Constant compression and random weights

Wolfgang Merkle, Jason Teutsch

Constant compression and random weights*

Wolfgang Merkle[1] and Jason Teutsch[2]

1,2 Ruprecht-Karls-Universität Heidelberg
Heidelberg, Germany
{merkle|teutsch}@math.uni-heidelberg.de

Abstract

Omega numbers, as considered in algorithmic randomness, are by definition real numbers that are equal to the halting probability of a universal prefix-free Turing machine. Omega numbers are obviously left-r.e., i.e., are effectively approximable from below. Furthermore, among all left-r.e. real numbers in the appropriate range between 0 and 1, the Omega numbers admit well-known characterizations as the ones that are Martin-Löf random, as well as the ones such that any of their effective approximation from below is slower than any other effective approximation from below to any other real, up to a constant factor. In what follows, we obtain a further characterization of Omega numbers in terms of Theta numbers.

Tadaki considered for a given prefix-free Turing machine and some natural number a the set of all strings that are compressed by this machine by at least a bits relative to their length, and he introduced Theta numbers as the weight of sets of this form. He showed that in the case of a universal prefix-free Turing machine any Theta number is an Omega number and he asked whether this implication can be reversed. We answer his question in the affirmative and thus obtain a new characterization of Omega numbers.

In addition to the one-sided case of the set of all strings compressible by at least a certain number a of bits, we consider sets that comprise all strings that are compressible by at least a but no more than b bits, and we call the weight of such a set a two-sided Theta number. We demonstrate that in the case of a universal prefix-free Turing machine, for given a and all sufficiently large b the corresponding two-sided Theta number is again an Omega number. Conversely, any Omega number can be realized as two-sided Theta number for any pair of natural numbers a and $b > a$.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases computational complexity, Kolmogorov complexity, algorithmic randomness, Omega number

Digital Object Identifier 10.4230/LIPIcs.STACS.2012.172

1 Universal prefix-free machines and random reals

An Omega number is the weight of the domain of a universal prefix-free machine U, i.e., a real number of the form $\Omega_U = \sum_{\sigma \in \mathrm{dom}\, U} 2^{-|\sigma|}$. Chaitin [3] introduced Omega numbers and demonstrated, after work of Zvonkin and Levin [10], that Omega numbers admit recursive approximations from below yet feature completely random binary expansions. That is, Omega numbers are left-r.e. and Martin-Löf random. Remarkably, Omega numbers are the only such numbers and therefore characterize the set of reals with these two properties. Calude, Hertling, Khoussainov, Wang [1] and Kučera, Slaman [5] proved this equivalence known as the Kučera-Slaman Theorem [4, p. 410].

* Research supported by Deutsche Forschungsgemeinschaft under grant ME 1806/3-1.

29th Symposium on Theoretical Aspects of Computer Science (STACS'12).
Editors: Christoph Dürr, Thomas Wilke; pp. 172–181

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Tadaki introduced Theta numbers Θ^a_M as the weight of the set of strings that can be compressed by a constant number a of bits relative to their length with respect to the coding given by some prefix-free Turing machine M, i.e.,

$$\Gamma^a_M = \{\sigma \in \{0,1\}^* : (\exists \tau)\, M(\tau) = \sigma \text{ and } |\tau| \leq |\sigma| - a\}, \qquad \Theta^a_M = \sum_{\sigma \in \Gamma^a_M} 2^{-|\sigma|}.$$

Tadaki showed that in the case of a prefix-free universal Turing machine any Theta number is an Omega number and he asked whether this implication can be reversed. We answer his question in the affirmative and obtain this way a new characterization of Omega numbers.

In addition to the one-sided case of the set Γ^a_M of all strings compressible by at least a of bits, we consider the two-sided case of the set $\Gamma^{a\setminus b}_M$ of all strings that are compressible by at least a but no more than b bits. We demonstrate that such sets cannot contain an r.e. set, hence are not r.e., but somewhat surprisingly, in the case of a universal prefix-free machine, for given a and for all sufficiently large b the corresponding two-sided Theta number $\Theta^{a\setminus b}_M$ is left-r.e. and, in fact, is again an Omega number. Conversely, any Omega number can be realized as two-sided Theta number for any pair of natural numbers a and $b > a$.

Note that due to space considerations in the sequel several results are stated without proof.

