* ** ***

THE BIT EXTRACTION PROBLEM OR t–RESILIENT FUNCTIONS

(Preliminary Version)

Benny Chor * Oded Goldreich ** Johan Hastad
Laboratory for Computer Science, MIT

Joel Freidmann Steven Rudich *** Roman Smolensky
EECS, UC-Berkeley

ABSTRACT – We consider the following adversarial situation. Let n, m and t be arbitrary integers, and let $f:\{0,1\}^n \mapsto \{0,1\}^m$ be a function. An adversary, knowing the function f, sets t of the n input bits, while the rest ($n-t$ input bits) are chosen at random (independently and with uniform probability distribution). The adversary tries to prevent the outcome of f from being uniformly distributed in $\{0,1\}^m$.

The *question* addressed is for what values of n, m and t does the adversary necessarily fail in biasing the outcome of $f:\{0,1\}^n \mapsto \{0,1\}^m$, when being restricted to set t of the input bits of f. We present various lower and upper bounds on m's allowing an affirmative answer. These bounds are relatively close for $t \leq n/3$ and for $t \geq 2n/3$. Our results have applications in the fields of fault-tolerance and cryptography.

1. INTRODUCTION

The *bit extraction* problem formulated above The bit extraction problem was suggested by Brassard and Robert [BRref] and by Vazirani [Vref].an be viewed as a three move game between a *user* and an *adversary*. The game is parametrized by the integers n, m and t; and proceeds as follows. First, the *user* picks a function $f:\{0,1\}^n \mapsto \{0,1\}^m$. (The function f will be applied to a n-bit string.) Next, the *adversary* picks t locations in the input n-bit string and sets the bit values of these locations. The *user* does not know which locations and what values were chosen by the *adversary*. The remaining $n-t$ bits of the string are set by the outcomes of independent unbiased coin tosses. Finally, the *user* applies the function f to the entire string. The *user*'s objective is to cause the output of the function to be uniformly distributed in $\{0,1\}^m$; while the objective of the *adversary* is to prevent this. The question is which of the parties (*user* or *adversary*) has a

* Supported in part by an IBM Graduate Fellowship and a Bantrell Postdoctoral Fellowship.
** Supported in part by a Weizmann Postdoctoral Fellowship. On leave from the CS Dept., Technion, Israel.
*** Supported in part by an IBM Graduate Fellowship and NSF Grant MCS 82-04506.
The bit extraction problem was suggested by Brassard and Robert [BRref] and by Vazirani [Vref]. c

winning strategy.

It is evident that the *user* has a winning strategy in the following two extreme cases:

1)$m = 1$ and $t \leq n - 1$ (by XORing all the bits).

2)$t = 1$ and $m \leq n - 1$ (by XORing every two adjacent bits).

In both cases $m \leq n - t$. On the other hand, the *adversary* has a winning strategy when $m > n - t$. *Can the user win whenever* $m \leq n - t$? We show that the answer is negative. In particular, the *adversary* has a winning strategy in the following two cases:

1)When $m = 2$ and $t \geq \lfloor 2n/3 \rfloor$.

2)When $t = 2$ and $m \geq n - \log_2(n+1)$.

Lower and Upper Bounds

Before proceeding any further, let us state the bounds we obtain on the number of extractable bits. Let n, m and t be as above. Let $Bit(n,t)$ denote the maximal m for which the *user* has a winning strategy (when playing against an adversary who fixes t out of the n bits). We now state the lower and upper bounds on $Bit(n,t)$ and approximate these expressions for $t = o(n)$.

$$\otimes Bit(n,t) \geq n - \log_2 \sum_{i=0}^{t-1} \binom{n-1}{i} \approx n - t \cdot \log_2 \frac{n}{t}$$

$$\otimes Bit(n,t) \leq n - \log_2 \sum_{i=0}^{\lfloor t/2 \rfloor} \binom{n}{i} \approx n - \lfloor \frac{t}{2} \rfloor \cdot \log_2 \frac{n}{t}$$

Relation to Error Correcting Codes

Note the similarity and difference between the "extraction game" and the "error correcting game" hereby presented. First the *user* chooses two functions $f_e : \{0,1\}^m \mapsto \{0,1\}^n$ and $f_d : \{0,1\}^n \mapsto \{0,1\}^m$, a m-bit string s and applies f_e to s. Next, the *adversary* may alter any t bits of $c = f_e(s)$ resulting in a string c'. Finally, the *user* applies f_d to c'. In the theory of error correcting codes, the objective of the *user* is to always retrieve s. Although the two games are different, we show that they have close relationship when *(in both games) the user is restricted to use linear* $GF(2)$ *transformations.* This relationship implies lower bounds on the number of extractable bits. We show that these lower bounds (obtained by linear transformations) are close to being optimal, by proving an upper bound on the number of extractable bits using general extraction functions.

