A Note on the Kolmogorov Data Complexity and Nonuniform Logical Definitions*

Jerzy Tyszkiewicz

Mathematische Grundlagen der Informatik,
RWTH Aachen, D-52074 Aachen, Germany.
jurek@informatik.rwth-aachen.de

Finite model theory; Descriptive complexity; Kolmogorov complexity; Combinatorial problems.

1 Introduction

Descriptive complexity is the theory emerging from investigating how difficult is to *describe* a given problem in a given logical formalism, in contrary to the classical complexity theory question, how difficult it is to decide this problem in a given computational model. Although these approaches seem quite different, they are sometimes even equivalent: the famous and seminal for descriptive complexity result of Fagin [4] states, that a problem, i.e., a class of finite structures, is computable in NP iff it can be defined by a sentence of existential second order logic Σ^1_1.

In the descriptive complexity the resources used to define problems, and thus to measure their difficulty, are e.g. kind of higher-order-order constructs used in the description. If we insist on using only first-order logic, this can be e.g. the number of variables/quantifiers. However, as first-order logic is known to be of severely limited power in the finite, one has often to use nonuniform approach, similar to this used in circuit complexity, and allow problems to be described by sequences $\{\varphi_n\}_{n\in\mathbf{N}}$ of sentences, where φ_n is intended to define the problem among structures of cardinality n. In this way a mere first-order-inexpressibility result can possibly be converted into a precise statement, how far it is from being definable. The well-known result of Cai, Fürer and Immerman asserts e.g. that the isomorphism of graphs requires cn variables for some $c > 0$ (they can be requantified as often as necessary), even if all counting quantifiers are allowed. A bit less cryptic is the result of Dawar [2], using essentially the same technique, that any first-order definition of 3-colourability of n-element graphs requires $\sqrt{n}$ variables to be used.

Another example: if we are going to define some PSPACE-complete problem nonuniformly in first-order logic, we have to use a sequence $\{\varphi_n\}_{n\in\mathbf{N}}$ complicated enough to have the problem: *given* $\mathfrak{A}$, *determine* $\varphi_{|\mathfrak{A}|}$ *and verify whether* $\mathfrak{A} \models \varphi_{|\mathfrak{A}|}$ PSPACE-complete—recall that for constant sentence sequences it is even in LOGSPACE.

In this paper we consider another resource: the quantity of information, included in the problem, which is made formal using Kolmogorov complexity. Thus the nonuniform definition must include the same amount of information, and this manifests in the length of sentences forming the definition.

*Research supported by the German Science Foundation DFG.

Applying this method, we reprove the nonexpressibility results of Hella, Luosto and Väänänen [5] and Hella, Kolaitis and Luosto [6] about hierarchies of generalized (so called Lindström) quantifiers, replacing their counting arguments by Kolmogorov complexity. This yields not only a cleaner and more intuitive proof, but, as a by-product, a sharp estimate of the length of sentences defining more complicated Lindström quantifiers in the terms of simpler ones.
The method we use seems interesting on its own, and can possibly be used in other contexts as well, so we present in more detail than necessary just to prove our theorem.

2 Kolmogorov Data Complexity

2.1 The overall picture

Any given logic L, if restricted to its sentences, can be represented in the finite structures in a form of an infinite binary matrix, whose rows correspond to all sentences of L and columns to all finite structures. A 1 in row n and column m means that $\mathfrak{A}_m \models \varphi_n$, and 0 that $\mathfrak{A}_m \not\models \varphi_n$.

	$\mathfrak{A}_0$	$\mathfrak{A}_1$	$\mathfrak{A}_2$	$\dots$
φ_0	0	1	0	$\dots$
φ_1	1	0	1	$\dots$
φ_2	0	1	1	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Figure 1: Logic represented as a binary matrix

