INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Online Adaptive Randomness Extraction for Physical and Quantum Entropy Sources: Composing Health-Test Soundness with the Leftover Hash Lemma

Junyop Choe[1], Indran Xavier[2]*

[1]Institute of Mathematics, State Academy of Sciences, Pyongyang, Democratic People's Republic of Korea

[2]Applied Security Technology (AST), Subang Jaya, Malysia

***Corresponding Author**

DOI: https://doi.org/10.51584/IJRIAS.2026.11070071

Received: 24 July 2026; Accepted: 29 July 2026; Published: 03 August 2026

ABSTRACT

Physical and quantum random number generators (QRNGs) do not emit uniform bits: detector bias, dead time, after-pulsing, and environmental drift leave their raw output biased, correlated, and partially predictable to an adversary who models the device. Deployable generators therefore include a conditioning (randomness-extraction) stage that distils near-uniform keys from a source of certified min-entropy, and the cryptographic workhorse of that stage is the Leftover Hash Lemma (LHL). This paper analyses two ways in which real deployments depart from the textbook LHL. First, a physical source has no fixed min-entropy: continuous health tests, as standardised in NIST SP 800-90B, produce a runtime lower-bound estimate that varies from block to block and can occasionally be optimistic. Driving the extractor output length from this estimate, we prove an online adaptive extraction theorem whose security bound cleanly composes the estimator soundness error with the extractor error, yielding a $(\delta + \varepsilon)$-secure key and separating how often the entropy estimate is optimistic from how uniform the hash output is, as two independently tunable budgets. Second, for settings that publish or reuse the extractor seed — public randomness beacons, reproducible pipelines, and seed-constrained devices — we prove that publishing the seed costs nothing, since this is the defining property of a strong extractor, and that reusing a single seed across m independent source blocks costs at most $m\varepsilon$ in statistical distance. A corollary bounds the security loss when the extracted key is consumed by a downstream symmetric primitive, and a worked instantiation with realistic QRNG parameters reports the achievable secret-key rate. A reproducible numerical simulation on a synthetic biased source corroborates all four claims. Throughout, we are explicit about the single load-bearing unproven premise — empirical min-entropy certification of the physical device — which belongs to entropy-source validation and lies outside the scope of the analysis.

Keywords: quantum random number generator, randomness extraction, min-entropy, Leftover Hash Lemma, entropy-source validation

INTRODUCTION

A random number generator built on a physical process — photon arrival times, vacuum-field fluctuations, laser phase noise, thermal noise — can in principle deliver information-theoretic unpredictability that no algorithm can reproduce [5, 9, 10], and device-independent variants can certify randomness from the violation of a Bell inequality [8]. In practice the raw output of such a device is far from uniform: detector bias, dead time, after-pulsing, quantisation, and slow environmental drift all leave exploitable structure. Every deployable generator therefore includes a post-processing or conditioning stage that maps raw samples of certified min-entropy to bits statistically indistinguishable from uniform. The mathematics of that stage is randomness extraction, and its cryptographic workhorse is the Leftover Hash Lemma (LHL) of Impagliazzo, Levin and Luby [6, 2]: a two-universal hash function compresses any source of sufficient min-entropy into a near-uniform string, even given a public description of the hash and any side information the adversary holds.

The textbook LHL assumes (i) a fixed, known min-entropy for the source, and (ii) a fresh, secret-or-irrelevant seed used once. Real deployments violate both, and this paper treats each violation as a separate, self-contained analysis problem.

Departure 1: the min-entropy is a runtime estimate. A physical source drifts, so its available entropy is not a fixed constant but a quantity that must be measured online. Standardised entropy sources run continuous health tests [13] that, per block, emit an assessed conditional min-entropy $\hat{k}$ intended as a lower bound. Occasionally — when the source degrades faster than the tests detect — that assessment is optimistic. A conditioning stage that fixes its output length in advance is therefore either wasteful (too conservative) or unsafe (too aggressive); the right design sets the output length from the runtime estimate, and its security must account for the estimate sometimes being wrong.

Departure 2: the seed may be public and reused. Several common settings expose or reuse the extractor seed: a public randomness beacon publishes its seed so that anyone can verify the output was correctly extracted; a reproducible scientific or engineering pipeline logs the seed so a run can be replayed deterministically; and a seed-constrained device amortises one sampled seed over many extractions to save storage or transmission. Naively, publishing or reusing the randomness that produced a key sounds dangerous. It is not, and quantifying exactly why is the second problem we address.

CONTRIBUTIONS

We are deliberately conservative about novelty. The extractor and the LHL are classical [6, 2, 3, 14], and we claim none of them. Our contributions are the following, each a small but — to our knowledge — not previously isolated consequence of the classical tools.

1. **Online adaptive extraction (Theorem 12).** With the output length driven by the runtime health-test estimate $\hat{k}$, the extracted key is $(\delta + \varepsilon)$-secure, where δ bounds the estimator's soundness error and ε is the extractor error. The bound cleanly composes the two error sources, exposing them as independently tunable budgets — statistical power of the health tests versus entropy loss of the hash.

