

Randomness extraction and asymptotic Hamming distance

Cameron E. Freer and Bjørn Kjos-Hanssen*

Abstract

We obtain a non-implication result in the Medvedev degrees by studying sequences that are close to Martin-Löf random in asymptotic Hamming distance. Our result is that the class of stochastically bi-immune sets is not Medvedev reducible to the class of sets having complex packing dimension 1.

1 Introduction

We are interested in the extent to which an infinite binary sequence X , or equivalently a set $X \subseteq \omega$, that is algorithmically random (Martin-Löf random) remains useful as a randomness source after modifying some of the bits. Usefulness here means that some algorithm (extractor) can produce a Martin-Löf random sequence from the result Y of modifying X . For further motivation see Subsection 1.2 and Section 3.

A set that lies within a small Hamming distance of a random set may be viewed as produced by an adaptive adversary corrupting or fixing some bits after looking at the original random set. Similar problems in the finite setting have been studied going back to Ben-Or and Linial [1].

If A is a finite set and $\sigma, \tau \in \{0, 1\}^A$, then the Hamming distance $d(\sigma, \tau)$ is given by

$$d(\sigma, \tau) = |\{n : \sigma(n) \neq \tau(n)\}|.$$

Let the collection of all infinite computable subsets of ω be denoted by \mathfrak{C} . Let $p : \omega \rightarrow \omega$. For $X, Y \in 2^\omega$ and $N \subseteq \omega$ we define a notion of proximity, or similarity, by

$$X \sim_{p,N} Y \iff (\exists n_0)(\forall n \in N, n \geq n_0)(d(X \upharpoonright n, Y \upharpoonright n) \leq p(n)).$$

We will study the effective dimension of sequences that are $\sim_{p,N}$ to certain algorithmically random reals for suitably slow-growing functions p .

*This material is based upon work supported by the National Science Foundation under Grants No. 0652669 and 0901020. Additionally, work on this publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

We use the following notation for a kind of neighborhood around X .

$$[X]_{p,N} = \{Y : Y \sim_{p,N} X\}.$$

Moreover, for a collection \mathcal{A} of subsets of ω ,

$$[\mathcal{A}]_{p,N} = \bigcup \{[X]_{p,N} : X \in \mathcal{A}\}.$$

Turing functionals as random variables. Since a random variable must be defined for all elements of the sample space, we consider a Turing functional Φ to be a map into

$$\Omega := 2^{<\omega} \cup 2^\omega.$$

Setting the domain of Φ to also be Ω allows for composing maps. Let

$$\Lambda^X(n) = X(n).$$

Thus $\Lambda : 2^\omega \rightarrow 2^\omega$ is the identity Turing functional.

We define a probability measure λ on Ω called *Lebesgue (fair-coin) measure*, whose σ -algebra of λ -measurable sets is

$$\mathfrak{F} = \{\mathcal{S} \subseteq \Omega : \mathcal{S} \cap 2^\omega \text{ is Lebesgue measurable}\},$$

by letting $\lambda(\mathcal{S})$ equal the fair-coin measure of $\mathcal{S} \cap 2^\omega$. Thus $\lambda(2^\omega) = 1$ and $\lambda(2^{<\omega}) = 0$, that is, λ is concentrated on the functions that are actually total.

The *distribution* of Φ is the measure $\mathcal{S} \mapsto \lambda\{X : \Phi^X \in \mathcal{S}\}$, defined on \mathfrak{F} . Thus the distribution of Λ is λ .

If $X \in 2^\omega$ then X is called a real, a set, or a sequence depending on context. If $I \subseteq \omega$ then $X \upharpoonright I$ denotes X , viewed as a function, restricted to the set I . We denote the cardinality of a finite set A by $|A|$. Regarding X, Y as subsets of ω and letting $+$ denote sum mod two, note that $(X + Y) \cap n = \{k < n : X(k) \neq Y(k)\}$ and generally for a set $I \subseteq \omega$, $(X + Y) \cap I = \{k \in I : X(k) \neq Y(k)\}$.

For an introduction to algorithmic randomness the reader may consult the recent books by Nies [10] and Downey and Hirschfeldt [4]. Let MLR denote the set of Martin-Löf random elements of 2^ω . For a binary relation R we use a set-theoretic notation for image,

$$R[A] = \{y : (\exists x \in A)((x, y) \in R)\}.$$

Let the *use* $\varphi^X(n)$ be the largest number used in the computation of $\Phi^X(n)$. We write

$$\Phi^X(n) \downarrow @s$$

if $\Phi^X(n)$ halts by stage s , with use at most s ; if this statement is false, we write $\Phi^X(n) \uparrow @s$. We may assume that the running time of a Turing reduction is the same as the use, because any X -computable upper bound on the use is a reasonable notion of use.

For a set $\mathcal{A} \subseteq 2^\omega$, let

$$\begin{aligned} \text{Interior}_{p,N}(\mathcal{A}) &= \{X : (\forall Y \sim_{p,N} X)(Y \in \mathcal{A})\} \\ &\subseteq \text{Interior}_*(\mathcal{A}) = \{X : (\forall Y =^* X)(Y \in \mathcal{A})\} \subseteq \mathcal{A} \end{aligned}$$

where $=^*$ denotes almost equality for all but finitely many inputs. It is easy to see that

$$\text{Interior}_{p,N}(\text{MLR}) = \emptyset$$

whenever $N \subseteq \omega$ and p is unbounded.

Definition 1.1 (Effective convergence). *Let $\{a_n\}_{n \in \omega}$ be a sequence of real numbers.*

- $\{a_n\}_{n \in \omega}$ converges to ∞ effectively if there is a computable function N such that for all k and all $n \geq N(k)$, $a_n \geq k$.
- $\{a_n\}_{n \in \omega}$ converges to 0 effectively if the sequence $\{a_n^{-1}\}_{n \in \omega}$ converges to ∞ effectively.

Definition 1.2. *For a sequence of real numbers $\{a_n\}_{n \in \omega}$, $\lim_{n \rightarrow \infty}^* a_n$ is the real number to which a_n converges effectively, if any; and is undefined if no such number exists.*

As a kind of effective big-O notation, $p_n = \omega^*(q_n)$ means $\lim_{n \rightarrow \infty}^* q_n/p_n = 0$, i.e., q_n/p_n goes to zero effectively.

