Miskolc Mathematical Notes
Vol. 12 (2011), No 1, pp. 11-23

HU e-ISSN 1787-2413
DOI: 10.18514/MMN.2011.325

Optimal combinatorial batch codes derived from dual systems

Csilla Bujtás and Zsolt Tuza

Miskolc Mathematical Notes
Vol. 12 (2011), No. 1, pp. 11–23

HU e-ISSN 1787-2413

OPTIMAL COMBINATORIAL BATCH CODES DERIVED FROM DUAL SYSTEMS

CSILLA BUJTÁS AND ZSOLT TUZA

Received 16 February, 2011

Abstract. Combinatorial batch codes with parameters n, k, m, and $t = 1$ may be viewed as set systems $\mathcal{F}$ consisting of n subsets over an m-element set (repetitions allowed), satisfying the following restricted version of Hall's Condition: for every $1 \le i \le k$, the union of any i members of $\mathcal{F}$ has cardinality at least i. An optimization problem is to determine $N(n,k,m)$, the minimum total size $\sum_{F\in\mathcal{F}} |F|$ in such systems. Beside its theoretical interest, the problem has strong practical motivation, too, concerning distributed storage and retrieval of data in a database.

Already the case $n = m+2$ turns out to be somewhat complicated. Here we give explicit optimal constructions and prove the following formulae: in the range $k \le m \le k+\sqrt{k}$

$$N(m+2,k,m) = 2m + \left\lfloor \frac{k}{m-k+1} \right\rfloor,$$

and if $m > k+\sqrt{k}$ then

$$N(m+2,k,m) = N(m+1,k,m-1)+1 = m+k-2+\lceil 2\sqrt{k+1}\,\rceil$$

for all $m \ge k \ge 1$. Our method is purely combinatorial, whereas the first proof by Brualdi *et al.* [Adv. Math. Commun., 4 (2010), 419–431 & 597] used the theory of transversal matroids. We also present an optimality-preserving transformation, by which a large family of non-isomorphic optimal constructions can be derived if one is already available. Moreover, we prove a new general upper bound on $N(n,k,m)$.

2000 *Mathematics Subject Classification:* 05A05; 05C65; 68R05

Keywords: combinatorial batch code, dual system, Hall's Condition, system of distinct representatives

1. INTRODUCTION

In this paper we study a discrete optimization problem concerning data storage and retrieval in distributed databases. In the considered model, a certain amount of data is to be stored on a given number of servers. For a fixed number k, the task is to distribute data storage on the servers in such a way that any k data can be retrieved by communicating simultaneously with k suitably chosen different servers, one server

Research supported in part by the Hungarian Scientific Research Fund, OTKA grant 81493.

per data. The overall goal is to minimize storage space, or equivalently the total amount of data replication.

This problem, whose abstract formulation is given in the next subsection, was studied in detail first by Paterson *et al.* [6], as a restricted version of the batch code problem introduced by Ishai *et al.* [5]. In the latter, which is more application-oriented, a limited amount t of communication is allowed with any one server, and even an encoding/decoding phase is allowed during data distribution/retrieval. The terminology *combinatorial batch code* refers to the fact that the encoding aspect is excluded. Our results for the general case $t \geq 1$ will be announced in [4]. In the present paper we will consider only the case $t = 1$.

1.1. *Several alternative formulations*

The problem has several equivalent formulations, which we list in this subsection. Most of them (except the last one) are taken from [6], as purely combinatorial versions of the more complex model of [5]. Throughout the paper we shall disregard the first form, since we find the dual systems and related concepts more handy. The last condition—the complementary version of the restricted Hall Condition—is a new one introduced here and it will turn out to be useful at several points of the proofs.

Combinatorial batch code. For positive integers n,k,m, a *combinatorial batch code* CBC(n,k,m) is a set system[1] $\mathcal{S}$ of $|\mathcal{S}| = m$ sets over an underlying set D of cardinality $|D| = n$, with the property that for every $1 \leq \ell \leq k$ and every ℓ-tuple $\{d_1,\ldots,d_\ell\} \subseteq D$, there exists a subsystem $\{S_1,\ldots,S_\ell\} \subset \mathcal{S}$ of ℓ distinct members of $\mathcal{S}$, such that $d_i \in S_i$ holds for every $1 \leq i \leq \ell$. The goal is to determine

$$N(n,k,m) := \min_{\mathcal{S}} \sum_{S\in\mathcal{S}} |S|$$

where the minimum is taken over all combinatorial batch codes $\mathcal{S}$ with parameters n,k,m.

In this setting the elements of D represent the items of the database, and the members of $\mathcal{S}$ correspond to the contents of the servers; i.e., for any $d \in D$ and $S \in \mathcal{S}$, server S stores item d if and only if $d \in S$. Hence, $N(n,k,m)$ determines the smallest possible storage space required.

In order to avoid trivialities, we shall assume that the following inequalities are valid.

- $k \geq 2$. (For $k = 1$ the only condition is that each $d \in D$ should occur in some $S \in \mathcal{S}$, hence $\mathcal{S}$ can be composed e.g. of D and $m-1$ copies of the empty set, and consequently $N(n,1,m) = n$ for all n,m.)