Many notions have been considered to characterize the expressive power of logics in this context. The absolute expressive power is the set of all rows of the matrix: if for two logics L and L' these sets are equal, then L and L' are of equal expressive power, in the sense that for every sentence $\varphi \in L$ there is a sentence $\varphi' \in L'$ such that $\varphi \leftrightarrow \varphi'$ is true in all finite models, and *vice versa:* for every $\varphi' \in L'$ there is $\varphi \in L$ such that $\varphi \leftrightarrow \varphi'$ in all finite models.
Other notions are often *a priori* weaker than the absolute one. Typically they rely on specific properties possessed by the sets of rows and/or columns of the matrix, the properties indicating somehow the strength or weakness of the expressive power.
Let us give three examples, the first of which is classical:
1. Computational complexity of rows. The indication of weakness of one logic would be that every row is in some low complexity class, say LOGSPACE. An indication of strength of another logic could be that there is a PSPACE-complete row. We can immediately deduce that the two logics are of different expressive power[1]. Using this kind of approach, we are working in the area of so called descriptive complexity. 2. Kolmogorov complexity of columns. Although it is not as easy to see like in the previous case, a properly chosen kind of Kolmogorov complexity of columns of the matrix gives rise to an expressiveness measure, as the author has shown in [8]. 3. Kolmogorov complexity of rows. This is the method used in this paper.
After Vardi [9] notions referring to rows of the matrix are said to be of *data type*, and those referring to columns to be of *expression type.* It is therefore e.g. legitimate to call the notion in Item 2. above Kolmogorov expression complexity and the last on ethe *Kolmogorov data complexity,* KD in short.

[1]Note that this kind of argument can lead to separations depending upon open problems in complexity theory.

In this paper, considering the Kolmogorov complexity of rows, we aim not only at nonexpressibility results per se, but at lower bounds concerning non-uniform definitions. And this tool allows us deducting lower bounds concerning length of formulas in the definition.

2.2 Logic and Kolmogorov complexity

Proviso. $\mathbf{N}$ is the set of nonnegative integers, identified with the set $\{0,1\}^*$ of finite binary strings, ordered first by length, and then lexicographically. Thus 0 is the empty word. We will use $|x|$ to denote the length of the word x.
We often use asymptotic notation in its standard meaning, such as O "big oh", o "small oh", etc. $\log n$ throughout the paper stands for the greatest natural m such that $2^m \le n$. We write $\exp(n)$ for 2^n.
Finally, for notational's convenience we assume $n = \{0, \ldots, n-1\}$.

Logic and structures. In this paper we will deal with finite structures. We fix an infinite signature ω, consisting of infinitely many relation symbols of each finite arity. A finite structure $\mathfrak{A}$ over this signature is a structure whose universe A is an initial segment of natural numbers, and in which only finitely many relations are nonempty. The cardinality of A is denoted, somehow nonconsequently, by $|\mathfrak{A}|$. $\mathbf{Fin}(\omega)$ is the set of all *isomorphism types* of such structures. If σ is a subsignature of ω, $\mathbf{Fin}(\sigma) \subseteq \mathbf{Fin}(\omega)$ is the set of all isomorphism types of structures such that $R = \emptyset$ for $R \notin \sigma$. For simplicity we will assume that members of $\mathbf{Fin}(\omega)$ are structures themselves, rather than formally equivalence classes of the isomorphism relation. Functions are in our model represented as restricted relations, and constants as 0-ary functions.
We assume $\mathbf{Fin}(\omega)$ to be ordered in a recursive way by $\preceq$, the particular choice of the ordering relation being immaterial for the moment.
We assume the first-order logic to be known. The only convention we must adopt is the length of formulae: they are written using the standard syntax, with the fixed alphabet consisting of $(,), \exists, \forall, \neg, \vee, \wedge, \leftrightarrow, x, _0, _1$. Variables are constructed using $x, _0$ and $_1$, the latter two serving as binary digits to create subscript numbers to variables. The length $|\varphi|$ of φ is then just the number of symbols appearing in it.
Sometimes, when writing down a formula, we use other names of variables as well, treating them as notational abbreviations, to make the result more readable.
We write $\varphi[\mathfrak{A}] \in \{0,1\}$ for the truth value of a sentence $\varphi \in L$ in $\mathfrak{A} \in \mathbf{Fin}(\omega)$. Sometimes we don't write $\varphi[\mathfrak{A}] = 1$, but traditionally $\mathfrak{A} \models \varphi$ instead.

