*Iranian Journal of Operations Research*
Vol. 16, No. 1, 2025, pp. 60-74

# Computing the Capacity of Sum-networks with Dependent Sources

Mehdi Ghiyasvand[1,*], Sepideh Ghazvineh[2]

*A sum-network is a directed acyclic network with multiple sources and multiple sinks, where each sink demands the sum of the independent information generated at the sources. The coding capacity of sum networks with independent sources has been investigated in Tripathy and Ramamoorthy (2015). This paper shows that dependencies between the sources can change the upper bound of the coding capacity of sum-networks. We prove that the upper bound of the coding capacity of a sum network with dependent sources is greater than 1 which is different from the results in Tripathy and Ramamoorthy (2015). It is also shown that a non-solvable sum-network with independent sources can be converted to a solvable sum-network when the sources have arbitrary dependencies.*

***Keywords***: *Network coding, Sum-network, Capacity region, Dependent sources, Balanced incomplete block design, Entropy function.*

## 1. Introduction

The work in the area of function computation over a communication network has been received attention in the last years [12, 13, 14, 18, 21]. In a function computation problem, each sink wants to compute some function of a subset of source messages. In recent years, the concept of network coding was proposed in the area of function computation problems [31, 32, 33]. The sum-network problems are a special kind of function computation problems because they need a simple function such as the sum [17, 30, 32, 34]. A sum-network is considered as a communication network with multiple sources and multiple sinks such that each sink requires the sum of all the messages generated at the source nodes. The first work in this area was done by Ramamoorthy in [32]. He showed that, for sum-networks with at most two sources or two sinks, the sum of source messages can be communicated to the sinks if and only if each source-sink pair is connected.

One of the fundamental problems in network coding problems is to characterize the capacity region of them. In [2], the weighted Hamming distance to measure the modification of the arc capacities is considered. Network routing capacity, network coding capacity and channel coding capacity were investigated in [4, 9, 11, 10, 15, 26]. In [1], it was shown that the network coding capacity of a multicast network is the minimum of the min-cuts from the source to the individual sinks. Li et al.(2003) showed that 1 the capacity of multicast networks is achieved by linear network codes. By the assumption that sources are independent, the network coding capacity of multi-source multi-terminal networks was studied in [5, 20, 16, 38].

Some factors can change the coding capacity of communication networks. For multiple unicast networks, Cannons et al. (2006) proved that the coding capacity is independent of the using alphabet [4]. Also, the linear coding capacity of networks over ring and module alphabets has been considered in [6]. Furthermore, under the assumption that sources are dependent, the capacity region of multi-

[1] Department of Mathematics, University of Bu-Ali Sina, Hamedan, Iran, Email: mghiyasvand@basu.ac.ir.
* Corresponding Author.

[2] Department of Mathematics, University of Bu-Ali Sina, Hamedan, Iran, Email: s.ghazvineh@sci.basu.ac.ir.

source multi-terminal network coding problems and line networks were investigated in [4, 19, 22]. In [23], the edge-removal problem and its connections to the δ-dependent problems is investigated.
For sum-networks with independent sources, the capacity region has been studied extensively and some bounds on the network coding capacity of these networks were presented in [7, 24, 27, 28, 29, 36, 37]. In [8, 27, 28, 37], sum-networks with arbitrary capacities were constructed. In [28], a ratio m/n (where m and n are non-negative integers) was considered and a sum-network with capacity m/n was constructed that has $2n^2 - n$ sources and $2n^2 - n + 1$ sinks. In [27, 37], the work of [28] was generalized and sum-networks of smaller size was constructed that had capacity m/n. Tripathy et al. (2015) showed that the coding capacity of the sum-networks depends on the characteristic of finite field F used as the message alphabet [36]. Moreover, they constructed a sum-network with independent sources that the coding capacity of it is at most 1.

**Main contributions:**

This paper considers sum-networks whose capacities depend on the dependency or independence of information generated by the sources. It constructs a sum-network which the upper bound of its capacity is 1 as the sources are independent, and it can be (strictly) greater than 1 as the sources are dependent. Also, an example of a non-solvable sum-network with independent sources is presented, it becomes solvable when the sources are dependent.

# 2. Preliminaries

## 2.1. System model

We consider a communication network as a directed, acyclic, and finite graph $G = (V, E, S, T)$, where V is the set of nodes, $E \subseteq V \times V$ is the set of links, $S \subset V$ is the set of source nodes and $T \subset V$ is the set of sink nodes. For any link $e = (u, v) \in E$, node $u \in V$ is called the tail of e, and node $v \in V$ is called the head of $e$, and are denoted by $u = tail(e)$ and $v = head(e)$, respectively. Moreover, we call $e$ an incoming link of $v$ and an outgoing link of $u$. For two links $e, e' \in E$, link $e$ is an incoming link of $e'$ (or $e'$ an outgoing link of $e$) if $tail(e') = head(e)$. The edges are delay-free and the capacity of each link is assumed to be one unit. For each $v \in V$ , the set of incoming edges of $v$ is denoted by $In(v)$. We assume that each source process is uniformly distributed over the finite field $\mathcal{F}$. Each source node does not have any incoming edge and each sink node does not have any outgoing edge.
By [36], a network code is an assignment of a local encoding function to each edge and a decoding function to each sink. A $(r, l)$ fractional network code is described as follows:

(1) For an edge e with $tail(e) = v$, a local encoding function is defined as

$\bar{\phi}_e: \mathcal{F}^r \to \mathcal{F}^l, \qquad if\ v \in S,$
and
$\bar{\phi}_e: \mathcal{F}^{l|In(v)|} \to \mathcal{F}^l, \qquad if\ v \notin S.$

(2) For a sink $t_i$ a decoding function is defined as

$\psi_{t_i}: \mathcal{F}^{l|In(t_i)} \to \mathcal{F}^r.$

In a sum-network each sink wants to recover the sum of the $r$-length vectors produced at the sources. If $r$ source symbols can be transferred to the sinks in $l$ units of time, then a network has a $(r, l)$ fractional network code solution over $\mathcal{F}$. The ratio $r/l$ is called as rate of the $(r, l)$ fractional network code. A rate $r/l$ is achievable if there is a $(r, l)$ fractional network code solution for the network. Moreover, the supremum of all achievable rates is called the capacity of the network. Also, a network is solvable if it has a $(1, 1)$ network coding solution.