2. **Public seeds are free (Proposition 9).** Because a strong extractor keeps its output near-uniform given the seed, publishing the seed (beacon, reproducibility, verifiability) incurs no security loss whatsoever.

3. **Seed reuse across independent blocks (Proposition 14).** Reusing a single seed to extract from m mutually independent source blocks costs at most $m\varepsilon$ in statistical distance, giving an explicit seed-rotation budget.

4. **Downstream composition (Corollary 16) and a worked instantiation.** The near-uniform key degrades any downstream keyed primitive's security by at most the extraction error; a numerical example with realistic QRNG parameters reports the secret-key rate.

5. **Reproducible simulation.** A numerical study on a synthetic biased source corroborates the four results above and exhibits the predicted quantitative behaviour of each bound.

Honest scope

The security of the construction rests on one premise we do not prove: that when the health tests pass, the raw block genuinely carries the certified conditional min-entropy (Assumption 10). Establishing this for a specific physical device is the task of entropy-source validation [13] — an empirical, statistical, and physical-modelling effort outside the scope of this paper. Everything downstream of that assumption is proved; we flag where it is load-bearing.

Organisation

The section Preliminaries recalls the entropy measures, statistical distance, universal hashing, the LHL, and a conditioning lemma. Source and Adversary Model fixes the source and adversary model. Results proves the four results. A Worked Instantiation gives the numerical instantiation and Numerical Simulation the reproducible simulation. Relation to Prior Work positions the work, and Conclusion concludes.

PRELIMINARIES

We collect the standard machinery; proofs of the classical facts are included for self-containment. Throughout, log is base 2 and all sets are finite.

Statistical distance and entropy

Definition 1 (Statistical distance). For distributions P, Q on $\mathcal{X}$, $\Delta(P, Q) = 1/2 \sum_x |P(x) - Q(x)|$. We say X is ε-close to Y if $\Delta(X, Y) \le \varepsilon$.

Lemma 2 (Distinguishing characterisation). $\Delta(P, Q) = \max_{S\subseteq\mathcal{X}}(P(S) - Q(S))$; equivalently, for every (possibly randomised) test D, $|\mathbb{P}[D(P) = 1] - \mathbb{P}[D(Q) = 1]| \le \Delta(P, Q)$, with equality for the optimal test.

Proof. Taking $S^\star = \{x: P(x) > Q(x)\}$ maximises $P(S) - Q(S)$, and $P(S^\star) - Q(S^\star) = 1/2 \sum_x |P(x) - Q(x)|$ because the positive and negative parts of $\sum_x (P(x) - Q(x)) = 0$ have equal magnitude. A test is a convex combination of indicators of such sets. □

Definition 3 (Min-entropy and conditional min-entropy). For X on $\mathcal{X}$, the min-entropy is $H_\infty(X) = -\log\max_x p_X(x)$ and the collision probability is $\mathrm{Col}(X) = \sum_x p_X(x)^2$, giving collision entropy $H_2(X) = -\log\mathrm{Col}(X)$. For jointly distributed (X, Z), the average conditional min-entropy is [3]

$$H_\infty(X \mid Z) = -\log \mathbb{E}_{z\leftarrow Z}\left[\max_x p_{X|Z}(x \mid z)\right],$$

so that $2^{-H_\infty(X|Z)}$ is the maximum probability with which an unbounded adversary observing Z guesses X.

Proposition 4. $H_\infty(X) \le H_2(X) \le 2H_\infty(X)$.

Proof. With $p_{\max} = \max_x p_X(x)$: $\mathrm{Col}(X) = \sum_x p_X(x)^2 \le p_{\max} \sum_x p_X(x) = p_{\max}$ gives the left inequality; $\mathrm{Col}(X) \ge p_{\max}^2$ gives the right. □

Two-universal hashing and the Leftover Hash Lemma

Definition 5 (Two-universal family). A family $\mathcal{H} = \{h: \{0,1\}^n \to \{0,1\}^\ell\}$ is two-universal if, for h drawn uniformly and every fixed $x \ne x'$, $\mathbb{P}_h[h(x) = h(x')] \le 2^{-\ell}$.

Example 6 (Field-multiplication family). Identify $\{0,1\}^n$ with $\mathbb{F}_{2^n}$. For seed $\theta = (a, b)$ with $a \in \mathbb{F}_{2^n}$ and $b \in \{0,1\}^\ell$, set $h_{a,b}(x) = \mathrm{msb}_\ell(a \cdot x) \oplus b$. The family is two-universal, with seed length $n + \ell$, and is attractive as a hardware conditioner because $\mathbb{F}_{2^n}$-multiplication is cheap and constant-time. (Toeplitz-matrix families are an equally standard choice with a shorter seed.)

