*Article*

# Effective Complexity of Stationary Process Realizations

**Nihat Ay** [1,3]**, Markus Müller** [1,2] **and Arleta Szkoła** [1,*]

[1] Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, Leipzig 04103, Germany;
  E-Mail: nay@mis.mpg.de (N.A.); mueller@math.tu-berlin.de (M.M.)

[2] Institute of Mathematics 7-2, Technical University Berlin, Straße des 17. Juni 136, Berlin 10623,
  Germany

[3] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

[*] Author to whom correspondence should be addressed; E-Mail: szkola@mis.mpg.de;
  Tel.: +49-(0)-341-9959-525; Fax: +49-(0)-341-9959-555.

**Abstract:** The concept of effective complexity of an object as the minimal description length of its regularities has been initiated by Gell-Mann and Lloyd. The regularities are modeled by means of ensembles, which is the probability distributions on finite binary strings. In our previous paper [1] we propose a definition of effective complexity in precise terms of algorithmic information theory. Here we investigate the effective complexity of binary strings generated by stationary, in general not computable, processes. We show that under not too strong conditions long typical process realizations are effectively simple. Our results become most transparent in the context of coarse effective complexity which is a modification of the original notion of effective complexity that needs less parameters in its definition. A similar modification of the related concept of sophistication has been suggested by Antunes and Fortnow.

## 1. Introduction

The concept of effective complexity has been initiated by Gell-Mann and Lloyd in [2], see also [3]. The main motivation was to define a complexity measure that distinguishes between regular and random aspects of a given object typically encoded as a binary string. This is in contrast to Kolmogorov complexity which is not sensitive to the source of incompressibility and in this sense fails to capture what is meant by complexity in the common language.

The main idea underlying the concept has been considered at different places in the literature, see [4–10]. It may be summarized as follows. One considers programs computing a given binary string as consisting of two parts: the implementation of an algorithm and a valid input for that algorithm. Then the corresponding measures of complexity refer to the algorithm part.

In [2] the algorithm part has been motivated as a description of a physical theory-represented by a probability distribution on finite binary strings, while the second part has been used to distinguish one among all possible objects contained in the (typical) support of the distribution. Effective complexity is equal to the length of the algorithm/theory part which is minimized over the set of programs that compute the string and that are almost minimal, *i.e.*, their length is close to the Kolmogorov complexity of the string.

In [1] we have proposed a definition of effective complexity in precise terms of algorithmic information theory. Our formalization allows to include the concept into the context of algorithmic statistics, which also deals with two-part codings of binary strings [5]. Instances of corresponding measures of complexity are Kolmogorov minimal sufficient statistics and sophistication [5,10]. Roughly speaking, while Kolmogorov minimal sufficient statistics of a binary string $x$ is the minimal algorithmic statistic of $x$ from the model class of finite sets, and sophistication refers to the model class of total programs, effective complexity mainly coincides with the length of algorithmic statistics of $x$ minimized over computable probability distributions.

More precisely, the minimization domain of effective complexity consists of computable probability distributions with total information which is approximately equal to the Kolmogorov complexity of the string: the tolerance level being specified by a parameter $\Delta$. Total information has been defined by Gell-Mann and Lloyd in [2,3] as the sum of Kolmogorov complexity and Shannon entropy of a given computable ensemble. It is worth mentioning that it is equivalent to the concept of physical entropy introduced by Zurek for large physical systems such as thermodynamic engines [8].

Restricting the minimization domain of effective complexity by intersecting with subsets corresponding to pre-knowledge about the object, which is subjective to the observer, one ends up with a version of effective complexity with constraints. As far as we know, there is no literature other than the papers by Gell-Mann and Lloyd [2], where the idea to incorporate subjective pre-knowledge into the measure of complexity has been considered explicitly.

Compared to the effective complexity without constraints, which we will refer to as plain effective complexity or simply effective complexity, this gives a larger value and is the reason why Gell-Mann and Lloyd suggest to use the constrained version instead of the plain one: "If we impose no other conditions, every entity would come out simple!", see [2], (p. 392). This statement has to be contrasted with the fact that there exist strings with large plain effective complexity, *cf.* Theorem 13 of our previous work [1]. See

also corresponding results in the context of algorithmic statistics and sophistication, Theorem 2.2 in [5] and Theorem 6.5 in [6], respectively. Hence, the above conviction can be substantiated only in a weaker version referring to some typical behaviour. In the present contribution, we find a framework to this end in the form of almost sure statements in terms of probability theory. The focus is on mathematical foundations for the concept of effective complexity. In particular, we extend the analysis of [1] to the context of asymptotic behaviour.

In more detail, we investigate discrete-time stochastic processes with binary state space in the context of effective complexity as it has been defined in [1]. In addition to proving that typical strings are simple with respect to the plain effective complexity, our results also allow a deeper understanding of the dependence of effective complexity on the parameter $\Delta$. Recall that this parameter determines the minimization domain consisting of computable ensembles with a total information that is $\Delta$-close to the Kolmogorov complexity of the string. A corresponding parameter also appears in the context of sophistication and, more generally, algorithmic sufficient statistics [5,10]. Conceptually, it also corresponds to the significance level of Bennett's logical depth defined in [11]. The relation between effective complexity and logical depth has been elaborated in [1].

