Sharper Upper Bounds for Unbalanced Uniquely Decodable Code Pairs

Per Austrin, Petteri Kaski, Mikko Koivisto, and Jesper Nederlof

Abstract—Two sets of 0–1 vectors of fixed length form a uniquely decodable code pair if their Cartesian product is of the same size as their sumset, where the addition is pointwise over integers. For the size of the sumset of such a pair, van Tilborg has given an upper bound in the general case. Urbanke and Li, and later Ordentlich and Shayevitz, have given better bounds in the unbalanced case, that is, when either of the two sets is sufficiently large. Improvements to the latter bounds are presented.

Index Terms—Additive combinatorics, binary adder channel, isoperimetric inequality, uniquely decodable code pair, zero-error capacity.

I. INTRODUCTION

A CANONICAL problem in multi-user communication theory is how to coordinate unambiguous communication through a multiple access channel, such that several independent senders can simultaneously send as much information as possible to a single receiver (see, e.g., the book by Cover and Thomas [1, Chapter 15]); this could for example occur when several satellites need to send their data to a single terminal.

Unfortunately, despite vast research in the last decades, even in some of the simplest models the zero-error capacity of such communication channels remains far from clear. An extensively investigated and fundamental example is the two-user binary adder channel (BAC). The zero-error capacity of the BAC is equal to the maximum size of the product of the code sizes of a uniquely decodable code pair (UDCP); a pair $A, B \subseteq \{0, 1\}^n$ such that $|A + B| = |A| \cdot |B|$ where $A + B$ denotes the sumset $\{a + b : a \in A, b \in B\}$, and $a + b$ denotes addition over $\mathbb{Z}^n$.

P. Austrin was supported by the Swedish Research Council, under Grant 621-2012-4546. P. Kaski was supported by the European Research Council, under Grant 338077. M. Koivisto was supported by the Academy of Finland, under Grant 276864. J. Nederlof was supported by the NWO VENI project 639.021.438. This paper was presented in part at the 2016 IEEE International Symposium on Information Theory.

P. Austrin is with the School of Computer Science and Communication, KTH Royal Institute of Technology, Sweden (e-mail: austrin@csc.kth.se).

P. Kaski is with the Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University, Finland (e-mail: petteri.kaski@aalto.fi).

M. Koivisto is with the Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland (e-mail: mikko.koivisto@helsinki.fi).

J. Nederlof is with the Department of Mathematics and Computer Science, Technical University of Eindhoven, The Netherlands (e-mail: j.nederlof@tue.nl).

Communicated by S. Mesnager, Associate Editor for Sequences.

Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purpose must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Most previous research on UDCPs has focused on constructions. A basic observation is that, if $A_1, B_1 \subseteq 2^{[n]}$ is a UDCP and $A_2, B_2 \subseteq 2^{[n]}$ is a UDCP, then $A_1 \times A_2, B_1 \times B_2$ is also a UDCP; here and henceforth, we freely interchange vectors with sets in the natural way. Therefore, for finding asymptotically good constructions for every $n$, it is sufficient to focus on finite $n$. Letting $\alpha$ and $\beta$ denote respectively $\log_2(|A|)/n$ and $\log_2(|B|)/n$, a natural and popular goal is to find a UDCP maximizing $\alpha + \beta$. The most direct construction is to let $A$ be all strings where the first $\beta n$ coordinates are fixed to 0, and $B$ be all strings which use only the first $\beta n$ coordinates. This yields any pair $(\alpha, \beta)$ with $\alpha + \beta = 1$. The simplest non trivial construction, $A = \{00, 01, 11\}$, $B = \{10, 01\}$ giving $\alpha + \beta = \log_2(3 + 1)/2 \approx 1.29248$, was presented by Kasami and Lin [2]. This was the best until 1985. Then it was improved to 1.30366 by van den Braak and van Tilborg [3], and after subsequent improvements by Ahlswede and Balakirsky [4], van den Braak [5], Urbanke and Li [6], the current record is 1.31781 by Mattas and Østergård [7]. Several of these results were obtain by computer searches for finite $n$. More relevant to our study is the important work by Kasami et al. [8], which shows that for sufficiently large $n$ there exist (somewhat surprisingly) UDCPs with $\alpha \geq 1 - o(1)$ and $\beta \geq 0.25$.