### 2.2. Constructing a sum-network

Constructing a sum-network using balanced incomplete block designs (BIBDs) was illustrated in [36]. In this section, we briefly explain this procedure. First, by [35], we define a $2-(v, k, \lambda)$ balanced incomplete block design as follows:

**Definition 2.1** [35] A $2-(v, k, \lambda)$ balanced incomplete block design (BIBD) is a set system $\mathcal{D} = (P, \mathcal{B})$ with the following two components.

(1) A set $P$ is formed from $v$ elements that are indexed in arbitrary order as $P = \{p_1, p_2, \dots, p_v\}$, where these $v$ elements are called points.
(2) A set $\mathcal{B}$ of size b whose elements are k-subsets of $P$ such that $\mathcal{B} = \{B_1, B_2, \dots, B_{\boldsymbol{b}}\}$. $\boldsymbol{\mathcal{B}}$ has the following regularity property. For $p_i, p_j \in P, i \neq j$,

$$\left|\{\boldsymbol{\mathcal{B}} \in \boldsymbol{B} : \boldsymbol{p_i} \in \boldsymbol{B}, \boldsymbol{p_j} \in \boldsymbol{B}\}\right| = \boldsymbol{\lambda}.$$

For any $p \in P$ and $B \in \boldsymbol{\mathcal{B}}$, by [35], two sets $< p >$ and $< \boldsymbol{\mathcal{B}} >$ are defined as follows:
$< p >= \{B \in \boldsymbol{\mathcal{B}} : \boldsymbol{p} \in \boldsymbol{B}\}$, $< \boldsymbol{B} >= \cup_{\boldsymbol{p} \in \boldsymbol{B}} < \boldsymbol{p} >= \cup_{\boldsymbol{p} \in \boldsymbol{B}} \{\boldsymbol{B'} \in \boldsymbol{\mathcal{B}} : \boldsymbol{p} \in \boldsymbol{B'}\}$.

By [36], a sum-network $G = (V, E)$ can be constructed from any BIBD $\boldsymbol{\mathcal{D}}$. At first, the vertex set $V$ is described as $V = S \cup T \cup M^H \cup M^T$, where $S = \{s_p : p \in P\} \cup \{s_B : B \in \mathcal{B}\}, T = \{t_p : p \in P\} \cup \{t_B : B \in \boldsymbol{\mathcal{B}}\}$, $M^H = \{m_{1h}, m_{2h}, \dots, m_{vh}\}$ and $M^T = \{m_{1t}, m_{2t}, \dots, m_{vt}\}$. Also, the edge set $E$ is denoted as $E = M \cup D$, where $M$ contains $v$ unit-capacity bottleneck edges $e_i = (m_{it}, m_{ih})$, $i \in \{1, 2, \dots, v\}$ and the following edges for all $p_i \in P$

- $(s_{p_i}, m_{it})$ *and* $(s_{B_j}, m_{it})$ *for all* $B_j \in < p_i >$,
- $(m_{ih}, t_{p_i})$ *and* $(m_{ih}, t_{B_j})$ *for all* $B_j \in < p_i >$.

Moreover, $D$ contains the following four groups of unit-capacity direct edges for every $p_i \in P$ and $B_j \in \boldsymbol{\mathcal{B}}$,

- $(s_{p_l}, t_{p_i})$ *for all* $p_l \neq p_i$,
- $(s_{B_l}, t_{p_i})$ *for all* $B_l \notin < p_i >$,
- $(s_{p_l}, t_{B_j})$ *for all* $p_l \notin B_j$,
- $(s_{B_l}, t_{B_j})$ *for all* $B_l \notin < B_j >$.

The next example constructs a sum-network from a given $2-(3, 2, 1)$ design.

**Example 2.1.** Consider a $2-(3, 2, 1)$ design D = ({1, 2, 3}, {A, B, C}), where A = {1, 2}, B = {2, 3} and C = {1, 3}. We can construct a sum-network G = (V, E) by defining the following sets:

$$S = \{s_1, s_2, s_3, s_A, s_B, s_C\},$$

$$T = \{t_1, t_2, t_3, t_A, t_B, t_C\},$$
$$M^H \cup M^T = \{m_{1h}, m_{2h}, m_{3h}, m_{1t}, m_{2t}, m_{3t}\},$$
$$M = \{e_1 = (m_{1t}, m_{1h}), e_2 = (m_{2t}, m_{2h}), e_3 = (m_{3t}, m_{3h})\}.$$

Also, there are direct edges that connect sources to bottleneck edges and some other direct edges that connect bottleneck edges to terminals. A part of the constructed sum-network is depicted in Figure 1.

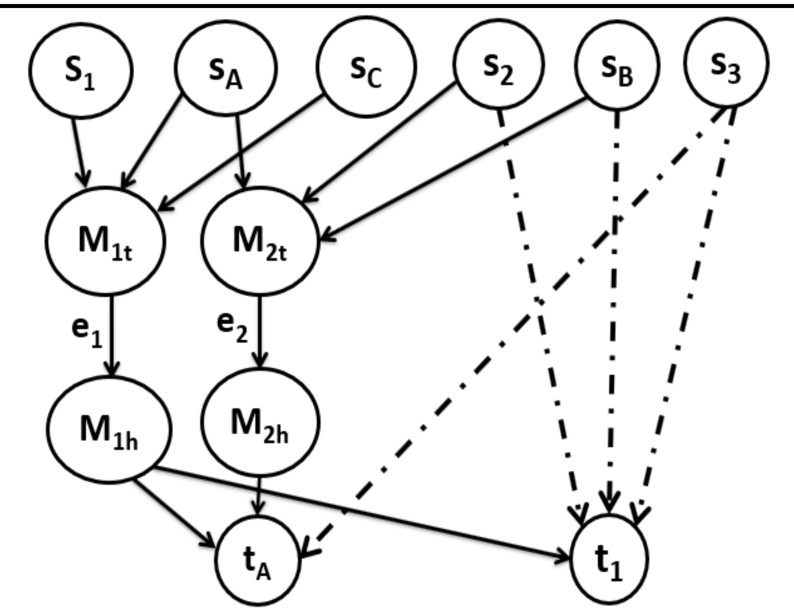


**Figure 1**. A part of the constructed sum-network from a 2 − (3, 2, 1) design.

## 2.3. Dependent sources

This section introduces some definitions and notations about dependent random variables. We show that the upper bound of the coding capacity of constructed sum-networks from BIBDs is increased when the sources are dependent. Let $\mathcal{D}$ be a $2-(v,k,1)$ design and $G$ be the constructed sum-network from it. Suppose that there exists a $(r,l)$ fractional network code solution with rate $r/l$ for constructed sum-network $G$ over $\mathcal{F}$. Then, $r$-length vector $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,r})$ is generated at the source $s_i$, where $X_{i,j} \in \mathcal{F}$ for $j = 1,2,\dots,r$. We assume that, $X_i$ is uniformly distributed over $\mathcal{F}^r$ and $H(X_i) = r\log_2|\mathcal{F}|$, where $H(X_i)$ is the entropy function for a random variable $X_i$ . For a subset A⊂ S, the notation $X_A$ is the vector of source random variables and is denoted as follows:

