How Much Randomness Can Be Extracted from Memoryless Shannon Entropy Sources?

Maciej Skorski(✉)

Cryptology and Data Security Group, University of Warsaw, Warsaw, Poland
maciej.skorski@mimuw.edu.pl

Abstract. We revisit the classical problem: given a memoryless source having a certain amount of Shannon Entropy, how many random bits can be extracted? This question appears in works studying random number generators built from physical entropy sources.

Some authors proposed to use a heuristic estimate obtained from the Asymptotic Equipartition Property, which yields roughly n extractable bits, where n is the total Shannon entropy amount. However best precise results of this form give only $n - O(\sqrt{\log(1/\epsilon)n})$ bits, where ϵ is the distance of the extracted bits from uniform. In this paper we show a matching $n - \Omega(\sqrt{\log(1/\epsilon)n})$ upper bound. Therefore, the loss of $\Theta(\sqrt{\log(1/\epsilon)n})$ bits is necessary. As we show, this theoretical bound is of practical relevance. Namely, applying the imprecise AEP heuristic to a mobile phone accelerometer one might overestimate extractable entropy even by 100 %, no matter what the extractor is. Thus, the "AEP extracting heuristic" should not be used without taking the precise error into account.

Keywords: Shannon entropy · Randomness extractors · Asymptotic equipartition property

1 Introduction

1.1 Entropy

Randomness Sources. Important computer applications, like generating cryptographic keys, building countermeasures against side-channel attacks or gambling, demand randomness of excellent quality, that is uniformly or almost uniformly distributed sequences of bits. Unfortunately, in practice we cannot generate pure randomness. Even best physical sources of randomness produce bits that are slightly biased or correlated. Sources which provide some (not maximal) amount of randomness are called *weak sources*. In practice, randomness can be gathered based on a physical phenomena (like radiation [hot], photons transmission, thermal noise [BP99], atmospheric noise [ran], jitters) or even from

M. Skorski–This work was partly supported by the WELCOME/2010-4/2 grant founded within the framework of the EU Innovative Economy Operational Programme.

H. Kim and D. Choi (Eds.): WISA 2015, LNCS 9503, pp. 75–86, 2016.
DOI: 10.1007/978-3-319-31875-2_7

a human-device interaction (like timing I/O disk and network events [dev], keystrokes or mouse movements [pgp], shaking accelerators in mobile phones [VSH11] and other ideas). Such raw randomness must be further post-processed before use, in order to eliminate bias and correlations between individual bits. While this task can be easily achieved by general-purpose tools called *randomness extractors* [BST03], the main problem is in evaluating the quality of a random source. One needs to ensure that enough randomness has been collected, depending on the chosen post-processing technique. This is the major concern in designing so called *true random number generators*, which combine randomness sources with postprocessing algorithms to generate random output of high quality from underlying weak sources. The design of a typical TRNG is illustrated in Fig. 1 below [BST03].

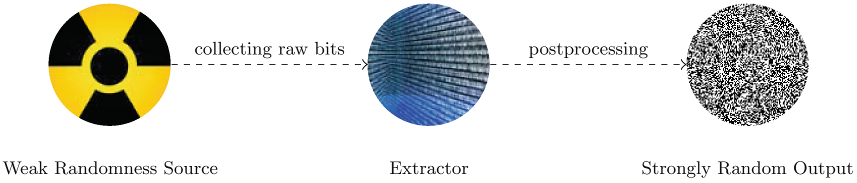

Fig. 1. True Random Number Generators. The scheme illustrates the typical design, where the building blocks are: (a) an entropy source (b) a harvesting mechanism and (c) a postprocessor (extractor). The main issue is how to ensure that enough raw bits have been collected?

Quantifying Randomness in Theory. From a theoretical point of view, to evaluate randomness within a *known* probability distribution one uses the notion of entropy. One uses different entropy definitions depending on the context. In information theory most widely used is Shannon entropy, which quantifies the encoding length of a given distribution. For any discrete random variable X its Shannon Entropy equals

$$H(X) = -\sum_{x} \Pr[X=x] \log \Pr[X=x]. \tag{1}$$

In turn, cryptographers use the more conservative notion called min-entropy, which quantifies unpredictability (in particular, min-entropy provides a bound on how much randomness can be extracted). The min-entropy of a probability distribution X is defined as

$$H_{\infty}(X) = \min_{x}(1/\log \Pr[X=x]). \tag{2}$$

In general, there is a large gap between these two measures: the min-entropy of an n-bit string may be only $O(1)$ whereas its Shannon entropy as big as $\Omega(n)$[1].

