),r_2) \neq G_2((s,r_1),r_2)] &= \Pr_{s',r_1'}[h_{f(s')}(r_2) \neq h_{f(s)}(r_2)] \\ &\leq \Pr_{s'}[f(s') \neq f(s)] \\ &= 1 - \Pr[Z = f(s)].\end{aligned}$$

Hence, letting S denote the support of Z, the influence of (s,r_1) on G_2 is upper-bounbded by $\max_{z\in S}\{\log_2(1/\Pr[Z=z])\}$. Recall that by the hypothesis $|S| \leq 2^t$, and note that if $\min_{z\in S}\{\Pr[Z = z]\} \geq 2^{-t}$, then we get a cross influence bound of t.

In the general case (where $\min_{z\in S}\{\Pr[Z=z]\} \geq 2^{-t}$ may not hold), the proposition follows by first omitting from the support of Z any element that occurs with probability smaller than $\epsilon/2^t$, and then applying the foregoing argument. Specifically, denoting by Z' the random variables Z conditioned on $Z \in \{z : \Pr[Z=z] \geq \epsilon/2^t\}$, we let (X',Y') be the joint distribution obtained by picking z according to Z' and outputting $(X|_{Z=z}, Y|_{Z=z})$. Note that (X',Y') is ϵ-close to (X,Y). Applying the foregoing argument to (X',Y'), we infer that (X',Y') has cross influence at most $\log_2(\epsilon/2^t)^{-1}$, and the claim follows. ■

Separating cross influence from bits of coordination. An immediate corollary of combining Theorem 3.2 and Proposition 4.3 is that the model of cross influence is strictly more liberal than the model of coordination. Specifically, this yields Part 2 of the following corollary, whereas Part 1 follows by combining Theorem 3.2 and Proposition 4.6.

Corollary 4.12 (cross influence may be much smaller than the amount of coordination):

1. *There exists a joint distribution that can be generated with at most six bits of cross influence, but is* 0.24*-far from any* $(n - O(\log n))$*-coordinated distribution.*
2. *There exists a joint distribution that can be generated with at most one bit of cross influence, but is* 0.49*-far from any* $(0.5n - O(1))$*-coordinated distribution.*

Furthermore, each part in the foregoing joint distributions has min-entropy at least $n - 4$.

Proof: For Part 1, observe that (by Theorem 3.2 and [6, Thm. 7(i)]) there exists an extractor $\texttt{EXT}$ (with error $o(1)$) for any $(n - O(\log n))$-coordinated distribution of min-entropy at least $n - 4$ (since [6, Thm. 7(i)] asserts that standard extractors exist for min-entropy $\Theta(\log n)$). But Proposition 4.6 asserts that there exists a joint distribution (X,Y) having six bits of cross influence and min-entropy $n-4$ (for each part) such that $|\Pr[\texttt{EXT}(X,Y) = 1] - 0.5| \geq 1/4$. Hence, (X,Y) cannot be 0.24-close to an $(n - O(\log n))$-coordinated distribution.

[25] In general, the inputs for f and the g_z's and h_z's may be distributed uniformly in arbitrary finite sets. But the same can be done for G; see brief discussion following Definition 4.1. Furthermore, we can approximate the general processes by processes that take uniformly distributed ℓ-bit long strings, where $\ell = O(n/\epsilon^2)$.

For Part 2, recall that the proof of Proposition 4.3 actually establishes that there exists a joint distribution (X,Y) having a single bit of cross influence and min-entropy $n-1$ on which the inner-product mod 2 always outputs 0. On the other hand, by Theorem 3.2 and [6, Thm. 9], the inner-product mod 2 is an extractor with error 0.01 for any $(0.5n-O(1))$-coordinated distribution of min-entropy at least $n-1$ (since [6, Thm. 9] asserts that the inner-product mod 2 is a standard extractors for min-entropy $0.5n+\Omega(1)$). Hence, (X,Y) cannot be 0.49-close to a $(0.5n-O(1))$-coordinated distribution. ■

4.3.3 On the converse of Proposition 4.2

Recall that Proposition 4.2 asserts that *if the joint distribution (X,Y) can be generated with at most t bits of cross influence, then* Eq. (6) *holds for every x and y*. We show that while the converse does not hold literally (i.e., perfectly), it does hold approximately. Actually, it will be instructive to rewrite Eq. (6) as follows.

$$\Pr[Y=y \mid X=x] \leq 2^t \cdot \Pr[Y=y]. \tag{11}$$

We call t the min-entropy loss.

Proposition 4.13 (small min-entropy loss implies low cross influence): *Suppose that the joint distribution (X,Y) has min-entropy loss at most t; that is,* Eq. (11) *holds for every x,y* (w.r.t this value of t). *Then, for every $\epsilon>0$, the joint distribution (X,Y) is ϵ-close to a joint distribution that can be generated with $t+O(\log(1/\epsilon))$ bits of cross influence. Furthermore, the first element is generated with no influence of the second part of the seed.*

As shown in Proposition 4.14, the relaxation allowing to generate a distribution that is only close to the original is essential to the foregoing result.

Proof: Loosely speaking, for a joint distribution (X,Y) that satisfies Eq. (11), we shall present a generator $G=(G_1,G_2)$ with $t+O(1)$ bits of cross influence such that $G(U_{\ell+\ell})$ is close to (X,Y). We will generate X with no influence of the second half of the seed, and select Y with some bounded influence of the first half of the seed. Specifically, for $r=(r_1,...,r_4)$ and $s=(s_1,....,s_{O(2^t)})$, we let $G_1(r,s)$ be a function of r_1 only, and $G_2(r,s)$ be one of $O(2^t)$ strings corresponding to $s_1,...,s_{O(2^t)}$. Furthermore, each of the latter strings will be output with probability at least $\Omega(2^{-t})$, where the probability is taken over the choice of r (and the bound holds for any fixed s). To be more specific, we need a few definitions.

Let $g_1,g_2:\{0,1\}^{\ell'}\to\{0,1\}^n$ be such that $X\equiv g_1(U_{\ell'})$ and $Y\equiv g_2(U_{\ell'})$. Then, $G_1((r_1,r_2,r_3,r_4),s)=g_1(r_1)$ and $G_2(r,(s_1,....,s_T))$ will always be in $\{g_2(s_i):i\in[T]\}$. To actually define G_2, we need a few auxiliary notations. For every x,y, let $p_x(y)\stackrel{\rm def}{=}\Pr[Y=y\mid X=x]$, and $p(y)\stackrel{\rm def}{=}\Pr[Y=y]=\Pr[g_2(U_{\ell'})]$. Letting $q_x(y)\stackrel{\rm def}{=}p_x(y)/p(y)$, recall that $q_x(y)\leq 2^t$ (by Eq. (11)), and note that $\mathrm{E}[q_x(Y)]=\sum_y p_x(y)=1$. For a parameter $\epsilon>0$ (e.g., $\epsilon=0.001$), let $f=O(\epsilon^{-2}\log(1/\epsilon))$ and $T=f\cdot 2^t$.

- For $r=(r_1,r_2,r_3,r_4)\in\{0,1\}^{\ell'}\times[1/\epsilon]\times[T]\times[0,1]$ and $s=(s_1,...,s_T)\in(\{0,1\}^{\ell'})^T$, define $\chi(r,s)=0$ if either $r_2=1$ or $\sum_{i\in[T]}q_{g_1(r_1)}(g_2(s_i))\notin[(1\pm\epsilon)\cdot T]$; otherwise, $\chi(r,s)=1$.

- Then, $G_2(r,s) = g_2(s_{r_3})$ if $\chi(r,s) = 0$, and otherwise G_2 uses r_4 to output $g_2(s_i)$ with probability proportional to $q_{g_1(r_1)}(g_2(s_i))$; that is, in the latter case, $g_2(s_i)$ is output with probability

$$\pi_{r_1,s}(i) \stackrel{\text{def}}{=} \frac{q_{g_1(r_1)}(g_2(s_i))}{\sum_{j\in[T]} q_{g_1(r_1)}(g_2(s_j))} \tag{12}$$

  Letting $\sigma_{r_1,s}(r_4)$ denote the selection made by r_4 based on r_1 and s, we have

$$\Pr_{r_4}[\sigma_{r_1,s}(r_4)=i] = \pi_{r_1,s}(i). \tag{13}$$

  In this case, $G_2(r,s) = g_2(s_{\sigma_{r_1,s}(r_4)})$.