In [10] Antunes and Fortnow suggested a modification of sophistication called coarse sophistication. In an analogous way, we introduce coarse effective complexity. It modifies the original concept of plain effective complexity by, roughly speaking, incorporating $\Delta$ into the definition as a further minimization argument. As a consequence, the definition becomes independent of the choice of this parameter. Our main results on effective complexity have direct implications on the asymptotic behaviour of coarse effective complexity. In particular, for an arbitrary stationary process the value of coarse effective complexity of a typical finite string is asymptotically upper bounded by any linear function of a string's length.

After fixing notations and the mathematical framework in Section 2, we formulate and prove our main result, Theorem 1, in Section 3. It states that sufficiently long typical strings generated by a stationary process are effectively simple. The proof relies on the observation that the total information of uniform distributions on universally typical subsets is upper bounded by a value that exceeds the Kolmogorov complexity of a typical string by any linear growing amount in the string's length. In Section 4 we introduce the concept of coarse effective complexity. We show that strings of moderate value of coarse effective complexity exist, see Theorem 3, and derive from our main theorem an upper bound on coarse effective complexity of long typical realizations of a stationary process, see Theorem 4. Finally, Section 5 contains some conclusions and an outlook for further analysis of effective complexity in its constrained version.

## 2. Notations and Preliminaries

We denote by $\{0,1\}^*$ the set of finite binary strings, *i.e.*, $\{0,1\}^* = \{\lambda\} \cup \bigcup_{n\in\mathbb{N}}\{0,1\}^n$, where $\lambda$ is the empty string, while the set of doubly infinite sequences $(\ldots, x_{-1}, x_0, x_1, \ldots)$ with $x_i \in \{0,1\}$, $i \in \mathbb{Z}$, is denoted by $\{0,1\}^\infty$. We write $\ell(x)$ for the length of $x \in \{0,1\}^*$. Finite blocks $x_m^n = (x_m, x_{m+1}, \ldots, x_n)$, $m \leq n$, of $x \in \{0,1\}^\infty$ are elements of $\{0,1\}^*$ of length $\ell(x_m^n) = n - m + 1$. We may identify them with cylinder sets $[x_m^n] := \{y \in \{0,1\}^\infty : y_i = x_i, m \leq i \leq n\}$. In a similar fashion, strings $x \in \{0,1\}^*$ are associated to cylinder sets of the form $[x] := \{y \in \{0,1\}^\infty : y_i = x_i, 1 \leq i \leq \ell(x)\}$. The

$\sigma$-algebra on $\{0,1\}^\infty$ generated by the cylinder sets $[x_m^n]$, $m, n \in \mathbb{Z}$, $m \leq n$, is denoted by $\Sigma$. We write $\mathcal{T}(\{0,1\}^\infty)$ for the convex set of probability measures on $(\{0,1\}^\infty, \Sigma)$, which are invariant with respect to the left-shift $T$ on $\{0,1\}^\infty$. The subset of ergodic $T$-invariant probability measures, *i.e.*, the extremal points in $\mathcal{T}(\{0,1\}^\infty)$, is denoted by $\mathcal{E}(\{0,1\}^\infty)$.

Let $P \in \mathcal{T}(\{0,1\}^\infty)$. The random variables $X_i$, $i \in \mathbb{Z}$, given by the coordinate projections $X_i(x) := x_i$, $x \in \{0,1\}^\infty$, respectively, represent a stationary process with values in $\{0,1\}$. Typical outcomes of such stochastic processes are the main focus in the present paper. The goal is to estimate their effective complexity. We will refer to elements of $\mathcal{T}(\{0,1\}^\infty)$ as stationary probability measures and stationary stochastic processes interchangeably.

Adopting the setup of [1] as far as possible, we refer to probability distributions on $\{0,1\}^*$ as *ensembles*.

For each $n \in \mathbb{N}$ we identify the joint distribution (alternatively called $n$-block distribution) $P^{(n)}$ of $n$ successive outcomes $(X_1, X_2, \ldots, X_n)$ of a stationary process with an ensembles $\mathbb{E}_n$ on $\{0,1\}^*$ through the relation: $\mathbb{E}_n(x) = P^{(n)}(x)$ if $\ell(x) = n$ and $\mathbb{E}_n(x) = 0$ otherwise.

Recall the definition of *prefix Kolmogorov complexity* $K(x)$ of a binary string $x \in \{0,1\}^*$:

$$K(x) := \min\{\ell(p): U(p) = x\}$$

where $U$ is an arbitrary but fixed universal prefix computer. For details concerning the basics as well as deeper results on Kolmogorov complexity theory we refer to the book by Li and Vitányi [12].