Proof (two-universality of Example 6). For $x \ne x'$, a collision requires $\mathrm{msb}_\ell(a(x - x')) = 0$. Since $d = x - x' \ne 0$ is invertible and a is uniform, $a \cdot d$ is uniform over $\mathbb{F}_{2^n}$, so its top ℓ bits are uniform and vanish with probability $2^{-\ell}$. □

Theorem 7 (Conditional Leftover Hash Lemma [6, 2, 3]). Let $\mathcal{H}$ be two-universal with output length ℓ, let θ (hence $h = h_\theta$) be a uniform seed independent of (X, Z), and suppose $H_\infty(X \mid Z) \ge k$. Then

$$\Delta((\theta, h(X), Z), (\theta, U_\ell, Z)) \;\le\; 1/2\; 2^{-\frac{1}{2}(k-\ell)}.$$

In particular, $\ell \le k - 2\log(1/\varepsilon)$ implies the left-hand side is at most ε.

Proof (sketch). For fixed z, bound the collision probability of $(\theta, h(X))$ under $X \sim X \mid Z = z$: independence of θ and two-universality give $\mathrm{Col} \leq |\mathcal{H}|^{-1}\left(2^{-\ell} + 2^{-H_\infty(X|Z=z)}\right)$. Convert to statistical distance by the Cauchy–Schwarz bound $\Delta(P, U) \leq 1/2\sqrt{N\,\mathrm{Col}(P) - 1}$ on a set of size $N = |\mathcal{H}|\,2^{\ell}$, obtaining $1/2\ 2^{-\frac{1}{2}(H_\infty(X|Z=z)-\ell)}$; average over z using Jensen's inequality (concavity of the square root) together with $H_\infty(X \mid Z) = -\log\mathbb{E}_z[2^{-H_\infty(X|Z=z)}]$. Full details are standard [3, 14]. □

The quantity $2\log(1/\varepsilon)$ is the entropy loss; it is optimal up to an additive constant for extractors that must work for all sources of a given min-entropy [11].

A conditioning lemma

We will combine a rare "bad" event with an extractor bound. The following is the convexity of total-variation distance.

Lemma 8 (Conditioning). For any random variables A, B, V and any event $\mathcal{G}$,

$$\Delta((A, V), (B, V)) \ \leq\ \mathbb{P}[\neg\mathcal{G}] \ +\ \Delta((A, V) \mid \mathcal{G},\ (B, V) \mid \mathcal{G}).$$

Proof. Write the law of (A, V) as $P = \mathbb{P}[\mathcal{G}]P_{\mathcal{G}} + \mathbb{P}[\neg\mathcal{G}]P_{\neg\mathcal{G}}$ and that of (B, V) as $Q = \mathbb{P}[\mathcal{G}]Q_{\mathcal{G}} + \mathbb{P}[\neg\mathcal{G}]Q_{\neg\mathcal{G}}$, conditioning the same event $\mathcal{G}$ on both sides. By convexity of Δ, $\Delta(P, Q) \leq \mathbb{P}[\mathcal{G}]\Delta(P_{\mathcal{G}}, Q_{\mathcal{G}}) + \mathbb{P}[\neg\mathcal{G}]\Delta(P_{\neg\mathcal{G}}, Q_{\neg\mathcal{G}})$; bound $\Delta(P_{\neg\mathcal{G}}, Q_{\neg\mathcal{G}}) \leq 1$ and $\mathbb{P}[\mathcal{G}] \leq 1$. □

SOURCE AND ADVERSARY MODEL

Entropy source. Time is divided into blocks $i = 1,2, \ldots$. In block i the generator emits a raw string $X_i \in \{0,1\}^n$. We assume nothing about independence or bias within a block; we assume only a bound on its conditional min-entropy, certified as below.

Adversary. The adversary E is computationally unbounded — the reason for using a physical source and information-theoretic conditioning in the first place. For block i, E holds side information Z_i, which we take to include any device or environmental leakage and the public health-test statistics S_i for that block. Thus E is at least as informed as any observer of the published diagnostics. The residual unpredictability of X_i to E is measured by $H_\infty(X_i \mid Z_i)$ (Definition 3).

Conditioner. A function $h_i: \{0,1\}^n \to \{0,1\}^{\ell_i}$ is drawn from the public two-universal family of Example 6 using a seed θ_i; the key is $K_i = h_{\theta_i}(X_i)$ truncated to ℓ_i bits, with ℓ_i chosen at runtime. Depending on the deployment, the seed θ_i may be secret, published, or reused; each case is covered below.

Security goal. A key K is η-secure against a view V if $\Delta((K, V), (U_\ell, V)) \leq \eta$, with U_ℓ uniform and independent of V. By Lemma 2, η upper-bounds the advantage of any attack, of any computational power, that uses V to distinguish K from an ideal key.

RESULTS

Publishing the seed is free

Proposition 9 (Public seed). Fix a block and suppose $H_\infty(X \mid Z) \geq k$ and $\ell \leq k - 2\log(1/\varepsilon)$. Let the seed θ be published (part of E's view). Then the key is ε-secure: $\Delta((K, V), (U_\ell, V)) \leq \varepsilon$ with $V = (\theta, Z)$.

