THE BULLETIN OF SYMBOLIC LOGIC
Volume 19, Number 2, June 2013

SHIFT-COMPLEX SEQUENCES

MUSHFEQ KHAN

Abstract. A Martin-Löf random sequence is an infinite binary sequence with the property that every initial segment σ has prefix-free Kolmogorov complexity $K(\sigma)$ at least $|\sigma| - c$, for some constant $c \in \omega$. Informally, initial segments of Martin-Löf randoms are highly complex in the sense that they are not compressible by more than a constant number of bits. However, all Martin-Löf randoms necessarily have contiguous *substrings* of arbitrarily low complexity. If we demand that all substrings of a sequence be uniformly complex, then we arrive at the notion of shift-complex sequences. In this paper, we collect some of the existing results on these sequences and contribute two new ones. Rumyantsev showed that the measure of oracles that compute shift-complex sequences is zero. We strengthen this result by proving that the Martin-Löf random sequences that do not compute shift-complex sequences are exactly the incomplete ones, in other words, the ones that do not compute the halting problem. In order to do so, we make use of the characterization by Franklin and Ng of the class of incomplete Martin-Löf randoms via a notion of randomness called *difference randomness*. Turning to the power of shift-complex sequences as oracles, we show that there are shift-complex sequences that do not compute Martin-Löf random (or even Kurtz random) sequences.

§1. Introduction. We use $K : 2^{<\omega} \to \omega$ to denote *prefix-free Kolmogorov complexity*. Informally, prefix-free Kolmogorov complexity is a measure of how complicated a string is to describe, or equivalently, how *incompressible* it is. A string of a million zeros ought to have a short description relative to its length and hence Kolmogorov complexity much smaller than a million. On the other hand, a string of random bits will, with high probability, have no description shorter than itself, hence Kolmogorov complexity roughly equal to its length.

One of the major threads in the study of algorithmic randomness has been to determine how information-theoretic formulations of string complexity (such as prefix-free Kolmogorov complexity) interact with measure-theoretic formulations of randomness of infinite sequences. For example, the class of Martin-Löf random sequences was originally defined in terms of *tests* that

Received June 14, 2012.
2010 *Mathematics Subject Classification.* 68Q30, 03D32, 03D28.
Key words and phrases. Kolmogorov complexity, subshifts, Martin-Löf randomness.
The author's research was partially funded by NSF Grant DMS-1001847.

© 2013, Association for Symbolic Logic
1079-8986/13/1902-0002/$2.70
DOI:10.2178/bsl.1902020

generalize the idea of statistical tests of randomness. But it also has a precise characterization in terms of prefix-free Kolmogorov complexity. If f and g are functions from some set S to ω, then we write $f \leq^+ g$ to indicate that there exists a constant $c \in \omega$ such that for all $x \in S$, $f(x) \leq g(x) + c$. Schnorr showed[1] that a sequence $X \in 2^\omega$ is Martin-Löf random if and only if $K(X \upharpoonright n) \geq^+ n$, where $X \upharpoonright n$ is the initial segment of X of length n.

Every Martin-Löf random sequence is *normal*, meaning that finite strings of equal length occur as substrings with equal asymptotic frequency. Therefore, every Martin-Löf random sequence must have substrings that are compressible by more than any given factor.[2] Because of Schnorr's theorem, we cannot hope to construct sequences with the property that for all substrings σ, $K(\sigma) \geq^+ |\sigma|$. However, it is possible to construct sequences that are uniformly *somewhat* complex everywhere. The following definitions make this idea precise:

DEFINITION 1.1. Fix $\delta \in (0, 1)$ and $b \in \omega$. A sequence $X \in 2^\omega$ is (δ, b)-*shift-complex* if for every substring σ of X, $K(\sigma) \geq \delta|\sigma| - b$.

A sequence is δ-*shift-complex* if it is (δ, b)-shift-complex for some $b \in \omega$.

A sequence is *shift-complex* if it is δ-shift-complex for some $\delta \in (0, 1)$.

In this paper, we collect some of the main results in the growing body of research on shift-complex sequences. In addition, we contribute two new results that shed light on how shift-complex sequences relate to Martin-Löf random sequences in the setting of the Turing degrees. Access to a sequence of the latter type implies access to arbitrarily long finite strings of high complexity, so can we exploit such resources to *effectively* produce shift-complex sequences? A result by Rumyantsev [13] states that this is not the case: if a sequence is "sufficiently random", then it cannot compute any shift-complex sequence. We calibrate precisely the level of randomness at which this theorem holds. In a similar vein, we ask if access to a shift-complex sequence enables us to effectively obtain a Martin-Löf random sequence. Again, the answer is no.

§2. Preliminaries and overview. We let ω denote the natural numbers. By a *sequence* we mean an element of 2^ω, i.e., an infinite sequence of binary digits indexed by ω. We use the term *real* to refer both to sequences and to elements of $\mathbb{R}$. By a *string* we mean a finite sequence of binary digits, i.e., an element of $2^{<\omega}$. If σ is a string, then $|\sigma|$ denotes the length of σ. In the previous section, we introduced the preordering $\geq^+$. The equivalence relation induced by this preordering is denoted by $=^+$.

A set of reals $X \subseteq 2^\omega$ is called *effectively open* (or a Σ^0_1 *class*) if there is a c.e. set W of finite strings such that X consists exactly of the reals that

[1] See, for example, [11] Section 3.2.

[2] As an extreme example, if 0^n denotes a string of zeros of length n, then $K(0^n) \leq^+ 2\log(n)$.

have an initial segment in W. We refer to W as a *presentation* of X. The complement of an effectively open set in 2^ω is called *effectively closed* (or a Π^0_1 *class*). A real X is of *PA degree* if it computes a member of every nonempty Π^0_1 class.

As remarked earlier, K denotes prefix-free Kolmogorov complexity. A *prefix-free machine* is a partial computable function from strings to strings that has prefix-free domain: if σ and τ are distinct strings in the domain, then neither is a prefix of the other. Let $\mathbb{U}$ be a *universal* prefix-free machine: it can simulate any other prefix-free machine. In particular, this means that every string has a $\mathbb{U}$-*description*, i.e., for each string σ, there is an input τ to $\mathbb{U}$ that produces σ as the output. Then, for a string σ, $K(\sigma)$ is the length of a shortest $\mathbb{U}$-description of σ. The string σ^* is a distinguished shortest description of σ: it is the lexicographically least of the strings of length $K(\sigma)$ on which the universal prefix-free machine halts and outputs σ in the least number of steps. If τ is a string, the *conditional prefix-free Kolmogorov complexity* $K(\sigma \mid \tau)$ is defined much as before, except now we imagine that $\mathbb{U}$ has access to τ. For an in-depth account of the theory of Kolmogorov complexity, we refer the reader to Downey and Hirschfeldt [3] or Nies [11].