$$X_A = (X_s : s \in A).$$

Moreover, let $X$ be the set of all source processes, $\phi_e$ is the corresponding global encoding function for edge $e$ and $\phi_e(X)$ is a $l$-length vector that is transmitted on edge $e$. So, for all $v \in V/S$, the set $\phi_{In(v)}(X)$ is defined as follows:

$$\phi_{In(v)}(X) = \{\phi_e(X) : e \in In(v)\}.$$

The definition of $\delta$-dependent sources is presented as follows:

**Definition 2.2.** [19] For coding at block length $m$, information sources $X_S$ are said to be $\delta$-dependent if $\sum_{s\in S} H(X_s) - H(X_S) \leq \delta r$. Independent random variables are 0-dependent.

We assume that the source random variables of sum-network $G$ are $\delta$-dependent. Similar to [19], for $\delta$-dependent sum-networks, we can define the probability of decoding error $P_{err}$ as follows:

**Definition 2.3.** For sum-network $G$ with $\delta$-dependent sources information, decoding error $P_{err}$ is defined as follows:

$$P_{err} \triangleq \Pr\left\{\exists t \in T : \psi_t\left(\phi_{In(t)}(X)\right) \neq Z\right\}.$$

By Definition 2.3, all the decoding functions $\psi t$ can recover the sum of sources indicated by $Z = \sum_{p\in P} X_p + \sum_{B\in\mathcal{B}} X_B$ with probability $1 - \epsilon$, where $\epsilon$ is the upper bound of decoding error Perr. When the sources are independent, by [35], $Z$ can be evaluated from $\phi_{In(t)}(X)$, which means $H\left(Z\middle|\phi_{In(t)}(X)\right) = 0$. If the sources be δ-dependent, then $Z$ can be evaluated from $\phi_{In(t)}(X)$ with probability $1 - \boldsymbol{\epsilon}$, which means $H\left(Z\middle|\phi_{In(t)}(X)\right) = 0$, for all $t \in T$ with probability $1 - \boldsymbol{\epsilon}$. Thus, we suppose that there exists a $((r, l), \epsilon, \delta)$ fractional network code solution for sum-network $G$, where $\epsilon$ is the upper bound of decoding error $P_{err}$.

The next proposition and lemma present some properties of the entropy function.

**Proposition 2.1. [39]**

(1) For random variables $X_1, X_2, \dots, X_n$, the chain rule for entropy is described as

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} H(X_i|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n).$$

(2) Conditioning decreases entropy, which means $H(Y|X) \leq H(Y)$, where $X$ and $Y$ are two random variables.

(3) For two random variables $X$ and $Y$, $H(X|Y) = H(X,Y) - H(Y)$.

(4) For two random variables $X$ and $Y$, $I(X,Y) = H(X) + H(Y) - H(X,Y)$, where $I(X,Y)$ is the mutual information between two random variables $X$ and $Y$.

The next lemma is a trivial consequence of the definition of $\delta$-dependence.

**Lemma 2.1.** Let $X_1$ and $X_2$ be two random variables at blocklength $r$. If $X_1$ and $X_2$ are $\delta$-dependent, then $H(X_1) - H(X_1|X_2) \leq \delta r$ and $H(X_2) - H(X_2|X_1) \leq \delta r$.

**Proof:** By Definition 2.2, we get

$$H(X_1) + H(X_2) - H(X_1, X_2) \leq \delta r. \qquad (1)$$

By Part (3) of Proposition 2.1, we have

$$H(X_1, X_2) = H(X_1|X_2) + H(X_2), \qquad (2)$$

also,

$$H(X_1, X_2) = H(X_2|X_1) + H(X_1). \qquad (3)$$

Hence, by (1) and (2), $H(X_1) - H(X_1|X_2) \leq \delta r$. Moreover, by (1) and (3), we get

$$H(X_2) - H(X_2|X_1) \leq \delta r,$$

which concludes the claim.

Tripathy et al. (2015) showed that if all source random variables are independent, then certain partial sums can be computed by observing subsets of the bottleneck edges. For example, in Lemma 1 in [36], we have

$H(Xp_i + \sum_{B\in<p_i>} X_B \,|\phi_{e_i}(X)) = 0,$

where $\phi_{e_i}(X)$ is a $l$-length vector that is transmitted on the bottleneck edge $e_i$ . The next lemma is for the case that the sources are not independent.

**Lemma 2.2.** Let $\mathcal{D}$ and $G$ be a $2-(v,k,1)$ BIBD and the constructed sum-network from it, respectively. If all source random variables are $\delta$-dependent and there exists a $((r,l),\epsilon,\delta)$ fractional network code solution for $G$, then

$H(X_{p_i}+\sum_{B\epsilon p_i} X_B\,|\phi_{e_i}(X)) \leq \delta r$,

where $\phi_{e_i}(X)$ is a $n$-length vector that is transmitted on the bottleneck edge $e_i$.

