Algebraic Barriers to Halving Algorithmic Information Quantities in Correlated Strings

Andrei Romashchenko
LIRMM Univ Montpellier & CNRS, France

Abstract

We study the possibility of scaling down algorithmic information quantities in tuples of correlated strings. In particular, we address a question raised by Alexander Shen: whether, for any triple of strings (a, b, c), there exists a string z such that each conditional Kolmogorov complexity $C(a|z), C(b|z), C(c|z)$ is approximately half of the corresponding unconditional Kolmogorov complexity. We provide a negative answer to this question by constructing a triple (a, b, c) for which no such string z exists. Our construction is based on combinatorial properties of incidences in finite projective planes and relies on recent bounds for point-line incidences over prime fields, obtained using tools from additive combinatorics and algebraic methods, notably results by Bourgain–Katz–Tao and Stevens–De Zeeuw. As an application, we show that this impossibility yields lower bounds on the communication complexity of secret key agreement protocols in certain settings. These results reveal algebraic obstructions to efficient information exchange and highlight a separation in information-theoretic behavior between fields with and without proper subfields.

2012 ACM Subject Classification Mathematics of computing → Information theory; Theory of computation → Communication complexity; Security and privacy → Information-theoretic techniques; Mathematics of computing → Combinatoric problems

Keywords and phrases Kolmogorov complexity, algorithmic information theory, communication complexity, discrete geometry

Digital Object Identifier 10.4230/LIPIcs.MFCS.2025.84

Related Version *Full Version*: https://arxiv.org/abs/2504.14408

Funding The work was supported in part by the ANR project FLITTLA ANR-21-CE48-0023.

Acknowledgements The author sincerely thanks the anonymous reviewers for their careful reading and insightful suggestions, which helped improve the clarity and presentation of the paper.

1 Introduction

Algorithmic information theory (AIT) (introduced and developed in the 1960s by Solomonoff [23–25], Kolmogorov [13] and Chaitin [4]) aims to define the amount of information in a discrete object and to quantify the information shared by several objects. The crucial difference with Shannon's information theory is that AIT is interested not in an *average* compression rate (for some distribution of probabilities) but in the optimal compression of some specific *individual* object. Speaking informally, the information content of an individual object (e.g., of a string, a text, and so on) is defined as the minimal length of a program that produces that object. The length of the shortest program producing a string x is called Kolmogorov complexity of x and denoted $C(x)$. Similarly, conditional Kolmogorov complexity of x given y, denoted $C(x\,|\,y)$, is the length of an optimal program producing a string x given y as an input. A string x is called random or incompressible if $C(x) \approx |x|$.

The value of Kolmogorov complexity depends on the chosen programming language. However, it is known that there exist optimal programming languages that make the complexity function minimal up to bounded additive terms. An extensive introduction to

50th International Symposium on Mathematical Foundations of Computer Science (MFCS 2025).
Editors: Paweł Gawrychowski, Filip Mazowiecki, and Michał Skrzypczak; Article No. 84; pp. 84:1–84:18
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

AIT and the theory of Kolmogorov complexity can be found, for example, in the classical paper [30] and in the textbooks [14, 22].

AIT is tightly connected with the classical Shannon's information theory. The technique of Kolmogorov complexity is used in various problems of theoretical computer science and discrete mathematics. Time-bounded Kolmogorov complexity has deep and interesting links with computational complexity and theoretical cryptography, see., e.g., the surveys [6] and [15].

One of the fundamental questions of the AIT is a characterization of the possible values of Kolmogorov complexity of a tuple of strings. For example, for any triple of strings x, y, z, we have seven values of Kolmogorov complexity (sometimes called *complexity profile* of (x, y, z)):

$$C(x),\ C(y),\ C(z),\ C(x, y),\ C(x, z),\ C(y, z),\ C(x, y, z).$$

Which vectors of seven positive numbers can be realized as Kolmogorov complexity of some x, y, z? What are the universal constraints connecting different components in such vectors of complexities? There are, for example, classical inequalities

$$C(x, y) \leq C(x) + C(y) + O(\log C(x, y))$$

(subadditivity, or non-negativity of the mutual information) and

$$C(x, y, z) + C(z) \leq C(x, z) + C(y, z) + O(\log C(x, y, z))$$

(submodularity, or non-negativity of the conditional mutual information). It is known that for triples of strings there are no substantially different linear inequalities for Kolmogorov complexity: any linear inequality for Kolmogorov complexity of x, y, z (valid up to an additive term $o(C(x, y, z))$) is necessarily a positive linear combination of several instances of subadditivity and submodularity, [9]. However, when four or more strings are involved (so we have $\geq 2^4 - 1 = 15$ quantities of Kolmogorov complexity), there also exist different linear inequalities (usually called *non Shannon type* inequalities) that are less intuitive in appearance and cannot be represented as linear combinations of subadditivity and submodularity, see e.g. the survey [29]. It is known that exactly the same linear inequalities that are valid for Kolmogorov complexity and for Shannon entropy, but the problem of precise characterization of these inequalities for $n \geq 4$ objects remains widely open.

While the questions on linear inequalities for Kolmogorov complexity and for Shannon's entropy are known to be equivalent, from some other perspectives, questions about Kolmogorov complexity appear more difficult than similar questions about Shannon's entropy. It is not known, for example, whether the complexity profiles can be scaled with any factor $\lambda > 0$ (let us say, up to a logarithmic additive term). More specifically, the following question is open:

▶ **Question 1.** Let λ be a positive real number. Is it true that for every k-tuple of strings $(x_1, \ldots, x_k)$ there exists another k-tuple $(x'_1, \ldots, x'_k)$ such that

$$C(x'_{i_1}, x'_{i_2}, \ldots, x'_{i_s}) = \lambda C(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) + O(\log C(x_1, \ldots, x_n))$$

for all tuples of indices $(i_1, \ldots, i_s)$, $1 \leq i_1 < i_2 < \ldots < i_s \leq k$?

The answer to this question is known to be positive for $k \leq 3$ and any λ and for any k and integer λ. For non-integer factors, e.g., for $\lambda = 1/2$, the question is open for all $k \geq 4$, see [20]. Alexander Shen posed another question (see a comment to Question 1 in [20]):

▶ **Question 2.** Let $\lambda < 1$ be a positive real number. Is it true that for every k-tuple of strings $(x_1, \ldots, x_k)$ there exists a string z such that

$$C(x_i \,|\, z) = \lambda C(x_i) + O(\log C(x_1, \ldots, x_n))$$

for $i = 1, \ldots, k$?

The positive answer to Question 2 would imply the positive answer to Question 1. Besides, Question 2 is interesting in its own right as a special case of the problem of *extension of complexity profiles*, generalizing the classical notions of *common information* (by Gács-Körner [7] and Wyner [28]):

▶ **Question 3** (informal). For each given k-tuple of strings $(x_1, \ldots, x_k)$, what can we say about possible values

$$\{C(x_{i_1}, x_{i_2}, \ldots, x_{i_s}, z)\}_{1 \le i_1 < \ldots < i_s \le k}$$

achievable with various strings z?