Central Limit Theorem. Let \mathcal{N} be the cumulative distribution function for a standard normal random variable; so

$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Let \mathbb{P} denote fair-coin probability on Ω . We may write

$$\mathbb{P}(\text{Event}) = \mathbb{P}(\{X : X \in \text{Event}\}) = \lambda\{X : X \in \text{Event}\}.$$

We will make use of the following quantitative version of the central limit theorem.

Theorem 1.3 (Berry-Esséen¹). *Let $\{X_n\}_{n \geq 1}$ be independent and identically distributed real-valued random variables with the expectations $\mathbb{E}(X_n) = 0$, $\mathbb{E}(X_n^2) = \sigma^2$, and $\mathbb{E}(|X_n|^3) = \rho < \infty$. Then there is a constant d (with $.41 \leq d \leq .71$) such that for all x and n ,*

$$\left| \mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{\sigma\sqrt{n}} \leq x\right) - \mathcal{N}(x) \right| \leq \frac{d\rho}{\sigma^3\sqrt{n}}.$$

We are mostly interested in the case $X_n = X(n) - \frac{1}{2}$, $X(n) \in \{0, 1\}$, for $X \in 2^\omega$ under λ , in which case $\sigma = 1/2$.

¹See for example Durrett [5].

1.1 New Medvedev degrees

Let \leq_s denote Medvedev (strong) reducibility and let \leq_w denote Muchnik (weak) reducibility. A recent survey of the theory behind these reducibilities is Hinman [8].

Definition 1.4 (see, e.g., [9]). *A set X is immune if for each $N \in \mathfrak{C}$, $N \not\subseteq X$. If $\omega \setminus X$ is immune then X is co-immune. If X is both immune and co-immune then X is bi-immune.*

Definition 1.5. *A set X is stochastically bi-immune if for each set $N \in \mathfrak{C}$, $X \upharpoonright N$ satisfies the strong law of large numbers, i.e.,*

$$\lim_{n \rightarrow \infty} \frac{|X \cap N \cap n|}{|N \cap n|} = \frac{1}{2}.$$

Definition 1.6. *Let $0 \leq p < 1$. A sequence $X \in 2^\omega$ is p -stochastically dominated if for each $L \in \mathfrak{C}$,*

$$\limsup_{n \rightarrow \infty} \frac{|L \cap n|}{n} > 0 \implies (\exists M \in \mathfrak{C}) \quad M \subseteq L \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{|X \cap M \cap n|}{|M \cap n|} \leq p.$$

The class of stochastically dominated sequences is denoted $\text{SD} = \text{SD}_p$. If $\omega \setminus X \in \text{SD}_p$ then we write $X \in \text{SD}^p$ and say that X is stochastically dominating.

Let IM denote the set of immune sets, CIM the set of co-immune sets, and W3R the set of weakly 3-random sets. Let K denote prefix-free Kolmogorov complexity.

Definition 1.7 (see, e.g., [4, Ch. 13]). *The effective Hausdorff dimension of $A \in 2^\omega$ is*

$$\dim_H(A) = \liminf_{n \in \omega} \frac{K(A \upharpoonright n)}{n}.$$

The complex packing dimension of $A \in 2^\omega$ is

$$\dim_{cp}(A) = \sup_{N \in \mathfrak{C}} \inf_{n \in N} \frac{K(A \upharpoonright n)}{n}.$$

The effective packing dimension of $A \in 2^\omega$ is

$$\dim_p(A) = \limsup_{n \in \omega} \frac{K(A \upharpoonright n)}{n}.$$

Proposition 1.8. *For all $A \in 2^\omega$,*

$$0 \leq \dim_H(A) \leq \dim_{cp}(A) \leq \dim_p(A) \leq 1.$$

Proof. The inequality $\dim_H(A) \leq \dim_{cp}(A)$ uses the fact that each cofinite set $N \subseteq \omega$ is in \mathfrak{C} . The inequality $\dim_{cp}(A) \leq \dim_p(A)$ uses the fact that each $N \in \mathfrak{C}$ is an infinite subset of ω . \square

By examining the complex packing dimension of reals that are $\sim_{p,N}$ to a Martin-Löf random real for p growing more slowly than $n/(\log n)$, we will derive our main result, which states the existence, for each Turing reduction Φ , of a set Y of complex packing dimension 1 for which Φ^Y is not stochastically bi-immune.

1.2 Relation of our results to other recent results.

Jockusch and Lewis [9] prove that the class of bi-immune sets is Medvedev reducible to the class of *almost diagonally non-computable functions* DNC^* , i.e., functions f such that $f(x) = \varphi_x(x)$ for at most finitely many x . Downey, Greenberg, Jockusch, and Milans [3] show that DNC_3 (the class of DNC functions taking values in $\{0, 1, 2\}$) and hence also its superset DNC^* , is not Medvedev above the class of Kurtz random sets. We do not know whether the class of stochastically bi-immune sets is Medvedev reducible to the class of DNC^* functions. We show in Theorem 4.3 below that from a set of complex packing dimension 1 one cannot uniformly compute a stochastically bi-immune set; on the other hand, to compute a DNC^* function from a set of complex packing dimension 1 one would apparently also need to know the witnessing set $N \in \mathfrak{C}$.

