# Optimal Codes in the Enomoto-Katona Space

Yeow Meng Chee, Han Mao Kiah, Hui Zhang and Xiande Zhang
School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

***Abstract*—Coding in a new metric space, the Enomoto-Katona space, is considered recently in connection to the study of implication structures of functional dependencies and their generalizations in relational databases. The central problem here is the determination of $C(n,k,d)$, the size of an optimal code of length $n$, weight $k$, and distance $d$ in the Enomoto-Katona space. The value of $C(n,k,d)$ is known only for some congruence classes of $n$ when $(k,d)\in\{(2,3),(3,5)\}$. In this paper, we obtain new infinite families of optimal codes in the Enomoto-Katona space. In particular, $C(n,k,2k-1)$ is determined for all sufficiently large $n$ satisfying either $n\equiv 1 \bmod k$ and $n(n-1)\equiv 0 \bmod 2k^2$, or $n\equiv 0 \bmod k$.**

## 1. Introduction

The problem we consider is motivated by implication structures of functional dependencies in relational databases.

Let $A$ be a set of $n$ *attributes*. Each attribute $x\in A$ is associated a set $\Omega_x$, called its *domain*. A *relation* is a finite set $R$ of $n$-tuples (called *data items*) so that $R\subseteq \times_{x\in A}\Omega_x$. A relation $R$ of $m$ data items may be visualized as an $m\times n$ array (called a *table*), with columns indexed by $A$, such that each row corresponds to a data item. Denote this table by $R(A)$. Formally, if $R=\{(\mathsf{d}_{i,x})_{x\in A} : 1\le i\le m\}$, then the cell in $R(A)$ with row index $i$ and column index $x$ has entry $\mathsf{d}_{i,x}$. A *relational database* is a set of tables, where tables may be defined over different attribute sets. Relational database, introduced by Codd [1], is the first database with a rigorous mathematical foundation, and remains the predominant choice for data storage and management today.

For a given table $R(A)$ and $X\subseteq A$, the $X$-*value* of a data item $\mathsf{d}=(\mathsf{d}_x)_{x\in A}$ in $R(A)$ is the $|X|$-tuple $\mathsf{d}\,|_X=(\mathsf{d}_x)_{x\in X}$. Let $X\subseteq A$ and $y\in A$ for a given table $R(A)$. We say that $y$ *(functionally) depends*[1] on $X$, written $X\to y$, if no two rows of $R(A)$ agree in $X$ but differ in $y$. In other words, if the $X$-value of a data item is known, then its $\{y\}$-value can be determined with certainty. Identifying functional dependencies is important in relational database design [2]–[5].

Demetrovics, Katona, and Sali [6] generalized functional dependencies as follows.

**Definition 1.1.** Let $X\subseteq A$ and $y\in A$ for a given table $R(A)$. Then for positive integers $p\le q$, we say that $y$ $(p,q)$-*depends* on $X$, written $X\xrightarrow{(p,q)} y$, if there do not exist $q+1$ data items (rows) $\mathsf{d}_1,\mathsf{d}_2,\ldots,\mathsf{d}_{q+1}$ of $R(A)$ such that

(i) $|\{\mathsf{d}_i|_{\{x\}}\colon 1\le i\le q+1\}|\le p$ for each $x\in X$, and
(ii) $|\{\mathsf{d}_i|_{\{y\}}\colon 1\le i\le q+1\}|=q+1$.

Our usual concept of functional dependency is equivalent to the special case of $(1,1)$-dependency. When functional dependencies are not known, $(p,q)$-dependencies identified in a relational database can still be exploited for improving storage efficiency [6]–[9].

[1]By definition, if $y\in X$, then $X\to y$ trivially.

Let $p\le q$ be positive integers. For a table $R(A)$, define the operation $J^{(p,q)}_{R(A)}:2^A\to 2^A$ so that for $X\subseteq A$, we have

$$J^{(p,q)}_{R(A)}(X)=\left\{y\in A : X\xrightarrow{(p,q)} y\right\}.$$

We call $J^{(p,q)}_{R(A)}$ the $(p,q)$-*implication structure* of $R(A)$, since it specifies the subsets of attributes that are implied by some $(p,q)$-dependency of $R(A)$. A function $J:2^A\to 2^A$ is said to be $(p,q)$-*representable* if there exists a table $R(A)$ such that $J^{(p,q)}_{R(A)}=J$.

The function $J^{(1,1)}_{R(A)}$ is a closure operator on $A$. Armstrong [2] showed that the converse is also true: any closure operator $J:2^A\to 2^A$ is $(1,1)$-representable. This is, however, not true for general $p$ and $q$ [6]. When a function $J$ is $(p,q)$-representable, there is interest in determining the table $R(A)$ with the least number of rows such that $J^{(p,q)}_{R(A)}=J$ [7]–[9]. Consideration of this problem, particularly when for fixed $k$, the function $J^k_n:2^A\to 2^A$ takes the form

$$J^k_n(X)=\begin{cases}X, & \text{if } |X|<k\\ A, & \text{otherwise,}\end{cases}$$

has led to coding-theoretic problems in a new metric space, called the *Enomoto-Katona space* [10].