It is know that the answer to Question 2 is positive for $k = 1$ and for $k = 2$ (see Section 2). The main result of this paper is the negative answer to this question for $k = 3$, even for $\lambda = 1/2$. We show there exists a triple of strings (a, b, c) such that there is no z which "halves" the complexities of each of them,

$$C(a \,|\, z) \approx \frac{1}{2}C(a),\;\; C(b \,|\, z) \approx \frac{1}{2}C(b),\;\; C(c \,|\, z) \approx \frac{1}{2}C(c).$$

We provide an example of a triple (a, b, c) such that for every z such that $C(a \mid z) \approx \frac{1}{2}C(a)$, $C(b \,|\, z) \approx \frac{1}{2}C(b)$, the value of $C(c \,|\, z)$ must be much smaller than $\frac{1}{2}C(c)$. We prove this result in Section 3.

1.1 The main construction

To prove our main result, we propose an explicit construction of a tuple that provides the negative answer to Question 2. This construction is based on *incidences in a finite projective plane*. We fix a finite field $\mathbb{F}$, take the projective plane over this field, and consider pairs (x, y), where x is a line in this plane and y is a projective line passing through a point. We call such pairs *incidences*. An incidence in a projective plane is a classical combinatorial object, and its properties were extensively studied in different contexts. Incidences were considered in AIT, see, e.g., [5, 17].

In a projective plane over $\mathbb{F}$ there are $\Theta(|\mathbb{F}|^2)$ points, $\Theta(|\mathbb{F}|^2)$ lines, and $\Theta(|\mathbb{F}|^3)$ incidences. For the vast majority of incidences (x, y) we have

$$C(x) \approx 2\log|\mathbb{F}|,\;\; C(y) \approx 2\log|\mathbb{F}|, C(x, y) \approx 3\log|\mathbb{F}|. \tag{1}$$

The upper bounds are trivial: to specify a point or a line in a projective plane, it is enough to provide two elements of $\mathbb{F}$; to specify together a point and a line incident to this point, it is enough to provide three elements of $\mathbb{F}$. The lower bound follows from a simple counting argument: the number of programs (descriptions) shorter than k is less than 2^k; therefore, for most incidences (x, y) there is no short description, and $C(x, y) \approx 3\log|\mathbb{F}|$. A similar argument implies $C(x) \approx 2\log|\mathbb{F}|$ and $C(y) \approx 2\log|\mathbb{F}|$. We call an incidence (x, y) *typical* if it satisfies (1).

Thus, for a typical incidence (x, y) the mutual information between x and y is

$$I(x:y) := C(x) + C(y) - C(x,y) \approx \log|\mathbb{F}|.$$

An. Muchnik observed in [17] that the mutual information of an incidence is hard to "materialize," i.e., we cannot find a z that "embodies" this amount of information shared by x and y. More formally, Muchnik proved that

$$\text{there is no } z \text{ such that } C(z\,|\,x) \approx 0,\ C(z\,|\,y) \approx 0,\ C(z) \approx I(x:y).$$

This insight did not close the question completely: the optimal trade-off between $C(z\,|\,x)$, $C(z\,|\,y)$, $C(z)$ is still not fully understood. Our work follows this direction of research. We prove that for *prime fields* $\mathbb{F}$, some specific values of $C(z\,|\,x)$, $C(z\,|\,x)$, and $C(z)$ are forbidden:

$$\begin{array}{l}\text{For a prime } \mathbb{F}, \text{ for a typical incidence } (x,y) \text{ there is no } z \text{ such that}\\ C(z) \approx 1.5\log|\mathbb{F}|,\ C(z\,|\,x) \approx 0.5\log|\mathbb{F}|,\ C(z\,|\,y) \approx 0.5\log|\mathbb{F}|,\ C(z\,|\,x,y) \approx 0,\end{array} \tag{2}$$

see a more precise statement in Theorem 3.8 on p. 11. This result contrasts with a much simpler fact proven in [19]:

$$\begin{array}{l}\text{If } \mathbb{F} \text{ contains a subfield of size } \sqrt{|\mathbb{F}|}, \text{ then for a typical incidence } (x,y) \text{ there exists}\\ \text{a } z \text{ such that } C(z) \approx 1.5\log|\mathbb{F}|,\ C(z\,|\,x) \approx C(z\,|\,y) \approx 0.5\log|\mathbb{F}|,\ C(z\,|\,x,y) \approx 0,\end{array} \tag{3}$$

see Theorem 3.7 on p. 10.

Our proof of (2) uses a remarkable result by Sophie Stevens and Frank De Zeeuw, which gives a non-trivial upper bound on the number of incidences between points and lines in a plane over a prime field [26]. The first theorem of this type was proven by Bourgain, Katz, and Tao, [2]. This result has been improved further in [10–12]. We use the bound from [26], the strongest to date.

Typical incidences in the projective plane over a prime field imply the negative answer to Question 2: if (x, y) is a typical incidence, we let

$$a := x,\ b := y,\ c := \langle x, y\rangle$$

and prove that for every string z satisfying the conditions $C(a\,|\,z) \approx \frac{1}{2}C(a)$ and $C(b\,|\,z) \approx \frac{1}{2}C(b)$, the value of $C(c\,|\,z)$ must be much smaller than $\frac{1}{2}C(c)$, see Corollary 3.9.

1.2 Application: impossibility results for secret key agreement

The main result (2) can be interpreted as a partial (very limited in scope) answer to Question 3, as it claims that for some specific pairs (x, y) (typical incidences) there exist limitations for realizable complexity profiles of triples (x, y, z). It is no surprise that this fact can be used to prove certain *no-go* results in communication complexity, for settings where the participants of the protocol are given such x and y as their inputs. We present an example of such result – a theorem on secret key agreement protocols, as we explain below.

Unconditional *secret key agreement* is one of the basic primitive in information-theoretic cryptography, [27]. In the simplest setting, this is a protocol for two parties, Alice and Bob. At the beginning of the communication, Alice and Bob are given some input data, x and y respectively. It is assumed that x and y are strongly correlated, i.e., the mutual information between x and y is non-negligible. Further, Alice and Bob exchange messages over a public channel and obtain (on both sides) some string w that is incompressible (i.e., $C(w)$ is close to its length) and has negligible mutual information with the transcript of the protocol, i.e.,

$$C(w\,|\,\text{concatenation of all messages sent by Alice and Bob}) \approx |w|.$$

Thus, Alice and Bob transform the mutual information between x and y into a common secret key (which can later be used, for example, into a one-time-pad or some other unconditionally secure cryptographic scheme). The *secrecy* of the key means that an eavesdropper should get (virtually) no information about this key, even having intercepted all communication between Alice and Bob. For a more detailed discussion of the secret key agreement in the framework of AIT we refer the reader to [8, 21].

▶ Remark 1.1. In this paper, we assume that the communication protocol is *uniformly computable*; that is, Alice and Bob exchange messages and compute the final result according to a single algorithmically defined rule that applies uniformly to inputs of all lengths. We also assume that the protocol is public (i.e., known to an eavesdropper), so no secret information can be hardwired into the protocol description; see [8, Remark 1] and [21, Remarks 2, 4, 13] for a more detailed discussion of the communication model.

