# 4 Physical Random-Number Generators

## 4.1 Generalities

The doctrine in cryptology is that the algorithm of encryption is known to the adversary (Eve) and that the only thing that is kept secret is the key, which normally is a bitsequence or a sequence of natural numbers or elements of a finite ring (e.g. a residue ring or a finite field). Mostly, such key sequences are produced by an algorithmic generator (i.e., they are so-called pseudo-random numbers), since these offer the following benefits: the sequence of numbers can be reproduced for debugging and testing; no special hardware is necessary; a large quantity of random numbers can be produced in a short time. In Chapter 7, we will provide several tests of "randomness" of such pseudo-random sequences. However, there is no practically implementable "universal" test of randomness: every test procedure just measures a certain aspect of "non-regularity". If one wants to have genuine random numbers, then they have to be produced by a physical device. A very drastic drawback of classical pseudo-random generators has been pointed out in the paper entitled "Random numbers fall mainly in the planes" by Marsaglia (1968). Possible physical random sources are electronic noise produced by a semiconducting diode (Richter (1993)) or the impulses of a Geiger counter in connection with a radioactive source (Inoue et al. (1983)). In the latter paper, the authors propose a hardware implementation of this device, the radioactive source consisting of a PG-508 pulse generator. Another device using a Geiger counter has been described in Nisley (1990), the RM-60 Micro Roentgen Radiation Monitor from Aware Electronics. Finally, there is HOT BITS (see Walker (1996)), a source of random bits available via the Internet, which uses beta radiation from the decay of Krypton-85.

The output of such a generator (which in the latter case leads directly to a Poisson (for the number of events) resp. exponential (for the inter-occurence waiting times) distribution) has to be processed further in order to obtain standard uniform random numbers (digits, or reals in $[0, 1]$). Since the parameters of the distribution of the data is not known exactly, only a small amount of this information is used (usually the last digit), to be on the safe side, and so the yield of this method is relatively small. However, physically generated random numbers are expensive and can not be produced in too

D. Neuenschwander: Prob. and Stat. Methods in Cryptology, LNCS 3028, pp. 47-55, 2004.

high quantities. For example, the HOT BITS hardware produces only about 240 bits per second.

Modifying an idea of von Neumann (1963), used to extract unbiased bits from a sequence of biased ones by comparison of two subsequent bits, we propose to obtain random numbers in $[0, 1]$ from a sequence $X_0, X_1, X_2, \ldots$ of independent exponentially distributed data by using $U_n := \frac{X_{2n}}{X_{2n}+X_{2n+1}}$. This gives us one real number for every two data values instead of only two bits, considerably increasing the output. If the distribution of the $X_n$ is exponential, the $U_n$ are uniform in $[0, 1]$. The question of the "rate of disappearing" of the bias (so-called extraction rate) is addressed in Section 4.3, in particular for rational biases $b$. It turns out that the size of $b$ does not influence the extraction rate, but that the latter is solely determined by the arithmetic properties of $b$. On the other hand, the extraction rate can be shown to be 0 in Lebesgue-almost all cases.

In the practical implementation of this method we have to take into account that the exponential times $X_n$ can only be measured up to a certain precision.

## 4.2 Construction of Uniformly Distributed Random Numbers from a Poisson Process

In this section, we will consider the output of a Geiger counter as source of randomness. The other examples mentioned in the previous section are of a similar nature. If the number of impulses during a fixed amount of time is counted, a variable with a Poisson distribution is the raw material that has to be processed further in order to obtain unbiased random bits. Usually, the length $t_0$ of the time interval is chosen large with respect to the mean time $1/\theta$ between two impulses; then the number $N$ of counts during this interval has a Poisson distribution $\pi_\lambda$ with $\lambda = t_0\theta$. In most cases, the last digit $X$ in the binary representation of $N$ is used as an approximation for a uniformly distributed random bit.

Another method (see Inoue et al. (1983)) makes use of the random waiting time $T$ between two consecutive impulses, which obeys an exponential distribution $\varepsilon_\theta$. Clearly, if the intensity $\theta > 0$ were known exactly, one could obtain a uniformly distributed random variable $U$ just by the usual transform method $U := \exp(-\theta T)$. But $\theta$ not being known exactly enough to guarantee that $U$ is "sufficiently" uniform, it has to be estimated. One can use two consecutive waiting times produced by the Geiger counter, one so to say to estimate $\theta$ and the other one to obtain a uniform random variable.