### A. The Enomoto-Katona Space

If $X$ is a finite set, the set of all $k$-subsets of $X$ is denoted $\binom{X}{k}$. Let $n$ and $k$ be positive integers such that $2k\le n$ and let $X$ be an $n$-set. Consider the set

$$\mathcal{E}(X,k)=\left\{\{A,B\}\subseteq\binom{X}{k} : A\cap B=\varnothing\right\}$$

of all unordered pairs of disjoint $k$-subsets of $X$. Elements of $\mathcal{E}(X,k)$ are called *set-pairs*. The function $\mathrm{d}_{\mathcal{E}}:\mathcal{E}(X,k)\times\mathcal{E}(X,k)\to\{0,1,\ldots,2k\}$ given by

$$\mathrm{d}_{\mathcal{E}}(\{A,B\},\{S,T\})=\min\{|A\setminus S|+|B\setminus T|,|A\setminus T|+|B\setminus S|\}$$

is a metric of $\mathcal{E}(X,k)$ and the finite metric space $(\mathcal{E}(X,k),\mathrm{d}_{\mathcal{E}})$ is called the *Enomoto-Katona space*.

An *Enomoto-Katona code* (or *EK code*, in short), is a set $\mathcal{C}\subseteq\mathcal{E}(X,k)$. More specifically, $\mathcal{C}$ is an EK code of *length* $n$, *weight* $k$, and *distance* $d$, or $(n,k,d)$-EK code, if $\mathrm{d}_{\mathcal{E}}(\mathsf{u},\mathsf{v})\ge d$ for all distinct $\mathsf{u},\mathsf{v}\in\mathcal{C}$.

The following example gives a construction of a table from an EK-code (see [8], [11]).

**Example 1.1.** Consider the following $(9,2,3)$-EK code $\mathcal{C}$, where $X = \mathbb{Z}/9\mathbb{Z}$.

$$c_1 = \{\{0,1\},\{2,4\}\},\ c_2 = \{\{1,2\},\{3,5\}\},\ c_3 = \{\{2,3\},\{4,6\}\},$$
$$c_4 = \{\{3,4\},\{5,7\}\},\ c_5 = \{\{4,5\},\{6,8\}\},\ c_6 = \{\{5,6\},\{7,0\}\},$$
$$c_7 = \{\{6,7\},\{8,1\}\},\ c_8 = \{\{7,8\},\{0,2\}\},\ c_9 = \{\{8,0\},\{1,3\}\}.$$

Let $A$ be a set of nine attributes, given by $\mathcal{C}$. We construct a table $R(A)$ with nine rows indexed by $X$ whose implication structure $J^{(1,1)}_{R(A)}$ is precisely $J^2_9$. Each set-pair $\{A,B\}$ constructs a column in the following manner: place 1 at rows indexed by elements of $A$, place 2 at rows by elements of $B$ and place distinct elements from $\mathbb{Z}_{\geq 3}$ for the remaining rows.

| | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $c_7$ | $c_8$ | $c_9$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 1 |
| 1 | 1 | 1 | 4 | 4 | 4 | 3 | 2 | 3 | 2 |
| 2 | 2 | 1 | 1 | 5 | 5 | 4 | 4 | 2 | 3 |
| 3 | 3 | 2 | 1 | 1 | 6 | 5 | 5 | 4 | 2 |
| 4 | 2 | 4 | 2 | 1 | 1 | 6 | 6 | 5 | 4 |
| 5 | 4 | 2 | 5 | 2 | 1 | 1 | 7 | 6 | 5 |
| 6 | 5 | 5 | 2 | 6 | 2 | 1 | 1 | 7 | 6 |
| 7 | 6 | 6 | 6 | 2 | 7 | 2 | 1 | 1 | 7 |
| 8 | 7 | 7 | 7 | 7 | 2 | 7 | 2 | 1 | 1 |

The maximum size of an $(n,k,d)$-EK code is denoted by $C(n,k,d)$. An $(n,k,d)$-EK code of size $C(n,k,d)$ is said to be *optimal*. The central problem is to determine $C(n,k,d)$.

### B. Problem Status

Trivially, $C(n,k,1) = \binom{n}{k}\binom{n-k}{k}/2$, $C(n,k,2k) = \lfloor n/2k \rfloor$, so we assume $2 \leq d \leq 2k-1$ for the rest of this paper.

General upper and lower bounds on the size of codes in the Enomoto-Katona space have been obtained by Brightwell and Katona [12]. In particular, they showed for $1 \leq d \leq 2k \leq n$,

$$C(n,k,d) \leq \frac{\prod_{i=n-2k+d}^{n} i}{2\left(\prod_{i=\lceil (d+1)/2\rceil}^{k} i\right)\cdot\left(\prod_{i=\lfloor (d+1)/2\rfloor}^{k} i\right)}. \quad (1)$$

Brightwell and Katona [12] also showed that $C(n,k,d) = \Theta(n^{2k-d+1})$ for fixed $k$ and $d$. Bollobás *et al.* [13] (see also [11]) subsequently established that the upper bound in (1) is asymptotically tight.