The challenges in secret key agreement are to (i) maximize the size of the secret key and (ii) to minimize the communication complexity of the protocol (the total length of messages sent to each other by Alice and Bob). It is known that the maximum size of the secret key is equal to the mutual information between x and y, i.e., $I(x : y) = C(x) + C(y) - C(x, y)$ (see [21] for the proof in the framework of AIT and [1, 16] for the original result in the classical Shannon's settings). There exists a communication protocol that allows to produce a secret of optimal size with communication complexity

$$\max\{C(x \,|\, y), C(y \,|\, x)\}, \tag{4}$$

see [21], and this communication complexity is tight, at least for some "hard" pairs of inputs (x, y), see [8]. Moreover, subtler facts are known:

- the standard protocol achieving (4) (the construction dates back to [1, 16]; see [21] for the AIT version) is highly asymmetric: all messages are sent by only one party (Alice or Bob);
- for some pairs of inputs (x, y), if we want to agree on a secret key of maximal possible size $I(x : y)$, not only the total communication complexity must be equal to (4), but actually *one of the parties* (Alice or Bob) must send $\max\{C(x \,|\, y), C(y \,|\, x)\}$ bits of information, [3];
- for some pairs of inputs (x, y), the total communication complexity $\max\{C(x \,|\, y), C(y \,|\, x)\}$ cannot be reduced *even if the parties need to agree on a sub-optimal secret key of size δn* (for any constant $\delta > 0$), see [8].

It remains unknown whether we can always organize a protocol of secret key agreement where the communication complexity (4) is shared evenly by the parties (both Alice and Bob send $\frac{1}{2}C(x \,|\, y)$ bits) if they need to agree on a key of sub-optimal size, e.g., $\frac{1}{2}I(x : y)$.

When we claim that communication complexity of a protocol is large *in the worst case*, i.e., Alice and Bob must send to each other quite a lot of bits *at least for some pairs of inputs*, it is enough to prove this statement of some specific pair of data sets (x, y). Such a proof may become simpler when we use (x, y) with nice combinatorial properties, even though these inputs may look artificial and unusual for practical applications. Such is the case with the mentioned lower bounds for communication complexity proven in [8] and [3]. Both these arguments employ as an instance of a "hard" input (x, y) a typical incidence in a finite projective plane. Thus, it is natural to ask whether, for these specific (x, y), it is possible to agree on a secret key of sub-optimal size using a *balanced* communication load – that is, when Alice and Bob each send approximately the same number of bits, roughly half the total communication complexity. We show, quite surprisingly, that the answer to this question depends on whether the field admits a proper subfield:

▶ **Positive result.** *If the field $\mathbb{F}_q$ contains a subfield of size $\sqrt{q}$, then there exists a* balanced *communication protocol with communication complexity* $\log q$ *where*

- *Alice sends to Bob* $\approx 0.5 \log q$ *bits,*
- *Bob sends to Alice* $\approx 0.5 \log q$ *bits,*

and the parties agree on a secret key of length $\approx 0.5 \log q$*, which is incompressible even conditional on the transcript of the communication between Alice and Bob.*

▶ **Negative result.** *If the field $\mathbb{F}_q$ is prime, then in every* balanced *communication protocol with communication complexity* $\log q$ *such that*

- *Alice sends to Bob* $\approx 0.5 \log q$ *bits,*
- *Bob sends to Alice* $\approx 0.5 \log q$ *bits,*

the parties cannot *agree on a secret key of length* $\approx 0.5 \log q$ *or even of any length* $> \frac{3}{7} \log q$ *(the secrecy of the key means that the key must remain incompressible even conditional on the transcript of the communication between Alice and Bob).*

For a more precise statements see Theorem 4.3 and Theorem 4.1 respectively.

1.3 Techniques

A projective plane is a classical geometric object, and combinatorial properties of discrete projective planes have been studied with a large variety of mathematical techniques. It is no surprise that, in the context of AIT, the information-theoretic properties of incidences in discrete projective planes have been studied using many different mathematical tools. In this paper we bring to AIT another (rather recent) mathematical technique that helps distinguish information-theoretic properties of projective planes over prime fields and over fields containing proper subfields.

As we mentioned above, we apply the new approach to the problem of secret key agreement: we consider the setting where Alice and Bob receive as inputs data sets x and y such that (x, y) is a "typical" incidence in a projective plane (x is a line and y is a point incident to this line) over a finite field $\mathbb{F}$ with $n = \lceil \log |\mathbb{F}| \rceil$. We summarize in Table 1 below several technical results concerning this communication problem, and the techniques in the core of these results.

Table 1 Bounds for secret key agreement in the framework of AIT.

for any protocol of secret key agreement, the size of the secret key $\lesssim I(x:y) \approx n$ [21]	information-theoretic techniques: intern.inform.cost $\le$ extern.inform.cost (not specific for lines and points)
$\lvert$Alice's messages$\rvert$ + $\lvert$Bob's messages$\rvert$ $\gtrapprox n$, even for a secret key of size ϵn [8]	spectral method, expander mixing lemma (applies to all fast-mixing graphs, including the incidence graph of a projective plane)
$\lvert$Alice's messages$\rvert$ $\gtrapprox n$ or $\lvert$Bob's messages$\rvert$ $\gtrapprox n$ if the parties agree on a secret key of size $\approx n$, [3]	combinatorics of a projective plane (applies to all projective planes)
for incidences in a plane over a prime field if $\lvert$Alice's messages$\rvert$ $\approx 0.5n$ and $\lvert$Bob's messages$\rvert$ $\approx 0.5n$ then the size of the secret key $\ll 0.5n$, **[this paper]**	additive combinatorics, algebraic and geometric methods [2, 10–12, 26] (**applies to only projective planes over prime fields**)

One of the motivations for writing this paper was to promote the notable results of [2, 10–12, 26], which presumably can find interesting applications in AIT and communication complexity.

1.4 Organization

The rest of the paper is structured as follows. In Section 2 we briefly discuss (3) (known from [19]). In Section 3 we formally prove our main result (2). In Section 4 we discuss an application of the main result: we show that the performance of the secret key agreement for Alice and Bob given as inputs an incident pair (x, y) (from a projective plane) differs between fields that do and do not contain proper subfields.

1.5 Notation

- $|\mathcal{S}|$ stands for the cardinality of a finite set $\mathcal{S}$
- we write $F(n) \ll G(n)$ if $G(n) - F(n) = \Omega(n)$ (e.g., $\frac{22n}{15} \ll \frac{3n}{2}$)
- for a bit string x we denote by $x_{k:m}$ a factor of x that consists of $m - k + 1$ bits at the positions between k and m (in particular, $x_{[1:m]}$ is a prefix of x of length m);
- we denote $\mathbb{FP}$ the projective plane over a finite field $\mathbb{F}$;
- $G = (R, L; E)$ stands for a bipartite graph where $L \cup R$ (disjoint union) is the set of vertices and $E \subset L \times R$ is the set of edges;
- $C(x)$ and $C(x \mid y)$ stand for Kolmogorov complexity of a string x and, respectively, conditional Kolmogorov complexity of x conditional on y, see [14, 22]. We use a similar notation for more involved expressions, e.g., $C(x, y \,|\, v, w)$ denotes Kolmogorov complexity of the code of the pair (x, y) conditional on the code of another pair (v, w)
- we also talk about Kolmogorov complexity of more complex combinatorial objects (elements of finite fields, graphs, points and lines in a discrete projective plane, and so on) assuming that each combinatorial object is represented by its *code* (for some fixed computable encoding rule)
- $I(x : y) := C(x) + C(y) - C(x, y)$ and $I(x : y \,|\, z) := C(x \,|\, z) + C(y \,|\, z) - C(x, y \,|\, z)$ stand for information in x on y and, respectively, information in x on y conditional on z