Considering upper bounds, the rather direct $\alpha + \beta \leq 1.5$ has been independently found by at least Liu [9], Ahlswede [10], Lindström [11] and van Tilborg [12]. Leaving a gap to the lower bound, 1.5 is still the best upper bound known on $\alpha + \beta$ in general. However, Urbanke and Li [6] managed to break through the 1.5 bound in the unbalanced case: assuming $\alpha \geq 1 - \epsilon$ for a sufficiently small value of $\epsilon$, they showed that $\beta \leq 0.4921$. On a high level, their approach works as follows: a result of van Tilborg [12] (see Lemma 1 below) shows there are not many pairs $(a, b) \in A \times B$ of small Hamming distance, and if $A$ and $B$ are sufficiently large, then the number of such pairs is bounded from below by an isoperimetric inequality for which the authors use Harper’s theorem. Later, this result was improved to $\beta \leq 0.4798$ by Ordentlich and Shayevitz [13].

Their proof idea is somewhat more involved: the authors give a procedure that, given a UDCP $A, B \subseteq \{0, 1\}^n$, constructs another UDCP $C, D \subseteq \{0, 1\}^{(1-\gamma)n}$ of comparable size for some $\gamma > 0$. This was achieved by proving the existence of a subset $L \subseteq [n]$ with $|L| = \gamma n$ such that for some $c \in \{0, 1, 2\}^{|L|}$, the projection $(a + b)_L$ equals $c$ for many pairs $a, b$. The existence of such a subset is proved using a variant of the Sauer–Perles–Shelah lemma. Unfortunately, both the referred bounds [6], [13] converge fast to $(1 - \epsilon) + \beta \leq 1.5$ as $\epsilon$ increases (see Figure 1 of Ordentlich and Shayevitz [13]).
The present authors [14] gave a novel and direct connection between UDCPs and additive number theory. Motivated by algorithm design for the Subset Sum problem, they observed the following: if \( w \in \mathbb{Z}^n, t \in \mathbb{Z} \) and \( A \subseteq \{0,1\}^n \) such that \( a \cdot w = a' \cdot w \) implies \( a = a' \) for every \( a, a' \in A \), and \( B = \{ b \in \{0,1\}^n : w \cdot b = t \} \), then \( A, B \) is a UDCP. Here \( \cdot \) denotes the inner product.

The channel capacity application has also inspired studies of several variants of the basic setting of this paper, for example, with both sets being the same [15], [16], with noise [17], or with more than two users [10], [18], [19].

Contributions

Motivated by the lack of progress on the large gap between the current lower and upper bounds for UDCPs, we propose to restrict attention to the case \(|A| \geq 2^{(1-\epsilon)n}\) for small values of \( \epsilon \): before we can understand the exact tradeoff between \( \alpha \) and \( \beta \), we first need to understand this tradeoff for large values of \( \alpha \). A natural question is whether \( \alpha \geq 1 - o(1) \) implies \( \beta \leq 0.25 + o(1) \); in other words, is the construction of Kasami et al. [8] optimal, or could it be improved? While the present work does not settle this question, we narrow the gap by pushing the upper bound closer to 0.25. Our main result is the following:

**Theorem 1 (Main):** Suppose \( A, B \subseteq \{0,1\}^n \) is a UDCP with \(|A| \geq 2^{(1-\epsilon)n}\) and \(|B| = 2^n\). Then \( \beta \leq 0.4228 + \sqrt{\epsilon} \).

Our proof combines ideas from both previous upper bounds [6], [13] with new ideas. We will present our proof by first providing a “warm-up” bound of \( \beta \leq 0.4777 + O(\sqrt{\epsilon}) \) (Theorem 2). To establish this bound, we study the joint probability \( \Pr[a \in A, b \in B] \) for two correlated random vectors \( a, b \in \{0,1\}^n \). We bound this probability from above and below using, respectively, van Tilborg’s lemma (Lemma 1) and an isoperimetric inequality due to Mossel et al. [20]. This approach is similar to that of Urbanke and Li [6], but improves their bound for small values of \( \epsilon \).

The intuition behind our main bound (and also, in part, the bounds of Urbanke and Li [6] and Ordentlich and Shayeivitz [13]) is as follows. The above strategy does not give a good bound if \( A \) and \( B \) are antipodal Hamming balls: the studied probability is very small in this case, so the upper bound is not really stringent. However, intuitively such a pair cannot form a large UDCP since the pairwise sums will be concentrated on the sum of the two centers of the Hamming balls. Our novel approach is that we use the encoding argument from van Tilborg’s lemma to show that if \( A \) is large enough, then \( B \) needs to be sufficiently spread out over the hypercube. Specifically, we show that there exists a set \( L \subseteq [n] \) of size close to \( n/2 \) such that \( L \) has an almost maximum number of projections on \( B \). Subsequently, we use this set \( L \) to define a refined distribution of the vectors \( x \) and \( y \). In the refined distribution, \( x, y \) are only correlated in the coordinates from \( L \), and for applying the isoperimetric inequality the large number of projections is then essential.