Definition 1.9 (see, e.g., [10, Def. 7.6.4]). *A sequence $X \in 2^\omega$ is Mises-Wald-Church (MWC) stochastic if no partial computable monotonic selection rule can select a biased subsequence of X , i.e., a subsequence where the relative frequencies of 0s and 1s do not converge to $1/2$.*

Definition 1.10. *A sequence $X \in 2^\omega$ is BI^2 (bi-immune for sets of size two) if there is no computable collection of disjoint finite sets of size 2 on which the set omits a certain pattern such as 01. More precisely, X is BI^2 if for each computable disjoint collection $\{T_n : n \in \omega\}$ where each T_n has cardinality two, say $T_n = \{s_n, t_n\}$ where $s_n < t_n$, and each $P \subseteq \{0, 1\}$, there is an n such $X(s_n) = P(0)$ and $X(t_n) = P(1)$.*

Each von Mises-Wald-Church stochastic (MWC-stochastic) set is stochastically bi-immune. Our main theorem implies that a set of complex packing dimension 1 does not necessarily uniformly compute a MWC-stochastic set. This consequence is not really new with the present paper, however, because the fact that DNC_3 is not Medvedev above BI^2 is implicit in Downey, Greenberg, Jockusch, and Milans [3] as pointed out to us by Joe Miller. The situation is diagrammatically illustrated in Figure 1, with notation defined in Figures 2 and 3. In the future we could hope to replace complex packing dimension by effective Hausdorff dimension in Theorem 4.3.

2 Hamming space

The Hamming distance between a point and a set of points is defined by $d(y, A) := \min_{a \in A} d(y, a)$. The r -neighborhood of a set $A \subseteq \{0, 1\}^n$ is

$$\Gamma_r(A) = \{y \in \{0, 1\}^n : d(y, A) \leq r\}.$$

In particular,

$$\Gamma_r(\{c\}) = \{y \in \{0, 1\}^n : d(y, c) \leq r\},$$

and

$$\Gamma_r(A) = \bigcup_{a \in A} \Gamma_r(\{a\}).$$

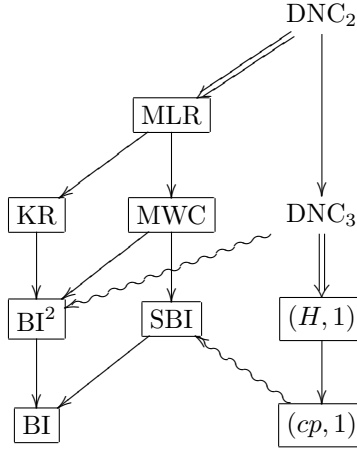


Figure 1: Some Medvedev degrees. The fact that $(cp, 1)$ is not Medvedev above SBI is Theorem 4.3.