We call an ensemble *computable* if there exists a program for the universal computer $U$ that, given $x \in \{0,1\}^*$ and $m \in \mathbb{N}$ as inputs, computes an approximation of the probability $\mathbb{E}(x)$ with accuracy of at least $2^{-m}$.

In [1] we have introduced an extension of the notion of Kolmogorov complexity to the case of computable ensembles $\mathbb{E}$ with computable and finite entropies $H(\mathbb{E})$. Here we mean by entropy $H(\mathbb{E})$ the Shannon entropy of the probability distribution $\mathbb{E}$ defined by $-\sum_{x \in \{0,1\}^*} \mathbb{E}(x) \log \mathbb{E}(x)$. Note that a computable ensemble does not necessarily have a computable entropy, such that the corresponding requirement is a restriction, see [1] for details. In what follows a distinction only between general ensembles and computable ones with computable and finite entropies is drawn. We will refer to the latter ones as computable for short.

The *Kolmogorov complexity* $K(\mathbb{E})$ of a computable ensemble $\mathbb{E}$ is defined as the length of the shortest computer program that, given $x \in \{0,1\}^*$ and $m \in \mathbb{N}$ as inputs, outputs both $\mathbb{E}(x)$ and $H(\mathbb{E})$ with an accuracy of at least $2^{-m}$.

Additionally, we need to define computability of stochastic processes. The following definition is a reformulation of the notion of a "computable measure" in [12].

A stationary process $P$ is called *computable* if there exists a program $p \in \{0,1\}^*$ for a universal computer $U$ that, given $x \in \{0,1\}^*$ and $m \in \mathbb{N}$ as inputs, computes the probability $P([x])$ up to accuracy $2^{-m}$.

## 3.  Effective Complexity of Stationary Processes

The goal is to show that under not too strong conditions long typical samples of stationary processes are effectively simple. Before we make rigorous statements we need a number of definitions. The first ones we adopt from our previous paper [1].

Let $\delta \geq 0$. We say that an ensemble $\mathbb{E}$ is $\delta$-*typical for a string* $x \in \{0,1\}^*$, or alternatively, we call $x$ $\delta$-typical for $\mathbb{E}$, if the Shannon entropy $H(\mathbb{E})$ of $\mathbb{E}$ is finite and

$$- \log \mathbb{E}(x) \leq H(\mathbb{E})(1+\delta)$$

In particular, the special case of an equidistributed ensemble is $\delta$-typical for all strings in the support and any $\delta \geq 0$ The *total information* $\Sigma(\mathbb{E})$ of a computable ensemble $\mathbb{E}$ is defined by

$$\Sigma(\mathbb{E}) := H(\mathbb{E}) + K(\mathbb{E})$$

For a motivation of the two definitions above, see [1].

Let $\delta \geq 0$ and $\Delta > 0$. *Effective complexity* $\mathcal{E}_{\delta,\Delta}(x)$ of a finite string $x \in \{0,1\}^*$ is defined by

$$\mathcal{E}_{\delta,\Delta}(x) := \min\{K(\mathbb{E}) : \ \mathbb{E} \in \mathcal{P}_{\delta,\Delta}(x)\}$$

where $\mathcal{P}_{\delta,\Delta}(x)$ denotes the *minimization domain* associated to $x$:

$$\mathcal{P}_{\delta,\Delta}(x) := \{\mathbb{E} : \ \mathbb{E} \text{ computable ensemble}, \mathbb{E} \ \delta\text{-typical for } x, \ \Sigma(\mathbb{E}) \leq K(x) + \Delta\} \tag{1}$$

We refer to elements of $\mathcal{P}_{\delta,\Delta}(x)$ as *effective ensembles* for $x$.

Taking the point of view of [2], which was reviewed in [1], effective ensembles represent theories that are judged to be good explanations for the appearance of $x$.

The more general notion of *effective complexity with constraints* has been suggested in [2] mainly to circumvent problems of plain effective complexity. We have discussed them shortly in the Introduction. The main idea is that the constraints reflect some pre-knowledge about the possible theory for $x$. In [1] we have proposed a formalization of the constrained version in the following manner:

$$\mathcal{E}_{\delta,\Delta}(x \mid \mathcal{C}) := \min\{K(\mathbb{E}) : \ \mathbb{E} \in \mathcal{P}_{\delta,\Delta}(x), \mathbb{E} \in \mathcal{C}\}$$

where $\mathcal{C}$ is a subset of $\mathcal{P}(\{0,1\}^*)$. Note that with $\mathcal{C} = \mathcal{P}(\{0,1\}^*)$ we have $\mathcal{E}_{\delta,\Delta}(x) = \mathcal{E}_{\delta,\Delta}(x|\mathcal{C})$, for all $x \in \{0,1\}^*$.

In what follows other essential concepts are that of typical and/or universally typical subsets.