[1] Consider simply a n-bit distribution X which puts the weight 0.5 on the string 0^n and is uniform elsewhere.

Quantifying Randomness in Practice. From a practical point of view, often it happens that not only the distribution of X is unknown, but it may be even hard to assign a fitting theoretical model with some degrees of freedom (for example if we knew that X comes from a particular paremetrized family of distributions, not knowing the concrete parameters). This is in particular the case when randomness is being gathered from different sources (for example the linux random number generator). In such cases, we have no understanding of the underlying physical process and cannot conclude anything about its characteristics (like the entropy), which would be the recommended solution [KMT15]. Instead, we can only observe output samples of the source, considering it a black-box device. In this setting, we need an *entropy estimator*, which guess the entropy of an *unknown* distribution X based on its samples.

1.2 Entropy Estimating

Motivations for Entropy Estimating. We have already seen that we need entropy estimators to evaluate the quality of a source and thus the quality of the extractor output. Also, entropy estimating is motivated in practice by the fact that actually many parties may be interested in evaluating the source entropy in the context of TRNGs [BL05,KMT15]: (a) designers, when they fail to fit a good model to the source (b) testing labs, when verifying quality claimed by manufacturers and (c) developers, especially when working with multiple sources.

Shannon Entropy or Min-Entropy? Technically speaking, extractable randomness is quantified in terms of min-entropy, not Shannon entropy. However, there are two reasons for why we actually makes sense to work with Shannon entropy

(a) *Shannon entropy is easier to be reliably estimated.*
(b) *For memoryless sources, Shannon and min-entropy are comparable*

Regarding (a), we note that to estimate Shannon entropy one can use fairly general source models based on Markov chains like Maurer-Coron tests [Cor99] or measures based on mutual information [YSK13]. Also, Shannon Entropy is much easier (and efficient) to estimate in an *online* manner, where the source distribution may change over time. Such estimators are an active research area and find important applications not only in cryptography [BL05,LPR11] but also in learning, data mining or network anomaly detection [HJW15].

To discuss (b), recall that *memoryless source* (called also *stateless*) is a source which produces consecutive samples independently. While this is a restriction, it is often assumed as a part of the model by practitioners working on random number generators (cf. [LRSV12,BKMS09,BL05,DG07]) or enforced be under some circumstances (so called *certification mode* which enforces fresh samples, see [BL05,DG07]). An important result is obtained from a more general fact called Asymptotic Equipartition Property (AEP). Namely, for a stateless source the min-entropy rate (min-entropy per sample) is close to its Shannon entropy per bit (conditionally with probability almost 1), for a large number of samples.

AEP and Extracting from Memoryless Shannon Sources. We start with the following general fact, which easily follows by the Weak Law of Large Numbers.

Theorem 1 (Asymptotic Equipartition Property). *The min entropy per bit in a sequence $X_1,\ldots,X_n$ of i.id. samples from X, for large n and with high probability, is* close *to the Shannon entropy of X. More precisely*

$$-\frac{1}{n}\log P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \longrightarrow H(X) \tag{3}$$

where the convergence holds in probability (over $(x_1,\ldots,x_n)\leftarrow X_1,\ldots,X_n$).

Intuitively, the AEP simply means that, conditionally with large probability, the product of many independent copies is flat and its min-entropy approaches Shannon entropy.

Thus, the AEP is a bridge connecting the heuristic use of Shannon entropy as a measure of extractable randomness (practice) and the provable security (randomness extractors theory). The best known quantitative form of Eq. (3) appears in [Hol06].