II. NOTATION AND PRELIMINARIES

A. Notation

Given reals \( a, b \) with \( b \geq 0 \), we write \( a \pm b \) for the interval \([a - b, a + b]\). If \( n \) is an integer, we denote by \([n] \) the set \( \{1, \ldots, n\} \). For a vector \( x \in \mathbb{R}^n \), we let \( x^{-1}(z) \subseteq [n] \) denote the set of coordinates \( i \) such that \( x_i = z \). For binary vectors, we apply the usual set operations in the obvious way, by interpreting a vector \( x \in \{0,1\}^n \) as the set \( x^{-1}(1) \subseteq [n] \). For example, \( x \setminus y \) is a vector whose \( i \)th entry is 1 if \( x_i = 1 \) and \( y_i = 0 \), and 0 otherwise; \( x \triangle y \) denotes the symmetric difference (or alternatively, the componentwise XOR) of \( x \) and \( y \); and \( |x| \) denotes the Hamming weight of \( x \). Given a vector \( x \in \{0,1\}^n \) and a subset \( P \subseteq [n] \), we let \( x_P \) denote the projection of \( x \) on \( P: x_P \in \{0,1\}^P \) such that \( x_P \) agrees with \( x \) on all coordinates in \( P \). For a family \( X \subseteq \{0,1\}^n \) we also write \( X_P \) for the family \( \{ x_P : x \in X \} \).

We write \( o(1) \) for all terms that tend to zero when \( n \) tends to infinity. Such terms can be safely ignored for our purposes as no other variables will depend on \( n \) and upper bounds for UDCPs of large dimension imply upper bounds for UDCPs of finite dimension due to the construction mention in Section I.

B. Entropy

For a real \( x \in [0,1] \) we denote by \( h(x) \) the binary entropy of \( x \), that is, \( h(x) = -x \log_2(x) - (1-x) \log_2(1-x) \). It is well known that \( h(x) \) is monotonically increasing for \( x \in [0,1/2] \), monotone decreasing for \( x \in [1/2,1] \), and that \((\gamma^n) \leq 2^{h(n)/2}\).

The following inequality can be shown by standard calculus:

**Observation 1:** For all \( x \in [0,1/2] \), \( h\left(\frac{1}{2} + x\right) < \frac{1}{2} \frac{3}{\pi^2} x^2 \).

This observation implies another useful bound:

**Observation 2:** Let \( \epsilon > 0 \) be a constant. Let \( X \subseteq \{0,1\}^n \) such that \( |X| \geq 2^{(1-\epsilon)n} \), \( z \in \{0,1\}^n \), and \( \gamma \geq \sqrt{\ln(2)/\epsilon} \). Then \( |\{x \in X : |x \triangle z| \leq \gamma n \}| \geq |X|/2 \) for all sufficiently large \( n \).

To see this, note that

\[
\{x \in X : |x \triangle z| \leq \gamma n \} \leq 2 \sum_{k=0}^{n\gamma} \binom{n}{k} \leq n^{2h(1/2 - \gamma)}.
\]

Since \( h\left(\frac{1}{2} - \gamma\right) < 1 - \frac{2}{\pi^2} \gamma^2 = 1 - \epsilon \), there is some \( \epsilon' > 0 \) (depending only on \( \epsilon \)) such that \( h\left(\frac{1}{2} - \gamma\right) = 1 - \epsilon - \epsilon' \). Then

\[
n^{2h(1/2 - \gamma)} = n^{2-\epsilon' n} 2^{(1-\epsilon)n} n^{2-\epsilon' n} |X|
\]

which, for all sufficiently large \( n \), is smaller than \( |X|/2 \).

C. UDCPs

We will use the following well known property of UDCPs that directly follows from noting that whenever \( a-b = a'-b' \) we have \( a+b = a'+b' \):

**Observation 3:** If \( A, B \) is a UDCP, then \( |A-B| = |A| \cdot |B| \).

We will also use the following bound. Since the proof is elegant and highly instructive for understanding our approach, we provide a (known) proof.
Lemma 1 (van Tilborg [12]): Let \( A, B \subseteq \{0,1\}^n \) be a UDCP and let \( W_d = \{ (a, b) \in A \times B : |a \Delta b| = d \} \).
Then \( |W_d| \leq \binom{n}{d} 2^{\min\{d,n-d\}} \).