Many natural equalities and inequalities for Kolmogorov complexity are valid only up to a logarithmic additive term, e.g., $C(x, y) = C(x) + C(y \,|\, x) \pm O(\log n)$, where n is the sum of lengths of x and y (this is the chain rule a.k.a. Kolmogorov–Levin theorem, see [30]). To simplify the notation, we write $A \overset{\log}{\le} B$ instead of $A \le B + O(\log N)$, where N is the sum of lengths of all strings involved in the expressions A and B. Similarly we define $A \overset{\log}{\ge} B$ (which means $B \overset{\log}{\le} A$) and $A \overset{\log}{=} B$ (which means $A \overset{\log}{\le} B$ and $B \overset{\log}{\le} A$). For example, the chain rule can be expressed as

$$C(x, y) \overset{\log}{=} C(x) + C(y \,|\, x);$$

the well known fact of symmetry of the mutual information can be expressed as

$$I(x : y) \overset{\log}{=} C(x) + C(y) - C(x, y).$$

2 Halving complexities of two strings

In this section we discuss the positive answer to Question 2 for $k = 1, 2$ and $\lambda = 1/2$. These results were proven in [19]. Here we recall the main ideas and technical tools behind this argument.

First of all, we observe that Question 2 for $k = 1$ and $\lambda = 1/2$ is pretty trivial. Given a string x of length N, we can try $z = x_{[1:k]}$ for $k = 0, \ldots N$. It is clear that for $k = 0$ we have $C(x \,|\, x_{[1:k]}) = C(x) + O(1)$, and for $k = N$ we obtain $C(x \,|\, x_{[1:k]}) = O(1)$. At the same time, when we add to the condition z one bit, the conditional complexity $C(x \,|\, z)$ changes by only $O(1)$. It follows immediately that for some intermediate value of k we obtain $z = x_{[1:k]}$ such that $C(x \,|\, z) = \frac{1}{2}C(x) + O(1)$.

This argument employ (in a very naive from) the same intuition as the intermediate value theorem for continuous functions. The case $k = 2$ is more involved, but it also can be proven with "topological" considerations.

▶ **Theorem 2.1.** *For all strings a, b of complexity at most n there exists a string z such that*

$$\left|C(a \,|\, z) - \frac{1}{2}C(a)\right| = O(\log n) \ \textit{and}\ \left|C(b \,|\, z) - \frac{1}{2}C(b)\right| = O(\log n).$$

In fact, [19] proved a tighter and more general statement:

▶ **Theorem 2.2.** *[19, Theorem 4] For some constant κ the following statement holds: for every two strings a, b of complexity at most n and for every integers α, β such that*

- $\alpha \le C(a) - \kappa \log n$,
- $\beta \le C(b) - \kappa \log n$,
- $-C(a \,|\, b) + \kappa \log n \le \beta - \alpha \le C(b \,|\, a) - \kappa \log n$,

there exists a string y such that $|C(a \,|\, z) - \alpha| \le \kappa$ and $|C(b \,|\, z) - \beta| \le \kappa$.

With $\alpha = \frac{1}{2}C(a)$ and $\beta = \frac{1}{2}C(b)$, this theorem implies the following corollary, which is (for non-degenerate parameters) a stronger version of Theorem 2.1:

▶ **Corollary 2.3.** *For some constant κ the following statement holds: for every two strings a, b such that $C(a \,|\, b) \ge \kappa \log n$ and $C(b \,|\, a) \ge \kappa \log n$ there exists a string z such that*

$$\left|C(a \,|\, z) - \frac{1}{2}C(a)\right| \le \kappa \ \textit{and}\ \left|C(b \,|\, z) - \frac{1}{2}C(b)\right| \le \kappa.$$

The proof of Theorems 2.2 proposed by A. Shen involves topological ideas. The argument in a nutshell: we build z by combining two parts, a piece extracted from a and a piece extracted from b; the only challenge is to choose the sizes of these two parts in a suitable way. It turns out that suitable sizes of these pieces can be chosen using the topological statement known as *the drum theorem*, [18], which is equivalent to the fact that a circle is not a retract of a closed disk, see [19] and the arXiv version of this paper for the complete proof.

3 Typical incidences in a projective plane

In this section we discuss typical pairs (line, point) in a finite projective plane; we study their information-theoretic properties, focusing on distinctions that arise depending on whether the underlying field contains a proper subfield.

3.1 Typical pairs: interface between the combinatorial language and AIT

In this section we introduce the framework that helps to translated information-theoretic questions in the combinatorial language.

▶ **Definition 3.1.** *Let $G = (L, R; E)$ with $E \subset L \times R$ be a simple non-directed bipartite graph. This graph is bi-regular if all vertices in L have the same degree (the same number of neighbors in L) and all vertices in R have the same degree (the same number of neighbors in L).*

To specify the quantitative characteristics of G we will use a triple of parameters (α, β, γ) such that

$$|L| = 2^{\alpha},\ \ |R| = 2^{\beta},\ \ |E| = 2^{\gamma}.$$

If G is bi-regular, then the degrees of vertices in L are equal to $|E|/|L| = 2^{\gamma-\alpha}$ and the degrees of vertices in R are equal to $|E|/|R| = 2^{\gamma-\beta}$.

▶ **Proposition 3.2.** *Let $G = (L, R; E)$ be a bi-regular with parameters (α, β, γ), as defined above. If the graph is given explicitly (the complete list of vertices and edges of the graph can be found algorithmically given the value of the parameters n), then the vast majority (let us say, for 99%) of pairs $(x, y) \in E$ we have*

$$\begin{cases} C(x) \overset{\log}{=} \alpha + O(\log n), \\ C(y) \overset{\log}{=} \beta + O(\log n), \\ C(x,y) \overset{\log}{=} \gamma + O(\log n). \end{cases} \tag{5}$$

Proof. This proposition follows from a standard counting, see e.g. [22]. ◀

▶ **Definition 3.3.** *For a graph $G = (L, R; E)$ with parameters (α, β, γ) we say that an edge $(u, v) \in E$ is* typical *if it satisfies (5).*

▶ **Proposition 3.4.** *Let $G = (L, R; E)$ be an explicitly given bi-regular bipartite graph with parameters (α, β, γ), as in Definition 3.1. Let $(x, y) \in E$ be a typical edge in this graph, as in Definition 3.3. And let z be a string satisfying:*

$$C(x \mid z) \le \alpha', \quad C(y \mid z) \le \beta', \quad C(x, y \mid z) \ge \gamma',$$

for some integers $(\alpha', \beta', \gamma')$ with $\alpha' \le \alpha$, $\beta' \le \beta$, and $\gamma' \le \gamma$. Then there exists an induced subgraph $H = (L', R'; E')$ of G,

$$L' \subset L, \quad R' \subset R, \quad E' = (L' \times R') \cap E,$$

such that $|L'| = 2^{\alpha' \pm O(\log n)}$, $\quad |R'| = 2^{\beta' \pm O(\log n)}$, $\quad |E'| \ge 2^{\gamma' - O(\log n)}$.