Let $P$ be a $T$-invariant probability measure on $(\{0,1\}^\infty, \Sigma)$. We call a sequence of subsets $M_n \in \Sigma$, $n \in \mathbb{N}$, *P-typical* if

$$\lim_{n\to\infty} P(M_n) = 1$$

We call $(M_n)$ *strongly P-typical* if for $P$-almost all $x$ there exists an $N_x \in \mathbb{N}$ such that

$$x \in M_n \qquad \text{for every } n \geq N_x$$

The above notions of typicality apply naturally to sequences $M_n \subseteq \{0,1\}^n$, $n \in \mathbb{N}$, if we identify subsets $M_n$, $n \in \mathbb{N}$, with cylinder sets $[M_n] \subseteq \{0,1\}^\infty$, respectively.

Let $\Lambda$ be a set of stationary processes with values in $\{0,1\}$. We call $M_n \subseteq \{0,1\}^n$, $n \in \mathbb{N}$, *universally typical for* $\Lambda$ if the sequence is $P$-typical for every $P \in \Lambda$, *i.e.*, $\lim_{n\to\infty} P^{(n)}(M_n) = 1$. We call the sequence *universally strongly typical for* $\Lambda$ if it is strongly $P$-typical for every $P \in \Lambda$.

For sets $\Lambda_r \subseteq \mathcal{E}(\{0,1\}^\infty)$ consisting of ergodic processes with entropy rate upper bounded by $r > 0$ there exist universally typical subsets $T_{r,n} \subseteq \{0,1\}^n$ with

$$|T_{r,n}| \leq 2^{rn} \tag{2}$$

for all $n \in \mathbb{N}$. Moreover, there are methods to construct such sequences of universally typical subsets for $\Lambda_r$. We will apply the Lempel-Ziv algorithm in the construction procedure below. This famous algorithm represents a universal sequential data compression, see [13–15]. The main point is that all we need to know about an ergodic process $P$ is its entropy rate $h_P$. This allows to prove the following theorem for stationary, in general *not computable* processes.

**Theorem 1.** Let $P$ be a stationary process, $\delta \geq 0$, $\epsilon > 0$ and $\Delta_n = \epsilon n$. Then $P$ is effectively simple in the sense that for $P$-almost every $x$,

$$\mathcal{E}_{\delta,\Delta_n}(x_1^n) \overset{+}{<} \log n + \mathcal{O}(\log\log n) \tag{3}$$

*Proof.* Firstly, assume that $P$ is an ergodic process with entropy rate $h_P$. We construct universally typical subsets $T_{r,n} \subseteq \{0,1\}^n$, $n \in \mathbb{N}$, such that for an appropriate choice of the parameter $r = r(h_P, \epsilon)$ the total information of the uniform distributions $\mathbb{E}_{r,n}$ on $T_{r,n} \subseteq \{0,1\}^n$, respectively, is upper-bounded by $K(x_1^n) + \Delta_n$ of $P$-almost every $x \in \{0,1\}^\infty$ and for $n$ large enough. Hence the Kolmogorov complexity of $\mathbb{E}_{r,n}$, which is approximately upper bounded by $\log n$, gives an estimate from above on the effective complexity $\mathcal{E}_{\delta,\Delta_n}(x_1^n)$ of sufficiently long $P$-typical strings $x_1^n$.

First, let $r > 0$ be arbitrary and define for each $n \in \mathbb{N}$ the subset $T_{r,n} \subseteq \{0,1\}^n$ as the set consisting of all binary strings $x_1^n$ which are mapped by the Lempel-Ziv (LZ) algorithm to a code word of length $\ell_{LZ}(x_1^n)$ lower than $nr$. Then the sequence $T_{r,n}$, $n \in \mathbb{N}$, is universally typical for the set $\Lambda_r$ of ergodic processes with entropy rates lower than $r$.

Recall the following remarkable property of the LZ algorithm: For every ergodic process $Q$ with entropy rate $h_Q$ it holds $\lim_{n\to\infty} \frac{1}{n}\ell_{LZ}(x_1^n) = h_Q$ for $Q$-almost all $x \in \{0,1\}^\infty$. This implies that indeed the subsets $T_{r,n}$ as constructed above are typical for any ergodic $Q$ with $h_Q < r$.

The upper bound (2) on the size $|T_{r,n}|$ follows from the fact that the LZ algorithm is a faithful coder. Hence the codeword lengths satisfy the Kraft inequality:

$$\begin{aligned}
1 &\geq \sum_{x_1^n \in \{0,1\}^n} 2^{-\ell_{LZ}(x_1^n)} \geq \sum_{x_1^n \in T_{r,n}} 2^{-\ell_{LZ}(x_1^n)} \\
&\geq \sum_{x_1^n \in T_{r,n}} 2^{-nr} = |T_{r,n}| 2^{-nr}
\end{aligned}$$