—————> Included in

~~~~~> Not Medvedev above

=====> Medvedev above

Figure 2: Meaning of arrows.

| Abbreviation | Unabbreviation                                   | Definition |
|--------------|--------------------------------------------------|------------|
| $DNC_n$      | Diagonally non-computable function in $n^\omega$ |            |
| MLR          | Martin-Löf random                                |            |
| KR           | Kurtz random (weakly 1-random)                   |            |
| MWC          | Mises-Wald-Church stochastic                     | 1.9        |
| SBI          | Stochastically bi-immune                         | 1.5        |
| BI           | bi-immune                                        | 1.4        |
| $BI^2$       | bi-immune for sets of size two                   | 1.10       |
| $(H, 1)$     | effective Hausdorff dimension 1                  | 1.7        |
| $(cp, 1)$    | complex packing dimension 1                      | 1.7        |

Figure 3: Abbreviations used in Figure 1.

A *Hamming-sphere*<sup>2</sup> with center  $c \in \{0, 1\}^n$  is a set  $S \subseteq \{0, 1\}^n$  such that for some  $k$ ,

$$\Gamma_k(\{c\}) \subseteq S \subseteq \Gamma_{k+1}(\{c\}).$$

**Theorem 2.1** (Harper [7]; see also Frankl and Füredi [6]). *For each  $n, r \geq 1$  and each set  $A \subseteq \{0, 1\}^n$ , there is a Hamming-sphere  $S \subseteq \{0, 1\}^n$  such that*

$$|A| = |S|, \quad \text{and} \quad |\Gamma_r(A)| \geq |\Gamma_r(S)|.$$

Following Buhrman et al. [2], we write

$$b(n, k) := \binom{n}{0} + \cdots + \binom{n}{k}.$$

Note that for all  $c \in \{0, 1\}^n$ ,  $|\Gamma_k(\{c\})| = b(n, k)$

If the domain of  $\sigma$  is an interval  $I$  in  $\omega$  rather than an initial segment of  $\omega$ , we may emphasize  $I$  by writing

$$B_r^I(\sigma) = \Gamma_r(\{\sigma\}) = \{\tau \in \{0, 1\}^I : d(\sigma, \tau) \leq r\}.$$

$\mathbb{P}$  denotes the uniform distribution on  $\{0, 1\}^I$ , so by definition

$$\mathbb{P}(E) = \frac{|E|}{2^{|I|}}.$$

Recall that  $D_m$  is the  $m^{\text{th}}$  canonical finite set. The intuitive content of Lemma 2.2 below is that a medium size set is unlikely to contain a random large ball. (Note that we do not assume the sets  $I_m$  are disjoint.)

**Lemma 2.2.** *Let  $\chi \in \omega^\omega$ . Suppose*

$$\lim_{n \rightarrow \infty}^* \chi(n)/\sqrt{n} = \infty. \tag{1}$$

*Let  $f \in \omega^\omega$  be a computable function. Let  $I_m = D_{f(m)}$  and  $n_m = |I_m|$ . Suppose*

$$\lim_{m \rightarrow \infty}^* n_m = \infty. \tag{2}$$

*For each  $m \in \omega$  let  $E_m \subseteq \{0, 1\}^{I_m}$ . Suppose  $\limsup_{m \rightarrow \infty} \mathbb{P}(E_m) \leq \mathfrak{p}$  where  $0 < \mathfrak{p} < 1$  is computable. Writing*

$$B_{\chi(n)}(X) \quad \text{for} \quad B_{\chi(n_m)}^{I_m}(X \upharpoonright I_m),$$

*we have*

$$\lim_{m \rightarrow \infty}^* \mathbb{P}(\{X : B_{\chi(n)}(X) \subseteq E_m\}) = 0. \tag{3}$$

*Moreover, for each  $m_0 \in \omega$  and computable  $\mathfrak{q} \in (\mathfrak{p}, 1)$  there is a modulus of effective convergence in (3) that works for all sets  $\{E_m\}_{m \in \omega}$  such that for all  $m \geq m_0$ ,  $\mathbb{P}(E_m) \leq \mathfrak{q}$ .*

---

<sup>2</sup>A Hamming-sphere is more like a ball than a sphere, but the terminology is entrenched.

*Proof.* Let  $X \in 2^\omega$  be a random variable with  $X =_d \Lambda$ , and

$$S^{(m)} = \sum_{i \in I_m} X(i).$$

Let

$$f_m(x) = \mathbb{P} \left( \frac{S^{(m)} - n/2}{\sqrt{n}/2} \leq x \right)$$

We have<sup>3</sup>

$$|f_m(x) - \mathcal{N}(x)| \leq \frac{d}{\sqrt{n_m}}.$$

Since  $\lim^*_{m \rightarrow \infty} n_m = \infty$ ,  $\lim^* \frac{d}{\sqrt{n_m}} = 0$ . So

$$\lim^*_{m \rightarrow \infty} \sup_x |f_m(x) - \mathcal{N}(x)| = 0. \quad (4)$$

Let  $r = r_m$  be such that

$$b(n, r) \leq |E_m| < b(n, r + 1).$$

Let

$$a_m = \frac{r_m - \frac{n}{2}}{\sqrt{n}/2},$$

and let

$$b_m = a_m + \frac{1}{\sqrt{n}/2} - \frac{\chi(n)}{\sqrt{n}/2}.$$

By (4),

$$\lim^*_{m \rightarrow \infty} |f_m(b_m) - \mathcal{N}(b_m)| = 0. \quad (5)$$

We have

$$\begin{aligned} \limsup_{m \rightarrow \infty} f_m(a_m) &= \limsup_{m \rightarrow \infty} \mathbb{P} \left( \frac{S^{(m)} - n/2}{\sqrt{n}/2} \leq \frac{r_m - \frac{n}{2}}{\sqrt{n}/2} \right) \\ &= \limsup_{m \rightarrow \infty} \mathbb{P}(S^{(m)} \leq r_m) = \limsup_{m \rightarrow \infty} \frac{b(n, r)}{2^n} \leq \limsup_{m \rightarrow \infty} \mathbb{P}(E_m) \leq \mathfrak{p}. \end{aligned}$$

Since  $f_m \rightarrow \mathcal{N}$  uniformly, it follows that

$$\limsup_{m \rightarrow \infty} \mathcal{N}(a_m) \leq \mathfrak{p},$$

---

<sup>3</sup>Indeed, let  $Y_i = X_i - \mathbb{E}(X_i)$  where  $\mathbb{E}(X_i) = \frac{1}{2}$  is the expected value of  $X_i$ , so  $\mathbb{E}(Y_i) = 0$ . By the Berry-Esséen Theorem 1.3, for all  $x$

