LABORATORY FOR
COMPUTER SCIENCE

MASSACHUSETTS
INSTITUTE OF
TECHNOLOGY

MIT/LCS/TM-283

UNBIASED BITS FROM SOURCES OF WEAK RANDOMNESS AND PROBABILISTIC COMMUNICATION COMPLEXITY

BENNY CHOR
ODED GOLDREICH

SEPTEMBER 1986

545 TECHNOLOGY SQUARE, CAMBRIDGE, MASSACHUSETTS 02139

Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity

September 12, 1986

Benny Chor * Oded Goldreich **

MIT – Laboratory for Computer Science
Cambridge, Massachusetts 02139

ABSTRACT – A new model for weak random physical sources is presented. The new model strictly generalizes previous models (*e.g.* the Santha and Vazirani model [26]). The sources considered output strings according to probability distributions in which *no single string is too probable.*

The new model provides a fruitful viewpoint on problems studied previously as:

- *Extracting almost perfect bits from sources of weak randomness:* the question of possibility as well as the question of efficiency of such extraction schemes are addressed.
- *Probabilistic Communication Complexity:* it is shown that most functions have linear communication complexity in a very strong probabilistic sense.
- *Robustness of BPP* with respect to sources of weak randomness (generalizing a result of Vazirani and Vazirani [29]).

Keywords and phrases: Randomness, imperfect random bits, discrete probability distributions, probabilistic communication complexity, robustness of randomized complexity classes.

---

* Research supported in part by an IBM Graduate Fellowship and a Bantrell Postdoctoral Fellowship. Current address: Aiken Computing Laboratory, Harvard University, Cambridge, MA 02138.

** Research supported in part by a Weizmann Postdoctoral Fellowship. On leave from the Computer Science Dept., Technion, Israel.

1. INTRODUCTION

The notion of randomness is central to the theory of computation. Thus the question of whether and how randomness can be implemented in a computer is of major importance. Our intention is not to address the metaphysical aspect of the above question. We rather assume that there are physical phenomenon which appear to be "somewhat random", and study the consequences of such assumption.

In reality, there is a variety of physical sources, the output of which appears to be unpredictable in some sense (e.g. noise diodes, Geiger counters, etc.). However, these sources do not seem to be perfect (i.e. they do not output a uniform distribution). This phenomena is amplified when trying to convert the analogue signal to a digital one, and in particular when sampling the physical source very frequently.

The main contribution of this paper is in presenting a general model for sources of weak randomness. This model not only generalizes previous models, but is also very convenient to manipulate and analyze. The new model provides a new viewpoint on several problems studied previously, and enables us to obtain interesting new results:

- *Extracting almost perfect bits from sources of weak randomness:* It is shown that almost all functions can be used for extracting many "almost unbiased" bits from two independent sources of "weak" randomness. An explicit function which performs almost as good is also presented. These results yield an extraction scheme which is efficient both in terms of output entropy and computational complexity.
- *Probabilistic Communication Complexity:* It is shown that most Boolean functions have linear communication complexity in a very strong probabilistic sense. This resolves an open problem of Yao [31].
- *Robustness of BPP* with respect to sources of weak randomness. It is shown that any probabilistic polynomial-time algorithm can be modified so that it works with bits supplied by a *single* source of weak randomness.

1.1 Previous Models

Previous works on extracting unbiased bits from non-perfect sources have implicitly or explicitly proposed models of "weak randomness". Von Neumann's classic algorithm [17] deals with sequences of bits generated by independent tosses of a single coin with fixed bias. This model is totally memoryless. Blum [5] models physical sources as finite state markov chains (with unknown transition probabilities). In this model, one can describe a dependency of the next bit (output by the source) on the previous c bits (for any fixed c).

Santha and Vazirani [26] have further relaxed the restrictions on the physical source. Their model, hereafter referred to as the *SV-model*, is the start point for our investigations. In the SV-model each bit in the output sequence is "slightly random" in the sense that it is 0 with probability at least δ and 1 with probability at least δ, where $\delta \leq 1/2$ is a constant. This allows to model a probabilistic dependency of the next bit (output by the source) on all previous bits. However, no bit of the output may be totally determined by the previous bits. It follows that in the SV-model, every bit sequence is output with some positive probability. This restriction could be violated by some "random" physical sources, which are constrained in a way that prevents certain bit sequences.

1.2 The New Model

We introduce and study a general model for physical sources, hereafter referred to as the *model of Probability-Bounded sources (PRB-sources)*. Loosely speaking, the probability that a PRB-source will output a particular *string* is bounded above by some parameter. This allows the source to be very imperfect, still it may not concentrate its probability mass on too few strings.

The PRB-model is formalized using two constants l (length parameter) and b (probability bound). A physical source S is a device which outputs an infinite sequence of bits. We say that S is a (l, b)-source if for every prefix α of the output sequence, and every l-bit string β, the conditional probability that the next l bits output by S equal β is at most 2^{-b} (i.e. $Pr(\beta|\alpha) \leq 2^{-b}$).

The PRB-model is a strict generalization of the SV-model. To see the inclusion, note that any SV-source with parameter δ is a $(1, \log_2(1-\delta)^{-1})$-source. To see that the inclusion is proper, consider the $(2,1)$-source which outputs 11 with probability $1/2$ and 10 with probability $1/2$. Clearly, this source is not a SV-source. Thus, *all positive results* (with respect to the PRB-model) presented *in this paper – apply also to the SV-model.*

1.3 Extracting Unbiased Bits From Sources of Weak Randomness

Algorithms for extracting unbiased bits from non-perfect sources depend on the underlying source model. Von Neumann's algorithm [17] for generating a sequence of unbiased bits by using a coin with fixed bias, is a well-known classic:

1) Toss the biased coin twice. Denote the outcome by $\sigma\tau \in \{HH, HT, TH, TT\}$.

2) If $\sigma = \tau$ then goto step (1). (nothing is output.)

3) If $\sigma\tau = HT$ output 0; If $\sigma\tau = TH$ output 1; Goto step (1).

Elias [10] improved upon von Neumann algorithm, showing how to nearly achieve the entropy of the one coin source. He also considered special type of visible finite Markov chains. His algorithm produces perfect bits from such sources.

Blum [5] has considered extracting (perfect) unbiased bits from general finite Markov chains with unknown structure and transition probabilities. He gave algorithms which work in linear expected time. Using Elias's techniques [10], the extracted bits reach the entropy of the source in the limit.

It seems that as far as extracting *perfect* unbiased bits, Blum schemes are optimal. However, as pointed out by Santha and Vazirani [26], for practical purposes one may lower the standards and settle for "almost" unbiased bits. Having this goal in mind, they further relaxed the restrictions on the physical source and introduced the SV-model (see sec 1.1). Santha and Vazirani showed that a single SV-source cannot be used to extract almost unbiased bits, while sufficiently many independent SV-sources can be used for this purpose. Vazirani [28] showed that by applying inner-product mod 2 to strings of length $C_\delta \cdot \log_2 \varepsilon^{-1}$ output by two independent SV-sources, a bit with bias $\leq \frac{1}{2} + \varepsilon$ is produced.

Summarizing the results in [26] and [28], we conclude that the SV-model presents a sufficient condition for the extraction of almost unbiased bits from two independent physical sources. We substantially relax this condition.

In this paper we show that almost all functions can be used to extract many independent unbiased bits from the output of any two independent (l, b)-sources. To be more specific, let $m = (b - 3 - \log l)/3 > 0$, and consider extraction functions from $l + l$ bits to m bits. The m extracted bits are *almost unbiased and independent* in the sense that each m-bit string appears with probability at least $(1 - \frac{1}{2^m}) \cdot 2^{-m}$ and at most $(1 + \frac{1}{2^m}) \cdot 2^{-m}$. This is achieved by a $1 - 2^{-2^b}$ fraction of all functions from $2l$-bit strings to m-bit strings. Notice

that the number of bits we extract from the two sources is within a constant factor ($\approx \frac{1}{6}$) of the information theoretic bound, a feature not achieved in previous works [26, 28].

We also prove that, for all $b_1 + b_2 \geq l + 2 + 2\log_2 \varepsilon^{-1}$, all functions corresponding to 2^l-by-2^l Hadamard matrices can be used to extract a single bit with bias $\leq \varepsilon$ from any two independent PRB-sources which are (l, b_1) and (l, b_2) distributed respectively.

A new result contained in this paper, resolves a problem left open in the preliminary version of this work [8]: *an extraction scheme which is efficient both in terms of information rate and computation complexity.* The core of the new method is the discrete logarithm function, and its analysis is based on the method of trigonometric sums.

1.4 Probabilistic Communication Complexity

Vazirani pointed out that "good" bit-extraction functions have high communication complexity [28]. We establish further connections between the two notions. We show that functions which can be used for extracting an almost unbiased bit from two probability-bounded sources have *linear* communication complexity in a very strong sense. It follows that almost all functions, and in particular all functions corresponding to Hadamard matrices, have linear communication complexity. This resolves Yao's open problem [31] regarding the probabilistic communication complexity of random functions and of the set intersection function. (Related lower bounds on the communication complexity of random functions were presented independently by Alon, Frankl and Rödl [4] and by Orlitsky and El-Gamal [19]. Our linear ($\Omega(n)$) lower bound on the inner product modulo 2 function, improves over Vazirani's $\Omega(n/\log n)$ bound presented in [28].)

Another contribution in the field of communication complexity is the presentation of definitions and results for the case that the inputs are taken from probability-bounded distributions (i.e. distributions in which no string is too likely). This contribution is in the spirit of Vazirani's suggestion to analyze the communication complexity with respect to inputs chosen by a SV-model [28]. However, we feel that probability-bounded distributions are more natural in the context of communication complexity. We consider *randomized* protocols where the objective is to guess the value of the function with *average* success probability exceeding $\frac{1}{2} + \varepsilon$. Both the average length of a run and the average success probability are taken with respect to the "best" (for the protocol) probability-bounded distribution. We show that, even with respect to such protocols and distributions, the

average communication complexity of almost all functions is linear in the probability bound b (where no input appear with probability greater than 2^{-b}).

1.5 On the Robustness of BPP

The class R [1] and its symmetric version BPP [12] consist of problems which can be solved with high probability in polynomial time. The probability is taken over the tosses of an unbiased coin. Umesh Vazirani raised the question whether BPP problems can be efficiently solved if a (single) SV-source is producing the coin tosses. Recently, Vazirani and Vazirani have answered this question affirmatively [29]. In this paper, we generalize their result by showing that BPP problems can be efficiently solved if a (single) PRB-source is producing the coin tosses. The underlying principles of our proof originate from Vazirani and Vazirani [29], but our proof is significantly simpler.

The main idea of the proof is that while a single PRB-source is useless for producing a *single* unbiased bit, it can nevertheless be used for producing polynomially many bits, most of which are unbiased. Our key observation is that *any* function which extracts almost unbiased bits from any two independent PRB-sources, can be used for this purpose.

1.6 Organization

In Section 2, we present our basic definitions and results concerning the extraction of unbiased bits from sources of weak randomness. These results are the basis for the rest of the paper. Subsection 2.1 consists of definitions. In subsection 2.2 we present impossibility results. In subsection 2.3 we introduce the notion of flat distributions and demonstrate its importance. In subsection 2.4 we show that almost all functions extract unbiased bits from any two independent PRB-sources, and in subsection 2.5 we show that functions corresponding to Hadamard matrices also perform well.

Each of the next three sections is based on Section 2 only, and can be read indepedently of the others. In Section 3, we further study the problem of extracting unbiased bits from probability-bounded sources. In subsection 3.1 we analyze extraction schemes with respect to two efficiency measures: rate and computation complexity. In subsection 3.2 we present and analyze the "discrete logarithm" extraction scheme. In subsection 3.3 we

consider extraction from slightly dependence sources. In subsection 3.4 we consider various extensions of our model and results.

In Section 4, we present results concerning probabilistic communication complexity. In subsection 4.1 we present old ans new definitions of probabilistic communication complexity. In subsection 4.2 we prove linear lower bounds on the communication complexity of functions, and in subsection 4.3 we present almost matching upper bounds. In subsection 4.4 we suggest and investigate a robust notion of communication complexity.

In Section 5, we deal with the robustness of BPP with respect to probability-bounded sources. Conclusions and open problems appear in Section 6.

2. EXTRACTING UNBIASED BITS - PART I

In this section we present our basic definitions and results concerning the extraction of unbiased bits from sources of weak randomness. These results will be the basis for our more advanced study of the efficiency of extraction schemes, as well as our results concerning communication complexity and the robustness of BPP. In subsection 2.1, we define probability bounded sources (distributions) and robust extraction schemes. In subsection 2.2, we present impossibility results which will be later used to demonstrate the optimality of our positive results. In subsection 2.3, we introduce the notion of flat distributions and demonstrate its importance. In subsection 2.4, we use a counting argument to prove the existence of good extraction schemes. In subsection 2.5, we show that functions corresponding to Hadamard matrices constitute good extraction schemes.

2.1 Definitions

The first two definitions are used to characterizes the PRB-sources.

Definition 1: Let l be a positive integer, and $b > 0$ a real number. Let X be a random variable assuming values in $\{0,1\}^l$. We say that X is (l,b)-*distributed* if for every $\alpha \in \{0,1\}^l$, the probability that $X = \alpha$ is $\leq 2^{-b}$.

Definition 2: Let $X_1, X_2, \ldots, X_t$ be a sequence of random variables each assuming values in $\{0,1\}^l$. The random variable X_t is (l,b)-*distributed given* $X_1, \ldots, X_{t-1}$ if for every $\alpha \in \{0,1\}^{(t-1)\cdot l}$ and $\beta \in \{0,1\}^l$, $Pr\,(X_t = \beta | X_1 \cdots X_{t-1} = \alpha) \leq 2^{-b}$.

An (l,b)-source is an infinite sequence of random variables $X_1, X_2, X_3 \ldots$ each assuming values in $\{0,1\}^l$ such that for every t, the random variable X_t is (l,b)-distributed given the values of the variables X_1 through X_{t-1}.† Unless otherwise stated, all distributions are conditioned on the entire past.

The next definitions will be used in evaluating the quality of the extracted bits.

Definition 3: Let Z be a random variable assuming values in $\{0,1\}^m$. Z is said to be ε-*robust* if for every $\alpha \in \{0,1\}^m$,

$$(1-\varepsilon)\cdot 2^{-m} \leq Pr(Z = \alpha) \leq (1+\varepsilon)\cdot 2^{-m}$$

† This definition is somewhat less restrictive from the one sketched in the introduction.

Definition 4: Let X_1, X_2,..., X_s, be s independent random variables, each assuming values in $\{0,1\}^l$. A function f: $\{0,1\}^{s \cdot l} \mapsto \{0,1\}^m$ is said to be *ε-robust on X_1, X_2,..., X_s* if the random variable $Z \stackrel{\text{def}}{=} f(X_1, X_2, ..., X_s)$ is ε-robust.

A function f: $\{0,1\}^{s \cdot l} \mapsto \{0,1\}^m$ is said to be *ε-robust with respect to properties P_1, P_2,..., P_s* if f is ε-robust on every s independent random variables X_1, X_2,..., X_s, satisfying P_1, P_2,..., P_s respectively.

2.2 Impossibility Results

It is of no surprise that one probability-bounded source cannot be used to generate unbiased bits, since probability-bounded sources include SV-sources for which an impossibility result was shown [26]. Yet, a stronger impossibility result holds for our model.

Theorem 1: Let $k \geq 1$ be an integer, and $f : \{0,1\}^{k \cdot l} \mapsto \{0,1\}$ be a Boolean function. Then there exists a $\sigma \in \{0,1\}$ and a sequence of k random variables $X_1, X_2, \ldots, X_k$, each $(l, l-1)$-distributed given the previous ones, such that $f(X_1 X_2 \cdots X_k)$ is identically σ.