**Proof:** There exists a $((r,l),\epsilon,\delta)$ fractional network code solution, so, $Z$ can be computed from $\phi_{In\left(t_{p_i}\right)}(X)$, for any $p_i \in P$ and all $i=1,2,\dots,v$ with probability $1-\epsilon$, which means

$H(Z|\phi_{In(t_{p_i})}(X))=0$ with probability $1-\epsilon$, where

$Z=X_{p_i}+\sum_{p\neq p_i} X_p+\ \sum_{B\in<p_i>} X_B+\sum_{B\notin<p_i>} X_B$,

and

$$\phi_{In(t_{p_i})}(X)=\{\phi_{e_i}(X):\ p_i\ \in\ P\}\ \cup\ \{Xp:\ p\neq p_i\}\ \cup\ \{XB:\ B\ \notin<p_i>\}.$$

Let

$Z_1=\sum_{p\neq p_i} X_p,\ \ Z_2=\sum_{B\in<p_i>} X_B\ \ and\ \ Z_3=\sum_{B\notin<p_i>} X_B,$

we have $H(X_{p_i}+Z_1+Z_2+Z_3|\phi_{e_i}(X),\{Xp:\ p\ \neq\ p_i\},\{X_B:\ B\ \notin<p_i>\})=0$ with probability $1-\epsilon$. Since $Z_1$ and $Z_3$ are a subset of $\phi_{In\left(t_{p_i}\right)}(X)$, we get

$$H(X_{p_i}+Z_2\,\big|\,\phi_{e_i}(X),\{Xp:\ p\ \neq\ p_i\},\{XB:\ B\ \notin<p_i>\}\,)=\ 0, \quad (4)$$

with probability $1-\epsilon$. Since all source random variables are $\delta$-dependent, by Lemma 2.1, we have

$H(X_{p_i}+Z_2|\phi_{e_i}(X))-H(X_{p_i}+\ Z_2|\phi_{e_i}(X),\{Xp:\ p\ \neq\ p_i\},\{X_B:\ B\ \notin<p_i>\})\leq\delta r.$

Thus, by (4),

$H(X_{p_i}+Z_2|\phi_{e_i}(X))\leq\delta r.$

The next example describes Lemma 2.2.

**Example 2.2.** Consider the constructed sum-network depicted in Figure 1. Since all source random variables are $\delta$-dependent and there exists a $((r,l),\epsilon,\delta)$ fractional network code solution for $G$, we get $H(Z|\phi_{In(t_{p_1})}(X))=0$ with probability $1-\epsilon$, where

$$Z=X_1+X_A+X_C+X_2+X_B+X_3,$$

and

$$\phi_{In(t_{p_1})}(X)=\{\phi_{e_1}(X):\ p_1\in P\}\cup\{X_2,\ X_3\}\cup\{X_B\}.$$

By the definition of set $<p>$, we have $<p_1>=\ \{A,C\}$. So, $Z_2=X_A+X_C, Z_3=X_B$ and $Z_1=\sum_{p\neq p_1} X_p=X_2+X_3$. Thus, $H(X_1+X_A+X_C|\phi_{e_1}(X),\{X_2,X_3\},\{X_B\})=0$ with probability $1-\epsilon$. All source random variables are $\delta$-dependent, so

$H(X_1+X_A+X_C|\phi_{e_1}(X))-H(X_1+X_A+X_C|\phi_{e_1}(X),\{X_2,X_3\},\{X_B\})\leq\delta r.$

Hence,

$$H(X_1+X_A+X_C|\phi_{e_1}(X))\leq\delta r.$$

Let $H(A)$ be the entropy function for a random variable $A$ and define, for any $v\ >\ 1$,

$$H(A_1,A_2,\dots,A_v)=H(\{A_1,A_2,\dots,A_v\})=H(\{A_i\}_1^v).$$

**Lemma 2.3.** For each $i \in \{1, \dots, v\}$, assume $H(A_i|A_i') \le \delta r$, where $A_i$ and $A_i'$ i are $\delta$ −dependent random variables. Then $H(\{A_i\}_1^v \,|\{A_i'\}_1^v) \le v\delta r$.

**Proof:** We proof the lemma by induction on $v$. For $v = 1$, it is obtained by the assumption. Suppose that it holds for $v = k$. We show that it is also true for $v = k + 1$. Then,

$$H(\{A_i\}_1^{v+1}|\{A_i'\}_1^{v+1}) = H(\{A_i\}_1^v, A_{v+1}|\{A_i'\}_1^{v+1}), = H(\{A_i\}_1^v \,|\{A_i'\}_1^{v+1}) + H(A_{v+1}|\{A_i\}_1^v, \{A_i'\}_1^{v+1}),$$
$$\le H(\{A_i\}_1^v \,|\{A_i'\}_1^v) + H(A_{v+1}|A_{v+1}'),$$
$$\le v\delta r + \delta r,$$
$$= (v + 1)\delta r.$$

Note that by Parts (1) and (2) of Proposition 2.1, the second equality and first inequality are concluded.

**Corollary 2.1.** Let D be a $2 - (v, k, 1)$ BIBD and $G$ be the constructed sum-network from it. If all source random variables are $\delta$-dependent and there exists a $((r, l), \epsilon, \delta)$ fractional network code assignment for $G$, then

$$H(\{X_{p_i} + \textstyle\sum_{B\in<p_i>} X_B\}_1^v \,|\{\phi_{e_i}(X)\}_1^v) \le v\delta r.$$

**Proof:** By Lemmas 2.2 and 2.3, the claim is concluded.

**Lemma 2.4**. Let $D$ be a $2 - (v, k, 1)$ BIBD and $G$ be the constructed sum-network from it. If all random variables in $\{\{X_{p_i}\}_1^v \cup \{X_{B_j}\}_1^b\}$ are $\delta$-dependent, then there exists a $\delta'$ such that the random variables $X_i' = X_{p_i} + \sum_{B\in<p_i>} X_B$ are $\delta'$ -dependent, for $i \in \{1, \dots, v\}$.

**Proof:** Since all random variables in $\{\{X_{p_i}\}_1^v \cup \{X_{B_j}\}_1^b\}$ are $\delta$ −dependent and

$X_i' = X_{p_i} + \sum_{B\in<p_i>} X_B$,

for $i \in \{1, \dots, v\}$, we conclude that the random variables $X_i'$ are dependent for $i \in \{1, \dots, v\}$. Thus,

$$H(\{X_i'\}_1^v) < \sum_{i=1}^{v} H(X_i').$$

Then, under $((r, l), \epsilon, \delta)$ fractional network code, there exists a $\delta'$ such that

$$\sum_{i=1}^{v} H(X_i') - H(\{X_i'\}_1^v) < \delta' r,$$

which concludes the claim.

In the following, we show that how does $\delta'$ depend on $\delta$.

**Corollary 2.2.**

1. By [36], when all source random variables are independent and uniformly distributed over $\mathcal{F}^r$, then $X_i'$ are also independent and uniformly distributed over $\mathcal{F}^r$. Thus, if $\delta = 0$, then $\delta'$= 0.

2. Since $X_i'$ are uniformly distributed over $\mathcal{F}^r$, by Lemma 2.4, we get $\delta'$ can not be large. Then, we have not a loose upper bound.