1.1 Fault-Tolerance Application :

How to agree on a shared random string

Consider a synchronous communication network consisting of n processors, each having a perfect source of random bits (i.e. the source's output is a sequence of independent unbiased coin flips). Suppose that the processors wish to share a common randomly selected bit string. This can be trivially achieved if one processor just transmits to all processors the output of his local source. Things become more difficult if there is a danger that some local sources are *faulty* and their output is no longer unbiased. Still a trivial solution exists: each processor can transmit the next k bits output by his local source, and then take the bit-by-bit exclusive-or of all the transmitted k-bit strings. This protocol yields a shared k-bit string with uniform probability distributed, as long as one of the local sources is not faulty. However, this solution is very wasteful. The ratio of the number of extracted bits over the number of transmitted bits is $\frac{1}{n}$.

Much more efficient solutions are implied by our results. For example, suppose that it is guaranteed that at most t of the local sources are faulty. Then using the function presented in Section 2, we can present a protocol which is both efficient in terms of rate and robust in the presence of at most t faults. Each processor randomly chooses and transmits a $\lceil \log_2 n \rceil$-bit string, and then applies the function to the concatenation of all the strings, resulting in a $(n-t) \cdot \lceil \log_2 n \rceil$ bit string. The ratio of extracted/transmitted bits is $\frac{n-t}{n}$, and the resulting bit string is uniformly distributed in $\{0,1\}^{(n-t) \cdot \log_2 n}$, as long as at most t local sources are faulty. This result is optimal in terms of rate versus number of faults, since we get as many unbiased global bits as the number of unbiased local bits. Our solution holds also in the more general fault model of *simultanous networks* [ACGMref].

1.2 Cryptographic Application :

Renewing a Partially Leaked Key

Suppose that two parties share a secret, randomly selected n-bit key, various parts of which they use for various purposes. Suppose that at some moment an eavesdropper has succeeded in finding out t of the bits of the key (but the parties do not know which t bits these are). As this may endanger tasks which rely only on t bits, the parties wish to have a completely new and secert key. A trivial solution is to let one party randomly choose a new key and secretly transmit it to the other. This requires randomization as well as communication resources. Our results allow solutions which cost nothing in terms of randomization and communication.

A new shared key can be determinstically computed from the old one, by each party, without any communication between them. The new key is completely secret, as its bits are independent and unbiased with respect to the eavesdropper who only knows t bits of the old key. It should be stated that the new key is shorter than the old one. In particular, for "small" t's, the length of the new key is $n - t \cdot \lceil \log_2 n \rceil$ (this is close to optimal).

Other cryptographic applications of the bit extracting problem were studied in [BRref].

1.3 Terminlogy

Definition 1: Let Z be a random variable assuming values in the set of m-bit strings. Z is said to be *unbiased* if it is uniformly distributed on $\{0,1\}^m$ (i.e. if for every $\alpha \in \{0,1\}^m$ $Pr(Z = \alpha) = 2^{-m}$).

Definition 2: Let $f : \{0,1\}^n \mapsto \{0,1\}^m$ be a function and $\{x_1, x_2, \ldots, x_n\}$ be a set of random variables assuming values in $\{0,1\}$. The function f is said to be *unbiased with respect to* $T \subset \{1,2,...,n\}$ if the random variable $f(x_1 x_2 \cdots x_n)$ is unbiased, when $\{x_i : i \notin T\}$ is a set of independent unbiased random variables and $\{x_i : i \in T\}$ is a set of constants. A function $f : \{0,1\}^n \mapsto \{0,1\}^m$ is said to be *t-resilient* if for every

$T \subset \{1, 2, ..., n\}$ of cardinality t, the function f is unbiased with respect to T.

The Bit Extraction Problem: Let n and t be integers. What is the maximum m such that there exist a t-resilient function $f : \{0,1\}^n \mapsto \{0,1\}^m$. We denote this number by $Bit(n,t)$.

1.4 Organization (Summary of the results)

In Section 2, we present an explicit t-resilient function from n-bit strings to $(n - t \cdot \log_2 n)$-bit strings, implying that $Bit(n,t) \geq n - t \cdot \log_2 n$. This is done by reducing the bit extraction problem to a related problem defined with respect to blocks of bits. The construction is conceptionally simple and is suitable for applications.

In Section 3, better lower bounds on $Bit(n,t)$ are derived using a relation we establish between the *linear* extraction problem and the theory of *linear* error correcting codes. Of special interest is the XOR-Lemma, stating that a necessary and sufficient condition for a set of random bits to be independent and unbiased is that each non-empty exclusive-or of these bits is unbiased.

In Section 4, we demostrate a general upper bound on $Bit(n,t)$ implying that the construction of Section 2 (as well as the lower bounds of Section 3) is reasonably good. Of special interest is the Uniform Projection Lemma, which provides a lower bound on any set of strings which has a uniform projection on every t coordinates.

In section 5, we show that 2 bits can be extracted if and only if $t \leq \lfloor 2n/3 \rfloor - 1$. In section 6, we consider linear schemes for $t > n/2$. In section 7, we consider the case where the function is symmetric.

In Section 8, we demonstrate a bound on techniques (à la Luby [Lref]) for converting efficient randomized algorithms based on k-wise independent choices, to efficient deterministic algorithms.