Proof. This is precisely the strong form of Theorem 7: the seed θ appears on both sides of the statistical distance, i.e., the output is near-uniform conditioned on the seed. Publishing θ therefore reveals nothing the guarantee has not already accounted for; the public statistics $S \subseteq Z$ are likewise already in the conditioning. □

Proposition 9 is the reason a randomness beacon can publish its seed, and a reproducible pipeline can log it, with no loss: strong extraction is designed to tolerate a public seed. The secret that must never be exposed is the raw source block X (and the key K), never the seed.

Online adaptive extraction

Continuous health tests [13] monitor the raw stream and, per block, output an assessed conditional min-entropy $\hat{k}_i$ intended as a lower bound on $H_\infty(X_i \mid Z_i)$. When the source degrades faster than the tests detect, the assessment can be optimistic. We model this with one soundness parameter.

Assumption 10 (Entropy-source soundness). There is $\delta \in [0,1)$ such that, for every block i, letting $\mathcal{G}_i = \{\hat{k}_i \le H_\infty(X_i \mid Z_i)\}$ be the event that the assessment is a valid lower bound, we have $\mathbb{P}[\neg\mathcal{G}_i] \le \delta$, and, conditioned on $\mathcal{G}_i$ and on the realised transcript $(\theta_i, \hat{k}_i, S_i)$, the block retains its assessed entropy: $H_\infty\big(X_i \mid Z_i, \theta_i, \hat{k}_i, \mathcal{G}_i\big) \ge \hat{k}_i$ (the seed θ_i is drawn independently of the source, so it does not reduce this quantity).

Remark 11. Assumption 10 is the load-bearing, unproven hypothesis. Its first clause is a statement about the statistical power of the health tests; its second says that a "passing" block genuinely has the certified entropy even after the public transcript is revealed. Both are exactly what an entropy-source validation regime such as NIST SP 800-90B [13] is designed to certify empirically for a specific device. We treat δ as a tunable budget: stricter tests lower δ at the cost of more frequent alarms (discarded blocks).

Theorem 12 (Online adaptive extraction). Fix a target extractor error $\varepsilon > 0$. In each block, draw a seed θ_i for the family of Example 6, set the output length

$$\ell_i \;=\; \max\big(0,\ \lfloor \hat{k}_i - 2\log(1/\varepsilon) \rfloor\big), \qquad (1)$$

and output $K_i = h_{\theta_i}(X_i)$ truncated to ℓ_i bits. Then, under Assumption 10, with the adversary view $V_i = (\theta_i, \hat{k}_i, S_i, Z_i)$,

$$\Delta\big((K_i, V_i), (U_{\ell_i}, V_i)\big) \;\le\; \delta + \varepsilon.$$

Proof. Apply Lemma 8 with $A = K_i$, $B = U_{\ell_i}$, $V = V_i$, and event $\mathcal{G} = \mathcal{G}_i$. The first term contributes $\mathbb{P}[\neg\mathcal{G}_i] \le \delta$. For the second, condition additionally on the realised transcript $w = (\theta_i, \hat{k}_i, S_i)$, which fixes the length $\ell_i = \ell_i(w)$ via (1). By the second clause of Assumption 10, on $\mathcal{G}_i$ the source satisfies $H_\infty(X_i \mid Z_i, w) \ge \hat{k}_i$, while (1) gives $\ell_i \le \hat{k}_i - 2\log(1/\varepsilon)$. The seed θ_i is uniform and independent of (X_i, Z_i), and truncation to a prefix is a deterministic map that does not increase statistical distance. Hence the Conditional Leftover Hash Lemma (Theorem 7) applies with $k = \hat{k}_i$ and yields $\Delta\big((K_i, V_i) \mid \mathcal{G}_i, (U_{\ell_i}, V_i) \mid \mathcal{G}_i\big) \le 1/2\ 2^{-\frac{1}{2}(\hat{k}_i - \ell_i)} \le 1/2\,\varepsilon \le \varepsilon$. Summing the two terms gives $\delta + \varepsilon$. □

Remark 13 (Why the composition matters). Theorem 12 separates two physically distinct, separately tunable error sources: δ is bought with statistical rigour in the health tests (how sure are we that the entropy is there?), while ε is bought with entropy loss $2\log(1/\varepsilon)$ in the hash (how uniform is the output given the entropy?). A designer can pick, say, $\delta = 2^{-40}$ and $\varepsilon = 2^{-64}$ independently, rather than conflating them into a single opaque safety margin.

Reusing one seed across independent blocks

Seed-constrained devices, and beacons that publish a single seed per epoch, reuse one seed across many blocks. The next result bounds the cost, provided the source blocks are independent — a reasonable model for a healthy generator sampled at separated times, and a property the health tests are meant to police.

Proposition 14 (Seed reuse). Let θ be one seed drawn once, reused to extract $K_j = h_\theta(X_j)$ for $j = 1, \dots, m$ from source blocks $X_1, \dots, X_m$ that are mutually independent given their side information, each with a valid assessment and each yielding ℓ-bit output with per-block extractor error ε as in Theorem 7. Then

$$\Delta\Big((\theta, K_1, \dots, K_m, \vec{Z}),\ (\theta, U_\ell^{(1)}, \dots, U_\ell^{(m)}, \vec{Z})\Big) \ \leq\ m\,\varepsilon,$$

where $U_\ell^{(1)}, \dots, U_\ell^{(m)}$ are independent uniform strings and $\vec{Z} = (Z_1, \dots, Z_m)$.