Theorem 2 (Quantitative Asymptotic Equipartition Property *[Hol06]***).** *Let $X^n = X_1,\ldots,X_n$ be a sequence of i.i.d. samples from a distribution X of Shannon entropy k. Then the sequence $(X_1,\ldots,X_n)$ is ϵ-close in the variational distance to a distribution of min entropy $kn - O\left(\sqrt{kn\log(1/\epsilon)}\right)$.*

Now we restate the same result in language of randomness extractors

Corollary 1 (Extracting from Memoryless Shannon Sources, Lower Bound). *In particular, in the above setting, one can extract at least*

$$m = kn - O\left(\sqrt{kn\log(1/\epsilon)}\right) - 2\log(1/\epsilon) \tag{4}$$

bits which are ϵ-close to uniform (e.g. using independent hash functions [HILL99] as an extractor). Since in most settings we have[2] $\epsilon \gg 2^{-kn}$, we extract

$$m = H(X^n) - O\left(\sqrt{H(X^n)\log(1/\epsilon)}\right) \tag{5}$$

bits, that is we extract all the Shannon entropy but $O\left(\sqrt{kn\log(1/\epsilon)}\right)$ bits.

1.3 Problem Statement

We have already seen that in case of many independent copies the amount of extractable bits approaches asymptotically the Shannon entropy. We note that some works, including works on entropy estimating [BL05,LPR11] suggest to use a simplified (asymptotic) version of Eq. (5), namely

[2] Because $\epsilon \approx 2^{-kn}$ provides exponential security which is already overkill in most cases.

$$m \approx H(X^n) \tag{6}$$

bits ϵ-close to uniform, ignoring a smaller order term $O\left((H(X^n)\log(1/\epsilon))^{\frac{1}{2}}\right)$. The question which naturally arises is how much do we lose by this approximation. Is it safe to assume (heuristically) that the equality (6) holds in practical parameter regimes?

> **Question**: What is the exact error of the AEP heuristic Eq. (6)?

We rewrite this problem as the task of finding upper bounds on the extraction rate of Shannon entropy memoryless sources.

> **Question (Reformulated)**: Suppose that we have a source which produces i.i.d samples $X_1, X_2, \ldots$ each of Shannon entropy k. How much almost uniform bits can be extracted from n such samples?

This question is well-motivated as no upper bounds to Theorem 2 have been known so far (though some other works [Hol11] also address lower bounds), and because of the popularity of the AEP herustic (6).

1.4 Our Results and Applications

The Tight No-Go Result. We answer the posted question, showing that the convergence rate in Eq. (3) given in Theorem 2 is optimal.

Theorem 3 (An Upper Bound on the Extraction Rate from Shannon Sources). *From any sequence of i.i.d. binary random variables $X^n = X_1, \ldots, X_n$ we no extract can get more than*

$$m = H(X^n) - \Theta(\sqrt{H(X^n)\log(1/\epsilon)}) \tag{7}$$

bits which are ϵ-close (in the variation distance) to uniform. This matches the lower bound in [Hol06] (the constant under $\Theta(\cdot)$ depends on the source X).

Corollary 2 (A Significant Entropy Loss in the AEP Heuristic Estimate). *In the above setting, the gap between the Shannon entropy and the number of extractable bits ϵ-close to uniform equals at least $\Theta(\sqrt{\log(1/\epsilon)kn})$. In particular, for the recommended security level ($\epsilon = 2^{-80}$) we obtain the loss of $kn - m \approx \sqrt{80kn}$ bits, no matter what an extractor we use.*

An Application to TRNGs: Not To Overestimate Security. Imagine a mobile phone where the accelerometer is being used as an entropy source. Such a source was studied in [LPR11] and the Shannon entropy rate was estimated to be roughly 0.125 per bit. Since the recommended security level for almost random bits is $\epsilon = 2^{-80}$. According to the heuristic (3) we need roughly $m = 128/0.125 = 1024$ samples to extract a 128-bit key. However taking into account the true error in our Theorem 3 we see that we need at least $m \approx 2214$ bits!

An Application to the AEP: Convergence Speed. Let $X^n = X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. bit random variables. By the standard AEP we know that with probability $1-\epsilon$ over $x \leftarrow X^n$ we have $P_{X^n}(x) \leqslant 2^{-nH(X_1)+O\left(\sqrt{nH(X_1)\log(/\epsilon)}\right)}$. Our result implies that for *any* event E of probability $1-\epsilon$ for some $x \in E$ we have $P_{X^n}(x) \geqslant 2^{-nH(X_1)+\Omega\left(\sqrt{nH(X_1)\log(1/\epsilon)}\right)}$. This proves that the error term for the convergence is really $\Theta\left(\sqrt{nH(X_1)\log(1/\epsilon)}\right)$.