Sketch of the proof. We let

$$L' = \{x' \in L\ :\ C(x' \,|\, z) \le \alpha'\},\ R' = \{y' \in R\ :\ C(y' \,|\, z) \le \beta'\}.$$

Observe that $x \in L'$ and $y \in R'$.

▶ **Lemma 3.5.** $|L'| = 2^{\alpha' \pm O(\log n)}$, $|R'| = 2^{\beta' \pm O(\log n)}$.

Proof of lemma. This lemma is a standard translation between the combinatorial and the information-theoretic languages. The upper bound for $|L'|$ follows from the fact that each element of L' is obtained from z by a program of length at most α'. The lower bound follows from the observation that L' contains, among other elements, the $2^{\alpha' - O(\log n)}$ smallest elements of L in lexicographic order. The argument for R' is similar. A more detailed proof can be found, e.g., in [3, lemma 1 and lemma 2]. ◀

It remains to prove a bound on the cardinality of E'. Given a string z, we can run in parallel all programs of length α' and β' on input z and enumerate the results that they produce. These results will provide us with the lists of elements L' and R' revealing step by step. Accordingly, we can enumerate edges of E'. Every pair $(x', y') \in E'$ can be specified by (i) the binary expansion of the numbers α, β and (ii) by the ordinal number of (x', y') in the enumeration of E'. This argument applies in particular to the pair (x, y), which belongs to E'. Therefore, $C(x, y\,|\,z) \le \log|E'| + O(\log n)$. Reading this inequality from the right to the left, we obtain

$$|E'| \ge 2^{C(x,y|z)-O(\log n)} = 2^{\gamma'-O(\log n)},$$

and we are done. ◀

3.2 Typical incidences in a projective plane

Now we instantiate the framework discussed above and discuss the central construction of this paper – typical incidences in a finite projective plane.

▶ **Example 3.6.** Let $\mathbb{F}$ be a finite field and $\mathbb{FP}$ be the projective plane over this field. Let L be the set of points and R be the set of lines in this plane. A pair $(x, y) \in L \times R$ is connected by an edge iff the chosen point x lies in in the chosen line y. Hereafter we denote this graph by $G^{\mathrm{PL}}_{\mathbb{F}}$.

We proceed with a discussion of properties of (x, y) from Example 3.6 that differ depending on whether $\mathbb{F}$ possesses a proper subfield.

▶ **Theorem 3.7** (see [5]). *Let $\mathbb{F}$ be a field with a subfield of size $\sqrt{|\mathbb{F}|}$. Then for a typical edge (x, y) of $G^{\mathrm{PL}}_{\mathbb{F}}$ (i.e., a typical incident pair $(line, point)$ on the plane $\mathbb{FP}$) we have*

$$C(x) \overset{\log}{=} 2n,\ C(y) \overset{\log}{=} 2n,\ C(x, y) \overset{\log}{=} 3n,$$

and there exists a z such that

$$C(x\,|\,z) \overset{\log}{=} n,\ C(y\,|\,z) \overset{\log}{=} n,\ C(x, y\,|\,z) \overset{\log}{=} 1.5n$$

or, equivalently

$$C(x\,|\,y, z) \overset{\log}{=} 0.5n,\ C(y\,|\,x, z) \overset{\log}{=} 0.5n,\ I(x : y\,|\,z) \overset{\log}{=} 0.5n,$$

for $n = \lceil|\mathbb{F}|\rceil$, as shown in the diagram in Fig. 1.

Sketch of the proof. The first claim of the theorem (the values of unconditional Kolmogorov complexity) follows from Proposition 3.2 and from typicality of (x, y). The second claim, concerning the values of conditional Kolmogorov complexity, is subtler and requires a construction. The statement statement boils down to the following combinatorial claim: the graph $G^{\mathrm{PL}}_{\mathbb{F}}$ can be covered by a relatively small family[1] of induced subgraphs $H_i = (L_i, R_i; E_i)$, where

$$|L_i| = |R_i| = 2^n,\ |E_i| = 2^{1.5n}.$$

This combinatorial claim, in turn, follows from the facts that $G^{\mathrm{PL}}_{\mathbb{F}}$ is edge-transitive and $\mathbb{F}$ contains a subfield of size $\sqrt{|\mathbb{F}|}$, see [5, Theorem 9] for details. ◀

[1] More technically, we need $2^{1.5n+O(\log n)}$ such subgraphs, and every edge of $G^{\mathrm{PL}}_{\mathbb{F}}$ is covered by at most poly(n) subgraphs H_i.

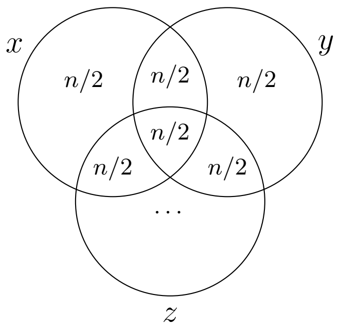

Figure 1 Complexity profile for (x, y, z) from Theorem 3.7.

Theorem 3.7 contrasts with Theorem 3.8.

▶ **Theorem 3.8.** *Let $\epsilon \geq 0$ be a small enough real number and $\mathbb{F}$ be a field of a prime cardinality p, and $n := \lceil \log p \rceil$. Than for a typical edge (x, y) of $G_{\mathbb{F}}^{\mathrm{PL}}$ (i.e., a typical incident pair (line, point) on the plane $\mathbb{FP}$) we have*

$$C(x) \overset{\log}{=} 2n,\ C(y) \overset{\log}{=} 2n,\ C(x, y) \overset{\log}{=} 3n,$$

and for every z such that

$$C(x\,|\,z) \overset{\log}{\leq} (1+\epsilon)n,\ C(y\,|\,z) \overset{\log}{\leq} (1+\epsilon)n \tag{6}$$

we have $C(x, y\,|\,z) \overset{\log}{\leq} \frac{22}{15}(1+\epsilon)n \ll \frac{3n}{2}$.