Notation A STRING is a finite binary sequence, the length of a string σ is denoted by $|\sigma|$, where $|\cdot|$ will also denote cardinality for sets. A set of strings is PREFIX-FREE if no string in the set is a proper prefix of another string in the set. We let $\operatorname{dom} M = \{\sigma : M(\sigma)\downarrow\}$ be the domain of a Turing machine M, where $M(\sigma)\downarrow$ and $M(\sigma)\uparrow$ denote convergence and divergence of the computation of M on input σ. A prefix-free Turing machine or, for short, a PREFIX-FREE MACHINE is a Turing machine that has prefix-free domain. The PREFIX-FREE KOLMOGOROV COMPLEXITY of a string σ with respect to a prefix-free machine M, denoted K_M, is the length of the shortest input to M which results in output σ. A prefix-free machine U is UNIVERSAL if for any other prefix-free machine M, there exists a constant c such that for all strings σ, $\mathrm{K}_U(\sigma) \leq \mathrm{K}_M(\sigma) + c$. Universal prefix-free machines exist [4, 6]. We fix some reference universal prefix-free machine U and write K in place of K_U. Furthermore, let $\mathrm{K}(\sigma|\rho) = \min\{|\tau| : U(1^{|\rho|}0\rho\tau) = \sigma\}$ denote the PREFIX-FREE COMPLEXITY OF σ GIVEN ρ.

We will identify strings and natural numbers via the order morphisms between the length-lexicographical ordering on strings and the usual order on the natural numbers, and accordingly the function K, in addition to strings, may take natural numbers as arguments or even integers, where the latter are viewed as a coded pair of a natural number and the sign. For a natural number n, we let n^* denote a code for n of minimum length, i.e., $U(n^*) = n$ and $|n^*| = \mathrm{K}(n)$, where among all codes of minimum length the code n^* is the one with the least running time on U, breaking ties by choosing the least string in lexicographical order. For $n = 0, 1, \ldots$, we let $\bar{n}$ denote an encoding of the natural number n with respect to U, i.e, $U(\bar{n}) = n$, that has length at most $2\log n + c$ for some constant c [4], where log denotes logarithm to base 2. We choose the mapping $n \mapsto \bar{n}$ to be recursive, while this would not be possible for the mapping $n \mapsto n^*$.

Unless explicitly specified otherwise, the term SEQUENCE refers to an infinite binary sequence. A sequence $x_1x_2\ldots$ can be viewed as the real that has binary expansion $0.x_1x_2\ldots$, and the notation sequence and real will be used interchangeably. A real number $\alpha \in [0,1]$ is called LEFT-R.E. if it is the limit of an effectively given sequence of nonnegative dyadic rationals, i.e., nonnegative rationals with denominators that are a power of 2. For a real α

equal to $0.x_1x_2\ldots$, the string $\alpha \upharpoonright n$ consists of the first n bits $x_1x_2\ldots x_n$ of α after the decimal point. A real α is MARTIN-LÖF RANDOM if there exists a constant c such that $\mathrm{K}(\alpha \upharpoonright n) \geq n - c$ for all n. This definition coincides with our intuition that random objects do not compress too much. A real that is left-r.e. and Martin-Löf random is called an OMEGA NUMBER. For further background on notions discussed in this section, see the monograph by Downey and Hirschfeldt [4], which contains also a detailed account of the Kraft-Chaitin theorem to be used in the sequel.

2 Sets of compressible strings

Kolmogorov complexity comes in several flavors [4]. Besides the prefix-free Kolmogorov complexity K introduced in Section 1, one can consider the PLAIN version defined similarly but without any requirements on machines being prefix-free. The plain Kolmogorov complexity of a string of length n never exceeds n plus an additive constant, and a straightforward combinatorial argument shows that for some positive constant d and all natural numbers a and n, at most a fraction of 2^{-a+d} of all strings of length n have plain Kolmogorov complexity of at most $n - a$ [4, p. 112]. For prefix-free Kolmogorov complexity, the situation is similar but somewhat more involved because the upper bound of n has to be replaced by $n + \mathrm{K}(n)$. Prefix-free Kolmogorov complexity for strings of length n may achieve but never exceeds $n + \mathrm{K}(n)$, up to an additive constant [4, p. 128]. Furthermore, Chaitin's celebrated Counting Theorem asserts that the number of strings describable by codes shorter than this upper bound minus a constant has a simple upper bound reminiscent of the one for plain complexity.

▶ **Counting Theorem** (Chaitin [2, 4, 6]). *For some positive constant d and all natural numbers a and n, it holds that*

$$|\{\sigma \in \{0,1\}^n \colon \mathrm{K}(\sigma) \leq n + \mathrm{K}(n) - a\}| \leq 2^{n-a+d} \,,$$

i.e., at most a fraction of 2^{-a+d} of all strings of length n have prefix-free Kolmogorov complexity of no more than $n + \mathrm{K}(n) - a$.

When working with plain Kolmogorov complexity, it is suggestive to call an n-bit string a-compressible in case the plain Kolmogorov complexity of the string is at most $n - a$ and to call a string compressible in case it is 1-compressible. Following the literature conventions, we extend this notation to the prefix-free setting. Note that indicating compression relative to n and not relative to the upper bound $n + \mathrm{K}(n)$ avoids having to count bits of compression relative to the nonrecursive latter bound. We will, by slight abuse of notation, permit our notation to carry over to negative values of a because some of our results extend to this case.