2. A Simple t-Resilient Function

We reduce the bit extraction problem to the block extraction problem, defined below. The block extraction problem is identical to the bit extraction problem except that the variables assume bit-strings values, instead of assuming bit values.

Definition 3: Let $f : \{0,1\}^{n \cdot k} \mapsto \{0,1\}^{m \cdot k}$ be a function, and $\{y_1, y_2, \ldots, y_n\}$ be a set of random variables assuming values in $\{0,1\}^k$. The function f is said to be *k-unbiased with respect to* $T \subset \{1, 2, ..., n\}$ if the random variable $f(y_1 y_2 \cdots y_n)$ is unbiased, when $\{y_i : i \notin T\}$ is a set of independent unbiased random variables and $\{y_i : i \in T\}$ is a set of constants. (Note that the y_i's are variables assuming values in $\{0,1\}^k$.) A function $f : \{0,1\}^{n \cdot k} \mapsto \{0,1\}^{m \cdot k}$ is said to be (t,k)-*resilient* if for every $T \subset \{1, 2, ..., n\}$ of cardinality t, the function f is k-unbiased with respect to T.

The Block Extraction Problem: Let k, n and t be integers. What is the maximum m such that there exist a (t,k)-resilient function $f : \{0,1\}^{n \cdot k} \mapsto \{0,1\}^{m \cdot k}$. Let us denote the answer by $Block_k(n,t)$.

Note that $k \cdot Block_k(n,t)$ is the number of *bits* which can be extracted by a (t,k)-resilient function. Evidently,

Lemma 1: Let k be an integer. Then
1)$k \cdot Block_k(n,t) \geq Bit(n \cdot k, t \cdot k)$.
2)$Bit(n \cdot k, t) \geq k \cdot Block_k(n,t)$.

The block extraction problem has a direct application to fault-tolerance (see section 1.1). We now show that it has an optimal solution, when $n < 2^k$. Namely, $Block_k(n,t) = n - t$ $(n \leq 2^k - 1)$.

Construction: Consider the field $GF(2^k)$ and the arithmetic in it. Suppose that $n \leq 2^k - 1$ and let a_1 a_2 $\cdots$ a_n be n *distinct* nonzero elements in this field. Define

$$r_i(y_1, y_2, ..., y_n) = \sum_{j=1}^{n} a_j^i \cdot y_j \ , \text{ for } 1 \leq i \leq n - t.$$

Lemma 2: Fixing any t of the y_i's but allowing the rest to be independent random variables (with uniform probability distribution over $GF(2^k)$), the r_i's are independent unbiased random variables.

sketch of proof: Consider the equations $r_i(y_1, y_2, ..., y_n) = \sum_{j=1}^{n} a_j^i \cdot y_j$, $1 \leq i \leq n - t$. Evaluate the terms which correspond to variables fixed by the adversary and move these values to the left hand side of the equations. The right hand side of the equations is a linear system with $n - t$ variables and $n - t$ rows. Note that the resulting matrix is the transpose of the Vandermonde matrix, which is non-singular. Therefore the system has a unique solution for every distinct value of its left hand side column. The Lemma follows. **QED**

Combining Lemma 2 and an elementary counting argument (to get the upper bound), we get

Theorem 1: Let $n \leq 2^k - 1$. Then $Block_k(n,t) = n - t$.

Returning to the Bit Extraction Problem, we combine Lemma 1 and Theorem 1 to get

Corollary 1: $Bit(n,t) > n - (t+1) \cdot \log_2 n$.

3. Linear Extraction Scheme and Linear Error Correcting Codes

In this section we reduce the problem of extracting independent unbiased bits through a *linear* extraction scheme to the well studied problem of *linear* error correction codes. A similar reduction was proven independently by Brassard and Robert [BRref] and by Odlyzko [Oref].

3.1 Preliminaries

Convention: By a *random bit* we mean a random variable with arbitrary probability distribution which assumes values 0 or 1. Throughout the rest of the paper $x = x_1 x_2 \cdots x_n$ will denote the concatenation of the random bits $x_1, x_2, \ldots, x_n$ and $a = a_1 a_2 \cdots a_n$ will denote the concatenation of the bit values $a_1, a_2, \ldots, a_n$. We take the liberty of associating n-bit strings with vectors in $GF(2^n)$, in the obvious manner.

We say that a set of m random bits $\{x_i\}_{i=1}^m$ is unbiased and independently distributed, when for every $a \in \{0,1\}^m$, $Pr(x = a) = \prod_{i=1}^m Pr(x_i = a_i)$ and $Pr(x_i = a_i) = \frac{1}{2}$. An equivalent condition is proven below.

XOR-Lemma: A set $\{x_i\}_{i=1}^m$ of random bits is unbiased and independently distributed iff the exclusive or of any non-empty subset of the bits is unbiased.

The *only if* direction is trivial. The other direction is proved by using the following two lemmas.