Finally, if X is a sequence in 2^ω, the *effective Hausdorff dimension* of X, denoted by $\dim(X)$, is

$$\liminf_{n\to\infty} \frac{K(X \upharpoonright n)}{n},$$

while the *effective packing dimension* of X, denoted by $\mathrm{Dim}(X)$, is obtained by replacing the lim inf with a lim sup.

We begin by surveying the known constructions of shift-complex sequences in Section 3, and show that they lead easily to the existence of bi-infinite shift-complex sequences (i.e., sequences indexed by ξ, the order type of the integers) in Section 4. In Section 5, we provide a proof of a result by Rumyantsev that shows that the measure of oracles that compute shift-complex sequences is 0. We also answer the following question that arises from Rumyantsev's theorem:

QUESTION 2.1. How random does an oracle have to be to ensure that it does *not* compute a shift-complex sequence?

In Theorem 5.7, we show that every Martin-Löf random that computes a shift-complex sequence is Turing complete, using the characterization by Franklin and Ng [6] of this class via a notion of randomness they term *difference randomness*. The difference randoms are those sequences that escape being captured by *difference tests*, which are a generalization of Martin-Löf tests where each test element is a *difference* of Σ^0_1 sets. They comprise a class properly between the weak 2-randoms and the Martin-Löf randoms. We remark that Theorem 5.7 could be viewed as a generalization of Stephan's result in [14] that every Martin-Löf random of PA degree is Turing complete.

In Section 6, we turn to the question of whether shift-complex sequences can compute random sequences. In Theorem 6.7, we show that there are shift-complex sequences that compute no Kurtz random real (and hence no Martin-Löf random real). In order to do so, we adapt the technique employed by Greenberg and Miller [7] of using slow-growing diagonally noncomputable functions to avoid computing random reals.

§3. Constructions of shift-complex sequences. At the time of writing, there are three existence proofs of shift-complex sequences in the literature. The first is due to Durand, Levin and Shen [4]:

THEOREM 3.1 (Durand, Levin and Shen [4]). *For any* $0 < \delta < 1$, *there exists a* δ*-shift-complex sequence.*

PROOF. The construction proceeds by building a sequence using extensions of an appropriately chosen fixed length m (which depends only on δ), such that each extension is sufficiently complex relative to the initial segment constructed so far.

First, note that

$$K(xy) + K(|y|) \geq^+ K(x, y)$$

since there is a prefix-free machine that, from a description of xy concatenated with a description of $|y|$, can recover both x and y, and $\mathbb{U}$ can simulate this machine.

By the prefix-free symmetry of information theorem,[3] $K(x, y) =^+ K(x) + K(y \mid x^*)$. Since $K(|y|) \leq^+ 2\log(|y|)$, we have $K(xy) - K(x) \geq^+ K(y \mid x^*) - 2\log(|y|)$. Let c_0 be such that $K(xy) - K(x) \geq K(y \mid x^*) - 2\log(|y|) - c_0$.

Now for any length n, and for any x, we can choose a string y of length n so that $K(y \mid x^*) \geq n$. Let m be large enough so that $m - 2\log m - c_0 \geq \delta m$. Then for any string x, there exists a string y of length m such that $K(xy) - K(x) \geq \delta m$.

We construct A in blocks of size m, each time choosing a block such that the complexity of the string built so far increases by at least δm. It follows immediately that the desired property holds for initial segments of A of lengths that are multiples of m. We claim that $K(\sigma) \geq \delta|\sigma|$ also holds for substrings σ of A that start and end at indices that are multiples of m. Note that there is a constant c_1 such that

$$K(xy) \leq K(x) + K(y) + c_1.$$

If $\alpha\sigma$ is an initial segment of A where σ starts and ends at indices that are multiples of m, $K(\alpha\sigma) - K(\alpha) \geq \delta|\sigma|$. By the inequality above, $K(\sigma) \geq \delta|\sigma| - c_1$.

[3] For more details, see Downey and Hirschfeldt [3], chapter 3.

Next, if σ is an arbitrary subsequence of A, we can pad σ at either end to obtain a string $\alpha\sigma\beta$ that starts and ends at indices that are multiples of m. Since both α and β have length less than m, there is a constant c_2 such that

$$K(\sigma) + c_2 \geq K(\alpha\sigma\beta) \geq \delta|\alpha\sigma\beta| - c_1 \geq \delta|\sigma| - c_1,$$

and so $K(\sigma) \geq \delta|\sigma| - (c_1 + c_2)$. Thus A is $(\delta, c_1 + c_2)$-shift-complex. ⊣

Rumyantsev and Ushakov in [12] provide an alternative existence proof. They use the Lovász Local Lemma to show that for each δ, there is a $b \in \omega$ such that there exist arbitrarily long *finite* (δ, b)-shift-complex strings. The existence of an infinite (δ, b)-shift-complex sequence then follows from the compactness of 2^ω.

The final construction is due to Miller [10]. We say that $X \in n^\omega$ *avoids* a set $S \subseteq n^{\leq\omega}$ if no $\sigma \in S$ is a substring of X. The set of all sequences in n^ω that avoid a given set of finite strings is called a *subshift*. Miller provides a condition on the lengths of strings in S that guarantees that the subshift of S is nonempty. We outline the proof of this result since an effective version of this construction is required in the proof of Theorem 6.3. Since we are interested only in binary sequences, we present it for the case $n = 2$:

THEOREM 3.2 (Miller [10]). *Let $S \subseteq 2^{<\omega}$. If there is a $c \in (1/2, 1)$ such that*

$$\sum_{\tau \in S} c^{|\tau|} \leq 2c - 1,$$

then there is an $X \in 2^\omega$ that avoids S.