Proof. Again, the first claim of the theorem (the values of unconditional Kolmogorov complexity) follows from Proposition 3.2 and from typicality of (x, y). We proceed with the second claim. From Proposition 3.4 it follows that in $G_{\mathbb{F}}^{\mathrm{PL}}$ there is a subgraph $G' = (L', R', E')$ such that

$$\begin{array}{rcl} |L'| & = & 2^{(1+\epsilon)n+O(\log n)}, \\ |R'| & = & 2^{(1+\epsilon)n+O(\log n)}, \end{array} \tag{7}$$

and

$$|E'| \geq 2^{C(x,y|z)-O(\log n)}. \tag{8}$$

If $\epsilon < 1/7$, then the cardinalities of L' and R' are less than $|\mathbb{F}|^{8/7}$. It was shown in [26] that for every subgraph G' in $G_{\mathbb{F}}^{\mathrm{PL}}$ for a prime $\mathbb{F}$ satisfying the constraints

$$|L'|^{7/8} < |R'| < |L'|^{8/7} \text{ and } \max\{|L'|, |R'|\} \leq |\mathbb{F}|^{8/7}$$

we have

$$|E'| \leq (|L'| \cdot |R'|)^{11/15}.$$

We plug in this inequality (7) and (8) and obtain

$$C(x, y\,|\,z) \overset{\log}{\leq} \frac{22}{15}(1+\epsilon)n \ll \frac{3n}{2},$$

provided that ϵ is small enough. ◀

▶ **Corollary 3.9.** *For every n there exists a triple of strings (a, b, c), each one of complexity $\Theta(n)$, such that there is no z satisfying*

$$\begin{array}{rcl} C(a\,|\,z) & = & \frac{1}{2}C(a) + O(\log n), \\ C(b\,|\,z) & = & \frac{1}{2}C(b) + O(\log n), \\ C(c\,|\,z) & = & \frac{1}{2}C(c) + O(\log n). \end{array}$$

More precisely, for all z such that $C(a\,|\,z) \overset{\log}{=} \frac{1}{2}C(a)$ and $C(b\,|\,z) \overset{\log}{=} \frac{1}{2}C(b)$, we have

$$C(c\,|\,z) \leq \frac{22}{45}C(c) + O(\log n) \ll \frac{1}{2}C(c).$$

Proof. We fix an integer n and the minimal prime number p such that $2^n < p < 2^{n+1}$, and let (x, y) be a typical edge in $G^{\text{PL}}_{\mathbb{F}_p}$, as in Theorem 3.8. Then we define $a := x,\ b := y,\ c := \langle x, y\rangle$ and apply Theorem 3.8. ◀

4 Secret key agreement

In this section we study communication complexity of the protocol of unconditional (information-theoretic) secret key agreement. Let us recall the settings of the unconditional *secret key agreement.* We deal with two parties, Alice and Bob. Alice and Bob receive input data, x and y respectively. It is assumed that the mutual information between x and y is non-negligible, and its value is known to Alice and Bob, as well as to the adversary. Further, Alice and Bob exchange messages over a public channel and obtain (on both sides) some string w that must be incompressible (i.e., $C(w)$ is close to its length) and must have negligible mutual information with the transcript of the protocol, i.e.,

$$C(w\,|\,\text{concatenation of all messages sent by Alice and Bob}) \approx |w|.$$

Thus, Alice and Bob use the mutual information between x and y to produce a common secret key w using a communication via a non-protected channel. The protocol succeed if Alice and Bob obtain one and the same w, and an eavesdropper gets only negligible information about this key, even having intercepted all messages sent to each other by Alice and Bob. In this paper we assume that the communication protocols are deterministic. All arguments easily extends to randomized communication protocols with a public[2] source of randomness (accessible to Alice, Bob, and the eavesdropper). A more detailed discussion of the settings of secret key agreement problem in the framework of AIT can be found in [8, 21].

The optimal size of the secret key is known to be equal to the mutual information between x and y, and communication complexity of the protocol is at most (4), see [21] (in what follows we discuss pairs (x, y) with a symmetric complexity profile where $C(x\,|\,y) = C(y\,|\,x)$).

4.1 Specific input data: secret key agreement with a typical incidence from a finite plane

Let us focus on the case where the inputs (x, y) represent a pair of typical incidences in a projective plane over a finite field $\mathbb{F}$ (we denote $n := \lceil \log |\mathbb{F}| \rceil$). In this case the upper bound (4) (which rewrites in this case to to n) is tight, the communication complexity cannot be made better than $n - O(\log n)$, [8]. Moreover,

[2] The case of private sources of randomness is a more complex setting. We leave the consideration of this type of protocols for further research.

(i) for every communication protocol, for its transcript t we have

$$C(t) \overset{\log}{\geq} I(t:x\,|\,y) + I(t:y\,|\,x) \overset{\log}{\geq} n,$$

(the first inequality is known from [21] and the second one from [8]);

(ii) this bound remains valid *even if the parties agree on a sub-optimal secret key of size δn* for any $\delta > 0$, [8];

(iii) if Alice and Bob agree on a secret key w of maximal possible size $I(x:y) = n$, then not only the total communication complexity must be equal to n but actually *one of the parties* (Alice or Bob) must send $\max\{C(x\,|\,y), C(y\,|\,x)\} \overset{\log}{=} n$ bits of information, [3].

We summarize:

- *even for a suboptimal key size* communication complexity of the protocol $\overset{\log}{\geq} n$;
- *for an optimal key size* the communication is very asymmetric – all n bits are sent by one of the participants.

There remained a question: Does there exist a protocol with a symmetric communication load (both Alice and Bob send $\approx n/2$ bits) with a suboptimal key size? In what follows we show that the answer to this question depends on whether the underlying field contains a proper subfield.

4.2 Prime field: a negative result

▶ **Theorem 4.1.** *Let q be a prime number and $\mathbb{F}_q$ be the field with q elements. Let $\mathbb{FP}$ be the projective plane over $\mathbb{F}_q$, and (x, y) be a typical incidence in this plane (x is a line in this projective plane, y is a point in this line, and $C(x, y) \overset{\log}{=} 3 \log q$). Let us denote $n = \lceil \log q \rceil$.*

We consider communication protocols where Alice is given as her input x and Bob is given as his input y. Assume that there exists a communication protocol where

- *Alice sends messages of total length $(\frac{1}{2} + \epsilon)n$ bits to Bob,*
- *Bob sends messages of total length $(\frac{1}{2} + \epsilon)n$ bits to Alice,*
- *at the end of the communication, Alice and Bob agree on a secret key w of length k, satisfying $C(w\,|\,t) \overset{\log}{=} C(w) \overset{\log}{=} k$, where t is the transcript of the protocol (the sequence of all messages exchanged between Alice and Bob during the protocol); in other words, the protocol reveals virtually no information about the secret to the eavesdropper.*

We claim that for small enough ϵ the size of the secret key is much less than $\frac{1}{2}I(x:y)$, i.e., $k \ll n/2$.

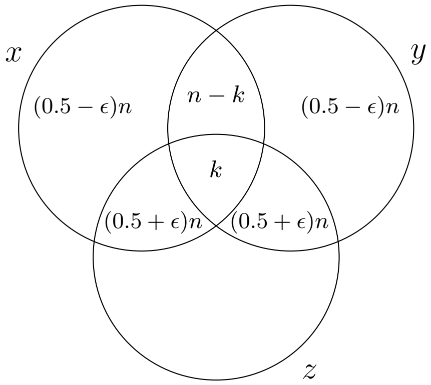

Figure 2 Complexity profile for (x, y, z) from Theorem 4.1, cf. Fig. 1.

Proof. Let Alice and Bob agree on a secret key w in protocol with transcript t. The fact that both Alice and Bob compute w at the end of the protocol means that $C(w\,|\,t,x)$ and $C(w\,|\,t,y)$ are negligibly small. Security of the key means that $I(w:t)$ is negligible, i.e., the transcript divulges virtually no information about the key. Keeping in mind these observations, we define $z=\langle t,w\rangle$. We have $C(z\,|\,x,y)=O(\log n)$ (given both x and y, we can simulate the protocol and compute the transcript and the key). We may assume that $C(t)\overset{\log}{\geq} n$ (otherwise the size of the key is negligibly small, [8]). On the other hand, since Alice and Bob each send at most $(0.5+\epsilon)n$ bits, we have $C(t)\overset{\log}{\leq}(1+2\epsilon)n$ and, moreover, $C(t\,|\,x)\overset{\log}{\leq}(0.5+\epsilon)n$ and $C(t\,|\,y)\overset{\log}{\leq}(0.5+\epsilon)n$.