Kolmogorov complexity. We recall briefly the main definitions and notions of Kolmogorov complexity, assuming notation from the Li and Vitányi's book [7]. This book contains a detailed introduction of this notion.
Let $\phi(\cdot,\cdot)$ be a *universal* partial recursive function $\phi : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}^*$. The *Kolmogorov complexity of a string $x \in \{0,1\}^*$ relative to a string $y \in \{0,1\}^*$* is

$$C(x|y) = \min\{|z| : \phi(z, y) = x\}.$$

$C(x|y)$ says how many bits we must add to y in order to describe x uniquely, where the method of understanding descriptions is given by ϕ. It can be shown, that this value does not change by more than an additive constant, if another universal function in place of ϕ is used.

Proposition 1 *For every $y \in \{0,1\}^*$ and arbitrary $n \in \mathbf{N}$ there is a string $x \in \{0,1\}^n$ such that $C(x|y) \ge n$.*

The above proposition is often rephrased as "there are incompressible strings", i.e., strings which cannot be described in any way shorter then the string itself. Such incompressible objects are often used as "difficult" inputs in lower bound proofs of various kinds. Many examples can be found in the already mentioned book by Li and Vitányi [7]. We will see another application of this idea in this paper.
We need a small additional technicality about Kolmogorov complexity: It is possible to define a pairing function $\langle \_, \_ \rangle$ for binary strings such that $|\langle x, y\rangle| = |x| + |y| + O(\log \min(|x|, |y|)$.

2.3 The definition of KD

If $\mathcal{A} \subseteq \mathbf{Fin}(\omega)$ is finite, then the binary string $\langle \varphi[\mathfrak{A}]\rangle_{\mathfrak{A}\in\mathcal{A}}$ we denote $\varphi[\mathcal{A}]$, by analogy to $\varphi[\mathfrak{A}]$ used for a single structure. The set $\mathcal{A}$ itself is represented as a binary string, resulting from the characteristic function of $\mathcal{A}$ in $\mathbf{Fin}(\omega)$ by cutting all 0's after the last 1 in this sequence.
For each row in the matrix, say corresponding to $\varphi \in L$, we consider the function

$$\mathrm{KD}_\varphi : \mathcal{P}_{fin}(\mathbf{Fin}(\omega)) \to \mathbf{N}$$

defined as follows:

$$\mathrm{KD}_\varphi(\mathcal{A}) = C(\varphi[\mathcal{A}] \mid \mathcal{A}).$$

The function measures how much information we need in order to evaluate the sentence φ in models in $\mathcal{A}$, when $\mathcal{A}$ is given. Certainly if our logic has a fixed recursive semantics, or, in other words, if the relation $\models$ between models and structures is recursive, then this quantity remains uniformly bounded for all $\mathcal{A}$, and the bound depends on φ only. But if there is no recursive semantics, it can be used to measure the complexity of logics.

2.4 The invariance of KD

We have to show that the definitions we have given are robust in a reasonable sense. I.e., the values of KD do not depend on the choices we have made: actually only the choice of the ordering $\preceq$ of $\mathbf{Fin}(\omega)$ matters for it. And it appears the values do not change by *more than an additive constant,* when the ordering of $\mathbf{Fin}(\omega)$ changes. On the other hand, the Kolmogorov complexity of strings can itself change by an additive constant, depending on the choice of the universal partial-recursive function, so the additional indeterminacy introduced by our choices doesn't spoil more than has been already spoiled by the indeterminacy of the Kolmogorov complexity itself.