PROOF. For $\sigma, \tau \in 2^{<\omega}$, we define

$$T_{\sigma,\tau} = \{\rho \in 2^{<\omega} : \ |\rho| < |\tau| \text{ and } \sigma\rho \text{ ends in } \tau\}$$

and let

$$w(\sigma) = \sum_{\tau \in S} \sum_{\rho \in T_{\sigma,\tau}} c^{|\rho|}.$$

It is helpful to think of $w(\sigma)$ as a measure of the danger of an extension of σ ending in a forbidden string. Note that if σ ends in a string $\tau \in S$, then $\lambda \in T_{\sigma,\tau}$, and so $w(\sigma) \geq 1$. We build the sequence X a bit at a time, ensuring that for each initial segment σ of X, $w(\sigma) < 1$. Then X avoids S.

Suppose that we have built a string σ that avoids S and that $w(\sigma) < 1$. Note that $w(\sigma 0) + w(\sigma 1) = w(\sigma)/c + (\sum_{\tau \in S} c^{|\tau|})/c$. The second term on the right corresponds to the new threats that emerge as a result of extending σ by either a 0 or a 1, while the first term corresponds to the existing threats to σ, magnified by a factor of $1/c$. So we have

$$w(\sigma 0) + w(\sigma 1) = \frac{w(\sigma)}{c} + \frac{1}{c} \sum_{\tau \in S} c^{|\tau|} < \frac{1}{c} + \frac{2c - 1}{c} = 2,$$

from which it follows that either $w(\sigma 0)$ or $w(\sigma 1)$ is strictly less than 1. ⊣

Theorem 3.1 follows as a corollary.

COROLLARY 3.3 (Miller [10]). *Fix $\delta \in (0,1)$. Let $b = -\log(1-\delta)+1$ and $S = \{\tau \in 2^{<\omega} \mid K(\tau) < \delta|\tau| - b\}$. Then there is an $X \in 2^\omega$ that avoids S, and is therefore δ-shift-complex.*

PROOF. Let $c = 2^{-\delta}$. Then

$$\sum_{\tau\in S} c^{|\tau|} = \sum_{\tau\in S} 2^{-\delta|\tau|} < \sum_{\tau\in S} 2^{-(K(\tau)+b)} = 2^{-b}\sum_{\tau\in S} 2^{-K(\tau)} \leq 2^{-b}$$

where the last inequality follows from Kraft's inequality.[4] Now $2^{-b} = (1-\delta)/2$, which for $\delta \in (0,1)$ is less than $2^{1-\delta}-1 = 2c-1$, so we can apply Theorem 3.2. ⊣

An advantage of the construction in Theorem 3.2 is that it can be effectivized to yield the following:

PROPOSITION 3.4. *Suppose $S \subseteq 2^{<\omega}$ satisfies the condition of Theorem* 3.2 *and is computable. Then there is a computable $X \in 2^\omega$ that avoids S.*

Before proceeding with the proof, we establish some terminology. We say $c \in \mathbb{R}$ is a *computable real number* if there is an algorithm which, when given a rational ε, outputs a rational d such that $|d-c| \leq \varepsilon$. The computable real numbers form a field. We say $c \in \mathbb{R}$ is *left-c.e.* if there is an algorithm that enumerates an increasing sequence $(q_i)_{i\in\omega}$ of rationals such that $\lim_{i\to\infty} q_i = c$. It is not difficult to show that if the sum of two left-c.e. real numbers is computable, then both are computable.

PROOF OF PROPOSITION 3.4. Let $a_n = |S\cap 2^n|$, and let $f(x) = \sum_{n\in\omega} a_n x^n$. We first argue that if there is a $c \in \mathbb{R}$ such that c is computable, $f(c)$ is computable, and $f(c) \leq 2c-1$, then $w(\sigma)$ is computable, uniformly[5] in $\sigma \in 2^{<\omega}$. We proceed by induction on the length of σ. First, note that $w(\lambda) = 0$ (hence computable). Now assume that $w(\sigma)$ is computable. As observed in the proof of Theorem 3.2,

$$w(\sigma 0) + w(\sigma 1) = \frac{w(\sigma)}{c} + \frac{f(c)}{c},$$

where the right hand side is computable by hypothesis. Both $w(\sigma 0)$ and $w(\sigma 1)$ are left-c.e. reals, and since their sum is uniformly computable, both are computable (uniformly). It follows that we can now *effectively* build X bit by bit as in Theorem 3.2.

Next, we show that if the coefficients a_n are computable, then there is a computable c such that $f(c)$ is also computable. The function $f(x)-2x+1$ is concave up for $x > 0$ and so one of the following cases must hold:

[4] If A is a prefix-free set of strings, then $\sum_{\tau\in A} 2^{-|\tau|} \leq 1$.

[5] We mean that there is a single procedure that, given σ and a rational error ε as input, computes $w(\sigma)$ to within ε.

Case 1: There is an interval $[p, q] \subseteq (1/2, 1)$ such that for all $x \in [p, q]$, $f(x) \leq 2x - 1$. In this case, we can choose any rational $c \in [p, q)$. Since $f(q)$ converges,

$$a_i c^i = a_i q^i \frac{c^i}{q^i} \leq f(q) \left(\frac{c}{q}\right)^i \leq L d^i,$$

where L is any rational greater than $f(q)$ and d is any rational in $[c/q, 1)$. In other words, $f(c)$ is dominated term-by-term by the convergent geometric series $\sum_{i \in \omega} L d^i$. It follows that $f(c)$ is computable: to compute it to within ε, choose $k \geq N$ such that $(Ld^k)/(1 - d) < \varepsilon$ and compute $\sum_{i=0}^{k-1} a_i c^i$.

Case 2: There is a unique $c \in (1/2, 1)$ such that $f(c) = 2c - 1$. The function $f(x) - 2x + 1$ is computably approximable from below by $f_s(x) - 2x + 1$, where $f_s(x) = \sum_{n \leq s} a_n x^n$. For s large enough, $f_s(x) - 2x + 1$ has two roots, a_s and b_s, both of which are computable. Note that c is between a_s and b_s, so to compute c to within ε, we simply search for s such that $|a_s - b_s| < \varepsilon$. Now $f(c)$, being equal to $2c - 1$, is also computable. ⊣

The constructions above produce sequences that are *at least* δ-shift-complex. An *exactly δ-shift-complex sequence* is one that is δ-shift-complex but not δ'-shift-complex for any $\delta' > \delta$, while an *almost δ-shift-complex sequence* is one that is δ'-shift-complex for every $\delta' < \delta$, but not δ-shift-complex. Hirschfeldt and Kach have shown that it is possible to adapt the construction in Theorem 3.1 to produce such sequences. Recall that for a sequence X, $\mathrm{Dim}(X)$ denotes the effective packing dimension of X.