▶ **Definition 1.** Let a be any integer. A string σ is a-COMPRESSIBLE WITH RESPECT TO A PREFIX-FREE MACHINE M if $\mathrm{K}_M(\sigma) \leq |\sigma| - a$. Furthermore, a string σ is a-COMPRESSIBLE if $\mathrm{K}(\sigma) < |\sigma| - a$.

▶ **Definition 2.** Let M be a prefix-free machine and let a and b be integers. The SET OF a-COMPRESSIBLE STRINGS WITH RESPECT TO M, denoted Γ^a_M, is

$$\Gamma^a_M = \{\sigma \in \{0,1\}^* : (\exists \tau)\, M(\tau) = \sigma \text{ and } |\tau| \leq |\sigma| - a\},$$

and the SET OF $[a,b)$-COMPRESSIBLE STRINGS WITH RESPECT TO M is

$$\Gamma^{a\setminus b}_M = \Gamma^a_M - \Gamma^b_M \,.$$

We will refer to sets of the form Γ^a_M and $\Gamma^{a\setminus b}_M$ as ONE-SIDED and TWO-SIDED GAMMA SETS, respectively, and we will call such Gamma sets UNIVERSAL in case M is a universal prefix-free machine.

By the Counting Theorem, there exists a constant d such that for all integers a and string lengths n,

$$|\Gamma^a_U \cap \{0,1\}^n| \leq 2^{n-\mathrm{K}(n)-a+d} \,. \tag{1}$$

In particular, this shows that for any integer a the fraction of strings of length n that are a-compressible goes to 0 when n goes to infinity.

Miller and Yu [7] refined Chaitin's Counting Theorem [4, Section 3.7]. Using their result, we can improve (1) to lower and upper bounds that match up to a constant factor. We state Miller and Yu's result in a slightly altered form where we replace one occurrence of $\mathrm{K}(\sigma)$ by $\mathrm{K}_U(\sigma)$ for an arbitrary universal prefix-free machine U. One can resolve the differences introduced in our alternate version via appropriately chosen values for the constant d; details are left to the reader.

▶ **Improved Counting Theorem** (Miller and Yu [7]). *Let U be a universal prefix-free machine. There is a constant d such that for all natural numbers c and n it holds that*

$$2^{n-c-\mathrm{K}(c|n^*)-d} \leq |\{\sigma \in \{0,1\}^n : \mathrm{K}_U(\sigma) \leq n + \mathrm{K}(n) - c\}| \leq 2^{n-c-\mathrm{K}(c|n^*)+d} \,.$$

Note that the bounds given by the Improved Counting Theorem are false in general for negative values of c. By the Improved Counting Theorem we obtain in Corollaries 3 and 4 bounds for the number of strings of length n in sets of the form $\Gamma^{a\setminus b}_U$ and $\Gamma^{a\setminus b}_U$ for a universal prefix-free machine U. Proposition 5 then shows that the assertion of Corollary 4 cannot be strengthened to hold for all b instead of just all sufficiently large b.

▶ **Corollary 3.** *Let U be any universal prefix-free machine. There is a constant d such that for all natural numbers a and all n, as well as for all integers a and for all sufficiently large natural numbers n it holds that*

$$2^{n-\mathrm{K}(n)-a-\mathrm{K}(a|n^*)-d} \leq |\Gamma^a_U \cap \{0,1\}^n| \leq 2^{n-\mathrm{K}(n)-a-\mathrm{K}(a|n^*)+d} \,. \tag{2}$$

▶ Remark. Tadaki states the special case of Corollary 3 where a is equal to 1 and attributes this result to Solovay [9, Theorem 5]. For this special case, as with any constant value of a, the additive terms a and $\mathrm{K}(a|n^*)$ in the exponents of the bounding terms in (2) can be subsumed into the constant d.

▶ **Corollary 4.** *Let U be any universal prefix-free machine and let a be any integer. Then for any real $\varepsilon > 0$, for all sufficiently large integers b, and for all n it holds that*

$$(1-\varepsilon)\,|\Gamma^a_U \cap \{0,1\}^n| \leq \left|\Gamma^{a\setminus b}_U \cap \{0,1\}^n\right| \leq |\Gamma^a_U \cap \{0,1\}^n| \,. \tag{3}$$

▶ **Proposition 5.** *For every pair of integers a and b there is a universal prefix-free machine U such that $\Gamma^{a\setminus b}_U$ is empty.*