[1] Throughout, 'set system' is meant as a 'multisystem'. Among the members of the system repetitions are allowed; that is, distinct members of the system may correspond to the same set.

- $n > m$. (For $n \leq m$, we may take the elements of D as 1-element members (singletons) of $\mathcal{S}$, together with $m-n$ copies of $\emptyset$, hence we again have $N(n,k,m) = n$ for all $n \leq m$ and all $k \geq 1$.)
- $m \geq k$. (Assuming $n > m$ by the previous observation, the case $m < k$ would lead to the contradiction for $\ell = m+1 \leq k$ that no ℓ-tuple can satisfy the requirement; that is, no such CBC(n,k,m) exists.)

In formulating the results, however, we shall sometimes allow $n = m$ and/or $k = 1$ to hold when it fits the generality of presentation, despite that these are obvious cases.

Dual system. A combinatorial batch code as a set system can be represented with its *dual system*. We say that a set system $\mathcal{F}$ is a CBC$^*(k)$-system if, for every $1 \leq \ell \leq k$ and every ℓ-element subsystem $\mathcal{F}' = \{F_1, \ldots, F_\ell\} \subseteq \mathcal{F}$ there exist ℓ mutually different elements $x_1, \ldots, x_\ell$ of the underlying set, such that $x_i \in F_i$ holds for every $1 \leq i \leq \ell$. A set system $\mathcal{F}$ over the underlying set X is called a CBC$^*(n,k,m)$ if $|\mathcal{F}| = n$, $|X| = m$, and $\mathcal{F}$ is a CBC$^*(k)$-system.

Asterisk in the notation refers to the usage of dual systems in the sense of hypergraph theory. If $\mathcal{H} = (V, \mathcal{E})$ is a hypergraph with vertex set $V = \{v_1, \ldots, v_n\}$ and edge set $\mathcal{E} = \{E_1, \ldots, E_m\}$ (i.e., $\mathcal{E}$ is a family of sets over the underlying set V, multiple sets allowed), then its *dual* is a hypergraph $\mathcal{H}^* = (V^*, \mathcal{E}^*)$ where $V^* = \{v_1^*, \ldots, v_m^*\}$, $\mathcal{E}^* = \{E_1^*, \ldots, E_n^*\}$, and $v_i^* \in E_j^*$ if and only if $v_j \in E_i$ $(1 \leq i \leq m,\ 1 \leq j \leq n)$.

The condition on ℓ-tuples of sets in a dual system clearly corresponds to the condition on ℓ-element subsets of D given in definition of CBC(n,k,m). Hence, $\mathcal{S}$ is a CBC(n,k,m) if and only if its dual system $\mathcal{S}^*$ is a CBC$^*(n,k,m)$. By this equivalence we have

$$N(n,k,m) = \min_{\mathcal{F}} \sum_{F \in \mathcal{F}} |F|$$

where the minimum is taken over all CBC$^*(n,k,m)$ systems.

We say that $\mathcal{F}$ is *optimal* if it is a CBC$^*(n,k,m)$ attaining equality for $N(n,k,m)$.

k-Restricted Hall Condition (k-HC). The definition of CBC$^*(k)$-system can equivalently be expressed by a Hall-type condition, in the following way. We say that a set system $\mathcal{F}$ satisfies the *k-restricted Hall Condition*, abbreviated as k-HC, if for every $\ell \leq k$ and every family $\mathcal{F}'$ of ℓ members of $\mathcal{F}$ the inequality $|\bigcup_{F \in \mathcal{F}'} F| \geq \ell$ holds. (Certainly, $|\mathcal{F}|$-HC means precisely the classical Hall Condition for the existence of a System of Distinct Representatives.)

Restricting our attention to the range $n \geq m \geq k \geq 1$, let us introduce the notation

$$s := m-k, \qquad q := n-m.$$

In this way the goal is to determine $N(k+s+q, k, k+s)$.

We will consider the definition of CBC$^*(k)$-system in the following three equivalent formulations. From now on, each of them will be referred as k-HC.

(k1) The set system $\mathcal{F}$ satisfies k-HC.

(k2) For every $0 \le j \le k-1$, every j-element subset of the underlying set contains at most j members from the system $\mathcal{F}$.
(k3) For every $t > s$, every t-element subset of the underlying set meets at least $t+q$ members of $\mathcal{F}$.

For the equivalence of the last formulation, we note that writing $j := k+s-t$, the condition is equivalent to the requirement (k2). Property (k3) can be considered as the complementary version of (k1) and (k2), viewing the problem 'from the other side.'

Remark 1. For $k \ge 2$, in any CBC$^*(k)$, each singleton set $\{x\}$ may occur with multiplicity at most 1. This is the particular case $j = 1$ of the requirement (k2).