$$\left| \mathbb{P} \left( \frac{\sum_{i \in I_m} Y_i}{\sigma \sqrt{n}} \leq x \right) - \mathcal{N}(x) \right| \leq \frac{d\rho}{\sigma^3 \sqrt{n}} = \frac{d}{\sqrt{n}},$$

where  $\rho = 1/8 = \mathbb{E}(|Y_i|^3)$ , and  $\sigma = 1/2$  is the standard deviation of  $X_i$  (and  $Y_i$ ).



and so as  $\mathcal{N}$  is strictly increasing,

$$\limsup_{m \rightarrow \infty} a_m \leq \mathcal{N}^{-1}(\mathbf{p}) \quad (= 0 \text{ if } \mathbf{p} = 1/2).$$

Let  $m_0$  be such that for all  $m \geq m_0$ ,

$$a_m + \frac{1}{\sqrt{n_m}/2} \leq \mathcal{N}^{-1}(\mathbf{p}) + 1.$$

Since by assumption  $\lim_{n \rightarrow \infty}^* \chi(n)/\sqrt{n} = \infty$ , we have that  $b_m$  is the sum of a term that goes effectively to  $-\infty$ , and a term that after  $m_0$  never goes above  $\mathcal{N}^{-1}(\mathbf{p}) + 1$  again. Thus

$$\lim_{m \rightarrow \infty}^* b_m = -\infty.$$

It is this rate of convergence that is transformed in the rest of the proof. Now

$$\lim_{m \rightarrow \infty}^* \mathcal{N}(b_m) = 0.$$

Hence by (5),  $\lim_{m \rightarrow \infty}^* f_m(b_m) = 0$ .

Let us write

$$B_t(X) := B_t^{I_m}(X \upharpoonright I_m),$$

considering  $X \upharpoonright I_m$  as a string of length  $n$ . By Harper's Theorem 2.1, we have a Hamming sphere  $H$  with

$$|H| = |\neg E_m| \quad \text{and} \quad |\Gamma_{\chi(n)}(\neg E_m)| \geq |\Gamma_{\chi(n)}(H)|.$$

Then

$$\mathbb{P}(\{X : X \in \Gamma_{\chi(n)}(\neg E_m)\}) \geq \mathbb{P}(\{X : X \in \Gamma_{\chi(n)}(H)\}).$$

Therefore

$$\mathbb{P}(\{X : X \notin \Gamma_{\chi(n)}(\neg E_m)\}) \leq \mathbb{P}(\{X : X \notin \Gamma_{\chi(n)}(H)\}).$$

Let  $\widehat{H}$  be the complement of  $H$ . If the Hamming sphere  $H$  is centered at  $c \in \{0, 1\}^n$  then clearly  $\widehat{H}$  is a Hamming sphere centered at  $\bar{c}$ , where  $\bar{c}(k) = 1 - c(k)$ . Since

$$|\widehat{H}| = |E_m| < b(n, r + 1),$$

we have  $\widehat{H} \subset \Gamma_{r+1}(\{\bar{c}\})$ . So we have:

$$\begin{aligned} & \mathbb{P}(\{X : B_{\chi(n)}(X) \subseteq E_m\}) \leq \mathbb{P}(\{X : B_{\chi(n)}(X) \subseteq \widehat{H}\}) \\ & < \mathbb{P}(\{X : B_{\chi(n)}(X) \subseteq \Gamma_{r+1}(\{\bar{c}\})\}) = \frac{b(n, r + 1 - \chi(n))}{2^n} \\ & = \mathbb{P}[S^{(m)} \leq r + 1 - \chi(n)] = \mathbb{P}\left[\frac{S^{(m)} - \frac{n}{2}}{\sqrt{n}/2} \leq \frac{r + 1 - \frac{n}{2} - \chi(n)}{\sqrt{n}/2}\right] \end{aligned}$$

$$= f_m \left( a_m + \frac{1}{\sqrt{n}/2} - \frac{\chi(n)}{\sqrt{n}/2} \right) = f_m(b_m).$$

Since we showed that  $\lim_{m \rightarrow \infty}^* f_m(b_m) = 0$ , and since by assumption  $\lim_{m \rightarrow \infty}^* n_m = \infty$ ,

$$\lim_{m \rightarrow \infty}^* \mathbb{P}(\{X : B_{\chi(n)}(X) \subseteq E_m\}) = 0.$$

□

### 3 Turing reductions that preserve randomness

The way we will obtain our main result Theorem 4.3 is by proving essentially that for any “randomness extractor” Turing reduction, and any random input oracle, a small number of changes to the oracle will cause the extractor to fail to produce a random output. This would be much easier if we restricted attention to Turing reductions having disjoint uses on distinct inputs, since we would be working with independent random variables. Indeed, one can give an easy proof in that case, which we do not include here. The main technical achievement of the present paper is to be able to work with overlapping use sets; key in that respect is Lemma 3.3 below. The number of changes to the random oracle that we need to make is small enough that the modified oracle has complex packing dimension 1. We were not able to set up the construction so as to guarantee effective Hausdorff dimension 1 (or even greater than 0); this may be an avenue for future work.

For a set of pairs  $E$ , we have the projections  $E^x = \{y : (x, y) \in E\}$  and  $E_y = \{x : (x, y) \in E\}$ .

**Lemma 3.1.** *Let  $\mu_1$  and  $\mu_2$  be probability measures on sample spaces  $\Omega_1$  and  $\Omega_2$  and let  $E$  be a measurable subset of  $\Omega_1 \times \Omega_2$ . Suppose that  $\eta$ ,  $\alpha$ , and  $\delta$  are positive real numbers such that*

$$\mu_1 E_y > \eta \quad (\forall y \in \Omega_2), \quad \text{and} \tag{6}$$

$$\mu_1 \{x : \mu_2 E^x \leq \alpha\} \geq 1 - \delta. \tag{7}$$

Then  $\eta < \alpha + \delta$ .

*Proof.* By Fubini’s theorem,

$$\begin{aligned} \eta &< \int_{\Omega_2} \mu_1(E_y) d\mu_2(y) = \iint_{\Omega_1 \times \Omega_2} E(x, y) d\mu_1(x) d\mu_2(y) = \int_{\Omega_1} \mu_2(E^x) d\mu_1(x) \\ &\leq \alpha \cdot \mu_1 \{x : \mu_2(E^x) \leq \alpha\} + 1 \cdot \mu_1 \{x : \mu_2(E^x) \geq \alpha\} \leq \alpha \cdot 1 + 1 \cdot \delta. \end{aligned}$$

□

**Definition 3.2.** For a real  $X$  and a string  $\sigma$  of length  $n$ ,

$$(\sigma \searrow X)(n) = \begin{cases} \sigma(n) & \text{if } n < |\sigma|, \\ X(n) & \text{otherwise,} \end{cases}$$