Lemma 3: A set $\{x_i\}_{i=1}^m$ of m random bits is unbiased and indepently distributed if and only if

$$E(f(x)) = 2^{-m} \sum_{\alpha \in \{0,1\}^m} f(\alpha)$$

for all $f : \{0,1\}^m \mapsto \mathbf{R}$.

Proof: The *only if* direction is trivial. For the *if* direction assume that there is an $a \in \{0,1\}^m$ such that $Pr(x = a) \neq 2^{-m}$. Then take as f the singleton function which is 1 at a and 0 elsewhere. This f violates the condition. **QED**

Given a subset $\{x_i : i \in S\}$ of the varibles we have a natural function $\psi_S : \{0,1\}^m \mapsto \{0,1\}$ which is the exclusive-or of these variables (i.e. $\psi_S(x) = \oplus_{i \in S} x_i$). Redifine this function slightly by making it into $\{-1,1\}$ by replacing 0 by 1, and 1 by -1. If S is the empty set define ψ_S to be identically 1.

Lemma 4: Let f is an arbitrary function from $\{0,1\}^m$ to $\mathbf{R}$. Then there are uniqly determined $c_S \in \mathbf{R}$ such that $f = \sum_{S \subset \{1,2,\ldots,m\}} c_S \psi_S$.

sketch of proof: Identify the given function space with $\mathbf{R}^{2^m}$, by letting the i-th coordinate correspond to $f(i)$. One may readily verify that the ψ_S's are 2^m mutually orthogonal vectors and hence they span the space. **QED**

Proof of the *if* direction of the XOR-Lemma

By Lemma 3, it suffices to show $E(f(x)) = 2^{-m} \sum_{\alpha \in \{0,1\}^m} f(\alpha)$ for all $f : \{0,1\}^m \mapsto \mathbf{R}$. By the additivity of the expectation operator and Lemma 4, it suffices to show this for all ψ_S. By our hypothesis, $\psi_S(x)$ is unbiased for every nonempty S and therefore $E(\psi_S(x)) = 0 = 2^{-m} \sum_{\alpha \in \{0,1\}^m} \psi_S(\alpha)$. Also note that $E(\psi_\emptyset(x)) = 1 = 2^{-m} \sum_{\alpha \in \{0,1\}^m} \psi_\emptyset(\alpha)$. The XOR-Lemma follows. **QED**

3.2 The Reduction

Let us recall the basic definitions of linear codes that we need. Further details can be found in [McWSref, ch. 1].

Definition: Let $V \subset \{0,1\}^n$ be a linear subspace of $GF(2)^n$ with cardinality 2^m. Then V is *a linear code with information words of length m and code words of length n*. The *distance* of V is the minimum Hamming distance of two vectors in V. The m-by-n matrix M is a *generator matrix* of V if the rows of M form a basis of V.

Discussion: The information word $a \in \{0,1\}^m$ is encoded by the code word $aM \in V$. The distance of the code equals the minimum Hamming weight of V's nonzero vectors. (A code of distance $t+1$ can correct $\lfloor t/2 \rfloor$ errors.)

Theorem 2: Consider arithmetic in $GF(2)$ and let M be an m-by-n zero-one matrix. M is a generator matrix of a linear error corrcting code with distance $t+1$ if and only if $f(x) = Mx^T$ is t-resilient.

sketch of proof: First, we prove that if the code has distance $t+1$ then the function is t-resilient.

By the virtue ot the XOR-Lemma we only need to check that the exclusive-ors are unbiased. An exclusive-or of some of the bits of $f(x)$ corresponds to the bit aMx^T for an appropriate nonzero vector a. Note that $b = aM$ is the codeword corresponding to the information vector a, and hence has at least $t+1$ *one*'s. Then at least one of the bits in the sum $bx^T = \sum_{i=1}^{n} b_i x_i$ is truely random and the result is unbiased.

For the converse, suppose that the code has distance at most t. That is, there exist an a such that aM has at most t ones. Then the adversary can bias the corresponding exclusive-or. **QED**

3.3 Implications

Theorem 2 imposes both upper and lower bounds on the number of extractible bits in the case of linear schemes.

Corollary 2: $Bit(n,t) \geq n - \log \sum_{i=0}^{t-1} \binom{n-1}{i}$.

This follows by combining Theorem 2 with the Gilbert-Varshamov bound for linear codes [McWSref, ch. 1, p. 34]. This is an existential result. Explicit constructions, which almost achieve this value, are known for $t = \Omega(n)$. In fact, the explicit construcion of section 2 is analogous to the well known Reed-Solomon codes [McWSref, ch. 10].

Corollary 3: Linear t-resilient functions cannot extract more than $n - \log \sum_{i=0}^{\lfloor t/2 \rfloor} \binom{n}{i}$ bits.

This follows by combining Theorem 2 with the Hamming Bound [McWSref, ch. 1, p. 19]. In the next section, we will show that a similar upper bound holds also for general t-resilient functions.

4. An Upper Bound on the Number of Extractible Bits by a General Scheme

In this section we demonstrate an upper bound on the number of independent unbiased bits extractable by a *general* scheme.