1.2. *Results and brief history*

There are not too many ranges of (n,k,m) for which an exact formula for $N(n,k,m)$ is known. In [6] the problem was solved for arbitrary k,m and $n \ge (k-1)\binom{m}{k-1}$. For smaller $n \ge \binom{m}{k-2}$ the optimum was determined by the present authors in [3] and independently by Bhattacharya *et al.* in [1]. These two results together yield $N(n,3,m)$ for all n and m, as computed in [3]. The case $N(n,2,m)$ follows already from the quoted theorem of [6]. But, in contrast, $k = 4$ which was settled in [3] is quite complex and the solution is composed of four different ranges of (n,m).

In the paper [6], also $N(n,k,k) = k(n-k+1)$ and $N(m+1,k,m) = m+k$ are proven. But already the determination of $N(m+2,k,m)$ turns out to be quite hard and requires a deep insight. This was done first by Brualdi *et al.* in [2], using the heavy machinery of transversal matroids. Our major point here is to explore the structure of batch codes with $n = m+2$, and to develop purely combinatorial methods to obtain a transparent derivation of the formula for $N(m+2,k,m)$, stated as Theorem 2. The main structural observations are made in Section 3, and the computation proving optimality is given in Section 4.

We also design a class of optimal batch codes with transparent structure, and describe a transformation which generates further families of codes while preserving optimality. These results are presented in Section 2 in a general setting, where a general upper bound on $N(n,k,m)$ is also proved. Some further remarks and open problems are mentioned in the concluding section.

1.3. *Hypergraph preliminaries*

Given a set system $\mathcal{F}$, we denote

$$S(\mathcal{F}) := \sum_{F \in \mathcal{F}} |F|.$$

If $\mathcal{F}$ is over the set X, for $x \in X$ we denote by $d(x)$ the number of sets $F \in \mathcal{F}$ containing x and call it the *degree* of x. Counting in two ways the pairs (x,F)

satisfying the incidence relation $x \in F$, we obtain $\sum_{F\in\mathcal{F}} |F| = \sum_{x\in X} d(x)$ and hence

$$S(\mathcal{F}) = \sum_{x\in X} d(x)$$

is valid. Moreover, if $d(x) \geq 2$ for all $x \in X$, we partition X into two sets:

$$X_2 = \{x \in X \mid d(x) = 2\}, \qquad X_3 = \{x \in X \mid d(x) \geq 3\}.$$

Then from the degree-sum expression above we obtain

$$S(\mathcal{F}) \geq 2|X_2| + 3|X_3| = 2|X| + |X_3| = 3|X| - |X_2|. \tag{1.1}$$

The following fact can be applied in arguments proving that a construction is a CBC$^*(k)$.

Lemma 1. *If a set system $\mathcal{F}$ violates the k-restricted Hall Condition but $\mathcal{F}$ is minimal in the sense that each of its proper subsystems satisfies k-HC, then all elements of $\bigcup_{F\in\mathcal{F}} F$ are contained in at least two members of $\mathcal{F}$.*

Proof. If some $F \in \mathcal{F}$ is the unique member of $\mathcal{F}$ containing an element x, then x can be chosen to represent F in every subfamily involving F. This has no effect on k-HC in $\mathcal{F} \setminus \{F\}$. □

Corollary 1. *Let $n \geq m \geq k \geq 2$, and suppose that $\mathcal{F}$ is an optimal CBC$^*(n,k,m)$. Then*

(*i*) *no elements have degree 0;*
(*ii*) *if $d(x) = 1$, then $\{x\} \in \mathcal{F}$.*

Proof. (*i*) This part was proved in [6] and [2], too; but with reference to Lemma 1 we can simply say that if x had degree 0, then from any $n \geq m$ subsets of $X \setminus \{x\}$, a non-singleton F could be replaced with the singleton $\{x\}$, contradicting the optimality of $\mathcal{F}$. (Such an $F \in \mathcal{F}$ must occur whenever x has degree 0 and $k \geq 2$, by Remark 1.)

(*ii*) If F is the unique member of $\mathcal{F}$ containing x, and $|F| > 1$ holds, then by Lemma 1, $\mathcal{F}$ remains a CBC$^*(k)$-system when F is replaced with $\{x\}$. But applying this replacement we would decrease $S(\mathcal{F})$, contradicting the assumption concerning optimality. □

Further structural simplification will be presented for elements of degree 2 in the next section.

2. General upper bounds and transformations

In this section we present constructions, general upper bounds on $N(n,k,m)$ and some transformations applicable when elements of degree 2 occur, aiming at structural simplification under which the optimality of systems remains unchanged.

2.1. *Constructions and upper bounds*

We present two constructions, one recursive and one explicit, providing upper bounds on $N(n,k,m)$, which will turn out to be optimal in the entire range of m if $n = m+2$. Actually, the recursive step is very simple.

Construction A.. For any three integers $k \geq 1$, $s \geq 1$ and $q \geq 0$, take a system $\mathcal{F}'$ which is a CBC*$(k+s+q-1,k,k+s-1)$, and add a further vertex x^* and the singleton edge $\{x^*\}$ to it. Clearly, the system remains a CBC*(k). Hence, a CBC*$(k+s+q,k,k+s)$ is obtained with $S(\mathcal{F}) = S(\mathcal{F}')+1$.