THEOREM 3.5 (Hirschfeldt and Kach [8]). *For any $\delta \in (0, 1)$, there is a sequence X that is almost δ-shift-complex and a sequence Y that is exactly δ-shift-complex. Moreover, X and Y can be chosen so that* $\mathrm{Dim}(X) = \mathrm{Dim}(Y) = \delta$.

It is easy to see that the effective Hausdorff dimension of a δ-shift-complex sequence is at least δ. Since the effective packing dimension of a sequence bounds its effective Hausdorff dimension, for each of the sequences in Theorem 3.5, the effective Hausdorff dimension is δ.

Hirschfeldt and Kach have also observed that the effective packing dimension of a shift-complex sequence is always less than 1. In fact, this is true of all sequences that have the property that some string never occurs in the sequence.

PROPOSITION 3.6 (Folklore). *Suppose a finite string σ never occurs as a substring of $X \in 2^\omega$. Then* $\mathrm{Dim}(X) < 1$.

PROOF. Let $m = |\sigma|$. For $n \in \omega$, let $\pi(n)$ denote the number of strings of length n that occur as substrings of X. Since $\pi(m) \leq 2^m - 1$, for all $j \geq 1$, $\pi(mj) \leq (2^m - 1)^j$. If j is large enough, $\pi(mj) \leq 2^{mj-1}$. So we can represent substrings of X of length mj using an alphabet consisting of strings of length $mj - 1$, which allows us to uniformly compress substrings

of X by a factor of $mj/(mj-1)$. In particular,

$$K(X\upharpoonright n) \leq^+ K(n) + \frac{mj-1}{mj}n \leq^+ 2\log(n) + \frac{mj-1}{mj}n,$$

and so $\mathrm{Dim}(X) \leq (mj-1)/mj < 1$. ⊣

A similar argument can be used to show:

THEOREM 3.7 (Hirschfeldt and Kach [8]). *Fix a δ-shift-complex sequence A. Then for some $\varepsilon > 0$, there is a $(\delta+\varepsilon)$-shift-complex sequence $B \leq_T A$.*

We close the section with a couple of observations:

PROPOSITION 3.8. *For every $\delta \in (0,1)$, there is a nonempty Π^0_1 class of δ-shift-complex sequences.*

PROOF. We assume that δ is a rational. By any of the constructions described above, for some b, the class of (δ, b)-shift-complex sequences is nonempty. For each string σ, the set of descriptions of σ is uniformly c.e., by which we mean there is a single program that, given an arbitrary string σ, enumerates its descriptions. It follows that the set W of finite strings that are not (δ, b)-shift-complex is c.e., since we can construct a program that enumerates a finite string γ if any substring of γ has a description that is too short. A real X is (δ, b)-shift-complex if and only if it has no initial segment in W. Hence, the set of (δ, b)-shift-complex sequences is a Π^0_1 class. ⊣

COROLLARY 3.9. *Every oracle of PA degree computes a δ-shift-complex sequence for every $\delta \in (0,1)$.*

§4. Bi-infinite Shift-complex Sequences. Given a shift-complex sequence $A \in 2^\omega$, it is easy to produce a bi-infinite sequence (in 2^ξ, where ξ denotes the order type of the integers) of lower complexity by treating the even and odd bits of A as two distinct sequences and reversing one of them. If B and C are sequences in 2^ω, we denote by $B \oplus C$ the *join* of B and C, i.e., the sequence in 2^ω formed by interleaving the bits of B and C.

PROPOSITION 4.1. *Fix $\varepsilon > 0$. Every $(1-\varepsilon)$-shift-complex sequence uniformly computes a bi-infinite $(1-2\varepsilon)$-shift-complex sequence.*

PROOF. Suppose an infinite sequence A is $(1-\varepsilon)$-shift-complex. Letting $A = B \oplus C$, we claim that $Z = \overleftarrow{B}\,C$ (where $\overleftarrow{B}$ is just B reversed) is $(1-2\varepsilon)$-shift-complex.

First, we consider a substring σ of Z that lies completely in either the left or the right half of Z (i.e., it does not overlap the index 0). Let τ be the corresponding substring of the other half of Z of the same length such that $\tau \oplus \sigma$ is a substring of A. By assumption, $K(\tau \oplus \sigma) \geq^+ (1-\varepsilon)2n$, where $n = |\sigma|$. It is easy to see that there is a prefix-free machine that witnesses the inequality $K(\sigma) + K(\tau \mid \sigma) \geq^+ K(\tau \oplus \sigma)$. A similar argument shows that $K(\tau \mid \sigma) \leq^+ K(n \mid \sigma) + n$. So we have:

$$K(\sigma) \geq^+ (1-\varepsilon)2n - K(n \mid \sigma) - n.$$

But there is a constant c such that $K(|\sigma| \mid \sigma) \leq c$, so we obtain:

$$K(\sigma) \geq^+ (1-\varepsilon)2n - n = (1-2\varepsilon)n.$$

Next, we consider the case where σ overlaps 0. Let $\sigma = \tau\gamma$ where τ is the part of σ to the left of 0 and γ the part to the right. If $|\tau| = |\gamma|$, then $K(\sigma) \geq^+ (1-\varepsilon)|\sigma| > (1-2\varepsilon)|\sigma|$, so suppose (without loss of generality) that $n = |\tau| > |\gamma| = m$. We can pad σ on the right by ρ, so that $|\sigma\rho| = 2n$ and $\sigma\rho$ is a substring of Z. Then, $K(\sigma\rho) \geq^+ (1-\varepsilon)2n$, since it corresponds to an initial segment of A of length $2n$. We also have the inequality $K(\sigma\rho) \leq^+ K(\sigma) + K(\rho \mid \sigma)$. So

$$\begin{aligned} K(\sigma) &\geq^+ (1-\varepsilon)2n - K(\rho \mid \sigma) \\ &= (1-2\varepsilon)(n+m) + (n-m) + 2\varepsilon m - K(\rho \mid \sigma). \end{aligned}$$