4.1 Preliminaries

Definition: Let $S \subset \{0,1\}^n$ be a set of strings and $I = (i_1, i_2, ..., i_t)$ be a monotonely increasing sequence of t integers from $\{1, 2, ..., n\}$. For $a \in \{0,1\}^t$, we denote

$$S_{I,a} = \{x_1 x_2 \cdots x_n \in S : x_{i_j} = a_j, 1 \leq j \leq t\} \ .$$

The set $S \subset \{0,1\}^n$ has a uniform projection onto the i_1-st, i_2-nd,..., i_t-th coordinates if for every $a \in \{0,1\}^t$, $|S_{I,a}| = \frac{|S|}{2^t}$.

Let us show first show that sets having this property for every t coordinates, must be of large cardinality.

The Uniform Projection Lemma :

If $S \subset \{0,1\}^n$ has uniform projection on any t coordinates then $|S| \geq \sum_{i=0}^{\lfloor \frac{t}{2} \rfloor} \binom{n}{i}$.

sketch of proof: Let $k = |S|$. For convienience change all 1 to -1 and 0 to 1. Now taking the exclusive-or of two vectors corresponds to coordinatewise multiplication. Let H be the $k \times n$ matrix with the elements of S as rows.

Consider j arbitrary columns of H, when $j \leq t$. Let H' be the matrix consisting of the corresponding columns of H. Since the rows of H have uniform projection onto these coordinates, all possible j-tuples appear as rows of H' with the same frequency. Thus, exactly half of the rows of H' have an even number of -1. It follows that the exclusive-or of the columnvectors of H' has as many 1's as -1's.

Let V be the set of vectors which result by taking the exclusive-or of i distinct columnvectors of H ($i \leq \lfloor \frac{t}{2} \rfloor$). The vectors in V are distinct and mutually orthogonal when considered as real vectors (since by the above paragraph the coordinatewise multiplication of any pair of distinct vectors in V has as many 1's as -1's). Therefore, V spans a subset of $\mathbf{R}^k$, and $|V| \leq k$ follows. Noting that $|V| = \sum_{i=0}^{\lfloor \frac{t}{2} \rfloor} \binom{n}{i}$, the Lemma follows. **QED**

Observe that one can do slightly better when t is odd by considering also all columnvectors which are xor's of $\frac{t-1}{2}$ arbitrary vectors and the first columnvector.

In coding theory, the matrix H is called an orthogonal array of strength t. It is likely that the above Lemma has already been proven.

4.2 The Upper bound

Theorem 3: $Bit(n,t) \leq n - \log \sum_{i=0}^{\lfloor \frac{t}{2} \rfloor} \binom{n}{i}$.

sketch of proof: Let $f : \{0,1\}^n \mapsto \{0,1\}^m$ be a t-resilient function. One can easily verify that $f^{-1}(0,0\ldots,0)$ is a set which has a uniform projection onto any t coordinates. Applying the Uniform Projection Lemma, we get $\left|f^{-1}(0,0,\ldots,0)\right| \geq \sum_{i=0}^{\lfloor \frac{t}{2} \rfloor} \binom{n}{i}$. On the other hand $\left|f^{-1}(0,0\ldots,0)\right| = 2^{n-m}$, and the theorem follows. **QED**

The proof of Theorem 3 makes use of the fact that a t-resilient function $f : \{0,1\}^n \mapsto \{0,1\}^m$ yields an orthogonal arrays of strength t. In fact, f yields 2^m such arrays whose rows fill the entire n-dimentional space. Thus, such a function is a much more complicated object than an orthogonal array.

By Theorem 2, this bound can be reached if perfect linear codes, with n-bit code words and distance $t+1$, do exist. Perfect codes are quite rare and hence we do not know whether the optimal scheme is linear in the general case.

5. Tight Bounds for Extracting Two Bits

5.1 Preliminaries

Recall that by Lemma 4 (section 3), any Boolean function $f(x)$ can be written as a sum of the exclusive-or functions (that is the functions $\psi_S(x) = \oplus_{i \in S} x_i$ for $S \subset \{1, 2, \ldots, n\}$). Furthermore, it was implicitly stated that expressing f as a sum of the $\psi_S(x)$'s can be done in a unique way. We now show that when testing the resiliency of a function it suffices to test the resiliency of the $\psi_S(x)$'s with nonzero coefficients in this expression. Clearly, a $\psi_S(x)$ is t-resilient if and only if $|S| > t$. This proves the *if* direction of the following proposition.

Proposition: Let $f : \{0,1\}^n \mapsto \{0,1\}$ be a non-trivial Boolean function, and let $f(x) = \sum_S c_S \psi_S(x)$. The function f is t-resilient if and only if there is no $S \subset \{1, 2, \ldots, n\}$ such that both $c_S \neq 0$ and $|S| \leq t$.