Proof: Let us bound the number of possibilities for \( a+b \) and \( b-a \) for pairs \((a, b) \in W_d\). Note that
\[
(a \triangle b) = (a+b)^{-1}(1) \setminus (b-a)^{-1}(0).
\]
Thus, since \(|a \triangle b| = d\), fixing \( a \triangle b \) (in one of the \( \binom{n}{d} \) possible ways) leaves either \( 2^{n-d} \) possible choices for \((a+b)^{-1}(0)\) and \((a+b)^{-1}(2)\), or \( 2^d \) possible choices for \((b-a)^{-1}(1)\) and \((b-a)^{-1}(1)\). By the UDCP property, either of these two completely determines \((a, b) \in W_d\), and the bound follows.

D. \( \rho \)-Correlation and Isoperimetry

For \( x \in \{0,1\}^U \), we write \( y \sim_{\rho} x \) for a \( \rho \)-correlated copy of \( x \), i.e., a vector where, independently for each \( e \in U \),
\[
y_e = \begin{cases} x_e, & \text{with probability } \frac{1+\rho}{2}, \\ 1-x_e, & \text{with probability } \frac{1-\rho}{2}. \end{cases}
\]
If \( x \) is not fixed, we use \( y \sim_{\rho} x \) to denote the joint distribution over \((x,y)\) where \( x \) is a uniformly random vector and \( y \) is a \( \rho \)-correlated copy of \( x \). Our bounds will rely on the reverse small-set expansion theorem, an isoperimetric inequality of the noisy Boolean hypercube [20].

Lemma 2 (Reverse Small-Set Expansion [20, Th. 3.4]): For all \( \rho \in [0,1) \) the following holds. Let \( F,G \subseteq \{0,1\}^U \) with \( |F| \geq 2^{\ell(U)} \) and \( |G| \geq 2^{\ell(U)} \).
Then
\[
\Pr_{y \sim_{\rho} x} [x \in F, y \in G] \geq 2^{-\ell(U)} \left( \frac{1+\rho}{2} + 2\sqrt{\rho(1-\rho)} \right). 
\]

IV. PROOF OVERVIEW OF MAIN BOUND

The proof of our main bound follows the same blueprint as the proof of Theorem 2, but we use a more refined version of the noise distribution. In particular, we only apply the noise on a subset of \([n]\) where both \( A \) and \( B \) are sufficiently dense, e.g. have sufficiently many projections to that subset.

Definition 1: Fix \( L \subseteq [n] \). Given \( x \in \{0,1\}^n \) let \( y \sim_{\rho}^L x \) denote that \( y \in \{0,1\}^n \) is the random variable distributed as follows:
\[
y_i \sim_{\rho} x_i, \quad \text{if } i \in L,
y_i \sim_{0} x_i, \quad \text{if } i \notin L.
\]
In other words, \( y \) is a \( \rho \)-correlated copy of \( x \) on the coordinates of \( L \), and uniformly random outside \( L \).

We proceed to give upper and lower bounds on the quantity \( \Pr_{y \sim_{\rho}^L} [a \in A, b \in B] \). In order for these bounds to hold, we need a mild density condition on \( A \) with respect to the split \((L, [n] \setminus L)\). In particular, we make the following definition.

Definition 2: We say that \( A \subseteq \{0,1\}^n \) is \( \epsilon \)-dense with respect to \( L \subseteq [n] \) if \( |A_L| \geq 2^{|L|-\epsilon n-1} \) and for every \( a \in A \), the number of \( a' \in A \) such that \( a_L = a'_L \) is at least \( 2^{|L|-\epsilon n-1} \).

As the following simple claim shows, our set \( A \) is guaranteed to have a dense subset.

Claim 1: Let \( A \subseteq \{0,1\}^n \) such that \( |A| \geq 2^{(1-\epsilon)n} \). Then for any \( L \subseteq [n] \), there is an \( A' \subseteq A \) that is \( \epsilon \)-dense with respect to \( L \).

Proof: For \( a, a' \in A \) note that the condition \( a_L = a'_L \) is an equivalence relation partitioning \( A \) into at most \( 2^{|L|} \) equivalence classes, each of size at most \( 2^{n-|L|} \). It follows that there must be at least \( |A|/2^{n-|L|+1} \geq 2^{|L|-\epsilon n-1} \) equivalence classes of size at least \( |A|/2^{|L|+1} \geq 2^{n-|L|-\epsilon n-1} \) and we can take \( A' \) to be the union of these.