**Example 2.3.** Consider three binary random variables

$$X_1 = (b_0, b_1), X_2 = (b_0, b_2) \text{ and } X_3 = (b_1, b_2),$$

such that they are uniformly distributed over $\mathcal{F}^2$ , where $|\mathcal{F}| = 2$.

We see that $X_1$ and $X_2$ are $1/2$ −dependent. Similarly, we can show that $X_2$ and $X_3$ are also $1/2$ −dependent. Let $X'_1 = X_1 + X_2$ and $X'_2 = X_1 + X_3$, so $X'_1 = (b_0 + b_0,\ b_1 + b_2)$ and $X'_2 = (b_0 + b_1,\ b_1 + b_2)$. Thus, $X'_1$ and $X'_2$ are dependent because $H(X'_1) = H(X'_2) = 2$ and $H(X'_1, X'_2) = 3$. Hence $H(X'_1) + H(X'_2) - H(X'_1, X'_2) = 4 - 3 = 1 < \frac{2}{3}(2)$ which concludes $\delta' = 2/3$.

**Lemma 2.5**. Let $\mathcal{D} = (P, B)$ be a set system of a BIBD and

$$X'_i = X_{p_i} + \sum_{B\in<p_i>} X_B,$$

where $i \in \{1, \ldots, v\}$. If all random variables in $\{\{X_{p_i}\}_1^v \cup \{X_{B_j}\}_1^b\}$ are $\delta$-dependent, then there exists a $\delta'$ such that $H(\{X'_i\}_1^v) > r(v \log_2 q - \delta')$.

**Proof :** Since all random variables in $\{\{X_{p_i}\}_1^v \cup \{X_{B_j}\}_1^b\}$ are $\delta$-dependent, by Lemma 2.4, there exists a $\delta'$ such that $\sum_{i=1}^v H(X'_i) - H(\{X'_i\}_1^v) < \delta' r$. Moreover, under a $(r, l)$ fractional network code, we get $H(X'_i) = r \log_2 q$, for all $i \in \{1, \ldots, v\}$, which means

$$\sum_{i=1}^v H(X'_i) = vr \ \log_2 q.$$

Hence,

$$H(\{X'_i\}_1^v) > vr \log_2 q - \delta' r = r(v \log_2 q - \delta').$$

## 3. The coding capacity of sum-networks with dependent sources

In this section, we obtain an upper bound on the coding capacity of sum-networks when the sources are $\delta$-dependent. Let $\mathcal{D}$ be a $2 - (v, k, 1)$ design and $G$ be the constructed sum-network using the given construction $\mathcal{D}$. Under the assumption that sources are independent, the capacity of the sum-network $G$ is at most 1 [36]. In the following theorem, we obtain an upper bound for the network coding capacity of the sum-networks when the sources are $\delta$-dependent.

**Theorem 3.1.** Let $\boldsymbol{\mathcal{D}}$ be a $2 - (v, k, 1)$ design and $G$ be the constructed sum-network using the given construction $\boldsymbol{\mathcal{D}}$. Supposing that all random variables in $\{\{X_{p_i}\}_1^v \cup \{X_{B_j}\}_1^b\}$ are $\delta$ −dependent and $X'_i = X_{p_i} + \sum_{B\in<p_i>} X_B$ . Then, the upper bound of the network coding capacity of $G$ is at most $A = \frac{v\ \log_2 q}{v \log_2 q - (v\delta + \delta')}$ , where $\delta'$ is used to quantify the dependency among the random variables $X'_i$ .

**Proof:** Since $\phi_{e_i}(X)$ is a $l$-length vector that is transmitted on the bottleneck edge $e_i$ , under a $((r, l), \epsilon, \delta)$ fractional network code, we have $H(\phi_{e_i}(X)) \leq l \log_2 q$. Thus,

$$H\left(\left\{\phi_{e_i}(X)\right\}_1^v\right) \leq \sum_1^v H\left(\phi_{e_i}(X)\right) \leq vl \log_2 q. \tag{5}$$

By (3) and (4) of Proposition 2.1, we get

$$I\left(\left\{\phi_{e_i}(X)\right\}_1^v, \{X'_i\}_1^v\right) = H\left(\left\{\phi_{e_i}(X)\right\}_1^v\right) - H\left(\left\{\phi_{e_i}(X)\right\}_1^v \left|\{X'_i\}_1^v\right.\right). \tag{6}$$

Also, we have

$$I(\{\phi_{e_i}(X)\}_1^v, \{X'_i\}_1^v) = H(\{X'_i\}_1^v) - H(\{X'_i\}_1^v | \{\phi_{e_i}(X)\}_1^v). \tag{7}$$

Hence, by (6) and (7),

$$\begin{aligned} H(\{\phi_{e_i}(X)\}_1^v) &= I(\{\phi_{e_i}(X)\}_1^v, \{X'_i\}_1^v) + H(\{\phi_{e_i}(X)\}_1^v | \{X'_i\}_1^v), \\ &= H(\{X'_i\}_1^v) - H(\{X'_i\}_1^v | \{\phi_{e_i}(X)\}_1^v) + H(\{\phi_{e_i}(X)\}_1^v | \{X'_i\}_1^v), \end{aligned}$$

So, by (5),

$H(\{X_i'\}_1^v) - H(\{X_i'\}_1^v \,|\{\phi_{e_i}(X)\}_1^v) + H(\{\phi_{e_i}(X)\}_1^v \,|\{X_i'\}_1^v) \leq vl \log_2 q.$

Also, by Lemma 2.5 and Corollary 2.1,

$$r(v \log_2 q - \delta') - vr\delta + H(\{\phi_{e_i}(X)\}_1^v \,|\{X_i'\}_1^v) \leq vl \log_2 q.$$

By $H(\{\phi_{e_i}(X)\}_1^v \,|\{X_i'\}_1^v) \geq 0$, we get

$$r(v \log_2 q - \delta') - vm\delta \leq vl \log_2 q.$$

Thus,

$\frac{r}{l} \leq \frac{v \log_2 q}{v \log_2 q - (v\delta + \delta')} = A.$

**Corollary 3.1.** Let $\boldsymbol{\mathcal{D}}$ be a $2-(v,k,1)$ design and $G$ be the constructed sum-network using the given construction $\boldsymbol{\mathcal{D}}$. Supposing that all random variables in $\{\{X_{p_i}\}_1^v \cup \{X_{B_j}\}_1^b\}$ are $\delta$ −dependent and $X_i' = X_{p_i} + \sum_{B\in<p_i>} X_B$. Then, the upper bound of the network coding capacity of G is increased by the factor of $\frac{(v\delta+\delta')}{v \log_2 q-(v\delta+\delta')}$, where $\delta'$ is used to quantify the dependency among the random variables $X_i'$, for $i = 1,\ldots,v$.