Theorem 2 *If $\preceq_1$ and $\preceq_2$ are two different recursive orderings of $\mathbf{Fin}(\omega)$, and KD_1 and KD_2 two associated Kolmogorov data complexity functions, then $|\mathrm{KD}_1 - \mathrm{KD}_2| \leq c$ for some constant c.*

Proof. Since both $\preceq_1$ and $\preceq_2$ are recursive, the (unique) bijection α mapping n onto the $\preceq_2$-number of the $\preceq_1$-nth structure in $\mathbf{Fin}(\omega)$ is recursive, and its inverse α^{-1} is recursive, too. Now given $\varphi \in L$ and $\mathcal{A} \in \mathcal{P}_{fin}(\mathbf{Fin}(\omega))$, we can translate the strings $\varphi[\mathcal{A}]$ according to $\preceq_1$ and $\preceq_2$ into one another by repeated applications of either α or α^{-1} to the indices of structures in $\mathcal{A}$ and then sorting the string according to the obtained values. Application of such a translation does not increase the length of shortest programs by more than an additive constant. ■
Note the reason we need to have this theorem: it is not the case, that "complicated" ordering of $\mathcal{A}$ can require longer program to reconstruct $\varphi[\mathcal{A}]$, than a "simpler" ordering, since $\mathcal{A}$ together with ordering is always given to the reconstruction program. It is the other way round:

"complicated" ordering of $\mathcal{A}$ could accidentally contain some information about $\varphi[\mathcal{A}]$, and we have had to make sure, that this amount of accidental information is small and doesn't really affect the results.
The next theorem follows directly from the definitions, but it can be seen as the basis of what we do in this paper, therefore we give the proof. It establishes that KD is an invariant of expressive power of logics.

Theorem 3 *Suppose $\mathcal{A}$ is a nonempty finite subset of* $\mathbf{Fin}(\omega)$. *Let φ, ψ be two sentences such that $\mathcal{A} \models \varphi \leftrightarrow \psi$. Then* $\mathrm{KD}_\varphi(\mathcal{A}) = \mathrm{KD}_\psi(\mathcal{A})$.

Proof. If $\mathrm{KD}_\varphi(\mathcal{A}) \neq \mathrm{KD}_\psi(\mathcal{A})$, i.e., $C(\varphi[\mathcal{A}] \mid \mathcal{A}) \neq C(\psi[\mathcal{A}] \mid \mathcal{A})$, then in particular $\varphi[\mathcal{A}] \neq \psi[\mathcal{A}]$, which means $\mathcal{A} \not\models \varphi \leftrightarrow \psi$. ■
Although this theorem isn't particularly deep, it says that our method makes sense. Another nice property of this fact, that some weaker forms of equivalence of formulas, like "equivalence in great majority of models" lead to other versions of it, with the requirement $\mathrm{KD}_\varphi(\mathcal{A}) = \mathrm{KD}_\psi(\mathcal{A})$ weakened to "the difference between $\mathrm{KD}_\varphi(\mathcal{A})$ and $\mathrm{KD}_\psi(\mathcal{A})$ is not too large".

Theorem 4 *Suppose $\mathcal{A}_n$ is a sequence of nonempty finite subsets of* $\mathbf{Fin}(\omega)$. *Let φ_n, ψ_n be two sequences of formulas such that*

$$\lim_{n\to\infty} \frac{|\{\mathfrak{A} \in \mathcal{A}_n \mid \mathfrak{A} \models \varphi_n \leftrightarrow \psi_n\}|}{|\mathcal{A}_n|} = 1. \tag{1}$$

Then

$$|\mathrm{KD}_{\varphi_n}(\mathcal{A}_n) - \mathrm{KD}_{\psi_n}(\mathcal{A}_n)| = o(|\mathcal{A}_n|).$$

Proof. First note the full symmetry of the problem concerning φ's and ψ's. So if we prove just one of the two inequalities hidden in the absolute value inequality, we are done.
Let $a_n = |\mathcal{A}_n|$. The string $\varphi_n[\mathcal{A}_n]$ can be described by $\psi_n[\mathcal{A}_n]$ plus the information where the strings differ. Let the number of places where they differ be f_n. Then this information can be represented as a binary string of length a_n with f_n many ones, marking places of difference. To provide this information, it is enough to give the lexicographic number of it among all strings of this length with *at most* that many ones—note that we cannot assume f_n to be given. Equation (1) means that $f_n/a_n \to 0$ when $n \to \infty$. In the following calculation we skip subscripts in a_n and f_n, treating n as fixed.
There are $\binom{a}{i}$ binary strings of length a with precisely $i \leq f$ 1's. The number of bits we will need is then