3 Compressible strings and enumerability

We note that by definition every one-sided Gamma set is r.e., and every two-sided Gamma set is the difference of two r.e. sets, or D.R.E., for short (see the monographs cited in the

references [4, 6, 8] for background on r.e. and d.r.e. sets). Furthermore, in general sets of the form Γ^a_M and $\Gamma^{a\setminus b}_M$ can be rather simple and may for example be empty or may be infinite and recursive, where the latter can be achieved by choosing M to be a prefix-free machine where $M(0^k) = 0^{k+1}$ while M is undefined, otherwise. In contrast to this, complements of one-sided universal Gamma sets cannot even be r.e because any infinite r.e. set must contain highly compressible strings. For two-sided Gamma sets $\Gamma^{a\setminus b}_U$ a similar assertion holds for sufficiently large b according to Proposition 6. We conclude this section by Lemma 7 which provides the technical machinery for Theorems 10 and 11.

▶ **Proposition 6.** *Let U be a universal prefix-free machine, and let a be any integer. For any integer b, the set $\Gamma^{a\setminus b}_U$ does not contain an infinite r.e. set. For almost all integers $b > a$, the complement of the set $\Gamma^{a\setminus b}_U$ is not r.e.*

▶ **Lemma 7.** *Let U be a universal prefix-free machine and let a and b be any integers where $a < b$. Suppose that for each integer t an enumeration without repetitions of the set Γ^t_U is given uniformly effectively in t and let $\sigma_0, \sigma_1, \sigma_2, \ldots$ and $\tau_0, \tau_1, \tau_2, \ldots$ be the corresponding enumerations of Γ^a_U and Γ^b_U, respectively. Furthermore, let d be any natural number and let r be any recursive function. Then for all sufficiently large b there is a strictly increasing recursive function g such that for all i,*

(I) $\big|\sigma_{g(i)}\big| = |\tau_i| - d$,

(II) $g(0) > r(0)$ *and* $g(i+1) > r(g(i))$,

(III) $\sigma_{g(i)} \neq \tau_j$ *for* $j = 0, \ldots, r(i)$.

Proof. Each of the b-compressible strings τ_j occurs exactly once in the sequence $\sigma_0, \sigma_1, \ldots$, thus there is a computable function h such that for all i, each of the strings $\tau_0, \ldots, \tau_i$ occurs among the strings $\sigma_0, \ldots, \sigma_{h(i)}$, hence does not occur among $\sigma_{h(i)+1}, \sigma_{h(i)+2}, \ldots$. For further use note that $h(i) \geq i$.

We define inductively functions γ and g, which *a priori* are not necessarily total. For a start, we set m_0 to $h(r(0)) = \max\{r(0), h(r(0))\}$, let $\gamma(0)$ be the least string of length $|\tau_0| - d$ that differs from σ_0 through σ_{m_0}, and let $g(0)$ be the least (in fact possibly undefined but if defined unique) index j such that $\gamma(0) = \sigma_j$. Assuming that γ and g have already been defined for all arguments up to i, let

$$\begin{aligned} m_{i+1} &= \max\{g(i), r(g(i)), h(r(i))\}, \\ \gamma(i+1) &= \min\Big\{\eta \in \{0,1\}^{|\tau_{i+1}|-d} \colon \eta \neq \sigma_j \text{ for } j \in \{0, \ldots, m_{i+1}\}\Big\}, \\ g(i+1) &= \min\{j \colon \sigma_j = \gamma(i+1)\}, \end{aligned}$$

that is, $\gamma(i+1)$ is the lexicographically least string of length $|\tau_{i+1}| - d$ that differs from all the strings $\sigma_0, \ldots, \sigma_{m_{i+1}}$, while $g(i+1)$ is the index of $\gamma(i+1)$ in the enumeration $\sigma_0, \sigma_1, \ldots$.

Now consider any i such that $\gamma(i)$ and $g(i)$ are both defined. Then assertion (I) holds true by choice of $\gamma(i)$ and because $\gamma(i)$ and $\sigma_{g(i)}$ are the same. Furthermore, assertions (II) and (III) hold true because of $g(i) > m_i$ and because by choice of h and m_i, $\sigma_{g(i)}$ differs from τ_0 through $\tau_{r(i)}$. Since the functions γ and g are partial recursive, in order to prove the lemma, it remains to show that g is total for all sufficiently large b.

By the Counting Theorem, for some n_0 and all $n \geq n_0$ there exists a string of length $n-d$ that is not a-compressible. In case $b > n_0$, the b-compressible strings $\tau_0, \tau_1, \ldots$ must all have length at least n_0, hence when trying to define $\gamma(i+1)$ there will always be a string of length $|\tau_{i+1}| - d$ that differs from the a-compressible strings σ_0 through $\sigma_{m_{i+1}}$. So in case $b > n_0$, the only way g might avoid being total is that there is a least index i such that

the functions g and γ are defined on all values up to i, the string $\gamma(i+1)$ is defined, too, but the value $g(i+1)$ is undefined. That is, the string $\gamma(i+1)$ is defined but does not occur in the enumeration $\sigma_0, \sigma_1, \ldots$ of all a-compressible strings, which we show is impossible.