and

$$(\sigma \frown X)(n) = \begin{cases} \sigma(n) & \text{if } n < |\sigma|, \\ X(n - |\sigma|) & \text{otherwise.} \end{cases}$$

Thinking of  $\sigma$  and  $X$  as functions we may write

$$\sigma \searrow X = \sigma \cup (X \upharpoonright \omega \setminus |\sigma|)$$

and thinking in terms of concatenation we may write

$$\sigma \frown X = \sigma X.$$

**Lemma 3.3.** Let  $\Phi$  be a Turing reduction such that

$$\lambda(\Phi^{-1} \llbracket \text{SD}_{\mathfrak{p}} \rrbracket) = 1 \tag{8}$$

and let  $\Phi_{\sigma}^X = \Phi^{\sigma \searrow X}$ . Then for any finite set  $\Sigma \subseteq 2^{<\omega}$ ,

$$(\forall \varepsilon > 0)(\forall i_0)(\exists i > i_0)(\forall \sigma \in \Sigma)$$

$$\mathbb{P}(\{X \mid \Phi_{\sigma}^X(i) = 1\}) \leq \mathfrak{p} + \varepsilon.$$

*Proof.* First note that for all  $\sigma \in 2^{<\omega}$ ,  $\lambda(\Phi_{\sigma}^{-1} \llbracket \text{SD}_{\mathfrak{p}} \rrbracket) = 1$  as well.

Suppose otherwise, and fix  $\varepsilon$ ,  $i_0$  and  $\Sigma$  such that

$$(\forall i > i_0)(\exists \sigma \in \Sigma) \quad \mathbb{P}(\Phi_{\sigma}(i) = 1) > \mathfrak{p} + \varepsilon.$$

By density of the rationals in the reals we may assume  $\varepsilon$  is rational and hence computable. Since there are infinitely many  $i$  but only finitely many  $\sigma$ , it follows that there is some  $\sigma$  such that

$$(\exists^{\infty} k > i_0) \quad \mathbb{P}(\Phi_{\sigma}(k) = 1) > \mathfrak{p} + \varepsilon \tag{9}$$

and in fact

$$\limsup |\{k < n : \mathbb{P}(\Phi_{\sigma}(k) = 1) > \mathfrak{p} + \varepsilon\}| / n > 0.$$

Fix such a  $\sigma$  and let  $\Psi = \Phi_{\sigma}$ . Let  $\{\ell_n\}_{n \in \omega}$  be infinitely many values of  $k$  in (9) listed in increasing order; note that  $L = \{\ell_n\}_{n \in \omega}$  may be chosen as a computable sequence.

For an as yet unspecified subsequence  $\mathcal{K} = \{k_n\}_{n \in \omega}$ ,  $\mathcal{K} \subseteq L$ , let

$$E = \{(X, n) : \Psi^X(k_n) = 1\}. \tag{10}$$

We obtain then also projections  $E_n = \{X : \Psi^X(k_n) = 1\}$ ,  $E^X = \{n : \Psi^X(k_n) = 1\}$ . By (9) we have for all  $n \in \omega$ ,

$$\lambda E_n > \mathfrak{p} + \varepsilon. \tag{11}$$

The fraction of events  $E_n$  that occur in  $N = \{0, \dots, N-1\}$  for  $X$  is denoted

$$e_N^X = \frac{|E^X \cap N|}{N}$$

By assumption (8),

$$\lambda \left\{ X : (\exists \mathcal{K} \subseteq L)(\exists M)(\forall N \geq M) \left( e_N^X \leq \mathfrak{p} + \frac{\varepsilon}{2} \right) \right\} = 1.$$

Thus there is an  $M$  and a  $\mathcal{K}$  (using that  $\mathfrak{C}$  is countable) such that

$$\lambda \left\{ X : e_M^X \leq \mathfrak{p} + \frac{\varepsilon}{2} \right\} \geq \lambda \left\{ X : (\forall N \geq M) e_N^X \leq \mathfrak{p} + \frac{\varepsilon}{2} \right\} \geq 1 - \frac{\varepsilon}{3}. \quad (12)$$

Let  $\Omega_1$  be the unit interval  $[0, 1]$ . Let  $\Omega_2 = M = \{0, 1, \dots, M-1\}$ . Let  $\mu_1 = \lambda$ . Let  $\mu_2 = \text{card}$  be the counting measure on the finite set  $M = \{0, 1, \dots, M-1\}$ , so that for a finite set  $A \subset M$ ,  $\text{card}(A)$  is the cardinality of  $A$ .<sup>4</sup> Let  $\eta = \mathfrak{p} + \varepsilon$ ,  $\alpha = \mathfrak{p} + \varepsilon/2$ , and  $\delta = \varepsilon/3$ , and note that  $\eta > \alpha + \delta$ . By (11), (12) and Lemma 3.1,  $\eta < \alpha + \delta$ , a contradiction.  $\square$

## 4 Extraction and Hamming distance

**Theorem 4.1.** *Let  $\mathfrak{p} < 1$  be computable. Let  $p : \omega \rightarrow \omega$  be any computable function such that  $p(n) = \omega^*(\sqrt{n})$ . Let  $\Phi$  be a Turing reduction. There exists an  $N \in \mathfrak{C}$  and an almost sure event  $\mathcal{A}$  such that*

$$\mathcal{A} \cap \text{Interior}_{p,N}(\Phi^{-1}[\mathcal{A}]) = \emptyset$$

*Proof.* Let

$$\mathcal{A} := \text{W3R} \subset \text{MLR} \subset \text{CIM} \cap \text{IM} \cap (\text{SD}_{\mathfrak{p}} \cup \text{SD}^{\mathfrak{p}}).$$

We show

1. If  $\lambda(\Phi^{-1}[\text{SD}_{\mathfrak{p}}]) = 1$ , or  $\lambda(\Phi^{-1}[\text{SD}^{\mathfrak{p}}]) = 1$ , then

$$\text{MLR} \cap \text{Interior}_{p,N}(\Phi^{-1}[\text{CIM}]) = \emptyset, \quad \text{or}$$

$$\text{MLR} \cap \text{Interior}_{p,N}(\Phi^{-1}[\text{IM}]) = \emptyset, \quad \text{respectively.}$$

2. Otherwise; then

$$\text{W3R} \cap \text{Interior}_*(\Phi^{-1}[\text{SD}_{\mathfrak{p}}]) = \emptyset$$

and

$$\text{W3R} \cap \text{Interior}_*(\Phi^{-1}[\text{SD}^{\mathfrak{p}}]) = \emptyset$$

---

<sup>4</sup>In this case,  $\int \mu_1(E_y) d\mu_2(y) = \int \mu_1(E_n) d\mu_2(n) = \int \lambda(E_n) d \text{card}(n) = \sum_{n \in \Omega_2} \lambda(E_n) \text{card}(\{n\}) = \sum_{n=0}^{M-1} \lambda(E_n) \cdot 1$ .