Proof: The proof is by induction on k. The base case $k = 1$ is easy. Without loss of generality, f attains the value 0 on at least half the inputs. Setting X_1's probability distribution to be uniform on these inputs and 0 otherwise, $f(X_1)$ is identically 0. By the induction hypothesis, for every $\alpha \in \{0,1\}^l$, there is a $\sigma \in \{0,1\}$ such that the function $f_\alpha(X_2, \ldots, X_k) = f(\alpha, X_2, \ldots, X_k)$ can be made identically σ. Without loss of generality, for at least half the α's, f_α can be made identically 1. Setting X_1's probability distribution to be uniform on these α's and 0 elsewhere, the Theorem follows. □

While a single source cannot be used at all, there is a lower bound on the robustness of functions applied to the output of two probability-bounded sources. We start with a combinatorial Lemma

Lemma 2: Let M be a L-by-N Boolean matrix. Then there exist a $\sigma \in \{0,1\}$ and a $L/16$-by-$N/2$ submatrix of M containing at least $\frac{1}{2} \cdot \frac{L}{16} \left(\frac{N}{2} + \sqrt{\frac{N}{2}} \right)$ σ-entries.

Proof: Without loss of generality, at least half the rows contain at least half 1's. We restrict ourselves to these $L/2$ rows. Fix any such row, pick $N/2$ columns at random, and let P denote the probability that at least $T = \frac{1}{2}(N/2 + \sqrt{N/2})$ of the corresponding entries are 1's. Clearly, P is minimized when each of these rows contains exactly $N/2$ ones. In that case P is the tail of a hypergeometric distribution, and by Uhlmann (see [15, ch. 6,

sec. 5, p. 151]) is bounded below by the corresponding tail of the binomial distribution. That is

$$P \geq 2^{-N/2} \sum_{i=T}^{N/2} \binom{N/2}{i}.$$

This last expression can be approximated by the normal distribution, and in particular is bounded below by $1 - \Phi(1) \geq 1 - 0.8413 > 1/8$. By standard probabilistic arguments, this implies that there is a choice of $N/2$ columns and $(1/8) \cdot L/2 = L/16$ rows which have the desired proportion of 1's. □

Theorem 3: Let $b \leq l - 1$, and $f : \{0,1\}^{2l} \mapsto \{0,1\}$ be an arbitrary Boolean function. Then there exist a $\sigma \in \{0,1\}$ and two independent random variables X and Y, such that X is $(l, l-4)$-distributed and Y is (l,b)-distributed, and $Pr(f(X,Y) = \sigma) > \frac{1}{2}\left(1 + 2^{-b/2}\right)$.

Proof: View f as a 2^l-by-2^l Boolean matrix, with the (i,j)-th entry specifying $f(i,j)$. Let $L = 2^l$ and $N = 2^{b+1}$. Applying Lemma 2 to an arbitrary L-by-N submatrix S, there exist a σ and a 2^{l-4}-by-2^b submatrix S' of S with a fraction $\frac{1}{2}\left(1 + 2^{-b/2}\right)$ of σ-entries. Making X flat on the rows of S', and Y flat on its columns, we get the desired result. □

The above argument was based on estimating the probability that the number of ones in randomly selected columns, is at least one standard deviation away from the mean. One can consider the probability that this number is several standard deviations away from the mean. This yields a bigger bias but fewer rows, and thus a more concentrated X. Thus, for every constant ν and sufficiently large b, there exist $\sigma \in \{0,1\}$ such that $Pr(f(X,Y) = \sigma) > \frac{1}{2}\left(1 + 2^{-(b-\nu)/2}\right)$.

When b is very small, the situation is even worse.

Theorem 4: Let $b \leq \log_2(l - \log_2 l) - 1$, and $f : \{0,1\}^{2l} \mapsto \{0,1\}$ be an arbitrary Boolean function. Then there exist a $\sigma \in \{0,1\}$ and two independent random (l,b)-distributed variables X and Y, such that $Pr(f(X,Y) = \sigma) = 1$.

Proof: Consider (arbitrarily) the first $r \stackrel{\text{def}}{=} 2^{b+1}$ columns in the 2^l-by-2^l matrix of f. This defines a 2^l-by-r submatrix. There is a r-bit string which occurs in at least $2^l/2^r$ rows of the submatrix. Pick these rows. Let $\sigma \in \{0,1\}$ be a bit which occurs $t \geq r/2$ times in each of these rows. Picking the t columns containing σ, we get a 2^{l-r}-by-t submatrix with identical entries σ. As $2^{l-r} > 2^b$ and $t \geq 2^b$, this submatrix corresponds to a pair of (l,b)-distributed variables. □

2.3 Flat Distributions

In this subsection we introduce the notion of flat distributions. The importance of this notion steams from the two fact. First, as we will shortly show, the worse behaviour of extraction functions occurs on flat distributions. Second, as demonstrated through the paper, flat distributions are very easy to deal with.

Definition 5: Let X be a random varible assuming values in $\{0,1\}^l$, and $S \subset \{0,1\}^l$. We say that X is *equi-probable on* S if for every $\alpha, \beta \in S$,

$$Pr(X = \alpha) = Pr(X = \beta) .$$

We say that X is *flat on* S if X is equi-probable on S and for $\alpha \notin S$, $Pr(X = \alpha) = 0$. We say that X is (l,b)-*flat* if X is (l,b)-distributed and there exist some S such that X is flat on S.

For simplicity, we assume throughout this section that 2^b, 2^{b_1} and 2^{b_2} are integers. Flat distributions are interesting because the "worst case behaviour" of a function occurs on them. Namely,

Lemma 5: For every function f: $\{0,1\}^{2l} \mapsto \{0,1\}^m$ and every $\alpha \in \{0,1\}^m$

$$\sup_{\substack{X,Y \text{ are independent} \\ X \text{ is } (l,b_1)\text{-distributed} \\ Y \text{ is } (l,b_2)\text{-distributed}}} \{Pr\left(f(X,Y) = \alpha\right)\} = \max_{\substack{X,Y \text{ are independent} \\ X \text{ is } (l,b_1)\text{-flat} \\ Y \text{ is } (l,b_2)\text{-flat}}} \{Pr\left(f(X,Y) = \alpha\right)\}$$

and

$$\inf_{\substack{X,Y \text{ are independent} \\ X \text{ is } (l,b_1)\text{-distributed} \\ Y \text{ is } (l,b_2)\text{-distributed}}} \{Pr\left(f(X,Y) = \alpha\right)\} = \min_{\substack{X,Y \text{ are independent} \\ X \text{ is } (l,b_1)\text{-flat} \\ Y \text{ is } (l,b_2)\text{-flat}}} \{Pr\left(f(X,Y) = \alpha\right)\}$$

Proof: Denote $p_i = Pr(X = i)$ and $q_j = Pr(Y = j)$. Let $f_\alpha(i,j) = 1$ if $f(i,j) = \alpha$ and 0 otherwise. Then

$$\begin{aligned} P_\alpha(X,Y) &\stackrel{\text{def}}{=} Pr\left(f(X,Y) = \alpha\right) \\ &= \sum_{i,j} Pr(X = i, Y = j) \cdot f_\alpha(i,j) \\ &= \sum_{i,j} p_i q_j f_\alpha(i,j) . \end{aligned}$$

The last equality follows from the independence of X and Y. $P_\alpha(X,Y)$ is a function of the variables p_i, q_j, and it attains a global maximum in the range $0 \le p_i \le 2^{-b_1}$, $0 \le q_j \le 2^{-b_2}$, $\sum_i p_i = \sum_j q_j = 1$. We look for a characterization of this global maximum. Fixing the probability distribution of X (i.e. fixing the p_i's), $P_\alpha(X,Y)$ is a linear program in the q_j's, subject to the constraints $0 \le q_j \le 2^{-b_2}$, $\sum_j q_j = 1$. Applying linear programming techniques [21, ch. 2], one can verify that every basic feasible solutions has exactly 2^{b_2} non-zero variables q_j, each equal 2^{-b_2}. Thus we have shown that for every fixed X the distributions which maximize/minimize the value $P_\alpha(X,Y)$, over all possible choices of (l,b)-distributions, are flat distributions. The same obviously holds for fixed Y. Now let X_0, Y_0 be the pair of (l,b_1)-distribution and (l,b_2)-distribution where $P_\alpha(X,Y)$ attains its maximum. Then both X_0 and Y_0 must be flat. Note that the characterization holds for *any* function f. □

We demonstrate the utility of Lemma 5 by using it to argue that the following Boolean function $f : \{0,1\}^2 \times \{0,1\}^2 \mapsto \{0,1\}$ (tabulated below) is $\frac{1}{2}$-robust with respect to all pairs of indepedent $(2,1)$-distributed variables.

$X \backslash Y$	00	01	10	11
00	0	0	1	1
01	1	0	0	1
10	0	1	0	1
11	1	0	1	0

Using Lemma 5, it suffices to consider the behaviour of the function on all pairs of indepedent $(2,1)$-flat variables. There are $6^2 = 36$ possibilities altogether, each corresponding to a 2-by-2 submatrix of this table. It is readily verified that no such submatrix contain all 1's (or all 0's). Thus, for every pair of independent $(2,1)$-distributed variables, X and Y, we have $\frac{1}{4} \le Pr(f(X,Y) = 1) \le \frac{3}{4}$.

2.4 A Counting Argument

In this subsection we show that two independent probability-bounded sources can be used to get almost unbiased bits. In fact we show that almost all functions can be used for this purpose. To this end, we use the characterization of the distributions on which the "worst

case behaviour" of any function occurs (i.e. Lemma 5), and apply a counting argument to estimate the fraction of functions which are good with respect to all flat distributions.

It is helpful to note that there is a natural correspondance between flat distributions and the set of strings on which they are concentrated. This suggests the following notation: Let Z be an arbitrary fixed (l,b)-flat variable. We write $z \in S_Z$ if $Pr(Z=z) = 2^{-b}$.

In the next lemma we consider two fixed and independent flat variables, and bound below the fraction of functions which are ε-robust for these two *specific* variables.

Lemma 6: Let X and Y be two independent distributions, such that X is (l,b_1)-flat and Y is (l,b_2)-flat, and $0 < \varepsilon < 1/2$. The fraction of functions $f : \{0,1\}^{2l} \mapsto \{0,1\}^m$, which are ε-robust on X and Y, is at least $1 - 2^{m-\varepsilon^2 2^{b_1+b_2-m-2}}$.

Proof: We say that a function $f : \{0,1\}^{2l} \mapsto \{0,1\}^m$ is *ε-bad on the string α* ($\alpha \in \{0,1\}^m$) if

$$\frac{|\{(x,y) \in S_X \times S_Y : f(x,y) = \alpha\}|}{2^{b_1+b_2}} \notin \left[(1-\varepsilon)\cdot 2^{-m}, (1+\varepsilon)\cdot 2^{-m}\right] .$$

Let $P_{\alpha,\varepsilon}$ denote the fraction of functions which are ε-bad on the string α. To study $P_{\alpha,\varepsilon}$ we consider the probability space of the functions in $f : \{0,1\}^{2l} \mapsto \{0,1\}^m$ taken with uniform distribution. For each $i \in S_X$ and $j \in S_Y$, let the random variables $\varsigma_{i,j}$ be defined as follows:

$$\varsigma_{i,j} = \begin{cases} 1 & \text{if } f(i,j) = \alpha \\ 0 & \text{otherwise} \end{cases}$$

Then

$$P_{\alpha,\varepsilon} = Pr\left(\frac{1}{2^{b_1+b_2}} \left| \sum_{i \in S_X, j \in S_Y} \left(\varsigma_{i,j} - \frac{1}{2^m} \right) \right| \geq \frac{\varepsilon}{2^m} \right)$$

Recall the Chernoff Bound [25, ch. VII, sec. 4, Th. 2]: Let $\varsigma_1, \varsigma_2, \ldots, \varsigma_t$ be independent random variables with $Pr(\varsigma_i = 1) = p$ and $Pr(\varsigma_i = 0) = 1-p$, where $p \leq 1/2$. Then for all $0 < \delta \leq p(1-p)$ we have

$$Pr\left(\frac{1}{t} \cdot \left| \sum_{i=1}^{t} (\varsigma_i - p) \right| \geq \delta \right) \leq 2 \cdot \exp\left[- \frac{t\delta^2}{2p(1-p)\left(1 + \frac{\delta}{2p(1-p)}\right)^2} \right]$$

Letting $p = 2^{-m}$, $t = 2^{b_1+b_2}$ and $\delta = 2^{-m} \cdot \varepsilon$, we get

$$P_{\alpha,\varepsilon} < \exp\left[-\varepsilon^2 2^{b_1+b_2-m}/4.5\right] .$$

Switching to base 2 and summing over all possible α's, the probability that a function $f : \{0,1\}^{2l} \mapsto \{0,1\}^m$ is ε-bad on some $\alpha \in \{0,1\}^m$ is at most

$$\sum_{\alpha \in \{0,1\}^m} P_{\alpha,\varepsilon} < 2^{m-\varepsilon^2 2^{b_1+b_2-m-2}} .$$

The Lemma follows. □

We are now ready to prove that almost all functions work.

Theorem 7: Let $1 \leq m \leq b$ be an integer, and $0 < \varepsilon \leq 1/2$. Let F be the set of functions $\{f : \{0,1\}^{2l} \mapsto \{0,1\}^m\}$.

1) Let $G_{b,\varepsilon} \subset F$ the set of functions which are ε-robust for *any* two independent (l,b)-distributed random variables. If $m + 2\log_2 \varepsilon^{-1} \leq b - 2 - \log_2(2l+1)$ then

$$\frac{|G_{b,\varepsilon}|}{|F|} > 1 - 2^{-2^b} .$$

2) Let $H_{b,\varepsilon} \subset F$ the set of functions which are ε-robust for *any* two independent random variables which are (l,b_1)-distributed and (l,b_2)-distributed respectively, where $b_1 + b_2 \geq 2b$. If $m + 2\log_2 \varepsilon^{-1} \leq 2b - l - 5$ then

$$\frac{|H_{b,\varepsilon}|}{|F|} > 1 - 2^{-2^l} .$$

Proof: We start with (1) and then sketch the necessary modifications for the proof of (2). By Lemma 5, $f \in G_{b,\varepsilon}$ if and only if it is ε-robust for every two independent (l,b)-flat random variables. By Lemma 6, the fraction of functions in F which fail on a *particular* pair of independent (l,b)-flat variables is double-exponentially vanishing ($< 2^{m-\varepsilon^2 2^{2b-m-2}}$). Evidently, the fraction of functions which could fail on *some* pair of independent (l,b)-flat variables, is at most the number of pairs of (l,b)-flat variables times the above fraction. Let N_b denote the number of (l,b)-flat variables. Clearly

$$N_b = \binom{2^l}{2^b} < \frac{2^{l \cdot 2^b}}{2^b}$$

Thus,

$$\frac{|G_{b,\varepsilon}|}{|F|} \geq 1 - N_b^2 \cdot 2^{m-\varepsilon^2 2^{2b-m-2}} > 1 - 2^{2^b \cdot (2l - \varepsilon^2 2^{b-m-2})} .$$

Since $m + 2\log_2 \varepsilon^{-1} \leq b - 2 - \log_2(2l+1)$, we get $2l - \varepsilon^2 2^{b-m-2} \leq -1$, and (1) follows.