Next, we show that if $r$ is chosen to be a positive rational number satisfying $0 < r - h_P < \epsilon/4$ then for $P$-almost every $x \in \{0,1\}^\infty$ there exists an $N_x \in \mathbb{N}$ such that

$$\Sigma(\mathbb{E}_{r,n}) \leq K(x_1^n) + \Delta_n \qquad \text{for every } n \geq N_x \tag{4}$$

where again $\mathbb{E}_{r,n}$ denotes the uniform distribution on the universally typical subset $T_{r,n}$. First note that for all $n \in \mathbb{N}$

$$H(\mathbb{E}_{r,n}) = \log|T_{r,n}| \leq rn$$

Secondly, we prove that there is a constant $c \in \mathbb{N}$ such that for all $n \in \mathbb{N}$

$$K(\mathbb{E}_{r,n}) \leq K(n) + K(r) + c \tag{5}$$

This is derived from the existence of a program $p$ of length $c$ which expects as inputs $n \in \mathbb{N}$, $r \in \mathbb{Q}$ and $x \in \{0,1\}^*$ and outputs the value $\frac{1}{|T_{r,n}|}$ if $x \in T_{r,n} \subseteq \{0,1\}^n$ and 0 otherwise. Thus for fixed inputs $n$ and $r$ it gives a description of the uniform distribution $\mathbb{E}_{r,n}$ on $T_{r,n}$.

Indeed, $p$ may be constructed on the base of a program $p_{LZ}$ implementing the Lempel-Ziv (LZ) algorithm on the given reference universal computer $U$. For given inputs $n$ and $r$ let $p$ apply $p_{LZ}$ as a subroutine in order to determine elements of $T_{r,n}$. Then for fixed $n \in \mathbb{N}$ the number $|T_{r,n}|$ and hence the probability value $1/|T_{r,n}|$ of each $x \in T_{r,n}$ may be calculated easily.

To specify $r \in \mathbb{Q}$ and $n \in \mathbb{N}$ a number of $K(r) + K(n)$ bits is sufficient. With $c = \ell(p)$ the estimate (5) follows.

Next, fix an $N \in \mathbb{N}$ such that $K(n) + K(r) + c \leq \frac{\epsilon}{4}n$ for all $n \geq N$. Then

$$\Sigma(\mathbb{E}_{r,n}) \leq nr + K(n) + K(r) + c \leq n\left(r + \frac{\epsilon}{4}\right) \leq n\left(h + \frac{\epsilon}{2}\right) \tag{6}$$

where the last inequality holds by assumption $r - h < \frac{\epsilon}{4}$. According to the theorem by Brudno, see [16], for $P$-almost all $x$ there exists an $N_{x,\epsilon} \in \mathbb{N}$ such that $K(x_1^n) \geq n(h - \frac{\epsilon}{2})$ for all $n \geq N_{x,\epsilon}$. It follows for $\Delta_n = \epsilon n$

$$K(x_1^n) + \Delta_n \geq n\left(h + \frac{\epsilon}{2}\right), \qquad n \geq N_{x,\epsilon} \tag{7}$$

Relations (6) and (7) together imply (4) for $P$-almost all $x$ and $n \geq N_x := \max\{N_{x,\epsilon}, N\}$. It follows that $P$-almost surely the effective complexity $\mathcal{E}_{\delta,\Delta_n}(x_1^n)$ is upper bounded by the Kolmogorov complexity of $\mathbb{E}_{r,n}$ for $n \geq N_x$:

$$\begin{aligned}
\mathcal{E}_{\delta,\Delta_n}(x_1^n) &\leq K(\mathbb{E}_{r,n}) \leq K(n) + K(r) + c \\
&\overset{+}{<} \log n + \mathcal{O}(\log\log n)
\end{aligned}$$

Now let $P$ be an arbitrary stationary process. Recall that there is a unique ergodic decomposition of $P$

$$P = \int_{\mathcal{E}(\{0,1\}^\infty)} Q\, d\mu(Q)$$

Moreover, to $P$-almost every $x \in \{0,1\}^\infty$ we may associate an ergodic component $Q_x$ of $P$ such that $x$ is a typical element of $Q_x$. Then there exists an $N_{x,\epsilon}$ such that

$$K(x_1^n) + \Delta_n \geq n\left(h_x + \frac{\epsilon}{2}\right), \qquad n \geq N_{x,\epsilon}$$

where $h_x$ denotes the entropy rate of $Q_x$. Hence the proof for the stationary case reduces to the ergodic situation considered in the first part above.

Finally, we remark that Theorem 1 applies to the large class of stationary processes. This covers, in particular, the case of independent identically distributed processes, which represent the simplest and at the same time one of the best studied classes of stochastic processes. Further, our theorem is valid for stationary Markov chains. These represent another important and rather feasible class of stationary processes.

## 4. Coarse Effective Complexity

Our main result becomes most transparent if presented in the context of *coarse effective complexity*. This is a modification of plain effective complexity which incorporates the parameter $\Delta$ as a penalty into the original formula. It is inspired by a corresponding modification of sophistication, called coarse sophistication, which has been introduced by Antunes and Fortnow in [10].