Proof. Hybrid argument. Define $H_0, \dots, H_m$, where in H_t the first t outputs are the real extractions $K_1, \dots, K_t$ and the remaining $m - t$ are replaced by fresh independent uniforms; H_0 is the all-ideal distribution and H_m the all-real one. Adjacent hybrids H_t and H_{t+1} differ only in coordinate $t + 1$: uniform versus $K_{t+1} = h_\theta(X_{t+1})$. Condition on θ and on all other blocks and their side information; since X_{t+1} is independent of them given Z_{t+1}, and θ is a valid seed for the two-universal family, the Conditional LHL (Theorem 7, seed-fixed strong form) bounds the distance between the two hybrids by ε. The triangle inequality over the m adjacent pairs gives $\Delta(H_0, H_m) \leq m\varepsilon$. □

Remark 15. Proposition 14 turns seed rotation into an explicit budget: to keep the aggregate distance below a target η, rotate the seed after at most $m = \lfloor \eta/\varepsilon \rfloor$ blocks. Folding in Assumption 10 over m blocks (a union bound adds $m\delta$) gives an end-to-end guarantee of $m(\delta + \varepsilon)$ for a reused-seed window. The independence hypothesis is essential; where blocks may be correlated, the seed must be rotated per block and Theorem 12 applies directly.

Consuming the extracted key

Corollary 16 (Downstream composition). Suppose K is η-secure (e.g. $\eta = \delta + \varepsilon$ from Theorem 12) and is used as the key of any downstream cryptographic scheme Π (for instance AES-GCM authenticated encryption). Then, for every adversary $\mathcal{A}$ (of any computational power) against Π using E's view,

$$\big|\mathbb{P}[\mathcal{A} \text{ wins with key } K] - \mathbb{P}[\mathcal{A} \text{ wins with an ideal uniform key}]\big| \ \leq\ \eta.$$

Hence the game-based security of Π under the real extracted key equals its security under an ideal key, degraded by at most η.

Proof. Running Π and the attack $\mathcal{A}$ is a randomised function of the key together with E's view. Statistical distance cannot increase under such a function (data processing), so the two winning probabilities differ by at most $\Delta((K, V), (U, V)) \leq \eta$ (Lemma 2). □

A Worked Instantiation

We instantiate the pipeline with representative — and deliberately conservative — QRNG parameters. All numbers are illustrative design targets, not measurements of any specific device; a real deployment must replace the assessed entropy with a validated figure per [13].

Source and assessment. Suppose the generator is sampled into 8-bit symbols and a validated continuous health test certifies an assessed conditional min-entropy of $H_{sym} = 3.6$ bits per byte (entropy rate 0.45; typical of a biased but healthy physical source after modelling detector imperfections). Aggregating a raw block of $n = 8192$ bits (1024 bytes) gives a per-block assessment

$$\hat{k} \ =\ 1024 \times 3.6 \ =\ 3686 \text{ bits (rounded down).}$$

Security budget. Target extractor error $\varepsilon = 2^{-64}$ and soundness budget $\delta = 2^{-40}$, so each key is $(\delta + \varepsilon) \leq 2^{-40}$-secure by Theorem 12.

Extracted key length. By (1),

$$\ell \;=\; \hat{k} - 2\log(1/\varepsilon) \;=\; 3686 - 2\cdot 64 \;=\; 3558 \text{ bits}$$

per 8192-bit raw block: an extraction efficiency of $3558/8192 \approx 43.4\%$, i.e. about 0.434 near-uniform key bits per raw bit (3.47 per byte) — essentially the full assessed entropy minus the fixed 128-bit loss, amortised to negligibility over the block.

Seed cost and reuse. The seed for Example 6 is $n + \ell = 8192 + 3558 = 11750$ bits (approximately 1.43 KB). Under Proposition 14, reusing one seed for m independent blocks keeps the window within $m(\delta + \varepsilon) \le m \cdot 2^{-40}$; to stay below an aggregate 2^{-20} the seed may be reused for up to $m = 2^{20} \approx 10^{6}$ blocks — more than 8 Gbit of raw source per seed — before rotation.

Downstream keys. A single 3558-bit extracted block furnishes thirteen fresh 256-bit symmetric keys ($13 \times 256 = 3328$ bits) with room to spare, each of which (Corollary 16) is within 2^{-40} of ideal. Table 1 shows the trade-off as the security budget varies.

Table 1. Extracted key length and efficiency for the worked source ($n = 8192$ raw bits, assessed $\hat{k} = 3686$). Efficiency is ℓ/n. Stronger security costs a fixed additive $2\log(1/\varepsilon)$ per block, negligible at this block size.