Proof: For the *only if* direction, let S_0 denote a set S of minimum cardinality for which $c_S \neq 0$, and $n_0 = |S_0|$. Assume, on the contrary, that $n_0 \leq t$. Now, suppose that the adversary fixes the value 1 for all the variables in $\{x_i : i \in S_0\}$ (and lets the rest be independent unbiased bits). Let A_0 denote the set of all possible outcomes for the n-bit string when the adversary acts so; and let x be a random variable with uniform probability distribution in A_0. Equivalently, $Pr(x_i = 1) = 1$ if $i \in S_0$ and $Pr(x_i = 1) = \frac{1}{2}$ if $i \notin S_0$. Let $P(n)$ denote the power set of $\{1, 2, \ldots n\}$.

$$
\begin{aligned}
E(f(x)) \otimes &= \sum_{\alpha \in A_0} 2^{-(n-n_0)} \cdot f(\alpha) \\
\otimes &= \sum_{\alpha \in A_0} 2^{-(n-n_0)} \cdot \sum_{S \in P(n)} c_S \psi_S(\alpha) \\
\otimes &= \sum_{S \in P(n) - \{S_0\}} 2^{-(n-n_0)} \cdot \sum_{\alpha \in A_0} c_S \psi_S(\alpha) \\
\otimes &\quad + 2^{-(n-n_0)} \cdot \sum_{\alpha \in A_0} c_{S_0} \psi_{S_0}(\alpha) \\
\otimes &= \sum_{S \in P(n) - \{S_0\}} c_S \cdot E(\psi_S(x)) \\
\otimes &\quad + c_{S_0} \cdot E(\psi_{S_0}(x)) \\
\otimes &= 0 + c_{S_0} \\
\otimes &\neq 0 .
\end{aligned}
$$

Thus there is a way to fix at most t variables which makes b biased. **QED**

5.2 The Bounds

Lemma 5: Let $n = 3l$ be a multiple of 3. Let $b_1(x) = \oplus_{i=1}^{2l} x_i$, $b_2(x) = \oplus_{i=l+1}^{3l} x_i$ and $f(x) = b_1(x) b_2(x)$. Then f is $(2l-1)$-resilient.

Proof Note that b_1 and b_2 satisfy $b_1 \oplus b_2 = \left(\oplus_{i=1}^{l} x_i\right) \oplus \left(\oplus_{i=2l+1}^{3l} x_i\right)$. So if the adversary is allowed to fix at most $2l-1$ of the n bits, both b_1, b_2 and their exclusive-or are unbiased. By the XOR-lemma (see section 3), b_1 and b_2 are two independent random bits.**QED**

Similarly we get

Lemma 6 Let $n = 3l + 2$, $b_1(x) = \oplus_{i=1}^{2l+1} x_i$, $b_2(x) = \oplus_{i=l+2}^{3l+2} x_i$. Then $b_1(x)b_2(x)$ is $2l$-resilient.

On the other hand

Lemma 7: Let $\mu \in \{0,1\}$. Then, there exists no $2l$-resilient function $f : \{0,1\}^{3l+\mu} \mapsto \{0,1\}^2$.

Proof: Assume, on the contrary, that f is $2l$-resilient, and interpret f as a function from $\{1,-1\}^{3l}$ to $\{1,-1\}^2$. Let $b_1(x)$ denote the first bit of $f(x)$, and $b_2(x)$ denote the second bit of $f(x)$. By the XOR-Lemma (section 3), both b_1 and b_2 as well as $b_1 \oplus b_2$ must be $2l$-resilient. Thus using the Proposition, for these three Boolean functions the ψ_S's corresponding to nonzero coefficients must have $|S| > 2l$. We now show that this condition cannot be met. Let

$$\otimes b_1(x) = \sum_{S \subset \{1,2,\ldots,n\}} c_S \psi_S(x) \text{ and}$$

$$\otimes b_2(x) = \sum_{T \subset \{1,2,\ldots,n\}} d_T \psi_T(x) \ .$$

Then $b_1(x) \oplus b_2(x)$ corresponds to

$$b_1(x) \cdot b_2(x) \otimes = \sum_{S,T \subset \{1,2,\ldots,n\}} c_S d_T \psi_S(x) \psi_T(x)$$

$$\otimes$$

$$\otimes = \sum_{S,T \subset \{1,2,\ldots,n\}} c_S d_T \psi_{S \Delta T}(x)$$

(where $S \Delta T = S \cup T - S \cap T$ is the symmetric difference). Recall that $|S|, |T| \geq 2l + 1$ for all S, T where $c_S \cdot d_T \neq 0$. Thus, all non-zero coefficients of $b_1 \cdot b_2$ correspond to subsets of cardinality $\leq 2(3l + \mu - (2l+1)) \leq 2(l + \mu - 1) \leq 2l$. **QED**

Similarly,

Lemma 8: Let $n = 3l + 2$. Then, there exists no $(2l+1)$-resilient function $f : \{0,1\}^n \mapsto \{0,1\}^2$.

Combining the above four Lemmas, we get the following result conjectured by Vazirani [Vref].

Theorem 4: There exist a t-resilient function $f : \{0,1\}^n \mapsto \{0,1\}^2$ if and only if $t < \lfloor 2n/3 \rfloor$.