With these definitions in place, we are ready to state the precise upper and lower bounds on the refined noise probability.
Lemma 3: Fix $L \subseteq [n]$ and let $\lambda = |L|/n$. Then for any $0 \leq \rho \leq 1$ and UDCPs $(A, B)$ such that $A$ is $\epsilon$-dense with respect to $L$, we have
\[
\log_2 \Pr_{a \sim \mu_B}[a \in A, b \in B] n \leq \sqrt{\frac{\ln(2)\epsilon}{2}} - \frac{1}{2} + \lambda \left(\log_2(3 - \rho) - \frac{3}{2}\right) + o(1).
\]

The proof appears in Section VI.

Lemma 4: Fix $L \subseteq [n]$ and let $\lambda = |L|/n$. Then for any $0 \leq \rho < 1$ and UDCPs $(A, B)$ such that $A$ is $\epsilon$-dense with respect to $L$ and $|B_L| = 2^{\rho n}$ for some $0 \leq \pi \leq \lambda$, we have
\[
\log_2 \Pr_{a \sim \mu_B}[a \in A, b \in B] n \geq \pi - \lambda - \epsilon - 2\rho \sqrt{\epsilon(\lambda - \pi)} - 1 - \rho^2 + \lambda - 1 - \epsilon - o(1).
\]

The proof appears in Section VII.

The bound on the lower bound depends on the size of $|B_L|$ and in particular we would like to find a split $L$ such that $|B_L| \approx |B|$. At the same time we would like $|L|$ to be as small as possible. The following lemma shows that we can take $|L| \approx n/2$ and still have $|B_L| \approx |B|$.

Lemma 5: For sufficiently large $n$ and UDCPs $(A, B)$ such that $|A| \geq 2^{(1-\epsilon)n}$, $|B| = 2^{\rho n}$, there exists $L \subseteq [n]$ such that $|L| \in \left[\frac{n}{2} \pm \sqrt{\ln(2)\epsilon/2}\right]$ and $|B_L| \geq 2^{(\beta - \epsilon)n-1}$.

Proof: Let $P \subseteq A \times B$ consist of all pairs $(a, b)$ such that $|a \Delta b| \in \left[\frac{1}{2} \pm \sqrt{\ln(2)\epsilon/2}\right]$. We have that
\[
|P| = \sum_{b \in B} \left|\left\{ a \in A : |a \Delta b| \in \left[\frac{1}{2} \pm \sqrt{\ln(2)\epsilon/2}\right]\right\}\right| \\
\geq \sum_{b \in B} |A|/2 = |A| \cdot |B|/2,
\]
where the inequality is by Observation 2. Similarly as in the proof of Lemma 1, consider the encoding
\[
\eta : (a, b) \mapsto (a \Delta b, b \setminus a).
\]

By Observation 3, $|A - B| = |A| \cdot |B|$, and since $a - b$ can be computed from $\eta(a, b)$, it follows that $\eta$ is injective and $|\eta(P)| = |P|$. We now bound $|\eta(P)|$ from above. To this end, note that $b \setminus a \subseteq a \Delta b$, and so $b \setminus a \in B_{a \Delta b}$. (More precisely, $b \setminus a$ projected to $a \Delta b$ is only needed that $b \setminus a$ can be described by a single element of $B_{a \Delta b}$.) Therefore, by summing over the possible values of $X = a \Delta b$ we have that
\[
|\eta(P)| \leq \sum_{X \subseteq [n]} |B_X|.
\]

Thus there must be an $X \subseteq [n]$ with $|X| \in \left[\frac{1}{2} \pm \sqrt{\ln(2)\epsilon/2}\right]$, and $|B_X| \geq |\eta(P)|/2^n = |P|/2^n \geq |A| \cdot |B|/2^{n+1} \geq 2^{(\beta - \epsilon)n-1}$, as we claimed.

V. COMBINING THE BOUNDS: PROOF OF THEOREM 1

We prove our main theorem by combining Lemmata 3, 4, and 5. To this end, let $A, B \subseteq [0, 1]^n$ be a UDCP with $|A| \geq 2^{(1-\epsilon)n}$ and $|B| = 2^{\rho n}$. We will show that $\beta \leq 0.4228 + \sqrt{\epsilon}$.

Without loss of generality, we may assume that $n$ is sufficiently large for all estimates to hold, since a lower bound for large $n$ also holds for small $n$: if $(A_1, B_1)$ and $(A_2, B_2)$ are UDCPs, then so is $(A_1 \times A_2, B_1 \times B_2)$.