1.5 Organization

The remainder of the paper is structured as follows. In Sect. 2 we give some basic facts and auxiliary technical results that will be used later. The proof of the main result, that is Theorem 3, is given in Sect. 3. Finally, Sect. 4 concludes the work.

2 Preliminaries

2.1 Basic Definitions

The most popular way of measuring how two distributions are close is the statistical distance.

Definition 1 (Statistical Distance). *The statistical (or total variation) distance of two distributions X, Y is defined as*

$$\mathrm{SD}\left(X;Y\right) = \sum_x \left|\Pr[X=x] - \Pr[Y=x]\right| \tag{8}$$

We also simply say that X and Y are ϵ-close.

Below we recall the definition of Shannon entropy and min entropy. The logarithms are taken at base 2.

Definition 2 (Shannon Entropy). *The Shannon Entropy of a distribution X equals $H(X) = -\sum_x \Pr[X=x]\log\Pr[X=x]$.*

Definition 3 (Min Entropy). *The min entropy of a distribution X equals $H_\infty(X) = \min_x(1/\log\Pr[X=x])$.*

2.2 Extractors

Extractors are functions which transform inputs of some required min-entropy amount into an almost uniform string of known length. To extract from every high-entropy source (that is, to have an extractor of general purpose), one needs to allow extractors to use small amount of auxiliary randomness, which can be "reinvested" as in the case of catalysts in chemistry. Also, one has to accept some small deviation of the output from being uniform (small enough to be acceptable for almost every application) and some entropy loss [RTS00]. Good extractors, simple, provable-secure and widely used in practice, are obtained from universal hash families [CW79]. We refer the reader to [Sha11] for a survey.

Definition 4 (Randomness Extractors). *A function* $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ *is a* (k,ϵ)*-extractor*

$$\mathsf{SD}\left(\mathsf{Ext}(X,U_d),U_d;U_{m+d}\right) \leqslant \epsilon$$

for any distribution X *of min-entropy at least* k *and independent* d*-bit string* U_d*.*

Definition 5 (Extractable Entropy, [RW04]). *We say that* X *has* k *extractable bits within distance* ϵ*, denoted* $H^{\epsilon}_{\text{ext}}(X) \geqslant k$*, if for some randomized function* Ext *we have* $\text{SD}\left(\text{Ext}(X,S);U_k,S\right) \leqslant \epsilon$*, where* U_k *is a uniform* k*-bit string and* S *is an independent uniform string (called the seed).*

The so called Leftover Hash Lemma [HILL88] ensures that $H^{\epsilon}_{\text{ext}}(X) \geqslant H_\infty(X) - 2\log(1/\epsilon)$.

2.3 Technical Facts

Our proof uses the following characterization of "extractable" distributions.

Theorem 4 (An Upper Bound on Extractable Entropy, [RW04]). *If* $H^{\epsilon}_{\text{ext}}(X) \geqslant k$ *then* X *is* ϵ*-close to* Y *such that* $H_\infty(Y) \geqslant k$*.*

The second important fact we use is the sharp bound on binomial tails.

Theorem 5 (Tight Binomial Tails [McK]). *Let* $B(n,p)$ *be a sum of independent Bernoulli trials with success probability* p*. Then for* $\gamma \leqslant \frac{3}{4}q$ *we have*

$$\Pr\left[B(n,p) \geqslant pn + \gamma n\right] = Q\left(\sqrt{\frac{n\gamma^2}{pq}}\right) \cdot \psi\left(p,q,n,\gamma\right) \tag{9}$$

with the error term satisfies

$$\psi\left(p,q,n,\gamma\right) = \exp\left(\frac{n\gamma^2}{2pq} - n\text{KL}\left(p+\gamma \parallel p\right) + \frac{1}{2}\log\left(\frac{p+\gamma}{p}\cdot\frac{q}{q-\gamma}\right) + O_{p,q}\left(n^{-\frac{1}{2}}\right)\right) \tag{10}$$

where $\text{KL}\left(a \parallel b\right) = a\log(a/b) + (1-a)\log((1-a)/(1-b)$ *is the Kullback-Leibler divergence, and* Q *is the complement of the cumulative distribution function of the standard normal distribution.*

3 Proof of Theorem 3

3.1 Characterizing Extractable Entropy

We state the following fact with an explanation in Fig. 2.