The following lemma is easy:

**Lemma 4.1.** *Let $X$ and $Y$ be independent random variables with common exponential distribution $\varepsilon_\theta$ (where $\theta > 0$). Then*

$$U := \frac{X}{X+Y}$$

*obeys a uniform law on* $[0,1]$.

Therefore if the raw material is a stream of independent exponential random times $X_n$ $(n \in I\!N)$, a sequence of independent uniform variables can be obtained by setting $U_n := X_{2n}/(X_{2n} + X_{2n+1})$.
Unfortunately the waiting time between two impulses of the Geiger counter is not measured as a real number, but only in multiples of the length $\mathit{\Delta}$ of the clock cycle (w.l.o.g. we may assume $\mathit{\Delta} = 1$). If two impulses occur during one clock cycle, then they are counted as one. Hence the $n$-th observation of an impulse occurs at the time $S'_n$ defined recursively by $S'_0 := 0$,

$$S'_n := \min\{k \in I\!N : N_k \geq N_{S'_{n-1}} + 1\},$$

where the Poisson process $\{N_t\}_{t\geq 0}$[1] indicates the number of impulses up to time $t$. Instead of the sequence $\{X_n\}_{n\geq 1}$ of exactly exponentially distributed waiting times between two impulses, we can only observe the sequence $\{X'_n\}_{n\geq 1}$, where $X'_n := S'_n - S'_{n-1}$.

**Proposition 4.1.** *The* $X'_1, X'_2, \ldots$ *are i.i.d. such that* $X'_n - 1$ *obeys a geometric distribution with parameter* $1 - \exp(-\theta)$.

**Proof:** Let $\{\mathcal{F}_t\}_{t\geq 0}$ denote the canonical filtration of the Poisson process $\{N_t\}_{t\geq 0}$. Then $S'_1 < S'_2 < \ldots$ is a sequence of stopping times. Assume $n \geq 2$. Since the Poisson process is stationary with independent increments, the process $\{N'_t\}_{t\geq 0}$ with $N'_t := N_{t+S'_{n-1}} - N_{S'_{n-1}}$ is again a Poisson process with parameter $\theta$. Therefore the distribution of

$$X'_n = S'_n - S'_{n-1} = \min\{k \in I\!N : N'_k \geq 1\}$$

is the same as that of $X'_1$. The latter law can easily be calculated to be the geometric distribution with parameter

$$P(X'_1 - 1 = 0) = P(X_1 < 1) = 1 - \exp(-\theta).$$

[1] A stochastic process $\{N_t\}_{t\geq 0}$ is called a Poisson process with intensity $\lambda > 0$ if $N_t$ obeys a Poisson distribution with parameter $\lambda t$ $(t > 0)$. This is equivalent to the fact that for the "jump times"

$$\mathit{\Gamma}_0 = 0 < \mathit{\Gamma}_1 < \mathit{\Gamma}_2 < \ldots$$

(where $\mathit{\Gamma}_k := \inf\{t \geq 0 : N_t \geq k\}$) we have that the "inter-occurence times" $\mathit{\Gamma}_{k+1} - \mathit{\Gamma}_k$ are i.i.d exponentially distributed random variables as

$$P(\mathit{\Gamma}_{k+1} - \mathit{\Gamma}_k > x) = e^{-\lambda x}$$

for $x \geq 0$.

Furthermore, the process $\{N'_t\}_{t\geq 0}$ and hence the random variable $X'_n$ is independent of $\mathcal{F}_{S'_{n-1}}$. On the other hand, the random variables $X'_1, X'_2, \ldots, X'_{n-1}$ are $\mathcal{F}_{S'_{n-1}}$-measurable and hence independent of $X'_n$. □

Denote by $F_X$ the distribution function of a random variable $X$.