We know that $K(\rho \mid \sigma) \leq^+ K(|\rho| \mid \sigma) + |\rho| = K(n-m \mid \sigma^*) + (n-m)$. Now, since $|\sigma| = n+m$, $K(n-m \mid \sigma) \leq^+ K(m) \leq^+ 2\log(m)$. Combining all of the above, we have:

$$K(\sigma) \geq^+ (1-2\varepsilon)(n+m) + 2(\varepsilon m - \log(m)) \geq^+ (1-2\varepsilon)(n+m). \qquad \dashv$$

It follows that for any $\delta \in (0,1)$, by starting out with a sequence of sufficiently high shift-complexity, we can apply the proposition above to obtain a bi-infinite δ-shift-complex sequence:

COROLLARY 4.2. *For every $\delta \in (0,1)$, there exists a bi-infinite δ-shift-complex sequence.*

However, the answer to the following question is as yet unknown:

QUESTION 4.3. Does every δ-shift-complex sequence compute a bi-infinite δ-shift-complex sequence?

§5. Extracting shift-complexity from randomness. Initial segments of Martin-Löf random reals have high prefix-free Kolmogorov complexity. Might it not be possible to effectively obtain a shift-complex sequence from any Martin-Löf random real? A theorem by Rumyantsev shows that this is not the case. In this section we give a proof of Rumyantsev's theorem and show something stronger: The Martin-Löf random reals which compute shift-complex sequences do so not because they are random, but because they are of PA degree.

THEOREM 5.1 (Rumyantsev [13]). *The set of reals that compute shift-complex sequences has measure* 0.

We need a preliminary definition and a lemma.

DEFINITION 5.2. We say that a shift-complex sequence Y is *abundant* if there is an $n \geq 2$ such that Y is δ-shift-complex for some $\delta > 1/n$ and further, for every m, Y contains at least $2^{m(n-1)/n}$ different substrings of length m.

LEMMA 5.3. *Every shift-complex sequence computes an abundant shift-complex sequence.*

PROOF. We begin by observing that if a sequence Y is δ-shift-complex for some $\delta > 1/n$ but is not abundant with witness m (i.e., Y has fewer than $2^{m(n-1)/n}$ different substrings of length m), then it has fewer than $(2^{m(n-1)/n})^n = 2^{m(n-1)}$ strings of length mn. Thus, substrings of Y of length mn can be represented using an alphabet consisting of strings of length $m(n-1)$, giving us a uniform way to compress substrings of Y whose lengths are a multiple of mn by a factor of $n/(n-1)$.

We show by induction on n that if a sequence X is δ-shift-complex for $\delta > 1/n$, then X computes an abundant shift-complex sequence. First note that if X is δ-shift-complex for some $\delta > 1/2$, then X itself is abundant. For if it is not, fix m such that X has fewer than $2^{m/2}$ strings of length m. By the observation above, we can compress substrings of X whose lengths are a multiple of $2m$ by a factor of 2, which contradicts the fact that X is δ-shift-complex for δ strictly greater than $1/2$.

Next, suppose X is δ-shift-complex for $\delta > 1/n$ and is not abundant with witness m. Again by the observation at the beginning of the proof, $X \equiv_T Y$, where Y is obtained from X by coding segments of X of length mn by strings of length $m(n-1)$. It is not hard to see that Y is δ'-shift-complex for $\delta' > 1/(n-1)$. By the induction hypothesis, Y computes an abundant shift-complex sequence. ⊣

In the proof of Theorem 5.1 we appeal to the Kraft–Chaitin Theorem, which allows us to compress a set of strings subject to some constraints.

DEFINITION 5.4. A c.e. set $W \subseteq \omega \times 2^{<\omega}$ is a *Kraft–Chaitin set* if

$$\sum_{(n,\tau)\in W} 2^{-n} \leq 1.$$

We refer to elements of W as *requests*.

THEOREM 5.5 (Kraft–Chaitin). *For each Kraft–Chaitin set W, there is a constant d such that for all $(n,\tau) \in W$, $K(\tau) \leq n + d$. The constant d is called the* coding constant *of W. Further, one can effectively obtain d from a procedure for enumerating W.*

In conjunction with the Kraft–Chaitin Theorem, we use the Recursion Theorem from computability theory, which allows us to know the coding constant of a Kraft–Chaitin set while we are enumerating it. For a detailed exposition of this technique, we refer the reader to Chapter 2 of Nies [11].

PROOF OF THEOREM 5.1. The strategy is to assume, for a contradiction, that a positive measure set of reals computes shift-complex sequences, hence abundant shift-complex sequences by Lemma 5.3. We will argue that since every oracle in this positive measure set computes *lots* of strings of fairly high complexity, some of these strings must be computed by a large measure of

oracles. Using the Kraft–Chaitin Theorem, we will compress these strings, thereby invalidating those oracles that compute them.

More precisely, each shift-complex sequence computes an abundant shift-complex sequence for some n, and there are countably many choices for n. So we can assume that for some $n \geq 2$ there is a positive measure set A of oracles every member of which computes an abundant shift-complex sequence for n. By a similar argument, we can further assume that there is a single Turing functional Γ, a rational $\delta > 1/n$ and an integer b such that if $X \in A$, then Γ^X is abundant for n and (δ, b)-shift-complex.

Suppose $\mu(A) > \varepsilon$ where ε is rational. Let α be a rational in the interval $(1/n, \delta)$. We build a Kraft–Chaitin set W, and by the Recursion Theorem, we assume that we know the coding constant d of W in advance. Let m be chosen so that:

1. αm is an integer,
2. $\alpha m < \delta m - (b + d)$,
3. $2^{(\alpha - \frac{1}{n})m}\varepsilon > 1 - \varepsilon$.

We remark here that the choice of m is effective in ε, δ, n and b. This fact will be relevant to the proof of Theorem 5.7.

If σ is a string of length m, let

$$S_\sigma = \{X \in 2^\omega \mid \Gamma^X \text{ contains } \sigma \text{ as a substring}\}.$$

We say that we *compress* a string τ of length m when we enumerate the pair $(\alpha m, \tau)$ into W. Note that by compressing a string, we ensure that it has a description no longer than $\alpha m + d$. We can compress $2^{\alpha m}$ many strings. If we ensure that $\mu(\bigcup_{(n,\sigma)\in W} S_\sigma) > 1 - \varepsilon$ then we will have obtained a contradiction, since if an oracle X computes a string that has a description strictly smaller than $\delta m - b$, then $X \in \bar{A}$.