Corollary 2. *For every $k \geq 1$, $s \geq 1$ and $q \geq 0$,*

$$N(k+s+q,k,k+s) \leq N(k+s+q-1,k,k+s-1)+1.$$

Construction B.. For every fixed $k \geq 1$, $s \geq 0$ and $q \geq 2$, let us write s and k in the following form, where a, b, p and r are integers:

$$s = a(q-1)+b \quad \text{with} \quad 0 \leq b \leq q-2, \quad \text{hence} \quad a = \left\lfloor \frac{s}{q-1} \right\rfloor;$$

$$k = p(a+1)+r \quad \text{with} \quad 0 \leq r \leq a, \quad \text{hence} \quad p = \left\lfloor \frac{k}{a+1} \right\rfloor = \left\lfloor \frac{k}{\left\lfloor \frac{s}{q-1} \right\rfloor + 1} \right\rfloor.$$

Let $X = V_0 \cup V \cup U \cup R \cup Y$, where

$$V = V_1 \cup \cdots \cup V_a \quad \text{with} \quad |V_1| = \ldots = |V_a| = p,$$

$$U = U_1 \cup \cdots \cup U_{q-1} \quad \text{with} \quad |U_1| = \ldots = |U_{q-1}| = a,$$

$$|V_0| = p, \quad |R| = r, \quad |Y| = b.$$

The sets $V_0, V_1, \ldots V_a, U_1, \ldots U_{q-1}, R$ and Y are pairwise disjoint. This means

$$|X| = p + pa + (q-1)a + r + b = k+s.$$

Then $\mathcal{F}$ is composed of the $k+s-p$ singletons of $V \cup U \cup R \cup Y$, any p mutually disjoint $(a+1)$-tuples each having precisely one element in each V_i $(i = 0,1,\ldots,a)$, and the following q sets: $F_j = U_j \cup V_0$ $(1 \leq j \leq q-1)$ and $F_q = R \cup V_0$. Hence, $\mathcal{F}$ has precisely $k+s+q$ members and the sum of their cardinalities is equal to

$$k+s-p+p(a+1)+(q-1)(a+p)+r+p = 2k+2s+(q-1)p-b =$$

$$= 2k+2s+(q-1)\left\lfloor \frac{k}{\left\lfloor \frac{s}{q-1} \right\rfloor + 1} \right\rfloor - b.$$

Depending on the fixed parameters k, s and q, some of the sets V_0, V, U, R and Y may be empty, but no member of $\mathcal{F}$ is empty. In particular, if $s < q-1$ then $a = 0$, and hence $X = V_0 \cup Y$.

Lemma 2. *For every $k \geq 1$, $s \geq 0$ and $q \geq 2$, Construction B yields a CBC*$^*(k)$*-system.*

Proof. Consider the system $\mathcal{F}$ constructed as described above with parameters k, s and q. To prove that $\mathcal{F}$ is a CBC$^*(k)$, condition k-HC will be checked in the form (k3); that is, we will prove that any $z \geq s+1$ elements of X intersect at least $z+q$ members of $\mathcal{F}$.

Consider a subset $Z \subseteq X$ with $|Z| = z \geq s+1$. We have two cases:

- If $Z \cap V_0 \neq \varnothing$ then the elements in $Z \cap V_0$ together intersect precisely $|Z \cap V_0| + q$ members of the system $\mathcal{F}$. Moreover, each element of $Z \setminus V_0$ is a singleton. Thus, Z meets at least $z+q$ members.
- If $Z \cap V_0 = \varnothing$, each element of Z is a singleton. Moreover, in this case $a \geq 1$ has to hold. (Indeed, otherwise $Z \subseteq Y$ would imply $z \leq b \leq s$.) Let us denote the cardinalities of $Z \cap V$, $Z \cap U$, $Z \cap R$ and $Z \cap Y$ by z_1, z_2, z_3 and z_4, respectively. Since $z_4 \leq b$,

$$z_1 + z_2 + z_3 \geq z - b \geq s + 1 - b = a(q-1) + 1$$

holds. By construction, each non-singleton member of $\mathcal{F}$ meets at most one of V, U and R, and it does not meet Y, moreover it contains at most a elements from V or U and at most $r \leq a$ elements from R. Hence, the number of sets $F \in \mathcal{F}$ intersected by Z is at least

$$z + \left\lceil \frac{z_1}{a} \right\rceil + \left\lceil \frac{z_2}{a} \right\rceil + \left\lceil \frac{z_3}{a} \right\rceil \geq z + \left\lceil \frac{z_1 + z_2 + z_3}{a} \right\rceil \geq z + \left\lceil \frac{a(q-1)+1}{a} \right\rceil = z + q,$$

as claimed.