**Proof:** If the sources are independent, then the coding capacity of the constructed sum-network is at most 1 [36]. On the other hand, according to Theorem 3.1, if the sources are $\delta$ −dependent, then the coding capacity of the constructed sum-network is upper bounded by $\frac{v \log_2 q}{v \log_2 q-(v\delta+\delta')}$.

Thus, by subtracting the value 1 of $\frac{v \log_2 q}{v \log_2 q-(v\delta+\delta')}$, the claim is concluded.

Therefore, by Corollary 3.1, if the sources are δ-dependent, then the coding capacity upper bound of sum-networks is increased.

### 3.1. Numerical results

In this section, we present some examples of communication networks with dependent sources. We calculate the coding capacity of communication networks with dependent sources, then we compare the obtained results with previous works (the coding capacity of communication networks with independent sources). We show that a special kind of dependency among the sources can increase the capacity region of communication networks. The next two examples show that how dependent sources can increase the routing capacity of communication networks.

**Example 3.1.** Consider network $N$ shown in Figure 2, in which two sources $s_1$ and $s_2$ produce messages $X$ and $Y$, respectively. Also, two sinks $t_1$ and $t_2$ demand the both of two messages. By [4], the routing capacity of this network is $1/2$ (for more detail see Example III.2 in [4]). Supposing that the sources are dependent such that $X = Y$. Then, $t_1$ and $t_2$ can receive $X$ and $Y$ through two edges $(s_1, t_2)$ and $(s_2, t_1)$, respectively. Since sinks $t_1$ and $t_2$ demand both of two messages $X$ and $Y$, it is sufficient that only one of those two messages is passed through the edge $(v_1, v_2)$. Thus, the routing capacity of $N$ is $1$, which means it is increased by the factor of $1/2$.

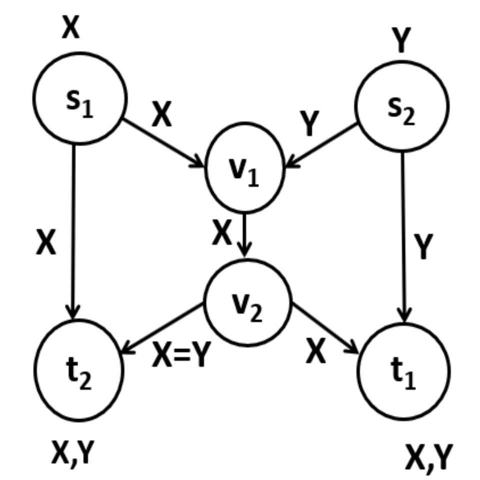


**Figure 2.** The network $N$, where only one of the two messages $X$ and $Y$ passes through the edge $(v_1, v_2)$.

- The previous example demonstrates a trivial statement. In the following, a much simpler example is presented.

**Example 3.2.** Consider a "$Y$" network with two sources $s_1$ and $s_2$ on top and one sink t at the bottom such that these two sources are connected to the sink through a shared link. Let messages $X$ and $Y$ be produced with $s_1$ and $s_2$, respectively. Moreover, sink $t$ demands the both of two messages. Then, the routing capacity of this network is $1/2$. Now, supposing that the sources are linear dependent such that $X = Y$ . Then, the routing capacity of this network is 1.

In the next example, we consider a sum-network with dependent sources. We show that the capacity region of a sum-network change when the sources be linearly dependent. This result is shown by the following example.

**Example 3.3.** Consider a network with two sources $s_1$ and $s_2$ and two sinks $t_1$ and $t_2$. Let two messages $X$ and $Y$ be produced by $s_1$ and $s_2$, respectively. Supposing that this network only has 10 two edges $(s_1, t_1)$ and $(s_2, t_2)$. When the sources are independent, then the sum-capacity is zero. If $X = Y$, then the sum-capacity is 1. Moreover, if $X = -Y$, then the sum-capacity is infinite.

By [28], the linear coding capacity of sum-network $S_3$, depicted in Figure 3, is at least $\frac{2}{3}$ . We show that a special kind of dependency among the sources can increase the capacity lower bound of this sum-network.

**Example 3.4.** Consider network $S_3$ shown in Figure 3. Supposing that $X_i = (X_{i,1}, X_{i,2})$ is generated at source si, for $i = 1, 2, 3$. Consider the following dependency among the sources:

$$X_{1,1} = X_{2,1}, \qquad X_{1,2} = X_{3,1}, \qquad X_{2,2} = X_{3,2}.$$

We define two sums $Sum_1$ and $Sum_2$ as follows:

$$Sum_1 = Sum_2 = X_{1,1} + X_{1,2} + X_{2,2}.$$

• If $\boldsymbol{Sum_1} + \boldsymbol{Sum_2}$ is transmitted to terminals $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ in the first time slot and $\boldsymbol{Sum_1} + \boldsymbol{Sum_2}$ is transmitted to terminal $\boldsymbol{t_3}$ in the second time slot, then $\boldsymbol{Sum_1}$ and $\boldsymbol{Sum_2}$ can be transmitted to all the terminals in two time slots. Thus, the linear coding capacity of the sum-network $\boldsymbol{S_3}$ is at least 1.

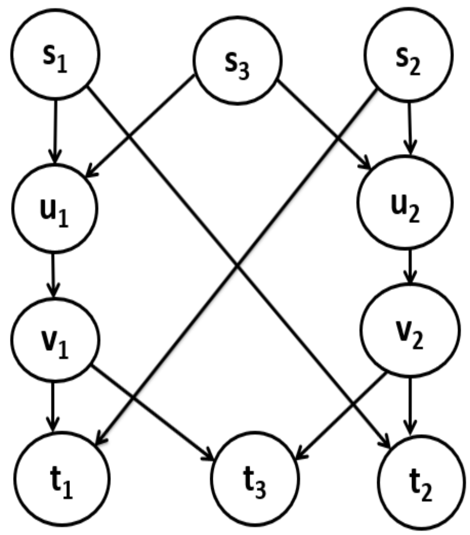


**Figure 3.** The network $S_3$.

By Example 3.4, the lower bound of the network coding capacity of the sum-network $S_3$ is increased from $\frac{2}{3}$ to 1 because the sources are dependent. Thus, by Corollary 3.1, we conclude that the dependency between the sources can change the capacity region of sum-networks. In the next example, we show that dependency among the sources can convert a non-solvable sum-network to a solvable one. According to [31], if the sources are independent, then the coding capacity of the network $S_3'$ (depicted in Figure 3) is $\frac{2}{3}$. Moreover, by [31], the network $S_3'$ $is$ non-solvable, which means all the terminals can not receive the sum of sources at rate 1.