Actually, the $\pi_{r_1,s}(i)$'s (and so the $\sigma_{r_1,s}(r_4)$'s) depend only on $g_1(r_1)$ amd s (i.e., they are independent of the specific r_1 as long as $g_1(r_1)$ is fixed). These variables also determine whether or not $\sum_{i\in[T]} q_{g_1(r_1)}(g_2(s_i)) \notin [(1\pm\epsilon)\cdot T]$ holds. As we shall see, this condition holds with high probability, over the random choice of s. The variable r_2 is an alternative way of setting χ to 0; in this case $r_3 \in [T]$ is used to select the output among the $g_2(s_i)$'s. Typically $\chi = 1$, and in this case r_4 is used to select the outcome (according to Eq. (12); see also Eq. (13)). Either way, the output is always one of the $g_2(s_i)$'s.

The output distribution of the generator. Observe that, for every fixed $r = (r_1,r_2,r_3,r_4)$, and uniformly distributed $s = (s_1,...,s_T) \in (\{0,1\}^{\ell'})^T$ the $q_{g_1(r_1)}(g_2(s_j))$'s are IIDs in $[0,2^t]$ with expectation 1. Using the multiplicative Chernoff bound, we get

$$\Pr_{s_1,...,s_T\in\{0,1\}^{\ell'}}\left[\sum_{j\in[T]} q_{g_1(r_1)}(g_2(s_j)) = (1\pm\epsilon)\cdot T\right] \;>\; 1-\exp(-\epsilon^2\cdot T/2^t) \;=\; 1-\epsilon.$$

It follows that $\Pr_{r,s}[\chi(r,s)=0] < 2\epsilon$. We upper-bound the statistical distance betwen (X,Y) and $G(U_{2\ell})$ by considering an arbitrary set $S \subset \{0,1\}^{n+n}$ and observing that

$$\begin{aligned}
&\left|\Pr[(X,Y)\in S] - \Pr_{r,s\in\{0,1\}^\ell}[G(r,s)\in S]\right| \\
&\quad< \quad 2\epsilon + |\Pr[(X,Y)\in S] - \Pr_{r,s}[G(r,s)\in S \mid \chi(r,s)=1]| \\
&\quad\le \quad 2\epsilon + 0.5\cdot\sum_{x,y} |\Pr[(X,Y)=(x,y)] - \Pr_{r,s}[G(r,s)=(x,y) \mid \chi(r,s)=1]| \\
&\quad= \quad 2\epsilon + 0.5\cdot\sum_{x}\Pr[X=x]\cdot\sum_{y} |p_x(y) - \Pr_{r,s}[G_2(r,s)=y \mid \chi(r,s)=1 \;\&\; G_1(r)=x]|
\end{aligned}$$

where the equality uses $G_1(U_{2\ell}) \equiv X$ and $p_x(y) = \Pr[Y=y \mid X=x]$. Using $G_1((r_1,r_2,r_3,r_4),s) = g_1(r_1)$, for each (x,y), we analyze the corresponding term, while letting $A_x(s) = 1$ if and only if (the "approximation condition") $\sum_{j\in[T]} q_x(g_2(s_j)) = (1\pm\epsilon)\cdot T$ holds.

$$\begin{aligned}
&\Pr_{r,s}[G_2(r,s)=y \mid \chi(r,s)=1 \;\&\; g_1(r_1)=x] \\
&\quad= \quad \sum_{i\in[T]} \Pr_{s,r_4}[g_2(s_i)=y \;\&\; \sigma_{r_1,s}(r_4)=i \mid A_x(s)=1] \qquad\qquad \text{(for any } r_1 \in g_1^{-1}(x)) \\
&\quad\le \quad \sum_{i\in[T]} \Pr_{s_i}[g_2(s_i)=y]\cdot \max_{s:A_x(s)=1}\left\{\frac{q_x(y)}{\sum_{j\in[T]} q_x(g_2(s_j))}\right\}
\end{aligned}$$

$$\begin{aligned} &\leq \sum_{i\in[T]} p(y)\cdot\frac{p_x(y)/p(y)}{(1-\epsilon)\cdot T} \\ &< (1+2\epsilon)\cdot p_x(y). \end{aligned}$$

Similarly, we obtain a lower bound of $(1-2\epsilon)\cdot p_x(y)$. Hence, for every x, y, it holds that $\Pr_{r,s}[G_2(r,s)=y \,|\, \chi(r,s)=1 \ \&\ G_1(r)=x]$ is in $[(1\pm 2\epsilon)\cdot p_x(y)]$. It follows that the output of (G_1,G_2) is 3ϵ-close to (X,Y).

The cross influence of the generator. Recall that s has no influence on the output of G_1. So all that is left is to upper-bound the influence of r on G_2. This is easy to do by relying only on the case that $r_2=1$. Specifically, recall that for every (r,s) it holds that $G_2(r,s)=g_2(s_i)$ for some $i\in[T]$. Hence, for every (r,s) and such i, we have

$$\begin{aligned} \Pr_u[G_2(u,s)=G_2(r,s)] &= \Pr_u[G_2(u,s)=g_2(s_i)] \\ &\geq \Pr_{u=(u_1,u_2,u_3,u_4)}[u_2=1 \ \&\ u_3=i] \\ &= \frac{1}{1/\epsilon}\cdot\frac{1}{T} \end{aligned}$$

which implies an influence upper bound of $\log_2(T/\epsilon)=t+O(\log(1/\epsilon))$. ■

Proposition 4.14 (exact versus approximate generation with bounded cross influence): *There exists a joint distribution (X,Y) that is 2^{-n}-close to a product distribution but cannot be generated with less than one bit of cross influence. Furthermore, (X,Y) has min-entropy loss at most $O(2^{-n})$.*

Proof: Let (X,Y) be uniform on the set $S \stackrel{\text{def}}{=} \{(x,y)\in\{0,1\}^{n+n} : x\neq y\}$. Then, (X,Y) is 2^{-n}-close to the uniform distribution on $\{0,1\}^{n+n}$, which is a product distribution. We also observe that the min-entropy loss of (X,Y) is

$$\begin{aligned} \max_{(x,y)\in S}\left\{\log_2\left(\frac{\Pr[Y=y\,|\,X=x]}{\Pr[Y=y]}\right)\right\} &= \log_2\left(\frac{2^n}{2^n-1}\right) \\ &= \Theta(2^{-n}). \end{aligned}$$

On the other hand, generating (X,Y) requires at least one bit of cross influence, because *generating any joint distribution that is not a product distribution requires at least one bit of cross influence.* The latter assertion is proved by considering an arbitrary generator $G=(G_1,G_2)$ and observing that

$$\begin{aligned} \min_{r,s\in\{0,1\}^\ell}\left\{\Pr_{u\in\{0,1\}^\ell}[G_1(r,u)=G_1(r,s)]\right\} &= \min_{r\in\{0,1\}^\ell}\left\{\min_{y\in\mathrm{Supp}(G_1(r,U_\ell))}\{\Pr[G_1(r,U_\ell)=y]\}\right\} \\ &\leq \min_{r\in\{0,1\}^\ell}\left\{\frac{1}{|\mathrm{Supp}(G_1(r,U_\ell))|}\right\} \end{aligned}$$

where the last inequality is an equality in the case that $\mathrm{Supp}(G_1(r,U_\ell))$ is a singleton. Hence, if any of the sets $\mathrm{Supp}(G_1(r,U_\ell))$ is not a singleton, then the influence of the second part of the seed on G_1 is at least $\log_2|\mathrm{Supp}(G_1(r,U_\ell))|\geq 1$. The same holds, of course, for G_2. ■

Comment. The joint distribution (X,Y) used in the proof of Proposition 4.14 can be generated using a single bit of cross influence. Recall that (X,Y) is uniform on $S = \{(x,y) \in \{0,1\}^{n+n} : x \neq y\}$. For $r = (r',b) \in \{0,1\}^{n+1}$ and $s = (s_0,s_1) \in S$, consider the generator $G = (G_1,G_2)$ such that $G_1((r',b),s) = r'$ and $G_2((r',b),(s_0,s_1)) = s_b$ if $s_b \neq r'$ and $G_2((r',b),(s_0,s_1)) = s_{b\oplus 1}$ otherwise (i.e., when $s_b = r'$). The reader may verify that for every $(x,y) \in S$ it holds that

$$\begin{aligned}
&\Pr_{((r',b),(s_0,s_1))\in\{0,1\}^{n+1}\times S}[G_2((r',b),(s_0,s_1))=y \mid G_1((r',b),(s_0,s_1))=x] \\
&= \Pr_{(b,(s_0,s_1))\in\{0,1\}\times S}[G_2((x,b),(s_0,s_1))=y] \\
&= \Pr_{(b,(s_0,s_1))\in\{0,1\}\times S}[s_b=x \ \&\ s_{b\oplus 1}=y] + \Pr_{(b,(s_0,s_1))\in\{0,1\}\times S}[s_b=y] \\
&= \frac{1}{|S|} + 2^{-n} \\
&= \frac{1}{2^n-1}
\end{aligned}$$

which implies that G generates (X,Y). As for the influence of r on G_2, note that for every $r = (r',b)$ and $s = (s_0,s_1)$, letting $i \in \{0,1\}$ such that $G_2(r,s) = s_i$, it holds that

$$\begin{aligned}
&\Pr_{(u,c)\in\{0,1\}^{n+1}}[G_2((u,c),(s_0,s_1))=G_2((r',b),(s_0,s_1))] \\
&= \Pr_{(u,c)\in\{0,1\}^{n+1}}[G_2((u,c),(s_0,s_1))=s_i] \\
&\geq \Pr_{(u,c)\in\{0,1\}^{n+1}}[b=i \ \&\ u\neq s_i] + \Pr_{(u,c)\in\{0,1\}^{n+1}}[b=i\oplus 1 \ \&\ u=s_{i\oplus 1}] \\
&= \frac{2^n-1}{2^{n+1}} + \frac{1}{2^{n+1}}
\end{aligned}$$

and the claim follows.