**Theorem 1.1** (Bollobás et al. [13])**.**

$$\lim_{n\to\infty} \frac{C(n,k,d)}{n^{2k-d+1}} = \frac{1}{2\cdot\left(\prod_{i=\lceil (d+1)/2\rceil}^{k} i\right)\cdot\left(\prod_{i=\lfloor (d+1)/2\rfloor}^{k} i\right)}.$$

The best known upper bound is due to Quistorff [14].

**Theorem 1.2** (Quistorff Bound [14])**.** Suppose $k-d+1 \leq e \leq \min\{k, 2k-d\}$. Then

$$C(n,k,d) \leq \left\lfloor \frac{\binom{n}{e}}{2\binom{k}{e}} \left\lfloor \frac{\binom{n-e}{2k-d-e+1}}{\binom{k}{2k-d-e+1}} \right\rfloor \right\rfloor.$$

Only the following exact values of $C(n,k,d)$ are known.

**Theorem 1.3** (Bollobás et al. [13])**.**

$$C(n,2,3) = \frac{n(n-1)}{8}, \qquad \text{if } n \equiv 1 \text{ or } 9 \bmod 72,$$
$$C(n,3,5) = \frac{n(n-1)}{18}, \qquad \text{if } n \equiv 1 \text{ or } 19 \bmod 342.$$

### C. Contributions

Our contributions in this paper are as follows.

**Main Theorem.** For any fixed $k \geq 2$, we have

$$C(n,k,2k-1) = \left\lfloor \frac{n}{2k} \left\lfloor \frac{n-1}{k} \right\rfloor \right\rfloor$$

for all sufficiently large $n$ satisfying

(i) $n \equiv 1 \bmod k$ and $n(n-1) \equiv 0 \bmod 2k^2$, or
(ii) $n \equiv 0 \bmod k$.

Previous asymptotic results are known only when $k \in \{2,3\}$. In addition,

(i) We determine the exact value of $C(n,2,d)$ completely. Previously, the value of $C(n,2,2)$ is unknown and $C(n,2,3)$ is determined only when $n \equiv 1$ or $9 \bmod 72$.
(ii) The exact value of $C(n,3,5)$ is determined for $n$ belonging to a set of density $4/9$. Previously, the exact value of $C(n,3,5)$ is known only for $n \equiv 1$ or $19 \bmod 342$, a set of density $1/171$.

These results are obtained by constructing EK codes (or their equivalent combinatorial objects) whose sizes meet the Quistorff bound. Owing to space constraints, we prove the Main Theorem and determine $C(n,2,2)$ in this paper, leaving the proofs for the remaining results to the full paper.

## 2. EK Packings and Designs

Our approach is based on combinatorial design theory. In this section, we introduce necessary concepts and establish connections to EK codes.

Throughout the rest of this paper, $X$ denotes a set of size $n$. For a positive integer $k$, $[k]$ denotes the set of integers $\{1,2,\ldots,k\}$, while $\mathbb{Z}_{\geq k}$ denotes the set of integers at least $k$. The set of all (ordered) $k$-tuples of a finite set $X$ with distinct components is denoted $\overline{\binom{X}{k}}$.

We use angled brackets $\langle$ and $\rangle$ for multisets. We sometimes use the exponential notation to describe multisets so that a multiset where an element $g_i$ appears $s_i$ times, $i \in [t]$, is denoted $g_1^{s_1} g_2^{s_2} \cdots g_t^{s_t}$.

A *set system* is a pair $\mathfrak{S} = (X, \mathcal{A})$, where $X$ is a finite set of *points* and $\mathcal{A} \subseteq 2^X$. Elements of $\mathcal{A}$ are called *blocks*. The *order* of $\mathfrak{S}$ is the number of points in $X$, and the *size* of $\mathfrak{S}$ is the number of blocks in $\mathcal{A}$. Let $K \subseteq \mathbb{Z}_{\geq 0}$. The set system $(X,\mathcal{A})$ is said to be *$K$-uniform* if $|A| \in K$ for all $A \in \mathcal{A}$.

Let $2 \leq t < 2k$ and $0 \leq e \leq \min\{k, \lfloor t/2 \rfloor\}$. We say that the tuple $(x_1,x_2,\ldots,x_t) \in \overline{\binom{X}{t}}$ is *$(e,t)$-contained* in a set-pair $\{A,B\} \in \mathcal{E}(X,k)$ if either $\{x_1,x_2,\ldots,x_e\} \subseteq A$ and $\{x_{e+1},x_{e+2},\ldots,x_t\} \subseteq B$, or $\{x_1,x_2,\ldots,x_e\} \subseteq B$ and $\{x_{e+1},x_{e+2},\ldots,x_t\} \subseteq A$.