**Example 3.5.** Consider the network $S_3'$ depicted in Figure 4. By [31], the network $S_3'$ has a $(2, 3)$ fractional network coding solution which means it is not solvable. Supposing that the sources are dependent such that $X_1 = X_3$, where $X_1$ and $X_3$ are the generated messages at two sources $S_1$ and $S_3$, respectively. Then, there exists a network coding scheme for $S_3'$ such that all terminals can receive the sum of the sources at rate 1. Figure 4 shows this network coding solution. The depicted coding scheme in Figure 4 shows that two terminals $t_1$ and $t_2$ can receive the sum of the source messages through their incoming edges. Also, the terminal $t_3$ can receive $X_3$ and $X_2 + X_3$ through incoming edges. Since $X_1 = X_3$, the terminal $t_3$ can receive the sum of the source messages. Thus, the network $S_3'$ is solvable.

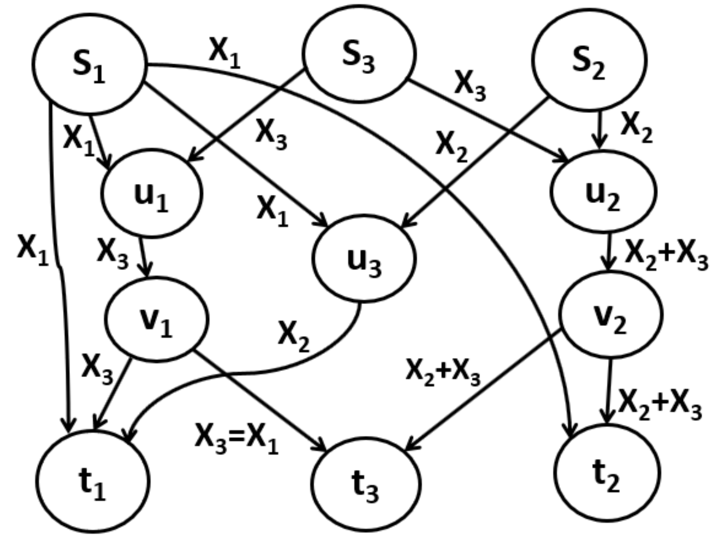


**Figure 4.** The network $S_3'$ and its corresponding coding scheme. Terminal $t_3$ can receive message $X_1$ through the incoming edge $(v_1, t_3)$.

## 4. Conclusion

This work considers a sum-network with $\delta$ −dependent sources. It evaluates the upper bound of the coding capacity of this network for the case where $\delta \neq 0$. We conclude that the dependency between the sources can alter the capacity region of sum-networks. By Theorem 3.1, if the value of $\delta$ is increased, the capacity upper bound of the sum-network also increases. In more detail, the relationship between $\delta$ and $A$ (the upper bound obtained in Theorem 3.1) is as follows:

• If $\delta = 0$, then $A = 1$. In other words, when the sources are independent, the upper bound of the coding capacity is 1, which coincides with the upper bound presented in [36] (see Theorem 1 in [36]).

• If $\delta > 0$ and $|\mathcal{F}| = q \geq 2$, then $v \log_2 q > v \log_2 q - (v\delta + \delta')$. Therefore, by Theorem 3.1, we have $A > 1$. Hence, when the sources of the considered sum-network are $\delta$ −dependent, the upper bound of its coding capacity is greater than 1.

This work has investigated the coding capacity of a sum-network employing $\delta$ −dependent sources. Our primary contribution is the characterization of an upper bound on the capacity for the general case where $\delta \neq 0$, demonstrating that statistical dependency between sources can significantly alter the capacity region of such networks.

The key insight, formalized in Theorem 3.1, is that the upper bound $A$ is a non-decreasing function of the dependency parameter $\delta$. Specifically, our analysis reveals the following precise relationship:

- Independent Sources ($\delta = 0$): The upper bound simplifies to $A = 1$. This result perfectly coincides with and reinforces the established bound for independent sources given in [35], serving as a sainty check for our generalized model.

- Dependent Sources ($\delta > 0$): For any finite field size $q \geq 2$, the derived upper bound yields $A > 1$. This establishes that any positive source dependency strictly increases the upper bound on the coding capacity compared to the independent case.

These findings imply that the correlation between sources introduces a new dimension to the network coding problem, potentially enabling higher achievable rates. This challenges the conventional design principle based on the assumption of independent sources and suggests the leveraging source dependency could be a powerful tool for enhancing network performance.

For future research, several directions emerge naturally. First, the tightness of this upper bound should be investigated by constructing achievable coding schemes that match it for specific value of $\delta$. Second, it would be valuable to explore whether similar dependency-exploiting gains in other types of network problem beyond sum-networks. Finally, analyzing more complex, non-linaer dependency structures between sources presents a challenging but fruitful for the further study.

## References


[1] Ahlswede, R., Cai, N., Li, S. Y., & Yeung, R. W. (2000). Network information flow. *IEEE Transactions on information theory*, *46*(4), 1204-1216.

[2] Aman, M., & Tayyebi, J. (2014). Capacity inverse minimum cost flow problem under the weighted hamming distances. *Iranian Journal of Operations Research*, *5*(2), 12-25.

[3] Bakshi, M., Effros, M., Gu, W., & Koetter, R. (2007, June). On network coding of independent and dependent sources in line networks. In *2007 IEEE International Symposium on Information Theory* (pp. 1096-1100). IEEE.

[4] Cannons, J., Dougherty, R., Freiling, C., & Zeger, K. (2006). Network routing capacity. *IEEE Transactions on Information Theory*, *52*(3), 777-788.

[5] Chan, T., & Grant, A. (2010, June). On capacity regions of non-multicast networks. In *2010 IEEE International Symposium on Information Theory* (pp. 2378-2382). IEEE.

[6] Connelly, J., & Zeger, K. (2018). Capacity and achievable rate regions for linear network coding over ring alphabets. *IEEE Transactions on Information Theory*, *65*(1), 220-234.