**Theorem 4.1.** *Let $X'-1, Y'-1$ be independent geometric random variables with parameter $\theta' = 1-\exp(-\theta)$ and denote by $U$ a random variable distributed uniformly on the interval $[0,1]$. Then $U' := \frac{X'-0.5}{X'+Y'-1}$ satisfies*

$$\frac{1}{2}\tanh\bigl(\frac{\theta}{2}\bigr) \leq ||F_{U'} - F_U||_\infty \leq 1 - \exp\bigl(-\frac{\theta}{2}\bigr).$$

**Proof:** The lower bound follows from the observation that $F_U$ is continuous at $\frac{1}{2}$ whereas $F_{U'}$ has a jump of size $P\{U' = \frac{1}{2}\} = \sum_n P\{X' = n\}P\{Y' = n\} = \sum_{n=0}^{\infty} \theta'^2 (1-\theta')^{2n} = \frac{\theta'}{2-\theta'} = \frac{1-\exp(-\theta)}{1+\exp(-\theta)} = \tanh\bigl(\frac{\theta}{2}\bigr)$.

We will now assume w.l.o.g. that $X'$ and $Y'$ are of the form $X' = \lfloor X \rfloor$ and $Y' = \lfloor Y \rfloor$ with $X$, $Y$ independent and with exponential distribution with parameter $\theta$. Let $U := X/(X+Y)$. Since the distribution of $U'$ is symmetric about $\frac{1}{2}$ it is easy to see that for the upper bound it is sufficient to show $|F_{U'}(t) - t| \leq 1 - \exp(-\theta/2)$ only for $t \in ]0, \frac{1}{2}]$. For such $t$ we have

$$\begin{aligned} F_{U'}(t) - t &= E\Bigl(1_{[0,t]}(U') - 1_{[0,t]}(U)\Bigr) \\ &= \sum_{m,n\in I\!N} E\bigl(\bigl[1_{[0,t]}\bigl(\frac{m+0.5}{m+n+1}\bigr) - 1_{[0,t]}(U)\bigr] \\ &\qquad\qquad\qquad\qquad \cdot 1_{\{m<X<m+1,\ n<Y<n+1\}}\Bigr) \\ &= S_+ - S_-, \end{aligned}$$

where

$$S_+ = \sum_{\frac{m+0.5}{m+n+1}\leq t,\, \frac{m+1}{m+n+1}>t} P(m < X < m+1,\ n < Y < n+1,\ \frac{X}{X+Y} > t)$$

and

$$S_- = \sum_{\frac{m+0.5}{m+n+1}>t,\, \frac{m}{m+n+1}<t} P(m < X < m+1,\ n < Y < n+1,\ \frac{X}{X+Y} \leq t).$$

The last equality follows from the fact that the random variable in the expectation takes only the values $-1, 0$, or $1$, and all summands with either $\frac{m+1}{m+n+1} \leq t$ or $\frac{m}{m+n+1} \geq t$ vanish, since the square $]m, m+1[\times]n, n+1[$ lies completely on one side of the line $\{\frac{x}{x+y} = t\}$ in these cases.

We now collect the summands in $S_\pm$ that belong to the same $m$. Let $a := (1-t)/t \geq 1$; then $\frac{x}{x+y} > t$ if and only if $ax > y$.

From this we obtain

$$S_+ = \sum_{m=0}^{\infty} P(X < m+1,\ Y > n_m,\ aX > Y)$$

and

$$S_- = \sum_{m=0}^{\infty} P(X > m,\ Y < n_m,\ aX \leq Y),$$

where $n_m$ is the smallest $n \in I\!N$ with $\frac{m+0.5}{m+n+1} \leq t$. Considering a single summand we have

$$\begin{aligned} &P(X > m,\ Y < n_m,\ aX \leq Y) \\ &\leq P(m < X < m + (1/2),\ am < Y < n_m) \\ &= c \cdot P(m < X < m+1,\ am < Y < n_m) \end{aligned}$$

with

$$c := P(m < X < m + (1/2))/P(m < X < m+1) = \big(1 + \exp\big(-\frac{\theta}{2}\big)\big)^{-1}.$$

In order to prove the latter inequality we have used the fact that the density of $P$ is a decreasing function of $x+y$ and an elementary geometric argument in the $m - n_m$-plane. A similar argument yields

$$P(X < m{+}1,\ Y > n_m,\ aX > Y) \leq c{\cdot}P(m < X < m{+}1,\ n_m < Y < a(m{+}1)).$$