Lemma 1 (An Uppper Bound on the Extractable Entropy). *Let* X *be a distribution. Then for* $\overline{\textit{every}}$ *distribution* Y *which is* ϵ*-close to* X*, twe have* $H_\infty(Y) \leqslant -\log t$ *where* t *satisfies*

$$\sum_x \max(\mathbf{P}_X(x) - t, 0) = \epsilon. \tag{11}$$

Proof (of Lemma 1). The proof follows easily by observation that the optimal mass rearrangement (which maximizes $H_\infty(Y)$) is to decrease probability mass at biggest points. Indeed, we have to find the optimal value of the following optimization program.

$$\begin{array}{ll} \text{maximize} & H_\infty(Y) \\ \text{s.t.} & \mathrm{SD}(X;Y) \leqslant \epsilon \end{array} \tag{12}$$

Note that we can write $P_Y = \epsilon(x) + P_X$ where $\sum_x \epsilon(x) = 0$ and $\sum_x |\epsilon(x)| = 2\epsilon$, transforming program (12) into

$$\begin{array}{ll} \text{maximize} & \min_x \left(\log \frac{1}{P_X(x)+\epsilon(x)} \right) \\ \text{s.t.} & \begin{cases} \sum_x \epsilon(x) = 0 \\ \sum_x |\epsilon(x)| = 2\epsilon \end{cases} \end{array} \tag{13}$$

where the optimization runs over the numbers $\epsilon(x)$. Since $u \to \log u^{-1}$ is an decreasing function when $u \in (0,1)$, the program has the same maximizer as the solution of

$$\begin{array}{ll} \text{maximize} & \min_x (P_X(x) + \epsilon(x)) \\ \text{s.t.} & \begin{cases} \sum_{x'} \epsilon(x') = 0 \\ \sum_{x'} |\epsilon(x')| \leqslant 2\epsilon \end{cases} \end{array} \tag{14}$$

For the set $S = \{x : \epsilon(x) < 0\}$, we claim that the optimal solution satisfies $P_X(x) + \epsilon(x) = \text{const}$ on $x \in S$. Indeed, otherwise we have

$$P_X(x_1) + \epsilon(x_1) < P_X(x) + \epsilon(x) \leqslant P_X(x_2) + \epsilon(x_2)$$

for some $x_1, x_2 \in S$ and every $x \in S \setminus \{x_1\}$. Then replacing $\epsilon(x_1), \epsilon(x_2)$ by $\frac{\epsilon(x_1)+\epsilon(x_2)}{2}$ increases the objective keeping the constraint, a contradiction. Thus $P_X(x) + \epsilon(x) = t_0$ whenever $\epsilon(x) < 0$. Similarly, we prove that $P_X(x) + \epsilon(x) \leqslant t_0$ for any x. Going back to Eq. (13), it suffices to observe that we have

$$\min_x \left(\log \frac{1}{P_X(x) + \epsilon(x)} \right) \leqslant -\log t_0.$$

Finally, since $-\epsilon(x) = P_X(x) - t_0$ when $x \in S$, we get $\epsilon \geqslant -\sum_x \epsilon(x) = \sum_x \max(P_X(x) - t_0, 0)$. In particular, $t_0 \geqslant t$ which gives $-\log t_0 \leqslant -\log t$ and the result follows. □

Without losing generality, we assume from now that $X \in \{0,1\}$ where $\Pr[X = 1] = p, q = 1 - p$. Define $X^n = (X_1, \ldots, X_n)$. For any $x \in \{0,1\}^n$ we have

$$\Pr[X^n = x] = p^{\|x\|} q^{n - \|x\|}. \tag{15}$$

According to the last lemma and Theorem 4, we have

$$H_{\text{ext}}^{\epsilon}(X^n) \leqslant -\log t \tag{16}$$

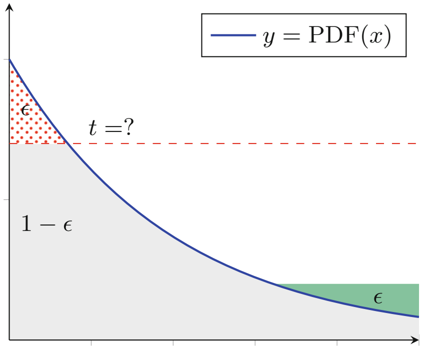

Fig. 2. The Entropy Smoothing Problem. For a given probability density function, we want to cut a total mass of up to ϵ above a possibly highest threshold (in dotted red) and rearrange it (in green), to keep the upper bound smallest possible.