Let $\mathcal{C} \subseteq \mathcal{E}(X,k)$. Then $(X,\mathcal{C})$ is an EK packing of *strength* $t$, or more precisely a *$t$-$(n,k)$ EK packing*[2], if for $0 \leq e \leq \lfloor t/2 \rfloor$, every $t$-tuple in $\overline{\binom{X}{t}}$ is $(e,t)$-contained in at most one set-pair in $\mathcal{C}$. A *$t$-$(n,k)$ EK design* is a $t$-$(n,k)$ EK packing satisfying the condition that for $e = \lfloor t/2 \rfloor$, every $t$-tuple in

[2]Note that $\mathcal{C} \subseteq \mathcal{E}(X,k)$, while $\mathcal{A} \subseteq 2^X$.

$\overline{\binom{X}{t}}$ is $(e,t)$-contained in exactly one set-pair in $\mathcal{C}$. It is easy to see that if $(X,\mathcal{C})$ is a $t$-$(n,k)$ EK design, then

$$|\mathcal{C}| = \frac{\binom{n}{t}\binom{t}{\lfloor t/2 \rfloor}}{2\binom{k}{\lfloor t/2 \rfloor}\binom{k}{\lceil t/2 \rceil}}.$$

EK packings of strength $t$ are equivalent to EK codes of distance $2k-t+1$, while EK designs of strength $t$ give rise to optimal EK codes of distance $2k-t+1$.

**Proposition 2.1.** Let $\mathcal{C} \subseteq \mathcal{E}(X,k)$. Then $(X,\mathcal{C})$ is a $t$-$(n,k)$ EK packing if and only if $\mathcal{C}$ is an $(n,k,2k-t+1)$-EK code. Furthermore, if $(X,\mathcal{C})$ is a $t$-$(n,k)$ EK design, then $\mathcal{C}$ is an optimal $(n,k,2k-t+1)$-EK code.

*Proof:* Suppose $(X,\mathcal{C})$ is a $t$-$(n,k)$ EK packing and $\{A,B\}$, $\{S,T\} \in \mathcal{C}$. We claim that $d_{\mathcal{E}}(\{A,B\},\{S,T\}) \geq 2k-t+1$. Suppose otherwise. Then without loss of generality, $|A \setminus S| + |B \setminus T| \leq 2k-t$ and there exists a nonnegative $e \leq \lfloor t/2 \rfloor$, $I \in \binom{X}{e}$, $J \in \binom{X}{t-e}$ such that $I \subseteq A \cap S$ and $J \subseteq B \cap T$. If $I = \{x_1, x_2, \ldots, x_e\}$ and $J = \{x_{e+1}, x_{e+2}, \ldots, x_t\}$, we see that $(x_1, x_2, \ldots, x_t)$ is $(e,t)$-contained in $\{A,B\}$ and $\{S,T\}$, contradicting the fact that $(X,\mathcal{C})$ is a $t$-$(n,k)$ EK packing.

Conversely, suppose $\mathcal{C}$ is an $(n,k,2k-t+1)$-EK code. If $(X,\mathcal{C})$ is not a $t$-$(n,k)$ EK packing, then there exists a nonnegative $e \leq \lfloor t/2 \rfloor$, $(x_1, x_2, \ldots, x_t) \in \overline{\binom{X}{t}}$, and $\{A,B\}$, $\{S,T\} \in \mathcal{C}$ such that $(x_1, x_2, \ldots, x_t)$ is $(e,t)$-contained in $\{A,B\}$ and $\{S,T\}$. Without loss of generality, $\{x_1, x_2, \ldots, x_e\} \subseteq A \cap S$ and $\{x_{e+1}, x_{e+2}, \ldots, x_t\} \subseteq B \cap T$. Hence, $|A \setminus S| + |B \setminus T| \leq 2k-(e+t-e) = 2k-t$, and consequently $d_{\mathcal{E}}(\{A,B\},\{S,T\}) \leq 2k-t$, contradicting the fact that $\mathcal{C}$ is an $(n,k,2k-t+1)$-EK code.

Finally, when $(X,\mathcal{C})$ is a $t$-$(n,k)$ EK design, $\mathcal{C}$ is an optimal $(n,k,2k-t+1)$-EK code, since $|\mathcal{C}|$ meets the Quistorff bound with $e = \lfloor t/2 \rfloor$. ■

In view of Proposition 2.1, our strategy in constructing optimal EK codes (and hence determining $C(n,k,d)$) is to construct equivalent EK packings and designs of sizes meeting the Quistorff bound. We introduce next EK group divisible designs and their connections to EK codes and EK packings.

### A. EK Group Divisible Designs

Let $\mathcal{G} = \{G_1, G_2, \ldots, G_s\}$ be a partition of an $n$-set $X$ and $\mathcal{C} \subseteq \mathcal{E}(X,k)$. Then $(X,\mathcal{G},\mathcal{C})$ is an *EK group divisible design* (or EKGDD, in short) if for all $(x,y) \in \overline{\binom{X}{2}}$ such that $\{x,y\} \not\subseteq G_i$ for all $i \in [s]$, we have

(i) $(x,y)$ is $(1,2)$-contained in exactly one set-pair $\{A,B\}$,

(ii) $(x,y)$ is $(0,2)$-contained in at most one set-pair $\{A,B\}$.