Proof of (2): If we are not in case (1) then  $\lambda\{X \mid \Phi^X \in \text{SD}_{\mathfrak{p}}\} < 1$ , so by the 0-1 Law,  $\lambda\{X \mid (\forall Y =^* X)(\Phi^Y \in \text{SD}_{\mathfrak{p}})\} = 0$ . This is (contained in) a  $\Pi_3^0$  null class, so if  $X \in \text{W3R}$  then  $(\exists Y =^* X)(\Phi^Y \notin \text{SD}_{\mathfrak{p}})$  hence we are done.  
Proof of (1): By Lemma 3.3,

$$(\exists \mathfrak{p} < 1)(\forall \varepsilon > 0)(\forall n)(\forall i)(\exists i' > i)(\forall \sigma \in 2^{=n})$$

$$\mathbb{P}(\{Z : \Phi^{\sigma \setminus Z}(i') = 1\}) \leq \mathfrak{p} + \varepsilon; \quad (13)$$

Since  $\Phi$  is total for almost all oracles, it is clear that  $i'$  is a computable function  $f(k, n)$  of  $\varepsilon = 1/k$  and  $n$ . Let  $g : \omega \rightarrow \omega$  be the computable function with  $\lim_{n \rightarrow \infty} g(n) = \infty$  given by  $g(s) = 2s$ . Let  $n_0 = 0$  and  $i_0 = 0$ . Assuming  $s \geq 0$  and  $n_s$  and  $i_s$  have been defined, let

$$i_{s+1} = f(g(s), n_s),$$

and let  $n_{s+1}$  be large enough that

$$(\forall \sigma \in 2^{=n_s}) \quad \lambda\{Z \mid \Phi^{\sigma \setminus Z}(i_{s+1}) \uparrow @n_{s+1}\} \leq \frac{1}{2s}, \quad (14)$$

$$\lim_{s \rightarrow \infty}^* \frac{p(n_{s+1} - n_s)}{\sqrt{n_{s+1} - n_s}} = \infty, \quad \text{and} \quad \sum_{k=0}^s p(n_{k+1} - n_k) \leq p(n_{s+1}).$$

Note that since  $i' > i$  in (13), we have  $i_{s+1} > i_s$  and hence  $R := \{i_0, i_1, \dots\}$  is a computable infinite set. We now have

$$(\forall s)(\forall \sigma \in 2^{=n_s}) \quad \mathbb{P}(\{Z : \Phi^{\sigma \setminus Z}(i_{s+1}) = 1\}) \leq \mathfrak{p} + \frac{1}{2s} \quad (15)$$

so

$$\mathbb{P}(\{Z : \Phi^{\sigma \setminus Z}(i_{s+1}) \downarrow = 1 @n_{s+1}\}) \leq \mathfrak{p} + \frac{1}{2s}. \quad (16)$$

Note  $[a, b) = b \setminus a$ .

Let  $X \in \text{MLR}$ . We aim to define  $Y \sim_p X$  such that  $\Phi^Y \notin \text{MLR}$ . We will in fact make  $Y \leq_T X$ , so we define a reduction  $\Xi$  and let  $Y = \Xi^X$ . Since we are defining  $Y$  by modifying bits of  $X$ , the use of  $\Xi$  will be the identity function:  $\xi^X(n) = n$ .

Since  $n_0 = 0$ ,  $Y \upharpoonright n_0$  is the empty string. Suppose  $s \geq 0$  and  $Y \upharpoonright n_s$  has already been defined. The set of “good” strings now is

$$\mathcal{G} = \{\tau \succ Y \upharpoonright n_s \mid \Phi^{\tau \upharpoonright n_{s+1}}(i_{s+1}) = 0\}.$$

Define the “cost” of  $\tau$  to be the additional Hamming distance to  $X$ , i.e.,  $d(\tau) = |(X + \tau) \cap [n_s, n_{s+1})|$ .

**Case 1.**  $\mathcal{G} \neq \emptyset$ . Then let  $Y \upharpoonright n_{s+1}$  be any  $\tau_0 \in \mathcal{G}$  of length  $n_{s+1}$  and of minimal cost, i.e., such that  $d(\tau_0) = \min\{d(\tau) \mid \tau \in \mathcal{G}\}$ . That is, let

$$Y \upharpoonright n_{s+1} \in \arg \min_{\tau \in \mathcal{G}} d(\tau).$$

**Case 2.** Otherwise. Then make no further changes to  $X$  up to length  $n_{s+1}$ , i.e., let  $Y \upharpoonright_{n_{s+1}} = Y \upharpoonright_{n_s} \searrow X \upharpoonright_{n_{s+1}}$ .

This completes the definition of  $\Xi$  and hence of  $Y$ . It remains to show that  $\Phi^Y \notin \text{MLR}$ . For any string  $\sigma$  of length  $n_s$  let

$$\begin{aligned} E_{s+1}^\sigma &= \left\{ Z \in \{0, 1\}^{[n_s, n_{s+1})} : \neg (\Phi^{\sigma \searrow Z}(i_{s+1}) \downarrow = 0 @ n_{s+1}) \right\} \\ &= \left\{ Z : \Phi^{\sigma \searrow Z}(i_{s+1}) \downarrow = 1 @ n_{s+1} \right\} \\ &\quad \cup \left\{ Z : \Phi^{\sigma \searrow Z}(i_{s+1}) \uparrow @ n_{s+1} \right\} \end{aligned}$$

Since (14) and (16) hold for all strings of length  $n_s$ , in particular they hold for  $\sigma = \Xi^X \upharpoonright n_s$ , so

$$(\forall s) \quad \mathbb{P}(E_{s+1}^\sigma) \leq \mathfrak{p} + \frac{1}{2s} + \frac{1}{2s} = \mathfrak{p} + \frac{1}{s}, \quad \text{hence} \quad \limsup_{s \rightarrow \infty} \mathbb{P}(E_{s+1}^{X \upharpoonright n_s}) \leq \mathfrak{p}. \quad (17)$$

Let

$$U_s^{X \upharpoonright n_s} = \{Z \in \{0, 1\}^{[n_s, n_{s+1})} : B_{p(n_{s+1}-n_s)}^{[n_s, n_{s+1})}(Z) \subseteq E_{s+1}^{X \upharpoonright n_s}\}.$$

Since

$$\frac{p(n_{s+1} - n_s)}{\sqrt{n_{s+1} - n_s}} \rightarrow^* \infty.$$

we can apply Lemma 2.2 and there is  $h(s)$  with  $\lim_{s \rightarrow \infty}^* h(s) = 0$  and

$$\mathbb{P}(U_s^{X \upharpoonright n_s}) \leq h(s)$$

that only depends on an upper bound for an  $s_0$  such that for all  $s \geq s_0$ ,  $\mathbb{P}(E_{s+1}) \leq \mathfrak{q}$  (where  $\mathfrak{p} < \mathfrak{q} < 1$  and  $\mathfrak{q}$  is just some fixed computable number). Since by (17) such an upper bound can be given that works for all  $X$ , actually  $h(s)$  may be chosen to not depend on  $X$ . Let

$$V_s = \{Z : Z \in U_s^{Z \upharpoonright n_s}\},$$