where $t \in (0,1)$ is such that

$$\sum_x \max\left(\mathbf{P}_{X^n}(x) - t, 0\right) = \epsilon. \tag{17}$$

From now we assume that

$$t = p^{pn+\gamma n} q^{qn-\gamma n}. \tag{18}$$

3.2 Determining the Threshold t

The next key observation is that t is actually small and can be omitted. That is, we can simply cut the $(1-\epsilon)$-quantile. This is stated in the lemma below.

Lemma 2 (Replacing the Threshold by the Quantile). *Let $x_0 \in \{0,1\}^n$ be a point such that $\|x_0\| = pn + \gamma n$. Then we have*

$$\sum_{x\colon\ \|x\| \geqslant \|x_0\|} \max\left(\mathbf{P}_{X^n}(x) - \mathbf{P}_{X^n}(x_0)\right) \geqslant \frac{1}{2} \sum_{x\colon\ \|x\| \geqslant \|x_0\|} \mathbf{P}_{X^n}(x) \tag{19}$$

To prove the lemma, note that from Theorem 5 it follows that setting

$$\gamma' = \gamma + n^{-1}\log\left(\frac{p}{q}\right) \tag{20}$$

we obtain

$$\sum_{j \geqslant pn+\gamma' n} \binom{n}{j} \geqslant \frac{3}{4} \cdot \sum_{j \geqslant pn+\gamma n} \binom{n}{j} \tag{21}$$

when γ is sufficiently small comparing to p and q (formally this is justified by calculating the derivative with respect to γ and noticing that it is bigger by at most a factor of $1+\frac{\gamma}{\sqrt{npq}}$). But we also have

$$p^j q^{n-j} \geqslant 2 \cdot p^{(p+\gamma)n} q^{(q-\gamma)n} \quad \text{for } j \geqslant \gamma' n \tag{22}$$

Therefore,

$$\begin{aligned}\sum_{j \geqslant pn+\gamma n} \binom{n}{j} p^j q^{n-j} &\geqslant \sum_{j \geqslant pn+\gamma' n} \binom{n}{j} p^j q^{n-j} \\ &\geqslant 2 \cdot p^{(p+\gamma)n} q^{(q-\gamma)n} \cdot \sum_{j \geqslant pn+\gamma' n} \binom{n}{j} \\ &\geqslant 2 \cdot \frac{3}{4} \cdot p^{(p+\gamma)n} q^{(q-\gamma)n} \cdot \sum_{j \geqslant pn+\gamma n} \binom{n}{j} \end{aligned} \tag{23}$$

which finishes the proof.

3.3 Putting This All Together

Now, by combining Lemmas 1 and 2 and the estimate $Q(x) \approx x^{-1}\exp(-x^2/2)$ for $x \gg 0$ we obtain

$$\epsilon \geqslant \exp\left(-n\mathrm{KL}\left(p+\gamma \parallel p\right) - \log\left(\frac{n\gamma^2}{2pq}\right) + O_{p,q}(1)\right) \tag{24}$$

which, because of the Taylor expansion $\mathrm{KL}\left(p+\gamma \parallel p\right) = \frac{\gamma^2}{2pq} + O_{p,q}(\gamma^3)$, gives us

$$\gamma \geqslant \Omega\left(\sqrt{\frac{\log(1/\epsilon)}{pqn}}\right) \tag{25}$$

Setting $\gamma = c \cdot \sqrt{\frac{\log(1/\epsilon)}{pqn}}$, with sufficiently big c, we obtain the claimed result.

4 Conclusion

We show an upper bound on the amount of random bits that can be extracted from a Shannon entropy source. Even in the most favourable case, that is for independent bits, the gap between the Shannon entropy and the amount of randomness that can be extracted is significant. In practical settings, the Shannon entropy may be even 2 times bigger than the extractable entropy. We conclude that the hard error term in the AEP needs to be taken into account when extracting from memoryless Shannon sources.

