A Brief History of the Constructions of Independent Source Extractors

Xin Li
Johns Hopkins University
lixints@cs.jhu.edu

Randomness extractors are functions that transform biased random sources (i.e., non-uniform probability distributions, hereafter called *weak random sources*) into almost uniform random bits. Their original motivation is to bridge the gap between the uniform random bits required in standard applications (such as in randomized algorithms, distributed computing, and cryptography), and practical random sources (such as weather pattern and noise) which are almost always biased. However, researchers quickly find out that they have many other applications in areas such as complexity theory, coding theory and combinatorics.

When talking about randomness extractors, a parameter of particular interests is the *entropy* of the input weak random source. This is used to measure the quality of the random source, and the smaller the entropy is, the weaker the source is. Thus it is important to understand what is the entropy requirement for an extractor to exist, and give corresponding explicit constructions. It turns out to be impossible to construct a randomness extractor for a general weak random source if the only thing known is that it has some entropy, even if the entropy is as high as $n-1$ where n is the length of the source in binary bits, and even if one only wishes to extract a single random bit. Given this negative result, a popular and most studied model is the *independent source* model, where one has access to several (at least two) independent weak random sources. This model is popular for several reasons. First, it can be shown that with just two independent weak random sources, it becomes possible to construct randomness extractors. Furthermore, in practice it is reasonable to assume that some random sources are independent (e.g., those that are far away from each other). Second, such extractors have natural applications in distributed computing and multiparty computation which involve several parties [KLRZ08, KLR09]. Finally, this question is intimately connected to the question of constructing optimal Ramsey graphs, a central problem in graph theory which has been open since the existence of very good Ramsey graphs was first proved by Erdős in his seminal paper which inaugurated the probabilistic method [Erd47].

Historically, Chor and Goldreich [CG88] first studied this model in 1985, and showed the existence of an extractor for two independent sources with entropy $\log n + O(1)$, where n is the length of each source. Such a construction would also give a Ramsey graph on N vertices with no clique or independent set of size $O(\log N)$, matching the existence result shown by Erdős. Since then, constructing extractors for a small number of independent sources with small entropy has been a major open problem, and one which turns out to be quite difficult despite extensive research. Indeed, until 2003 there has been essentially no progress, and the best known result is still the result given by Chor and Goldreich, which only works for two sources with entropy $k > n/2$.[1]

[1] There are constructions with better parameters based on unprove conjectures.

Starting from 2004, a long line of new research has resulted in significant progress on this problem. This can be divided into roughly three generations.

Generation 1. An exciting work by Barak, Impagliazzo, and Wigderson [BIW06] introduced new techniques from additive combinatorics into the problem of constructing independent source extractors. Using these sophisticated techniques, they constructed explicit extractors for $O(1/\delta)$ independent sources with entropy $k \geq \delta n$. This is the first real progress since the work of Chor and Goldreich. This result was later improved by Barak, Kindler, Shaltiel, Sudakov, and Wigderson [BKS$^+$05] to give an explicit extractor for three independent sources with entropy $k \geq \delta n$, any $\delta > 0$. Using different techniques, Raz [Raz05] gave a two source extractor where one source needs to have entropy $k > n/2$ while the other source can have entropy as small as $k = O(\log n)$. Remarkably, again by using techniques from additive combinatorics, Bourgain [Bou05] gave a two source extractor for entropy $k \geq (1/2 - \alpha)n$, where $\alpha > 0$ is some universal constant. This improves the entropy requirement of explicit two-source extractors for the first time. However, these techniques seem insufficient to give extractors for a constant number of sources with substantially sub-linear entropy (e.g., $\sqrt{n}$ entropy).

Generation 2. The second generation of new constructions begins with the beautiful work of Rao [Rao06]. In this paper, instead of using sophisticated techniques from additive combinatorics, Rao used the idea of *somewhere random source* and compositions of extractors to give an explicit extractor that works for $O(1/\delta)$ independent sources with entropy $k \geq n^{\delta}$. His extractor thus works for a constant number of independent sources even with polynomially small entropy. Extending his techniques and using a sophisticated challenge-response mechanism, Barak, Rao, Shaltiel, and Wigderson [BRSW06] gave an explicit construction of a two-source disperser (a weaker version of extractor) for entropy $k \geq 2^{\log^{1-\alpha} n}$, where $\alpha > 0$ is some absolute constant. This also gives an explicit Ramsey graph on N vertices with no clique or independent set of size $2^{\log^{o(1)} N}$, which for the first time improves the best know construction of $2^{\widetilde{O}(\sqrt{\log N})}$ by Frankl and Wilson in 1981. However, these techniques also have their limits. They seem insufficient to give extractors for a constant number of sources with sub polynomially small entropy.