5 Dependence as mutual information

This model is the most general model considered in the current work, and it offers the weakest positive results regarding the extraction of randomness.

5.1 Definition and Observations

The mutual information of the pair (X,Y), denoted $I(X;Y)$, is defined as $H(X)+H(Y)-H(X,Y)$, where $H(Z) \stackrel{\text{def}}{=} \sum_z \Pr[Z=z] \cdot \log_2(1/\Pr[Z=z])$ is the entropy of the distribution Z. A natural proposal is to measure the dependence between a pair of sources via their mutual information.

Definition 5.1 (the model of bounded mutual information): *The model of sources of bounded mutual information, denoted* $\mathtt{MI}_n(k,t)$*, consists of all joint distribution* (X,Y) *over* $\{0,1\}^{n+n}$ *such that* $I(X;Y) \leq t$ *and each of the* n*-bit long sources has min-entropy at least* k*.*

Relation to other models. It seems that an upper bound on the amount of cross influence implies an upper bound on mutual information. This is indeed the case, as shown next while relying on Proposition 4.2.

Proposition 5.2 (low cross influence implies low mutual information): *Suppose that the joint distribution (X,Y) can be generated with at most t bits of cross influence. Then, (X,Y) has mutual information at most t* (i.e., $I(X;Y) \leq t$).

Combinined with Proposition 4.11, it follows that a joint distribution that is t-coordinated is ϵ-close to having mutual information $t + \log_2(1/\epsilon)$. A stronger bound can be obtained directly (see Proposition 5.3).

Proof: Combining the definition of mutual definition[26] with Proposition 4.2 (see Eq. (6)), we have

$$\begin{aligned} I(X;Y) &= \sum_{x,y} \Pr[(X,Y)=(x,y)] \cdot \log_2 \left(\frac{\Pr[(X,Y)=(x,y)]}{\Pr[X=x] \cdot \Pr[Y=y]} \right) \\ &\leq \log_2(2^t). \end{aligned} \tag{14}$$

The claim follows. ■

Proposition 5.3 (low coordination implies low mutual information): *Every joint distribution that is t-coordinated has mutual information at most t.*

Proof: Let (X,Y) be a joint distribution that is t-coordinated, and let Z be the related distribution guaranteed by Definition 3.1. Recall that $I(A;(B,C)) = H(A) - H(A|(B,C)) \geq H(A) - H(A|B) = I(A;B)$. Hence, $I((X,Z);(Y,Z)) \geq I(X;Y)$ and

$$I(X;Y) \;\leq\; I((X,Z);(Y,Z)) \;=\; H(X,Z) + H(Y,Z) - H(X,Y,Z) \tag{15}$$

follows. Next, recall that

$$H(X,Y,Z) = H(Z) + H((X,Y)|Z) \tag{16}$$

and note that

$$\begin{aligned} H((X,Y)|Z) &= \sum_z \Pr[Z=z] \cdot H((X,Y)|Z=z) \\ &= \sum_z \Pr[Z=z] \cdot (H(X|Z=z) + H(Y|Z=z)) \\ &= H(X|Z) + H(Y|Z), \end{aligned}$$

where the second equality is due to the independence of X and Y when conditioned on any value of Z. Hence, we have

$$H((X,Y)|Z) = H(X,Z) - H(Z) + H(Y,Z) - H(Z). \tag{17}$$

[26] Indeed, mutual definition is often defined as Eq. (14). To see the equality to $H(X) + H(Y) - H(X,Y)$, use

$$\log_2 \left(\frac{\Pr[(X,Y)=(x,y)]}{\Pr[X=x] \cdot \Pr[Y=y]} \right) = \log_2 \left(\frac{1}{\Pr[X=x]} \right) + \log_2 \left(\frac{1}{\Pr[Y=y]} \right) - \log_2 \left(\frac{1}{\Pr[(X,Y)=(x,y)]} \right)$$

Combining Eq. (15), Eq. (16), and Eq. (17), we get

$$\begin{aligned} I(X;Y) &\leq H(X,Z) + H(Y,Z) - H(X,Y,Z) \\ &= H(X,Z) + H(Y,Z) - (H(Z) + H(X,Y|Z)) \\ &= H(X,Z) + H(Y,Z) - (H(Z) + H(X,Z) - H(Z) + H(Y,Z) - H(Z)) \\ &= H(Z), \end{aligned}$$

and the claim follows since $H(Z) \leq t$. ■

5.2 Extraction and condensing

The main result of this section is a condenser for joint distributions of bounded mutual information. This result, presented in Theorem 5.5, is inferior to the condensing result provided for the case of bounded cross influence (in Theorem 4.4). Unfortunately, as shown in Proposition 5.7, this is essentially the best one can do in the current model (i.e., Theorem 5.5 is essentially optimal). But before getting there, we present negative results regarding extraction and condensing (for low mutual information) that are sharper than those presented for low cross influence (cf. Propositions 4.6 and 4.7).

Proposition 5.4 (on the impossibility of extraction and limitation of condensing for joint distribution with low mutual information): *For every function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ and every $t \leq m$, there exists a joint distribution (X,Y) having at most t bits of mutual information such that both X and Y have min-entropy at least $n-t$, but the first t bits of $F(X,Y)$ are constant.*

Recall that Proposition 4.7, which refers to joint distributions with t bits of cross influence, only implies that the first t bits of $F(X,Y)$ have min-entropy at most $0.75t + O(1)$. (As for Proposition 4.6, it implies that if $t \geq 6$, then the first bit of $F(X,Y)$ has a bias at least $1/2$.) Furthermore, the following proof is much simpler than the proof of Proposition 4.7 (and is also is simpler than the proof of Proposition 4.6).

Proof: We let $F'(x,y)$ denote the t-bit long prefix of $F(x,y)$, and focus on the analysis of F'. For each v, we define $S_v \stackrel{\text{def}}{=} \{(x,y) \in \{0,1\}^{n+n} : F'(x,y) = v\}$, and let $p_v \stackrel{\text{def}}{=} \Pr_{x,y \in \{0,1\}^n}[(x,y) \in S_v]$. We fix v such that $p_v \geq 2^{-t}$, and let S be a subset of S_v that has exactly $2^{-t} \cdot 2^{2n}$ elements. We consider the joint distribution (X,Y) that is uniform over S. Then, for every $x \in \{0,1\}^n$, we have

$$\begin{aligned} \Pr[X=x] &= \Pr_{r,s \in \{0,1\}^n}[r = x \mid (r,s) \in S] \\ &= \frac{\Pr_{r,s \in \{0,1\}^n}[(r,s) \in S \mid r = x] \cdot \Pr_{r \in \{0,1\}^n}[r = x]}{\Pr_{r,s \in \{0,1\}^n}[(r,s) \in S]} \\ &= \frac{\Pr_{s \in \{0,1\}^n}[F'(x,s) = v] \cdot 2^{-n}}{2^{-t}} \end{aligned}$$

which is at most 2^{t-n}. This implies that X has min-entropy at least $n-t$, and the same considerations hold for Y. In addition, we observe that (X,Y) has mutual information at most t, since $H(X,Y) = 2n - t$ (by virtue of being uniform over S), whereas $H(X) + H(Y) \leq 2n$. The claim follows, since $F'(X,Y) \equiv v$. ■

Condensing (with inferior parameters). While we cannot meet the condensing guarantees provided in the case of cross influence (see Theorem 4.4), we can obtain meaningful (but inferior) condensing for joint distributions of low mutual information. The key observation is that the proof of Theorem 4.4 (i.e., condenser for sources with low cross influence) only relies on Eq. (6), which Proposition 4.2 asserts to hold for joint distribution of cross influence t. (Recall that Eq. (6) says that the min-entropy of (X,Y) is at most t units smaller than the sum of the min-entropies of X and Y.) We obtain a somewhat similar result for sources with low mutual information by showing that they are close to sources that satisfy Eq. (6), albeit with a larger min-entropy loss. Specifically, we shall show that if $I(X;Y) \leq t$ then, for every $\beta > 0$, the joint distribution $(1X, 1Y)$ is $O(\beta)$-close to a joint distribution that satisfy Eq. (6) with parameter $t' = O(t/\beta)$ (rather than with t).