Acknowledgments. The author is grateful to the organizers of the 8th Annual North American School of Information Theory in San Diego, and the organizers of the conference LATINCRYPT 2015 in Guadalajara, for the opportunity to present preliminary versions of this work.

References

[BKMS09] Bouda, J., Krhovjak, J., Matyas, V., Svenda, P.: Towards true random number generation in mobile environments. In: Jøsang, A., Maseng, T., Knapskog, S.J. (eds.) NordSec 2009. LNCS, vol. 5838, pp. 179–189. Springer, Heidelberg (2009)

[BL05] Bucci, M., Luzzi, R.: Design of testable random bit generators. In: Rao, J.R., Sunar, B. (eds.) CHES 2005. LNCS, vol. 3659, pp. 147–156. Springer, Heidelberg (2005)

[BP99] Benjamin, J., Paul, K.: The intel random number generator (1999)

[BST03] Barak, B., Shaltiel, R., Tromer, E.: True random number generators secure in a changing environment. In: Walter, C.D., Koç, Ç.K., Paar, C. (eds.) CHES 2003. LNCS, vol. 2779, pp. 166–180. Springer, Heidelberg (2003)

[Cor99] Coron, J.-S.: On the security of random sources (1999)

[CW79] Carter, J.L., Wegman, M.N.: Universal classes of hash functions. J. Comput. Syst. Sci. **18**(2), 143–154 (1979)

[dev] http://www.cs.berkeley.edu/~daw/rnd/linux-rand

[DG07] Dichtl, M., Golić, J.D.: High-speed true random number generation with logic gates only. In: Paillier, P., Verbauwhede, I. (eds.) CHES 2007. LNCS, vol. 4727, pp. 45–62. Springer, Heidelberg (2007)

[HILL88] Hstad, J., Impagliazzo, R., Levin, L.A., Luby, M.: Pseudo-random generation from one-way functions. In: Proceedings of the 20th STOC, pp. 12–24 (1988)

[HILL99] Hastad, J., Impagliazzo, R., Levin, L.A., Luby, M.: A pseudorandom generator from any one-way function. SIAM J. Comput. **28**(4), 1364–1396 (1999)

[HJW15] Han, Y., Jiao, J., Weissman, T.: Adaptive estimation of shannon entropy. CoRR abs/1502.00326 (2015)

[Hol06] Holenstein, T.: Pseudorandom generators from one-way functions: A simple construction for any hardness. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 443–461. Springer, Heidelberg (2006)

[Hol11] ——: On the randomness of repeated experiment

[hot] Hotbits project homepage. www.fourmilab.ch/hotbits/

[KMT15] Kelsey, J., McKay, K.A., Turan, M.S.: Predictive models for min-entropy estimation. IACR Cryptology ePrint Arch. **2015**, 600 (2015)

[LPR11] Lauradoux, C., Ponge, J., Röck, A.: Online Entropy Estimation for Non-Binary Sources and Applications on iPhone. Rapport de recherche, Inria, June 2011

[LRSV12] Lacharme, P., Röck, A., Strubel, V., Videau, M.: The linux pseudorandom number generator revisited, Cryptology ePrint Archive, Report 2012/251 (2012). http://eprint.iacr.org/

[McK] McKay, B.D.: On littlewood's estimate for the binomial distribution. In: Advances in Applied Probability

[pgp] Pgp project homepage. http://www.pgpi.org

[ran] Random.org project homepage. www.random.org

[RTS00] Radhakrishnan, J., Ta-Shma, A.: Bounds for dispersers, extractors, and depth-two superconcentrators. SIAM J. Discrete Math. **13**, 2000 (2000)

[RW04] Renner, R., Wolf, S.: Smooth Renyi entropy and applications. ISIT **2004**, 232 (2004)

[Sha11] Shaltiel, R.: An introduction to randomness extractors. In: Aceto, L., Henzinger, M., Sgall, J. (eds.) ICALP 2011, Part II. LNCS, vol. 6756, pp. 21–41. Springer, Heidelberg (2011)

[VSH11] Voris, J., Saxena, N., Halevi, T.: Accelerometers and randomness: Perfect together. In: WiSec 2011, pp. 115–126. ACM (2011)

[YSK13] Bong, H., Young, C., Kim, S., Yeom, Y.: Online test based on mutual information for true random number generators. J. Korean Math. Soc. **504**, 879–897 (2013)