then  $V_s$  is uniformly  $\Delta_1^0$ . To find the probability of  $V_s$  we note that for each of the  $2^{n_s}$  possible beginnings of  $Z$ , there are at most  $(h(s) \cdot 2^{n_{s+1}-n_s})$  continuations of  $Z$  on  $[n_s, n_{s+1})$  that make  $Z \in V_s$ ; so we compute

$$\begin{aligned} \mathbb{P}(V_s) &= |\{Z \in \{0, 1\}^{n_{s+1}} : Z \upharpoonright [n_s, n_{s+1}) \in U_s^{Z \upharpoonright n_s}\}| 2^{-n_{s+1}} \\ &\leq 2^{n_s} (h(s) \cdot 2^{n_{s+1}-n_s}) 2^{-n_{s+1}} = h(s) \end{aligned}$$

so since  $\lim_{s \rightarrow \infty}^* h(s) = 0$ ,  $\{V_s\}_{s \in \omega}$  is a Kurtz randomness test. Let  $\{m_s\}_{s \in \omega}$  be a computable sequence such that  $\sum_{s > t} h(m_s) \leq 2^{-t}$ . Let  $W_t = \bigcup_{s \geq t} V_{m_s}$ . Then  $\mathbb{P}(W_t) \leq 2^{-t}$  and  $W_t$  is uniformly  $\Sigma_1^0$  and hence it is a Martin-Löf randomness test. Since  $X \in \text{MLR}$ ,  $X \notin W_t$  for some  $t$  and hence  $X \notin V_{m_s}$  for all but finitely many  $s$ . So  $\Phi^Y(m_s) = 0$  for all but finitely many  $s$ , hence  $\Phi^Y \notin \text{CIM}$ .

By construction, we have

$$|(X + Y) \cap [n_s, n_{s+1})| \leq p(n_{s+1} - n_s)$$

for all but finitely many  $n$ . Therefore

$$|(X + Y) \cap [0, n_{s+1})| \leq \sum_{k=0}^s p(n_{k+1} - n_k) \leq p(n_{s+1})$$

so  $X \sim_{p,N} Y$  where  $N = \{n_s : s \in \omega\}$ . □

#### 4.1 Main result

**Lemma 4.2.** *Let  $p(n) = o(\frac{n}{\log n})$  and let  $N \in \mathfrak{C}$ . If  $X \in \text{MLR}$  and  $X \sim_{p,N} Y$  then  $\dim_{cp}(Y) = 1$ .*

*Proof.* Suppose there are at most  $p(n)$  many bits changed to go from  $X \upharpoonright n$  to  $Y \upharpoonright n$ , in positions  $a_1, \dots, a_{p(n)}$ . (In case there are fewer than  $p(n)$  changed bits, we can repeat  $a_i$  representing the bit 0 which we may assume is changed.) Let  $(Y \upharpoonright n)^*$  be a shortest description of  $Y \upharpoonright n$ . From the code

$$0^{|K(Y \upharpoonright n)|} \frown 1 \frown K(Y \upharpoonright n) \frown (Y \upharpoonright n)^* \frown a_1 \cdots a_{p(n)}$$

we can effectively recover  $X \upharpoonright n$ . Thus

$$\begin{aligned} n - c_1 &\leq K(X \upharpoonright n) \leq 2 \log[K(Y \upharpoonright n)] + 1 + K(Y \upharpoonright n) + p(n) \log n + c_2 \\ &\leq 2 \log[n + 2 \log n + c_3] + 1 + K(Y \upharpoonright n) + p(n) \log n + c_2. \end{aligned}$$

Hence

$$\begin{aligned} n &\leq^+ 3 \log n + K(Y \upharpoonright n) + p(n) \log n, \quad \text{and} \\ n - (p(n) + 3) \log n &\leq^+ K(Y \upharpoonright n). \end{aligned}$$

□

**Theorem 4.3.** *For each Turing reduction procedure  $\Phi$  there is a set  $Y$  with  $\dim_{cp}(Y) = 1$  such that  $\Phi^Y$  is not stochastically bi-immune.*

*Proof.* Let  $p(n) = n^{2/3}$ , so that  $p(n) = o(n/\log n)$  and  $p(n) = \omega^*(\sqrt{n})$ . By the proof of Theorem 4.1 and since the sequence of numbers  $n_s$  is computable, for each weakly 3-random set  $X$  there is a set  $Y \sim_{p,N} X$  (for some  $N \in \mathfrak{C}$ ) such that  $\Phi^Y$  is not both co-immune and in  $\text{SD}_{1/2}$ , in particular  $\Phi^Y \notin \text{SBI}$ . By Lemma 4.2, each such  $Y$  has complex packing dimension 1. □

## References

- [1] Michael Ben-Or and Nati Linial, *Collective coin flipping*, Randomness and Computation (S. Micali, ed.), Academic Press, New York, 1989, pp. 91-115. ↑1
- [2] Harry Buhrman, Lance Fortnow, Ilan Newman, and Nikolai Vereshchagin, *Increasing Kolmogorov complexity*, STACS 2005, Lecture Notes in Comput. Sci., vol. 3404, Springer, Berlin, 2005, pp. 412–421. MR2151636 (2006c:68091) ↑7
- [3] Rodney G. Downey, Noam Greenberg, Carl G. Jockusch Jr., and Kevin G. Milans, *Binary subtrees with few labeled paths*, *Combinatorica* **31** (2011), no. 3, 285–303, DOI 10.1007/s00493-011-2634-3. MR2835294 (2012i:05336) ↑5
- [4] Rodney G. Downey and Denis R. Hirschfeldt, *Algorithmic randomness and complexity*, Theory and Applications of Computability, Springer, New York, 2010. MR2732288 (2012g:03001) ↑2, 4
- [5] Rick Durrett, *Probability: theory and examples*, 4th ed., Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2010. MR2722836 (2011e:60001) ↑3
- [6] Peter Frankl and Zoltán Füredi, *A short proof for a theorem of Harper about Hamming-spheres*, *Discrete Math.* **34** (1981), no. 3, 311–313, DOI 10.1016/0012-365X(81)90009-1. MR613409 (83a:05004) ↑7
- [7] Lawrence H. Harper, *Optimal numberings and isoperimetric problems on graphs*, *J. Combinatorial Theory* **1** (1966), 385–393. MR0200192 (34 #91) ↑7
- [8] Peter G. Hinman, *A survey of Mučnik and Medvedev degrees*, *Bull. Symbolic Logic* **18** (2012), no. 2, 161–229, DOI 10.2178/bsl/1333560805. MR2931672 ↑4
- [9] Carl G. Jockusch Jr. and Andrew E.M. Lewis, *Diagonally non-computable functions and bi-immunity*, *J. Symbolic Logic.* to appear. ↑4, 5
- [10] André Nies, *Computability and randomness*, Oxford Logic Guides, vol. 51, Oxford University Press, Oxford, 2009. MR2548883 ↑2, 5