ε	**δ**	**loss $2\log(1/\varepsilon)$**	**key bits ℓ**	**efficiency**
2^{-32}	2^{-32}	64	3622	44.2%
2^{-64}	2^{-40}	128	3558	43.4%
2^{-96}	2^{-48}	192	3494	42.7%
2^{-128}	2^{-64}	256	3430	41.9%

Reading the numbers. The pipeline converts a biased physical stream carrying approximately 0.45 bits of assessed entropy per raw bit into keys within 2^{-40} of perfectly uniform. The dominant lever on the key rate is the device's assessed entropy, not the cryptography — the correct engineering conclusion, directing effort to entropy-source validation, the one thing this analysis assumes rather than proves.

Numerical Simulation

We corroborate the four results with a reproducible numerical study. We emphasise at the outset that this is a simulation on a synthetic source, not a measurement of any physical device: the source is a software model with a known entropy, which is precisely what lets us check the theory against ground truth. Hardware validation of a real generator is a separate empirical task (Assumption 10) and is not attempted here. All experiments use a single fixed pseudorandom seed and about 120 lines of NumPy; the numbers below are directly reproducible.

Setup. The synthetic source emits n independent bits, each $\mathrm{Bernoulli}(q)$, so its per-bit collision and min-entropies are $h_2(q) = -\log(q^2 + (1-q)^2)$ and $h_\infty(q) = -\log(\max(q, 1-q))$, and the block values $H_2 = n\,h_2(q)$ and $H_\infty = n\,h_\infty(q)$ are known exactly. The conditioner is a Toeplitz two-universal family (an $\ell \times n$ binary Toeplitz matrix applied over $\mathbb{F}_2$), a standard and seed-efficient alternative [4] to the field family of Example 6. Statistical distance to uniform is estimated by histogramming the ℓ-bit output over 2^{ℓ} bins across N blocks; because a finite sample of a truly uniform source also yields a nonzero estimate, we report alongside each measurement a uniform control (the same estimator applied to genuine uniform draws) as the sampling-noise floor.

Extractor uniformity versus the LHL bound

Table 2 isolates the regime where the true distance is large enough to measure: a small-entropy source ($n = 32$, $q = 0.173$, so $H_2 = 15.5$), with the empirical distance averaged over 16 random seeds to match the seed-averaged quantity the LHL bounds. In every row the measured mean distance lies below the LHL bound

$1/2\,2^{-(H_2-\ell)/2}$ (the guarantee holds) while sitting clearly above the uniform-control floor (the residual non-uniformity is genuinely present and being controlled), and it grows with ℓ as the bound predicts.

Table 2. Seed-averaged empirical statistical distance of the ℓ-bit output versus the LHL bound, for a small-entropy source ($n = 32$, $q = 0.173$, $H_2 = 15.5$, $H_\infty = 8.76$; $N = 2^{20}$ blocks, 16 seeds). The mean distance is below the bound in every row and above the sampling-noise floor, confirming that the bound is a valid, non-vacuous upper bound.

$\boldsymbol{\ell}$	**mean Δ (16 seeds)**	**± s.d.**	**uniform floor**	**LHL bound**
2	0.00432	0.0060	0.00063	4.6×10^{-3}
4	0.00351	0.0011	0.00138	9.2×10^{-3}
6	0.01399	0.0059	0.00311	1.8×10^{-2}
8	0.02628	0.0122	0.00653	3.7×10^{-2}
10	0.04927	0.0115	0.01248	7.3×10^{-2}
12	0.07690	0.0067	0.02463	1.5×10^{-1}

In the complementary high-loss regime ($n = 64$, $H_2 = 31.1$), the true distance is astronomically small and the measurement is dominated by the sampling floor: Table 3 shows the extracted output is indistinguishable from the uniform control to 3–4 significant figures, with the actual guarantee (the LHL bound) orders of magnitude tighter than anything the sample size can resolve.

Table 3. High-loss regime ($n = 64$, $q = 0.173$, $H_2 = 31.1$; $N = 2^{21}$). The extracted output matches the uniform control to within sampling noise; the true distance is bounded far below the measurement floor by the LHL.

$\boldsymbol{\ell}$	**loss $\mathbf{H_2 - \boldsymbol{\ell}}$**	**empirical Δ**	**uniform floor**	**LHL bound**
8	23.1	0.00453	0.00429	1.7×10^{-4}
12	19.1	0.01761	0.01741	6.7×10^{-4}
16	15.1	0.07016	0.06995	2.7×10^{-3}

Statistical test battery: raw versus extracted

Table 4 runs a standard battery on a long stream before and after extraction ($n = 64$, $q = 0.173$, output $\ell = 24$ bits per block). The raw source fails every test decisively — a frequency of 0.173, a byte χ^2 of 3.4×10^7 (against a null mean of 255), and a most-common-value min-entropy of only 2.18 of 8 bits per byte. After extraction the stream passes: balanced frequency, monobit $p = 0.83$, byte $\chi^2 = 248$ (255 degrees of freedom), negligible serial correlation, and a min-entropy estimate of 7.87 of 8 bits per byte.