Theorem 5.5 (condensers for sources with low mutual information): *Let $k \leq n$ and $\mathtt{CND} : \{0,1\}^{n+1} \times \{0,1\}^{n+1} \to \{0,1\}^m$ be a condenser with error ϵ and min-entropy deficiency d for pairs of independent sources that have min-entropy at least k each. Then, for every $\beta \in (0, 0.25)$ and for every $(X,Y) \in \{0,1\}^{n+n}$ such that $I(X;Y) \leq t$ and both X and Y have min-entropy at least k, it holds that $\mathtt{CND}(1X, 1Y)$ is $(4\beta + 2^{t'} \cdot \epsilon)$-close to a distribution that has min-entropy at least $m - d - t'$, where $t' \stackrel{\text{def}}{=} (t + O(1))/\beta$.*

That is, if $\mathtt{CND}$ is a condenser of error ϵ and deficiency d for the model $\mathtt{STD}_{n+1}(k)$, then $\mathtt{CND}'(x,y) = \mathtt{CND}(1x, 1y)$ is a condenser of error $4\beta + 2^{t'} \cdot \epsilon$ and deficiency $d + t'$ for the model $\mathtt{MI}_n(k,t)$, where $t' = (t + O(1))/\beta$. (In Theorem 5.6 we show that $\mathtt{CND}$ itself is a condenser of error $4\beta + 2^{t'} \cdot \epsilon$ and deficiency $d + t'$ for the model $\mathtt{MI}_{n+1}(k+1,t)$.) Indeed, this result is weaker than the analogous result for low cross influence (i.e., Theorem 4.4), since we lose a $t' = \beta^{-1} \cdot (t + O(1))$ term in the min-entropy guarantee for the output (whereas in Theorem 4.4 we only lost a t term). As is shown in Proposition 5.7, this loss is inevitible.

Proof: As hinted above, the proof boils down to showing that $(1X, 1Y)$ is 4β-close to a joint distribution (X', Y') such that both X' and Y' have min-entropy at least k and for every x and y it holds that

$$\Pr[(X',Y')=(x,y)] \leq 2^{t'} \cdot \Pr[X'=x] \cdot \Pr[Y'=y], \tag{18}$$

where $t' = O(t/\beta)$. Once Eq. (18) is established (with X' and Y' that have min-entropy at least k each), it follows that $\mathtt{CND}(X', Y')$ is $2^{t'} \cdot \epsilon$-close to having min-entropy at least $m - d - t'$, so $\mathtt{CND}(1X, 1Y)$ is $(4\beta + 2^{t'} \cdot \epsilon)$-close to a distribution having min-entropy at least $m - d - t'$. (We use $(1X, 1Y)$ rather than (X,Y) in order to simplify the proof; this "feature" is removed in Theorem 5.6.)

Towards proving Eq. (18), we consider the set of pairs (x,y) that violate this inequality; specifically,

$$B \stackrel{\text{def}}{=} \left\{ (x,y) : \Pr[(X,Y)=(x,y)] > 2^{t'} \cdot \Pr[X=x] \cdot \Pr[Y=y] \right\}. \tag{19}$$

Recalling that

$$I(X;Y) \;=\; \sum_{(x,y)} \Pr[(X,Y))=(x,y)] \cdot \log_2 \left(\frac{\Pr[(X,Y))=(x,y)]}{\Pr[X=x] \cdot \Pr[Y=y]} \right) \tag{20}$$

we observe that the contribution of pairs in B to Eq. (20) is large (i.e., each $(x,y) \in B$ contributes more than t' units). Assuming that the contribution of all other pairs is non-negative (which is not

the case), the foregoing would imply that $\Pr[X \in B]$ is small (i.e., smaller than t/t'). The actual argument requires addressing the foregoing objection.

Claim 5.5.1 (upper-bounding $\Pr[(X,Y) \in B]$): *For $t' = (t+16)/\beta$, it holds that $\Pr[(X,Y) \in B] < \beta$.*

Proof: The contribution of B to $I(X;Y)$ is given by the following sum

$$\sum_{(x,y) \in B} \Pr[(X,Y)) = (x,y)] \cdot \log_2 \left(\frac{\Pr[(X,Y)) = (x,y)]}{\Pr[X=x] \cdot \Pr[Y=y]} \right) \;>\; \Pr[(X,Y) \in B] \cdot t', \tag{21}$$

where the inequality is due to the definition of B. Now, if we could ignore the negative contribution to $I(X;Y)$, then we would have derived $\Pr[(X,Y) \in B] < t/t' < \beta$, and the claim would follow. But we need to justify ignoring the negative contribution or rather bound its effect, as we do next.

Letting $\rho_{x,y} \stackrel{\text{def}}{=} \frac{\Pr[(X,Y)) = (x,y)]}{\Pr[X=x] \cdot \Pr[Y=y]}$ we upper-bound the (magnitude of the) *negative* contribution of elements in the set $\{(x,y) : \rho_{x,y} < 1\}$ to $I(X;Y)$ by partitioning this set according to the approximate value of $\rho_{x,y}$. Specifically, for any $b > 1$ (e.g., $b = 2$), we consider the sets $S_i \stackrel{\text{def}}{=} \{(x,y) : \rho_{x,y} \in [b^{-i}, b^{-i+1})\}$, for $i \in \mathbb{N}$. We first note that

$$\begin{aligned} \Pr[(X,Y) \in S_i] &= \sum_{(x,y) \in S_i} \Pr[(X,Y) = (x,y)] \\ &< \sum_{(x,y) \in S_i} b^{-i+1} \cdot \Pr[X=x] \cdot \Pr[Y=y] \\ &\leq b^{-i+1}. \end{aligned}$$

Using $\Pr[(X,Y) \in S_i] < b^{-i+1}$ (and the definition of S_i), the *magnitude of the negative contribution* (of the pairs in $\bigcup_{i \in \mathbb{N}} S_i$) equals

$$\begin{aligned} N &\stackrel{\text{def}}{=} -\sum_{i \in \mathbb{N}} \sum_{(x,y) \in S_i} \Pr[(X,Y)) = (x,y)] \cdot \log_2 \rho_{x,y} \\ &= \sum_{i \in \mathbb{N}} \sum_{(x,y) \in S_i} \Pr[(X,Y)) = (x,y)] \cdot \log_2 (1/\rho_{x,y}) \\ &\leq \sum_{i \in \mathbb{N}} \Pr[(X,Y) \in S_i] \cdot \log_2 (b^i) \\ &< \sum_{i \in \mathbb{N}} b^{-i+i} \cdot (i \cdot \log_2 b) \\ &= \frac{b^4}{(b-1)^2} \cdot \log_2 b \end{aligned}$$

which equals 16 when $b = 2$. (Actually, the expression is minimized at $b \approx 1.40457$, and its value there is approximately $11.6546 < 12$.) Combining this fact with Eq. (21), we get

$$\begin{aligned} I(X;Y) &> \sum_{(x,y): \rho_{x,y} \in [1, 2^{t'}]} \Pr[(X,Y)) = (x,y)] \cdot \log_2 \rho_{x,y} + \Pr[(X,Y) \in B] \cdot t' - N \\ &\geq \Pr[(X,Y) \in B] \cdot t' - 16. \end{aligned}$$

Letting $t' = (t+16)/\beta$, it follows that $\Pr[(X,Y) \in B] < (I(X;Y)+16)/t' = \beta$. ■

Defining the joint distribution (X',Y'). One may be tempted to define (X',Y') as (X,Y) conditioned on not residing in B (i.e., $\Pr[(X',Y')=(x,y)] = \Pr[(X,Y)=(x,y)|(X,Y) \notin B]$ for every $(x,y) \notin B$), but this cause new violation of Eq. (20) (i.e., a pair (x,y) that was not in B may violate Eq. (20) because the probability of x (or y) was reduced by the modification). Hence we define (X',Y') by modifying the joint distribution (X,Y) in a more careful manner. Intuitively, compansate for removing pairs (x,y) that hit B by moving the probability weight to other pairs of the form $(x,\cdot)$ and $(\cdot,y)$. Specifically, we consider the following randomized sampling procedure.