Therefore, k-HC is satisfied and $\mathcal{F}$ is a CBC$^*(k)$, indeed. □

Corollary 3. *Let $k \geq 1$, $s \geq 0$, $q \geq 2$, and let b denote the residue of s modulo $q-1$. Then*

$$N(k+s+q, k, k+s) \leq 2k + 2s + (q-1)\left\lfloor \frac{k}{\left\lfloor \frac{s}{q-1} \right\rfloor + 1} \right\rfloor - b.$$

If $q = 2$ then $b = 0$, and we have the following consequence:

Corollary 4. *For every $k \geq 1$ and $s \geq 0$*

$$N(k+s+2, k, k+s) \leq 2k + 2s + \left\lfloor \frac{k}{s+1} \right\rfloor.$$

2.2. *Optimality-preserving transformations*

By Corollary 1, the situation with elements of degree smaller than 2 is clear. Here we handle the elements of degree 2. In this subsection $n > m$ is assumed.

Transformation C.. Suppose that $\mathcal{F}$ is a CBC$^*(k)$, and some $x \in X$ has degree 2. Let $F_1, F_2 \in \mathcal{F}$ be the two members of $\mathcal{F}$ containing x. Suppose further that $|F_2| \geq 2$ and $y \in F_2 \setminus \{x\}$. Define $\mathcal{F}_{x,y}$ as the system obtained from $\mathcal{F}$ by replacing F_1 and F_2 with $F_1' := F_1 \cup \{y\}$ and $F_2' := F_2 \setminus \{y\}$, respectively.

Let us put some comments about the flexibility of this transformation. First, there is no condition about the degree of y. Second, it is also allowed that $y \in F_1$, in which case $S(\mathcal{F})$ decreases by 1; and otherwise $S(\mathcal{F})$ remains unchanged. Third, F_1 is allowed to be a singleton, although $|F_1'| \geq 2$ always holds because $x \neq y$.

Lemma 3. *The system $\mathcal{F}_{x,y}$ is a CBC$^*(k)$ whenever so is $\mathcal{F}$.*

Proof. Assume for a contradiction that the assertion is not valid, and let $\mathcal{F}' \subset \mathcal{F}_{x,y}$ be a minimal subsystem violating k-HC. By Lemma 1, x cannot have degree 1 in $\mathcal{F}'$; hence, either both or none of F_1, F_2 belong to $\mathcal{F}'$. But if $F_1, F_2 \notin \mathcal{F}'$ then $\mathcal{F}' \subset \mathcal{F}$ and must satisfy k-HC, whereas for $F_1, F_2 \in \mathcal{F}'$ the union $\bigcup_{F \in \mathcal{F}'} F$ is the same in $\mathcal{F}'$ as in $\mathcal{F}$. Thus, $\mathcal{F}'$ cannot violate k-HC. □

Corollary 5. *Every CBC$^*(n,k,m)$ system $\mathcal{F}$ can be transformed to some CBC$^*(n,k,m)$ system $\mathcal{F}'$ such that $S(\mathcal{F}') \leq S(\mathcal{F})$, the degree of each element is at most as large in $\mathcal{F}'$ as that in $\mathcal{F}$, and every element x of degree 2 in $\mathcal{F}'$ is a singleton member $\{x\} \in \mathcal{F}'$.*

Proof. A singleton can be obtained from F_2 after $|F_2| - 1$ applications of Transformation C. This singleton will always remain in the system if we do not choose it as F_1. Hence, the number of singleton sets can be increased as long as there exists an element of degree 2 which is not a singleton member of the system.

Since the transformation does not increase the degrees, it does not increase the value of $S(\mathcal{F})$ either. □

Using the conditions and notation given for transformation C, we can also conclude:

Corollary 6. *If $\mathcal{F}$ is an optimal CBC$^*(k)$, then*

(i) $\mathcal{F}_{x,y}$ is optimal;
(ii) $\mathcal{F}$ can also be obtained from $\mathcal{F}_{x,y}$ by Transformation C;
(iii) if $F', F'' \in \mathcal{F}$ share an element z of degree 2, then $F' \cap F'' = \{z\}$.

Transformation C can also be applied to derive many non-isomorphic optimal families starting from a known one.

3. Lower bounds for $n = m + 2$

From now on, we concentrate on the case of $q = 2$. Here we prove two lower bounds on $N(m+2,k,m)$, which will match the constructions of the previous section for $q = 2$, and also correspond to the two ranges of m (with respect to k) in the main theorem. We again begin with the simpler one, for vertices of degree 1.

Lemma 4. *Let $k \geq 2$ and $s \geq 1$. If $\mathcal{F}$ is a CBC$^*(k+s+2,k,k+s)$ with minimum vertex degree 1, then*

$$S(\mathcal{F}) \geq N(k+s+1,k,k+s-1)+1.$$

Proof. Consider a vertex x of degree 1 from $\mathcal{F}$. Let the only edge F incident with x replaced by the singleton $F^* = \{x\}$. After this modification the system $\mathcal{F}^*$ is obtained. Consider any at most $(k-1)$-element subset Y of the underlying set X. If $x \notin Y$, then Y contains exactly as many members from $\mathcal{F}^*$ as from $\mathcal{F}$, hence at most $|Y|$. In the other case, when $x \in Y$, the vertex set $Y \setminus \{x\}$ contains at most $|Y|-1$ members from $\mathcal{F}$, hence Y contains at most $|Y|$ members from $\mathcal{F}^*$. This proves that $\mathcal{F}^*$ is a CBC$^*(k+s+2,k,k+s)$. Moreover, omitting the vertex x and the singleton $\{x\}$ from $\mathcal{F}^*$, the obtained system $\mathcal{F}^-$ still satisfies the CBC$^*(k)$-property. This implies that $N(k+s+1,k,k+s-1) \leq S(\mathcal{F}^-) = S(\mathcal{F}^*)-1$. By our construction, also $S(\mathcal{F}^*) \leq S(\mathcal{F})$ holds. Hence, the lemma follows. □