Kolmogorov complexity of $z=\langle t,w\rangle$ is equal to $C(t)+C(w)$ (the mutual information between w and t is negligible since protocol reveals no information about the secret). However, conditional on x and conditional on y, Kolmogorov complexities of z and t are essentially the same (given the transcript t and the input of one of the parties, we can obtain the secret key w for free). It follows that

$$\begin{aligned} C(x\,|\,z) &\overset{\log}{=} C(x,z)-C(z)\overset{\log}{=}C(x)+C(z\,|\,x)-C(z)\\ &\overset{\log}{=} C(x)+C(t\,|\,x)-C(t,w)\overset{\log}{=}C(x)+C(t\,|\,x)-C(t)-C(w)\\ &\overset{\log}{\leq} 2n+(0.5+\epsilon)n-n-k\overset{\log}{=}(1.5+\epsilon)n-k. \end{aligned}$$

Similarly we obtain $C(y\,|\,z)\overset{\log}{\leq}(1.5+\epsilon)n-k$ and

$$\begin{aligned} C(x,y\,|\,z) &\overset{\log}{=} C(x,y,z)-C(z)\overset{\log}{=}C(x,y)+C(z\,|\,x,y)-C(t)-C(w)\\ &\overset{\log}{\geq} 3n+0-(1+2\epsilon)n-k\overset{\log}{=}(2-2\epsilon)n-k. \end{aligned}$$

If we assume now that $k=\frac{n}{2}\pm O(\epsilon n)$, we obtain

$$C(x\,|\,z)\overset{\log}{\leq}n+O(\epsilon n),\ C(y\,|\,z)\overset{\log}{\leq}n+O(\epsilon n),\ C(x,y\,|\,z)\overset{\log}{\geq}1.5n-O(\epsilon n),$$

which for small enough ϵ contradicts Theorem 3.8. ◀

▶ Remark 4.2. Theorem 4.1 states that, for the given setting, in a communication protocol in which each party sends approximately $\frac{1}{2}C(x\,|\,y)+O(\epsilon n)=\frac{1}{2}C(y\,|\,x)+O(\epsilon n)=n/2+O(\epsilon n)$ bits of information, the size of the secret key cannot attain $\frac{1}{2}I(x:y)=n/2$. Our proof (application of Theorem 3.8) actually implies a stronger bound: the size of the key cannot be greater than $3n/7+O(\epsilon n)\ll\frac{1}{2}I(x:y)$.

4.3 Field with a large subfield: a positive result

▶ **Theorem 4.3.** *Let $\mathbb{F}_q$ be a field with q elements, and $q=p^2$ for some integer p (e.g., p is prime and q is a square of this prime number, or $p=2^k$ and $q=2^{2k}$).*

Let $\mathbb{FP}$ be the projective plane over $\mathbb{F}_q$, and (x,y) be a typical incidence in this plane (x is a line in this projective plane, y is a point in this line, and $C(x,y)=3\log q\pm O(\log n)$). We consider communication protocols where Alice is given as her input x and Bob is given as his input y. We claim that there exists a communication protocol where

- *Alice sends a message m_A of length $n/2$ bits to Bob,*
- *Bob sends a message m_B of length $n/2$ bits to Alice,*

- *then Alice and Bob compute a secret key w of length $n/2$ such that*

$$C(w\,|\,\langle m_A, m_B\rangle) \overset{\log}{\geq} n/2,$$

where $n = \lceil \log q \rceil$, i.e., the protocol reveals virtually no information about the secret to the eavesdropper.

Proof. A point x and a line y in the projective plane $\mathbb{FP}$ can be specified by their projective coordinates $(x_0 : x_1 : x_2)$ and $(y_0 : y_1 : y_2)$ respectively. Without loss of generality, we assume $x_0 \neq 0$ and $y_2 \neq 0$ and denote

$$x_1' := x_1/x_0,\ x_2' := -x_2/x_0 \text{ and } y_0' := y_0/y_2,\ y_1' := y_1/y_2.$$

The incidence of x and y means that $x_0y_0 + x_1y_1 + x_2y_2 = 0$, or equivalently

$$y_0' + x_1'y_1' - x_2' = 0. \tag{9}$$

Since $q = p^2$, the field $\mathbb{F}_q$ contains a subfield $\mathbb{G}$ of size p, and there exists an element $\xi \in \mathbb{F}_q$ such that every element $\alpha \in \mathbb{F}_q$ can be represented as $\alpha = a_0 + a_1 \cdot \xi$ for some $a_0, a_1 \in \mathbb{G}$. So we may represent x_i' and y_i' as follows:

$$x_1' = f + r\xi,\quad y_0' = g + t\xi,\ y_1' = h + s\xi$$

for some $f, g, h, r, s, t \in \mathbb{G}$. In this notation, (9) rewrites to

$$(g + t\xi) + (f + r\xi)(h + s\xi) = x_2'.$$

It follows that

$$x_2' = g + fh + (t + fs + hr)\xi + rs\xi^2 \tag{10}$$

(The value ξ^2 can be represented as $u + v\xi$ for some $u, v \in \mathbb{G}$, but we do not need to specify these parameters.) Let us recall that Alice knows all parameters derived from x (including f, r, x_2'), and Bob knows all parameters derived from y (including g, h, s, t). We use the following protocol.

Communication protocol

Round 1 Bob sends to Alice the value $m_1 := s$ (this message consists of $\log|\mathbb{G}| = n/2$ bits of information)

Round 2 Alice computes $m_2 := g + fh$ and sends it to Bob (this message also consists of $\log|\mathbb{G}| = n/2$ bits of information)

Post-processing Both participants compute the value f and take it as the final result (the secret key, which also consists of $\log|\mathbb{G}| = n/2$ bits of information).

▷ Claim 4.4. Alice has enough information to compute m_2.

Proof of claim. Initially, Alice is given the values of $x_1' = f + r\xi$ and $x_2' = u' + v'\xi$, where f, r, u', v' are elements of $\mathbb{G}$. When she receives from Bob s, she gets all information to compute $rs\xi^2 = u'' + v''\xi$ (for some $u'', v'' \in \mathbb{G}$). From (10) it follows that $g + fh = u' - u''$. ◁

Alice is given the secret key f as a part of her input. Bob, however, needs to do some computation to get this value.

▷ Claim 4.5. Bob has enough information to compute the final result f.

Proof of claim. Initially, Bob was given the values $g, t, h, s \in \mathbb{G}$. Bob receives from Alice the value $g + fh$, which is another element of the field $\mathbb{G}$. This allows him to compute f as $((g + fh) - g) \cdot h^{-1}$. ◁

It remains to show that we reveal no information to the eavesdropper. The adversary can intercept the messages $m_1 = s$ and $m_2 = g + fh$. We need to show that these messages give no information about the produced secret key:

▷ Claim 4.6. $I(f : \langle m_1, m_2 \rangle) = O(\log n)$.