Summing up, we obtain

$$\begin{aligned} S_+ + S_- &\leq \sum_{m=0}^{\infty} c \cdot P(m < X < m+1,\ am < Y < n_m) \\ &\qquad\qquad + c \cdot P(m < X < m+1,\ n_m < Y < a(m+1)) \\ &= c \cdot \sum_{m=0}^{\infty} P(m < X < m+1,\ am < Y < a(m+1)) \\ &= c \cdot \sum_{m=0}^{\infty} \exp(-\theta(a+1)m)(1-\exp(-\theta))(1-\exp(-a\theta)) \\ &= \frac{(1-\exp(-\theta))(1-\exp(-a\theta))}{(1+\exp(-\theta/2))(1-\exp(-(a+1)\theta))}. \end{aligned}$$

But since $1 - \exp(-a\theta) \leq 1 - \exp(-(a+1)\theta)$, we finally get the bound

$$|F_{U'}(t) - t| = |S_+ - S_-| \leq S_+ + S_- \leq 1 - \exp\big(-\frac{\theta}{2}\big),$$

and this proves Theorem 4.1. □

The upper and lower bounds $1 - \exp\frac{-\theta}{2} \leq \frac{\theta}{2}$ and $\frac{1}{2}\tanh\frac{\theta}{2} \approx \frac{\theta}{4}$ in Theorem 4.1 differ by a factor of approximately 2. As one can see from numerical experiments, the lower of these is the true value, but the proof of this fact is more complicated.

## 4.3 *The Extraction Rate for Biased Random Bits

In this section, it will make things a little simpler (e.g., as we will see, we can work with expectations) if we replace $\mathbb{B} = \{0,1\}$ by $\mathcal{B} := \{1,-1\}$. Since there will be no danger of misunderstanding, also the elements of $\mathcal{B}$ will be called (random) "bits".

We want to investigate the following question: Given $n$ i.i.d. random bits with common bias $b$ (i.e. $P(X_1 = 1) - P(X_1 = -1) = E(X_1) = b \in ]0,1[$ (w.l.o.g.)), how is it possible to construct from them an "as unbiased as possible" random bit? It turns out that a good method is to multiply[2] the $X_i \in \mathcal{B}$, for if

$$P_n := \prod_{i=1}^{n} X_i,$$

then the bias of $P_n$ turns out to be only $b^n$, i.e. $P(P_n = 1) - P(P_n = -1) = b^n$. One may ask if there are functions $f : \mathcal{B}^n \to \mathcal{B}$ that behave better (in the sense of bias reduction) than multiplication. Let us define, for $f$ and $b$ as defined before, the quantity

$$\xi_{f,n}(b) := |E(f(X_1, X_2, \dots, X_n))|$$

and

$$\Xi_n(b) := \min_{f:\mathbb{B}^n \to \mathbb{B}} \xi_{f,n}(b).$$

The relation $\xi_{\cdot,n}(b) = b^n$ (as mentioned before) can be interpreted as follows: For each new (independent $b$-biased) bit source $X_{n+1}$ combined with the sources $X_1, X_2, \dots, X_n$, the multiplication-function "extracts" another factor $b$ in the output bit $P_{n+1}$ (compared with $P_n$). So if we replace the multiplication-function by an (asymptotically (as $n \to \infty$)) optimal function $f$, we should have at least the extra multiplicative factor $b$ for every step $n \to n+1$ (i.e. by taking one additional bit source). Therefore, we define the so-called extraction rate of $b$ by

$$\Xi(b) := \lim_{n\to\infty} \sqrt[n]{\Xi_n(b)}.$$

The extraction rate can be interpreted as the optimal asymptotic multiplicative effect of each new input bit source on the resulting bias of the output bit. Or - in other words - it is the asymptotical (as $n \to \infty$) speed of the diminution of the bias per new random bit source, when the final output bit is produced by adding (mod.2) (in $\mathbb{B}$) $n$ independent biased random bit sources. It can be shown that for Lebesgue-almost all $b \in ]0,1[$ we have $\Xi(b) = 0$ (see Näslund, Russell (2001), Theorem 21). For rational $b$ we have the following:

[2] If we identify $\mathcal{B}$ and $\mathbb{B}$ in the natural way, then multiplication in $\mathcal{B}$ corresponds to addition mod.2 in $\mathbb{B}$.