Remark 2. The above lemma and proof are valid for any $q \geq 0$. That is, if $\mathcal{F}$ is a CBC$^*(k+s+q,k,k+s)$ with minimum vertex degree 1 where $k \geq 2$ and $s \geq 1$, then

$$S(\mathcal{F}) \geq N(k+s+q-1,k,k+s-1)+1.$$

The crucial tool for the proof of the main theorem is the following lower bound. We will use the notation X_2 and X_3 as they were introduced in Section 1.3. Let us also recall from there that $S(\mathcal{F}) = \sum_{x \in X} d(x) \geq 2k+2s+|X_3|$ holds by (1.1), whenever no vertices of degree less than 2 occur.

Theorem 1. *If $\mathcal{F}$ is a CBC$^*(k+s+2,k,k+s)$ with minimum vertex degree at least 2, then*

$$S(\mathcal{F}) \geq 2k+2s+\left\lfloor \frac{k}{s+1} \right\rfloor.$$

Proof. By Corollary 5, we can assume that every $x \in X_2$ is a singleton member $\{x\} \in \mathcal{F}$. On the other hand, referring the condition k-HC in the form (k3), any $\ell \geq s+1$ elements of X_2 have to meet at least $\ell+2$ members of $\mathcal{F}$. Therefore, each $F \in \mathcal{F}$ contains at most s elements from X_2. This yields the following lower bound on the number $|\mathcal{F}| = n = k+s+2$ of sets:

$$n \geq |X_2| + \frac{|X_2|}{s} = |X_2| \frac{s+1}{s}.$$

This implies an upper bound on $|X_2|$, and hence a lower bound on $|X_3|$.

$$|X_3| = k+s-|X_2| \geq k+s-n\frac{s}{s+1} = \frac{k+s+2}{s+1} - 2 = \frac{k-s}{s+1}.$$

Therefore, by (1.1) we have that

$$S(\mathcal{F}) \geq 2k+2s+\left\lceil \frac{k-s}{s+1} \right\rceil = 2k+2s+\left\lfloor \frac{k}{s+1} \right\rfloor,$$

as it was claimed. □

Remark 3. The analogous general bound

$$|X_2| \le \frac{n\left\lfloor \frac{s}{q-1} \right\rfloor + b}{\left\lfloor \frac{s}{q-1} \right\rfloor + 1},$$

where b denotes the residue of s modulo $q-1$, is valid for any $q \ge 2$ and any $s \ge 0$.

4. Exact values for $n = m+2$

Theorem 2. *For every two integers $k \ge 1$ and $s \ge 0$,*

(i) $N(k+s+2,k,k+s) = 2k+2s+\left\lfloor \frac{k}{s+1} \right\rfloor$, *if $k \ge s^2$; and*

(ii) $N(k+s+2,k,k+s) = 2k+s-2+\lceil 2\sqrt{k+1} \rceil$, *if $k < s^2$.*

Proof. For every $n \ge m$ and $k \ge 1$ there exists an optimal CBC$^*(k)$ without isolated vertices. Then, due to Corollaries 2 and 4, Lemma 4, and Theorem 1, for every $k \ge 1$ and $s \ge 1$

$$N(k+s+2,k,k+s) = \min\left(N(k+s+1,k,k+s-1)+1, 2k+2s+\left\lfloor \frac{k}{s+1} \right\rfloor \right)$$

holds. Let us introduce the notation

$$a(k,s) := N(k+s+2,k,k+s), \qquad b(k,s) := 2k+2s+\left\lfloor \frac{k}{s+1} \right\rfloor.$$

Hence, for any $k \ge 1$ and $s \ge 1$ we have

$$a(k,s) = \min(a(k,s-1)+1, b(k,s)). \tag{4.1}$$

Before proving the statements of the theorem, let us observe that

$$b(k,s) \le b(k,s-1)+1 \quad \text{if} \quad k \ge s^2 \ge 1. \tag{4.2}$$

Indeed, $\frac{k}{s+1} \le \frac{k}{s} - 1$ holds whenever $k \ge s^2+s$. Moreover, if $s^2 \le k < s^2+s$, then $\left\lfloor \frac{k}{s+1} \right\rfloor = s-1$ and $\left\lfloor \frac{k}{s} \right\rfloor = s$ are valid, yielding (4.2) with equality in this range of k. On the other hand, a similar computation shows that