For every fixed b_1 and b_2, the fraction of functions in F which fail on a *particular* pair of independent variables which are (l, b_1)-flat and (l, b_2)-flat resp., is $< 2^{m-\varepsilon^2 2^{b_1+b_2-m-2}}$. We multiply this fraction by $\left(2^{2^l}\right)^2 = 2^{2^{l+1}}$, which is an obvious upper bound on the number of possible pairs of flat variables. We get

$$\frac{|H_{b,\varepsilon}|}{|F|} \geq 1 - 2^{2^{l+1}} \cdot 2^{m-\varepsilon^2 2^{2b-m-2}} > 1 - 2^{2^{l+1}+m-\varepsilon^2 2^{2b-m-2}} .$$

Since $m + 2\log_2 \varepsilon^{-1} \leq 2b - l - 5$, we get $2^{l+1} + m - \varepsilon^2 2^{2b-m-2} \leq -2^l$, and (2) follows. □

There is a trade-off between m, the number of extracted bits, and ε, the robustness of these bits. Some cases of special interest are listed below:

1) Setting $m = b - 4 - \log_2(2l+1)$ and $\varepsilon = 1/2$, we convert two independent (l, b)-sources to a single $(m, m-1)$-source. Intuitively, this convertion is very efficient in terms of rate: even if the entropy of the input sources is b units per each block, we extract a block of $\approx b$ bits with entropy $\approx b$.

2) Setting $m = \left(b - 2 - \log_2(2l+1)\right)/3$ and $\varepsilon = 2^{-m}$, we see that most functions can be used to extract many high quality bits per each block of the two independent (l, b)-sources.

3) Setting $m = 1$ and $\varepsilon = 2^{-(b-3-\log_2(2l+1))/2}$, we see that all but a 2^{-2^b} fraction of the Boolean functions are ε-robust with respect to two independent (l, b)-sources. This bias is almost optimal: Theorem 3 states that no Boolean function can be $2^{-(b-O(1))/2}$-robust with respect to such sources. Theorem 4 asserts that "nothing can be extracted" if $b < \log_2(l - \log_2 l) - 1$.

4) Setting $m = 1$ and $\varepsilon = 2^{-(2b-6-l)/2}$, we see that all but a 2^{-2^l} fraction of the Boolean functions are ε-robust for any pair of (l, b_1), (l, b_2) sources satisfying $b_1 + b_2 = 2b$.

2.5 Hadamard Matrices

In the subsection 2.3 we showed that the bias of a Boolean function $f : \{0,1\}^{2l} \mapsto \{0,1\}$ with respect to two independent (l, b)-sources, can be estimated by considering flat distributions only. Viewing f as a 2^l-by-2^l ± 1 matrix, this corresponds to taking all 2^b-by-2^b submatrices of f (not necessarily consecutive), and estimating the maximum submatrix

elements' sum. While in the previous section we showed that most functions have small submatrix sum, this section deals with a specific class of functions, whose matrices are Hadamard matrices.

A *Hadamard matrix* is a ± 1 matrix in which every two distinct rows (columns) are orthogonal (see [13, ch. 14] and [16, ch. 2, sec. 3]). Hadamard matrices are a subject of rich literature. In particular it is well known that submatrices of any Hadamard matrix are "balanced". In order to make the paper self contained, we present a proof of this fact, following Erdös and Spencer [11, p. 88].

Lemma 8 (J.H. Lindsey): Let $H = \left(h_{i,j}\right)$ be a t-by-t Hadamard matrix. Then the sum of elements in every r-by-s submatrix of H is at most $\sqrt{s \cdot r \cdot t}$.

Proof: Since orthogonality is preserved under any row and column permutation, it suffices to consider $\left|\sum_{i=1}^{r}\sum_{j=1}^{s} h_{i,j}\right|$, the sum of elements in the leftmost/uppermost $r \times s$ submatrix. Let $\vec{h}_i$ denote the i-th row of H, and

$$\vec{I} \stackrel{\text{def}}{=} (\overbrace{1,1,\ldots,1}^{s \text{ ones}},\overbrace{0,0,\ldots,0}^{t-s \text{ zeros}}) .$$

Then by Cauchy-Schwartz inequality

$$\begin{aligned}\left|\sum_{i=1}^{r}\sum_{j=1}^{s} h_{i,j}\right| &= \left|\sum_{i=1}^{r} \vec{h}_i \cdot \vec{I}\right| \\ &\leq \left\|\sum_{i=1}^{r} \vec{h}_i\right\|_2 \cdot \left\|\vec{I}\right\|_2 \\ &= \left\|\sum_{i=1}^{r} \vec{h}_i\right\|_2 \cdot \sqrt{s} .\end{aligned}$$

Since the $\vec{h}_i$ are orthogonal,

$$\left\|\sum_{i=1}^{r} \vec{h}_i\right\|_2 = \sqrt{\sum_{i=1}^{r} \left\|\vec{h}_i\right\|_2^2} \leq \sqrt{r \cdot t}$$

and the bound on $\left|\sum_{i=1}^{r}\sum_{j=1}^{s} h_{i,j}\right|$ follows. □

Theorem 9: Let M be an 2^l-by-2^l Hadamard matrix corresponding to the Boolean function f (i.e. $f(i,j) = \frac{1+M_{i,j}}{2}$). Suppose $b_1 + b_2 = l + 2 + 2\log_2 \varepsilon^{-1}$, where $\varepsilon < 1$. Then

the function f is ε-robust with respect to any pair of independent random variables X, Y which are (l, b_1)-distributed and (l, b_2)-distributed, respectively.

Proof: By Lemma 5, it suffices to show that every 2^{b_1}-by-2^{b_2} submatrix has relatively small elements' sum. Substituting in Lemma 8, $r = 2^{b_1}$, $s = 2^{b_2}$ and $t = 2^l$, the submatrix sum is at most $\frac{\varepsilon}{2} \cdot 2^{b_1+b_2}$. □

Subsequently Noga Alon proved a more general statement (without using Lemma 5) [3].

Remarks:

1) The case where $b_2 = l - 1$ will be useful in Section 9. We get that any Hadamard matrix is $2^{-(b-3)/2}$-robust with respect to any pair of independent random variables which are $(l, l-1)$-distributed and (l, b)-distributed, respectively.

2) Inner-product modulo 2 corresponds to a special form of Hadamard matrices, known as Sylvester matrices. This provides an alternative proof for Vazirani's Theorem [28], for the case $\delta > 1 - \sqrt{1/2} \approx 0.293$ (but not for smaller δ's).

For inner-product modulo 2, Theorem 9 cannot be significantly improved (with respect to probability bounded sources).

Proposition 10: Let $b_1 + b_2 \leq l - 4 + 2\log_2 \varepsilon^{-1}$, where $\varepsilon < 1$. Then the inner-product modulo 2 function is NOT ε-robust on some pair of independent, (l, b_1)-distributed and (l, b_2)-distributed variables.

Proof: First, consider the case where $b_1 + b_2 \leq l$. Picking X to be flat on strings of the form $0^{l-b_1}\{0,1\}^{b_1}$ and Y to be flat on $\{0,1\}^{b_2}0^{l-b_2}$ the inner product of X and Y is identically 0. For the case $b_1 + b_2 > l$, repeating exactly the same construction does not yield the desired bias. However, we can modify it using Theorem 2. Let $\Delta \stackrel{\text{def}}{=} b_1 + b_2 - l$. Consider the following family of (l, b_1)-distributed variables $\mathcal{X}$. Each variable $X \in \mathcal{X}$ is the concatenation of three independent variables X_1, X_2, X_3, where X_1 is uniformly distributed over $\{0,1\}^{b_1-\Delta-4}$, X_2 is $(\Delta+8, \Delta+4)$-distributed, and X_3 has $l - b_1 - 4$ bits which are identically 0. Similarly, $Y = Y_1Y_2Y_3 \in \mathcal{Y}$ satisfies Y_1 is identically 0^{l-b_2-4}, Y_2 is $(\Delta+8, \Delta+4)$-distributed, and Y_3 is uniformly distributed over $\{0,1\}^{b_2-\Delta-4}$. For every pair $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$, the inner-product of X and Y equals the inner-product of X_2 and Y_2. Since both X_2 and Y_2 could be any $(\delta+8, \delta+4)$-distributed variables, by Theorem 3 their inner-product may have bias $> \frac{1}{2}(1 + 2^{-(\delta+4)/2}) = \frac{1}{2}(1 + 2^{(l-b_1-b_2-4)/2})$. Thus, for $\varepsilon \leq 2^{(l-b_1-b_2-4)/2})$, the function is not ε-robust. □

The last proposition demonstrates an inherent limitation of the inner-product function with respect to (l, b)-distributions when $b \leq l/2$. This limitation need not be shared by all Hadamard matrices. In fact, a simple construction, known as *the Paley Graph*, is conjectured in the combinatorial folklore to have a stronger imbalance (small submatrix sum) property.

Let p be a prime, and $\left(\frac{i}{p}\right)$ be the Legendre symbol of the residue $i \bmod p$. The matrix M with $M_{i,j} = \left(\frac{i-j}{p}\right)$ is "almost" Hadamard [16, p. 47], as for any $0 \leq a < b \leq p-1$,

$$\sum_{i=0}^{p-1} \left(\frac{i-a}{p}\right) \cdot \left(\frac{i-b}{p}\right) = -1 \,.$$

Thus, with minor modifications, Theorem 9 applies also to the matrix M.

Conjecture: For *any* constant $0 < \mu < 1$, there exists a constant $1.5\mu < c_\mu < 2\mu$ such that every p^μ-by-p^μ submatrix of M has elements' sum at most p^{c_μ}, for large enough p.

Remark: By Theorem 3, the constant c_μ must satisfy $c_\mu > 1.5\mu$.

Corollary 11: Let $f(i,j) = \frac{1}{2} \cdot \left(1 + \left(\frac{i-j}{p}\right)\right)$. Under the Paley Graph conjecture, the function f is $p^{c_\mu - 2\mu}$-robust with respect to any pair of independent $(\log_2 p, \mu \cdot \log_2 p)$-distributed random variables.

3. EXTRACTING UNBIASED BITS - PART II

In this section we further investigate the problem of extracting unbiased bits from probability-bounded sources. In subsection 3.1 we introduce two efficiency measures: rate and computation complexity and consider extraction schemes, arising from our results, with respect to these measures. In subsection 3.2 we present and analyze the discrete logarithm extraction scheme (this result did not appear in the preliminary version of this paper [8]). In subsection 3.3 we consider extraction from slightly dependence sources. In subsection 3.4 we further extend the probability-bounded model: we consider sources with lower bound on entropy, sources with varying block length and probability bound.

3.1 Efficiency Considerations

Before going any further, let us discuss the significance of the results presented in Section 2. A first moral is that it is possible to generate a sequence of almost unbiased and independent bits from the output of two independent probability-bounded sources. Once the question of possibility is resolved, we are interested in the efficiency of the extraction schemes. We consider two measures – *rate* and *computational complexity*, both with respect to the desired robustness.

3.1.1 Efficiency Measures

In Section 2, we have considered functions which operate on corresponding l-bit blocks of two (l,b)-sources. More generally, we now consider deterministic algorithms (families of functions) which may use several blocks from each (l,b)-source at a time. In order to decrease the bias $\varepsilon(n)$ as n increases, the number of blocks used by the algorithm depends on the security parameter n. Of special interest is the case where $\varepsilon^{-1}(n)$ grows faster than any polynomial. In this case the extracted bits are as good as perfect bits for all "$poly(n)$ purposes".

In the following defintion of an extraction scheme, $\varepsilon(n)$ denotes the robusteness of the scheme, $s(n)$ denotes the number of bits taken from each source, $e(n)$ denotes the number of extracted bits, and c denotes the number of (l,b)-sources used.

Definition 6: Let ε be a function from integers to the interval $(0,1)$, and s,e be functions from integers to integers. Let c,l be integers and b $(0 \leq b \leq l)$ be a real. An

$(\varepsilon(\cdot), s(\cdot), e(\cdot), l, b, c)$-*extraction scheme* is a family of functions $\{f_n\}$ such that for every n the following holds:

1) For every $\alpha_1, \alpha_2, ..., \alpha_c \in \{0,1\}^{s(n)}$ $f_n(\alpha_1, \alpha_2, ..., \alpha_c) \in \{0,1\}^{e(n)}$.

2) The function $f_n : \{0,1\}^{c \cdot s(n)} \mapsto \{0,1\}^{e(n)}$ is $\varepsilon(n)$-robust with respect to any c independent variables $X_1, X_2, ..., X_c$, where each of the X_i's is the first $s(n)$ bits output by some (l, b)-source.

The efficiency of an algorithmic scheme should be evaluated with respect to the resources it uses. In the setting of randomness extraction schemes the resources to be considered are the input "randomness" and the deterministic computation required to effect the extraction. We measure the efficiency with respect to the randomness resource by the ratio of the entropy entering the extraction scheme and the entropy leaving it.

Definition 7: The *rate* of an $(\varepsilon(\cdot), s(\cdot), e(\cdot), l, b, c)$ extraction scheme is defined as

$$r(n) \stackrel{\text{def}}{=} \frac{e(n) \cdot H_O}{c \cdot s(n) \cdot H_I} ,$$

where H_O is the entropy of each output bit ($H_O \approx 1 - \varepsilon^2(n)$), and $H_I = b/l$ is a lower bound on the average entropy of each input bit. If there exist a constant $r > 0$ such that for every n, $r(n) > r$ then we say that the extraction scheme has *constant rate.*

For every n, the rate $r(n)$ is the ratio of the input and output entropies to f_n. The entropy of the input is taken by the worse possible one since the extraction scheme cannot "adapt" to better sources without an explicit guarantee.

Definition 8: The *computational complexity* of an extraction scheme, $\{f_n\}$, is defined as the complexity of a family of circuits $\{C_n\}$ such that for every n, C_n implements f_n.

3.1.2 Efficient Extraction Schemes

We now present several extraction schemes, which follow immediately from the results presented in Section 2, and analyze their performance with respect to the above efficiency measures. Following is a simple but important observation used in developing these schemes: For every integer $q > 1$, an (l, b)-source is also a $(q \cdot l, q \cdot b)$-source.

A Rate Efficient Scheme which is not computationally efficient

One consequence of Theorem 7 is that for every l and b and every desired bias $\varepsilon(\cdot)$, there is a non-uniform circuit family $\{C_n\}$ which extracts bits at constant rate from any two independent (l,b)-sources. This is obtained by letting $s(n) = \frac{3l}{b}\log_2 \varepsilon^{-1}(n)$, $e(n) = \log_2 \varepsilon^{-1}(n)$, and using Theorem 7 with respect to two $(s(n), \frac{b}{l}\cdot s(n))$-sources. Since the entropy per input block may be b, the rate is $\approx \frac{1}{6}$. The size of C_n is $\varepsilon^{-6l/b}(n)\cdot\log_2 \varepsilon^{-1}(n)$.

Computationally Efficient Schemes which do not have constant rate

By Theorem 9, for every l and $b > l/2$ and every desired bias $\varepsilon(\cdot)$, taking the binary inner-product of $l\cdot(b-\frac{l}{2})^{-1}\log_2\varepsilon^{-1}(n)$ bits from two independent (l,b)-sources, a single $\varepsilon(n)$-robust bit is extracted. While this yields an efficient algorithm, its rate is $1/\Theta\left(\log\varepsilon^{-1}(n)\right)$.

Under the number theoretic conjecture of subsection 2.5, efficient algorithms exist for *any* l and b. For every desired bias $\varepsilon(\cdot)$, let $p > \varepsilon^{-C(l/b)}(n)$ be a prime (where $C(\varrho) = O(\varrho)$). Taking $log_2 p$ bits from each of the two independent (l,b)-sources, and computing the Legendre symbol of their integer difference modulo p, we get an $\varepsilon(n)$-robust bit. The extracted bit can be computed by an algorithm running in time polynomial in $\log_2\varepsilon^{-1}(n)$ (and l/b).

A direct consequence of Theorem 7 is that for every l and $b > 5 + \log_2 l$ there exist a table of size 2^{2l} which transforms two independent (l,b)-sources into one $(m, m-2^{-m})$-source, where $m = (b-3-log_2 l)/3$. In other words, we can transform two independent but very weak sources into one source which is quite good (although it is far from being "almost perfect"). For every bias $\varepsilon(n)$, using the inner-product function on the output of the virtual $(m, m-2^{-m})$-source and a third independent (l,b)-source, we get the desired bias. We conclude that for every $0 < b \leq l$ and $\varepsilon(\cdot)$, there is a fast algorithm (running in time $O(\log_2\varepsilon^{-1}(n)$) that on input n and access to three independent (l,b)-sources, generates $\varepsilon(n)$-robust bits.