$$\log(\sum_{i\leq f} \binom{a}{i}) \leq \log(f \binom{a}{f}),$$

Indeed, for $i \leq f < a/2$, and this is the case almost always, the value of $\binom{a}{i}$ grows monotonically with the growth of i. For the same reason we may assume that $f \to \infty$. Now

$$\begin{aligned}\binom{a}{f} &= \frac{a!}{f!(a-f)!} \\ &= (1+o(1))\frac{a^a e^{-a}\sqrt{2\pi a}}{f^f e^{-f}\sqrt{2\pi f}(a-f)^{(a-f)}e^{f-a}\sqrt{2\pi(a-f)}} \quad \text{[by Striling's formula]}\end{aligned}$$

$$
\begin{aligned}
&= (1+o(1))\frac{a^a}{f^f\sqrt{2\pi f}(a-f)^{(a-f)}} \\
&= \frac{1}{\sqrt{2\pi f}}\cdot\frac{a^f}{f^f}\cdot\frac{a^{a-f}}{(a-f)^{a-f}} \\
&= \frac{1}{\sqrt{2\pi f}}\cdot\epsilon^{-\epsilon a}\cdot(1-\epsilon)^{(\epsilon-1)a}. \quad \text{[we substitute } f := \epsilon a\text{]}
\end{aligned}
$$

Now the upper bound of $\log f\binom{a}{f}$ we need is

$$
\begin{aligned}
\log f\binom{a}{f} &\le \log(f/\sqrt{2\pi f}) + (-\epsilon a)\log\epsilon + (\epsilon-1)a\log(1-\epsilon) \\
&= \frac{1}{2}\log f + a(-\epsilon\log\epsilon + (\epsilon-1)\log(1-\epsilon)) + O(1) \\
&= O(\log a) + o(a),
\end{aligned}
$$

because $\lim_{\epsilon\to 0^+}\epsilon\log\epsilon + (1-\epsilon)\log(1-\epsilon) = 0$.
Now the string $\varphi_n[\mathcal{A}_n]$, given n, can be described by $\psi_n[\mathcal{A}_n]$ plus a string of $o(a)$ additional bits, plus $O(\log a)$ bit separator, which finishes the proof. ∎

3 Application: Hierarchies of Lindström Quantifiers

3.1 Lindström Quantifiers

For a textbook introduction of Lindström quantifiers, see e.g. the book by Ebbinghaus and Flum [3].
Let σ be a finite subsignature of ω with r_1 unary symbols, r_2 binary symbols, …, r_s s-ary symbols, and no symbols of higher arities. A *Lindström quantifier Q of type* $(r_1, r_2, \ldots, r_s)$ is a subset $Q \subseteq \mathbf{Fin}(\sigma)$. Now we order the types of quantifiers: let for $\sigma = (r_1, r_2, \ldots, r_s)$ th symbol cp_σ denote the polynomial $\sum_{i=1}^{s} r_i \cdot x^i$. Now the type σ' precedes the type σ iff $(\mathrm{cp}_{\sigma'}(n) - \mathrm{cp}_\sigma(n)) \to_{n\to\infty} \infty$.
Rather than defining the introduction of the quantifier Q quite formally, we do it on an example of type $(1, 2)$, which is then easily extendable to the general case. We won't distinguish Q from its syntactic counterpart.
To fix the attention, we choose the set Q to consist of structures which are (by necessity) directed graphs with an unary relation, and such that the vertices in the unary relation form a maximal clique in the graph. The way it is used in logics is as follows: Given a logic L, we define $L(Q)$, the extension of L by the Lindström quantifier Q. The syntax of L is extended by adding to the formula formation rules of L the following formation rule: if $\varphi(x, y, \bar{u})$ and $\psi(z, \bar{u})$ are formulas of $L(Q)$, then the expression $Qx, y, z(\psi(z, \bar{u}), \varphi(x, y, \bar{u}))$ is a formula of $L(Q)$. x, y, z are the variables bound by the quantifier, and they are necessarily all distinct. $\psi(z, \bar{u}), \varphi(x, y, \bar{u})$ are the formulas the quantifier is applied to.
The semantics is as follows:
Suppose $\mathfrak{A} \in \mathbf{Fin}$ and $\bar{d} \in A^k$. Then $\mathfrak{A}, \bar{u} : \bar{d} \models Qx, y, z(\psi(z, \bar{u}), \varphi(x, y, \bar{u}))$ iff the structure $\langle A, \{c \in A : \psi(c, \bar{d})\}, \{(a, b) \in A^2 : \varphi(b, \bar{d})\}\rangle$ is in Q, i.e., if, assuming the values $\bar{d}$ of $\bar{u}$ as parameters, the subset of A defined by ψ is a maximal clique in the directed graph defined by φ.
It remains to fix the rule concerning length of formulas from $\mathrm{FO}(Q)$: we adopt the rule, that for each Q we introduce a new letter to our alphabet which has been used for describing first order