In addition, $|G_i \cap (A \cup B)| \leq 1$ for all $i \in [s]$ and $\{A,B\} \in \mathcal{C}$. Such an EKGDD is more precisely called a $(k,T)$-EKGDD, where $T = \langle |G_i| : i \in [s] \rangle$.

A 2-$(n,k)$ EK design can be regarded as a $(k,1^n)$-EKGDD, where each group contains just a single point. Furthermore, a $(k, g_1 g_2 \cdots g_s)$-EKGDD can be regarded as a 2-$(k, \sum_{i=1}^{s} g_i)$ EK packing, and hence as a $(\sum_{i=1}^{s} g_i, k, 2k-1)$-EK code. In addition, as the following shows, certain classes of EKGDD give optimal EK codes.

**Proposition 2.2.** Suppose there exists a $(k,k^s)$-EKGDD $(X,\mathcal{G},\mathcal{C})$. Then $\mathcal{C}$ is an optimal $(ks,k,2k-1)$-EK code.

*Proof:* Observe $\mathcal{C}$ is a $(ks,k,2k-1)$-EK code since $(X,\mathcal{C})$ is an 2-$(ks,k)$ EK packing. There are $(ks)\cdot(ks-k)$ ordered pairs $(x,y) \in \overline{\binom{X}{2}}$ where $\{x,y\}$ does not belong to any group. In addition, we have $2k^2$ ordered pairs in $\overline{\binom{X}{2}}$ that are $(1,2)$-contained in each set-pair. Hence, the code $\mathcal{C}$ is of size $s(s-1)/2$, which meets the Quistorff bound. ■

## 3. $C(n,k,2k-1)$ for Sufficiently Large $n$

We show that a 2-$(n,k)$ EK design and a $(k,k^n)$-EKGDD exist when $n$ belongs to certain congruence classes, provided $n$ is sufficiently large. Our proof is an application of decompositions of edge-colored directed graphs (digraphs).

An *edge-colored directed graph* is a triple $G = (V,C,E)$, where $V$ is a finite set of *vertices*, $C$ is a finite set of *colors* and $E$ is a subset of $\overline{\binom{V}{2}} \times C$. Members of $E$ are called *edges*. The *complete edge-colored digraph* on $n$ vertices with $r$ colors, denoted by $K_n^{(r)}$, is the edge-colored digraph $(V,C,E)$, where $|V| = n$, $|C| = r$, and $E = \overline{\binom{V}{2}} \times C$.

A family $\mathcal{F}$ of edge-colored subgraphs of an edge-colored digraph $K$ is a *decomposition* of $K$ if every edge of $K$ belongs to exactly one member of $\mathcal{F}$. Given an edge-colored digraph $G$, a decomposition $\mathcal{F}$ of $K$ is a *$G$-decomposition of $K$* if each edge-colored digraph in $\mathcal{F}$ is isomorphic to $G$.

Lamken and Wilson [15] studied the existence of $G$-decompositions of $K_n^{(r)}$ and showed that for fixed $G$ and $r$, a $G$-decomposition exists for sufficiently large $n$ under certain conditions. To state the theorem, we require more concepts.

Consider an edge-colored digraph $G = (V,C,E)$ with $|C| = r$. Let $((u,v),c) \in E$ denote a directed edge from $u$ to $v$, colored by $c$. For any vertex $u$ and color $c$, define the *indegree* and *outdegree* of $u$ with respect to $c$ as follows:

$$\mathrm{in}_c(u) = |\{v : ((v,u),c) \in E\}|,$$
$$\mathrm{out}_c(u) = |\{v : ((u,v),c) \in E\}|.$$

Then for vertex $u$, we define the *degree vector* of $u$, denoted by $\boldsymbol{\delta}(u)$, to be the vector of length $2r$. That is, $\boldsymbol{\delta}(u) = (\mathrm{in}_c(u), \mathrm{out}_c(u))_{c \in C}$. Define $\alpha(G)$ to be the least positive integer $t$ such that $(t,t,\ldots,t)$ is an integral linear combination of the vectors in $\{\boldsymbol{\delta}(u) : u \in V\}$. The following is due to Lamken and Wilson [15].

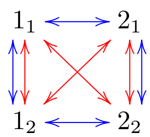
**Theorem 3.1** (Lamken and Wilson [15, Theorem 1.1])**.** Let $G$ be an edge-colored digraph with $r$ colors and $m$ edges of each of $r$ different colors. There exists a constant $n_0$ such that there is $G$-decomposition of $K_n^{(r)}$ for all $n \geq n_0$ satisfying both

$$n(n-1) \equiv 0 \bmod m \text{ and } n-1 \equiv 0 \bmod \alpha(G).$$

Now for fixed $k \geq 2$ define the edge-colored digraph $G_k = (V_k, C_k, E_k)$, where

$$\begin{aligned} V_k &= \{i_j : i \in [k], j \in [2]\}, \\ C_k &= \{{\color{red}\bullet}, {\color{blue}\bullet}\}, \\ E_k &= \{((i_r, j_s), {\color{red}\bullet}) : i, j \in [k], (r, s) \in \{(1,2),(2,1)\}\} \\ &\quad \cup \{((i_r, i_s), {\color{blue}\bullet}) : i \in [k], (r, s) \in \{(1,2),(2,1)\}\} \\ &\quad \cup \left\{((i_r, j_r), {\color{blue}\bullet}) : (i,j) \in \overline{\binom{[k]}{2}}, r \in [2]\right\}. \end{aligned}$$

**Example 3.1.** The edge-colored graph $G_2$ is given by




where ${\color{blue}\longleftrightarrow}$ denotes two directed edges of color ${\color{blue}\bullet}$ (one in each direction), and ${\color{red}\longleftrightarrow}$ denotes two directed edges of color ${\color{red}\bullet}$ (one in each direction).