By Lemma 5, there exists a partition $L, R$ of $[n]$ such that $\lambda = |L|/n \in \left[\frac{1}{2} \pm \sqrt{\ln(2)\epsilon/2}\right]$, and $2^{\rho n} := |B_L| \geq 2^{(\beta - \epsilon)n-1}$. By Claim 1, there is an $A' \subseteq A$ such that $A$ is $\epsilon$-dense with respect to $L$.

Applying Lemmata 3 and 4 to the UDCP $(A', B)$ we then obtain that
\[
\pi - \lambda - \epsilon - 2\rho \sqrt{\epsilon(\lambda - \pi)} + \lambda - 1 - \epsilon - o(1) \\
\leq \log_2 \Pr_{a \sim \mu_B}[a \in A', b \in B] n \\
\leq \sqrt{\frac{\ln(2)\epsilon}{2}} - \frac{1}{2} + \lambda \left(\log_2(3 - \rho) - \frac{3}{2}\right) + o(1).
\]

Simplifying, we get
\[
\pi \leq \left(\sqrt{\frac{\ln(2)\epsilon}{2}} + \frac{3}{2} + \epsilon + \lambda \cdot \left(\log_2(3 - \rho) - \frac{3}{2}\right)(1 - \rho^2) + 2\rho \sqrt{\epsilon(\lambda - \pi)} + \epsilon + \lambda + o(1)\right).
\]

We now set $\rho = 0.654$. Plugging in this value and simplifying, (1) becomes
\[
\pi \leq 0.2861421 + 0.2733156\lambda + 1.573\epsilon + 0.33691\sqrt{\epsilon + 1.308\sqrt{\epsilon(\lambda - \pi)}} + o(1).
\]

Using $\lambda \leq \frac{1}{2} + \sqrt{\ln(2)\epsilon/2}$ and simplifying further, we get
\[
\pi \leq 0.4228 + 1.573\epsilon + o(1) \\
+ \left(0.4979 + 1.308\sqrt{\frac{1}{2} + \sqrt{\ln(2)\epsilon/2} - \pi}\right)\sqrt{\epsilon}.
\]

Since $\beta \leq \pi + \epsilon + o(1)$, we would like to show that $\pi < 0.4228 + \sqrt{\epsilon}$. Assume for the sake of contradiction that $\pi \geq 0.4228 + \sqrt{\epsilon}$. Plugging this into (2) gives
\[
0 < 2.573\epsilon + o(1) + \left(0.4979 - 1 + 1.308\sqrt{0.0772 + \sqrt{\ln(2)\epsilon/2} - \pi - \epsilon}\right)\sqrt{\epsilon}.
\]

For $0 \leq \epsilon \leq 0.01$, it can be verified using a computer that the right-hand side of (3) is non-positive, yielding the desired contradiction (for sufficiently large $n$), and proving that $\beta < 0.4228 + \sqrt{\epsilon}$. For $\epsilon > 0.01$, we have $\beta < 0.5 + \epsilon < 0.4228 + \sqrt{\epsilon}$ (the first inequality being the classic $|B| \leq 2^{1.5n}/|A|$ upper bound). This completes the proof.

VI. UPPER BOUND: PROOF OF LEMMA 3

In this section, we prove the upper bound on the refined noise probability stated in Lemma 3. Fix $L \subseteq [n]$ and let $\lambda = |L|/n$. Furthermore, let $0 \leq \rho \leq 1$ and let $(A, B)$ be a UDCP such that $|A|$ is $\epsilon$-dense with respect to $L$.

Let $R = [n] \setminus L$ be the coordinates not in $L$. Let $W_d$ be the set of pairs $a_L a_R \in A, b_L b_R \in B$ such that $|a_L \Delta b_L| = d$.

Claim 2: For sufficiently large $n$, we have that
\[
|W_d| \leq \left(\frac{|L|}{d}\right)^2 2^{d(2.5 |L| / \ln(2)\epsilon/2 + 1)}.
\]

Copyright (c) 2017 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.
Proof: Let \( \epsilon' = \sqrt{(\epsilon \ln 2)/2(1-\lambda)} \), and let \( W'_d \subseteq W_d \) be all pairs from \( W_d \) such that \( |aR_b|/|R| \) \( \geq \frac{\epsilon}{2} \). Similarly as in the proof of Lemma 5, we see that

\[
|W'_d| = \sum_{b_1, b_2, \ldots, b_{|R|} \in b} \left| \{aR : aR \in A, \sum_{b_i \in A} (a_i, a_{i+1}, \ldots, a_{i+|R|-1}) \} \right| 
\geq \sum_{b_1, b_2, \ldots, b_{|R|} \in b} \frac{1}{2}(|aR \in \{0, 1\}^{R} : aR, aR \in A, \sum_{b_i \in A} (a_i, a_{i+1}, \ldots, a_{i+|R|-1}) = \frac{1}{2}|W_d|.
\]