formulae, but they *don't count* when measuring the length of formulae. What counts, however, are all the parentheses, variables, formulas, etc., which appear right after Q, as well as the other new symbol we use: ,.
We need an information about the cardinalities of sets $\mathbf{Fin}_n(\sigma) = \{\mathfrak{A} \in \mathbf{Fin}(\sigma) : |\mathfrak{A}| = n\}$. The following well known theorem of Fagin [4] gives an asymptotic estimate of this value:

Theorem 5 *For any sequence $\{\sigma_n\}_{n\in\mathbf{N}}$ of non-unary signatures with fixed maximal arity holds $|\mathbf{Fin}_n(\sigma)| = (1 + o(1))\mathrm{cp}_{\sigma_n}(n)/n!$.* ■

The value $|\mathbf{Fin}_n(\sigma)|$ we denote by $\mathrm{ns}_\sigma(n)$.

Lemma 6 *For every nonunary type σ there is a Lindström quantifier Q of type σ and a sentence $\varphi \in \mathrm{FO}(Q)$ with*

$$\mathrm{KD}_\varphi(\mathbf{Fin}_n(\sigma)) \geq \mathrm{ns}_\sigma(n)/n! - O(1).$$

Proof. First we define classes $Q_n \subseteq \mathbf{Fin}_n(\sigma)$. Then we will set $Q = \bigcup_{n\in\mathbf{N}} Q_n$.
Let us fix a binary string w_n of length $|\mathbf{Fin}_n(\sigma)|$ such that $C(w_n \mid n) \geq |w_n|$. This means that w_n cannot be reconstructed from any string shorter than $|w_n|$, even if n is known. Now let $\mathfrak{A} \in Q_n$ iff $\mathfrak{A}$ is i-th in the ordering of $\mathbf{Fin}_n(\sigma)$ and the i-th bit of w_n is 1.
Consider the sentence $\varphi \equiv Q\bar{x}(R_1, \ldots, R_n)$ obtained by applying the quantifier to the signature relations from σ.
Given n, the string w_n can be easily reconstructed from $\varphi[\mathbf{Fin}_n(\sigma)]$ as follows: for each $i = 0, \ldots, n-1$ we have $\mathfrak{A}_i \models \varphi$ iff the i-th bit of w_n is 1, where $\mathfrak{A}_i$ is the i-th structure in $\mathbf{Fin}_n(\sigma)$.
This immediately implies the thesis. ■
And now the first hierarchy result, extending that of Hella, Luosto and Väänänen [5]:

Theorem 7 *For any nonunary type σ there exists a Lindström quantifier Q of this type such that if $\{\varphi_n\}_{n\in\mathbf{N}}$ is a sequence of formulas of* FO *extended by all Lindström quantifiers of types lower than the type of Q, and for each n the formula φ_n defines Q in all finite structures $\mathfrak{A}$ of cardinality n, then $\liminf_{n\to\infty} \frac{|\varphi_n|}{n \log n} > \frac{1}{5}$.*

Proof. Let $\sigma = (r_1, r_2, \ldots, r_s)$. According to the last lemma there is a quantifier Q of type σ and a sentence φ of $\mathrm{FO}(Q)$ with

$$\mathrm{KD}_\varphi(\mathbf{Fin}_n(\sigma) = \mathrm{ns}_\sigma(n)/n! - O(1) = (1 + o(1))\mathrm{cp}_\sigma(n)/n!. \tag{2}$$