Suppose by stage s, we have compressed a set W_s of strings. Let C_s be the set of oracles that we have seen by stage s that compute a string in W_s via Γ. The measure of C_s should be thought of as the measure of oracles that we have eliminated by stage s by determining that they belong to $\bar{A}$. The key is that we compress a string when we observe that the additional measure eliminated by doing so is large. Note that since $\mu(A) > \varepsilon$ and every element of A computes at least $2^{m(n-1)/n}$ strings of length m, there is a string σ of length m such that

$$\mu(S_\sigma \cap A) > \varepsilon \left(\frac{2^{m(n-1)/n}}{2^m} \right) = \varepsilon \left(2^{-\frac{m}{n}} \right).$$

Since A is disjoint from C_s, eventually we will see a string σ such that $\mu(S_\sigma \setminus C_s) > \varepsilon \cdot 2^{-\frac{m}{n}}$. When we encounter such a string, we compress it.

Each time we compress a string, we ensure that C_s grows by a measure greater than $\varepsilon \cdot 2^{-\frac{m}{n}}$, so by compressing $2^{\alpha m}$ strings, the measure of oracles

we will have eliminated will be greater than

$$2^{\alpha m} \cdot \varepsilon \cdot 2^{-\frac{m}{n}} = 2^{(\alpha - \frac{1}{n})m}\varepsilon,$$

which exceeds $1 - \varepsilon$ by our choice of m above. ⊣

Because any property that holds of almost all oracles must hold of sufficiently random oracles, Rumyantsev's theorem could be rephrased as follows: If a sequence X is sufficiently random, then it does not compute a shift-complex sequence. In the remainder of the section, we establish precisely how random an oracle must be for this property to hold. Franklin and Ng in [6] introduce a notion called difference randomness and show it to be strictly stronger than Martin-Löf randomness and strictly weaker than weak 2-randomness.

DEFINITION 5.6. A *difference test* is a uniform sequence of pairs (U_i, V_i) of Σ^0_1 classes such that for all $i \in \omega$, $\mu(U_i \setminus V_i) \leq 2^{-i}$. A real *passes* a difference test $((U_i, V_i))_{i\in\omega}$ if it is not contained in $\bigcap_{i\in\omega}(U_i \setminus V_i)$. A real is *difference random* if it passes all difference tests.

THEOREM 5.7. *No difference random real computes a shift-complex sequence.*

PROOF. Suppose a real Y computes a shift-complex sequence X via Γ. Without loss of generality we can assume X is abundant for some n and that it is (δ, b)-shift-complex for some $b \in \omega$ and $\delta > 1/n$. For an arbitrary real A, Γ^A may be partial. We say σ is a substring of Γ^A if Γ^A converges on a contiguous set of positions to σ. Let G_m be the set of reals A such that

1. Γ^A contains no substring σ such that $K(\sigma) < \delta|\sigma| - b$,
2. Γ^A contains at least $2^{m(n-1)/n}$ different substrings of length m.

Then G_m is $U_m \setminus V_m$, where U_m is the set of reals A such that Γ^A converges enough to produce $2^{m(n-1)/n}$ substrings of length m and V_m is the set of reals B such that Γ^B converges on a substring σ such that $K(\sigma) < \delta|\sigma| - b$. Clearly, both U_m and V_m are Σ^0_1.

As remarked in the proof of Theorem 5.1, for fixed δ, b and n, given a rational $\varepsilon > 0$ we can effectively find an m such that $\mu(G_m) \leq \varepsilon$. In other words, there is a computable f such that $\mu(G_{f(i)}) \leq 2^{-i}$. Then $((U_{f(i)}, V_{f(i)}))_{i\in\omega}$ is a difference test that captures Y. ⊣

Franklin and Ng also provide the following characterization of the difference random reals:

THEOREM 5.8 (Franklin and Ng [6]). *The difference random reals are precisely the incomplete Martin-Löf random reals.*

Since the halting problem has PA degree, any Martin-Löf random real that is not difference random computes a shift-complex sequence, by Corollary 3.9. Together, Theorem 5.7 and Theorem 5.8 imply the following:

COROLLARY 5.9. *A Martin-Löf random real computes a shift-complex sequence if and only if it is complete.*

We will see shortly that there are shift-complex sequences that are not of PA degree, so Corollary 5.9 can be viewed as a generalization of the following well-known result:

THEOREM 5.10 (Stephan [14]). *A Martin-Löf random real is of PA degree if and only if it is complete.*

§6. Extracting randomness from shift-complexity. We turn our attention now to the strength of shift-complex sequences as oracles. In this section we show that not all shift-complex sequences compute Martin-Löf random reals. The plan is similar to Greenberg and Miller's construction in [7] of a real of effective Hausdorff dimension 1 that computes no Martin-Löf random real. The separation is achieved through an analysis of the computational power of slow-growing diagonally noncomputable (DNC) functions.

Recall that the (partial computable) *jump function* J is defined by

$$J(n) := \varphi_n(n).$$

DEFINITION 6.1. DNC is the class of functions $f \in \omega^\omega$ such that for all n, if $n \in \mathrm{dom}(J)$, then $f(n) \neq J(n)$. For $a \geq 2$, DNC_a denotes $\mathrm{DNC} \cap a^\omega$.

DEFINITION 6.2. An *order function* is a function $h : \omega \to \omega \setminus \{0, 1\}$ that is computable, non-decreasing and unbounded. For an order function h, let DNC_h denote $\{f \in \mathrm{DNC} : (\forall n) f(n) < h(n)\}$.

Our main goal in this section is to show that for any $\delta \in (0, 1)$, all sufficiently slow-growing DNC functions compute δ-shift-complex sequences:

THEOREM 6.3. *Fix $\delta \in (0, 1)$. There is an order function h such that every DNC_h function computes a δ-shift-complex sequence.*

Then, to see that for every $\delta \in (0, 1)$, there are δ-shift-complex sequences that compute no Martin-Löf random real, we could appeal to a result by Greenberg and Miller that there are arbitrarily slow-growing DNC functions that compute no Martin-Löf random real:

THEOREM 6.4 (Greenberg and Miller [7]). *For every order function h, there is an $f \in \mathrm{DNC}_h$ that does not compute a Martin-Löf random real.*

However, we can prove something stronger. J. Miller and the author have shown that Theorem 6.4 can be strengthened to hold for a weaker notion of randomness called Kurtz randomness.