The inequality follows from Observation 2 combined with the \( \epsilon' \)-dense property:

\[
|\{aR \in \{0, 1\}^{R} : aR, aR \in A, \sum_{b_i \in A} (a_i, a_{i+1}, \ldots, a_{i+|R|-1}) = \frac{1}{2}|W_d| \} \geq 2^{|R| - \epsilon n - 1} = 2^{(1 - \frac{\epsilon}{\ln 2})|R| - 1}.
\]

We proceed to bound \( |W'_d| \) from above. Similarly as in the proof of Lemma 1, define an encoding \( \eta \) on elements \((a, b)\) of \( W'_d \):

\[
\eta: (aR, aL, bL) \mapsto (aL, bL, aL \triangle bL, aR \triangle bR, aR \triangle bR).
\]

Since the image \( \eta(a, b) \) directly gives \( a - b \) and \( |A - B| = |A||B| \) by Observation 3, we have that \( \eta \) is injective and thus

\[
|W'_d| = |\eta(W'_d)| \leq \left( \frac{|R|}{d} \right)^{2d} \sum_{\epsilon \in (0.5, \epsilon')} |\eta(R)|^{2i},
\]

where the inequality follows by bounding the number of possibilities in every coordinate of \( \eta(\cdot) \). The claim is then implied for sufficiently large \( n \) from the easy observation that

\[
\sum_{\epsilon \in (0.5, \epsilon')} (|\eta(R)|^{2i}) \leq 2^{(1.5 + \epsilon')|R|} \leq 2^{1.5|R| + n \sqrt{\ln(2)}/2}.
\]

By the refined definition of \( \sim_{\rho} \) we have that

\[
Pr_{a \sim_{\rho} b}[a \in A, b \in B] = 2^{-n} \sum_{d=0}^{|L|} \left( \frac{1 + \rho}{2} \right)^{d} \left( \frac{1 - \rho}{2} \right)^{d} 2^{-|R||W_d|}. \tag{4}
\]

To see this, note that \( |W_d| \) counts exactly the pairs \( a \in A, b \in B \) satisfying \( |aL \triangle bL| = d \), and that the probability that such a pair is picked can be computed as the probability that \( a \) is picked (which is \( 2^{-n} \)) times the probability that \( b \) is picked given that \( a \) is given. The probability that \( bR \) is picked is simply \( 2^{-|R|} \) since it is picked uniformly at random, and the probability that \( bL \) is picked is \( \left( \frac{1 + \rho}{2} \right)^{d} \left( \frac{1 - \rho}{2} \right)^{d} \), similarly as in the proof of Theorem 2.

Using Claim 2, we bound (4) from above by

\[
Pr_{a \sim_{\rho} b}[a \in A, b \in B] \leq 2^{-2n} \sum_{d=0}^{|L|} (1 + \rho)^{|L| - d} (1 - \rho)^{d} \left( \frac{|L|}{d} \right)^{2d} 2^{1.5|R| + \ln(2)\sqrt{n}/2} + 1
\]

\[
= 2^{-2n + 1.5|R| + \ln(2)\sqrt{n}/2} + 1 \sum_{d=0}^{|L|} (1 + \rho)^{|L| - d} (2 - 2\rho)^{d} \left( \frac{|L|}{d} \right)^{2d}
\]

\[
= 2^{\left( \ln(2)\sqrt{n}/2 - 1 \right) - \frac{|L|}{2}} 2^{n + 1.5|L| + 1} (3 - \rho)^{|L|}.
\]

where the last equality follows from the binomial theorem. Using \( |R| = n - |L| \), taking logs, and dividing by \( n \), we get

\[
\log_2 Pr_{a \sim_{\rho} b}[a \in A, b \in B] \leq \sqrt{\frac{\ln(2)}{2} - \frac{1}{2}}
\]

\[
+ \lambda \left( \log_2(3 - \rho) - \frac{3}{2} \right) + 1/n.
\]

VII. LOWER BOUND: PROOF OF LEAK 4

In this section, we prove the lower bound on the refined noise probability stated in Lemma 4. Fix \( L \subseteq [n] \) and let \( \lambda = |L|/n \). Furthermore, let \( 0 \leq \rho < 1 \), and let \((A, B)\) be a UDCP such that \( A \) is \( \epsilon \)-dense with respect to \( L \) and \( |B_L| = 2^{\pi n} \) for some \( 0 \leq \pi \leq \lambda \).