Suppose to the contrary that there is a sequence $\{\varphi_n\}_{n\in\mathbf{N}}$ of sentences of $\mathrm{FO}(\mathbf{Q})$ such that $\mathbf{Fin}_n(\sigma) \models \varphi_n \leftrightarrow \varphi$, where $\mathbf{Q}$ is the collection of all quantifiers of type lower than r, and such that $|\varphi_n| \leq \frac{1}{4.5} n \log n$ for almost all n.
Note that any occurrence of a quantifier binding m formulas must have length at least $\frac{1}{2} m \log m$. Indeed, by syntactical necessity all variables bound by this quantifier are different, there are at least m of them, so they are indexed by at least m pairwise distinct binary sequences. The estimate follows now by an elementary calculation.
By the estimate we have just made, no quantifier occurring in φ_n can bind more than, say, $\frac{1}{2}n$ formulas, therefore its type must be as follows for some $1 \leq i \leq s$: $(\leq \frac{1}{2}n, \ldots, \leq \frac{1}{2}n, \leq r_i - 1, r_{i+1}, \ldots, r_s)$.
Fix a large n. We get

$$\begin{aligned}
\mathrm{KD}_{\varphi_n}(\mathbf{Fin}_n(\sigma)) &\leq 4n\log n + O(n^{s+2}) + \\
&\quad 4n\log n\sum_{i=1}^{s}(1+o(1))\exp\Big(\sum_{j>i} r_j n^j + (r_i-1)n^i + \frac{1}{2}n\sum_{j<i} n^j\Big)/n! \\
&\leq O(n^{s+2}) + 5n\log n\sum_{i=1}^{s}\exp\Big(\sum_{j>i} r_j n^j + (r_i-\frac{1}{2})n^i + \frac{1}{2}\sum_{j<i-1} n^{j+1}\Big)/n! \\
&\leq 4n\log n + 5n\log n\sum_{i=1}^{s-1}\exp\Big(\sum_{j>i} r_j n^j + (r_i-\frac{1}{3})n^i\Big)/n!.
\end{aligned}$$

Indeed, it is possible to reconstruct $\varphi_n[\mathbf{Fin}_n(\sigma)]$, given n, from φ_n itself plus descriptions of all at most $|\varphi_n|$ quantifiers appearing in it. More precisely, from each quantifier we need only the information about structures of size n which belong to it. To make this description a binary word, each such class can be described by a string of bits of length $\mathrm{ns}_\tau(n)$, where τ is the type of the quantifier, telling which structures are in it, and which aren't. Additionally we need bits necessary to separate all these strings, and their number is logarithmic in the length of strings to be separated times the number of them. Applying the estimates of of $\mathrm{ns}_\tau(n)$ from Theorem 5 we get the first inequality. (The mysterious constant 4 comes from the fact, that Kolmogorov complexity is about binary strings, whereas our formulas are written using 12 symbols—the conversion is done by describing each of them by four bits.) The remaining two inequalities follow then by simple calculation.
Now, using (2), we immediately see that

$$\mathrm{KD}_{\varphi_n}(\mathbf{Fin}_n(\sigma)) = o(\mathrm{KD}_{\varphi}(\mathbf{Fin}_n(\sigma))). \tag{3}$$

Thus $\mathrm{KD}_{\varphi}(\mathbf{Fin}_n(\sigma)) \neq \mathrm{KD}_{\varphi_n}(\mathbf{Fin}_n(\sigma))$ if n is sufficiently large, and we get a contradiction by Theorem 3. ■

We can apply the same idea to prove lower bounds concerning *almost sure* definability. We achieve this by application of Theorem 4. This allows us to improve the second hierarchy theorem for Lindström quantifiers, due to Hella, Kolaitis and Luosto [6]. We say, that the sequence $\{\varphi_n\}_{n\in\mathbf{N}}$ *is almost surely equivalent to* φ iff

$$\lim_{n\to\infty}\frac{|\mathfrak{A}\in\mathbf{Fin}_n(\sigma) : \mathfrak{A}\models\varphi\leftrightarrow\varphi_n\}|}{|\mathbf{Fin}_n(\sigma)|} = 1,$$

where σ is the signature made of all symbols appearing in φ.