Generation 3. The final turning point came in my work around 2013 [Li13b]. In that work, I introduced completely new techniques and formed a general framework to tackle this problem. The new framework is based on the following two observations: (1) Similar to [Rao06], one can first obtain a *somewhere random source*, where some part of the source is already uniform (but we don't know which part) and the other part of the source is biased and can depend on the uniform part. Furthermore, one can model this as a set of players where some players are good and the others are malicious, and they want to jointly elect a good leader or generate a uniform random string. This establishes connections to the well studied problems of *leader election* and *collective coin flipping* in distributed computing, so one can borrow from a rich set of tools in that literature. (2) One can use a special class of extractors known as *non-malleable extractors* (introduced by Dodis and Wichs [DW09] in the context of a cryptographic problem known as *privacy amplification*) to create special structures in the players in the previous model, which makes the problem easier to handle. Based on this work, I gave the first explicit extractor for a constant number of independent sources with entropy $k \geq \mathrm{polylog}(n)$ in [Li13a]. Later, I improved this result and gave a three-source extractor for entropy $k \geq \mathrm{polylog}(n)$ [Li15], which is one step away from obtaining a two-source extractor for such small entropy.

The final missing piece was provided by the elegant work of Chattopadhyay and Zuckerman [CZ16]. Based on my previous work, they showed how to combine techniques from pseudorandom generators and resilient functions (a well studied primitive in leader election and collective coin flipping) to achieve a two-source extractor for entropy $k \geq \mathrm{polylog}(n)$. This also gives an explicit Ramsey graph on N vertices with no clique or independent set of size $2^{\mathrm{poly}\log\log N}$, although the exponent in poly is relatively large. Around the same time, Gil Cohen [Coh16b] also achieved an explicit two-source disperser with roughly the same parameters, using my work [Li15] and the challenge-response mechanism. This also gives the Ramsey graph result. Since then, the focus has been to push the limit of existing techniques to try to achieve nearly optimal entropy. Many researchers have contributed to this line of work [Li16, CS16, CL16, Coh16a, BADTS17, Coh17, Li17, Li18]. Until the time of writing, the best known two-source extractor (in terms of entropy requirement) appears in my work [Li18], which works for entropy $k = o(\log n \log\log n)$. This also gives the best known explicit Ramsey graph on N vertices, which has no clique or independent set of size $(\log N)^{o(\log\log\log N)}$.

References

[BADTS17] Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Explicit two-source extractors for near-logarithmic min-entropy. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, 2017.

[BIW06] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness using few independent sources. *SIAM Journal on Computing*, 36:1095–1118, 2006.

[BKS$^+$05] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 1–10, 2005.

[Bou05] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1:1–32, 2005.

[BRSW06] Boaz Barak, Anup Rao, Ronen Shaltiel, and Avi Wigderson. 2 source dispersers for $n^{o(1)}$ entropy and Ramsey graphs beating the Frankl-Wilson construction. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[CL16] Eshan Chattopadhyay and Xin Li. Explicit non-malleable extractors, multi-source extractors and almost optimal privacy amplification protocols. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[Coh16a] Gil Cohen. Making the most of advice: New correlation breakers and their applications. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[Coh16b] Gil Cohen. Two-source dispersers for polylogarithmic entropy and improved ramsey graphs. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, 2016.

[Coh17] Gil Cohen. Two-source extractors for quasi-logarithmic min-entropy and improved privacy amplification protocols. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*, 2017.

[CS16] Gil Cohen and Leonard Schulman. Extractors for near logarithmic min-entropy. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

[CZ16] Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, 2016.

[DW09] Yevgeniy Dodis and Daniel Wichs. Non-malleable extractors and symmetric key cryptography from weak secrets. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing*, pages 601–610, 2009.

[Erd47] P. Erdős. Some remarks on the theory of graphs. *Bulletin of the American Mathematics Society*, 53:292–294, 1947.

[KLR09] Yael Kalai, Xin Li, and Anup Rao. 2-source extractors under computational assumptions and cryptography with defective randomness. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 617–628, 2009.

[KLRZ08] Yael Tauman Kalai, Xin Li, Anup Rao, and David Zuckerman. Network extractor protocols. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 654–663, 2008.

[Li13a] Xin Li. Extractors for a constant number of independent sources with polylogarithmic min-entropy. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science*, pages 100–109, 2013.

[Li13b] Xin Li. New independent source extractors with exponential improvement. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 783–792, 2013.

[Li15] Xin Li. Three source extractors for polylogarithmic min-entropy. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, 2015.

[Li16] Xin Li. Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, 2016.

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

[Li18] Xin Li. Non-malleable extractors and non-malleable codes: Partially optimal constructions. Technical Report TR18-028, ECCC, 2018.

[Rao06] Anup Rao. Extractors for a constant number of polynomially small min-entropy independent sources. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, 2006.

[Raz05] Ran Raz. Extractors with weak random seeds. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 11–20, 2005.