**Proposition 3.1.** If a $G_k$-decomposition of $K_n^{(2)}$ exists, then a 2-$(n, k)$ EK design exists.

*Proof:* Let $\mathcal{F}$ be a $G_k$-decomposition of $K_n^{(2)}$. Then for a subgraph $G \in \mathcal{F}$, let $\phi_G : G_k \to G$ be a graph isomorphism and define

$$A_G = \{\phi_G(i_1) : i \in [k]\}, \quad B_G = \{\phi_G(i_2) : i \in [k]\}.$$

Let $X$ be the vertex set of $K_n^{(2)}$ and

$$\mathcal{C} = \{\{A_G, B_G\} : G \in \mathcal{F}\}.$$

We claim that $(X, \mathcal{C})$ is a 2-$(n, k)$ EK design. Since $|\mathcal{C}| = n(n-1)/(2k(k-1))$, it suffices to check that for $e \in \{0, 1\}$, each $(x, y) \in \overline{\binom{X}{2}}$ is $(e, 2)$-contained in at most one set-pair in $\mathcal{C}$.

Suppose otherwise. Then there exist $(x, y) \in \overline{\binom{X}{2}}$, $G, H \in \mathcal{F}$ and $e \in \{0, 1\}$ such that $(x, y)$ is $(e, 2)$-contained in $\{A_G, B_G\}$ and $\{A_H, B_H\}$.

If $e = 0$, then assume that $\{x, y\} \subset A_G \cap A_H$. Hence, the edge $((x, y), {\color{red}\bullet})$ belongs to both $G$ and $H$, contradicting the fact that $\mathcal{F}$ is a $G_k$-decomposition of $K_n^{(2)}$.

If $e = 1$, then assume that $x \in A_G \cap A_H$ and $y \in B_G \cap B_H$. Hence, the edge $((x, y), {\color{blue}\bullet})$ belongs to $G$ and $H$, contradicting the fact that $\mathcal{F}$ is a $G_k$-decomposition of $K_n^{(2)}$. ■

Observe there are $2k^2$ edges of each color in $G_k$ and $\boldsymbol{\delta}(v) = (k, k, k, k)$ for all $v \in V_k$. Hence, $\alpha(G_k) = k$. The following is immediate from Propositions 2.1, 3.1, and Theorem 3.1.

**Theorem 3.2.** Fix $k \geq 2$. Then

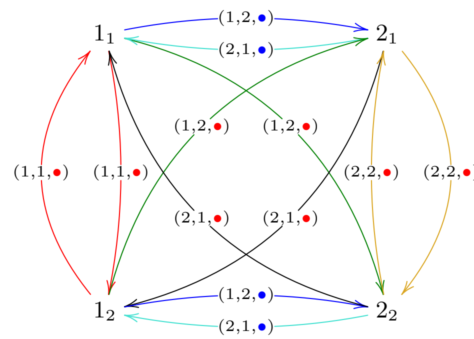
$$C(n, k, 2k-1) = \frac{n(n-1)}{2k^2}$$

for all sufficiently large $n$ satisfying $n \equiv 1 \bmod k$ and $n(n-1) \equiv 0 \bmod 2k^2$.

To determine $C(n, k, 2k-1)$ when $n \equiv 0 \bmod k$, consider the following graph. Fix $k \geq 2$ and define the edge-colored digraph $H_k = (V_k, C_k, E_k)$, where

$$\begin{aligned} V_k &= \{i_j : i \in [k], j \in [2]\}, \\ C_k &= ([k] \times [k] \times \{{\color{red}\bullet}\}) \cup \left(\overline{\binom{[k]}{2}} \times \{{\color{blue}\bullet}\}\right), \\ E_k &= \{((i_r, j_s), (i, j, {\color{red}\bullet})) : i, j \in [k], (r, s) \in \{(1,2),(2,1)\}\} \\ &\quad \cup \left\{((i_r, j_r), (i, j, {\color{blue}\bullet})) : (i, j) \in \overline{\binom{[k]}{2}}, r \in [2]\right\}. \end{aligned}$$

**Example 3.2.** The graph $H_2$ is given by




**Proposition 3.2.** If an $H_k$-decomposition of $K_n^{(2k^2-k)}$ exists, then a $(k, k^n)$-EKGDD exists.

*Proof:* Let $\mathcal{H}$ be an $H_k$-decomposition of $K_n^{(2k^2-k)}$. Then for a subgraph $H \in \mathcal{H}$, let $\phi_H : H_k \to H$ be a graph isomorphism and define

$$A_H = \{\phi_H(i_1)_i : i \in [k]\}, \quad B_H = \{\phi_H(i_2)_i : i \in [k]\}.$$