$$b(k,s) \ge b(k,s-1)+1 \quad \text{if} \quad k < s^2. \tag{4.3}$$

Part (i) states that $a(k,s) = b(k,s)$ if $k \ge s^2$. For $s = 0$, our statement is the particular case $n = k+2$ of Theorem 4 in [6]. Indeed, the result $N(n,k,k) = k(n-k+1)$ implies $a(k,0) = 3k = b(k,0)$. We proceed by induction on s. Assume that $s \ge 1$, $k \ge s^2$, and that $a(k,s') = b(k,s')$ is true for all $0 \le s' < s$ and all $k \ge s^2$. Since $k > (s-1)^2$, the induction hypothesis and (4.2) imply $a(k,s-1)+1 = b(k,s-1)+1 \ge b(k,s)$. Hence, $a(k,s) = b(k,s)$ follows by (4.1). This proves (i) for every $k \ge 1$ and $s \ge 0$.

Concerning Part (*ii*), we first prove that in case of $k < s^2$ the recursion $a(k,s) = a(k,s-1)+1$ applies. Due to (4.1), this precisely means to verify $b(k,s) \geq a(k,s-1)+1$. This inequality is valid indeed, because $a(k,s-1) = \min(a(k,s-2)+1, b(k,s-1))$ and hence, applying (4.3), we obtain

$$a(k,s-1)+1 \leq b(k,s-1)+1 \leq b(k,s)$$

as required.

The rest of the proof for $s > \sqrt{k}$ is a purely technical matter of computation. Namely, from the '$a(k,s-1)+1$' branch of the recursion we obtain

$$\begin{aligned} N(k+s+2,k,k+s) &= b(k,\lfloor\sqrt{k}\,\rfloor)+(s-\lfloor\sqrt{k}\,\rfloor) \\ &= 2k+2\lfloor\sqrt{k}\,\rfloor+\left\lfloor\frac{k}{\lfloor\sqrt{k}\,\rfloor+1}\right\rfloor+(s-\lfloor\sqrt{k}\,\rfloor) \\ &= 2k+s+\lfloor\sqrt{k}\,\rfloor+\left\lfloor\frac{k}{\lfloor\sqrt{k}\,\rfloor+1}\right\rfloor. \end{aligned}$$

Observing $\lfloor\sqrt{k}\,\rfloor = \lceil\sqrt{k+1}\,\rceil - 1$, the above formula yields

$$\begin{aligned} N(k+s+2,k,k+s) &= 2k+s-1+\left\lceil\sqrt{k+1}\,\right\rceil+\left\lfloor\frac{k}{\lceil\sqrt{k+1}\,\rceil}\right\rfloor \\ &= 2k+s-2+\left\lceil 2\sqrt{k+1}\,\right\rceil \end{aligned}$$

where the last step follows from the identity

$$\lceil\sqrt{n}\,\rceil+\left\lfloor\frac{n-1}{\lceil\sqrt{n}\,\rceil}\right\rfloor = \lceil 2\sqrt{n}\,\rceil - 1.$$

Indeed, for any integer $t \geq 1$, if $t^2+1 \leq n \leq t^2+t$ then $\lceil\sqrt{n}\,\rceil = t+1$, $\left\lfloor\frac{n-1}{\lceil\sqrt{n}\,\rceil}\right\rfloor = t-1$, and $\lceil 2\sqrt{n}\,\rceil = 2t+1$; and if $t^2+t+1 \leq n \leq t^2+2t+1$ then $\lceil\sqrt{n}\,\rceil = t+1$, $\left\lfloor\frac{n-1}{\lceil\sqrt{n}\,\rceil}\right\rfloor = t$, and $\lceil 2\sqrt{n}\,\rceil = 2t+2$. □

5. Concluding remarks

We have developed some methods of combinatorial nature to design batch codes and to prove their optimality in some cases. It is very likely that the class of systems described in Constructions A and B is optimal for many ranges of (n,k,m). We have proved this for all $n = m+2$.

A novel contribution to methodology here is that the matroid approach of [2] is now completely replaced with a combinatorial one. By this switch we expect that some of the techniques may turn out to be applicable for larger values of $n-m$, too.

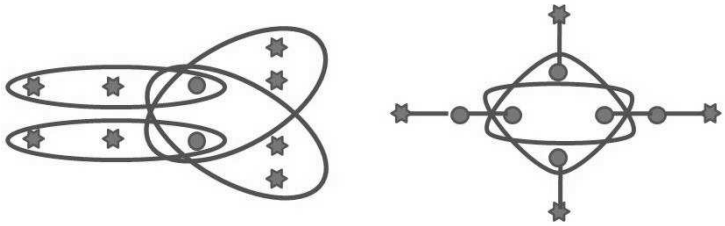

FIGURE 1. Two optimal constructions verifying $N(12,8,10) = 22$. Star vertices indicate singleton members of the system.

Currently the smallest unexplored subclasses of the general problem are $n = m+3$ and $k = 5$.