- With probability β, the procedure selects uniformly $r,s \in \{0,1\}^n$, and outputs $(0r,0s)$.
- Otherwise (i.e., with probability $1-\beta$), the procedure acts as follows:
  - Samples (X,Y), obtaining $(x,y) \leftarrow (X,Y)$.
  - If $(x,y) \notin B$, it outputs $(1x,1y)$ with probability $1/2$ and halts without output otherwise.
  - Otherwise (i.e., when $(x,y) \in B$), the procedure selects uniformly $r \in \{0,1\}^n$, outputs $(1x,0r)$ with probability $1/2$, and outputs $(0r,1y)$ otherwise.

Letting p denote the probability that this procedure produces output, and $\beta' \stackrel{\text{def}}{=} \Pr[(X,Y) \in B] < \beta < 1/2$, observe that

$$p \;=\; \beta + (1-\beta)\cdot((1-\beta')\cdot 0.5 + \beta') \;=\; \beta + 0.5\cdot(1+\beta')\cdot(1-\beta), \tag{22}$$

which implies

$$0.5 \;<\; 0.5+0.5\beta \;\leq\; p \;<\; 0.5+\beta \;<\; 0.75, \tag{23}$$

where the last inequality uses $\beta < 0.25$. The joint distribution (X',Y') is obtained by applying rejection sampling to the foregoing procedure; that is, $\Pr[(X',Y')=(x',y')]$ equals $1/p$ times the probability that the foregoing procedure outputs (x',y').

Claim 5.5.2 (properties of the joint distribution (X',Y')):

1. *(X',Y') is 4β-close to $(1X,1Y)$.*
2. *(X',Y') satisfies $\Pr[(X',Y')=(x',y')] < 2^{t'+1}\cdot\Pr[X'=x']\cdot\Pr[Y'=y']$ for every x',y'.*
3. *X' and Y' have each min-entropy at least k.*

Proof: The following facts are each readily verifable.

1. For every $r,s \in \{0,1\}^n$, it holds that $\Pr[(X',Y')=(0r,0s)] = \beta\cdot 2^{-2n}/p$.
2. For every $(x,y) \in \{0,1\}^{2n} \setminus B$, it holds that

   $$\Pr[(X',Y')=(1x,1y)] \;=\; \frac{(1-\beta)\cdot\Pr[(X,Y)=(x,y)]\cdot 0.5}{p} \;=\; \frac{1-\beta}{2p}\cdot\Pr[(X,Y)=(x,y)],$$

   which is smaller than $\Pr[(X,Y)=(x,y)]$, since $\frac{1-\beta}{2p} < \frac{1-\beta}{1+\beta} < 1$.

In particular, combined with Claim 5.5.1, *this establishes Part 1 of the current claim.* Specifically, observe that the total variation distance between (X',Y') and $(1X,1Y)$ equals

$$\begin{aligned}
&\sum_{w:\Pr[(1X,1Y)=w]>\Pr[(X',Y')=w]} \big(\Pr[(1X,1Y)=w]-\Pr[(X',Y')=w]\big)\\
&= \sum_{(x,y)\in\{0,1\}^{n+n}} \big(\Pr[(1X,1Y)=(1x,1y)]-\Pr[(X',Y')=(1x,1y)]\big)\\
&= \Pr[(X,Y)\in B]+\sum_{(x,y)\in\{0,1\}^{2n}\setminus B} \big(\Pr[(1X,1Y)=(1x,1y)]-\Pr[(X',Y')=(1x,1y)]\big)\\
&= \Pr[(X,Y)\in B]+\sum_{(x,y)\in\{0,1\}^{2n}\setminus B} \left(\Pr[(X,Y)=(x,y)]-\frac{1-\beta}{2p}\cdot\Pr[(X,Y)=(x,y)]\right)\\
&= \Pr[(X,Y)\in B]+\Pr[(X,Y)\notin B]\cdot\left(1-\frac{1-\beta}{2p}\right)\\
&\leq \Pr[(X,Y)\in B]+\left(1-\frac{1-\beta}{2p}\right)
\end{aligned}$$

where the second equality is due to the fact that $\Pr[(X',Y')=(1x,1y)]=0$ if $(x,y)\in B$ and is smaller than $\Pr[(X,Y)=(x,y)]$ otherwise. Using $p<0.5+\beta$, we have $\left(1-\frac{1-\beta}{2p}\right)<\left(1-\frac{1-\beta}{1+2\beta}\right)<3\beta$. Using Claim 5.5.1, the total variation distance is upper-bounded by 4β, which establishes Part 1 of the current claim.

3. For every $x,r\in\{0,1\}^n$, it holds that

$$\Pr[(X',Y')=(1x,0r)]=\frac{1-\beta}{2p}\cdot\sum_{y:(x,y)\in B}\Pr[(X,Y)=(x,y)]\cdot 2^{-n}.$$

   Ditto for $\Pr[(X',Y')=(0r,1y)]$.

4. Using Facts 2 and 3, for every $x\in\{0,1\}^n$, we have

$$\Pr[X'=1x] \;=\; \frac{1-\beta}{2p}\cdot\sum_{y}\Pr[(X,Y)=(x,y)] \;=\; \frac{1-\beta}{2p}\cdot\Pr[X=x]$$

   which resides in $(0.5\cdot\Pr[X=x],\Pr[X=x])$, since $p\in(0,5,0.5+\beta)$ and $\beta<1/4$. In particular, $\Pr[X'=1x]<\Pr[X=x]\leq 2^{-k}$. Ditto for $\Pr[Y'=1y]$.

5. Using Facts 1 and 3, for every $r\in\{0,1\}^n$, we have

$$\begin{aligned}
\Pr[X'=0r] &= \frac{\beta\cdot 2^{-n}}{p}+\frac{1-\beta}{2p}\cdot\sum_{(x,y)\in B}\Pr[(X,Y)=(x,y)]\cdot 2^{-n}\\
&= \left(\beta+\frac{(1-\beta)\cdot\beta'}{2}\right)\cdot\frac{2^{-n}}{p}
\end{aligned}$$

   Hence, this probability resides in $[\beta\cdot 2^{-n}/p, 1.5\beta\cdot 2^{-n}/p)$, which is a subset of $(\beta\cdot 2^{-n}, 3\beta\cdot 2^{-n})$. In particular, $\Pr[X'=0r]<3\beta\cdot 2^{-n}<2^{-n}$, since $\beta<1/4$. Ditto for $\Pr[Y'=0r]$.

   Combining Facts 4 and 5, it follows that the min-entropy of X' is $\min(k,n)=k$. Ditto for Y'. *This establishes Part 3 of the claim.*

6. Using Facts 2 and 4 (as well as the definition of B), for every $(x,y) \in \{0,1\}^{2n} \setminus B$, we have

$$\begin{aligned}
\Pr[(X',Y')=(1x,1y)] &= \frac{(1-\beta)}{2p} \cdot \Pr[(X,Y)=(x,y)] \\
&\leq \frac{(1-\beta)}{2p} \cdot 2^{t'} \cdot \Pr[X=x] \cdot \Pr[Y=y] \\
&= \frac{(1-\beta)}{2p} \cdot 2^{t'} \cdot \frac{\Pr[X'=1x]}{(1-\beta)/2p} \cdot \frac{\Pr[Y'=1y]}{(1-\beta)/2p} \\
&= \frac{2p}{1-\beta} \cdot 2^{t'} \cdot \Pr[X'=1x] \cdot \Pr[Y'=1y] \\
&< 2^{t'+1} \cdot \Pr[X'=1x] \cdot \Pr[Y'=1y]
\end{aligned}$$

since $2p/(1-\beta) < 2$.

Recall that $\Pr[(X',Y')=(1x,1y)] = 0$ for every $(x,y) \in B$.

7. Combining Fact 3 with Facts 4 and 5, for every $(x,r) \in \{0,1\}^{2n}$, we have

$$\begin{aligned}
\Pr[(X',Y')=(1x,0r)] &= \frac{(1-\beta)}{2p} \cdot \sum_{y:(x,y)\in B} \Pr[(X,Y)=(x,y)] \cdot 2^{-n} \\
&\leq \frac{(1-\beta)}{2p} \cdot \Pr[X=x] \cdot 2^{-n} \\
&< \frac{(1-\beta)}{2p} \cdot \frac{\Pr[X'=1x]}{(1-\beta)/2p} \cdot \frac{\Pr[Y'=0r]}{\beta} \\
&= \beta^{-1} \cdot \Pr[X'=1x] \cdot \Pr[Y'=0r],
\end{aligned}$$

where the second inequality is due to $\Pr[Y'=0r] > \beta \cdot 2^{-n}$. Ditto for $\Pr[(X',Y')=(0r,1y)]$.