Consider a prefix-free machine M that assumes its input to be of the form $\bar{a}\bar{b}\rho$ where a and b are integers and ρ is a prefix-free code for some b-compressible string, i.e., $U(\rho) = \tau_i$ for some index i. In case M is able to verify this assumption, M simulates the inductive definition of γ and g in order to compute $\gamma(i+1)$, and outputs $\eta = \gamma(i+1)$. But then there exists a c such that for all large enough b and for an optimal code ρ for τ_i,

$$\mathrm{K}_U(\eta) \leq \left|\bar{a}\bar{b}\rho\right| + c \leq \left|\bar{a}\bar{b}\right| + |\tau_i| - b + c = |\tau_i| - d - a - (b-a) + \left|\bar{a}\bar{b}\right| + c + d \leq |\eta| - a,$$

where the inequalities follow, first, by universality of U, second, by choice of ρ as a code of length at most $|\tau_i| - b$, third, by rearranging terms, and, last, because η has length $|\tau_i| - d$ and because for any b that is large enough the difference $b - a$ will be larger than $\left|\bar{a}\bar{b}\right|$ plus the constant $c + d$. Hence for sufficiently large b, for all i the string $\gamma(i)$ is a-compressible, hence $g(i)$ is defined. ◀

4 Left-r.e. approximations for Theta numbers

In the following definition, we review and slightly extend Tadaki's [9] concept of Theta number, which is central for this exposition.

▶ **Definition 8.** The WEIGHT of a (not necessarily finite) set A of strings is the value of the sum $\sum_{\sigma\in A} 2^{-|\sigma|}$, and the WEIGHT of a singleton string σ is $2^{-|\sigma|}$. For a prefix-free machine M and integers a and b let

$$\Theta^a_M = \sum_{\sigma\in\Gamma^a_M} 2^{-|\sigma|} \quad \text{and} \quad \Theta^{a\setminus b}_M = \sum_{\sigma\in\Gamma^{a\setminus b}_M} 2^{-|\sigma|}$$

be the weights of the set Γ^a_M of a-compressible strings and of the set $\Gamma^{a\setminus b}_M$ of $[a,b)$-compressible strings with respect to M.

We will refer to reals of the form Θ^a_M and $\Theta^{a\setminus b}_M$ as ONE-SIDED and TWO-SIDED THETA NUMBERS, respectively. A Theta number is UNIVERSAL if its underlying prefix-free machine is universal. Note that for any prefix-free machine M and any integers a and b, in case $a \leq b$, we have Γ^b_M is a subset of Γ^a_M and therefore

$$\Theta^{a\setminus b}_M = \Theta^a_M - \Theta^b_M\,,$$

whereas $\Theta^{a\setminus b}_M = 0$, otherwise. Furthermore, for any prefix-free machine M and any integers a and b, the Theta numbers Θ^a_M and $\Theta^{a\setminus b}_M$ are both finite since both can be at most as large as 2^{-a} times the weight of the domain of the prefix-free machine M, where the latter weight is at most 1 by the Kraft inequality, i.e.,

$$\Theta^{a\setminus b}_M \leq \Theta^a_M = \sum_{\tau\in\Gamma^a_M} 2^{-|\tau|} \leq \sum_{\sigma\in\mathrm{dom}\,M} 2^{-(|\sigma|+a)} \leq 2^{-a}\,. \tag{4}$$

As observed by Tadaki [9], the real Θ^1_U and indeed all one-sided Theta numbers, or reals of the form Θ^a_M for integers a, are the weight of some r.e. set, which is equivalent to being left-r.e. Proposition 6, which says that $\Gamma^{a\setminus b}_M$ need not be r.e., now comes back to haunt us because in contrast to the one-sided case, a two-sided Theta number may fail to be left-r.e.

▶ **Proposition 9.** *Let a be any integer. There exists a prefix-free machine M such that for all $b > a$ the real $\Theta_M^{a\setminus b}$ is not left-r.e.*

The following theorem asserts that two-sided Theta numbers $\Theta_U^{a\setminus b}$ are indeed left-r.e for all sufficiently large b in the case of a universal prefix-free machine U. This result comes as a slight surprise since for all sufficiently large b the set $\Gamma_U^{a\setminus b}$ is not r.e. according to Proposition 6.