Proof of claim. To specify the incidence (x, y), it is enough to provide the values f, g, h, r, s, t in $\mathbb{G}$. Thus, we have

$$\begin{aligned}
C(x,y) &\overset{\log}{=} C(f,g,h,s,r,t)\\
&\overset{\log}{\leq} C(m_1) + C(m_2) + C(f,g,h,s,r,t \,|\, m_1,m_2)\\
&\overset{\log}{\leq} C(m_1) + C(m_2) + C(s \,|\, m_1,m_2) + C(f \,|\, m_1,m_2) + C(h)\\
&\quad + C(g \,|\, m_1,m_2,f,h) + C(r) + C(t)\\
&\overset{\log}{\leq} C(m_1) + C(m_2) + C(f \,|\, m_1,m_2) + C(h) + C(r) + C(t)\\
&\overset{\log}{\leq} 5\log|\mathbb{G}| + C(f \,|\, m_1,m_2)\\
&\overset{\log}{=} \tfrac{5}{2}n + C(f \,|\, m_1,m_2)
\end{aligned}$$

(in this calculation, $C(s \,|\, m_1, m_2)$ vanishes since $m_1 = s$, and $C(g \,|\, m_1, m_2, f, h)$ vanishes since we can compute g given f, h and the value of $g + fh$).

Since the incidence (x, y) is typical, i.e., $C(x, y) \overset{\log}{=} 3n$, we obtain $C(f \,|\, m_1, m_2) \overset{\log}{\geq} \frac{n}{2}$. Thus, $C(f \,|\, m_1, m_2) \overset{\log}{\geq} C(f)$, and the claim is proven. ◁

◀

5 Conclusion

We have shown that the Kolmogorov complexities of a triple of correlated strings cannot, in general, be reduced by a constant factor through conditioning. This impossibility relies on algebraic structures arising in incidence geometry over fields without proper subfields. Our argument connects recent advances in bounds on point-line incidences with previously developed techniques in algorithmic information theory and communication complexity. This suggests broader prospects for the algebraic and combinatorial methods developed in [2, 10–12, 26] for studying fundamental barriers in information theory and communication complexity. We emphasize that Question 1 on page 11 (see also [20]) remains widely open.

References

1 Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography. i. secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, 1993. doi:10.1109/18.243431.

2 Jean Bourgain, Nets Katz, and Terence Tao. A sum-product estimate in finite fields, and applications. *Geometric and Functional Analysis*, 14:27–57, 2004.

3 Geoffroy Caillat-Grenier, Andrei Romashchenko, and Rustam Zyavgarov. Common information in well-mixing graphs and applications to information-theoretic cryptography. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 181–186, 2024. doi:10.1109/ITW61385.2024.10806994.

4 Gregory J. Chaitin. On the simplicity and speed of programs for computing infinite sets of natural numbers. *Journal of the ACM (JACM)*, 16(3):407–422, 1969. doi:10.1145/321526.321530.

5 Alexei Chernov, Andrej Muchnik, Andrei Romashchenko, Alexander Shen, and Nikolai Vereshchagin. Upper semi-lattice of binary strings with the relation "x is simple conditional to y". *Theoretical Computer Science*, 271(1-2):69–95, 2002. doi:10.1016/S0304-3975(01)00032-9.

6 Lance Fortnow. Kolmogorov complexity and computational complexity. In *Complexity of Computations and Proofs*, volume 13 of *Quaderni di Matematica*. De Gruyter, 2004.

7 Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2:149–162, 1973.

8 Emirhan Gürpınar and Andrei Romashchenko. Communication complexity of the secret key agreement in algorithmic information theory. *ACM Transactions on Computation Theory*, 16(3):1–37, 2020.

9 Daniel Hammer, Andrei Romashchenko, Alexander Shen, and Nikolai Vereshchagin. Inequalities for shannon entropy and kolmogorov complexity. *Journal of Computer and System Sciences*, 60(2):442–464, 2000. doi:10.1006/jcss.1999.1677.

10 Harald Andrés Helfgott and Misha Rudnev. An explicit incidence theorem in $\mathbb{F}_p$. *Mathematika*, 57(1):135–145, 2011.

11 Timothy G. F. Jones. Further improvements to incidence and beck-type bounds over prime finite fields, 2012. arXiv:1206.4517.

12 Timothy G. F. Jones. An improved incidence bound for fields of prime order. *European Journal of Combinatorics*, 52:136–145, 2016. doi:10.1016/j.ejc.2015.09.004.

13 Andrei N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1(1):1–7, 1965.

14 Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 4 edition, 2019.

15 Zhenjian Lu and Igor C. Oliveira. Theory and applications of probabilistic kolmogorov complexity. *Bulletin of EATCS*, 137(2):44, 2022. URL: http://bulletin.eatcs.org/index.php/beatcs/article/view/700.

16 Ueli M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, 39(3):733–742, 1993. doi:10.1109/18.256484.

17 Andrej A. Muchnik. On common information. *Theoretical Computer Science*, 207(2):319–328, 1998. doi:10.1016/S0304-3975(98)00070-X.

18 Mikhail M. Postnikov. *Lectures in Geometry. Smooth Manifolds*. Mir pulishers, Moscow, 1989.

19 Andrei Romashchenko and Alexander Shen. Topological arguments for kolmogorov complexity. *Theory of Computing Systems*, 56(3):513–526, 2015. doi:10.1007/s00224-015-9606-8.

20 Andrei Romashchenko, Alexander Shen, and Marius Zimand. 27 open problems in kolmogorov complexity. *ACM SIGACT News*, 52(4):31–54, 2022. doi:10.1145/3510382.3510389.

21 Andrei Romashchenko and Marius Zimand. An operational characterization of mutual information in algorithmic information theory. *Journal of the ACM (JACM)*, 66(5):1–42, 2019. doi:10.1145/3356867.

22 Alexander Shen, Vladimir Uspensky, and Nikolay Vereshchagin. *Kolmogorov Complexity and Algorithmic Randomness*, volume 220. American Mathematical Society, 2017.

23 Ray J. Solomonoff. A preliminary report on a general theory of inductive inference. Technical report, Zator Company, Cambridge, MA, 1960.

24 Ray J. Solomonoff. A formal theory of inductive inference. Part I. *Information and Control*, 7(1):1–22, 1964. doi:10.1016/S0019-9958(64)90223-2.

25 Ray J. Solomonoff. A formal theory of inductive inference. Part II. *Information and Control*, 7(2):224–254, 1964. doi:10.1016/S0019-9958(64)90131-7.

26 Sophie Stevens and Frank De Zeeuw. An improved point-line incidence bound over arbitrary fields. *Bulletin of the London Mathematical Society*, 49(5):842–858, 2017.

27 Himanshu Tyagi and Shun Watanabe. *Information-theoretic Cryptography.* Cambridge University Press, 2023.

28 Aaron D. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975. doi:10.1109/TIT.1975.1055346.

29 Raymond W. Yeung. Facets of entropy. *Communications in Information and Systems*, 15(1):87–117, 2015. doi:10.4310/cis.2015.v15.n1.a6.

30 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.