Theorem 8 *For any nonunary type σ there exists a Lindström quantifier Q of this type and a sentence $\varphi\in\mathrm{FO}(Q)$ such that if $\{\varphi_n\}_{n\in\mathbf{N}}$ is a sequence of formulas of* FO *extended by all Lindström quantifiers of types lower than σ, and $\{\varphi_n\}_{n\in\mathbf{N}}$ is almost surely equivalent to φ, then* $\liminf_{n\to\infty}\frac{|\varphi_n|}{n\log n} > \frac{1}{5}$.

Proof. Combine Equation (3) with Theorem 4. ■

And now we show, that the bounds we have obtained are fairly tight. The example type $(0,0,3)$ can be replaced by any nonunary one, without changing the argument.

Proposition 9 *Every Lindström quantifier Q of type $(0,0,3)$ can be defined in n element structures by a first order formula of length $O(n\log n)$ using only Lindström quantifiers of types (k,l) for $k,l\in\mathbf{N}$.*

Proof. The idea is to represent a ternary relation $R \subseteq A^3$ as n slices $\{(a,b) \in A^2 : (a,b,c) \in R\}$, $c \in A$. We implement this method defining the equivalent of $Qx,y,z(\varphi(x,y,z,\bar{u}))$, where Q is a class of structures with one ternary relation, in n element structures as follows:

$$\exists z_1 \ldots \exists z_n Q' v_1, \ldots, v_n, x_1, y_1, \ldots, x_n, y_n \left(\begin{array}{c} z_1 = v_1, \ldots, z_n = v_n, \\ \varphi(x_1, y_1, z_1, \bar{u}), \ldots, \varphi(x_n, y_n, z_n, \bar{u}) \end{array} \right),$$

where Q' consists of structures $\langle n, \{i\}_{i=0,\ldots,n-1}, \{(a,b) \in n^2 : (a,b,i) \in R\}_{i=0,\ldots,n-1} \rangle$ over all $\langle n, R \rangle \in Q$. Note the rôle of unary relations in structures in Q': they guarantee that all z's are different (written in first-order fashion it would require about $n^2/2$ inequalities), and allow identifying the slices with elements they correspond to—the order of slices is identical as the order of singleton unary relations identifying elements.
A simple analysis reveals that the length of the constructed sentence is $O(|\varphi| \cdot n \log n)$. ■

Acknowledgment. I would like to thank Kerkko Luosto for very interesting discussions we had on the topic and helpful comments ofn the earlier versions of this note.

References

[1] J. Cai, M. Fürer and N. Immerman, An optimal lower bound on the number of variables for graph identification, *Combinatorica* **12**(1992), pp. 389–410.

[2] A. Dawar, A restricted second-order logic for finite structures, *Logic and Computational Complexity* (ed. D. Leivant), Lecture Notes in Computer Science Vol. 960, Springer-Verlag (1995), pp. 393–413

[3] H.-D. Ebbinghaus and J. Flum, *Finite model theory*, Perspectives in Mathematical Logic, Springer Verlag, 1995.

[4] R. Fagin, Probabilities on finite models, *Journal of Symbolic Logic* **41**(1976), pp. 50–58.

[5] L. Hella, K. Luosto and J. Väänänen, The hierarchy theorem for generalized quantifiers, to appear in *Journal of Symbolic Logic.*

[6] L. Hella, Ph. Kolaitis and K. Luosto, Almost sure equivalence of logics, *manuscript.*

[7] M. Li and P.M.B. Vitányi, *An introduction to Kolmogorov complexity and its applications*, Springer Verlag, New York, 1993.

[8] J. Tyszkiewicz, The Kolmogorov expression complexity of logics, submitted to *Information and Computation.*

[9] M.Y. Vardi, Complexity of relational query languages, in: *Proc. 14th Symposium on Theory of Computation* 1982, pp. 137–146.