Due to the chain rule

\[
Pr_{a \sim_{\rho} b}[a \in A, b \in B] = Pr_{a \sim_{\rho} b}[a \in A, b \in B, a_\pi L \in A_L, b_\pi L \in B_L] \times Pr_{a \sim_{\rho} b}[a \in A, b \in B, a_\pi L \in A_L, b_\pi L \in B_L]. \tag{5}
\]

We proceed by giving lower bounds for the two factors in the product (5). Let \( R = [n] \setminus L \). For the first factor, note that if \( b_R \in B_R \), there is at least one \( b_R \) such that \( b_R, b_R \in B \) by the definition of \( B_L \), and such a \( b_R \) is picked with probability \( 2^{-|R|} \) since it is uniformly distributed over \( 2^{|R|} \). Similarly, if \( a_L \in A_L \), there are at least \( 2|L| - \epsilon n / 2 \) sets \( a_R \subseteq R \) such that \( a_\pi R \in A' \) by the definition of \( A' \), and so such an \( a_R \) is picked with probability at least \( 2^{-\epsilon n / 2} \). In summary,

\[
Pr_{a \sim_{\rho} b}[a \in A, b \in B, a_L \in A_L, b_L \in B_L] \geq 2^{-|R| - \epsilon n / 2}
\]

\[
= g(\lambda - \epsilon - o(1)).
\]

For the second term, apply Theorem 2 with \( U = L \) and

\[
F = A_L, \quad f = \left( \left| \{L\} - \epsilon n - 1 \right| \right) / |L| = 1 - \epsilon - o(1),
\]

\[
G = B_L, \quad g = \frac{\pi}{\lambda},
\]

which gives that

\[
\log_2 \Pr_{a \sim_{\rho} b}[a_L \in A_L, b_L \in B_L] \geq -|L| \left( \frac{1 - \frac{\pi}{\lambda}}{\lambda} + 2 \rho \sqrt{(1 - \frac{\pi}{\lambda})(\frac{\pi}{\lambda} + o(1))} \right)
\]

\[
= n \left( \frac{\pi - \lambda - \epsilon - 2\rho \sqrt{\epsilon \lambda - \epsilon^2}}{1 - \rho^2} - o(1) \right).
\]

The statement now follows by multiplying the two lower bounds into a lower bound for the product (5).

VIII. CONCLUSION

We presented a new upper bound for UDCPs, considerably strengthening previous bounds. We obtained the bound by combining an isoperimetric inequality, which was not used before in the UDCP literature, with an extension of van Tilborg’s bound that works well if the set families are clustered.

Two outstanding open questions that are of main interest remain. In our setting \( (\alpha \geq 1 - \epsilon) \), there is still a big gap between the best construction (achieving \( \beta \geq 1/4 \)) and our new upper bound of \( 0.4228 + \sqrt{\epsilon} \). Narrowing this gap from either direction would be very interesting. In the general case, a major unresolved problem is whether the classic upper bound of \( |A| \cdot |B| \leq 2^{1.5n} \) is tight.
REFERENCES


Per Austrin is an Associate Professor in Computer Science at KTH Royal Institute of Technology, Stockholm, Sweden. He received the Ph.D. degree from KTH in 2008 with his Ph.D. thesis “Conditional Inapproximability and Limited Independence”, following which he did post-docs at Institut Mittag-Leffler, New York University, University of Toronto, and Aalto University.

Petteri Kaski is an Associate Professor in Computer Science at Aalto University, Helsinki, Finland. He received the M.Sc.(Tech.) and D.Sc.(Tech.) degrees from Helsinki University of Technology in 2001 and 2005, respectively. His research interests range from algorithm design and analysis to classification problems in discrete mathematics.

Mikko Koivisto is an Associate Professor of Computer Science at the University of Helsinki, Finland, where he also received the M.Sc. and Ph.D. degrees in 2001 and 2004, respectively. His main research interests include exact exponential algorithms, counting problems, and learning and inference in graphical models.

Jesper Nederlof is an Assistant Professor in the Discrete Mathematics section at Eindhoven University of Technology. He received the M.Sc. degree in applied computing science from Utrecht University, the Netherlands, in 2008, and successfully defended his Ph.D. thesis titled “Space and Time Efficient Structural Improvements of Dynamic Programming Algorithms” in December 2011 at the University of Bergen, Norway.