Let $\delta \geq 0$. The *coarse effective complexity* $\mathcal{E}_\delta(x)$ of a finite string $x \in \{0,1\}^*$ is defined by

$$\mathcal{E}_\delta(x) := \min\{K(\mathbb{E}) + \Sigma(\mathbb{E}) - K(x) : \mathbb{E} \text{ is computable ensemble, } \mathbb{E} \text{ is } \delta - \text{ typical for } x\}$$

The term $\Sigma(\mathbb{E}) - K(x)$ accounts for the exact value by which the total information of an ensemble $\mathbb{E}$ exceeds the Kolmogorov complexity of $x$. By definition of total information $\Sigma(\mathbb{E})$ an equivalent expression for $\mathcal{E}_\delta(x)$ reads:

$$\mathcal{E}_\delta(x) = \min\{2K(\mathbb{E}) + H(\mathbb{E}) - K(x) : \mathbb{E} \text{ is computable ensemble, } \mathbb{E} \text{ is } \delta - \text{ typical for } x\}$$

We derive the basic properties of coarse effective complexity similarly as it has been done in [10] in the context of coarse sophistication. That is, firstly, in the proposition below, we prove an upper bound on coarse effective complexity. Secondly, we show existence of strings which are close to saturate this bound.

**Proposition 2.** Let $\delta \geq 0$. There is a constant $c$ such that for all $x \in \{0,1\}^*$ we have

$$\mathcal{E}_\delta(x) \leq \frac{n}{2} + \log n + c \tag{8}$$

where $n = \ell(x)$.

*Proof.* Suppose that $K(x) \leq \frac{n}{2} + \log n$. Let $\mathbb{E}_x$ denote the ensemble with $\mathbb{E}(x) = 1$ and $\mathbb{E}(y) = 0$ for $y \neq x$. Note that $\mathbb{E}_x$ is trivially $\delta$-typical for $x$ for any $\delta \geq 0$ and obviously $H(\mathbb{E}_x) = 0$. Moreover, there is a constant $c_1$ such that it holds $K(\mathbb{E}_x) \leq K(x) + c_1$. This implies the upper bound

$$\begin{aligned} \mathcal{E}_\delta(x) &\leq 2K(\mathbb{E}_x) + 0 - K(x) \\ &\leq K(x) + 2c_1 \\ &\leq \frac{n}{2} + \log n + 2c_1 \end{aligned}$$

where the last line holds by assumption.

Now, suppose that $K(x) > \frac{n}{2} + \log n$. Let $\mathbb{E}_n$ be the ensemble on $\{0,1\}^*$ given by $\mathbb{E}_n(y) = \frac{1}{2^n}$ for all $y$ with $\ell(y) = n$ and vanishing elsewhere. Then $H(\mathbb{E}_n) = n$ and there exists a constant $c_2$, independent of $n$, such that $K(\mathbb{E}_n) \leq \log n + c_2$. It follows

$$\begin{aligned} \mathcal{E}_\delta(x) &\leq 2\log n + 2c_2 + n - K(x) \\ &\leq \frac{n}{2} + \log n + 2c_2 \end{aligned}$$

where, again, the second line holds by assumption on $K(x)$. Setting $c := \max\{2c_1, 2c_2\}$ completes the proof.

**Theorem 3.** Let $\delta \geq 0$. For every sufficiently large $n \in \mathbb{N}$ there exists a string $x \in \{0, 1\}^n$ with

$$\mathcal{E}_\delta(x) \geq (1 - 3\delta)\frac{n}{2} - (2 + 3\delta) \log n - 2 \log \log n + C \tag{9}$$

where $C$ is a global constant.

*Proof.* For $x \in \{0, 1\}^*$ and $\Delta \geq 0$ denote by $\mathbb{E}_x^\Delta$ the minimal ensemble associated to $\mathcal{E}_{\delta,\Delta}(x)$. Due to Lemma 22 in [1] for every $\epsilon > 0$ there exists a subset $S_x^\Delta$ of $\{0, 1\}^*$ such that

$$\log |S_x^\Delta| \leq H(\mathbb{E}_x^\Delta)(1 + \delta) + \epsilon \tag{10}$$
$$K(S_x^\Delta) \leq K(\mathbb{E}_x^\Delta) + c_1 \tag{11}$$

where $c_1$ is a global constant. In [1] we have proven the relation

$$K(x|S_x^\Delta, K(S_x^\Delta)) \geq \frac{\log |S_x^\Delta|}{1 + \delta} - \log n - 2 \log \log n - \Lambda_\Delta \tag{12}$$

which holds for arbitrary $x \in \{0, 1\}^n$, $n \in \mathbb{N}$. The term $\Lambda_\Delta$ is constant in $x \in \{0, 1\}^*$ and monotonically increasing in $\Delta$, *cf.* (32) in [1]. Now, let $K_n := \max\{K(t)| t \in \{0, 1\}^n\}$ and define