The problem of finding an extraction scheme which combines both rate and computational efficiency was left open in our preliminary report [8]. This was true even for the SV-model. In the following subsection we present a solution to that problem.

3.2 An Extraction Scheme Efficient in Both Measures

In this subsection we present an extraction scheme based on k-th power residues modulo a prime, which is efficient both in terms of information rate and computation complexity. This scheme is a generalization of the Paley graph construction, and was developed through conversations with László Babai. We begin this subsection by presenting the scheme. We then use results of Ajtay, Babai, Hanjal, Komlos, Pudlàk, Rödl, Szemerédi, and Turán [2] to show that the scheme has high robustness. To guarantee high information rate, our scheme uses large values of k, and is related to computing (partial) discrete logarithms in Z_p. We investigate the conditions under which the scheme is efficiently computable, and show that primes satisfying these conditions can be precomputed in expected polynomial time, given access to two probability-bounded sources.

Definition: Let p be a prime, g a primitive element of Z_p, and $k > 1$ an integer dividing $p-1$. We define $f_k : Z_p \times Z_p \mapsto \{0,1,\ldots,k-1\}$ by $f_k(x,y) = \left(\log_g(x-y)\right) \bmod k$.

Comments:

1) For $\alpha \in \{0,1,\ldots,k-1\}$, let $R_\alpha = \{g^{\alpha+ik} : 0 \le i \le (p-1)/k\}$. Then $f_k(x,y) = \alpha$ iff $x - y = z$ for some $z \in R_\alpha$.

2) By restricting the function to a subset of Z_p, f_k can be viewed as a function from $\{0,1\}^l \times \{0,1\}^l$ to $\{0,1,\ldots,k-1\}$, for $l = \lfloor \log_2 p \rfloor$.

3) The range of f_k is $\{0,1,\ldots,k-1\}$. Taking $m = \lfloor \log_2 k \rfloor$, the range of f_k can be viewed as $\{0,1\}^m \cup \{\perp\}$ (in case of $\perp$, the function is undefined). This causes at most a factor 2 loss in entropy.

To evaluate the robustness of f_k, we'd like to have upper and lower bounds on $Pr\left(f(X,Y) = \alpha\right)$ for all pairs of independent (l,b_1)-distributed, (l,b_2)-distributed sources X,Y. By lemma 5, it suffices to consider flat distributions. Therefore, we're interested in bounds on the number of solutions $x - y = z$ for $x \in A$, $y \in B$, $z \in R_\alpha$, where $A,B \subset Z_p$ are arbitrary subsets of size $2^{b_1}, 2^{b_2}$, respectively. Let $\nu(A,B,\alpha)$ denote this number. Clearly, $\nu(A,B,\alpha) = \nu(g^{-\alpha}A, g^{-\alpha}B, 0)$, and thus it suffices to consider R_0, the set of k-th residues modulo p.

Let $\omega \in C$ be a primitive p-th root of unity. Let $\varphi_k(j) = \sum_{x \in R_0} \omega^{jx}$, and $\Phi_k = \max_{1 \le j \le p-1} |\varphi_k(j)|$. A result of [2] relates $\nu(A,B,\alpha)$ to the size of A,B via Φ_k.

Lemma 12 [2]: Let $A, B \subset Z_p$ be two arbitrary subsets, and α an integer, $0 \leq \alpha \leq (p-1)/k$. Then

$$\left| \nu(A, B, \alpha) - \frac{|A| \cdot |B|}{k} \right| \leq \Phi_k \sqrt{|A| \cdot |B|} \; .$$

The following bound on Φ_k was given by László Babai (private communication).

Lemma 13: $\Phi_k < \sqrt{p}$.

Proof: The proof uses methods of trigonometric sums over finite fields (see [27]). We start by presenting some definitions and notations. An additive character of Z_p is a mapping $\psi : Z_p \mapsto C$ (C denotes the complex numbers) satisfying $\psi(a+b) = \psi(a) \cdot \psi(b)$; the unit character satisfies $\psi_0(\cdot) = 0$. A multiplicative character of Z_p is a mapping $\chi : Z_p^* \mapsto C^*$ satisfying $\chi(a \cdot b) = \chi(a) \cdot \chi(b)$; the unit character satisfies $\chi_0(\cdot) = 1$ ($\chi(0) \stackrel{\text{def}}{=} 0$ unless $\chi = \chi_0$). A Gaussian sum $S(\chi, \psi)$ is defined as $\sum_{x \in Z_p} \chi(x)\psi(x)$. It is well known [27, p. 47, thm. 3A] that for $\chi \neq \chi_0$, $\psi \neq \psi_0$, $|S(\chi, \psi)| = \sqrt{p}$, while $S(\chi_0, \psi) = 0$.

Let $\xi \in C$ be a primitive k-th root of unity. For $0 \leq t \leq k-1$, define $\chi_t : Z_p^* \mapsto C^*$ by $\chi_t(x) = \xi^{t \log_g x}$, then χ_t is a multiplicative character, and

$$x \in R_0 \Rightarrow \log_g x = ki \Rightarrow \chi_t(x) = 1 \tag{1}$$

$$0 \neq x \notin R_0 \Rightarrow \log_g x = ki + \alpha \text{ for some } 1 \leq \alpha \leq k-1$$

$$\Rightarrow \sum_{t=0}^{k-1} \chi_t(x) = \sum_{t=0}^{k-1} (\xi^\alpha)^t = 0 \tag{2}$$

$$x = 0 \Rightarrow \sum_{t=0}^{k-1} \chi_t(x) = 1 \tag{3}$$

For $1 \leq j \leq p-1$, define $\psi_j : Z_p \mapsto C$ by $\psi_j(x) = \omega^{jx}$, then ψ_j is an additive character $\neq \psi_0$. Using (1), (2), and (3) we get

$$\begin{aligned}
\sum_{t=0}^{k-1} S(\chi_t, \psi_j) &= \sum_{t=0}^{k-1} \sum_{x \in Z_p} \chi_t(x)\psi_j(x) \\
&= \sum_{x \in R_0} \omega^{jx} \sum_{t=0}^{k-1} \chi_t(x) + \sum_{0 \neq x \notin R_0} \omega^{jx} \sum_{t=0}^{k-1} \chi_t(x) + \sum_{t=0}^{k-1} \chi_t(0) \\
&= k \sum_{x \in R_0} \omega^{jx} + 1 \; .
\end{aligned}$$

Using the equalities for Gaussian sums, the last equality implies that for every $1 \leq j \leq p-1$

$$\begin{aligned}\left|\sum_{x \in R_0} \omega^{jz}\right| &= \frac{1}{k}\left|-1 + \sum_{t=0}^{k-1} S(\chi_t, \psi_j)\right| \\ &\leq \frac{1}{k}\left(1 + \sum_{t=0}^{k-1} |S(\chi_t, \psi_j)|\right) \\ &= \frac{1}{k}(1 + 0 + (k-1)\sqrt{p}) \\ &< \sqrt{p} \qquad \square\end{aligned}$$

Reformulating these bounds in terms of sources and robustness, we get

Theorem 14: Let $p(n), k(n), l(n) = \lfloor \log_2 p(n) \rfloor, f_{k(n)}$ be as above, and $\varepsilon(n) > 0$. Suppose $b_1(n) + b_2(n) \geq l(n) + 1 + 2\log_2 \varepsilon^{-1}(n) + 2\log_2 k(n)$. Then $f_{k(n)} : Z_{p(n)} \times Z_{p(n)} \mapsto \{0, 1, \ldots, k(n)-1\}$ is $\varepsilon(n)$-robust for any pair of independent, $(l(n), b_1(n))$-distributed, $(l(n), b_2(n))$-distributed sources.

Proof: It suffices to consider flat sources X, Y. Let $A \subset Z_{p(n)}$ be the set where $Pr(X = a) = 1/2^{b_1(n)}$ (similarly for $B, Y, b_2(n)$). Then $|A| = 2^{b_1(n)}$, $|B| = 2^{b_2(n)}$, and for every $\alpha \in \{0, 1, \ldots, k(n)-1\}$ we have

$$Pr(f(X,Y) = \alpha) = \frac{\nu(A, B, \alpha)}{2^{b_1(n)} \cdot 2^{b_2(n)}} .$$

Combining Lemmas 12 and 13, we have

$$\frac{1}{k(n)}\left(1 - k(n)\sqrt{\frac{p(n)}{2^{b_1(n)+b_2(n)}}}\right) < Pr(f(X,Y) = \alpha) < \frac{1}{k(n)}\left(1 + k(n)\sqrt{\frac{p(n)}{2^{b_1(n)+b_2(n)}}}\right) .$$

Substituting $p(n) < 2^{l(n)}$, $b_1(n) + b_2(n) \geq l(n) + 1 + 2\log_2 \varepsilon^{-1}(n) + 2\log_2 k(n)$, the claim follows. $\square$

We now establish some relations between the various quantities above. We denote by n the security parameter, and parametrize by it the the block length, the probability bounds, the prime p, the divisor of $p-1$, k, and the bias guarantee ε (that is, they will be denoted by $l(n)$, $b_1(n)$, $b_2(n)$, $p(n)$, $k(n)$, and $\varepsilon(n)$ respectively). Typically, $\varepsilon(n) = \frac{1}{n^{h(n)}}$ where $h(n) \geq 0$ is either a constant (the case of a polynomial bias) or a function tending to ∞ with $n \to \infty$ (the case of a subpolynomial bias).

The information rate of the scheme is $\approx \frac{\log k(n)}{2 \log p(n)}$. Therefore, in order to guarantee a constant rate, $k(n)$ must be $> p(n)^d$ for some constant d, $0 < d < 1$. We will typically use $3/16 \leq d \leq 1/4$. Assuming $b_1(n) + b_2(n) \geq 1.75\, l(n)$ (an assumption we'll later justify), and substituting $\log_2 \varepsilon^{-1}(n) = g(n) \log_2 n$, $\log_2 k(n) = d \cdot l(n)$ in the equality of the last theorem, we see that $l(n) \geq \frac{h(n)}{3/8 - d} \log_2 n$ is a necessary condition for the scheme to produce $\varepsilon(n) = n^{-h(n)}$-robust bits. The case of equality in the last equation implies $\varepsilon^{-1}(n) \leq k(n) \leq \varepsilon^{-2}(n)$ (for $3/16 \leq d \leq 1/4$), an expression which relates the size of $k(n)$ to the robustness of bits produced by $f_{k(n)}$.

The computation complexity of the scheme equals the (deterministic) complexity of finding discrete logarithms modulo $k(n)$ in the field $Z_{p(n)}$. We first assume that $p(n)$, $k(n)$ and g, a primitive element of $Z_{p(n)}$, are given, and analyze the run time of the scheme. We then turn to the complexity of the preprocessing stage, in which $p(n)$, $k(n)$ and g are produced.

Given $p(n)$, $k(n)$ and g, a primitive element of $Z_{p(n)}$, then by essentially trying all possible candidates, we can compute $\log_g(\cdot) \bmod k(n)$ in time $O\left(k(n) \log^3(p(n))\right)$ (see [18]). Thus if the bias $\varepsilon(n)$ is required to be just polynomial in n ($\varepsilon(n) = n^{-O(1)}$), then by employing our scheme with brute-force discrete logarithm subroutine, the computational complexity is polynomial in n.

In order to generate subpolynomially biased bits ($\varepsilon(n) = n^{-h(n)}$, with $h(n) \to \infty$), we need more efficient ways of computing discrete logarithms (modulo $k(n)$) in $Z_{p(n)}$. There are known algorithms with complexity $2^{O\left(\sqrt{\log p(n) \log\log p(n)}\right)}$, and this run-time would suit our needs, but unfortunately these algorithms are randomized, so we cannot use them to (deterministically) evaluate $f_{k(n)}$. Instead, we look for $p(n)$'s with *smooth* $k(n)$, and use an algorithm due to Pohlig and Hellman [23] which is sufficiently fast in such circumstances.

A natural number z is called y-smooth if all its prime factors are $\leq y$. Suppose $p(n)$ is a prime in the range $n^{\sqrt{\log\log n}} \leq p(n) \leq n^{2\sqrt{\log\log n}}$, so that $k(n)$ ($k(n) \,|\, p(n) - 1$) is n-smooth. Given a primitive element g of $Z_p(n)$, the Pohlig and Hellman algorithm [23] finds $\log_g(\cdot) \bmod k(n)$ in $O(n \log^3 n)$ deterministic time.

Given $l(n)$, every pair of (l, b_1)-distributed, (l, b_2)-distributed sources can be viewed as $(l(n), l(n) \cdot b_1/l)$, $(l(n), l(n) \cdot b_2/l)$ sources respectively. Using "nice" primes as above, we have

Theorem 15: Suppose $p(n)$ is a prime in the range $n^{\sqrt{\log\log n}} \leq p(n) \leq n^{2\sqrt{\log\log n}}$, so that $k(n) \mid p(n)-1$, $p(n)^{3/16} \leq k(n) \leq p(n)^{1/4}$, and $k(n)$ is n-smooth. Let $g \in Z_{p(n)}$ be a given primitive element. Furthermore, let l, b_1, b_2 satisfy $b_1 + b_2 \geq 1.75\,l$. Then for any pair of independent (l, b_1)-distributed and (l, b_2)-distributed sources, the function $f_{k(n)} : Z_{p(n)} \times Z_{p(n)} \mapsto \{0, 1, \ldots, k(n)-1\}$ produces $1/n^{\sqrt{\log\log n}}$-robust bits, with information rate $3/32 \leq r \leq 1/8$ and computational complexity $O(n \log^3 n)$.

We now turn to the preprocessing stage, starting with the question of finding appropriate primes. Let $\Psi(x, y)$ denote the number of natural numbers not exceeding x which are y-smooth. Canfield, Erdös and Pomerance proved the following theorem concerning $\Psi(x, y)$:

Theorem 16 [7]: For $y \geq log^2 y$, $\Psi(x, y) = xu^{-u+o(u)}$, where $u = \log x / \log y$.

In particular, for large enough x, $\Psi(x, y) \geq xu^{-2u}$. Choosing $h(n) = 2\sqrt{\log\log n}$, $y = n$, and $x = n^{h(n)}$, we have $u = \log x / \log y = h(n)$. Substituting these quantities in the last theorem, we conclude that for large enough n, the probability that a randomly chosen $z \leq n^{h(n)}$ will be n-smooth is bounded below by $h(n)^{-2h(n)} = 1/\left(2\sqrt{\log\log n}\right)^{4\sqrt{\log\log n}} > \log^{-1/3} n$. Obviously, the same lower bound holds for the probability that a random z has an n-smooth divisor d such that $z^{3/16} < d < z^{1/4}$. However, for our purposes it is not enough for z to have such divisor, but $z+1$ must be a prime as well. Carl Pomerance (private communication) has provided us with an estimate of the probability of this event.

Theorem 17: For a randomly chosen $n^{\sqrt{\log\log n}} \leq z \leq n^{2\sqrt{\log\log n}}$,

$$Pr\left(z \text{ has an } n\text{-smooth divisor in the range } [z^{3/16}, z^{1/4}] \text{ and } z+1 \text{ is prime}\right) \geq \frac{1}{\log^2 n}\,.$$

By choosing random integers in the above range, an appropriate prime $p(n)$ with a large n-smooth divisor $k(n)$ and a generator for Z_p can be found in expected polynomial time, given access to an unbiased independent coin. This is done as follows. First, we choose $p(n)$ at random, factor $p(n)-1$ and look for a sufficiently large n-smooth divisor $k(n)$. For factoring $p(n)$, we use Dixon's algorithm [9], that runs in expected time $2^{O\left(\sqrt{\log p(n) \log\log n}\right)}$ (which is polynomial in n). Next, we verify that $p(n)$ is a prime, using Pratt's algorithm [24] (again using Dixon's algorithm as a factoring subroutine, and trying to find primitive elements by choosing elements at random). In case $p(n)$ is indeed a prime, Pratt's algorithm yields

a a primitive element of $Z_{p(n)}$. We now substitute the unbiased independent coin used in the preprocessing, by bits extracted from two probability-bounded sources, using any of the computationally efficient (but not necessarily rate efficient) schemes of subsection 3.1.