[7] Das, N., & Rai, B. K. (2017, March). Sum-networks: Dependency on characteristic of the finite field under linear network coding. In *2017 Twenty-third National Conference on Communications (NCC)* (pp. 1-6). IEEE.

[8] Das, N., & Rai, B. K. (2015, February). On the number of sources and terminals of sum-networks with capacity p/q. In *2015 Twenty First National Conference on Communications (NCC)* (pp. 1-6). IEEE.

[9] Dougherty, R., Freiling, C., & Zeger, K. (2006). Unachievability of network coding capacity. *IEEE Transactions on Information Theory*, *52*(6), 2365-2372.

[10] Fereydounian, M., Jamali, M. V., Hassani, H., & Mahdavifar, H. (2019, August). Channel coding at low capacity. In *2019 IEEE Information Theory Workshop (ITW)* (pp. 1-5). IEEE.

[11] Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., ... & Stern-Ginossar, N. (2021). The coding capacity of SARS-CoV-2. *Nature*, *589*(7840), 125-130.

[12] Fragouli, C., Le Boudec, J. Y., & Widmer, J. (2006). Network coding: an instant primer. *ACM SIGCOMM Computer Communication Review*, *36*(1), 63-68.

[13] Ghazvineh, S., & Ghiyasvand, M. (2024). An $O(|V||E|^2)$ Time Algorithm to Diagnose the Solvability of Single Rate n-Pair Networks with Common Bottleneck Links. *Iranian Journal of Operations Research*, *15*(2), 122-132.

[14] Giridhar, A., & Kumar, P. R. (2005). Computing and communicating functions over sensor networks. *IEEE Journal on selected areas in communications*, *23*(4), 755-764.

[15] Harvey, N. J., Kleinberg, R., & Lehman, A. R. (2006). On the capacity of information networks. *IEEE Transactions on Information Theory*, *52*(6), 2345-2364.

[16] Huang, S., & Ramamoorthy, A. (2013). On the multiple-unicast capacity of 3-source, 3-terminal directed acyclic networks. IEEE/ACM Transactions on Networking, 22(1), 285-299.

[17] Jha, A. V., & Gupta, D. K. (2018, October). The 4s/3t Sum-Networks and Solvability. In *2018 International Conference on Applied Electromagnetics, Signal Processing and Communication (AESPC)* (Vol. 1, pp. 1-5). IEEE.

[18] Kamal, A. E., & Mohandespour, M. (2014). Network coding-based protection. Optical Switching and Networking, 11, 189-201.

[19] Kim, W., Langberg, M., & Effros, M. (2016, July). A characterization of the capacity region for network coding with dependent sources. In *2016 IEEE International Symposium on Information Theory (ISIT)* (pp. 1764-1768). IEEE.

[20] Kramer, G., & Savari, S. A. (2006). Edge-cut bounds on network coding rates. *Journal of Network and Systems Management*, *14*(1), 49-67.

[21] Krithivasan, D., & Pradhan, S. S. (2008, July). An achievable rate region for distributed source coding with reconstruction of an arbitrary function of the sources. In *2008 IEEE International Symposium on Information Theory* (pp. 56-60). IEEE.

[22] Langberg, M., & Effros, M. (2012, September). Source coding for dependent sources. In *2012 IEEE Information Theory Workshop* (pp. 70-74). IEEE.

[23] Langberg, M., & Effros, M. (2019). The Edge-Removal Problem's Connections to the Zero-Error and $\delta $-Dependence Problems in Network Coding. *IEEE Transactions on Information Theory*, *66*(2), 900-913.

[24] Langberg, M., & Ramamoorthy, A. (2009, June). Communicating the sum of sources in a 3-sources/3-terminals network. In *2009 IEEE International Symposium on Information Theory* (pp. 2121-2125). IEEE.

[25] Li, S. Y., Yeung, R. W., & Cai, N. (2003). Linear network coding. *IEEE transactions on information theory*, *49*(2), 371-381.

[26] Nawaz, F., & Jeoti, V. (2016). GREEN protocol: gradient based energy efficient routing with network coding capacity. *Telecommunication Systems*, *62*(1), 135-147.

[27] Rai, B. K., & Das, N. (2013, September). On the capacity of ms/3t and 3s/nt sum-networks. In *2013 IEEE Information Theory Workshop (ITW)* (pp. 1-5). IEEE.

[28] Rai, B. K., & Das, N. (2013, October). On the capacity of sum-networks. In *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (pp. 1545-1552). IEEE.

[29] Rai, B. K., & Das, N. (2017). On characteristics dependent linear coding capacity of sum-networks. *IET Networks*, *6*(5), 133-137.

[30] Rai, B. K., & Dey, B. K. (2012). On network coding for sum-networks. *IEEE Transactions on Information Theory*, *58*(1), 50-63.

[31] Rai, B. K., Dey, B. K., & Shenvi, S. (2010, January). Some bounds on the capacity of communicating the sum of sources. In *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)* (pp. 1-5). IEEE.

[32] Ramamoorthy, A., & Langberg, M. (2013). Communicating the sum of sources over a network. *IEEE Journal on Selected Areas in Communications*, *31*(4), 655-665.

[33] Sánchez-Cauce, R., Paris, I., & Díez, F. J. (2021). Sum-product networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3821-3839.

[34] Shenvi, S., & Dey, B. K. (2010, June). A necessary and sufficient condition for solvability of a 3s/3t sum-network. In *2010 IEEE International Symposium on Information Theory* (pp. 1858-1862). IEEE.

| | |
|---|---|
| [35] | Stinson, D. R. (2008). Combinatorial designs: constructions and analysis. *ACM SIGACT News*, *39*(4), 17-21. |
| [36] | Tripathy, A., & Ramamoorthy, A. (2015, June). Capacity of sum-networks for different message alphabets. In *2015 IEEE International Symposium on Information Theory (ISIT)* (pp. 606-610). IEEE. |
| [37] | Tripathy, A., & Ramamoorthy, A. (2017). Sum-networks from incidence structures: construction and capacity analysis. *IEEE Transactions on Information Theory*, *64*(5), 3461-3480. |
| [38] | Yan, X., Yeung, R. W., & Zhang, Z. (2007, June). The capacity region for multi-source multi-sink network coding. In *2007 IEEE International Symposium on Information Theory* (pp. 116-120). IEEE. |
| [39] | Yeung, R. W. (2008). *Information theory and network coding*. Springer Science & Business Media. |