$$k := n - \delta(K_n + \Delta_n + \epsilon) + \log n + 2 \log \log n - \Lambda_{\Delta_n} - c_2$$

where $\Delta_n := \frac{n}{2} + \log n + c$ is the upper bound on $\mathcal{E}_\delta(x)$ obtained in the previous proposition, and $c_2$ is a global constant from Theorem IV.2 in [5], see also Lemma 12 in [1]. If $n$ is large enough then $0 < k < n$ holds, and Theorem IV.2 in [5] applies: There is a string $x_k \in \{0, 1\}^n$ such that

$$K(x_k|S, K(S)) < \log |S| - n - k + c_2 \tag{13}$$

for every set $S \ni x_k$ with $K(S) < k - c_3$, where $c_3$ is another global constant. Let $\mathbb{E}_x$ denote the minimizing ensemble associated to coarse effective complexity $\mathcal{E}_\delta(x)$ and $\Delta_x := K(\mathbb{E}_x) + H(\mathbb{E}_x) - K(x)$ such that $\mathcal{E}_\delta(x) = K(\mathbb{E}_x) + \Delta_x$. Further, define $S_x := S_x^{\Delta_x}$. We have the inequality

$$
\begin{aligned}
-\delta(K_n + \Delta_n + \epsilon) &\leq -\delta(K(x_k) + \Delta_x + \epsilon) \\
&\leq -\delta \left( H(\mathbb{E}_{x_k}) + \epsilon \right) \\
&\leq -\delta \left( H(\mathbb{E}_{x_k}) + \frac{\epsilon}{1 + \delta} \right) \\
&= \frac{-\delta}{1 + \delta} \left( H(\mathbb{E}_{x_k})(1 + \delta) + \epsilon \right) \\
&\leq \left( \frac{1}{1 + \delta} - 1 \right) \log |S_{x_k}|
\end{aligned}
$$

where the last upper bound holds by (10). Now suppose that $K(S_{x_k}) < k - c_1$. Then

$$
\begin{aligned}
K(x_k | S_{x_k}, K(S_{x_k})) \quad &< \quad \log |S_{x_k}| - n + k + c_2 \\
&\leq \quad \log |S_{x_k}| - \log n - 2 \log \log n \\
&\qquad -\Lambda_{\Delta_n} - \delta(K_n + \Delta_n + \epsilon) \\
&\leq \quad \frac{\log |S_{x_k}|}{1 + \delta} - \log n - 2 \log \log n \\
&\qquad -\Lambda_{\Delta_n} \\
&\leq \quad \frac{\log |S_{x_k}|}{1 + \delta} - \log n - 2 \log \log n \\
&\qquad -\Lambda_{\Delta_{x_k}}
\end{aligned}
$$

But the strict inequality is a contradiction to (12). Hence our assumption must be false and we instead have $K(S_{x_k}) \geq k - c_3$. By $\mathcal{E}_\delta(x) = K(\mathbb{E}_x) + \Delta_x$ and using both (11) and the bound $K_n \leq n + 2 \log n + \gamma$, where $\gamma$ is a global constant, we finally obtain

$$
\begin{aligned}
\mathcal{E}_\delta(x_k) \quad &= \quad K(\mathbb{E}_{x_k}) + \Delta_{x_k} \\
&\geq \quad K(S_{x_k}) - c_1 + \Delta_{x_k} \\
&\geq \quad k - c_3 - c_1 + \Delta_{x_k} \\
&\geq \quad n - \delta \left( \frac{3}{2} n + 3 \log n + \gamma + c + \epsilon \right) - \log n \\
&\qquad -2 \log \log n - \frac{n}{2} - \log n - c - c_2 - 1 - c_3 - c_1 \\
&= \quad (1 - 3\delta) \frac{n}{2} - (2 + 3\delta) \log n - 2 \log \log n + C
\end{aligned}
$$

where $C := -\delta(\gamma + \epsilon) - 1 - (1 + \delta)c - c_1 - c_2 - c_3$.

Although, according to the above theorem, for arbitrary large $n$ the existence of strings of length $n$ with moderate coarse effective complexity is ensured, the coarse effective complexity of sufficiently long prefixes of a typical stationary process realization becomes small. This is a direct implication of Theorem 1.

**Theorem 4.** Let $P$ be a stationary process, $\delta \geq 0$ and $\epsilon > 0$. Then for $P$-almost every $x$

$$
\mathcal{E}_\delta(x_1^n) \leq \epsilon n + \log n + \mathcal{O}(\log \log n) \tag{14}
$$

*Proof.* By definition of coarse effective complexity, it holds $\mathcal{E}_\delta(x) \leq \Delta + \mathcal{E}_{\delta,\Delta}(x)$, for all $x \in \{0,1\}^*$ and $\Delta > 0$. We set $\Delta_n = \epsilon n$. Then the conditions of Theorem 1 are satisfied, and applying (3), we arrive at (14).