**Theorem 4.2.** *If $b \in \mathbb{Q}$, $b = \frac{r}{s}$, $r, s \in \mathbb{N}$, $r, s$ relatively prime, then $\Xi(b) = \frac{1}{s}$.*

So interestingly enough, it is not the size of $b$, but rather its arithmetic properties that determine its extraction rate!
**Proof of Theorem 4.2:** 1. We first prove that

$$\Xi_n(b) \geq \frac{1}{s^n}. \tag{4.1}$$

Let us fix some notation. For a subset $C \subset \mathcal{B}^n$, define its weight by

$$w(C) := P((X_1, X_2, \dots, X_n) \in C),$$

and put

$$f(X_1, X_2, \dots, X_n) := 2(\mathbf{1}(C)(X_1, X_2, \dots, X_n) - \frac{1}{2}).$$

Now consider a collection (subset) $C \subset \mathcal{B}^n$ with $w(C) = \frac{1+\delta}{2}$, where $|\delta|$ is the bias of $f$. W.l.o.g. we may suppose that $(-1, -1, \dots, -1) \notin C$. Then we may calculate

$$\begin{aligned} w(C) &= \frac{1+\delta}{2} \\ &= \sum_{i=1}^{n} t_i (\frac{r}{s})^i (1 - \frac{r}{s})^{n-i} \\ &= \frac{1}{s^n} \sum_{i=1}^{n} t_i r^i (s-r)^{n-i} \end{aligned}$$

for some integers $t_i \in \{0, 1, \dots, \binom{n}{i}\}$, or - equivalently -

$$s^n (1+\delta) = 2 \sum_{i=1}^{n} t_i r^i (s-r)^{n-i}.$$

Since $\delta \neq 0$ and $b > 0$ we have that $r > 1$ is a divisor of the right-hand side of the above equality. Furthermore, we have supposed that $r$ and $s$ are relatively prime. So the left-hand side must be an integer (since the right-hand side is) and inequality (4.1) follows.
2. Now we turn to the other direction. We will construct a family of functions $f_n : \mathcal{B}^n \to \mathcal{B}$ with the property that

$$\sqrt[n]{|E(f_n(X_1, X_2, \dots, X_n))|} \to \frac{1}{s}. \tag{4.2}$$

For this, we will prove the following lemma, which is also of some independent interest. Then (4.1) and (4.2) will yield the result of Theorem 4.2. □

**Lemma 4.2.** *If $b$ is as in Theorem 4.2, we have that*

$$\Xi(b) \leq \frac{1}{s}.$$

*More precisely, for $n > 2r+1$ we obtain*

$$\Xi_n(b) \leq \frac{2r(s-r)^2}{s^n}$$

*and there exists a (deterministic) polynomial-time algorithm for finding an optimal $f$, such that*

$$\xi_{f,n}(b) \leq \frac{2r(s-r)^2}{s^n}.$$

**Proof:** Define $q := s - r$, so that we have $\frac{q}{s} + \frac{r}{s} = 1$. Since we have supposed $b > 0$, it follows that $r > q$. Let $\mathcal{B}_i^{(n)}$ be the $i$-th level of $\mathcal{B}^n$, i.e. those elements of $\mathcal{B}^n$ with Hamming weigth (number of ones) $i$. Let $P_i^{(n)}(b)$ denote the probability that an element of $\mathcal{B}^n$ is equal to some fixed element of $x \in \mathcal{B}_i^{(n)}$. This probability is indeed independent of the specific $x$ and given by

$$P_i^{(n)}(b) = b^i(1-b)^{n-i}.$$

Hence in our case, we have

$$P_i^{(n)}(b) = \frac{r^i q^{n-i}}{s^n}.$$

We want to find collections $C_n \subset \mathcal{B}^n$ such that $s^n w(C_n)$ is "close" to $\frac{s^n}{2}$. Then for the function

$$f_n(X_1, X_2, \ldots, X_n) := 2(\mathbf{1}(C_n)(X_1, X_2, \ldots, X_n) - \frac{1}{2})$$