Finally, in order to satisfy $b_1 + b_2 \geq 1.75\,l$, we start with *four* independent (l_0, b_0)-distributed sources (where the ratio between the constants b_0 and l_0 can be arbitrarily small). Using the techniques of subsection 3.1, every pair of these sources is converted into a single $(l, 0.9\,l)$-distributed source (where $l = 2l_0 \frac{l_0}{b_0} \log_2 \frac{l_0}{b_0}$). This conversion is rate efficient, and its complexity does not depend on $\varepsilon(n)$.

3.3 Slightly Dependent Sources

In the previous section, we showed how to extract unbiased bits from the output of two independent probability-bounded sources. A natural question is whether the independence requirement can be relaxed, and if so – to what extent. We suggest the following definition and investigate its ramifications.

Definition 9: Let $\delta \geq 0$. We say that two variables X and Y *are δ-dependent* if, for every $\alpha, \beta \in \{0,1\}^l$ with $Pr(X = \alpha) \cdot Pr(Y = \beta) \neq 0$ the following holds

$$(1+\delta)^{-1} \leq \frac{Pr(X = \alpha \text{ and } Y = \beta)}{Pr(X = \alpha) \cdot Pr(Y = \beta)} \leq (1+\delta) \ .$$

Thus, 0-dependence identifies with independence. Also, notice that this is a more refined measure of dependency than correlation. A different definition of slightly dependent SV-sources was presented in [28], and does not seem to extend to PRB-sources. The following Lemma can be easily verified.

Lemma 18: Suppose that f is ε-robust for any two independent variables satisfying properties P_1 and P_2 respectively. Then f is $(\delta + (1+\delta)\varepsilon)$-robust for any two δ-dependent variables satisfying properties P_1 and P_2 respectively.

Applications to extracting unbiased bits from slightly dependent functions follow immediately, by combining Lemma 18 with Theorems 7 or 9. Lemma 18 may seem weak at first glance. It only guarantees that, for small δ, the added bias introduced by the δ-dependency does not exceed δ. However, this result is almost optimal ! We will show that δ-dependency may causes an added $\Omega(\delta)$ bias.

Theorem 19: Let $0 < \delta \leq 30$ and $f : \{0,1\}^{2l} \mapsto \{0,1\}$ be an arbitrary Boolean function. Then at least one of the following two statements hold:

1) There exist a $\sigma \in \{0,1\}$ and a pair of δ-dependent $(l, l-2)$-distributed variables X and Y such that $Pr(f(X,Y) = \sigma) \geq \frac{1}{2} \cdot (1 + \frac{\delta}{64})$.

2) There exist a $\sigma \in \{0,1\}$ and a pair of independent $(l, l - 7 - \log_2 \delta^{-1})$-distributed variables X and Y such that $Pr(f(X,Y) = \sigma) \leq \frac{1}{3}$.

Proof: Without loss of generality, we assume that $|\{(i,j) \in \{0,1\}^{2l} : f(i,j) = 1\}| \geq \frac{1}{2} \cdot 2^{2l}$. The function f is represented as a bipartite graph $G(V,E)$, where $V = A \cup B$ ($A = \{a_i : i \in \{0,1\}^l\}$ and $B = \{b_j : j \in \{0,1\}^l\}$) is the bipartition and the edge set $E \subset A \times B$ satisfies

$$(a_i, b_j) \in E \text{ iff } f(i,j) = 1 \ .$$

The idea of the proof is to show the existence of a large regular subgraph with relatively high degree. If we succeed, the variables are defined to be flat on the vertex sets of the subgraph, and the dependency allowance is used to make the edges of the subgraph "heavy". Thus, the probability mass is concentrated more on entries on which the function has value 1, and the function is bias towards 1 as required in statement (1) of the Theorem. We actually give an algorithm for finding a large regular subgraph. The algorithm is carried out in stages, where at each stage a new perfect matching is found. It will be shown that if the algorithm fails then statement (2) of the Theorem holds.

In the following we will assume that statement (2) of the Theorem is false and will show that statement (1) follows.

Finding a large regular subgraph

Let $n = 2^l$, $m \geq n/4$, $k = m/32$. We present an algorithmic proof that the graph $G(V,E)$ contains a k-regular subgraph with m vertices on each side. The argument proceeds in two phases. First, we find a subgraph G' of G, which has $m \geq n/4$ vertices on each side, average degree $\geq (\frac{1}{2} - \frac{\delta}{128}) \cdot m$, and minimum degree $\geq m/8$. Next we find a spanning k-regular subgraph of the latter.

The subgraph G' is found by applying the following procedure.

1. **procedure 1:** FIND LARGE SUBGRAPH G'
; A vertex $a \in A$ in a bipartite graph $G((A,B),E)$ is called *bad* if its degree is $< \frac{1}{4} \cdot |B|$.
2. INPUT $\leftarrow G(V,E)$

; **Step 1 – Omitting bad vertices from both sides**
3. $G' \leftarrow G$
4. *While* both sides of $G'(V', E')$ contain bad vertices *do begin*
; Let L (resp. R) be the set of bad vertices in the left (resp. right) side of G'.
5. $\beta \leftarrow \min\{|L|, |R|\}$;
6. Omit β vetices of L and β vertices of R from G'.
; The resulting graph is refered to as G'.
7. *end*;
; **Step 2 – Omitting bad vertices from the remaining side**
; Let L (resp. R) be the set of bad vertices in the left (resp. right) side of G'.
[By the above, either L or R is empty. Wlog let $R = \emptyset$.]
8. Omit all remaining bad vertices (i.e. L) from the graph.
9. Omit, from the right side of the remaining graph, $|L|$ vertices of minimum degree.
10. **return** G'.

Throughout the execution of Step (1) of the above procedure, we only omit vertices with degree not exceeding half the current average degree. Thus, at the end of Step (2) average degree in the remaining graph is no less than half the number of vertices in one side of the graph.

Let $2t$ denote the number of vetices omitted in Step (1). Then the number of edges deleted in Step (1) is at most

$$t \cdot \frac{n}{4} + t \cdot (n - t) \ .$$

(We charge egdes with a bad leftpoint to vertices omitted from the left side of the graph, while charging all other omitted edges to the vertices omitted from the right.) Using the above upper bound on $|E - E'|$, we get

$$\frac{n^2}{2} \leq |E| = |E - E'| + |E'| \leq t \cdot \frac{n}{4} + t \cdot (n - t) + (n - t)^2$$

which yields $t \leq \frac{2}{3} \cdot n$.

Now we claim that in Step (2), L could not be too big. If $|L| \geq \frac{\delta}{32} \cdot r$, where r is the number of vertices in the right side of G' after Step (1), we reach contradiction by constructing two independent $(l, l - 7 - \log_2 \delta^{-1})$-sources. (The sources we construct are flat on L and the right side of G' respectively, and applying the function to their output yields a bias of $\frac{1}{4}$.) It is easy to see that after Step (2) the number of vertices on each side of the graph, denoted m, is $\geq (1 - \delta/32)n/3 > n/4$. Also, the minimum degree in the remaining graph is $\geq \frac{r}{4} - \delta \cdot r/32 \geq m/8$, and the average degree is $\geq (\frac{1}{2} - \frac{\delta}{32 \cdot 4}) \cdot m$.

The second phase of our construction consist of finding a spanning k-regular graph of $G_0 = G'$. This is done by applying the following procedure.

1. **procedure 2:** FIND SPANNING k-REGULAR SUBGRAPH
2. INPUT $\leftarrow G_0((A_0, B_0), E_0)$
3. $R \leftarrow \emptyset$
4. *For* $i = 1$ *to* k *do begin*
5. Find a perfect matching M_i in G'.
6. $R \leftarrow R \cup M_i$.
7. Omit M_i from G_{i-1}, resulting in G_i.
8. *end*;
9. **return** R.

We now show that Procedure 2 does not fail. Assume on the contrary that in some iteration $i+1 \leq k$ a perfect matching is not found. Namely, G_i does not contain a perfect matching. By Hall's Theorem [6, sec. 5.2, p. 72], the left side of G_i (i.e. A_0) contains a set of vertices A' such that the neighbourhood of A' (denoted B') has cardinality smaller than $|A'|$. Since the residual degree of G_i is $\geq \frac{m}{8} - i$, we get

$$|A'| > |B'| \geq \frac{m}{8} - k \ .$$

Consider the neighbourhood of a node in $B_0 - B'$ (such a node does exist since $|B'| < |A'| \leq |B_0|$). This neighbourhood has cardinality $\geq \frac{m}{8} - k$ and does not intersect with A'. We conclode that

$$|B_0 - B'| > |A_0 - A'| \geq \frac{m}{8} - k \ .$$

It should be noted that there are no edges in G_i between A' and $B_0 - B'$. Thus, G_0 (or G for this matter) contain at most $i \cdot \min\{|A'|, |B_0 - B'|\}$ edges between A' and $B_0 - B'$. This is at most one third of $|A'| \cdot |B_0 - B'|$ (since $i < k = m/32$ and $\min\{|A'|, |B_0 - B'|\} \geq \frac{m}{8} - k = 3m/32$). Letting one source be flat on A' and the other be flat on $B_0 - B'$ we get bias $\frac{1}{3}$ and reach contradiction (as these sources are $(l, l-6)$-distributed).

Using the dependency allowance to bias the function

We now use the regular subgraph $((A_0, B_0), R)$ to present a pair of δ-dependent probability bounded sources, X and Y, which make the function bias. X will be flat on A_0 and Y will be flat on B_0; that is

$$Pr(X = i) \stackrel{\text{def}}{=} \begin{cases} |A_0|^{-1} & \text{if } a_i \in A_0 \\ 0 & \text{otherwise} \end{cases}$$

$$Pr(Y = j) \stackrel{\text{def}}{=} \begin{cases} |B_0|^{-1} & \text{if } b_j \in B_0 \\ 0 & \text{otherwise} \end{cases}$$

The dependency allowance is used as follows:

$$Pr(X=i,Y=j) \stackrel{\text{def}}{=} \begin{cases} (1+\delta)\cdot Pr(X=i)\cdot Pr(Y=j) & \text{if } (a_i,b_j)\in R \\ (1-\delta/31)\cdot Pr(X=i)\cdot Pr(Y=j) & \text{otherwise} \end{cases}$$

The reader may verify the validity of the above definition, by noting that $\frac{1}{32}\cdot(1+\delta)+\frac{31}{32}\cdot(1-\delta/31)=1$. It follows that

$$\begin{aligned} Pr(f(X,Y)=1) &\geq \frac{1}{32}\cdot(1+\delta)+(\frac{15}{32}-\frac{\delta}{128})\cdot(1-\delta/31) \\ &> \frac{1}{2}+\frac{\delta}{128} \end{aligned}$$

The Theorem follows. □

3.4 Variations: Entropy, Varying Length

We conclude our investigation into the problem of extracting unbiased bits from weak sources of randomness, by two remarks. The first remark concerns the probability bound b, while the second remark concerns both b and l.

We have defined probability bounded variables as having an upper bound (2^{-b}) on the probability for each individual l-bit string. A more "natural" but less convinient definition considers variables with a lower bound on the (information theoretic) entropy. Every (l,b)-variable has entropy $\geq b$, but the converse does not hold. Nevertheless, every source which has non-zero entropy is in fact a probability-bounded source (a quantative statement is omitted).

So far, we have considered sources where for some fix integer l, given the history, the next l bits have a particular distribution. A natural question, raised by Jeff Lagarias, is what happends when the number of next bits is a variable. More precisely, let $l(\cdot)$ and $b(\cdot)$ be functions, and consider a source S with the following output distribution: for every integer $n>0$, for every $\alpha\in\{0,1\}^n$ and every $\beta\in\{0,1\}^{l(n)}$, the conditional probability that the next $l(n)$ bits output by S equal β given that the first n bit output by S are α does not exceed $2^{-b(n)}$. We call this a varying probability-bounded source (VPRB-source). Extending Theorems 7 and 4, we get an almost sharp threshold for the value of $b(\cdot)$ which allows the extraction of almost unbiased bits from two VPRB-sources. Such extraction is possible whenever $b(n) > 4+\log_2 l(n)$, and is impossible whenever $b(n) < \log_2(l(n)-\log_2 l(n))-1$.

4. COMMUNICATION COMPLEXITY

In this section, we present results concerning probabilistic communication complexity. In subsection 4.1 we recall the common definitions of communication complexity, present new definition and compare them. In subsection 4.2 we prove lower bounds on the communication complexity of the functions considered in Section 2. In subsection 4.3 we demonstrate the tightness of our results by presenting upper bounds on the communication complexity of all functions. In subsection 4.4 we suggest a robust notion of communication complexity and extend our lower bounds to it.

Consider two interactive parties A and B, such that A knows an input $x \in \{0,1\}^n$ and B knows an input $y \in \{0,1\}^n$. The inputs are randomly and independently chosen, each with uniform probability distribution. Let $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$ be a function and assume that A and B wish to compute $f(x,y)$. To this end they use a possibly *randomized* protocol P. As commonly assumed, the messages sent at each round are prefix free. The protocol is terminated by party A and the last bit B sent to A, is their *joint guess of the value of* $f(x,y)$. A natural question is how many bits should be exchanged among the party so that their joint guess is significatly better than the a-priori guess. The answer depends on the exact definitions of the notions "number of bits" and "success probability".

4.1 Definitions

Let $x,y \in \{0,1\}^n$. We consider the probability space defined by the coin tosses of the parties A and B. Let $l_P(x,y)$ be the random variable denoting the number of bits A and B exchange on the pair (x,y), using the protocol P. Let $L_P(x,y)$ denote the expected value of $l_P(x,y)$ and $l_P^*(x,y)$ denote the supremum of $l_P(x,y)$. Let $s_{P,f}(x,y)$ be the random variable denoting the success of P with respect to f on the pair (x,y); and let $S_{P,f}(x,y)$ denote the expected value of $s_{P,f}(x,y)$. That is, $S_{P,f}(x,y)$ is the probability that the last bit exchanged by A and B on inputs x and y equals $f(x,y)$.

The *average operator* (denoted $\mathcal{A}ve$), and the *minimum* and *maximum operators* (denoted $\mathcal{M}in$ and $\mathcal{M}ax$ resp.) are defined in the obivious manner. These operators are used in defining the various measures. For example, $\mathcal{A}ve(L_P) = 2^{-2n}\sum_{x,y\in\{0,1\}^n} L_P(x,y)$ is the average number of bits exchanged in the protocol P; $\mathcal{M}ax(L_P)$ is the expected number of bits on the worst pair of inputs; and $\mathcal{M}ax(l_P^*)\max_{x,y\in\{0,1\}^n}\{l_P^*(x,y)\}$ is the maximum

number of bits taken over all possible executions and pairs. Two measures for success are $Min(S_{P,f})$ (worst pair) and $Ave(S_{P,f})$ (averaged over all pairs).

Previous Definitions

Various definitions of randomized communication complexity have appeared. We present some of them† (other definitions can be found in [20], [14], and [28]):

- *Yao's definition of randomized communication complexity* [31] (hereby denoted $C_\varepsilon(f)$) is thus the infimum of $Max(L_P)$, when taken over all *randomized* protocols P satisfying $Min(S_{P,f}) \geq \frac{1}{2} + \varepsilon$.
- *Yao's definition of distributed communication complexity* [32] (hereby denoted $D_\varepsilon(f)$) is the infimum of $Ave(L_P)$, when taken over all *deterministic* protocols P satisfying $Ave(S_{P,f}) \geq \frac{1}{2} + \varepsilon$.
- Orlitsky and El-Gamal [19] measure the *average communication complexity* (hereby denoted $\overline{C}_\varepsilon(f)$) as the infimum of $Ave(L_P)$, when taken over all *randomized* protocols P satisfying $Min(S_{P,f}) \geq \frac{1}{2} + \varepsilon$.
- Paturi and Simon [22] define the *unbounded communication complexity* (hereby denoted $U(f)$) to be the infimum of $Max(l_P^*)$, when taken over all randomized protocols P satisfying $Min(S_{P,f}) > \frac{1}{2}$.