6. On Extracting Few Bits when $t > n/2$

In this section we show that k independent unbiased bits can be extracted if the adversary can determine less than $2^{k-1} \cdot \lfloor \frac{n}{2^k - 1} \rfloor$ of the original $n \geq 2^k - 1$ bits. We also show that this is close to the best possible performance as far as linear extraction schemes are concerned.

6.1 Possibility Result

Theorem 5: Let $k \leq \lfloor \log_2 n \rfloor$. Then there exist a $(\lfloor \frac{n}{2^k - 1} \rfloor \cdot 2^{k-1} - 1)$-resilient scheme extracting k bits out of n.

sketch of proof: Assume that $\lfloor \frac{n}{2^k-1} \rfloor = 1$. For $1 \leq i \leq k$, let $J_i \subset \{1, 2, \ldots, 2^k - 1\}$ be the subset of integers j such that the i-th least significant bit in the binary expansion of j equals 1. Let $b_i(x_1 x_2 \cdots x_n) = \oplus_{j \in J_i} x_j$. Let $f(x) = b_1(x) b_2(x) \cdots b_k(x)$. We will show that the function $f : \{0,1\}^n \mapsto \{0,1\}^k$ is $(2^{k-1} - 1)$-resilient.

Note that each of the b_i, as well as each exclusive or of any non-empty subset of the b_i's, is a random variable depending on 2^{k-1} of the x_i's. (In particular, consider the set S and the random variable $r_S(x) = \oplus_{i \in S} b_i(x)$. Then $r_S(x) = \oplus_{j \in J_S} x_j$, where J_S is the bit-by-bit exclusive or of the k-bit strings which correspond to the binary expansion of the integers in S. Note that $|J_S| = 2^{k-1}$.) For general n, make $\lfloor \frac{n}{2^k-1} \rfloor$ copies of the above construction.**QED**

6.2 Impossibility Result

Theorem 6: Let $k \leq \lfloor \log_2 n \rfloor$. Then there exist no *linear* $(\frac{2^{k-1}}{2^k-1} \cdot n)$-resilient extraction scheme for extracting k bits.

sketch of proof: Suppose that $f : \{0,1\}^n \mapsto \{0,1\}^k$ is a linear t-resilient function. Note the correspondence between linear extraction schemes and schemes in which each extracted bit is the exclusive or of some subset of the original bits. Consider an $2^k - 1$ by n matrix M in which each row correspond to an exclusive or of a non-empty subset of the bits of $f(x)$. By the fact f is t-resilient, each row must have at least $t+1$ non-zero entries. On the other hand, each column contains exactly 2^{k-1} ones if it corresponds to a variable which appears in some extracted bit, and contains no ones otherwise. Therefore, we have $n \cdot 2^{k-1} \geq (2^k - 1) \cdot (t+1)$ and $t < \frac{2^{k-1}}{2^k-1} \cdot n$. The Theorem follows.**QED**

An alterrnative proof of Theorem 6 can be derived by combining Plotkin Bound [McWSref, ch. 2, pp. 41-42] and our Theorem 2.

We conclude by suggesting the following

Conjecture: Let $k \leq \lfloor \log_2 n \rfloor$. Then there exist no *general* $(\frac{2^{k-1}}{2^k-1} \cdot n)$-resilient extraction scheme for extracting k bits.

7. On Symmetric Predicates

A Boolean predicate $f : \{0,1\}^n \mapsto \{0,1\}$ is called *symmetric* if for every permutation $\pi : \{1, 2, \ldots, n\} \mapsto \{1, 2, \ldots, n\}$,

$$f(x_1, x_2, \ldots, x_n) = f(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}) \ .$$

Let $w(x)$ denote the Hamming weight of x. Then for every symmetric predicate f there exists an $S \subset \{1, 2, \ldots, n\}$ such that

$$f(x) = \begin{cases} \otimes 1 & \text{if } w(x) \in S \\ \otimes 0 & \text{otherwise} \end{cases}$$

Thus, an unbiased symmetric predicate on n Boolean variables correspond to an equal partition of the n-th row in Pascal's triangle (i.e. the set S corresponding to the predicate satisfies $\sum_{i \in S} \binom{n}{i} = \sum_{i \notin S} \binom{n}{i}$). Fixing a variable in a symmetric predicate, corresponds to sliding the partition up one row to the right or left.

We have obtained the following results:

1)The exclusive-or of all n variables and its negation, are the only $2n/3$-resilient symmetric predicates.

2)For sufficiently large n, the exclusive-or of all n variables and its negation, are the only $7n/100$-resilient symmetric predicates.

An interesting open problem is to prove or disprove the following Conjecture: *The exclusive-or of all n variables and its negation, are the only 1-resilient symmetric predicates.*

8. On k-wise Independence

In [Lref], Luby demonstrates how to convert a randomized algorithm that uses pairwise independent choices into a parallel deterministic algorithm of the same depth. In this section, we consider generalizations of his technique to the case of k-wise independent choices, and show that polynomiality can be maintained only if k is a constant.