Let $V$ be the vertex set of $K_n^{(2k^2-k)}$ and

$$\begin{aligned} X &= \{v_i : v \in V, i \in [k]\}, \\ \mathcal{G} &= \{\{v_i : i \in [k]\} : v \in V\}, \\ \mathcal{C} &= \{\{A_H, B_H\} : H \in \mathcal{H}\}. \end{aligned}$$

We claim that $(X, \mathcal{G}, \mathcal{C})$ is a $(k, k^n)$-EKGDD. Suppose otherwise. Since $|\mathcal{C}| = n(n-1)/2$, it suffices to consider the following two cases.

(i) There exist $v \in V$ and $H \in \mathcal{H}$ such that $|\{v_i : i \in [k]\} \cap (A_H \cup B_H)| \geq 2$. This contradicts the fact that $H$ is isomorphic to $H_k$.

(ii) There exist $(x, y) \in \overline{\binom{X}{2}}$, $G, H \in \mathcal{H}$ and $e \in \{0, 1\}$ such that $(x_i, y_j)$ is $(e, 2)$-contained in $\{A_G, B_G\}$ and $\{A_H, B_H\}$.
If $e = 0$, then assume that $\{x_i, y_j\} \subset A_G \cap A_H$. Hence, the edge $((x, y), (i, j, {\color{red}\bullet}))$ belongs to both $G$ and $H$, contradicting the fact that $\mathcal{H}$ is an $H_k$-decomposition.
Similarly, if $e = 1$, then assume that $x_i \in A_G \cap A_H$ and $y_j \in B_G \cap B_H$. Hence, the edge $((x, y), (i, j, {\color{blue}\bullet}))$ belongs to both $G$ and $H$, contradicting the fact that $\mathcal{H}$ is an $H_k$-decomposition of $K_n^{(2k^2-k)}$.

■

Observe there are two edges of each color in $H_k$ and $\sum_{i\in[k]} \boldsymbol{\delta}(i_1) = (1,1,\ldots,1)$. Hence, $\alpha(H_k) = 1$. From Propositions 2.2, 3.2, and Theorem 3.1, we have the following.

**Theorem 3.3.** Fix $k \geq 2$. Then

$$C(n,k,2k-1) = \frac{n(n-k)}{2k^2}$$

for all sufficiently large $n$ satisfying $n \equiv 0 \bmod k$.

Theorems 3.2 and 3.3 combine to give the Main Theorem.

## 4. The Value of $C(n,2,2)$

In this section, we give a complete solution for $C(n,2,2)$. Our proof makes use of $t$-wise balanced designs.

**Definition 4.1.** A *$t$-wise balanced design*, or a $t$-BD$(v,K)$, is a $K$-uniform set system $(X,\mathcal{A})$ of order $v$ such that every $t$-subset of $X$ is contained in exactly one block of $\mathcal{A}$.

The following existence result for 3-BDs is known.

**Theorem 4.1** (Hanani [16])**.** A 3-BD$(v,\{4,6\})$ exists for all even $v \geq 4$.

The following proposition gives a recursive construction for EK designs of strength $t$.

**Proposition 4.1** (Filling in Blocks)**.** Let $K \subseteq \mathbb{Z}_{\geq 1}$ and suppose that a $t$-BD$(v,K)$ exisits. If a $t$-$(h,k)$ EK design exists for all $h \in K$, then a $t$-$(v,k)$ EK design exists.

*Proof:* Let $(X,\mathcal{A})$ be a $t$-BD$(v,K)$. For each $A \in \mathcal{A}$, let $(A,\mathcal{C}_A)$ be a $t$-$(|A|,k)$ EK design. Then $(X,\cup_{A\in\mathcal{A}}\mathcal{C}_A)$ is a $t$-$(v,k)$ EK design. ■

We first determine $C(n,2,2)$ when $n$ is even.

**Proposition 4.2.** A 3-$(n,2)$ EK design exists for even $n \geq 4$.

*Proof:* When $n=4$, the pair $(X,\mathcal{C})$, where

$$\begin{aligned} X &= \mathbb{Z}/4\mathbb{Z}, \\ \mathcal{C} &= \{\{\{0,1\},\{2,3\}\},\{\{0,2\},\{1,3\}\},\{\{0,3\},\{1,2\}\}\}, \end{aligned}$$

is a 3-$(4,2)$ EK design.

When $n=6$, let

$$\begin{aligned} X &= \mathbb{Z}/6\mathbb{Z}, \\ \mathcal{C}_0 &= \{\{\{0,1\},\{2,4\}\},\{\{0,1\},\{3,5\}\},\{\{0,2\},\{3,1\}\}, \\ &\qquad \{\{0,3\},\{1,4\}\},\{\{0,5\},\{1,2\}\}\}, \\ \mathcal{C} &= \{\{\{a+i,b+i\},\{c+i,d+i\}\} : \\ &\qquad \{\{a,b\},\{c,d\}\} \in \mathcal{C}_0, i \in \{0,2,4\}\}. \end{aligned}$$

Then $(X,\mathcal{C})$ is a 3-$(6,2)$ EK design.