DEFINITION 6.5. A real is *Kurtz random* if it is not contained in any Π^0_1 class of measure 0.

Note that Martin-Löf random reals and 1-generics (even weakly 1-generics) are Kurtz random.

THEOREM 6.6 (Khan and Miller [9]). *For every order function h, there is an $f \in \mathrm{DNC}_h$ that does not compute a Kurtz random real.*

Theorem 6.3 and Theorem 6.6 then imply:

THEOREM 6.7. *For every $\delta \in (0,1)$ there exists a δ-shift-complex sequence that does not compute a Kurtz random real.*

Before proving Theorem 6.3 we need a preliminary result.

PROPOSITION 6.8. *Let $S \subseteq 2^{<\omega}$ be computable and suppose that for some $\alpha \in (0,1)$, $|S \cap 2^n| \leq 2^{\alpha n}$ for all n. Then there is a computable $X \in 2^\omega$ that avoids S except for finitely many strings.*

PROOF. Pick a rational $c \in (1/2, 1)$ such that $\log(c) < -\alpha$. Then

$$\sum_{\tau \in S} c^{|\tau|} \leq \sum_{n \in \omega} 2^{\alpha n} c^n = \sum_{n \in \omega} 2^{(\alpha + \log(c))n},$$

where the sum on the right converges geometrically. Let $N \in \omega$ be such that

$$\sum_{n \geq N} 2^{(\alpha + \log(c))n} \leq 2c - 1.$$

Now $S' = S \setminus 2^{<N}$ and c satisfy the hypotheses of Theorem 3.2. By Proposition 3.4, there is a computable $X \in 2^\omega$ that avoids S'. ⊣

If $S \subseteq 2^{<\omega}$ covers $\{\sigma \in 2^{<\omega} : K(\sigma) < \delta|\sigma|\}$, then any sequence which avoids S except for finitely many strings is δ-shift-complex. Relativizing Proposition 6.8 to such an S we obtain the following:

COROLLARY 6.9. *Fix $\delta \in (0,1)$. Suppose $S \subseteq 2^{<\omega}$ contains $\{\sigma \in 2^{<\omega} : K(\sigma) < \delta|\sigma|\}$ and for some $\alpha \in (0,1)$, $|S \cap 2^n| \leq 2^{\alpha n}$ for all n. Then S computes a δ-shift-complex sequence.*

The next step is to show that for every $\delta \in (0,1)$, every sufficiently slow-growing DNC function computes an $S \subseteq 2^{<\omega}$ satisfying the hypotheses of the corollary above.

LEMMA 6.10. *Suppose $\delta \in (0,1)$ and $\alpha \in (\delta, 1)$ are rational. Let $\pi(n) := \lfloor 2^{(\alpha - \delta)n} \rfloor$. Then, uniformly in n, given access to a function $g \in \mathrm{DNC}_{\pi(n)}$, we can compute a set $S_n \subset 2^n$ such that*

1. $|S_n| \leq 2^{\alpha n}$,
2. $\{\sigma \in 2^n : K(\sigma) < \delta n\} \subseteq S_n$.

Moreover, there is a computable function $\theta : \omega \to \omega$ (independent of the oracle g) such that the use of the computation is bounded by $\theta(n)$.

Instead of working directly with DNC functions, we work with a related class. Let $J(n,m)$ denote $\varphi_n(m)$.

DEFINITION 6.11. For $a \geq 2$ and $c > 0$, let $\mathcal{P}_a^c$ be the class of functions $f \in a^\omega$ such that for all n and all $x < c$, if $(n,x) \in \mathrm{dom}(J)$, then $f(n) \neq J(n,x)$.

We use the following fact, originally due to Cenzer and Hinman [1], in the form presented in Greenberg and Miller [7]:

THEOREM 6.12 (Cenzer and Hinman [1]). *For each $a \geq 2$ and $c > 0$, there is a functional Γ such that if $g \in \mathrm{DNC}_a$, then $\Gamma^g \in \mathcal{P}^c_{ac}$. Further, the functional Γ can be obtained effectively from a and c.*

Note that since DNC_a is a Π^0_1 class in a^ω, the functional Γ can be assumed to be total on a^ω, and hence a truth-table functional.

PROOF OF LEMMA 6.10. Let $c = \lfloor 2^{\delta n} \rfloor$. By the preceding theorem, we can obtain $f = \Gamma^g \in \mathcal{P}^c_{c\pi(n)}$. Let B_n denote $\{\sigma \in 2^n : K(\sigma) < \delta n\}$. We construct $S_n \subset 2^n$ by eliminating strings that are not in B_n one at a time. A simple counting argument shows that $|B_n| \leq 2^{\lfloor \delta n \rfloor} \leq c$. The key is that for an appropriately chosen m, B_n can be covered by the values of $J(m, x)$ for $x \leq c$. Since $f(m)$ avoids these values, it picks out a string that is not in B_n.

More formally, for $T \subseteq 2^n$, we can computably find an $m_T \in \omega$ such that $J(m_T, x) = y$ if and only if the x^{th} element seen in an enumeration of B_n is the y^{th} lexicographically least element of T. If $|T| > 2^{\alpha n}$, then

$$|T| > 2^{(\alpha-\delta)n} \cdot 2^{\delta n} \geq \lfloor 2^{(\alpha-\delta)n} \rfloor \lfloor 2^{\delta n} \rfloor = c\pi(n).$$

Therefore, $f(m_T)$ corresponds to an string in $T \setminus B_n$. So let $T_0 = 2^n$ and, for $0 \leq i < 2^n - \lfloor 2^{\alpha n} \rfloor$, let $T_{i+1} = T_i \setminus \sigma_i$, where σ_i is the $f(m_{T_i})^{\text{th}}$ lexicographically least element of T_i. Finally, let $S_n = T_{2^n - \lfloor 2^{\alpha n} \rfloor}$.