Convention: Let A be a set. We write $a \in_R A$ to abbreviate "a is picked at random with uniform probability distribution in A".

Suppose that in the original *polynomial-time* algorithm, elements are picked randomly with uniform distribution in a set E, and that the correctness of the algorithm is only based on the fact that these choices are *pairwise independent*. Assume that $|E|$ is polynomial in the size of the input n. By change of parameters, we can assume that the algorithm makes at most n random choices at each round. Luby's (efficient) transformation is based on the construction of a set of sequences S which combines the following properties.

0)$s \in S$ is a n-long sequence of elements in E.

1)A sequence $s \in_R S$ defines a sequence of *pairwise independent* random variables each uniformly distributed in E.

2)The set S has a polynomially bounded cardinality.

Once such a set S is constructed, one may substitute the pairwise independent random choices in the algorithm by the elements of a sequence $s \in_R S$. Furthermore, instead of picking randomly $s \in_R S$ one can exhaust all possible $s \in S$, and run them all in parallel.

In [ACGSref], a simple construction that satisfies the above conditions was presented, and used in a different context. This construction easily extends to allow the n elements be k-wise independent Let $|E| = p$ be a prime power, and let $a_1, a_2, \ldots, a_n$ be n distinct non-zero elements in the field $GF(p)$. Consider the sequence $s_i(x) = \sum_{j=1}^{k} a_i^j x_j \mathrm{mod} p$ $(1 \leq i \leq n)$. If the x_i's are independent random variables (and each $x_i \in_R E$), then the $s_i(x)$'s are k-wise independent variables each uniformly distributed in E. Finally note that the set $S = \{(s_1(\alpha), s_2(\alpha), \ldots, s_n(\alpha)) : \alpha \in GF(p)^k\}$ can be deterministically constructed in $p^k \cdot n \cdot k$ $GF(p)$-operations. When k is a fixed constant, this construction is polynomial in $|E|$ and n. Similar constructions of k-wise independent elements were used in [Lref, Aref, AWref, KUWref].

Let $|E| = p$ be a prime power, and let $a_1, a_2, \ldots, a_n$ be n distinct non-zero elements in the field $GF(p)$. Consider the sequence $s_i(x) = \sum_{j=1}^{k} a_i^j x_j \mathrm{mod} p$ $(1 \leq i \leq n)$. If the x_i's are independent random variables (and each $x_i \in_R E$), then the $s_i(x)$'s are k-wise independent variables each uniformly distributed in E. Finally note that the set $S = \{(s_1(\alpha), s_2(\alpha), \ldots, s_n(\alpha)) : \alpha \in GF(p)^k\}$ can be deterministically constructed in $p^k \cdot n \cdot k$ $GF(p)$-operations. .

A natural question is whether such techniques can be extended, while maintaining polynomiality in $|E|$ and n, to k's which are not fixed. More generaly, how large should a set $S \subset E^n$ be so that the elements of a sequence $s \in_R S$, are k-wise independent random variables with uniform probability distributed in E.

Using the Uniform Projection Lemma of Section 4, one can verify that such S must satisfy

$$|S| \geq \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \binom{n}{i} \approx n^{\frac{k}{2}} .$$

Thus, a deterministic simulation of a $k(n)$-wise independent n-bit sequence cannot be done in $poly(n)$-time, when $\lim_{n\to\infty} k(n) = \infty$.

9. Open Problem

Prove or disprove the following claim: *for every n and t there exist a linear t-resilient function from $\{0,1\}^n$ to $\{0,1\}^{Bit(n,t)}$.*

Acknowledgments

We would like to thank Richard Anderson, Richard Karp and Ernst Mayr for a discussion which led to the application discussed in Section 8. We also wish to thank Baruch Awerbuch, Peter Elias, Shimon Even, Abraham Lempel and Andrew Odlyzko for their valuable comments.

References

[AWref]Ajtai, M., and A. Wigderson, "Deterministic Simulation of Probabilistic Constant Depth Circuits", these proceedings.

[ACGSref]Alexi, W., B. Chor, O. Goldreich, and C.P. Schnorr, "RSA/Rabin Bits are $\frac{1}{2} + \frac{1}{poly(\log n)}$ Secure", *Proc. 25th FOCS*, Oct. 1984, pp. 449-457.

[Aref]Anderson, R., "Set Splitting", manuscript, (1985).

[ACGMref]Awerbuch, B., B. Chor, S. Goldwasser, and S. Micali, "How to Implement Verifiable Secret Sharing and Simultaneous Networks", these proceedings.

[BRref]Brassard, G., and JM Robert, "How to Reduce your Enemy's Information", to appear in the proceedings of *Crypto85*.

[KUWref]Karp, R.M., E. Upfal and A. Wigderson, "The Complexity of Parallel Computation on Matroids", these proceedings.

[Lref]Luby, M., "A Simple Parallel Algorithm for the Maximal Independent Set Problem", *Proc. 17th STOC*, May 1985, pp. 1-10.

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

[Oref]Odlyzko, A.M., private communication (1985).

[Vref]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. 366-377.