▶ **Theorem 10.** *Let U be a universal prefix-free machine, and let a be any integer. For all sufficiently large integers b, the real $\Theta_U^{a\setminus b}$ is left-r.e.*

Proof. Apply Lemma 7 to U and a where d is equal to 0 and r is the identity function. Fix any b that is so large that there are enumerations $\sigma_0, \sigma_1, \sigma_2, \dots$ and $\tau_0, \tau_1, \tau_2, \dots$ of Γ_U^a and Γ_U^b, respectively, and a recursive function g as in Lemma 7. Recall that the function g is strictly increasing, hence is one-to-one and its range R is recursive. Then $\Theta_U^{a\setminus b}$ is left-r.e. because we have

$$\begin{aligned}\Theta_U^{a\setminus b} &= \sum_{\sigma\in\Gamma_M^{a\setminus b}} 2^{-|\sigma|} = \sum_{\sigma\in\Gamma_M^a} 2^{-|\sigma|} - \sum_{\tau\in\Gamma_M^b} 2^{-|\tau|}\\ &= \sum_{k\in\mathbb{N}\setminus R} 2^{-|\sigma_k|} + \sum_{k\in\mathbb{N}\cap R} 2^{-|\sigma_k|} - \sum_{k\in\mathbb{N}} 2^{-|\tau_k|}\\ &= \sum_{k\in\mathbb{N}\setminus R} 2^{-|\sigma_k|} + \sum_{k\in\mathbb{N}} \underbrace{2^{-|\sigma_{g(k)}|} - 2^{-|\tau_k|}}_{=0}.\end{aligned}$$

◀

5 Theta numbers and Martin-Löf randomness

Tadaki [9] demonstrated that every one-sided universal Theta number is Martin-Löf random. Using Theorem 10, we extend Tadaki's result to show that two-sided universal Theta numbers are Martin-Löf random. As just mentioned, the first statement in the following theorem is due to Tadaki [9].

▶ **Theorem 11.** *Let U be a universal prefix-free machine and let a be a natural number.*

(I) *The real Θ_U^a is Martin-Löf random.*

(II) *For all sufficiently large natural numbers b, the real $\Theta_U^{a\setminus b}$ is Martin-Löf random.*

Proof of (II). Fix any natural number $b > a$. Assuming that $\Theta_U^{a\setminus b}$ is not Martin-Löf random, we will obtain a contradiction if b is sufficiently large. In order to apply Lemma 7, take $d = 1$ and let r be equal to the identity function. Furthermore, let $\sigma_0, \sigma_1, \sigma_2, \dots$ and $\tau_0, \tau_1, \tau_2, \dots$ be enumerations of Γ_U^a and of Γ_U^b, respectively, as in the assumption of the lemma. Then for sufficiently large b there is a strictly increasing function g as in the lemma, i.e., for all i, the string $\sigma_{g(i)}$ is one bit shorter than the string τ_i and differs from $\tau_0, \tau_1, \dots, \tau_i$. Next let for any natural number s,

$$I_s = \{i \le s \colon \sigma_i \notin \{\tau_0, \dots \tau_s\}\} \quad \text{and} \quad \Theta_{U,s}^{a\setminus b} = \sum_{i\in I_s} 2^{-|\sigma_i|}\,.$$

Observe that the sequence $\{\Theta_{U,s}^{a\setminus b}\}$ converges to $\Theta_U^{a\setminus b}$, but not necessarily monotonically so. Similarly to the one-sided case, let M be a prefix-free machine that, on input η, first tries to compute the string $U(\eta)$ and its length n. If successful, M next searches for the least s such

that the length n initial segment of the binary expansion of $\Theta^{a\setminus b}_{U,s}$ is equal to $U(\eta)$. If such a number s is found, M outputs the least string of length $n-2$ that differs from $\sigma_0, \ldots, \sigma_s$, where such an output string exists for all sufficiently large n by the Counting Theorem.

By letting d_0 be equal to the coding constant for M with respect to U, we can fix a sufficiently large length n such that the following holds. The initial segment of $\Theta^{a\setminus b}_U$ of length n is equal to $U(\eta)$ for some code η of length at most $n-a-d_0-2$, and the string $M(\eta)$ exists, has length $n-2$, and satisfies $\mathrm{K}(M(\eta)) \leq |\eta| + d_0 \leq n-a-2$. It follows that $M(\eta)$ is a-compressible, hence is equal to σ_t for some index $t > s$.