Our Definitions

We say that the protocol P has *average ε-advantage in guessing f* if $Ave(S_{P,f}) \geq \frac{1}{2} + \varepsilon$. In the following definitions we consider protocols with *average* ε-advantage.

- We define the *average randomized communication complexity of function f* (denoted $A_\varepsilon^R(f)$) as the infimum of $Ave(L_P)$, when taken over all *randomized* protocols P which have average ε-advantage in guessing f.
- The *worst-case randomized communication complexity of function f* (denoted $W_\varepsilon^R(f)$) is defined as the infimum of $Max(l_P^*)$, when taken over all *randomized* protocols P which have average ε-advantage in guessing f.

† The role of ε in our notations differs from its role in [31,32,19]. Here ε denotes the advantage of the protocol over $1/2$, while originally it was used to denote the error probability.

- The *deterministic communication complexities* ($A_\varepsilon^D(f)$ and $W_\varepsilon^D(f)$) of the function f are defined similarly, for deterministic protocols.

Comparison of Definitions

For all functions $f \in F_n$, the following inequalities are immediate from the definitions:

$$A_\varepsilon^R(f) \leq \overline{C}_\varepsilon(f) \leq C_\varepsilon(f)$$
$$A_\varepsilon^R(f) \leq A_\varepsilon^D(f) = D_\varepsilon(f) \leq W_\varepsilon^D(f)$$

Yao showed that $C_{\frac{1}{2}-\eta}(f) \geq \frac{1}{2} D_{\frac{1}{2}-2\eta}(f)$ [30,32].

There are however functions for which $A_\varepsilon^R(f) \ll C_\varepsilon(f), D_\varepsilon(f)$. One such function is the ordering function g defined by $g(x,y) = 1$ iff $x \leq y$. Yao showed that for any fixed $\varepsilon > 0$, $C_\varepsilon(g) = \Omega(\log n)$ [31]. The protocol in which A sends the most significant bit of x has a $1/4$-advantage in guessing g, and thus $A_{1/4}^R(g) = 1$ (in fact $W_{1/4}^D(g) = 1$). (Paturi and Simon [22] showed that $U(g) = 2$.) The three measures $W_\varepsilon^R(f)$, $U(f)$ and $C_\varepsilon(f)$ are not always comparable.

4.2 Lower Bounds

We begin this subsection by stating our lower bounds, and comparing them to recent results of other researchers.

Theorem 20: Let $0 < \varepsilon \leq 1/2$.

1) For at least a $1 - 2^{-2^n}$ fraction of the Boolean functions $f \in F_n$,

$$W_\varepsilon^R(f) > n - 7 - 3\log_2 \varepsilon^{-1}$$
$$A_\varepsilon^R(f) > 2\varepsilon \cdot \left(n - 7 - 3\log_2 \varepsilon^{-1} n\right) - 1 \ .$$

2) For every $f \in F_n$ representable by a Hadamard matrix, the following holds.

$$W_\varepsilon^R(f) > n - 3 - 3\log_2 \varepsilon^{-1}$$
$$A_\varepsilon^R(f) > 2\varepsilon \cdot \left(n - 3 - 3\log_2 \varepsilon^{-1} n\right) - 1 \ .$$

In particular, this holds for the inner-product function.

Comparison to Other Works

Our result implies that for almost all $f \in F_n$, $\overline{C}_\varepsilon(f) \geq 2\varepsilon\left(n - 7 - 3\log_2 \varepsilon^{-1} n\right) - 1$. This is related to a recent independent result of Orlitsky and El Gamal [19], who showed that almost all $f \in F_n$ have $\overline{C}_\varepsilon(f) \geq 2\varepsilon(n - 1 - \log_2 n)$. (Actually, they showed that for all $2^{-n} < r \leq 1/2$, almost all functions f with $r \cdot 2^{2n}$ 1's in their table, have $\overline{C}_\varepsilon(f) \geq 2\varepsilon(n - \log_2 r^{-1} n)$.)

Our bound on $W_\varepsilon^R(f)$ (for almost all $f \in F_n$) is related to a recent independent result of Alon, Frankl and Rödl [4] who showed that almost all $f \in F_n$ have $U(f) \geq n - 5$. Their result implies $C_\varepsilon(f) > 2\varepsilon(n - 5)$.

All three works resolve Yao's open problem [31]: *What is $C_\varepsilon(f)$ for a random $f \in F_n$?*

Proof of Theorem 20

Proposition 21: Let $0 < \delta \leq 1$. Then $A_\varepsilon^R(f) \geq (2 - \delta)\varepsilon \cdot W_{\delta\varepsilon/2}^R(f)$. (In particular, $A_\varepsilon^R(f) \geq \varepsilon \cdot W_{\varepsilon/2}^R(f)$.)

Proof: Consider runs of protocol P which has an ε-advantage in guessing f and average length a. Truncate runs of P which exceeding $\varepsilon^{-1}a/(2-\delta)$ bits. In the event of a long run, flip a coin to determine the final guess. Such runs occur with probability $\leq (2 - \delta)\varepsilon$, and by guessing at random we lose at most $(2 - \delta)\varepsilon/2$ of the average success probability. This yields a protocol with $\frac{\delta}{2} \cdot \varepsilon$-advantage, the runs of which are no more than $\varepsilon^{-1}a/(2 - \delta)$ bits long. □

Proposition 22: $W_\varepsilon^R(f) = W_\varepsilon^D(f)$.

Proof: For a randomized protocol, the average advantage is a sum over both the inputs and the coin tosses. There must be at least one sequence of coin tosses which does at least as well as the average. Using this string, we get a deterministic protocol with the same length and at least the same average advantage. □

Notice that the above argument holds since we are interested in average advantage; the advantage on individual pairs may decrease. Using Proposition 22, we may concentrate in studying $W_\varepsilon^D(f)$.

Theorem 23: Let k, n be integers, and $0 < \varepsilon \leq 1$. Suppose that for every $b_1 + b_2 \geq 2n - k - 1 + \log_2 \varepsilon$, the Boolean function $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$ is ε-robust with respect

to any pair of independent random variables X, Y satisfying: X is (n, b_1)-distributed and Y is (n, b_2)-distributed. Then

$$W_\epsilon^D(f) > k \ .$$

Proof: Suppose, towards a contradiction, that P is a deterministic protocol with average ϵ-advantage in guessing f, such that $Max(l_P^*) \leq k$. Consider all possible executions of P and assume, without loss of generality, that A and B exchange exactly k bits on each pair of inputs.

For every $\gamma \in \{0,1\}^k$, denote by $C(\gamma)$ the set of (x,y) pairs on which A and B's communication consists of γ. Note that by prefix freeness, the parsing of γ is unique. Let $A(\gamma) = \{x : \exists y \text{ s.t. } (x,y) \in C(\gamma)\}$, and $B(\gamma) = \{y : \exists x \text{ s.t. } (x,y) \in C(\gamma)\}$. By a cut-and-paste argument of Yao [30], $C(\gamma) = A(\gamma) \times B(\gamma)$.

Denote by $last(\gamma)$ the last bit exchanged in the communication γ, and let $G(\gamma) = \{(x,y) \in C(\gamma) : last(\gamma) = f(x,y)\}$. Since P has ϵ-advantage on f, we have $\sum_{\gamma \in \{0,1\}^k} |G(\gamma)| \geq (\frac{1}{2} + \epsilon) \cdot 2^{2n}$. Let us say that $C(\gamma)$ is *small* if $|C(\gamma)| < \epsilon \cdot 2^{2n-k-1}$. Since there are at most 2^k rectangles $C(\gamma)$, the number of points in all small rectangles is at most $\epsilon \cdot 2^{2n-1}$. Thus

$$\sum_{\gamma \text{ s.t. } |C(\gamma)| \geq \epsilon \cdot 2^{2n-k-1}} |G(\gamma)| \geq (\frac{1}{2} + \frac{\epsilon}{2}) \cdot 2^{2n} \ .$$

This implies that there exist a $\gamma \in \{0,1\}^k$ such that both $|C(\gamma)| \geq \epsilon \cdot 2^{2n-k-1}$ and $|G(\gamma)| \geq (\frac{1}{2} + \frac{\epsilon}{2}) \cdot |C(\gamma)|$ (i.e. $C(\gamma)$ is sufficiently large, and the protocol has non-negligible advantage on the pairs in it).

Set X and Y to be two independent random variables, flat on $A(\gamma)$ and $B(\gamma)$, respectively. Then $Pr\,(f(X,Y) = last(\gamma)) \geq \frac{1}{2} \cdot (1 + \epsilon)$. Let $b_1 = |A(\gamma)|$ and $b_2 = |B(\gamma)|$. Then X is (n, b_1)-distributed, Y is (n, b_2)-distributed, and $b_1 + b_2 \geq \log_2 (\epsilon \cdot 2^{2n-k-1}) = 2n - k - 1 + \log_2 \epsilon$. This contradicts the ϵ-robustness of f. □

Theorem 20 (above) is a consequence of combining Theorem 23 (and Propositions 21 and 22) with Theorems 7 and 9 (of sections 4 and 5 respectively). The arithmetic details are as follows.

1) Let $0 < \epsilon \leq 1/2$. By Theorem 7 (see special case 4), all but at most 2^{-2^n} of the functions $f : \{0,1\}^{2n} \mapsto \{0,1\}$ are ϵ-robust for any pair of (n, b_1), (n, b_2) sources satisfying $b_1 + b_2 \geq n + 6 + 2\log_2 \epsilon^{-1}$. Setting $k = n - 7 - 3\log_2 \epsilon^{-1}$, these functions

f satisfy the condition of Theorem 23 (being ε-robust for sources with $b_1 + b_2 \geq 2n - k - 1 - \log_2 \varepsilon^{-1}$). Thus, these functions f have $W_\varepsilon^D(f) > n - 7 - 3\log_2 \varepsilon^{-1}$. To get the bound on $A_\varepsilon^R(f)$, we use Propositions 22 and 21:

$$\begin{aligned} A_\varepsilon^R(f) &\geq \max_{0<\delta\leq 1} (2-\delta)\varepsilon \cdot W^R_{\delta\varepsilon/2}(f) \\ &\geq (2 - 2/n)\varepsilon \cdot W^D_{\varepsilon/n}(f) \\ &> 2\varepsilon \cdot \left(n - 7 - 3\log_2 \varepsilon^{-1} n\right) - 1 \ . \end{aligned}$$

2) Similarly, by using Theorem 9, and setting $k = n - 3 - 3\log_2 \varepsilon^{-1}$.

QED (Theorem 20)

4.3 Upper Bounds

The lower bound on $A_\varepsilon^R(f)$ for almost all f's is nearly optimal, since Orlitsky and El Gamal showed that most $f \in F_n$ have $C_\varepsilon(f) \leq 2\varepsilon(n + 6\log_2 \varepsilon^{-1} n)$ [19] (recall $A_\varepsilon^R(f) \leq C_\varepsilon(f)$). The lower bound on $W_\varepsilon^R(f)$ is also nearly optimal, since we have the following upper bounds

Theorem 24:

1) For every $f \in F_n$ and every $2^{-\frac{n}{2}+1} < \varepsilon < 1/2$, $W_\varepsilon^D(f) \leq n + 11 - 2\log_2 \varepsilon^{-1}$.

2) For all $f \in F_n$, $W^D_{2^{-\frac{n}{2}+5.5}}(f) \leq 2$.

Proof: For part (1), we use the following protocol. Party B sends the $n + 9 - 2\log_2 \varepsilon^{-1}$ most significant bits of y to party A. This defines a 2^n-by-$2 \cdot (32\varepsilon)^{-2}$ strip in f's table. By Lemma 2 each such strip contains a 2^{n-4}-by-$(32\varepsilon)^{-2}$ submatrix S with $\frac{1}{2} + 32\varepsilon$ fraction of identical entries σ in it. In addition, party B sends a bit specifying whether y corresponds to a column in S. Party A replies by σ if (x, y) is in S, and by the outcome of a coin flip otherwise. This way, we get an average ε-advantage.

For part (2), let S be a 2^{n-4}-by-2^{n-1} submatrix containing a $\frac{1}{2} + 2^{(n-1)/2}$ fraction of identical entries σ in it (Lemma 2 guarantees the existence of S). Party B sends a bit specifying whether y corresponds to a column in S. Party A replies by σ if (x, y) is in S, and by the outcome of a coin flip otherwise. This way, we get an average $\frac{1}{32} \cdot 2^{(n-1)/2}$-advantage. □

4.4 Extension to (n,m)-distributions

In the definitions and results presented above, we have assumed that the inputs to the protocol are uniformly distributed in $\{0,1\}^n$. A natural question is what happens if we allow the inputs to be (n,m)-distibuted. In particular, we consider protocols which have advantage with respect to some (n,m)-distributions, and study the infimum average number of bits they exchanged under these "advantageous" distributions. We show that most functions in F_n have $\Theta(m)$ complexity, even in this weak measure.

Definition 10: Let D_1 and D_2 be two (n,m)-distributions, and let $P_i(z)$ be the probability that D_i assigns to z. For any protocol P, a function $f \in F_n$, and a metric $M_{P,f}$ over runs of P, we define the *average operator on* $D_1 \times D_2$

$$Ave_{D_1,D_2}(M_{P,f}) = \sum_{x,y\in\{0,1\}^n} Pr_1(x) \cdot Pr_2(y) \cdot M_{P,f}(x,y) \ .$$

Similarly, we define the *maximum operator on* $D_1 \times D_2$

$$Max_{D_1,D_2}(M_{P,f}) = \max_{x\in \mathrm{supp}(D_1), y\in \mathrm{supp}(D_2)} \{M_{P,f}(x,y)\} \ ,$$

where $\mathrm{supp}(D_i) = \{z \in \{0,1\}^n \ Pr_i(z) > 0\}$.

We say that the protocol P has $D_1 \times D_2$-*average* ε-*advantage in guessing* f if $Ave_{D_1,D_2}(S_{P,f}) \geq \frac{1}{2} + \varepsilon$.

- We define the *average randomized communication* (n,m)-*complexity of function* f (denoted (n,m)-$A_\varepsilon^R(f)$) as the infimum of $Ave_{D_1,D_2}(L_P)$, when taken over all (n,m)-distributions D_1, D_2 and all *randomized* protocols P which have $D_1 \times D_2$-average ε-advantage in guessing f.
- The *worst-case randomized communication* (n,m)-*complexity of function* f (denoted (n,m)-$W_\varepsilon^R(f)$) is defined as the infimum of $Max_{D_1,D_2}(l_P^*)$, when taken over all (n,m)-distributions D_1, D_2 and all *randomized* protocols P which have $D_1 \times D_2$-average ε-advantage in guessing f.
- The *deterministic communication* (n,m)-*complexities* of the function f are defined similarly, for deterministic protocols.

The key to dealing with (n,m)-complexities, is the fact that they are minimized on flat distributions (the proof is analogous to Lemma 5). Using the proof techniques of Theorem

20, the problem reduces to the existence of large submatrices with significant advantage inside the submatrix specified by the pair of flat distributions. We get

Theorem 25: Let $0 < \varepsilon \leq 1/2$.