Many optimal systems. Transformation C described in Section 2.2 generates a large class of optimal systems. To illustrate this, we take the case of $(n,k,m) = (12,8,10)$. Figure 1 exhibits two set systems which are in some sense the two extremes. The first one is given in Construction B, where all but two of the elements are singleton members of the system. The prize of including many singletons is of course that a considerable number of larger sets must also occur. In the other system, there are only few singletons, but practically all sets are fairly small. Adding 10 new elements as singletons, we obtain the optimal $(22,8,20)$ system exhibited in [2] in the form of a 20×22 matrix of 0s and 1s.

One can easily verify that the two systems exhibited in Figure 1 can be obtained from each other (in either direction) via a sequence of step-by-step applications of Transformation C.

The case $m = k$. It has been proved in [6, Theorem 4] that $N(n,k,k) = k(n-k+1)$ holds for all $n \geq k \geq 1$. An optimal system is obtained easily by taking the k elements of X as singletons, and $n-k$ copies of X for the remaining members of $\mathcal{F}$.

Here we describe an alternative optimal construction that replaces the k singletons[2] and two copies of X if $n-2 \geq k \geq 3$. Let $X = \{x_1,\ldots,x_k\}$, and consider any non-trivial partition $F' \cup F'' = X$. Beside F' and F'' we take the k pairs $\{x_1,x_2\}$, $\{x_2,x_3\}$, …, $\{x_{k-1},x_k\}$, $\{x_k,x_1\}$. The sum of cardinalities clearly remains unchanged.

Since any s pairs with $s < k$ form a union of paths in the terminology of graphs, they have at least $s+1$ elements in their union, therefore Hall's Condition is satisfied for any at most k members of the set system obtained.

Let us note that the side condition $n-2 \geq k \geq 3$ is natural, because otherwise (if $n \leq k+1$ or $k = 2$) we do not obtain a new system. As a matter of fact, for $k = 2$ the unique optimal system has two (distinct) singletons and $n-2$ identical pairs; this

[2]The case of $k = 3$ is slightly different from $k \geq 4$, because then one singleton will remain.

follows by Remark 1. On the other hand, for $n = k+1$ the starting configuration of k singletons together with $\{X\}$ is 2-regular, hence the transformation described in Section 2.2 generates a bunch of non-isomorphic optimal systems. Complementing any of these systems with $n-k-1$ copies of $\{X\}$ if $n \geq k+2$, many further optimal configurations can be obtained.

Using parameters n, k and m. In the proofs of this paper we found simpler the usage of parameters q, k, s instead of n, k, m. It is worth, however, formulating our estimates in terms of the original parameters, too. In this way, Corollaries 2 and 3 and Theorem 2 can be transcripted as follows:

- $N(n,k,m) \leq N(n-1,k,m-1)+1$ for all $n \geq m \geq k+1 \geq 2$;
- $N(n,k,m) \leq 2m + (n-m-1)\left\lfloor \frac{k}{\lfloor \frac{m-k}{n-m-1} \rfloor + 1} \right\rfloor - b$ for all $m \geq k \geq 1$ and $n \geq m+2$, where b is the residue of $m-k$ modulo $n-m-1$;
- $N(m+2,k,m) = 2m + \left\lfloor \frac{k}{m-k+1} \right\rfloor$ for all $k \leq m \leq k+\sqrt{k}$;
- $N(m+2,k,m) = m+k-2+\lceil 2\sqrt{k+1} \rceil$ for all $m > k+\sqrt{k}$.

Acknowledgement. The authors thank Srimanta Bhattacharya for a correction regarding Remark 3.

REFERENCES

[1] S. Bhattacharya, S. Ruj, and B. Roy, "Combinatorial Batch Codes: A Lower Bound and Optimal Constructions," *arXiv:1102.4951v1*, 2011.

[2] R. A. Brualdi, K. P. Kiernan, S. A. Meyer, and M. W. Schroeder, "Combinatorial batch codes and transversal matroids," *Adv. Math. Commun.*, vol. 4, no. 3, pp. 419–431, 2010.

[3] C. Bujtás and Z. Tuza, "Optimal batch codes: Many items or low retrieval requirement," *manuscript*, 2010.

[4] C. Bujtás and Z. Tuza, "Combinatorial batch codes: Extremal problems under Hall-type conditions," *Electronic Notes in Discrete Math.*, p. to appear, 2011.

[5] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Batch codes and their applications," in *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*. New York: ACM, 2004, pp. 262–271.

[6] M. B. Paterson, D. R. Stinson, and R. Wei, "Combinatorial batch codes," *Adv. Math. Commun.*, vol. 3, no. 1, pp. 13–27, 2009.

Authors' addresses

Csilla Bujtás

University of Pannonia, Department of Computer Science and Systems Technology, Egyetem u. 10, H–8200 Veszprém, Hungary

E-mail address: bujtas@dcs.vein.hu, tuza@dcs.vein.hu

Zsolt Tuza

Computer and Automation Institute, Hungarian Academy of Sciences, Kende u. 13–17, H–1111 Budapest, Hungary

E-mail address: tuza@sztaki.hu