For the length n indicated in the previous paragraph, consider the corresponding values of s, η and t and the corresponding set I_s, as well as the set $I_s^+ = I_s \cup \{t\}$. By choice of η, the set I_s^+ contains only indices of a-compressible strings and the sum of the weights of these strings is strictly larger than $\Theta^{a\setminus b}_U$. More precisely, since $\Theta^{a\setminus b}_{U,s}$ differs from $\Theta^{a\setminus b}_U$ by at most 2^{-n}, we have

$$\sum_{i\in I_s^+} 2^{-|\sigma_i|} = \left(\sum_{i\in I_s} 2^{-|\sigma_i|}\right) + 2^{-|\sigma_t|} = \Theta^{a\setminus b}_{U,s} + 4\cdot 2^{-n} \geq \Theta^{a\setminus b}_U + 3\cdot 2^{-n}\,. \tag{5}$$

Having the weight of strings indexed by I_s^+ to be greater than $\Theta^{a\setminus b}_U$ is not a contradiction because some of these strings may in fact be b-compressible and hence do not contribute to $\Theta^{a\setminus b}_U$. However, whenever a string ρ is b-compressible, i.e., is equal to some string τ_j, then $\sigma_{g(j)}$ is another a-compressible string with strictly greater weight than ρ. The string $\sigma_{g(j)}$ may be b-compressible in turn, in which case there is another a-compressible string of weight strictly larger than $\sigma_{g(j)}$. Iterating this process, we eventually reach a TERMINAL a-compressible string that is not b-compressible and contributes its weight to $\Theta^{a\setminus b}_U$. In the remainder of the proof, we argue that the terminal strings that are reached by such cascades starting from strings with indices in I_s^+ have a total weight that is strictly larger than $\Theta^{a\setminus b}_U$, which is then indeed a contradiction.

Formally, define a partial function h such that σ_i is equal to $\tau_{h(i)}$ in case σ_i is indeed b-compressible, and h is undefined otherwise. Let $f = g \circ h$. Then for all i such that $h(i)$ is defined, we have

$$\sigma_{f(i)} = \sigma_{g(h(i))}, \quad \text{hence} \quad |\sigma_{f(i)}| = |\sigma_i| - 1 \quad \text{by choice of } h \text{ and } g.$$

▶ Claim 1. The function f is one-to-one in the sense that if $f(i)$ and $f(j)$ are both defined and are the same, then i is equal to j.

Proof. In case $f(i) = g(h(i))$ and $f(j) = g(h(j))$ are both defined and are the same, then $h(i)$ and $h(j)$ must both be defined and the same because g is strictly increasing and hence one-to-one. Consequently, $h(i)$ and $h(j)$ are indices of identical strings σ_i and σ_j, hence i and j must be the same. ◀

▶ Claim 2. For all $i \in I_s$, either $f(i)$ is undefined or $s < f(i)$.

Proof. Fix i in I_s. In case $h(i)$ is defined, by definition of h we have $\sigma_i = \tau_{h(i)}$, hence $h(i) > s$ by definition of I_s. Then also $f(i) = g(h(i)) > s$ because g is strictly increasing. ◀

▶ Claim 3. For all i it holds that $f(i) < f(f(i))$ whenever both values are defined.

Proof. It suffices to show that $h(i)$ is strictly less than $h(g(h(i)))$ because g is strictly increasing and maps these two indices to $f(i)$ and $f(f(i))$, respectively. In case the string $\sigma_{g(h(i))}$ occurs among $\tau_0, \tau_1, \ldots$ at all, then the corresponding index $h(g(h(i)))$ must be strictly larger than $h(i)$ because by choice of g, the string $\sigma_{g(h(i))}$ differs from τ_0 through $\tau_{h(i)}$. ◀

For every i there is a maximum natural number m such that $f^{[m]}(i)$ is defined because the strings $\sigma_{f^{[0]}(i)}, \sigma_{f^{[1]}(i)}, \ldots$ are mutually distinct and have all length at most $|\sigma_i|$. Given i and such maximum m, we let the i-CASCADE be the sequence

$$i, f(i), f(f(i)), \ldots, f^{[m]}(i)$$

and we call i, $f^{[m]}(i)$, and m respectively the STARTING POINT, the END POINT, and the LENGTH of this cascade. The minimum possible length of a cascade is 0, in which case starting point and end point coincide. Note that by choice of f, the length of an i-cascade can be equivalently defined as the least k such that $\sigma_{f^{[k]}(i)}$ is not b-compressible, i.e., an index j occurring in a cascade is the end point of the cascade if and only if σ_j is not in Γ^b_U, or equivalently, is in $\Gamma^{a\setminus b}_U$.

▶ Claim 4. If two cascades have the same end point, then the starting point of one cascade occurs in the other.

Proof. For a proof, consider an i-cascade of length k and a j-cascade of length $l \leq k$ that have the same end point. In case the j-cascade has length 0, there is nothing to prove. Otherwise, since f is one-to-one, the indices $f^{[l-1]}(i)$ and $f^{[k-1]}(i)$ must be the same, and by an easy induction argument we obtain

$$i = f^{[0]}(i) = f^{[k-l]}(j)\ . \qquad \blacktriangleleft$$

▶ Claim 5. Any two distinct starting points which belong to I_s have distinct end points for their respective cascades.