Table 4. NIST-style statistical battery [13] and most-common-value (MCV) min-entropy estimate, raw versus extracted ($n = 64$, $q = 0.173$, $\ell = 24$). The biased raw source fails every test; the extracted stream passes.

Test (statistic)	**Raw source**	**Extracted**
Length (bits)	1.68×10^7	6.29×10^6
Frequency of ones	0.1727	0.5000

Monobit p-value	0 (fail)	0.83 (pass)
Byte χ^2 (df 255)	3.43×10^7 (fail)	248.4 (pass)
Serial correlation	5.9×10^{-5}	4.7×10^{-4}
MCV min-entropy (bits/byte)	2.18	7.87

Online adaptive extraction: the soundness budget is tunable

We simulate a source whose bias drifts over time, run a finite-sample most-common-value health test [13, 7] on a sliding window of 1536 raw bits, and set the output length from its one-sided upper-confidence estimate of H_∞ (confidence multiplier z). Table 5 sweeps z over 60,000 blocks ($n = 256$, mean true $H_\infty = 88.3$). The empirical soundness-failure rate — how often the estimate exceeds the true min-entropy, i.e. the parameter δ of Assumption 10 — tracks the predicted $\Phi(-z)$ closely and is driven down by four orders of magnitude as the test is made stricter, at the cost of a modest reduction in mean output length. This is the trade-off between δ and key rate in Theorem 12 made concrete.

Table 5. Online adaptive extraction under drift ($n = 256$, $N = 60{,}000$ blocks, entropy loss $2\log(1/\varepsilon) = 40$). The measured soundness-failure rate δ tracks $\Phi(-z)$ and falls as the health test is tightened, trading against mean extracted length — exactly the composition of Theorem 12.

confidence z	**measured δ**	**predicted $\Phi(-z)$**	**mean output ℓ**
1.00	0.1549	0.1587	43.0
1.645	0.0468	0.0500	40.0
2.326	0.0077	0.0100	36.9
2.576	0.0036	0.0050	35.9
3.090	0.0006	0.0010	33.7

Seed reuse scales linearly

Finally, we reuse a single Toeplitz seed to extract from m independent blocks and measure the aggregate distinguishing proxy (the sum of per-block distances). Table 6 shows that it grows exactly linearly in m — the per-block average is constant at 0.070 across a 32-fold range — with no super-linear penalty from reusing the seed. This is the empirical signature of the $m\varepsilon$ bound of Proposition 14.

Table 6. Seed reuse across m independent blocks ($n = 64$, $q = 0.173$, $\ell = 14$, $N = 2^{19}$ per block, one shared seed). The aggregate distance is linear in m, matching the $m\varepsilon$ scaling of Proposition 14.

blocks m	**aggregate distance**	**per-block average**
1	0.0699	0.0699
2	0.1401	0.0700
4	0.2820	0.0705
8	0.5647	0.0706

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

16	1.1254	0.0703
32	2.2522	0.0704

Summary. Across the four experiments the simulation reproduces every qualitative and quantitative prediction: the extractor output is uniform to the resolution of the sample and provably beyond it; the guarantee is a valid non-vacuous bound; a biased source that fails all tests passes after conditioning; the soundness budget δ is tunable and matches $\Phi(-z)$; and seed reuse degrades security only linearly. The synthetic-source caveat remains the whole point — ground truth is known here, whereas for a real device it must be certified empirically.

RELATION TO PRIOR WORK

Randomness extraction and the LHL are classical [6, 2, 3]; Vadhan's monograph [14] is the standard reference and Radhakrishnan and Ta-Shma [11] establish optimality of the $2\log(1/\varepsilon)$ loss. Quantum and physical RNGs and their trust models are surveyed in [5, 9, 10], with device-independent certification in [8]. Post-processing of such generators by universal (typically Toeplitz) hashing is established practice, and seed-efficient constructions that reduce the published seed length are known [4]; validation of physical entropy sources — including the continuous health tests that supply the runtime estimate $\hat{k}$ of Theorem 12, and the min-entropy estimators behind them [7] — is codified in NIST SP 800-90B [13], with the surrounding generator constructions in the SP 800-90C draft [1]. Our aim is not to improve any of these components but to account cleanly for two ways in which deployments depart from the textbook LHL: a runtime-varying, occasionally optimistic entropy estimate, and a seed that is public and/or reused. To our knowledge the $(\delta + \varepsilon)$ composition of estimator soundness with extractor error (Theorem 12) and the paired seed-publication and seed-reuse accounting (Propositions 9 and 14) have not been isolated together in this form, though each step is elementary given the classical tools. We claim the assembly and its accounting, not the tools.

CONCLUSION

We gave a self-contained, provably secure conditioning analysis for a physical or quantum entropy source. Two small results carry the practical content: an online adaptive extractor whose security cleanly composes entropy-estimator soundness δ with extractor error ε into a $(\delta + \varepsilon)$ guarantee (Theorem 12), and a seed-reuse bound of $m\varepsilon$ across m independent blocks (Proposition 14); together with the observation that publishing the seed is security-free (Proposition 9) and the downstream composition of the key into any keyed primitive (Corollary 16). The analysis is explicit about its one unproven premise — empirical min-entropy certification of the physical device (Assumption 10) — and localises all remaining risk there.