8. Lastly, using Facts 1 and 5, for every $(r,s) \in \{0,1\}^{2n}$, we have

$$\begin{aligned}
\Pr[(X',Y')=(0r,0s)] &= \frac{\beta}{p} \cdot 2^{-2n} \\
&< \frac{\beta}{p} \cdot \frac{\Pr[X'=0r]}{\beta} \cdot \frac{\Pr[Y'=0s]}{\beta} \\
&= \beta^{-1} \cdot \Pr[X'=0r] \cdot \Pr[Y'=0s].
\end{aligned}$$

Combining Facts 6–8, we get $\Pr[(X',Y')=(x',y')] < \max(2^{t'+1}, \beta^{-1}) \cdot \Pr[X'=x'] \cdot \Pr[Y'=y']$ for every $x', y' \in \{0,1\}n+1$. Using $t' = (t+16)/\beta > 1/\beta > \log_2(1/\beta)$, *this establishes Part 2 of the claim.* Recalling that we have already established Parts 1 and 3 (see Facts 2 and 5, respectively), the claim follows. ■

Conclusion. Recall that, by our hypothesis, $\mathtt{CND}$ is a condenser of error ϵ and deficiency d for the model $\mathtt{STD}_{n+1}(k)$. The proof of Theorem 4.4 actually establishes that *if U and V have min-entropy k and $\Pr[(U,V)=(u,v)] \leq 2^{\tau} \cdot \Pr[U=u] \cdot \Pr[V=v]$ for all $u,v \in \{0,1\}^{n+1}$, then $\mathtt{CND}(X',Y')$ is $2^{\tau} \cdot \epsilon$-close to a distribution that has min-entropy $m-d-\tau$.* These two conditions are established for (X',Y') and $\tau = t'+1$ by Parts 3 and 2 of Claim 5.5.2, respectively. Hence, $\mathtt{CND}(X',Y')$ is $2^{t'+1} \cdot \epsilon$-close to a distribution that has min-entropy $m-d-(t'+1)$. Using Part 1 of Claim 5.5.2, it follows that $\mathtt{CND}(1X,1Y)$ is $(4\beta + 2^{t'+1} \cdot \epsilon)$-close to a distribution that has min-entropy $m-d-(t'+1)$. Recalling that $t' = (t+16)/\beta$, the theorem follows (by replacing t' with $t'+1$). ■

A small variation. As shown next, the fact that in Theorem 5.5 the standard condenser $\mathtt{CND}$ is applied to $(1X, 1Y)$, rather than to the joint distribution (X, Y) itself, is immaterial; it was done only in order to make the proof more transparent.

Theorem 5.6 (Theorem 5.5, revisited): *Let $k \leq n$ and $\mathtt{CND} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ be a condenser with error ϵ and min-entropy deficiency d for pairs of independent sources that have min-entropy at least $k-1$ each. Then, for every $\beta \in (0, 0.25)$ and for every $(X, Y) \in \{0,1\}^{n+n}$ such that $I(X;Y) \leq t$ and both X and Y have min-entropy at least k, it holds that $\mathtt{CND}(X, Y)$ is $(4\beta + 2^{t'} \cdot \epsilon)$-close to a distribution that has min-entropy at least $m - d - t'$, where $t' \stackrel{\text{def}}{=} (t + O(1))/\beta$.*

That is, if $\mathtt{CND}$ is a condenser of error ϵ and deficiency d for the model $\mathtt{STD}_n(k-1)$, then it constitutes a condenser of error $4\beta + 2^{t'} \cdot \epsilon$ and deficiency $d + t'$ for the model $\mathtt{MI}_n(k, t)$, where $t' = (t + O(1))/\beta$.

Proof: Recall that, given a joint distribution $(X, Y) \in \{0,1\}^{n+n}$, we have defined in the proof of Theorem 5.5 a distribution $(X', Y') \in \{0,1\}^{2\cdot(n+1)}$, which was shown to be 4β-close to $(1X, 1Y)$. Here we define a distribution (X'', Y'') so that X'' (resp., Y'') is the n-bit long suffix of X' (resp., Y'). We establish the following properties of (X'', Y''), while relying on the corresponding properties of (X', Y') (as asserted in Claim 5.5.2):

1. (X'', Y'') is 4β-close to (X, Y).

   This holds because the application of a function can only decrease the total variation distance between distributions. Specifically, consider the function $f : \{0,1\}^{2(n+1)} \to \{0,1\}^{2n}$ defined by $f(x', y') = (x'', y'')$ such that x'' (resp., y'') is the n-bit long suffix of x' (resp., y'). Then, $f(X', Y')$ is 4β-close to $f(1X, 1Y)$, whereas $(X'', Y'') = f(X', Y')$ and $(X, Y) = f(1X, 1Y)$.

2. (X'', Y'') satisfies $\Pr[(X'', Y'') = (x'', y'')] < 2^{t'+1} \cdot \Pr[X'' = x''] \cdot \Pr[Y'' = y'']$ for every x'', y''.

   This is beacuse

   $$\begin{aligned} \Pr[(X'', Y'') = (x'', y'')] &= \sum_{\sigma,\tau \in \{0,1\}} \Pr[(X', Y') = (\sigma x'', \tau y'')] \\ &\leq \sum_{\sigma,\tau \in \{0,1\}} 2^{t'+1} \cdot \Pr[X' = \sigma x''] \cdot \Pr[Y' = \tau y''] \\ &= 2^{t'+1} \cdot \Pr[X'' = x''] \cdot \Pr[Y'' = y''].\end{aligned}$$

3. X'' and Y'' have each min-entropy at least $k - 1$.

   This is because

   $$\begin{aligned} \Pr[X'' = x''] &= \sum_{\sigma \in \{0,1\}} \Pr[X' = \sigma x''] \\ &\leq 2 \cdot 2^{-k}.\end{aligned}$$

As in the proof of Theorem 5.5, it follows that $\mathtt{CND}(X'', Y'')$ is $2^{t'} \cdot \epsilon$-close to a distribution that has min-entropy $m - d - (t' + 1)$, and consequently $\mathtt{CND}(X, Y)$ is $(4\beta + 2^{t'} \cdot \epsilon)$-close to a distribution that has min-entropy $m - d - (t' + 1)$. The theorem follows. ■

On the optimality of Theorem 5.5. As mentioned already a couple of times, the min-entropy loss suffered by Theorem 5.5 is inevitable.

Proposition 5.7 (on the limitation of condensing for joint distribution with low mutual information): *For every function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ and every $t \leq m$ and $\epsilon \in (0,1)$, there exists a joint distribution (X', Y') having at most t bits of mutual information such that, for $t' \stackrel{\text{def}}{=} ((t-2)/\epsilon)$, both X' and Y' have min-entropy at least $n - t'$, but there exists $v \in \{0,1\}^{t'}$ such that the first t' bits of $F(X', Y')$ equal v with probability greater than ϵ.*

Actually, $t' = (t - H_2(\epsilon))/\epsilon$, where H_2 is the binary entropy function. Hence, for any $\epsilon' < \epsilon$, any distribution W that is ϵ'-close to $F(X,Y)$ satisfes $\Pr[W = v] \geq \epsilon - \epsilon'$. It follows that W has min-entropy at least $\log_2(2^{m-t'}/(\epsilon - \epsilon')) = m - t' - \log_2(\epsilon - \epsilon')$, which means that *there is no condenser with deficiency $t' - \log_2(1/(\epsilon - \epsilon'))$ and error ϵ' for sources of cross influence t that each have min-entropy at least $n - t'$*. Hence, the deficiency obtained by Theorem 5.5 is optimal up to a constant factor. Specifically, recalling that the deficiency bound obtained in Theorem 5.5 for error ϵ' is $(t + O(1))/(0.25 \cdot \epsilon' - o(1))$, whereas $t' - \log_2(1/(\epsilon - \epsilon')) = (t - 2 - \epsilon^{-1} \cdot \log_2(1/(\epsilon - \epsilon'))/\epsilon$, we conclude that the constant factor is arbitrary close to 4 (by choosing $\epsilon' < \epsilon$ arbitrarily close to ϵ while assuming $\log_2(1/(\epsilon - \epsilon')) + O(1) = o(t)$).