Proof. By Claims 2 and 3, the numbers that occur in a cascade that starts at any point in I_s are all strictly larger than s, except for the starting point, which has size at most s. So given two distinct indices $i, j \in I_s$, i cannot occur in the j-cascade and vice versa, hence the cascades starting at i and at j must have distinct end points by Claim 4. $\blacktriangleleft$

Let E be the set of all indices i that are end points of a cascade starting at some index in I^+_s. The cardinality of E is then equal to the cardinality of either I_s or I^+_s since by Claim 5 the end points of the cascades starting at indices in I^+_s are mutually distinct except that there may be a unique index $j \in I_s$ such that the j-cascade and the t-cascade have the same end point. In the latter case, the index t is equal to $f^{[k]}(j)$ for some $k > 0$ by $s < t$ and Claims 2, 3, and 4, hence the length of σ_j is at least $|\sigma_t| + 1$. Furthermore, in case there is such an index j, we have

$$\sum_{i\in E} 2^{-|\sigma_i|} \geq \sum_{i\in I^+_s\setminus\{j\}} 2^{-|\sigma_i|} \geq \Theta^{a\setminus b}_U + 3\cdot 2^{-n} - 2^{-|\sigma_j|} > \Theta^{a\setminus b}_U,$$

where the inequalities hold, first, by choice of the index j and because the strings indexed by a cascade decrease in length, hence increase in weight, second, by (5) and, third, because of $|\sigma_j| \geq |\sigma_t| + 1 = n - 1$. Otherwise, in case the end points of the cascades starting at some index in I^+_s are mutually distinct, we can argue similarly and infer rather directly from (5) that the weight of the strings with index in E is strictly larger than $\Theta^{a\setminus b}_U$. So we obtain in both cases a contradiction to the definition of $\Theta^{a\setminus b}_U$ because the end point of any cascade is the index of a string that is in Γ^a_U but not in Γ^b_U, hence this string contributes its weight to $\Theta^{a\setminus b}_U$. This concludes the proof of Theorem 11. $\blacktriangleleft$

6 Universal Theta numbers and Omega numbers

Finally, we ask which reals can be realized as one-sided or two-sided Theta numbers. Tadaki [9] demonstrates that one-sided universal Theta numbers are always Omega numbers. He then asks whether conversely every Omega number can be realized as one-sided universal Theta number. Similarly, by Theorems 10 and 11, which assert that any two-sided universal Theta number is indeed an Omega number in case the corresponding larger compression bound b is sufficiently large, it is suggesting to ask whether all Omega numbers can be realized as two-sided universal Theta numbers. We give a positive answer to both question in Theorem 12. Together with the results mentioned above this yields a new characterization of the Omega numbers: a real is an Omega number if and only if the real is a one-sided universal Theta number.

▶ **Theorem 12.** *Let a and $b > a$ be natural numbers and let α be a nonnegative left-r.e. Martin-Löf random real where $\alpha < 2^{-a}$. Then there are universal prefix-free machines V and V' such that $\alpha = \Theta_V^{a\backslash b} = \Theta_{V'}^{a}$.*

Acknowledgements The authors are grateful to anonymous referees of the STACS conference for pointing out an erroneous statement of one of the results and for their comments and corrections in general.

References

1 Cristian S. Calude, Peter H. Hertling, Bakhadyr Khoussainov, and Yongge Wang. Recursively enumerable reals and Chaitin Ω numbers. *Theoretical Computer Science*, 255(1-2):125–149, 2001.
2 Gregory J. Chaitin. A theory of program size formally identical to information theory. *Journal of the ACM*, 22:329–340, 1975.
3 Gregory J. Chaitin. Incompleteness theorems for random reals. *Advances in Applied Mathematics*, 8(2):119–146, 1987.
4 Rodney G. Downey and Denis R. Hirschfeldt. *Algorithmic randomness and complexity*. Springer, New York, 2010.
5 Antonín Kučera and Theodore A. Slaman. Randomness and recursive enumerability. *SIAM Journal on Computing*, 31(1):199–211, 2001.
6 Ming Li and Paul Vitányi. *An introduction to Kolmogorov complexity and its applications*. Springer, New York, third edition, 2008.
7 Joseph S. Miller and Liang Yu. Oscillation in the initial segment complexity of random reals. *Advances in Mathematics*, 226(6):4816–4840, 2011.
8 Robert I. Soare. *Recursively enumerable sets and degrees*. Springer-Verlag, Berlin, 1987.
9 Kohtaro Tadaki. A new representation of Chaitin Ω number based on compressible strings. In Cristian Calude, Masami Hagiya, Kenichi Morita, Grzegorz Rozenberg, and Jon Timmis, editors, *Unconventional Computation*, volume 6079 of *Lecture Notes in Computer Science*, pages 127–139. Springer, Berlin Heidelberg, 2010.
10 Alexander K. Zvonkin and Leonid A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83–124, 1970.