Open directions. (i) Replacing the field-multiplication family with a seed-efficient Toeplitz or Trevisan-type extractor to shrink the seed while quantifying the effect on ε; (ii) extending the model to quantum side information, where the smooth quantum conditional min-entropy replaces $H_\infty(X \mid Z)$ and the LHL continues to hold [12]; (iii) a sequential (martingale) refinement of Assumption 10 that models the health-test estimator as an online predictor with adversarially drifting entropy, tightening the union bound behind the $m\delta$ term.

Reproducibility and verification. The simulation uses a single fixed pseudorandom seed and about 120 lines of NumPy; all reported figures are directly reproducible from it. The theoretical results are elementary consequences of classical tools, but subtle errors in information-theoretic security proofs are easy to make. Before any deployment, the theorems and the modelling of Assumption 10 should be checked by an independent expert in randomness extraction and entropy-source validation.

DECLARATIONS

Ethical approval. This study involved no human participants and no animal subjects; ethical approval was therefore not required.

Conflict of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability. No experimental datasets were generated. All numerical results reported in the section Numerical Simulation are produced by a short NumPy simulation with a single fixed pseudorandom seed and are fully reproducible; the simulation script is available from the corresponding author on reasonable request.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

1. Barker, E., Kelsey, J., McKay, K., Roginsky, A., & Turan, M. S. (2022). Recommendation for Random Bit Generator (RBG) Constructions (NIST Special Publication 800-90C, 3rd public draft). National Institute of Standards and Technology. https://csrc.nist.gov/pubs/sp/800/90/c/3pd
2. Bennett, C. H., Brassard, G., Crépeau, C., & Maurer, U. M. (1995). Generalized privacy amplification. IEEE Transactions on Information Theory, 41(6), 1915–1923. https://doi.org/10.1109/18.476316
3. Dodis, Y., Ostrovsky, R., Reyzin, L., & Smith, A. (2008). Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. SIAM Journal on Computing, 38(1), 97–139. https://doi.org/10.1137/060651380
4. Hayashi, M., & Tsurumaru, T. (2016). More efficient privacy amplification with less random seeds via dual universal hash function. IEEE Transactions on Information Theory, 62(4), 2213–2232. https://doi.org/10.1109/TIT.2016.2526018
5. Herrero-Collantes, M., & Garcia-Escartin, J. C. (2017). Quantum random number generators. Reviews of Modern Physics, 89(1), 015004. https://doi.org/10.1103/RevModPhys.89.015004
6. Impagliazzo, R., Levin, L. A., & Luby, M. (1989). Pseudo-random generation from one-way functions. In Proceedings of the 21st Annual ACM Symposium on Theory of Computing (pp. 12–24). Association for Computing Machinery. https://doi.org/10.1145/73007.73009
7. Kelsey, J., McKay, K. A., & Turan, M. S. (2015). Predictive models for min-entropy estimation. In Cryptographic Hardware and Embedded Systems — CHES 2015 (Lecture Notes in Computer Science, Vol. 9293, pp. 373–392). Springer. https://doi.org/10.1007/978-3-662-48324-4_19
8. Liu, Y., Zhao, Q., Li, M.-H., Guan, J.-Y., Zhang, Y., Bai, B., Zhang, W., Liu, W.-Z., Wu, C., Yuan, X., Li, H., Munro, W. J., Wang, Z., You, L., Zhang, J., Ma, X., Fan, J., Zhang, Q., & Pan, J.-W. (2018). Device-independent quantum random-number generation. Nature, 562, 548–551. https://doi.org/10.1038/s41586-018-0559-3
9. Ma, X., Yuan, X., Cao, Z., Qi, B., & Zhang, Z. (2016). Quantum random number generation. npj Quantum Information, 2, 16021. https://doi.org/10.1038/npjqi.2016.21
10. Mannalath, V., Mishra, S., & Pathak, A. (2023). A comprehensive review of quantum random number generators: Concepts, classification and the origin of randomness. Quantum Information Processing, 22(12), 439. https://doi.org/10.1007/s11128-023-04175-y
11. Radhakrishnan, J., & Ta-Shma, A. (2000). Bounds for dispersers, extractors, and depth-two superconcentrators. SIAM Journal on Discrete Mathematics, 13(1), 2–24. https://doi.org/10.1137/S0895480197329508
12. Renner, R. (2005). Security of quantum key distribution [Doctoral dissertation, ETH Zürich]. https://doi.org/10.3929/ethz-a-005115027
13. Turan, M. S., Barker, E., Kelsey, J., McKay, K., Baish, M., & Boyle, M. (2018). Recommendation for the entropy sources used for random bit generation (NIST Special Publication 800-90B). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-90B
14. Vadhan, S. P. (2012). Pseudorandomness. Foundations and Trends in Theoretical Computer Science, 7(1–3), 1–336. https://doi.org/10.1561/0400000010