Proof: Let $F'(x,y)$ denote the first t' bits of $F(x,y)$, and let (X,Y) be a joint distribution as guaranteed by Proposition 5.4; that is, (X,Y) has at most t' bits of mutual information, both X and Y have min-entropy at least $n - t'$, and the first t' bits of $F'(X,Y)$ are constant, denoted v. Define (X', Y') to equal (X,Y) with probability ϵ, and equal the uniform distribution over $\{0,1\}^{2n}$ otherwise. Then:

1. The mutual information of (X', Y') is at most $\epsilon \cdot t' + 2 = t$; actually, $I(X', Y') \leq \epsilon \cdot I(X,Y) + 2 \cdot H_2(\epsilon)$, where H_2 is the binary entropy function.

   This holds because $H(X') \leq H_2(\epsilon) + \epsilon \cdot H(X) + (1-\epsilon) \cdot n$, and ditto for Y', whereas $H(X', Y') \geq \epsilon \cdot H(X,Y) + (1-\epsilon) \cdot 2n$. Hence,

   $$\begin{aligned} I(X', Y') &= H(X') + H(Y') - H(X', Y') \\ &\leq 2 \cdot H_2(\epsilon) + \epsilon \cdot (H(X) + H(Y)) + 2 \cdot (1-\epsilon) \cdot n - (\epsilon \cdot H(X,Y) + (1-\epsilon) \cdot 2n) \\ &= 2 \cdot H_2(\epsilon) + \epsilon \cdot I(X,Y). \end{aligned}$$

2. The min-entropy of X (resp., Y) is at least $n - t'$, since $\Pr[X = x] \leq \epsilon \cdot 2^{-(n-t')} + (1-\epsilon) \cdot 2^{-n}$ for every x.

3. $\Pr[F'(X', Y') = v] = \epsilon + (1-\epsilon) \cdot 2^{-2n} > \epsilon$.

The claim follows. ■

Separating cross influence from mutual information. An immediate corollary of combining Theorem 4.4 and Proposition 5.7 is that the model of mutual information is strictly more liberal than the model of cross influence.

Corollary 5.8 (mutual information may be much smaller than cross influence): *For every $\epsilon' \in [\Omega(1/n), 0.5]$, there exists a joint distribution that that has mutual information at most three, but is ϵ'-far from any distribution that can be generated with $1/2\epsilon'$ bits of cross influence. Furthermore, each part in the foregoing joint distribution has min-entropy at least $n - (1/2\epsilon')$.*

Equivalently, for every $t' < n/O(1)$, there exists a joint distribution has mutual information at most three, but is $(1/2t')$-far from any distribution that can be generated with t' bits of cross influence.

Proof: We contrast the following two facts.

1. By Theorem 4.4 and [6, Thm. 7(i)], for any $\epsilon' > 0$, there exists a condenser $\mathtt{CND} : \{0,1\}^{n+n} \to \{0,1\}^{t'+\log_2(1/\epsilon')}$ with error ϵ' and deficiency t' for any joint distribution having t' bits of cross influence and min-entropy $3 \cdot (t' + \log(9n/\epsilon'))$.

   This is the case because [6, Thm. 7(i)] asserts that standard extractors of error $2^{-t'} \cdot \epsilon'$ that output m bits exist for min-entropy $k = m + 2 \cdot (t' + \log_2(9n/\epsilon'))$, whereas by Theorem 4.4 such condensers have error ϵ' and deficiency t' for any joint distribution having t' bits of cross influence and min-entropy k.

2. On the other hand, instatiating Proposition 5.7 with $t = 3$ and $\epsilon = 1/t'$ implies that there exists a joint distribution (X, Y) having three bits of mutual information and min-entropy $n - t'$ (for each part) for which $\mathtt{CND}(X, Y)$ is $(\epsilon - \epsilon')$-far from having min-entropy at least $\log_2(1/\epsilon')$.

Hence, if $n - t' \geq 3 \cdot (t' + \log_2(9n/\epsilon'))$ and $\epsilon - \epsilon' \geq \epsilon'$, then (X, Y) cannot be generated with t' bits of cross influence, because otherwise Fact 1 would imply that $\mathtt{CND}(X, Y)$ is ϵ'-close to having min-entropy at least $\log_2(1/\epsilon')$ (since this corresponds to deficiency t' on an output of length $t' + \log_2(1/\epsilon')$), whereas Fact 2 says that $\mathtt{CND}(X, Y)$ is $(\epsilon - \epsilon')$-far from any such distribution. Using $\epsilon' = \epsilon/2$ and $t' = 1/\epsilon < n/5$, the claim follows. ∎

6 Detour: On the communication complexity of sampling

The communication complexity of sampling problems, as defined in Definition 3.3, seems to have emerged in [2]. While our intention in making this definition was to capture a natural notion of dependence between sources (viewed as the outcomes of the two parties in such a protocol), we observe that our results regarding extraction from such sources yield communication complexity lower bounds. Specifically, combining Theorems 3.2 and 3.4, we derive the following lower bound.

Theorem 6.1 (communication complexity lower bounds): *Let* $\mathtt{EXT} : \{0,1\}^{n-1} \times \{0,1\}^{n-1} \to \{0,1\}$ *be a standard two-source extractor with error* 0.12 *for sources of min-entropy* k*. Then, the joint distribution* $(X, Y) = ((X', 1), (Y', \mathtt{EXT}(X', Y')))$*, where* (X', Y') *is uniformly distributed in* $\{0,1\}^{n-1} \times \{0,1\}^{n-1}$*, has communication complexity greater than* $n - k - 7$*.*

A couple of corollaries of Theorem 6.1 are stated below. As stated upfront, the communication complexity of such problems seems to have emerged in [2], and we do not know if Theorem 6.1 was known before. For sure, the current proof is simple. We also note that although [6, Thm. 24] showed that computing a standard extractor for min-entropy k requires $n - k - O(1)$ bits of communication, the sampling problem may be easier.

Proof: Note that the joint distribution $(X, Y) = ((X', 1), (Y', \mathtt{EXT}(X', Y')))$ generalizes the distribution used in the proof of Proposition 4.3, where we used the inner product (mod 2) function in the role of $\mathtt{EXT}$. On the one hand, the combination of Theorems 3.2 and 3.4 implies if (X, Y) has communication complexity $n - k - O(1)$, then applying any standard two-source extractor to

(X,Y) yields an output bit with bounded bias. On the one hand, we show that a specific standard extractor that is derived from $\texttt{EXT}$ yields a constant bit when applied to (X,Y). We comment that for the special case of the inner product (mod 2) extractor, the transformation of $\texttt{EXT}$ to a related standard extractor is not needed, and the proof reduces to invoking Proposition 4.3 (as is) and contrasting it with Theorems 3.2 and 3.4 (see Footnote 27).

Seeking to establish the more general result, we first modify the extractor $\texttt{EXT}$ in a way that maintains its ability to extract from independent sources, while allowing us to show (later) that it fails to extract on (X,Y). Specifically, we first show that the function $F:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ defined by $F(x'a,y'b)=\texttt{EXT}(x',y')+b$, where $a,b\in\{0,1\}$, is a standard two-source extractor with error 0.24 for sources of min-entropy $k+4$.

Consider two arbitrary independent sources U and V, each having min-entropy $k'=k+\log_2(1/0.12)$. Then, the $(n-1)$-bit prefix of U, denoted U', has min-entropy $k'-1>k$. The same holds for the $(n-1)$-bit prefix of V, denoted V'. Furthermore, for $V=V'B$ (i.e., $B\in\{0,1\}$), if $\Pr[B=b]\geq 0.12$, then the *min-entropy of V' conditioned on $B=b$ is at least* $k'-\log_2(1/0.12)=k$ (i.e., $\Pr[V'=v'|B=b]\leq 2^{-k'}/0.12$). Using the hypothesis regarding $\texttt{EXT}$ (and noticing that the case of $\Pr[B=b]<0.12$ only adds 0.12 to the extraction error of F), we infer that

$$\begin{aligned}\Pr[F(U'A,V'B)=0] &= \Pr[\texttt{EXT}(U',V')+B=0]\\ &= \sum_{b\in\{0,1\}}\Pr[B=b]\cdot\Pr[\texttt{EXT}(U',V')+B=0\,|\,B=b]\\ &= 0.5\pm 2\cdot 0.12,\end{aligned}$$

where one 0.12 term is due to the extraction error of $\texttt{EXT}$ (on sources of min-entropy k) and the other term is due to the case $\Pr[B=b]<0.12$ (whenever such a b exists). Hence, F is a standard two-source extractor with error 0.24 for sources of min-entropy $k+\log_2(1/0.12)$.