For $n \geq 8$, there exists a 3-BD$(n,\{4,6\})$ by Theorem 4.1. The result now follows from Proposition 4.1. ■

**Proposition 4.3.** There exists a 3-$(n,2)$ EK packing of size $n(n-1)(n-3)/8$ for all odd $n \geq 5$.

*Proof:* By Proposition 4.2, there exists a 3-$(n+1,2)$ EK design $(X,\mathcal{C})$. Fix any point $x \in X$ and define

$$X' = X \setminus \{x\}, \quad \mathcal{C}' = \{\{A,B\} \in \mathcal{C} : x \notin A \cup B\}.$$

Since $x$ is contained in exactly $n(n-1)/2$ set-pairs in $\mathcal{C}$, we have $|\mathcal{C}'| = n(n+1)(n-1)/8 - n(n-1)/2 = n(n-1)(n-3)/8$. ■

Propositions 2.1, 4.2, 4.3, and Theorem 1.2 combine to give the following.

**Theorem 4.2.** Let $n \geq 4$. Then

$$C(n,2,2) = \begin{cases} \dfrac{n(n-1)(n-2)}{8}, & \text{if } n \text{ is even,} \\ \dfrac{n(n-1)(n-3)}{8}, & \text{if } n \text{ is odd.} \end{cases}$$

## 5. Conclusion

New infinite families of optimal codes in the Enomoto-Katona space are obtained in this paper. In particular, we show that $C(n,k,2k-1)$ attains the Quistorff bound for infinitely many $n$. The value of $C(n,2,2)$ is also completely determined.


## Acknowledgement

Research of the authors is supported in part by the Singapore National Research Foundation under Research Grant NRF-CRP2-2007-03. The authors are grateful to the anonymous reviewers and the TPC member whose comments greatly improved the presentation of the paper.


## References


[1] E. F. Codd, "A relational model of data for large shared data banks," *Comm. ACM*, vol. 13, no. 6, pp. 377–387, 1970.

[2] W. W. Armstrong, "Dependency structures of data base relationships," in *IFIP Conference Proceedings*. North-Holland, 1974, pp. 580–583.

[3] P. A. Bernstein, "Synthesizing third normal form relations from functional dependencies," *ACM Trans. Database Syst.*, vol. 1, no. 4, pp. 277–298, 1976.

[4] C. Beeri, R. Fagin, and J. H. Howard, "A complete axiomatization for functional and multivalued dependencies in database relations," in *SIGMOD '77 – Proceedings of the 1977 ACM SIGMOD International Conference on Management of Data*, 1977, pp. 47–61.

[5] J. Rissanen, "Independent components of relations," *ACM Trans. Database Syst.*, vol. 2, no. 4, pp. 317–325, 1977.

[6] J. Demetrovics, G. O. H. Katona, and A. Sali, "The characterization of branching dependencies," *Discrete Appl. Math.*, vol. 40, no. 2, pp. 139–153, 1992.

[7] ——, "Minimal representations of branching dependencies," *Acta Sci. Math. (Szeged)*, vol. 60, no. 1–2, pp. 213–223, 1995.

[8] ——, "Design type problems motivated by database theory," *J. Statist. Plann. Inference*, vol. 72, no. 1–2, pp. 149–164, 1998.

[9] G. O. H. Katona and A. Sali, "New type of coding problem motivated by database theory," *Discrete Appl. Math.*, vol. 144, no. 1–2, pp. 140–148, 2004.

[10] H. Enomoto and G. O. H. Katona, "Pairs of disjoint $q$-element subsets far from each other," *Electron. J. Combin.*, vol. 8, no. 2, p. R7, 2001.

[11] A. Sali, "Coding theory motivated by relational databases," in *SDKB '10 – Proceedings of the 4th International Conference on Semantics in Data and Knowledge Bases*, ser. Lecture Notes in Comput. Sci., vol. 6834. Springer-Verlag, 2011, pp. 96–113.

[12] G. Brightwell and G. O. H. Katona, "A new type of coding problem," *Studia Sci. Math. Hungar.*, vol. 38, pp. 139–147, 2001.

[13] B. Bollobás, G. O. H. Katona, and I. Leader, "A coding problem for pairs of subsets," in preparation.

[14] J. Quistorff, "Combinatorial problems in the Enomoto-Katona space," *Studia Sci. Math. Hungar.*, vol. 46, no. 1, pp. 121–139, 2009.

[15] E. R. Lamken and R. M. Wilson, "Decompositions of edge-colored complete graphs," *J. Combin. Theory Ser. A*, vol. 89, no. 2, pp. 149–200, 2000.

[16] H. Hanani, "Truncated finite planes," in *Combinatorics (Proc. Sympos. Pure Math., Vol. XIX, Univ. California, Los Angeles, Calif., 1968)*, T. S. Motzkin, Ed. Providence, R.I.: Amer. Math. Soc., 1971, pp. 115–120.