1) For at least a $1 - 2^{-2^m}$ fraction of the Boolean functions $f \in F_n$,

$$(n,m)\text{-}W_\varepsilon^R(f) > m - 7 - 3\log_2 \varepsilon^{-1}$$
$$(n,m)\text{-}A_\varepsilon^R(f) > 2\varepsilon \cdot \left(m - 7 - 3\log_2 \varepsilon^{-1} m\right) - 1 \ .$$

2) For every $f \in F_n$ representable by a Hadamard matrix, the following holds.

$$(n,m)\text{-}W_\varepsilon^R(f) > 2m - n - 3 - 3\log_2 \varepsilon^{-1}$$
$$(n,m)\text{-}A_\varepsilon^R(f) > 2\varepsilon \cdot \left(2m - n - 3 - 3\log_2 \varepsilon^{-1} m\right) - 1 \ .$$

In particular, this holds for the inner-product function.

5. ON THE ROBUSTNESS OF BPP

The class R [1] and its symmetric version BPP [12] consist of problems which can be solved with high probability in polynomial time, with the use of an unbiased coin. Recently, Vazirani and Vazirani [29] showed that all BPP problems can be efficiently solved even if a single SV-source is producing the coin tosses. In this section, we generalize their result by showing that BPP problems can be efficiently solved if a (single) PRB-source is producing the coin tosses.

The main idea of the proof is that *any* function which is robust with respect to *two independent* PRB-sources, can be used to produce *polynomially many* bits such that almost all of them are unbiased. Repeating this process m times, we get $poly(m)$ strings of length m each. Most of these strings are almost uniformly distributed, and thus the fraction of these strings which hit the witness set $W \subset \{0,1\}^m$ is close to the density of W. If W's density is large enough (say ≥ 0.8) then with probability bounded away from 0.5 (e.g. ≥ 0.55), the majority of the generated strings hit W. This argument need careful formalization which is carried out below. The final observation is that there are explicit and efficiently computable functions which are appropriate for the above procedure (e.g. the inner product or the Paley graph functions).

The key technical lemma used in the proof is

Lemma 26: Let $0 < \varepsilon < 1$ be a real and $f : \{0,1\}^l \times \{0,1\}^l \mapsto \{0,1\}$ be a Boolean function. Define $f_i : \{0,1\}^l \mapsto \{0,1\}$ by $f_i(j) = f(i,j)$ for every $i,j \in \{0,1\}^l$. Suppose that the f is ε-robust with respect to any two independent random variables which are $(l, l-1)$-distributed and (l,b)-distributed respectively. Then for every (l,b)-distributed Y, all but $\sqrt{\varepsilon}$ fraction of the f_i's are $4\sqrt{\varepsilon}$-robust on Y.

The reader may find it convenient to picture the two-argument Boolean function $f : \{0,1\}^l \times \{0,1\}^l \mapsto \{0,1\}$, as a table where the (i,j)-entry corresponds to $f(i,j)$. The lemma can be stated (informally) as follows: *if a function can be used for extracting almost unbiased bits from the output of any two independent PRB-sources, then most of its "rows" can be used for extracting an almost unbiased bit from a single PRB-source.* The identity of these "good" rows depends on the specific PRB-source, but for each source most of the rows will work. The proof is by contradiction, showing that if the conclusion of the lemma is violated then it is possible to find a pair of probability bounded sources which falsify

the robustness of f. While the proof of this lemma is rather simple, it seems much harder to prove a similar statement based merely on robustness with respect to SV-sources.

Proof: Let Y be an arbitrary (l,b)-distributed random variable. This defines a probability space on the strings in $\{0,1\}^l$. We'll show that the number of rows which are biased too much towards 1 is small (rows with high 0 bias are treated identically). Let B denote the set of these rows, that is

$$B = \left\{ i : Pr\left(f_i(Y) = 1\right) > \frac{1}{2}\left(1 + 4\sqrt{\varepsilon}\right) \right\} ,$$

and let k denote B's size.

We first show that $k \leq 2^{l-1}$. Assume on the contrary that $k > 2^{l-1}$. We will reach a contradiction by defining a $(l, l-1)$-distributed source X_1 to be flat on B. Then

$$\begin{aligned} Pr\left(f(X_1,Y) = 1\right) &= k^{-1} \sum_{i \in B} Pr\left(f_i(Y) = 1\right) \\ &> \frac{1}{2} \cdot \left(1 + 4\sqrt{\varepsilon}\right) \\ &> \frac{1}{2} \cdot (1 + \varepsilon) \ . \end{aligned}$$

Now that we know $k \leq 2^{l-1}$, we define a $(l, l-1)$-distributed source X_2 to be flat on $\{0,1\}^l - B$. Applying the ε-robustness of f to the uniform source X_0, we have

$$\begin{aligned} \frac{1}{2^l} \sum_{i \in \{0,1\}^l} Pr\left(f_i(Y) = 1\right) &= Pr\left(f(X_0,Y) = 1\right) \\ &< \frac{1}{2}\left(1 + \varepsilon\right) \ . \end{aligned}$$

In order to bound k from above, we first derive an upper bound on $Pr\left(f(X_2,Y) = 1\right)$.

$$\begin{aligned} Pr\left(f(X_2,Y) = 1\right) &= \frac{1}{2^l - k} \cdot \sum_{i \in \{0,1\}^l - B} Pr\left(f_i(Y) = 1\right) \\ &= \frac{\sum_{i \in \{0,1\}^l} Pr\left(f_i(Y) = 1\right) - \sum_{i \in B} Pr\left(f_i(Y) = 1\right)}{2^l - k} \\ &< \frac{2^l \cdot \left(\frac{1}{2}(1+\varepsilon)\right) - k \cdot \left(\frac{1}{2}(1+4\sqrt{\varepsilon})\right)}{2^l - k} \\ &= \frac{1}{2}\left(1 + \frac{2^l\varepsilon - k \cdot 4\sqrt{\varepsilon}}{2^l - k}\right) \ . \end{aligned}$$

Applying the ε-robustness of f to X_2 we get

$$Pr\left(f(X_2,Y)=1\right) > \frac{1}{2}\left(1-\varepsilon\right) \ .$$

Combining the upper and lower bounds on $Pr\left(f(X_2,Y)=1\right)$, we have $\frac{1}{2}\left(1+\frac{2^l\varepsilon-4k\sqrt{\varepsilon}}{2^l-k}\right) \geq \frac{1}{2}\left(1-\varepsilon\right)$. By a simple manipulation, $k \leq \frac{1}{2}\sqrt{\varepsilon}\cdot 2^l$ follows. □

With Lemma 26 at our disposal, we can use a single source to generate many strings, most of which are almost unbiased. These strings are generated one bit at a time (i.e. in step t, the t-th bit of all strings is generated). In the BPP application, the generated strings are tested for membership in a witness set W. A careful probabilistic analysis shows that if W is dense enough, there is a fairly large probability that the *majority* of these strings will hit W.

Proposition 27:

1) Let m be an integer, and $W \subset \{0,1\}^m$ be an *arbitrary* set. Denote by p the density of W (i.e. $p \stackrel{\text{def}}{=} |W|/2^m$), and let $q \stackrel{\text{def}}{=} 1-p$.

2) Let l be an integer, and $0 < b < l$. Let $0 < \varepsilon < 1/2$ and $f : \{0,1\}^l \times \{0,1\}^l \mapsto \{0,1\}$ be a Boolean function. Suppose that f is ε-robust with respect to any two independent random variables which are $(l, l-1)$-distributed and (l,b)-distributed respectively.

3) Let $Y_1, Y_2, \ldots Y_m$ be a sequence of *arbitrary* random variables assuming values in $\{0,1\}^l$ such that for every $1 \leq t \leq m$ the variable Y_t is (l,b)-distributed given $Y_1, Y_2, \ldots, Y_{t-1}$. Let Y denote the concatenation of the Y_t's.

4) For every $i \in \{0,1\}^l$, let $f_i : \{0,1\}^l \mapsto \{0,1\}$ be defined as in Lemma 26 (i.e. $f_i(j) = f(i,j)$ for every $i,j \in \{0,1\}^l$). Let $h_i(Y)$ be a random variable assuming values in $\{0,1\}^m$, such that $h_i(Y)$ is the concatenation of the random variables $f_i(Y_1)$, $f_i(Y_2)$,..., $f_i(Y_m)$.

Then the probability that a majority of the $h_i(Y)$'s *miss* W, does not exceed $2q + 8m\sqrt{\varepsilon}$. That is,

$$Pr\left(\left|\left(\left\{h_i(Y) : i \in \{0,1\}^l\right\} \bigcap W\right)\right| \leq 2^{l-1}\right) \leq 2q + 8m\sqrt{\varepsilon} \ .$$

5.1 Proof of Proposition 27

Convention: Throughtout the proof, the probability space is the cartisian product of Y (of item 3 above) with an (m,m)-distributed random variable Z. The random variable $Y = Y_1Y_2\cdots Y_m$ assumes values in $\{0,1\}^{m\cdot l}$, and $Z = Z_1Z_2\cdots Z_m$ is uniformly distributed in $\{0,1\}^m$ independently of Y. A value which Y may assume will be denoted by $\alpha = \alpha_1\alpha_2\cdots\alpha_m$, where $\alpha_t \in \{0,1\}^l$. A value which Z may assume will be denoted by $\beta = \beta_1\beta_2\cdots\beta_m$, where $\beta_t \in \{0,1\}$. We will also use the notation $Y'_t = Y_1Y_2\cdots Y_t$ and $\alpha'_t = \alpha_1\alpha_2\cdots\alpha_t$ to denote the prefix consisting of first t elements of Y and α respectively.

Definitions:

1) For every $i \in \{0,1\}^l$, $0 \le t \le m$, we define a Boolean function $\xi_{i,t} : \{0,1\}^{lm} \times \{0,1\}^m \mapsto \{0,1\}$ as follows:

$$\xi_{i,t}(\alpha,\beta) = \begin{cases} 0 & \text{if} \quad f_i(\alpha_1)f_i(\alpha_2)\cdots f_i(\alpha_t)\beta_{t+1}\beta_{t+2}\cdots\beta_m \in W \\ 1 & \text{otherwise} \end{cases}$$

2) For every $i \in \{0,1\}^l$, $0 \le t \le m$ we define a Boolean function $\eta_{i,t} : \{0,1\}^{lm} \mapsto \{0,1\}$ as follows:

$$\eta_{i,t}(\alpha) = \begin{cases} 0 & \text{if} \quad \frac{1}{2} - 2\sqrt{\varepsilon} \le Pr\left(f_i(Y_{t+1}) = 1 | Y'_t = \alpha'_t\right) \le \frac{1}{2} + 2\sqrt{\varepsilon} \\ 1 & \text{otherwise} \end{cases}$$

Explanation: Letting α assume values in $\{0,1\}^{lm}$ according to the random variable Y, and β assume values in $\{0,1\}^l$ according to the uniform distribution Z, the two functions above induce two random variables $\xi_{i,t}(Y,Z)$ and $\eta_{i,t}(Y)$. These random variables correspond to hybrids of the (l,b)-source and truely unbiased, independent coin tosses. The random variable $\xi_{i,t}(Y,Z)$ equals 0 ("a success") when an hybrid element, generated by applying the function f_i to the first t blocks output by the source Y and letting the rest be truely random, hits the set W. The random variable $\eta_{i,t}(Y)$ equals 0 ("a good bit") if, given the first t blocks of the source, the bit generated by applying f_i to the $(t+1)$-st block is almost unbiased.

Elementary Observations

Fact 1: For every $i \in \{0,1\}^l$, we have

$$Pr\left(\xi_{i,0}(Y,Z)=0\right)=p\ .$$

Proof: By definition $\xi_{i,0}(Y,Z)=0$ iff $Z \in W$, and $|W|=p2^m$. □

Fact 2:

$$\sum_{i\in\{0,1\}^l}\sum_{t=0}^{m-1} Exp\left(\eta_{i,t}(Y)\right) \leq 2^l m\sqrt{\varepsilon}\ .$$

Proof: By Lemma 26, for every $0 \leq t < m$ and every $\alpha \in \{0,1\}^{m\cdot l}$ we have

$$\sum_{i\in\{0,1\}^l} \eta_{i,t}(\alpha) \leq 2^l\sqrt{\varepsilon}\ .$$

Thus for every $0 \leq t \leq m-1$,

$$\sum_{i\in\{0,1\}^l} \eta_{i,t}(Y) \leq 2^l\sqrt{\varepsilon}\ .$$

The sum of m such expressions (for the m values of t) is thus bounded above by $2^l m\sqrt{\varepsilon}$, and so is the expected value. Changing the order of summation, we get the claimed bound. □

The next Fact formulates the intuition that, *when the $(t+1)$-st bit produced in the i-th row is almost unbiased*, then the $(t+1)$-th hybrid of this row has almost the same success probability as the t-th hybrid to hit W.

Fact 3: For every $i \in \{0,1\}^l$ and $0 \leq t < m$, we have

$$Pr\left(\xi_{i,t+1}(Y,Z)=0|\eta_{i,t}(Y)=0\right) \geq Pr\left(\xi_{i,t}(Y,Z)=0|\eta_{i,t}(Y)=0\right) - 2\sqrt{\varepsilon}\ .$$

Proof: Consider an arbitray $\alpha \in \{0,1\}^{lm}$ such that $\eta_{i,t}(\alpha)=0$. Let $r = Pr\left(f_i(Y_{t+1})=1|Y'_t=\alpha'_t\right) - \frac{1}{2}$. Then $|r| \leq 2\sqrt{\varepsilon}$. Let $s = Pr\left(\xi_{i,t}(Y,Z)=0|Y'_t=\alpha'_t\right)$, $s_0 = Pr\left(\xi_{i,t}(Y,Z)=0|Y'_t=\alpha'_t, Z_{t+1}=0\right)$, $s_1 = Pr\left(\xi_{i,t}(Y,Z)=0|Y'_t=\alpha'_t, Z_{t+1}=1\right)$. By definition $s=\frac{1}{2}s_0+\frac{1}{2}s_1$. Then

$$\begin{aligned} Pr\left(\xi_{i,t+1}(Y,Z)=0|Y'_t=\alpha'_t\right) &= \left(\frac{1}{2}-r\right)\cdot s_0 + \left(\frac{1}{2}+r\right)\cdot s_1 \\ &= s - r(s_0-s_1) \\ &\geq s - 2\sqrt{\varepsilon} \end{aligned}$$

Summing over all α'_t's, Fact 3 follows. □

Probability Calculation

The next fact is cruicial to our proof. It expresses the (unconditional) success probability of the $(t+1)$-st hybrid of the i-th row, in terms of the t-th hybrids of this row. The difference between the the $(t+1)$-st and t-th hybrids is bounded by the sum of a small error probability ($\leq 2\sqrt{\varepsilon}$), introduced by runs in which the $(t+1)$-st bit is almost unbiased, and the probability that the $(t+1)$-st bit is biased.

Fact 4: For every $i \in \{0,1\}^l$ and $0 \leq t < m$, we have

$$Pr\left(\xi_{i,t+1}(Y,Z)=0\right) \geq Pr\left(\xi_{i,t}(Y,Z)=0\right) - 2\sqrt{\varepsilon} - 2 \cdot Pr\left(\eta_{i,t}(Y)=1\right)$$

Proof: By following manipulation, using Fact 3 (when passing from the first line to the second line), and the inequality $Pr(A|B) \geq Pr(A) - Pr(\overline{B})$ (passing from the third line to the fourth line).

$$\begin{aligned}
Pr\left(\xi_{i,t+1}(Y,Z)=0\right) &\geq Pr\left(\xi_{i,t+1}(Y,Z)=0 | \eta_{i,t}(Y)=0\right) \cdot Pr\left(\eta_{i,t}(Y)=0\right) \\
&\geq \left(Pr\left(\xi_{i,t}(Y,Z)=0 | \eta_{i,t}(Y)=0\right) - 2\sqrt{\varepsilon}\right) \cdot \left(1 - Pr\left(\eta_{i,t}(Y)=1\right)\right) \\
&\geq Pr\left(\xi_{i,t}(Y,Z)=0 | \eta_{i,t}(Y)=0\right) - 2\sqrt{\varepsilon} - Pr\left(\eta_{i,t}(Y)=1\right) \\
&\geq Pr\left(\xi_{i,t}(Y,Z)=0\right) - 2\sqrt{\varepsilon} - 2 \cdot Pr\left(\eta_{i,t}(Y)=1\right) \quad \square
\end{aligned}$$

This yields an upper bound on the probability that the i-th row does not hit W, when being generated from the blocks of the (l,b)-source using the function f_i.