Since Γ is a truth-table functional, its use γ is computable. We can define $\theta(n)$ to be

$$\max\{\gamma(m_T) : T \subseteq 2^n, |T| > 2^{\alpha n}\}. \qquad \dashv$$

The uniformity in Lemma 6.10 allows us to string together the constructions of S_n for varying n. At the same time, the fact that as n increases the DNC strength needed to compute S_n *decreases* (π is exponential in n) implies that instead of using the computational power of a function in DNC_a for a fixed $a \in \omega$, we can use a function whose range is allowed to grow unboundedly.

LEMMA 6.13. *For every $\delta \in (0, 1)$, there is an order function h such that every $f \in \mathrm{DNC}_h$ computes a set $S \subseteq 2^{<\omega}$ that contains $\{\sigma \in 2^{<\omega} : K(\sigma) < \delta|\sigma|\}$ and satisfies the condition that for some $\alpha \in (\delta, 1)$, $|S \cap 2^n| \leq 2^{\alpha n}$ for all n.*

PROOF. Without loss of generality, assume δ is rational and choose a rational $\alpha \in (\delta, 1)$. Let θ be the computable bound on the use of the computation in Lemma 6.10. Now, for $i \in \omega$, let $h(i) = \lfloor 2^{(\alpha-\delta)m} \rfloor$, where $m = \min\{n : \theta(n) \geq i\}$. Then h is clearly total, non-decreasing, unbounded, and computable. Moreover, by Lemma 6.10, if $f \in \mathrm{DNC}_h$, then we can use the values of f on the interval $[0, \theta(n)]$ to compute S_n, since $h(i) \leq \lfloor 2^{(\alpha-\delta)n} \rfloor$ for $i \in [0, \theta(n)]$. $\dashv$

To complete the proof of Theorem 6.3, let h be the order function of Lemma 6.13. By Corollary 6.9, every $f \in \mathrm{DNC}_h$ computes a δ-shift-complex sequence.

§7. **Two questions on dimension extraction.** A shift-complex sequence X has positive effective packing dimension, so by a result of Fortnow, Hitchcock, Pavan, Vinochandran, and Wang [5], we know that for every $\varepsilon > 0$, X computes a real Y such that $\mathrm{Dim}(Y) > 1 - \varepsilon$. On the other hand, Conidis [2] has shown that there is a real of positive effective packing dimension that computes no real of effective packing dimension 1. However, the added assumption of shift-complexity might allow us to circumvent the limitation posed by Conidis's result.

QUESTION 7.1 (Kach). Does every shift-complex sequence compute a real of effective packing dimension 1?

We can also ask the analogous question for effective Hausdorff dimension:

QUESTION 7.2. Does every shift-complex sequence compute a real of effective Hausdorff dimension 1? If not, then for every $\varepsilon > 0$, does every shift-complex sequence compute a real of effective Hausdorff dimension $1 - \varepsilon$?

Acknowledgments. The author would like to thank his PhD supervisor, Joseph S. Miller, who introduced him to shift-complex sequences and provided invaluable guidance during the course of the research that went into this paper. The author would also like to thank an anonymous referee for their exceedingly thorough and helpful commentary on the preprint.

REFERENCES

[1] DOUGLAS CENZER and PETER G. HINMAN, *Degrees of difficulty of generalized r.e. separating classes*, ***Archive for Mathematical Logic***, vol. 46 (2008), no. 7-8, pp. 629–647.

[2] CHRIS CONIDIS, *A real of strictly positive effective packing dimension that does not compute a real of effective packing dimension one*, ***The Journal of Symbolic Logic***, vol. 77 (2012), no. 2, pp. 447–474.

[3] RODNEY G. DOWNEY and DENIS R. HIRSCHFELDT, ***Algorithmic randomness and complexity***, Theory and Applications of Computability, Springer, 2010.

[4] BRUNO DURAND, LEONID A. LEVIN, and ALEXANDER SHEN, *Complex tilings*, ***The Journal of Symbolic Logic***, vol. 73 (2008), no. 2, pp. 593–613.

[5] LANCE FORTNOW, JOHN M. HITCHCOCK, A. PAVAN, N. V. VINODCHANDRAN, and FENGMING WANG, *Extracting Kolmogorov complexity with applications to dimension zero-one laws*, ***Information and Computation***, vol. 209 (2011), no. 4, pp. 627–636.

[6] JOHANNA N. Y. FRANKLIN and KENG MENG NG, *Difference randomness*, ***Proceedings of the American Mathematical Society***, vol. 139 (2011), no. 1, pp. 345–360.

[7] NOAM GREENBERG and JOSEPH S. MILLER, *Diagonally non-recursive functions and effective Hausdorff dimension*, ***The Bulletin of the London Mathematical Society***, vol. 43 (2011), no. 4, pp. 636–654.

[8] DENIS HIRSCHFELDT and ASHER KACH, *Shift complex sequences*, Talk presented at the AMS Eastern Sectional Meeting, 2012.

[9] MUSHFEQ KHAN and JOSEPH S. MILLER, *Bushy tree forcing*, in preparation.

[10] Joseph S. Miller, *Two notes on subshifts*, ***Proceedings of the American Mathematical Society***, vol. 140 (2012), no. 5, pp. 1617–1622.

[11] André Nies, ***Computability and randomness***, Oxford Logic Guides, vol. 51, Oxford University Press, Oxford, 2009.

[12] A. Yu. Rumyantsev and M. A. Ushakov, *Forbidden substrings, Kolmogorov complexity and almost periodic sequences*, ***STACS 2006*** (Bruno Durand and Wolfgang Thomas, editors), Lecture Notes in Computer Science, vol. 3884, Springer, 2006, pp. 396–407.

[13] Andrey Yu. Rumyantsev, *Everywhere complex sequences and the probabilistic method*, ***STACS 2011*** (Thomas Schwentick and Christoph Dürr, editors), Leibniz International Proceedings in Informatics (LIPIcs), vol. 9, Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2011, pp. 464–471.

[14] Frank Stephan, *Martin-Löf random and PA-complete sets*, ***Logic Colloquium '02*** (Zoé Maria Chatzidakis, Peter Koepke, and Wolfram Pohlers, editors), Lecture Notes in Logic, vol. 27, Association for Symbolic Logic, La Jolla, CA, 2006, pp. 342–348.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN
MADISON, WI 53706-1388, USA
E-mail: khan@math.wisc.edu