we will have that $E(f_n(X_1, X_2, \ldots, X_n))$ will be close to 0.
For this construction we proceed as follows: Define an initial collection

$$\tilde{C}_n := \mathcal{B}_n^{(n)} \cup \mathcal{B}_{n-1}^{(n)} \cup \mathcal{B}_{n-2}^{(n)} \cup T,$$

where $T$ is a maximal subset of $\bigcup_{i<n-2} \mathcal{B}_i^{(n)}$ for which $|\mathcal{B}_j^{(n)} \backslash T| \geq r-1$ (for $1 \leq j \leq n-3$) and $s^n w(\tilde{C}_n) \leq \frac{s^n}{2}$. Now let us adjust this collection suitably to bring its weight (multiplied by $s^n$) closer to $\frac{s^n}{2}$. Since $r > q$ and $P_i^{(n)}(b) < P_j^{(n)}(b)$ (for $i < j$) and by the maximality of $T$ we get

$$|s^n w(\tilde{C}_n) - \lfloor \frac{s^n}{2} \rfloor| < s^n P_{n-2}^{(n)}(b) = r^{n-2} q^2.$$

Now consider the cyclic group $\mathbb{Z}_{r^{n-2}q^2}$ and denote by $\pi : \mathbb{Z} \to \mathbb{Z}_{r^{n-2}q^2}$ the canonical projection. Since $r$ and $q$ are relatively prime, it follows that for every $i > 2$, the element $\pi(r^{n-i}q^i)$ has order $r^{i-2}$ in $\mathbb{Z}_{r^{n-2}q^2}$, so that we obtain the following chain (or "tower") of subgroups of $\mathbb{Z}_{r^{n-2}q^2}$:

$$0 = \langle\pi(r^{n-2}q^2)\rangle \subset \langle\pi(r^{n-3}q^3)\rangle \subset \ldots \subset \langle\pi(rq^{n-1})\rangle$$

($\langle x\rangle$ denotes the cyclic subgroup generated by $x$). All the groups in the above chain have index $r$ in the next one and the last one ($\langle\pi(rq^{n-1})\rangle$) has index $rq^2$ in $\mathbb{Z}_{r^{n-2}q^2}$. Thus the group $\langle\pi(rq^{n-1})\rangle$ can be used to approximate every element of $\mathbb{Z}_{r^{n-2}q^2}$ to within an additive error of $rq^2$. In particular for $\Delta := \pi(\lfloor\frac{s^n}{2}\rfloor - s^n w(\tilde{C}_n))$, there exists an element $\Delta' \in \langle\pi(rq^{n-1})\rangle$ such that

$$\Delta - \Delta' \in \{\pi(0), \pi(1), \ldots, \pi(rq^2 - 1)\}.$$

On the other hand, we of course may write $\Delta' =: c\pi(rq^{n-1})$, so by well-known algebraic facts (see Näslund, Russell (2001), p. 308) one has an equation

$$\Delta' = \sum_{i=1}^{n-3} t_i \pi(r^i q^{n-i}) \in \langle\pi(rq^{n-1})\rangle$$

with integers $t_i \in \{0, 1, \ldots, r-1\}$. As $r > q$, we may "lift" this equation to

$$\lfloor\frac{s^n}{2}\rfloor = \sum_{i=1}^{n-3} t_i r^i q^{n-i} - mr^{n-2}q^2 + w(\tilde{C}_n)s^n + E$$

(where $m \leq nr$ and $E \in \{0, 1, \ldots, rq^2 - 1\}$ represents the error term). Now, if we add $t_i$ elements of $\mathcal{B}_i^{(n)}$ to $\tilde{C}_n$ and, on the other hand, remove $m$ elements of $\mathcal{B}_{n-2}^{(n)}$ from $\tilde{C}_n$ (which is possible as long as $m < \binom{n}{n-2}$, i.e. $r < \frac{n-1}{2}$), we indeed obtain a new collection $C_n$ with

$$s^n w(C_n) - \lfloor\frac{s^n}{2}\rfloor < E.$$

Dividing this equation by $s^n$ yields

$$\Xi_n(b) \leq \frac{2rq^2}{s^n}$$

and the result follows (since each step of the above-described algorithm can be carried out in polynomial time).□