Fact 5: For every $i \in \{0,1\}^l$

$$Exp\left(\xi_{i,m}(Y,Z)\right) \leq q + 2\sqrt{\varepsilon}m + 2\sum_{t=0}^{m-1} Exp\left(\eta_{i,t}(Y)\right)$$

Proof: By combining Fact 1 with repeated use of Fact 4, and using the fact that for any $0-1$ random variable V, $Exp(V) = Pr(V=1)$. □

Conclusion

We now bound the probability that the majority of rows produce elements which *do not hit* the set W.

Fact 6:

$$Pr\left(\sum_{i\in\{0,1\}^l}\xi_{i,m}(Y,Z)\geq 2^{l-1}\right)\leq 2q+8\sqrt{\epsilon}m$$

Proof: Applying Markov inequality the sum of the $\xi_{i,t}(Y,Z)$, using Fact 5 (when passing from the first line to the second line) and Fact 2 (when passing from the second line to the third line).

$$\begin{aligned}Pr\left(\sum_{i\in\{0,1\}^l}\xi_{i,m}(Y,Z)\geq 2^{l-1}\right) &\leq \frac{Exp\left(\sum_{i\in\{0,1\}^l}\xi_{i,m}(Y,Z)\right)}{2^{l-1}}\\ &\leq 2^{-l+1}\cdot\left(2^l\cdot(q+2\sqrt{\epsilon}m)+2\cdot\sum_{i\in\{0,1\}^l}\sum_{t=0}^{m-1}Exp(\eta_{i,t}(Y))\right)\\ &\leq 2q+4\sqrt{\epsilon}m+4\cdot\sqrt{\epsilon}m\\ &= 2q+8\sqrt{\epsilon}m \quad \square\end{aligned}$$

Since $\xi_{i,m}(Y,Z)=1$ is just a fancy way of writing $h_i(Y)\notin W$, we get

$$Pr\left(\left|\left(\{h_i(Y):i\in\{0,1\}^l\}\bigcap W\right)\right|\leq 2^{l-1}\right)\leq 2q+8\sqrt{\epsilon}m\ .$$

QED (proposition 27)

5.2 The Transformation of BPP Algorithms

It will be convenient to consider randomized algorithms as deterministic algorithms with an auxiliary random input. The performance of such an algorithm (on input x) will be evaluated with respect to the distribution of the auxiliary random input (denoted y). The issue of the robustness of the class BPP is then stated as follows: *can a BPP algorithm be converted to a polynomial-time algorithm which has an advantage bounded away from 1/2 even when its random input is generated by a single probability-bounded source ?*

Definition: The class (l,b)-BPP consists of all decision problems $D:\{0,1\}^*\mapsto\{0,1\}$ for which there exist polynomials P,Q and an algorithm $A:\{0,1\}^*\times\{0,1\}^*\mapsto\{0,1\}$ such that *for every (l,b)-source Y* the following holds:

1) On each input of length n, algorithm A runs at most $P(n)$ steps, and then stops, outputing a single bit.

2) Let $x \in \{0,1\}^m$ be an input, and $y \in \{0,1\}^{P(n)}$ be the auxiliary random input generated by the (l,b)-source Y. Then

$$Pr\left(A(x,y) = D(x)\right) \geq \frac{1}{2} + Q^{-1}(n) \ .$$

Clearly, (l,l)-BPP is just a fancy way of writing BPP. Theorem 28 states that so is (l,b)-BPP.

Theorem 28: For every integer l and real $0 < b \leq l$,

$$(l,b)\text{-BPP} = \text{BPP} \ .$$

Proof: Let D be a decision problem in BPP, and A_0 be a randomized polynomial-time algorithm for D. Let $P(n)$ be the number of random bits used by the algorithm A_0 on inputs of length n. Without loss of generality, we may assume that for every input $x \in \{0,1\}^m$, the witness set $W(x)$ for x (i.e. the y's satisfying $A_0(x,y) = D(x)$) contains $p \geq 0.8$ of the strings in $\{0,1\}^{P(n)}$.

Given l and b, let $\varepsilon(n) \stackrel{\text{def}}{=} (160 \cdot P(n))^{-2}$, $B(n) \stackrel{\text{def}}{=} 3 + 2\log_2 \varepsilon^{-1}(n)$, and $L(n) \stackrel{\text{def}}{=} \lceil lB(n)/b \rceil$. By Theorem 9, every function corresponding to a Hadamard matrix is $\varepsilon(n)$-robust with respect to any pair of independent random variables X, Y which are $(L(n), L(n)-1)$-distributed and $(L(n), B(n))$-distributed, respectively. Furthermore, some of these families of functions, such as the inner-product modulo 2 or the quadratic resiuuocity modulo a prime (the Paley Graph function), can be computed by $poly(n)$-time algorithms. Let f be one of these functions. Let F be an algorithm that on inputs n and $i,j \in \{0,1\}^{L(n)}$, outputs $f(i,j)$.

By Proposition 27, we can use a single $(L(n), B(n))$-source to efficiently generate $2^{L(n)}$ strings such that for every $x \in \{0,1\}^n$, the majority of these strings hit the witness set $W(x)$, with probability greater than $1 - 2q - 8P(n)\sqrt{\varepsilon(n)} \geq 0.55$ $(q \leq 0.2)$. We remind the reader that the (l,b)-source can be used as an $(L(n), B(n))$-source. Consider the following algorithm A for deciding membership in D (with a two-sided error bounded above by 0.45).

1. Algorithm A
2. INPUT $\leftarrow x$
; Let $n = |x|$, and $m = P(n)$.
3. AUXILIARY INPUT: $y \in \{0,1\}^{L(n)m}$ generated by an arbitrary (l,b)-source.

; Let $y = y_1 y_2 \cdots y_m$, where each $y_t \in \{0,1\}^{L(n)}$.
4. For every $i \in \{0,1\}^{L(n)}$ and $1 \leq j \leq m$,
compute $f(i, y_t)$, by invoking $F(n, i, y_t)$.
; For every $i \in \{0,1\}^{L(n)}$, let w_i denote the concatenation $f(i,y_1) \cdot f(i,y_2) \cdots f(i,y_m)$.
5. For every $i \in \{0,1\}^{L(n)}$, compute $v_i \stackrel{\text{def}}{=} A_0(x, w_i)$.
6. If $\sum_{i \in \{0,1\}^{L(n)}} v_i \leq 2^{L(n)-1}$ then $d \leftarrow 0$ else $d \leftarrow 1$.
7. OUTPUT: d.

By the above discussion, $Pr\,(A(x,y) = D(x)) \geq 0.55$. The running time of $A(x,y)$ is polynomial in $2^{L(n)} = n^{O(1)}$, and in the running time of $A_0(x,\cdot)$ and $F(|x|,\cdot,\cdot)$. The Theorem follows. □

Remarks:

1) Clearly, the same algorithm works also for the class R. It produces one sided error, since for $x \notin D$ the original algorithm A_0 never errs.

2) Proposition 27 can be viewed as providing a method for using a single PRB source to distinguish "high density" sets from the complement of "high density" sets. That is, given $K \subset \{0,1\}^k$ so that either $|K| \geq p2^k, |\overline{K}| \leq q2^k$ or $|\overline{K}| \geq p2^k, |K| \leq q2^k$, determine which of the two cases occur, with success probability about $1 - 2q$.

   This view point is helpful in solving the following additive approximation problem: For any $\epsilon, \delta > 0$, and any set $S \subset \{0,1\}^m$, find an *additive* (δ, ϵ) approximation of the S density, ϱ, using a single probability-bounded source (when we have an oracle for deciding membership in S). By an *additive* (δ,ϵ)*-approximation* we mean that with probability $\geq 1 - \epsilon$, $|\varrho - a| \leq \delta$ where a is the approximated value and $\varrho \stackrel{\text{def}}{=} |S|/2^m$ is the true density. To get this approximation, we first transform the problem of additive approximation into $2/\delta$ problems of the form "P_j: is $\varrho \in [j \cdot \frac{\delta}{2}, j \cdot \frac{\delta}{2} + \delta]$?" where $0 \leq j \leq (2 - 2\delta)/\delta$ (notice that every pair of consecutive intervals overlaps by $\delta/2$). By sampling $k = O(\delta^{-2} \log(\epsilon\delta)^{-1})$ points in $\{0,1\}^m$, and counting the number of times S is hit, every P_j is translated into a subset $S_j \subset \{0,1\}^{mk}$. It is easy to see that for at most two consecutive j's in the above range, S_j has more than $(1 - \epsilon\delta)2^{kn}$ points, while all other j 's (execpt possibly another consecutive j) have fewer than $\epsilon\delta 2^{kn}$ points. Now we use the ideas above to try and hit all S_j's by strings generated from a single probability-bounded source. With probability $> 1 - \epsilon$, we get positive answers only for j's in a δ neighbourhood of ϱ. In case we get positive answers for several P_j's, we choose the median j, and estimate ϱ as being in the middle of the jth interval.

6. CONCLUSIONS

We have presented a new model of sources of weak randomness, and distilled its "hard core": the class of probability-bounded distributions. Probability-bounded distributions constitute a natural and wide class, which is convenient to analyze and yields strong results.

ACKNOWLEDGMENTS

We wish to thank Baruch Awerbuch for collaborating with us at the early stages of this research. We are very grateful to Noga Alon, László Babai and Carl Pomerance for sharing with us some of their mathematical knowledge. Noga Alon has referred us to the properties of Hadamard matrices, which turned out to be very suitable for our purposes. László Babai has referred us to the k-th residues construction used in section 3, and Carl Pomerance has provided us with the estimates on the density of primes suitable for this construction. Their help enabled us to get the extraction scheme which is efficient in both rate and computation measures.

We are grateful to Jeff Lagarias for suggesting to consider "extraction at the limit", Nati Linial for pointing out Theorem 4, Mike Luby for pointing out that the entropy bound does suffice, Andrew Odlyzko for referring us to the Paley Graph Conjecture, and Avi Wigderson for his interpretation of the result in [29].

We also wish to thank Reuven Bar-Yehuda, Shimon Even, Michael Fischer, Mihali Gereb, Shafi Goldwasser, Abraham Lempel, Leonid Levin, Silvio Micali, Ron Rivest, David Shmoys, and Umesh Vazirani for very helpful discussions.

REFERENCES

[1] Adleman, L., "Two Theorems on Random Polynomial Time", *Proc. 19th FOCS*, Oct. 1978, pp. 75-83.

[2] Ajtai, M., L. Babai, P. Hajnal, J. Komolós, P. Pudlák, V. Rödl, E. Szemerédi, and G. Turán, "Two Lower Bounds for Branching Programs", *Proc. 18th STOC*, May 1986, pp. 30-38.

[3] Alon, N., private communication (1985).

[4] Alon, N., P. Frankl, and V. Rödl, "Geometric Realization of Set Systems and Probabilistic Communication Complexity", *Proc. 26the FOCS*, Oct. 1985, pp. 277-280.

[5] Blum, M., "Independent Unbiased Coin Flips from a Correlated Biased Source: a Finite State Markov Chain", *Proc. 25th FOCS*, Oct. 1984, pp. 425-433.

[6] Bondy, J.A., and U.S.R. Murty, *Graph Theory with Applications*, American Elsevier Publishing Co., Inc, (1976).

[7] Canfield, E.R., P. Erdös, and C. Pomerance, "On a Problem of Oppenheim Concerning "Factorisatio Numerorum"," *Jour. of Number Theory*, Vol. 17, No. 1, 1983, pp. 1-28.

[8] Chor, B., and O. Goldreich, "Unbiased Bits from Sources of Weak Randomness and Probabilistic Communication Complexity", *Proc. 26th FOCS*, Oct. 1985, pp. 429-442.

[9] Dixon, J.D., "Asymptotically Fast Factorization of Integers", *Math. Comp.*, 36, 1981, pp. 255-260.

[10]Elias, P., "The Efficient Construction of an Unbiased Random Sequence", *Ann. Math. Statist.*, Vol. 43, No. 3, 1972, pp. 865-870.

[11]Erdös, P., and J. Spencer, *Probabilistic Methods in Combinatorics*, Academic Press, (1974).

[12]Gill, J., "Complexity of Probabilistic Turing Machines", *SIAM Jour. on Computing*, Vol. 6, No. 4, 1977, pp. 675-695.

[13]Hall, M. Jr., *Combinatorial Theory*, Blaisdell Publishing Co., (1967).

[14]Ja'Ja', J., V.K. Prasanna Kumar, and J. Simon, "Information Transfer under Different Sets of Protocols", *SIAM Jour. on Computing*, Vol. 13, No. 4, 1984, pp. 840-849.

[15]Johnson, N.J., and S. Kotz, *Distributions in Statistics*, Vol. 1, *Discrete Distributions*, John Wiley & Sons, 1969.

[16]McWilliams, F.J., and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland Publishing Company, 1977.

[17]von Neumann, J., "Various Techniques Used in Connection with Random Digits", notes by Forsythe G.E., National Bureau of Standards, Applied Math. Series, 1951, Vol. 12, pp. 36-38. Reprinted in *von Neumann's Collected Works*, Vol. 5, Pergamon Press (1963), pp. 768-770.

[18]Odlyzko, A.M., "Discrete Logarithms in Finite Fields and their Cryptographic Significance", *Advances in Cryptology: Proceedings of EuroCrypt84*, T. Beth et al., eds., Springer–Verlag, 1985, pp. 224–314.

[19]Orlitsky, A., and A. El Gamal, "Randomized Communication Complexity", preprint, (1985).

[20]Papadimitriou, C.H., and M. Sipser, "Communication Complexity", *Jour. of Comp. and Sys. Sci.*, Vol. 28, No. 2, 1984, pp. 260-269.

[21]Papadimitriou, C.H., and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Inc. (1984).

[22]Paturi, R., and J. Simon, "Probabilistic Communication Complexity", *Proc. 25th FOCS*, Oct. 1984, pp. 118-126.

[23]Pohlig, R.C. and M. Hellman, "An improved algorithm for computing logarithms over $GF(p)$ and its cryptographic significance", *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 106–110, 1978.

[24]Pratt, V., "Every Prime has a Succient Certificate", *SIAM J. Comt.*, 1975, pp. 214-220.

[25]Rényi, A., *Probability Theory*, North-Holland Publishing Company (1970).

[26]Santha, M., and U.V. Vazirani, "Generating Quasi-Random Sequences from Slightly-Random Sources", *Proc. 25th FOCS*, Oct. 1984, pp. 434-440.

[27]Schmidt, W.M., *Equations over Finite Fields: An Elementary Approach*, Lecture Notes in Mathematics, Vol. 536, Springer-Verlag, Berlin, 1976.

[28]Vazirani, U.V., "Towards a Strong Communication Complexity Theory or Generating Quasi-Random Sequences from Two Communicating Slightly-Random Sources", *Proc. 17th STOC*, May 1985, pp. 366-378.

[29]Vazirani, U.V., and V.V. Vazirani, "Random Polynomial Time is Equal to Slightly-random Polynomial Time", *Proc. 26th FOCS*, Oct. 1985, pp. 417-428.

[30]Yao, A.C., "Probabilistic Computations: Towards a Unified Measure of Complexity", *Proc. 18th FOCS*, Oct. 1977, pp. 222-227.

[31]Yao, A.C., "Some Complexity Questions related to Distributive Computing", *Proc. 11th STOC*, April 1979, pp. 209-213.

[32]Yao, A.C., "Lower Bounds by Probabilistic Arguments", *Proc. 24th FOCS*, Nov. 1983, pp. 420-428.