Recall that $(X,Y)=((X',1),(Y',\texttt{EXT}(X',Y')))$, where (X',Y') is uniformly distributed in $\{0,1\}^{n-1}\times\{0,1\}^{n-1}$. Now, suppose that (X,Y) has communication complexity at most t, and consider the following two conflicting facts regarding $F(X,Y)$.

1. On the one hand, $F(X,Y)$ is 0.48-close to a uniformly distributed bit, provide that X (resp., Y) has min-entropy at least $(k+\log_2(1/0.12)+t+\log_2(1/0.24)=k+t+2\log_2(1/0.12)-1$.

   This follows by combining Theorem 3.4, which implies that (X,Y) is t-coordinated, with Theorem 3.2, which asserts that any extractor of error ϵ for $\texttt{STD}_n(k'')$ constitutes an extractor of error 2ϵ for $\texttt{COOR}_n(k''+t+\log_2(1/\epsilon),t)$.

2. On the other hand, $F(X,Y)\equiv 0$, although X (resp., Y) has min-entropy at least $n-1$.

   This is the case because $F((x',1),(y',\texttt{EXT}(x',y')))=\texttt{EXT}(x',y')+\texttt{EXT}(x',y')=0$, whereas $(X,Y)=((X',1),(Y',\texttt{EXT}(X',Y')))$.

These two facts stand in contradiction if $k+t+2\log_2(1/0.12)-1\leq n-1$. Hence, $k+t+2\log_2(1/0.12)>n$ must hold, which yields $t>n-k-7$. ■

Corollary 6.2 (corollaries of Theorem 6.1):

1. *Let $\texttt{IP}_2:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ denote inner product* (mod 2) *function. Then, the joint distribution $(X,Y)=((X',1),(Y',\texttt{IP}_2(X',Y')))$ has communication complexity greater than*

$0.5n - 12$.[27]

2. *For more than a* $1-(1/n)$ *fraction of the functions* $F : \{0,1\}^{n-1} \times \{0,1\}^{n-1} \to \{0,1\}$*, it holds that the joint distribution* $(X,Y) = ((X',1),(Y',F(X',Y')))$ *has communication complexity greater than* $n - \log_2 n - 19$*.*

In both parts, (X',Y') *is uniformly distributed in* $\{0,1\}^{n-1} \times \{0,1\}^{n-1}$*,*

Proof: For Part 1, we use [6, Thm. 9] which asserts that the inner product (mod 2) is an extractor with error 0.12 for independent sources that have min-entropy at least $0.5n + 1 + \log_2(1/0.12)$. For Part 2, we use [6, Thm. 7(i)]) which asserts that more than a $1 - (1/n)$ fraction of the functions $F : \{0,1\}^{n-1} \times \{0,1\}^{n-1} \to \{0,1\}$ constitute extractors with error 0.12 for independent sources that have min-entropy at least $\log_2 n + 5 + 2\log_2(1/0.12)$. ■

Acknowledgments

Marshall Ball was supported by an IBM Research PhD Fellowship. Oded Goldreich was partially supported by an ISF grant number (Nr. 1146/18), and the research was conducted while he enjoyed the hospitality of the computer science department at Columbia University.

This work was supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) via Contract No. 2019-1902070006. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either express or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

[1] D. Aggarwal, M. Obremski, J. Ribeiro, L. Siniscalchi, and I. Visconti. How to Extract Useful Randomness from Unreliable Sources. Cryptology ePrint Archive: Report 2019/1156.

[2] A. Ambainis, L.J. Schulman, A. Ta-Shma, U.V. Vazirani, and A. Wigderson. The Quantum Communication Complexity of Sampling. *SIAM Journal on Computing*, Vol. 32 (6), pages 1570–1585, 2003. Preliminary version in *39th FOCS*, 1998.

[3] A. Ben-Aroya, G. Cohen, D. Doron, and A. Ta-Shma. Two-Source Condensers with Low Error and Small Entropy Gap via Entropy-Resilient Functions. In *23rd RANDOM*, LIPIcs 145, pages 43:1–43:20, Springer, 2019.

[27] A slightly stronger statement can be proved by modifying the proof of Theorem 6.1 to the current case (of $\mathtt{EXT} = \mathtt{IP}_2$). Specifically, we can use $F = \mathtt{IP}_2$ (rather than use $F(x'a, y'b) = \mathtt{IP}_2(x', y') + b$), and note that $F(x'a, y'b) = \mathtt{IP}_2(x', y') + ab$. Using the fact that $\mathtt{IP}_2$ is an extractor with error 0.24 for independent sources that have min-entropy at least $k = 0.5n + 1 + \log_2(1/0.24)$ [6, Thm. 9], it follows that the same holds for F (with no loss in parameters). Using Theorems 3.4 and 3.2 it follows that $F(X,Y)$ is 0.48-close to a uniformly distributed bit, provide that X (resp., Y) has min-entropy at least $k + t + \log_2(1/0.24)$, where t is the communication complexity of the joint distribution $(X,Y) = ((X',1),(Y',\mathtt{IP}_2(X',Y')))$. Hence, we reach a contradiction if $n - 1 \geq k + t + \log_2(1/0.24)$, which implies that $t > n - k - 1 - \log_2(1/0.24) = 0.5n - 2 - 2\log_2(1/0.24) > 0.5n - 7$ must hold.

[4] E. Chattopadhyay, J. Goodman, V. Goyal, an X. Li. Extractors for Adversarial Sources via Extremal Hypergraphs. *ECCC*, TR19-184, 2019.

[5] E. Chattopadhyay and D. Zuckerman. Explicit Two-Source Extractors and Resilient Functions. *ECCC*, TR15-119, 2015.

[6] B. Chor and O. Goldreich. Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity. *SIAM Journal on Computing*, Vol. 17 (2), pages 230–261, 1988. Preliminary version in *26th FOCS*, 1985.

[7] Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz. Improved Randomness Extraction from Two Independent Sources. In *8th RANDOM*, Lecture Notes in Computer Science (Vol. 3122), pages 334–344, Springer, 2004.

[8] Y. Dodis, T. Ristenpart, an S. Vadhan. Randomness Condensers for Efficiently Samplable, Seed-Dependent Sources. In *9th TCC*, Lecture Notes in Computer Science (Vol. 7194), pages 618–635, Springer, 2012.

[9] E. Gat and S. Goldwasser. Probabilistic Search Algorithms with Unique Answers and Their Cryptographic Applications. *ECCC*, TR11-136, 2011.

[10] O. Goldreich. Another Motivation for Reducing the Randomness Complexity of Algorithms. In *Studies in Complexity and Cryptography*, Lecture Notes in Computer Science (Vol. 6650), pages 555–560, Springer, 2011.

[11] O. Goldreich. In a World of P=BPP. In *Studies in Complexity and Cryptography*, Lecture Notes in Computer Science (Vol. 6650), pages 191–232, Springer, 2011.

[12] O. Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.

[13] O. Goldreich. On the Complexity of Estimating the Effective Support Size. *ECCC*, TR19-088, 2019.

[14] X. Li. Improved Constructions of Two-Source Extractors. *ECCC*, TR15-125, 2015.

[15] A. Rao. A 2-source almost-extractor for linear entropy. In *8th RANDOM*, Lecture Notes in Computer Science (Vol. 5171), pages 549–556, Springer, 2008.

[16] R. Raz and O. Reingold. On recycling the randomness of states in space bounded computation. In *31st STOC*, pages 159168, 1999.

[17] J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, anddepth-two superconcentrators. *SIAM Journal on Disc. Math. and Alg.*, Vol. 13 (1), pages 2–24, 2000.

[18] M. Santha and U.V. Vazirani. Generating Quasi-Random Sequences from Slightly-Random Sources. *Journal of Computer and System Science*, Vol. 33 (1), pages 75–87, 1986. Preliminary version in *25th FOCS*, 1984.

[19] R. Shaltiel. Recent Developments in Explicit Constructions of Extractors. *Bulletin of the EATCS*, Vol. 77, pages 67–95, 2002.

[20] R. Shaltiel. An Introduction to Randomness Extractors. In *38th ICALP*, Part II, Lecture Notes in Computer Science (Vol. 6756), pages 21–41, Springer, 2011.

[21] S. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, Vol. 7 (1-3), pages 1–336, 2012.

Authors' addresses

***Marshall Ball*:** Computer Science Department, Columbia University, New York.

E-mail: marshall@cs.columbia.edu

***Oded Goldreich*:** Department of Computer Science, Weizmann Institute of Science, Israel.

E-mail: oded.goldreich@weizmann.ac.il

***Tal Malkin*:** Computer Science Department, Columbia University, New York.

E-mail: tal@cs.columbia.edu

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il