## 5. Conclusions

In this contribution, we studied the notion of plain effective complexity, which is assigned to a given string, within the context of an underlying stochastic process as model of the string generating mechanism. In [1] we have shown that strings which are called "non-stochastic" in the context of Kolmogorov minimal sufficient statistics have large value of plain effective complexity. The existence of

such strings has been proven by Gács, Tromp and Vitányi in [5]. Here, our aim was to understand how properties of the stochastic process such as ergodicity and stationarity influence the effective complexity of corresponding typical realizations. Is it possible that the prefixes of a typical process realization represent a sequence of finite strings in increasing length $n$ that eventually have a high or moderate value of effective complexity? Our main theorem refers to stationary and in general *non-computable* processes. It proves that modelling the regularities of strings by computable ensembles with total information that is allowed to excess the string's Kolmogorov complexity up to a linearly growing amount $\epsilon n$ with an *arbitrary small* $\epsilon > 0$ is sufficient for typically generating non-complex strings.

The value $\epsilon n$ plays the role of a parameter in the concept of effective complexity. In order to have a notion that is independent of this parameter we introduced coarse effective complexity. It corresponds to coarse sophistication introduced by Antunes and Fortnow in [10] and modifies effective complexity by incorporating the parameter as a further minimization argument. Our result on effective complexity has a direct implication on the asymptotic behaviour of coarse effective complexity of typical realizations of a stationary process. The main statement in this context demonstrates the utility of the linear parameter scaling which we have considered. Moreover, it allows to analyze the interplay between the complexity of a stochastic process and the complexity of its typical realizations. In particular, it demonstrates that, in order to have a notion of effective complexity that also reflects the complexity of a stochastic process, further modifications of plain effective complexity are necessary, for instance introduction of appropriate constraints. This possibility is in line with Gell-Mann and Lloyd's suggestion in [2] which we discussed in the Introduction.

Finally, we point out that continuing our previous work [1] we have formulated our results for the concept of effective complexity only. However, in line with the general equivalence statements obtained in the literature, *cf.* Section V in [6] or Lemma 20 in [1], it should be possible to reformulate our main theorem in the more general context of algorithmic statistics. Indeed, our upper bound on effective complexity of typical process realizations is derived in terms of computable ensembles that are uniform distributions on finite sets (universally typical subsets). This demonstrates the close relation in particular to the concept of Kolmogorov minimal sufficient statistics which refers to the model class of finite sets.

## References

1. Ay, N.; Müller, M.; Szkoła, A. Effective complexity and its relation to logical depth. *IEEE Trans. Inform. Theory* **2010**, *56*, 4593–4607.
2. Gell-Mann, M.; Lloyd, S. Effective complexity. In *Nonextensive Entropy-Interdisciplinary Applications*; Gell-Mann, M., Tsallis, C., Eds.; Oxford University Press: New York, NY, USA, 2004; pp. 387–398.

3. Gell-Mann, M.; Lloyd, S. Information measures, effective complexity, and total information. *Complexity* **1996**, *2*, 44–52.

4. Rissanen, J. Stochastic complexity in statistical inquiry. In *Series in Computer Science 15*; World scientific publishing Co.: Singapore, 1988.

5. Gács, P.; Tromp, J.T.; Vitányi, P.M. Algorithmic statistics. *IEEE Trans. Inform. Theory* **2001**, *47*, 2443–2463.

6. Vitányi, P.M. Meaningful information. *IEEE Trans. Inform. Thoery* **2006**, *52*, 4617–4626.

7. Koppel, M. Structure. In *The Universal Turing Machine: A Half-Century Survey*; Herken, R., Ed.; Oxford University Press: Oxford, UK, 1988; pp. 235–252.

8. Zurek, W.H. Algorithmic randomness and physical entropy. *Phys. Rev. A* **1989**, *40*, 4731–4751.

9. Vereshchagin, N.K.; Vitányi, P.M.B. Kolmogorov's structure function and model selection. *IEEE Trans. Inform. Theory* **2004**, *50*, 3265–3290.

10. Antunes, L.; Fortnow, L. Sophistication revisited. *Theory Comput. Syst.* **2009**, *45*, 150–161.

11. Bennett, C. Logical depth and physical complexity. In *The Universal Turing machine-a Half-Century Survey*; Herken, R., Ed.; Oxford University Press: Oxford, UK, 1988.

12. Li, M.; Vitanyi, P. *An Introduction to Kolmogorov Complexity and Its Applications*; 2nd Ed.; Springer-Verlag: New York, NY, USA, 1997.

13. Kieffer, J. A unified approach to weak universal source coding. *IEEE Trans. Inform. Theory* **1978**, *24*, 674–682.

14. Ziv, J. Coding of sources with unknown statistics-I: Probability of encoding error. *IEEE Trans. Inform. Theory* **1972**, *18*, 384–389.

15. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: New York, NY, USA, 1991.

16. Brudno, A.A. Entropy and the complexity of the trajectories of a dynamical system. *Trans. Moscow Math. Soc.* **1983**, *2*, 127–151.